LABOR SUPPLY AND TAXATION

RICHARD BLUNDELL

Edited by Andreas Peichl and Klaus F. Zimmermann

IZA PRIZE IN LABOR ECONOMICS

IZA

OXFORD

Labor Supply and Taxation

IZA Prize in Labor Economics Series

Since 2002, the Institute for the Study of Labor (IZA) has awarded the annual IZA Prize in Labor Economics for outstanding contributions to policy-relevant labor market research and methodological progress in this sub-discipline of economic science. The IZA Prize is the only international science prize awarded exclusively to labor economists. This special focus acknowledges the global significance of high-quality basic research in labor economics and sound policy advice based on these research findings. As issues of employment and unemployment are among the most urgent challenges of our time, labor economists have an important task and responsibility. The IZA Prize in Labor Economics is today considered one of the most prestigious international awards in the field. It aims to stimulate further research on topics that have enormous implications for our future. All prize-winners contribute a volume to the IZA Prize in Labor Economics Series published by Oxford University Press, which has been established to provide an overview of the laureates' most significant findings.

The IZA Prize in Labor Economics has become an integral part of the institute's manifold activities to promote progress in labor market research. Based on nominations submitted by the IZA Research Fellows, a high-ranking IZA Prize Committee selects the prize-winner. In conjunction with the Award Ceremony the IZA Prize Conference brings together a number of renowned experts to discuss topical labor market issues.

lt is not by coincidence that the IZA Prize in Labor Economics Series is published by Oxford University Press. This well-reputed publishing house has shown a great interest in the project from the very beginning as this exclusive series perfectly complements their range of publications. We gratefully acknowledge their excellent cooperation.

Klaus F. Zimmermann, IZA Director

Winners of the IZA Prize in Labor Economics

2012	Richard Blundell (University College London and IFS)
2011	George J. Borjas (Harvard University) Barry R. Chiswick (George Washington University)
2010	Francine D. Blau (Cornell University)
2009	Richard A. Easterlin (University of Southern California)
2008	Richard Layard (London School of Economics) Stephen J. Nickell (Nuffield College)
2007	Richard B. Freeman (Harvard University)
2006	David Card (University of California, Berkeley) Alan B. Krueger (Princeton University)
2005	Dale T. Mortensen (Northwestern University) Christopher A. Pissarides (London School of Economics)
2004	Edward P. Lazear (Stanford University)
2003	Orley C. Ashenfelter (Princeton University)
2002	Jacob Mincer (Columbia University)

Richard Blundell
2012 IZA Prize Laureate

Labor Supply and Taxation

Richard Blundell

Edited by
Andreas Peichl
Klaus F. Zimmermann

OXFORD
UNIVERSITY PRESS

OXFORD
UNIVERSITY PRESS

Great Clarendon Street, Oxford, OX2 6DP,
United Kingdom

Oxford University Press is a department of the University of Oxford.
It furthers the University's objective of excellence in research, scholarship,
and education by publishing worldwide. Oxford is a registered trade mark of
Oxford University Press in the UK and in certain other countries

First Edition published in 2016

Impression: 1

Published in the United States of America by Oxford University Press
198 Madison Avenue, New York, NY 10016, United States of America

British Library Cataloguing in Publication Data
Data available

Library of Congress Control Number: 2015940149

ISBN 978–0–19–874980–6
Printed in Great Britain by Clays Ltd, St Ives plc

Award Statement of the IZA Prize Committee

The 2012 IZA Prize in Labor Economics is awarded to Richard Blundell (University College London and IFS) for his path-breaking contributions to the econometric analysis of labor markets and public policy reforms. Professor Blundell has played a major role in the development of modern labor and policy analysis. His work is characterized by a focus on pressing questions in economic policy, the development of cutting-edge econometric methods to tackle such questions, and painstaking analysis of micro data. Blundell's contributions laid the foundations for the modern economic analysis of labor supply, consumer behavior, and policy reforms. His work has been enormously influential in the worldwide debates about tax and welfare reforms over the past decades.

Richard Blundell has provided some of the most significant contributions to the econometric analysis of labor supply. Among his early work was a seminal paper with Richard Smith in which they provided a novel test for endogeneity in econometric models (Econometrica, 1986). The test signified a crucial advancement in addressing important empirical questions, for instance, regarding the relationship between household income and female labor supply. Richard Blundell's work has greatly enhanced our understanding of how taxation and social policy influence individuals' labor market behavior. Blundell did not only help develop a new theoretical framework for understanding labor supply reactions to policy changes. He also complemented the theoretical analysis by developing the necessary instruments for careful empirical analysis, and he applied these tools to real-world policy questions. His paper with Alan Duncan and Costas Meghir (Econometrica, 1998) is exemplary of this approach. The authors developed novel methods to examine how household labor supply responds to tax reforms, and used these methods to estimate the reactions to the UK policy reforms in the 1980s.

Richard Blundell also made fundamental contributions to studying labor supply and consumer demand in an integrated, unifying framework. His approach offered a dynamic perspective, taking into account that consumption smoothing over the life-cycle and labor supply are inherently linked. More recently, Blundell has also been concerned with economic inequality and the changes in income distributions. He made important methodological contributions that have advanced the econometric modeling of such dynamic issues. Furthermore, Blundell was among the first to explicitly take individual-level heterogeneity into account when assessing actual policy problems.

In addition to his academic contributions, Richard Blundell has made a compelling case for more evidence-based research and policy advice. He has taken a leading role in demonstrating the importance of giving researchers enhanced access to micro data for their research. Throughout his career, Blundell has intensely engaged in the policy debate. From 2006 until 2011, he served as co-editor of the influential Mirrlees Review. In this role, he provided practical recommendations for improving UK tax policy and helped develop the pillars of an efficient modern tax system. His enduring commitment to research and policy make Blundell a role model for combining academic strength with policy relevance.

Richard Blundell currently holds the David Ricardo Chair of Political Economy at University College London and is Research Director of the Institute for Fiscal Studies, where he is also Director of the ESRC Centre for the Microeconomic Analysis of Public Policy. He is a Commander of the British Empire, Fellow of the Econometric Society and of the British Academy, and Honorary Member of the American Economic Association and the American Academy of Arts and Science. In 1995, he was awarded the Yrjö Jahnsson Prize for his work in microeconometrics, labor supply and consumer behavior; in 2000, he received the Frisch Medal by the Econometric Society for his 1998 Econometrica paper with Alan Duncan and Costas Meghir. Blundell has delivered numerous distinguished lectures, and he holds honorary doctoral degrees from the Universities of St. Gallen, Bergen, and Mannheim. He is the current president of the Royal Economic Society, and past president of the Econometric Society, the European Economic Association, and the Society of Labor Economists. He also served as co-editor for journals such as Econometrica, the Journal of Econometrics, and the Review of Economic Studies.

George A. Akerlof	University of California, Berkeley; IZA
Marco Caliendo	University of Potsdam; IZA
Richard Portes	London Business School; President CEPR
Jan Svejnar	University of Michigan, Ann Arbor; IZA
Klaus F. Zimmermann	IZA; University of Bonn

Acknowledgements

This volume draws on more than three decades of research, much of it co-authored with the brilliant students and colleagues of whom I have been lucky enough to work with over the years. In particular, I would like to thank Antoine Bozio, Mike Brewer, Monica Costa-Dias, Alan Duncan, Marco Francesconi, John Ham, Guy Laroque, Julian McCrae, Costas Meghir, Andrew Shephard, John Van Reenen and Ian Walker.

The early work in this volume was written while I was a lecturer in the Econometrics Department at the University of Manchester. Since 1984 I have been Professor of Economics at University College London and since 1986 also Research Director at the Institute for Fiscal Studies. I am very grateful for the creative environment afforded to me by these institutions.

The bulk of the research reported here is part of the program of research of the ESRC Centre for the Microeconomic Analysis of Public Policy at IFS. Funding from the ESRC, grant number R000239865 is gratefully acknowleged.

The data used in these studies was made available by the ONS through the ESRC Data Archive and has been used by permission of the controller of HMSO. Neither the ONS nor the ESRC Data Archive bear responsibility for the analysis or the interpretation of the data reported here.

I would like to thank Andreas Peichl and Klaus Zimmermann, who edited this volume for their advice and comments.

Finally, I would like to thank IZA for awarding me the IZA Prize in Labor Economics and for the support it has offered me, and many other researchers in Labor Economics, over the years.

Richard Blundell

Contents

Part I

Introduction by the Editors: Taxation and Labor Supply – Revisiting the Contributions by Richard Blundell

Andreas Peichl and Klaus F. Zimmermann

The question of how individuals adapt their behavior in response to policy changes is one of the most investigated topics in empirical labor and public economics. Do people reduce their working hours if governments decide to raise taxes? Might they even withdraw completely from the labor market? Even if these questions are not new, they are still topical – maybe more topical as ever before. Labor supply estimations are extensively used for various policy analyses and economic research. Labor supply elasticities are key information when evaluating tax-benefit policy reforms and their effect on tax revenue, employment and redistribution. For instance, the question whether welfare programs should be directed to the workless poor, through traditional demogrant policies, or the working poor, via in-work support (Saez 2001) depends on whether participation decisions (the extensive margin) systematically prevail over responses in terms of work hours (the intensive margin). Large participation responses may subsequently lead to large elasticities in the lower part of the income distribution, which is crucial for welfare analysis (see Eissa, Kleven and Kreiner 2008). Finally, the optimal taxation of couples, and notably the issue of joint versus individual taxation, critically relies on the knowledge of cross-wage elasticities of spouses (Immervoll et al. 2011).

Understanding labor supply behavior, both in a static and dynamic context, and using this evidence to guide policy has been on the research agenda of Richard Blundell for many years. He has made several path-breaking contributions in the field of labor supply and taxation for which he was awarded the 2012 IZA Prize in Labor Economics. Every year, the Institute for the Study of Labor awards the Prize to an individual who has made outstanding contributions to policy-relevant research and methodological progress in the field of labor economics. Since its establishment in 2002, the Prize has been given to several distinguished labor economists; two of them later even received the Nobel Prize in Economics. The 2012 Prize honors Richard Blundell's path-breaking contributions to the modern econometric analysis of labor markets and public policy reforms. Professor Blundell is an eminent scholar and has worked on a range of issues in labor economics, econometric theory and tax policy. His research is unified by a focus on household behavior and the way various labor market and tax policy institutions affect this behavior. His empirical work is driven by policy questions and motivated by economic theory. Many of his contributions show how frontier level econometric methods can be put to work in studying applied problems.

Richard Blundell holds the David Ricardo Chair of Political Economy at University College London. He is a graduate of the University of Bristol and London School of Economics. At the age of 23, he became a faculty member at the University of Manchester without obtaining a PhD first. Since 1986 he has been Research Director of the Institute for Fiscal Studies (IFS), where he is also Director of the ESRC Centre for the Microeconomic Analysis of Public Policy. He is a Commander of the British Empire, Fellow of the Econometric Society and of the British Academy, and Honorary Member of the American Economic Association and the American Academy of Arts and Science. In 1995, he was awarded the Yrjö Jahnsson Prize for his work in microeconometrics, labor supply and consumer behavior; in 2000, he received the Frisch Medal by the Econometric Society. Blundell has delivered numerous distinguished lectures, and he holds honorary doctoral degrees from the Universities of St. Gallen, Bergen, and Mannheim. He is the current president of the Royal Economic Society, and past president of the Econometric Society, the European Economic Association, and the Society of Labor Economists. He also served as co-editor for journals such as Econometrica, the Journal of Econometrics, or the Review of Economic Studies. Professor Blundell has edited nine books and published more than 180 papers – many in the most prestigious economics journals. Many of his papers are very influential as evidenced from their large number of citations. In total, his work was cited more than 33,500 times according to Google Scholar (as of August 2013); his Repec score is 8500 ranking him in the top 20 of economists world-wide. His h-index is 72 – implying that 72 of his publications have been cited at least 72 times. 175 publications have been cited at least 10 times, 120 of those have received at least 10 citations in the last 5 years. For instance, his 1998 Journal of Econometrics Paper with Steve Bond on the GMM Dynamic Panel Estimator has received more than 7,000 citations and his 1999 Handbook of Labor Economics survey on labor supply with Thomas MaCurdy has received more than 1,600 citations.

In various publications, Richard Blundell has provided some of the most significant contributions in the economic and econometric analysis of labor supply and taxation. He was involved in developing a new theoretical framework for our understanding of labor supply reactions to policy changes and complemented the theoretical analysis by developing the necessary instruments for careful empirical analysis. He made fundamental contributions to studying labor supply and consumption smoothing over the life cycle in an integrated framework. In addition to his academic contributions, Richard Blundell has been at the forefront in pushing for more evidence-based research and policy advice. He has taken a leading role in demonstrating the importance of enabling researchers an enhanced access to micro data for their research.

Throughout his career, Professor Blundell has been and still is intensely engaged in the policy debate by applying his empirical methods to real-world policy questions. For instance, he is one of the editors of the Mirrlees Review of Tax Reform. In this role, he provided practical recommendations for improving UK tax policy and helped developing the pillars of an efficient modern tax system. This project drew together insights on labor supply and commodity demand, along with basic principles of tax design, to offer suggestions for reforming the UK tax system. Its findings have been discussed in both academic and policy circles in the UK and elsewhere.

In a similar spirit, IZA recently proposed a tax reform concept for Germany, which is based largely on methodology developed by Richard Blundell. The key element of IZA's concept for a simple, efficient and fair tax system are plans to integrate regressive social insurance contributions into the progressive system of income tax, meaning that social insurance contributions and income tax are no longer levied separately, instead that there is consequently only one payment to be made to the state (see Löffler et al. 2012). An integrated system of income tax would not only finance the country's social security system but also other state services, including investment in education and infrastructure. Simulations for Germany show that such a reform would reduce the burden above all on low- and middle-income earners, while increasing the fiscal burden on the top 10% of earners. Importantly, the proposal would have a considerable effect on the labor market, leading to the creation of an estimated half a million new jobs due to the more efficient tax structure.

IZA's special focus on Labor Policy reflects the great importance of evidence-based policy making. With its team of policy experts, IZA analyzes and evaluates existing policies on the national and international level, utilizes the results of the IZA program areas, and develops concrete policy proposals. With its own behavioral techniques, *IZAΨMOD* (IZA Policy SImulation MODel), IZA is able to analyze the labor supply reactions and employment effects of tax and benefit reforms in Germany (see Peichl, Schneider and Siegloch 2010). *IZAΨMOD* enables the institute to give policymakers information about appropriate modifications of planned initiatives even prior to their adoption. *IZAΨMOD* consists of three main components. The basis is a static microsimulation model for the German tax and benefit system. The second module is an econometrically estimated labor supply model, which takes into account behavioral reactions to tax reforms. The estimation and modeling in this step relies heavily on methods developed and applied by Richard Blundell. The third component is a labor demand module, which completes the analysis of the labor market and allows a global assessment of the effects of policy measures.

In addition to making own reform proposals, *IZAΨMOD* is used to evaluate current policies. For instance, Eichhorst et al. (2012) analyze the recent trend of increasing marginal employment at the expenses of low-wage full time jobs. It finds substantial disincentives set by the current legislation and calls for a reform of the mini-job regulations as well as income taxation (married couple tax splitting) and finds that these would have a large positive impact on the labor market.

The connection between IZA and Richard Blundell is very long. He joined IZA as a Research Fellow in October 1999, shortly before IZA's first anniversary. But even before this time, Richard was already connected to IZA Director Klaus F. Zimmermann, who had introduced him and his co-author Costas Meghir to the labor group at the University of Mannheim to which he was affiliated with in the 1980s, to discuss Richard's early work on labor supply and taxation. Later on, they both contributed to a volume on *"Welfare and Work Incentives"* edited by Tony Atkinson and Gunnar Mogensen in 1993. While Richard discussed *"Taxation and Labour-Supply Incentives in the UK"* (Blundell 1993), Zimmermann (1993) focused on Germany. Both studies conclude that tax incentives matter but that the resulting labor supply elasticities are usually small. They also highlight the importance of using micro data, microsimulation models and microeconometric techniques when analyzing the effects of tax reforms on labor supply.

When assessing the effects of policy reforms on the labor market, most studies only focus on labor supply. The interaction of supply and demand side is not explicitly modeled, which might lead to biased estimates of potential labor market outcomes. This is discussed in Chapter 5 of this volume based on Blundell, Ham and Meghir (1987). At IZA, Peichl and Siegloch (2012) propose a straightforward method to remedy this shortcoming. The authors use information on firms' labor demand behavior and feed them into a structural labor supply model, completing the partial analysis of the labor market on the microdata level. They show the performance and relevance of our extension by introducing a pure labor supply side reform, the workfare concept[1], in Germany and simulating the labor market outcome of the reform. The results show that demand effects offset about 25 percent of the positive labor supply effect of the policy reform.

Discrete choice models have become the workhorse in labor supply analyses. However, they are often criticized for being a black box due to numerous underlying modeling assumptions, with respect to, e.g., the functional form, unobserved error components or several exogeneity assumptions. In a methodological contribution, IZA researchers open the black box and show how these assumptions affect the statistical fit of the models and the estimated labor supply elasticities (Löffler, Peichl and Siegloch 2014). In total, the authors estimate more than 3,250 different model specifications. The results show that the specification of the utility function is not crucial for performance and predictions of the model but that the estimates are extremely sensitive to the treatment of wages – a neglected dimension so far. Estimated labor supply elasticities vary greatly depending on whether wages and preferences are estimated separately or jointly; and whether and how wage prediction errors are taken into account (not at all, single draw or full distribution integrated out). As a consequence, the authors propose a new estimation strategy which overcomes the highly restrictive but commonly made assumption of independence between wages and the labor supply decision.

Static models of labor supply are very useful to predict ex ante the effect of tax-benefit policy reforms or more generally to provide an order of magnitude of the short-term response to financial incentives. Several excellent surveys report evidence on labor supply elasticities for different countries and different periods.[2] However, the literature only reaches a consensus on few aspects, establishing that own-wage elasticities are largest for married women and small or sometimes negative for men. In terms of magnitude, a large variation in labor supply elasticities is found in the literature. For instance, Blundell and MaCurdy (1999) report uncompensated wage elasticities ranging from –0.01 to 2.03 for married women, while Evers, de Mooij and van Vuuren (2008) indicate huge variation in elasticity estimates. Admittedly, much of the variation across studies is due to different methodological choices, including the type of data used (tax register data or interview-based surveys), selection (e.g. households with or without children), the period of observation (see Heim 2007) and estimation method.

Bargain and Peichl (2013) have collected empirical evidence focusing on 15 European countries and the US. For each demographic group (e.g. married or single individuals with or without children), they observe a large variance in estimates across all available studies, pointing to data year and estimation methods as the main sources of variation. The authors show that international comparisons based on existing evidence are generally imperfect and incomplete, with insufficient common support across studies to conclude about genuine differences in labor supply responsiveness between countries. The only clear pattern in the literature is that elasticities are larger for women in countries where their participation rate is lower. However, estimates are missing or scarce for several EU countries and also some demographic groups, such as childless single individuals.

To close some of the gaps in the literature, Bargain, Orsini and Peichl (2014) provide labor supply elasticity estimates for 17 European countries and the US based on a harmonized approach using a structural discrete choice model as used in well-known contributions for Europe (van Soest 1995; Blundell et al. 2000) or the US (Hoynes 1996; Keane and Moffitt 1998). They report compensated and uncompensated own-wage and income elasticities separately for men and women, both single and in couples, with and without children, as well as cross-wage elasticities for individuals in couples. The authors find that that own-wage elasticities, both compensated and uncompensated, are relatively small and much tighter across countries than suggested by previous results in the literature. In particular, estimates for married women lie in a narrow range between 0.2 and 0.6, with significantly larger elasticities obtained for countries in which female participation is lower (e.g., Greece, Spain, Ireland). Elasticities for married men are smaller and even more concentrated, while elasticities for single individuals show substantial variation with income levels. Consistent results are also found across countries, with important implications for welfare and optimal tax analysis: the extensive margin systematically dominates the intensive margin; for single individuals, whereas income elasticities are extremely small. Using a decomposition analysis, Bargain, Orsini and Peichl (2014) rule out differences in tax policy, wage/hours levels and demographics as explanations for cross-country differences in labor supply responses. Accordingly, the results are consistent with Western countries having genuinely different individual and social preferences, e.g. different preferences for work and childcare institutions.

IZA is not only applying the methods developed by Richard Blundell and others but also helping in advancing the field. It was in order to demonstrate the state of the art in labor supply modeling and to compare the performance of quasi-experimental techniques with the established structural models that IZA invited about 20 international top researchers to a workshop in Dublin in 2012. Of course, Richard Blundell contributed to this workshop as well. In his keynote speech, he emphasized the importance of labor supply for consumption smoothing as a response to (unemployment) shocks. Overall, the workshop clearly showed that research on labor supply behavior has made tremendous progress recently. Different approaches exist to evaluate the labor supply effects of policy reforms and researchers need to scrutinize carefully which identification strategy is best suited for their research question. High unemployment rates in many European labor markets point to the need to improve labor market institutions, including tax and transfer systems. Credible research results can help policy makers to improve labor market regulations and to achieve the intended policy goals. IZA will continue leading the research in this area and organizing future workshops on this topic.

IZA, together with 28 partner institutes and universities, is a key contributor to the "NEUJOBS" research project which is financed by the European Commission under the 7th Framework Program. The objective of NEUJOBS is to analyze possible future developments of the European labor markets under the main assumption that European societies are now facing or preparing to face main transitions that will have a major impact on employment. IZA's contribution to the NEUJOBS project is to model structural and behavioral aspects of labor supply and labor demand in order to shed light on the important question how employment in European labor markets might look like in 2030 or even beyond. The labor force is projected to shrink in most EU countries while becoming older and more educated in composition. Dolls et al. (2013) add a behavioral dimension to this trend and find divergent implications across countries. The behavioral changes (hours worked per worker) can worsen or dampen the structural changes (the size of the labor force). Notable challenges are expected to arise in Austria, Germany and Spain.

In order to provide researchers interested in labor supply a comprehensive reference, this volume brings together a number of key papers which were the basis for awarding Richard Blundell the 2012 IZA Prize in Labor Economics. Many of the papers that are included in this collection are widely read classics. Of course, selecting papers for such a volume can only be incomplete given Richard Blundell's impressive list of publications. For instance, Blundell, Chiappori and Meghir (2005) and Blundell et al. (2007) analyze collective models of labor supply. They find that the estimates of the sharing rule show that male wages and employment have a strong influence on bargaining power within couples. Blundell, Walker and Bourguignon (1988) analyze the optimal taxation of family income when accounting for labor supply incentives. The authors argue that the tax system should discriminate between individuals in the same household if their labor supply responses are different. Blundell, Meghir and Neves (1993) suggest an intertemporal model for labor supply and consumption. Blundell et al. (1988) investigate the practical importance of the functional specification of labor supply equations for the analysis of tax/benefit reforms. In addition to methodological contributions, Richard has also written widely read and cited surveys. For instance, Blundell and MaCurdy (1999) reviewed different approaches to modeling labor supply.

Being selective was not easy. We had to focus on the topic of labor supply and taxation for which Richard was awarded the IZA Prize 2012. We also tried to achieve a good balance between methodological contributions and policy applications from different periods of Richard's career. The resulting book consists of three parts. In the first part, the topic is introduced followed by two recent and overarching chapters on labor supply and policy. Part II is more analytical and conceptual. This section presents Richard Blundell's early work, arising from demand theory and the integration of labor supply and commodity demand modeling. These initial papers are largely concerned with tackling methodological issues or with grounding the analysis of labor supply in an appropriate theoretical model. The third and last part illustrates how these analytical techniques can be put to good use through various policy applications and their impact on household behavior. The papers in this section showcase the wide range of policies that bear on labor market activities, particularly those of low-income households. The careful analysis of welfare-to-work policies, job search programs, and tax policy provides not only insight on each of these initiatives in particular, but also a reminder of the "choice of instrument" problem when tackling redistribution while preserving work incentives. The different chapters will be introduced in the following.

Part II of this book gives an overview of the topic labor supply and taxation. Chapter 1 highlights the role of empirical evidence in tax policy design. The context for the discussion is the recently published *Mirrlees Review* of tax reform. This chapter highlights the taxation of earnings and also comments on earnings taxation in the context of VAT base-broadening reforms and the taxation of capital. Five different topics are discussed: (i) key margins of adjustment, (ii) measurement of effective tax rates, (iii) the importance of information and complexity, (iv) evidence on the size of responses, and (v) implications from theory for tax design.

Chapter 2 proposes a systematic way of examining the importance of the extensive and the intensive margins of labor supply in order to explain the overall movements in total hours of work over time. This analysis is applied to the evolution of hours of work in the US, the UK, and France and shows that both the extensive and intensive margins matter in explaining changes in total hours.

Part III summarizes Richard's methodological contributions. Chapter 3 deals with modeling the joint determination of household labor supply and commodity demand. In addition to choosing the allocation of total expenditure between commodities, households may also be able to make decisions over the allocation of their time between market work and leisure. In both theoretical and empirical work it has often been the case that these decisions have been analyzed separately. This chapter shows the theoretical attractions of considering the joint determination of the allocation of time between work and leisure and the allocation of total expenditure between commodities in a utility maximizing framework. The empirical importance of the joint determination model over the separate determination of labor supplies and commodity demands is then evaluated using micro data.

Chapter 4 is rather technical and estimates a utility maximizing model of the joint determination of male and female labor supplies. The emphasis is on the estimation of within period preferences that are consistent with inter-temporal two-stage budgeting under uncertainty. Chapter 5 extends the standard model of labor supply by accounting for unemployment. In the basic empirical model of female labor supply any woman reporting zero hours of work is assumed not to want to work. Labor supply parameters are estimated from a likelihood function where the probability of an individual recording zero hours is equivalent to the probability of her not having positive desired hours of work. All non-participants in this 'Tobit' model are assumed not to want to work. The assumptions of the Tobit model stand in sharp contrast to those made in calculating the labor force statistics. In particular, those reporting zero hours of work but seeking work are considered to be labor market participants, and a measure of the unemployment rate is often formed from data on such individuals. The purpose of this chapter is to modify the standard model of female labor supply to allow for unemployed workers who want to work at their perceived market wage but cannot find a job.

Chapter 6 is based on the 1998 Econometrica paper with Alan Duncan and Costas Meghir which was awarded the Frisch Medal of the Econometric Society in 2000. It provides a methodology for estimating labor supply responses using tax reforms from the 1980s. However, changing sample composition, aggregate shocks, the changing composition of the tax paying population and discontinuities in the tax system create serious identification and estimation problems. The paper develops grouping estimators that address these issues. The results reveal positive and moderately sized wage elasticities and as well as negative income effects for women with children. In addition, the developed 'grouping IV estimator' has become a standard tool in empirical labor economics.

Chapter 7 investigates single women's labor supply changes in response to three reforms that affected individuals' work incentives using British panel data. It is found that only the 1999 reform led to a significant increase in single mothers' hours of work. The mechanism by which the labor supply adjustments were made occurred largely through job changes rather than hours changes with the same employer.

Chapter 8 is the first of Part IV on policy applications and is a compilation of several papers on increasing work incentives in the UK. It deals with the labor market impact of the Working Families Tax Credit (WFTC) which, in 1999, replaced the Family Credit as the main package of in-work support for families with children in the UK. This chapter analyzes the impact of WFTC on hours and participation. To simulate labor supply responses, the authors use a discrete behavioral model of household labor supply with controls for fixed and childcare costs, and unobserved heterogeneity. They also simulate a number of scenarios regarding the take-up of the credit, entry wage level and hourly childcare price. The results show that the WFTC increased participation rates among single mothers by around 2.2 percentage points, while participation rates for married women decreased. As a consequence of both effects, the simulation results indicate a small increase in overall participation of around 30,000 individuals.

In-work benefit reforms seek to reduce poverty and promote employment among low-income families. Using evidence from similar policies in the USA and Canada, Chapter 9 reviews the likely impact of recent UK reforms. The focus is on employment and hours. In particular, the paper examines the effectiveness of the new Working Families Tax Credit in the UK in increasing employment among low-income families. It presents evidence suggesting modest increases in employment for single parents and workless married couples with children, but with some off-setting reductions in employment in two-earner couples with young children.

Alongside the growth in overall employment and the steady rise in average real incomes over the 1990s, the UK experienced a concentration of worklessness and low pay among certain groups in society. This was particularly acute for families with children, but was also reflected in the frequency of spells out of work by the young and by the falling attachment to the labor market of older men. In response, the focus of welfare policy shifted towards "making work pay". The Working Families Tax Credit and the New Deal were central among the policy options that were implemented. Chapter 10 considers the validity of the arguments underlying this shift in welfare policy and asks: which policies work and why? It examines two broad classes of policy options that are motivated by the "making work pay" objective: (i) active labor market programs that involve wage subsidies together with improved job matching; and (ii) earned income tax credits that supplement wages for working low-income families. These programs have many features in common. Using evaluations of UK reforms this chapter brings empirical evidence into the debate on the effectiveness of these programs and assesses which aspects of the design of welfare-to-work programs work well and which aspects could be improved.

Chapter 11 also compares active labor market policy and employment tax credits based on evidence from UK reforms. The background motivation is that many welfare-to-work programs in both North America and Europe are directed at making work pay for the low skilled. This chapter identifies two alternative policies that are motivated by this same objective: active labor market programs that involve wage subsidies together with improved job matching; and earned income tax credits that supplement wages for working low-income families. Although sharing similar concerns over labor market incentives for low-skilled workers, these alternative policies typically differ in many important ways. This chapter evaluates the impacts of two such programs designed to enhance the labor market attachment of low-wage workers in the UK.

Chapter 12 exploits area-based piloting and age-related eligibility rules to identify treatment effects of a labor market program for mandatory job search assistance – the New Deal for Young People in the UK. A central focus is on substitution/displacement effects and on equilibrium wage effects. The program includes extensive job assistance and wage subsidies to employers. The results show that the impact of the program significantly raised transitions to employment by about 5 percentage points. The impact is robust to a wide variety of non-experimental estimators.

Chapter 13 is based on the Adam Smith Lecture 2005 presented at the joint EALE/SOLE World Meeting in San Francisco, June 2005. In this paper, the impact and optimality of Earned Income Tax Policies to tackle the low labor market attachment and high incidence of poverty among certain groups is analyzed. EITC policies have taken a central position in recent EU labor market policy debate and the focus of this chapter is on actual reforms over the last decade in the UK.

Chapter 14 examines the optimal design of low-income support using a structural labor supply model. The approach incorporates unobserved heterogeneity, fixed costs of work, childcare costs and the detailed non-convexities of the tax and transfer system. The analysis considers purely Pareto improving reforms and also optimal design under social welfare functions with different degrees of inequality aversion. The authors explore the gains from tagging and also examine the case for the use of hours-contingent payments. Using the tax schedule for lone parents in the UK as policy environment, the results point to a reformed non-linear tax schedule with tax credits only optimal for low earners. The results also suggest a welfare improving role for tagging according to child age and for hours-contingent payments, although the case for the latter is mitigated when hours cannot be monitored or recorded accurately by the tax authorities.

Finally, in Part V Richard Blundell summarizes and concludes his research on labor supply and taxation over the last 30 years and provides an outlook for future research – explicitly accounting for human capital decisions, the dynamics of life cycle family labor supply decisions and restrictions on job offers and choices of hours of work. Richard concludes by stating that analyzing labor supply and taxation is clearly set to remain an active and emerging field of economic research for many years to come.

The future of labor is being set today through the decisions of individuals and policy makers. Today's labor market and today's policies set in motion decisions about participation, education and retirement which cast a long shadow into the future, as individuals progress through their careers over the life cycle. With the future of labor program area, IZA seeks to improve our understanding of the effects of today's labor market and labor policies on future labor market outcomes in the hope that better understanding will lead, eventually, to better policy. Analyzing labor supply and how it is shaped by policies is a key element of this research area and Richard Blundell is among the most important contributors. Continuing this line of research is high on IZA's research agenda in order to provide evidence-based policy advice in the spirit of Richard Blundell's work. The IZA Prize in Labor Economics 2012 honors the work of an eminent scholar who has greatly shaped our view on labor market analysis and economic policy. Richard is a role model to many researchers across the globe. His work is a remarkable combination of academic excellence and policy relevance.

Part II
Overview

Introduction

What is it that makes economic research so compelling? My short answer would be the challenges of a continually changing economic environment and the interplay between policy questions and empirical research. Nowhere is this more the case than in the study of labor supply and taxation. How should we develop the foundations for rigorous empirically-based policy analysis? How should we balance policy and methodological research? New evidence, new (applied) theory and a new economic environment, keep these questions alive and move the frontier forward. Research has to run just to keep up!

The studies in this volume are dedicated to addressing these challenges. In the first chapter I begin with the question: what is the role of empirical research in policy analysis? To answer this I consider the role of evidence loosely organised under five headings - my 'five steps to heaven' – heaven as seen by an addict of empirical economic research!

I have chosen these five steps reflecting on the long history of studies presented in this volume. Many of these draw on research and discussions with many wonderful co-authors, students and colleagues from University College London over the years. Much of this research has also been driven by the continuing flurry of key policy questions raised in day-to-day work at the unique Institute for Fiscal Studies. The building of the UCL-IFS research and policy base over nearly three decades has been the bedrock of this research and I have been a lucky fellow traveller in this journey. This work is team research and I thank my many co-authors, mentors and students in this venture.

The first of my *five steps* is entitled 'Key margins of adjustment to reform'. In other words – getting the facts straight. Where are there important differences in behaviour that may have resulted from policy reform? From a labor supply perspective this is answered in the second chapter that investigates the key differences across countries and demographic groups in key measures of labor supply.

The second of the steps is 'Measurement of effective tax incentives'. Exactly how does the policy change the incentives faced by individuals? If we see little response it may be that individuals are not responsive, but it may also be that there was nothing for them to respond to! Getting the measurement of incentives right requires a huge investment of integrating the various aspects of the tax and welfare benefit system together with the data on labor supply – something my colleagues at IFS have been involved in over many years. Indeed they have created an unparalleled research resource on which I have drawn on extensively in many chapters in this volume.

Next comes 'The importance of information and complexity'. We, as economists, are not always so proficient at figuring out effective incentives in tax systems, so we might well ask how well do the agents themselves understand them. What is the role of information, stigma and cognition? Governments too can exploit 'less than salient' taxes. Transparency seems a first principle in reform but understanding responses to existing reforms requires a careful modelling of this aspect of behaviour. With the cocktail of multiple benefits and taxes facing low-wage workers in the UK, this has been a key component in reform simulation models but perhaps this behaviour should attract more attention by public finance economists more generally.

Forth on the list is the microeconometrics core – 'Evidence on the size of responses'. The aim is to understand causal impact of policy on behaviour. To recover the 'deep' parameters that reflect preferences and constraints. Here I argue that an eclectic mix of experimental and structural approaches can provide a powerful evidence base for reform.

Finally, the theory of mechanism design and optimal taxation comes into play – 'Implications form theory from optimal policy design'. Insights from theory for constrained efficient policy under different modelling assumptions on social welfare functions, individual preferences, expectations and credit markets are central here.

In sum, I hope the studies in this volume tee-up a large and exciting agenda for future research. In the concluding chapter I take a look toward the future. Where is the action likely to be? Not only is there room to improve what we have done already, there is also a continual requirement to adapt to new economic environments, new insights from theory, new policy ideas and new data measurements. But that's what makes this area of research one of the most compelling in applied economics!

As a final precursor I would like to thank the numerous colleagues and students who, over the years, have helped me with the work in this volume, especially those who have co-authored some of the papers presented here. Much of the research would have not been possible without the funding from the Economic and Social Research Council of the UK through the ESRC Centre for the Microeconomic Analysis of Public Policy at IFS. For this I am eternally grateful. Finally, I also want to thank two talented economics undergraduates at UCL, Fabien Eckert and Hannes Ansorg, who helped prepare this volume.

1

Tax Policy Reform: The Role of Empirical Evidence

1.1 Introduction

How should evidence be used in the study of tax design? What is the appropriate balance between theory and empirics? These questions lay at the heart of the *Mirrlees Review*. Motivated by the aim to develop a broad set of principles for what makes a *good tax system*, the *Review* was an attempt to base tax reform on the large body of economic theory and empirical evidence. It was inspired by the Meade Report (1978) with the idea to review tax design from first principles for modern open economies in general and for the UK in particular. The UK over the past thirty years would be the working laboratory.

The *Mirrlees Review* was published in two volumes: *Dimensions of Tax Design* (Mirrlees et al. 2010) bringing together expert evidence across a wide range of aspects of tax reform, and *Tax by Design* (Mirrlees et al. 2011) setting out the conclusions and recommendations. This paper examines the role of evidence used in the derivation of the recommendations for reform. It also examines the linkages between theory and empirical evidence. To maintain consistency and coherence in the discussion, the focus here will be on the taxation of earnings although the *Review* itself concerned all aspects of the tax system. The discussion is organized loosely under five related headings:

The original version of this chapter was published as: Blundell, R. (2012). Tax Policy Reform: The Role of Empirical Evidence, in: Journal of the European Economic Association, 10(1): 43–77.

I would like to thank the Editor, referee and participants at the JEEA Lecture at the AEA meetings and at the FBBVA Lecture in Madrid, for helpful comments. I am also grateful to my co-authors and co-editors on the Mirrlees Review for the many discussions over the course of the Review. All figures and tables not directly sourced are from the second volume of the Mirrlees Review, Tax by Design (2011).

(i) Key margins of adjustment.
(ii) Measurement of effective tax rates.
(iii) The importance of information and complexity.
(iv) Evidence on the size of responses.
(v) Implications from theory for tax design.

The first of these headings highlights the importance of establishing empirical facts about key aspects of behavior where we think taxes could have an impact. The second reinforces a pervasive theme of the *Review* which was to consider the tax system as a whole and examine the *wedge* created by all aspects of the tax system, including the implicit tax rates in the benefit and tax-credit systems. This also naturally motives the third heading which relates to the understanding of the incentives implicit in the tax and benefit system by the individuals, households, and firms themselves and the stigma and hassle costs involved by those accessing the system. The forth heading is the core of any rigorous empirical analysis and concerns the robust measurement of the causal impact of tax reforms. Here I suggest the use of a mix of (quasi-)experimental and structural approaches with the experimental approaches acting as a 'reality check' on the structural model. Structural models allow the study of behavior in counterfactual environments and it is difficult to envisage a complete empirical analysis of tax design analysis that does not draw on such counterfactuals.

Under the final heading, these empirical relationships are brought together with the structure of mechanism design from economic theory to determine efficiency costs, overall optimality, and improvements to tax design. There are three key ingredients to any optimal tax analysis: the accurate measurement of response elasticities, the detailed description of the distribution of income, and some view of social welfare weights. The first two of these are positive and can be learned from a careful evidence-based analysis. The last is normative and therefore something over which reasonable people may differ. The aim here is to draw broad evidence-based conclusions while making fairly weak assumptions on social welfare weights, perhaps assuming no more than that they are declining in some measure of equivalized income.

Why the focus on earnings taxation? Earnings taxation is ideally suited for examining the role of evidence in tax design. There are substantial empirical results on labor supply responses to tax reform for individuals and families, see Blundell and MaCurdy (1999) and Meghir and Phillips (2010) for surveys. This research has emphasized the need to distinguish between the intensive and extensive margins of labor supply – that is between the decision of whether to work or not and how much to work, respectively. It has also shown clear differences in responses by age, gender, and family composition. Both of these observations are central to tax design. Further, tax return information provides additional evidence on taxable income elasticities, highly relevant for the design of earnings taxation, see Gruber and Saez (2002), for example. We will argue that this evidence naturally supplements and extends work on employment and hours of work responses to tax reform.

The next five sections of this paper reflect these five aspects of the empirical analysis of reform. This is not meant to imply that the taxation of earnings should stand separately from the design of the rest of the tax system. As the *Review* recommendations volume *Tax by Design* makes clear, any comprehensive reform must bring together all aspects of taxation. Indeed, the taxation of earnings bears the brunt of much of the tax reform proposals through the need to adjust for changes in redistribution and work incentives induced by other aspects of the reform package. Therefore, to round off this paper, the discussion turns to the interplay between earnings tax design and base-broadening reforms to VAT, as well as to the taxation of capital and reforms that seek to align effective tax rates across all sources of income.

1.2 Key Margins of Adjustment

With the focus on earnings tax reforms, our analysis begins with the key changes in lifetime employment patterns over the last three decades. This sets the scene for understanding where, over their working life, individuals and families are most likely, and most able, to respond to tax reform.

The recent history of variation in hours and employment has been made up of three key trends which we will argue also point to the three key margins where responses to tax reform are most likely to occur: a decline in employment among men especially at older ages, a strong rise in employment and total hours of work for women, and a decline in employment among those in their late teens and early 20s reflecting the increase in educational attainment over this period.

As has already been noted, an important distinction in analyzing labor supply responses is between the extensive (whether to work) and intensive (how much to work) margins of labor supply. Although it is the case that hours of work are often found to respond less than employment decisions, Blundell, Bozio and Laroque (2011a) show that the intensive and extensive margins both matter in explaining the broad changes in total hours over the last three decades in the UK, France and the United States. But they matter in different ways for different

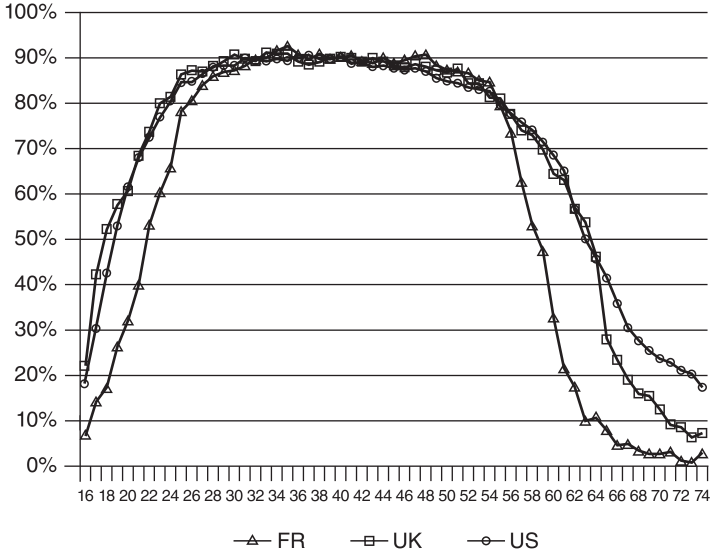


Figure 1.1. Male Employment in the UK, 2007
Source: Blundell, Bozio and Laroque (2011b).

age and demographic groups. For men, variations in the extensive margin occur mainly at the beginning and at the end of their working lives. These are the schooling-work margins and the early retirement margins: Figure 1.1, from Blundell, Bozio and Laroque (2011b), provides a broad view of employment rates by age for the UK, France and the United States in 2007 (just before the onset of the most recent recession). The similarity of average employment rates in 2007 for men aged 30–54 in these three economies is striking. It suggests that differences in employment are concentrated at early and later points in the working life. Heckman (1993), Prescott (2004), Ohanian, Raffo and Rogerson (2008), and Gruber and Wise (1999) have all pointed to the importance of the extensive margin at these points of the life cycle.

The extensive margin is not the end of the story. Figure 1.2 points out that hours differences, conditional on employment, matter too for men and they matter across the working life. Although it is unlikely that tax and benefit systems alone explain all these differences, in any discussions of tax reform it would seem unwise to play down the intensive margin too much.[1]

For women, Figures 1.3 and 1.4 show that hours conditional on employment and employment itself both vary across the working lives.

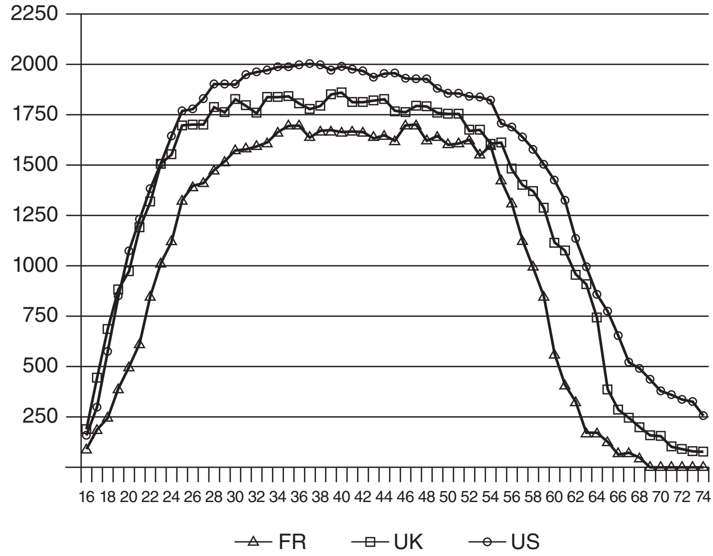


Figure 1.2. Male Total Hours Worked in the UK
Source: Blundell, Bozio and Laroque (2011b).

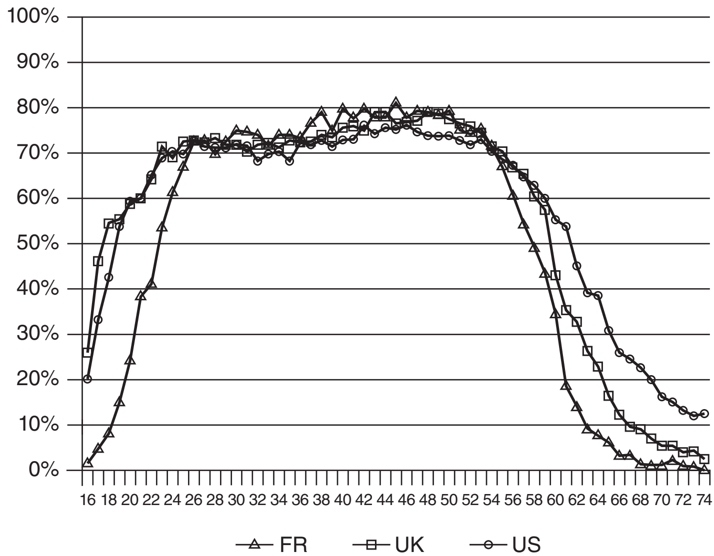


Figure 1.3. Female Employment in the UK, 2007
Source: Blundell, Bozio and Laroque (2011b).

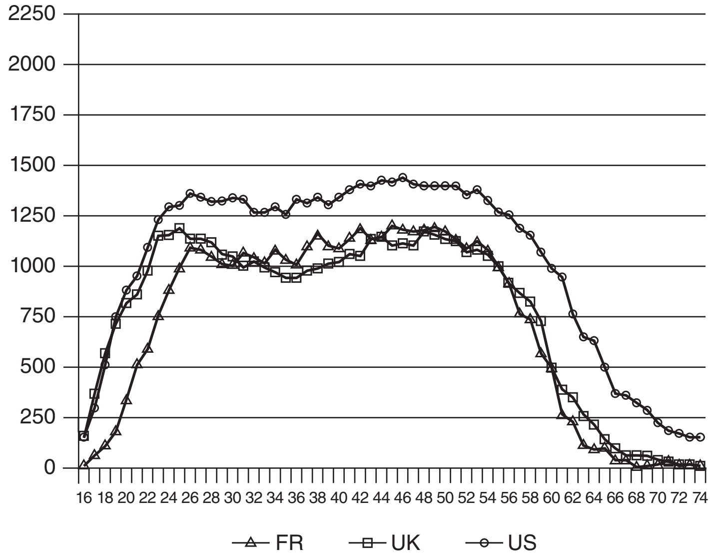


Figure 1.4. Female Total Hours Worked in the UK, 2007
Source: Blundell, Bozio and Laroque (2011b).

As was the case for men, average employment rates in 2007 were surprisingly close at ages between the late 20s and early 50s. Again, it is at the early and later periods in the working life where the extensive margin choices become important. We will also point to important variation at the extensive margin for mothers with pre-school children and with lower levels of education. Hours of work conditional on employment for women show more variation over the life cycle, especially in the UK where there still remains a dip around child- bearing ages. For women with younger children it is not usually just an employment decision that is important, it is also whether to work part-time or full-time. Some of this variation in the UK we will be able to attribute to the specific design of the tax and benefit system.

In the sections that follow we focus a little more in detail on what has happened to the labor supply of women over the recent past and relate it to some of the key changes in tax and benefit policy. To wrap up the descriptive discussion in this section it is worth examining the overall changes in labor supply in France, the United States and the UK over the three decades leading up to 2007. Figure 1.5 from Blundell, Bozio and Laroque (2011a) presents such a breakdown of total hours worked by age and gender. The huge declines in total hours

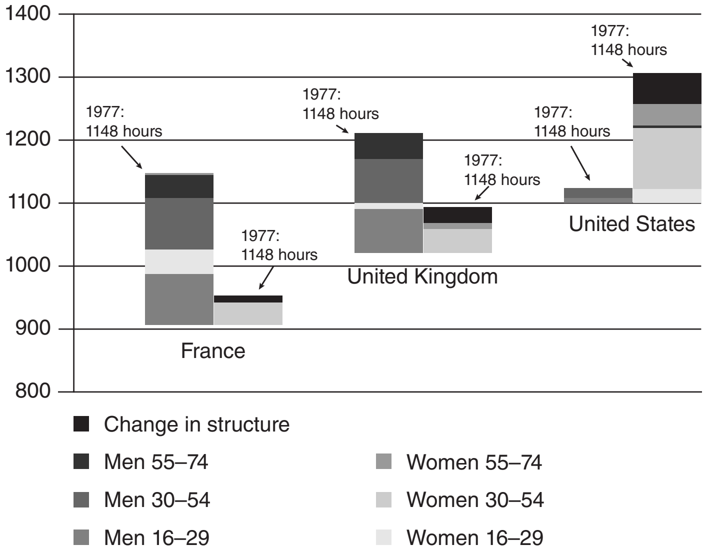


Figure 1.5. The Change in Total Hours by Age and Gender
Source: Blundell, Bozio and Laroque (2011a).

among men in the UK and France and the large rise in women's labor supply in the United States dominate the picture. These changes in total hours mask somewhat the key changes which have occurred at the extensive and intensive margin. For example, it turns out the expansion for women at the extensive margin (employment) over this period is quite similar in the United States and France – what differs is the distinctly different paths at the intensive margin.

To allocate total hours changes between the extensive and intensive margins is not a trivial exercise. While we observe the changes in hours per worker and employment, we do not know exactly how these changes contribute to the changes in total hours worked. Blundell, Bozio and Laroque (2011a) address this by developing bounds on the changes at the extensive and intensive margins which allow such a decomposition. They consider how the overall average hours worked H per person varies over time and across countries. Of course, this quantity differs across a person's characteristics, age and gender for instance. Suppose there are $j = 1,\ldots,J$ broad categories. The overall statistic H_t is computed in any year t as an average of the total hours in category j, H_{jt}, with weights equal to the population shares q_{jt}:

$$H_t = \sum_{j=1}^{J} q_{jt} H_{jt}$$

They then write total hours of work H_{jt} as the product of hours per worker h_{jt} and employment in the labor market p_{jt}:

$$H_{jt} = h_{jt} p_{jt}$$

When we observe a change in yearly hours worked per person, $H_t - H_{t-1}$, we would like to be able to know how much of the change is due to the intensive or extensive margins. First define a structural effect S_t due to the change in the composition of the population:

$$S_t = \sum_{j=1}^{J} H_{jt}[q_{jt} - q_{j,t-1}]$$

Then measure the change due to the behavior of category j, holding the population structure constant as in date t – 1, as in a Laspeyres index:

$$\Delta_{jt} = q_{j,t-1}[H_{jt} - H_{j,t-1}], \tag{1}$$

then the total change across all J categories of workers is simply

$$\Delta_t = \sum_{j=1}^{J} \Delta_{jt}, \tag{2}$$

and we have by construction

$$H_t - H_{t-1} = S + \Delta_t. \tag{3}$$

The change in total hours for any category of workers reflecting changes at the intensive margin (hours per worker), and at the extensive margin (employment) satisfies two polar exact statistical decompositions:

$$\Delta_{jt} = q_{j,t-1}\left\{\left[h_{jt} - h_{jt-1}\right] p_{jt} + \left[p_{jt} - p_{jt-1}\right] h_{jt-1}\right\} \tag{4}$$

or

$$\Delta_{jt} = q_{j,t-1}\left\{\left[h_{jt} - h_{jt-1}\right] p_{jt-1} + \left[p_{jt} - p_{jt-1}\right] h_{jt}\right\} \tag{5}$$

The first term on the right-hand side is the intensive margin, weighted in the top formula (4) with the final employment rate (as in a Paasche index) and in the bottom formula (5) with the initial employment rate (as in a Laspeyres index). The second term is the extensive margin (Laspeyres in (4), Paasche in (5)). The empirical counterparts to these are given in Table 1.1.

The indices examine what part of any overall change in hours is attributable to changes at the extensive or intensive margin for any particular subgroup of the population. The row [$I – L, I – P$] shows the bounds on the intensive margin, L standing for Laspeyres (the change in hours being weighted by the initial employment rate), P for Paasche (final employment rate). Similarly, the Laspeyres index for the extensive margin ($E - L$) (resp. $E - P$), given by the second term in equation (4) (resp. (5)), is equal to the change in employment multiplied by average hours worked at the initial (resp. final) date.[2]

Turning first to prime-age workers, the steep decline at the *intensive* margin for prime-aged men in France and the UK relative to the United States is striking. For this group the bounds are quite narrow and leave little room for ambiguity. These changes represent an enormous shift in the relative position of these countries. Table 1.1 tells us that the *extensive* margin for prime-age men in Britain and in France also falls more than in the United States, although there are declines in the United States too. As we have noted, for prime-age women it is the increase at the extensive margin that is so extraordinary, especially in the United States and in France where the bounds in Table 1.1 suggest a very similar change and one that is nearly twice the size of that experienced in the UK. Intensive margins provide an interesting picture here, falling back strongly in France, staying put in the UK while growing in the United States.

For older men and women there is a large decrease in hours per worker in France, similar in the UK, contrasting with an increase in the United States. There are falls at the extensive and intensive margins for UK men but increases at the extensive margin for UK women. This phenomenon is replicated to some extent across all countries and offsets the stronger incentives to retire earlier in the UK and in France. The contrast with the United States is stark, where at all margins and for both genders the bounds point to positive changes for older workers. The changes among the young are also sizable and predominantly

Table 1.1. The Extensive and Intensive Margins Between 1977 and 2007

Year	Youth (16–29)		Prime aged (30–54)		Old (55–74)		All
	Men	Women	Men	Men	Men	Women	(16–74)
Δ	−82	−38	−82	−82	−36	−3	−195
[I-L, I-P]	[−37,−28]	[−23,−19]	[−59,−56]	[−59,−56]	[−11,−8]	[−9,−10]	[−185,−183]
[E-L, E-P]	[−54,−45]	[−19,−16]	[−27,−23]	[−27,−23]	[−28,−25]	[7,6]	[−12,−10]
Δ	−71	−9	−70	−70	−42	10	−118
[I-L, I-P]	[−42,−36]	[−23,−26]	[−48,−45]	[−48,−45]	[−22,−19]	[−6,−8]	[−161,−167]
[E-L, E-P]	[−35,−29]	[17,14]	[−25,−22]	[−25,−22]	[−23,−20]	[17,15]	[50,43]
Δ	−19	22	−19	−19	6	38	165
[I-L, I-P]	[−6,−6]	[1,1]	[−5,−5]	[−5,−5]	[3,3]	[3,5]	[15,17]
[E-L, E-P]	[−13,−13]	[21,21]	[−14,−14]	[−14,−14]	[3,3]	[33,35]	[148,150]

Note: I-P designs the Paasche measure of the intensive margin, I-L the Laspeyres measure, with E-P and E-L respectively for the extensive margin, as described by equations (4) and (5).

Source: Blundell, Bozio and Laroque (2011a).

negative. In France and the UK there are large falls for young men at both the extensive and intensive margin.

These changes inform us as to where labor supply is likely to be most responsive to reform. They also set up the key question in the analysis of tax incentives and labor supply: How well do structural economic models explain these changes in observed behavior? For this we have to turn first to the measurement of the effective tax rates in the tax and benefit system.

1.3 Effective Tax Rates

What of effective tax rates? To understand how taxes and benefits might affect labor supply choices, we need to measure the effective work incentives implicit in the tax and benefit system. To describe the distribution of incentives implicit in the tax and benefit system, there are two summary measures that are useful to document: the effective marginal tax rate (EMTR, that is the proportion of a small increase in earnings taken in tax and withdrawn benefits) and the participation tax rates (PTR, the incentive to be in paid work at all) defined by the proportion of total earnings taken in tax and withdrawn benefits.

Perhaps the main (perceived) defects in current welfare/benefit systems is that participation tax rates at the bottom remain very high. This is certainly the case in the UK where effective marginal tax rates are well over 80% for some low-income working families. As we will see, this is mainly due to the phasing-out of means-tested benefits and tax credits. But high implicit tax rates at low incomes can be optimal for welfare functions that place a high weight on redistribution.

Consider a typical budget constraint for a single mother. A complete analysis of the effective tax rate will combine the implicit tax rates in the benefit system, the tax credit system and the income tax system. Figure 1.6 provides such a case study and shows the complexity arising from the cocktail of taxes and benefits. This constraint assumes all eligible benefits are accessed.[3]

One component of particular interest in the taxation of earnings is the tax credit system which has become an increasingly important part of the effective tax system facing low-earning families in many countries. In the UK, the earned income tax credit (called the Working Tax Credit (WTC), previously the Working Families Tax Credit) scheme has certain unique features. As with other tax credit systems, the UK system is designed to enhance income in work for those facing low rates of pay and/or higher costs of work. Figure 1.6 provides a case-study

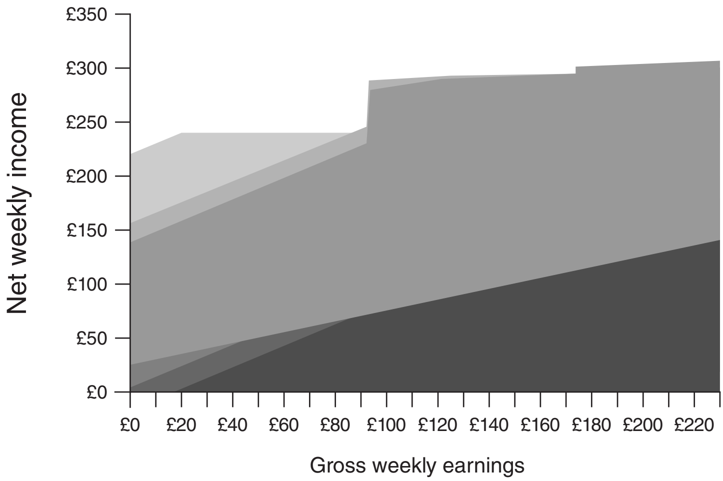


Figure 1.6. The Interaction Between Taxes and Benefits in the UK

Note: Lone parent, with one child aged between one and four, earning the minimum wage (£5.80 per hour), with no other private income and no childcare costs, paying £80 per week in rent to live in a council tax Band B property in a local authority setting council tax rates at the national average.

Source: Blundell and Shephard (2010).

budget constraint for an example low-income single parent. In the UK eligibility depends on an hours of work condition which consists of a minimum hours rule at 16 hours per week with an additional hours-contingent payment at 30 hours. There is also a family eligibility criterion which requires children in full-time education or younger. The tax credit then consists of an adult credit plus amounts for each child. There is a family net income eligibility threshold, above which the credit tapers away at 55%. Taken together with Income Support and other benefits, low-income earners in the UK can face a complex rate schedule with relatively high effective tax rates. Indeed, families in receipt of other benefits would gain less from the WTC than otherwise equivalent families not receiving these benefits.

The distribution of these tax rates by income and family type in the UK is presented in Figures 1.7 and 1.8. In an important sense it is the participation tax rate that is relevant for the employment margin, and the marginal tax rate for the effort margin. The EMTRs and the PTRs can be negative as well as positive, but they are typically positive and often high at lower incomes.

Couples with one earner and lone parents are the two distinct groups in the picture of average marginal tax rates by gross earnings in Figure 1.7.

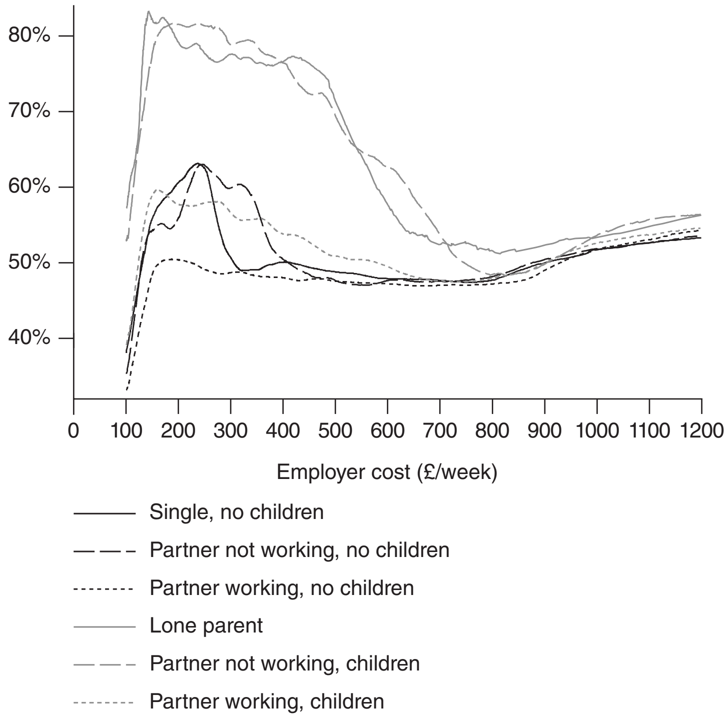


Figure 1.7. Average EMTRs for Different Family Types

They face high effective marginal tax rates when their earnings are low. High tax rates at low earnings are a distinctive feature of many tax systems and have led some commentators to question why lower-earning individuals face the highest tax rates.

But any system that redistributes income by targeting benefits towards families with low earnings and high needs will induce high effective tax rates as a natural by-product.

The effective tax rates in Figures 1.7 and 1.8 indicate the strong redistribution towards low-income families with children in the current UK tax system. Indeed, the more accurately the tax system targets low income, the higher the effective marginal tax rate on low earnings is likely to be. Not surprisingly therefore, tax schedules can easily possess the feature of high effective marginal tax rates at low earnings. It is simply the result of means-testing which is the flip-side of targeted redistribution. Whether it is optimal or not will depend, as we shall see in what follows, on the responsiveness of labor supply to these implicit

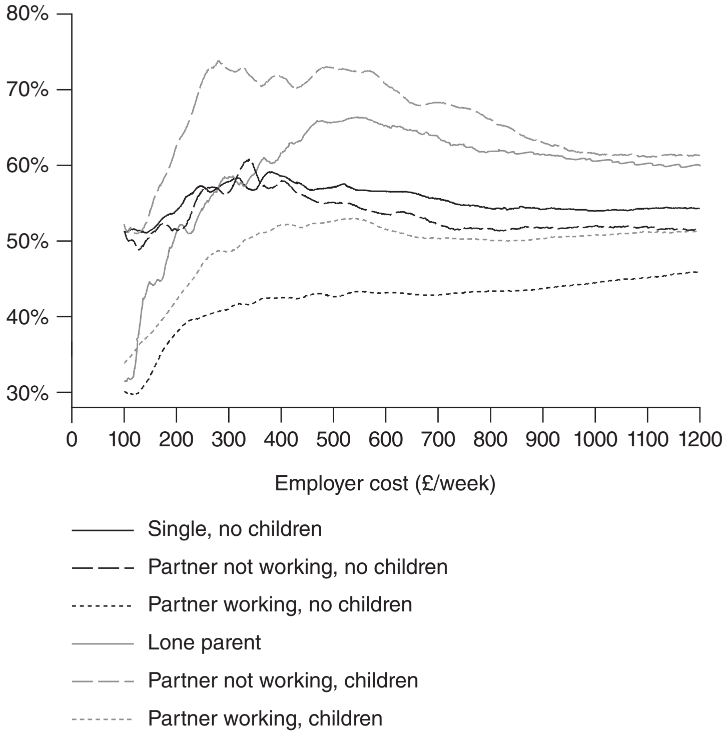


Figure 1.8. Average PTRs for Different Family Types

tax rates, on the distribution of income and on the desire to redistribute to low-income families of a particular composition.

1.4 The Importance of Information and Complexity

The EMTRs and PTRs in Figures 1.7 and 1.8 are just *local* averages at each gross earnings level. As evidenced in our discussion of the single-parent budget constraint in Figure 1.6, the current structure in the UK, and elsewhere, of multiple benefits with an array of overlapping means-tests leaves some people facing effective marginal tax rates of over 90% (see Chapter 5 of Mirrlees et al. 2011 for further details). This degree of complexity can leave some individuals unwilling and/or unable to access all the benefits and tax credits to which they are eligible.

One way to formalize some of the issues surrounding information and complexity in earnings taxation is to allow individuals who are eligible for certain benefits and tax credits not to participate or not to *take-up*. This reflects the idea that individuals may not understand the system or they may find the stigma or hassle costs involved in participating in the benefit or tax credit program too high to be worthwhile.

Typically, take-up is an increasing function of the eligible amount of benefit or tax credit but full take-up is rarely achieved, and rates of take-up can be quite low, especially for families eligible to small amounts of benefit. In Figure 1.9 the take-up rate for married couples and lone parents in the UK is plotted according to the eligible amount of Working Families Tax Credit (or its predecessor, Family Credit). This measures the proportion taking-up among those we estimate to be eligible for some benefit. It is plotted against the amount we estimate the family is eligible to. The figure suggests that the stigma or information costs are increasingly overcome as the value of take-up increases. This provides some insight into how to model take-up decisions.

To rationalize incomplete take-up of a benefit or tax credit program we follow Moffitt (1983), and the subsequent developments in Blundell et al. (2000b) and Brewer et al. (2006), and assume the presence of some *stigma* or *hassle* cost. This will provide an interpretation of Figure 1.9 and will also feed into our structural econometric specification. In turn it will help separate preferences from information and stigma costs.

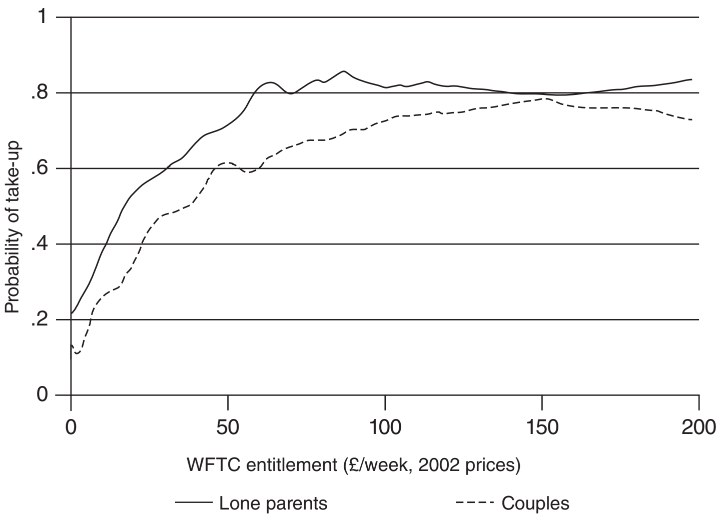


Figure 1.9. The Take-up of Tax Credits in the UK

As an example, we can imagine representing the take-up cost as an amount subtracted from the overall utility of families who claim tax credit. Suppose η represents this utility cost, then we could write

$$\eta = X'\beta_\eta + \varepsilon_\eta, \tag{6}$$

where X represents a vector of demographic and other household characteristics, β_η unknown parameter vector and ε_η unobserved heterogeneity. If the individual takes-up then his utility would be reduced by η. Families who are entitled to the credit or benefit will claim if the utility gain derived from the higher income exceeds the utility cost η.

The actual amount of taxes and benefits any individual worker receives (or pays) T will now depend on whether take-up occurs or not: we write $P = 1$ if take-up occurs and $P = 0$ if not. Suppose an individual faces an hourly wage rate w and works h hours, then the net taxes paid will not only be a function of hours of work h and total earnings wh but also whether $P = 1$ or 0. The tax function becomes

$$T = T\ (wh, h, P, X) \tag{7}$$

The net financial gain in work will depend on take-up, wages, and the choice of hours.

This framework will be further developed when we return to the structural econometric specification in Section 4.3. For now it is worth noting the practical difficulty of incorporating many different benefits. Moreover, because benefits and tax credits are based on family income, incorporating several workers in each family is also challenging.[4]

It is difficult to argue against any policy reform that clarifies which benefits and tax credits individuals are eligible to, and what the effective tax rates in the system are. One clear implication for reform is that, to be effective, the tax and benefit system requires some integration of the various benefits and tax credits. This discussion of take-up and the integration may indicate how to reduce the complexity of the tax and benefit system but it does not tell us about the appropriate tax rate structure. For this we need to know about the size of behavioral responses. We now turn to the robust measurement of the impact of tax reform.

1.5 Evidence on the Size of Responses to Tax Reform

In Section 1.1, we documented the growth of female labor supply, changes in youth employment and changes in *early retirement* behavior which must form the backdrop to any earnings tax reform agenda.

To these we add changes in demography, including growth in single person and single-parent households.

An important distinction in analyzing labor supply responses for the purposes of earnings tax design is between the extensive (whether to work) and intensive (how much to work) margins of labor supply. The microeconometric literature highlights differences between extensive/intensive responses (e.g. Heckman 1993; Blundell and MaCurdy 1999; Blau and Kahn 2007). Knowing precisely where the largest labor supply responses to incentives are to be found is a key ingredient in achieving a good earnings tax design.

Of course, other responses affecting taxable income matter for earnings tax design, certainly for the rich and self-employed. Although for many workers the employment and hours margins are the key measures of their labor supply, for other workers it is the level of effort, for any hour of work, that they can use to respond to tax incentives. For others still there will be exemptions and deductions which will allow them to change their taxable income with little change in their overall earnings. Acknowledging this, in Section 1.5, we will use the impact of taxation on taxable income to examine tax rate reform for top earnings.

1.5.1 *Alternative Approaches to Measuring the Size of Responses*

There are three dominant empirical approaches to the measurement of responses and all can prove useful in understanding earnings tax reform: the *experimental* approach using randomized control trials (RCTs), the *quasi-experimental* approach using historic reforms, and the *structural* approach based on a formal optimization model of individual and family choices. There are many comprehensive reviews of quasi-experimental approaches, see Blundell and Costa-Dias (2009) and references therein. Although few in number, there are also some influential control-trial experiments on labor supply which we will also briefly discuss.

It is difficult to envisage a full-fledged tax reform analysis that does not draw on a structural model. Policy simulation, and understanding the impact of particular rate structures, requires a model of decision-making and of the budget constraint. Herein though lies the difficulty; to fully specify the choice problem and the budget constraint inevitably requires assumptions for which we have relatively little empirical foundation.

In much of the literature on structural labor supply models, the complexity of the budget constraint has led researchers to approach estimation as a general discrete choice model with (unobserved) heterogeneity, see Blundell and MaCurdy (1999) for a review. For the most part these models have been estimated on cross-section surveys and do not attempt to directly model key *changes* in employment and hours, such as those documented in Section 1.1. Integrating the structural approach to estimation with the observed changes in incentives and policy reforms is an important step in deriving reliable structural models. There is an increasing recognition that changes in incentives over time can, and should, be used directly to estimate labor supply parameters, see Chetty (2009). In an early example of this approach, Blundell, Duncan and Meghir (1998) use exogenous changes in work incentives to estimate structural parameters for female labor supply preferences at the intensive margin. They deal directly with endogenous selection and unobserved heterogeneity and, in doing so, are able to bring to bear the insights on transparent identification from the broader microeconometric and evaluation literature.

As we will argue, deriving convincing response effects from structural models is the key to tax design analysis. Experimental and quasi-experimental analyses have an important role to play in gauging the overall size of responses and in validating structural models. Further, we will see that some mechanism design problems in taxation can be expressed in terms of a small number of sufficient statistics, see Chetty (2008), some of which may be recoverable from quasi-experimental or experimental analysis. We suggest, at a minimum, quasi-experimental analyses of policy reforms should be used as a method of validating structural models. Inevitably, the more comprehensive and the more robust is the empirical evidence, the better we can address the tax design problem.

1.5.2 *Randomized Control Trials*

Experimental evaluations of tax and benefit policies are relatively rare, although the active use of RCTs in tax policy has existed for some time. For example, the Seattle-Denver Income Maintenance Experiment (SIME/DIME) was one of a small number of large-scale income maintenance/negative income tax experiments undertaken in the late 1960s and early 1970s in the United States. The idea was to measure the disincentive effects of cash transfers on the market work of those eligible for them, see Hurd (1976), Johnson and Pencavel (1980) and West (1979).

Ashenfelter and Plant (1990) use the SIME/DIME experiment to estimate nonparametric labor supply effects of the negative tax experiment. Nonetheless, in their comprehensive review of field experiments in economics, Card, DellaVigna and Malmendier (2011) find very few experimental studies that are directed towards recovering structural economic parameters. Even fewer directed at tax and benefit reform. They note that

> as a result of the frustrations in dealing with the complex designs of the negative income tax experiments (and with the confusing message that emerged from such designs) many respected analysts adopted the view that social experiments should be designed as simply as possible. This shift away from designs that explicitly attempt to model response variation to multiple treatments and toward a single manipulation has led to a new round of criticism that the social experiments are often 'black boxes' that '... contribute next to nothing to the cumulative body of social science knowledge...' (Heckman and Smith 1995: 108).

Even so, experiments that are simply a single control treatment contrast can be useful even in respect of a more theory-based analysis such as the tax design problem addressed here. Experiments are typically designed to test a hypothesis of no effect of some particular policy against a one-sided alternative of a positive effect. They estimate an average effect of the policy and do so under relatively weak assumptions, provided the experiment is carried out correctly. They can therefore be used to assess the reliability of theoretical prediction and/or gauge the overall size of incentive effects in some tax or benefit reform.

Of the more recent randomized control trials in tax and welfare policy, perhaps the most successful and most influential has been the Canadian Self Sufficiency Project (SSP), see Card and Robins (1998). This was designed to answer the question as to whether financial incentives could encourage work among low-skilled lone parents who had spent time on welfare. The aim of the SSP was to encourage employment among this group. The reform consisted of a 50% earnings supplement – effectively a tax credit – for acquiring a job with at least 30 hours per week. This was paid on earnings up to an annual limit of $36,000 and, as in the case of earned income tax credits, the SSP award was provided directly to the individual. The individual had to have at least twelve months on welfare before they could be eligible.

The SSP has many design aspects that are similar to the British tax credits and the US EITC. There were however three big differences – eligibility required at least twelve months on welfare receipt, the eligibility criterion of full-time employment has to be satisfied within twelve months of program entry, and the program only lasted for 36 months. All of this information was fairly well disseminated to the treatment group.

The impact on the employment rate for eligibles and controls is presented in Figure 1.10 – see Card and Hyslop (2005) for more details

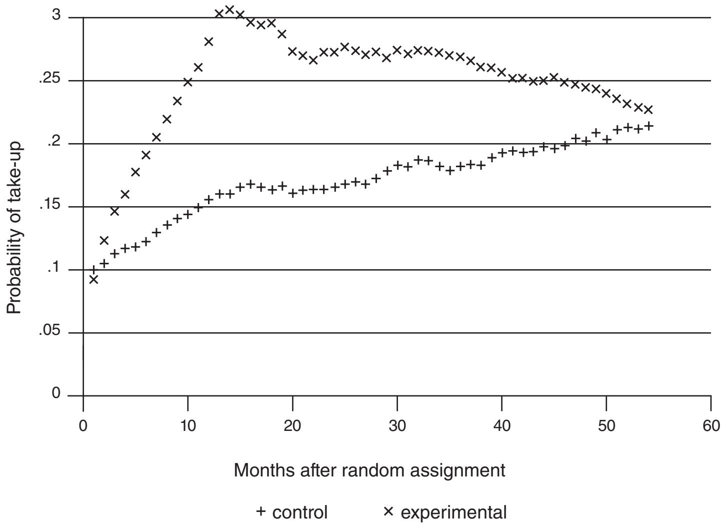


Figure 1.10. SSP, Employment Rate by Months After RA

and for further references. The increase in monthly employment, at least at the twelve-month window, was substantial. It left no doubt that financial incentives could impact quite strongly on the employment behavior of welfare mothers. However, by the end of the program the employment effect has all but disappeared. We return to this latter point in Section 4.4 in the discussion of the dynamic effects of tax and benefit programs.

1.5.2.1 QUASI-EXPERIMENTS

Quasi-experiments are by definition ex-post analyses. They evaluate the impact of an historic policy through the comparison with a control group who are ineligible but in many other aspects similar to the group targeted by the policy. Typically, this takes the form of a difference-in-differences analysis, comparing outcomes of eligibles and non-eligibles before and after the reform, estimating *average* impact of the reform on the targeted group – treatment on the treated. As in the case of an experiment, what is learned is typically only indirectly related to what is needed for optimal design. But again, quasi-experiments form an excellent base from which to (partially at least) validate structural models.

As an example of quasi-experimental validation analysis consider the analysis of earnings taxation among low-wage workers. In particular, the Working Families Tax Credit (WFTC) reforms in the UK described in Section 1.2. These were aimed at improving net income in work for low-wage parents. Our running example will again be tax reform for single mothers. This was a group who, as we saw in Figure 1.5, could face substantially improved work incentives in WFTC, certainly if their housing benefit and council tax payments were not too large.

The WFTC policy reform was enacted at the end of 1999 and Figure 1.11 shows the employment rates for working-age women 1978 through 2005. A comparison between single mothers (who are eligible) and single women without children (who are not eligible) is the contrast of direct interest. The large fall in employment in the early 1980s is finally turned around by the late 1990s, somewhat coincident with the WFTC reform. A more systematic analysis would control for observable differences in age, education, etc., between these two groups. This is what a matched difference-in-differences does (see Blundell and Costa-Dias 2009, for example) and is what we will use to validate our structural model in the next section.

Table 1.2 shows the average marginal impact of the WFTC reform for two data sets: the Labour Force Survey (used in Figure 1.11) and the smaller, but more detailed, Family Resources Survey. The Family Resources Survey is what we use in policy simulation analysis due to its comprehensive income measure. Both data sets point to a four to five percentage point increase in employment on a base of around 45%

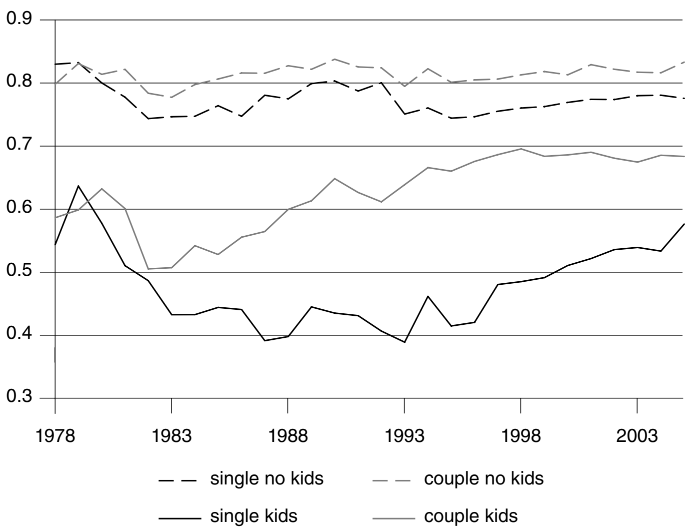


Figure 1.11. The Female Employment Rate in the UK by Demographic Type
Source: Blundell and Shephard 2010.

Table 1.2. Average Impact of WFTC Reform: Percentage Employment

Single mothers	Average effect	Standard error	Sample size
Family Resources Survey	4.5	1.55	25,163
Labour Force Survey	4.7	0.56	233,208

Notes: Data for Spring 1996 to Spring 2002. Matching covariates: age, education, region, and ethnicity.

for single mothers in comparison to women without children. As we will see, this lines up quite well with the predictions of our preferred structural model specification.

Of course, the difference-in-differences estimator of the average impact does not come for free. It relies on two key assumptions: a common trends assumption which states that time-varying unobservable differences must be common across the two groups, and a no-composition-shift assumption which states that cross-section differences in the composition of the two groups should not vary systematically. Both are strong assumptions, even given the matching covariates. But at least these are useful descriptive statistics from which to judge the predictions of any structural model.

Before moving to the structural approach it is worth noting that tax credit reforms do not just change the incentives for employment, they also change incentives for hours of work. In the UK this is especially the case given the minimum weekly hours of work requirement for eligibility. This allows a further quasi-experimental contrast. First we note that the minimum weekly hours eligibility condition was moved from 24 hours to 16 hours in 1992. Figures 1.12 and 1.13 provide the histogram of hours before and after reform and show a strong shift towards a spike at 16 hours. The incentive at 16 hours clearly has an impact on behavior (for further comparisons see Blundell and Shephard 2010). Indeed the change in hours of work over this period is clear from Figure 1.14, and can be seen to follow the hours condition in the tax credit system moving from 30 to 24 in the mid-1980s, and then to 16 in 1992. Average hours for working single mothers fall systematically over this period.

We would be unhappy with any structural model that could not reproduce these important contrasts before and after the various policy reforms. This places a stronger requirement on the model than is often adopted in structural studies. The model not only has to fit the cross-section distribution of hours and employment, it also has to be able to explain the time series *changes* in the distribution of hours and employment induced by changes in the rules of the tax and benefit system.

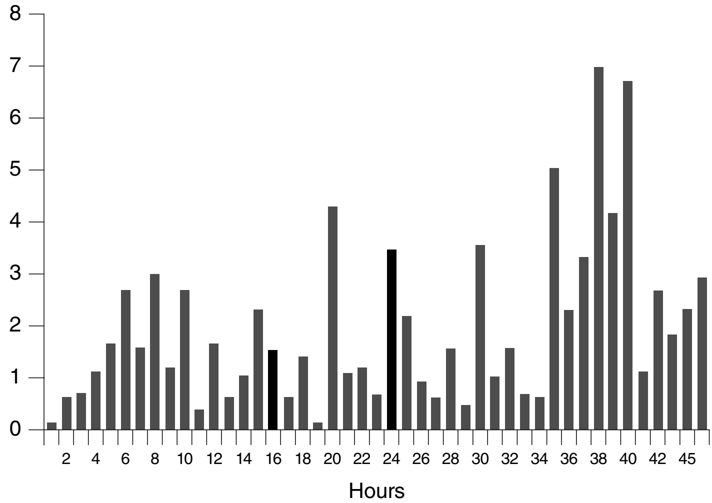


Figure 1.12. The Hours Distribution for Single Mothers, Before the 16 Hour Rule

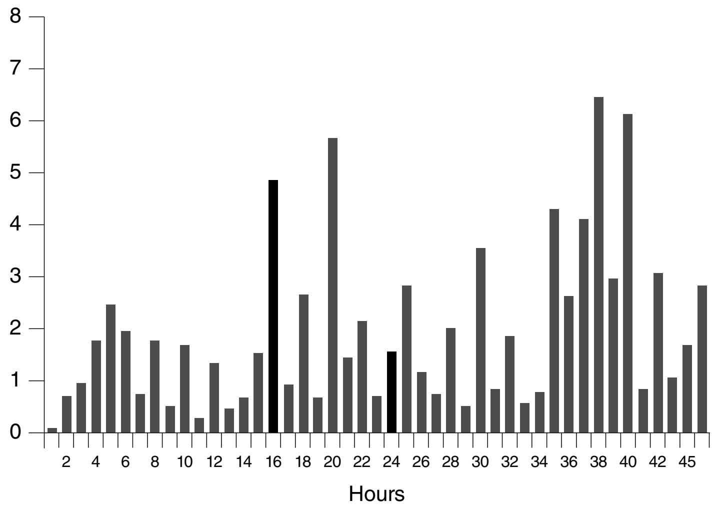


Figure 1.13. The Hours Distribution for Single Mothers, After the 16 Hour Rule
Source: Blundell and Shephard 2010.

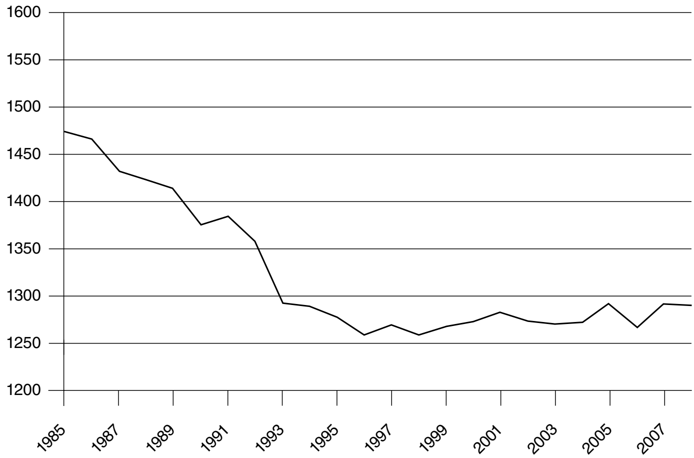


Figure 1.14. Average Annual Hours of Work: Single Mothers

Source: Blundell and Shephard 2010.

1.5.3 *Structural Models*

The structural modeling approach is useful because it fully specifies all the elements of the choice problem and the budget constraint. It therefore allows us to simulate the impact of actual or hypothetical reforms. It also allows the comparison of reforms in terms of their predicted behavior and deadweight loss. It will be a key component of any optimal tax design too.

As we have already noted, these advantages are also the potential undoing of the structural approach. The specification of constraints and choice probabilities will typically be built on strong assumptions – about the distribution of unobserved heterogeneity, about the budget constraint for each worker, and about the optimizing framework from which choices are made. A complete structural model for hours and employment choices will have to allow for unobserved work-related fixed costs, childcare costs, observed and unobserved heterogeneity, program participation *take-up* costs.

There is now a long history in the specification of such models, see Hoynes (1996), Keane and Moffitt (1998), and van Soest (1995). There are many alternative specifications for preferences, see Blundell and MaCurdy (1999). As an illustration, I follow the model developed in Blundell and Shephard (2010).[5] Section 1.2 has already detailed the complexity of the budget constraint and Section 1.3 has stressed the importance of stigma and hassle costs in modeling benefits and tax credits. Following the earlier discussion we focus on the employment and hours decision of a single worker with children and the tax credit reform in the UK. As before, we represent the *heterogeneity* across individuals in observed characteristics by X and in unobserved characteristics by ε. In Section 1.3 above, we let P be a binary indicator for the take-up of tax credit. The participation cost of taking up is $P_\eta(X, \varepsilon_\eta)$ which is subtracted from utility.

We need now to be more specific about the form of *utility* over working hours h and consumption c for each individual. Consider a utility specification:

$$u(c, h, P, X, \varepsilon) = \alpha_y(X, \varepsilon_y)\frac{c^{\theta_y}-1}{\theta_y} + \alpha_l(X, \varepsilon_l)\frac{(1-h/H)^{\theta_l-1}}{\theta_l} \\ -P\eta(X, \varepsilon_\eta), \qquad (8)$$

where θ_y and θ_l describe the shape of the marginal utility over consumption and hours respectively. The set of functions $\alpha_y(X, \varepsilon_y)$, $\alpha_l(X, \varepsilon_l)$, and $\eta(X, \varepsilon_\eta)$ capture observed and unobserved preference heterogeneity. H is the total hours available for work and leisure.

To make estimation tractable we need to make some assumptions over the various definitions and functions in (8). Blundell and Shephard (2010) assume hours of work h are chosen from some finite set H, which in our main empirical results will correspond to the discrete weekly hours points $H = \{0, 10, 19, 26, 33, 40\}$. These hours points correspond to the empirical hours ranges 0, 1–15, 16–22, 23–29, 30–36 and 37+ respectively.[6] They also set $\ln \alpha_y(X, E) = X_y'\beta_y + \varepsilon_y$ and $\ln \alpha_l(X, \varepsilon_l) = X_l'\beta_l$. This is clearly restrictive but, as we shall see in what follows, it appears to provide a reasonably accurate description of observed employment and hours for single parents in the UK.

As we noted in Section 1.3, the function $\eta(X, \varepsilon_\eta)$ is included to reflect the possible disutility associated with claiming in-work tax credits ($P = 1$), and its presence allows us to rationalize less than complete take-up of tax credit programs. We denote $P^*(h) \in \{0, E(h; X, \varepsilon)\}$ as the optimal choice of program participation for given hours of work h, where $E(h; X, \varepsilon) = 1$ if the individual is eligible to receive tax credits at hours h, and zero otherwise. Assuming eligibility, it then follows that $P^*(h) = 1$ if and only if

$$u(c(h, P=1; T, X, \varepsilon), h, P=1; X, \varepsilon) \geq \\ u(c(h, P=0; T, X, \varepsilon), h, P=0; X, \varepsilon).$$

The choice of hours of work h affects consumption c through two main channels. Firstly, through its direct effect on labor market earnings and its interactions with the tax and transfer system; secondly, working mothers purchase childcare for their children which varies with maternal hours of employment.

Individuals face a budget constraint, determined by a fixed gross hourly wage rate (generated by a log–linear relationship of the form $\log w = X_w{}'\beta_w + \varepsilon_w$ and the tax and transfer system $T(wh, h, P; X)$. Non-labor income, such as child maintenance payments, enter the budget constraint through the dependence of the tax and transfer schedule T on demographic characteristics X. To arrive at a measure of consumption c, in the absence of saving, we subtract both childcare expenditure and fixed work-related costs, $\alpha_f(X, \varepsilon_f) \times 1\{h > 0\}$, from net income, $wh - T(wh, h, P; X)$.

The hours of childcare h_c is assumed to vary stochastically with hours of work and demographic characteristics. Total weekly childcare expenditure is then given by $p_c h_c$, with p_c denoting the hourly price of childcare. Empirically, this is modeled by assuming that p_c follows some distribution $p_c \sim F_c(\cdot; X_c)$ which varies with demographic characteristics.

Incorporating endogenous take-up of tax credits through cost $\eta(X, \varepsilon_\eta)$, it then follows that the optimal choice of hours $h^* \in H$ maximizes

$$U(c(h, P^*(h); T, X, \varepsilon), h, P^*(h); X, \varepsilon),$$

subject to the various constraints already detailed.

This brief outline of the key features of a structural model illustrates some of the key ingredients and assumptions required, and this for a single worker decision. Family labor supply models require further assumptions, in particular the modeling of joint hours and employment choices, see Blundell et al. (2000b) and Brewer et al. (2006). Blundell and MaCurdy (1999) and Meghir and Phillips (2010) give further insights into the structural modeling of labor supply.

Blundell and Shephard (2010) take this structural model specification to the UK Family Resources Survey data and argue that it does a good job of describing observed behavior. For example, the model is used to simulate the WFTC reform. This is then compared with the simulated average response with the quasi-experimental estimate described in Table 1.1. The simulated difference-in-differences parameter from the structural evaluation model is precise and does not differ significantly from the difference-in-differences estimate itself.

1.5.4 *Dynamics and Frictions?*

Finding that a structural model does a reasonable job of predicting the changes of reforms over time does not imply that it is the correct model, simply that it is not rejected. One area where we might expect the model to perform poorly is in capturing the dynamics of labor supply. In particular, experience effects and adjustment frictions. In their study of the hours and employment changes around the tax credit reforms in the UK, Blundell, Brewer and Francesconi (2008) already note that changes seem to take place over a relatively short time, within a year or two, suggesting relatively small adjustment costs. Of course, for the most part, these reforms occurred in a period of economic expansion and we would not expect our structural model to provide a good description of labor supply choices during a recession. Moreover, these reforms were large and well announced. Smaller tax reforms may have less impact, see Chetty et al. (2009).

But what about experience effects? That is the dynamic pay-off in terms of earnings and employment of being in the labor market. For some evidence on this for low-skill workers we turn to the Canadian SSP. Figures 1.15 and 1.16 from the previously described Canadian SSP control-trial experiment tell an interesting story.

Recall that the SSP program is complete after 36 months from eligibility, and that eligibility can take up to twelve months. By 50 months from baseline all the treatment group would have completed the program. Indeed we saw, in Figure 1.10, that the treated had, on average, a sizable increase in employment. But did this translate into higher wages or earnings after the program was finished? Not according to these figures. There is little noticeable difference after 50 months in hourly wages or earnings, see also Card and Hyslop (2005). It seems

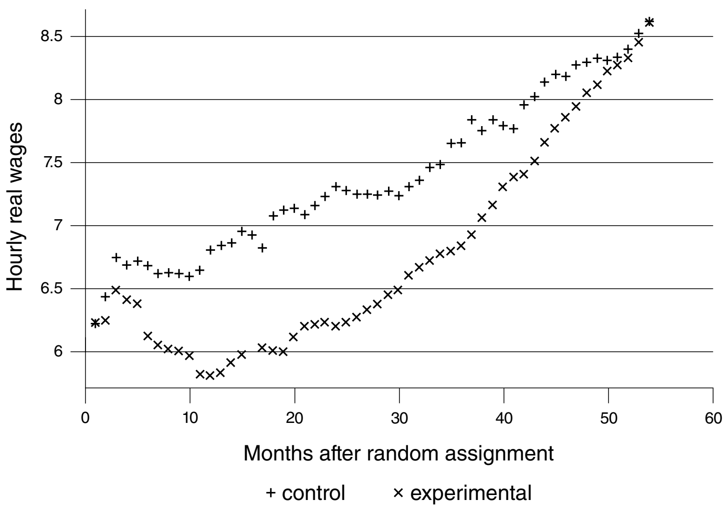


Figure 1.15. SSP, Hourly Wages by Months After RA

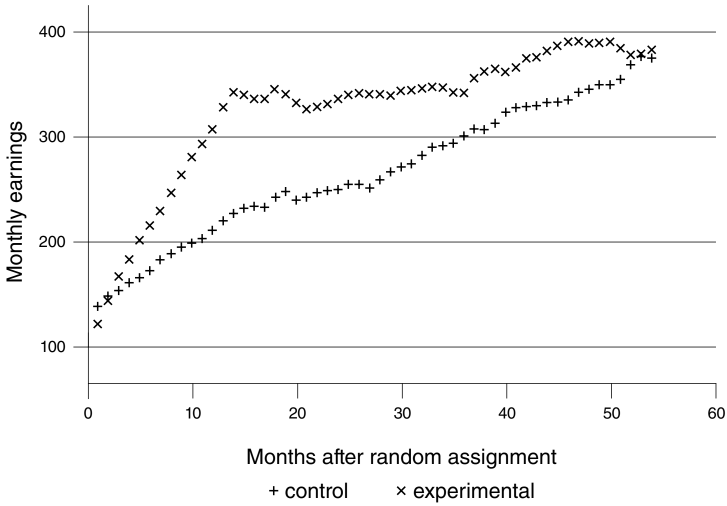


Figure 1.16. SSP, Monthly Earnings by Months After RA

that experience effects, for this group at least, are negligible. In general, they may well be small for low-skill workers, explaining why relatively simple models that ignore experience effects explain labor supply adjustments quite well for these types of workers.

1.6 Implications from Theory for Tax Design

We have argued earlier that there are three key ingredients to any optimal tax analysis: the accurate measurement of response elasticities, the detailed description of the distribution of income, and some view of social welfare weights. The first two of these are positive and can be learned from a careful evidence-based analysis. The last is normative and therefore something where opinions may differ. In this section we will see how these come into play in the design of earnings taxation.

An important distinction we took from our discussion of labor supply responses in Sections 1.2 and 1.5 is between the extensive and intensive margins of labor supply. Why is this also important for tax design? The optimal taxation literature explores the consequences for tax design (e.g. Diamond 1980; Saez 2002; Laroque 2005). Once individuals are allowed to respond to changes in the tax schedule by deciding whether or not to work, as well as how hard to work, then the optimal tax schedule can change dramatically. In particular, when this employment decision becomes relatively more important, optimal marginal tax rates can be lower (and perhaps even negative) for those with low-earnings capacity. As Brewer, Saez and Shephard (2010) note:

> a striking implication is that, if the government values redistribution then the participation tax rate should be negative for low earnings – in other words, low-income workers should receive an earnings subsidy. Hence, in sharp contrast to the intensive model, the extensive model implies that earnings subsidies or work-contingent credits (such as the earned income tax credit or the working tax credit) should be part of an optimal tax system.

This is one of the key lessons from recent optimal tax design. A *large* extensive elasticity can *turn around* the impact of declining social weights, implying a higher transfer to low-wage workers than those out of work and a role for earned income tax credits.

A further key consideration in tax design is the way in which responses differ across individuals of different characteristics. Unless there are good redistributive reasons to do otherwise, tax rates will generally be lower on those types of individuals with more elastic responses.

Finally, the degree of inequality and of income uncertainty will also matter for earnings tax design. The past three decades or more has seen strong growth in earnings inequalities and a change in the nature of earnings risks. The redistributive element of the earnings tax and benefit system acts, in part, as an insurance to earnings risks. As the nature of these risks changes and as underlying inequality grows, the balance between inequality and work incentives gets harder to balance. Designing an efficient structure to the earnings tax and benefit system such that it achieves the desirable distributional objectives and becomes ever more salient.

1.6.1 *Optimal Design for Low-Income Workers*

How should we think about an optimal design? We will assume that the government seeks to maximize *social* welfare subject to revenue constraints. Following on from our running illustration of tax design for low-income single mothers, assume we want to redistribute '£R' to this group. What is the *optimal* way to do this?

Our aim will be to recover optimal tax/credit schedule in terms of earnings. There are two related approaches. The first to use the Diamond-Saez approximation in terms of extensive and intensive elasticities at different earnings. The second approach involves a *complete* Mirrlees optimal tax computation requiring a complete specification of choices and constraints.

The first approach is exemplified in the work of Saez (2002). He provides an intuitive expression for the tax rate schedule in terms of the extensive and intensive elasticities, and simple summary measures of the distribution of earnings and the social welfare weights. The formula is only approximate and assumes away income effects, however as a guide to the setting of tax rates it is extremely informative, see Immervoll et al. (2007). We return to these optimal tax formulas in our discussion of the optimal top rate of taxes in what follows.

The latter approach might be labeled the *structural microeconometric approach* to tax design – effectively, a stochastic mechanism design problem. In this case the optimal tax model *is* the labor supply model. Consequently, all of the assumptions concerning behavior are also required for this analysis. The distribution of earnings, fixed costs of work, childcare, demographic differences, and unobserved heterogeneity described in the previous section all influence choice of tax rate schedule.

Suppose we assume earnings (and certain characteristics) are all that is observable to the tax authority.[7] Social welfare W is represented by the sum of transformed utilities:

$$W = \int_{X,\in}\int_{\varepsilon} \Upsilon(U(c(h^*;\, T,\, X,\, \varepsilon),\, h^*;\, X,\, \varepsilon))dF(\varepsilon)dG(X,\, \varepsilon),\, \varepsilon$$

where function Υ captures redistributive preferences.[8]

The government seeks to maximize W subject to revenue constraint:

$$\int_{X,\in}\int_{\varepsilon} T(wh^*;\, h^*;\, X)dF(\varepsilon)dG(X,\, \varepsilon) \geq \overline{T}\,(\equiv -R).$$

Blundell and Shephard (2010) control the preference for equality in social welfare by the transformation function

$$\Upsilon(U\,|\,\theta) = \tfrac{1}{\theta} \times ((\exp U)^{\theta} - 1) \tag{9}$$

where $\theta < 0$ favors equality of utilities.

The objective is to find robust tax rate schedules for fairly general social welfare weights. Given the structural parameter estimates, we can solve for optimal schedules. In their application Blundell and Shephard restrict to piecewise linear tax schedule (out-of-work income, nine marginal rates at breaks of £50 up to £400), with possible hours contingent payments.

The key findings of the Blundell–Shephard analysis, under range of values for θ (i.e. allowing different degrees of preference for inequality), are that optimal marginal rates are broadly increasing in earnings for all groups. The results also point to a shift of out-of-work support towards families with younger children. This suggests an optimal tax schedule with tagging according to the age of children. Moreover, pure *tax credits* are found to be optimal at low earnings but only for those with school-aged children.

The analysis also found that hours contingent payments can improve design: if hours are accurately observed a full-time bonus is desirable for low-wage mothers with older children. But measurement error and the possibility of hours manipulation are found to weaken the argument for hours rules.

In the *Review* recommendations we not only stress reforms for lone parents, but also for married parents, and older workers pre-retirement. From the examination of response elasticities we recommended 'tagging' tax rates by age of (youngest) child for mothers/parents and also at pre-retirement ages.

1.6.2 *Tax Rates at the Top*

At the top of the income distribution we take a rather different approach. Hours and employment may not be the only, nor the most important, ways to change earnings in response to tax changes. When it comes to the taxation of top incomes, concerns about the tax base come back into play. Feldstein (1995; 1999) makes a convincing case for looking directly at taxable income. The more opportunities for exemptions and deductions and the possibility to pass income through other lower tax jurisdictions, the more difficult it is to raise revenue from the top-income earners. Consequently, we require a more general elasticity measure that captures these other avenues for response. The taxable income elasticity does just that.

A higher tax rate on a smaller base will raise less revenue and will probably be harder to sustain. To quote Slemrod and Kopczuk (2002):

> When personal tax rates on ordinary income rise, evasion may increase, businesses may shift to corporate form, there may be a rise in the consumption of deductible activities such as charitable giving, and individuals may rearrange their portfolios and compensation packages to receive more income as tax-preferred capital gains. These responses to higher taxes, and all others, will show up in declines in taxable income, and there is a growing body of evidence, that, at least for high-income individuals, the elasticity of taxable income to the marginal tax rate is substantial.

It is hardly surprising therefore to find that the responsiveness of taxable income to the tax rate is a key parameter on which the setting of top tax rates depends. What we have to bear in mind is that the responsiveness itself will be affected by the tax base. This elasticity can be expected to be larger the narrower the tax base. Given the need to capture all these margins and the fact that effort is very hard to quantify, the behavioral effect will require a different kind of measurement from that used to gauge hours and employment responses.

Consider an *optimal* top tax rate and suppose the welfare weight on top bracket incomes is negligible. The optimal rate will be the revenue maximizing rate – the Laffer rate. We first note that the top of the taxable income distribution is well approximated by a Pareto distribution. Figure 1.17 shows this to be the case. Suppose we write e as the taxable income elasticity, then Brewer, Saez and Shephard (2010) show that the revenue maximizing rate is given by

$$t = 1/(1 + a \cdot e),$$

where a is the Pareto parameter. For the UK, a is approximately 1.67, see Brewer and Browne (2009).

To estimate e reliably is fraught with difficulties. Typically a difference-in-differences methodology is used, see Gruber and Saez (2002). When this approach is applied to past changes in tax rates among the top 1% in the UK, using the 2%–5% group as a control, the evidence to the *Mirrlees Review* suggested a preferred estimate of e of 0.46 with a standard error of 0.13.

Exploring various formulations of the difference-in-differences specification for the UK, the estimate of e remains in the 0.35–0.55 range with a central value around 0.45, but is clearly quite fragile. An

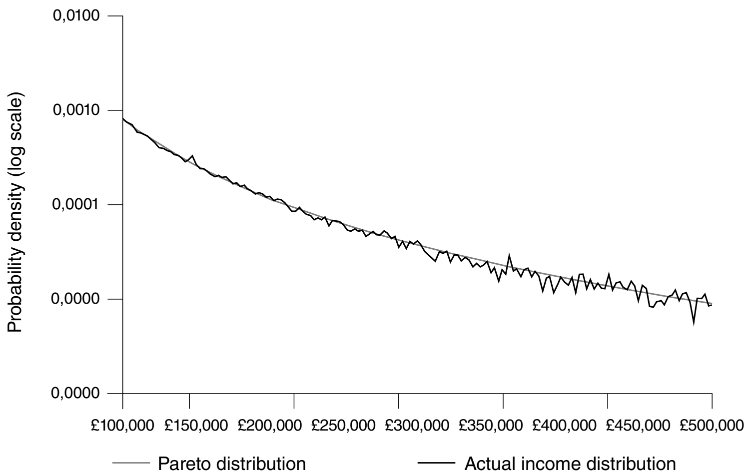


Figure 1.17. The Pareto Distribution and the Taxable Income Distribution at the Top

estimate of this magnitude would suggest the optimal top 1% bracket rate of around 57%, close to the current top rate (including other taxes on earned income). This analysis suggests little room for any further raising of the top rate of income tax in the UK without changes to the tax base for earned income itself.

1.7 Work Incentives, Redistribution and Base-Broadening Reforms

Earnings taxation plays a major role in getting the structure of tax design right across the whole tax system. For example, a major recommendation from the *Review* is a broadening of the VAT tax base. This is especially important in the UK which has one of the narrowest bases for VAT and has also just raised the standard rate to 20%. With many goods (food at home, children's clothing, financial services) being zero rated or exempt, the potential welfare cost created by differential commodity taxation is large. Indeed, empirical evidence suggests that current indirect tax rates do not line up with any reasonable justification. But broadening the base of VAT is not trivial. On its own it would be regressive and would harm work incentives. To reap welfare gains from base-broadening, distributional concerns have to be addressed and reduced work incentives redressed. Both of these require the careful redesign of earnings taxation. Earnings taxation becomes an integral part of the VAT reform.

Three key empirical observations about consumer behavior come into play when thinking about justifying differential rates of VAT. First, some commodities are luxuries and some are necessities. Differential commodity taxation can act as a redistributive mechanism. But they are an inefficient way of delivering redistribution given the other direct (earnings) tax instruments available. Secondly, nonseparabilities with labor supply are important. Certainly this is a key justification in the optimal tax literature for differential commodity taxes with goods that are complementary to work bearing a heavier tax. Empirically these relate mainly to childcare and work-related expenditures.

Current VAT rates are quite different from these, see Crawford, Keen and Smith (2010) for example. Finally, price elasticities differ with total expenditure and wealth. That is responses, and therefore welfare costs, differ across the income distribution. To value welfare losses and calculate compensation these microeconomic differences matter a lot.

The welfare gain from broadening the VAT is based on assumptions that are unlikely to hold in reality: weak separability between goods and leisure, common preferences, and competitive pricing by suppliers. Nonetheless, a broad uniform base seems likely to be a good baseline from which to judge reform. Non-separabilities are clearly evident but mainly in relation to work-related expenses. Preferences are heterogeneous but often differ by characteristics that are in the tax system, like family composition and family income. Differentials in tax rates across commodities seem to be more motivated by redistributive concerns and the power of certain pressure groups: with food and children's clothing belonging to the first and financial services to the second.

On its own, the base broadening of VAT in the UK would be regressive and weaken work incentives. Can a practical package avoid this? *Tax by Design* provides an illustration of how this can be done, implementing a reform package that achieves compensation while also avoiding significant damage to work incentives. Working with the existing set of UK direct tax and benefit instruments, the *Review* simulated removing almost all zero and reduced rates in the UK. This raises £24bn (with a 17.5% VAT rate) if no behavioral response. With responses, the empirical results suggest (in principle) every household could be compensated and still leave a £3bn to £5bn welfare gain.

Summary results from the *Review* show the key interaction between earnings taxation and the base-broadening reform. These are summarized in Figures 1.18, 1.19 and 1.20. Turning first to distributional

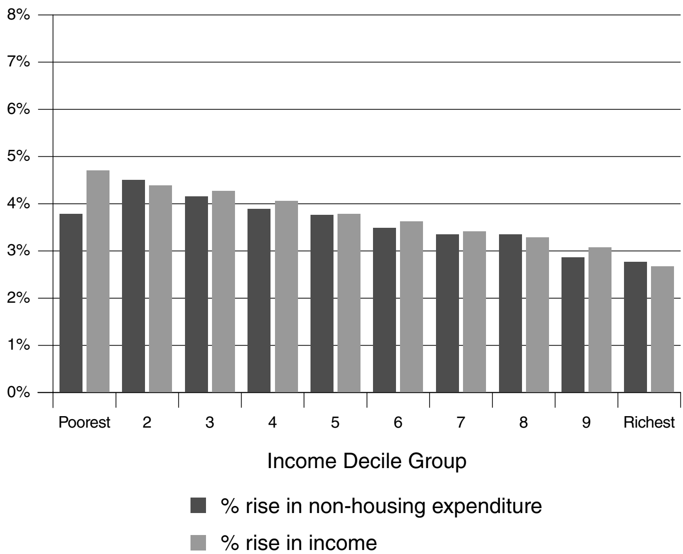


Figure 1.18. VAT Reform, Effects by Income Decile

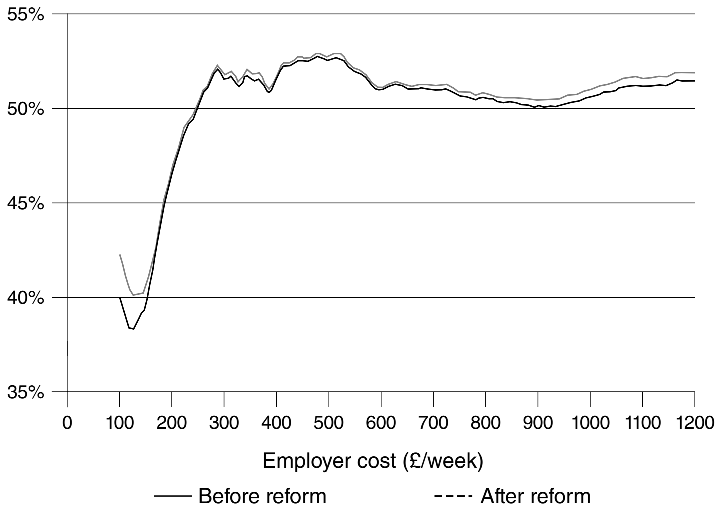


Figure 1.19. VAT Reform, Impact on Participation Tax Rates

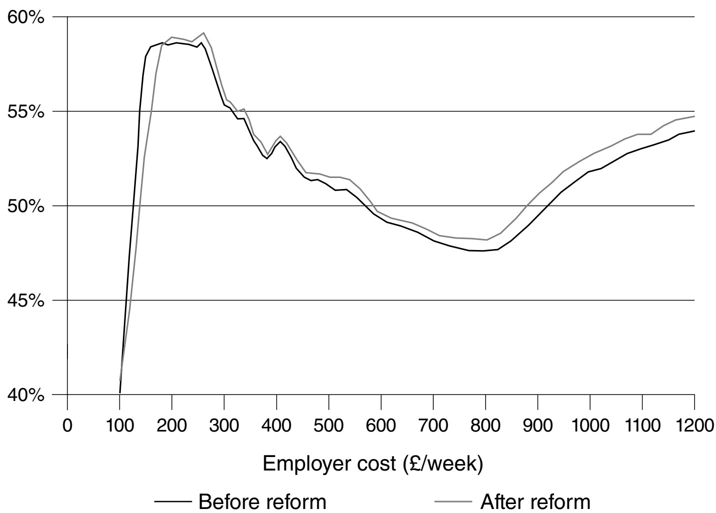


Figure 1.20. VAT Reform, Impact on Effective Marginal Tax Rates

concerns, Figure 1.18 shows that average percentage rises in the cost of living by income decile are more than compensated for by increases in income. But what about work incentives? Figures 1.19 and 1.20 show that effective tax rates on earnings are also left largely unchanged at both the extensive and intensive margin.

Other base-broadening reforms also require care in their interactions with earnings taxation. For example, in the discussion of capital taxation, the *Review* suggests moving towards an expenditure tax treatment of saving by providing an allowance for income saved which effectively eliminates the tax on the normal return to saving. This provides a framework for the integration of capital income taxation with corporate taxation. Exempting the normal rate also gives neutrality between debt and equity capital gains and dividends are treated in the same way and the resulting capital tax structure overcomes the *lock-in* incentive from Capital Gains Tax. However, this also requires alignment between tax rates on corporate income, shareholder income, and labor income. With progressive tax rates on labor income, progressive rates are also required on shareholder income to avoid differential tax treatments of incorporated and unincorporated firms – a lower progressive rate structure on shareholder income than on labor income reflecting the corporate tax already paid.

1.8 Summary and Conclusions

In this paper I have used the *Mirrlees Review* as an illustration of the use of evidence in the development of a tax reform program. In developing the recommendations for the reform, the *Review* attempted to draw on empirical evidence wherever possible.

The aim here was to show how to make the best use of all available evidence, from broadly descriptive evidence to that gleaned from quasi/experimental evaluations and also from structural model estimation.

Much of the discussion has focused on earnings tax design, partly for illustrative reasons but also because there is a large body of evidence on labor supply and taxable income responses to tax/benefit reform and to policy design. Wherever possible we have argued for the use of well specified and carefully validated structural microeconometric models as the basis for design. In cases where we have less in the way of detailed structural models to draw on, we have to use more general information on likely size response elasticities. In this we follow closely the lead in Saez (2002) and Brewer, Saez and Shephard (2010).

Across the board, we have documented a key role for labor supply responses at the extensive and intensive margins. Both matter but differ by gender, age, education, and family composition. We found that labor supply responses for families with children vary by age of the youngest child. We also found different responses for older workers in *pre-retirement* years.

The results of our analysis suggested changing the rate structure to match lessons from *new* optimal tax analysis. It pointed to lower marginal rates at the bottom of the earnings distribution. Means-testing should be less aggressive, at least for some key groups. Tax credits should be better targeted to lower incomes and to the families where labor supply is most responsive. We particularly stress reforms for lone parents, married parents, and older workers pre-retirement. From the examination of response elasticities we recommended 'tagging' tax rates by age of (youngest) child for mothers/parents and also at pre-retirement ages.

When put together, this reform agenda can be interpreted through a lifetime view of taxation, implying a *life cycle* rearrangement of tax incentives and welfare payments to match elasticities and early years investments, effectively redistributing across the life cycle, distinguishing by age of (youngest) child for mothers/parents and at pre-retirement ages. The simulation results in *Tax by Design* suggested significant employment and earnings increases from such a reform package. The evidence on taxable income elasticities implies limits to tax rises at the top of the income distribution. Tax reforms for this group are better directed towards base broadening to address tax avoidance and revenue shifting.

Finally, we noted how earnings taxation is also called on to undo the impact on distributional and work incentives of the rest of any tax reform package.

2

Labour Supply and the Extensive Margin

Richard Blundell, Antoine Bozio and Guy Laroque

Forty years ago the Europeans (here French and British) used to work more than the Americans. They now work less. The aim of this paper is to provide a coherent picture of these changes. To do so we split the overall level of work activity into the number of individuals in work and the intensity of work supplied by those in work. This reflects the distinction between whether to work and how much to work at the individual level and is referred to, respectively, as the extensive and intensive margin of labour supply.

The difference between the extensive and intensive margins has been highlighted in recent research attempting to resolve differences between micro and macro responses of labour supply to tax reform. For example, Rogerson and Wallenius (2009), following the work of Prescott (2004), argue that the responsiveness of the extensive margin of labour supply to taxation plays a major role in explaining aggregate differences in total hours worked across countries. They show that an economy with fixed technology costs for firms and an inverted U-shape life cycle productivity for workers can produce large aggregate extensive labour supply responses driven by movements in employment at either end of the working life. This, they argue, can reconcile the

The original version of this chapter was published as: Blundell, R./Bozio, A./Laroque, G. (2011). Labor Supply and the Extensive Margin, in: American Economic Review, 101(3): 482–86. *Blundell: University College London (UCL) and Institute for Fiscal Studies (IFS), 7 Ridgmount Street WC1E 7AE London United Kingdom, r.blundell@ucl.ac.uk. Bozio: UCL and IFS, Antoine_b@ifs.org.uk. Laroque: UCL, IFS and CREST, g.laroque@ucl.ac.uk.* *We thank Steve Davis and participants at the AEA session for comments, and the Data archive UK, INSEE and IUPMS for data access.*

small micro-based elasticities of hours worked with the large responses required if taxes and social security are to explain cross-country differences in total hours of work.

The distinction between the extensive and intensive margin has long been recognised in microeconometric studies of labour supply, especially for women with children (Heckman 1974; Blundell and MaCurdy 1999), and in studies of older workers (Gruber and Wise 1999). The relative size of labour supply responses at the intensive and extensive margin has also been a key parameter in the public economics literature on earnings tax design (Saez 2002; Laroque 2005, Brewer, Saez and Shephard 2010).

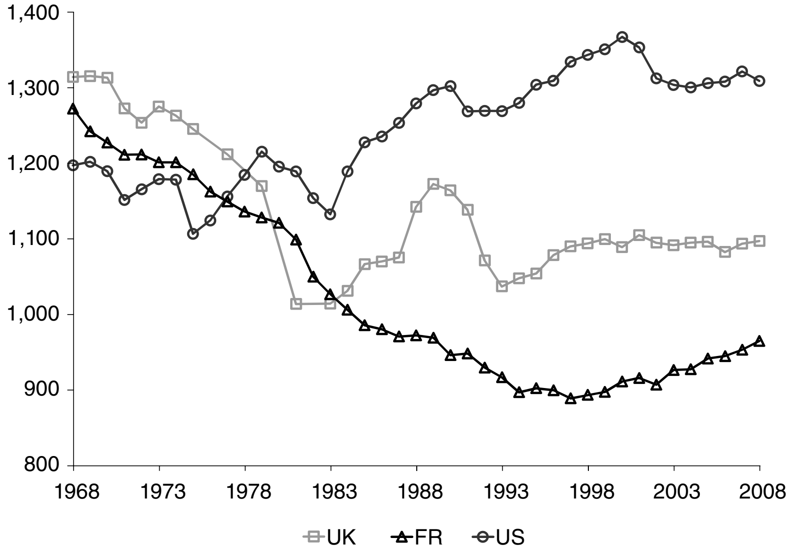
But what do we know about the importance of these margins for different types of workers? How well does the extensive margin explain changes in aggregate hours? In this paper we provide a decomposition of the evolution of aggregate hours of work into changes at the extensive and intensive margin. More details are available in the companion paper (Blundell, Bozio and Laroque 2011b), where we also develop a life cycle model that delivers an aggregate total hours elasticity in terms of the distribution of intensive and extensive elasticities at the micro level.

We examine three key countries, the US, the UK and France, over the forty year period up to 2008. These three countries stand respectively at the top, middle and bottom of Prescott's table of labour supply. The UK provides an interesting comparison with the polar cases of France and the US. They are also countries where we can access nationally representative detailed microdata over a long period of time (see supplementary material) so as to examine the relationship between the extensive and intensive margin across different individual types.

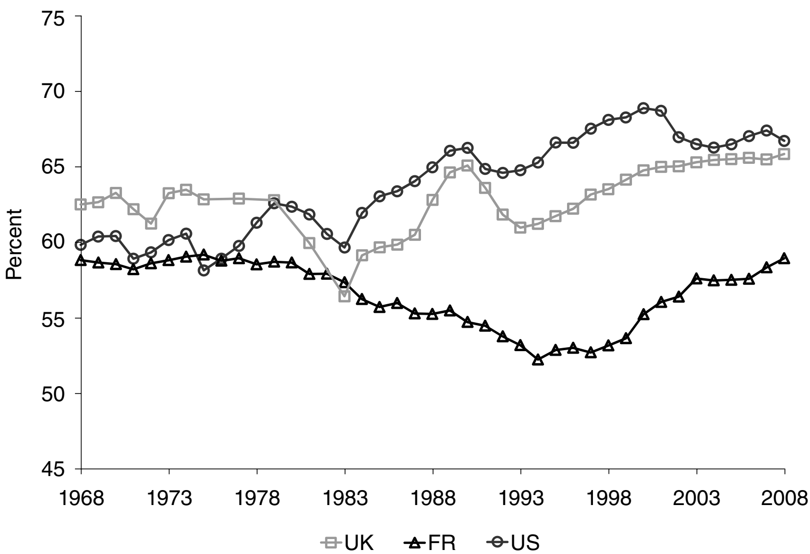
Figure 2.1.a highlights the key piece of evidence used to motivate our analysis. It charts the evolution of the average annual hours of work per individuals aged 16 to 74 from 1968 to 2008. The pattern of total hours per individual shows evidence of a three way split after 1980 in the evolution of total hours across the three countries, hiding different evolutions of employment and hours per worker.

Overall employment rates in the UK and the US have moved somewhat in line with each other showing an increase over this period. Employment rates in France have progressed very differently. Figure 2.1.b shows a strong decline in employment in France until the mid-1990s with recovery thereafter but leaving a large difference in current employment rates.

Changes in hours per worker tell a different story. Figure 2.1.c shows the UK and France following each other with strong declines over



a) Mean Annual Hours Per Individual



b) Employment Rate (Per Population)

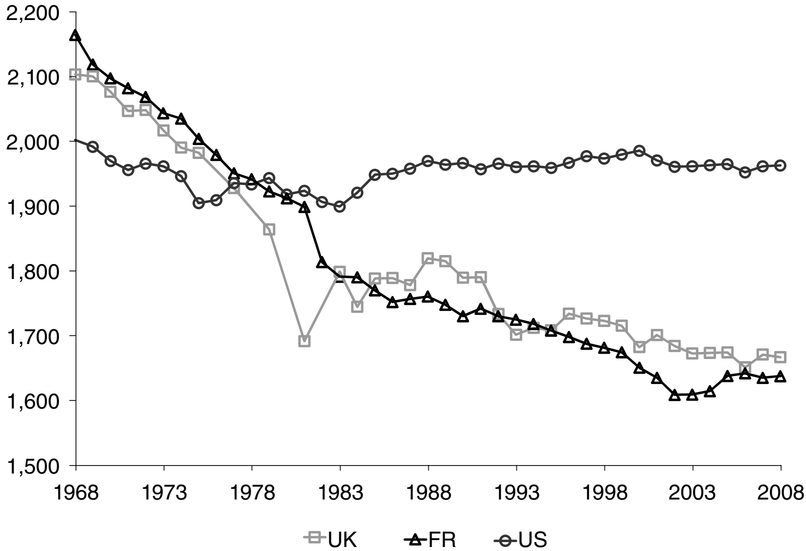


Figure 2.1. Measures of Market Work. c) Mean Annual Hours Per Worker
Source: Enquete Emploi, Labour Force Survey, Current Population Survey.

this period stabilizing somewhat in the 2000s. In contrast, the US has retained a stable pattern of hours per worker over the entire period apart from a dip in the late 1970s and early 1980s.

2.1 Bounding Changes at the Extensive and Intensive Margins

Our central interest is how the overall average hours worked per person in any year t, H_t, varies over time and across countries, in particular along the intensive and extensive margins. Of course, average hours worked differs across different people, by age and gender for instance. Suppose there are $j = 1, \ldots, J$ such broad categories. H_t is computed as the sum of average hours per person for each category, H_{jt}, weighted by their population shares q_{jt}. We decompose H_{jt} as the product of hours per worker h_{jt} and participation in the labour market p_{jt}.

When we observe a change in yearly hours worked per person, $H_t - H_{t-1}$, we would like to be able to know how much of the change is due to the intensive or extensive margins. To achieve this we propose a statistical decomposition: First we measure the change due to the behavior of category *j*, holding the population structure constant as in date $t-1$, as in a Laspeyres index

$$\varDelta_{jt} = q_{j,t-1}\left[H_j - H_{j,t-1}\right]. \tag{1}$$

The total change across all *J* categories of workers is simply

$$\varDelta_t = \sum_{j=1}^{J} \varDelta_{jt} \tag{2}$$

and we have, by construction

$$H_t - H_{t-1} = S_t + \Delta_t \tag{3}$$

where S_t is a structural effect due to the change in the composition of the population given by $\sum_{j=1}^{J} H_{jt}\left[q_{jt} - q_{jt-1}\right]$.

There is no obvious way to decompose the change in total hours experienced by category *j* into the sum of an extensive E_j and an intensive I_j components. It is however natural to suppose that any plausible measure I_j of the intensive margin would have the same sign as the difference of the hours worked per worker at date $t-1$ and t: $\Delta_{hj} = h\_jt - h_{j,t-1}$. Assuming linearity, we can then express the change Δ_j as the sum of an intensive component $I_j = p_{Ij}\Delta_{hj}$ and an extensive component $E_j = h_{Ej}\Delta_{pj}$. Supposing the fraction p_{Ij} is in the interval $[p_{j,t-1}, p_{jt}]$, we get the intensive bounds

$$I_j \in \left[p_{j,t-1}(h_{jt} - h_{j,t-1}), p_{j,t}(h_{jt} - h_{j,t-1})\right].$$

From the identity $\Delta_{jt} = I_j + E_j$, the extensive bounds are given by

$$E_j \in \left[h_{j,t-1}(p_{jt} - p_{j,t-1}), h_{j,t}(p_{jt} - p_{j,t-1})\right].$$

At the limits, the change in total hours for any category of workers satisfies two polar exact statistical decompositions:

$$\varDelta_{jt} = q_{j,t-1}\left\{\left[h_{jt} - h_{jt-1}\right]p_{jt} + \left[p_{jt} - p_{jt-1}\right]h_{jt-1}\right\}, \tag{4}$$

$$\varDelta_{jt} = q_{j,t-1}\left\{\left[h_j - h_{jt-1}\right]p_{jt-1} + \left[p_{jt} - p_{jt-1}\right]h_{jt}\right\}. \tag{5}$$

The first term on the right-hand side of both expressions is the intensive margin, weighted in the top formula (4) with the final participation rate (as in a Paasche index) and in the bottom formula (5) with the initial participation rate (as in a Laspeyres index). The second term is the extensive margin (Laspeyres in (4), Paasche in (5)).

2.2 Decomposing Total Hours Worked

In this section we turn to the empirical analysis and first examine the evolution of h_{jt} and p_{jt} for different age and gender groups. We then

Table 2.1. The Evolution of Hours of Work Between 1977 and 2007 by Gender and Age Groups

	Year	Youth Men	(16–29) Women	Prime age Men	(30–54) Women	Old Men	(55–74) Women	Residual	All (16–74)
FR	1977	1,402	871	2,010	951	827	367		1,148
	2007	858	627	1,639	1,116	508	344		953
	Δ	−82	−38	−82	36	−36	−3	10	−195
UK	1977	1,707	938	2,117	873	1,107	323		1,212
	2007	1,219	876	1,786	1,055	790	385		1,094
	Δ	−71	−9	−70	39	−42	10	25	−118
US	1977	1,344	835	2,018	947	1,025	447		1,156
	2007	1,236	956	1,922	1,373	1,084	754		1,321
	Δ	−19	22	−19	90	6	38	46	165

Note: Δ are computed following equation (1).

use (4) and (5) to provide bounds on the importance of intensive and extensive margins in the evolution of hours worked across these various groups. We focus on the comparison between 1977 and 2007, a period for which we are more certain of the reliability of our data.

The lines Δ of Table 2.1 show the contributions of the categories and the effect of demographic structure, according to equations (1), (2) and (3). Two clear features emerge: First, for nearly all the categories the US and France are at the extremes with the UK in between. The contribution to the aggregate of the hours worked by the young and prime age men is negative in all countries. The decline in absolute value is smallest in the US, and increases when one goes from the US to the UK, and then to France. Second, the increased participation of women works against the general trend. This is particularly obvious for women aged 30–54, who all work more in 2007 than in 1977, but also for the old and young.

Using the statistical bounds framework developed in the previous section we can go further and, in Table 2.2, we examine what part of any overall change in hours is attributable to changes at the extensive or intensive margin for any particular subgroup of the population. The row [*I–L, I–P*] shows the bounds on the intensive margin, (*L* standing for Laspeyres, *P* for Paasche), while the row [*E–L, E–P*] shows the bounds on the extensive margin.

As a concrete example, examine the first entry in the top left of Table 2.2, French men aged 16–29. The impact on total hours for this group is −82. The *I–L* index of −37 tells us that the intensive margin does a good bit but not the majority of the work in explaining total hours changes for this group. The *E–L* estimate of −54 confirms the relative importance of the extensive margin for this group.

Table 2.2. The Extensive and Intensive Margins between 1977 and 2007

	Year	Youth (16–29)		Prime age (30–54)		Old (55–74)		All
		Men	Women	Men	Women	Men	Women	(16–74)
FR	Δ	−82	−38	−82	36	−36	−3	−195
	[I-L, I-P]	[−37,−28]	[−23,−19]	[−59,−56]	[−35,−49]	[−11,−8]	[−9,−10]	[−185,−183]
	[E-L, E-P]	[−54,−45]	[−19,−16]	[−27,−23]	[85,71]	[−28,−25]	[7,6]	[−12,−10]
UK	Δ	−71	−9	−70	39	−42	10	−118
	[I-L, I-P]	[−42,−36]	[−23,−26]	[−48,−45]	[−2,−3]	[−22,−19]	[−6,−8]	[−161,−167]
	[E-L, E-P]	[−35,−29]	[17,14]	[−25,−22]	[41,41]	[−23,−20]	[17,15]	[50,43]
US	Δ	−19	22	−19	90	6	38	165
	[I-L, I-P]	[−6,−6]	[1,1]	[−5,−5]	[14,19]	[3,3]	[3,5]	[15,17]
	[E-L, E-P]	[−13,−13]	[21,21]	[−14,−14]	[72,77]	[3,3]	[33,35]	[148,150]

Note: *I–P* designs the Paasche measure of the intensive margin, *I–L* the Laspeyre measure, respectively, *E–P* and *E–L* for the extensive margin, as described by equations (4) and (5).

Turning first to prime-age workers, the steep decline at the *intensive* margin for prime-aged men in France and the UK relative to the US is striking. For this group the bounds are quite narrow and leave little room for ambiguity. These changes represent an enormous shift in the relative position of these countries. Increases in effective marginal tax rates and/or the regulation of working hours could explain these patterns. In terms of regulation Britain has seen fewer changes than France and yet has experienced similar changes.

Income effects could be part of the explanation. There are two potential sources for these. First, as the economy grows individuals may prefer to take some of the gains in income in terms of increased leisure, cutting back their hours of work. However, given overall growth has been somewhat similar across all three countries, it would have to be that Europeans take more leisure in response to rises in income. A second source of income effect for prime age men is the increased participation by women. This is often termed the added-worker effect. Prime-aged women have certainly seen a strong increase in participation. Indeed, the bounds on the extensive margin changes in Table 2.2 for women aged 30–54 are the largest positive change to be found in any country-age cell and at any margin. But the largest overall increase, especially when the intensive margin is taken into account, is for US women. Yet the change in hours is the least for US men. Again responses would have to be different in Europe.

Table 2.2 tells us that the *extensive* margin for prime-age men in Britain and in France also falls more than in the US, although there are declines in the US too. Increases in relative employment costs or out-of-work benefits in France and Britain could explain such changes.

Income effects may also play a role at the extensive margin as individuals cut back on their overall life cycle labour supply. However, this seems more likely at either end of the life cycle.

As we have noted, for prime age women it is the increase at the extensive margin that is so extraordinary, especially in the US and in France where the bounds in Table 2.2 suggest a very similar change and one that is nearly twice the size of that experienced in the UK. Intensive margins provide somewhat of a puzzle here, falling back strongly in France, staying put in the UK while growing in the US.

For older men and women there is a large decrease in hours per worker in France, similar in the UK, contrasting with an increase in the US. There are falls at the extensive and intensive margins for UK men but increases at the extensive margin for UK women. This surely is linked to the strong increase in women's participation. This phenomenon is replicated to some extent across all countries and offsets the stronger incentives to retire earlier in the UK and in France. The contrast with the US is stark, where at all margins and for both genders the bounds point to positive changes for older workers.

The changes among the young are sizable and predominantly negative. In France and the UK there are large falls for young men at both the extensive and intensive margin.

2.3 Conclusions

In this paper we have proposed a systematic way of examining the importance of the extensive and the intensive margins of labour supply in explaining the overall movements in total hours of work over time. We have shown how informative bounds can be developed on each of these margins. We have applied this analysis to the evolution of hours of work in the US, the UK and France. We have shown that both the extensive and intensive margins matter in explaining changes in total hours.

An objective of this research is to link up the changes at the extensive and intensive margins to movements in the distribution of taxes, relative wages, demographics and other incomes. This will allow us to draw implications for the aggregate hours elasticity. Davis and Henrekson (2004) note the importance of household production in interpreting these effects. In Blundell, Bozio and Laroque (2011b), we develop a life cycle aggregation framework and apply it to the UK using a consistent series on marginal taxes, incomes, hours of work, wages and consumption for a representative sample of households. We focus on individuals aged 30 to 54. Following Blundell, Duncan and Meghir (1998), we use the large changes in relative growth of after-tax wages and other incomes across different education, age and gender groups over the years 1978, 1987, 1997 and 2007, to identify Marshallian and Hicksian elasticities. Frisch elasticities can also be estimated using consumption data.

In line with previous empirical studies we find that elasticities for women at both margins are larger than those for men. But we also note that the key determinant of these differences across gender is the age composition of children in the family. For this subpopulation, the median Marshallian extensive elasticity for women is .34, for men is .25, and the distribution has a large spread. The corresponding intensive elasticity ranges between .09 and .23. Using the empirical distribution of the wages and estimated unobserved heterogeneity, we find the aggregate total hours elasticity lies in the range .3 to .44.

Part III

Conceptual Contributions to Labor Supply Modelling

Introduction

The research in this part of the volume addresses some of the key aspects of labor supply modelling. It develops approaches to modelling family labor supply decisions and the interactions of these decisions with consumption decisions in families. It provides a framework for placing these decisions in a life cycle setting. It extends the labor supply model to allow for unemployment and studies the process by which individuals actually make their hours of work adjustments. At all times an eye is kept on casting the empirical analysis in a robust setting, relying on only the minimum assumptions required to provide an economic interpretation of the parameter estimates.

The first chapter "Modelling the Joint Determination of Household Labor Supplies and Commodity Demands," recognizes that labor supply choices are not made in isolation. They are taken along with other labor supply and consumption choices in the family. Perhaps most importantly these decisions interact in a relatively complicated way with some being complementary to each other while others are substitutes. Blundell, Chiappori and Meghir (2005) and Blundell, Chiappori, Magnac and Meghir (2007) extend these family labor supply models to the collective framework with non-participation and with children. Using a sample of households from the UK Family Expenditure Survey this is shown to differ across households with different family composition – most notably the number and ages of children.

Children follow a life cycle pattern, and this naturally leads us to ask whether we can easily place these family labor supply and consumption choices in a life cycle setting. In the second chapter "A Life Cycle Consistent Empirical Model of Family Labor Supply using Cross-Section Data," this is considered formally. Consumption data is shown to allow the standard labor supply model to be framed in a life cycle consistent setting. Consumption is a sufficient statistic for expectations, assets and credit constraints. Consequently the use of consumption data together with employment and hours of work data permit us to study within period choices under quite general assumptions about life cycle behavior.

Of course, this generalization to a life cycle setting does not come for free. Intertemporal weak separability is required. This life cycle consistent framework therefore rules out habits and labor market experience dynamics. Nonetheless, the Marshallian elasticities derived are a key part of any tax policy analysis and allow us to study quite complicated interactions between consumption and labor supply behavior. It does allow us to study non-separability between consumption and family labor supply. These, in turn, provide a key input into the discussion of differential rates of commodity taxation. Moreover, Blundell, Meghir and Neves (1993) and Blundell, Browning and Meghir (1994) show how to extend this analysis to fully study life cycle decisions.

One key assumption in these, and most other, labor supply models is a lack of frictions. That is, observed employment reflects a choice by each individual given their wages, their assets and their socio-demographic characteristics. In the chapter "Unemployment and Female Labor Supply" the life cycle consistent model of labor supply model is developed so as to formally incorporate unemployment. Here the decision to work not only results from a comparison of the reservation wage to the market wage but also to the probability of being offered that wage.

Much of empirical labor supply analysis rests on strong functional form and exogeneity assumptions. This is largely unwarranted and leaves estimated elasticities, and the tax policy reform analysis based on them, subject to legitimate concerns about fragility. The chapter "Estimating Labor Supply Responses using Tax Reforms," develops a robust grouping estimator that allows the estimation of life cycle consistent labor supply models with theoretically interpretable parameters to be estimated in a robust way. It uses the differential growth in wages for different birth-cohorts and education groups of women to carefully pin down hours of work elasticities for married women, controlling for the endogeneity of wages, other income and selection into work. The grouping estimator controls for the endogeneity of selection into work, endogeneity of marginal (after tax) wages, and other income (and saving). It is implemented as a control function approach, developing on from the ideas in Blundell and Smith (1986; 1994).

These labor supply models estimate how much hours and employment adjust to wage and income incentives. In final chapter the process of adjustment is studied more precisely. How fast do adjustments take place and do they take place in the same job or do they require job changes? Entitled 'Job changes, hours changes and the path of labor supply adjustment' it constructs a robust grouping method applied to panel data to uncover female labor supply responses to a sequence of tax credit reforms in the UK. The period studied, the late 1990s and early 2000s, is a buoyant period in the labor market. It is found that adjustment of hours is fairly quick, within the year. Even so these adjustments in hours largely involve changing jobs. This suggests that although different workplaces organize labor supply with similar hours arrangements, differences in these arrangements across workplaces allowed the women subject to tax credit reforms to significantly change their hours in response to incentives.

3

Modelling the Joint Determination of Household Labor Supplies and Commodity Demands

Richard Blundell and Ian Walker

In addition to choosing the allocation of total expenditure between commodities, households may also be able to make decisions over the allocation of their time between market work and leisure. In both theoretical and empirical work it has often been the case that these decisions have been analyzed separately. In this paper we stress the theoretical attractions of considering the *joint* determination of the allocation of time between work and leisure and the allocation of total expenditure between commodities in a utility maximizing framework. Using a sample of individual households we attempt to evaluate the empirical importance of the joint determination model over the separate determination of labor supplies and commodity demands. Our approach follows that of Abbott and Ashenfelter (1976), Phlips (1978), Barnett (1979), Deaton and Muellbauer (1980), and Atkinson and Stern (1980). Here we pay particular attention to the following four important aspects of household decision-making over commodity demand and labor supply.

The original version of this chapter was published as: Blundell, R./Walker, I. (1982). Modelling the Joint Determination of Household Labour Supplies and Commodity Demands, in: The Economic Journal, 92(366): 351–64. © 1982 by the Royal Economic Society. This work was partly supported by SSRC research grant HR 7456 for a project entitled 'Equivalence Scales in Models of the Demand for Goods and Leisure'. We are grateful to the SSRC Survey Archive at Essex University and the Department of Employment for making the household data from the Family Expenditure Survey available to us. The paper has benefited from helpful discussions with Charles Blackorby, Dave Ryan and Terry Wales, and from the comments of an anonymous referee and an editor of this JOURNAL. Responsibility for remaining errors rests with the authors.

The first concerns the commonly assumed restriction on the household's preferences of (weak) separability between goods and leisure. This assumption allows the estimation of commodity demand systems and Engel curves that exclude wage-rate variables. In cross-section household budget data with large variation in wage rates across the sample the invalidity of this assumption would involve a serious mis-specification. Any correlation between the excluded wage variable and the included price, income and demographic variables would lead to biased parameter estimates and hence biased estimated elasticities. Similarly, this assumption allows the estimation of labor supply curves that exclude relative price variables and its invalidity could produce biased labor supply elasticities.

A second aspect of household decision-making which we wish to highlight arises from the suspicion, commonly alluded to in labor economics texts, that primary male workers may not be free to choose their hours of work. Thus we estimate a matched pair of rationed and unrationed systems following the work of Neary and Roberts (1980) and Deaton and Muellbauer (1981). The separability restriction is not unrelated to this point, since without it we will see that the rationed hours of work enter demand systems in a complex fashion. In effect, in the absence of separability, rationing has direct substitution effects in addition to the indirect income effects.

The third aspect we wish to emphasize concerns the leisure decision. Here female leisure is considered as distinct from male leisure both with respect to commodity consumption and to its response to changes in household composition. Moreover, if male labor supply is predetermined because of rationing, the household's income will still be endogenous through its dependence on the female labor supply decision. Thus we regard it as important to model female labor supply behaviour despite the fact that it complicates the econometric analysis.

Finally, we introduce demographic variables to capture the effects of household composition on both labor supplies and commodity demands. We do this by extending the translation method of Pollak and Wales (1978), but other methods could equally be used, to incorporate the effect of household composition on leisure as well as commodity expenditures. The empirical significance of the effect of young children on female labor supply decisions has been demonstrated by Barton, Layard and Zabalza (1980) and here we attempt to distinguish between the number of children and their ages by allowing for economies of scale. Estimates of the effects of household composition on expenditure have often been used to compute a cost of living subindex defined over commodities alone and then used to make welfare comparisons between households (see Blundell (1980) and Muellbauer (1974)). However, Pollak and Wales (1979) point out that such a comparison ignores the utility which children yield to their parents. Indeed, if composition is endogenous, the logic of their argument suggests that children should leave the household's cost of living unaffected since the equivalent income generated by their presence must be at least as great as the extra expenditure entailed. The empirical issue of endogeneity we leave for subsequent research but note that when making welfare comparisons within the traditional framework the cost in terms of foregone female time available for work or leisure is likely to be at least as important as the cost in terms of commodity expenditures. In addition, a cost of living sub-index defined over commodities alone will only be defined when preferences are separable between goods and leisure (see Blackorby, Primont and Russell (1978: 327)).

Since the separability assumption is crucial to the purpose of this paper, we choose a model which allows this assumption to be tested against the data. The *a priori* unreasonableness of the separability assumption can be demonstrated by noting that if all goods are normal and leisure is a substitute for at least one commodity then separability requires that it be a substitute for all commodities. This is an unattractive restriction to impose as a maintained hypothesis since at least some consumption activities are likely to be complementary to leisure. Thus, by choosing a model which does not impose separability we avoid these unattractive properties and we can discover which goods are complements to leisure and which are substitutes. The usefulness of such a classification stems from the role of complementaries between goods and leisure in optimal taxation theory. For example, if in a two good world, good 1 is a substitute for leisure and good 2 is a complement to leisure and the tax rates are initially the same on both goods, then by decreasing the tax rate on good 1 and increasing the tax rate on good 2, holding tax revenue constant, the supply of labor will increase. This is because taxing complements to leisure is effectively an indirect way of taxing leisure itself (see, for example, Atkinson and Stiglitz (1972) and Sandmo (1976)).

The assumption of separability has been rejected by Barnett (1979) using a time series of US data. However, separability is a hypothesis about individual preferences, so it is clearly preferable to test it against micro data rather than against aggregate time series data. In this study we utilise family budget and labor supply data and the consequent large variation in wage rates across households should give greater power to our separability test.

As we have suggested, it is not clear that all leisure decisions of the household are freely made. Hours worked may be determined by employer preferences, union pressure, legislation or involuntary unemployment. Given the lack of information in our data set concerning the presence or absence of such constraints we allow for both possibilities in our estimation. When male labor supply is rationed all variability in household income arises from female labor supply behavior. This contrasts significantly to the leisure-goods model of Atkinson and Stern (1980) where male labor supply is assumed to be unrationed and female labor supply has no direct effect on leisure-goods decisions. To estimate both rationed and unrationed models on an equal footing we require a matched pair of leisure-goods demand systems. Our model is a generalization of the familiar Linear Expenditure System which, although implying quasi-homothetic preferences, no longer imposes the unrealistic additive separability restrictions and is the most general specification for which a matched pair of demand systems is easily obtainable.

Our data is a sample of two adult manual worker families drawn from the UK Family Expenditure Survey of 1974. All households selected contained married women and this raises the problem of modelling the female participation decision. The decision to participate or not depends on the difference between the offered wage and the value of time spent at home. Obviously this difference will depend on the female's personal characteristics, and on household composition, as well as on wages and prices. However, we cannot observe the offered wage for non-participating females so that our model is left with a limited dependent variable and lack of data on an explanatory variable when the limit is observed. Such a model is a generalization of the limited dependent variable type of Tobin (1958). Wales and Woodland (1980) consider several possible estimation methods for such models. Since we are dealing with a system of utility maximizing consistent demand equations it proves too difficult to integrate out the unobserved female wage variable. Further, we have few variables with which to predict female wage and so we select households with a participating female only correcting for the resulting selectivity bias using a technique due to Amemiya (1974) and Heckman (1979).

The paper proceeds as follows. Section 3.1 sets up the problem using constrained and unconstrained household cost (expenditure) functions in a way which emphasises the similarities between the problems of constraints on male leisure decisions and the implicit constraint on female leisure decisions due to their inability to choose more than T hours leisure, where T is the maximum number of hours available for work and leisure. That is, the participation decision is viewed as one of constrained decision making. A functional form is chosen for the cost function which allows for non-separability and enables us to estimate demand equations for a sufficient number of goods to make non-separability a possibility. Section 3.2 outlines the technique used to correct for the selection bias caused by the exclusion of households with non-participating females from the sample. Section 3.3 describes the data, discusses some further estimation problems, and presents the estimates. We conclude with a discussion of the merits of the estimates and some comments on directions for future research in Section 3.4.

3.1 Specification of the Model

We suppose that each household faces the same prices $p = (p_1 \ldots, p_n)$ for n goods $q = (q_1 \ldots, q_n)$ and different marginal values $w = (w_m, w_1)$ for their male and female leisure time $l = (l_m, l_f)$. If T represents the total time available, and p' is unearned income, then the household budget constraint may be written as

$$p'q + w'l = w_m T + w_f T + \mu' = \mu,$$

where μ is known as full income.

Assuming that the utility function for each household, $U(q, l)$, is strictly quasi-concave, then the minimised cost of attaining a given utility level $\bar{U}$ defines the full cost function.

$$C(p, w, \bar{U}) = \min_{q,l}\left[p'q + w'l \mid U(q,l) \geq \bar{U}\right]. \tag{1}$$

The form of (1) chosen for our analysis is a generalization, due to Muellbauer (1981), of the Gorman Polar form given by

$$C(p, w, \bar{U}) = a(p) + w_m d_m(p) + w_f d_f(p) + b(p)^{1-\theta_f-\theta_m} w_f^{\theta_f} w_m^{\theta_m} \bar{U}, \tag{2}$$

where $a(p)$ and $b(p)$ are concave linear homogeneous functions and $d_m(p)$ and $d_l(p)$ are homogeneous of degree zero. The form of (2) is convenient since we can derive an explicit functional form for its rationed counterpart. In addition, the system of demand equations derived from (2) allows separability between goods and leisure to be tested. As with the Linear Expenditure System the interpretation of the first three terms in (2) is the necessary expenditures; in this case, necessary expenditures out of full income on commodities, male leisure and female leisure respectively. Notice that the necessary quantities are not assumed to be constants as in the Linear Expenditure System but are allowed to be general functions of prices. Since, however, our data are derived from a single cross section so that all commodity prices are constant the particular functional forms of $a(p)$, $d_m(p)$, $d_f(p)$ and $b(p)$ are not important provided they satisfy the homogeneity conditions. Indeed, they could be any general second-order flexible forms and therefore we are not imposing any separability restrictions between groups of commodities.

For purposes of exposition however, it is convenient to choose the following functional forms. We let the necessary costs of commodities, $a(p)$, have the form

$$a(p) = \sum_{i=1}^{n} p_i \gamma_i m_i,$$

where m_i are the number of equivalent adults in each household with respect to good i. For necessary male leisure hours we specify

$$d_m(p) = \gamma_m \prod_{i=1}^{n} p_i^{\delta_{mi}}, \gamma_m > 0, \sum_i \delta_{mi} = 0$$

and for necessary female leisure we add a term m_f which captures the direct effect of household composition on female leisure. Thus

$$d_f(p) = \gamma_f \prod_{i=1}^{n} p_i^{\delta_{fi}} + m_f, \gamma_f > 0, \sum_i \delta_i = 0$$
$$d_f(p) = \bar{d}_f(p) + m_f.$$

Finally, we write

$$b(p) = \prod_{i=1}^{n} p_i^{b_i}, \sum_i b_i = 1$$

Notice that household composition enters only via necessary commodity expenditures $a(p)$, and necessary female leisure time $d_f(p)$ (we assume that there is no direct effect of composition on the necessary male leisure time requirement). This method of incorporating household composition is an extension of the translation approach of Pollak and Wales (1978) to the leisure-goods model, and is explored in more detail in Blundell (1980). The precise specification and interpretation of m_i and m_l are left until Section 3.3.

The compensated commodity and leisure demand equations are obtained from the price and wage derivatives of (2). Eliminating $\bar{U}$ from these equations using (2) yields both the system of uncompensated demand equations given by

$$p_i q_i = p_i \gamma_i m_i + \delta_{mi} w_m d_m + \delta_{fi} w_f d_f + (1 - \theta_m - \theta_f) b_i \left[\mu' + (T - d_f) w_f + (T - d_m) w_m - a \right], \tag{3}$$

and, letting $h = (T - l)$ denote hours worked, the following labor supply equations

$$w_f h_f = w_f (1 - \theta_f)(T - d_f) - \theta_f \left[\mu' + (T - d_m) w_m - a \right], \tag{4}$$

$$w_m h_m = w_m (1 - \theta_m)(T - d_m) - \theta_m \left[\mu' + (T - d_f) w_f - a \right]. \tag{5}$$

Since prices are constant across households the terms d_m, $\bar{d}_l$ and b_i can be estimated as single parameters.

Turning to the separability issue, the results of Goldman and Uzawa (1964: 392) imply that for a group of commodities I to be separable from good $j \notin I$ requires

$$C_{ij} = kC_{iu} \text{ for all } i \in I,$$

where k is some constant function for all i, and the subscripts refer to the derivatives of the cost function. In the context of our leisure-goods model, separability of goods from male leisure requires

$$\delta_{mi} = 0 \text{ for all } i = 1, \dots, n. \quad (6)$$

However, the restrictions $\Sigma b_i = 1$ and $\Sigma \delta_{mi} = 1$ forces k to be zero so that (6) becomes

$$\delta_{mi} = 0 \text{ for all } i = 1, \dots, n.$$

Similarly, separability of goods from female leisure requires

$$\delta_{fi} = 0 \text{ for all } i = 1, \dots, n.$$

Inspection of our commodity demand system given by (3) shows that the omission of w_m and w_l would imply separability of goods from both male and female leisure. Thus our chosen cost function allows a particularly simple test of separability and, moreover, the test is not data dependent. Turning now to the problem posed by constraints on male labor supply we write the cost minimizing problem for the household subject to a given male labor supply as

$$C^R(w, p, \bar{l}_m, \bar{U}) = \min_{q, l_f} (w_m \bar{l}_m + w_f l_f + p'q \mid U \geq \bar{U}). \quad (7)$$

C^R is the minimum cost of achieving $\bar{U}$ at prices w_m, w_l and p given that the male is constrained to take exactly $\bar{l}_m$ leisure hours. The properties of the rationed cost function (7) are described in detail by Neary and Roberts (1980) and Deaton and Muellbauer (1980). For our purposes it is sufficient to outline the relationship between the rationed and unrationed cost functions and to derive the system of rationed demand equations.

Since $\bar{l}_m$ is not a choice variable we can define the conditional cost function as

$$\widetilde{C}(w_f, p, l_m, \bar{U}) = \min_{q, l_f} (w_f l_f + p'q \mid U \geq \bar{U}) = C^R - \bar{l}_m w_m. \quad (8)$$

To relate the rationed and unrationed functions we define the *virtual* wage, w_m, to be that wage which would just induce the ration to be freely chosen; that is

$$\frac{\partial c(w_f, \bar{w}_m, p, \bar{U})}{\partial w_m} = \bar{l}_m \tag{9}$$

As shown in Deaton and Muellbauer (1981) the convenience of our chosen cost function is that $\bar{w}_m$ can be solved for explicitly from (9). At $\bar{w}_m$ the minimum cost of achieving $\bar{U}$ is the same whether the ration is imposed or not; that is,

$$C(w_f, \bar{w}_m, p, \bar{U}) = C^R(w_f, w_m, p, \bar{w}_m, \bar{U}). \tag{10}$$

We can rewrite (10) in terms of the conditional cost function, yielding

$$C(w_f, \bar{w}_m, p, \bar{U}) = \tilde{C} + \bar{w}_m \bar{l}_m,$$

which after eliminating $\tilde{C}$ using (8) gives

$$C(w_f, \bar{w}_m, p, \bar{U}) = C^R - \bar{l}_m(w_m - w_m),$$

and rearranging gives

$$C^R = \bar{C} + \bar{l}_m(w_m - \bar{w}_m), \tag{11}$$

where $\bar{C}$ is the unrationed cost function, C, evaluated at the virtual wage $\bar{w}_m$. From the unrationed cost function (2) we can use (9) to derive the following expression for the virtual wage

$$\bar{w}_m = \left(\frac{\theta_m b^{1-\theta_f-\theta_m} w_f^{\theta_f} \bar{U}}{\bar{l}_m - d_m} \right)^{1/(1-\theta_m)} \tag{12}$$

Substituting (12) into (11) gives our functional form for the rationed counterpart to (2) as

$$C^R = a + d_f w_f + \bar{l}_m w_m + b^{1-\rho_1} w_f^{\rho_1} (\bar{l}_m - d_m)^{-\rho_2} \theta_m^{\rho_2} (1-\theta_m) \bar{U}^{1/(1-\theta_m)}, \tag{13}$$

where $\rho_1 = \theta_f/(I - \theta_m)$ and $\rho_2 = \theta_m/(I - \theta_m)$. The system of rationed demand equations are obtained from the derivatives of C^R with respect to p_i giving

$$\begin{aligned} p_i q_i = {} & p_i \gamma_i m_i + d_f \delta_{fi} w_f + \left[(1-\rho_1) b_i + \rho_2 d_m \delta_{mi} / (\bar{l}_m - d_m) \right] \\ & \times \left[\mu' + w_f (T - d_f) + wm(T - \bar{l}_m) - a \right]. \end{aligned} \tag{14}$$

Separability of goods from male and female leisures still requires that $\delta_{mi} = 0$ and $\delta_{fi} = 0$ for all i. Notice that unless goods are separable from male leisure, the ration affects not only the level of full supernumerary income (the final term in square brackets in (14)) but also each marginal propensity to consume (the first term in square brackets in (14)). Thus, for example, if a male worker is rationed to work less than he desires, full income is reduced, *ceteris paribus*, which with constant marginal propensities would simply have income effects on expenditures. But, in the absence of separability, the rationed worker will substitute away from substitutes to leisure into complements to leisure, that is he reduces his marginal propensities to consume substitutes to leisure and increases his marginal propensities to consume complements to leisure.

The problem of non-participating females is now easily incorporated into the analysis by regarding non-participation as the constraint $l_f = T$. When $l_f = T$ we can write the rationed cost function corresponding to (11) as

$$C^R(w_f > w_m, p, T, \bar{U}) = C^* + T(w_f - w_f^*),$$

where $C^* = C(w_m, w_f^*, p, \bar{U})$ is the unrationed cost function evaluated at w_f^*, the virtual price of female leisure at which $l_f = T$ would be freely chosen. This wage is more commonly referred to as the reservation wage. Females will choose to participate if the cost of doing so is less than the cost of not participating; that is, if $C^* < C^R$. Thus participation occurs when the offered wage, w_f, is greater than the reservation wage, w_f^*.

3.2 Econometric Analysis

As is usually the case in demand analysis, all equations in either the unrationed or rationed system contain the same explanatory variables. In addition, since both systems satisfy the usual adding up restrictions exactly, we can delete one equation from each system without loss of information. To describe the estimation method used in this study we concentrate on the unrationed system (3) and (4) and delete equation (5), but exactly the same technique is used on the rationed system (14). We write a stochastic version of (3) and (4) as

$$Y_i = X'\beta_i + \varepsilon_i (i = 1, \ldots, n, f), \tag{15}$$

where ε_i is normally distributed with zero mean and constant variance σ_i^2. Expression (15) is a system of $n + 1$ expenditure and female earnings equations where the β_i terms are nonlinear functions of the underlying parameters and for $i = 1, \ldots, n$, $Y_i = p_i q_i$, while for $i = f$, $Y_f = w_f h_f$. All dependent variables in this, or any other, expenditure system are constrained to be non-negative but, as we will show, this is unimportant provided the probability of attaining the zero limit is very small. However, for female labor income Y_f, this is not the case since a zero dependent variable will be observed whenever the offered wage is less than the reservation wage.

Estimating a system like (15) using joint least squares on a selected sample where $Y_f > 0$ gives rise to inconsistency. This can be seen from examining the expectation of the disturbances given $Y_f > 0$. For female participants we have

$$E(\varepsilon_f \mid Y_f > 0) = E(\varepsilon_f \mid \varepsilon_f > -X'\beta_f) \neq 0, \tag{16}$$

and similarly

$$E(\varepsilon_i \mid Y_f > 0) = E(\varepsilon_i \mid \varepsilon_f > -X'\beta_f) \neq 0 \text{ for all } i \neq f \tag{17}$$

provided $E(\varepsilon_i\, \varepsilon_f) \neq 0$. Assuming normality for the disturbances the probability that $\varepsilon_i > -X'\beta_f$ is given by $1 - G(L)$, where $L = -X'\beta_f/\sigma_f$ and G is the cumulative normal density function. Following Tallis (1961) the conditional expectations in (16) and (17) can now be written as

$$E(\varepsilon_i \mid \varepsilon_f > -X'\beta_f) = \lambda\sigma_{if} / \sigma_f \text{ for all } i = 1,\ldots,n,f,$$

where $\lambda = g(L)/[1 - G(L)]$, g being the standard normal density function.

By including λ linearly in all equations in (15) we can obtain consistent estimates using joint least squares. Unfortunately λ is unknown unless β_f/σ_f is known. In this study we estimate λ by first estimating β_f/σ_f using an instrumental variable estimator due to Amemiya (1973).

This ingenious estimator is derived from squaring equation (15) and taking expectations conditional on the female participating to give

$$E(Y_f^2 \mid \varepsilon_f > -X'\beta_f) = -X'\beta_f E(Y_f \mid \varepsilon_f > -X'\beta_f) + \sigma_f^2.$$

Thus we estimate

$$Y_f^2 = Y_f X'\beta_f + \sigma_f^2 + \eta,$$

where $\eta = \varepsilon_f^2 - \sigma_f L \varepsilon_f - \sigma_f^2$, using the instruments $(X\hat{Y}_f : 1)$, where $\hat{Y}_f$ is the least-squares prediction from (15). This consistently estimates β_f and σ_f^2, from which we can derive $\hat{\beta}_f/\hat{\sigma}_f$ and finally

$$\hat{\lambda} = \frac{g\left(\dfrac{X'\hat{\beta}_f}{\hat{\sigma}_f}\right)}{1 - G\left(-\dfrac{X'\hat{\beta}_f}{\hat{\sigma}_f}\right)}$$

While it is true that the properties of this estimator are heavily dependent on the normality assumption, this is also the case for all alternative procedures such as the probit method of Heckman (1979).

The introduction of $\hat{\lambda}$ into each equation of the system (15) produces consistent estimates of all parameters when joint least squares is used on the selected sample.

The variance-covariance matrix for the resulting estimates of the underlying preference parameters can be obtained from an extension of the results of Lee, Maddala and Trost (1980) and its derivation is available from the authors on request.

3.3 Empirical Results

Only households containing two married adults of working age with the head of household a male manual employee were selected. Choosing a sample with working wives reduced the number of households from 208 to 115 implying a participation rate (in this sample of working age females) of a little over 55%. Following the procedure of Atkinson and Stern (1980) the marginal wage rates were calculated by multiplying the normal gross hourly earnings by one minus the basic tax rate which includes an adjustment for national insurance contributions. Unearned income was then defined simply by the linear budget constraint. In order to reduce the importance of unexpected fluctuations in hours worked, normal rather than actual hours worked were used for both males and females.

The form of our cost function (2) assumes quasi-homothetic preferences and in order to make this assumption more palatable we made a further selection of households, choosing those with total weekly expenditure (on all goods except housing) in the range of £35 to £55 per week, the pre-selection sample average being £42.60. The resulting sample contained 103 observations. This selection on the basis of a sum of dependent variables will, in general, lead to inconsistent parameter estimates. However, as 90% of the sample fell within this expenditure range the resulting inconsistency is likely to be small, and for this reason we only correct for the selection bias caused by the selection of female participants. Finally to reduce the possibility of heteroscedasticity all dependent variables were defined as expenditure shares.

The effect of household composition on the leisure-goods choices of the household was entered via the necessary expenditure terms of our cost functions as described in Section 3.1. The commodity specific composition effects, m_i, were allowed to be general continuous functions of household age structure and size, the estimated parameters of which give the underlying continuous commodity equivalence scales as discussed in Blundell (1980). These commodity specific effects have an *indirect* effect on labor supply through the function $a(p)$ which enters the supernumerary full income terms in the labor earnings equations. An increase in the number or age of children tends to increase $a(p)$ and have a positive effect on both male and female labor supply (if male and female leisure are normal goods). We also allow composition to enter female labor supply decisions *directly* through the term m_f in the $d_f(p)$ equation. Although there are many possible ways of entering composition, this is not the primary aim of this paper and we choose a particularly parsimonious form which allows for age and scale effects. In particular we write

$$m_f = \gamma_a n + \gamma_b n^2,$$

where $n = \sum_{t=0}^{18} \alpha^t n_t$ and where n_t is the number of children of age t and α is a parameter which captures the depreciating effect of age of child on female labor supply. The parameter α was estimated by grid search across the range 0.5–1.0. While we recognize that a continuously declining age effect will not capture perfectly the structural changes due to children attending school with varying day lengths as they get older, we feel that it should act as a good approximation.

This direct influence of composition on female labor supply will tend to counteract the indirect expenditure effect that enters through $a(p)$. We would expect the direct (indirect) effect to dominate for households with younger (older) children and hence lead to a decrease (increase) in female labor supply. The interplay between these composition effects is obviously of some interest for horizontal equity and provides another reason for analysing expenditure and leisure decisions simultaneously.

The estimates of the unrationed model with the correction for selection bias are given in Table 3.1. The nonlinear routine RESIMUL of Wymer (1978) was used to generate the parameter estimates having deleted the male labor supply equation in order to remove the singularity of the system. Since all parameters automatically satisfy the adding-up restrictions across equations the estimates are invariant to the equation deleted.

Overall the parameters look plausible. We note that the crucial restriction of separability between goods and leisure, for which Wald test statistics are given by χ_f^2 and χ_m^2, can be rejected.[1] To consider the importance of complementarities between goods and leisure it is useful to note that for our cost function the compensated substitution effects between good i and, say, female leisure is given by

$$C_{if} = \delta_{if} d_f + \theta_f C_{iu} \bar{U}.$$

Table 3.1. Parameter Estimates of the Unrationed Leisure-Goods Model[a]

Commodity group	γ_i	$(1-\theta_m-\theta_f)b_i$	$d_m\delta_{mi}$	δ_{fi}	σ_{fi}/σ_f
Food	4.8198	0.0613	−0.7809	−1.1603	0.0484
	(0.9261)	(0.0401)	(1.2011)	(1.7758)	(0.0458)
Energy	0.9261	0.0021	−0.6302	1.4301	0.0251
	(0.2311)	(0.0148)	(0.7491)	(0.7654)	(0.0182)
Clothing	0.9721	0.0813	−0.9496	2.5208	−0.0201
	(0.3628)	(0.0357)	(1.1666)	(1.0566)	(0.0408)
Durables	1.3857	0.0628	−2.8993	−0.8381	0.0817
	(0.6243)	(0.0312)	(1.0681)	(1.2879)	(0.0357)
Transport	2.0103	0.0621	2.6688	1.3184	0.0853
	(0.7285)	(0.0373)	(1.2011)	(1.8795)	(0.0542)
Services	2.6200	0.1465	2.5914	−4.9470	−0.0306
	(1.0586)	(0.0472)	(1.2684)	(2.1302)	(0.0466)

$T-d_f = \frac{36.585}{(2.4952)}$ $\theta_f = \frac{0.2212}{(0.0177)}$ $\gamma_a = \frac{13.9176}{(2.9591)}$ $\alpha = 0.9$ $\chi^2_m = 19.16$

$T-d_m = \frac{47.6861}{(3.1185)}$ $1-\theta_f-\theta_m = \frac{0.4159}{(0.0586)}$ $\gamma_b = \frac{-3.1871}{(1.0407)}$

$\sigma^2_f = \frac{0.1070}{(0.0465)}$ $\chi_{f\wedge 2} = 22.13$

Note: [a]Asymptotic standard errors in parentheses. The critical $\chi^2_{0.01}$ value for 5 degrees of freedom is 15.09.

Since θ_f, and $\bar{U}$ are both positive, as is C_{iu} if female leisure is a normal good, a strongly negative estimate of $\delta_{if}d_f$ indicates complementarity, while a positive $\delta_{if}d_f$ indicates substitutability. Services and transport are strong substitutes for male leisure, whereas clothing, food, energy and our definition of durables tend to be complements to male leisure.[2] As might be expected these goods do not necessarily have the same relationships with female leisure. Services tend to be complementary to female leisure, clothing is a substitute and energy tends to be a complement. The presence of children in the household has a pronounced effect on female labor supply, since γ_a is highly significant and large, the birth of a first child reducing the time available for female work by nearly 14 hours per week. γ_b is significantly negative indicating economies of scale in the care of children.

The estimates of the parameters of the rationed model are presented in Table 3.2. Again the Wald tests indicate the rejection of separability which implies that even if labor income is given it is not sufficient to let leisure time effect commodity demands simply through the income term. The estimate of the labor time available for females, $T-\bar{d}_f$, has increased over the unrationed estimate such that now all females in the

Table 3.2. Parameter Estimates of the Rationed Leisure-Goods Model[a]

Commodity group	γ_i	$(1-\rho_1)b_i$	$\rho_2 d_m \delta_{mi}$	$\bar{d}_f \delta_{fi}$	σ_{fi}/σ_f
Food	3.3696	0.1512	−0.1076	−0.1842	−0.0976
	(0.6032)	(0.0313)	(0.0567)	(0.4641)	(0.0516)
Energy	0.7109	0.0150	−0.0233	0.8247	0.0148
	(02.551)	(0.0129)	(0.0164)	(0.6107)	(0.0216)
Clothing	0.7777	0.1114	0.0137	1.3065	−0.2117
	(0.6266)	(0.0284)	(0.0401)	(0.6015)	(0.0499)
Durables	1.4496	0.0976	−0.0263	−1.1148	0.0244
	(0.4560)	(0.0258)	(0.0521)	(0.2435)	(0.0446)
Transport	2.0420	0.1821	−0.0800	0.9311	−0.0289
	(0.9602)	(0.0307)	(0.0454)	(1.5401)	(0.0521)
Services	1.3911	0.2491	0.2195	−1.7633	−0.1265
	(0.8649)	(0.0366)	(0.0852)	(1.0613)	(0.0620)

$T - d_f = \frac{40.8869}{(2.7300)}$ $\quad 1-\rho_1 = \frac{0.8066}{(0.0410)}$ $\quad \gamma_a = \frac{20.7869}{(3.6308)}$ $\quad \alpha = 0.9$

$\chi^2_m = 18.65 \quad \chi^2_f = 21.01$

Note: [a]Asymptotic standard errors in parentheses. The critical $\chi^2_{0.01}$ value for 5 degrees of freedom in 15.09.

sample work less than this maximum. The relationships between goods and female leisure remains unchanged, but transport now tends to be a complement to male leisure. The effect of children on female labor supply is now even more pronounced and economies of scale are greater.

Finally it is important to note that the unrationed and rationed models cannot be compared statistically since the two systems contain different numbers of dependent variables and have quite different stochastic specifications.[3] Without further information on the degree of rationing in the sample (see, for example, Ham (1982)) it is difficult to test the rationing hypothesis. The most that can be said from a comparison of Table 3.2 with Table 3.1, is that there are no indications that rationing is not at work. More positive conclusions await improved data.

Since we are working with a single cross-section exhibiting no price variation, we are unable to identify the own price elasticities for goods. However, wages do vary across the sample and we are able to identify the labor supply elasticities and evaluate these at the average hours worked (39.6 per week for males, 20.2 for females). For the unrationed case these are given by

$$\frac{w_j}{h_j}\frac{\partial h_j}{\partial w_j} = \frac{(1-\theta_j)(T-d_j)}{h_j} - 1 \; (j = m, f),$$

and for the rationed model, the female labor supply elasticity is given by

$$\frac{w_f}{h_f}\frac{\partial h_f}{\partial w_f} = \frac{(1-\rho_1)(T-d_f)}{h_f} - 1.$$

Table 3.3. Labor Supply Elasticities

	No. of children	Unrationed	Rationed
Male		−0.2863	
	2	0.4274	0.6489
Female	1	0.1074	0.0889
	0	−0.1926	−0.3010

Table 3.3 presents these elasticities for no children, for one child aged 3 years, and for two children aged 3 and 6 years. Notice that for zero children the data here are consistent with much previous evidence of backward bending male labor supply and forward sloping female labor supply. As we would intuitively expect the labor supply of a female with a small child becomes more inelastic as it becomes more difficult to substitute into market time. But with two children the income effect begins to dominate the substitution effect and the female labor supply elasticity becomes negative. Contrasting the rationed and unrationed elasticities we can see some evidence that, in the presence of constraints on male labor supply, female labor supply becomes more responsive to changes in the female wage.

3.4 Conclusions

This paper has attempted to analyze the interactions between household decisions over labor supplies and commodity demands in a utility maximizing framework. These decisions have been modelled under the assumptions that either male labor supply is freely chosen or the observed male hours of work are exogenously determined so effectively imposing a ration on the household. We demonstrated the importance of the separability assumption in this framework and tested the restrictions implied by separability using data from the UK Family Expenditure Survey. Under both the unrationed and rationed assumptions these restrictions were rejected.

The incorporation of female labor supply into the analysis required that we face the important problem of non-participation. For estimation purposes we selected a sample of households with participating females only and then corrected for the resulting selectivity bias. We found evidence of strong household composition effects on female labor supply entering both through necessary female leisure time and through necessary commodity expenditures. These effects would typically result in higher (lower) levels of participation for females with older (younger) children compared with a household with no children. We also found evidence of economies of scale in the rearing of children.

The results of this work suggest a number of interesting avenues for further research. The implications of rejecting separability for optimal taxation theory are clear, but the computation of an optimal taxation system requires estimates of all the parameters of the model which needs data from at least two cross sections for identification. Secondly we have shown that rationing may be important for household decision-making through its income effects, and, given the rejection of separability, through its effects on marginal propensities to consume. Our empirical work has been only suggestive and further information on labor supply constraints will facilitate the pooling of our rationed and unrationed models. In principle, for example, our methods could be used to analyze involuntary unemployment along the lines pursued by Ashenfelter (1980) using times series data. Finally, we noted the importance of household size and composition for expenditure and leisure decisions. It would be particularly useful to attempt to incorporate composition and size effects into a life cycle model since, if intertemporal separability is not true, our estimates may be picking up life cycle phenomena. Moreover, we have assumed that children are exogenous and we may have overestimated the 'cost' of children to the extent that fertility decisions are part of a life cycle plan.

4

A Life-Cycle Consistent Empirical Model of Family Labor Supply Using Cross-Section Data

Richard Blundell and Ian Walker

4.1 Introduction

Although the empirical study of labor supply has generally been cast in a static framework, a number of important results have recently been obtained within an explicitly intertemporal decision-making model (see, for example, Heckman 1974b; Ghez and Becker 1975; Smith 1977; Heckman and MaCurdy 1980, MaCurdy 1981; 1983; Browning, Deaton and Irish 1985 and Ham 1986). However, underlying each of these studies are implicit restrictions on within period preferences, which are usually associated with the empirical approach adopted rather than with the life cycle theory itself. For example, the empirical framework chosen by Heckman and MaCurdy (1980) and MaCurdy (1981) imposes within period additivity which clearly implies strong restrictions on behaviour (see Deaton 1974). In view of the variety of labor supply behavior observed in micro-data it is particularly important to allow a flexible representation of preferences as simple labor

The original version of this chapter was published as: Blundell, R./Walker, I. (1986). A Life-Cycle Consistent Empirical Model of Family Labour Supply Using Cross-Section Data, in: Review of Economic Studies, 53(4): 539–58. © 1986 by Oxford University Press. This paper estimates a utility maximising model of the joint determination of male and female labor supplies using a sample of married couples from the UK Family Expenditure Survey. The emphasis is on the estimation of within period preferences that are consistent with intertemporal two-stage budgeting under uncertainty. However, the approach we adopt provides an alternative method of estimating certain aspects of life-cycle behaviour to the fixed effects λ-constant approach of Heckman and MaCurdy (1980), MaCurdy (1981) and Browning, Deaton and Irish (1985). Moreover, it relaxes some of the underlying restrictions that are implicit in these λ-constant models under uncertainty.

supply models often impose quite implausible restrictions on within period behaviour (see Stern 1986).

Through the combination of intertemporal two stage budgeting and a dual representation of within period preferences, we generate a family labor supply model which is consistent with the life cycle theory and nevertheless relaxes a number of the important underlying restrictions on within period preferences inherent in previous empirical models. Using a single cross-section of data we can only retrieve intertemporal elasticities with the addition of some identifying assumptions on intertemporal preferences (see MaCurdy 1983), but these assumptions are common to a number of the models referred to above. Moreover, our estimates of within period preferences are invariant to such assumptions. In an intertemporal context the static specification, which has current labor supplies determined by the current marginal real wages and the current level of unearned income, is incorrect unless asset levels are planned or constrained to stay constant throughout the life cycle. Current labor supply will depend not only on current assets and current real wages, but also on all future real wages. Thus the measurement of unearned income, which is a problem for the static model, does not arise directly in the intertemporal model. What is a problem, however, is the definition of a life cycle consistent model in a form suitable for econometric estimation using available data. The source of this problem is that the whole life cycle of real wages and their expectations is not available.

Ghez and Becker (1975) and Smith (1977) attempt to overcome the missing data problem by constructing synthetic cohorts using estimated life cycle wage profiles from cross-section data on individuals of different ages. The synthetic cohorts are assumed to depict the life cycle of a representative individual. The major difficulty with this approach is that it tends to suffer from cohort bias and hence confounds cohort effects with true life cycle effects. Many of the problems associated with the use of synthetic cohorts can be overcome by exploiting the repeated observations available in panel data, together with the important theoretical insights of Heckman (1974a), Heckman and MaCurdy (1980) and MaCurdy (1981). They show that under certain assumptions the individual's marginal utility of money (or λ) is, after suitable discounting, constant over the life cycle, and that labor supply functions which condition on λ provide a suitable framework for both interpretation and estimation of life cycle behavior. The wage derivatives of such functions, for example, pick out exactly the responses to *anticipated* wage changes. Although unobservable, λ can be eliminated in estimation with panel data using the repeated observations available on each individual.

In all these models preferences are assumed intertemporally separable, so that individual decision-making can be viewed as a two-stage budgeting process, which makes the intuition behind the λ-constant approach clear. In the first stage the household allocates full life cycle wealth across the lifetime so as to equalise the marginal utility of (suitably discounted) money in all periods of the life cycle. At the second stage, the current period's allocation of full income out of life cycle wealth is distributed between consumption and non-market time depending on the level of the current real wages – the influence of all past and expected future variables is captured by the level of λ determined at the first stage.

This approach has been further developed by Browning, Deaton and Irish (1985) who consider the theoretical background to this λ-constant specification by relating it to the derivatives of the profit function representation of household preferences described in Gorman (1976). The attractions of their study are the generation of functional forms that break the within period additivity of the Heckman-MaCurdy specification and the use of pseudo panel data constructed from age cohort means across successive random cross-sections. This procedure avoids the cohort bias inherent in synthetic cohorts and has an advantage over true panel data to the extent that pseudo panels do not suffer attrition. However, averaging across cohorts inevitably reduces the underlying variation in dependent and explanatory variables as well as causing some difficulty in appropriately capturing the participation decision, especially important in the labor supply of married women which is a vital element of our family labor supply model.

As a vehicle for interpreting life cycle labor supply behavior the λ constant (or Frisch) framework is clearly very appealing. However in the estimation of life cycle decisions under uncertainty or with replanning $\ln \lambda$ follows a random walk and is eliminated from the labor supply equation by first differencing. The way in which λ is allowed to enter the labor supply equation is consequently restricted. As a result not only are limitations imposed on the form of within period preferences, but also on the form of the "cardinalization" of utility. Indeed the estimates of within period preferences from the first differenced model will not be invariant to the form of the chosen cardinalization. Furthermore all current explanatory variables *not* fully anticipated become correlated with the stochastic error term and some instrumental variable procedure is required for estimation.

Despite the relaxation of the within period additivity restriction of the Heckman-MaCurdy specification that is achieved in Browning, Deaton and Irish (1985) there remains some strong restrictions on within period preferences.

These points are developed in Section 2 along with an alternative approach which uses a flexible model imposing few restrictions on preferences and yet is consistent with the same underlying life cycle optimizing model described above. The specification is generated from a dual representation of within period preferences allowing a general structure for wage and demographic variables. Household labor supplies that condition on the current period allocation out of life cycle wealth, rather than on the marginal utility of money, are derived. Given the intertemporal separability assumptions this turns out to be a natural conditioning variable under two-stage budgeting and can easily be seen to capture all future anticipations and past decisions just as λ does in the λ-constant specification.

A similar approach is adopted by Altonji (1986) in his study of male labor supply where current food consumption is used to replace the unobservable marginal utility of money. Food expenditure then effectively summarizes all future anticipations. MaCurdy (1983), using an additive specification for within period preferences, exploits expenditure in the DIME panel data to estimate a male marginal rate of substitution function which is consistent with life cycle planning. In our study we use the consumption expenditure data available in the UK Family Expenditure Survey.

In Section 4.3 the precise specification of our empirical application is described with particular attention to female participation, demographic variation and cohort effects. In Section 4.4 we consider the appropriate econometric estimation strategy and the results of applying it to a sample of married couples from the 1980 UK Family Expenditure Survey.

4.2 Two-Stage Budgeting and λ-constant Frisch Demands

A principal objective of this section is to compare the various alternative empirical parameterizations of the life cycle decision making model and point to their implicit assumptions. In order to do so we employ a common optimising framework. Initially we shall assume that households have perfect foresight and choose current labor supplies and commodity demands so as to maximize discounted lifetime utility subject to budget, time and asset accumulation constraints. Following previous empirical studies in this area (see for example Heckman and MaCurdy 1980 and Browning, Deaton and Irish 1985) we shall also assume intertemporal additive separability of life cycle utility. Although additivity is not necessary for the application of two stage budgeting, it is invaluable when we come to relax the perfect foresight assumption. Separability of current from future decisions, on the other hand, is crucial throughout, enabling the influence of all past decisions over labor supply and commodity consumption to be summarized through the current level of assets.

Defining x^s to be the choice vector in period s, containing female and male non-market time (l_{fs}, l_{ms}) and commodity consumption q_s, lifetime utility viewed from period t is written as the following discounted sum of concave and twice differentiable period by period utility indices $U_s(x^s)$,

$$V_t = \sum_{s=t}^{L} \delta^{s-t} U_s(x^s) \tag{1}$$

where L is the lifetime horizon and δ represents a subjective time discount factor. The direct dependence of period by period utility on "s" reflects the influence of predetermined taste shifter variables, such as family size and composition variables, on life cycle preferences.

Corresponding to x^s there is a price vector containing female wage (w_{fs}), male wage (w_{ms}) and a commodity price index (p_{qs}) which define a within period budget identity

$$p_s x_s = y_s \tag{2}$$

where full income y_s has the form

$$y_s = w_{js} T_f + w_{ms} T_m + \mu_s \tag{3}$$

with T_f and T_m as the total hours of market and nonmarket time available for women and men respectively. The variable μ_s is a measure of end of period s net dissaving and provides the crucial connection between current decisions and those in other periods. This point is easily seen by defining A_{s-1} to be the level of assets at end of period $s-1$ which earn interest at a rate r_s paid at end of period s and writing

$$\mu_s = r_s A_{s-1} - \Delta A_s \tag{4}$$

where ΔA_s is the additional asset allocation to period s. A negative value for ΔA_s, for example, reflects a movement of income out of period s and only for the case where $\Delta A_s = 0$ is the usual unearned income measure appropriate. This observation provides the important distinction between "static" and "dynamic" models under intertemporal separability.

Clearly, where $\Delta A_s \neq 0$ the use of unearned income in the within period budget constraint is incorrect and can lead to a misinterpretation of within period behaviour (see MaCurdy 1985).

The two constraints (2) and (4) are usefully combined into a lifetime wealth constraint

$$\sum_{s=t}^{L} \hat{p}^{s} \cdot x^{s} = W_t \tag{5}$$

where $\hat{p}^s = \rho_s p_s$ and ρ_s is the following market discount factor

$$\rho_s = \frac{1}{(1+r_s)(1+r_{s-1})(1+r_{t+1})} \text{ for all } s > t \text{ and } \rho_s = 1 \text{ for all } s = t.$$

Similarly total life cycle wealth at end of period t, W_t is given by[1]

$$W_t = (1+r_t)A_{t-1} + \sum_{s=t}^{L} \hat{y}_s. \tag{6}$$

In addition to (5) there are inequality constraints on time: the upper bound of T_f on l_{fs} merits particular attention as a binding constraint in any period reflects female nonparticipation in market work.

The form of (1) and (5) is ideal for the application of two stage budgeting results (see Blackorby, Primont and Russell (1978), or Gorman 1968) where at a first stage y, is chosen so as to equalize the marginal utility of money in each period and at the second stage x^t is chosen *conditional* on y_t. These second stage demands can then be compared directly with the corresponding λ-constant Frisch demands of Heckman and MaCurdy (1980) and Browning, Deaton and Irish (1985) which condition on the marginal utility of money λ_t rather than on y_t. Although each is simply a reparameterization of the same underlying optimizing model described above, their corresponding empirical representations will be shown to differ in important ways especially when the perfect foresight assumption is relaxed.

Under two-stage budgeting the allocation of y_t is given by

$$y_t = \phi_t \left(p^t, \hat{p}^{t+1}, \ldots, \hat{p}^{L}, W_t\right) \tag{7}$$

where ϕ_t is homogeneous of degree zero in discounted prices $\hat{p}^s$ and wealth W_t. It is clear from (7) that y_t summarizes the influence of all future economic and demographic variables on current period decisions as well as the influence of past decisions through A_{t-1} in W_t. The second stage allocation then determines within period demands according to

$$x_t = g^t(p^t, y_t) \tag{8}$$

where g^t is a vector of demand equations homogeneous of degree zero in the price vector p^t and the conditioning variable y_t.

The forms of (7) and (8) are only correct if there are no binding constraints on x^t in period t. Should, for example, the upper bound on female time bite, the form of g^t will generally change. This switch in preferences has been considered in Blundell and Walker (1983) and will be discussed further in Section 4.3 below. It is sufficient to note that the effect of this corner in the budget constraint is precisely the same as that observed in the standard static female participation model so long as the conditioning variables y_t is correctly measured irrespective of the regime. Binding constraints on x^s where $s > t$ will simply alter the form of (7) and will have no direct impact on (8). In each of these cases the measurement of y_t is crucial and where consumption data is available this can be achieved through the within period budget constraint without direct knowledge of the form of (7). Since the emphasis here is on estimating within period preferences and as we are able to measure y_t directly, the precise form that (7) takes across regimes is unimportant. Moreover, we shall see that under certain conditions intertemporal substitution elasticities can be recovered without knowing the future variables in (7).

As an alternative parameterization consider the familiar first order conditions (see Heckman and MaCurdy 1980) for the maximization of (1) subject to (5)

$$\frac{\partial U_t}{\partial x^t} = \lambda_t p^t, \tag{9}$$

and

$$\lambda_{s+1} = \frac{1}{\delta(1+r_{s+1})}\lambda_s, s = t,\ldots,L. \tag{10}$$

The relationship (10) between the marginal utility of money in each period provides the link between the current and other period decisions analogous to (7). The variable λ_t acts as a summary of between period allocations and is a suitable conditioning variable for λ-constant demands which can be viewed as a rearrangement of (9) generating

$$x^t = f^t(p^t, \lambda_t), \tag{11}$$

which are homogeneous of degree zero in p^t and λ_t.[2] The general properties of demand equations (11) are described in detail in Browning, Deaton and Irish (1985) and clearly provide a useful interpretation of life cycle behavior. For example, the wage elasticities identify the effect of (fully anticipated) movements along the household's lifetime wage profile. The drawback of working with (11) directly is that λ_t, is unobservable and therefore information in (10) has to be exploited for empirical implementation. Once perfect foresight is relaxed this approach can generate implicit restrictions on the type of within period preferences that are permitted.

To illustrate these various points and to motivate our own empirical illustration consider representing within period preferences by the following general Gorman Polar Form indirect utility function

$$V_t = F_t y_t - \left[\frac{a_t(p^t)}{(b_t(p^t)} \right] \tag{12}$$

where F_t is some concave transformation and the functions $a_t(p^t)$ and $b_t(p^t)$ are linear homogeneous functions of p^t. Assuming there are no binding constraints the allocation of life cycle wealth described by (7) may now be written as

$$y^t = a_t(p^t) + \theta_t \left[b_t(p^t), b_{t+1}(\hat{p}^{t+1}), \ldots, b_L(\hat{p}^L), W_t - \sum_{s=t}^{L} a_s(\hat{p}^s) \right] \tag{13}$$

where $a_s(p^s)$ and $b_s(p^s)$ act as price aggregators for across period allocations. From the application of Roy's identity to (12) within period demands will have the form

$$x_{it} = a_{it}(p^t) + \frac{b_{it}(p^t)}{b_t(p^t)} \left[y_t - a_t(p^t) \right], i = f, m, q \tag{14}$$

where a_{it} and b_{it} are the derivatives of a_t and b_t. The particular forms of $a_t(p^t)$ and $b_t(p^t)$ will be given in the next section where emphasis is placed on their flexibility and the way in which they depend on demographic variables. However, it is worth pointing out that by choosing $a_t(p^t)$ to contain linear wage terms the T_f and T_m parameters can be subsumed into estimated parameters of the labor supply equations avoiding the need to choose them arbitrarily. While estimates of the parameters of (14) are independent of the choice of F_t in (12), in order to generate the corresponding λ-constant demands a particular cardinalization must be chosen. For example if F_t is chosen to be log linear and independent of t the marginal utility of income from (12) is simply

$$\lambda_t = \frac{1}{y_t - a_t(p^t)}, \tag{15}$$

and the λ-constant demands corresponding to (14) take the form[3]

$$x_{it} = a_{it}(p^t) + \frac{b_{it}(p^t)}{b_t(p^t)} \frac{1}{\lambda_t}. \tag{16}$$

Thus all λ-constant elasticities can be retrieved from the estimation of (14) given the choice of F_t. Clearly as λ_t is unobservable, (16) cannot be estimated directly but if panel data were available then λ_t, could be written in terms of λ_0 from (10) which becomes a fixed effect for each household over the panel. Where panel data are not available but consumption data are, the alternative parameterization (14) becomes invaluable. Indeed, when perfect foresight is relaxed the use of λ-constant demands for estimation can become quite restrictive.

The introduction of replanning or uncertainty leaves the y_t-conditional model almost unchanged since all uncertainty is captured through y_t in (14). For any choice of F_t the λ-constant wage elasticities, now reflecting fully anticipated movements along the wage profile, can be derived from (16). However, the evolution of λ_s in (10) is replaced by the following stochastic specification which ensures that λ_s is positive for all s,

$$\lambda_{s+1} = \frac{1}{\delta(1+r_{s+1})}\lambda_s(1+\varepsilon_{s+1}) \tag{17}$$

where ε_{s+1} reflects all unanticipated "news" gathered in period $s + 1$ which, if exploited rationally, satisfies the condition[4]

$$E_s(\varepsilon_{s+1}) = 0. \tag{18}$$

Rewriting (2.17) as

$$\ln\lambda_{s+1} - \ln\lambda_s = \ln\left(\frac{1}{\delta(1+r_{s+1})}\right) + \xi_{s+1} \tag{19}$$

allows λ_s to be differenced out of λ-constant demands *provided* they are written linear in the logarithm of λ_t. Alternatively the left-hand side of (19) may be expressed as the marginal rate of intertemporal substitution as in Hansen and Singleton (1982). A general form for λ_t could be derived from (12) and estimation using (19) could proceed by generalizing the approach of MaCurdy (1983). However, in order to generate the more popular linear differenced models restrictions have to be imposed on the form of F_t and the form of within period preferences.

Working with direct within period utility, the λ-constant linear differenced model requires that within period preferences be *explicitly* additive (see Heckman and MaCurdy 1980; MaCurdy 1981). Browning, Deaton and Irish (1985), on the other hand, working with indirect preferences (12) generates a λ-constant linear differenced model that allows within period additivity to be relaxed. However the price aggregator, $b_t(p^t)$ is required to be Leontief restricting substitution possibilities for those with a high $y_t - a_t(p^t)$, and once again the form of F_t is chosen prior to estimation. In these λ-constant models ξ_{t+1} becomes part of the disturbance term in the differenced demand equations so that all anticipated components of price, wage and demographic variables dated $t + 1$ become correlated within the disturbances and suitable instruments for consistent estimation are required.

The alternative y_t-conditional parameterization adopted here, overcomes a number of these problems since within period separability can be relaxed quite generally. Moreover, in our random effects models, specified in Section 4, there is no *theoretical* implication that any of the explanatory variables, in particular y_t, be correlated with the disturbances. Indeed, even though y_t is a choice variable, it may be reasonable to assume that disturbances on (8), (relating to choices within a period) are independent of those on (7) (relating to allocations across time), although clearly it would be desirable to test this assumption. The problem with developing such a test, as the discussion following (7) and (8) indicates, is that the form for y_t in (7) not only depends on future unobservable variables but also switches in the presence of binding current or future period constraints. From (15) it can be seen that the form of λ_t switches in a similar manner. Since the decision not to participate in the labor market can be represented by a binding constraint on available time, the form of y_t in (17) will exhibit an endogenous switching reduced form. Although some progress has been made in testing the independence assumption in the standard limited dependent variable model (see Smith and Blundell 1983), the switching reduced form further complicates both testing this hypothesis and estimation under dependence.

4.3 An Empirical Specification of Family Labor Supplies

For the empirical work presented here it is most convenient to work with the household's cost function which can be derived for our specification of preferences by inverting the indirect utility function (12). Dropping the time subscripts and ignoring for the moment the taste shifter variables, this is given by

$$C(w_f, W_m, p_q, U) = a(w_f, w_m, p_q) + b(w_f, w_m, p_q)G(U), \tag{20}$$

where $U = F^{-1}V$, $G(0) = 0$ and $C(\ ) = y$ at the optimum.

For interior solutions, where both partners choose positive hours of work the cost function is of the above form with the usual properties (see Deaton and Muellbauer 1980). The wage derivatives of (20) yield Marshallian male and female labor supply equations on substituting the indirect utility function for U. However, should the female not participate in the current period then the household is at a corner solution with $l_f = T_f$ and the cost function takes on its "rationed" form.

Following the literature on household decision-making under rationing (see, for example, Latham 1980; Neary and Roberts 1980 and Deaton and Muellbauer 1981) the relationship between the rationed and unrationed cost functions is given by

$$C^R(w_f, w_m, p_q, T_f, U) = C(w_f^*, w_m^*, p_q, U) + (w_f - w_f^*)T_f \qquad (21)$$

where $C(\,)$ is (20) evaluated at the female reservation wage w_f* at which the female would just choose to work. In Blundell and Walker (1983) we demonstrated that only when current female time is separable from other current choice variables is it the case that the functional form of the cost function does not change when a corner solution is attained. In the absence of within period separability, the cost function switches to its rationed form when nonparticipation occurs, even when the time path of w_f has been perfectly anticipated. Since the cost function switches at a corner solution, so too will the form of the male labor supply equation. That is, the derivatives of (21) with respect to w_m generates a male labor supply equation subject to the constraint that $h_f = 0$ which is different from that obtained from the derivative of (20) for $h_f > 0$. However, under both behavioral regimes the use of the conditioning variable y to produce labor supply equations that are consistent with life cycle optimization remains legitimate. Thus the model is effectively one of switching regimes where female labor supply, itself an endogenous variable, acts as an indicator variable determining the sample separation.

A similar problem arises in Hausman and Ruud (1984) in the context of male and female labor supply with nonparticipation, and in Blundell and Walker (1982), in the context of labor supply and commodity expenditures. In Blundell and Walker (1982), rationing on male labor supply is analyzed and the problem is simplified by using a preference specification that allows the virtual or reservation wage to be solved explicitly. In Hausman and Ruud (1984) rationing takes the form of female nonparticipation and the preference specification is such that the reservation wage, w_f*, is defined implicitly by a quadratic equation which permits a relatively simple solution. Here rationing also occurs through nonparticipation but the preference specification is such that the reservation wage is defined by an implicit equation. However, our estimation technique is to select a sample of households for which $h_f > 0$ and use only the unrationed cost function to represent preferences. The resulting sample selection bias is then overcome through the appropriate sample likelihood outlined in Section 4.4. This avoids the need to estimate a matched pair of rationed and unrationed male labor supply equations and yet still identifies all the parameters of the model. It also has the advantage that it does not invoke in estimation the assumption that nonparticipation occurs purely because the (predicted) wage is less than the reservation wage. The form of the cost function (20) used here corresponds to a flexible version of Gorman Polar Form preferences (see Blackorby, Boyce and Russell 1978) where quasi-homotheticity is imposed but not within period separability. The specification of $a(w_f, w_m, p_q)$, the cost of living at $U = 0$, is chosen to be second-order flexible but also allows for the possibility of fixed coefficients or zero substitution, and takes the generalized Leontief form

$$a(w_f, w_m, p) = \gamma_{ff} w_f + \gamma_{mm} w_m + \gamma_{qq} p_q + 2\gamma_{fm}(w_f w_m)^{1/2} + 2\gamma_{fq}(w_f p_q)^{1/2} + 2\gamma_{mq}(w_m p_q)^{1/2}.$$

In contrast, the form of $b(w_f, w_m, p_q)$ is chosen to be second order flexible nesting the substitution possibilities of the Cobb-Douglas case, and has the following Translog form

$$\ln b(w_f, w_m, p_q) = \beta_f \ln w_f + \beta_m \ln w_m + \beta_q \ln p_q + \frac{1}{2}\left[\beta_{ff} \ln\left(\frac{w_f}{p_q}\right)^2 + \beta_{mm} \ln\left(\frac{w_m}{p_q}\right)^2 + 2\beta_{fm} \ln\left(\frac{w_f}{p_q}\right) \ln\left(\frac{w_m}{p_q}\right)\right]$$

where $\beta_f + \beta_m + \beta_q = 1$. The γ and β coefficients are preference parameters, some or all of which may depend on taste shifter variables. The forms for $a(\,)$ and $b(\,)$ are such that the resulting y-conditional demand system nests the familiar Linear Expenditure System (by setting γ_{ij} for $i \neq j$ and all β_{ij} to zero), yet retains its convenient linearity in transformed variables. Written as labor supply equations, the system has the form

$$\begin{aligned} h_f &= \hat{\gamma}_{ff} - \gamma_{fm}\left(\frac{w_m}{w_f}\right)^{\frac{1}{2}} - \gamma_{fq}\left(\frac{p_q}{w_f}\right)^{\frac{1}{2}} - \frac{\tilde{\beta}_f}{w_f}\left[\mu - \tilde{a}(w_f, w_m, p_q)\right], \\ h_m &= \tilde{\gamma}_{mm} - \gamma_{fm}\left(\frac{w_f}{w_m}\right)^{\frac{1}{2}} - \gamma_{mq}\left(\frac{p_q}{w_m}\right)^{\frac{1}{2}} - \frac{\tilde{\beta}_m}{w_m}\left[\mu - \tilde{a}(w_f, w_m, p_q)\right] \end{aligned} \tag{22}$$

Where

$$\tilde{\gamma}_{ff} = T_f - \gamma_{ff}, \tilde{\gamma}_{mm} = T_m - \gamma_{mm}, \; \tilde{a}(w_f, w_m, p_q) = a(\,) - w_f T_f - w_m T_m,$$

$$\tilde{\beta}_f = \beta_f + \beta_{ff} \ln\left(\frac{w_f}{p_q}\right) + \beta_{fm} \ln\left(\frac{w_m}{p_q}\right),$$

and

$$\tilde{\beta}_m = \beta_m + \beta_{fm} \ln\left(\frac{w_f}{p_q}\right) + \beta_{mm} \ln\left(\frac{w_m}{p_q}\right).$$

Since μ is directly observable through the budget identity (22) can be estimated. Moreover estimation of (22) can proceed without making any assumptions about the values of T_f and T_m. Indeed if values for T_f and T_m are assumed, the estimated preference parameters will be independent of those assumptions.

Following the procedure in Section 4.2 the λ-constant labor supply functions corresponding to (16) are given by

$$\begin{aligned} h_f &= \tilde{\gamma}_{ff} - \gamma_{fm}\left(\frac{w_m}{w_f}\right)^{\frac{1}{2}} - \gamma_{fq}\left(\frac{p_q}{w_f}\right)^{\frac{1}{2}} - \frac{\tilde{\beta}_f}{w_f}\left(\frac{1}{\lambda}\right), \\ h_m &= \tilde{\gamma}_{mm} - \gamma_{fm}\left(\frac{w_f}{w_m}\right)^{\frac{1}{2}} - \gamma_{mq}\left(\frac{p_q}{w_m}\right)^{\frac{1}{2}} - \frac{\tilde{\beta}_{mm}}{w_m}\left(\frac{1}{\lambda}\right), \end{aligned} \tag{23}$$

Under the particular choice of F in (12) the elasticities of the λ-constant functions are those with respect to evolutionary wage change over the life cycle and are of the form

$$\begin{aligned} E_{ff} &= \frac{1}{h_f}\left[\tilde{\gamma}_{ff} - \frac{1}{2}\gamma_{fm}\left(\frac{w_m}{w_f}\right)^{\frac{1}{2}} - \frac{1}{2}\gamma_{fq}\left(\frac{p_q}{w_f}\right)^{\frac{1}{2}} - \frac{\beta_{ff}}{w_f}\left(\frac{1}{\lambda}\right)\right] - 1, \\ E_{mm} &= \frac{1}{h_m}\left[\tilde{\gamma}_{mm} - \frac{1}{2}\gamma_{fm}\left(\frac{w_f}{w_m}\right)^{\frac{1}{2}} - \frac{1}{2}\gamma_{mq}\left(\frac{p_q}{w_m}\right)^{\frac{1}{2}} - \frac{\beta_{mm}}{m}\left(\frac{1}{\lambda}\right)\right] - 1. \end{aligned} \tag{24}$$

These correspond to male and female labor supply intertemporal substitution elasticities which theory dictates should be positive. The parameters required for the computation of (24) can be obtained from the estimation of (22), the Marshallian labor supply equations.

The deterministic specification of (22) is completed by allowing for the effects of taste shifter variables. The fact that our cross section consists of individuals of different ages suggests that some allowance should be made for different cohorts. That is, for instance, older women in the sample may have different leisure preferences simply because they were brought up at a time when female participation was less common than during the formative years of younger women in the sample. This suggests that β_f should be allowed to vary with female age, A. The role of dependent children is inevitably more complicated. The age of dependent children is likely to be extremely important as well as the number of children. Moreover it seems likely that there may be some economies of scale involved especially in their effect on the allocation of female time. With these aspects in mind, and with a view to economizing on parameters while attempting to capture the wide variety of behavior in the data we allow β_f to depend linearly on A, A^2, and dummy variables indicating the presence of the youngest child in each of three age groups; 0 to 4, 5 to 10, and 11 to 18. Thus we have

$$\beta_f = \beta_f^0 + \beta_f' D' + \beta_f'' D'' + \beta_f''' D''' \beta_f^A (A - 40) + \beta_f^{AA} (A - 40)^2$$

where $D' = 1$ if n', the number of children in the youngest group, is positive, $D'' = 1$ if $n'' > 0$ and $n' = 0$, $D''' = 1$ if $n''' > 0$ and $n'' = n' = 0$. Such a specification ought to allow us to separate the cohort effect of female age from the effect of children on the marginal value of time. The role of the taste shifter variables in the γ parameters is given by

$$\gamma_{ij} = \begin{cases} \gamma_{ff} = \gamma_{ff}^0 + \gamma_{ff}' n' + \gamma_{ff}'' D'' + \gamma_{ff}''' \, D''', \\ \gamma_{mm} = \gamma_{mm}^0 + \gamma_{ff}' n' + \gamma_{mm}'' D'' + \gamma_{mm}''' \, D''', \\ \gamma_{qq} = \gamma_{qq}^0 + \gamma_{qq}' n' + \gamma_{qq}'' n'' + \gamma_{qq}''' \, n''', \end{cases}$$

and

$$\gamma_{ij}^0 + \gamma_{ij}' n' + \gamma_{ij}'' (n'' D'')^{\frac{1}{2}} + \gamma_{ij}''' (n''' D''')^{\frac{1}{2}} \text{ for } i = m \text{ or } f, j = q$$
$$\gamma_{ij}^0 + \gamma_{ij}' n' + \gamma_{ij}'' D'' + \gamma_{ij}''' D''' \text{ for } i = f, j = m.$$

Such a specification allows for the possibility of economies of scale on male and female time but not directly on consumption. Further details of this demographic specification can be found in Blundell and Walker (1984).

4.4 Estimation, Data and Results

4.4.1 *A Maximum Likelihood Estimator*

For estimation purposes, a sample of families with working wives was selected and additive disturbances (random effects) assumed for the earnings and expenditure equations derived in Section 4.3. As these disturbances will very likely be heteroskedastic with a variance related to the overall level of income in the household, each equation was deflated by full income so that the dependent variables become budget shares. As is usual in demand analysis the system of estimating equations is overdetermined and all information on preferences can be recovered after the arbitrary deletion of one equation. However, the selection rule for households invalidates the use of seemingly unrelated estimation procedures since such estimators would suffer from selectivity bias as described in Heckman (1979). Indeed, the adding-up condition implies that selectivity bias will affect equations other than the female time equation over which the selection takes place (see Blundell and Walker 1983).

Estimation methods that correctly account for selectivity fall into two categories. The first are the two-stage methods of Heckman (1979), Lee, Madalla and Trost (1980) and Hanoch (1980). These are inefficient in comparison with full maximum likelihood procedures and require reasonably complicated algebraic manipulation to derive asymptotic covariances. Given that our system contains only two equations (after deletion), full maximum likelihood estimation which comprises the second group of estimators is feasible. The form of the likelihood for truncated samples is described in Hausman and Wise (1977), and Wales and Woodland (1980). It is a mixture of density and distribution functions with similar properties to the Tobit likelihood described in Tobin (1958) and Amemiya (1973).

If we let $\phi > (u_{fh}, u_{mh})$ represent the joint density of the disturbances on the female time and male time equations in share form, then the likelihood for a sample of H households is given by:

$$\prod_{h=1}^{H} \phi(u_{fh}, u_{mh}) / Pr(l_{fh} < T_f) \tag{25}$$

where $Pr(l_{fh} < T)$ is the probability that household h is selected. To maximise the log of (25) a nonlinear iterative technique is required: the method adopted here is a version of the quasi-Newton algorithm, E04JBF from the NAG Fortran Library, described in Gill and Murray (1972). As the truncated likelihood (25) cannot be made globally concave, particular care was taken to ensure that a global maximum was attained in all cases.

4.4.2 *Data*

In order to implement the specification outlined in Section 4.3 we require a source of data that contains hours worked, wages and either consumption expenditure directly or saving so as to construct consumption expenditure via the budget and asset accumulation constraints. Such information is available from the UK Family Expenditure Surveys and here we use a subset from the 1980 FES. The subset was chosen so that all households consisted of two married working employees (FES code A201 = 1) with the head of household either a manual worker (code A210=6, 7 and 8), a shop assistant (A210 = 5) or a clerical worker (A210 = 4). The resulting sample contained 1378 households giving a female participation rate of around 64%. The hourly gross wages, assumed exogenous, for both husband and wife were constructed as the gross earnings to hours ratio where the "normal" definitions of these variables were used in an attempt to minimise the measurement error problem.[5] The method used to calculate the households' marginal (aftertax) hourly wage rates is briefly described in the Data Appendix along with summary statistics for each variable in the sample.

4.4.3 *Empirical Results*

The parameter estimates, their standard errors and corresponding log likelihood values for two models of particular interest are presented in Tables 4.1a and 4.1b. Model 1(a) refers to the preferred specification chosen by the usual asymptotic likelihood ratio criterion from among those that adopt a Cobb-Douglas form for the price index $b(w_f, w_m, p)$ while Model 1(b) is the preferred specification from among those that adopt a Translog form. From the log likelihood values it is clear that Model 1(b) is the overall preferred specification. However, it is clear from both specifications that certain behavioral features of the household allocation of time and goods are robust to the change in specification.

Table 4.1. a) Model Estimates with $b(p)$ Cobb-Douglas

	Female time		Male time		Goods expenditure	
β_i^0	0.122	(0.021)	0.089	(0.008)	0.789	(0.021)
β_i'	0.191	(0.039)	0		−0.191	(0.039)
β_i''	0.160	(0.037)	0		−0.160	(0.037)
β_i'''	0		0		0	
β_i^A	0.00067	(0.001)	0		−0.0067	(0.001)
β_i^{AA}	0.0001	(0.0001)	0		−0.0001	(0.0001)
β_{fi}	0		0		0	
	61.52	(1.472)	0		−8.48	(1.894)
γ_{ij}^0			49.47	(0.249)	0	
					50.64	(2.489)
	9.70	(1.400)	0		3.25	(1.964)
γ_{ij}'			0		0	
					7.77	(2.974)
	4.59	(2.480)	0		8.15	(2.747)
γ_{ij}''			0		0	
					1.62	(1.783)
	8.73	(1.701)	0		0.72	(1.663)
γ_{ij}'''			0		0	
					11.84	(1.546)
	0.0028	(0.0001)	0.0003	(0.00006)	−0.0025	(0.0001)
σ_{ij}			0.0015	(0.00006)	−0.0012	(0.00009)
					0.0037	(0.0001)

b) Model Estimates with $b(p)$ Translog

	Female time		Male time		Goods expenditure	
β_i^0	0.144	(0.022)	0.089	(0.009)	0.767	(0.022)
β_i'	0.198	(0.039)	0		−0.198	(0.039)
β_i''	0.162	(0.038)	0		−0.162	(0.038)
β_i'''	0		0		0	
β_i^A	0.0064	(0.001)	0		−0.0064	(0.001)
β_i^{AA}	0.001	(0.0001)	0		−0.0001	(0.0001)
β_{fi}	0.249	(0.031)	0		−0.249	(0.031)
	52.03	(2.477)	0		2.33	(2.641)
γ_{ij}^0			49.47	(0.247)	0	
					41.25	(5.092)
γ_{ij}'	13.10	(3.519)	0		2.88	(4.296)
			0		0	
					12.02	(6.540)
	6.88	(2.630)	0		3.67	(2.542)
γ_{ij}''			0		0	
					0.52	(2.445)
	7.62	(2.645)	0		0.93	(2.383)
γ_{ij}'''			0		0	
					11.90	(2.319)
	0.0026	(0.0001)	0.0003	(0.00006)	−0.0023	(0.0001)
σ_{ij}			0.0015	(0.00006)	−0.0012	(0.00009)
					0.0035	(0.0002)

Turning first to the β coefficients which determine the marginal values of time and goods we found that the β_m coefficients are not significantly sensitive to composition and cohort changes and are therefore restricted to zero in these preferred specifications. Thus, the strong positive effects of the presence of pre-school and junior school children in the household on the marginal value of female time is reflected in the marginal value of goods expenditure. Similarly, the strong, almost linear, positive effect of female age on the marginal value of female time is reflected in its negative effect on the marginal value of goods. Thus, a 60-year old female will have a marginal value of time of approximately 0.26 in excess of that for an otherwise identical 20-year old; a difference which we interpret as a reflection of the change in tastes across cohorts but could also be attributed to vintage effects. Strong cohort effects on female labor supply are no surprise and have been reported elsewhere; for example in Greenhalgh (1977). In Table 4.1b the index $b(w_f, w_m, p_q)$ takes the Translog form and the coefficient of 0.249 is that on $\ln(w_f/\bar{w}_f)$ so as to maintain the comparability of the β coefficients. This coefficient implies a large wage impact on the marginal value of time; for example, a wage of 10% above the mean would add almost 0.1 to the marginal value of time. The most important differences between the two specifications lies in the contrasting ways in which the female real wage enters the equations. In Table 4.1b with the Translog specification, w_f enters via both the $a(w_f, w_m, p_q)$ and $b(w_f, w_m, p_q)$ functions, while in Table 4.1a w_f enters only via the $a(w_f, w_m, p_q)$ function. Both specifications imply that preferences between goods and female time are non-separable. Notice that the inclusion of $\ln(w_f)$ in the marginal budget shares of Table 4.1b appears to be strongly significant while the γ^0_{fq} parameter is small and insignificant. Thus, the Translog specification rejects separability through the significance of $(\ln w_f/\bar{w})^2$ in $b(w_f, w_m, p_q)$ while the Cobb-Douglas specification rejects separability through the significance of w_f in $\partial a(w_f, w_m, p_q)/\partial p_q$. Since the inclusion of $\ln(w_f)$ in the consumption function turns γ^0_{fq} from being significantly negative to insignificantly different from zero the diagonal elements γ^0_{ff} and γ^0_{qq} are correspondingly smaller.

The specification of the male equation is, in both tables, such that male time is additively separable from goods and female time so that this is essentially a Linear Expenditure System form. Thus, the interpretation of γ^0_{mm} is the usual one of subsistence leisure time. Since there is little variation in male labor supply to be explained such a simple specification does not seem inappropriate.

Analogous to the Linear Expenditure System we can evaluate subsistence quantities. Each row or column of the γ^0 matrix can, after suitable weighting, be summed to give the predicted expenditure on female time, male time and goods evaluated at $U = 0$ for households with no children. Thus, using Table 4.1a and mean wages, a two adult household subsistence female time is 54.2 hours ($T_f = 90$), subsistence male time is 49.5 hours ($T_m = 90$) and subsistence goods expenditure is £40.80.

Male labor supply is restricted to be independent of the presence and number of children except to the extent that they affect supernumary income, $y - a(w_f, w_m, p_q)$. The effect of children on subsistence costs is allowed through the γ', γ'', and γ''' matrices which show that the number of children have a large effect on 'necessary' expenditure and that the presence of children has a large effect on 'necessary' female time. These have an unambiguously positive effect on male hours of work indirectly through increasing $a(w_h, w_m, p_q)$. However, the effects of children are more complicated in the consumption and female time functions. The positive effect of the increase in subsistence goods expenditure on female labor supply tends to be offset by negative effects of the increase in subsistence female time. The net effect is likely to be negative for a young child and positive for an old child, and the negative effect of a young child is reinforced by its large effect on the marginal value of female time.

Finally, Tables 4.1a and 4.1b include the estimated variance-covariance matrices. Both sets of covariances indicate that given the underlying distributional assumption the selection bias in the female equation significantly affects the male and goods equations. Although these models were selected on the basis of the standard likelihood ratio criterion after appropriately adjusting for selectivity, the statistical properties of our estimated parameters depend critically on the appropriateness of the stochastic assumptions in our random effects specification. In particular, the joint normality of the error distribution across the equation system and the independence of explanatory variables with these errors. Thus an important area for future research is the development of reliable diagnostics for these assumptions in systems of this type.

Turning to the underlying properties of the models it is clear that the interactions in the models are complicated and since the elasticities are sensitive to the points at which they are evaluated we have selected a number of subsamples from the full sample (the means of the subsamples data are presented in Table 4.2a). The subsamples were selected in an attempt to represent a typical life cycle by selecting successively older women with a time path of fertility involving two children born five years apart, the first in her mid-twenties. The children grow older and eventually leave the household when the mother is in her early fifties. Since the households conforming to the snapshots of such a life cycle profile were selected from the single cross section the subsamples inevitably confound life cycle and cohort effects; and since many households do not conform to such a profile the subsamples do not exhaust the complete sample used in estimation. Nevertheless, this procedure is useful to check the predictive performance of the estimated equations and to examine the sensitivity of elasticities. In Table 4.2b the predicted hours of males and females are both corrected for the selectivity bias in the female equation, the predicted consumption figures are derived from the budget constraint. All predictions were obtained by calculating them for each relevant data point and taking the mean, rather than calculating them for the mean of the relevant subsamples. In general both sets of estimates tend to underpredict hours and hence underpredict consumption also. For the mean of the whole sample the error is greatest for female hours, but this is not surprising given the larger variation compared with male hours and consumption. Since there are relatively large changes in μ and w_m as well as female age across the subsamples it is difficult to conclude

Table 4.2. a) Sample Statistics For Selected Household Types

	Age range	Mean age	Household composition						Sample size	Marginal female wage (£)	Marginal male wage (£)	Female hours	Male hours	Net dissaving (£)	Consumption (£)
		$\bar{A}$	n'	n''	n'''	D'	D''	D'''	—	w_f	w_m	h_f	h_m	μ	q
1	16–30	24.2	0	0	0	0	0	0	251	1.39	1.90	36.1	39.2	–59.89	64.67
2	20–35	26.9	1	0	0	1	0	0	48	1.46	1.88	20.1	39.3	–34.22	69.98
3	25–40	31.0	1	1	0	1	0	0	69	1.40	2.08	17.9	40.2	–35.14	73.65
4	30–45	37.0	0	1	1	0	1	0	85	1.30	2.31	23.0	40.0	–38.83	83.34
5	35–50	42.0	0	0	1	0	0	1	78	1.35	2.14	24.4	40.9	–34.78	85.72
6	40–60	52.1	0	0	0	0	0	0	268	1.35	1.99	29.5	39.6	–48.72	69.75
Total sample	16–60	37.1	0.182	0.463	0.458	0.152	0.253	0.169	1378	1.38	2.08	26.6	39.8	–45.30	74.37

b) Mean Predictions and Elasticities For Selected Household Types

		Selectivity adjusted predictions			λ constant elasticities			Compensated elasticities						Income elasticities					
		$\hat{h}_f$	$\hat{h}_m$	$\hat{q}$	E_{ff}	E_{mm}	E_{gg}	e_{ff}	e_{fm}	e_{fg}	e_{mm}	e_{mg}	e_{gg}	η_f	η_m	η_q	$y - \hat{a}$	$\hat{\beta}_f$	$\hat{\beta}q$
Model (a)	1	35.3	39.4	61.04	−0.089	0.029	−0.110	−0.080	−0.002	−0.083	0.028	−0.026	0.022	−0.206	−0.280	3.281	26.01	0.04	0.87
Cobb-	2	18.5	40.0	67.99	0.084	0.010	−0.098	0.083	−0.007	−0.059	0.013	−0.008	−0.047	−2.542	−0.314	2.746	11.59	0.24	0.67
Douglas	3	18.1	39.7	72.78	0.143	0.013	−0.134	0.135	−0.013	0.103	0.016	−0.010	−0057	−3.13	−0.297	2.607	15.99	0.26	0.65
model	4	21.6	39.5	80.50	0.191	0.020	−0.171	0.165	−0.019	−0.132	0.020	−0.013	−0.068	−2.867	−0.277	2.387	22.33	0.26	0.65
	5	24.2	39.1	81.56	−0.012	0.033	−0.152	−0.013	−0.012	0.033	0.033	−0.027	−0.028	−1.203	−0.295	2.743	34.20	0.14	0.78
	6	30.1	38.7	68.93	0.080	0.046	−0.175	0.046	−0.019	−0.021	0.041	−0.031	−0.058	−1.367	−0.290	2.588	39.19	0.22	0.69
Whole Sample		25.9	39.5	72.60	0.028	0.024	−0.137	0.032	−0.010	−0.014	0.024	−0.018	−0.038	−1.323	−0.287	2.800	23.81	0.17	0.74
Model (b)	1	34.8	39.5	63.53	−0.055	0.027	−0.457	−0.057	−0.003	0.061	0.025	−0.022	0.013	−0.369	−0.279	3.141	31.13	0.08	0.84
Translog	2	18.8	39.9	68.24	0.129	0.013	−0.292	0.102	−0.011	−0.077	0.014	−0.008	−0.053	−2.930	−0.315	2.481	12.51	0.29	0.52
model	3	17.9	39.6	72.29	0.162	0.019	−0.344	0.102	−0.019	−0.075	0.018	−0.011	−0.045	−3.519	−0.298	2.423	17.92	0.30	0.61
	4	21.1	39.3	79.38	0.150	0.025	−0.441	0.079	−0.025	−0.045	0.025	−0.017	−0.043	−2.977	−0.278	2.354	27.72	0.27	0.54
	5	24.3	39.2	81.91	−0.042	0.033	−0.463	−0.061	−0.013	0.079	0.031	−0.025	−0.005	−1.305	−0.295	2.637	32.26	0.15	0.76
	6	29.8	38.9	68.92	0.018	0.041	−0.546	−0.029	−0.019	0.052	0.038	−0.028	−0.009	−1.466	−0.288	2.505	35.29	0.23	0.58
Whole sample		25.6	39.5	72.19	0.033	0.026	−0.433	0.009	−0.012	0.007	0.024	−0.018	−0.026	−1.495	−0.287	2.669	24.81	0.20	0.71

Note: All elasticities are for labor supplies or expenditure. Income elasticities are full income elasticities. Predictions $\hat{h}_f$, $\hat{h}_m$ *and* $\hat{q}$ are adjusted for the selection rule $h_f > 0$.

from the relative sizes of the errors that the model is misspecified in any particular systematic way, and the fact that our predictions are at most 8% low is reassuring. The male equation does not seem to track the variation in actual hours particularly well but this is not too important given the limited variation in the data. The consumption equation seems to capture the subsample data reasonably well except that it doesn't entirely capture the large increases for groups 4 and 5. Table 4.2b also presents labor supply and expenditure elasticities for the various samples using both specifications. The elasticities were calculated by computing the elasticities at each relevant data point and taking means. The full income elasticities confirm previous work and show that male time is a normal good, and female time more strongly so except where the female is young and without children, when the two elasticities take broadly similar magnitudes. The compensated own price/wage elasticities take the appropriate sign when averaged over the whole sample but occasionally contradict economic theory when averaged over the subsamples. The compensated cross elasticities are negative except for e_{fq} which change sign across subsamples for both sets of parameter estimates. Thus male and female time are complements while female time and consumption as one would expect are often substitutes. Such a pattern of substitution could not arise for normal goods if separability were a maintained hypothesis. Through the Slutsky equation we can retrieve the uncompensated elasticities which hold μ_t constant. The female labor supply elasticity is generally positive for both specifications.

Given a particular choice of the cardinalization, F, in the indirect utility function (12) we can recover λ-constant intertemporal substitution elasticities from the estimated y-conditional equations. For example, choosing F to be log linear generates the elasticities in Table 4.2b. Under the assumptions of a perfect credit market and rational expectations the λ-constant labor supply elasticities are required to be positive and the λ-constant consumption elasticity is required to be negative. Unlike intertemporal elasticities derived in the λ-constant approach of Heckman and MaCurdy (1980) and MaCurdy (1981) our specification allows intertemporal substitution effects to be data dependent. The variations in the data across the subsamples results in quite large variations in the elasticities. Again, the female own intertemporal elasticity violates economic theory for some subsamples, but otherwise the elasticities take the appropriate sign.

Finally, in Table 4.3 we provide the comparative statics of a large change in net dissaving on the behavior of six hypothetical households.

Table 4.3. Model Simulation for Hypothetical Households

a) w_f = £1.00, w_m = £2.00, μ = –£10.00

Age	Composition			Selectivity adjusted predictions						
	n'	n''	n'''	$\hat{h}_f$	$\hat{h}_m$	$\hat{q}$	$\hat{\beta}_f$	$\hat{\beta}_q$	$y - \hat{a}$	$\hat{\varphi}_f$
25	0	0	0	34.2	37.6	99.17	0.05	0.87	65.85	0.99
25	1	0	0	16.3	38.3	79.70	0.24	0.68	31.88	0.80
30	1	1	0	16.2	38.4	79.63	0.26	0.66	30.26	0.79
35	0	1	1	19.4	38.6	84.58	0.25	0.66	31.49	0.88
40	0	0	1	23.7	38.3	89.33	0.12	0.79	43.83	0.95
50	0	0	0	25.1	37.4	89.04	0.20	0.71	65.85	0.96

b) w_f = £1.00, w_m = £2.00, μ = –£60.00

Age	Composition			Selectivity adjusted predictions						
	n'	n''	n'''	$\hat{h}_f$	$\hat{h}_m$	$\hat{q}$	$\hat{\beta}_f$	$\hat{\beta}_q$	$y - \hat{a}$	$\hat{\varphi}_f$
25	0	0	0	36.3	39.8	55.88	0.05	0.87	15.85	1.00
25	1	0	0	23.8	41.3	45.95	0.24	0.68	–18.12	0.98
30	1	1	0	24.5	41.3	46.87	0.26	0.66	–19.74	0.99
35	0	1	1	29.0	41.3	51.55	0.25	0.66	–18.51	0.99
40	0	0	1	28.4	40.8	49.86	0.12	0.79	–6.17	0.99
50	0	0	0	33.8	39.8	53.45	0.20	0.71	15.86	1.00

One interpretation of the table is the effects of a decrease in unearned income of £50 in the static model – that is, the model with perfect credit rationing. In this case the behavioral changes given by the predicted hours, consumption, and female participation $\hat{\phi}_f$ are the result of pure income effects. The life cycle model offers two alternative interpretations. Low net saving in Table 4.3a may arise from either high initial assets or in anticipation of high future wage growth. Thus, under the life cycle interpretation the households in Table 4.3a are currently identical to the corresponding households in Table 4.3b but differ in respect to either their pasts or their anticipated futures. In the first case the behavioral responses are the result of pure property income effects. In the second case, where the households in Table 4.3a anticipate high future wages relative to the current wage and those in Table 4.3b anticipate low future wages, the behavioral effects are the result of both intertemporal substitution and wealth effects. Under additive intertemporal separability such substitution effects are proportional to income effects.

4.5 Summary and Conclusion

This paper presents estimates of a model of family labor supply which allows for quite general effects of relative wages and demographic variables on within period behaviour and yet is consistent with life cycle optimising behavior under intertemporal separability and uncertainty. Furthermore, providing some identifying assumptions are made on the cardinalization of utility, certain useful intertemporal substitution elasticities can be retrieved. These identifying assumptions turn out to be no more restrictive than those imposed in many of the popular λ-constant life cycle labor supply models. Moreover, our model relaxes a number of the underlying restrictions on within period preferences implicit in the empirical formulation of these alternative specifications and our estimates of within period preference parameters are invariant to the chosen cardinalization. Indeed, one of the motivations for this study was to highlight the importance of allowing for general substitution and demographic effects in the within period allocation of time and goods. These effects were found to be critically important in the analysis of family labor supply presented here.

For our empirical application we chose a large sample of working couples from the 1980 UK Family Expenditure Survey. The advantage of this survey is that it not only has reliable earnings and hours data but also collects information on commodity expenditures which, following the important work of MaCurdy (1983; 1985), allow the unobservable marginal utility of wealth in the λ-constant formulation to be replaced by an observable full-income variable. It is this replacement that permits the relaxation of the implicit restrictions on within period preferences underlying the empirical formulations of the λ-constant model. The estimated models confirm the need for a flexible representation of preferences over time and goods. As one would expect, the hours of work of women in the sample appear to be much more responsive to changes in marginal wage and income variables than is the case for men. Simple models such as the Linear Expenditure System, which implies linear earnings equations, are easily rejected by the data as are models that restrict the interaction of demographic and economic variables. The resulting estimates for our preferred models were found to track the large variety of observed behaviour across different subsamples relatively well, and the estimated elasticities were generally found to be consistent with economic theory.

Data Appendix

Gross wages were calculated from the ratio of normal earnings (A080) to normal hours (A220). The amount of taxable income was deduced from: the date of interview (since tax allowances varied across the year), interest payments on mortgages (code 130 or 150), life insurance premium (codes 196 and 199) and superannuation (pension) contributions (code 318). National insurance contributions were calculated from the date of interview and individual earnings. Entitlement to Family Income Supplement was calculated from the date of interview, the number of children and gross household income excluding child benefit. The appropriate marginal deduction rate was then applied to the gross hourly wage as described in Blundell et al. (1986). On the accuracy of the FES data on earnings and hours see Atkinson, Micklewright and Stern (1981; 1982). On the overall degree of nonresponse, see Kemsley (1977). Total household expenditure is defined as expenditure on energy, food, clothing and footwear, services and other goods (the sum of codes 368P, 369P, 376P, 372P and 374P), which excludes housing, alcohol, tobacco, durables, transport and vehicles, and miscellaneous. Finally, since the data are drawn from households surveyed throughout the year, expenditure and the marginal wages were adjusted for inflation over the year using a price index constructed from the relevant components of the retail price index.

		Mean	Standard deviation
Female leisure share	S_f	0.328	0.102
Male leisure share	S_m	0.397	0.098
Female hours of work per week	h_f	26.64	11.575
Male hours of work per week	h_m	39.79	4.771
Female net marginal wage (£)	w_f	1.384	0.791
Male net marginal wage (£)	w_m	2.081	0.853
Energy expenditure (£/week)		6.51	6.44
Food expenditure (£/week)		29.55	11.22
Clothing and footwear expenditure (£/week)		11.73	14.25
Services expenditure (£/week)		14.58	32.10
Other goods expenditure (£/week)		10.61	10.15
Number of children, 0–4	n'	0.182	0.461
Number of children, 5–10	n''	0.463	0.737
Number of children, 11–18	n'''	0.458	0.779
Female age	A	37.11	10.48

5

Unemployment and Female Labor Supply

Richard Blundell, John Ham and Costas Meghir

In the basic empirical model of female labor supply (see Heckman 1974; Hausman 1981; Layard, Barton and Zabalza 1980, for example) any woman reporting zero hours of work is assumed not to want to work. Labor supply parameters are estimated from a likelihood function where the probability of an individual recording zero hours is equivalent to the probability of her not having positive desired hours of work. All non-participants in this 'Tobit' model are assumed not to want work.

The assumptions of the Tobit model stand in sharp contrast to those made in calculating the labor force statistics. In particular, those reporting zero hours of work but seeking work are considered to be labor market participants, and a measure of the unemployment rate is often formed from data on such individuals. Our purpose in this paper is to modify the standard model of female labor supply to allow for unemployed workers who want to work at their perceived market wage but cannot find a job.

We believe this extension to the standard labor supply model is important for two reasons. First, it provides a more realistic model of the labor market. Second, in the presence of unemployed workers, the standard Tobit model produces inconsistent estimates of labor supply parameters, since these individuals are mis-specified as non-participants. Thus our

The original version of this chapter was published as: Blundell, R./Ham, J./Meghir, C. (1987). Unemployment and Female Labour Supply, in: The Economic Journal, 97(388a): 44–64. © 1987 by the Royal Economic Society. We should like to thank George Jakubson, Ian Walker and a referee for helpful comments. Javad Emami provided excellent research assistance. Finance for this research was provided by the ESRC under project B0023 2060. All errors remain our responsibility.

work is a natural complement and extension of previous work on unemployment and labor supply, which has focused on prime aged males (e.g. Ham 1982; 1986) or utilized aggregate data (e.g. Ashenfelter 1980).

In any labor supply study, assumptions concerning functional form and the distribution of unobservables play a critical role. In this paper, we are careful to specify a general functional form for preferences and to employ diagnostic tests for distributional assumptions. We do this to minimize the possibility that differences between our model and the standard model represent mis-specification in the standard model unrelated to unemployment.

We should note that the fixed cost literature developed by Cogan (1981) and Hausman (1980) provides an alternative mechanism for relaxing the Tobit model. Thus both the model presented here and the fixed cost model provide an explanation for the results reported in Mroz (1987), who finds that the Tobit model is rejected in favour of a model with an unrestricted specification for the participation decision.[1] In principle, fixed costs can be incorporated into the unemployment model estimated in this paper. This would provide an extremely useful extension for future work. However, combining fixed costs and unemployment in one model is a difficult task, and here we concentrate solely on the unemployment issue.

Allowing for unemployment provides a natural mechanism for the introduction of demand side variables which vary across time and regions.[2] Moreover, the use of such demand side variables generates a fairly direct link between observables and the economic model; this link appears much closer than that between the theory and the variables that we usually have to proxy fixed costs. Finally, the introduction of unemployment into our labor supply model offers the potential to increase considerably the stability of our estimates across different years of data, since the demand variables provide a natural means of controlling for a changing economic environment.

Our application is to a sample of women in the UK Family Expenditure Survey. This survey has the distinct advantage of allowing regional and time variation in the date of interview. It also has been subjected to a number of favorable reliability checks (Atkinson, Micklewright and Stern 1982).

The paper proceeds as follows. In Section 5.1 we review the standard Tobit model of labor supply and participation. In Section 5.2 we generalize the Tobit model to allow for unemployment. Section 5.3 contains our empirical results. We find that the Tobit model is decisively rejected by the data. In Section 5.4 we conclude the paper and discuss topics for further research.

5.1 The Standard Labor Supply Model

In this subsection we briefly review the standard labor supply model where an individual '*i*' is assumed to be able to work as many hours as is desired at her market wage w_i. Given other income y_i, demographic characteristics d_i and unobservable characteristics u_i, desired hours of work are described by some function

$$h_i = g(w_i, y_i, d_i; \beta) + u_i \tag{1}$$

derived from the first order conditions for utility maximization. In (1) β represents the vector of unknown preference parameters and u_i represents unobservable taste variation.

In this model anyone with positive desired labor supply works. It is important for the analysis that follows to note that *conditional* on the wage there are no 'demand side' variables entering the hours or participation decision. Hence the probability of observing zero hours of work is

$$\begin{aligned} Pr(h_i \leq 0) &= 1 - Pr\,(h_i \geq 0) \\ &= 1 - F_i. \end{aligned} \tag{1a}$$

In order to define (1) and (1*a*), one needs to choose an explicit form for h_i and a distribution for u_i. It is quite common (see Heckman 1974; Hausman 1980, for example) to adopt the following normal linear specification

$$h_i = \beta_1 x_i + \beta_2 w_i + u_i \tag{2}$$

where $u_i \sim N(0, \sigma^2)$ and $x_i = (1, y_i, d_i)'$. Although convenient for estimation purposes, the linear model rules out some quite plausible types of labor supply responses (see Stern 1986). In particular, the linear model excludes the joint occurrence of forward sloping and backward bending labor supply behavior. In our empirical work we utilize a more general specification which is consistent with life cycle utility maximization and which allows for considerably greater flexibility in substitution behavior.

Given a general functional form for $g(\cdot)$ in (1), and making a normality assumption on u_i, the probability of observing a woman not working in this model is simply

$$\begin{aligned} Pr(h_i \leq 0) &= Pr[u_i - g(w_i, x_i; \beta)] \\ &= F[-g(w_i, x_i; \beta)/\sigma_i] \\ &= 1 - F[g(w_i, x_i; \beta)/\sigma_i], \end{aligned} \tag{3}$$

where $F(\cdot)$ is the cumulative normal distribution function. We show below that our choice of preference structure and the random taste interpretation of u_i implies a particular form of heteroscedasticity in σ_i. In what follows below we shall simply write the probability of an individual not wanting to work as

$$Pr(h_i \leq 0) = 1 - F(w_i, x_i; \beta, \sigma_i), \tag{4}$$

where $F(\cdot)$ will depend on the particular functional form chosen for $g(\cdot)$.

When w_i is known for all individuals, or when the decision to work is made on the basis of an expected wage which can be predicted perfectly, the likelihood for a random sample is described by the standard Tobit model. Identifying the workers by '+' and the non-participants by '0', the Tobit likelihood is

$$L_1 = \Pi + f(h_i \mid w_i, x_i; \beta, \sigma_i).\Pi_0 [1 - F^R(w_i, x_i, \beta, \sigma_i)], \tag{5}$$

where $f(\ )$ is the normal density for u_i.

Consider, as an alternative, the case where individuals recording zero hours of work know their market wage but the analyst does not. The parameters of a wage equation are now required to specify the sample likelihood. Assume that the wage equation takes the form

$$\log(w_i) = \delta.r_i + r_i, \tag{6}$$

where r_i represents a vector of individual characteristics, 0 represents the corresponding unknown parameters and $T_i \sim N(0, \sigma^2)$. For simplicity we assume T_i to be independent of u_i. The likelihood function is now

$$\begin{aligned} L_1^* = \Pi + f(h_i \mid w_i, x_i; \beta, \sigma_i)\ .\ q(w_i \mid r_i; \delta, \sigma_\tau) \\ \times \Pi_0[1 - F^R(x_i, r_i, \delta, \sigma_i, \sigma_\tau)], \end{aligned} \tag{7}$$

where $q(\cdot)$ is the density of w_i and $F^R(\cdot)$ is the cumulative distribution function of the reduced form equation for h_i given x_i and r_i. In the linear labor supply model with a linear wage equation $F^R(\cdot)$ represents the cumulative normal distribution of $u_i + \beta_2\tau_i$, and (7) involves explicit restrictions between the parameters of $f(\cdot)$, $q(\cdot)$ and $F^R(\cdot)$. For the non-linear labor supply function, such as that considered here, imposing such cross equation restrictions is, as a practical matter, computationally infeasible. Thus in our empirical work we maximize the likelihood (5) using predicted wages for the non-workers from a wage equation based on (6).

5.2 Unemployed Workers

In Section 5.1 we made the standard labor supply assumption that an individual can find employment at her market wage with certainty.

In this section we relax this assumption and consider the case where some individuals who want to work at their market wage do not find employment. We introduce an index function, depending on a variety of macroeconomic and microeconomic factors z_i, that determines whether an individual who wishes to work is in employment. If

$$E_i = \theta.z_i + v_i > 0 \tag{8}$$

the individual would obtain employment whereas if

$$E_i = \theta.z_i + v_i \leq 0 \tag{8'}$$

she would not. The probability of finding employment is then simply

$$P^E(z_i;\theta) = Pr(v_i > -\theta.z_i) = Pr(v_i < \theta.z_i \tag{9}$$

assuming v_i is drawn from a symmetric distribution with mean zero. Since the time at which individuals are interviewed varies (by year or month), elements of z_i representing demand factors in the economy also will vary over time.

It is worth noting that $P^E(\cdot)$ is a reduced form equation and does not represent explicitly a layoff rate or a rate of job arrivals. To see this, consider a simple example of an individual who leaves school at the end of period $t - 2$. Assume that a represents the probability of receiving an acceptable job offer in a given period, that y represents the probability of being laid off from a job in a given period and that if the individual is laid off in a period she must stay unemployed for at least that period. Then there are two ways in which this person can be employed at the end of period t. The first possibility occurs if she finds a job in period $t - 1$ and is not laid off in period t. This event has probability $\alpha(1 - \gamma)$. The second possibility occurs if she does not find a job in period $t - 1$ but does find one in period t. This event has probability $(1 - \alpha)\alpha$. (This assumes that you cannot be laid off in the same period that you accept a job.) Then in period t the probability P^E equals $\alpha(1 - y) + (1 - \alpha)\alpha$ which is a combination of both the lay-off and job arrival rates.

In the presence of unemployment, two groups of individuals will record zero hours of work. The first group consists of those who do not want to work at their market wage. The second group consists of those who want to work but remain unemployed. The first event has probability $(1 - F_i)$ while the second event has probability $(1 - P_i^E)F_i$. The probability of an individual reporting zero hours is then given by

$$(1 - F_i) + (1 - P_i^E)F_i = 1 - P_i^E F_i. \tag{10}$$

In writing down this probability we have assumed a normal distribution for v_i and u_i and independence between their distributions.

However, in our empirical application we test this assumption as well as the other assumptions that affect the consistency of our results. The likelihood for a random sample is the following generalization of the Tobit model

$$L_2 = \Pi + P_i^E f_i . \Pi_0 (1 - P_i^E F_i). \tag{11}$$

This model relates directly to Cragg's Double-Hurdle model (Cragg 1971) and has been considered by Atkinson, Gomulka and Stern (1984) and by Deaton and Irish (1984) in the context of modelling individual household commodity expenditures.

When L_2 is the correct model specification, the standard Tobit model L_1 attributes the wrong probability to unemployed and working women. This will generate inconsistent labor supply parameters. Moreover, by modelling P_i^E it is possible to introduce aggregate and local demand factors directly into the labor market model. The impact of these factors over time and regions provides us with potentially useful information on the degree to which zero hours of work reflect unemployment as opposed to desired non-participation. It is interesting to note that under independence of v_i and u_i the truncated model (Blundell and Walker 1986; Meghir 1985; Blundell and Meghir 1986), which considers positive hours of work only, provides consistent estimates of the labor supply parameters whether L_1 or L_2 is the appropriate likelihood for the full sample.

It is useful to contrast the probability defined in (10) with that in a standard probit equation for participation, and in particular the influence of demand variables in the respective probabilities. The probit equation implicit in (11) takes the form

$$L_p = \Pi + P_i^E F_i \; \Pi_0 (1 - P_i^E F_i) = \Pi + P_i^* \; \Pi_0 (1 - P_i^*) \tag{12}$$

where $P_i{*} = P_i^E F_i = Pr(h_i > 0)$. Demand variables will affect $P_i{*}$ directly through P_i^E, and indirectly through their effect on the wage rate. For example, a negative demand shock may reduce market wages and decrease the probability that a woman has positive desired hours. Secondly, it will reduce the probability that she finds a job. Thus dropping the subscript *i*, we know that

$$\frac{\partial p^*}{\partial x_k} = \frac{\partial F}{\partial w} \frac{\partial w}{\partial x_k} p^E + \frac{\partial p^E}{\partial x_k} F, \tag{13}$$

where x_k is a demand variable that influences labor supply only through the wage. A standard probit equation on participation which includes demand variables captures *both* effects in (13) and does not allow one to separate these effects. Equations (11) and (12), because they incorporate labor supply behavior conditional on market wages, allow one to estimate the two effects in (13) separately. Thus use of (11) allows one to assess the impact of demand variables on the probability of working, above and beyond that due to changing wages. Further, the presence of significant demand effects in $P^E(\cdot)$ is not (at least to a first approximation) compatible with using a fixed cost model to relax the Tobit specification since demand variables should enter P^* in a fixed costs model only through the wage effect on labor supply.

5.3 Empirical Results

The empirical results presented here refer to a sample of 2011 married women from the 1981 UK Family Expenditure Survey. Precise details of the sample selection and information on each of the variables used are provided in the Data Appendix. The sample split is 1076 recording positive hours and 935 recording zero hours.

As noted above we believe that it is crucial to provide a general specification of preferences. Otherwise, one faces the problem that allowing for unemployment 'makes a difference' simply because the assumed hours equation is too restrictive. The overall specification of the hours of work equation is taken from Blundell and Meghir (1986). That study was concerned only with individuals currently in work and used the truncated model. The labor supply model is a generalization of the Linear Expenditure System that possesses a number of desirable features. In particular, it allows for quite general substitution effects and demographic interactions. The specification is derived from the following indirect utility function defined over wages (w), prices (p), other income (y) and is conditional on several demographic (and taste shifter) variables (d)

$$V(w,p,y;d) = \frac{y + a(w,p;d)}{b(w,p;d)}. \tag{14}$$

Wage rates for participants refer to marginal rates after accounting for the impact of the tax and benefit system using the procedure described in Blundell et al. (1986). Wage rates for nonparticipants are derived from the wage equation presented in Appendix A.[3] The other income variable y is defined as the difference between expenditures and net earnings to ensure the life cycle consistency of the estimated labor supply model (see Blundell and Walker 1984). This variable may be negative or positive although $y + a(\cdot)$ is strictly positive for all individuals in our estimated model. Choosing the form for $a(w, p; d)$ to be

$$a(w,p;d)=\alpha_{ff}(d)w-2\alpha_{fc}(d)w^{½}p^{½}-\alpha_{cc}(d)p \tag{15}$$

and by writing

$$\ln b(w,p;d)=\beta_f(d)\ \ln w+[1-\beta_f(d)]\ \ln p \tag{16}$$

we ensure the zero homogeneity of $V(\cdot)$ in w, p and y. As economic theory tells us very little about the way demographic characteristics affect preferences over paid work and consumption, we shall allow the parameters of $V(\cdot)$ to depend fairly generally on the demographic variables d.

Using Roy's identity, the corresponding hours of work equation takes the form

$$h=\alpha_{ff}(d)-\alpha_{fc}(d)p^{½}/w^{½}-\beta_f(d)\left[y+a(w,p;d)\right]/w \tag{17}$$

The specifications of $a(w, p; d)$ and $b(w, p; d)$ in (15) and (16) ensure a wide variety of shapes for this function. For example, when α_{fc} is zero, preferences are represented by the popular Stone-Geary utility function and generate a linear earnings function corresponding to the Linear Expenditure System. The Stone-Geary model provides a useful illustration of the restrictiveness of simple models of labor supply since the sign of the wage coefficient (somewhat crucial in policy analysis) is determined by the signs of $(y-\alpha_{cc}p)$ and β_f. If leisure is a normal good, β_f is positive, the labor supply curve is forward sloping everywhere for $y>\alpha_{cc}p$ and backward bending everywhere for $y<\alpha_{cc}p$. Our more general specification allows for both forward and backward bending labor supply to coexist independently of y. The flexibility of our general form is probably best illustrated through the wage derivative

$$\frac{\partial h}{\partial w}=\frac{1}{2}\propto_{fc}(1-2\beta_f)p^{\frac{1}{2}}/w^{\frac{3}{2}}+\beta_f(y-\alpha_{cc}p)/w^2. \tag{18}$$

In the additive Linear Expenditure System specification, the wage derivative consists only of the second term.

In order to allow a random preference interpretation to the stochastic specification of the labor supply model derived in the previous section, we introduce random variation in the β_f parameter. This parameter represents the marginal budget share of non-market time or the marginal value of time. We assume that a normal distribution approximates unobservable taste variation; this assumption is tested below. In modelling unobservable random preference errors, we cannot simply add a homoscedastic error to (17), since the model will no longer integrate back to the underlying indirect utility function. After suitable rearrangement, our labor supply function for individual i corresponding to (1) with random unobservable taste errors may be written

$$h_i = \alpha_{ff}(d_i) - \alpha_{fc}(d_i)p^{1/2} / w_i^{1/2} - \left[\beta_f(d_i) - \mu_i\right]\left[y_i + a(w_i, p_i; d_i)\right] / w_i, \quad (19)$$

where $\mu_i \sim N(0, \sigma^2)$. Hence u_i in (1) takes the form $\mu_i[y_i + a(\cdot)]/w_i$.

It is important to subject the distributional assumptions on μ_i (the labor supply disturbance) and v_i (the disturbance on the employment equation) to diagnostic testing. Following Bera, Jarque and Lee (1984) and Blundell and Meghir (1986), we use the Pearson family to derive score tests against skewness and kurtosis. These are equivalent to the score tests using generalized residuals developed in Chesher and Irish (1984) and Gourieroux et al. (1984).[4] A similar score test based on generalized residuals is used to test independence between u_i and v_i.

As described above demographic characteristics are allowed to enter through both the α and β parameters. That is, we specify

$$\begin{aligned}
\alpha_{ff}(d) &= \propto_{ff}^0 + n_1 \propto_{ff}^1 + D_2 \propto_{ff}^2 + D_3 \propto_{ff}^3, \\
\alpha_{fc}(d) &= \propto_{ff}^0 + n_1 \propto_{fc}^1 + \sqrt{(n_2 D_2)} \propto_{fc}^2 + \sqrt{(n_3 D_3)} \propto_{fc}^3, \\
\alpha_{cc}(d) &= \propto_{cc}^0 + n_1 \propto_{cc}^1 + n_2 \propto_{cc}^2 + n_3 \propto_{cc}^3, \\
\beta_f(d) &= \beta_{ff}^0 + \beta_f^1 D_1 + \beta_f^2 D_2 + \beta_f^3 D_3 + \beta_f^A AGE \\
&\quad + \beta_f^{AA} AGE^2 + \beta_f^{ed} education,
\end{aligned}$$

where n_1 = number of children with $0 < \text{age} < 2$,
n_2 = number of children with $3 \le \text{age} < 10$,
n_3 = number of children with $11 \le \text{age} < 18$,

and where D_i are conditional dummies such that

$D_1 = 1$ if $n_1 > 0$, zero otherwise,
$D_2 = 1$ if $n_2 > 0$, and $n_1 = 0$, zero otherwise,
$D_3 = 1$ if $n_3 > 0$, and $n_1 = n_2 = 0$, zero otherwise.

As described in the data appendix, age is the female age while education is her age at the end of schooling. The conditional dummies D_1, D_2 and D_3 are used in some parts of the specification rather than the actual number of children to account for economies of scale in the time requirement of children.

Given the specification of the hours equation, we can turn to the estimation of the standard Tobit model L_1 described in Section 5.1. This model does not allow for the possibility of unemployment and therefore does not require a specification for the employment probability index function $P^E(\cdot)$. Model L_1 is simply based on the desired hours equation of (19) and the reservation wage condition. The wage equation is provided in Appendix A and the model estimates are presented in Table 5.1.

Table 5.1. Parameter Estimates[a] and Diagnostics for the Tobit Model L_1

α_{ff}^{0}	38.466	(2.878)	$\beta_f^{AA} \times 100$	0.046	(0.008)
α_{ff}^{1}	11.296	(13.978)	β_f^{1}	0.544	(0.079)
α_{ff}^{2}	–4.929	(4.677)	β_f^{2}	0.219	(0.044)
α_{ff}^{3}	–1.147	(3.997)	β_f^{3}	0.086	(0.043)
α_{fc}^{0}	–8.909	(3.403)	β_f^{ed}	–0.014	(0.004)
α_{fc}^{1}	1.037	(11.821)	β_f^{servs}	0.093	(0.018)
α_{fc}^{2}	–6.803	(3.345)	β_f^{manu}	–0.025	(0.017)
α_{fc}^{3}	–2.720	(2.249)	σ	0.274	(0.012)
α_{cc}^{0}	31.065	(3.134)	Log L	–5201.68	
α_{cc}^{1}	–6.383	(10.218)	Skewness (I)[a]	85.345	
α_{cc}^{2}	2.199	(2.170)	Kurtosis (I)	2.547	
α_{cc}^{3}	10.017	(1.242)	–	–	
β_f^{0}	0.338	(0.0318)	–	–	
$\beta_f^{A} \times 10$	0.054	(0.0085)	–	–	

Note: [a] Standard errors in parentheses throughout the following Tables.
[b] Under the null hypothesis all test statistics are distributed as χ^2 with degrees of freedom as indicated in parentheses.

On the whole the α and β estimates appear quite plausible. The β_f parameters are generally well determined and show an expected (albeit quite large) increase in the value of female time with age and young children. However, turning to the diagnostic tests we see that normality is very strongly rejected through the skewness test. From the results in Blundell and Meghir (1986), it is interesting to note that skewness often picks up general model mis-specification. To provide an insight into which directions this standard labor supply model may be misleading we turn next to the general 'unemployment' specification based on (11).

The labor supply estimates in Table 5.2a refer to this more general unemployment or 'Double-Hurdle' model and stand in contrast to the simple Tobit model estimates. The estimates of the parameters of the employment probability index function $P^E(\cdot)$ are provided in Table 5.2b. In specifying the arguments of this probability function we have tried to allow for both individual and demand side characteristics. Precise definitions of the variables are given in the Data Appendix.

Turning first to the parameters of this probability index in Table 5.2b, we find a strong impact of the demand side characteristics directly related to the female both through the age specific unemployment

Table 5.2. a) Labor Supply Estimates and Diagnostics for Model L_2

α_{ff}^0	29.328	(2.104)	$\beta_f^{AA} \times 100$	0.013	(0.005)
α_{ff}^1	–20.112	(7.495)	β_f^1	0.256	(0.067)
α_{ff}^2	–15.441	(4.677)	β_f^2	0.095	(0.032)
α_{ff}^3	–2.838	(3.997)	β_f^3	0.079	(0.033)
α_{fc}^0	–16.242	(3.403)	β_f^{ed}	–0.007	(0.003)
α_{fc}^1	–27.527	(11.821)	β_f^{servs}	0.082	(0.011)
α_{fc}^2	–16.591	(3.345)	β_f^{manu}	0.059	(0.011)
α_{fc}^3	–6.342	(2.249)	σ	0.156	(0.008)
α_{cc}^0	29.206	(3.134)	Log L	–4835.760	
α_{cc}^1	26.630	(10.218)	Skewness (I)[a]	5.267	
α_{cc}^2	4.520	(2.170)	Kurtosis (I)	9.192	
α_{cc}^3	11.968	(1.242)	–	–	
β_f^0	0.161	(0.0318)	Independence (I)	20.896	
$\beta_f^A \times 10$	0.025	(0.0085)			

b) Estimated Parameters of the Employment Probability Index for Model L_2

Intercept	3.3372	(0.8467)
Education (female)	–0.0249	(0.0291
Age (female)	–0.0319	(0.0068)
Regional unemployment	–0.0132	(0.0249)
Unemployment by age (female)	–0.0675	(0.0197)
Unemployment by industry (female)	–0.3218	(0.0206)
Vacancy by region	–0.0591	(0.1162)
Redundancy by region	–0.0030	(0.0284)
Vacancy by industry (female)	0.9149	(0.0722)
Redundancy by industry (female)	0.0055	(0.0031)
D_1	–1.5060	(0.1495)
D_2	–0.5869	(0.1162)
London	0.2241	(0.2450)
Other south	0.1050	(0.1673)
Manual worker (female)	–0.5466	(0.1467)
Services industry (female)	–5.2387	(0.3856)
Unemployment by industry (male)	0.0217	(0.0090)
Manual worker (male)	–0.0932	(0.1099)
Services industry (male)	0.1461	(0.1240)
Skewness (i)	41.086	
Kuirtosis (I)	14.479	

c) Predictive Performance

	(i) Tobit	
	Actual	
Predicted	Non-workers	Workers
Non-workers	601	210
Workers	335	866
	(ii) Double-Hurdle	
Non-workers	822	203
Workers	114	873

rate, the industry specific unemployment rate and the industry specific vacancy rate. Among those non-participants that record an industry (the majority of these also report that they are actively seeking work) the vacancy rate is on average 20% lower and the redundancy rate is 50% higher than that for participants. Interestingly, there is also a significant direct effect of young children (D_1 and D_2) possibly reflecting the importance of fixed costs associated with child care (not observed in our data) or alternatively the impact of recently having been out of the labor market. Both age and the industry and occupation specific dummies are significant.[5] The range of $P^E(\cdot)$ for non-working women in the sample is 0.031-0.993 with a mean and standard deviation of 0.313 and 0.236 respectively while for workers the corresponding figures are 0.003-0.999, 0.822 and 0.254 respectively. The standard 'Tobit' labor supply model assumes that this probability is identically unity for all individuals and on the basis of a likelihood ratio test ($-2 \log (L_1/L_2) = 732.26$) the Tobit model is strongly rejected. In Table 5.2c, a classification of individuals according to whether their predicted probability is higher as a non-participant or participant points also to a superior performance from the Double-Hurdle model. In this classification an individual is predicted to work if P_i^* from (12) exceeds 0.5. For the Tobit model this reduces to exceeding 0.5.

The impact of introducing this employment probability function on the determination of hours of work and desired participation is quite dramatic. Many of the labor supply parameters in Table 5.2a have changed significantly from those reported for the Tobit model in Table 5.1. The additivity assumption behind the LES preference specification (measured by non-zero β_f^1 terms) is now more strongly rejected. Moreover, the impact of young children, although still very important on participation through a combination of effects on F_i and P_i^E, has reduced its impact through the β_f^1 term from the rather large figure in Table 5.1. The diagnostic tests now appear more reasonable - the skewness test statistic has fallen dramatically. The independence test on the stochastic processes determining the two hurdles is quite large and would suggest re-estimation under dependence.

Estimates for the Double-Hurdle under dependence are presented in Appendix B and display broadly similar properties to those in Table 5.2. The likelihood ratio test and the Wald test for a zero correlation ($p = 0$) between the stochastic terms in the two hurdles takes the values 20.14 and 22.79 respectively. As in the case of the score test this implies a rejection of independence. If the two latent variables underlying the Double-Hurdle model were both continuously observed then a non-zero correlation would only affect the efficiency of the estimates. However, in this model with the two hurdles jointly determining the observability of working hours, consistency will be affected. Despite this, it is apparent from Appendix B that the labor supply parameters change only marginally moreover, the significant parameters of the employment probability index remain of the same sign and display a similar pattern to those reported in Table 5.2b.[6]

In Table 5.3 we present a description of the distribution of the wage elasticities (for non-market time) over the whole sample in the estimated Tobit and Double-Hurdle models from Tables 5.1 and 5.2a. Participants are evaluated at their actual labor supply hours by including the estimated random preference error in the marginal value of time β_f For non-participants a random preference error is drawn from the appropriately truncated normal distribution.

Despite the fact that we use the same flexible preference structure in each case, the Double-Hurdle model displays considerably more backward bending behaviour, denoted by a positive elasticity in Table 5.3. These results could have important implications for policy analysis in

Table 5.3. Wage Elasticities (None Market Time)[a]

	Participants		Non-Participants	
	Tobit	Double-Hurdle	Tobit	Double-Hurdle
Mean	0.040	0.087	−0.207	−0.095
S.D.	0.147	0.125	0.116	0.123
Minimum	−1.299	−1.937	−0.671	−0.593
M_{25}	−0.045	0.013	−0.270	−0.151
M_{50}	0.40	0.096	−0.189	−0.095
M_{75}	0.122	0.155	−0.128	−0.038
Maximum	1.254	0.541	0.180	1.460
Proportion positive	0.610	0.784	0.010	0.153

Note: [a] M_i is the *i*th percentile.

general and in particular for the analysis of the incentive effects of reforms to the tax and benefit system. These differences occur because the Tobit model forces the reservation wage to exceed the market wage for all non-participants and misallocates the observations on non-participants.

We also can contrast the Tobit and Double-Hurdle models through the effect that demand side variables have on actual employment. While the Tobit model accounts for demand side effects only via their influence on the wage rate, in the Double-Hurdle model these variables also have a direct impact through the first hurdle employment probability index P^E. The effects may counteract or reinforce each other, and thus it is of some interest to consider the two effects explicitly and analyze their relative size and magnitude. Noting that log $[pr(h > 0)] = \log[P^E(\cdot)] + \log[F(\cdot)]$ the elasticity of the positive hours probability with respect to some demand side variable z_k can be conveniently decomposed as

$$\frac{\partial \log pr(h>0)}{\partial \log z_k} = \frac{\partial \log F}{\partial \log w}\frac{\partial \log w}{\partial \log z_k} + \frac{\partial \log p^E}{\partial \log z_k}, \tag{20}$$

where the first term on the right hand side is the elasticity of participation while the second term is the elasticity of the employment probability index. The Tobit model sets this second term to zero. Rewriting the labor supply model (1) as

$$h_i = g(w_i, y_i, d_i; \beta) + \mu_i[y_i + a(\cdot)] / w_i \tag{21}$$

allows the desired participation probability to be expressed as

$$F(\cdot) = Pr\{\mu_i < w_i g(w_i, y_i, d_i; \beta) / [y + a(\cdot)]\} \tag{22}$$

Consequently, the derivative of this participation probability is given by

$$\begin{aligned}\partial F(\cdot)/\partial w_i = \phi(\cdot)\{&w_i \partial g(\cdot)/\partial w_i + g(\cdot) - w_i g(\cdot)[y_i + a(\cdot)]^{-1}[\partial a(\cdot)/\partial w_i]\}\\ &[y_i + a(\cdot)]^{-1}\end{aligned} \tag{23}$$

Here $\phi(\cdot)$ is the standard normal density function evaluated at the same point as $F(\cdot)$ and $\partial g(\cdot)/\partial w_i$ is defined by the labor supply derivative (18). From (23) it is clear that the sign of the first term on the right-hand side of (20) is not necessarily positive for those with positive desired hours, giving rise to the possibility that the wage derivative of the participation probability is negative for some workers.[7]

In Tables 5.4a,b we present a description of the distribution of these elasticities over all participants and non-participants in our sample for two important demand side variables: female unemployment by age

Table 5.4. Employment Elasticities With Respect to Demand Side Variables.

a) Female Unemployment by Age

	Participation	pE	Total
	(i) Participants		
Double–Hurdle			
Mean	–0.003	–0.148	–0.151
S.D.	0.039	0.247	0.252
M_{25}	–0.001	–0.172	–0.175
M_{50}	0.000	–0.047	–0.049
M_{75}	0.000	–0.014	–0.016
Tobit			
Mean	–0.009	0	–0.009
S.D.	0.041	0	0.041
M_{25}	–0.010	0	–0.010
M_{50}	–0.001	0	–0.001
M_{75}	0.000	0	0.000
	(ii) Non Participants		
Double-Hurdle			
Mean	–0.003	–0.680	–0.682
S.D.	0.007	0.508	0.508
M_{25}	–0.004	–0.951	–0.951
M_{50}	–0.001	–0.525	–0.530
M_{75}	0.000	–0.291	–0.293
Tobit			
Mean	–0.037	0	–0.037
S.D.	0.041	0	0.041
M_{25}	–0.052	0	–0.052
M_{50}	–0.021	0	–0.021
M_{75}	–0.008	0	–0.008

b) Female Unemployment by Industry

	Participation	pE	Total
	(i) Participants		
Double-Hurdle			
Mean	–0.005	–0.985	–0.990
S.D.	0.071	2.084	2.087
M_{25}	–0.003	–1.048	–1.043
M_{50}	0.000	–0.220	–0.227
M_{75}	0.000	–0.072	–0.075
Tobit			
Mean	–0.016	0	–0.016
S.D.	0.074	0	0.074
M_{25}	–0.018	0	–0.013
M_{50}	–0.001	0	–0.001
M_{75}	0.000	0	0.000

(continued)

Table 5.4. (b) (Continued)

	Participation	pE	Total
	(ii) Non Participants		
Double-Hurdle			
Mean	–0.007	–3.346	–3.853
S.D.	0.014	1.633	1.632
M_{25}	–0.010	–5.227	–5.237
M_{50}	–0.003	–3.654	–3.671
M_{75}	0.000	–2.942	–2.951
Tobit			
Mean	–0.068	0	–0.063
S.D.	0.062	0	0.062
M_{25}	–0.100	0	–0.100
M_{50}	–0.050	0	–0.050
M_{75}	–0.022	0	–0.022

and female unemployment by industry. The effect of the demand side variables on the wage rate was obtained by a regression of the (gross) wage rate which controls only for the variables appearing in the employment probability index P^E. The column 'participation' refers to the first term on the right hand side of (20), the column P^E refers to the second term on the right hand side of the column 'Total' to the left hand side of (20). In the Double-Hurdle model most of the impact of the demand side variables comes through their direct effect on the first hurdle probability P^E. The indirect effect which works through the wage rate is quite small. In contrast the Tobit model sets the direct effect to zero and forces the demand variables to work entirely through the wage rate. For this reason the Tobit model gives a misleading picture of both the channels through which demand side variables affect the number of people in employment and the magnitude of these effects. The Double-Hurdle model allows us both to obtain a more accurate picture of the *total* effect of a change of a macro variable on the expected numbers of people in work as well as to decompose the effect into a direct one via the employment probability index and an indirect one via the wage in the labor supply function.

5.4 Conclusions

In this paper we discuss ways of extending the standard model of labor supply and participation. In particular we introduce the possibility that zero hours of work represent unemployment and not simply non-participation. This was achieved by modelling the probability that a particular individual obtains a job at his/her perceived market wage. The resulting Double-Hurdle model offers a framework for studying the effects of macroeconomic variables on labor force participation. Moreover, the model tests the traditional Tobit model and hence the hypothesis underlying the basic labor supply model can easily be tested using the likelihood ratio test.

We apply our proposed methodology to a sample of married women drawn from the UK Family Expenditure Survey of 1981. We find that demand variables have an important impact in the P^E function, which in turn plays a role in determining the probability of working. Moreover, the Tobit model is strongly rejected using a likelihood ratio test and the non-normality diagnostics for the labor supply model improve considerably when this model is generalized.

Although the Double-Hurdle model presented here seems to offer an improved framework over the Tobit model for analyzing both participation and hours of work, there is room for improvement in future research. Firstly, the model would benefit from qualitatively better information relating to demand side constraints. This is suggested by the relatively large normality statistic reported in Table 5.2 (b) that points to some remaining mis-specification. Secondly, in future, it would be fruitful to exploit the information separating the non-workers into job seekers and non-seekers and explicitly accounting for discouraged workers. Thirdly, our approach may offer a useful framework for analyzing the stability of the parameter estimates from several years of FES data. Finally, we could investigate the challenging problem of incorporating fixed costs and unemployment in a single model.

Data Appendix

The data were drawn from the Family Expenditure Survey 1981. The following sample selection was applied:

(a) Occupation: FES variable A2I0 with value 2, 3, 4, 5, 6, 7, 8.
(b) Age: women with $16 \leq \text{Age} < 60$ men with $16 \leq \text{Age} < 65$.
(c) Hours: female hours ≤ 80, male hours ≤ 80.
(d) Two adult households with the two adults being a married couple.

Sample Size:
Households with working women 1076 (A201 = 1).
Households with non-working women 935 (A201 = 2, 3, 4, 5, 7).

Analysis for Households with Working Women

Variable	Mean	Standard deviation	Minimum	Maximum
Female hours	25.961	11.835	1	80.00
Male hours	38.762	3.939	9.00	77.00
Children				
0–2	0.068	0.269	0	2
3–4	0.064	0.259	0	2
5–10	0.437	0.709	0	3
11+	0.474	0.767	0	4
Female wage	1.363	1.206	0.012	10.817
Male wage	2.022	0.747	0.295	5.392
Expenditure	71.601	34.307	18.748	331.093
Other Income	39.292	34.682	−37.025	232.761
Female age	36.930	10.554	17.000	59.0
Education (female)	15.674	1.959	14.000	24.0

Note: Wages are in £/hour. This unit of measurement was used in estimation. The low minimum wage reflects the impact of the tax and benefit system. Education: Female age at the end of education (FES Variable A010).

Breakdown of Households by Numbers and Age of Children

Age	Number of children				
	0	1	2	3	4
0–2	1008	63	5	0	0
3–4	1011	61	4	0	0
5–10	735	219	114	8	0
11+	728	205	125	17	1

Breakdown of Households by Number of Children and Female Hours

			Female hours		
Children (age < 3)	1–10	11–20	21–30	31–40	41+
0	115	265	178	426	24
1	21	30	5	5	2
2	1	4	0	0	0
			Female hours		
Children (age 3–4)	1–10	11–20	21–30	31–40	41+
0	113	276	178	421	23
1	20	23	5	10	3
2	4	0	0	0	0
			Female hours		
Children (age 5–10)	1–10	11–20	21–30	31–40	41+
0	59	151	129	373	23
1	45	97	35	41	1
2	31	47	18	16	2
3	2	4	1	1	0
			Female hours		
Children (age > io)	1–10	11–20	21–30	31–40	41+
0	1001	116	114	327	20
1	20	86	39	57	3
2	14	41	28	40	2
3	1	6	2	7	1
4	1	0	0	0	0

Female Occupation

Occupation	Frequency
Clerical & Professional	510
Shop Assistants	79
Skilled Manual	60
Semi-Skilled Manual	278
Unskilled Manual	149

Female Industry by 32 FES Sub Grouping

FES Code	Frequency	FES Code	Frequency
1	6	17	5
2	3	18	27
3	37	19	9
4	2	20	16
5	13	21	10
6	5	22	1
7	13	23	41
8	3	24	179
9	24	25	84
11	20	26	245
12	9	27	18
13	14	28	170
14	2	30	33
15	32	31	5
16	7	32	40

Regional Breakdown

Region	Frequency
Northern	67
Yorks	103
East Midlands	91
East Anglia	43
Greater London	115
South East	196
South West	61
Wales	48
West Midlands	117
North Western	130
Scotland	105

Analysis for Households with Non–working Wives (A201 = 2, 3, 4, 5, 7)

Variable	Mean	Standard deviation	Minimum	Maximum
Male hours	35.906	14.065	0	80
Children				
0–2	0.475	0.635	0	4
3–4	0.251	0.465	0	2
5–10	0.525	0.786	0	4
11+	0.344	0.711	0	4
Female wage	1.254	0.389	0.339	5.307
Male wage	2.227	1.040	0.149	6.294
Expenditure	65.398	36.175	14.817	390.631
Female age	34.667	10.853	16.0	59.0
Education	15.634	2.177	14.0	24.0

Breakdown of Households by Numbers and Age of Children

	Number of children				
Age	0	1	2	3	4
0–2	560	308	66	0	1
3–4	713	209	13	0	0
5–10	588	227	100	16	4
11 +	722	123	74	13	3

Regional Breakdown

Region	Frequency
Northern	68
Yorks	105
East Midlands	75
East Anglia	42
Greater London	61
South East	163
South West	54
Wales	59
West Midlands	91
North West	99
Scotland	18

Appendix A

Dependent Variable: Log (gross wage rate in tenth of pence/hour)

	Variable	Parameter	S.E.
1.	Intercept	5.271	0.903
	Children		
2.	(0–2)	–0.182	0.132
3.	(3–4)	–0.101	0.077
4.	(5–10)	–0.069	0.026
5.	(11+)	–0.036	0.019
6.	Regional unemployment[a]	0.052	0.030
7.	London	0.596	0.217
8.	South East	0.485	0.241
9.	Yorkshire	0.260	0.122
10.	East Midlands	0.360	0.192
11.	East Anglia	0.407	0.240
12.	South West	0.340	0.180
13.	Wales	0.008	0.052
14.	West Midlands	0.184	0.86
15.	North West	0.063	0.061
	Tenure		
16.	Dummy 1[b]	–0.024	0.023
17.	Dummy 2	–0.019	0.043
	Male occupation		
18.	Dummy 1[c]	–0.045	0.036
19.	Dummy 2	–0.083	0.028
20.	Dummy 3	–0.061	0.032
21.	Male age	0.026	0.012
22.	Male age squared	–0.00028	0.0001
23.	Male education	0.00074	0.0262
24.	Male education squared	–0.00023	0.0008
25.	Female education	–0.087	0.032
26.	Female education squared	–0.002	0.001
27.	Female age	–0.043	0.019
28.	Female age squared	0.00016	0.0002
	Conditional child		
29.	Dummy 1[d]	0.064	0.151
30.	Dummy 2	–0.109	0.099
31.	Dummy 3	–0.037	0.049
32.	Dummy 4	–0.086	0.042
	Male industry		
33.	Dummy 1[e]	0.0503	0.0307
34.	Dummy 2	0.0119	0.068
35.	Dummy 3	0.0396	0.031
36.	Dummy 4	0.016	0.032
	Female industry		
37.	Dummy 1[f]	0.069	0.075
38.	Dummy 2	–0.072	0.085
39.	Dummy 3	–0.036	0.053
40.	Dummy 4	–0.484	0.079

(continued)

Dependent Variable: Log (gross wage rate in tenth of pence/hour) (Continued)

	Variable	Parameter	S.E.
	Female occupation		
41.	Dummy 1[g]	–0.333	0.032
42.	Dummy 2	–0.398	0.052
43.	Dummy 3	–0.514	0.034
44.	January	0.045	0.090
45.	February	–0.023	0.086
46.	March	0.046	0.079
47.	April	–0.025	0.075
48.	May	–0.019	0.067
49.	June	–0.039	0.056
50.	July	–0.075	0.048
51.	August	–0.091	0.041
52.	September	–0.063	0.040
53.	October	–0.013	0.042
54.	Female unemployment by age	–0.011	0.008
55.	Female unemployment by industry	–0.015	0.005
56.	Vacancy (region)	0.162	0.073
57.	Redundancy (region)	0.026	0.014
58.	Vacancy (industry/female)	0.052	0.013
59.	Redundancy (industry/female)	–0.001	0.001
60.	Male unemployment (age)	–0.006	0.004
61.	Female age x education	0.002	0.001

Definitions:
Vacancy (redundancy) by region: Proportion of vacancies (redundancies) to regional population per/1000.
Vacancy (redundancy) by Industry[h]: Proportion of vacancies (redundancies) to the employee in industry per/1000.
Source for population: Annual Abstract of Statistics (1984).
Source for vacancies & redundancies: Employment Gazette (1983).

Notes:
[a] Male unemployment by region and month.
[b] Tenure dummy 1: council rented accommodation.
Tenure dummy 2: other rented accommodation.
[c] Occupation dummy 1: clerical workers and shop assistants.
Occupation dummy 2: skilled manual worker.
Occupation dummy 3: semi-skilled and unskilled manual worker.
[d] Conditional child dummies constructed as described in the text for the four age groups.
[e] Male industry dummies:
1. FES (1981) industry codes 3–12.
2. FES (1981) industry codes 13–15.
3. FES (1981) industry codes 16–23.
4. FES (1981) industry codes 24–28.
[f] Female industry dummies as in footnote e above.
[g] Female occupation dummies as in footnote c above.
[h] For those non-workers where industry was not recorded the weighted average figure was used.

Appendix B

Table B. a) Labor Supply Estimates and Diagnostics For Mode L_2 with Dependence

α_{ff}^{0}	28.527	(2.105)	$\beta_f^{AA} \times 100$	0.012	(0.006)
α_{ff}^{1}	–19.770	(7.502)	β_f^{1}	0.231	(0.069)
α_{ff}^{2}	–14.767	(3.371)	β_f^{2}	0.101	(0.033)
α_{ff}^{3}	–3.306	(3.551)	β_f^{3}	0.076	(0.034)
α_{fc}^{0}	–17.036	(2.790)	β_f^{ed}	–0.005	(0.003)
α_{fc}^{1}	–27.510	(7.610)	β_f^{servs}	0.082	(0.011)
α_{fc}^{2}	–16.959	(3.043)	β_f^{manu}	0.068	(0.012)
α_{fc}^{3}	–6.494	(2.698)	σ	0.159	(0.009)
α_{cc}^{0}	30.198	(2.318)	r	–0.335	(0.070)
			Log L	–4825.688	—
α_{cc}^{1}	25.612	(8.048)	—	—	—
α_{cc}^{2}	5.152	(2.301)	—	—	—
α_{cc}^{3}	12.118	(1.638)	—	—	—
β_f^{0}	0.146	(0.025)	—	—	—
$\beta_f^{A} \times 10$	0.023	(0.007)	—	—	—

b) Estimated Parameters of the Employment Probability Index For Mode L_2 With Dependence

Intercept	3.0344	(0.8163)
Education (female)	–0.0156	(0.0274)
Age (female)	–0.0304	(0.0067)
Regional unemployment	–0.0089	(0.0238)
Unemployment by age (female)	–0.0646	(0.0194)
Unemployment by industry (female)	–0.3189	(0.0199)
Vacancy by region	–0.0406	(0.1109)
Redundancy by region	0.0110	(0.0268)
Vacancy by industry (female)	0.8568	(0.0705)
Redundancy by industry (female)	–0.0079	(0.0030)
D_1	–1.4534	(0.1550)
D_2	–0.5338	(0.1165)
London	0.2951	(0.2348)
Other south	0.1033	(0.1608)
Manual worker (female)	–05015	(0.1446)
Services industry (female)	–4.8604	(0.3830)
Unemployment by industry (male)	0.0203	(0.0087)
Manual worker (male)	–0.0964	(0.1053)
Services industry (male)	0.1331	(0.1186)

6

Estimating Labor Supply Responses Using Tax Reforms

Richard Blundell, Alan Duncan, and Costas Meghir

6.1 Introduction

The large number of tax policy reforms in the UK over the 1980s provides an ideal opportunity to evaluate labor supply responses. Indeed, since some working individuals will have been exempt from any direct impact of these reforms due to the progressive nature of the tax system, it may be thought that a control group suitable for evaluating reforms over time could be constructed.

Labor supply effects have been notoriously difficult to estimate in a robust and generally accepted way.[1] The difficulties that researchers typically face relate to the treatment of (non-linear) tax schedules, the fact that individuals have different tastes over non-market time and consumption for reasons that cannot be controlled for using observable information, and the fact that individuals' observed decisions represent intertemporal allocations as well as within period allocations. These issues lead to difficult simultaneity problems with the wage rate and other household income. Thus for example, all else

The original version of this chapter was published as: Blundell, R./Duncan, A./Meghir, C. (1998). Estimating Labor Supply Responses Using Tax Reforms, in: Econometrica, 66(4): 827–62. © 1998 by the Econometric Society. We thank two anonymous referees and a co-editor for detailed comments. We are grateful to Joe Altonji, Andrew Dilnot, Amanda Gosling, James Heckman, Paul Johnson, Tom MaCurdy, Thierry Magnac, Ian Preston, John van Reenen, and seminar participants at Berkeley, Chicago, Harvard, INSEE, Northwestern, Stanford, and Warwick for useful comments. This study is part of the research program of the ESRC Centre for the Microeconomic Analysis of Fiscal Policy at the IFS. Data from the FES made available from the CSO through the ESRC Data Archive has been used by permission of the controller of HMSO. We are responsible for all errors and interpretations.

being equal, "hard workers" will be facing higher marginal tax rates and hence lower hourly wages. This biases wage effects downwards. Instrumental variables based on arbitrary exclusion restrictions (such as excluding education) may provide no solution since these variables are probably correlated with tastes for work. However, tax reform can lead to exogenous changes in after-tax wages and incomes. Thus, the potential for direct evaluation of labor supply effects based on a comparison of responses over time by groups of individuals affected differentially by the reforms is evident.

Our analysis concerns the labor supply responses of married or cohabiting women. In the UK, the weekly hours of work distribution for this group is very dispersed with individuals observed working anything from 1 to over 60 hours a week. Individuals in this group may be expected, more than any other group, to be able to change hours of work in response to changing economic conditions or to changes in household composition. Our idea is to combine a structural approach together with instrumental variables, exploiting the variability induced by the tax reforms and the changing wage structure in order to circumvent the simultaneity problems mentioned above. The structural side of the analysis is crucial; this allows us to distinguish between income and substitution effects which are at the center of the policy debate on incentives and on welfare effects of taxation. Our model will be consistent with life cycle behavior and the estimation method will allow for the presence of fixed costs of work.

Thus in this paper we derive the conditions on grouping estimators required for the identification and estimation of wage and income elasticities. We relate this to the standard difference of differences approach and consider whether grouping according to tax status itself is likely to provide a reliable guide to labor supply responses. For these purposes the UK tax system has the advantage that it is quite simple, with most people being either basic rate taxpayers or non-taxpayers (because their earnings are below an exogenously given threshold). We argue that composition changes between these two groups, partly induced by the changes in tax policy itself, invalidate grouping according to taxpayer status.

Our identification strategy relies on comparing the labor supply responses over time for different groups defined by cohort and education level. Thus our approach exploits the differential growth of marginal wages between these groups. These differential changes reflect both the differential impact of the reforms on these groups as well as the differential growth in real wages; the latter is due to the well documented increases in the returns to education and to the cohort effects on the wage distribution.

Hours of work is just one aspect of work behavior. Another one is, of course, labor force participation, on which the tax reforms and the change in wage structure are likely to have important effects. Thus, in the presence of fixed costs of work or other factors that differentiate the participation model from labor supply, our behavioral elasticities cannot give the complete picture of the incentive effects of changes in taxes.

The data used in our empirical analysis come from the repeated cross-sections of the UK Family Expenditure Survey over the period 1978 to 1992. The FES data provide detailed information on wages, hours, consumption, and household composition. Although not panel data, they provide consistent and accurate micro level information over a long period of time. These data have also been the subject of considerable empirical application to date and have the distinct advantage of collecting accurate information on hours worked, earned income, and consumption expenditures across all household members, consumption data being particularly important in placing the labor supply decision in a life cycle context. The data also contain detailed information on the demographic structure of the household.

We begin in Section 2 with a description of the tax reforms of the 1980s. Section 3 discusses the identification of labor supply effects using tax reforms. In this context we develop suitable difference of differences estimators. Section 4 contains the empirical results. It begins with a description of the data. We provide a contrast between the response parameters estimated using our grouping estimator and those using taxpayer status as a grouping instrument. Our estimates provide small but positive uncompensated wage elasticities with income elasticities that are also small. These are shown to differ by household composition, but the general picture remains the same. In contrast, grouping by taxpayer status gives negative uncompensated wage elasticities. This is shown to result from the systematic change in composition in the taxpayer groups over time. Section 5 concludes. Further information on the data and intermediate results are presented in Appendix A. In Appendix B we present the way we estimate the covariance matrix of our estimator.

6.2 The UK Tax Policy Reforms

In the UK all individuals, irrespective of the total level of household income or consumption, have a tax allowance. Tax is paid only on earnings above this allowance. This aspect of the British tax system implies that about 30% of working married women do not pay tax on earnings. We refer to this group as non-taxpayers. Although this allowance has in the past been different for married men than for (married or single) women, it is totally independent of expenditures; this makes it known to the researcher and the individual very clearly. Beyond this allowance a basic rate of tax is paid. Almost always taxes are collected at source through the Pay as You Earn (PAYE) system. Over this period, reforms announced on budget day and implemented immediately. Any changes to the tax system were widely publicized.

Beyond a certain level of income a higher rate of tax is paid. In addition to income tax, individuals also pay national insurance contributions (NI). These are paid on the entire income by individuals earning above a threshold, the lower earnings limit (LEL). No national insurance contributions are payable over the "upper earnings limit." This system creates a discontinuity in the budget set. To obtain a correct measure of the marginal tax rate for individuals earning more than the tax allowance, we need to add the income tax and NI rates. The budget constraint for British workers over most of our sample period had the form shown in Figure 6.1, where we have omitted the higher rates, which are faced by practically no women in our sample. The NI kink and the tax kink are very close to each other in practice.

Finally almost all goods are subject to indirect tax (VAT). Thus, for example, in a partial equilibrium setting and on the assumption that

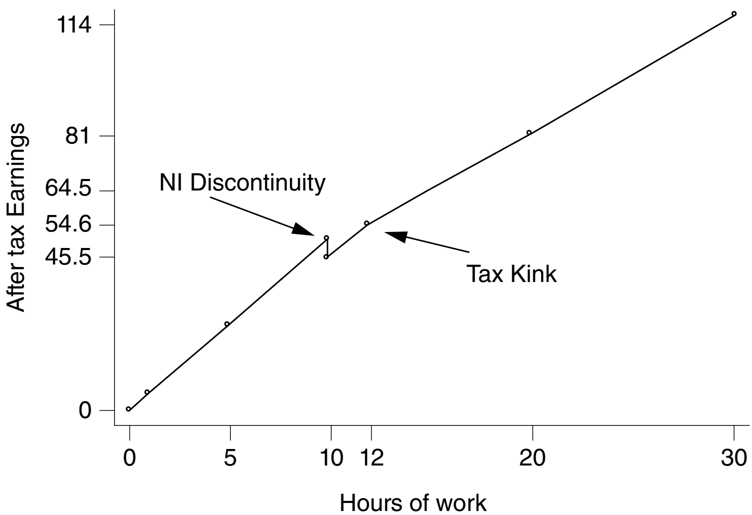


Figure 6.1. The Budget Constraint (illustrated for NI rate 9%, tax rate 25%, pre-tax wage £5)

an extra pound of earnings is spent on the consumption good, the marginal tax rate for a taxpayer is (income tax + NI + VAT)/(1 + VAT).

Table 6.1 presents the main changes to the tax system that are relevant for the marginal wages of married women: The overwhelming majority of taxpaying women face the “basic rate” (plus the NI rate). Nevertheless, there has been more reform activity, which has been affecting mainly incomes earned by men, who face higher rates more often. These reforms are important for our study since they affect what is other (or unearned) income for the wife. The reforms greatly simplified the tax system. In 1978 there were 12 tax bands (a “lower rate” 25%, the “basic rate” 33%, and the “higher rates” 40–75% in 5% steps and 83%). In 1979–1980, the basic rate was reduced to 30% from 33% and all rates higher than 60% were abolished. In 1980/81 the lower rate of 25% was abolished. In 1988 the basic rate was reduced to 25% and rates higher than 40% were abolished. Nevertheless it must be stressed that the earnings threshold for 40% has been falling in real terms and certainly not keeping up with earnings growth. As a result while only 3% of taxpayers (male and female) in 1978 were facing a rate higher than 33%, the equivalent figure now is 10% (at 40% rate). The other important aspect of the reforms has been the phasing out of the NI kink discontinuity. Since 1989 the drop in income at the NI kink is only 2% as opposed to 6.5% and 9% up until then. Nevertheless, the 9% contribution rate for earnings above the LEL remains.

Although the reform agenda can be summarized by saying that there was an overall restructuring with a shift away from direct and towards indirect taxation, the timing of the individual reforms has been such that effective tax rates have been increasing as well as decreasing over this time period. In addition the tax base, as defined by the size of the non-taxable allowances, has both increased and decreased over the sample period (1978–1992). This is evident in Table 6.1, where the allowance changes are given in real terms. Table 6.2 presents the average marginal rates faced by our sample of women married to employed men from the UK Family Expenditure Survey broken down by cohort and level of education. The tax rates change differentially across groups. In fact about 33% of the variation in the table is explained by cohort/education/time interactions, the remaining being explained by the primary cohort/education effects and the time effects.

Households where the husband is out of work will typically be entitled to means-tested benefits, which implies that the wife will face marginal tax rates close to 100% and a highly non-convex budget set. We decided to simplify the analysis by concentrating on women with

Table 6.1. Tax Reforms for the Years 1978 to 1992

Year	78	79	80	81	82	83	84	85	86	37	88	89	90	91	92
Basic rate	33	30	30	30	30	30	30	30	29	27	25	25	25	25	25
Top rate	33	60	60	60	60	60	60	60	60	60	40	40	40	40	40
Tax All. (% Δ)	0	8	0	–10	0	8	2	2	0	0	2	0	0	–3	–2
NI	6.50	6.50	6.75	7.75	8.75	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00	9.00
VAT	8	15	15	15	15	15	15	15	15	15	15	15	15	17.5	17.5

Note: "Tax All. (% Δ)" denotes real percentage change in the tax allowance. The basic rate, the top rate, NI, and VAT are in percentage terms. The top rate is the maximal applicable income tax marginal rate. From 1988 the top rate was the only rate higher than the basic rate.

Table 6.2. Marginal Tax Rates by Financial Year, Education, and Cohort

	Compulsory Education				Post-compulsory Education				Total
	<1940	1940–49	1950–59	1960+	<1940	1940–49	1950–59	1960+	
Financial Year									
1978/79	0.29	0.25	0.31	—	0.37	0.31	0.35	—	0.29
1979/80	0.28	0.24	0.26	—	0.32	0.29	0.32	—	0.27
1980/81	0.29	0.24	0.27	—	0.30	0.26	0.34	—	0.28
1981/82	0.29	0.24	0.28	0.31	0.33	0.28	0.33	—	0.28
1982/83	0.27	0.23	0.25	0.36	—	0.30	0.33	—	0.27
1983/84	0.26	0.23	0.24	0.32	—	0.29	0.29	—	0.26
1984/85	0.28	0.21	0.22	0.31	0.30	0.29	0.31	—	0.26
1985/86	0.29	0.24	0.21	0.32	—	0.26	0.30	0.37	0.27
1986/87	0.27	0.23	0.23	0.31	—	0.27	0.30	0.35	0.27
1987/88	0.24	0.23	0.22	0.28	—	0.30	0.30	0.31	0.26
1988/89	0.23	0.22	0.20	0.24	—	0.25	0.26	0.31	0.24
1989/90	0.23	0.25	0.21	0.23	—	0.29	0.26	0.29	0.25
1990/91	0.24	0.25	0.22	0.24	—	0.27	0.26	0.30	0.25
1991/92	0.24	0.26	0.22	0.24	—	0.29	0.27	0.29	0.25
1992/93	0.25	0.27	0.23	0.25	—	0.27	0.26	0.28	0.26
Total	0.27	0.24	0.24	0.27	0.33	0.28	0.30	0.30	0.26

Note: Cells with a hyphen denote either empty cells or cells that were excluded because the number of observations was less than 50.

employed husbands only. For them the budget set is much simplified. Given that most married men are employed, the potential selection bias is likely to have a minor impact on the results.[2]

In summary the large number of policy reforms over this same period have provided shifts in the tax system, sometimes increasing taxes and sometimes decreasing them, that enhance our ability to identify labor supply responses to tax reform. Tax reform is not the unique source of identifying information. There has been a large increase in wage dispersion mainly due to the increase in the returns to education and to increases in wage dispersion across cohorts (see Gosling, Machin and Meghir (2000) and Schmitt (1993)). Such variation will be a source of identifying information.

6.3 Identifying Labor Supply Responses From Tax Policy Reforms

6.3.1 *A Specification for Labor Supply*

We specify the equation for weekly hours of work (h) to be

$$h = \alpha + \beta \ln w + \gamma \mu \tag{1}$$

where w is the post-tax hourly wage rate and μ is other income defined by the difference between consumption (c) and wh, i.e. $\mu = c - wh$. This definition of other income is consistent both with intertemporal two-stage budgeting in the absence of liquidity constraints and with the presence of liquidity constraints (MaCurdy 1983; Blundell and Walker 1986; Arellano and Meghir 1992). Given that expenditures are collected using diary records, this is a good way of reducing measurement error in the computation of other income in any case. Unobserved taste variation can be introduced by allowing a to vary in the population.

6.3.2 *A Grouping Estimator*

To see the issues involved in estimating labor supply responses using tax reforms, we simplify the notation and use a labor supply equation with no income effect. Suppose this takes the form

$$h_{it} = a + b \ln w_{it} + u_{it} \tag{2}$$

where h_{it} are hours of work and w_{it} is the post-tax hourly wage rate for individual i in financial year t. The error term u_{it} in general will be serially correlated, correlated with the observables, and may be dependent across individuals, reflecting common (macroeconomic) shocks.

The presence of common shocks have a number of implications. First, tax reforms may no longer be exogenous for labor supply: Governments may time their reforms based on their predictions on how the aggregate labor supply is likely to shift over time. Second, even if this were not the case, the mere fact that the number of time periods we have is fixed will imply that the aggregate shocks do not average to zero and hence are a potential source of bias. With a large number of time periods, if tax reforms were predictable by instrumental variables exogenous to labor supply, we could use time series methods on the aggregated data to estimate the wage effects. In general this is not the case; data will consist of a number of repeated cross sections over a relatively small number of time periods. In order to control for the presence of common shocks in this context we need to use some cross section variation. Estimation will have to rely on comparing otherwise similar groups of individuals who have been affected in different ways by the reform, for reasons that are exogenous to labor supply.

The problems in estimating the wage effect b in (2) are the following. We need to control for (i) the common shocks, (ii) for the correlation of w_{it} with u_{it}, and (iii) for self selection into employment.

Suppose individuals can be categorized in one of, say, two groups, $g = \{u, d\}$, each sampled for at least two time periods.[3] For any variable x_{it}, define

$$D_x^{gt} = E(x_{it} \mid P_{it}, g, t) - E(x_{it} \mid P_{it}, g) - E(x_{it} \mid P_{it}, t), \tag{3}$$

where P_{it}, indicates that the individual is observed working. We start by making the following assumptions:

Assumption $A1.1$: $E(u_{it} \mid P_{it}, g, t) = a_g + m_i$.

Assumption $A1.1$: $E[D_w^{gt}]^2 \neq 0$.

Assumption $A1.1$ summarizes the exclusion restrictions for identification; it states that the unobserved differences in average labor supply across groups can be summarized by a permanent group effect (a_g) and an additive time effect (m_t). In other words differences in average labor supply across groups, given the observables, remain unchanged over time. It also says that any self selection into employment (the conditioning on P_{it}) can be controlled by group effects and time effects additively. Assumption $A2.1$ is equivalent to the rank condition for identification; it states that wages grow differentially across groups; this is because the assumption requires that after we have taken away time and group effects there is still some variance of wages left. If there is a tax reform between two periods, affecting the post-tax wages of the two groups in different ways, and assuming that tax incidence does not fully counteract the effects of the reforms, identification of the wage elasticity will be guaranteed.

With these assumptions we can implement a generalized Wald estimator (see Heckman and Robb (1985) for an extensive discussion of grouping estimators). Defining the sample counterpart of D_x^{gt} as $\tilde{x}_{gt} = \bar{x}_{gt} - \bar{x}_g - \bar{x}_t$, i.e. the time-group cell mean minus the overall mean for group g over time and minus the mean at time t over all groups (all defined over workers only), we can write the estimator as

$$\hat{b} = \frac{\sum_g \sum_t [\tilde{h}_{gt}][\ln(\tilde{w}_{gt})] / n_{gt}}{\sum_g \sum_t (\ln(\tilde{w}_{gt}))^2 / n_{gt}} \tag{4}$$

where n_{gt} is the number of observations in cell (g, t). The implementation of this estimator is simple; group the data for workers by g and by time and regress by weighted least squares the group average of hours of work on the group average of the log wage, including a set of time dummies and group dummies. An alternative that gives numerically identical results is as follows: regress using OLS the log after-tax wage rate on time dummies interacted with the group dummies, over the sample of workers only and compute the residual from this regression. Then use the original data to regress hours of work on the individual wage, a set of time dummies and group dummies, and the wage residual. The *t* value on the coefficient of the latter is a test of exogeneity, once the standard errors have been corrected for generated regressor bias (see Pagan 1986) and intra-group dependence. This is the approach that we follow. Finally note that for two time periods and two groups equation (4b) is the difference of differences estimator.

A potential problem with the approach above is that it assumes that the composition effects from changes in participation can be fully accounted for by the additive time and group effects, $a_g + m_t$. Firstly changes in m_t will cause individuals to enter and leave the labor market. Second, with non-convexities, tax reforms beyond the non-taxable allowance may lead to changes in participation. This will be particularly true if fixed costs are large relative to the non-taxable allowance.[4] The presence of composition effects is equivalent to saying that $E(u_{it} \mid P_{it}, g, t)$ is some general function of time and group and does not have the additive structure assumed in *A*1.1.

To control for the possibility that $E(u_{it} \mid P_{it}, g, t)$ may vary over time, we require structural restrictions. A parsimonious specification which we will use is to make the assumption of linear conditional expectation. We now extend *A*1.1 and *A*1.2 by assuming the following.

Assumption *A*1.1: $$E(u_{it} \mid P_{it}, g, t) = a_g + m_t + \delta\lambda_{gt}.$$

Assumption *A*1.2: $$[D_w g_t - \delta_w \lambda_{gt}]^2 \neq 0.$$

Here, λ_{gt} is the inverse Mill's ratio evaluated at $\phi^{-1}(L_{gt})$, ϕ^{-1} being the inverse function of the normal distribution and L_{gt} being the proportion of group *g* working in period *t*.[5] δ_w is a fixed but unknown parameter and δ_w is the (population) partial regression coefficient defined by $\delta_w = E\left[D_w^{gt} D_\lambda^{gt}\right] / E\left[D_\lambda^{gt}\right]^2$. Since we now have an extra parameter to estimate, we need an extra reform. Assumption *A*1.2 models the way composition changes affect differences in the observed labor supplies across groups. It implies that

$$E(h_{it} \mid P_{it}, g, t) = bE(\ln w_{it} \mid P_{it}, g, t) + a_g + m_t + \delta\lambda_{gt} \tag{5}$$

where all expectations are over workers only. Assumption *A*2.2 states that wages must vary differentially across groups over time over and above any observed variation induced by changes in sample composition. We have also implicitly assumed that $E[D_\lambda^{gt}]^2 \neq 0$. If this is not the case, there is no selection bias on the coefficients of interest (here the wage effect) and we can simply use (4). Otherwise we can now estimate the wage effect using a generalization of (4), i.e.

$$\hat{b} = \frac{\sum_g \sum_t [\tilde{h}_{gt} - \hat{\delta}_h \hat{\lambda}_{gt}][\ln(\tilde{w}_{gt}) - \hat{\delta}_w \hat{\lambda}_{gt}] / n_{gt}}{\sum_g \sum_t (\ln(\tilde{w}_{gt}) - \hat{\delta}_w \hat{\lambda}_{gt})^2 / n_{gt}} \tag{6}$$

where $\hat{\lambda}_{gt}$ is an estimate of λ_{gt} and where the partial regression coefficients $\hat{\delta}_x(x = h, w)$ are defined by $\hat{\delta}_x = \left(\sum_g \sum_t x_{gt}\hat{\lambda}_{gt} / n_{gt}\right) / \left(\sum_g \sum_t \hat{\lambda}_{gt}^2 / n_{gt}\right)$ where n_{gt} is the number of observations in cell (g, t). As before this estimator can be implemented using a residual addition technique. We can add an estimate of λ_{gt} as well as the residual of the wage equation estimated on the sample of workers (with no correction for sample selection bias as implied by (5)) to an OLS regression of individual hours on individual wages, time dummies, and group dummies.

To determine whether (6) or (4) should best be used, we can test the null hypothesis that $E[D_\lambda^{gt}]^2 = 0$, which implies that the group effects (a_g) and the time effects (m_t) adequately control for any composition changes (given our choice of groups). If we do not reject this we can use (4).

The assumption in $A1.2$ is worth some discussion. First note that where all regressors are discrete and a full set of interactions are included in the selection equation, use of the normal distribution to compute $\hat{\lambda}_{gt}$, imposes no restrictions. However, the linear conditional expectation assumption implies that a term linear in $\hat{\lambda}_{gt}$ is sufficient to control for selection effects and is potentially restrictive. Using the results in Lee (1984) in general we have that

$$E(u_{it} \mid P_{it}, g, t) = a_g + m_t + \sum_{k=1}^{K} \delta_K \lambda_{gt}^{(k)} \tag{7}$$

where $\lambda_{gt}^{(k)}$ are generalized residuals of order k. The linearity reduces the number of parameters to be estimated and hence the number of periods over which we require exogenous variability in wages. If it is found that $E[D_\lambda^{gt}]^2 \neq 0$, then one can experiment by including higher order generalized residuals after checking that they display sufficient independent variability.

6.3.2.1 ALLOWING FOR INCOME EFFECTS

In general, income effects are important for labor supply and we need to take them into account for at least two reasons. First, the wage elasticity cannot in general be interpreted as an uncompensated wage elasticity, unless we control for other income. Second, income effects are important if we wish to compute compensated wage elasticities for the purpose of evaluating the welfare effects of tax reforms. It is straightforward to extend the estimator in (6) to allow for extra regressors, such as other income. This involves regressing $\tilde{h}_{gt} - \hat{\delta}_w \hat{\lambda}_{gt}$ on $\ln(\tilde{w}_{gt}) - \hat{\delta}_h \hat{\lambda}_{gt}$ and $\tilde{\mu}_{gt} - \hat{\delta}_\mu \tilde{\lambda}_{gt}$ where μ is households' other income. The rank condition for identification is now more stringent: It requires that the covariance matrix $V = Ez_g tz'_{gt}$ has full rank, where $Z_{gt} = [D_w^{gt} - \delta_w \lambda_{gt}, D_\mu^{gt} - \delta_\mu \lambda_{gt}]$. This is equivalent to requiring that the matrix of coefficients on the excluded exogenous variables in the reduced forms of log wage and other income, after taking into account composition effects, has rank 2. A necessary but not sufficient condition for this to be true is that these coefficients be non-zero in each of the reduced forms – i.e., that $E(D_w^{gt} - \delta_w \lambda_{gt})^2$ and $E(D_\mu^{gt} - \delta_w \lambda_{gt})^2$ be nonzero. As before if we accept the hypothesis that $E(Df)2 = 0$ we need to consider whether the rank of $V^* = Ez^*_{gt} z^{*\prime}_{gt}$ is two, where $z^*_{gt} = [D_w^{gt}, D_\mu^{gt}]$. In this case we estimate the model sing the sample counterparts of z^*_{gt} as regressors.

6.3.3 *Discontinuities in the Budget Set*

The budget set in the UK up to 1989 had a major discontinuity at the level of earnings where individuals must start paying national insurance contributions. The contributions are payable on all earnings, leading to a drop in income at that point. In addition there is a kink in the budget set at the level of earnings beyond which individuals must start paying tax. Other kinks are unimportant for our sample. Both the tax kink and the NI discontinuity are close to each other in terms of earnings. The basic structure of the tax system is depicted in Figure 6.1. In the data there is evidence of bunching at the discontinuity.

There is a close link between the statistical coherency of labor supply models with non-linear taxes and the integrability conditions (see Heckman 1978; Gourieroux, Laffont and Monfort 1980; MaCurdy, Green and Paarsch 1990). Nevertheless, imposing the integrability condition at the kinks within the context of a model with a limited number of parameters risks distorting the effects elsewhere in the budget constraint. Ignoring the issue is also a problem since the results may be uninterpretable from a preference point of view. Moreover, the wage effects would probably be biased downwards since for people on the kink we would attribute their inertia to preferences rather than to the structure of the budget constraint. To overcome the problem we need to increase the flexibility of the model (by adding extra parameters); the easiest way to achieve this is to condition out observations close (in a range of 5 hours) to the kink.[6] To correct for this potentially endogenous selection we include an additional selectivity term. This is the first order generalized residual from an ordered probit with three groups: the working non-taxpayers, those close to the kink, and those above the kink. Taking the λ_{gt} terms in (6) to be a vector associated with two stepwise regression coefficients, nothing else changes. We discuss the identification issues that arise and implementation of the extended estimator below.

6.3.4 *The Identifying Assumptions*

To identify our model we need to define the groups whose post-tax wages and other income have changed differentially over time. One might be tempted to split the sample up into taxpayers and non-taxpayers. However, this separation is probably invalid because under very general conditions the composition of the two groups will change over time in a non-random way, in response to tax reforms. In an interesting paper, Eissa (1995) applies the difference of differences estimator and compares the behavior of wives married to high earning husbands to that of wives of lower earning husbands. The two groups were affected differentially by the 1986 tax reform she was analyzing. Her approach requires that the composition of the two groups vis a vis preferences for work not change as a result of the reform. This imposes implicitly restrictions on behavior, since the household's position in the income distribution is to an extent endogenous.

We group the data based on the year of birth and the age the person left full-time education, both interacted with the tax year. To make sure that the number of observations is large enough in each group/year cell we take four cohorts, each born in a ten year interval and only two education groups: Those who left education at the minimum legal age and those who continued beyond the minimum. The four cohorts consist of individuals born in 1930–1939, 1940–1949, 1950–1959, and 1960–1969 respectively. This defines eight groups. Our data extends over 15 financial years. Hence there are substantially more groups than parameters to estimate; this will allow us to construct a test of overidentifying restrictions. The identifying assumption we make is that the average differences in labor supply (given the wage, other income, and demographics) between the groups we defined above be constant over time, as implied by Assumptions *A*1.1 or *A*1.2.[7]

Hence, the identifying assumption does not require that the education choice be unrelated to preferences for hours of work or unrelated to the economic environment facing a cohort. It requires (i) the relationship of the unobservables and education can be described by a fixed group effect depending on education and cohort only and a time effect which is the same across groups; this is the meaning of equation (5); (ii) once 20 years old (when we start including individuals in the sample) individuals with just the statutory level of education cannot switch groups by returning to full-time education. The proportion of workers who left school after the minimum age has increased over time. This is a cohort effect and partly reflects the increase in the statutory years of education. Within a cohort our education measure remains constant apart from sampling variability. To illustrate this we take a ten year cohort of individuals born from 1945 to 1954, which is observed over the entire sample time period and we regress the proportion of those in the cohort who had post-compulsory education (post 16 years of age) on a linear trend increasing by one each year. The coefficient is -0.0016 with a standard error of 0.0034. Hence our measure of the education level used for grouping did not change over the 15 year period for this cohort. This does not mean that workers do not join any training courses, only that these do not imply a change in group vis-à-vis our education classification; these courses are part time or, when full time, they are attended by individuals with post-compulsory education. Evidence from the 1958 NCDS cohort confirms this (see Blundell, Dearden and Meghir 1996).

The reason we expect the groups by which we classify individuals to be affected differentially by the tax reforms is because the cohort/experience effects on wages and other income (essentially husbands' earnings) and the returns to education ensure that the wage and other income distribution will be different across groups. Moreover, the substantial increase in the education returns over these years provides another important source of identifying information.

Finally it is possible that the government was targeting taxation so as to exploit the increased returns to education. If this did happen, it could reduce the explanatory power of the instruments since the two effects would counteract each other. This does not seem to be the case, since taxation at the top of the income distribution fell relatively to the bottom, rather than increased.

In any case all these arguments call for a careful evaluation of the relevance and validity of our instruments. We discuss the way we do this below.

At this point, however, it is worth considering what distinguishes equation (1), a within period marginal rate of substitution equation, from an intertemporal Euler condition once we include a full set of time dummies, given the latter will be collinear with a common real interest rate. In this case identification requires that interest rates differ across individuals as well as time and that this variation is correlated with the education/cohort indicator. This will be true both because the relevant after tax interest rates are different and because of liquidity constraints. Thus the exclusion restriction distinguishing equation (1) from an intertemporal Euler condition can be thought of as the average group interest rate; this is assumed to vary over time and across groups. Such variation in the interest rate can be induced by the tax system as well as by liquidity constraints.

6.3.5 *Household Composition and Labor Supply*

In the labor supply model we include household composition variables. These are dummies that point to the age band of the youngest child in the family. The age bands for the children are 0–2 (*DK02*), 3–4 (*DK34*), 5–10 *(DK510)*, and 11 + (*DK11* +).

Potentially, demographics could be used like the other grouping instruments; the discrete demographic variables could be interacted with the other group indicators to form cells of data. The resulting cells would be too small; i.e., we would have an excessively large number of instruments relative to sample size, which would lead to overfitting in the reduced forms. Thus we restrict the reduced forms to include the set of demographic characteristics linearly. This is equivalent to imposing cross cell restrictions in computing the grouped averages. Finally, we assume that the relationship of demographics and labor supply is constant over time.

6.3.6 *The Relevance and Validity of the Instruments*

A number of recent papers have discussed the adverse effects of using weak instruments (see Staiger and Stock 1997; Bound, Jaeger and Baker 1995) as well as invalid exclusion restrictions. Thus, we need to evaluate whether indeed the post-tax wages and other incomes of the various groups (defined by cohort and education) do change differentially over time. In practice this amounts to evaluating the rank of the matrix of reduced form coefficients on the excluded cohort/education/time interactions. This is after accounting for the time effects, group effects, and demographics that are included in the labor supply equation. We also need to evaluate the validity of these overidentifying restrictions.

To evaluate the rank of the coefficient matrix on the excluded instruments in the reduced form, we use the extension of Anderson's (1951) eigenvalue-based test provided by Robin and Smith (2000). That is, let $\hat{\Pi}$ be a consistent and asymptotically normal estimator of a $p \times k$ $(p \geq k)$ reduced form parameter matrix Π on the excluded instruments (i.e. there are k endogenous variables and p excluded instruments). Let Ω, be the covariance matrix of $\sqrt{N}\text{vec}(\hat{\Pi})$ where N is the sample size. Assume that Ω, is full rank. Define $\hat{\tau}_1 \geq \ldots \geq \hat{t}_k$ to be the eigenvalues of $\hat{\Pi}'\hat{\Pi}$. Under the null hypothesis that the rank of the matrix Π is r, the smallest $k - r$ eigenvalues should be zero. Robin and Smith show that under this null, $N\sum_{i=r+1}^{k}\hat{\tau}_1$ has for limiting distribution a mixture of $(p - r)(k - r)$ one-degree-of-freedom chi-square distributions. The weights can be computed as the non-zero ordered characteristic roots of the matrix $(D'_{k-r} \otimes C'_{p-r})\,()$ where D_{k-r} (respectively C_{p-r}) is a $k \times (k - r)$ matrix (respectively $p \times (p - r)$) formed by the eigenvectors corresponding to the k lowest eigenvalues of $\hat{\Pi}'\hat{\Pi}$ (respectively $\hat{\Pi}\hat{\Pi}'$).

We first evaluate whether the effects of changes in participation and selection away from the NI kink can be explained individually and jointly by the group and time effects. To preempt, we find that this is the case for participation, while for the selection away from the kink the test is borderline. We then evaluate whether the matrix of reduced form coefficients on the excluded interactions in the log wage and other income equations is rank two. We then add the term relating to the selection away from the kink and consider the case for rank three, in the reduced form coefficient matrix for the log after-tax wage, other income, and the ordered probit.

This rank test procedure can also be used to construct a test of overidentifying restrictions: Suppose there are $k + 1$ endogenous variables, including the left-hand side. If we add to the set of endogenous variables the “left-hand-side one” (labor supply here), then the rank of the reduced form coefficients on the excluded instruments must be no more than k. A test of the null hypothesis that the rank is in fact k against the hypothesis that it is $k + 1$ is the test of overidentifying restrictions that we present. We present the most stringent version of the test where we test for rank two against rank three in the reduced form labor supply, other income, and log wage equations.

6.3.7 *Implementation of the Estimator*

First the four reduced forms are estimated on the individual data. In the reduced forms the right-hand-side variables include a complete set of group and time interactions as well as linearly the demographic variables DK_{it}. The estimation sample for the log wage and other income equations excludes non-workers and those working within five hours from the NI or “basic rate of tax” kink. The participation probit is estimated on the entire sample. Finally an ordered probit is used to correct for selection away from the tax and NI kink. This is estimated on the entire sample of workers. For this reduced form, workers are classified as those below the kink, those in close range to the kink, and those above. The reduced forms are used to evaluate the relevance of the instruments.

The labor supply equation is then estimated using OLS on

$$h_{it} = a_g + m_t + \theta' DK_{it} + \beta \ln w_{it} + \gamma \mu_{it} + \delta^w \hat{v}_{it}^w + \delta^\mu \hat{v}_{it}^\mu \\ + \delta^P \hat{v}_{it}^P + \delta^T \hat{v}_{it}^T + e_{it}, \tag{8}$$

where a_g are group dummies, m_t are time dummies, DK_{it} are the demographic variables, and $\ln w_{it}$ and μ_{it} are the individual level of log after-tax wages and other income ($consumption_{it} - w_{it}h_{it}$). The $\hat{v}$ are the residuals from reduced forms to control for the endogeneity of wages ($\hat{v}_{it}^w$), other income ($\hat{v}_{it}^\mu$) participation ($\hat{v}_{it}^p$ an inverse Mill's ratio), and selection away from the tax and NI kink ($\hat{v}_{it}^T$) a generalized residual from an ordered probit). This computational approach gives identical results to grouping but provides directly tests of exogeneity; these are the t statistics on the δ parameters (see Smith and Blundell 1986).

We estimate a version of the model that allows β and γ to vary with demographic composition. This is estimated simply by adding to equation (8) the interactions of $\log w$ and μ with the demographic characteristics described earlier.

The average group cell size is 142 observations and we exclude nine cells with less than 50 observations that occur at the higher ages. The estimator of the asymptotic covariance matrix that we use accounts for the generated regressors (the residuals) and for heteroscedasticity. Moreover, even after time and group effects are controlled for, it is still possible that there remains some limited intragroup correlation in the unobservables. Our estimate of the asymptotic covariance matrix takes this into account. The details of the computation of the covariance matrix are provided in Appendix B.

6.4 Labor Supply Responses

6.4.1 *The Data*

The data are drawn from the repeated cross sections of the UK Family Expenditure Survey (FES) for the years 1978–1992 and consist of married or cohabiting women in the age range 20–50, whose husbands/partners are employed.[8] The survey is continuous and individuals are uniformly distributed across all months of the year. There are 24626 women of which 16781 work. Of these 2970 are within five hours of the NI or tax kink. A brief summary of the data is presented in Table

6.15 of Appendix A. Hours of work are "usual weekly hours, including usual overtime" and the pre-tax wage is constructed by dividing "usual weekly earnings, including usual overtime pay" by "usual weekly hours, including usual overtime." Note that expenditures are not deductible for tax purposes, which makes the calculation of marginal tax rates much more straight-forward in the UK, since the earnings allowance is known explicitly. Finally for consumption we use total weekly non-durable household consumption.

In Figure 6.2 we show a histogram of hours of work by taxpayer status as well as overall. This shows that there is indeed a great deal of variability to be explained. A possible implication of this is that there is ample opportunity for women who wish to change their hours of work to do so. In Figure 6.3 we show the evolution of average hours for the workers which shows the aggregate number of hours per week does vary substantially.

More to the point though, in Figure 6.4 we plot the difference of hours worked by the taxpayers to those worked by the non-taxpayers. This shows a marked decline. We can compare this to the time series pattern of the difference in the after-tax log wage between taxpayers and non-taxpayers. Figure 6.5 shows that taxpayers' after-tax relative wages have increased quite impressively as we would expect given the increase in wage dispersion. The implied wage effect on hours worked

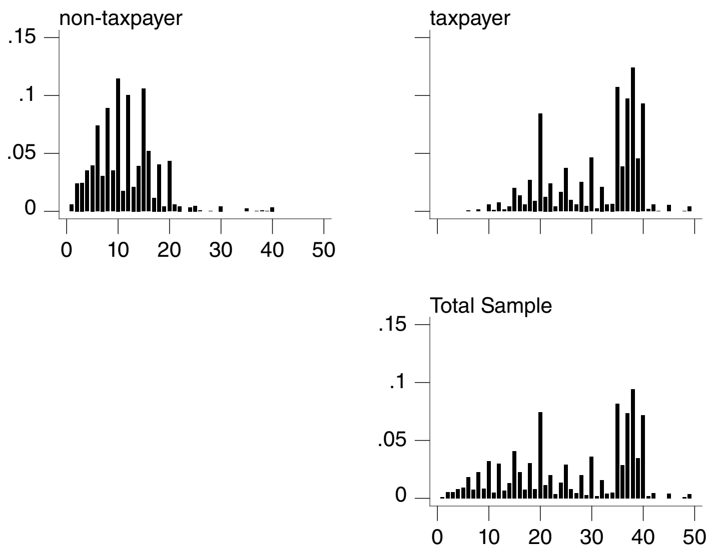


Figure 6.2. Hours of Work by Taxpayer Status

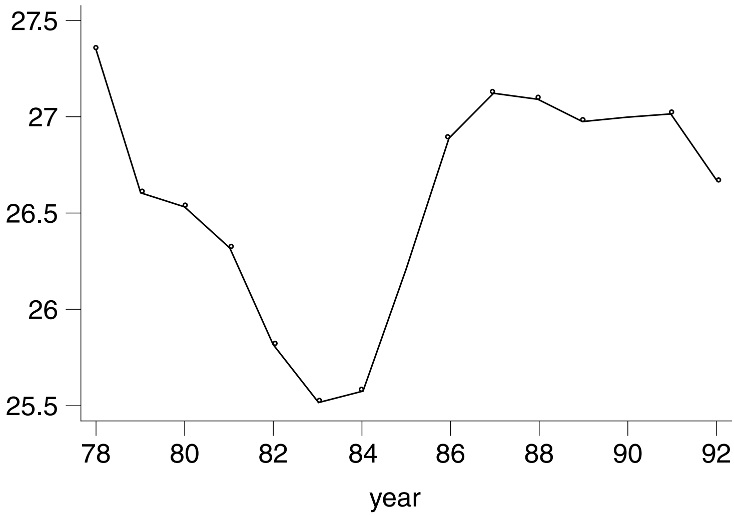


Figure 6.3. Female Hours of Work Over Time

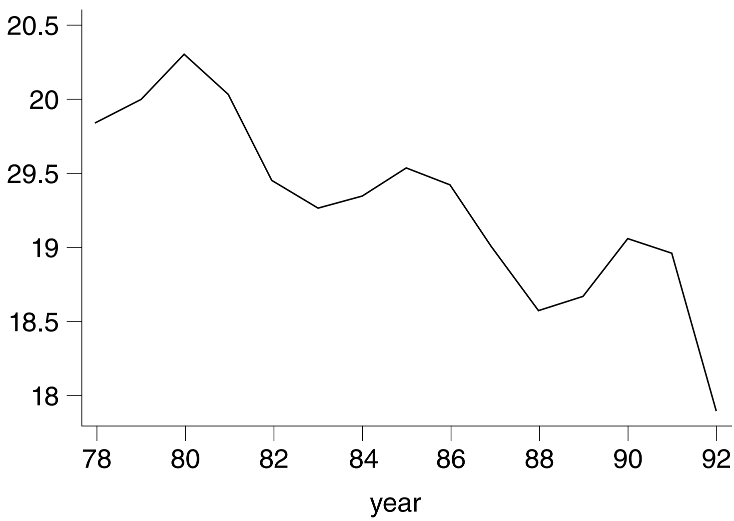


Figure 6.4. Differences of Female Hours of Work Between Taxpayers and Non-taxpayers Over Time

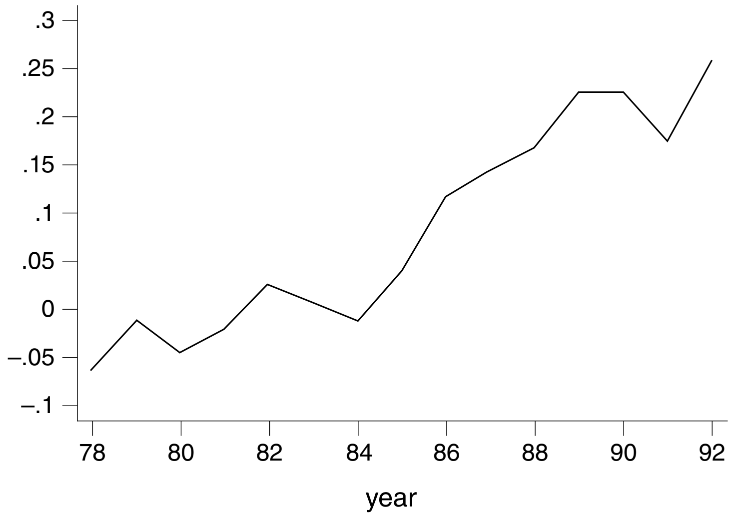


Figure 6.5. Differences in Female Log Wages Between Taxpayers and Non-taxpayers Over Time

is equivalent to an estimate obtained from a simple difference of differences estimator where the groups being compared are the taxpayers and the non-taxpayers. This wage effect on labor supply is negative (reported in detail later). Nevertheless, such an inference is only justified if we can assume that the composition of the two groups, vis-à-vis tastes for work has remained constant over time. We return to this below.

6.4.2 *The Reduced Forms and the Validity of Instruments*

Before we present any further results we report on the relevant rank tests. The reduced forms are presented in Appendix A. We first test the null hypothesis that each of the endogenous variables has not been changing differentially over time across education and cohort groups. The p-values for the four individual reduced forms are given in Table 6.3 in percentage terms. The reduced forms are presented in the Appendix.

This null hypothesis is clearly rejected for the wage and other income variables, indicating that the instruments are indeed highly significant for these two variables. The p-values for the other two equations (participation and ordered probit) are somewhat high. Nevertheless, in the participation equation, a number of interaction effects for the older low educated cohorts in particular are significant. The hypothesis that the rank of the reduced form coefficient matrix (on the excluded instruments) for these two latter reduced forms (participation

Table 6.3. *p*-Values for the Significance of the Excluded Instruments in the Reduced Forms

Log Wage	Other Income	Participation	Ordered Probit
0%	0.038%	10.4%	1.4%

and ordered probit) is zero has a p-value as high as 36.8% (based on the Robin-Smith test). In contrast, the hypothesis that the rank of the reduced form coefficient matrix for other income and the log wage is one is 0.63%, implying rank two as required for identification. Finally, the rank test for rank two against three in the reduced form coefficient matrix obtained from the log wage, other income, and the ordered probit (where the individual p-value is relatively low) is 8.0%.

These results can be interpreted as follows: The composition effects due to changes in participation are explainable by the included time and group effects (i.e. the cohort education indicators and the demographic variables). The result for the composition effects due to selection away from the kink are borderline. On the other hand, the excluded instruments have very strong explanatory power jointly for the after-tax wage and other income variable. These effects are clearly very well identified by our instruments. From our theoretical analysis the implication is that we can ignore the corrections for selection into work and (possibly) around the kink, in identifying the wage and income effects. There is however a question mark as to whether we can really drop the correction for selection away from the kink. The *p*-value for the excluded instruments in that reduced form is 1.4% and the rank test is not far from rejecting the rank two hypothesis in favor of rank three (*p*-value 8%). Thus we also present estimates which include corrections for participation and exclusion of individuals on the tax/NI kink.

Finally we also carried out the test of the overidentifying restrictions as described in Section 6.3.6. The overidentifying restrictions arise from the fact that we have a larger number of groups multiplied by time periods than parameters to estimate. It tests for the absence of time group interactions from labor supply over and above the number of exclusions needed for exact identification. The test has a *p*-value of 0.9% which is quite acceptable.

6.4.3 *Labor Supply Elasticities and Parameter Estimates*

We organize the presentation of the remaining results as follows. First we present a table of elasticities relating to the model including all endogeneity corrections which also allows the coefficients to vary by demographic group. We then show parameter estimates with and without interaction effects for demographics. Here we perform a sensitivity analysis where we assume that participation and the selection around the kink are exogenous, we compare the results to what happens when we include individuals around the kink and finally we compare the results to ones obtained by OLS. All standard errors have been corrected for generated regressor bias and for heteroscedasticity induced by including the generalized residuals, as well as for intra-group dependence as described in Appendix B.

6.4.3.1 THE ELASTICITIES

To start off, in Table 6.4 we present the elasticities implied by the estimates in Table 6.5 presented later. All wage elasticities are positive and highest for women with children at pre-school age, as we would expect. The income elasticities are all negative, except for those women with no children, where it is zero. As a result the compensated wage effects, which matter for welfare, are all positive and the model is consistent with standard theory everywhere in the data. As we report below these elasticities are lower than some recent US estimates, although the latter relate to annual hours of work. In any case our substitution effects imply that taxation does have efficiency costs since taxation will cause substitution leading to reductions in hours of work. Moreover this is only part of the story: Taxation may have important participation effects and corresponding welfare effects. Looking at hours of work is not sufficient to evaluate this, because of fixed costs of work. We now evaluate the robustness of these results.

Table 6.4. Elasticities Grouping Instruments: Cohort and Education

	Compensated			Group Means:		
	Wage	Wage	Other Income	Hours	Wage	Income
No Children	0.140 (0.075)	0.140 (0.038)	0.000 (0.041)	32	2.97	88.63
Youngest Child 0–2	0.205 (0.123)	0.301 (0.144)	−0.135 (0.104)	20	3.36	129.69
Youngest Child 3–4	0.371 (0.150)	0.439 (0.159)	−0.173 (0.139)	13	3.10	143.64
Youngest Child 5–10	0.132 (0.117)	0.173 (0.127)	−0.102 (0.109)	21	2.36	151.13
Youngest Child 11+	0.130 (0.107)	0.160 (0.117)	−0.063 (0.084)	25	2.33	147.31

Note: Asymptotic standard errors in parentheses.

6.4.3.2 THE PARAMETER ESTIMATES

In Table 6.5 we present the parameter estimates from six different specifications for the model including interactions with demographic variables. In Table 6.6 we present the same sets of models but with demographics included only in the intercept. All specifications include a full set of time dummies and group dummies cohort/education indicators

Table 6.5. Parameter Estimates – Groups Defined by Cohort and Education

	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Constant	33.147	33.339	32.261	40.947	29.635	29.558
	3.439	3.362	3.022	0.693	3.843	3.280
DK02	11.797	−11.684	−12.055	−10.138	−11.394	−11.424
	1.971	1.939	1.916	1.706	1.754	1.645
DK34	15.960	−16.012	−16.597	−15.048	−15.412	−15.402
	2.217	2.214	2.168	1.941	1.812	1.701
DK510	−8.466	−8.531	−9.196	−8.132	−9.242	−9.231
	1.381	1.364	1.240	0.897	1.410	1.300
DK110	−3.164	−3.183	−3.889	−3.198	−3.808	−3.810
	1.187	1.180	1.086	0.991	1.165	1.074
Wage Effects						
No Children	4.493	4.579	2.795	−2.377	4.196	4.155
	2.390	2.364	2.082	0.400	2.745	2.336
DK02	4.105	4.110	2.976	−2.148	1.766	1.749
	2.558	2.531	2.267	1.134	2.809	2.419
DK34	6.686	6.739	5.467	1.314	4.185	4.158
	2.707	2.683	2.405	1.109	2.912	2.533
DK510	2.777	2.841	1.520	−3.661	1.338	1.309
	2.448	2.426	2.178	0.606	2.781	2.383
DK11+	3.260	3.337	1.992	−3.230	2.308	2.275
	2.685	2.664	2.430	0.655	3.001	2.629
Other Income						
No Children	0.000	0.000	0.013	−0.008	0.018	0.018
	0.015	0.015	0.013	0.001	0.015	0.013
DK02	−0.028	−0.028	−0.016	−0.037	−0.004	−0.004
	0.016	0.016	0.014	0.005	0.016	0.014
DK34	−0.022	−0.021	−0.008	−0.030	0.002	0.002
	0.017	0.017	0.016	0.009	0.016	0.015
DK510	−0.014	−0.014	−0.001	−0.023	0.010	0.011
	0.015	0.015	0.013	0.003	0.015	0.013
DK11 +	−0.011	−0.010	0.002	−0.019	0.009	0.009
	0.014	0.014	0.012	0.003	0.014	0.012
Residuals Wage	−6.699	−6.758	−5.246		−7.435	−7.405
	2.482	2.455	2.204		2.820	2.426
Other Income	−0.008	−0.009	−0.021		−0.029	−0.029
	0.015	0.015	0.013		0.015	0.013
Tax Kink	0.336	0.321				
	0.082	0.083				
Participation	0.258				−0.071	
	0.450				0.347	

Note: Asymptotic standard errors in italics. Complete set of cohort/education and time dummies included.

Table 6.6. Estimates With No Demographic Interactions

	(i)	(ii)	(iii)	(iv)	(v)	(vi)
Constant	34.551	34.630	33.213	41.661	31.687	31.800
	3.386	*3.324*	*2.947*	*0.689*	*4.299*	*3.182*
DK02	–15.221	–15.211	–14.953	–13.079	–16.499	–16.492
	1.200	*1.200*	*1.137*	*0.509*	*1.397*	*1.074*
DK34	–16.033	–16.061	–16.112	–14.622	–16.945	–16.977
	1.214	*1.185*	*1.099*	*0.490*	*1.381*	*1.046*
DK510	–11.746	–11.774	–11.997	–11.025	–12.776	–12.805
	1.091	*1.067*	*0.971*	*0.325*	*1.233*	*0.945*
DK110	–5.433	–5.443	–5.706	–5.118	–6.624	–6.632
	0.883	*0.875*	*0.794*	*0.347*	*1.039*	*0.810*
Log Wage	4.254	4.273	2.635	–2.446	2.851	2.894
	2.349	*2.341*	*2.054*	*0.346*	*3.062*	*2.265*
Other Income	–0.010	–0.010	0.004	–0.016	0.009	0.009
	0.015	*0.015*	*0.013*	*0.001*	*0.017*	*0.013*
Residuals Wage	–6.779	–6.795	–5.153		–7.371	–7.410
	2.405	*2.396*	*2.135*		*3.113*	*2.334*
Other Income	–0.006	–0.006	–0.020		–0.026	–0.026
	0.015	*0.015*	*0.013*		*0.017*	*0.013*
Tax Kink	0.337	0.332				
	0.076	*0.075*				
Participation	0.083				0.092	
	0.436				*0.356*	

Note: Asymptotic standard errors in italics. Complete set of cohort/education and time dummies included.

fully interacted). Indicatively we present a set of cohort education and time effects in the Appendix, following the reduced form tables. These correspond to column (i) of Table 6.6. Age effects are accounted for by the combination of time and cohort effects. It is straightforward to compute the corresponding elasticities to compare with the results in Table 6.4 by using the group means for hours, the wage, and other income reported in that table.[9] The coefficients are presented in three groups: the intercept coefficients, followed by the wage effects for each demographic group, followed by the other income effects.

In the first column we correct for the endogeneity of the wage rate, other income, participation, and selection away from the kink. In the next column we drop the correction for participation, since the rank test suggested that the changes in participation can be controlled by the included group and time effects. As expected, the results between these two columns are virtually identical. In the third column we also drop the correction for selection away from the NI/tax kink. The wage elasticities do become somewhat smaller but the effects are not dramatic. In contrast, when we use OLS in the fourth column the results are completely different; the implied wage elasticities become negative and the income elasticities larger in absolute value. This reflects the large and negative coefficient on the wage residual in the first three columns; we return to this below.

As already noted the standard errors are corrected for heteroscedasticity, generated regressors, and intra-group dependence, over and above that accounted for by the group effects. It turns out that the latter correction has large effects. For example, the standard error for the wage effect for women without children in column (i) of Table 6.5 is increased from 2.086 to 2.390 as a result of the correction for intra-group dependence; the standard error for the wage effect for women with the youngest child aged 3–4 increases from 2.173 to 2.707. We argued earlier that leaving individuals on or close to the kink in the data would reduce the elasticities, since those who are on the kink are less likely to react to policy changes according to the labor supply model. The effects of including the entire sample and ignoring this non-linearity can be seen by comparing columns one and five (or two and six which do not include corrections for participation) of Table 6.5. When we use the entire sample the coefficients on the wage rate are always lower, but the effects are more marked for women with young children, who tend to work a low number of hours and hence are more likely to be close to the kink. Finally we can repeat the comparison using column (iii), where we exclude the observations close to the kink, but we do not correct for this selection. There we find marginally larger effects for women with children when we exclude the observations close to the kink vis-à-vis the comparable column (vi). For women without children the elasticity is larger when we include all the sample points. Even if we take this comparison to be credible, we should note that childless women rarely work so few hours as to be affected by the NI discontinuity.

In Table 6.6 we report the same set of experiments but excluding the interaction effects with demographics. The intercept of the model contains, as before, a full set of time and group dummies as well as the demographic characteristics. The same broad conclusions follow. In particular, the OLS results imply negative wage elasticities and larger other income elasticities. When we use the entire sample the wage elasticity is much smaller when we correct for the selection (columns (i) and (ii) compared to (v) and (vi) respectively) but virtually the same when compared to the case when we exclude the observations without correcting for the selection (compare columns (iii) with (vi)).

We now return to what feature of the data leads to the OLS results being so different from the IV ones. To understand whether the difference from the OLS results originates primarily from the endogeneity of the pre-tax wage or from differential changes in the composition of the taxpaying group we re-estimate (1) including as a grouping instrument taxpayer status. The model includes a full set of time effects and group effects, where the groups here are defined by education, cohort, and taxpaying status. The estimator is a difference of differences estimator with control group the non-taxpayers and treatment group the taxpayers.[10] In order to keep the cell sizes comparable to the previous results we aggregate the four date-of-birth cohorts we use to two larger cohorts. These estimates include no corrections for participation or for selection on the kink. We also do not include any demographic interactions; the model is most comparable to the results presented in Table 6.6. The results are presented in Table 6.7.

The estimates are very similar to those obtained by OLS which implies that the main source of endogeneity is in fact the changing composition of the taxpaying group. The estimates are a reflection of Figures 6.4 and 6.5.

To interpret the results, note that in Table 6.7 although we control for the endogeneity of individual pre-tax wages by grouping, we assumed that taste differences between taxpayers and non-taxpayers can be modelled as a group fixed effect and time effects; the basic difference between the results in Tables 6.7 and 6.6 (columns (i) and (ii) and (iii)) is that in the latter we allow for changes in taste composition between the two groups over time. Why might this be important?

Figure 6.6 shows how female participation rates have changed over time. After 1982 there is a rapid increase in the proportion of women working. At the same time the proportion of women paying taxes has varied substantially. In Figure 6.7 we show that this proportion fell quite dramatically up to 1984 but rose fast thereafter. This is partly a reflection of the effect of the reforms. Thus for example in 1983/84 there was a large increase in the nontaxable allowance. If, in addition, women entering the labor force in the 1980s are relatively well paid part-timers, as is considered to be the case, the average unobserved taste for

Table 6.7. Using Non-Taxpayers as a Control Croup

Wage		Other Income	
Coeff	Elasticity[a]	Coeff	Elasticity[a]
−2.877	−0.115	−0.0147	−0.0764
1.122	0.0449	0.0069	0.0359

Note: Elasticities evaluated at 25 hours and £130.00 other income.

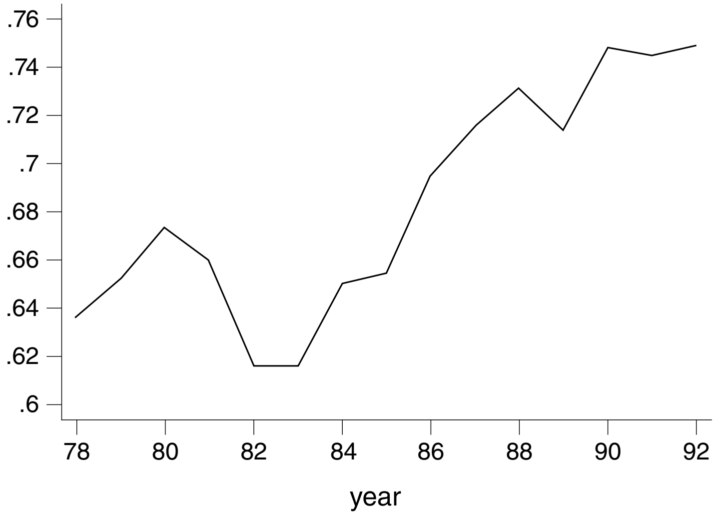


Figure 6.6. Female Participation Over Time

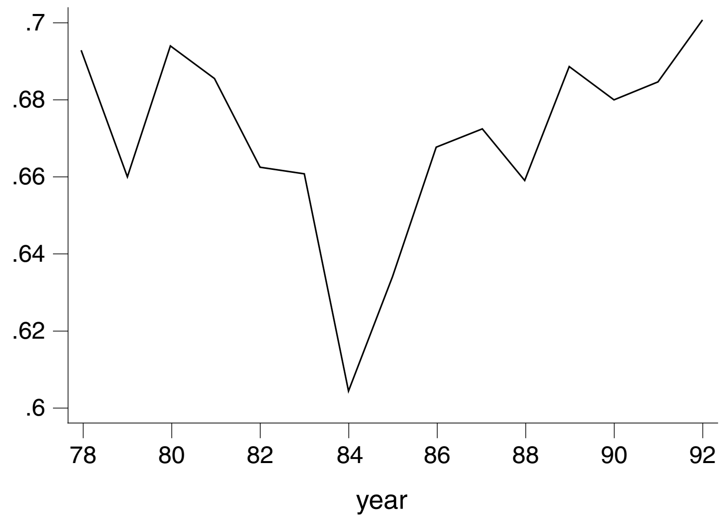


Figure 6.7. Proportion of Taxpayers Over Time

work will be falling among the taxpaying group. This would be consistent with the decline of relative hours for the taxpaying group as shown in Figure 6.4 and leads to the negative wage elasticity in Table 6.7 as well as for OLS. Regrouping the data by groups whose composition cannot change (date of birth and whether statutory education was received) reverses the results and reveals moderate but positive substitution elasticities as well as negative income effects for women with children.

6.4.3.3 SENSITIVITY TO THE NUMBER OF INSTRUMENTS AND TO RESTRICTIONS OF TIME EFFECTS

We conclude our analysis by carrying out some further sensitivity analysis vis-à-vis the number of instruments. For brevity we present results based on the model with no interactions of demographics. All the following experiments include all four residuals.

In the first experiment we reduce the number of instruments by using the prevailing value of five selected tax parameters interacted with the cohort/eduation indicators; this is instead of using time dummies in these interactions. We still include a full set of time dummies and cohort/education indicators additively in the reduced forms and the labor supply function. The tax parameters we use are: the basic tax rate, the higher tax rate, the VAT rate (value added tax, i.e. the indirect tax rate), the NI rate, and the non-taxable earnings allowance. The effect of using these instruments is not only to increase the number of observations per cell but also to increase the weight given to the tax reforms relative to the changing wage structure in identifying the labor supply effects. The results are very similar to what was obtained before: The wage and other income elasticity evaluated at the means is given in the first row of Table 6.8. The wage elasticity is still quite high but the income effect is effectively zero.

In the next experiment we restrict the time effects both in the reduced forms and in the labor supply equation to be a cubic time trend. As with

Table 6.8. Elasticities with Alternative Instrument Sets and Exclusion Restrictions

	Wage		Other Income	
	Coeff	Elasticity[b]	Coeff	Elasticity[b]
Tax parameters as instruments	4.540	0.182	0.00036	0.002
	2.426	0.094	0.014	0.082
Cubic Trends	5.732	0.229	0.008	0.042
	2.252	0.090	0.013	0.078
No Time Effects[a]	3.163	0.127	−0.016	−0.083
	1.537	0.061	0.014	0.085
No Time Effects &	8.680	0.347	−0.063	−0.330
no cohort/educa	1.137	0.045	0.012	0.072

Notes: [1] Includes age, age[2], and a dummy for age > 40.

[2] Elasticities evaluated at 25 hours and £130.00 other income.

our earlier results we include in the reduced form a full set of time effects interacted with cohort/education indicators. This effectively improves the precision of the explanatory power of the excluded instruments. The wage coefficient is larger and the income effect essentially zero.

In the last two experiments we assess the effect of excluding the time effects.

When we do this we do have to control for age, since there are important life cycle effects of age on hours worked. In the first of these two experiments we still include the full set of cohort education interactions. This implicitly means that the time effects are constrained and not completely suppressed. The wage elasticity becomes somewhat smaller and the income effect remains very small, but this time negative as in most of our earlier results. In the last row we exclude the cohort/education effects. This makes the approach similar to traditional cross section studies, such as those reviewed by Mroz (1987) except that the data contain a large number of time periods. Both the other income and wage elasticities now become much larger. The result is very similar to those reported in Arellano and Meghir (1992) where education is used as an identifying instrument.

Two papers known to us use broadly comparable methods although they are different in a number of respects. One is Angrist (1991); he groups PSID data on annual hours worked for a number of years but does not distinguish cross-sectional groups. He interprets his elasticities as intertemporal ones which are always at least as large as the within period ones which we report. He finds an elasticity of 0.634. When using OLS he finds -0.063. These results are consistent with ours. The other paper, by Eissa (1995), evaluates the effects of the 1986 tax reform on female labor supply. Her reported wage elasticities are at least 0.6 and some higher. She also considers the participation effects to derive a total elasticity. Thus our elasticities for weekly hours in the UK are lower than some of those estimated recently in the US. It is important however to emphasize that our paper differs from these studies in important methodological respects as well as in the hours measure we use. Tracing the precise reason for the differences in the estimates is an interesting project.

6.5 Conclusions

The aim of this paper is to investigate the responsiveness of labor supply to exogenous changes in wage rates and non-labor income. To estimate the model we use for our basis the numerous tax reforms of the

1980s whose effect at different times was both to raise and to reduce marginal tax rates. Moreover at the same point in time taxes went up (or down) for some individuals but remained unaffected for others. In addition there have been important changes in the dispersion of pre-tax wages leading to further variation over time in after-tax wages. These changes seem to form an ideal setting for identifying labor supply responses and appear to avoid the need for hard-to-justify exclusion or exogeneity restrictions.

We develop extensions to the difference of differences estimator that account for the effects of changes in labor force composition and for the effects of the discontinuity in the British tax system. Our estimates are based on comparing the evolution of post-tax wages, other income, and hours, of different date-of-birth cohort and education groups. These groups will have been affected differently by the reforms because they occupy different points in the income distribution. Moreover, the increase in wage dispersion favored some groups more than others. The reforms and the change in dispersion affect both after-tax wages and other income since the latter is comprised to a great extent of husband's earnings. We illustrate the explanatory power and validity of the grouping instruments using rank tests.

Using our approach we show that wage elasticities are positive and moderately sized. Other income elasticities are quite small and for women without children these are zero. The OLS results are very different from IV, implying negative wage elasticities. We trace the cause of this discrepancy to changes in the composition of the taxpayer group over time. On the other hand, we find that changes in labor force participation can be explained by common time effects across all groups. Once these are included in the model no further correction is necessary. Our results are very robust to a number of restrictions on the instrument set which effectively increases the number of observations per cell. In particular when we use the values of five key parameters of the tax system as instruments, interacting these with the cohort/education dummies, we find virtually the same results as when we use time dummies in these interactions. Our conclusion is that major tax reform should take into account behavioral effects since our compensated substitution elasticities suggest that the welfare effects are not negligible.

Appendix A: The Reduced Forms and Descriptive Statistics

In Tables 6.9, 6.10, 6.11, and 6.12 we present the reduced forms used in estimation. In each table the first row is the cohort/education effect and

Table 6.9. Reduced Form for Log Wage

	Time Effects	Compulsory Schooling Year of Birth				Post Compulsory Year of Birth			
		<1940	1940–49	1950–59	1960+	<1940	1940–49	1950–59	1960+
Group Effect:	Effects	1.279 (0.059)	1.145 (0.033)	1.316 (0.024)	1.130 (0.022)	1.749 (0.075)	1.608 (0.057)	1.544 (0.022)	1.343 (0.024)
Year = 78	–0.501 (0.029)	0.035 (0.069)	0.275 (0.050)	E	-	0.133 (0.092)	0.142 (0.074)	E	-
Year = 79	–0.444 (0.028)	0.043 (0.068)	0.239 (0.049)	E	-	–0.032 (0.096)	0.136 (0.074)	E	-
Year = 80	–0.387 (0.029)	0.033 (0.069)	0.210 (0.049)	E	-	0.011 (0.093)	0.161 (0.074)	E	-
Year = 81	–0.351 (0.033)	–0.055 (0.072)	0.153 (0.053)	–0.061 (0.041)	E	-	–0.056 (0.074)	E	-
Year = 82	–0.306 (0.033)	–0.121 (0.072)	0.148 (0.053)	–0.038 (0.042)	E	-	0.057 (0.076)	E	-
Year = 83	–0.243 (0.033)	–0.086 (0.073)	0.137 (0.054)	–0.091 (0.043)	E	-	–0.028 (0.077)	E	-
Year = 84	–0.264 (0.032)	–0.042 (0.072)	0.198 (0.054)	–0.041 (0.041)	E	–0.175 (0.101)	0.057 (0.078)	E	-
Year = 85	–0.286 (0.038)	–0.047 (0.076)	0.224 (0.058)	0.026 (0.049)	E	-	0.040 (0.081)	0.096 (0.049)	E
Year = 86	–0.209 (0.033)	–0.017 (0.074)	0.101 (0.056)	–0.003 (0.045)	E	-	0.165 (0.078)	0.142 (0.045)	E
Year = 87	–0.139 (0.032)	–0.043 (0.075)	0.074 (0.056)	–0.066 (0.044)	E	-	0.121 (0.078)	0.094 (0.044)	E
Year = 88	–0.080 (0.031)	–0.103 (0.075)	0.164 (0.055)	–0.050 (0.043)	E	-	0.073 (0.076)	0.090 (0.045)	E
Year = 89	–0.027 (0.030)	–0.106 (0.075)	0.105 (0.053)	–0.082 (0.043)	E	-	–0.045 (0.079)	–0.016 (0.044)	E
Year = 90	–0.032 (0.030)	–0.172 (0.077)	0.092 (0.054)	–0.094 (0.044)	E	-	–0.021 (0.074)	–0.052 (0.045)	E
Year = 91	0.069 (0.029)	–0.231 (0.077)	0.006 (0.053)	–0.177 (0.043)	E	-	–0.099 (0.077)	–0.018 (0.044)	E
Child Aged:	-	0–2	3–4	5–10	11+	-	-	-	-
	-	0.046 (0.011)	–0.020 (0.013)	–0.084 (0.009)	–0.096 (0.010)	-	-	-	-

Note: Asymptotic standard errors in parentheses.

the first column is the time effect. To obtain the predicted wage (say) the group effect is added to the time effect and to the number in the cell which represents the interaction effect. Thus the predicted log wage for a low-education individual born 1950–1959 in 1981 is 1.316 – .351 – 0.061 = 0.904 (Table 6.9). For the participation probit and the ordered probit these calculations provide index values which need to be converted to probabilities using the normal distribution. For the ordered probit the

Table 6.10. Reduced Form for Other Income

	Time Effects	Compulsory Schooling Year of Birth				Post Compulsory Year of Birth			
		<1940	1940–49	1950–59	1960+	<1940	1940–49	1950–59	1960+
Group Effect:		140.13 (14.51)	123.85 (9.36)	109.39 (6.52)	95.22 (5.74)	220.72 (18.56)	160.62 (15.89)	110.60 (5.95)	90.88 (6.21)
Year = 78	–40.16	6.34	1.85	E	-	–70.75.	–34.82	E	-
	(7.49)	(17.12)	(13.23)			(22.88)	(19.67)		
Year = 79	–24.61	–5.31	0.95	E	-	–36.30	–19.02	E	-
	(7.31)	(17.16)	(13.08)			(24.00)	(19.95)		
Year = 80	–33.30	0.32	0.06	E	-	–16.97	–12.65	E	-
	(7.42)	(17.25)	(13.26)			(23.51)	(19.97)		
Year = 81	–41.34	13.64	10.29	3.58	E	-	–16.53	E	-
	(8.61)	(17.92)	(14.05)	(10.37)			(20.17)		
Year = 82	–33.71	4.00	20.92	4.57	E	-	–5.26	E	-
	(8.46)	(18.02)	(14.00)	(10.05)			(20.26)		
Year = 83	–34.36	17.14	14.95	9.15	E	-	–6.44	E	-
	(8.26)	(18.12)	(14.23)	(10.26)			(20.42)		
Year = 84	–20.63	2.92	6.96	–2.30	E	–38.64	– 16.74	E	-
	(8.10)	(18.10)	(14.32)	(10.13)		(25.52)	(20.96)		
Year = 85	–23.32	2.39	8.06	0.53	E	-	3.83	12.78	E
	(9.64)	(19.11)	(15.46)	(12.21)			(21.62)	(12.33)	
Year = 86	–10.73	–11.55	–2.19	3.71	E	-	–6.22	7.69	E
	(8.84)	(18.78)	(15.26)	(11.62)			(21.47)	(11.83)	
Year = 87	–20.84	5.39	17.97	1.48	E	-	21.55	16.38	E
	(8.31)	(18.86)	(15.17)	(11.27)			(21.44)	(11.36)	
Year = 88	–12.06	3.96	15.61	5.01	E	-	–10.16	37.56	E
	(8.02)	(19.05)	(14.81)	(11.21)			(20.93)	(11.48)	
Year = 89	–16.16	0.89	15.76	3.75	E	-	–18.15	23.10	E
	(7.69)	(18.86)	(14.56)	(11.13)			(21.75)	(11.33)	
Year = 90	–24.68	22.29	6.22	8.69	E	-	32.16	51.34	E
	(7.81)	(19.31)	(14.82)	(11.42)			(20.71)	(11.60)	
Year = 91	–17.39	9.25	22.32	12.45	E	-	14.24	33.93	E
	(7.54)	(19.73)	(14.62)	(11.30)			(21.65)	(11.51)	
Child Aged:	-	0–2	3–4	5–10	11+	-	-	-	-
	-	78.50	75.16	64.74	49.37	-	-	-	-
	-	(2.36)	(3.03)	(2.44)	(2.88)	-	-	-	-
R^2	0.086		-	-	-	-	-	-	-

Note: Asymptotic standard errors in parentheses.

thresholds are given at the bottom of the table. Finally the linear demographic effects are presented at the bottom of each reduced form.

Whenever a cell had to be dropped because of exact multicollinearity this is denoted by "E". The interaction effect then is zero. The base year is 1992. Finally cells with a hyphen denote either empty cells or cells that were excluded because the number of observations were less than 50.

Table 6.11. Reduced Form Participation Probit

	Time Effects	Compulsory Schooling Year of Birth				Post Compulsory Year of Birth			
		<1940	1940–49	1950–59	1960+	<1940	1940–49	1950–59	1960+
Group Effect:		0.328	1.043	1.590	1.582	0.754	1.316	1.751	1.846
		(0.153)	(0.114)	(0.082)	(0.071)	(0.205)	(0.199)	(0.075)	(0.079)
Year = 78	–0.343	0.718	0.496	E	-	0.413	0.284	E	-
	(0.092)	(0.187)	(0.160)			(0.253)	(0.241)		
Year = 79	–0.154	0.511	0.380	E	-	0.099	0.168	E	-
	(0.091)	(0.188)	(0.159)			(0.265)	(0.245)		
Year = 80	–0.248	0.568	0.495	E	-	0.270	0.342	E	-
	(0.092)	(0.189)	(0.161)			(0.262)	(0.245)		
Year = 81	–0.293	0.459	0.348	0.075	E	-	0.316	E	-
	(0.107)	(0.197)	(0.171)	(0.128)			(0.248)		
Year = 82	–0.404	0.567	0.404	–0.077	E	-	0.243	E	-
	(0.104)	(0.198)	(0.169)	(0.122)			(0.247)		
Year = 83	–0.318	0.511	0.339	–0.117	E	-	0.119	E	-
	(0.101)	(0.198)	(0.171)	(0.123)			(0.248)		
Year = 84	–0.414	0.759	0.527	0.147	E	0.288	0.277	E	-
	(0.098)	(0.198)	(0.172)	(0.121)		(0.281)	(0.254)		
Year = 85	–0.351	0.540	0.373	–0.032	E	-	0.106	0.130	E
	(0.120)	(0.212)	(0.188)	(0.149)			(0.263)	(0.152)	
Year = 86	–0.071	0.252	0.231	–0.077	E	-	0.133	–0.080	E
	(0.115)	(0.210)	(0.189)	(0.146)			(0.269)	(0.150)	
Year= 87	–0.036	0.363	0.158	–0.131	E	-	–0.088	–0.043	E
	(0.105)	(0.209)	(0.186)	(0.139)			(0.264)	(0.139)	
Year = 88	–0.012	0.431	–0.010	0.016	E	-	0.038	–0.244	E
	(0.100)	(0.211)	(0.179)	(0.139)			(0.261)	(0.139)	
Year = 89	–0.030	0.287	0.195	–0.031	E	-	–0.084	–0.144	E
	(0.094)	(0.206)	(0.179)	(0.136)			(0.268)	(0.136)	
Year = 90	–0.015	0.182	0.191	–0.080	E	-	0.086	–0.177	E
	(0.096)	(0.210)	(0.185)	(0.139)			(0.261)	(0.140)	
Year = 91	0.106	0.321	–0.189	–0.213	E	-	–0.090	–0.220	E
	(0.093)	(0.219)	(0.177)	(0.139)			(0.273)	(0.139)	
Child Aged:	-	0–2	3–4	5–10	11+	-	-	-	-
	-	–1.831	–1.284	–0.650	–0.149	-	-	-	-
		(0.030)	(0.035)	(0.030)	(0.035)				

Note: Asymptotic standard errors in parentheses.

Following the reduced forms we present the cohort/education effects and time effects for the model in column (i) of Table 6.6 in Tables 6.13 and 6.14. The base year is 1992 and the base cohort consists of those born in the 1960s with post-compulsory education.

In Table 6.15 describing the data, child is a dummy for the age of the youngest child, education is the age at which the individual left full-time

Table 6.12. Reduced Form Ordered Probit For Selection Away From The Kink

	Time Effects	Compulsory Schooling Year of Birth				Post Compulsory Year of Birth			
		<1940	1940–49	1950–59	1960+	<1940	1940–49	1950–59	1960+
Group Effect:		–0.494	–0.319	0.312	0.448		0.381	0.711	0.787
		(0.327)	(0.270)	(0.252)	(0.247)		(0.312)	(0.241)	(0.253)
Year = 78	0.109	–0.471	0.045	E	-	0.428	0.680	E	-
	(0.116)	(0.393)	(0.250)			(0.174)	(0.312)		
Year = 79	–0.102	0.209	0.198	E	-	0.689	0.386	E	-
	(0.108)	(0.265)	(0.247)			(0.166)	(0.319)		
Year = 80	–0.020	0.337	0.212	E	-	0.510	0.227	E	-
	(0.106)	(0.261)	(0.247)			(0.166)	(0.325)		
Year = 81	0.044	0.185	–0.036	0.286	E	-	–0.089	E	-
	(0.144)	(0.252)	(0.267)	(0.193)			(0.172)		
Year = 82	0.028	–0.072	–0.096	0.199	E	-	–0.205	E	-
	(0.143)	(0.264)	(0.268)	(0.192)			(0.169)		
Year = 83	–0.029	0.209	–0.179	0.296	E	-	–0.230	E	-
	(0.131)	(0.267)	(0.263)	(0.188)			(0.164)		
Year = 84	–0.347	0.346	0.270	0.398	E	0.043	0.404	E	-
	(0.123)	(0.274)	(0.260)	(0.181)		(0.150)	(0.340)		
Year = 85	–0.015	0.258	–0.003	0.124	E	-	–0.260	–0.198	E
	(0.154)	(0.268)	(0.280)	(0.208)			(0.189)	(0.279)	
Year = 86	–0.108	–0.195	–0.037	0.215	E	-	–0.031	0.275	E
	(0.134)	(0.193)	(0.270)	(0.198)			(0.170)	(0.276)	
Year = 87	–0.136	0.122	–0.110	0.421	E	-	–0.001	0.282	E
	(0.131)	(0.190)	(0.270)	(0.196)			(0.168)	(0.278)	
Year = 88	–0.079	0.215	0.010	0.214	E	-	0.019	0.236	E
	(0.121)	(0.173)	(0.271)	(0.189)			(0.160)	(0.264)	
Year = 89	–0.062	0.135	–0.093	0.332	E	-	0.023	0.259	E
	(0.116)	(0.176)	(0.268)	(0.185)			(0.157)	(0.279)	
Year = 90	0.048	0.132	–0.142	0.216	E	-	–0.204	–0.041	E
	(0.118)	(0.169)	(0.276)	(0.188)			(0.161)	(0.259)	
Year = 91	0.070	–0.134	–0.283	0.076	E	-	–0.224	0.245	E
	(0.112)	(0.167)	(0.273)	(0.185)			(0.158)	(0.276)	
Child Aged:	-	0–2	3–4	5–10	11+	-	-	-	-
	-	–1.561	–1.643	–1.335	–0.718	-	-	-	-
		(0.040)	(0.044)	(0.033)	(0.035)				
Thresholds	-	no tax/NI	NI/tax	-	-	-	-	-	-
	-	–1.748	–1.075	-	-	-	-	-	-
		(0.248)	(0.248)						

Note: Asymptotic standard errors in parentheses.

Table 6.13. Cohort/Education Effects

Cohort:	Cohort Education Effects on Labor Supply (from Table 6.6, column (i))						
	Low Education				High Education		
	<1940	1940–49	1950–59	1960+	<1940	1940–49	1950–59
Coef.	−9.434	−6.335	−3.942	−1.107	−7.978	−5.336	−2.694
Stand. Err.	1.229	0.966	0.818	0.836	1.970	1.280	0.959

Table 6.14. Time Effects

Fin. Year	Time Effects on Labor Supply. Base Financial Year 1992 (from Table 6.6, Column (i))													
	1978	1979	1980	1981	1982	1983	1984	1985	1986	1937	1988	1989	1990	1991
Coef.	3.79	2.74	2.46	1.26	0.95	0.91	0.02	0.55	0.77	0.90	1.38	0.85	1.06	1.27
Stand. Err.	1.16	1.04	0.89	1.08	0.90	0.87	1.03	0.79	0.61	0.81	0.60	0.61	0.78	0.59

education, wages and other income are in 1992 prices, and year denotes the financial year that starts in April.

Appendix B: The Computation of the Standard Errors

The model we estimate has the form

$$y_i = \beta' x_i + \delta' \hat{z}_i + v_i$$

where i denotes individuals, x_i contains all the regressors including the time effects and the group effects. The are the estimated residuals. Let the kth estimated residual be defined by $\hat{z}_i^k = s(m_j' \hat{\gamma}_k)$ where $s(.)$ could represent a generalized residual or just a residual from a linear reduced form and where $\hat{\gamma}_k$ is the $q_k \times 1$ vector of coefficients in the kth reduced form. The $q_k \times 1$ vector of variables included in the reduced form for observation i is denoted m_i. Let z_i represent the residuals evaluated at the true parameter estimates. Finally $v_i = u_i + \delta'(z_i - \hat{z}_i)$. In computing the standard errors we need to account for the effect of using estimated rather than actual values for γ_k. Dependence within groups and time is mainly accounted for by the presence of the group and time effects (see, for example, Moulton 1986). However there may still be some

Table 6.15. Descriptive Statistics for the Sample of Workers

Year	78	79	80	81	82	83	84	85
Hours	27.34	26.62	26.53	25.89	25.81	26.04	25.17	26.41
	11.75	12.00	11.95	12.12	11.97	12.07	11.99	11.72
Log Wage	0.90	0.94	0.99	0.96	0.99	1.04	1.06	1.29
	0.39	0.39	0.40	0.40	0.41	0.44	0.40	0.45
Other Inc.	109.55	121.88	115.29	110.41	116.08	115.15	121.82	127.71
	87.51	105.22	104.81	97.44	102.83	97.65	106.84	112.30
Child 0–2	0.08	0.09	0.08	0.09	0.09	0.10	0.08	0.14
Child 3–4	0.08	0.07	0.07	0.05	0.06	0.06	0.08	0.09
Child 5–10	0.23	0.25	0.25	0.27	0.22	0.22	0.23	0.20
Child 11+	0.12	0.12	0.14	0.12	0.14	0.15	0.15	0.15
Age	38.25	37.97	38.47	38.10	38.04	37.99	38.50	38.60
	7.62	7.29	6.71	7.09	6.89	6.92	6.92	6.04
Education	15.96	15.98	16.05	16.23	16.26	16.24	16.24	16.86
	2.21	2.06	2.19	2.25	2.26	2.09	2.10	2.14
Educ >16	0.23	0.23	0.23	0.27	0.24	0.27	0.26	0.39
Year	86	87	88	89	90	91	92	
Hours	26.51	26.93	27.22	27.14	27.03	27.03	26.97	
	12.14	11.89	12.00	12.40	11.74	12.06	12.42	
Log Wage	1.08	1.14	1.18	1.24	1.24	1.23	1.29	
	0.40	0.44	0.46	0.46	0.43	0.45	0.47	
Other Inc.	114.45	124.38	120.47	131.23	118.22	117.14	121.59	
	109.31	153.74	125.51	170.55	126.58	126.00	116.23	
Child 0–2	0.10	0.10	0.14	0.13	0.15	0.12	0.14	
Child 3–4	0.07	0.08	0.08	0.08	0.07	0.09	0.09	
Child 5–10	0.22	0.20	0.20	0.22	0.21	0.19	0.19	
Child 11+	0.12	0.12	0.10	0.12	0.10	0.16	0.15	
Age	37.50	37.44	37.27	37.62	37.80	38.11	38.27	
	6.36	6.58	6.80	6.43	6.50	6.30	6.36	
Education	16.51	16.68	16.67	16.71	16.64	16.91	16.90	
	2.17	2.28	218	2.17	2.07	2.35	2.37	
Educ >16	0.32	0.34	0.35	0.36	0.36	0.39	0.37	

limited dependence between the errors within a group even after removing these main effects. We use a White (1982) approach to allow for this problem. There are N_{gt}, individuals within each group in period t.

Let the $N_{gt} \times N_{gt}$, covariance matrix of the errors within a group g in time period t be denoted by Ω_{gt}. The off-diagonal elements represent intragroup covariances. Let X_{gt}, and Z_{gt}, represent the matrix of N_{gt}, observations of the variables in x and z respectively for group g in period t. Define the $N_{gt} \times p$ matrix $Q_{gt} = [X_{gt}Z_{gt}]$, p being the total number of regressors including the residuals z_i. Let Q represent the entire matrix of observations over the whole sample for the x and z variables.

We assume that

$$\text{plim}N_{gt\to\infty}\frac{Q'_{gt}\Omega_{gt}Q_{gt}}{N_{gt}}=P_{gt} \tag{9}$$

where P_{gt} is a $p \times p$ positive definite matrix. This assumption effectively limits the amount of intra-group dependence and implies that the model can be consistently estimated with a fixed number of time periods and the number of individuals going to infinity, which is our framework. Denote the $N_{gt} \times 1$ vector of estimated residuals within a group g in period t by $\hat{v}_{gt}$ Denote $\zeta = (\beta'\delta')$. There are G groups over T time periods and K generated regressors. Denote by Γ_{gt}^{k} the $N_{gt} \times q_k$ matrix whose ith row is given by the derivative of $s(m_i'\hat{\gamma}_k)$ with respect to γ_k. Finally denote by $V(\hat{\gamma}_k)$ the covariance matrix of $\hat{\gamma}_k$. We assume that the number of time periods and the number of groups is fixed but that the number of individuals within each group is large and goes to infinity. Given the above assumptions we can estimate consistently the asymptotic covariance of the estimated parameters ζ by

$$V(\hat{\zeta}\ )=(Q'Q)^{-1}\sum_{g=1}^{G}\sum_{t=1}^{T}Q'_{gt}\hat{v}_{gt}\hat{v}'_{gt}\,Q_{gt}+\sum_{k=1}^{K}\hat{\delta}\ _{k}^{2}Q'_{gt}\Gamma_{gt}^{k}V(\hat{\gamma}_{k})\Gamma_{gt}^{k}Q_{gt}(Q'Q)^{-1}.$$

This covariance matrix allows for the effects of estimated residuals, for heteroscedasticity, and for dependence within groups consistent with assumption (9). The formula we use ignores, for computational simplicity, the covariance of the coefficients $\hat{\gamma}_k$ across the $k = 1,\ldots,K$ reduced forms. However, note that in our case the correction for generated regressors (the second term in the square brackets) accounts only for a small component of the above covariance matrix.

7

Job Changes and Hours Changes: Understanding the Path of Labor Supply Adjustment

Richard Blundell, Mike Brewer and Marco Francesconi

7.1 Introduction

The use of the canonical model of labor supply for policy analysis is pervasive. A central tenet of this model is that workers have flexible choices over hours of work, selecting their desired utility-maximizing level at any given wage. A number of studies have cast some doubt on this model by arguing that there is not free choice of hours within a job and limited choice across jobs, and providing evidence of job "packages" whereby wage and hours are tied together.[1] Most of the contributions in this literature however identify hours constraints by relying on observed individual characteristics (e.g., number and age of children, or job mobility) or stated labor supply preferences (Ham 1982; Moffitt 1984; Lundberg 1985; Altonji and Paxson 1988; Stewart

The original version of this chapter was published as: Blundell, R./Brewer, M./Francesconi, M. (2008). Job Changes and Hours Changes: Understanding the Path of Labor Supply Adjustment, in: Journal of Labor Economics, 26(3): 421–53. © 2008 by the University of Chicago. Financial supports from the Department of Trade and Industry 'Labor Market Flexibility Small Grant Fund 2004' and from the ESRC Centre for the Microeconomic Analysis of Public Policy (CPP) at the IFS are gratefully acknowledged. We thank Alison Booth, Mark Bryan, Jan van Ours, Arthur van Soest, the Editor (Jeff Smith), two anonymous referees, and seminar participants at the Universities of Essex and Tilburg, Institute for Fiscal Studies, Bank of England, Bank of Spain, UK Department for Education and Skills (London), the 2005 European Society for Population Economics (Paris) and the Econometric Society World Congress 2005 (London) for helpful suggestions. We are responsible for all remaining errors and shortcomings of the paper.

and Swaffield 1997; Euwals 2001). These two strategies are problematic because changes in labor supply preferences or other individual variables may not be exogenous to hours levels or changes.

Our strategy is to use a sequence of policy reforms that directly affected the labor supply incentives of specific groups of individuals while leaving the incentives faced by others unchanged. Our objective is to use these reforms to assess the degree of flexibility of hours changes within and across jobs. The emphasis is more on the extent of within- and between-job flexibility – whether it is large or small and for which type of workers it is larger or smaller – rather than on the question of whether hours flexibility is complete or not. Specifically, we analyze transitions from positive hours of work to positive hours of work made by single women in response to (exogenous) tax and benefit policy changes that occurred in Britain in the 1990s. We use three different reforms to highlight likely actual movements along the labor supply curve, and combine these with information on stated preferences and job mobility to assess whether and how women adjust their labor supply in response to changes in the incentives to work a given number of hours.

Many of the tax and benefit reforms in the United Kingdom, Canada, and the United States has been directed at increasing the labor market attachment of the lower skilled workers, in particular those facing high fixed costs of work such as childcare (Blundell 2002). A significant part of the rise in employment among single mothers in the US over the late 1980s and 1990s has been attributed to the expansion of the Earned Income Tax Credit (Eissa and Liebman 1996; Meyer and Rosenbaum 2001). Similarly, it has been argued that much of the rise in the participation of single mothers in the UK has been due to increases in the generosity of the tax credit policies, namely Family Credit (FC) and Working Families Tax Credit (WFTC).[2] The Self-Sufficiency Project (SSP) in Canada provided further experimental evidence on the effectiveness of financial incentives on the working decisions of low-income single-parents (Card and Robins 1998). An interesting feature of the UK reforms has been the changing incentive structure towards part-time and full-time work engendered by these reforms. Not only has employment responded to these reforms but so has the distribution of weekly hours of work (Blundell et al. 2000b; Brewer 2001). However, the mechanism for these adjustments in labor supply has not been studied. Are adjustments to hours made by moving jobs or do workers adjust their hours of work over time with the same employer? This mechanism of adjustment is the focus of this paper.

For such an analysis panel data are essential, as it is necessary to know the employment position and hours worked of each specific individual before and after adjustment takes place. Since 1991 a high-quality panel data survey, the British Household Panel Survey (BHPS), has been collected annually for Britain, and that is the data source we use in our analysis covering the period 1991–2002. The BHPS also has the attraction of recording individuals' *stated* preferences toward hours of work, so that actual movements can be examined alongside changes in stated preferences.

Even if hours were completely fixed *within* jobs but mobility *between* jobs was costless, we would still expect workers to be located on their labor supply curve, i.e., at their most preferred level of hours given the market wage. But if there are individual costs to moving between jobs, or firms collectively require a given number of hours due to facing fixed costs or technology-related coordination requirements,[3] then workers will face immobility (at least in the short run) on the hours they can work. This has implications for the interpretation of data on actual and preferred hours of work, rates of mobility between jobs, and for estimating models of labor supply. Various strands of research have suggested models of hours choice where hours are fixed within jobs. One strand, which dates back to Barzel (1973) and Rosen (1976), grounds its analysis in models where jobs are packages of fixed hours-wage combinations (Ham 1982; Moffitt 1984; Lundberg 1985; Biddle and Zarkin 1989; Kahn and Lang 1991; Altonji and Paxson 1988; 1992; Dickens and Lundberg 1993). Another more recent strand is developed within a monopsonistic environment, where employer preferences play a key role in determining hours of work in a given job (Manning 2003).

In this study we are interested in examining if and how employed single mothers vary their hours in response to exogenous changes in the incentives to work a given level of hours. For this purpose, we use reforms to the tax and benefit system that changed the hours conditions for FC in 1992 and 1995 and the attractiveness of work through WFTC in 1999 to assess the "canonical" model of hours flexibility. We also look at how changes in hourly wages both within and between jobs relate to the introduction of the reforms. Although this analysis can be biased by the usual endogeneity problems, it is likely to give us a more exhaustive picture of the British labor market and an indication of the possible presence of imperfections or technological rigidities.

Besides providing us with relatively "clean" experiments to test hours constraints, these three tax/benefit reforms (especially the WFTC program) have also been widely analysed in previous studies (Bingley and Walker 1997; Blundell et al. 2000b; Blundell and Hoynes 2004; Gregg and Harkness 2007; Francesconi and van der Klaauw 2007; Brewer et al. 2006). These studies have come up with broadly consensual evaluations of the reforms' effects on a number of outcomes, including employment and wages. None of these studies, however, focuses on changes in worked hours. Stewart and Swaffield (2004) examine the working hours of low-wage employees in the UK, but analyze the impact of the introduction of the National Minimum Wage in April 1999 rather than the impact induced by reforms that potentially changed the incentive to work a given number of hours per week. Their results indicate that the minimum wage had a negative effect on hours worked by low-wage women, although they do not show how single women with and without children have been differentially affected. In addition, neither these studies nor the earlier research on wage-hours packages analyse job changing behaviour as a mechanism to adjust hours of work or address the broader issue of labor supply adjustment.[4]

We find that the introduction of the WFTC reform in 1999 led to a substantial increase in single mothers' hours of work. The primary mechanism for this adjustment was through job changes rather than labor supply adjustments within a job. There is a good deal of heterogeneity in the effects of the WFTC reform, with evidence of even less adjustment within jobs emerging among single mothers whose youngest child was aged 0–4, and who worked in larger firms, service industries, and the public sector. The presence of some hours inflexibility within jobs is confirmed when we look at hours changes by stated labor supply preferences. Women who stated that they were unconstrained in their job showed the largest *upward* adjustments after the WFTC reform if they changed job. Similarly, and again in line with their stated preferences, overemployed women showed the largest *downward* adjustments after the 1992 FC reform (which reduced the minimum work requirement to receive FC from 24 to 16 hours a week) only if they changed job. Finally, we find relatively little effect on wages. However, there is some weak evidence that certain groups of women (especially single mothers who lived in London and the South East) operated under monopsonistic conditions, whereby changing job led to significantly lower wages after the introduction of WFTC.

Our research is likely to be relevant for many aspects of labor market policy, especially for the design of tax credit and benefit policies which specify a minimum number of hours of work per week as a precondition for entitlement to a given payment (e.g., the Working Tax

Credit, and the current pilot for the Employment Retention and Advancement Scheme in the UK). From the result that hours are not very flexible within jobs, we can infer that changes to the tax/benefit incentives to work a given number of minimum hours are likely to influence rates of job-to-job transitions for the affected groups of workers.

The next section briefly explains the rules and structure of the FC/WFTC programs, and discusses our estimation approach and identification strategy. Section 7.3 introduces the data, and describes the variables used in the analysis. Section 7.4 presents the empirical results, and Section 7.5 summarizes our main results.

7.2 "In-Work" Benefit Reforms in the United Kingdom

7.2.1 *Institutional Background*

Programmes to support low-income working families with children (hereafter called "in-work benefits", even though the more recent programmes are officially designated tax credits) have a long history in the United Kingdom. A peculiar feature of the UK's in-work benefits is that awards depend not just on the earned and unearned income and family characteristics, but also directly on (weekly) hours of work: since their inception, in-work benefits have only been available to families with children who usually work some minimum number of hours a week.[5]

Two in-work benefits were in operation during our sample period: Family Credit (FC), which existed from April 1988 until September 1999, and the Working Families Tax Credit (WFTC), which existed from October 1999 until March 2003.[6] In April 1992, the minimum work requirement in FC fell from 24 to 16 hours a week. This occurred between the first two waves of the BHPS. The impact of this reform on single parents' labor supply is ambiguous: those working more than 16 hours had an incentive to cut hours to (no less than) 16, while those previously working fewer than 16 hours had an incentive to increase their labor supply to (at least) the new cut-off. In 1995, there was another reform to Family Credit, in the form of an additional (small) credit for those adults working full time (i.e., 30 or more hours a week). This reform affected the labor supply decisions of lone parents in obvious ways: there was an increased incentive for those working less than 30 hours to increase their hours to 30, but an income effect meant that those already working at least 30 hours had an incentive to cut their hours worked to no less than 30.

The 1999 WFTC reform has a more complicated impact on labor supply. WFTC was more generous than FC in three ways: it had higher credits, particularly those for young children, families could earn more before the benefit began to be withdrawn, and it had a lower withdrawal/taper rate. Overall, the reform increased the attractiveness of working 16 or more hours a week compared to working fewer hours. But the last of the three aspects of the reform meant that the biggest income gains were experienced by families just at the end of the FC taper (i.e., families whose earnings had reduced their entitlement to FC just to zero), who tended to be working full time (Blundell et al. 2000b). The expected impact of the WFTC reform on lone parents' labor supply, conditional on working 16 or more hours, is as follows: (i) people receiving the maximum FC award will face an income effect away from work, but not below 16 hours a week; (ii) people working more than 16 hours and not on maximum FC will face an income effect away from work (but not below 16 hours a week), and a substitution effect towards work; (iii) people working more than 16 hours and earning too much to be entitled to FC but not WFTC ("windfall beneficiaries") will face income and substitution effects away from work if they claim WFTC (see Blundell and Hoynes 2004; Brewer et al. 2006).[7]

The occurrence of such reforms (i.e., the 1992 fall in hours requirement for FC, the 1995 additional credit for working full time, and the introduction of the WFTC program in 1999) means that we can divide our sample into three periods: (a) Autumn 1991 to (March) 1995, with the post-reform period (which in our analysis we label FC, i.e., under the in-work benefit regime of FC) covering the years 1992–94; (b) April 1995 to September 1999, with the post-reform period (labelled FC+) being defined over the years 1995–97; and (c) October 1999 to the end of the sample, with the post-reform period (labelled WFTC) being between 1999 and 2002.[8] In our empirical analysis we take advantage of each of these separate reforms: not only did they have the potential to affect single mothers' hours of work, but they also could have done so in opposite directions. However, although we use this three-group categorization, most of our analysis will only isolate the 1992 and 1999 reforms (as the additional credit under FC+ was small), and focus on the few years immediately following the introduction of each policy change.

7.2.2 Analytical Framework and Identification Issues

To assess whether female labor supply adjustments operate through job changes in response to exogenous changes in the incentives to work a given number of hours, we estimate four different specifications of a simple model of hours changes. We perform this assessment using a difference-in-difference method (Ashenfelter 1978; Heckman and Robb 1985): that is, we identify the FC and WFTC effects on single mothers' behaviour through the differential tax and benefit treatment that they receive as compared to a control group, which is given by single women without children.[9] The main identification condition underlying this approach is that, other than the introduction of the changes in in-work benefits, there are no contemporaneous shocks that affect the *relative* outcomes of the treatment and control groups. Therefore, identification relies on the assumption that variation in labor supply preferences of single parents be independent of the reforms conditional on the observed covariates and time effects.[10]

At the time of the introduction of the 1999 reform, however, there were other shocks that might have influenced single mothers' and childless women's labor supply differently. Three policy changes in particular could have interacted with the WFTC effects. First, there was an increase in basic child benefits under Income Support (the main welfare benefit, similar to the Aid to Families with Dependent Children or Temporary Assistance for Needy Families in the United States) between 1998 and 1999. In terms of labor supply, however, this increase implies a negative income effect that could lead to a downward bias in our effect estimates. Our estimates may then represent a lower bound of the true effect. Second, the National Minimum Wage (NMW) was introduced in April 1999 (Dickens and Manning 2004; Stewart 2004). The NMW might have affected both the extensive margin of labor supply (inducing inactive women to get a job) and the intensive margin (increasing the incentives for working women to work more hours). But such incentives presumably had the same impact on single mothers' behaviour as they did on single childless women's. The NMW-related shock, therefore, is not likely to have changed the employment outcomes of the treatment group differently from those of the control group.

Third, between July 1997 and October 1998, the British government launched a series of New Deal programmes intended to help different groups of low-income people move from welfare into work using a combination of intensive job search assistance and small basic skills courses (Blundell et al. 2004; Van Reenen 2004). One of such initiatives, the New Deal for Lone Parents (NDLP), was aimed at all lone parents in receipt of Income Support with children under 16 and whose youngest child was over five years and three months (from April 2000 this lower age cut-off was dropped to three).[11] Under NDLP, lone parents were assigned to a personal advisor, whom they were supposed to meet once every two weeks to receive advice on job vacancies, in-work benefits, childcare arrangements, training, and job search techniques. One interesting aspect of NDLP, which was largely shared with other New Deal programmes, was that involvement in the scheme and searching for work were entirely voluntary, and benefit entitlements did not depend on whether people decided to enter the scheme or not.[12] Single women without children were not involved in a similar initiative, unless they too were longer-term unemployed and had low income. Therefore, NDLP and any of the other New Deal schemes were likely to affect women – whether single mothers or not – only to the extent that they were unemployed. But since unemployed and out-of-the-labor-force women are excluded from our analysis (in fact, women must be employed for at least two consecutive years to be included in our sample; see Section 7.3), the influence of NDLP on hours *changes* is likely to be limited. In any case, as single mothers, on average, have less education and are more likely to be unemployed, we performed sensitivity checks that will be discussed in Section 7.4.3 by replicating our analysis using a more restricted control group, consisting of single childless women with low educational attainment.

Turning to the model specifications, let Δh_{it} denote the change in total (usual and overtime) weekly hours of work between year $t-1$ and year t; let d_{it-1} be a dummy variable that is equal to 1 if woman i is a lone mother at time $t-1$, and 0 otherwise; and let Q_{it} be equal to 1 if woman i changes a job between year $t-1$ and t, and zero otherwise. The four specifications are as follows:

$$\begin{aligned}\Delta h_{it} &= \alpha_0 + \alpha_1 d_{it-1} + \alpha_2 Q_{it} + \beta_{FC} d_{it-1} Q_{it} I(1992 \le t \le 1994) \\ &\quad + \beta_{WFTC} d_{it-1} Q_{it} I(1999 \le t \le 2002) + X'_{it}\gamma + \varepsilon_{it}\end{aligned} \tag{1}$$

$$\begin{aligned}\Delta h_{it} &= \alpha_0 + \alpha_1 d_{it-1} + \alpha_2 Q_{it} + (\alpha_{31} + \alpha_{32} d_{it-1})\delta(t) + \beta_{FC} d_{it-1} Q_{it} \\ &\quad I(1992 \le t \le 1994) + \beta_{WFTC} d_{it-1} Q_{it} I(1999 \le t \le 2002) + X'_{it}\gamma + \varepsilon_{it}\end{aligned} \tag{2}$$

$$\begin{aligned}\Delta h_{it} &= \alpha_0 + \alpha_1 d_{it-1} + \alpha_2 Q_{it} + (\alpha_3 + b_{FC} d_{it-1}) I(1992 \le t \le 1994) \\ &\quad + (\alpha_4 + b_{WFTC} d_{it-1}) I(1999 \le t \le 2002) + \beta_{FC} d_{it-1} Q_{it} I(1992 \le t \le 1994) \\ &\quad + \beta_{WFTC} d_{it-1} Q_{it} I(1999 \le t \le 2002) + X'_{it}\gamma + \varepsilon_{it}\end{aligned} \tag{3}$$

$$\begin{aligned}\Delta h_{it} &= \alpha_0 + \alpha_1 d_{it-1} + \alpha_2 Q_{it} + \alpha_{21} Q_{it} I(1992 \le t \le 1994) \\ &\quad + \alpha_{22} Q_{it} I(1999 \le t \le 2002) + (\alpha_3 + b_{FC} d_{it-1}) I(1992 \le t \le 1994) \\ &\quad + (\alpha_4 + b_{WFTC} d_{it-1}) I(1999 \le t \le 2002) + \beta_{FC} d_{it-1} Q_{it} I(1992 \le t \le 1994) \\ &\quad + \beta_{WFTC} d_{it-1} Q_{it} I(1999 \le t \le 2002) + X'_{it}\gamma + \varepsilon_{it},\end{aligned} \tag{4}$$

where $I(w)$ is a function indicating that the event w occurs; $\delta(t)$ in equation (2) is a linear time trend; X_{it} is a vector of individual characteristics measured either at $t - 1$ or between $t - 1$ and t; and ε_{it} is an i.i.d. error term. The variables in X described in detail in the next section, are a cubic polynomial in total work experience, dummy variables for race, educational qualification, firm size, public sector, region of residence, housing tenure, union coverage, and industry, the number and changes in the number of children by age group, and changes in health status and local unemployment rate.[13] The treatment effects for movers are captured by β_{FC} and β_{WFTC}, while b_{FC} and b_{WFTC} respectively capture the *FC* and *WFTC* treatment effects for workers who did not change job (stayers).[14]

The key differences across equations (1)–(4) involve the specification of time trends. In equation (1), time trends are not modelled, except those operating through β_{FC} and β_{WFTC}. Equation (2) instead allows for group-specific linear time trends (captured by α_{31} and α_{32}), while in equation (3), we have a more flexible specification with group-specific discrete jumps for stayers after both the 1992 and 1999 reforms (b_{FC} and b_{WFTC}). Finally, equation (4) introduces even greater flexibility by allowing different trends in job changing behaviour after each reform (through α_{21} and α_{22}). If $\hat{b}_j = \hat{\beta}_j$ (with $j = FC$, $WFTC$) we cannot statistically reject the hypothesis of within-job flexibility in hours choice, while if $\hat{b}_j$ is statistically smaller than $\hat{\beta}_j$ there is evidence of hours constraints within jobs.

Estimation of (1)–(4) is performed using ordinary least squares (OLS). However, because our regressions are in changes,[15] all individual time-invariant permanent unobservables that enter additively in the determination of hours levels are eliminated from the estimation. In computing the standard errors we take advantage of the fact that we have multiple observations over time, and thus we allow for arbitrary serial correlation.

7.3 Data

The data we use come from the first twelve waves of the British Household Panel Survey (BHPS) collected over the period 1991–2002. Since Autumn 1991, the BHPS has annually interviewed a representative random stratified sample of about 5,500 households covering more than 10,000 individuals. All adults and children in the first wave are designated as original sample members. On-going representativeness of the non-immigrant population has been maintained by using a "following rule" typical of household panel surveys: at the second and subsequent waves, all original sample members are followed (even if they moved house or if their households split up), and there are interviews, at approximately one-year intervals, with all adult members of all households containing either an original sample member, or an individual born to an original sample member whether or not they were members of the original sample. The sample therefore remains broadly representative of the population of Britain as it changes over time.[16]

Our estimation sample includes employed unmarried non-cohabiting females (separated, divorced, widowed and never married) who are at least 16 years old and were born after 1941 (thus aged at most 60 in 2002). Because equations (1)–(4) refer to changes in hours worked, we measure hours changes conditional on being in work in period $t - 1$ and remaining in work in period t. We exclude any female who was long-term ill or disabled, and in school full-time or self-employed or out of the labor force in a given year. The sample includes 2,284 women who have been observed working at least two consecutive times over the sample period and at some point were living alone, of whom 1,122 are lone mothers and the remaining 1,162 are childless. In line with the Inland Revenue's definition, a child must be aged 16 or less (or be under the age of 19 and in full-time education) to count as a dependent child for whom the single mother is responsible. Although only 16 percent of the women are observed in the same marital state for all the 12 years of the panel, about 60 percent of them are observed for at least 7 years in the same state. The resulting sample size, after pooling all available years for both groups of women, is 12,359 observations (4,585 on lone mothers and 7,774 on childless women). Of the 1,280 single women in the 1999 wave of interviews, 25 lone mothers and 32 childless women (about 4.5 percent of the sample in that year) were interviewed just before the day in which the 1999 reform was implemented (October 5th). To limit problems of interpretation, they were dropped from the estimating sample. Their inclusion however does not alter any of our main results.

Table 7.1 presents summary statistics of the outcomes and characteristics of the two groups of women, which we use as covariates in the empirical analysis below. There are some noticeable differences between the two groups.[17] Lone mothers are younger (30 versus 38 years), less educated (56 percent have qualifications short of A level versus 48 percent among childless women, and only 6 percent of lone

Table 7.1. Summary Statistics

Variable	Unpartnered women without children	Lone mothers
Total weekly hours of work	34.74	25.61
	(13.25)	(14.20)
Change in worked weekly hours[a]	0.39	2.25
	(11.40)	(12.31)
Absolute change In worked weekly hours[a]	6.00	6.91
	(9.62)	(10.28)
Hourly pay	7.06	5.85
	(6.01)	(5.19)
Monthly labor income conditional on working positive hours (in 2002 pounds)	1,110	694
	(911)	(629)
Age (years)	38.1	30.00
	(15.0)	(11.36)
Nonwhite	0.043	0.090
Registered disabled	0.049	0.023
Number of children by age group:[b]		
0–4		0.231
		(0.510)
5–10		0.588
		(0.755)
11–18		0.798
		(0.771)
House owner	0.578	0.541
In social housing	0.229	0.377
A level or higher educational qualification	0.520	0.438
No qualification	0.152	0.144
University degree or more	0.144	0.060
Total work experience (years)	14.33	8.67
	(11.47)	(7.88)
Employed in a firm with fewer than 50 workers	0.660	0.746
Employed in service industries[c]	0.838	0.820
Employed in the public sector	0.247	0.171
Union covered	0.514	0.530
Changed job during previous year	0.167	0.179
Local unemployment rate[d]	0.065	0.063
	(0.032)	(0.031)
Number of person-wave observations	7,774	4,585
Number of women	1,162	1,122

Notes: The figures are means (standard deviations in parentheses) computed over all person-wave observations for which two consecutive years of data are available.

[a] The change is measured over two consecutive years.

[b] Averages are computed over the entire subsample of lone mothers. If computed over the three specific subsamples of lone mothers with children in each child group, the means (standard deviations) are 1.172 (0.448), 1.318 (0.582), and 1.321 (0.548), respectively.

[c] 'Service industries' refer to banking, finance and insurance, distribution, hotels and catering, transport and communication, and other services (which include education and sanitary services).

[d] Computed over 306 travel to work areas.

mothers have a university degree versus 14.4 percent),[18] more likely to be non-white (9 versus 4.3 percent) and in social housing (38 versus 23 percent), less likely to be employed in the public sector (17 percent versus 25 percent), and have fewer years of work experience. The two groups of women are instead relatively similar in terms of job changing behaviour with 17 percent of childless women and 18 percent of single mothers moving across employers in any two given years.[19] Systematic differences emerge again in the case of labor market outcomes. Compared to unmarried women without children, lone mothers work about 9 fewer hours per week, earn £1.20 less per hour, and nearly £420 less per month, and report a larger change in worked hours from one year to the next (an increase of 2 hours and a quarter per week versus less than 25 minutes).[20]

To gain a greater insight into how the reforms might have changed the distribution of hours worked among all single women, Figure 7.1 plots histograms of total weekly hours of work for all women in the sample by survey year (with the vertical line in each panel indicating the eligibility hours cutoff).[21] Women can be found working any number of hours from 1 to 60 per week in any given year. In most years, we observe a great deal of variability with bunching at about 20, 30, and 35–40 hours and, depending on the year, at some hours between 40 and 50 (see also Blundell, Duncan and Meghir 1998). A striking feature is that, in every year up to 1996 (perhaps with the exclusion of 1991), there was only a small fraction of workers below the eligibility cutoff (accounting for about 15–20 percent of the single women in the sample), while in the two years prior to the 1999 reform there was clear evidence of bunching just below the 16-hour cutoff. From 1999 onwards, the fraction of female workers with total weekly hours between 16 and 20 was around 12 percent, almost twice as large as the fraction of workers in the same hours range between 1992 and 1998. These features of the data provides some quantitative indication of the hours effect associated with WFTC. From 1998 onwards, also discernible is a greater concentration of workers at 30 hours but we cannot detect any substantial change around the 30-hour cutoff immediately after the introduction of FC+.

We now turn to mean hours changes. Figure 7.2 plots the time trends for the year-on-year average changes in total hours worked over the sample period (with the dotted lines around the averages displaying the corresponding one-standard-deviation bands). Panel (a) shows the trends for all working women distinguishing lone mothers (straight line) from single childless women (dashed line), while panels (b) and (c) display the trends for female workers who moved between

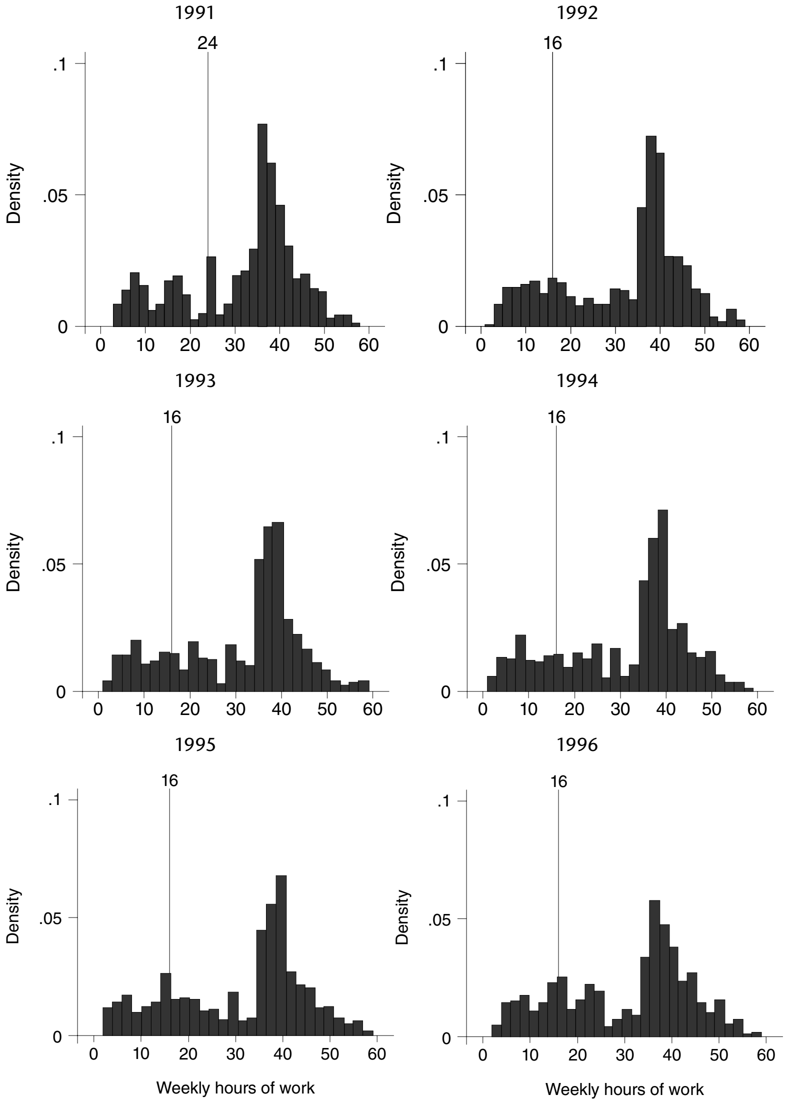


Figure 7.1. Female Weekly Hours of Work by Survey Year

Note: The vertical line indicates hours eligibility cutoff.

jobs and for workers who stayed with the same employer respectively. The data reveal that changes in hours worked among unmarried women without children are small and stable, ranging between 0 and 1 hour per week over the entire period (panel (a)). The mean hours changes for

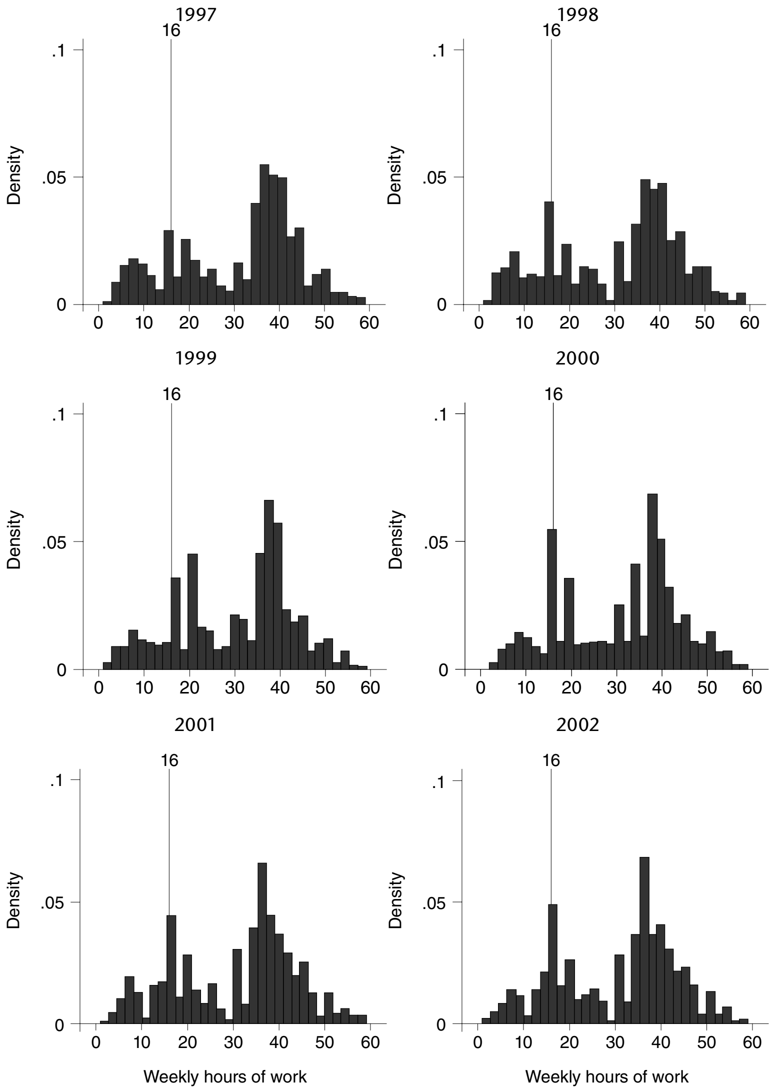
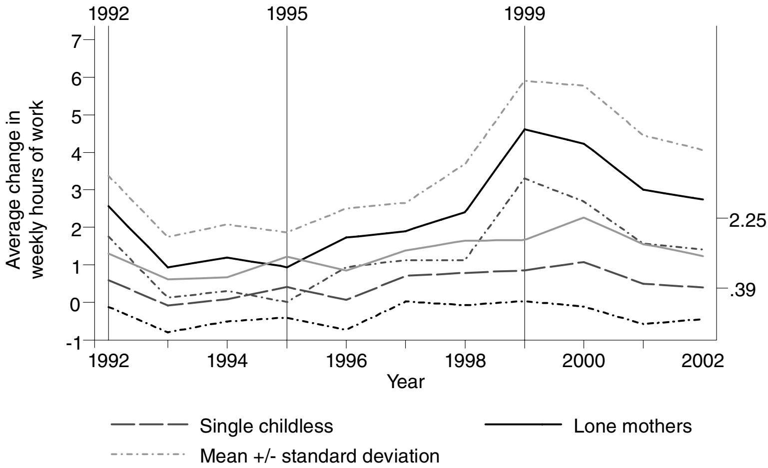
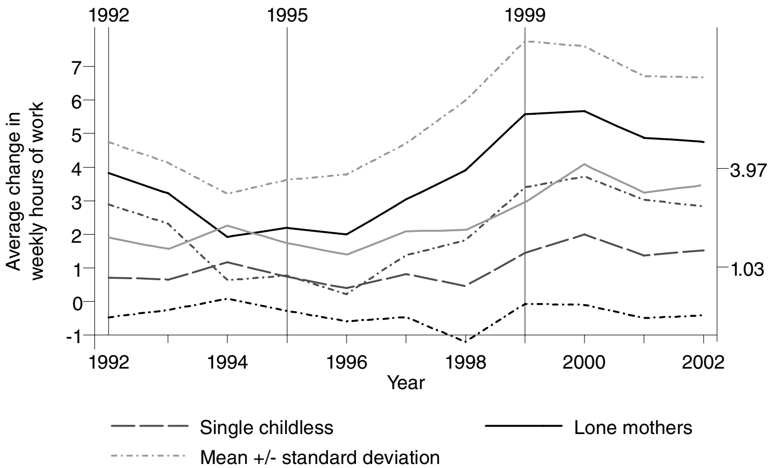


Figure 7.1. Continued

lone mothers instead are greater and their time variability is higher too. The largest hours change is observed after the introduction of the WFTC between 1998 and 1999, when lone mothers reported on average an increase of about 4.5 hours of work per week.[22] But, after 1999, lone mothers seem to have adjusted their hours changes downward. The



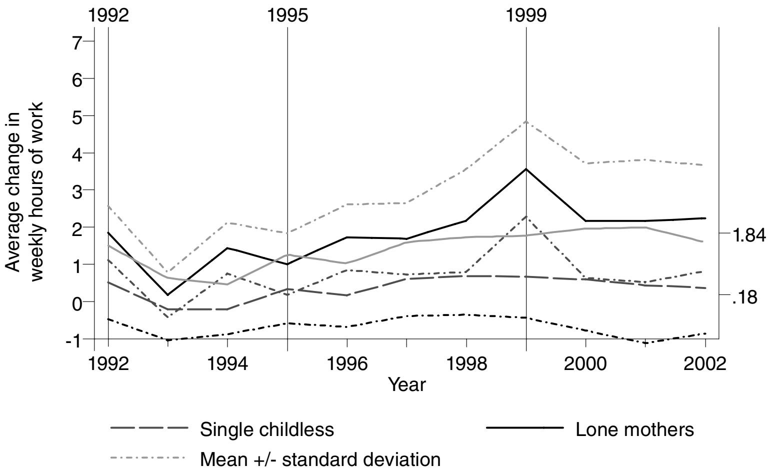
a) All Workers



b) Job Movers

Figure 7.2. Average Changes in Total Weekly Hours of Work Single Childless Women and Lone Mothers by Job Changing Status

1992 reform, which reduced the hours requirement for FC eligibility from 24 to 16 per week, increased single mothers' labor supply by about 2.5 hours, but again this increase was not followed by further increases in subsequent years. The additional FC for those working 30 or more hours does not appear to be associated with substantial changes in hours



c) Stayers

Figure 7.2. Continued

worked immediately after its introduction in 1995, but it is followed by a steady increase even before the peak between 1998 and 1999.

Panel (b) shows that the largest changes are experienced by women who moved between jobs, with lone mothers reporting an average change in hours of about 4 per week over the whole sample period and unmarried women without children of 1 per week. The time patterns for lone mothers are similar to those reported in the previous panel, although the peak in 1998–99 is followed by a further increase over the subsequent year. Lone mothers' increase in hours between 1991 and 1992 is also sizable, with an average close to 4 hours per week. Hours changes among those who stayed with the same employer instead are much smaller for both groups of women, especially for women without children (panel (c)).

7.4 Results

7.4.1 *Benchmark Estimates*

The estimates of the impact of job changing behaviour on hours changes are shown in Table 7.2. These are presented for the four specifications described in Section 7.2.2, and separately for the cases in which the variables in X are excluded or included.[23]

The regressions without controls indicate that changing job is associated with increases in women's labor supply by less than one hour per

Table 7.2. The Impact of the In-Work Benefit Reforms and Job Changes on Hours Changes

	Without controls				With controls			
	(1)	(2)	(3)	(4)	(1)	(2)	(3)	(4)
α_1	1.58	1.27	1.54	1.36	0.34	0.44	0.24	0.19
	(7.94)	(2.68)	(3.51)	(2.98)	(1.08)	(0.85)	(0.59)	(0.32)
α_2	0.77	0.74	0.73	−0.19	−0.30	−0.32	−0.31	−0.45
	(1.86)	(1.78)	(1.76)	(0.28)	(0.64)	(0.70)	(0.71)	(0.82)
b_{FC}			−0.25	0.03			−0.03	−0.21
			(0.40)	(0.07)			(0.01)	(0.44)
b_{WFTC}			0.16	0.45			0.20	0.56
			(0.33)	(0.89)			(0.42)	(0.94)
β_{FC}	0.11	0.72	0.95	0.48	0.21	0.83	0.89	0.44
	(0.08)	(0.51)	(0.67)	(0.29)	(0.15)	(0.59)	(0.62)	(0.28)
β_{WFTC}	2.56	2.66	2.48	3.39	2.54	2.65	2.60	3.42
	(2.46)	(2.56)	(2.29)	(2.82)	(2.51)	(2.63)	(2.47)	(2.92)
Number of observations	12,359				12,359			

Notes: Absolute values of t-statistics (obtained from standard errors that are adjusted to reflect multiple observations per person) are in parentheses. The labelling of columns (1)–(4) corresponds to equations (1)–(4) described in the text.

Source: British Household Panel Survey 1991–2002.

week (α_2) although this effect is only significant at the 10 percent level in the first three specifications, while single mothers experience significantly larger changes of about 1.5 hours per week (α_1) The treatment effects for stayers (b_{FC} and b_{WFTC}) are small and never statistically significant, and so are the average treatment effects for job movers after the 1992 reduction in hours requirement under $FC(\beta_{FC})$. But the introduction of *WFTC* had a strong impact on job movers with a significant increase of their labor supply by 2.5–3 hours per week on average. Importantly, from specification (4) we can reject the hypothesis that $\hat{b}_{WFTC} = \hat{\beta}_{WFTC}$ at the 5 percent level (the *p*-value of the *t*-test of equality is 0.024), which provides evidence of hours inflexibility within jobs. Most of these results are robust to the inclusion of the control variables *X*, with the only exception of α_1 which now becomes statistically insignificant.[24] With 16 and 20 percent of women changing job after the 1992 and 1999 reforms respectively, we can derive their overall effects on hours changes conditional on working: *FC* had virtually no overall impact, while *WFTC* increased single mothers' weekly hours of work by about 1.1 hours (specification (4)).

Because Figure 7.2 reveals that stayers also increased their worked hours immediately after the 1999 reform, the previous analysis was repeated after excluding the last two years of the sample. Indeed, the *WFTC* effect for stayers is now larger and close to one extra hour per week, but its *p*-value is never below 0.11. In any case, even after this selection, all other results are confirmed, including the rejection of the hypothesis of flexibility in hours within jobs.[25] Thus, in response to the exogenous change in work incentives given by the *WFTC* program, changing job seemed to have been the strongest mechanism of labor supply adjustment among single mothers after 1999.

We repeated the previous analysis with a different subsample, in which the control group of single childless women is limited to those with educational qualifications below A level. In the spirit of the discussion in Section 7.2.2, this allows us to see whether our results are concentrated in specific subgroups of the population who might have been affected by other policy initiatives (such as the New Deal for Lone Parents) that were introduced at approximately the same time as the *WFTC* reform. It also provides us with an important sensitivity check. Restricting our analysis to this different control group reduces only slightly the treatment effect for movers under *WFTC* (β_{WFTC}) to 3.28 (*t*-ratio = 2.86), changing neither the estimated effect for movers under FC nor the treatment effects for stayers in any significant way. The results illustrated so far, therefore, are robust to this change in the definition of the control group.

7.4.2 *Heterogeneous Responses*

It is possible that the labor supply responses to the policy reforms vary by observable characteristics of the women in the treatment and control groups. To allow for this, we look for heterogeneous responses by estimating models that distinguish women separately by individual attributes (such as education and number and age of children), work related attributes (such as firm size and industry), and stated labor supply preferences. The results from these regressions (based on specification (4) only) are reported in Table 7.3.

The estimates in the first panel of the table reveal that the increase in hours worked after the 1999 reform was predominantly observed among single mothers who changed job and used to work fewer than 16 hours per week. Women in this group experienced a labor supply increase in excess of 6 hours per week while women who already worked 16 or more hours experienced a more modest growth of about 2 additional hours.[26] Both effects are significantly different from zero, and they are statistically different from each other at conventional levels (*p*-value = 0.006).

Table 7.3. Heterogeneous Responses by Selected Observed Groups – Specification (4)

	α_1	α_2	b_{FC}	b_{WFTC}	β_{FC}	β_{WFTC}	Obs.
A. Previous hours worked							
Fewer than 16 per week	−0.17	1.88	−0.19	0.26	2.84	6.39	3,117
	(0.14)	(3.54)	(0.31)	(0.18)	(1.92)	(2.52)	
16 per week or more	0.32	−1.24	−0.23	0.68	−0.57	2.27	9,242
	(1.25)	(2.89)	(0.34)	(1.47)	(1.55)	(2.51)	
B. Education[a]							
Less educated	−0.31	−0.64	0.51	0.53	3.61	1.92	6,297
	(0.61)	(0.82)	(0.85)	(0.88)	(2.15)	(1.26)	
More educated	0.56	−0.25	−0.45	0.58	−2.65	4.89	6,062
	(0.83)	(0.39)	(0.23)	(0.77)	(1.52)	(2.88)	
C. Number of children[b]							
One child	0.60	−0.97	0.47	0.39	−0.47	4.15	6,427
	(1.07)	(1.49)	(0.78)	(0.64)	(0.44)	(2.95)	
Two or more children	−0.29	−0.16	−0.57	0.53	1.98	1.37	5,932
	(0.66)	(0.27)	(0.88)	(0.84)	(0.92)	(1.31)	
D. Age of youngest child[c]							
0–4 years	−0.33	−0.27	−0.59	0.30	−0.51	3.82	5,438
	(0.37)	(0.36)	(0.78)	(0.19)	(0.45)	(2.84)	
5 years or more	0.62	−0.76	0.03	0.64	0.97	2.80	6,921
	(1.09)	(1.45)	(0.82)	(1.05)	(0.56)	(2.63)	
E. Firm size							
Fewer than 50 employees	−0.01	−0.38	0.47	0.61	0.37	2.18	8,553
	(0.02)	(0.61)	(0.84)	(1.08)	(0.25)	(1.84)	
50 employees or more	0.74	−0.93	−0.38	−0.16	0.72	4.20	3,806
	(1.03)	(1.52)	(0.38)	(0.13)	(0.41)	(2.64)	
F. Industry[d]							
Services	0.72	−0.84	−0.48	0.33	−0.82	3.74	9,262
	(1.12)	(0.97)	(0.28)	(0.51)	(0.88)	(3.15)	
Manufacturing	−0.40	0.23	0.30	0.86	1.57	2.95	3,097
	(0.74)	(0.15)	(0.48)	(1.27)	(1.91)	(2.77)	
G. Sector[e]							
Private	−0.08	−0.26	0.09	0.55	1.08	3.09	9,659
	(0.19)	(0.27)	(0.16)	(0.91)	(0.54)	(2.43)	
Public	0.94	−0.88	0.43	−0.18	−0.52	4.10	2,700
	(1.14)	(1.52)	(0.68)	(0.15)	(0.76)	(2.71)	
H. Labor supply preferences[f]							
SAME=1	0.07	−1.21	0.61	0.84	0.43	4.20	7,539
	(0.17)	(1.58)	(0.90)	(1.52)	(0.74)	(2.97)	
OVER=1	1.65	−3.92	−0.53	−0.35	−6.74	1.09	3,090
	(1.19)	(2.86)	(0.86)	(0.66)	(1.84)	(0.31)	
UNDER=1	0.31	4.37	−0.24	−0.50	2.97	2.87	1,730
	(0.42)	(2.73)	(0.15)	(0.27)	(1.44)	(1.64)	

Notes: Absolute values of t-statistics (obtained from standard errors that are adjusted to reflect multiple observations per person) are in parentheses. All regressions include the control variables used in Table 2, except for: education (panel B); number and change in the number of children by age group (panels C and D); firm size (panel E); industry (panel F); sector (panel G). The variables defining each of the observed groups are measured at time t–1. For other definitions see the notes to Table 2.

These results suggest that an important part of the overall treatment effect of the 1999 reform was driven by greater entry into *WFTC* eligible employment of already working single mothers. As the differences between $\hat{b}$ and $\hat{\beta}$ document, there is evidence of a greater degree of hours inflexibility within jobs after the *WFTC* reform for both groups of women and, for women who worked fewer than 16 hours per week, also after the 1992 *FC* reform (but this effect is significant only at the 6 percent level).

The treatment effects for stayers do not differ between more educated women and less educated women (panel B). There are however asymmetric responses among movers. Less educated single mothers increased their labor supply by 3–4 hours per week after the 1992 reform, while more educated single mothers' supply increased by 4–5 hours after the 1999 reform.[27] Equality tests of the estimated b and β coefficients can be rejected at the 5 percent level during the *WFTC* regime among the more educated, and at the 10 percent during the *FC* regime among the less educated.

The next two panels demonstrate that the post-*WFTC* upward adjustment in single mothers' labor supply is primarily experienced by mothers of one child aged 0–4. Albeit smaller, the effect observed for mothers of children aged 5 or more is still sizeable and significant (panel D). If we pool all women as we did in Table 7.2 and interact the variable on b_{FC} with the indicator of the youngest child being aged 0–4, this interaction term is negative and statistically significant (b_{FC} and s.e. = 0.48), while the interaction with the indicator of the youngest child being older is never significant. This provides evidence that the 1992 reform induced some groups of workers (in this case, single mothers of young children who did not change job) to reduce their hours worked over the 1992–94 period.

The UK in-work benefit system interacts with other welfare benefits (Blundell and Hoynes 2004). One of these is Housing Benefit,

[a] 'Less educated' is defined as having less than A level qualifications; 'More educated' is defined as having A-level or higher qualifications.

[b] 'One child' and 'Two or more children' pertain to lone mothers.

[c] 'Youngest child aged 0–4' and 'Youngest child aged 5 or more' refer to lone mothers.

[d] 'Services' includes banking, finance and insurance, distribution, hotels and catering, transport and communication, and other services (which include education and sanitary services). 'Manufacturing' includes energy, extraction, metal goods, other manufacturing industries, construction, and primary industries.

[e] 'Public sector' includes civil service, central and local government, National Health Service, education, and non-profit organizations.

[f] OVER = 1 if the respondent indicated that she would like to work fewer hours "assuming that [she] would be paid the same amount per hour", and equals 0 otherwise; UNDER = 1 if the respondent indicated that she would like to work more hours "assuming that [she] would be paid the same amount per hour", and equals 0 otherwise; SAME = 1 if the respondent indicated that she would like to continue to work the same number of hours "assuming that [she] would be paid the same amount per hour", and equals 0 otherwise.

which works as a rent subsidy. If a single mother receives Housing Benefit, she would benefit less from a given amount of tax credit because this is treated as income in other means-tested programmes. Rents in some parts of the country (in particular, London and the South East) are high and have rapidly increased over the 1990s, while owner-occupiers are not eligible to Housing Benefit. To capture part of the relationship between Housing Benefit and the tax credits of interest here, we stratified our sample by region of residence (London and the South East in one group and the rest of the country in the other) and by housing tenure (owner-occupier or not), both measured at $t-1$. For the sake of brevity, the results are not shown but are available from the authors. From this analysis it emerges that labor supply adjustments observed after the 1999 reform were greater for single mothers who lived *outside* the London/South-East region (where house rents are lower and the interaction with Housing Benefit is likely to be more modest), and who were not owner-occupiers.[28]

Job specific characteristics provide other important sources of heterogeneity for the impact of job changes on hours changes after the 1999 reform. The strongest treatment effects are found for single mothers employed in relatively larger establishments (of the order of 4 additional hours per week, panel E), in service industries (about 3 extra hours, panel F),[29] and equally for those employed in the private sector or the public sector (between 3 and 4 additional hours per week, panel G). Strong evidence of hours inflexibility emerges among lone mothers who work in larger firms, service industries, and the public sector.

Another important dimension along which we expect to see heterogeneous responses is given by stated labor supply preferences. At each interview, the BHPS asks respondents whether they would like to work fewer hours, or more hours, or continue to work the same number of hours "assuming that they would be paid the same amount per hour". We use this information to construct three labor supply preference variables for any given year of the sample period, labelled OVER (=1 if a worker would like to work fewer hours, and zero otherwise), UNDER (=1 if a worker would like to work more hours, and zero otherwise) and SAME (=1 if a worker would like to continue to work the same number of hours, and zero otherwise).[30] We expect that workers who are overemployed/underemployed at one point in time reduce/increase their worked hours over time, and those who want to continue working the same number of hours do not change their labor supply. The estimates on α_2 reported in panel H of Table 7.3 confirm such expectations, with overemployed workers reducing their labor supply by 3 hours per week on average, underemployed workers increasing it by about 4 hours, and the remaining group of workers showing no significant change. The 1992 and 1999 in-work benefits reforms did not affect hours worked by women who would have liked to keep working the same number of hours and did not change job. But single mothers who wanted to continue working the same number of hours showed large upward labor supply adjustments of about 4 hours per week if they changed job after the *WFTC* reform.[31] Thus, initially "unconstrained" (i.e., neither over- nor under-employed) lone mothers did respond to the greater work incentives of the *WFTC* program but only through a change of job.[32] This upholds our previous finding that there is evidence of hours inflexibility within jobs.

The 1999 reform also led to increases of 1–3 hours per week among both overemployed and underemployed workers who changed job, although none of such increases is statistically significant at conventional levels. After the 1992 reform, instead, we observe large (and significant at the 10 percent level) reductions of about 7 hours per week among overemployed single mothers who changed job. This lines up very well with the 8-hour fall in the minimum work requirement to receive *FC* (from 24 to 16 hours a week). Again, this labor supply adjustment occurs through movements across (rather than within) jobs, although equality tests of the estimated b and β coefficients can be rejected only at the 10 percent level, irrespective of the specification. Underemployed workers seem to be unable to adjust their labor supply upward if they did not change job. But those who moved did manage to increase their worked hours even after the 1992 reform by about 3 hours per week (although this increase is not statistically significant).[33]

We re-estimated variants of equations (1)–(4) over the whole sample of women which included interaction terms between the variables on b_j and $\beta_j (j = FC, WFTC)$ and stated labor supply preferences. The results from this analysis (not shown) confirm those previously discussed. In particular (from specification (4)), unconstrained single mothers who changed job after the 1999 reform increased labor supply by about 4 hours (t-value = 4.61), and overemployed single mothers who moved across jobs after the 1992 reform reduced their hours by about 7 a week (t-value = 2.41). Further interactions with indicators of the age of the youngest child reveal that mothers of younger children (aged 0–4) who moved jobs experienced the greatest changes in hours conditional on working. In particular, after the introduction of WFTC, unconstrained mothers whose youngest child was aged 0–4 and who changed job worked nearly 5.5 extra hours (t-value = 3.27) as opposed to 3 among unconstrained mothers whose youngest child was aged 5–18. Similarly, after the FC reform, overemployed single mothers with younger children reduced their labor supply by 9 hours a week as compared to 5.6 among mothers of older children.

7.4.3 *Sensitivity Analysis*

We performed a number of sensitivity analyses to demonstrate the robustness of the results. For the sake of brevity Table 7.4 presents the results only from three exercises using specification (4). The results obtained from the other specifications are qualitatively similar to those discussed here.[34]

First, we performed our analysis accounting for the 1995 *FC* reform that provided extra credit for full-time work. The estimates in

Table 7.4. Robustness checks – Specification (4)

	Accounting for 1995 reform that provided extra credit for fulltime work (FC+)	Length of time in the panel[a]		Propensity score matching models[b]	
		6 years or more	9 years or more	Biweight kernel matching	Local linear regression matching
	(i)	(ii)	(iii)	(iv)	(v)
α_1	0.30	0.57	0.66	0.97	1.05
	(0.69)	(1.20)	(1.13)	(0.89)	(1.36)
α_2	−0.56	−0.81	−0.60	−1.12	−0.20
	(0.88)	(1.56)	(0.88)	(1.53)	(0.51)
b_{FC}	0.12	−0.45	−0.27	−0.13	−0.38
	(0.11)	(0.80)	(0.40)	(0.11)	(0.75)
b_{FC}+ (extra credit for FT work)	−0.22				
	(0.44)				
b_{WFTC}	0.24	0.81	0.75	−0.49	0.97
	(0.48)	(1.54)	(1.02)	(0.90)	(1.13)
β_{FC}	0.81	0.65	0.17	0.36	0.53
	(0.52)	(0.50)	(0.23)	(0.48)	(0.62)
β_{FC}+ (extra credit for FT work)	1.18				
	(1.32)				
β_{WFTC}	3.48	4.52	2.99	3.12	3.07
	(3.13)	(3.34)	(2.39)	(3.07)	(2.74)
Number of observations	12,359	8,314	5,153	12,359	12,359

Notes: Absolute values of t-statistics (obtained from standard errors that are adjusted to reflect multiple observations per person) are in parentheses.
All regressions include the control variables used in Table 7.2. For other definitions, see the note to Table 7.2.

[a] The categories '6 years or more' and '9 years or more' include only women who have been observed for at least 6 years and 9 years consecutively in the panel respectively.
[b] Absolute values of t-statistics (with standard errors obtained from 500 bootstrapped replications) are in parentheses. For the local linear regression matching regression, the estimates are obtained after imposing a tricube kernel.

column (i) confirm our previous findings, and document that the 1995 reform was followed by no sizeable change in worked hours irrespective of whether women changed employer or stayed in the same job.

As mentioned in Section 7.3, there may be concerns with changing sample composition over time, differential attrition, and missing data. Besides using weighted data, which provided similar results to those presented so far, we addressed these concerns by re-estimating our models only on women who have been successfully interviewed for a given number of times (for example, six or more waves). If attrition or changing sample composition are important, the results from such selected subsamples are expected to differ from those discussed earlier. Columns (ii) and (iii) of Table 7.4 report the estimates found from two subsamples, one in which we include only women who have been observed for six or more years (i.e., at least half of the time between 1991 and 2002) and the other in which women have to be observed for at least 9 consecutive times. In general, the estimates from both subsamples are relatively close to the corresponding figures reported in Table 7.2. For example, β_{WFTC} one of the key parameters in our study, is estimated to be 32 percent greater (column (ii)) and 13 percent smaller (column (iii)) than its counterpart of Table 7.2. Despite such differences in magnitude, therefore, these estimates tend to support our previous results, suggesting that missing data problems are likely to have only minor consequences for our analysis.

Finally, we estimated the effects using propensity score matching (biweight kernel and local linear regression matching). Although, like standard OLS regressions, matching methods rely on a selection-on-observables assumption (Angrist and Krueger 1999), they limit the potential bias due to differences in the support of *X* between single mothers and women without children and the bias due to the difference between the two groups of women in the distribution of *X* over its common support (Heckman et al. 1998). The estimates in the last two columns of Table 7.4 display patterns that are very similar to those illustrated above in this section.

7.4.4 *Wage Estimates*

The evidence so far indicates that British single mothers responded to the greater work incentives of the 1999 in-work benefits reform by substantially increasing their hours of paid work, whereas the two previous reforms to FC seemed to have induced only minor labor supply effects. The strong labor supply adjustment in conjunction with the introduction of *WFTC* was primarily achieved through a change of employer rather than changes in hours within the same job. This finding suggests that single mothers face some form of hours inflexibility within jobs. Against this background, we analyse wage responses. Of course, in-work benefits reforms were directly designed to change the incentive to work specific hours leaving wages unaltered, while wage determination was affected more explicitly by the introduction of the National Minimum Wage, and both hours and wages were (and still are) under employers' control, and so our partial-equilibrium analysis is likely to provide biased estimates. Nonetheless, gauging wage responses is important because it gives us a more complete picture of the British labor market and some indication of the possible presence of labor market imperfections or rigidities in the matching technology. We therefore estimated equations (1)–(4) with log hourly wages (expressed in 2002 prices) as dependent variable and the same set of explanatory variables used before. A number of checks, which were performed to test the robustness of such specifications, led to results that have the same qualitative implications as those reported here.

For both job movers and stayers and both the 1992 and 1999 reforms, we find no significant wage effect. There is also relatively little effect heterogeneity across different groups of women. Two important exceptions however are single mothers who lived in London and the South East and those who worked in small establishments. Among the former group of women, changing job after the introduction of *WFTC* implied not only a labor supply increase of almost 3 hours per week (t-ratio = 3.11), but also a wage reduction of 2.7 percent (t-value = 2.23). Among the latter, changing job after the 1999 reform led to 1.5 percent lower wages (t-value = 1.51) and modest positive hours changes. Thus, despite the presence of hours inflexibility, the labor market generally operates quite competitively, although there is an indication of monopsony among some groups of single mothers.

7.5 Conclusions

By using three in-work benefits reforms during the 1990s in the UK, which either changed hours requirements to be eligible for the benefits or increased the attractiveness of working a given number of hours, we are able to assess the mechanism of labor supply adjustment among single women with children – the main target of these in-work benefit reforms. We find that the 1992 and 1995 *FC* reforms had modest impacts on single mothers' hours of work, but the introduction of the *WFTC* reform in 1999 had large positive effects on their number of hours of work. This increase is largely driven by women who changed job, suggesting that the mechanism of labor supply adjustments is between rather than within jobs. This lines up well with the estimates we get when we look at hours changes by stated labor supply preferences: unconstrained women who changed job showed the largest hours increases after the 1999 reform, and overemployed women substantially reduced their hours worked after the 1992 reform (which did reduce the minimum work requirement to receive *FC* from 24 to 16 hours a week) only if they moved across jobs. There is evidence of considerable heterogeneity in the effects of the *WFTC* reform for different groups of women. The strongest evidence of hours inflexibility within jobs emerged among single mothers whose youngest child was aged 0–4. This was especially the case for those who worked in larger firms, service industries, and the public sector. Although there is little in the way of overall wage effects, we do find that after the introduction of *WFTC* hourly wages decreased significantly for single women who lived in London and the South East and moved jobs and, to a lesser extent, for movers who worked in small firms.

So what remains of the canonical labor supply model? We have shown that adjustments in hours of work are made primarily by movements *between* jobs and there is little evidence of systematic labor supply induced hours movements *within* jobs. Our analysis of stated preferences confirms this further, showing that responses are greater among those who say they are unconstrained as well as among those who are constrained but state that they would like to move in the direction suggested by the incentives. Thus, a labor supply model emerges in which hours adjustments are largely made by moving between workplaces. This could be achieved within an "adapted" canonical model in which establishments are organized around hours requirements and individuals move jobs to achieve hours flexibility. Of course, it could also be supported by theories that emphasize the importance of labor market frictions and imperfections, such as job search, wage-job packages, and/or dynamic monopsony. However, if there were such 'imperfections' we would expect these to be displayed in wage responses. The evidence is that such responses are not large and overall not statistically significant. Consequently, at least to a first approximation, an adapted canonical labor supply model with hours flexibility across jobs cannot be rejected. Nonetheless, our results by region and firm size suggest that production technology or employer preferences may not only reduce labor supply flexibility within firms but may also place constraints on hours mobility across firms.

Part IV
Policy Applications

Introduction

One of the key motivations for the empirical analysis of individual and family labor supply behavior is to gain a better understanding of the impact and welfare consequences of tax and welfare reform. The importance of employment and hours responses to policy reform has remained a central aspect of policy debate over many years and there is no sign of any diminution in this interest.

In many developed economies there has been a move to introduce earned income tax credits or in-work benefits. The idea is to use an earnings subsidy to encourage work among those with low skills and/or high costs of work. Parents with low education would be a good example. Typically, their return to work in terms of hourly wages is low reflecting the low-education investment and they face relatively high 'fixed' costs of work, especially childcare costs. The UK has been a testing ground for the design and implementation of in-work benefits and the first three chapters of this part of the book address these policies.

One of the most important expansions and redesigns of an earnings tax credit system was the introduction of the Working Families Tax Credit (WFTC) in the UK in 2000. This replaced the existing tax credit called Family Credit (FC), which provided means-tested support for working families with children. Effective tax rates in the FC could be very high. Above a threshold, FC was tapered away at a rate of 70%. To be eligible to FC, families had to have at least one adult working 16 or more hours a week and have at least one dependent child. The minimum hours requirement is somewhat unusual and was retained in WFTC, placing a strong incentive on part-time work for low-wage working parents in the UK, something I return to in later chapters. The maximum credit depended on family circumstances and hours of work.

WFTC was effectively the same benefit as FC, but substantially more generous. This was for three main reasons: maximum awards were higher, the means-testing threshold was higher (rising in real terms by 10%) and awards were tapered away more slowly (55% rather than 70%). The increase in maximum awards was particularly large. For example, for a lone parent working 20 hours at the minimum wage with one child aged 4 and no childcare expenditure, the maximum rose by 25% in real terms. There was also a more generous treatment of childcare. The FC childcare disregard was replaced by a childcare credit worth 70% of childcare expenditure up to a limit of 130 pounds per week. This meant that the maximum award rose enormously for parents spending considerable amounts on childcare. The combined effect of these changes was to substantially increase awards for existing claimants and extend entitlement to new (richer) families.

The chapter entitled "The Labor Market Impact of the Working Families' Tax Credit" presents an ex-ante analysis of the labor supply impact of this policy reform. The model developed here used exclusively pre-reform data and with only the parameters of the reform known. There was no availability of any ex-post quasi-experimental analysis so developing a structural behavioral model was essential. It required the labor supply of single parents and parents in couples to be modeled, allowing for childcare, the details of the reform incentives and other costs of work. With the particular incentive at 16 hours per week, it was also essential to allow for part-time labor supply. Finally, not all eligible families take up the tax credit so a model of participation among eligible families had to be layered on top of the multinomial framework. These elements of the tax credit system and the comparisons with similar systems elsewhere are provided in the second chapter in Part IV "Work Incentives and 'In-Work' Benefit Reforms: A Review". The model is a mixed multinomial model that respects all the complexities of the tax and benefit system.

The estimated model was used to provide an ex-ante simulation of the impact of the reform on labor supply. It pointed to a significant but not large positive impact on the labor supply of single mothers while an offsetting reduction in the labor supply married women where the husband was in work. Overall the labor supply impact on total employment and hours was limited but there was a clear increase in the number of families with at least one working adult and a boost to incomes in low-earning families. That these predictions turned out to be very close indeed to the actual outcomes for these groups was somewhat comforting, given the pure ex-ante analysis. Moreover the model could explain important differences with the expansion of EITC in the US where a stronger increase in employment had been found. Essentially it was the expansion of out-of-work income at the same time as the WFTC reform together with the interaction between the tax credit system and the benefit system in the UK that produced a smaller impact in the UK (see Blundell and Hoynes 2004).

One clear comparison is between active labor market policies and earned income tax credits. Interestingly, the UK experienced both policies and has developed a mixed policy strategy of active labor market programs and tax credits. Typically active labor market programs have mixed positive incentives through job search to help those out of employment and time-limited employment subsidies to firms, with job search requirements and penalties. For families with children, tax credits for parents seem better structured and targeted to deal with the costs of work and the different incentives created by other parts of the tax and welfare system. For younger adults job search requirements matched with time-limited wage subsidies are argued to provide a better design.

In ‘Evaluating the Employment Impact of a Mandatory Job Search Assistance Program’, a careful analysis of an active labor market program involving job search help and firm based time-limited training subsidies are examined. This is the New Deal for Young People (NDYP) introduced alongside WFTC in the UK in the late 1990s. The study used a novel area based design to develop an evaluation method that can examine the importance of general equilibrium effects and spillovers to similar untargeted workers. It finds little evidence of spillover effects. It also finds little evidence of delaying tactics by unemployed workers who can only access the scheme’s incentive after a six month period of unemployment. There are some significant impacts on inflows into employment. But these are found to be largely to subsidized jobs, limiting the overall effectiveness of the NDYP policy.

This empirical evidence formed key information in trying to assess how different policies toward the employment of low-skilled workers, especially those with families and children. The next chapter “Welfare-to-Work: Which Policies Work and Why?” takes stock of these issues and lays out some of the key differences between a range of different policies towards low-income workers. In particular, are firm based subsidies or individual work based subsidies to be preferred? Should there be an incentive for part-time work or just full-time work? Should these benefits and tax credits be time-limited? Overall, in-work benefits targeted toward full-time work have a clear and important role in tax and welfare policy. Nonetheless, part-time work incentives, although supporting low earnings, provide little in the way of longer-term pay-off.[1]

The policy part of this volume closes with two chapters on optimal policy design. These are inspired by the Mirrlees approach to tax design. This approach also provided much of the theoretical underpinning for the Mirrlees Review, a substantive study of tax reform for which I was a commissioning editor and author (Mirrlees et al. 2011). In the chapter entitled "Earned Income Tax Credit Policies: Impact and Optimality: The Adam Smith Lecture 2005", I lay out the broad principles of this approach and place it in the context of the UK WFTC reform to the tax rate schedule for low-income workers. This analysis used the path-breaking work of Saez (2002) to line up the tax rate structure with estimated labor supply elasticities, drawing out the key distinction between responses at the extensive and intensive margins. Of course, in reality elasticities are not constant and vary by family type. Moreover the Saez analysis assumes negligible income effects whereas income effects for low-education women with children can be substantial. In the final chapter of the volume 'Employment, Hours of Work and the Optimal Taxation of Low Income Families', the optimal tax analysis is brought together with a detailed structural model of labor supply choices. Here the specification of non-linear tax incentives, costs of work, stigma effects of take-up, and unobserved heterogeneity is made explicit. The analysis suggests a role for tax credits but only for low-wage women with children around school entry age where the extensive responses are found to be largest.

8

The Labor Market Impact of the Working Families' Tax Credit

Richard Blundell, Alan Duncan, Julian McCrae and Costas Meghir

8.1 Introduction

In the March 1998 Budget, the UK Chancellor, Gordon Brown, announced the introduction of the Working Families Tax Credit (WFTC) as a replacement for Family Credit (FC), the UK's main in-work benefit. The structure of WFTC was modelled closely on the FC system, with the exception that WFTC was to be packaged as a refundable tax credit rather than as a welfare benefit. Among a range of stated aims, the government claims that WFTC 'will improve work incentives, encouraging people without work to move into employment …' (HM Treasury Press Release, 17 March 1998). This is to be achieved by boosting the in-work incomes available to families with children in low-wage jobs. Most of the extra resources are delivered by reducing the withdrawal rate of WFTC to 55 percent, down from the 70 percent taper found in FC. Additionally, WFTC contains generous provision to subsidize childcare costs.

The aim of this paper is to consider the effects on the labor market of the WFTC reform. In particular, we aim to study the impact of WFTC on hours of work and participation. There are effectively two target groups

The original version of this chapter was published as: Blundell, R./Duncan, A./McCrae, J./Meghir, C. (2000). The Labour Market Impact of the Working Families' Tax Credit, in: Fiscal Studies, 21(1): 75–103. © 2000 by the Institute for Fiscal Studies. The authors are grateful to Andrew Dilnot and John Vickers for useful comments, and to the Bank of England for financial support for this study. The paper forms part of the research programme of the ESRC Centre for the Microeconomic Analysis of Fiscal Policy at IFS. Data from the Family Resources Survey and the Family Expenditure Survey, made available by the Office for National Statistics through the ESRC Data Archive, have been used by permission of the Controller of Her Majesty's Stationery Office. The authors are responsible for all errors and interpretations.

for WFTC: single parents and married couples with children. Nearly 50 percent of currently working single parents are in receipt of some FC. For married couples with children, this proportion is smaller, at around 16 percent. However, the latter group is more than two-and-a-half times the size of the former. The behaviour of these groups, and their economic circumstances, are sufficiently different to warrant separate analysis.

These three requirements and the way they are developed to address the impact of WFTC are discussed in more detail below. However, it is worth pointing out two other key features of our simulations. First, we construct the budget constraint facing each individual, which accounts for different levels of tax allowances and marginal tax rates as well as the detailed interactions of the benefit and WFTC systems. Second, we account for childcare costs and availability in the simulation of the WFTC reform. This posed a particularly difficult challenge. Our approach, detailed further in the text and in the appendix, calculates entitlements to the childcare credit element of WFTC at different hours levels on the basis of observed patterns of childcare use and cost. We do not allow the childcare market to adapt to the introduction of WFTC. However, we do allow usage to vary with hours worked and with the demographic composition of the household. We nevertheless recognize that the move to WFTC may affect the childcare market in a manner that could have implications for the costs of the new program. To give some feel for these effects, we experiment with a number of alternative assumptions concerning the take-up of the childcare credit and the use of childcare.

Putting all these aspects together, we present what we believe to be a reasonably accurate analysis of the likely labor market impact of the WFTC reform. Of course, there are many other labor market reforms currently taking place, some of which will have a direct impact on these target individuals. These include the changes to National Insurance contributions, the minimum wage legislation and the working hours directive. To obtain a full picture, these should be accounted for. However, given the relatively large impact of the WFTC reform on the budget constraints of the target groups, we believe that this is the dominant reform and that our results will not be altered greatly by the spillover effects from other reforms.

The layout of the rest of the paper is as follows. In the next section, we outline the structure of the WFTC reform and contrast it with the existing FC system. An analysis of the characteristics of the WFTC target groups is followed in Section 8.4 by a detailed analysis of the potential incentive effects of WFTC, both for lone parents and for married couples with children. We then describe the data and report the results of our simulations of the impact of the WFTC reform for a range of household types. Finally, Section 8.7 concludes.

8.2 The Structure of Family Credit and the WFTC Reform

Family Credit is the main UK in-work benefit. It is designed to provide support for low-wage families with children that are working. A family with children needs to have one adult working 16 hours or more per week to qualify for FC. Each family is potentially eligible to a maximum amount, which increases with the number of children in the household and a small addition if they work full-time. This maximum amount is payable if the family's net income (after income tax and National Insurance contributions) is lower than a threshold (£79 per week in 1998–99). Net income in excess of this threshold reduces entitlement to FC from the maximum by 70p for every £1 of excess income. The basic structure of FC is shown in Figure 8.1.

Family Credit is payable on a six-monthly flat rate, regardless of changes in the claimant's circumstances, in order to minimize administrative and compliance costs and to hide the effects of the high withdrawal rate over the period of the fixed payment. It is paid to mothers even when the eligibility is in respect of the father's earnings. The average

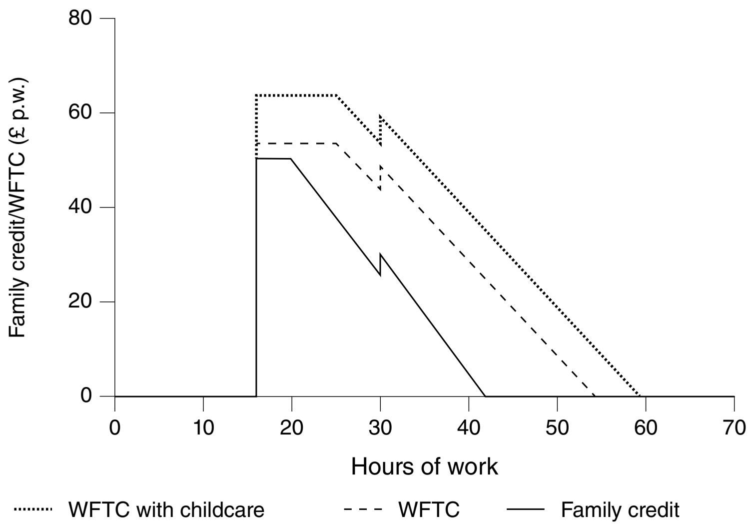


Figure 8.1. Increased Generosity of In-Work Support

payment at the end of 1996 was £57 per week. Take-up of FC after its introduction in 1988 was initially low but it has since increased. The present estimated take-up rates are that 69 percent of eligible individuals take up 82 percent of the potential expenditure.

The Working Families Tax Credit will be substantially more generous than FC. By the end of the century, the government expects to be spending £5 billion per year on WFTC, which is £1.5 billion more than was expected under FC. WFTC will increase the generosity of in-work support relative to the FC system in four ways:

- an increase in the credit for children under 11 from £12.35 to £14.85 per child;
- an increase in the threshold from £79 to £90 per week;
- a reduction in the taper from 70 percent to 55 percent;
- a childcare credit of 70 percent of actual childcare costs up to £150 per week.

The effects of these changes relative to FC are shown in Figure 8.1. Those currently receiving the maximum payment see a small increase in the level of their payment if they have children under 11. Those with net incomes between £79 and £90 move from being on the taper to receiving maximum support, and those who remain on the taper following the introduction of WFTC see their withdrawal rate fall from 70 percent to 55 percent. The largest cash gains from WFTC go to those people who are currently just at the end of the taper under FC but for whom the introduction of WFTC will create an entitlement to in-work support.

The final element of the increased generosity of the WFTC reform is the childcare credit. This replaces a childcare disregard in FC, which has suffered from very low levels of take-up. The credit increases the maximum amount of WFTC by 70 percent of childcare costs up to a maximum of £100 per week for those with one child or £150 per week for those with two or more children. The credit is available to lone parents and couples where both partners work more than 16 hours per week. The effect of the credit is also illustrated in Figure 8.1.

It should be noted that housing benefit and council tax benefit may also be payable at the same time as FC/WFTC. In these cases, the increase in the disposable income of a family may not be as large as the increase in the level of FC/WFTC payments, since entitlement to other benefits may be reduced. Such interactions with other benefits are taken into account in the budget constraint examples in Section IV and in all the modelled results.

8.3 Demographics

8.3.1 *Population Characteristics*

This section outlines some of the basic demographic characteristics of the target groups for the WFTC reform – lone parents and couples with children. Table 8.1 shows the number of each group that are in the population. Certain groups are excluded from the labor supply model set out in Section 8.5. These are the self-employed, students, the retired, those aged under 17 and those aged over 64. The self-employed are by far the most important of the excluded groups. They account for 850,000 of the excluded cases where the man works in a couple.

The participation rates for the target groups are 40 percent for lone parents, 82 percent for men in couples and 57 percent for women in couples. Figure 8.2 shows the distribution of hours for these groups. For men, the basic decision is one of whether to participate or not – virtually no men work part-time. Women, on the other hand, work at a wide range of hours levels. There are also clear indications of the impact of the benefit system on female labor supply, with a spike in the distribution of hours for lone parents at 16 hours – the cutoff for eligibility for Family Credit.

8.3.2 *Childcare Expenditure*

One key feature of the WFTC reform is the childcare element which could potentially have a large impact on labor supply decisions. For this paper, we concentrate on the childcare used by families where the youngest child is below school age.

Table 8.1. Numbers of Families with Children

Group	Number in population (thousands)
Lone parents	
Total population	1,600
Modelled population	1,550
Couples, man working	
Total population	4,550
Modelled population	3,500
Couples, man not working	
Total population	850
Modelled population	820

Source: Family Resources Survey, 1994–95.

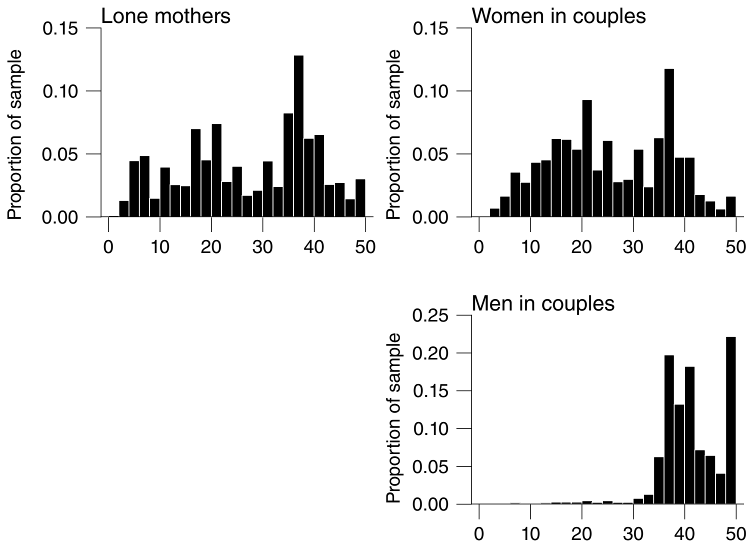


Figure 8.2. Hours of Work by Family Type
Source: Family Resources Survey, 1994–95.

Table 8.2. Type of Childcare Used When Youngest Child is Under 5

Type of care	Couples	Lone parents	All
No care reported	35.4	9.3	32.9
Relatives only	28.7	44.0	30.1
Relatives and friends combined	1.1	4.4	1.4
Friends only	3.0	9.8	3.6
Childminders only	11.2	11.1	11.2
Nursery care only	7.1	6.7	7.1
Childminders and informal combined	2.5	3.4	2.6
Nursery care and informal combined	4.3	7.5	4.6
Multiple formal care sources	3.4	1.0	3.1
Other forms of care	3.3	2.6	3.3
Total	100.0	100.0	100.0

Source: Family Resources Survey, 1994–95 and 1995–96.

Table 8.2 shows the type of childcare used by such families. For couples, over one-third report that they use no childcare, while just over 25 percent use relatives and 25 percent use formal childminders or nursery care. For lone parents, the largest difference is in the numbers reporting no childcare (below 10 percent) and a corresponding increase in those depending on relatives.

Table 8.3. Weekly Childcare Expenditure by Hours of Mother and Type of Care Couples

Type of care	1–10	11–20	21–30	31–40	>40	Total
Relatives only	1.25	2.50	6.20	13.41	15.26	7.00
Relatives and friends combined	15.00	5.17	9.39	25.82	a	13.06
Friends only	5.91	14.78	14.50	32.91	23.33	17.43
Childminders only	17.21	35.62	54.21	72.70	72.16	59.33
Nursery care only	40.57	47.53	60.58	82.81	66.58	67.56
Childminders and informal combined	15.33	29.05	41.41	53.85	55.00	41.36
Nursery care and informal combined	12.81	27.96	30.80	48.09	46.87	34.10
Multiple formal care sources	34.61	49.30	67.18	100.90	88.36	70.23
Other forms of care	35.00	64.69	33.27	124.19	119.69	108.85
Total	3.09	10.74	22.63	41.11	33.82	22.51

Source: Family Resources Survey, 1994–95 and 1995–96.

Table 8.4. Weekly Childcare Expenditure by Hours of Mother and Type of Care: Lone Parents

Type of care	Mother's hours of work (banded)					
	1–10	11–20	21–30	31–40	> 40	Total
Relatives only	0.82	2.34	5.27	15.94	9.00	5.75
Relatives and friends combined	3.33	6.33	18.00	50.00	—	16.94
Friends only	6.09	7.86	15.42	30.45	0.00	14.46
Childminders only	—	36.87	42.33	65.13	110.23	61.66
Nursery care only	15.75	9.67	48.03	64.53	66.50	48.39
Childminders and informal combined	—	34.13	13.67	55.60	48.00	40.38
Nursery care and informal combined	8.67	8.57	18.45	60.52	0.00	30.37
Multiple formal care sources	—	48.00	—	69.92	—	64.44
Other forms of care	0.00	42.50	71.00	57.50	138.00	60.85
Total	1.88	8.18	17.11	37.70	35.55	19.65

Note: Some cell size are too small for reliable figures to be produced.

Source: Family Resources Survey, 1994–95 and 1995–96.

Tables 8.3 and 8.4 show, for couples and for single parents, the weekly expenditure on childcare by the mother's hours of work and by type of childcare. Clearly, the expenditure rises with hours worked by the mother. It is also highly variable across types of childcare. For those on low hours, the total cost of care provided by relatives is minimal. The costs escalate for more formal types of childcare, such as nursery care and childminders, and also tend to be higher for couples than for lone parents.

8.4 Potential Incentives of WFTC

The Working Families Tax Credit is designed to influence the work incentives of those with low potential returns in the labor market. It does this via the increased generosity of in-work means-tested benefits. It is vital to be clear about the likely direction of the incentives that such changes generate. Any such increase unambiguously increases the financial returns to working a given number of hours relative to not working. But this does *not* imply that the reform will unambiguously increase either the total number of hours worked or the total number of labor market participants. It is well known that a change to marginal tax rates has an ambiguous effect on the number of hours worked by those currently in work – the income and substitution effects of the change move in opposite directions. Less widely appreciated is the fact that such changes also have an ambiguous effect on the numbers participating in the labor market. This is because in-work benefits, although nominally paid to one member of a family, are assessed on the basis of household rather than individual income. For second earners in couples, an increase in the generosity of in-work benefits can therefore introduce an incentive to stop participating in the labor market.

The aim of this section is to examine how we would expect WFTC to influence work incentives by examining hypothetical changes in budget constraints faced by 'typical' individuals. This will provide an intuitive explanation for the results generated by the labor supply model in Section 8.6. One should nevertheless guard against generalizing too widely at this stage on the basis of what are merely illustrative examples. We later report the results of a simulation study based on a large and representative sample of potential in-work benefit recipients.

8.4.1 *Incentives For Lone Parents*

For lone-parent families, the impact of WFTC is relatively straightforward, as we only have to consider the labor supply decision of one adult. As an illustration of the likely impact of WFTC on the work incentives of lone parents, Figure 8.3 presents the change in the budget constraint of a lone parent with one child. This constraint is constructed at the median hourly wage rate for working lone parents. We additionally assume that the lone parent is a social renter facing the median rent.

It is clear that the reform unambiguously enhances the probability of participation, as the financial returns to working any given hours level above 16 hours per week are greater post-reform than pre-reform.

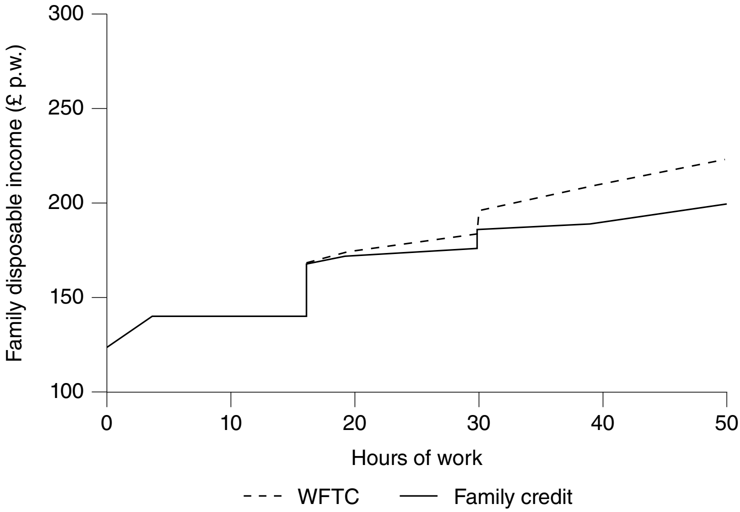


Figure 8.3. Budget Constraint for Example Lone Parent without Childcare Costs
Note: One child aged under 11. Hourly wage £4.39 (median for lone parents). Rent £41.10 p.w. (median for social renters with children). No childcare costs.

However, the increase in net income is small below 25 hours of work, due principally to the interaction of WFTC with the housing benefit system. For higher hours levels, the reduction in the WFTC taper starts to increase the returns to working.

For those already working, the labor supply response to the introduction of WFTC is less clear. The marginal tax rate is unambiguously reduced at all hours levels under the reform, though even with WFTC it remains high, at almost 70 percent. This increases the 'price' of non-market time (the marginal wage rate), causing individuals to consume less non-market time and therefore increase their hours of work – the standard substitution effect. The income effect, on the other hand, will be negative in hours of work (assuming non-market time is a normal good). The combination of the two effects leaves us with an ambiguous overall effect on hours of work.

Figure 8.4 shows the effect of the childcare credit element of WFTC. For illustration, we assume that childcare costs £1.96 per hour (the average hourly cost for those households in the mid-range in Table A1 of the appendix) and that the relationship between hours of work and hours of childcare is linear.[1] This graph is presenting *net* of childcare costs, and includes an entitlement to the childcare credit addition to WFTC. Two points are worth noting. First, the introduction of childcare costs *flattens* the budget constraint, since working more hours

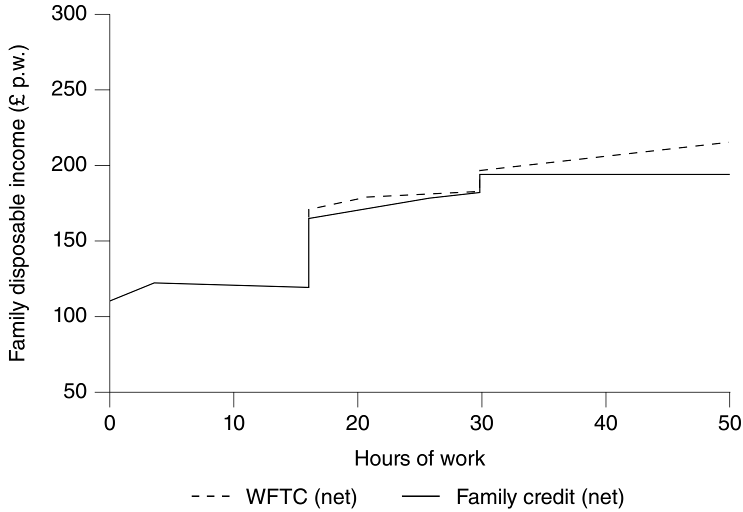


Figure 8.4. Budget Constraint for Example Lone Parent without Childcare Costs
Note: One child aged under 11. Hourly wage £4.39 (median for line parents). Rent £41.10 p.w. (median for social renters with children). Childcare at £1.96 per hour.

tends to imply greater childcare costs on average. Second, the relative reward to working at or beyond 16 hours increases once the WFTC compensation for childcare costs is introduced. One would therefore expect there to be an unambiguously positive participation response to the childcare credit element of WFTC over and above the standard credit, and an ambiguous hours response for those in work.

8.4.2 *Incentives for Men in Couples*

Figure 8.5 shows the effect on an example man in a single-earner couple. Again, the incentives are unambiguously to move into work. Indeed, the gains are far larger than for our lone-parent example, as the largest cash gains from the WFTC reform accrue to those at the end of the current taper. The incentives to change hours of work are again ambiguous. Nevertheless, there is a marked increase in the effective marginal tax rate for those who become eligible to WFTC as a result of the reform. This group face an increase in their marginal tax rates from 33 percent, produced by income tax and National Insurance, to just under 70 percent, produced by the interaction of the 55 percent WFTC taper on post-tax income. In the example, the marginal tax rate rises from 33 percent to just under 70 percent above 40 hours of work.

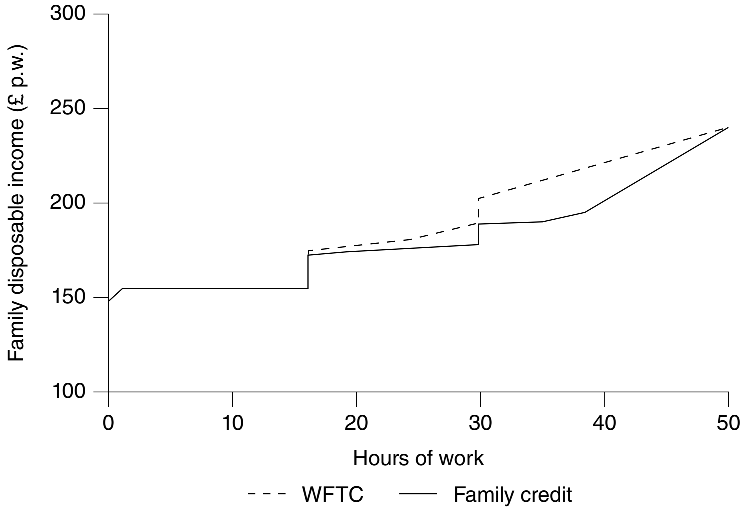


Figure 8.5. Budget Constraint for Example Man in Couple Without Childcare Costs

Note: Spouse not working. One child aged under 11. Hourly £5.87 (25th percentile for men in couples with children). Rent £41.10 p.w. (median for social renters with children). No childcare costs.

8.4.3 *Incentives for Women in Couples*

One point that is often neglected when considering the impact of increasing in-work means-tested benefits is that they can lead to incentives to move out of work altogether. Such an outcome relates particularly to secondary earners in two-adult households, many of whom are women. For illustration, Figure 8.6 shows the budget constraint for the partner of the man in Figure 8.5, conditional on him working 40 hours a week. Family income for a household in which the woman does not work therefore corresponds to the level of income at 40 hours on the man's budget constraint. The WFTC reform will increase household income for a non-working woman in a couple by an amount equal to the income difference at 40 hours for the man. As she moves into work (and household earnings increase), the income differential will fall as WFTC is withdrawn.[2] For the illustrative example shown here, any woman working more than 10 hours will have an increased incentive to reduce her labor supply or move out of work altogether.

The situation is a little different when we allow additionally for childcare costs. For couples, the childcare credit is only available if *both* partners work more than 16 hours per week. Figure 8.7 repeats Figure 8.6, but this time allows for the purchase of childcare at a cost of £1.96

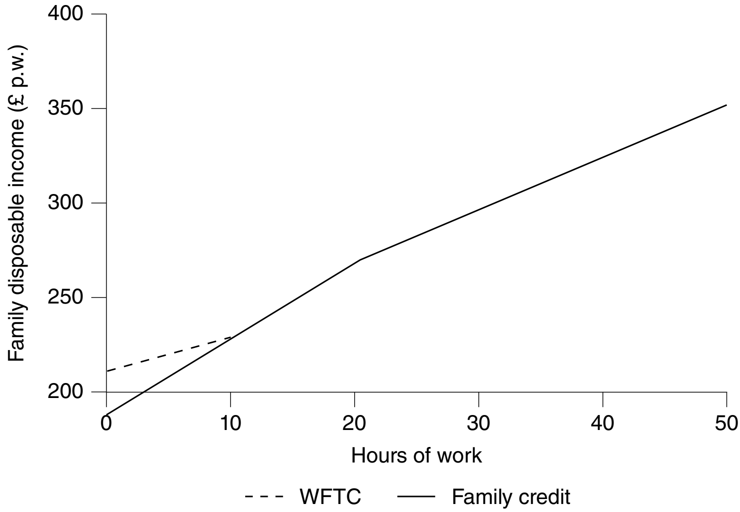


Figure 8.6. Budget Constraint for Example Woman in Couple without Childcare Costs

Notes: Spouse working 40 hours p.w. £5.87 per hour. One child aged under 11. Hourly wage £3.72 (25th percentile for women in couples with children). Rent £41.10 p.w. (median for social renters with children). No children costs.

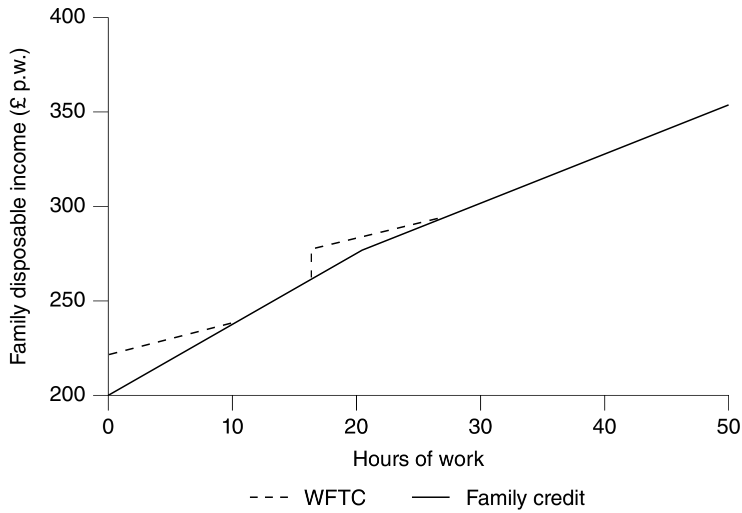


Figure 8.7. Budget Constraint for Example Woman in Couple with Childcare Costs

Notes: Spouse working 40 hours p.w. at £5.87 per hour. One child aged under 11. Hourly wage £3.72 (25th percentile for women in couples of children). Rent £41.10 p.w. (median for social renters with children). Childcare at £1.96 per hour.

per hour. Here, there is an additional incentive for the woman to work just over 16 hours to take advantage of the childcare credit.

8.5 WFTC Reform Simulations

8.5.1 *Data*

The Family Resources Survey (FRS) is used to examine the income and incentive impacts of tax credits. It is an annual budget survey of individual, family and household incomes, characteristics and labor market attachment. It is a new survey in the UK and contains a much larger and more representative sample of households than the surveys – principally the Family Expenditure Survey – on which the vast majority of previous work has been carried out. For this analysis, we use the 1994–95 and 1995–96 FRS datasets, together combining information on over 50,000 UK households. We select two samples for our empirical work – single-parent households and married or cohabiting couples. Excluding self-employed and retired households, together with students and those in HM Forces, that leaves samples of 1,807 single parents and 4,694 two-adult households for use in estimation.

To generate net incomes, we use the IFS tax and benefit model, TAXBEN. TAXBEN is a microsimulation model of the UK tax and benefit system which calculates taxes due and benefit entitlements for the Family Resources Survey data. It calculates the financial returns for each working-age individual to employment at all possible hours by calculating gross and net incomes at these levels. For workers, we assume their current wage remains unchanged; for non-workers, we estimate wages using a wage equation based on their characteristics. The use of TAXBEN combined with varying the potential hours of individuals allows us to generate highly accurate budget constraints for each individual in the survey in order to estimate the labor supply effect of the reforms to taxation.

8.5.2 *Modelling Childcare Costs*

The childcare credit component of WFTC could potentially offer generous benefits for those women who purchase some form of registered formal childcare. It is therefore important, both in estimation and in our simulations of the work incentive impact of the WFTC reform, that we account in some way for childcare expenditures. Ideally, we would like to observe actual childcare expenditures and childcare use among the full sample of women drawn from the FRS. Previous work[3] has established that up to 30 percent of non-working households may purchase formal childcare to some degree, which suggests that any model designed to assess work incentives in the presence of childcare ought to ideally include childcare expenditures among non-working households. However, these data are not available to us. Instead, we are forced to impute childcare expenditures because the FRS neglects to record childcare expenditures among non-working households.

Our strategy uses sample information on hourly prices of childcare and the relationship between formal hours of childcare and hours of work described earlier and set out in Table A2 of the appendix. We allow for the fact that similar households may purchase different types or levels of childcare by using information on the *distribution* of childcare prices paid by specific demographic groups directly in our estimation procedure. For each household in our sample, we calculate childcare expenditures at *each* price in the six-point distribution of childcare prices. By doing so, we are able to generate an average behavioural response over a range of possible values for the childcare credit component of WFTC.

8.5.3 *Non-Behavioural Effects*

The preceding analysis throws some light on the potential responses among different household types but fails to indicate the expected level of income gain following the introduction of WFTC. As with most benefit reforms, any increase in the generosity of either Family Credit or WFTC tends to be offset by consequent reductions in entitlement to other means-tested benefits.[4] The ultimate potential for reforms to improve work incentives may therefore be less than initially suggested by the particular structure of the reform.

To give some indication of the impact of introducing WFTC, Table 8.5 reports the proportion of households that are modelled to gain from the new credit among a sample of families drawn from the 1994–95 and 1995–96 FRSs. Data are split according to observed hours of work, marital status and age of children. We base these simulations on patterns of childcare expenditure observed in the FRS sample. Thus the reported figures *do not take account of* labor market responses to WFTC. Neither do they account for possible changes in the pattern of childcare use or cost. They nevertheless provide a benchmark against which we can judge the full behavioral impact of WFTC including the childcare credit element.

It is immediately evident that working lone parents are most likely to benefit from the WFTC reform. For example, nearly 80 percent of lone

Table 8.5. Proportion of Gainers from WFTC

	Hours of work (banded)					
	0	1–10	11–20	21–30	31–40	41+
Lone parents						
No pre-school children	—	—	62.1	74.0	52.2	51.1
One or more pre-school children	—	—	75.0	87.9	61.5	61.5
All women	—	—	65.2	78.2	53.8	53.4
Married, partner working						
No pre-school children	30.6	19.0	10.2	4.9	3.6	3.1
One or more pre-school children	35.9	12.7	11.7	5.3	4.4	4.1
All women	33.9	16.2	10.9	5.0	3.9	3.4
Married, partner not working						
No pre-school children	—	—	38.6	53.3	36.7	66.7
One or more pre-school children	—	—	73.1	80.0	45.0	33.3
All women	—	—	51.4	60.0	39.1	61.9

Note: Data are grouped according to observed hours of work for all household members and conditioned on observed childcare expenditure patterns.

Source: TAXBEN, based on Family Resources Survey, 1994–95 and 1995–96.

parents in part-time paid employment (of between 21 and 30 hours) will benefit from the new tax credit.[5] This proportion falls for those women in full-time paid employment, as fewer will have incomes low enough to qualify for FC or WFTC. No income gains are simulated among women on zero or low hours, since eligibility for FC or WFTC is contingent on working at least 16 hours per week.

Table 8.5 reveals an interesting pattern of gainers amongst women in couples where the male partner is in work. We find WFTC to be most generous to households in which the woman is not in paid employment: around one-third of this group will benefit from the introduction of the tax credit. For women in part-time work, the figure falls to around 5 percent. This feature of the reform may give rise to *negative* work incentive effects among women in couples, given that those in part-time employment (who are unlikely to see a financial benefit from the tax credit) are more likely to see an increase in their out-of-work incomes under WFTC. For women whose partner is unemployed, the pattern of increased eligibility is closer to that for the lone-parent population.

Among those lone-parent households that do gain, it is instructive to note from Table 8.6 that the greatest increases in income (of around £20 per week) fall to those in full-time employment, many of whom are new WFTC recipients not previously eligible for Family Credit. For married women, on the other hand, the income gains are spread more

Table 8.6. Average Income Gains from WFTC among Gainers

	Hours of work (banded)					
	0	1–10	11–20	21–30	31–40	41+
Lone parents						
No pre-school children	—	—	8.70	14.17	18.91	21.74
One or more pre-school children	—	—	12.67	19.59	26.70	20.67
All women	—	—	9.82	16.00	20.48	21.47
Married, partner working						
No pre-school children	19.12	19.32	17.95	17.57	19.86	19.72
One or more pre-school children	23.87	17.57	21.82	19.99	20.60	18.30
All women	22.27	18.73	19.72	18.44	20.13	19.24
Married, partner not working						
No pre-school children			12.42	16.45	20.57	24.33
One or more pre-school children	—	—	10.99	16.77	18.36	39.65
All women	—	—	11.66	16.56	19.83	25.51

Notes: As for Table 8.5. Each cell reports average gains (in pounds per week) among households who gain from WFTC. To recover average gains over all women, multiply each figure by the probability in the corresponding cell for Table 8.5.

Source: TAXBEN, based on Family Resources Survey, 1994–95 and 1995–96.

evenly across the hours distribution. Indeed, the greatest income gain falls to women not in paid employment whose partner is in work. Combined with the evidence from Table 8.5, this confirms the view that negative work incentive effects are entirely possible for this group of women.

8.5.4 *Simulating Labor Supply Responses*

To translate these potential income gains into simulated hours responses requires an economic model of labor market behavior among a range of household types. Our goal is to anticipate how families will respond to the change in economic circumstances brought about by the introduction of WFTC. For each household in our sample, we are able to generate the net income that would be associated with any choice of hours of work, both under the current tax and benefit system and following the introduction of WFTC. This information is used to simulate the labor supply responses to WFTC. Since our study involves the *simulation* of a tax reform that has yet to be implemented, rather than an *evaluation* of a tax reform already in operation, we require a structural model for household labor supply in order to proceed with the policy simulation.[6]

Budget sets faced by households in the UK are notoriously non-convex. The model specification chosen therefore needs to be able to deal adequately with non-linear taxes in estimation. The model ideally ought to account for labor supply decisions at the level of the household rather than at the level of the individual, and should include controls for fixed costs and child-related costs. These considerations lead us to choose a structural model of household preferences defined over a *subset of discrete hours points* rather than over a continuous hours range.[7] A precise specification of this model, together with detailed discussion of empirical estimates, can be found in Blundell et al. (2000a).

8.6 Simulation Results

We use the estimated models of household labor supply to simulate the work incentive impact of the WFTC reform. The IFS tax and benefit model, TAXBEN, generates household net incomes for each member of our FRS samples at each of the range of discrete hours choices {0,10,20,30,40}. For each individual, we compute the probability that they would be at each discrete hours point both before and after the WFTC reform, using the estimated models of preferences over work and household net incomes.[8] This enables us to simulate both the proportion of new workers and the change in the hours distribution.

We report in the main body of the text our simulations of the labor market impact of WFTC assuming 100 percent take-up of the childcare credit component of WFTC. We do, however, consider a number of alternative scenarios to establish to what degree the 100 percent take-up assumption dominates our results. Specifically, we examine alternative scenarios that assume zero take-up of the childcare credit component and a take-up rate of 50 percent. We must emphasize, however, that there is no model of take-up driving this latter scenario – we simply nominate a random 50 percent of our sample to receive the childcare credit component of WFTC.

Tables 8.7 to 8.10 report the results of our simulations for single-parent households, women in couples (separated according to the employment status of their partner) and men in couples under the assumption of 100 percent take-up of WFTC. These results are presented in the form of a matrix of simulated transitions between no work, part-time work and full-time employment under the two systems for women, and transitions between no work and (implicitly full-time) work for men. We also report summary measures of changes in participation rates and average changes in hours among workers and among the full sample. The simulated movements in hours for male and female members of a two-adult household derive from a joint labor market model, the estimates for which are reported by Blundell et al. (2000a). We report the *marginal* hours movements for each household member, on the basis that the likelihood of a *joint* labor market response to WFTC was found to be negligible in simulation.

8.6.1 *Policy Simulations: Lone Parents*

In Table 8.7, we report the simulated work responses to WFTC among our sample of single parents. The most common simulated transition takes around 2.2 percent of the sample from no work to either part-time or full-time work, with no offsetting movements out of the labor market. To take account of sampling variability, we simulate a standard error of 0.42 percent around this figure, which would admit the possibility that the actual increase could be as much as 3 percent. One can clearly see the reason for this shift in our earlier graphs of the potential impact of WFTC on single parents' budget constraints. At or above 16 hours per week, the single parent becomes eligible for

Table 8.7. Simulated Transitions among Single Parents (100% Take-up of WFTC)

a) Transitions

Pre-reform	Post-reform			Pre-reform %
	Out of work	Part-time	Full-time	
Out of work	58.0	0.7	1.5	60.2
Part-time	0.0	18.6	0.5	19.1
Full-time	0.0	0.2	20.6	20.7
Post-reform %	58.0	19.4	22.6	100
Change (% points)	–2.2	0.3	1.9	

b) Summary

	Mean	Standard deviation
Change in participation	+2.20%	[0.42%]
Average change in hours (all)	+0.75	[0.16]
Average change in hours (workers only)	+0.22	[0.04]
Average hours before reform (all)	10.20	
Average hours before reform (workers only)	25.70	

Notes: Transitions tables built by drawing 100 times from the distribution of unobserved heterogeneity and allocating each observation to the cell that yields maximum utility for each draw. Standard deviations for each transitions cell and summary measure are simulated by drawing 100 times from the estimated asymptotic distribution of the parameter estimates and, for each of those 100 parameter draws, applying the method described above to build transitions matrices.

WFTC (with any childcare credit addition to which she may be entitled). For some women, this extra income makes a transition to part-time employment attractive. Nevertheless, the level of the aggregate behavioral response is perhaps lower than one might have anticipated, given the potential cost of the WFTC reform; the 2.2 percent of single parents who are simulated to move into the labor market would gross up to around 30,000 women in the population.

We see a minor offsetting reduction in labor supply through a simulated shift from full-time to part-time employment among 0.2 percent of the sample. This is consistent with a small (negative) income effect among some full-time single women, for whom the increase in income through WFTC encourages a reduction in labor supply. Nevertheless, the predominant incentive effect among single parents could be said to be small but positive.

8.6.2 *Policy Simulations: Women with Employed Partners*

For married women, the simulated incentive effect is quite different. In Table 8.8, we report estimates of the transitions following WFTC among a sub-sample of women with employed partners. What we find is a significant overall *reduction* in the number of women in work, of

Table 8.8. Simulated Transitions among Married Women with Employed Partners (100% take-up of WFTC)

a) Transitions

Pre-reform	Post-reform			Pre-reform %
	Out of work	Part-time	Full-time	
Out of work	32.2	0.1	0.1	32.4
Part-time	0.3	31.6	0.0	32.0
Full-time	0.4	0.1	35.0	35.6
Post-reform %	33.0	31.3	35.2	100
Change (% points)	0.6	−0.1	−0.4	

b) Summary

	Mean	Standard deviation
Change in participation	−0.57%	[0.06%]
Average change in hours (all)	−0.18	[0.02]
Average change in hours (workers only)	−0.03	[0.005]
Average hours before reform (all)	17.34	
Average hours before reform (workers only)	25.65	

Notes: As for Table 8.7.

around 0.57 percent, equating to a grossed-up figure of around 20,000 in the population. This overall reduction comprises around 0.2 percent who move into the labor market following the reform and 0.8 percent who move from work to non-participation. The number of hours worked by women with employed partners is predicted to fall slightly, by 0.18 hours on average over the full sample.

The predominant negative response is clearly not one that is intended, but from our earlier budget constraint analysis one can easily see why it happens. There will be a proportion of non-working women whose low-earning partners will be eligible for WFTC. The greater generosity of the tax credit relative to the current system of Family Credit increases household income. This increase in income would be lost if the woman in the household were to work. For those women currently in the labor market, WFTC increases the income available to the household if she were to stop working.

8.6.3 *Policy Simulations: Women with Unemployed Partners*

In Table 8.9, we look at incentives for a sub-sample of women whose partners do not work. For this group, we simulate a significant overall

Table 8.9. Simulated Transitions among Married Women with Unemployed Partners (100% take-up of WFTC)

a) Transitions

Pre-reform	Post-reform			Pre-reform %
	Out of work	Part-time	Full-time	
Out of work	56.8	0.4	0.9	58.1
Part-time	0.0	22.2	0.4	22.6
Full-time	0.0	0.1	19.2	19.3
Post-reform %	56.8	22.8	20.5	100
Change (% points)	–1.3	0.2	1.1	

b) Summary

	Mean	Standard deviation
Change in participation	+1.32%	[0.16%]
Average change in hours (all)	+0.46	[0.067]
Average change in hours (workers only)	+0.14	[0.017]
Average hours before reform (all)	10.04	
Average hours before reform (workers only)	23.96	

Notes: As for Table 8.7.

increase of 1.32 percent in the number of women who work, equating to a grossed-up figure of around 11,000 in the population.

The reason for this shift is more straightforward, and stems from the increased generosity of basic WFTC relative to the current Family Credit system for those women who choose to move into work. Note that, for this group, the generosity of the childcare credit component of WFTC is not an issue, since households only qualify for the childcare credit if both adult household members work 16 hours or more. There is, of course, potential for both members of an unemployed couple to move into work in order to qualify for WFTC including the childcare credit, but the joint simulation shows that such an outcome is virtually non-existent, comprising less than 0.1 percent of the full sample of couple households.

8.6.4 *Policy Simulations: Men in Couples*

Turning to male labor supply responses to WFTC, we find the simulated net effect to be small (at around 0.1 percent overall). This nevertheless conceals a greater level of response once we disaggregate further. Table 8.10 reports the simulation results for men in couples and highlights

Table 8.10. Simulated Transitions among Married Men (100% take-up of WFTC)

a) Transitions

	Post-reform		Pre-reform %
Pre-reform	Out of work	Part-time	
Out of work	19.6	0.4	20.0
Working	0.3	79.8	80.0
Post-reform %	19.9	80.1	100
Change (% points)	−0.1	0.1	

b) Summary

	Mean	Standard deviation
Change in participation	+0.07%	[0.03%]
Average change in hours (all)	+0.03	[0.01]
Average hours before reform (all)	32.11	
Average hours before reform (workers only)	39.06	

Notes: As for Table 8.7.

two distinct responses which broadly balance out in aggregate. First, we see 0.4 percent of the male sample (grossing up to around 13,000 in the population) moving into work to take advantage of the increased generosity of the new tax credit. Since the majority of unemployed men in couples have a non-working partner (around 58 percent in our sample), this response contributes to an overall increase in the proportion of single-earner households. The second major response sees around 0.3 percent of the sample of men (equivalent to a grossed-up figure of 10,500) moving out of work. All in this group have an employed partner and are therefore choosing to move out of employment to take advantage of the increased generosity of WFTC on the basis of the female partner's earnings.

8.6.5 *Policy Simulations: A Sensitivity Analysis*

To gain some feel for the sensitivity of our simulation results, we consider a number of alternative modelling assumptions and scenarios. The simulations presented above assume 100 percent take-up of Family Credit, the Working Families Tax Credit and the additional childcare credit component of WFTC. Although the take-up of FC has been on the increase, these assumptions are certainly open to question. With this in mind, we consider two alternative scenarios. As a first experiment, we assign a random 50 percent of our sample to receive the childcare credit component of WFTC. The second alternative forces zero take-up of the childcare credit component of WFTC for the full sample. In some sense, this alternative scenario gives some feel for the work incentive effects of basic WFTC relative to FC rather than the cumulated incentive effect of full WFTC including the childcare component. A third scenario responds to the suggestion that the childcare credit might have an impact on the childcare market. We are currently unable to model an expansion in demand for childcare places following the WFTC reform. However, we can conduct a rough experiment on the supposition that the childcare credit increases demand for childcare places and bids up the price of childcare. This we do by altering the distribution of prices faced by each member of the sample, to an extent that increases the average hourly price paid for childcare by 50 percent. This hypothetical shift in no sense represents our view of what will actually happen in the childcare market; it merely serves to illustrate how price changes might affect behavioural responses to the childcare credit element of the WFTC reform.

As a final sensitivity test, we consider the use of alternative measures of the hourly wage for those observed out of employment. In a standard selectivity-adjusted model, the wage offers for those out of employment are, on average, below the observed wages for those in work.[9] As an alternative, we consider the recent work by Gregg, Johnson and Reed (1999). This studied the impact of WFTC on labor supply using a transitions model of labor market entry in which the relevant wage rates assumed for current non-workers had been predicted from a sample of labor market entrants drawn from the Labour Force Survey. Although their modelling strategy differs in a number of respects from ours,[10] the sensitivity of simulated responses to the assumed wage among potential labor market entrants is clearly an issue worthy of closer scrutiny. With this in mind, we compare simulation results from the bench-mark model with those from an alternative empirical model estimated using entry wage predictions for non-workers along the lines suggested in Gregg, Johnson and Reed (1999).[11]

In Table 8.11, we present simulations of labor market transitions under these alternative scenarios for each of the three groups of women (single parents, women with employed partners and women whose partners remain out of the labor market). For single parents, we find an increase in participation of around 1.3 percent with no take-up of the childcare credit compared with 2.2 percent assuming full take-up. For women with employed partners, we see a greater proportion (0.81 percent) moving out of work. This suggests that the additional childcare credit component does improve the incentive to enter the labor market.

Our sensitivity analysis also shows that an increase in the price of childcare serves to offset the effectiveness of WFTC in encouraging labor market participation. This is because the price rise increases the cost of working, with the childcare credit component only partially offsetting that change in cost.

The comparative results under the alternative entry wage assumption are interesting. We find an increase of around 0.8 of a percentage point in the proportion of single parents moving into work in response to WFTC when (typically lower) entry wage rates are assumed for non-workers, with smaller effects among women with an unemployed partner. This result can be rationalized on the basis that higher wages among those who have chosen not to work tend to imply less elastic preferences and less responsiveness to changes in incomes in work. For women with employed partners, the negative response to WFTC is also amplified.

Table 8.11. Simulated Response to WFTC: Alternative Scenarios

	Simulated responses to WFTC (%)				Summary measures		
	Non-work to work	Work to non-work	Part-time to full-time	Full-time to part-time	Δ partic. (%)	Δ in hours, all	Δ in hours, workers
Single parents							
1. Full take-up of childcare credit	2.2	0.0	0.5	0.2	+2.20	+0.75	+0.22
2. 50% take-up of childcare credit	1.8	0.0	0.5	0.2	+1.75	+0.60	+0.13
3. Zero take-up of childcare credit	1.4	0.0	0.4	0.2	+1.34	+0.47	+0.15
4. Increased childcare price	1.9	0.0	0.5	0.2	+1.76	+0.64	+0.20
5. Entry wage for non-workers	3.0	0.0	0.6	0.1	+2.99	+1.04	+0.30
Women, partner employed							
1. Full take-up of childcare credit	0.2	0.7	0.0	0.1	−0.57	−0.18	−0.03
2. 50% take-up of childcare credit	0.2	0.9	0.0	0.1	−0.73	−0.23	−0.04
3. Zero take-up of childcare credit	0.2	0.9	0.0	0.1	−0.81	−0.25	−0.04
4. Increased childcare price	0.2	1.9	0.0	0.4	−1.79	−0.59	−0.15
5. Entry wage for non-workers	0.2	1.1	0.0	0.2	−0.94	−0.28	−0.05
Women, partner unemployed							
1. Full take-up of childcare credit	1.3	0.0	0.4	0.1	+1.32	+0.46	+0.14
2. 50% take-up of childcare credit	1.3	0.0	0.3	0.1	+1.26	+0.43	+0.13
3. Zero take-up of childcare credit	1.3	0.0	0.3	0.1	+1.24	+0.43	+0.12
4. Increased childcare price	1.0	1.3	0.3	0.3	−0.30	+0.02	+0.04
5. Entry wage for non-workers	1.6	0.0	0.5	0.1	+1.61	+0.59	+0.27

Notes: As for Table 8.7. Summary measures are as follows: net change in participation (Δ partic.); average change in hours across the entire (sub-)sample (Δ in hours, all); and average change in hours among the working sub-sample (Δ in hours, workers). The 'bench-mark' state (1) involves full take-up of WFTC including the childcare credit (as presented in the main text in Tables 8.7 to 8.10). Alternative scenarios to this bench-mark include: a 50 percent (randomly assigned) chance of take-up of the childcare credit component of WFTC (2); zero take-up of the childcare credit component of WFTC (3); a simulated increase in the price of childcare (of around 50 percent on average), implemented by a progressive shift in the distribution of childcare prices (4); and a replacement of the standard wage rate predictions for non-workers by entry wage estimates derived from Labour Force Survey data along the lines suggested by Gregg, Johnson and Reed (1999) (5).

8.7 Conclusions

Using the Family Resources Survey, we estimate a model of family labor supply for married couples and individual labor supply for single parents. The model allows for childcare costs that vary with hours of work and takes into account the existence of many different types of childcare. Having validated the model in terms of its consistency with basic economic principles, we use it to simulate the labor supply effects of the new Working Families Tax Credit, assuming that the structure of the childcare market will not change. Given full take-up for the existing Family Credit and other benefits and full take-up of the new tax credit, we find that:

- the participation rate for single mothers increases by 2.2 percentage points, which corresponds to 34,000 individuals;
- the participation rate for married women with employed partners decreases by 0.57 percentage points (corresponding to 20,000 individuals) because of an income effect arising from the improved benefit eligibility of their husbands;
- the above behavioral effects combined with those for men and for women with non-working partners imply a small increase in overall participation of about 30,000 individuals.

These behavioural effects act to reduce the cost of the WFTC program compared with the no-behavioral, full take-up outcome, by 14 percent. When we simulate other take-up scenarios, we find small shifts in the level of participation. The question of the potential for changes in the childcare market remains an open one. It is not likely to lead to very large participation effects but could have important expenditure implications.

Appendix

Table A1. Hourly Childcare Costs among Women in Paid Employment

	No cost	Percentage in price range (lower bound, upper bound):				
		£0.00, £1.25	£1.25, £1.75	£1.75, £2.25	£2.25, £3.25	£3.25+
Married women						
One child, youngest <3	37.00	11.91	18.18	19.46	8.71	4.74
One child, youngest 3+	45.45	12.17	14.96	12.02	9.68	5.72
Two children, youngest <3	37.83	19.14	16.54	15.93	7.81	2.76
Two children, youngest 3+	49.22	15.37	12.46	11.11	7.37	4.47
Three+ children, youngest <3	43.98	25.90	11.45	7.83	5.42	5.42
Three+ children, youngest 3+	54.46	15.69	11.38	7.69	7.38	3.38
Single parents						
One child, youngest <3	46.15	21.98	15.38	6.59	6.59	3.30
One child, youngest 3+	53.22	15.79	13.16	9.94	4.68	3.22
Two children, youngest <3	52.73	27.27	10.91	1.82	7.27	0.00
Two children, youngest 3+	60.20	16.84	11.73	6.63	3.06	1.53
Three+ children, youngest <3	48.28	24.14	17.24	6.90	0.00	3.45
Three+ children, youngest 3+	57.14	28.57	1.30	5.19	1.30	6.49
Total	45.85	15.92	14.27	12.45	7.39	4.13
Average price within range	—	£0.79	£1.48	£1.96	£2.62	£5.20

Table A2. Relationship between Hours of Work and Hours of Childcare (per Child)

	Constant	t-value	Slope	t-value
Married women				
One child, youngest <3	1.589	1.57	0.885	26.75
One child, youngest 3+	4.335	3.03	0.384	8.84
Two children, youngest <3	2.321	2.43	0.674	20.17
Two children, youngest 3+	2.196	3.41	0.222	10.50
Three+ children, youngest <3	0.094	0.06	0.503	9.22
Three+ children, youngest 3+	3.103	3.82	0.139	5.16
Single parents				
One child, youngest <3	1.264	0.61	0.984	13.76
One child, youngest 3+	7.825	5.23	0.264	5.67
Two children, youngest <3	2.210	1.05	0.702	9.36
Two children, youngest 3+	6.900	5.65	0.141	3.63
Three+ children, youngest <3	5.228	1.72	0.350	3.31
Three+ children, youngest 3+	2.813	1.35	0.242	3.14

9

Work Incentives and 'In-Work' Benefit Reforms: A Review

9.1 Introduction

Any policy designed to alleviate poverty among families is open to the potentially harmful disincentive effects of the welfare trap. The typically steep withdrawal rate of benefit income as family income increases, creates a high implicit tax rate on earned income for the targeted group. Consequently, financial incentives to work are reduced, often considerably so. Indeed, in many income-support systems in developed countries there is effectively a pound for pound loss of welfare income as earnings rise. In addition, in-kind transfers such as free medical services, free dental care, free medical prescriptions, and subsidized housing are often lost with a move into employment. Thus, although supporting low incomes, welfare systems typically reduce the economic incentives to seek work. Recent decades have seen little increase in the real return to work for low-skill individuals. This again acts to reduce the financial incentives to work for such individuals. Low attachment to the labor market is then enhanced through low earnings in work and a benefit system in which benefits are reduced rapidly on entering work.

In-work benefits or earned income tax credits are designed as a method of poverty relief that does not create adverse work incentives. They do this by targeting low-income families but enforcing a work-contingent eligibility rule. Typically, eligibility is also based on the presence of

The original version of this chapter was published as: Blundell, R. (2000). Work Incentives and 'In-Work' Benefit Reforms: A Review, in: Oxford Review of Economic Policy, 16(1): 27–44. © 2000 by Oxford University Press. This research is part of the program of research at the ESRC Centre for the Microeconomic Analysis of Fiscal Policy at the Institute for Fiscal Studies. I am grateful to Alan Duncan, Hilary Hoynes, Andrew Glyn, Costas Meghir, Tim Jenkinson, Paul Gregg, Julian McCrea, and Ian Walker for helpful discussions and comments.

children, reflecting in part the higher welfare benefits for families with children and partly the costs of childcare etc. Consequently they are most heavily targeted toward single parents, but are also available to low-income couples with children. Increasingly, they are also being discussed for low-income workers with or without children.

In terms of poverty alleviation they are much more effective and better targeted than minimum wages and, in contrast to the usual welfare instruments, have fewer adverse work incentives. In fact, the positive incentives to work created by in-work benefits are often a central motivation for their introduction. However, with eligibility based on family income there are also likely to be some adverse work incentives for individuals in families with more than one potential worker. Also, in common with negative income tax schemes, the benefit reduction rate can create incentives for a lower effort or a lower supply of hours among workers.

In this review attention is focused on three specific in-work benefit systems that together cover most of the aspects and issues surrounding 'in-work' benefits: the Earned Income Tax Credit (EITC) in the USA; the Family Credit (FC) and Working Families Tax Credit (WFTC) in the UK; and the Self-Sufficiency Project (SSP) in Canada. It is worth emphasizing that in-work benefits are but one part of the welfare and tax system in each of these countries and should not be analyzed in isolation. In the UK the interaction between in-work benefits and housing benefit is crucial in understanding the overall impact of any reform.[1] Similarly, the overall impact of any reform will depend on the interactions with other benefits and taxes, as well as other concurrent reforms. For example, in the UK the introduction of the minimum wage and changes to National Insurance contributions are both likely to have an impact.[2]

The EITC is one of the oldest of these programs and was originally developed in the 1970s as a way of introducing a negative income tax for poor US families, which, in contrast to negative income programs, involved a work test. The EITC provides a tax credit supplement to earnings, which increases proportionally with earned income until a maximum credit or maximum income limit is reached. With the falling real wages of the low-educated in the USA over the 1980s, and the increasing level of welfare dependency among certain demographic groups, the EITC took on a new role in welfare policy as a mechanism for encouraging work by supplementing the working wage for low-wage workers.

This dual motivation of providing an income supplement for low-wage families, together with an incentive to work, also was behind the introduction of FC in the UK in 1988 and also the evolution of FC into WFTC in the late 1990s. The SSP in Canada is purely an experimental policy reform which not only requires a work test, but is also only available to long-term welfare recipients and is time-limited. The experimental nature of the SSP makes it particularly useful in assessing the effectiveness of these reforms in inducing welfare recipients to move into work.

The design and implementation of 'in-work' benefits is not altogether straightforward. There are a number of important issues: the overall size of the credit; whether or not some minimum hours limit should be placed on eligibility; how rapidly to reduce the credit as earnings or income rise beyond the maximum limit; whether the amount of credit should be based on individual or family earnings; and how to offset childcare costs. These issues are addressed in this paper within the overall discussion of the workings of the EITC, FC/WFTC, and the SSP.

The nature of these three in-work benefits also highlights the alternative evidence-based methods economic researchers have developed for evaluation. In many ways the most convincing method of evaluation is a randomized social experiment in which there is a control (or comparison) group which is a randomized subset of the eligible population. Such is the design of the SSP and the recent work of Card and Robbins (1998) has used the experimental nature of this reform to assess its effectiveness in inducing welfare recipients into work.

Of course, experiments have their own drawbacks. First, they are rare and typically expensive to implement. Second, they are not amenable to extrapolation. That is, they cannot easily be used in the ex ante analysis of policy reform proposals. Finally, they require the control group to be completely unaffected by the reform, typically ruling out spillover, substitution, and equilibrium effects on wages etc. Nonetheless, they have much to offer in enhancing our knowledge of the possible impact of such reform.

Another popular method of evaluation is the natural experiment approach. This considers the reform itself as an experiment and tries to find a naturally occurring comparison group that can mimic the properties of the control group in the properly designed experimental context. This method is also often labeled 'difference-in-differences' since it is usually implemented by comparing the difference in average behavior before and after the reform for the eligible group with the before and after contrast for the comparison group. In the absence of a randomized experiment, this approach can be seen to recover the average treatment effect on the treated by removing unobservable individual effects and common macro effects. It does, however, rest on strong assumptions and, again, cannot easily be used to extrapolate policy reform proposals.[3] The impact of the EITC reforms in the USA, very influential on recent reforms in the UK, has been studied extensively by this method. In particular, the important studies by Eissa and Liebman (1996) and Eissa and Hoynes (1998) purport to show important impacts of the reforms on the labor market.

Reforms to Family Credit in the UK provide some basis for ex-post evaluation of in-work benefit reform in the UK and below some evidence is presented on labor supply responses in the UK using these past reforms. However, to evaluate new reform proposals, such as the WFTC in the UK, requires an ex-ante evaluation. Although we can draw on the quasi-experimental and natural experiment findings, the evaluation of the WFTC reform requires a model that can simulate the choices of individuals as their earnings opportunities in work change. For this we draw on the recent WFTC evaluation study by Blundell et al. (2000b). This develops a statistical simulation model of family labor supply behavior based on individual survey data and incorporates a complete description of each family's budget constraint and childcare availability. To do this requires an accurate description of the budget constraint for each individual, whether they currently work or not. In turn this requires an assessment of the likely market wage for individuals who may now enter work.[4]

The remainder of this paper provides a brief, but systematic, review of these issues. In Section 9.2 some of the patterns of wages and employment among in-work benefit target groups are discussed. Section 9.3 then provides a summary of 'in-work' policy reforms using the quasi-experimental evidence from Canada, the natural experiment evidence from North America, and the existing evidence on Family Credit reforms in the UK. Section 9.4 presents the evaluation of the WFTC reforms and describes some of the issues in the design of such reforms. Finally, Section 9.5 concludes.

9.2 Earnings and Employment Among Low-Skilled Workers

Before discussing the particular types of reform, it is useful to place these reforms within the context of the current economic environment. Important changes have occurred in the low-skill labor market in many countries and it is only within this new economic environment that one can begin to understand how this kind of welfare reform is going to operate.

There has been a remarkable shift in returns to education and skill in many countries. For example, in the USA real earnings for the lowest education groups have fallen yearly since the late 1970s. This characteristic is quite exaggerated in the USA, but it is nonetheless common to most developed countries and is reflected in a growing concern for poverty among lower-skilled working families. This fall in returns also has the effect of making work relatively less attractive, especially where such individuals are eligible for relatively generous out-of-work benefits.

Whether or not the low levels of employment among certain target groups, such as single parents, can be attributed to this low pay-off to work, it certainly cannot be argued to have helped. Although there has been an increase in employment among women, this is misleading and covers up the dramatic rise in non-employment of single women with children.

Figure 9.1 shows the relative rise in non-employment among single women with children. Moreover, as a proportion of the working-age population, this group has risen threefold to 17 percent in the UK over the last three decades. At the same time there has been an increase in non-employment for low-skilled men, especially those in families and aged over 45 (see Blundell and Johnson (1998) for example). The declines in employment of these two groups have focused attention on the growing number of children in 'worker-less' families.

It is these simple labor-market facts that have focused policy attention on 'in-work' benefits for low-wage families. The aim is to make work more attractive for families with children whose current labor market opportunities (the low real wages and the relatively

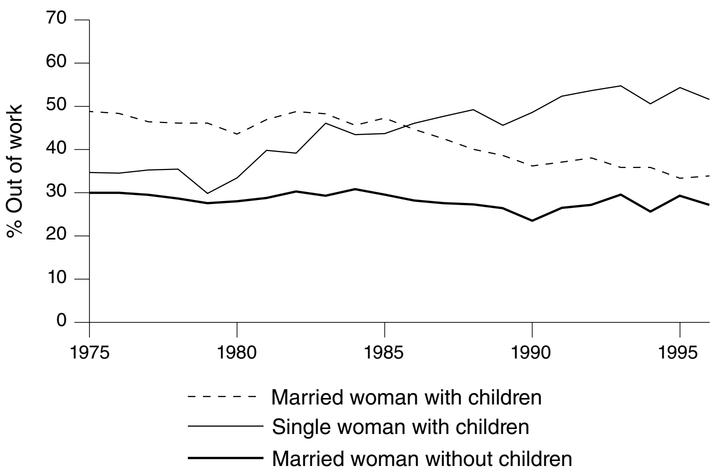


Figure 9.1. Non-Employment by Women with Children in the UK

high cost of childcare) are not sufficient to induce them to work. Indeed, WFTC in the UK is an attempt to directly address both of these concerns.

9.3 The Impact of Existing In-Work Benefit Programs

In-work benefits are designed to counter the low wages and the high implicit tax rates faced by those individuals on welfare. The idea is to modify the incentive structure so that a larger fraction of welfare recipients take jobs and leave welfare. The programs discussed here all share very similar characteristics and aims. They are designed with slightly different labor markets and slightly different target groups in mind, but nonetheless they have very strong similarities.

9.3.1 *The Canadian Self Sufficiency Project: A Randomized Experiment*

As an experimental reform the SSP in Canada is an ideal basis for evaluating the impact of targeted in-work benefits. The initial results from this program are thoroughly discussed and analyzed in Card and Robbins (1998).

Figure 9.2 shows a typical budget constraint for a Canadian welfare recipient on income assistance. It gives the budget set that an individual would face if they were earning the minimum wage in British Columbia, which was $6 an hour in 1993. Taking a job at a few hours a week attracts an earnings disregard of around $200, thereafter all earned income is effectively lost in a dollar-for-dollar transfer back to the income assistance program. So, until income assistance is exhausted – that is working nearly 50 hours a week – they would get no return, with an implicit tax rate of 100 percent on their earnings.

The SSP is available to a single parent with 12 months welfare history and who finds a job averaging 30 hours a week over a period of a month. This is calculated on a monthly rolling period. The minimum-hours criterion is interesting as the UK system also has a minimum-hours eligibility rule. It is a generous system and does not change the income assistance level; so it is not, for example, leaving those individuals who do not find employment on lower incomes. It is giving a supplement to those who move into work.

The program has been evaluated by a social experiment. This entailed following 6,000 families for 5 years starting in 1993. One-half of the group of 6,000 eligible single parents on welfare were offered the program and the others were not – they are the controls. The ones that

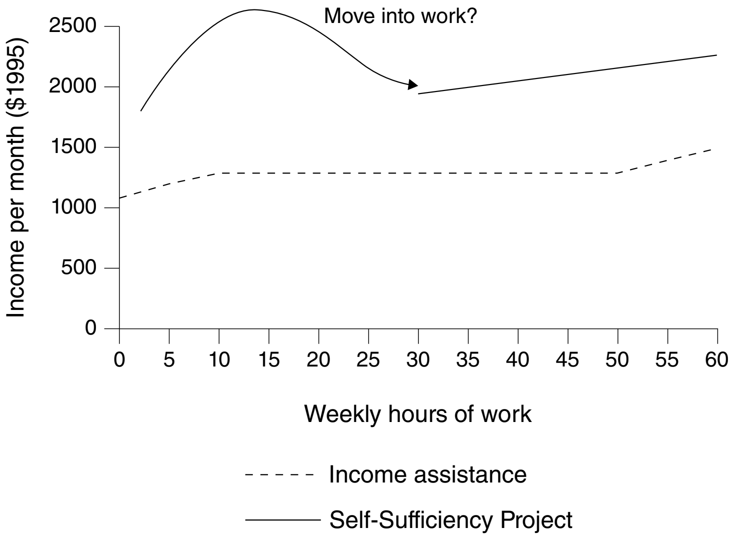


Figure 9.2. The Canadian SSP Experiment (Monthly Income for a Single Parent with One Child, British Columbia)

are on the program are the treatments – and we can compare those two groups.

This is a very well designed social experiment. The control and the treatment groups look very similar before the experiment takes place. That means that effectively the controls are really quite a good match for the treatment group. There is almost a doubling in employment for the treatment group. This is displayed in Figure 9.3, which also shows the close relationship between employment rates across the control and treatment group before the experiment began. Card and Robbins (1998) report many more results. The impact on hours and employment is very similar. The treatment group increased its hours of work, more or less, twofold over the control group. So it is having quite a big effect on the hours of work chosen by these individuals.

Although this is a very specific target group and a program with many individual idiosyncrasies, these are all single parents on welfare and it may be considered somewhat of a surprise that there is such an effect of financial incentives for those individuals. This type of quasi-experimental evidence certainly suggests that in-work benefits can have quite significant effects on labor market behavior even among lone parents on welfare – one of the central target groups for the UK and US reforms.

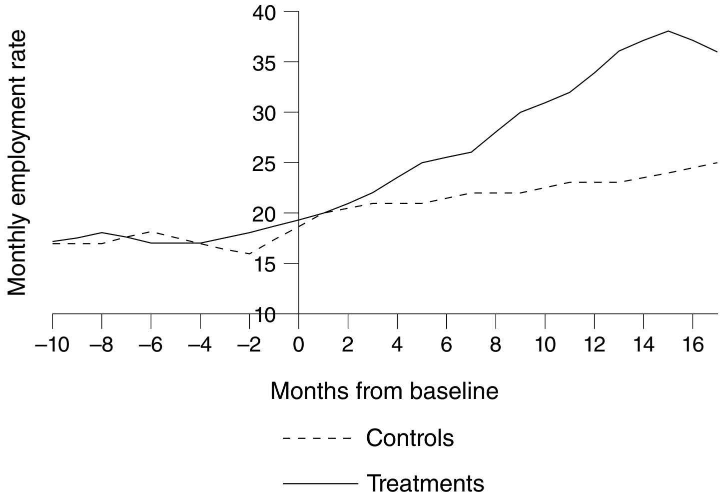


Figure 9.3. The Canadian SSP Experiment (Monthly Employment for a Single Parent with One Child BC)

9.3.2 *The EITC Reforms in the USA: A Natural Experiment?*

The Earned Income Tax Credit began in 1975 as a modest program aimed at offsetting the social security payroll tax for low-income families with children.[5] After major expansions in the tax acts of 1986, 1990, and 1993, federal spending on the EITC (including both tax expenditures and outlays) is projected to be 1.7 times as large as federal spending on Temporary Assistance for Needy Families (TANF) in 1996.

A taxpayer's eligibility for the earned income tax credit depends on the taxpayer's earned income (or in some cases adjusted gross income) and the number of qualifying children who meet certain age, relationship, and residency tests. First, the taxpayer must have positive earned income, defined as wage and salary income, business self-employment and earned income below a specified amount (in 1996, maximum allowable income for a taxpayer with two or more children was $28,495). Second, a taxpayer must have a qualifying child, who must be under age 19.[6]

The amount of the credit to which a taxpayer is entitled depends on the taxpayer's earned income, adjusted gross income, and, since 1991, the number of EITC-eligible children in the household. There are three regions in the credit schedule. These are presented in Figure 9.4, which provides a description of the EITC in 1984. The initial phase-in

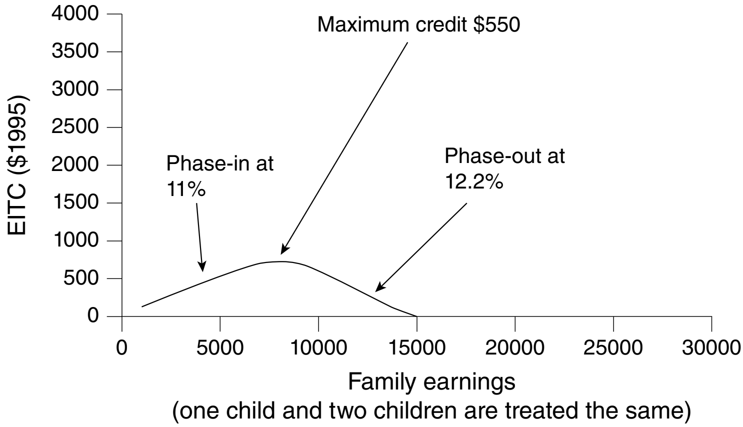


Figure 9.4. The US EITC in 1984

region transfers an amount equal to the subsidy rate times their earnings. In the flat region, the family receives the maximum credit. In the phase-out region, the credit is phased out at some given benefit reduction rate. There were expansions to the EITC in 1986 and 1991. By 1993, a family with two or more children could receive a maximum credit of $1,511, $777 more than a family with one child.

The largest single expansion over this period was contained in the 1993 legislation. The 1993 expansion of the EITC, phased in between 1994 and 1996, led to an increase in the subsidy rate from 19.5 percent to 40 percent (18.5 to 34 percent) and an increase in the maximum credit from $1,511 to $3,556 ($1,434 to $2,152) for taxpayers with two or more children (taxpayers with one child). This expansion was substantially larger for those with two or more children. A picture of these changes is presented in Figure 9.5. The phase-out rate was also raised, from 14 percent to 21 percent (13 to 16 percent) for taxpayers with two or more children (taxpayers with one child). Overall, the range of the phase-out was expanded dramatically, such that by 1996 a couple with two children would still be eligible with income levels of almost $30,000.

These EITC reforms are useful in providing a 'before and after' assessment of their effectiveness in changing labor market behavior. The idea of the 'natural experiment' or 'difference-in-differences' approach is to formalize this before-and-after contrast by finding a comparison group, not affected by the reform, which is likely to have shared a similar macro environment. This approach removes any common time

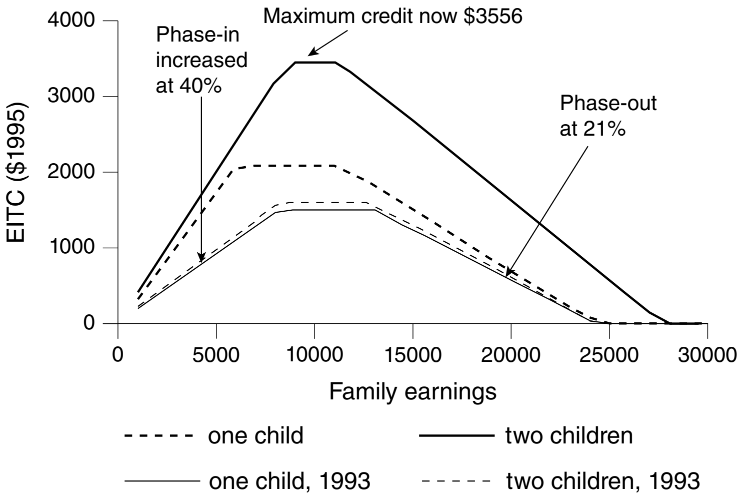


Figure 9.5. The US EITC in 1996

effects in participation across the groups. Consequently it strips out the effect of any common macro shocks that would otherwise be spuriously attributed to the reform.

The natural experiment approach relies on two important assumptions:[7]

(i) common time effects across target and comparison groups; and
(ii) no composition changes within each group over time.

Choosing a comparison group that satisfies these two assumptions is difficult. In their evaluation of the EITC policy reforms in the USA, Eissa and Liebman (1996) consider two contrasts from the repeated cross-sections of the Current Population Survey data. For the impact of the reform on single mothers, either the whole group of single women with children is used, with single women without children as controls, or the group of low-education single women with children is used, with the low-education single women without children as controls. Below we refer to results that use similar groups to evaluate Family Credit reforms in the UK.

The former control group can be criticized for not capturing the common macro effects – assumption (i). In particular, this control group is already working to a very high level of participation in the US labor market (around 95 percent) and, therefore, cannot be expected to increase its level of participation in response to the economy coming out of a recession. In this case all the expansion in labor market participation in the group of single women with children will be attributed to the reform itself. The latter group is, therefore, more appropriate as it targets better those single parents who are likely to be eligible for EITC and the control group has a participation rate of about 70 percent.

With these caveats in mind, there remain some relatively strong results on participation effects that come from the Eissa and Liebman study. For single parents there is evidence of a reasonable movement into work.[8] There is also some evidence of negative effect on hours for those in work, but this is rather small (see Scholtz 1996).

A more recent study (Eissa and Hoynes 1998) has considered the impact on married couples and finds some evidence of negative 'income' effect reducing the labor supply of married women. This is precisely the adverse effect that can be expected when a work-contingent tax credit is based on a family income and will also be found in our evaluation of the likely impact of the WFTC in the UK.

These studies of the EITC reforms in the USA, therefore, point to a reasonably strong positive effect on participation of single parents with offsetting effects on the labor supply of married women.

9.3.3 *The UK Family Credit Reforms*

9.3.3.1 FAMILY CREDIT ELIGIBILITY

Introduced in 1988, Family Credit shares many of the central features of the EITC in the USA. It was designed to provide support for low-wage families with children. An unusual feature of the Family Credit system is the minimum weekly hours eligibility criterion. A family with children needs to have one adult working 16 hours or more per week to qualify for FC. At its introduction this was set at 24 hours, but then reduced to 16 in April 1992 to encourage part-time work by lone parents with young children. To offset partially any adverse incentive effects for full-time work, a further supplementary credit at 30 hours per week was introduced in April 1995. These Family Credit reforms are interesting in their own right, but will be particularly useful as a basis for evaluating the impact of the WFTC proposals.

In the FC system each eligible family is paid a credit up to a maximum amount which depends on the number of children. There is also a smaller addition if they work 30 hours or more. Eligibility depends on family net income being lower than some threshold (£79 per week in 1998–9). As incomes rise the credit is withdrawn at a rate of 70 percent. In 1996 average payments were around £57 a week and take-up rates stand at 69 percent of eligible individuals and 82 percent of the potential expenditure.

9.3.3.2 THE 16-HOUR REFORM TO FAMILY CREDIT

The 16-hour reform took place in 1992. To examine the behavior of hours and participation before and after this reform we use the Family Expenditure Survey (FES) data source, a cross-section survey of some 7,000 British households per year. The data have been reorganized according to fiscal years to coincide with the reforms to the Family Credit system. The larger Family Resources Survey (FRS) data source, covering some 50,000 British households, which is used for the ex-ante evaluation of the WFTC reform below, is only available for the 1994–7 period and so is not useful in studying the impact of the 1992 budget reform. Both data sources collect sufficient income and earnings information to trace out accurately the budget constraint facing individual families.

The 1992 reform to Family Credit moved the hours eligibility rule from 24 hours per week to 16. A picture of the hours changes before and after the 1992 reform is presented in Figure 9.6. This figure relates to a lower-educated sample of women from the Family Expenditure Survey. The first histogram gives the distribution of hours of work for the fiscal year 1991 – before the reform. Notice that for single parents the spike at 24 hours tends to disappear and a spike at 16 hours becomes more pronounced. In Blundell et al. (2000a) a similar analysis is

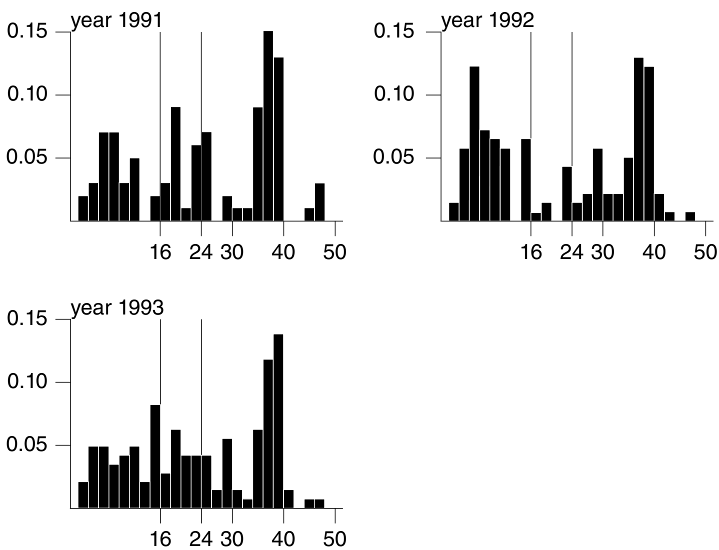


Figure 9.6. Weekly Hours of Work (single parents, lower education, UK Family Expenditure Survey)

Source: Blundell et al. (2000a).

provided for participation where there is a notable increase in the relative employment of lower-education women with children after the reform. The results suggest a 10 percent increase in employment from the 1992 reform. But they also point to a lower average hours worked among those eligible women in employment.

The analysis of the 1992 reforms, therefore, gives some support for positive employment effects and lower average hours worked resulting from such in-work benefit reforms.

9.3.3.3 THE 30-HOUR REFORM TO FAMILY CREDIT

Further support for financial incentive effects on employment and hours worked can be seen from the 1995 reform. This second reform to Family Credit added a further small credit at 30 hours. The larger FRS sample can be used to analyze hours of work before and after this reform, and in Figure 9.7 the hours of work among lower-educated, working, single parents is presented for each of the four financial years available in the FRS. Notice the pronounced spike in the hours distribution at 16 hours for single parents and the increasingly pronounced

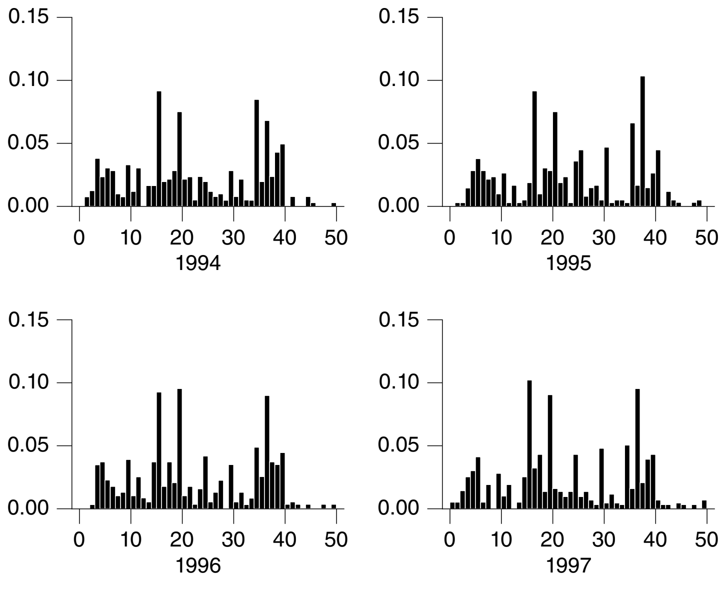


Figure 9.7. Weekly Hours Worked: Single Parents

Source: Blundell et al. (2000a).

spike at 30 hours. Again, these spikes are difficult to detect in other groups.

9.4 Evaluating The WFTC Proposals

9.4.1 *The WFTC Reform*

The replacement of FC – the WFTC – is substantially more generous and will be fully phased in by April 2000. It increases the generosity of in-work support relative to the FC system in four ways: by enhancing the credit for younger children; by increasing the threshold; by reducing the benefit reduction rate from 70 percent to 55 percent; and by incorporating a new childcare credit of 70 percent of actual childcare costs up to a quite generous limit. The effects of these changes relative to FC are shown in Figure 9.8. The largest cash gains go to those who are currently just at the end of the benefit reduction taper. The childcare credit increases the maximum amount of WFTC by 70 percent of childcare costs up to a maximum of £100 per week for those with one child or £150 per week for those with two or more children. The childcare credit component is available to lone parents and couples where both partners work more than 16 hours per week.

The childcare tax credit component could clearly have an important impact on labor supply behavior. In the simulations of the WFTC

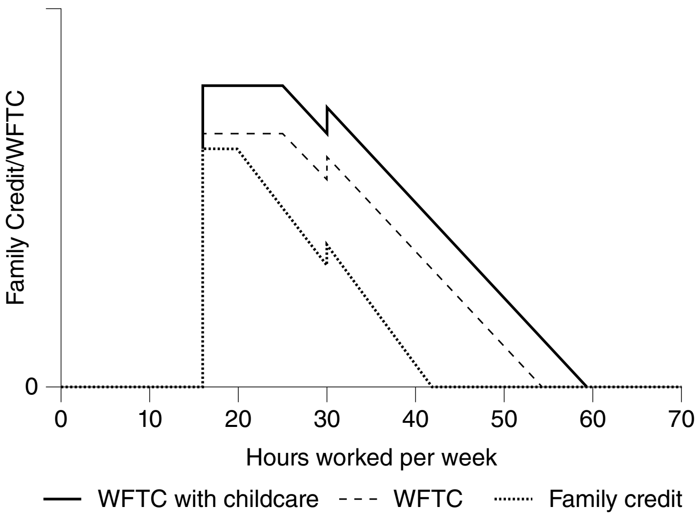


Figure 9.8. The UK WFTC Reform

reform reported here, it is assumed that the childcare market will not be affected by the reform. Indeed, the FRS data are used to compute the potential childcare costs of new entrants assuming that they will 'look like' existing childcare costs for those currently in work (see Blundell et al. 2000a). This may considerably underestimate the take-up of the credit and underestimate its impact on labor supply. Nonetheless, it seems a reasonable, if cautious, scenario to work with.

Although the before and after comparisons, discussed in Section 9.3, are indicative of what might happen, they do not provide sufficient information to simulate new reforms of this type. For that, a model that separates preferences from constraints is required. Such a model is developed in Blundell et al. (2000a). This is based on earlier work on structural simulation at the Institute for Fiscal Studies (see Blundell et al. 1988); it also develops recent work by Hoynes (1996) and provides a similar framework to Bingley and Walker (1997). In particular, it allows for childcare demands to vary with hours worked and it allows for fixed costs of work. It also accounts for take-up by incorporating welfare stigma following on from Keane and Moffitt (1998).[9]

There are two target groups for the WFTC reform: single parents and married couples with children. Two samples from the 1994–5 and 1995–6 British FRS are selected: single-parent households and married or de-facto married couples. Excluding self-employed and retired households, together with students and those in HM forces, leaves samples of 1,807 single-parents and 4,694 two-person households for use in estimation. Nearly 50 percent of currently working single parents were found to be in receipt of some Family Credit. For married couples with children this proportion is smaller, at around 16 percent. However, the latter group is more than two and a half times the size of the former. As we have seen, the WFTC reform is designed to influence the work incentives of those with low potential returns in the labor market. It does this via the increased generosity of in-work, means-tested benefits. For single parents the WFTC does unambiguously increase the incentive to work. For couples, however, the incentives created by the WFTC lead to lower participation in the labor market. Figure 9.9 shows the effect of the WFTC reform on the net income and hours schedule for a typical eligible single parent. Provided fixed costs of work are not too high, the financial incentive to move into work for a non-participant is clear. There is also an incentive to reduce hours of work among those single parents working full-time. The balance between these is purely an empirical matter, although the EITC analysis, discussed in the previous section, suggested this would not dominate the positive participation effect.[10]

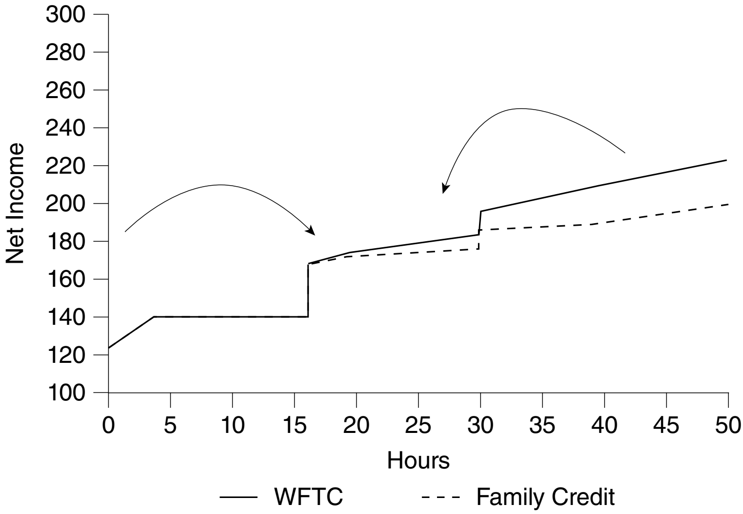


Figure 9.9. Lone Parent on Median Wage, WFTC in UK

Source: Blundell et al. (2000b).

Figure 9.10 presents a similar example of the financial incentives facing a male in a married couple where the partner does not work. For such couples where neither parent is working the incentives are unambiguously to move into work. Indeed, the gains are far larger than for our lone parent example, as the largest cash gains from the WFTC reform accrue to those at the end of the current taper. The incentives to change hours of work are ambiguous. But one interesting point is the marked increase in the effective marginal tax rate for those who become eligible for WFTC as a result of the reform. This group faces an increase in their marginal tax rates from 33 percent, produced by income tax and National Insurance, to just under 70 percent, produced by the interaction of the 55 percent WFTC taper on post-tax income. In the example, the marginal tax rate rises from 33 percent to just under 70 percent above 40 hours of work.

One final point, highlighted in the discussion of the EITC reforms in the USA, is the likely incentive for some workers in married couples to move out of work altogether. Figure 9.11 shows the budget constraint for the partner of the man in Figure 9.10. The figure is conditional on the man working 40 hours a week. Thus the family income of the woman, when she does not work, is that shown at the 40 hours point. This means that the income at zero hours has increased through the WFTC reform. In the example, anyone working more than 10 hours has an increased incentive to reduce their hours or move out of work altogether.

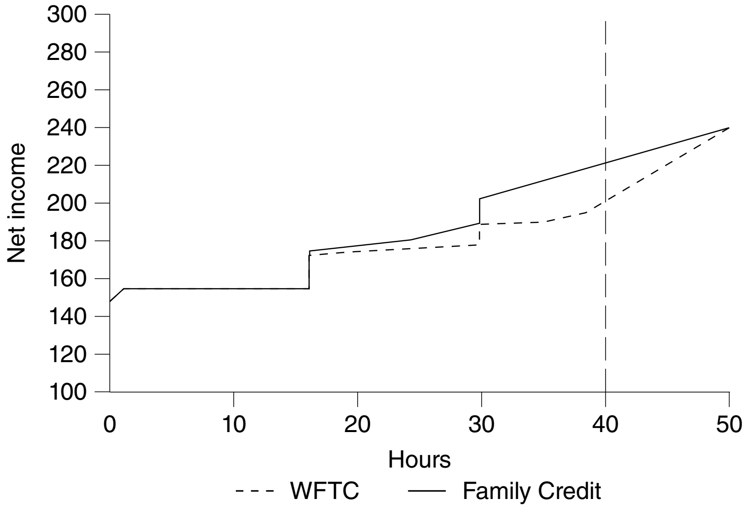


Figure 9.10. Single-Earner Male on 25th Percentile Wage, WFTC in UK
Source: Blundell et al. (2000b).

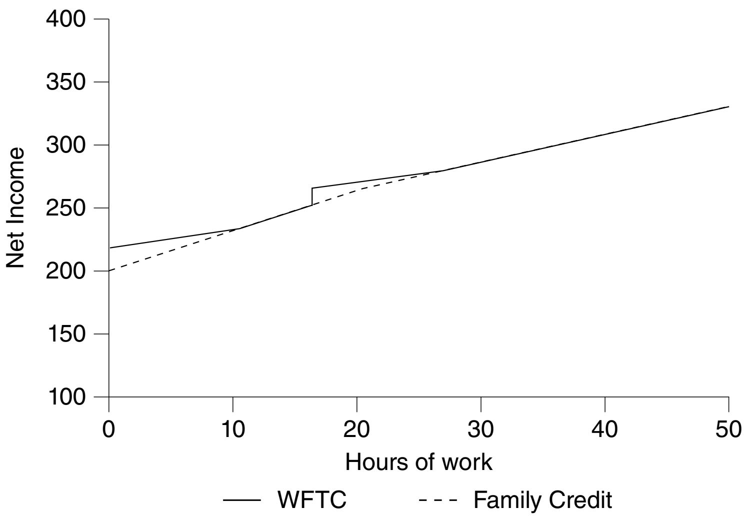


Figure 9.11. Partner in Two-Earner Couple on 25th Percentile Wage, WFTC in UK
Source: Blundell et al. (2000b).

The situation changes slightly when we allow for childcare costs at 16 hours. Here there is an additional incentive to work just over 16 hours to take advantage of the childcare credit. Thus the impact on partners in eligible families where there is already one worker is again ambiguous.

9.4.2 *Policy Simulations for the WFTC Reform*

The simulation model uses an estimated discrete choice structural labor supply model[11] to simulate the work incentive impact of the Working Families Tax Credit reform. Tables 9.1–9.3 report, respectively, the simulation results for single-parent households; women with employed partners; and women with unemployed partners. These results are presented in the form of a matrix of simulated transitions between no work, part-time work and full-time employment under the two systems.

9.4.2.1 LONE PARENTS

In Table 9.1 the simulated work responses to the WFTC among the sample of single parents is reported. The simulated transition takes around 2.2 percent of the sample from no work to either part-time or full-time work, with no offsetting movements out of the labor market. To take account of sampling variability, a standard error of 0.42 percent is placed around this figure, which would admit the possibility that the actual increase could be as much as 3 percent. One can clearly see the reason for this shift in the earlier graphs of the potential impact of the WFTC on single parents' budget constraints. At or above 16 hours per week the single parent becomes eligible for WFTC (with any childcare credit addition to which she may be entitled). For some women this extra income makes a transition to part-time employment attractive.

Table 9.1. WFTC Simulation Results: Single Parents

	Post-reform			
Pre-reform	Out of work	Part-time	Full-time	Pre-reform %
Out of work	53.0	0.7	1.5	60.2
Part-time	0.0	18.6	0.5	19.1
Full-time	0.0	0.2	20.6	20.7
Post-reform %	53.0	19.4	22.6	100
Change (%)	–2.2	0.3	1.9	

Source: Blundell et al. (2000b).

There is a minor offsetting reduction in labor supply through a simulated shift from full-time to part-time employment among 0.2 percent of the sample. This is consistent with a small (negative) income effect among some full-time single women, for whom the increase in income through the WFTC encourages a reduction in labor supply. Nevertheless, the predominant incentive effect among single parents is a positive effect on participation.

9.4.2.2 WOMEN WITH EMPLOYED PARTNERS

For married women the simulated incentive effect is quite different. In Table 9.2 estimates of the transitions following WFTC among a sub-sample of women with employed partners are reported. There is a significant overall reduction in the number of women in work of around 0.57 percent. This overall reduction comprises around 0.2 percent who move into the labor market following the reform, and 0.8 percent who move from work to non-participation. The number of hours worked by women with employed partners is predicted to fall slightly.

The predominant negative response is clearly not one that is intended, but from the earlier budget constraint analysis one can easily see why. There will be a proportion of non-working women whose low-earning partners will be eligible for the WFTC. The greater generosity of the tax credit relative to the current system of Family Credit increases household income. This increase in income would be lost if the woman in the household were to work. And for those women currently in the labor market, the WFTC increases the income available to the household if she were to stop working.

9.4.2.3 WOMEN WITH UNEMPLOYED PARTNERS

In Table 9.3 the incentives for a sub-sample of women whose partners do not work are presented. For this group there is a significant

Table 9.2. WFTC Simulation Results: Married Women with Employed Partners

	Post-reform			
Pre-reform	Out of work	Part-time	Full-time	Pre-reform %
Out of work	32.2	0.1	0.1	32.4
Part-time	0.3	31.6	0.0	32.0
Full-time	0.4	0.1	35.0	35.6
Post-reform %	33.0	31.3	35.2	100
Change (%)	0.6	–0.1	0.4	

Source: Blundell et al. (2000b).

Table 9.3. WFTC Simulation Results: Married Women with Partners Out of Work

	Post-reform			
Pre-reform	Out of work	Part-time	Full-time	Pre-reform %
Out of work	56.8	0.4	0.9	58.1
Part-time	0.0	22.2	0.4	22.6
Full-time	0.0	0.1	19.2	19.3
Post-reform %	56.8	22.8	20.5	100
Change (%)	–1.3	0.2	1.1	

Source: Blundell et al. (2000b).

Table 9.4. WFTC Simulation Results: Summary Table

Group	Number	%
Single parents	34,000	2.20
Married women, partner not working	11,000	1.32
Married women, partner working	–20,000	–0.57
Married men, partner not working	13,000	0.37
Married men, partner working	–10,500	0.30
Total effect	27,500	
Decrease in workerless families	57,000	

Source: Blundell et al. (2000b).

overall increase of 1.32 percent in the number of women who work. The reason for this shift is more straightforward, and stems from the increased generosity of the basic WFTC relative to the current Family Credit system for those women who choose to move into work. Note that for this group the generosity of the childcare credit component of the WFTC is not an issue, since households only qualify for the childcare credit if both household members work 16 hours or more. There is, of course, potential for both members of an unemployed household to move into work in order to qualify for the WFTC including the childcare credit, but a joint simulation (not reported here) shows that such an outcome is virtually nonexistent.

Table 9.4 provides an overall summary of the employment effects that could be expected from this reform. This table also provides the impact on male employment. Overall the effects on participation across these two groups of men roughly cancel each other out, leaving the major impact operating through the effects on women, mainly single parents. However, if we consider the impact on workerless households alone, then the overall figure covers up important effects and the impact on men and women in couples where neither is working is much more substantial.

9.4.3 *What Form Should the Tax Credit Take?*

There are a number of specific issues relating to the precise design of in-work benefit reforms, many of which have not been touched on in this review.[12]

There are practical issues, such as the period of assessment and the enforceability of eligibility. For example, Family Credit is only assessed every 6 months and hours of work are known to stray below the 16 hours requirement.[13] There are also within-household allocation issues which relate to who receives the credit. For example, with the switch from FC to WFTC there is a potential impact as the benefit is moved from the mother to the main earner who will receive the credit, which will more often than not be the male. There are issues concerning incidence. That is what happens to the gross wage for those in receipt of in-work benefits. It is often argued that a minimum wage can help offset any adverse incidence effects whereby the employer retains some of the credit by offering lower wages to those eligible. Then there are further issues concerning fertility incentives and marriage penalties.[14]

Probably the most important issue that relates to the WFTC reform concerns the childcare tax credit and its impact on the childcare market. This could provide the most important incentive effect if those currently not working face a constraint on low-cost childcare and can find childcare once the childcare credit is in place.[15] However, it could also be a very costly part of the reform package. There is little evidence on what might happen in this regard and the simulation results presented in this review could be an important underestimate if the childcare market for those not in work is really quite different to those currently in work.

A further important policy design issue, that has been evident from the discussion in this review, is the choice of credit (or benefit) reduction rate. As has been seen, the benefit reduction rate is 70 percent for FC, reducing to 55 percent for WFTC. It is interesting that in the EITC program it is quite low, at around 20 percent. In the Canadian system, it is about 50 percent. With the 70 percent rate currently in the UK system, it is almost inevitable that individuals working above this limit are going to find it relatively attractive to move down into more part-time work. There is a clear trade-off between individuals moving in, and others moving down into part-time work.

Exactly what benefit-reduction rate to choose is an important and difficult policy issue. If it is set too high, then there is a severe implicit tax rate. But if it is set too low, individuals on quite high incomes are able to attract some supplement from this credit system. The latter is precisely what has happened in the redesigned in-work benefit WFTC in the UK. So the exact slope of the benefit reduction rate is critical.

Given the hours-based eligibility requirement in the UK, there is a further issue as to whether a minimum hours limit should be set on eligibility. As we have seen, one interesting recent reform in the UK in 1992 was to move the minimum hours eligibility limit for in-work credit from 24 work hours down to 16.

This saw a pronounced change in the distribution of hours worked by single parents, with an increase in part-time work and a bunching at 16 hours.

A final issue that is worthy of consideration is that of time limits. Many welfare benefits in the USA are time limited. But these are not in-work benefits. The Canadian SSP, however, is time limited to 36 months. In the UK, the New Deal wage subsidies for those under 25 years of age are limited to 6 months. Limits are not only placed on wage subsidies to save money, but also to provide incentives to increase employability and training. If subsidies run out after a finite time, then the incentives for human capital are significantly greater. The same human capital (dis)incentives operate with in-work benefits and tax credits. With the recent move to unhook the link between child credits and Employment Tax Credits in the UK, there may be some pressure to make the employment tax credit component time limited.

9.5 Conclusions

This review has considered some of the issues surrounding the effectiveness of in-work benefit reforms. These programs are designed to target income to relatively poor families that suffer from low returns to work. Taking the evidence from reforms across a number of countries, the paper has argued that a careful design of these programs can significantly increase family incomes while providing reasonable incentives for parents to work. It also appears that any offsetting negative effect on hours worked by those already in employment is not strong enough to counter this overall positive increase in labor supplied. However, since these programs are generally based on family income, there is evidence of a negative offsetting effect on the labor supply of married women in households with young children. For the

WFTC reform in the UK, the overall impact on employment is expected to be modest, although, focusing on workerless households alone, the impact is likely to be more substantial.

The paper has considered evidence from quasi-experiments in Canada, from the before-and-after analysis of past reforms to the EITC in the USA and to the FC in the UK. All these different pieces of evidence tend to support this general conclusion.

One area in which we should have more to say, but for which there is little rigorous study, concerns the longer-run pay-off to labor market attachment for individuals of the type eligible for the WFTC. Eventually, one might hope that through the progression of wages and general increases in employability, workers move themselves out of the low-income group and on to a reasonable earnings level, then less likely to be drawing income assistance or any forms of tax credits. However, the incentives for training and human capital investment for low-skilled workers are typically reduced by in-work benefits. If the dynamic pay-off is relatively low, then individuals are likely to be stuck on tax credits and may remain continuously in this part of the welfare system.

10

Evaluating the Employment Impact of a Mandatory Job Search Program

Richard Blundell, Monica Costa Dias, Costas Meghir and John Van Reenen

10.1 Introduction

The literature on the evaluation of labor market programs is voluminous, growing and somewhat sobering. The sobering aspect is that we are learning that most programs appear to have very limited effects, especially those that focus on young low-skilled adults. This paper concerns the evaluation of a targeted active labor market program, "New Deal for the Young Unemployed", designed to move young unemployed individuals in the UK into work and away from welfare. This is a major program that has affected several million young people. It brings together many of the best features of other such initiatives, combining job search assistance in the first instance with subsidized job placement for those for who the initial treatment was not successful. As such, a rigorous evaluation of the program may lead to insights regarding the implementation of programs in other countries. In fact, we do find evidence that the program has successfully raised employment, although it is still an open question how long-lived these benefits will be.

The original version of this chapter was published as: Blundell, R./Costa Dias, M./Meghir, C./Van Reenen, J. (2004). Evaluating the Employment Impact of a Mandatory Job Search Program, in: Journal of the European Economic Association, 2(4): 569–606. *We thank David Card, Richard Freeman, James Heckman, Hide Ichimura, Richard Layard, Rebecca Riley, Garry Young and participants in seminars at Institute for Fiscal Studies, CEMFI, LSE, DfEE and St. Andrews. We are grateful to the Leverhulme foundation for funding this project. This research is also part of the program of research at the ESRC Centre for the Microeconomic Analysis of Public Policy at Institute for Fiscal Studies. The DfEE kindly provided access to NDED and the ESRC Data Archive to the JUVOS data. The second author acknowledges the financial support from Sub-Programa Ciência e Tecnologia do Segundo Quadro Comunitário de Apoio, grant number PRAXIS XXI/BD/11413/97. The usual disclaimer applies.*

The program we evaluate was piloted in certain areas before it was rolled out nationwide. Moreover, the program has age-specific eligibility rules. We use these area and age-based eligibility criteria that vary across individuals of identical unemployment durations to identify the program effects. We show how they allow us to examine the extent of substitution between eligible and non-eligible groups and also to assess whether there are significant general equilibrium (or "community-wide") effects of the program. We also exploit a number of additional features of our evaluation data to address fundamental problems that have affected non-experimental program evaluations. First, we use the long history of pre-program data at our disposal to evaluate the plausibility of the assumptions underlying our approach. Having longitudinal data on individuals for up to fifteen years prior to program introduction enables us to place bounds around the maximum and minimum effects of the program based on historical experience. Second, it has been suggested that results from non-experimental evaluations can be fragile – highly dependent on functional form assumptions and on the availability of suitable conditioning variables.[1] We use a variety of methods to provide sensitivity analysis on this issue and we find remarkably robust results. Moreover, as suggested by Heckman, Ichimura and Todd (1997) we are able to control for recent labor market history, which can be of central importance for the success of a non-experimental evaluation.

The program we study is directed toward individuals aged between eighteen and twenty-four and who have been claiming unemployment insurance (called "Jobseeker's Allowance"[2] in the UK) for six months. The whole program combines initial job search assistance followed by various subsidized options including wage subsidies to employers, temporary government jobs and full-time education and training. Prior to this program, young people in the UK could, in principle, claim unemployment benefits indefinitely. Now, after six months of unemployment, young people enter the "Gateway", which is the first period of intensive job search assistance. The program is mandatory, including the subsidized options part. In this paper we focus only on the job assistance and wage subsidy element of the New Deal as our data does not cover a sufficient period to analyze the other parts of the program (e.g. education and training).[3]

Our approach to evaluation consists of exploring sources of differential eligibility and different assumptions about the relationship between the outcome and the participation decision to identify the effects of the New Deal. On the "differential eligibility" side, we use two potential sources of identification: age and area. The fact that the program is age-specific implies that using slightly older people of similar unemployment duration is a natural comparison group. This is similar to the identification strategy in Katz (1998) who analyzed the withdrawal of a wage subsidy (the Targeted Job Tax Credit) from economically disadvantaged 23 and 24 year olds in 1989–90. He used a combination of age, economic disadvantage and time in order to construct different comparison groups, and identified a small but significant effect of the program on employment. Our study uses geographical areas as an additional source of identification to Katz (1998) by exploiting the fact that the program was first piloted in selected areas before being implemented nationwide.

Under a simple difference-in-differences approach, we show that the choice of the comparison group determines the parameter being estimated as various potential sources of biases are dealt with in different ways. We are especially concerned about substitution and equilibrium wage effects. Substitution occurs if participants take (some of) the jobs that non-participants would have got in the absence of treatment. Equilibrium wage effects may occur when the program is wide enough to affect wages through changes in the supply of labor. While studying the pilot period, we use a diversity of comparison groups who will be affected differentially by these types of indirect effects to obtain some indication on the importance of such biases.

We apply a number of different econometric techniques, all exploiting the longitudinal nature of the data set being used but making different assumptions about the structure of the disturbances. A general set up is developed, where all estimators can be interpreted in the light of combined difference-in-differences and matching methodologies. The conditions under which each estimator identifies and estimates the impact of treatment on the treated are derived.

The estimators being used in the present paper, as in many other evaluations, rely on the critical assumption that the evolution of employment in the two groups being compared would have been the same in the absence of the program.[4] One reason for this to be violated is the fact that individuals eligible for the New Deal program could react to it in anticipation of the program, i.e. before eligibility. We can test for this since we observe the complete inflow into unemployment and hence can assess whether the program induces differential behavior in the six months preceding eligibility. Other factors that could induce differential time trends relate to the slight differences either in location or age of the groups to be compared. We use past history to infer the extent to which this may affect our results.

We focus on the change in transitions from the unemployed claimant count to jobs during the first four months of treatment (the "Gateway" period), although we compare this with a slightly longer perspective. We find that the outflow rate to jobs has risen by about 20 percent for young men as a result of the New Deal. That is five percentage points more men finding jobs in the first four months of the New Deal above a pre-program level of twenty six percentage points. Similar results show up from the use of different adopted estimators, independently of the amount or type of structure imposed, and they appear to be robust to pre-program selectivity, changes in job quality and different cyclical effects. We obtain similar estimates from using across regional comparison groups (the pilot areas) as we do when using eligible vs. non-eligible age groups. Such an outcome suggests that either equilibrium wage and substitution effects are not very strong or they broadly cancel each other out.

The robustness of our results is reassuring, but we take care to judge how permanent the effects are likely to be. We do find evidence of an important "program introduction effect" in the sense that the impact of the program is much larger in the first quarter it is introduced than in subsequent quarters. However, there are reasons to suspect that a program such as the New Deal will have more sustainable effects than other labor market programs.[5] First, the program is mandatory: refusal to participate results in sanctions. Compulsory, sanction-enforced schemes have often been found to be more effective than voluntary schemes.[6] Secondly, the "disadvantaged youth" we consider are less disadvantaged than those typically treated in typical US programs often found to be ineffective (e.g. ex-offenders). The only entry requirement is six months unemployment benefit claim, which is not so uncommon for those under twenty-five years of age in Britain. Finally, we are evaluating the effects of job search assistance and wage subsidies where the US evidence has been more optimistic than in the case of training programs (see Sub-section 10.5.4 for a more detailed comparison of our results with other studies).[7]

The structure of the paper is as follows. We start in Section 10.2 with a more detailed description of the New Deal. Section 10.3 presents the methodology we apply. We discuss how the choice of the comparison group determines the parameter being identified along with the potential sources of bias in each case, and develop a combined difference-in-differences and matching set up where all the estimators being used can be interpreted. Section 10.4 describes the data and Section 10.5 details the empirical results. We separate the analysis of the pilot period of the program, where more detail is possible given the additional instruments we are able to explore to construct the counterfactual. Males and females are also discussed separately and we compare our UK results with experimental evaluations of similar US programs. Finally, Section 10.6 offers some concluding comments.

10.2 The Program

10.2.1 *Description of the New Deal for the Young Unemployed*

The New Deal for Young People is a recent initiative of the UK government to help the young unemployed gain work. The program is targeted at the eighteen to twenty-four years old longer-term unemployed. Participation is compulsory; every eligible individual who refuses to co-operate faces a loss of entitlement to benefits. The criteria for eligibility are simple: every individual aged between eighteen and twenty-four by the time of completion of the sixth month on Jobseeker's Allowance (JSA), equivalent to UI, is assigned to the program and starts receiving treatment.[8] Given the stated rules, the program can be classified as one of "global implementation", being administered to everyone in the UK meeting the eligibility criteria.[9] Indirect effects that spill over to other groups than the treatment group may occur and the nature of these effects will be discussed below.

The path of a participant through the New Deal is composed of three main steps (see Figure 10.1). On assignment to the program, the individual starts the first stage of the treatment called the "Gateway".

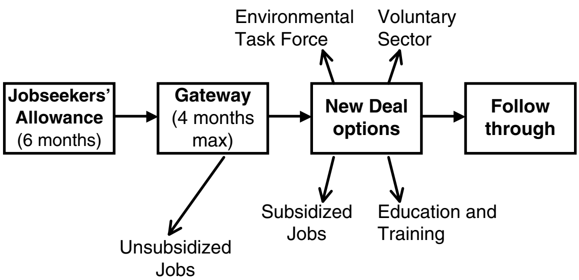


Figure 10.1. A Simplified Flow Diagram of the New Deal Program

Notes: The New Deal for the Young Unemployed is a mandatory welfare-to-work program. All young people (between the ages of 18 and 24) who had claimed unemployment insurance (JSA) for sixth months enter the program. During a "Gateway" period of at most four months participants are given extensive job search assistance. Those failing to find an unsubsidized job have four different options: entering employment with a six month wage subsidy to the employer, twelve months full-time education or training, working in the environmental taskforce (a public-sector job) or working in the voluntary sector. The individual faces the withdrawal of unemployment benefits if they do not co-operate. After the end of the options participants who rejoin the unemployed enter the "Follow Through" which is essentially the same as the Gateway. For more details see DfEE (1997).

This is the part of the program being evaluated in the present study. It lasts for up to four months and is composed of intensive job search assistance and small basic skills' courses. Each individual is assigned a "Personal advisor", a mentor who they meet at least once every two weeks to encourage/enforce job search.

The second stage is composed of four possible options. First, there is the "employer option" – a six-month spell on subsidized employment. For the subsidized employment option, the employer receives a £60 (about $90) a week wage subsidy during the first six months of employment plus an additional £750 (about $1125) contribution to finance a required minimum amount of job training equivalent to one day a week.[10] Under a second possible option, individuals can enroll in a stipulated full-time education or training course and receive an equivalent amount to the JSA payment for up to twelve months. Third, individuals can work in the voluntary sector for up to six months and are paid a wage or allowance of at least the JSA plus £400 ($600) spread over the six months. Finally, they may take a job on the "Environmental Task Force" – essentially a government job.[11]

Once the option period is over, if the individual has not managed to keep/find a job or leave the claimant count for any other reason, the third stage of the program is initiated, the "Follow Through". This is a process similar to the Gateway, taking up to thirteen weeks, where job search assistance is the main treatment being provided.

The program was launched in the whole UK in April 1998 (the "National Roll Out"). There was, however, a pilot from January to March 1998, when the program was implemented in twelve areas, called the Pathfinder pilots.[12] Clearly, identification of the treatment effect under these conditions requires stronger assumptions than when an experiment is run within regions using random assignment over a large number of areas. As will be discussed, the problem relates to the fact that the counterfactual must either be drawn from a different labor market or from a group with different characteristics operating in the same labor market. Below we explore what we can identify under different assumptions.

Given that the program has not been running for a long period, we focus in this paper on an evaluation of the Gateway. In particular, we are concerned with the degree to which enhanced job search assistance has lead to an increase in outflows to jobs. The evaluation is based on data provided by the Pathfinder areas before the National Roll Out of the program, as well as on data available following the National Roll Out.

In evaluating the impact of the program we need to consider the precise nature of the comparison group, and hence the definition of what is being measured, and the set of assumptions that underlie the interpretation of the parameter we estimate in each case. There are some important aspects covered within this discussion. One of them concerns the extent to which we can estimate the overall impact of the program on employment as opposed to the impact on the eligible individuals. Potential differences between the overall impact and the effect of treatment on the eligible group could arise because: (1) the impact of the program on eligible individuals may be at the expense of worsened labor market opportunities for similar but ineligible individuals or (2) the wider implementation of the program and the opportunities it offers to participants may affect the equilibrium level of wages and employment, affecting all workers.

10.2.2 *Choice of the Outcome Variables*

We focus on the impact of the program on the proportion leaving unemployment within four months of entering the "Gateway". The choice is mainly dictated by the desire to focus on the stated government targets and the paucity of data on individuals after they have finished the options.[13]

However an alternative outcome variable would have been the proportion leaving unemployment within, say, eight or ten months of entering the unemployment pool. This outcome variable would avoid the potential composition effects that may be induced by the anticipation of the program among eligible individuals. In particular, if the program is perceived as being able to improve placements, then individuals close to the Gateway and eligible for the program may reduce their search effort and wait for the program. In this case, the average individual among eligibles would be more prone to leave unemployment than its counterpart in the comparison group, leading to increased exit rates for this group. However, we can test this hypothesis by estimating the proportion of those who left unemployment by the end of the sixth month in the eligible and ineligible group. Such a comparison will provide an idea of how important such compositional effects are likely to be.

We will pay special attention to the outflows into employment, but we also examine total outflows from unemployment to all destinations. To assess the importance of the estimated effects, we interpret them in an historical perspective. We provide some lower and upper bounds for the treatment effect by using our methodology during other pre-program time periods. This can be done for total outflow for all years since 1982.

To summarize: treatment is understood as the job search assistance initiative of the New Deal and the treated are those who enroll in the program after completing a six month unemployment spell. We aim at measuring the impact of improved job search assistance on the probability of finding a job among the treated. To assess the robustness of our results, we also present estimates of other parameters that are informative about the adequacy of the underlying assumptions. Different definitions of treatment and the treated often characterize such parameters, and this is made clear on the following discussion.

10.3 Identification and Estimation Methods

Our approach to estimate the impact of the New Deal program relies on using information from the pilot period as well as information from the National Roll Out. The New Deal can affect employment of both eligible and ineligible individuals in a number of ways. First, the eligible individuals receive job search assistance that may enhance their ability to find a job. Second, some of the individuals who pass through the Gateway will receive the wage subsidy option, reducing the cost of employing them for an initial period of six months. This wage subsidy will expand the employment of such workers but may also lead to a substitution of other workers for these cheaper ones. The extent to which this may happen will depend on a number of factors. If the subsidy just covers the deficit in productivity as well as the costs of training, we would not expect any substitution; these workers are no cheaper than anyone else. Second, it will depend on the extent that these workers are substitutable in production for existing workers and on the extent that it is easy to "churn" workers, that is to replace a worker finishing a six-month subsidy with a new subsidized worker. The latter is an important point, since the subsidy only lasts six months. Moreover, the agencies implementing the New Deal are supposed to be monitoring the behavior of firms using wage subsidies and employing individuals on the New Deal. Of course, if job durations are generally short, firms will be able to use subsidized workers instead of the non-subsidized ones, without any extra effort.

The New Deal may also change the prices in a region or country as a whole as it affects a substantial number of people. For example, the increased search activities of the unemployed could lower the equilibrium wage for less skilled individuals and therefore increase aggregate employment through a higher job offer arrival rate. This will tend to increase employment for eligible and ineligible individuals and will counteract the effects of substitution on the non-treatment group. Randomized trials cannot account for these general equilibrium or "community wide" effects which have become an important issue in the program evaluation literature.[14]

Assessing the importance of substitution and of general equilibrium effects through wages or other channels is of central importance. Using the comparison between the pilot and control areas as described below, and assuming these areas are sufficiently separate labor markets from each other, we will be able to assess the extent to which substitution and other general equilibrium effects combined are likely to be important "side-effects" of the program, at least in the short run. Below we discuss the evaluation methodology, a central part of which is the choice of comparison group. This choice is to a large extent governed by the issues discussed above.

10.3.1 *The Choice of Comparison Group*

Define by Y^1_{it} the outcome for individual i in period t that they are exposed to the policy (treatment). The outcome for the same individual if not exposed to the policy is Y^0_{it}. Consequently, the impact for the i-th individual of the policy is $Y^1_{it} - Y^0_{it}$. The average policy impact for those going through the New Deal is $E(Y^1_{it} - Y^0_{it} \mid ND = 1)$.

This parameter will be the focus of our attention. $ND = 1$ denotes the areas assigned to the New Deal, $t = 0$ represents the period before implementation and $t = 1$ the period after. Quite clearly, the evaluation problem relates to the missing data that would allow us to estimate $E(Y^0_{it} \mid ND = 1)$ directly. In this section we define a number of alternative comparison groups that will allow us to estimate this counterfactual mean. As we will point out, the definition and interpretation of the estimated parameter will change in certain cases with the comparison group.

Consider first a contrast obtained by comparing employment growth in pilot areas to employment growth in control areas. Assume that

$$E(Y^0_{it} \mid ND = 1, t = 1) - E(Y^0_{it} \mid ND = 1, t = 0) = \\ E(Y^0_{it} \mid ND = 0, t = 1) - E(Y^0_{it} \mid ND = 0, t = 0). \quad (1)$$

This assumption means that the growth in employment in the New Deal areas would have been the same as in the non New Deal areas in the absence of the policy. In this case the missing counterfactual value can be replaced by

$$E(Y^0_{it} \mid ND = 1, t = 1) = E(Y^0_{it} \mid ND = 1, t = 0) + m_t.$$

This expression is simply the employment level in the New Deal areas before the policy was implemented, adjusted for aggregate employment growth, given by

$$m_t = [E(Y_{it}^0 \mid ND = 0, t = 1) - E(Y_{it}^0 \mid ND = 0, t = 0)].$$

This gives rise to a straightforward difference-in-differences estimator. Under the assumption in (1), such a comparison of growth rates estimates the impact of the New Deal on individuals residing in a pilot area, irrespective of whether they are eligible or not; hence, this comparison estimates the net effect of the program including any impact of general equilibrium effects and substitution.

We can, however, obtain a measure of the importance of substitution effects by comparing the growth of employment in pilot and control areas separately for eligible and ineligible individuals. Consider applying assumption (1) applied separately to eligible and ineligible individuals and computing the growth in the employment for the eligible individuals in the pilot and control areas separately. Substitution effects should increase the employment of eligible individuals at the expense of ineligible ones in the pilot areas. Area-specific general equilibrium effects due to the change in wage pressure from the increased supply of workers should tend to increase the employment of both eligible and ineligible individuals. The general equilibrium effects can be thought of as part of the program effect. The employment growth of eligible individuals will include the "pure" program effect, the general equilibrium effect and the presumably positive substitution effect. The employment growth of ineligible individuals will include a general equilibrium effect and a substitution effect of equal and opposite sign to that of the treatment group (assuming that the comparison group is the only group of workers displaced due to the wage subsidy). Thus a sum of the estimated "treatment" effects on eligibles and ineligibles in the pilot areas compared to the control areas (weighted by the size of each group) should provide us with an estimate of the program effect and the general equilibrium effect combined, but not of the substitution. If this is similar to an appropriately scaled version of the effect on eligibles alone we can infer that substitution effects are not an important issue.

The definition of the comparison group is of course central to the evaluation. The approach discussed above used as comparison group individuals in non-exposed areas during the pilot period. However, the pilot stage lasted three months only and it is possible that the impacts of the policy in this short first period are not generalizable to later periods (for example, because the administration of the program would have been in its infancy). So, we next consider using data from the National Roll Out, the term referring to the national implementation. Suppose we start by assuming that assumption (1) is valid when $ND = 1$ refers to eligible individuals following the national implementation and $ND = 0$ refers to "similar" but ineligible individuals, i.e. those unemployed for over 6 months whose age is just above 24. The choice of this group makes it most likely that their overall characteristics and behavior match that of the treatment group; i.e. that the growth rate of employment for the two groups would be similar in the absence of the program. Such an approach is similar to a regression discontinuity design.[15] By making assumption (1) with respect to these two groups, we are ruling out any substitution effects or equilibrium wage effects that have an impact on the groups in a differential way. In this case a comparison of the growth rates between eligible and ineligible individuals will provide an estimate of the impact of the program on the eligible ones.

The virtue of the comparison group – that it is very similar to the treatment group in terms of its characteristics and will therefore be expected to respond to shocks in similar ways – may be, in fact, its greatest disadvantage. The substitution effects are likely to be much more severe the closer the productivity characteristics of the two groups are. In the event of substitution, the impact of the program for the eligible group is biased upwards by the fact that the employment of the comparison group is decreasing. If such a decrease is, say, β, the "true" net increase in employment is 2β lower than the estimated increase in employment. However, the benefit in terms of employment for the target group would be β lower than our estimate. Within this framework of analysis, the only way we have of gauging the size of β is through the pilots, as discussed above. Alternatively a general equilibrium model would allow us to estimate β, at least in the long run, based on the substitutability of the two groups in production.

There are a number of additional issues that we need to address. First, there is the basic issue of whether we can assume that the two groups are subject to the same aggregate labor market trends. To the extent that the human capital of the two groups is similar and to the extent that preferences for work are the same, this assumption will be satisfied. Preferences for work between the eligible group in their early twenties and the ineligible group in their middle/late twenties may, however, not be the same as this is the age that many people get married and start to have children. This may generate differential aggregate trends across groups. We can address this issue by examining the trends in the exit rate from employment of the two groups for a number of years prior to the implementation of the New Deal. Over the preceding years there has been no major policy that explicitly discriminates between the two groups. This approach also suggests a method for bounding the impact of the policy using the historical trends in the two groups. In particular, we can identify the pre-program period within our data set that would maximize the estimated impact of the policy and the period that would minimize it. In the empirical section we show the historical trends for the two groups and we provide bounds for our estimates based on these fluctuations between the two groups.

The next important issue is whether the impact of the policy is heterogeneous with respect to observable characteristics. If this is the case, we should interpret the estimate we obtain as an average impact across different effects but must make sure that a suitable comparison group exists. One way to address this problem is to use propensity score matching adapted for the case of difference-in-differences. In this case, there are two assignments that are non-random. One assignment is to the eligible population and the other assignment is to the relevant time period (before or after the reform). For the evaluation to make sense with heterogeneous treatment effects, we must guarantee that the distribution of the relevant observable characteristics is the same in the four cells defined by eligibility and time. One way of achieving this is to extend propensity score matching by defining two propensity scores – one for eligibility and one for time periods. We then create a matched sample based on the two propensity scores. This approach ensures that the distribution of observed characteristics is balanced across all cells. In general, the assumption required to justify this approach is that where $ND = 1$ denotes eligibility and t the time period.

$$E(Y_{it}^0 \mid X, ND = 1, t = 0) - E(Y_{it}^0 \mid X, ND = 1, t = 1) =$$
$$E(Y_{it}^0 \mid X, ND = 0, t = 0) - E(Y_{it}^0 \mid X, ND = 0, t = 1)$$

This allows the time effects to differ by X. Following Ashworth et al. (2001), under this assumption it is possible to construct matched samples by conditioning on the propensity scores for eligibility, $P_{EX} = Pr(ND = 1 \mid X)$, and for being observed in time period $t = 1$, $P_{tX} = Pr(t = 1 \mid X)$

$$\begin{aligned} &E(Y_{it}^0 \mid P_{EX}, P_{tX}, ND = 1, t = 0) \\ &-E(Y_{it}^0 \mid P_{EX}, P_{tX}, ND = 1, t = 1) = \\ &E(Y_{it}^0 \mid P_{EX}, P_{tX}, ND = 0, t = 0) \\ &-E(Y_{it}^0 \mid P_{EX}, P_{tX}, ND = 0, t = 1) \end{aligned} \tag{2}$$

The observables we use include, among other things, labor market history. This approach, which can be implemented non-parametrically, ensures that the composition of the samples in the four cells being compared is kept constant with respect to these key determinants of employment outcomes. In addition we discuss simpler parametric methods that condition linearly on a number of observable characteristics. We further discuss these issues in the estimation section below. Finally, the discrete nature of our outcome variable may imply that the assumptions we make do not hold for the expectations (which are employment probabilities) but for some transformation thereof; in particular for the inverse of the probability function, which must be assumed known. In this case we assume that

$$\begin{aligned} f^{-1}[E(Y_{it}^{0} \mid X, ND=1, t=1)] - f^{-1}[E(Y_{it}^{0} \mid X, ND=1, t=0)] = \\ f^{-1}[E(Y_{it}^{0} \mid X, ND=0, t=1)] - f^{-1}[E(Y_{it}^{0} \mid X, ND=0, t=0)] \end{aligned}$$

where f^{-1} is the inverse of the probability function (e.g. the inverse logistic). This just says that the assumption we make is valid for the index rather than the probability itself. Define by Y_{it} the employment indicator for individual i in period t. In the New Deal areas in period t = 1, this will represent the outcome under treatment. In all other cases it will represent an outcome under non-treatment. The impact of the policy can then be evaluated as

$$\begin{aligned} I(X) = E(Y_{it} \mid X, ND=1, t=1) - \\ f[f^{-1}(E(Y_{it} \mid X, ND=1, t=1) - \alpha(X))] \end{aligned} \tag{3}$$

where

$$\begin{aligned} \alpha(X) = \left\{f^{-1}[E(Y_{it} \mid X, ND=1, t=1)] - f^{-1}[E(Y_{it} \mid X, ND=1, t=0)]\right\} - \\ \left\{f^{-1}[E(Y_{it} \mid X, ND=0, t=1)] - f^{-1}[E(Y_{it} \mid X, ND=0, t=0)]\right\} \end{aligned} \tag{4}$$

I(X) is then averaged using as weights the distribution of *X* among actually treated individuals. Despite the similarity to the linear case, the non-linear assumption stated above entails two additional restrictions on the nature of the error terms: only group-effects are allowed for and between groups homoscedasticity is required.

10.3.2 *Implementation*

Given a particular choice of comparison group, all methods we apply have the same structure as implied by (3) and (4). They differ only in the way that the expectations in these expressions are computed.

In the linear matching difference-in-differences estimator we run the following simple regression on the sample of comparison and treatment observations

$$Y_{it} = \theta_{ND} + d_t + \gamma' X_{it} + \alpha ND_{it} + \varepsilon_{it}$$

where Y_{it} is a discrete variable indicating whether the person is in employment or not, θ_{ND} is an eligibility specific intercept (may it be area or age defined or both, depending on the comparison group used), d_t reflects common/aggregate effects and where X is included to correct for differences in observable characteristics between individuals and areas registered at the eligibility point (completion of the 6th month in unemployment).

These procedures can be quite restrictive in a number of ways. First, they do not allow for α to depend on X. And second, they do not impose common support on the distribution of the Xs across all four cells. The first assumption can be relaxed under the parametric setting, and this is what we do within the non-linear logit specification. The effect of treatment is allowed to depend on the observable characteristics of the agents by applying the following estimation technique. A different relationship between the outcome and the observables is estimated by group of agents – eligibility status (area or age) interacted with time. Such relationships entail the particular behavior pattern of each group and the impact of treatment when it existed. By predicting the outcome of treated under the non-treated behavioral equation one obtains an estimate of how the treated would have been without the treatment would they belong to each of the other groups conditional on their observable characteristics. Applying difference-in-differences to such predictions using equation (3) produces an estimate of the expected impact of treatment on the treated.

To relax both assumptions simultaneously, we supplement the above results by propensity score matching. As mentioned above, this involves matching on two propensity scores, which balances the distribution of the X characteristics in the treatment and comparison samples, before and after the reform. The matching method we use smoothes the counterfactual outcomes either with a Kernel based method or with splines (see, Heckman, Ichimura and Todd 1997; Meghir and Palme 2001). We also present results based on the nearest neighbor weighting scheme. These, however, turn out to be much less precise. We provide details on the estimation method in Appendix C.[16]

10.3.3 *Other Estimation Issues*

10.3.3.1 THE CHOICE OF COMPARISON AREAS

As discussed above, the available options for the choice of the comparison group depend on the type of evaluation being performed.

When assessing the program from data on its National Roll Out, we are constrained to use ineligible individuals within the same area, for which we have chosen the age rule to define (in)eligibility. For the pilot study, however, the regional rule provides an additional instrument in the definition of the comparison group. We have used it in two ways, constructing two possible comparison groups: The first takes all eligible individuals living in all non-Pathfinder areas; the second selects all eligible individuals in the set of non-Pathfinder areas that most closely resemble the Pathfinder areas in a way detailed below.

The goal of a careful choice of the comparison area is to satisfy assumption (2), which requires that the time trend evolves in the same way for treatment and comparison groups. To assess how similar any two groups are, we compare men aged 19–24 years old that live in Pathfinder areas to those that live in all non-Pathfinder areas with respect to their recent history of conditional outflows from unemployment. It is clear that the Pathfinder areas have, on average, worse labor market conditions. However, for the purposes of evaluating eligibility rules, Figure 10.2 illustrates the evolution of the outcomes for the impact of the program based on these two groups. What is important is

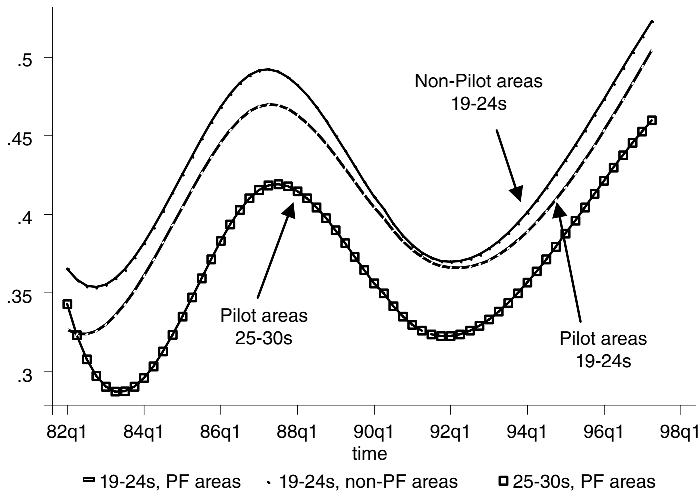


Figure 10.2. Outflows from JSA Conditional on Completing 6 Months: Effect by the End of the 10th Month on JSA

Notes: This graph illustrates the proportion of men leaving unemployment between the sixth and tenth months of unemployment 1982–1998. "PF" indicates that the men were living in a Pathfinder Pilot area (prior to New Deal introduction in 1998). The data have been smoothed by a cubic spline in time. Breakpoints were included at the first quarter of 1987 and the first quarter of 1990. No other covariates were considered.

that the difference between the two curves is kept nearly constant over time in order to guarantee that macro trends affect the two groups in similar ways. The older group aged 25–30 is also presented as a potential comparison.[17] This group tends to have lower outflows than their younger counterparts.

Nevertheless, this data shows that the size of the estimated impact can be sensitive to the choice of period for comparison and in the results section we are careful to test the sensitivity of the results to alternative timing assumptions.

When using all eligible individuals in non-Pathfinder areas as a comparison group (or a matched sub-sample of them), it is being assumed that the two curves represented in Figure 10.2 are indeed parallel so that similar individuals are similarly affected by macro trends, independently of where they live. One can, however, choose the areas that more closely follow the cycle pattern identified for the Pathfinder areas. This can be done either within each of the matching procedures described above, or prior to them, selecting the areas where the comparisons are to be drawn from. We have chosen to adopt this latter option, matching the areas in a first step and applying all types of estimators comparing eligibles in different areas to the sub-samples obtained. In this procedure, we have used a completely non-parametric technique, as described below.

The aim of matching the areas is to achieve a match as close as possible with respect to labor market characteristics. The procedure followed to match on labor market characteristics makes use of a quarterly time-series of the outcome variable from 1982 to just before the introduction of the New Deal, in January 1998. A measure of distance was then computed for each possible pair of Pathfinder and non-Pathfinder areas and the two nearest neighbors were chosen. Once the two nearest neighboring areas have been chosen based on similarity of the labor market trends, we carry out the estimation procedure as described earlier.

10.3.3.2 SENSITIVITY OF THE RESULTS

The relative size of the estimated impact of the program, when viewed in an historical perspective, can inform on how significant the result is. In order to do so, the series of year-by-year estimates of the impact of a fictitious program has been computed.[18] Given the lack of data on "destination when leaving JSA" before August of 1996, we use information on "exits to all destinations" to perform this analysis.

Suppose, for instance, that the estimated effect of the New Deal Gateway lies within typical values of the historical estimates. This might be an indication that such result is determined by some differential aggregate variation that is not being controlled for and is captured by the program dummy. In such a case, doubts are raised on whether the estimated effect is actually capturing the causal effect of the program alone. We can go further and bound the estimated impact of the Gateway using the distribution of year-by-year estimates to construct an upper and lower bound to the estimated effect. This is done by taking the percentiles on the tail of the distribution – say, percentiles five and ninety-five or ten and ninety – as being the expected value of the estimates in the absence of a program, and using them to re-scale the estimated impact up or down accordingly.

10.3.3.3 COMPOSITIONAL CHANGES IN THE TREATMENT GROUP

Such a large-scale program may have compositional effects on the group of eligible individuals. Having learned about the eligibility rules, potential participants may change their behavior in order to secure or avoid enrollment. If such a selection process is taking place, the estimated effects of the program will be affected because the groups being compared are not what they would have been in the absence of the program. We check for this selection bias by examining difference-in-differences estimates of individuals' probabilities of exiting unemployment in the pre-treatment period (i.e. in the months before reaching six months unemployment when the program begins).

10.4 Data

The data are drawn from the publicly available five percent longitudinal sample of the whole population claiming Jobseeker's Allowance (JSA) in the UK from 1982 to June 1999 (the JUVOS database). This is an administrative database that includes individual information on spells on JSA, the unemployment benefit available in the UK; the main focus being the starting and ending dates of the spells. Individuals can be followed through all their JSA spells since the same group of the population is followed over time. However, although we know the length of time in non-JSA spells, we have no information on any transitions between different jobs during these periods. Since 1996, however, the agencies have collected data on the destination when leaving the claimant count. There are twenty different destination codes, including exit to employment, training/education, other benefits, incarceration, etc. The JUVOS data set also includes a small number of other variables – age, gender, marital status, geographic location, previous occupation and sought occupation. Descriptive statistics on the treatment group and different comparison groups are presented in Appendix A, Table 10.1A.

We also make use of the New Deal Evaluation Database (NDED), an administrative data set that contains information on virtually all individuals that have gone through the New Deal, even if only briefly. For participants, very detailed information is available from the time they join the program, including the types of treatment being administered and the timing of each intervention, letters being sent and interviews being made, a long list of socio-demographic variables and the destination when leaving the program. Non-participants, however, are not included in the sample, which limits its use for evaluation purposes. Note that we only consider the flow at six months, so there is no direct problem with mixing the stock and flow.

The use of the evaluation dataset NDED is meant to complement the lack of information in benefit (JSA) administrative records about the take-up of New Deal options. Since starting an option implies dropping from the JSA claimant count, there is a potentially large group that is being re-classified as non-unemployed while simply being driven through the program according to its rules.

Unfortunately, we are unable to securely identify these types of exits from the JUVOS data set.[19] We use the NDED instead to know the proportion of participants that enroll in each type of option (in any given region-date) by length of the New Deal spell.

In drawing up the treatment groups we have used 19-24 year olds even though the New Deal also affects 18 year olds. This is because 18 year olds can still be in high school and in England high school is only compulsory up to the age of sixteen. Participation of 16 to 18 year olds in full-time education grew rapidly over this period so we decided to avoid any time varying composition effects by dropping 18 year olds. In any case, inclusion made no difference to the results.[20]

The historical period we are examining is partly dictated by the data. The current JUVOS data ends in July 1999. For the National Roll Out we consider all individuals who finished a 6-month JSA spell between April and December 1998 and then follow them up to four months later (so our end date is April 1999). We match this with the individuals who finished a 6-month JSA spell between April and December 1997. For the Pilot Study we compare individuals completing a 6-month JSA spell between the start of January and the end of March 1998 in the Pathfinder areas to the same group in January through March 1997. Ending the sample in April 1999 has the advantage that we avoid contaminating the New Deal effect with the introduction of the national minimum wage enforced from April 1999 onwards.[21]

Some information on the macroeconomic climate is given in Figure 10.3. The New Deal was introduced at a favorable point of the business cycle by historical standards. There was no rapid improvement in the labor market between Spring 1998 and 1999, however, unlike the previous 12 months. The changing business cycle illustrates the reason why we have to select our comparison groups carefully in implementing our approach to ensure that these macro trends are "differenced out".

Finally, it should also be pointed out that the effects of the program in this favorable climate may not be easily applied to less favorable periods. First, the pool of unemployed is likely to be of worse quality when the aggregate economy is booming. Opposing this is the fact that, in the presence of firing costs (formal or informal) hiring someone in boom may be less risky.

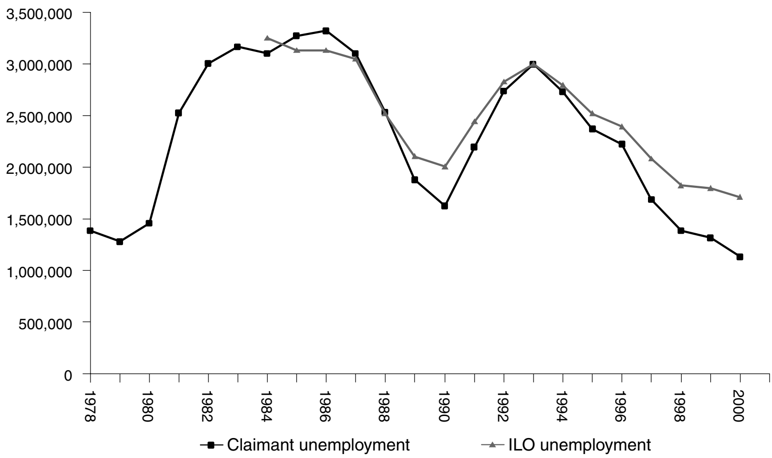


Figure 10.3. Unemployment – Claimant and ILO Measures

Notes: Data on ILO defined (International Labour Organization) unemployment were taken from the Labor Force Survey and claimant count unemployment taken from Labor Market Trends (various years). The ILO definition is based on asking out of work individuals whether they would be available and prepared to accept a job within two weeks. The claimant count is the number of people who are receiving unemployment benefit (called Jobseeker's Allowance since 1994). Although the series track each other relatively well, there will be some people who are ILO unemployed who will not be in the claimant count (e.g. if they left their job voluntarily this will disqualify them for benefit receipt for a period of time). Similarly some individuals could be claiming unemployment benefit without genuinely searching for a job.

10.5 Results

This section presents estimates of the impact of the Gateway on the flows into employment. We analyze men and women separately given the different composition of the two groups and characteristics of their behavior. We start by considering the men's case during the pilot period in Subsection 10.5.1, and discuss the different possible estimates and respective underlying assumptions available. Subsection 10.5.2 presents the results obtained for men during the National Roll Out, establishing a comparison with what the estimates were for the pilot period and assessing their robustness. Subsection 10.5.3 presents the results for women and Subsection 10.5.4 compares the magnitude of our results with those from similar US programs.

10.5.1 *Pilot Study: Men's Results*

Table 10.1 presents the main estimates of the impact of the Gateway on eligible men living in Pathfinder areas during the pilot period. We consider a number of different possible comparison groups, providing some insight on the possible size of indirect effects. Each row in the table corresponds to a different comparison, including different estimates, obtained under different methods, of the effects of the Gateway on outflows to employment after four months of treatment.[22]

The first row of Table 10.1 compares men aged 19 to 24 years old living in Pathfinder areas to other 19 to 24 year old men (with the same unemployment duration) living in all non-Pathfinder areas. After four months of treatment, it is estimated that the Gateway has improved participants' exits into employment very significantly – all the estimators point to an impact of about ten to eleven percentage points. This effect is even more impressive if compared with the outflow rates reported in Table 10.2. In the pre-program period only twenty-four percent of individuals in the treatment group obtained employment over the similar four months period (compared to thirty three percent afterwards). Thus, the improved job search assistance provided during the Gateway seems to have raised the probability of getting a job by about 42 percent (= 10%/24%) after four months of treatment.

Of course, this result should be contrasted with the information from the NDED concerning outflows into the employment option (the wage subsidy that may be offered to those who have not found employment through job assistance). It is estimated that the outflows into an employment option after 4 months of treatment sum up to 5.7 percent of men joining the Gateway (see Table 10.1). Subtracting this off the overall New Deal effect would give a "pure" Gateway impact

Table 10.1. (Men): Program Effect on Employment by the End of the Tenth Month After Starting an Unemployment Spell (conditional on being unemployed for six months); Pilot Period

				(1)	(2)	(3)	(4)
				Estimates based on the Difference In Differences combined with			
Experiment	Treatment group	Comparison group	Number of observations	Linear Matching (OLS/Linear probability model)	Non-linear matching with non-additive error term (Logit specification)	Propensity score matching using smoothing splines	Non-linear matching using smoothing splines (Logit specification)
(1)	19–24 year olds living in Pathfinder areas	19–24 year olds living in all non-Pathfinder areas	3,716	0.110** (0.039)	0.098** (0.039)	0.104** (0.046) (0.024; 0.182)	0.098** (0.044) (0.015; 0.176)
(2)	19–24 year olds living in Pathfinder areas	19–24 year olds living in matched non-Pathfinder areas	1.193	0.134** (0.053)	0.073 (0.060)	0.093 (0.073) (–0.015; 0.226)	0.080 (0.063) (–0.018; 0.190)
(3)	19–24 year olds living in Pathfinder areas	25–30 year olds living in Pathfinder areas	1,096	0.104* (0.055)	0.091 (0.057)	0.078 (0.079) (–0.050; 0.195)	0.074 (0.069) (–0.068; 0.182)
(4)	19–24 year olds living in Pathfinder areas	31–40 year olds living in Pathfinder areas	1.169	0.159** (0.050)	0.096 (0.062)	0.099* (0.078) (–0.015; 0.231)	0.082 (0.082) (–0.063; 0.205)
Outflow into the employment option (affecting 19–24 year olds in Pathfinder areas)			4,486	0.057			
(5)	25–30 year olds living in Pathfinder areas	25–30 year olds living in all other areas	3.180	0.016 (0.042)	–0.012 (0.043)	0.027 (0.049) (–0.058; 0.107)	0.031 (0.050) (–0.052; 0.109)
(6)	25–30 year olds living in Pathfinder areas	25–30 year olds living in matched non-Pathfinder areas	983	0.055 (0.058)	–0.027 (0.056)	–0.003 (0.066) (–0.107; 0.112)	–0.018 (0.078) (–0.144; 0.117)

(7)	19–30 year olds living in Pathfinder	19–30 year olds living in all other areas	6,896	0.066** (0.029)	0.052* (0.030)	0.058* (0.034) (0.004; 0.114)	0.051 (0.034) (–0.004: 0.109)
(8)	19–50 year olds living In Pathfinder areas	19–50 year olds living in all other areas	12,749	0.036* (0.021)	0.035* (0.021)	0.044* (0.023) (0.004; 0.080)	0.042* (0.023) (0.004; 0.078)

Notes: Each cell contains an estimate of the effects of the New Deal program using the JUVOS five percent longitudinal sample of all unemployed (JSA claimants). The "dependent variable" in each regression is whether an individual left unemployment between the sixth and eighth month of an unemployment spell. The average values of these are in Table 10.2 below. Estimates of the outflows into the employment option are from the New Deal Evaluation Database (NDED). The selected observations are individuals completing a six month spell of unemployment which began over a predefined time interval – this table considers inflows in the first quarters of 1997 and 1998. These individuals are then followed up to the end of the tenth month on unemployment to check whether they have found a job. The eligible group (defined by the age and area criteria) is compared with the selected control group before and after the start of the program. All estimates are from regressions that include a set of other controls: marital status, sought occupation, region and labor market history (the total number of JSA spells and the proportion of time on JSA over the two years preceding the start of the present spell). Age and the number of JSA spells since 1982 are also included when similar age groups are being compared. Propensity score matching is performed over the same covariates as the other estimates and the outcomes for the comparison groups are smoothed using cubic splines on the two propensity scores to achieve higher precision. Standard errors in parentheses: estimates for non-linear matching method (column 2) used the delta method and estimates for the propensity score matching (columns 3 and 4) used bootstrapping with 200 replications. Bias-corrected 90% confidence intervals in italics (estimation used the same bootstrap results).
** = significant at 0.05 level. * = significant at 0.10 level.

(on outflows to unsubsidized employment) of about four to five percentage points. But this is likely to be a lower bound. The calculation assumes that there is essentially no deadweight of the employer subsidy. This happens under the assumption that participants can be split into groups according to their ability to find a job, and that subsidized jobs are being attributed to those in need of a subsidy to leave unemployment. If, on the other extreme, it is believed that the subsidized jobs are being allocated to the most employable participants, then the amount of scaling down required might be small and the "true" effect would be closer to the full ten or eleven percentage points. Thus, four percentage points is a lower bound for the pure Gateway/job assistance effect. The method used to estimate the impact of treatment does not seem to substantially influence the results, reflecting some robustness of the estimates to the functional form assumptions.[23]

The rest of the rows in Table 10.1 present estimates for some of the other identifiable parameters discussed in Section 10.3, also providing some clues about the robustness of the results. We start by restricting the comparison group to be composed of eligible men living in matched non-Pathfinder areas in the second row. Depending on the method used, the

Table 10.2. (Men): Flows From the Claimant Count Into Employment by the End of the Tenth Month Since Starting an Unemployment Spell (conditional on being unemployed for six months)

	Flows by the end of the 10th month on JSA		
	Before the program	After the program	Difference
Pilot period			
Treatment group:			
19–24 year olds in Pathfinder areas	0.241	0.330	+0.089
Comparison group:			
19–24 year olds in all other areas	0.271	0.250	–0.021
Comparison group:			
19–24 year olds in matched non-Pathfinder areas	0.228	0.233	+0.005
Comparison group:			
25–30 year olds in Pathfinder areas National Roll Out	0.276	0.260	–0.016
Treatment group:			
19–24 year olds	0.258	0.281	+0.023
Comparison group:			
25–30 year olds	0.230	0.199	–0.031

Notes: The data are taken from the JUVOS five percent longitudinal sample of all unemployed (JSA claimants). Selected observations are those individuals completing a six month spell on JSA over a predefined time interval. Individuals satisfying this criterion are then followed up to the end of the eighth and tenth months of unemployment to check whether they have found a job. The present table considers the first quarters of 1997 and 1998 for the "Pilot period" estimates and the second to fourth quarters of 1997 and 1998 for the "National Roll Out" estimates. The eligible group (defined by the age and area criteria) is compared with the selected control group

estimated effect may rise or fall slightly, but not significantly so. This evidence supports the comparability of the two groups used in row 1.

The third row compares eligible and ineligible men aged 25 to 30 years old within the Pathfinder areas. Using an age-based eligibility criterion is our second main source of identification and is all that is available after the pilot period. The point estimates of the four months effect using age-based are very close and insignificantly different from those in row 1 using different areas. The linear matching estimator, for example, suggests a treatment effect of 10.4 percentage points when 25–30 year olds are used as the comparison group (row 3) compared to 11 percentage points when 19–24 year olds in non-Pathfinder areas are used as a comparison group (row 1). It was emphasized in Section 10.3 that this estimate is based on different assumptions from the estimates in rows 1 and 2. In fact, it may suffer from substitution more acutely and it is not immune to local labor market wide wage effects. However, it is informative to know that the obtained results are very similar, independently of the procedure used. We cannot reject the simple null hypothesis of a model without substitution and equilibrium wage effects. Alternatively, their effects may cancel out, the relative sizes of the substitution and wage effects being very similar. We further test for substitution using the older group of 31 to 40s living in Pathfinder areas as a comparison group. This group is expected to be less substitutable for 19–24 year olds than the younger 25–30 year old comparison group. Under this assumption, and given that substitution exacerbates the impact of the program, we would expect this estimate to be lower than the one presented in row 3. But the fourth row presents an estimate of the 4 months effect of the New Deal that, if anything is higher than the previously presented results. This is not consistent with large substitution effects. In rows 5 and 6 we compare ineligible individuals living in Pathfinder and non-Pathfinder areas. If there were significant substitution effects or differential trends across regions we may find differences in outflows in the New Deal period. In fact, no significant effects of the New Deal on non-eligibles are found.

Finally, rows 7 and 8 in Table 10.1 contain estimates of the employment effect in the "whole market". Men aged 19 to 30 and 19 to 50 years old and living in Pathfinder areas are compared with similar individuals living in non-Pathfinder areas. The results only confirm what has been established before: that, during the pilot period, the program had a very significant positive impact on outflows to employment in the markets it has been implemented. The point estimates are smaller because 19–24 year olds are only a fraction of the larger age range. For example, just over half the 19–30 year old group are 19–24 year olds. The linear matching estimator in row 7 implies a New Deal effect of 6.6 percentage points – as expected just over half the magnitude of the effect in row 1.

It is interesting to check how sensitive these results are to historical patterns. The lack of information about destinations when leaving the claimant count before 1996 imposes the use of a different variable, outflows to all destinations, to perform this analysis. Figure 10.4 considers different types of comparisons and plots the estimates of non-existent programs over time. The first panel in the chart compares eligible individuals living in Pathfinder areas with eligible individuals living in all other areas. The size of the New Deal effect, represented by the last point in the graph, is well above all other estimates for previous periods. This is just more evidence that the effects of the program on participants during the pilot period are very positive. Panel 2 compares participants with eligible individuals living in matched non-Pathfinder areas. It shows a similar pattern but with a stronger effect of the New Deal, which may be a consequence of the higher

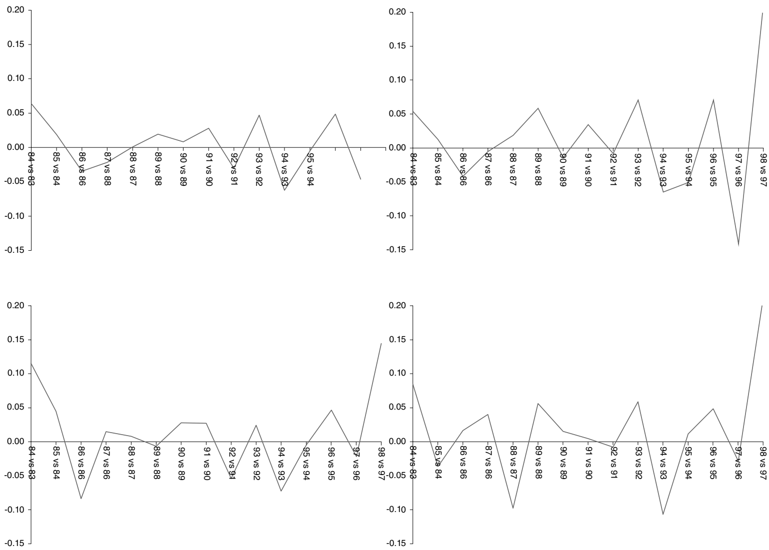


Figure 10.4. Difference-in-Differences Estimates Over Time. Outflows to All Destinations

Notes: Each panel presents the year-by-year difference in difference estimates of the impact of "fictional" programs on the total outflows from unemployment within four months of completion of the sixth month of unemployment. The total outflow is used because it is the only historic information available on a consistent basis for all years. The definition of the treatment and control groups follows the same rules as the ones used to estimate the New Deal program effect. The treatment group are all those aged 19 to 24 years old living in Pathfinder areas and are being compared with individuals of the same age group living in all other areas (Panel 1) or in matched areas (Panel 2), and with older groups in Pathfinder areas (Panel 3 for the 25 to 30 years old and Panel 4 for the 31 to 40 years old).

volatility observed. Panel 3 and 4 also confirm the importance of the estimated impact of the New Deal by comparing participants with older groups.

10.5.2 *National Roll Out: Men's Results*

Table 10.3 contains the main result from the National Roll Out. The first row shows an implied effect of around 5 percent on a pre-program base outflow (Table 10.2) of 25.8 percent, and once more, the method used does not seem to affect the result significantly. Although this is still a substantial impact, it is about half the magnitude estimated for the pilot period. These differences in size can be accounted for by a "program introduction" effect. In the first few months the program is operating, a very large increase in the flows to employment is observed, which then falls as the program matures. This is illustrated in the other rows of the table. The second and third rows report comparable estimates of the Gateway effect after 4 months of treatment for the first quarter the program operates in the Pathfinder and non-Pathfinder areas, respectively. As noticed before, estimates for the pilot period (first quarter in Pathfinder areas) are about twice the size of the effect over the whole period. The same is also true if one considers the estimates for the first quarter the New Deal operates in non-Pathfinder areas (see row 3). The fourth row presents estimates obtained using the following second and third quarters the program is operating and these are comparatively much lower and less significant.

There are, of course, many possible explanations for this. One explanation is that the agencies involved in delivering the program are initially very enthusiastic, but this naturally erodes over time. Another possibility is that the program diminishes welfare fraud. This would have particularly important effects during the first few months after the release of the program since potential participants are unlikely to be aware of the new claiming rules. Similar "cleaning up the register" effects have been noted of previous UK labor market reforms.[24]

There are many possible criticisms of the results. We shall now discuss some of the main ones – quality of job matches, selectivity and differential trends. How the program affects the women will be discussed in the next section.

First, there is the issue of whether the quality of job matches has improved (or deteriorated) under the New Deal. One of the benefits from the New Deal is said to be that job matches are of higher quality due to greater job assistance and mentoring of the personal advisor. For those who get onto the employer option there is a guarantee of one day a week training. On the other hand tougher monitoring may push claimants into low quality matches. Quality is difficult to measure without data on earnings and other job characteristics. One indicator of job match quality, however, is simply the longevity of a job. Following the government's preferred measure, we define a "sustained" job as one that lasts at least thirteen weeks. The first row of Table 10.4 Panel A repeats the analysis but uses the outflow to sustained jobs (instead of any job) as the outcome variable. The results are quite consistent with the earlier findings – the estimates point to an increase in the outflows to sustained jobs of 4.5% (in column 1 of Table 10.4), which compares to estimates of around 5% for the outflows to all employment (in column 1, first row of Table 10.3).

Table 10.3. (Men): Program Effects on Employment by the End of the Tenth Month Since Starting an Unemployment Spell (conditional on being on unemployed for six months); Comparing 19 to 24 Year Olds With 25 to 30 Year Olds Living in the Same Areas; Pilot Period and National Roll Out

Experiment	Type of estimate	Number of observations	(1)	(2)	(3)	(4)
			Estimates based on the Difference in Differences methodology combined with			
			Linear Matching	Non-linear matching with non-additive error term (Logit specification)	Propensity score matching using smoothing splines	Non-linear propensity score matching using smoothing splines (Logit specification)
(1)	Overall effect for the sample including the Pilot period and the National Roll Out (first three quarters the New Deal is operating in each region)	17,433	0.053** (0.013)	0.044** (0.015)	0.048** (0.015) (0.020; 0.069)	0.049** (0.016) (0.018; 0.072)
	Outflows to subsidized jobs	55,051		0.039		
(2)	Effect for the Pilot period – 1st quarter the program operates in Pathfinder areas	1,096	0.104* (0.055)	0.091 (0.057)	0.078 (0.079) (–0.050; 0.195)	0.074 (0.069) (–0.068; 0.182)
	Outflows to subsidized jobs	4,486		0.057		
(3)	Effect for the 1st quarter the program operates in non-Pathfinder areas	5,169	0.088** (0.025)	0.064** (0.027)	0.078** (0.031) (0.021; 0.131)	0.075** (0,831) (0.015; 0.128)
	Outflows to subsidized jobs	20,331		0.039		
(4)	Effect for the 2nd and 3rd quarters the program operates in all areas	11,161	0.031* (0.016)	0.023 (0.019)	0.024 (0.019) (–0.008; 0.051)	0.827 (0.819) (–0.005; 0.057)
	Outflows to subsidized jobs	30,234		0.036		

Notes: Estimates of the effects of the New Deal program used the JUVOS five percent longitudinal sample of all unemployed (JSA claimants). Estimates of the outflows into employment option used the New Deal Evaluation Database (NDED). Selected observations are those completing a six month spell on JSA over a predefined time interval: the present table compares 1997 with 1998. These individuals are then followed up to the end of the tenth month of unemployment (JSA) to check whether they have found a job. The eligible group (defined by the age criterion) is compared with the control group before and after the start of the program. All the estimates from regressions including a set of other controls: marital status, sought occupation, region and some information on the labor market history (the number of unemployment spells and the proportion of time on unemployment over the two years that preceded the start of the present spell). Propensity score matching is performed over the same covariates as the other estimates and the outcomes for the comparison groups are smoothed using cubic splines on the two propensity scores to achieve higher precision. Standard errors in parentheses: estimates for non-linear matching method (column 2) used the delta method and estimates for the propensity score matching (columns 3 and 4) used bootstrapping with 200 replications. Bias-corrected 90% confidence intervals in italics - estimation used the same bootstrap results.
** = significant at 0.05 level. * = significant at 0.10 level.

Secondly, there is the issue of selectivity. It may be that the introduction of the New Deal has an effect on the (unobserved) quality of the inflow of individuals reaching six month of JSA. The most likely route for this is that claimants in the fifth or sixth months of JSA may alter their behavior. If they believe the New Deal regime is "tougher" than the previous regime, they may be more likely to leave the unemployment rolls (this was one of the ways that RESTART, another job assistance program introduced in 1986 was deemed to have worked). On the other hand, if the New Deal is seen as a desirable thing (e.g. because of subsidies to "good jobs" or training), then claimants may delay exit. If the main effect is increased toughness, then we may underestimate the positive effects of the New Deal as there has been a decline in the unobserved quality of the stock (assuming the most job ready decide to leap into jobs before they are pushed off the unemployment rolls). If the New Deal is perceived as more attractive than the previous regime (as the qualitative evidence suggests) then we may actually be overestimating the effects of the Gateway period as the more job ready actually delay their exits prior to entering the Gateway.

To investigate these selectivity problems we examine outflows to employment during the fourth and fifth month of JSA, using the same methodology as before. The results are presented in rows 2 and 3 of Table 10.4, Panel B. The introduction of the New Deal had no significant impact on the outflows to employment prior to six months duration. All the estimates are small and insignificant at conventional levels.

Thirdly, we have not controlled for differential trends. Using the same method as before (see section V.1) we calculate upper and lower bounds for the New Deal effect on outflow rates. The average effect is again smaller than the estimates for the pilot period (see rows 5 and 6 of Table 10.4, Panel C). Nevertheless, even at the lower bound there is a significant effect of the program on the outflow rates to all destinations.

10.5.3 *The Impact of the Program on Women*

Finally, note that we have focused our results on male job outflow rates. Three quarters of all participants in the New Deal are men, but clearly the impact on women is also of great interest. The results for women are not as clear-cut as those for men. This is mainly because there is a systematic trend in the labor market behavior of older (25–30) compared to younger (19–24) women. The main problem, therefore, resides on the choice of the appropriate comparison group.

Table 10.4. (Men): Robustness of the Results. Comparing 19–24 Year Olds With 25–30 Year Olds in the Same Areas. Estimates for the First Three Quarters the Program is Operating in Each Area

	Outcome variable	Number of observa-tions	(1)	(2)	(3)	(4)
			Estimates based on the Difference in Differences methodology combined with			
			Linear Matching	Non-linear matching with non-additive error term (Logit specification)	Propensity score matching using smoothing splines	Non-linear propensity score matching using smoothing splines (Logit specification)
Panel A: Outflows to sustained jobs (conditional on being on JSA for 6 months						
(1)	Estimates	17.433	0.045** (0.011)	0.031** (0.013)	0.035** (0.013) (0.013; 0.055)	0.033** (0.016) (0.005; 0.054)
	Outflows to sustained subsidized jobs (affecting 19 to 24 year olds)	55,051			0.031	
Panel B: Outflows to employment before the start of the Gateway (conditional on being on JSA for 4 or 5 months)						
(2)	Effect between months 5 and 6 of JSA	20,957	0.004 (0.008)	0.000 (0.010)	0.004 (0.009) (−0.011; 0.019)	0.003 (0.010) (−0.013; 0.020)
(3)	Effect between months 4 and 6 of JSA	25,510	0.009 (0.010)	0.001 (0.011)	0.009 (0.011) (−0.011; 0.026)	0.009 (0.011) (−0.010; 0.027)
Panel C: Outflows to all destinations (conditional on being on JSA for 6 months)						
(4)	Estimates	17.433	0.108** (0.015)	0.093** (0.016)	0.095** (0.018) (0.061; 0.124)	0.095** (0.018) (0.060; 0.123)

(5)	Lower bound		0.084**	0.062**	0.048**	0.046**
			(0.019)	(0.020)	(0.023)	(0.022)
					(0.010; 0.087)	(0.010; 0.084)
(6)	Upper bound		0.143**	0.119**	0.126**	0.133**
			(0.019)	(0.020)	(0.024)	(0.026)
					(0.087; 0.164)	(0.091; 0.175)
	Outflows to all ND options (affecting 19 to 24 year olds)	55,051			0.137	

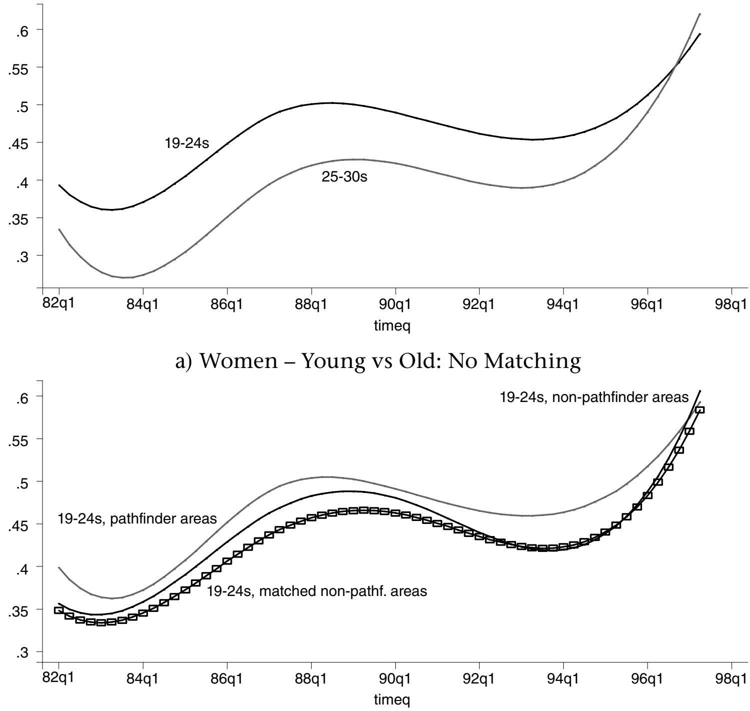
Notes: Estimates of the effects of the New Deal used the JUVOS five percent longitudinal sample of all unemployed (JSA claimants). Estimates of the outflows into employment option used the New Deal Evaluation Database. All estimates are based on the comparison between 1997 and 1998 and compare the eligible group (defined by the age criterion) with the selected control group before and after the start of the program to estimate its impact. Panel A refers to the stock of individuals completing a six month unemployment spell and follows them up to the end of the tenth month to check whether they have found a "sustained" job. An exit into employment is classified as sustained if it takes more than thirteen weeks for the individual to return to the claimant count.

Panel B uses the stock of individuals completing either a four or a five month spell on JSA and follows them up to the end of the sixth month on JSA to check whether they have found a job.

Panel C uses the stock of individuals completing six months of unemployment and follows them up to the end of the tenth month on JSA to check whether they have left unemployment. Upper and lower bounds are presented in Panel C using historical series of a similar parameter (see text for details).

All estimates from regressions including a set of other controls: marital status, sought occupation, region and some information on the labor market history (the number of unemployment spells and the proportion of time in unemployment over the two years that precede the start of the present spell). Propensity score matching is done over the same covariates as the other estimates and the outcomes for the comparison groups are smoothed using cubic splines on the two propensity scores to achieve higher precision. Standard errors in parentheses: estimates for non-linear matching method (column 2) used the delta method and estimates for the propensity score matching (columns 3 and 4) used bootstrapping with 200 replications. Bias-corrected 90% confidence intervals in italic – estimation used the same bootstrap results.

** = significant at 0.05 level. * = significant at 0.10 level.



b) Women – Young vs Old: No Matching

Figure 10.5. Outflows From JSA Conditional on Completing 6 Months. Effect by the End Month 10

Notes: This graph illustrates the proportion of women leaving unemployment between the sixth and tenth months of unemployment 1982–1998. "PF" indicates that the men were living in a Pathfinder Pilot area (prior to New Deal introduction in 1998). The data have been smoothed by a cubic spline in time. Breakpoints were included at the first quarter of 1987 and the first quarter of 1990.

Figure 10.5 illustrates the difficulties encountered by plotting the conditional exits to all destinations against time for treatments and different possible comparison groups. It is apparent from the upper panel of Figure 10.5 that an estimator based on different age groups can be severely contaminated by differential trends. Compared to the younger age groups, the older age groups seem to have systematically improved their position in the labor market over the 1982–99 period. If this trend extends to the treatment period, it is expected that such comparison under-estimates the impact of treatment on the treated. On the other hand, the lower panel of the graph suggests that the macro shocks seem to affect younger age groups living in different geographic regions much more similarly, making the Pathfinder – non-Pathfinder 19–24 year old groups comparable. Matching on regions improves the pattern, the two curves for treatment and comparisons being closer both in levels and slopes. The upshot of this is that using older women as a comparison group is not valid, and we should focus on the Pathfinder data to evaluate the effect of the New Deal for women.

Table 10.5 presents some estimates of the impact of the program on treated individuals using different comparison groups and estimation techniques. All estimates resulting from the comparison of similar age groups point to a positive effect of the program on the outflows to employment (see rows 1 and 2). These estimates are much less precise, more sensitive to the estimation technique used and generally smaller, but do not seem to reject the conclusions drawn for men. For example, the linear matching estimator in row 1 suggests an impact effect of 6.1 percent compared to 10.0 percent for men. The lack of precision is likely to be a consequence of the smaller sample sizes. Notice that the increased job taking-up rate seems to be mainly accounted for by the employment option, which ensured a job to almost 5 percent of the treated during this period. As expected, comparing different age groups changes the results drastically and in the predicted direction (see row 3): despite remaining statistically insignificant, the estimates are actually negative. Together with the pattern depicted in Figure 10.5, this explains why the women's case is not explored during the National Roll Out of the program. The only group we can draw comparisons from is composed of individuals older than the participants, and these are subject to very differential trends.

10.5.4 *Discussion of the Results: A Comparison with the Existing Literature*

How do our findings compare with the existing results? We overlap with several other program evaluation literatures: Unemployment insurance (UI) reform, wage subsidies, youth measures over education and training. Perhaps the most directly relevant are the recent program evaluations of mandatory job search associated with welfare-to work-reforms. Bloom and Michalopoulos (2001) survey 29 different initiatives that had demonstration projects. Eight of these schemes were job-focused (rather than education/training focused) and mandatory for welfare recipients. Table 10.6 summarizes the results from these studies and shows that although the precise impact effect differed probabilities was found in all eight cases. The median of the impacts in the final column of Table 10.6 is 0.23, which is not wildly out

Table 10.5. (Women): Gateway Employment Effects by the End of the Tenth Month (conditional on being on JSA for six months); Pilot Period

				(1)	(2)	(3)	(4)
				Estimates based on the Difference in Differences methodology combined with			
Experiment	Treatment group	Comparison group	Number of observations	Linear Matching	Non-linear matching with non-additive error term (Logit specification)	Propensity score matching using smoothing splines	Non-linear propensity score matching using smoothing splines (Logit specification)
(1)	19–24 year olds living in Pathfinder areas	19–24 year olds living in all non-Pathfinder areas	1,592	0.061 (0.058)	0.026 (0.060)	0.057 (0.084) (–0.073; 0.219)	0.051 (0.083) (–0.096; 0.19)
(2)	19–24 year olds living in Pathfinder areas	19–24 year olds living in matched non-Pathfinder	596	0.025 (0.071)	0.013 (0.077)	0.136 (0.151) (–0.106; 0.374)	0.113 (0.149) (–0.162; 0.334)
(3)	19–24 year olds living in Pathfinder areas	25–30 year olds living in Pathfinder areas	400	–0.047 (0.100)	–0.057 (0.101)	–0.053 (0.213) (–0.447; 0.270)	–0.080 (0.193) (–0.449; 0.219)
Outflow into the employment option (affecting 19–24 year olds living in Pathfinder areas)			1,653			0.046	

Notes: Estimates of the effects of the New Deal used the JUVOS five percent longitudinal sample of all unemployed (JSA claimants). Estimates of the outflows into the employment option used the New Deal Evaluation Database. Selected individuals are those completing a six month spell on unemployment (JSA) over a predefined time interval. The present table considers the first quarters of 1997 and 1998. These individuals are then followed up to the end of the tenth month of unemployment to check whether they have found a job. The eligible group (defined by the age and area criteria) is compared with the selected control group before and after the start of the program. All estimates from regressions including a set of other controls: marital status, sought occupation, region and some information on the labor market history (the number of unemployment spells and the proportion of time in unemployment over the two years that precede the start of the present spell). Age and the number of unemployment spells since 1982 are also included when similar age groups are being compared. Propensity score matching is performed over the same covariates as the other estimates and the outcomes for the comparison groups are smoothed using cubic splines on the two propensity scores to achieve higher precision. Standard errors in parentheses: estimates for non-linear matching method (column 2) used the delta method and estimates for the propensity score matching (columns 3 and 4) used bootstrapping with 200 replications. Bias-corrected 90% confidence intervals in italic – estimation used the same bootstrap results.
** = significant at 0.05 level. * = significant at 0.10 level.

Table 10.6. Evaluations of the Employment Impact of Welfare to Work Programs with Mandatory Employment Services (random assignment); Proportion Who Were Ever Employed in First Year After Treatment

1. Study	2. Sample size	3. Control group outcome (% employed)	4. Impact of program on treatment group (additional % employed)	Ratio of impact effect to control group proportion (=column 4 divided by column 3)
Job-search first programs				
SWIM	2,850	39%	11%	0.28
Atlanta LFA	3,783	48%	5%	0.1
Grand Rapids LFA	3,010	53%	10%	0.19
Riverside LFA	6,611	35%	17%	0.48
LA Jobs First GAIN	15,122	47%	11%	0.23
Employment-focused programs with mixed initial activities				
Project Independence	9785	50%	4%	0.08
Riverside GAIN	4640	31%	20%	0.64
Portland NEWWS	5442	47%	11%	0.23

Notes: These are all of the employment-focused programs evaluated by MDRC using random assignment. Education and training focused programs (MDRC's definition) are not included. The "impact" column (4) shows the difference in the proportion of the treatment group who got a job in the year after the program minus the same proportion in the control group. All impacts are statistically significant at the .05 level. The employment effect appears to diminish over time. With the exception of Portland, year three effects are all smaller than year one effect.

Source: Derived from Bloom and Michalopoulos (2001) Appendix Table C.1.

of line with our "central" estimate of a program impact of 0.2. Again we should note that 0.2 is probably an 'upper bound' measure since, as we have noted, a large part of this employment effect is towards subsidized jobs and also due to a "first quarter" effect.

Unlike the US welfare to work reforms where the affected groups are overwhelmingly females with children, the New Deal's main participants are men. Experiments over unemployment insurance reforms may, therefore, be more relevant. Meyer (1995) discusses five randomized trials and finds that increased job search assistance and monitoring significantly reduced the duration of unemployment claims (see also Katz and Meyer 1990; Meyer 1990). As with the New Deal it is unclear from these studies whether the "carrot" of job assistance or the "stick" of the tougher monitoring of job search played the most important role. Ashenfelter, Ashmore and Dechenes (1999) could find no increased benefit of stricter enforcement over job search in their examination of random trials, but Anderson (2001) and Abbring, van den Berg and van Ours (1996) do find evidence that sanctions and strict monitoring have important effects. Distinguishing between the relative importance of carrot and stick is an important area of ongoing research, but what seems to be less in doubt is that the combination of the two can be effective. It is interesting to note that in the study of worker profiling and reemployment services which involves mandatory employment and training services, Black et al. (2003) find most of the impact to be a sharp increase in early exits from UI coinciding associated with claimants finding out about their mandatory program obligations.

A feature of the New Deal is that it is youth-focused. Most evaluations of youth initiatives have been pessimistic, especially for young men (for example, Heckman, LaLonde and Smith 2000). Our study gives some room for optimism, but it should be remembered that the participant group for most US youth training programs are quite different from the British New Dealers. US schemes are focused on very disadvantaged youth – for example, long-term unemployment is rare in the US, but more common in Europe. It may be easier to help the young in the New Deal because they are far more "job-ready" than their US counterparts. In addition (unlike JTPA) we are not looking at the impact of the training/education aspects of the New Deal and have focused only on the mandatory job search and wage subsidy element.

Finally, there is an extensive literature on the role of financial incentives for employers and individuals in encouraging employment amongst the less skilled. Employer-based job subsidies of the kind discussed here are rarer than individual-based incentives such as EITC.[25] Both types of policies can be successful in raising employment,[26] but this conclusion depends very much on the exact program. A major problem with employer-based wage subsidies is that they have very low take-up by employers, perhaps due to stigma or administrative burdens.[27]

In summary, the finding of a small positive employment effect of the New Deal is not out of line with the results in the US literature. However, there remains the question of whether the social costs of the program justify the benefits. In this paper we do not embark on a full cost-benefit calculation since the longer-term effects of the program are unknown (especially the human capital raising elements). Nevertheless, Layard (2000) and Van Reenen (2004) make a preliminary attempt to gauge the costs using administrative data and assumptions over the size of earnings gains. They both find that the social benefits outweigh the social costs.

10.6 Conclusions

This paper has examined the labor market impact of the British New Deal for Young People. The New Deal is a compulsory program affecting all young people claiming unemployment benefit for at least six months. The program offers a combination of treatments, particularly job assistance for four months and a wage subsidy paid to employers. Two sources of identification are used to construct comparison groups in order to make inferences on the impact of the New Deal: a comparison between pilot areas and non-pilot areas and age-related eligibility criteria. Our results suggest similar quantitative effects whichever comparison group is chosen.

Based on the pilot period of the program we find an economically and statistically significant effect of the program on outflows to employment among men. The program appears to have caused an increase in the probability of young men (who had been unemployed for six months) finding a job in the next four months. On average, this increase is about 5 percentage points (relative to a pre-program baseline of about 26 percent). Part of this overall effect is the job subsidy element and part is a pure enhanced job search. We estimate that at least 1 percentage point of the 5 percentage points is due to the Gateway services, such as job search assistance (rather than the wage subsidy element). We also found that the treatment impact is much larger in the first quarter of introduction compared to the subsequent two quarters. This puts in question whether the effects of this aspect of the program will be sustained in the long run. Our findings are robust to a large number of experiments, including a number of different comparison groups.

Why are our non-experimental program evaluation results more robust than those seen elsewhere in the literature? We suspect that it is due to the combination of having a clear "before and after" design and matching our treatment group closely with a comparison group of similar duration on unemployment insurance. It is worthwhile recalling that both LaLonde (1986), and Fraker and Maynard (1987) found when using comparison groups based on benefit receipt (AFDC) experimental and non-experimental estimators gave much closer results than the "youth" group as a whole. Our results have a similar flavor.

There are at least three areas of further work. First, the main omission in our work is that we do not consider the longer-term effects of the New Deal. A full evaluation needs to consider whether individuals' employability has been enhanced by their experience of subsidized work and education and training. The data is only just becoming available to perform such an analysis. A second problem lies in untangling how robust our estimates are in the face of substitution and equilibrium wage changes. To take these into account involves putting more economic structure on the problem than we have done in this paper (e.g. Blundell, Costa Dias and Meghir 2003). It is reassuring, however, that the Pathfinder pilots vs. non-pilot comparisons yielded results that were quantitatively similar to the within Pathfinder analysis. Finally, we have eschewed a formal cost-benefit analysis given the uncertainty surrounding some of the benefits such as the training and education option. However, this is clearly an important next step that will be informed by some of the estimates obtained in this paper.

Appendix A: Data

Table 10.1A compares the mean values of some of the independent variables used in the analysis before and after matching on the propensity scores.[28] It can be observed that similar age groups are much more alike, at least with respect to the considered characteristics (compare columns 1 and 2 with 5 and 6). Moreover, matching on the propensity scores significantly improves the similarity between the groups (compare columns 3–4 with 1–2 or columns 7–8 with 5–6).

A more detailed diagnosis of the quality of the propensity score matching is presented in Figures 10.1A to 10.4A. These plots represent the distribution of the two propensity scores used in the matching process over the entire population and over specific subgroups. We compare 19 to 24 year olds living in pathfinder areas with 19 to 24 year olds in all non-pathfinder areas during the pilot period. All groups being included in the analysis are plotted: treatment and comparison groups, before and after the release of the New Deal. As expected, matching significantly improves the similarity between the curves – it can be observed that the curves on the right hand side of Figure 10.1A overlap almost precisely. Moreover, nearly all the initial support is maintained after matching. Figures 10.2A to 10.4A give some indications of how identical the distributions of the propensity scores are over sub-groups of the population. It is apparent that matching worked well even over sub-populations, making the distributions quite similar. Very similar results were obtained when using other groups and are available under request.

Table 10.1A. (Men): Descriptive Statistics for Different Treatment and Control Groups. Comparing the means of some variables used in the analysis

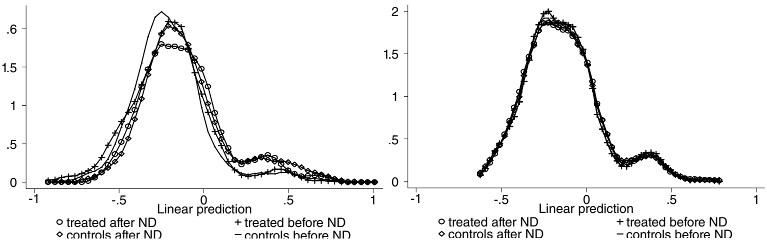
	19–24s in PF areas vs 19–24s in all other areas Pilot period				19–24s vs 25–30s in all areas First 3 quarters the program is operating			
	No matching		Matching on the propensity scores		No matching		Matching on the propensity scores	
	Treatment group (1)	Control group (2)	Treatment group (3)	Control group (4)	Treatment group (5)	Control group (6)	Treatment group (7)	Control group (8)
Number of observations	273	1,306	264	264	4,377	4,086	4,359	4,359
Marital status								
Married	.08	.10	.06	.08	.10	.23*	.10	.10
Time unemployed over the last 2 years								
Less than 6 months	.46	.48	.46	.49	.43	.33*	.43	46*
Less than 12 months	.64	.66	.64	.66	.64	.51*	.64	.64
Number of unemployment spells over the last two years								
0	.29	.26	.29	31	.20	.16*	.20	.20
1 to 2	.59	.56	.58	.57	.58	.64*	.58	.59
3 to 5	.12	.17*	.12	.10	.21	.18*	21	.19*
6 or more	.01	.01	.01	.01	.01	01	.01	.01
Sought occupation								
Manager	.03	.02	.03	.04	.02	.02*	.02	.02
Professional	.01	.02	.01	.02	.01	.03*	01	.01
Technical	.07	.07	.07	.11	.05	.07*	.05	.06
Clerical	.12	.17*	.12	.15	.18	.13*	.18	.13
Craft	.19	.12*	.19	.17	.14	.17*	.14	.14
Personal services	.11	.08	.11	.09	.07	.06*	.07	.06
Sales	.10	.10	.10	.10	.11	.07*	.10	.10

(continued)

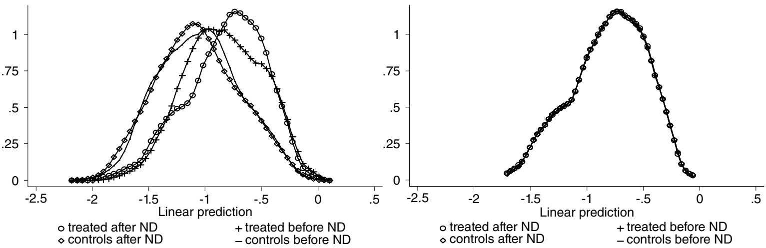
Table 10.1A. Continued

	19–24s in PF areas vs 19–24s in all other areas Pilot period				19–24s vs 25–30s in all areas First 3 quarters the program is operating			
	No matching		Matching on the propensity scores		No matching		Matching on the propensity scores	
	Treatment group (1)	Control group (2)	Treatment group (3)	Control group (4)	Treatment group (5)	Control group (6)	Treatment group (7)	Control group (8)
Machine operator	.07	.09	.07	.07	.10	.14*	.10	.11
Other	.29	.31	.28	.24	.31	.29*	31	.32
Region								
South East	.19	.26*	.19	.19	.24	.30*	.24	.26*
East Anglia	.00	.00	.00	.00	.02	.02	.02	.02
South West	.08	.06*	.08	.09	.05	.05	.05	.05
West Midlands	.17	.09*	.17	.19	.10	.08*	.10	.09
East Midlands	.04	.06*	.04	.04	.07	.07	.07	.07
York	.12	.11	.12	.13	.12	.11	.12	.11
North West	.07	.17*	.06	.04	.15	.14	.15	.15
North	.16	.07*	.16	.11	.08	.07	.08	.08
Wales	.13	.06*	.13	.16	.06	.05	.06	.05
Scotland	.05	.11*	.05	.05	.12	.11*	.12	.11

Note: *Estimated mean for treatments and controls are significantly different at a 5% level.

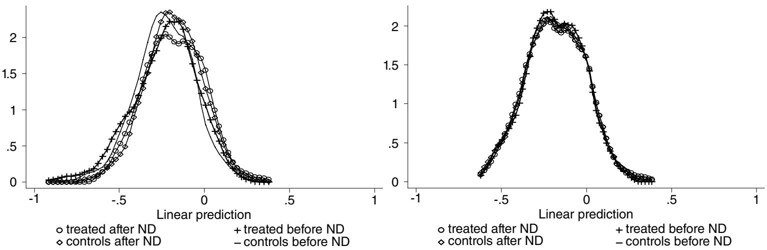


a) Before / After Comparison

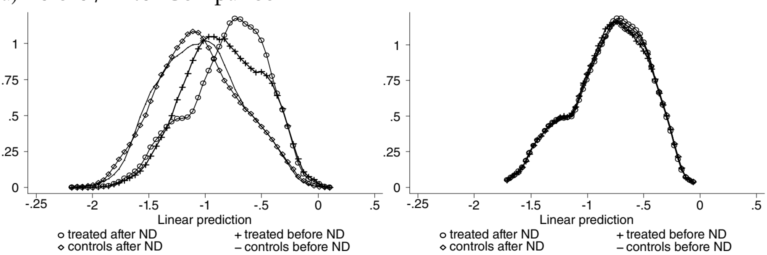


b) Treatment / Control Comparison

Figure 10.1A. Comparing 19–24s in PF Areas With 19–24s in All Non-PF Areas. Densities of the Propensity Scores Before and After Matching

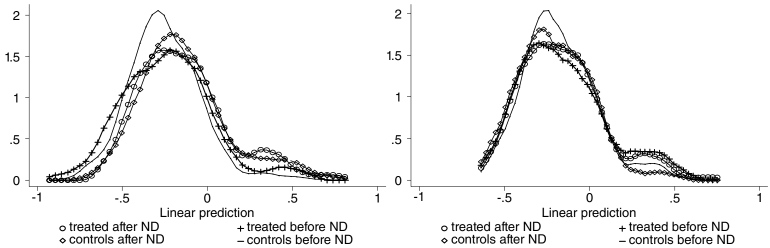


a) Before / After Comparison

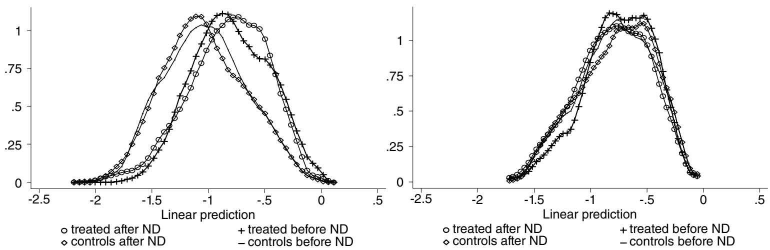


b) Treatment / Control Comparison

Figure 10.2A. Comparing 19–24s in PF Areas with 19–24s in All Non-PFs Areas. Densities of the Propensity Ores Before and After Matching: Single Individuals

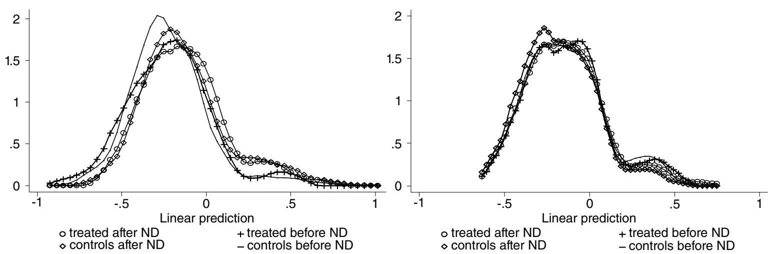


a) Before / After comparison

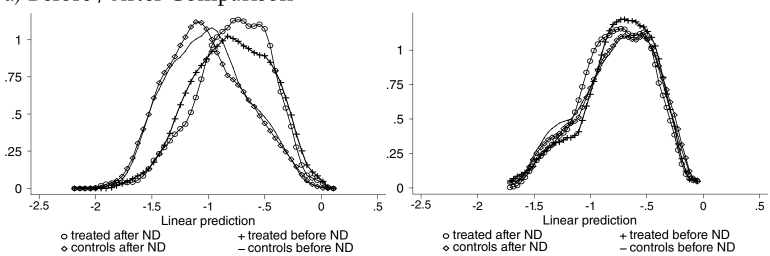


b) Treatment / Control comparison

Figure 10.3A. Comparing 19–24s in PF Areas With 19–24s in All Non-PF Areas. Densities of the Propensity Scores Before and After Matching: Total Time Unemployed Over the Last 2 Years is Less than 6 Months



a) Before / After Comparison



b) Treatment / Control Comparison

Figure 10.4A. Comparing 19-24s in PF Areas With 19-24s in All Non-PF Areas. Densities of the Propensity Scores Before and After Matching: 1 or Less Unemployment Spells Over the Last 2 Years

Appendix B: Gateway Employment Effects Under Different Propensity Score Matching Techniques

Table 10.2A presents estimates for the employment effects of the Gateway among men during the pilot period using three possible variations of the propensity score matching method under the linear specification assumption. Columns (1) to (3) present propensity score matching estimates of the parameters presented in Table 10.1 in the main text. Column (1) displays the estimates for the standard nearest neighbor propensity score method, where only one observation from each comparison group is chosen to match each observation in the treatment group – the closest one from the perspective of the two propensity scores at use. Column (2) uses the same method as in column (1) but smoothes the outcome of the comparison group. The same comparisons are chosen but the smoothed outcome is used to estimate the impact of the program. Column (3) uses kernel weights to select the counterfactual for each treatment observation: comparisons that are relatively near the treatment observation in terms of the propensity scores are given a weight depending on how close they are. These estimates used an Epanechnikov function with a diagonal matrix of bandwidths. The main result from Table 10.2A is that all methods produce similar estimates, and this remains true when comparing with the numbers in Table 10.1 in the main text. However, the precision of the estimates does change from method to method. The estimated standard errors presented in column (1) are much higher than similar estimates produced by other methods. The strong variation resulting from the fact that only one observation is being chosen as a comparison for each treated individual is in part to blame. The standard errors presented in column (3) are significantly lower but still too high to sustain a definitive conclusion. Estimates in column (2), however, are generally more precise, the result being due to the smoothing of the counterfactual outcomes.

Appendix C: Estimation Methods

The practical implementation of the completely parametric methods is discussed in the main text, and so we omit it here. We use propensity score matching based on two dimensions, time and eligibility, and using either the nearest neighbor method or smoothing the outcomes applying splines or kernel weights. With the same set of observables

Table 10.2A. (Men): Gateway Employment Effects by the End of the Tenth Month (conditional on being on JSA for 6 months); Pilot Period

Experiment	Treatment group	Comparison group	Nr of observ.	Estimates based on the Difference in Differences methodology combined with		
				(1)	(2)	(3)
				Propensity score matching (nearest neighbor)	Propensity score matching using smoothing splines nearest neighbor)	Propensity score matching (kernel weights)
(1)	19–24 year olds living in Pathfinder areas	19–24 year olds living in all non-Pathfinder areas	3,716	0.110 (0.083) (–0.028, 0.238)	0.104** 0.046) (0.024; 0.182)	0.078 (0.056) (–0.010, 0.170)
(2)	19–24 year olds living in Pathfinder areas	19–24 year olds living in matched non-Pathfinder areas	1,193	0.084 (0.100) (–0.076, 0.245)	0.093 0.073) (–0.015; 0.226)	0.070 (0.068) (–0.043, 0.183)
(3)	19–24 year olds living in Pathfinder areas	25–30 year olds living in Pathfinder areas	1,096	0.069 (0.112) (–0.117, 0.248)	0.078 0.079) (–0.050; 0.195)	0.054 (0.081) (–0.083, 0.191)
(4)	19–24 year olds living in Pathfinder areas	31–40 year olds living in Pathfinder areas	1,169	0.089 (0.129) (–0.116, 0.307)	0.099* 0.078) (–0.015; 0.231)	0.094 (0.078) (–0.034, 0.227)
(5)	25 30 year olds living in Pathfinder areas	25–30 year olds living in all other areas	3,180	0.016 (0.092) (–0.149, 0.164)	0.027 0.049) (–0.058; 0.107)	0.015 (0.063) (–0.079, 0.130)
(6)	25–30 year olds living in Pathfinder areas	25–30 year olds living in matched non-Pathfinder areas	983	–0.016 (0.126) (–0.220, 0.185)	0.003 0.066) (–0.107; 0.112)	–0.028 (0.081) (–0.167, 0.105)
(7)	19–30 year olds living in Pathfinder areas	19–30 year olds living in all other areas	6,896	0.033 (0.058) (–0.058, 0.132)	0.058* 0.034) (0.004; 0.114)	0.051 (0.041) (–0.019, 0.118)
(8)	19–50 year olds living in Pathfinder areas	19–50 year olds living in all other areas	12,749	0.025 (0.042) (–0.053, 0.094)	0.044* 0.023) (0.004; 0.080)	0.023 (0.026) (–0.025, 0.063)

Notes: Estimates of the employment effects of the New Deal program using the JUVOS five percent longitudinal sample of all unemployed (JSA claimants). Selected individuals are those completing a six month unemployment spell over a predefined time interval. The present table considers the first quarters of 1997 and 1998. These individuals are then followed up to the end of the tenth month of unemployment to check whether they have found a job. The eligible group (defined by the age and area criteria) is compared with the selected control group before and after the start of the program. Propensity score matching is performed over a set of controls: marital status, sought occupation, region and some information on the labor market history (the number of JSA spells and the proportion of time on JSA over the 2 years that precede the start of the present spell). Age and the number of JSA spells since 1982 are also included when similar age groups are being compared. Standard errors in parentheses: estimates for non-linear matching method (column 2) used the delta method and estimates for the propensity score matching (columns 3 and 4) used bootstrapping with 200 replications. Bias-corrected 90% confidence intervals in italic – estimation used the same bootstrap replications.
** = significant at 0.05 level. * = significant at 0.10 level.

used in the completely parametric estimates, we compute the two propensity scores,

$$P_{1X} = P(ND = 1 \mid X) \text{ and } P_{tX} = P(t = 1 \mid X).$$

In the nearest neighbor case, each treated individual is paired with one observation from each of the three comparison groups, the one that minimizes the Euclidean distance with respect to the two propensity scores conditional on two maximum distance restrictions, one for each dimension. Matching is done with replacement, meaning that each comparison observation may be chosen more than once and is weighted accordingly.

Under the smoothing splines method, we run a regression of the outcome of interest on a cubic polynomial of the two propensity scores for each of the comparison groups. Predictions of the outcome under the three non-treatment cases for each of the matched treated observations under the nearest neighbor method are then computed and used to estimate the impact of treatment.

The use of kernel weights to select each of the three comparison groups is based on the Epanechnikov function and a diagonal matrix of (constant) bandwidths, each element of the diagonal being given by $1.06\sigma_x n^{-1/5}$. Having constructed the three counterfactuals, the simple difference-in-differences method is applied to estimate the effect of the program under the assumption of separable additivity of the group and time effects. We also transform the outcome applying the logit transformation, as shown in equation (3.4), to estimate the impact of the ND under a non-linear specification.

Appendix D: UK Unemployment Benefit Rules

The main benefit available for unemployed young people is Jobseeker's Allowance (JSA). It was introduced in October 1996 to replace unemployment benefit. The level of JSA was about £40 a week throughout the New Deal period, though this amount depends on the age of the applicant, and the respective household income and needs. To be eligible for JSA, an unemployed person must: (i) Be "actively seeking work", which is assessed by a fortnightly short interview taking 5–10 minutes; and (ii) meet some conditions concerning the past two tax years working history, related to the amount of National Insurance contributions made while employed ("contributory JSA") or, alternatively, pass a "means test". Thus, it is possible for someone who never worked before to be entitled for the benefit. In a reform in 1986

(RESTART) more intensive job focused interviews took place at six monthly interviews.

If not before, receipt for JSA becomes "means-tested" after six months. Individuals with income from other sources (large assets or a partner bringing in income) have their JSA scaled down or taken away altogether. Prior to October 1996, this period of "non-means tested" unemployment benefit was one year. The JSA imposes no time limit: as long as the conditions are met, an applicant is entitled to it.

11

Welfare-to-Work: Which Policies Work and Why?

11.1 Introduction

This lecture considers the arguments behind the expansion in welfare-to-work programs that occurred over the last decade and reviews the effectiveness of alternative approaches to enhancing labor market attachment and earnings among the low skilled. It concerns the 'iron triangle' of welfare reform – that is the three, often conflicting, goals: raising the living standards of those on low incomes; encouraging work and economic self-sufficiency; and keeping government costs low. Many different policies can be cast in terms of these broad aims, albeit with different weights attached to each of the goals. In the UK there are active labor market programs like the New Deal and there are also financial incentive programs like the Working Families Tax Credit. Although the latter are often classed as welfare policies and the former as active labor market policies, both are motivated by similar concerns over low incomes and low labor market attachment and share many similar design features. The key organizing idea in this lecture is to provide an integrated view of the way the wide array of 'welfare to work' and 'make work pay' policies affect the earnings, incomes and incentives of working age individuals and their families. The aim is to assess their effectiveness in addressing low income, low earnings and low labor market attachment in the working age population.

The original version of this chapter was published as: Blundell, R. (2002). Welfare-to-Work: Which Policies Work and Why? Keynes Lecture in Economics: 2001. © 2002 The British Academy. I would like to thank Mike Brewer, Tom Clark, Monica Costa Dias, Alan Duncan, Hilary Hoynes, Costas Meghir, Michal Myck, Howard Reed, Barbara Sianesi, John Van Reenen and Ian Walker for their help and for extensive discussions. I would also like to thank the editor of these proceedings and my two discussants, John Ermisch and John Flemming for their helpful comments. This research is part of the program of research at the ESRC Centre for the Microeconomic Analysis of Public Policy at IFS.

Other countries, most notably the US and Canada, have implemented a similar array of policies and I will draw on the extensive evaluations of these in the discussion that follows. However, the UK over the last decade is, in many ways, an ideal test bed in which to examine such policy reforms since both the WFTC and the New Deal[1] were introduced and enhanced over this period. These policies are targeted at two groups: (1) low-income/low-educated families with young children, (2) low-skilled workers with long or repeat unemployment spells. In both cases the diagnosis was similar: relatively low hourly wages with little labor market experience implying little incentive for work.[2] However, the detail is different. In the first case it is the generosity of the out-of-work benefit system for families relative to potential earnings and childcare costs that are thought to provide the disincentive. For the second group it is employer matching and the low initial wages that are perceived as the central issue. Consequently, although the prescription for both is to enhance net earnings in work, the first involves a long-term income related supplement to earnings, possibly with a childcare component. While the second centers on job search assistance and short-term employer based employment subsidies. But to what extent are these differences in the design of welfare-to-work programs appropriate and could they be improved?

The 'in-work' structure of these two approaches is similar, relying on earnings credits or employment/wage subsidies. But again, they typically work rather differently. The wage subsidy is individually based, not means-tested and has limited duration. Eligibility is also typically dependent on a certain duration of unemployment insurance (or welfare) receipt. The tax credit, on the other hand, is typically subject to a family income based means-test and does not have a time limit. For the latter, the WFTC in the UK, the EITC in the US[3] and the In-Work Tax Credit in Belgium[4] are prime examples.

For the former, the New Deal in the UK and Work First[5] in the US are leading examples. There are, of course, many welfare-to-work policies that fall somewhere in between. For example, the Self-Sufficiency Project (SSP)[6] in Canada, although an in-work tax credit like the WFTC or EITC, has a three year time-limit and eligibility depends not only on overall family income and family composition but also on a minimum welfare duration and a minimum hours requirement. The New Hope[7] tax credit program in the US also has a three-year time limit and a minimum hours condition. Both programs provide job search assistance at least for some of the program participants.[8] The Minnesota Family Investment Program (MFIP)[9] is similar to the SSP, however, the job search assistance is mandatory as in the New Deal for Young People in the UK. An additional feature of these Canadian and US programs is that many were the subject of randomized experimental evaluation, the results of which provide a vital source of information in the discussion below. Finally, the earnings supplement and job search provisions within the many state run additions to the Temporary Assistance for Needy Families (TANF) program in the US have similar characteristics to the New Deal program (see Card and Blank 2000a).

So what is the best design for such policies? How should they differ with demographic characteristics? Does time limiting the in-work financial incentives help with human capital and wage progression? If so, how long should the time limit be set? Should there be a duration of welfare or a duration of UI recipiency requirement for eligibility. If so, for how long? Should there be mandatory job search assistance and/or accredited training? If so, how should sanctions apply? Should family income means testing be used to target incentives to those on low incomes? If so, at what level should the credit withdrawal or phase-out rate be set? Should the wage subsidy or tax credit be tied to a specific employer? Should there be a minimum hours requirement? Should childcare costs be incorporated in the financial incentive?

The recent proposal by the UK government[10] to separate the child component of WFTC from the adult component so as to form an integrated child credit (ICC) and an Employment Tax Credit (ETC), provides a further motivation for investigating the overall design features of in-work benefits and other make-work-pay policies. This is especially the case once it is recognized that the new ETC will be open to all adults irrespective of whether they have children.

There is also a growing theoretical literature examining the role of work requirements in the design of optimal income transfer programs. In a dynamic model the important issue relates to incentives for poverty reducing investments and investments in human capital. Besley and Coate (1992) derive conditions under which workfare can be optimal. Cossa, Heckman and Lochner (1999) develop a dynamic model with time limits and human capital investment. In a more static setting the recent contribution by Saez (2000) shows that, where labor supply responses are concentrated along the extensive margin (participation in work), an Earned Income Tax Credit system with transfers that increase with earnings at low levels can be optimal and justifies the move away from negative income tax schemes.

Examining the impact of such reforms on employment and on poverty requires a careful analysis. In any program of this type there will be those that are induced into employment by the program and those who benefit financially from it but are already in employment or who would have moved into employment anyway.[11] The distinction between these groups is key and we will draw on experimental and non-experimental evidence to gauge their likely size. Similarly any reliable evaluation requires a control group for comparison. This is in turn made more difficult when there are spillover effects (through displacement or more general equilibrium effects) on to groups that are not directly eligible. Again where possible we will pay attention to the importance of these effects, most notably in the analysis of the mandatory job search and wage subsidy elements of the New Deal policy.

To set the scene for our analysis we turn, in the next section, to the labor market trends over the last two decades that motivated the UK reforms. The cyclical volatility of employment for certain target groups and the secular changes in employment patterns for others is highlighted. Section 11.3 then considers a number of central design features, focussing on time limits, means testing and implicit tax rates, minimum hours requirements, welfare receipt eligibility, and wage progression. This is done in the context of the design of the New Deal and of the WFTC. In Section 11.4 we move on to evaluate specific aspects of the New Deal and WFTC reforms. We conclude, in Section 11.5, with an overview of these schemes and their effectiveness, and an assessment of the appropriate design of welfare-to-work and make-work-pay programs.

11.2 The Changing Structure and Economic Environment of Low-Wage Workers in the UK

This section considers the labor market trends that stimulated the New Deal and WFTC reforms in the UK. Turning first to the labor market for the young unskilled that motivated the New Deal (for Young People) NDYP program, we highlight the cyclical volatility of unemployment for this group and the frequency of short run transitions. We then move on to the corresponding employment trends for low-income families with children, which motivated the WFTC reform. Here non-employment rather than claimant unemployment or active job search is a more relevant measure of activity and we highlight the importance of both cyclical and secular trends.

11.2.1 *The Labor Market Background for the New Deal Reform*

The New Deal for Young People was directed at 18–24 year olds with more than 6 months unemployment. Across all countries youth unemployment is higher than unemployment for prime age individuals. In the late 1990s there was a relatively high proportion of young Britons in jobs and a low proportion of young people in full-time education. There was a large proportion of British youth that are neither in school nor in the labor force. Moreover, in the 1990s the UK had the highest numbers of 18-year-old men in this category and was second (after Italy) for 22-year-old men.[12] It also had the largest increase in the proportion of this group of youth.

Another feature of the youth labor market is its sensitivity to the business cycle.[13] The level of unemployment of the younger group, displayed in Figure 11.2, broadly mirrors the overall picture in Figure 11.1, but the cycle is, if anything, more pronounced. This is also true for employment rates as can be seen from Figure 11.3 (see also Bell, Blundell and Van Reenen 1999). The extent of cyclicality, and the differences across the cycle in unemployment and employment rates by age, is particularly important for the evaluation of the impact of welfare-to-work programs like the New Deal. For example, if a group of similar but older men were to be used as a comparison group for those entering the New Deal then adjustment for cyclical differences across the groups would be crucial. Otherwise the impact of cyclical differentials would be incorrectly attributed to the New Deal reform. This is highlighted in our examination of the impact of the New Deal on employment in Section 11.4 below.

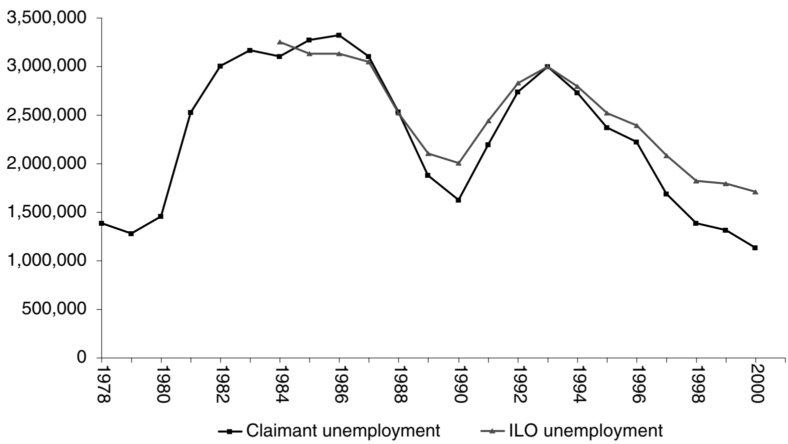


Figure 11.1. Unemployment – Claimant and ILO Measures

Source: Labor Market Trends and Employment Gazette, various issues.

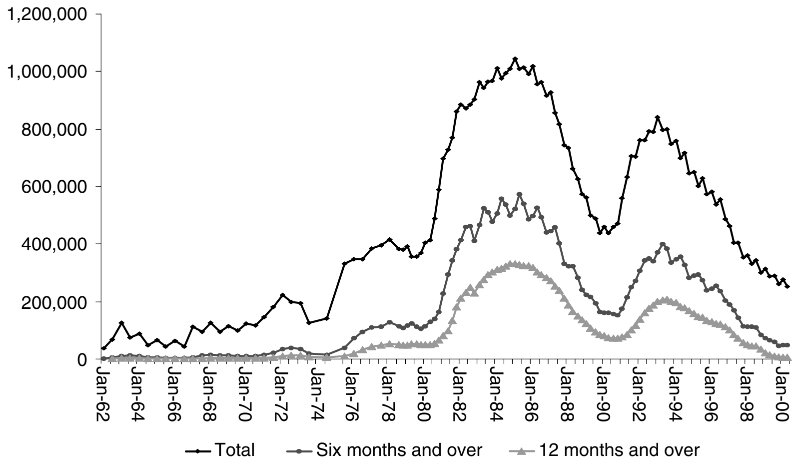


Figure 11.2. Claimant Unemployment Amongst 18–24 Year Olds
Source: Labor Market Trends and Employment Gazette, various issues.

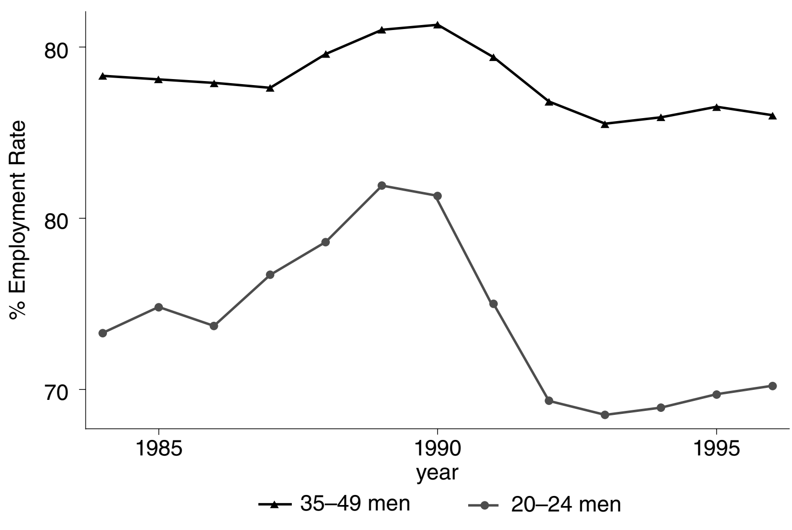


Figure 11.3. The Impact of the Cycle on Employment Rates by Age
Source: LFS.

11.2.2 *The Labor Market Background for the WFTC Reform*

The high levels of non-employment, experienced by certain specific demographic groups, were also the motivation for earned income tax credit reforms – or in-work benefit reforms. For example, one central stimulus for the introduction and subsequent expansion of the Working Families Tax Credit in the UK was the persistence in the low levels of attachment to the labor market by single mothers – at a time when for other groups of similar women attachment had generally been increasing. Figure 11.4 shows the secular change in female employment across four household types in the UK. The growth in the attachment by women in couples with children is as noticeable as is the fall for single women with children.[14] This is even more pronounced for those who left school at age 16 or before (age 16 being the minimum school leaving age for those born after 1960). Not only has attachment of lone mothers fallen but, at the same time, the size of this group has risen by more than twofold over the last twenty years. Blundell and Hoynes (2004) document this change and examine the similarities between demographic trends for single mothers in the UK and US.

Another distinguishing feature of the UK has been the growth in workless couples with children. This is documented in Figure 11.5 and provided a strong argument in the debate over the WFTC reform (see Gregg, Hansen and Wadsworth 1999). Indeed, for women in couples with unemployed partners employment rates have stayed no higher

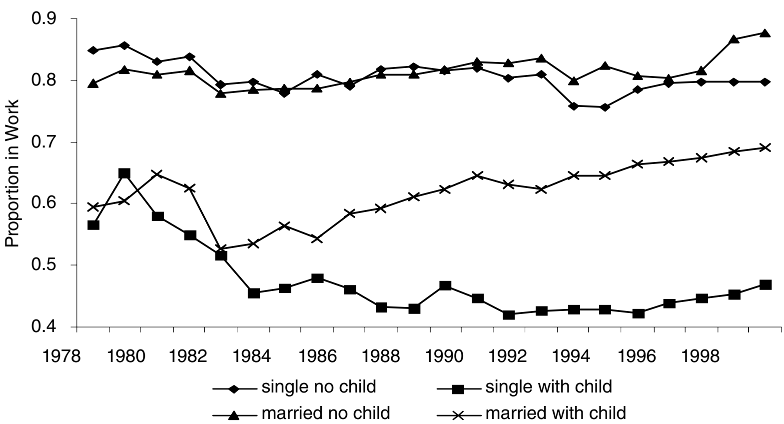


Figure 11.4. Employment Trends for Women in the UK: Proportion in Work
Notes: FES Data, working age.

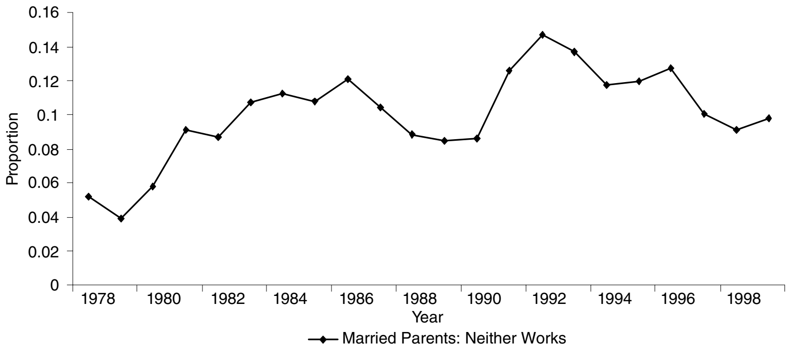


Figure 11.5. Proportion of Workless Couples in the UK

Notes: FES Data. Working age head.

than 30% over the past two decades – even lower than employment rates for the single parent group (see Blundell 2001). The (non-) employment rates for these two groups show clearly why they have been singled out as two target groups for tax and benefit reform.

11.2.3 *Inequality and the Real Wages of the Young Low Skilled*

It is not just the low employment rates among the low skilled that have attracted attention. So have the low real wages and the relatively slow growth in these wages over the past two decades.[15] Indeed, there have been well documented and remarkable shifts in returns to education and skill in many countries (see Gosling, Machin and Meghir (2000) for the UK and Katz and Autor (1999) for a survey of international evidence). For example, in the US real earnings for the lowest education groups fell yearly from the late 1970s to the mid 1990s. This characteristic was quite exaggerated in the US, but the overall pattern was common to most developed countries. Indeed, there is evidence that lower educated younger workers have seen a stronger decline in their wages relative to those with more education over the last two decades,[16] reducing further the incentives to take paid employment.

11.3 The New Deal and WFTC Reforms in Context

The simple but stark facts about the low-skill labor market, reviewed in the last section, focussed policy attention in the UK on ‘make work pay' policy reforms for the low skilled, the aim being to make work more attractive for those whose current labor market opportunities are not sufficient to induce work. As mentioned in the introduction the key organising idea of this lecture is to place the various 'welfare-to-work' or 'make work pay' reforms alongside each other, to focus on specific design features and to examine the importance of each of these features in addressing the objectives: raising the living standards of those on low incomes; and encouraging work and economic self-sufficiency. Before considering design issues, we first consider the specific characteristics of the New Deal and the WFTC policies in the UK. In our general discussion that follows we will then compare these features with those of similar reforms in North America, Canada and elsewhere.

11.3.1 *The Design of the New Deal*

The New Deal for Young People in the UK, which was launched in early 1998, is targeted at the 18 to 24 year olds with at least six months unemployment. Participation is compulsory, so that every eligible individual who refuses to participate risks losing their entitlement to benefits. The criteria for eligibility are simple: every individual aged between 18 and 24 by the time of completion of the sixth month on Jobseeker's Allowance (JSA) – the standard flat rate Unemployment Insurance in the UK – is assigned to the program and starts receiving treatment. Given the stated rules, the program can be classified as one of "global implementation", being administered to everyone in the UK meeting the eligibility criteria. Indirect effects that spill over to other groups than the treatment group may occur. The nature of these effects will be discussed below.

The path of a participant through the New Deal is composed of three main steps (see Figure 11.6). On assignment to the program, the individual starts the first stage of the treatment called the *Gateway*. It lasts for up to 4 months and is composed of intensive job search assistance and small basic skills' courses. Each individual is assigned a "Personal advisor", a mentor who they meet at least once every two weeks to encourage/enforce job search. The aim is to place individuals in unsubsidized employment (although there are a proportion who exited into subsidized jobs before exhausting the gateway period).

The second stage is composed of four possible options. First, there is the employer option – a six-month spell on a subsidized employment. For the subsidized employment option, the employer receives

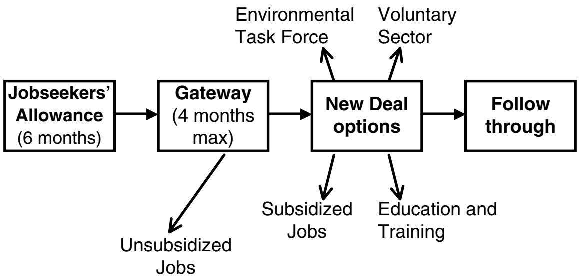


Figure 11.6. A Simplified Flow Diagram of the New Deal Program

a £60 a week wage subsidy during the first six months of employment plus an additional £750 payment for a required minimum amount of job training equivalent to one day a week.[17] Second, an individual can enroll in a stipulated full-time education or training course and receive an equivalent amount to the JSA payment for up to twelve months (and may be eligible for special grants in order to cover exceptional expenses). Third, individuals can work in the voluntary sector for up to six months (paid a wage or allowance at least equal to JSA plus £400 spread over the six months). Finally, they may take a job on the Environmental Task Force (essentially government jobs) and be paid a wage or allowance at least equal to JSA plus £400 (spread over the six months).[18]

The program was launched in the whole UK in April 1998. There was, however, a previous Pilot three months' period, from January to March 1998, when the program was implemented in 12 areas, called the *Pathfinder Pilots* (see Anderson, Riley and Young 1999). Clearly, identification of the treatment effect under these conditions requires stronger assumptions than when an experiment is run within regions using random assignment. The problem relates to the fact that the counterfactual must either be drawn from a different labor market or from a group with different characteristics operating in the same labor market. However, we are able to use the features of the pathfinder pilots in comparison to non-pathfinder areas to examine the impact of the policy and the potential issues concerning substitution effects and general equilibrium effects. These evaluation issues are discussed in detail in Blundell et al. (2004), in Section 11.4 below we simply summarize the results of that evaluation study and draw conclusions for the appropriateness of its design.

11.3.2 *The Design of the WFTC*

In-work benefits have existed in the UK in various forms since the 1970s. However, the current Working Families Tax Credit has its antecedents in the Family Credit system introduced in the late 1980s.[19] The Family Credit was designed to provide support for low-wage working families. In this system each eligible family was paid a credit up to a maximum amount which depended on the number of children. There was also a small addition if in full-time work. Eligibility depended on family net income being lower than some threshold (£79 per week in 1998–99). As incomes rose the credit was withdrawn at a rate of 70%. In 1996 average payments were around £57 a week and take-up rates stood at 69% of eligible individuals and 82% of the potential expenditure.

A striking feature of the Family Credit system, retained in the WFTC reform, is a minimum weekly hours eligibility condition. A family with children required one adult working 16 hours or more per week to qualify. At its introduction in 1988 this minimum hours cut-off was set at 24 hours but then reduced in 1992 to encourage part-time work by lone parents with young children (see Blundell et al. 2000a).

The WFTC reform increased the generosity of in-work support relative to the FC system in four ways: It increased the credit for younger children. It increased the threshold. It reduced the benefit-reduction rate from 70% to 55%. Finally, the reform incorporated a childcare credit. This was worth 70% of actual childcare costs up to £150 per week (for two children, £100 for one child). The largest cash gains went to those people who were currently just at the end of the benefit-reduction taper.

The credit was available to lone parents and couples where both partners work more than 16 hours per week. The transfers (excluding childcare credit) underlying the WFTC are illustrated in Figure 11.7.

Despite the dampening effect of these interactions with other benefits, there does seem to be some prima facie evidence of an impact on behavior. A look at the histogram of weekly hours worked for single parents presented in Figure 11.10a, for example, shows a strong peak in hours worked at 16 hours. This is not evident for ineligible groups such as single childless low-educated working women as reported in Figure 11.10b. Of course, there will be a large number of so called ‘windfall beneficiaries’ and there may also be those who decide to reduce their working hours in response to the incentive at 16 hours. These issues will be considered in the evaluation of the impact of WFTC reform on hours and employment in Section 11.4 below.

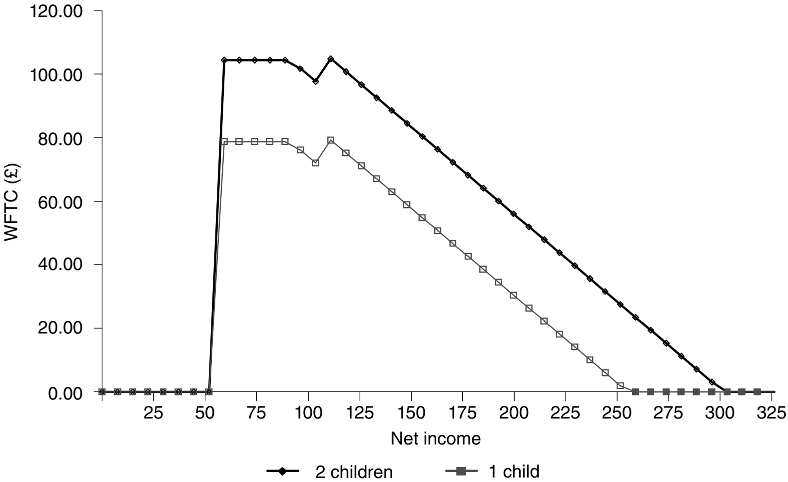


Figure 11.7. WFTC Weekly Award, June 2000

Source: Brewer (2000).

Figure 11.8. Single Mother before WFTC

Notes: Single parent, April 1997, earning £3.50 per hour (2000 prices).

It is worth noting at this stage that many of these design features are absent in other Employment/Earnings Tax Credit systems. The EITC in the US, for example, has no minimum working hours condition and the level of the credit is not counted as income in the computation of other taxes and benefits.[20] There is also a small EITC available to

Figure 11.9. Single Mother after WFTC

Notes: Single parent, April 2000, earning £3.50 per hour (2000 prices).

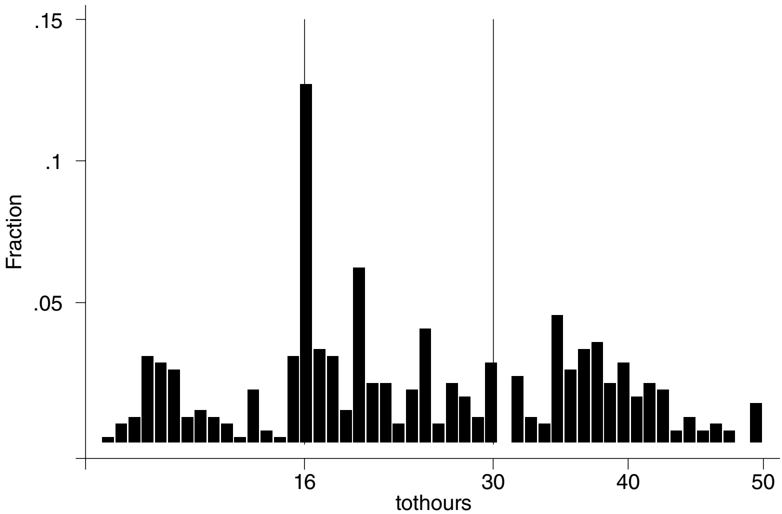
low-earning workers without children in the US. In the Canadian SSP there is a 30 weekly working hour condition (averaged over a month) but receipt of the credit is time limited to three years and eligibility requires a 12 month welfare duration, not simply a low family income as in the EITC and WFTC.

11.3.3 *Aspects of Design*

The discussion so far of the New Deal and the WFTC programs in the UK has highlighted certain central features in the design of these make-work-pay programs. Here we gather them together under the following seven headings:

11.3.3.1 TIME LIMITS AND WAGE PROGRESSION

There are a number of ways in which time limits have been incorporated in welfare-to-work programs. The US debate has focussed mainly on the time limits in the Temporary Assistance for Needy Families (TANF) program (see Moffitt and Pavetti 2000). In this program of income support the individual state can set a lifetime limit for receipt. Typically set at 60 months (the maximum allowed) and introduced in 1996, these time limits are just beginning to bind. Perhaps not surprisingly many individuals have left welfare before the limit and there is consequently some evidence that the limits themselves have helped in



a) Low Education Single Parents in the UK

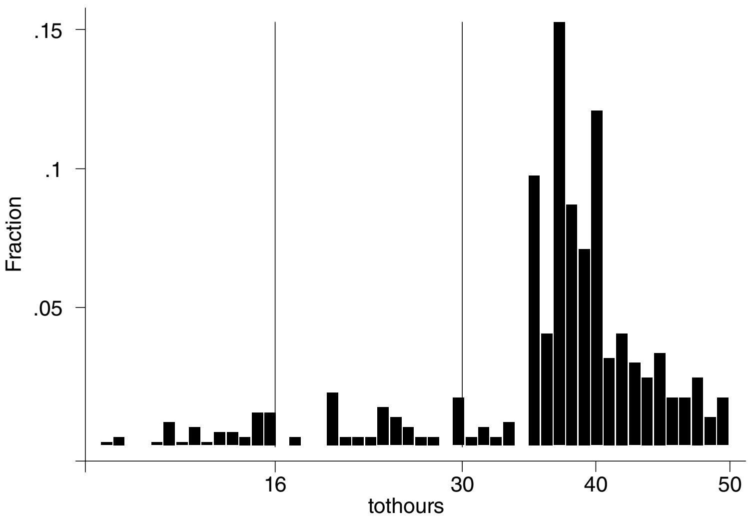


Figure 11.10. Weekly Hours Worked. b) Low Education Single Women Without Children in the UK

Notes: Family Resources Survey, 1998/99; Blundell and Hoynes (2004).

the dramatic reduction in welfare rolls in the US (Grogger 2000). Part of the success of the New Deal in the UK documented below is the effective time limit it places on receipt of JSA, although it is difficult, in the UK context, to separate the effect of this from the mandatory job search assistance and benefit sanctions that are included as part of the program.

Time limits can equally well be imposed on the receipt of financial incentives in work. This is not a feature of the WFTC or EITC. But it is part of the Canadian SSP tax credit system and it does feature in the earnings disregard programs that form part of the individual state specific features of the TANF program. These vary from 6 months for the New Deal and Work First and JOBS Plus[21] programs to three years in the case of SSP and many of the TANF based programs in the US.[22] The appropriate design of such time limits depends on the expected level of wage progression for program participants and the incentives for wage progression created by the time-limited system itself.

Incentives for wage progression are often enhanced by the provision of training – a central part of the New Deal program. With no time limit, tax credit systems can provide a strong negative incentive for wage progression and human capital investment, reducing the chance of longer run self-sufficiency. This will depend largely on the relative importance of the passive return to work experience, which occurs automatically once in work, in comparison with the return to 'active' human capital investment, which requires effort or time inputs by the individual. Cossa, Heckman and Lochner (1999) make this point forcibly. However, evidence of steep wage progression among low-skilled workers is rare. Most studies suggest that wage progression will be slow, no more than 3–4% per year, see Gladden and Taber (2000). This is further supported by the recent work by Card, Michalopoulos and Robins (2001) on the wage growth among the recipients of the Canadian SSP experiment. Consequently, a six-month time limit is unlikely to provide time for wage progression to result in self-sufficiency and could be counter productive. At the end of the subsidy either workers will move to lower wages, lose their employment or move into some other make-work-pay program. For example, the EITC in the US is used by many as a way of working themselves off time limited earnings supplements in TANF. But then in the EITC the incentives for active wage progression and human capital investment, once in work, are slight.

11.3.3.2 MEANS TESTING AND IMPLICIT TAX RATES

A key ingredient in understanding the structure of financial incentives underlying make-work-pay policies is their interaction with the tax and benefit system. Nowhere is this more pronounced than in the comparison between the EITC in the US and the WFTC in the UK. As we have seen above, in the UK the level of WFTC credit counts as income in means-tested benefit programs like Housing Benefit. This is deliberate and was part of the Family Credit reform in 1988. It ensures there are no implicit tax rates on earnings that exceed 100%. But implicit tax rates can be high, as is evident from Figures 11.8 and 11.9.

In contrast, the EITC shown in Figure 11.11, although providing a less generous credit, sits on top of the tax and benefit system. A consequence of this is that the lower withdrawal rates (phase-out rates) in EITC must be added to the implicit tax rates in TANF, Food Stamps and the income tax system. A typical budget constraint for a US EITC recipient is drawn in Figure 11.12, which should be contrasted with the similar UK system in Figure 11.13.

For couples, a further issue is whether the tax credit should be subject to an individual or a family means test. As is argued below, a family based system creates adverse incentives for labor supply of 'secondary' workers in the household. However, such a system is well targeted to family poverty and to the reduction of workless households. In contrast, individual tax credits can better target low-wage workers and low skills. A family means test means that work incentives can be improved for one and worsened for another partner, and can alter (usually worsen) the incentives to form a couple/marry. Both EITC and WFTC use a family income means test.

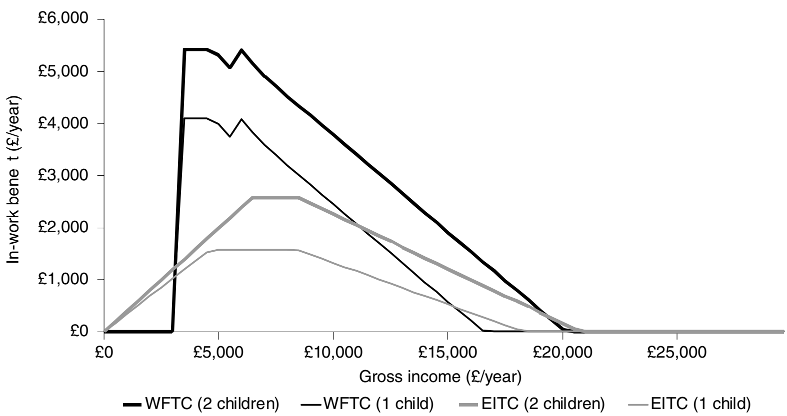


Figure 11.11. EITC Schedule and WFTC Weekly Award, 2000

Notes: £1 = $1.50. Assumes 2000 tax system in US, and 2000 tax system in UK.
Source: Brewer (2000).

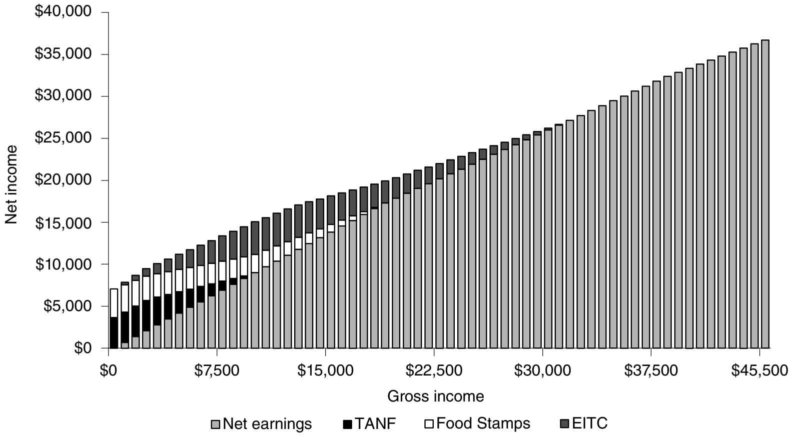


Figure 11.12. Gross and Net Incomes, Lone Parent with 2 Children, US

Notes: Assumes 2000 federal tax system and Florida's TANF system. Ignores housing and childcare costs and subsidies. Assumes all TANF and Food Stamps requirements are met, and that all income is earned.

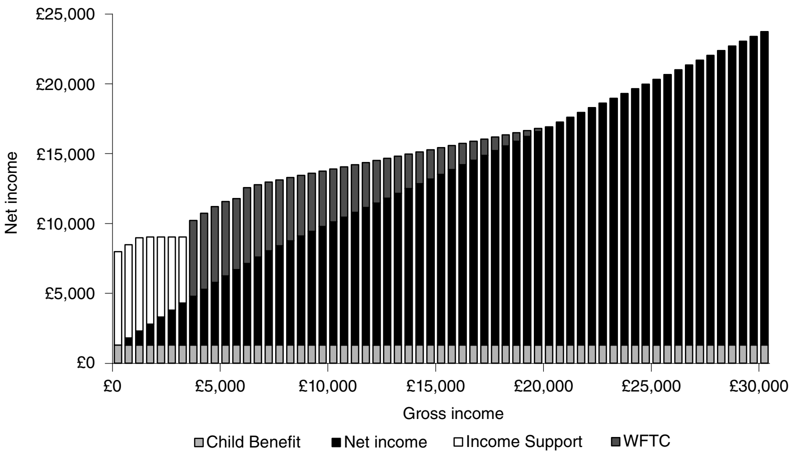


Figure 11.13. Gross and Net Incomes, Lone Parent with 2 Children, UK

Notes: Assumes 2000 tax and benefit system plus a Children's Tax Credit. Ignores housing and childcare costs and subsidies. Assumes 2 WFTC awards/year and minimum wage work, so that eligibility for WFTC occurs at £3,078 and 30 hour premium at £5,772.

Source: Brewer (2000).

11.3.3.3 SETTING THE LEVEL OF CREDIT OR SUBSIDY

The appropriate level of the credit or subsidy is intricately tied to whether it is to be means tested and whether it is time limited. The typical wage or employment subsidy, as in the New Deal, is a fixed weekly sum, time limited and independent of family income and composition. In contrast the credit in WFTC is means tested, varies with family composition and has no time limit. In some sense this reflects no more than the desire to achieve distributional objectives with the WFTC, in particular the desire to reduce the level of child poverty. Nonetheless the proposed separation of the child component in the WFTC in to an integrated child credit (ICC) (see Brewer, Clark and Myck 2001), leaves an adult Employment Tax Credit (ETC) that is available to those without children and whose level is much less about child poverty.

A higher level of credit implies a higher withdrawal rate, unless the credit is to extend high into the earnings and income distribution. Indeed the increased generosity underlying the WFTC reform together with the reduction in the withdrawal rate, extended eligibility and the phase-out region much higher into the income distribution than had previously been the case. Increasing the cost of the program and the number of recipients with relatively high incomes. The price for extending generosity at lower earnings, without increasing withdrawal rates, is a higher implicit tax rate further up the income distribution.[23]

11.3.3.4 MINIMUM HOURS CONDITIONS

Minimum hours conditions can reduce costs and remove the incentive to reduce hours to very low levels. However, if they are set too high they reduce the attractiveness of the program to individuals out of employment, especially those that have young children. The reduction in the hours condition, from 24 at the introduction of Family Credit to 16 in WFTC, can be seen to have encouraged a significant fraction of inactive single parents into work (see Dilnot and Duncan (1992) and Blundell and Hoynes (2004), for a discussion). It also reduced the number of hours worked by many single parents in employment. It should be noted that in 1995 a 30 hour 'full time' bonus of £10 was introduced.[24] In comparison, EITC has no minimum hours condition whereas SSP in Canada and New Hope in the US have a 30 hours condition. Some have argued for hourly wage based credits to address the adverse hours and effort incentives.[25]

Help with childcare costs can overturn some of these arguments. Indeed, the WFTC has a generous childcare credit. Also note that the proposed Employment Tax Credit in the UK is set to have a 30 hours condition for adults without children. It may well be true that wage progression in part-time low-skilled jobs is quite slow.[26]

11.3.3.5 TRAINING REQUIREMENTS AND HUMAN CAPITAL INCENTIVES

Many of the issues concerning individual incentives for human capital investment and wage progression have been made in the discussion of time limits and wage progression above. However, there are remaining issues concerning training provision. There is also strong evidence that workplace based training that leads to a vocational qualification is the most effective, at least for the lower skilled with relatively low formal education levels (see Blundell, Dearden and Meghir 1996 and references therein).

Is it possible to design an effective training incentive within an individually based tax credit system? Presumably, provided training is monitored and leads accredited qualifications, an individual incentive scheme can be as effective as one operated through the employer. It may have the added attraction allowing, or even enhancing, mobility. There is also no reason why accredited training should not be a condition of continuing receipt of an Employment Tax Credit or wage subsidy.

11.3.3.6 WELFARE, UI DURATION REQUIREMENTS AND PROGRAM TAKE-UP

Welfare receipt conditions are chosen so as to reduce costs and target the workless. SSP in Canada requires a 12-month duration on welfare for eligibility. But like WFTC and EITC, the New Hope program simply uses low income as an eligibility condition. The New Deal has a 6-month unemployment claimant condition. There are many other examples in other similar programs.

There are a number of counter arguments to such targeting. The first is the stigma impact perceived by both employer and employee. This is often cited as the reason for the low take up, especially among employer based subsidy schemes.[27] The second is the churning or cycling effect. Since eligibility depends on welfare receipt individuals have an incentive to churn or cycle through the system and the long run impact of such programs on employment will be mitigated (see Martin and Grubb 2001 and Meyer 1995, for example). Finally there is an entry effect where by those with short durations on welfare extend their spell to become eligible for the financial incentive.

It is clear that all these issues play a role. Indeed there is recent evidence, from the Swedish welfare-to-work programs,[28] that it may be important to act as soon as workers enter unemployment, or the welfare system, and not to wait. However, the argument in terms of reducing the number of so called 'windfall beneficiaries' often wins the day, see the discussion in Card and Blank (2000a). There is an important balance to be made and it may well be the case that a welfare recipient condition, as in the Canadian SSP, together with a relatively long time limit for receipt of the credit is an optimal schedule for helping those on welfare into work, supporting their income and leaving some incentive for wage progression and human capital investment. Once again though, if the only way to obtain the financial incentive is to have a period on welfare, there is an important issue of how to guard against inducing long welfare spells and cycling.

11.3.3.7 ACTIVE PROVISION OF JOB SEARCH ASSISTANCE

One important difference between various make-work-pay and welfare-to-work programs is whether they provide job search assistance. In some ways the EITC and WFTC programs, by focussing on workers, do not directly face this issue. But in so far as they are designed to enhance labor market attachment, job search assistance for new entrants and those likely to enter the program would seem quite plausible.[29] The New Deal for Lone Parents in the UK act in this way as once in work the Lone Parents become eligible to WFTC. On the job help in improving matching of workers could also be an important way of enhancing earnings through job mobility for such workers.

What kind of job search help should be given and should it be mandatory? The New Deal for Young People described above is mandatory and provides the participant with a personal advisor, with meetings at least once every two weeks to encourage/enforce job search. Missing a meeting can incur sanctions. This may be the effective part of the Gateway and builds on the apparent success of the Restart interviews.[30] However, it may well be the possibility of financial sanctions that had most impact.[31] Certainly the additional impact of voluntary job search advisors in the SSP randomised experiment had a relatively small impact over the financial incentives alone, on longer-term full-time jobs.[32] Mandatory job search assistance in the MFIP had a bigger effect.[33]

Before further discussion of what components of a welfare-to-work system are likely to work and for whom they work best, we turn our attention to the evaluation of the two UK programs on which we have focused: the New Deal and the WFTC.

11.4 Evaluating the Labor Market Impact of Reforms

The first concern of any evaluation is whether the appropriate statistical approach has been taken. There is a growing use of experimental evaluations and demonstration projects, especially in North America.[34] These clearly have some advantage over observational studies and they can provide important evidence to benchmark the discussion that follows.

However, experiments do not address all concerns and they do not adapt well to extrapolation and to the study of variations in policy design. Area based studies can also be attractive. As mentioned above the piloting of the New Deal in pathfinder areas provides some useful information on certain spillover effects.[35]

In some cases so called ‘natural experiments’ are useful. These occur when a control group appears naturally in the data rather than through a randomized experiment. For example, there may be a very similar group to the target group that is ineligible to the program. Provided they have the same macroeconomic trends and there are no systematic composition changes before and after the program,[36] a simple difference-in-differences methodology can provide a useful guide to the extent of a policy impact. Of course this is an ex-post evaluation.

Ex-ante evaluations either arise through an experimental demonstration project or through a structural econometric model in which the proposed reform can be simulated.[37] It is also sometimes possible to use matching on observables to mimic the controlled experiment. Where rich administrative data sources are available for evaluation, this is a particularly attractive approach.[38]

We will make use of all these alternative methods in what follows. We turn first to specific aspects of the New Deal and WFTC policies.

11.4.1 *The Impact of the New Deal for Young People*

Although there is now some evidence of the impact of employment of individuals completing the options in the New Deal, we focus here on an evaluation of the Gateway.[39] In particular, we are concerned with the degree to which enhanced mandatory job assistance has lead to more outflows to (unsubsidized) employment. The evaluation is based on data provided by the Pathfinder areas before the National Roll Out of the program, as well as on data available following the National Roll Out.

As mentioned above, there are two main issues that need to be considered in evaluating the impact of the program: the precise nature of the comparison group, and hence the definition of what is being measured, and the set of assumptions that underlie the interpretation of the parameter we estimate in each case. The clear understanding of these issues is an important input in an eventual cost-benefit analysis of the program since they determine the outcome from the program.[40] There are some important aspects covered within this discussion. One of them concerns the extent to which we can estimate the overall impact of the program on employment as opposed to the impact on the eligible individuals. Potential differences in the two outcomes may result from two main factors. First, the impact of the program on eligible individuals may be at the expense of worsened labor market opportunities for similar but ineligible individuals. Second, the wider implementation of the program and the opportunities it offers to participants may affect the equilibrium level of wages and employment, affecting all workers. We focus on the impact of the program on the proportion leaving unemployment within four months of entering the Gateway. We pay special attention to the *outflows into employment*, but we also examine total outflows from unemployment to all destinations.[41]

Our approach to estimating the impact of the New Deal program relies on using information from the pilot period as well as information from the *National Roll Out*. The New Deal can affect employment of both eligible and ineligible individuals in a number of ways. First the eligible individuals receive job search assistance, which may enhance their ability to find a job. Second, some of the individuals in the Gateway program receive wage subsidies, reducing the cost of employing them for an initial period of six months. This wage subsidy will expand the employment of such workers but may also lead to a substitution of other workers for these cheaper ones. The extent to which this may happen will depend on a number of factors. If the subsidy just covers the deficit in productivity and the reservation wage of the workers as well as the costs of training, we would not expect any substitution; these workers are no cheaper than anyone else. Second, it will depend on the extent that these workers are substitutable in production for existing workers and on the extent that it is easy to churn workers. The latter is an important point, since the subsidy only lasts six months. Moreover, the agencies implementing the New Deal are supposed to be monitoring the behavior of firms using wage subsidies and employing individuals on the New Deal. Of course, if job durations are generally short, firms will be able to use subsidized workers instead of the non-subsidized ones, without any extra effort.

An additional effect of the New Deal may be to decrease wage pressure through the increase in labor supply and through the presence of wage subsidies.[42] This will tend to increase employment for all types of workers and will counteract the effects of substitution on the non-treatment group.

Assessing the importance of substitution and of general equilibrium effects through wages or other channels is of central importance. Using the comparison between the pilot and control areas as described below, and assuming these areas are sufficiently separate labor markets from each other, we will be able to assess the extent to which substitution and other general equilibrium effects combined are likely to be important "side-effects" of the program, at least in the short run.

The available options for the choice of the comparison group depend on the type of evaluation being performed. When assessing the program from data on its National Roll Out, we are constrained to use ineligible individuals within the same area, for which we have chosen the age rule to define (in)eligibility. The Pilot Study, however, provides an additional instrument in the definition of the comparison group. We have used it in two ways, constructing two possible comparison groups: The first takes all eligible individuals living in all non-Pathfinder areas; the second matches eligible individuals in the set of non-Pathfinder areas that most closely resemble the Pathfinder areas. The goal of a careful choice of the comparison group is to satisfy a central assumption in non-experimental evaluation, which requires that the time trend evolve in the same way for treatments and controls.

The aim of matching the areas is to achieve a match as close as possible with respect to labor market characteristics. The procedure followed to match on labor market characteristics makes use of a quarterly time-series of the outcome variable from 1982 to just before the introduction of the New Deal, in January 1998. A measure of distance was then computed for each possible pair of Pathfinder and non-Pathfinder areas and the two nearest neighbors were chosen. Once the two nearest neighboring areas have been chosen based on similarity of the labor market trends, we carry out our estimation (see Blundell et al. 2004 for details of these procedures).

11.4.1.1 THE RESULTS FROM THE NEW DEAL PILOT AREAS

Table 11.1 presents the main estimate of the impact of the Gateway on eligible men living in Pathfinder areas during the Pilot period. Precisely the effects of the Gateway on *outflows to employment* after 4 months of treatment.[43] The estimate compares men aged 19 to 24 years old living in Pathfinder areas with a similar 19–24 year old age group living in all non-Pathfinder areas. After 4 months of treatment, it is estimated that the Gateway has improved participants' exits into employment very significantly – pointing to an impact of about 10–11 percentage points. In the pre-program period only 24 percent of individuals in the treatment group obtained employment over the similar four months period (compared to 33 percent afterwards). Thus, the improved job search assistance provided during the Gateway seems to have raised the probability of getting a job by about 42% (=10%/24%) after 4 months of treatment.

This result should be contrasted with the information from the New Deal Evaluation Database concerning outflows into the employment option. It is estimated that the outflows into an employment option after 4 months of treatment sum up to 5.7 percent of men joining the Gateway. Subtracting this off the overall New Deal effect would give a "pure" Gateway impact (on outflows to unsubsidized employment) of about 4 percent. But this is likely to be a lower bound. The calculation assumes that there is essentially no deadweight of the employer subsidy. This happens under the assumption that participants can be split into groups according to their ability to find a job, and that subsidized jobs are being attributed to those in need of a subsidy to leave unemployment. If, on the other extreme, it is believed that the subsidized jobs are being allocated to the most employable participants, then the amount of scaling down required might be small. Furthermore, the NDED will tend to find larger job outflows because of fewer missing values. Thus 4 percent is a lower bound for the pure Gateway/job assistance effect. The method used to estimate the impact of treatment does not seem to substantially influence the results, reflecting some robustness of the estimates to the functional form assumptions.[44]

Table 11.1. New Deal Gateway Employment Effects for Men

Experiment	Treatment group	Comparison group	Nr of observ.	Diff-in-Diff with Matching Estimator
(1)	19–24 year olds living in Pathfinder areas	19–24 year olds living in all non-Pathfinder areas	3,716	0.110** (0.039)

Notes: Estimates of the effects of the New Deal used the JUVOS 5% longitudinal sample of JSA claimants. By the end of the 10th month, conditional on being on JSA for 6 months. All estimates from regressions including a set of other controls, namely marital status, sought occupation, region and some information on the labor market history (comprising the number of JSA spells and the proportion of time on JSA over the 2 years that precede the start of the present spell). Age and the number of JSA spells since 1982 are also included when similar age groups are being compared. Propensity score matching is done over the same covariates as the other estimates and the outcomes for the comparison groups are smoothed using cubic splines on the two propensity scores to achieve higher precision. Standard errors in parentheses ** = significant at 0.05 level. * = significant at 0.10 level.

Source: Blundell et al. (2004).

Table 11.2 considers a number of different possible comparison groups, providing some insight on the possible size of indirect effects. Each row in the table corresponds to a different comparison, including different estimates, obtained under different methods, of the effects of the Gateway on *outflows to employment* after 4 months of treatment. These provide some clues about the robustness of the results. We start by restricting the comparison group to be composed of eligible men living in matched non-Pathfinder areas in the second row. Depending on the method used, the estimated effect may rise or fall slightly, but not significantly so. *This evidence supports the comparability of the two groups used in row 1 of Table 11.1.*

The next row compares eligible and ineligible men aged 25 to 30 years old within the Pathfinder areas. Using an age-based eligibility criterion is our second main source of identification and is all that is available after the pilot period. The point estimates of the 4 months

Table 11.2. Further Results for the New Deal Employment Effects for Men

Experiment	Treatment group	Comparison group	Nr of observ.	Diff-in-Diff with Matching Estimator
(2)	19–24 year olds living in Pathfinder areas	19–24 year olds living in matched non-Pathfinder areas	1,193	0.134** (0.053)
(3)	19–24 year olds living in Pathfinder areas	25–30 year olds living in matched Pathfinder areas	1,096	0.104* (0.055)
(4)	19–24 year olds living in Pathfinder areas	31–40 year olds living in matched Pathfinder areas	1,169	0.159** (0.050)
Outflow into the employment option (affecting 19–24 year olds living in Pathfinder areas)			*4,486*	
(5)	25–30 year olds living in Pathfinder areas	25–30 year olds living in all other areas	3,180	0.016 (0.042)
(6)	25–30 year olds living in Pathfinder areas	25–30 year olds living in matched non-Pathfinder areas	983	0.055 (0.058)
(7)	19–30 year olds living in Pathfinder areas	19–30 year olds living in all other areas	6,896	0.066*′ (0.029)
(8)	19–50 year olds living in Pathfinder areas	19–50 year olds living in all other areas	12,749	0.036* (0.021)

Notes: See Table 11.1.

Source: Blundell et al. (2004).

effect using age-based are very close and insignificantly different from those in row 1 using different areas.

The estimates suggest a treatment effect of 11.4 percentage points when 25–30 year olds are used as the comparison group (row 3) compared to 11 percentage points when 19–24 year olds in non-Pathfinder areas are used as a comparison group (row 1 in Table 11.1). This estimate may suffer from substitution more acutely and it is not immune to local labor market wide wage effects. However, it is informative to know that the obtained results are very similar, independently of the procedure used. We cannot reject the simple null hypothesis of a model without substitution and equilibrium wage effects.[45] Alternatively, their effects may cancel out, the relative sizes of the substitution and wage effects being very similar. We further test for substitution using the older group of 31 to 40 year olds living in Pathfinder areas as a control group. This group is expected to be less substitutable for 19–24 year olds than the younger 25–30 year old comparison group. Under this assumption, and given that substitution exacerbates the impact of the program, we would expect this estimate to be lower than the one presented in row 3. But the fourth row presents an estimate of the 4 months effect of the Gateway that, if anything is higher than the previously presented results. This is not consistent with large substitution effects. In rows 5 and 6 we compare ineligible individuals living in Pathfinder and non-Pathfinder areas. If there were significant substitution effects or differential trends across regions we may find differences in outflows in the New Deal period. In fact no significant effects of the Gateway are found.

Finally, rows 7 and 8 in Table 11.2 contain estimates of the employment effect in the "whole market". Men aged 19 to 30 and 19 to 50 years old and living in Pathfinder areas are compared with similar individuals living in non-Pathfinder areas. The results only confirm what has been established before: that, during the pilot period, the program had a very significant positive impact on outflows to employment on the markets it has been implemented. The point estimates are smaller because 19–24 year olds are only a fraction of the larger age range. For example, just over half the 19–30 year old group are 19–24 year olds. The linear matching estimator in row 7 implies a New Deal effect of 6.6 percentage points – as expected just over half the magnitude of the effect in Table 11.1.

11.4.1.2 RESULTS FROM THE NEW DEAL NATIONAL ROLL OUT

Table 11.3 contains the main result from the National Roll Out. The first row shows an implied effect of around 5 percent on a pre-program base outflow (Table 11.2) of 25.8 percent, and once more, the method used does not seem the affect the result significantly. Although this is still a substantial impact, it is about half the magnitude estimated for the Pilot period. These differences in size can be accounted for by a “program introduction” effect. In the first few months the program is operating, a very large increase in the flows to employment is observed, which then falls as the program matures. This is illustrated in the other rows of the table. The second and third rows report comparable estimates of the Gateway effect after 4 months of treatment for the first quarter the program operates in the Pathfinder and non-Pathfinder areas, respectively. As noticed before, estimates for the pilot period (first quarter in Pathfinder areas) are about twice the size of the effect over the whole period. The same is also true if one considers the estimates for the first quarter the New Deal operates in non-Pathfinder areas (see row 3). The fourth row presents estimates obtained using the following second and third quarters the program is operating and these are comparatively much lower and less significant.

In summary, the New Deal is a mandatory program affecting all young people claiming unemployment benefit for at least six months in the UK. The program offers a combination of treatments, particularly job assistance for four months and a wage subsidy paid to employers. Two sources of identification were used to construct comparison

Table 11.3. Employment Effects From the New Deal National Roll Out

Experiment	Type of estimate	Number of observ.	Diff-in-Diff with Linear Matching
(1)	Overall effect for the sample including the Pilot period and the National Roll Out (first three quarters the ND operating in each region)	17.433	0.053** (0.013)
	Outflows to subsidized jobs	55,051	0.039
(2)	Effect for the Pilot period – 1st quarter the programme operates in Pathfinder areas	1,096	0.104* (0.055)
	Outflows to subsidized jobs	4,486	0.057
(3)	Effect for the 1st quarter the programme operates in non-Pathfinder areas	5,169	0.088** (0.025)
	Outflows to subsidized jobs	20,331	0.039
(4)	Effect for the 2nd and 3rd quarters the programme operates in all areas	11,161	0.031* (0.016)
	Outflows to subsidized jobs	30,234	0.036

Notes: See Table 11.1.

Source: Blundell et al. (2004).

groups in order to make inferences on the impact of the New Deal: a comparison between Pilot and non-Pilot areas and an age-related eligibility criteria. Our results suggest similar quantitative effects whichever comparison group is chosen.

The main finding is of an economically and statistically significant effect of the program on outflows to employment among men. The program appears to have caused an increase in the probability of young men (who had been unemployed for 6 months) finding a job in the next four months. On average, this increase is about 5 percentage points (relative to a pre-program baseline of 26 percent). Part of this overall effect is the job subsidy element and part is a pure "Gateway" element (enhanced job search). We estimate that *at least* 1 percentage point of the 5 percentage points is due to the Gateway services, such as job search assistance. We also found that the treatment impact is much larger in the first quarter of introduction. These findings are robust to a large number of experiments.

Our results are more optimistic than many of the results from US studies of the effects of government labor market programs for male youth. There are three reasons. First, it is important to recognize that the program was mandatory. Refusal to participate results in sanctions. Mandatory, sanction-enforced schemes have often been found to be more effective than voluntary schemes. Secondly, the "disadvantaged youths" we consider are less disadvantaged than those treated in typical US programs (e.g. ex-offenders). To the extent that programs are more effective on those who are more job ready, one would expect to see more signs of a program effect in the UK than in the US. Finally, recall that we are evaluating the effects of job search assistance and wage subsidies. The US evidence here is less pessimistic than the evidence on public training schemes.[46]

11.4.2 *Evaluating the Impact of the WFTC Reform*

WFTC was introduced in October 1999. There was no piloting or randomized demonstrations to assist in the evaluation of the WFTC reform. To evaluate it we therefore adopt two approaches. The first uses an ex-ante simulation model developed in Blundell et al. (2000a).[47] This model was estimated using pre-reform household level data from the Family Resources Survey. In the second approach we use post-reform administrative figures to double-check our predictions.

We also use data from the Labour Force Survey before and after the reform. This before and after evaluation requires choosing a control group and here we follow the Eissa and Liebman (1996) study of the EITC reform and use higher educated women with children whose earnings are sufficiently high to render them largely ineligible to WFTC. To summarize: the structural model simulations appear to line up well with the ex-post data and can therefore be used with some confidence to assess many of the important aspects of designing and implementing a tax credit program of this type.

The simulations focus on two target groups for the WFTC reform: single parents and married couples with children. Nearly 50% of currently working single parents were found to be in receipt of some Family Credit. For married couples with children this proportion is smaller, at around 16%. However, the latter group is more than two and half times the size of the former.

As we have seen, the WFTC reform is designed to influence the work incentives of those families with low potential returns in the labor market. It does this via the increased generosity of in-work means-tested benefits. For single parents the WFTC does unambiguously increase the incentive to work. For couples, however, income effects from a working spouse created by the WFTC, can lead to a *lower* participation in the labor market. Table 11.4 presents an overall simulated impact of the reform.

In Table 11.5 are the detailed simulations. Panel (a) presents the responses for single parents. The simulation takes around 2.2% of the sample from no work to either part-time or full-time work, with no offsetting movements out of the labor market. One can clearly see the reason for this move into work in the earlier graphs of the potential impact of the WFTC on single parents' budget constraints. At or above 16 hours per week the single parent becomes eligible for WFTC (with any childcare credit addition to which she may be entitled). For some women this extra income makes a transition to part-time employment attractive. Nevertheless, the level of the aggregate behavioral response is perhaps lower than one might have anticipated.

Table 11.4. WFTC Reform Simulations: Summary Impact on Employment

Group	Increase in Employment	% point change
Single Parents	34,000	2.20
Women in couples (Partner not working)	11,000	1.32
Women in couples (Partner working)	–20,000	–0.57
Married men, partner not working	13,000	0.37
Married men, partner working	–10,500	0.30
Total Effect	27,500	
Decrease in Workerless Families	57,000	

Source: Blundell et al. (2000b).

Table 11.5. WFTC Reforms, Detailed Simulation Results
(a) Single Parents

pre-reform	Out of work	post-reform part-time	Full-time	pre-reform %
Out of work	32.2	0.1	0.1	32.4
Part-time	0.3	31.6	0.0	32.0
Full-time	0.4	0.1	35.0	35.6
Post-reform %	33.0	31.8	35.2	100
Change (%)	0.6	–0.1	–0.4	

(b) Women in Couples with Employed Partners

pre-reform	Out of work	post-reform part-time	Full-time	pre-reform %
Out of work	32.2	0.1	0.1	32.4
Part-time	0.3	31.6	0.0	32.0
Full-time	0.4	0.1	35.0	35.6
Post-reform %	33.0	31.8	35.2	100
Change (%)	0.6	–0.1	–0.4	

(c) Women in Couples with Partners Out-of-Work Post

pre-reform	Out of work	post-reform part-time	Full-time	pre-reform %
Out of work	56.8	0.4	0.9	58.1
Part-time	0.0	22.2	0.4	22.6
Full-time	0.0	0.1	19.2	19.3
Post-reform %	56.8	22.8	20.5	100
Change (%)	–1.3	0.2	1.1	

Source: Blundell et al. (2000b).

We see a minor offsetting reduction in labor supply through a simulated shift from full-time to part-time employment among 0.2% of the sample. This is consistent with a small (negative) income effect among some full-time single women, for whom the increase in income through the WFTC encourages a reduction in labor supply.

Nevertheless, the predominant incentive effect among single parents is positive. Given the low level of participation – a little over 40% – a 2.2 percentage point increase is important.

For women in couples the simulated incentive effect is quite different. In Table 11.5b we report estimates of the transitions following WFTC among a sub-sample of women with employed partners. There is a significant overall *reduction* in the number of women in work of around 0.57%, equating to a grossed-up figure of around 20,000 in the population.[48] This overall reduction comprises around 0.2% who move into the labor market following the reform, and 0.8% who move from work to non-participation. The number of hours worked by women with employed partners is predicted to fall slightly, by 0.18 hours on average over the full sample.

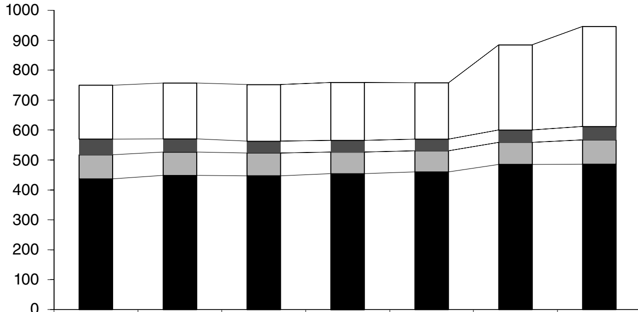
The predominant negative response is clearly not one that is intended, but from the earlier discussion one can easily see why. There will be a proportion of non-working women whose low earning partners will be eligible for the WFTC. The greater generosity of the tax credit relative to the current system of Family Credit increases household income. This increase in income would be lost if the woman in the household were to work. And for those women currently in the labor market, the WFTC increases the income available to the household if she were to stop working.

In panel (c) the incentives for a sub-sample of women whose partners do not work are presented. For this group there is a significant overall increase of 1.32% in the number of women who work, equating to a grossed-up figure of around 11,000 in the population. The reason for this shift is more straightforward, and stems from the increased generosity of the basic WFTC relative to the current Family Credit system for those women who choose to move into work. Note that for this group the generosity of the childcare credit component of the WFTC is not an issue, since households only qualify for the childcare credit if both household members work 16 hours or more. There is of course potential for both members of an unemployed household to move into work in order to qualify for the WFTC including the childcare credit, but a joint simulation (not reported here) shows that such an outcome is virtually non-existent.

11.4.2.1 SOME RECENT EX-POST EVIDENCE

The WFTC was introduced for all new recipients in October 1999 and fully phased in by April 2000. From recent administrative caseload data,[49] the introduction of the WFTC, and the substantial increase in generosity, appears to have had a marked effect on the number of people claiming in-work benefits. Indeed the caseload rose by 30% in the 12 months following May 1999.[50]

Some of the change in WFTC caseload will be due to the increased numbers of already working parents who qualify for WFTC due to its increased generosity – some of the so called 'windfall beneficiaries'. This alone cannot be taken as a measure of success in increasing employment, although it may be justified from a redistribution point of view. We can learn a little more by looking at administrative data on cross-benefit flows. Figure 11.14 breaks down the WFTC/FC caseload by their situation 12 months ago. It shows that a large component of



	Aug 98	Nov 98	Feb 99	Maz 99	Aug 99	Nov 99	Feb 00
□ New claimer (in own right)	180	186	189	193	188	285	334
■ From other MTBs	52	44	40	39	39	41	45
▫ Lone parents from IS	81	78	76	73	71	74	81
■ On WFTC already	437	449	447	454	460	485	486

Figure 11.14. Transitions of Families onto WFTC
Source: Blundell and Brewer (2000).

the caseload increase (around 75%, taking the last 4 quarters of FC as a baseline) since October 1999 came from people who were not claiming any means-tested benefits or tax credits 12 months before. Both these two facts are consistent with the increased entitlement of the WFTC compared with FC.

We also examine the impact on the relative employment rates of the main target groups. For example, Figure 11.15 shows the relative growth of low education to high education single parents. Using the high education group to control for common trends[51] the relative increase in employment since the introduction of WFTC in late 1999 shows about a 2.5% rate rise, very close to the prediction in Table 11.4.

Figure 11.16 presents the same comparison for women in couples with children who have working partners. Here there is evidence of a small relative decline in participation. Again much the same as predicted in Table 11.4.

Taken together with our simulation results these administrative statistics suggest that the impact of the WFTC reform on employment among low-income families in the UK has been positive but modest. This supports our overall view that the workings of the tax and benefit system in the UK together with the increased generosity to workless families with children, mean that changes to financial work incentives from in-work benefit reforms are significant but relatively small.[52]

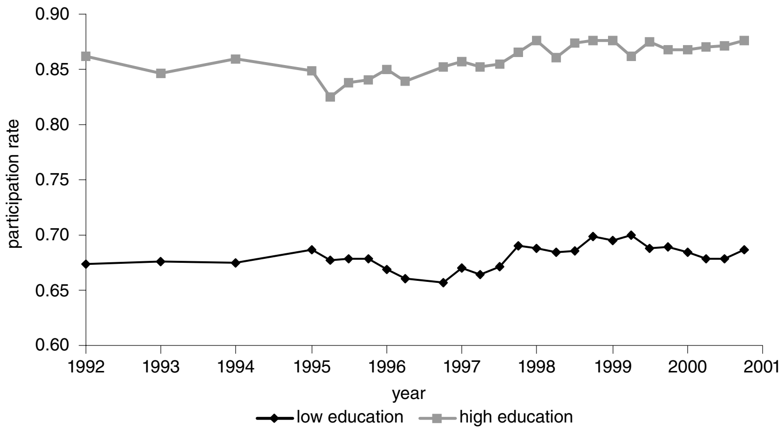


Figure 11.15. Single Parent Employment Rates by Education of Mother
Source: LFS.

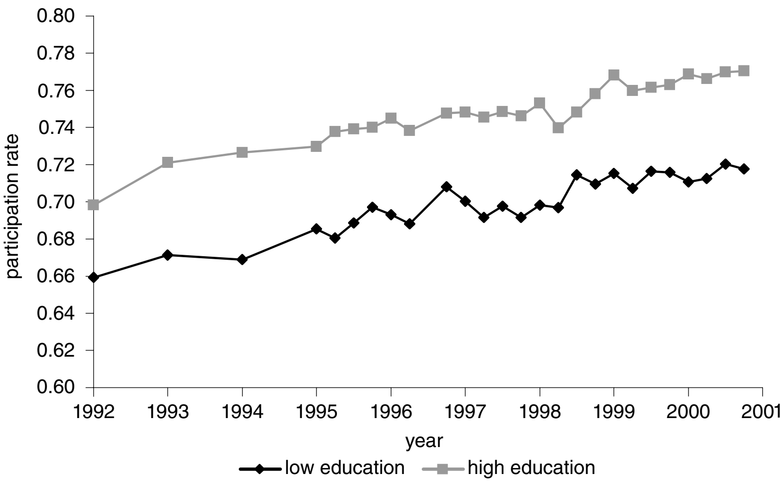


Figure 11.16. Employment Rate: UK Women in Couples with Children with Partner Working
Source: LFS.

11.4.3 *The Self Sufficiency Program (SSP) – An Experimental Evaluation*

To conclude this discussion we look more closely at an evaluation of a particularly interesting financial incentive program in Canada – the Self Sufficiency Program. This is purely an experimental or demonstration program, running in British Columbia and New Brunswick, and is examined in detail in Card and Robbins (1998). Figure 11.17 shows a typical budget constraint for a Canadian welfare recipient. It gives the budget set that an individual would face if they were earning the minimum wage in British Columbia, which was $6 an hour in 1993. Taking a job at a few hours a week attracts an earnings disregard of around $200, thereafter all earned income is effectively lost in a dollar-for-dollar transfer back to the income assistance program. So, until recipients have exhausted their income assistance – that is working nearly 50 hours a week – they would have no return, with an implicit tax rate of 100% on their earnings.

The SSP is available to a single parent with twelve months welfare history and who finds a job averaging 30 hours a week over a period of a month. This is calculated on a monthly rolling period. Providing employment is found within the first twelve months of the program, the participant remains eligible for three years. It is a generous system and does not change the income assistance level; so it is not, for example, causing more individuals who do not find employment to be on lower incomes. It is simply giving an earnings supplement to those who move into work.

The experimental nature of this reform makes it particularly attractive for evaluation reforms that rely on financial incentive to induce welfare recipients into work. The experiment entailed following 6,000 families for 5 years starting in 1993. One-half of the group of 6,000 eligible single parents on welfare were offered the program and the others were not – they are the controls. The individuals that are on the program are the treatments – and we can compare those two groups. As can be seen for Figure 11.18, the control and the treatment have very similar employment patterns before the experiment takes place. This is an indication of a well designed experiment and means that the controls are really quite a good match for the treatment group.

There is almost a doubling in employment for the treatment group. This is displayed in Figure 11.18, which also shows the close relationship between employment rates across the control and treatment group before the experiment began. Card and Robbins (1998) report many more results. In particular, the impact on hours and employment is very similar. These are low hour working individuals. The eligibility

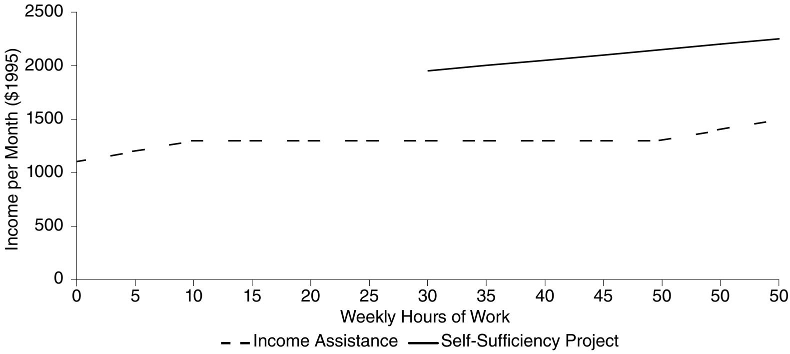


Figure 11.17. The SSP Budget Constraint

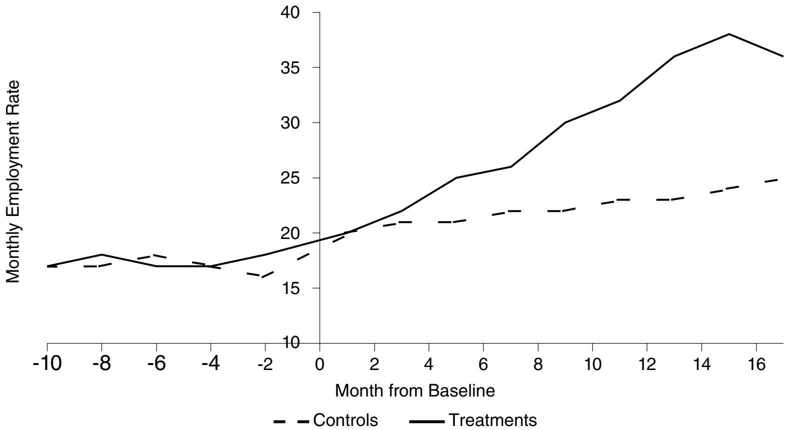


Figure 11.18. The SSP Impact on Monthly Employment for Single Parents

criterion is that recipients work at least one week of the month for 30 hours. For this experiment, the treatment group was found to have increased its hours of work, more or less, twofold over the control group.

Recent evidence on the SSP, see Bloom and Michalopoulos (2001), points to the control groups slowly catching up with the treatment group suggesting that these schemes enhance the speed with which individuals move off welfare but may not have such a large long run impact. It is also worth noting that the wages received by the participants were slightly lower than those received by the controls. Suggesting that the 'incentivised' group does, as one would expect, face less attractive labor market conditions.

11.5 Conclusions: Designing an Appropriate Welfare-to-Work Policy

11.5.1 *An Overview*

This lecture has identified several central aspects of the design of welfare-to-work and make-work-pay programs. It focussed on two broad types of schemes. The first is an individually based active labor market program that assists in job search and provides a wage or earnings subsidy once employment is found. Eligibility typically depends on a minimum duration of unemployment insurance or welfare; the subsidy is typically individually based and time limited. The second type of program is an earned income tax credit. This also provides a wage or earnings supplement. However, in this case the level of the supplement is typically means tested according to family income and varies with family size and composition. It is also typically not time-limited and has no welfare or UI duration eligibility. Although both operate through a supplement to earned income, they operate in very different ways. Is one design better than another? Is one more suited to a particular group? The analysis presented in this paper has highlighted five central design features: targeting, time limits, hours conditions, incentives for wage progression, and job search assistance.

Targeting can be by type of individual, by level of earnings and by family income. Each is designed to reduce cost and reduce deadweight. Targeting by type can increase substitution with 'close' types and, to be cost effective, it also typically requires some welfare/UI duration condition. But this in turn can lead to stigma effects. Targeting by earnings has the advantage of identifying low earners and the low skilled. But it can create a disincentive for effort and hours worked. It also reduces incentive for wage progression and skill formation in an effort based learning model. MaCurdy and McIntyre (2004) argue for an hourly wage based credit since this is more directly related to low skill, rather than simply low hours and creates less adverse effects on effort.[53] Targeting by income has the advantage of identifying poor families but often carries with it stigma effects and can create adverse family labor supply incentives. Operating through the tax return, as in the case of EITC (and WFTC), arguably reduce stigma effects.

Turning to time limits, these can refer to time limited unemployment or welfare benefits as well as time limited in-work tax credits and earnings supplements. Each seems to effectively reduce the disincentive effects that naturally occur in welfare and tax credit systems. An important issue is to what extent individuals move out of the system or simply cycle round the system.[54] These effects can be offset by wage progression. Indeed, time limits on earnings supplements and subsidies may enhance the incentive for wage progression in some cases. Of course, a phase-out rate with wage progression can act as a natural time limit. But a phase-out itself acts as a disincentive for wage progression.

The impact of these alternative designs on wage progression depends on the form of skill formation. There are typically two models used in the labor economics literature: (1) 'Passive' learning by doing models, in which wage progression itself provides a natural 'time limit'. However, low experience related learning for low skilled and low overall wage progression at some 2–3% per year suggests time limits that are too short, could be counter productive. (2) Effort based learning/investment models in which withdrawal rates act as a disincentive for human capital investment. In this case a time limit can help offset these adverse incentives. There is some evidence of important active wage progression and time limits will reduce the negative effect of the phase-out rate.

The imposition of a minimum hours requirement is designed to offset the incentive to reduce hours which underlies a high phase-out rate. However, quite different levels are used: for example, 35 hours in New Deal, 30 hours in SSP, 16 hours in WFTC, no minimum in EITC. If set too high they discourage work for those facing fixed costs of work – childcare costs, for example. It may be natural therefore to choose a 'low' limit for parents – 16 hours. Although the likelihood of wage progression in part-time low-skilled jobs may be slight. Adequate childcare support could mitigate against the need to set very low minimum hours conditions. In any case, there is strong evidence that the incentives underlying minimum hours conditions work, as noted by the peaks in working hours distributions.

A welfare or UI duration requirement reduces deadweight and targets those with low labor market experience. But it may induce longer welfare durations and increases stigma or labelling. It also is unable to adjust to 'shocks' or changes in earnings that occur without an unemployment spell. Again we have seen that schemes vary considerably. The New Deal choosing 6 months of unemployment claims and the SSP 12 months of welfare claims.

Job search assistance seems a natural supplement to any financial incentive to move from welfare-to-work. However, the evidence is mixed. In the New Deal the mandatory nature of the scheme seem to have a relatively important impact but this may also reflect the threat of sanctions.[55] In the MFIP mandatory job search assistance also seems to have worked well. And this was measured through a randomized experiment. However, a voluntary scheme in the SSP (also experimentally evaluated), although having high take-up had relatively small effects on longer-term employment.

11.5.2 *An Assessment: Designing an Appropriate Welfare to Work Policy*

The aim of this lecture was to open up the discussion of welfare-to-work or make-work-pay programs so as to define a broad set of design issues. The idea is to think in terms of an integrated set of policies designed to address the problems of subpopulations with low income, low human capital and low labor market attachment. The review of such schemes by Meyer (1995) finds a significant (and cost effective) impact of mandatory job search assistance schemes operating in the late 1980s in the US.

There are no magic solutions but several preferred design features emerge. There is strong evidence that financial incentives encourage work even among the low skilled, welfare dependent populations with little labor market experience. Time limiting seems to help with human capital incentives and self-sufficiency, but the length needs to be gauged to allow for the relatively slow rates of wage progression that are likely to occur. Indeed the evidence is that wage progression is likely to be low and time limits should be relatively generous. Targeting welfare dependent and unemployed populations is also more cost effective and probably reasonably equitable provided a longish time limit is set. That is not to say there should not remain some overall negative income tax or tax credit in place, but the generosity of this can be traded off against the need to target certain low-income populations. Indeed, it may not be beneficial on the mothers of young children to work, at least in comparison to pure income transfers. Here there is little reliable evidence.[56]

Financial incentives appear to work better when they are individually based rather than employer based. This may be because job mobility is an important route to wage progression but it also seems to be affected by issues relating to stigma. Moreover, mandatory job search assistance, together with sanctions, seems to play a useful role. Further there is no reason why help with job search and job matching should not extend into work. Progression and advancement in work can also be enhanced by training. This appears to bring higher returns if workplace located, at least for lower skilled individuals with low levels of prior education. This is particularly the case for training that leads to accredited qualifications to enhance transferability. Consequently, an additional training subsidy for the employed may be needed in addition to an individual financial incentive. Although, in principle, the financial incentive should allow the individual to bear the reduced wage during training.

There is lots we still need to know. But there is a growing evaluation literature, some of which has been referred to here, which is building a large array of results on a wide variety of programs, all of which try to address the problems of low income and low labor market attachment. This evidence is disparate and reforms are typically piecemeal. But they need not be. An integrated view of reform in this area needs to bring together welfare to work, tax credit, benefit and active labor market programs under one guise so that a complete picture can be drawn of the incentives for labor market attachment, income progression and redistribution.

The analysis so far suggests that an earnings tax credit program with time limits that are reasonably long and which targets welfare dependency, thereby focusing on low human capital and low labor market attachment, could form the basis for an integrated view of Employment Tax Credits and New Deal style programs. It could work as a relatively low cost way of enhancing earnings and self-sufficiency among these target populations.

12

Earned Income Tax Credit Policies: Impact and Optimality

"No society can surely be flourishing and happy in which part of the members are poor and miserable. Thus far at least seems certain, that, in order to bring up a family, the labor of the parent must be able to earn something more than what is precisely necessary. Indeed poverty, though it does not prevent the generation, is extremely unfavourable to the rearing of children."

Adam Smith: An Inquiry into the Nature and Causes of the Wealth of Nations

12.1 Introduction

The policies analyzed in this lecture are those directed towards addressing low labor market attachment and low wages of among certain groups of parents with children. The aim of the research agenda reported on in this lecture is to evaluate the impact and assess the optimality of Earned Income Tax Credits policies specifically for lone parents.[1] These policies have been at the center of welfare-to-work reforms in the UK, in the US and increasingly in continental Europe.

They are in the class of "make work pay" reforms highlighting a "work condition" in welfare policy. The objective is to balance poverty reduction in families with children and employment incentives. In the

Reprinted from Labour Economics 13(4), Blundell, R., 'Earned Income Tax Credit Policies: Impact and Optimality: The Adam Smith Lecture, (2005)', 423–443. © 2006, with permission from Elsevier. This paper is the text of the Adam Smith Lecture presented at the joint EALE/SOLE World Meeting in San Francisco, June 2005. I would like to thank my colleagues Mike Brewer and Andrew Shephard for allowing me to draw on our joint work. This study is part of the research program of the ESRC Centre for the Microeconomic Analysis of Fiscal Policy at the IFS. I am responsible for all errors and interpretations.

context of policy in the UK, lone parents have been a key group. This has also been the case for the US; see Blundell and Hoynes (2004), for example.

There are two key questions considered in this lecture. First, what is the impact of such in-work benefit reforms on labor supply and to what extent does a standard labor supply model capture the main impact? Second, to what extent are such tax credits policies optimal, that is do they constitute an optimal income transfer for low-income people?

As part of this lecture I also want to address a pronounced puzzle in the comparison of tax credit policies in the UK and the US. As we will see, on face value, the UK policy appears about twice as generous as the US policy. That is, the maximum transfer available in the UK system is twice in real terms that available through the US system. Yet the impact of this policy on labor supply responses among key eligible groups in the UK looks to be about half what it was among similar groups in the US. Why so? As we will show the puzzle can be convincingly resolved, and not in the trivial sense of attributing the differences to US workers being more responsive to incentives. Far from it, in fact, labor supply response elasticities across the two countries seem about the same. Rather, it is the design of the expansions in the generosity of these policies and the interaction of tax credits with other parts of the tax and benefit system that hold the answer. As we shall see the resolution of this puzzle highlights some of the key design issues in earned income tax credit policies both in terms of their impact and their optimality.

The growing popularity of earned income tax policies has stemmed from changes in the economic environment during the late 1980s and early 1990s. Specifically, the secular decline in the relative real wages of the low skilled,[2] and the resilience of child poverty rates in both the US and the UK.[3] To combat these two issues, welfare-to-work policies turned to in-work credits for lone parents. As a result the last decade has seen the increasing reliance in welfare policy on in-work benefits, and more specifically on earned income tax credits, see for example Inland Revenue (2001). The aim of such policies is to break the "iron triangle" of welfare policy – that is the three, often conflicting, goals: raising the living standards of those on low incomes; encouraging work and economic self-sufficiency; and keeping government costs low.

There is an expanding theoretical literature examining the role of work requirements in the design of optimal income transfer programs. In a dynamic model, the important issue relates to incentives for poverty reducing investments and investments in human capital. Besley and Coate (1992) derive conditions under which workfare can be optimal. Cossa, Heckman and Lochner (2002) develop a dynamic model with time limits and human capital investment. In a more static setting the recent contribution by Saez (2002) shows that, where labor supply responses are concentrated along the extensive margin (participation in work), an earned income tax credit system with transfers that increase with earnings at low levels can be optimal and justifies the move away from negative income tax schemes. Moffitt (2005) argues that paternalistic social welfare functions that include a social value placed on work are best able to motivate such reforms. It is these static optimality results that form the background to the analysis reported here, although I will return to the more dynamic aspects at the end of the lecture.

Using estimates of structural models of labor supply responses at the extensive and intensive margin, I will pose the question: can the existing tax credit systems we observe be considered "optimal" for reasonable social welfare weights? As a precursor to the analysis I will have to convince you of the validity of the structural model estimates. For this I will make a comparison with a simple difference-in-differences evaluation strategy. Although not providing sufficient information for policy simulation or the assessment of optimality, simple difference-in-differences evaluations can be valuable for validating the specification of more fragile microeconometric models. Provided the comparison groups can be reasonably argued to experience the same macroeconomic trends, and there are no systematic composition changes before and after the program, a difference-in-differences methodology can provide a useful guide to the extent of a policy impact. With a validated structural labor supply model I then turn to the optimality of the tax credit policies. It turns out that for the samples of lone parents we examine, an earned income tax credit is likely only to be optimal for those families on low incomes with children of school age. For those with pre-school age children and for reasonable social welfare weights it is much more difficult to justify allocating a larger transfer to those in work as a tax credit implies even if heavily means-tested.

The layout of the remainder of the discussion is as follows. In the next section I will briefly consider the relationship of earned income tax policies with other wage subsidy policies. Then I will move on to the policy context for the earned income tax credit policies. Section 12.4 will examine the nature of the UK reforms in comparison with the US policies. In Section 12.5 I will turn to the dual questions of impact and optimality. Finally, Section 12.6 will conclude and draw together some lessons for earned income tax policy design.

12.2 Earned Income Tax Credits and Their Relationship with Wage Subsidy Policies

To some extent earned income tax credit policies can be seen within the general set of wage subsidy policies but there are important differences. The wage subsidy is typically individually-based, not means-tested and has limited duration, see Katz (1998) and Phelps (1994). Where wage subsidies are provided to individuals, rather than directly to firms, eligibility is usually dependent on a certain duration of unemployment insurance (or welfare) receipt. The earned income tax credit, on the other hand, is typically subject to a family income based means-test and does not have a time limit. For the latter, the WFTC in the UK, the EITC in the US[4] and the In-Work Tax Credit in Belgium[5] are prime examples. For the wage subsidy case, the New Deal for Young People in the UK and Work First[6] in the US are leading examples.

There are, of course, many welfare-to-work policies that fall somewhere in between these two extremes. For example, the Self-Sufficiency Project (SSP)[7] in Canada, although an in-work tax credit like the WFTC or EITC, has a three year time-limit and eligibility depends not only on overall family income and family composition but also on a minimum welfare duration and a minimum hours requirement. The New Hope[8] tax credit program in the US also has a three-year time limit and a minimum hours condition. Both programs provide job search assistance at least for some of the program's participants.[9] The Minnesota Family Investment Program (MFIP),[10] is similar to the SSP, however, the job search assistance is mandatory as in the New Deal for Young People in the UK. An additional feature of these Canadian and US programs is that many were the subject of randomized experimental evaluation, the results of which provide a vital source of information in the discussion below. Finally, the earnings supplement and job search provisions within the many US state run additions to the Temporary Assistance for Needy Families (TANF) program in the US have similar characteristics to the New Deal program (see Card and Blank 2000b).

An essential characteristic of an earned income tax credit like the WFTC and EITC is the long term commitment to redistribution and the high implicit tax rates that are a consequence of the means-testing principle on which such tax credits are based. But what of their impact and their optimality? Can we think of means-testing combined with a work condition as an optimal component in a tax and benefit system? To address these issues we first turn briefly to the motivation for the introduction and expansion of earned income tax policies in the UK and the US during the 1990s and into the new millennium.

12.3 The Labor Market Background for the Shift Towards In-Work Benefits

The low levels of employment, experienced by certain specific demographic groups of working age in Europe and North America during the early 1990s, were a strong motivation for the introduction and expansion of earned income tax policies over this period. For example, one central stimulus for the Working Families Tax Credit in the UK was the stubbornly low levels of attachment to the labor market by single mothers in the 1990s – at a time when for other groups of similar women attachment had generally been increasing. Figure 12.1 shows the secular change in female employment across four household types in the UK over the 1980s and 1990s. The growth in the attachment by women in couples with children was as noticeable as is the fall for single women with children.[11] This low level of labor market attachment was even more pronounced for those with low levels of education. Blundell and Hoynes (2004) document this change and examine the similarities between demographic trends for single mothers in the UK and US.

Another distinguishing characteristic of the UK labor market over this period was the increase of workless couples with children. This

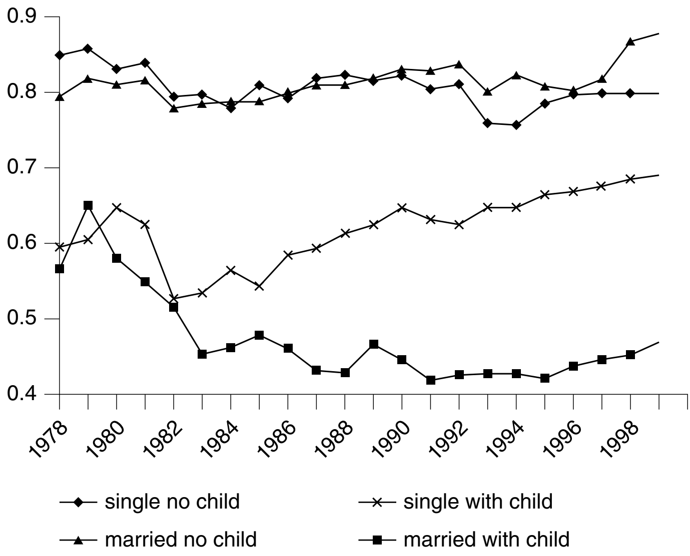


Figure 12.1. Employment Trends for Women in the UK: Proportion in Work

Notes: FES Data, working age.

Figure 12.2. Proportion of Workless Couples in the UK
Notes: FES Data. Working age head.

is documented in Figure 12.2 and provided a strong argument in the debate over the WFTC reform. Indeed, for women in couples with unemployed partners employment rates have stayed no higher than 30% over the past two decades – even lower than employment rates for the single parent group (see Gregg, Hansen and Wadsworth 1999). The (non-)employment rates for these two groups show clearly why they have been singled out as two target groups for tax and benefit reform. Unlike earned income tax reforms in the US, couples with children were given similar incentives to single parents. However, because the level of the credit was means-tested against family income, there were perverse incentives to work for individuals in couples where there was already one spouse in work.

In the US there is a relatively small credit for couples but nonetheless Eissa and Hoynes (2004) still document a perverse negative income effect on the wives of low income working men. This has been an important feature of the family income based means-testing component of earned income tax credits and there is strong evidence of negative employment effects in the UK among working wives in low-income families where both adults work, see Blundell et al. (2000b). Here I focus on lone parents, where this effect is irrelevant, although the issue of individual versus joint income assessment in the overall design of earned income tax credits for married couples is one that deserves serious attention.[12]

12.4 The Earned Income Tax Credit Reforms: WFTC and EITC

The 1999 reform, in which the UK government introduced WFTC, provided an increased generosity of the existing in-work benefits in the UK Family Credit. The Working Families Tax Credit had its antecedents in the Family Credit (FC) system introduced in the late 1980s. Indeed, earned income tax credit policies in the UK and the US date back more than two and a half decades.[13] The Family Credit policy in the UK was designed to provide modest support for low-wage working families.

There were three main eligibility criteria: work eligibility, which for lone parents required a job with 16 or more hours per week; family eligibility, which required children in full-time education or younger; income eligibility, which required a family's net income being below a certain threshold. In this system each eligible family was paid a credit up to a maximum amount which depended on the number of children. There was also a small addition in case of full-time work. As income increased above the threshold, the credit was withdrawn at a rate of 70%. In 1996, just before the WFTC reform, average FC payments were comparable to payments to those who were not working and take-up rates stood at 69% of eligible individuals and 82% of the potential expenditure.

The WFTC policy retained the main eligibility criteria of the Family Credit policy. However, the generosity of the system was expanded in a number of ways. It increased the credit for younger children and the overall income threshold, see Table 12.1.

Table 12.1. Adult and Child Elements of the WFTC

	Adult	Child awards by age		
		0 to 10	11 to 15	16 to 18
Mar-99	£58.80	£16.40	£22.60	£28.00
Oct-99	£56.60	£21.50	£22.60	£28.00
Mar-00	£56.60	£22.60	£22.60	£28.00
Jun-01	£61.90	£27.30	£27.30	£28.00
Jun-02	£64.40	£27.30	£27.30	£28.00
Increase	19.70%	66.40%	20.50%	0.00%

Note: All monetary amounts are expressed in April 2003 prices.

It reduced the benefit-reduction rate from 70% to 55%. Finally, the reform incorporated a childcare credit. This was worth 70% of actual childcare costs up to £150 per week (for two children, £100 for one child). The largest cash gains went to those people who were currently just at the end of the benefit-reduction taper. The transfers underlying the WFTC expansion are illustrated in Figure 12.3.

On the face of it, the UK system was and remains a very generous tax credit system, more generous than the US equivalent, as shown in Figure 12.4 which provides a direct comparison with EITC. As with the WFTC, eligibility for the EITC policy requires dependent children, positive earned income, and having income below the limit. By the end of the 1990s the EITC was phased in at a 34 (40) percent rate, phased out at a rate of 15.98 (21.06) percent for families with one child (two or more children). Figure 12.4 presents the two systems in terms of their gross transfers in 2001. These are evaluated for a minimum wage single parent with one and with two eligible children in both systems. Assuming that eligibility and receipt continued for a complete year. The broad similarities in the programs include larger credits for two child families and the phasing out of the benefits.

The differences are also clear from the figure. The vertical rise in eligibility in the UK system corresponds to the minimum-hours eligibility at 16 hours. At 16 hours the UK recipient receives the maximum she is eligible to. This contrasts with the US proportionate tax credit up to the

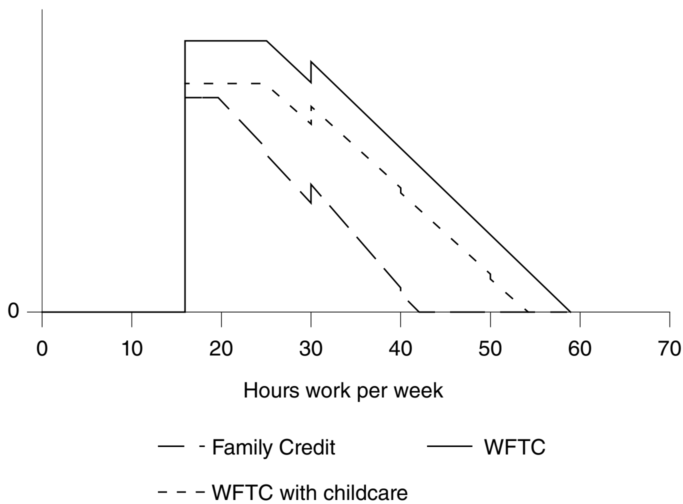


Figure 12.3. The WFTC Expansion

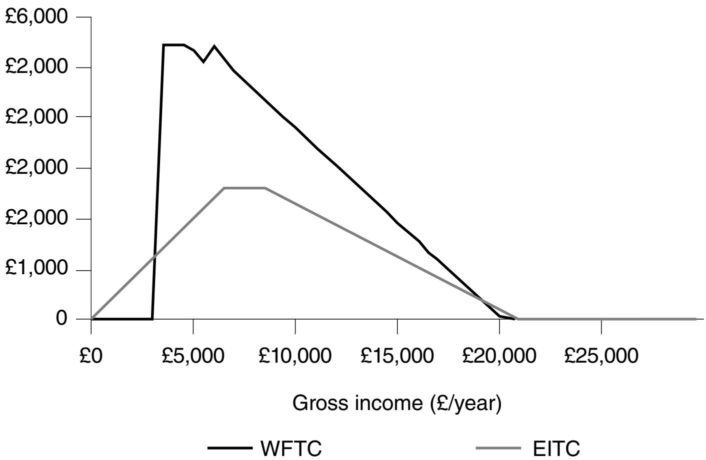


Figure 12.4. The WFTC and EITC Compared

Note: All monetary amounts are expressed in April 2003 prices.

Source: Brewer (2001).

maximum amount. The UK system also displays a much steeper withdrawal reflecting a higher benefit-reduction rate. This provides for a greater degree of targeting in the UK system but the potential for higher implicit tax rates. There are any additional specific idiosyncrasies to each of these systems (see Brewer 2001 for an in depth recent comparison). Overall, for low-earning families the UK system appears to be quite generous and significantly more so than the US system.

A key feature of the UK policy is that the tax credit is based on net (rather than gross) family income and we show that it is important when assessing the impact and design of the reform to allow for the interaction with other benefits and taxes. In contrast to the EITC in the US, the WFTC interacts fully with other benefits, most especially housing benefit. A majority of those individuals eligible to WFTC are also in receipt of housing benefit. Since income from the earned income tax credit is counted as income in the computation of Housing Benefit, the overall impact on net income of reforms to the system can be substantially reduced. This significantly reduces the incentive to work in the WFTC for families with large housing costs in the private or public rented sector, the large majority of single parents. Figure 12.5 illustrates these interactions for a "typical" single parent on the minimum wage post-WFTC.

Despite the dampening effect of these interactions with other benefits, there does seem to be some prima facie evidence of an impact on

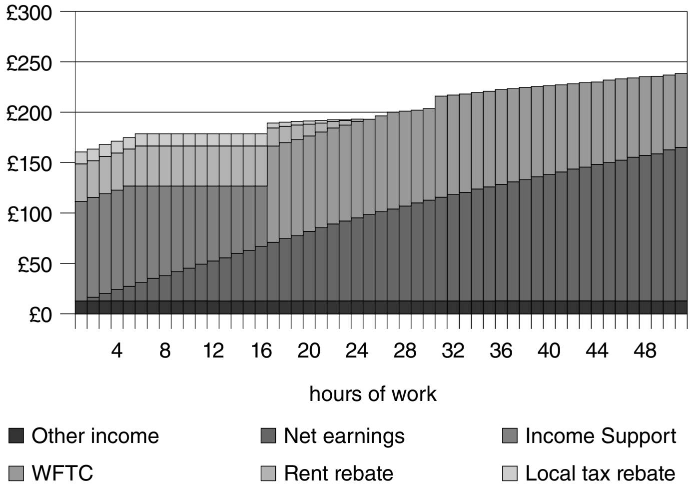
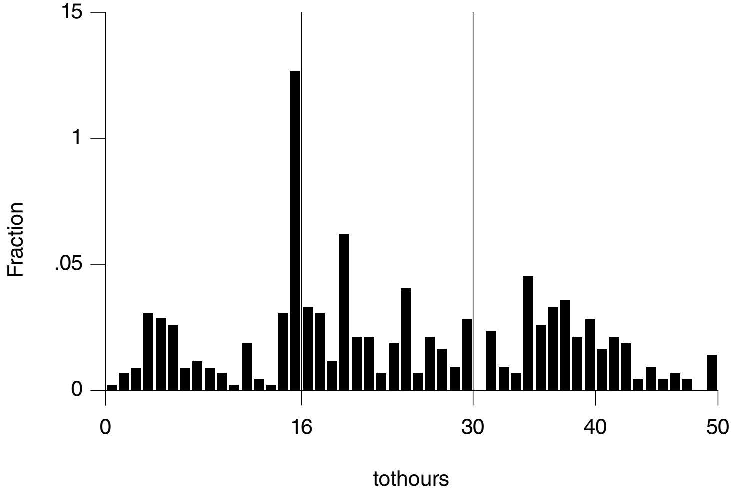


Figure 12.5. WFTC and the Interaction with Other Taxes and Benefits

behavior. A look at the histogram of weekly hours worked for single parents presented in Figure 12.6a, for example, shows a strong peak in hours worked at 16 hours. This is not evident for ineligible groups such as single childless low-educated working women as reported in Figure 12.6b. Of course, there will be a large number of so called "windfall beneficiaries" and there may also be those who decide to reduce their working hours in response to the incentive at 16 hours.

What actually happened when WFTC was introduced? The number of recipients increased markedly after its introduction in October 1999, and continued to rise at a much faster growth rate than seen under Family Credit (see Inland Revenue 2003; 2005). A year after its introduction, caseload had risen by 39%, and the majority of this increased caseload seems to have come directly from the increased generosity making more families entitled, rather than from families moving into work. The caseload of lone parents on out-of-work benefits (income support) has declined steadily and slowly since late 1996, with no discernible change in trend around 1999-2000. Analysis of administrative data that tracks individuals across income-related programs shows that the net inflow of lone parents from out-of-work benefits to WFTC in the 12 months from November 1999 to November 2000 was 50,000, 17,000 higher than the last 12 months of FC. Overall, the number of children in families on either out-of-work welfare benefits or FC/ WFTC has increased since early 1999.



a) Weekly Hours Worked: Low Education Single Parents in the UK

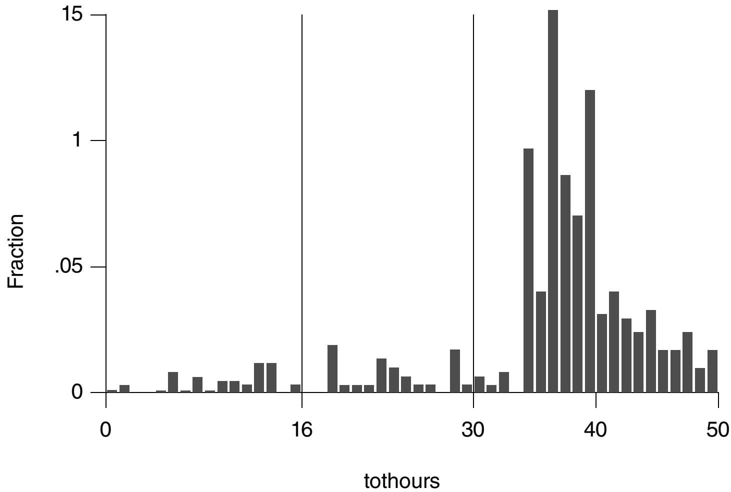


Figure 12.6. b) Weekly Hours Worked: Low Education Single Women Without Children in the UK

Notes: Family Resources Survey, 1998/99; Blundell and Hoynes (2001).

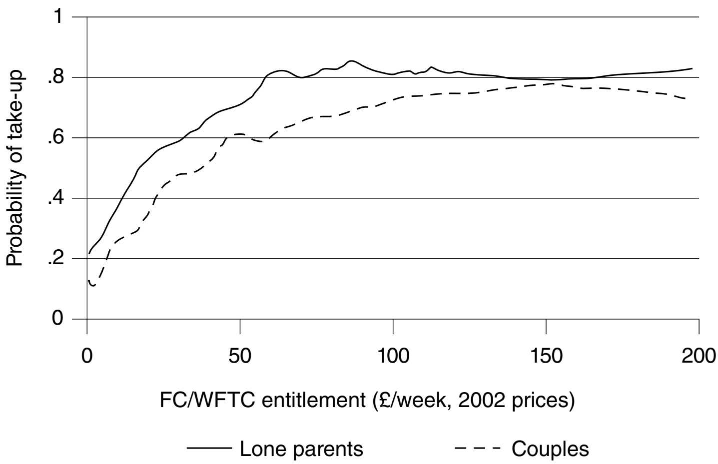


Figure 12.7. Take-Up Rates and the WFTC
Source: Adam and Brewer (2005).

The take-up of WFTC was roughly the same as it had been for FC for lone parents. Overall take-up among eligible lone parents was around 70% but varied in important ways with eligible entitlement level. Nonetheless as Figure 12.7 suggests take-up is an important characteristic of the WFTC system, with the take-up rate rising with the level of financial entitlement. In modelling any reform the trade-off between stigma and financial benefits in individual decision making is an important characteristic of behavior.[14]

As a final point it is important to note that other reforms occurred at the same time as the WFTC policy. This coincidence of reforms is crucial in understanding the impact of the reforms. It is also key in interpreting the degree to which child poverty relief as much as work incentives were an important design feature of the WFTC reform. At the time of the WFTC reform, there were three other main ways that the UK tax and transfer system provided support for children: Child Benefit, child allowances in Income Support, and a non-refundable income tax allowance.

The changes in the child rates of Income Support are documented in Table 12.2 and provide a further clue to the resolution of the puzzle in the comparison of impacts between the EITC and WFTC. Indeed the typical budget constraint for a single parent will have changed, but only as is documented in Figure 12.8. Thus providing some incentive for a move to full-time work but little overall strong incentives to work for many single parents facing relatively high costs of work.

Table 12.2. Child Rates of Income Support

	Child		
	0 to 10	11 to 15	16 to 18
Mar-99	£21.90	£28.00	£33.50
Oct-99	£27.00	£28.00	£33.50
Mar-00	£28.40	£28.40	£33.80
Mar-01	£33.00	£33.00	£33.80
Oct-01	£34.50	£34.50	£35.40
Mar-02	£34.50	£34.50	£35.40
Increase	57.50%	23.30%	5.70%

Notes: All monetary are expressed in April 2003 prices.

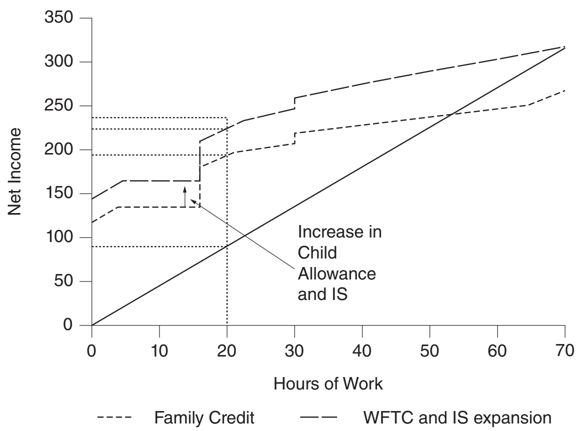


Figure 12.8. Transfers and Taxes under Family Credit (lone parent, min wage)

12.5 The Impact and Optimality

It is often the case that evaluation studies in modern labor economics restrict themselves to the simple average impact of any reform. In thinking about an earned income tax credit this seems too limited an objective. Especially as the reform itself is supposed to be balancing work disincentives and efficiency costs against redistribution and child poverty alleviation. In this lecture I want to pose a broader evaluation question: what is the impact of such policies and in what sense are they “optimal”? That is, do they constitute an optimal income transfer for low-income people?

In terms of optimal design, much theoretical work has focussed on the intensive margin of labor supply responses. However, recent work, notably Saez (2001; 2002), has developed a theory that combines decisions at the extensive and the intensive margins – employment and effort (hours).[15] If extensive and intensive elasticities differ, then the optimal structure of taxes and benefits can be to transfer more to those with low income but in work than those out of work, even with welfare weights that decline monotonically with income. Exactly the structure of an earned income tax credit.

To assess optimality we need robust estimates of elasticities at extensive (participation) and intensive (hours and weeks of work). These are structural parameters. Typically quasi-experimental or experimental provide estimates of average treatment effects from specific policy reforms which, while robust, bare only indirect relation to the elasticities needed.[16] On their own quasi-experimental approaches do not identify all the parameters necessary to assess optimality. But they can be argued to provide useful measures of average impact effects, see Blundell and Costa-Dias (2000; 2009) and Heckman, LaLonde, and Smith (2000), for example. As a consequence they can be used to assess the validity of structural estimates of the elasticity parameters. This is the approach taken in Blundell, Brewer and Shephard (2005) and Blundell, Chiappori and Meghir (2005) where a difference-in-differences estimator of the average impact of the WFTC reform is used to validate the structural simulation model required for the optimality analysis. The structural model is a stochastic choice model of labor supply and program participation building on the earlier work of Hoynes (2000), Keane and Moffitt (1998), Blundell et al. (2000b) and van Soest, Das and Gong (2002).

As we will see the structural evaluation results of the WFTC policy reform do show smaller impact effects than may have originally been expected given the generosity of the reform. But results appear robust – the quasi-experimental difference-in-differences estimate does not reject the structural model. To pre-empt: the small effects are due to interaction of WFTC with other taxes/benefits and the rise in family allowances which are given without a work condition, rather than "small" response elasticities. Moreover, under reasonable welfare weights the general design of the WFTC policy do line up with an optimal earned income tax credit design.

12.5.1 *Impact*

There was no piloting or randomized demonstrations to assist in the evaluation of the WFTC policy reform. To evaluate the impact, two approaches are adopted. The first uses the simulation model developed in the ex-ante structural evaluation study of Blundell et al. (2000b) and further developed in Brewer et al. (2006). This model was based on earlier structural labor supply research by Hoynes (2000) and by Keane and Moffitt (1998). In particular, it allows for childcare demands to vary with hours worked and it allows for fixed costs of work. It also accounts for take-up by incorporating welfare stigma costs following on from Keane and Moffitt (1998). In the second approach we use data from before and after the reform to provide a quasi-experimental difference-in-differences estimate of the average impact of the reform. A simple difference-in-differences methodology can provide a useful guide to the extent of a policy impact. Here we follow the Eissa and Liebman (1996) study of the EITC expansion in our choice of comparison group. This involves comparing outcomes of (potentially) eligible versus those single women without children who are not eligible. From this we identify average impact on eligibles by assuming a structure on unobservables. There are three key assumptions (i) separability, (ii) common trends across groups and (iii) invariance in group heterogeneity over time. In our implementation we follow Blundell et al. (2004) and adopt a "matching difference-in-differences" strategy which means that these assumptions only need to be valid conditional on a set of (matching) covariates.

The data on single mothers and childless single women used in the difference-in-differences analysis comes from two sources. The Family Resources Survey (FRS) is the data used also for the structural analysis. This is a cross-section household-based survey drawn from postcode records across Great Britain: around 30,000 families each year are asked detailed questions in face to face CAPI interviews about earnings, other forms of income, family composition and labor market status. It is the data set most often used to micro-simulate tax and benefit reforms in the UK, and was used to model labor supply in Blundell et al. (2000b). The second source is the Labour Force Survey (LFS) which is much bigger than the LFS but has much less accurate measures of income and hours and is not suitable for a structural analysis which requires careful measurement of the potential budget constraint for all individuals.

The data spans the period Spring 1996 – Spring 2003. We drop Summer 1999 – Spring 2000 inclusive as this covered the period when the WFTC policy was introduced. We also drop individuals aged over 45. The outcome variable reported in Table 12.3 is the employment rate expressed as a percentage. The matching covariates include age, education, region and ethnicity. Overall, Table 12.3 points to a 3.5 to 4 percentage point increase in single mothers' labor supply attributable

Table 12.3. Difference-in-Differences Results

Single women	Marginal effect	Standard error	Sample size
Family resources survey	3.57	0.81	74,959
Labour force survey	3.81	0.33	233,208

Data: Spring 1996 – Spring 2003.
Drop: Summer 1999 – Spring 2000 inclusive; individuals aged over 45.
Outcome: employment, Average impact × 100, employment percentage.
Matching covariates: age, education, region, ethnicity.
Source: Blundell et al. (2006).

to the WFTC policy. We also conducted a sensitivity analysis which considered alternative groups. For lower-education groups we found a slightly larger response from a lower base level of employment. These were the group most likely to be eligible to larger transfers under the WFTC policy. We also examined sensitivity to the choice of pre-treatment years. The results are robust to changing the pre-treatment time window and also choice of "hypothetical" reform on pre-reform years. For example, a hypothetical reform in spring 1997 would yield an impact effect of .07 (.11).

Next we turn to the structural analysis and the validation of the structural model. It should be noted that even if the difference-in-differences assumptions were not valid, the structural model will allow us to simulate the moments underlying the difference-in-differences estimator anyway and this in turn would still provide a validation of the structural model. The underlying variation which is used to identify the structural model comes from variation across location and time in taxes and benefits as well as an exploitation of the precise rules of the tax and benefit system. Specifically variation in housing costs and local taxation provides useful variation in the potential budget constraints across individuals in our sample. Of course this requires that individuals use this information in the same way as we do in constructing their counterfactual budget constraints.

There are a number of key features of the structural model.[17] There is the budget constraint which reflects tax and benefit interactions as well as take-up. There are also the modelling of preferences with discrete hours choices. Heterogeneity is allowed by demographic and ethnic group as well as a broad set of unobserved heterogeneity. The model pays particular attention to fixed costs of work, to the specification of stigma/hassle costs and to childcare costs.

The overall stochastic specification is a mixed-multinomial specification across discrete choices over ranges of hours. It builds directly on the work of Hoynes (2000) and Keane and Moffitt (1998). Individuals are assumed to maximize their utility subject to a budget constraint, determined by a fixed hourly wage and the tax and benefit system. The utility function is approximated with a second degree polynomial expansion in hours of work and net income with unobserved heterogeneity. In addition to preference heterogeneity in the marginal rate of substitution between work and consumption, the model allows for unobserved heterogeneity in program participation "costs", childcare costs and fixed costs of work.

Given the considerable non-convexities in the budget constraint generated by the tax and transfer system, individuals are assumed to choose from a small subset of hours corresponding to the hours ranges 0, 1–15, 16–22, 23–29, 30–36 and 37+ respectively. Blundell and MaCurdy (1999) give the arguments for modelling labor supply with a discrete choice model: the main advantage is that it easily permits the highly non-convex budget constraints created by welfare benefits and in-work support. For each choice of hours, there is also an additive stochastic component on the utility of each hours choice assumed to follow a standard (Type-I) extreme value distribution. This assumption is common, see Blundell et al. (2000b) and Keane and Moffitt (1998). Van Soest, Das and Gong (2002) discuss some possible interpretations of the errors (unobserved alternative-specific utility components, or errors in perception of the alternatives' utilities, for example); their main advantage is in providing positive probabilities that are continuous in the parameters.

Empirically, a number of studies have shown that estimating labor supply models without unobserved work-related costs is more likely to lead to estimates of preferences that are non-convex; conversely, allowing for work-related costs tends to lead to estimates of preferences which are convex (see references in Heim and Meyer 2004). Inferring parents' labor supply preferences from observed behavior without considering childcare is likely to lead to biased conclusions. As WFTC provides financial support for formal childcare costs for families where all adults are working, evaluating the impact of WFTC on labor supply requires us to specify the childcare costs of working parents. The model allows explicitly for childcare costs, assuming a linear relationship between hours of childcare per child and hours of work h. This relationship is allowed to vary with the number and age of children.

Unobserved heterogeneity enters in several places. Through the take-up or program participation cost, the childcare expenditure costs, and the fixed work-related costs. Unobserved heterogeneity also affects preferences directly through the linear income and hours terms in the quadratic utility specification. In estimation, the integrals in the log-likelihood are approximated using simulation methods (see Blundell, Brewer and Shephard 2006), integrating out the random preferences by drawing a number of times from the distribution, and computing the mean pseudo-likelihood across these realizations. The unobserved preference heterogeneity terms are assumed independently normally distributed, and we approximate the distribution of childcare prices with 6 discrete mass points. The estimated parameter values for the model are broadly consistent with economic theory. In particular, for lone parents 99.0% have positive marginal utility of net income at their observed state. Overall elasticities line up quite well too. With an average extensive elasticity of .81 (.13) and an intensive elasticity of .31 (.09).

The first step in simulating the WFTC policy reform is to estimate a choice probability distribution (over the combination of hours and program participation) for each individual under a given tax and transfer system: we do this by numerically averaging over the unobserved components in the model. To simulate the impact of a change in the tax and benefit system, the same numerical draws are used to compute the choice probabilities under both tax and benefit systems, and combine these into a matrix of transition probabilities over the choices. This gives the (estimated) expected value of the transition matrix given the parameter estimates, where the expectation is over all random components. Confidence intervals around these expectations are estimated by the bootstrap.

As noted above one important aspect of the WFTC reform was the accompanying increase in out-of-work incomes for families – income support. In fact, unlike with similar expansions of the EITC in the US there were almost matched increases in the generosity of income support for families with children. For single parents the WFTC reform did unambiguously increase the incentive to work. However, together with the interactions with other benefits outlined above, this considerably dampens the underlying incentive to work. For this reason we might expect relatively small impact measures.

Two main reforms to the tax and benefit system are simulated. Table 12.4a presents the effect on labor supply of moving from Family Credit to Working Families Tax Credit, holding all other things equal. It turns out to be important to disaggregate the simulation results according to the age of the youngest child.

To compute these impact transitions effects requires integrating over the unobserved heterogeneity in the structural model. In Table 12.4b we present the same transition table but for *all* reforms directed to

Table 12.4. a) Structural Evaluation Results: WFTC Expansion Alone by Age of Youngest Child

	All	y-child			
		0 to 2	3 to 4	5 to 10	11 to 18
Change in employment rate:	5.95	3.09	7.56	7.54	4.96
	0.74	0.59	0.91	0.85	0.68
Average change in hours:	1.79	0.71	2.09	2.35	1.65
	0.2	0.14	0.23	0.34	0.2

Source: Blundell, Brewer and Shephard (2006).

Table 12.4. b) Structural Evaluation Results: All Reforms by Age of Youngest Child

	All	y-child			
		0 to 2	3 to 4	5 to 10	11 to 18
Change in employment rate:	3.86	0.65	4.53	4.83	4.03
	0.84	0.6	0.99	0.94	0.71
Average change in hours:	1.02	0.01	1.15	1.41	1.24
	0.23	0.21	0.28	0.28	0.22

Notes: Simulated of FRS data; Standard errors in italics.
Source: Blundell, Brewer and Shephard (2006).

lone mothers and introduced during the WFTC reform period. The increases in participation are systematically lower compared to when WFTC was considered alone, and this is likely to reflect that the contemporaneous increases in income support dulled the positive labor supply impact of WFTC. This confirms that it is not the elasticities that are unusually small for the British case but simply the interactions with other taxes and benefits and the coincidence of off-setting reforms to those benefits accessible to lone parents who do not work.

Note that the simulated difference-in-differences parameter from the structural evaluation model does not differ significantly from the difference-in-differences estimate in Table 12.3. A difference-in-differences methodology cannot identify the labor market impact of WFTC alone because other taxes and benefits changed at the same time as its introduction. Comparing the simulated moment for the all reforms case with difference-in-differences moment the difference yields a p-value of .42. The simulated difference-in-differences parameter from the structural evaluation model is precise and does not differ significantly from the difference-in-differences estimate. We find similar results for comparisons with low education groups. So it appears that the structural model does present a reasonably accurate description of responses to the reform. As argued above the small effects of the reform are due to interaction of WFTC with other taxes and benefits and the rise in family allowances (all reforms) – which are given without a work condition, rather than "small" response elasticities.

12.5.2 *Is the Design Optimal?*

The structural labor supply model provides an appropriate framework for considering problems related to the optimality of the tax schedule. Using parameter estimates from a structural model of labor supply, such as that presented above, the labor supply behavior of individuals can be simulated as the parameters of the tax and transfer system are varied. With these endogenous and heterogeneous labor supply responses allowed for, the structural model provides all the necessary information to maximize any well-behaved social welfare function, subject to a government budget constraint.

Imagine that we want to redistribute some specific sum to low skilled lone parents. We can turn to a Mirrlees optimal tax computation and ask given the implied elasticities at extensive and intensive margin corresponding to estimated structural model, is the WFTC design "optimal" for reasonable social welfare weights? Rather than using the Saez (2002) approximation, Blundell, Brewer and Shephard (2006) work directly with the structurally estimated preferences and choose the optimal allocation by maximizing a welfare function that depends on the distribution of tastes and budget constraints.

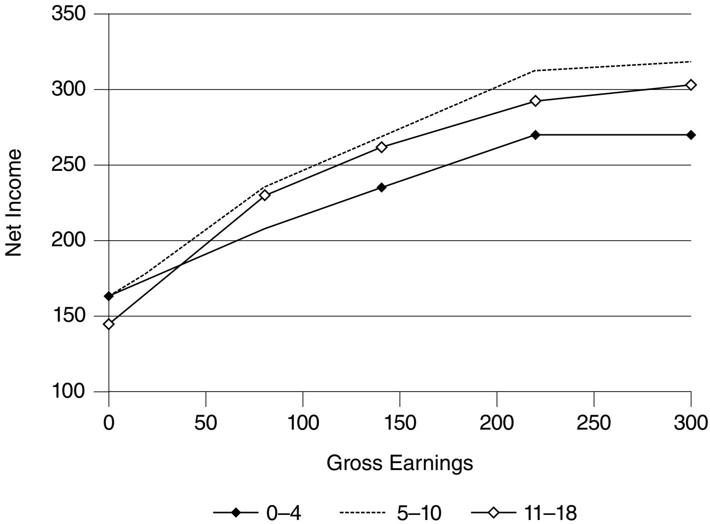
In this analysis, the social welfare function is given by the sum of individual (transformed) utilities, with the utility transformation function determining the governments relative preference for the equality (or otherwise) of utilities. We set the social welfare transformation function $\Gamma(U|\theta)=\frac{1}{\theta}\{(\exp U)\theta-1\}$. When θ is negative, the function favours the equality of utilities; when it θ is positive the reverse is true. θ = 0 corresponds to the linear case. The government then maximizes social welfare choosing a tax schedule.

The tax schedule is such that each individual chooses their hours of work to maximize their utility and the government satisfies its budget constraint. Conditioning on demographic characteristics the tax schedule will be parameterized by a level of out-of-work income (income support), and the different marginal tax rates. In this analysis four marginal tax rates are chosen corresponding to weekly earnings over five regions.[18] The parameters of the optimal tax schedule will be a function of demographic characteristics, the distributions of wages, preference parameters, preference errors, and the social welfare transformation function. Marginal tax rates are restricted to lie between –100% and 100%. Given flexible preferences and state specific errors, a non-decreasing budget constraint is a restriction, rather than a necessary consequence of the utility maximization process. The problem is particularly numerically intensive to solve and a grid in the parameter space is constructed to calculate the preferred labor market choice of each individual using the estimated structural model. The results of these simulations then allow the evaluation of both the government budget constraint and social welfare function(s). The government budget constraint is also used to restrict the parameter search space. The feasible set therefore contains many pair wise tax schedules that differ only in the marginal rate applied to the highest level of earnings. Blundell, Brewer and Shephard (2006) solve the schedule for parameter values $\theta = \{-.2, 0.0, 0.2\}$. Here we simply present those for –.2 which, as we will see, accords closest with the implicit weights underlying tax credit policy in the UK.

The optimal tax schedule is solved separately for three different groups on the basis of the age of youngest child. For each of these groups the value of government expenditure is set equal to the actual expenditure on this group within our sample. Conditioning upon this level of expenditure (which implicitly represents a preference for the relative welfare of the different groups) the tax schedule that maximizes social welfare is calculated.

Some initial results from this analysis are presented in Figure 12.9a and 12.9b and show some clear and interesting conclusions. The schedules are drawn over the five ranges of earned income described above. In all cases the welfare function has $\theta = -.2$ and therefore displays mild inequality aversion. As the age of the youngest child increases there is a shift toward relatively more support in-work than out of work. This can be interpreted as an increase in the extensive elasticity or a decline in the value of time spent at home by the lone parent. A comparison of the optimal constraint between a lone parent who has a child aged 0–4 and one whose child is aged 11–18 is probably most stark. The slope at low earned incomes is sharply steeper for the lone parent with the older child.

In all cases presented in Figure 12.9a the welfare weights on incomes are set with $\theta = -.2$ and the weights decline with income monotonically. These weights, displayed in Figure 12.9b are computed directly from the minimized social welfare function. These weights are reasonable and compare with the type of weights used in the Immervol et al. (2007) study. In Blundell, Brewer and Shephard (2006) a variety



a) Optimal Tax Schedules for Lone Parents by Age of Youngest Child

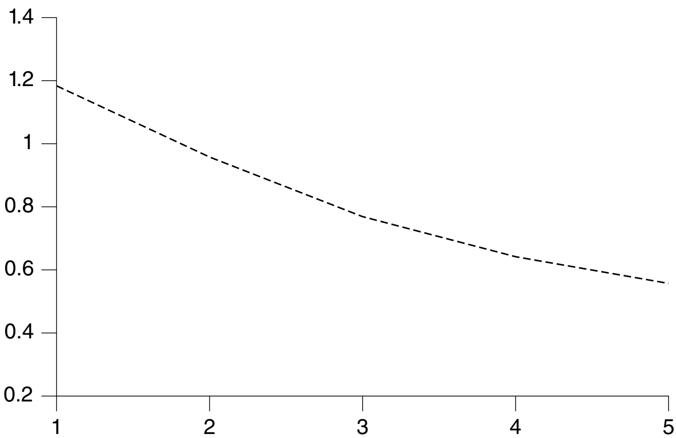


Figure 12.9. b) Social Welfare Weights for θ=–.2 for Lone Parent with Child Less than 4 Years Old

Source: Blundell, Brewer and Shephard (2006).

of θ values are considered. Even with $\theta = 0$ the overall conclusions in terms of optimal design remain.

Figure 12.10 derives the optimal constraint for a specific type of lone parent, defined by hourly wage, housing costs and childcare costs. Against this is plotted the actual WFTC policy with all its interactions

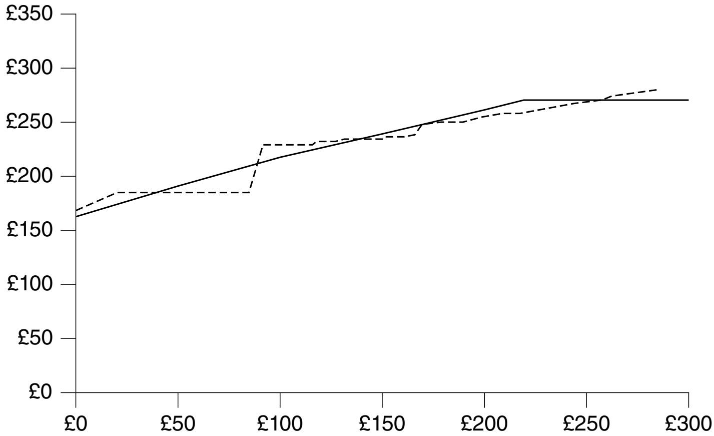


Figure 12.10. Optimal Tax Schedules in a Specific Case Relative to WFTC
Source: Blundell, Brewer and Shephard (2006).

with the rest of the tax and benefit system in the UK. Remarkably the optimal tax function and the WFTC constraint show a degree of similarity. Suggesting that the WFTC policy for social welfare weights with $\theta = -.2$ may well be an optimal design. A consequence of Figure 12.9a is that the existing WFTC policy probably does not contain a sufficiently strong tax credit for lone parents with older children. Another implication of this analysis is that the WFTC policy itself, without the simultaneous increases in Income Support could only have been optimal with much lower weight on redistribution. The comparison with the expansions in generosity of the EITC in the US suggest that, although the EITC expansion provided much of the motivation for the WFTC policy in the UK, the implicit social welfare weights were much more redistributive in the UK than those implicit in US welfare policy toward lone parents.

12.6 Conclusions: Designing a Welfare-to-Work Policy Using Tax Credits

This lecture has drawn on a new line of research that transcends the boundaries of labor economics and public finance. The aim has been to evaluate the responses to Employment Tax Credit reforms and the optimal design of such reforms. Specifically, a comparison of the reforms in the UK and the US. To gauge the optimality we need a specific set of treatment effect parameters from a structural model of economic responses. But structural models are fragile and I have argued for the need to validate them through comparison with experimental and quasi-experimental evaluations. I have also shown how a structural evaluation model with take-up and unobserved heterogeneity can provide a reasonably accurate description of labor supply behavior.

With empirically robust elasticities and knowledge of the full tax and benefit system we can easily reconcile EITC and WFTC "puzzle" – the smaller impact of the WFTC in the UK in comparison to its apparent generosity relative to the EITC policy in the US. We have also shown that empirically robust elasticities can easily justify an earned income tax credit policy even with social welfare weights that decline monotonically with income. This lines up well with the cross-country analysis in Immervol et al. (2007) and Eissa, Kleven and Kreiner (2004). Moreover, the UK reform is close to an optimal earned income tax credit policy, provided relatively high social welfare weights are placed on families with children.

There are many remaining questions concerning the adequacy of the empirical specification and the dimensions over which optimality is measured. The fact that the structural model has been shown to line up well with the quasi-experimental impact provides some comfort but what of more dynamic impacts. Cossa, Heckman and Lochner (2002) make a strong case for analyzing passive and active human capital responses as earned income tax credits place potentially important disincentives on human capital investment. This is surely deserving of further analysis and the broader dynamic benefits of encouraging work are often cited as important motivations for the expansion of earned income tax credit policies. Nonetheless, the evidence for some aspects of these dynamic effects is limited. The recent work by Gladden and Taber (2000) and Card and Hyslop (2005) find only small or insignificant impacts of work experience on wages.

Another important margin may well be the impact on fertility but the evidence for a significant impact of tax credits on fertility seems small, see Hoynes (1997a; 1997b). In line with the title of this lecture it seems appropriate to leave this last word on fertility to Adam Smith:

> *"Poverty, though it no doubt discourages, does not always prevent marriage. It seems even to be favourable to generation. A half-starved Highland woman frequently bears more than twenty children, while a pampered fine lady is often incapable of bearing any, and is generally exhausted by two or three. Barrenness, so frequent among women of fashion, is very rare among those of inferior station. Luxury in the fair sex, while it inflames perhaps the passion for enjoyment, seems always to weaken, and frequently to destroy altogether, the powers of generation."*
>
> Adam Smith (1776).

13

Employment, Hours of Work and the Optimal Taxation of Low-Income Families

Richard Blundell and Andrew Shephard

13.1 Introduction

This paper develops a structural approach to the optimal design of low-income support. The analysis concerns the optimal choice of the tax rate schedule in a Mirrlees (1971) framework extended to allow for unobserved heterogeneity, fixed costs of work, childcare costs and the detailed non-convexities of the tax and transfer system. Within this framework we consider Pareto improving reforms. We also explore the implications for the optimal tax rate schedule under social welfare functions with different degrees of inequality aversion.

The tax treatment of lone parents in the UK is used as the empirical environment for our analysis.[1] As in North America, this group has been the subject of a number of tax and benefit reforms, see Blundell and Hoynes (2004). These reforms can provide useful variation for assessing the reliability of structural models. In particular, we use the 1999 Working Families Tax Credit (WFTC) reform in the UK which considerably increased the generosity of in-work benefits/tax credits for lone parents, see Brewer (2001). The WFTC program uses hours-contingent

Bundell, R., Shephard, A. (2012): Employment Hours of Work and the Optimal Taxation of Low-Income Families, in: Review of Economic Studies, 79(2): 481–510. © 2012 by permission of Oxford University Press. We thank Stuart Adam, Mike Brewer, Roger Gordon, Guy Laroque, Ian Preston, Emmanuel Saez, Florian Scheuer, seminar participants at Harvard, MIT, Jerusalem, NIESR, SOLE and participants in the Mirrlees Review, the editor and the referees for helpful comments. We are alone responsible for all errors and interpretations.

payments.[2] Eligibility requires parents to be working in a job with at least 16 hours of work per week.

Designing low-income support is complicated. How should taxes and transfers depend on income when taking into account the labor supply responses for this group involving both intensive and extensive margin responses? Tagging has been suggested to improve the trade-off between equality and efficiency but how large are the potential gains? Hours are partially observable and used in practise for low-income support but how good is this “signal”? Optimal tax theory points to the relevant trade-offs but we need solid measurement of these trade-offs in order to move from theory to practical policy recommendations on how to reform actual tax schedules. The paper bridges the gap by setting up a structural model that is able to address all of these questions.

The microeconometric analysis here is based on an extension of the stochastic discrete choice labor supply approach (Hoynes 1996; Keane and Moffitt 1998; Blundell et al. 2000b; van Soest, Das and Gong 2002). This approach allows us to distinguish between the intensive margin of hours of work and the extensive margin where the work decision is made. As the empirical literature on labor supply has demonstrated, labor supply elasticities for certain groups of working age individuals appear to be much larger at the extensive margin, see Blundell and Macurdy (1999), for example. As Saez (2002) and Laroque (2005) have shown, empirical results on the responsiveness of different types of individuals at different margins of labor supply have strong implications for the design of earnings taxation.

Consistent with the empirical literature on the labor supply of the low paid, our structural estimation results show important differences in the responsiveness of labor supply at different margins. We use our estimated model to identify inefficiencies in the actual tax and transfer system and characterize Pareto improving reforms. This analysis points to relatively minor improvements in the tax schedule for lone parents. When imposing a social welfare function with reasonable social welfare weights, we obtain a reformed non-linear tax schedule with lower tax rates over a large range of earnings for many families, and with tax credits only optimal for low earners.

We also find that labor supply responses vary according to the age of children. We use this variation to quantify the potential welfare gains from tagging according to child age. Our results suggest a welfare improving role for age-based tagging, with tax credits being found to be most important for low-earning families with school age children. Our results also point to welfare gains from using hours-contingent payments. If the tax authorities are able to choose the lower limit on working hours that trigger eligibility for such families, we present an empirical case for using a full-time work rule rather than the main part-time rule currently in place for parents in the UK. However, the case is substantially mitigated when hours cannot be monitored or recorded accurately by the tax authorities.

The remainder of the paper proceeds as follows. In the next section we develop the analytical framework for optimal design within a stochastic structural labor supply model. In Section 13.3 we outline the WFTC reform in the UK and its impact on work incentives. Section 13.4 details the structural microeconometric model, while in Section 13.5 we describe the data and model estimates. Section 13.6 uses these model estimates to explore what normative conclusions may be derived from a weak Pareto improvement criteria. In Section 13.7 we then consider what additional results may be derived by imposing a specific social welfare function; we also demonstrate how these vary when we allow for tagging by age of children, and when hours of work are included in the tax base. Finally, Section 13.8 concludes.

13.2 The Optimal Design Problem

In this section we develop the analytical framework that will be used to explore both Pareto improving reforms to the actual tax and transfer system, as well as tax reforms that are optimal under some social welfare function.[3] In both cases we will allow for two scenarios. In the first only earnings and employment are observable by the tax authority, in the second we also allow for partial observability of hours of work. Hours of work h are chosen from the finite set $H = \{h_0, \ldots h_J\}$, with partial observability incorporated by allowing the tax authorities to additionally observe that hours belong to some closed interval $h = [\underline{h}, \overline{h}] \in H$ *with* $\underline{h} \leq h \leq \overline{h}$. For example, the tax authorities may be able to observe whether individuals are working at least h_B hours per week, but conditional on this, not how many.[4]

Work decisions by individuals (single mothers) are determined by their preferences over consumption c and labor hours h, as well as possible childcare requirements, fixed costs of work, and the tax and transfer system. Preferences are indexed by observable characteristics X, including the number and age of her children, and vectors of unobservable (to the econometrician) characteristics ϵ and ε. The vector ε is independent of both X and ϵ and corresponds to the additive hours (or state) specific errors in the utility function; we let $U(c, h; X, \epsilon, \varepsilon) =$

$u(c, h; X, \epsilon) + \varepsilon_h$ represent the utility of a single mother who consumes c and works h hours. We will assume that she consumes her net income which comprises the product of hours of work h and the gross hourly wage w plus non-labor income and transfer payments, less taxes paid, childcare expenditure, and fixed costs of work. In what follows we let F denote the cumulative distribution function of the state specific errors ε, and let G denote the joint cumulative distribution function of X and ϵ. We assume that ε is independent of both ϵ and X.

In our empirical analysis individual utilities $U(c, h; X, \epsilon, \varepsilon)$ will be described by a parametric utility function and a parametric distribution of unobserved heterogeneity (ϵ, ε). Similarly, a parametric form will be assumed for the process determining fixed costs of work and childcare expenditure. To maintain focus on the design problem, we delay this discussion regarding the econometric modelling until section 4. For now it suffices to write consumption c at hours h as $c(h; T, X, \epsilon)$[5] where $T(wh, h, X)$ represents the tax and transfer system. Non-labor income, such as child maintenance payments, enter the tax and transfer schedule T through the set of demographics X, and for notational simplicity we abstract from the potential dependence of the tax and transfer system on childcare expenditure. Taking T as given, each single mother is assumed to choose her hours of work $h^* \in H$ to maximize her utility. That is:

$$h^* = arg\max_{h \in H} U(c(h; T, X, \epsilon), h; \epsilon, \varepsilon). \tag{1}$$

13.2.1 *Pareto Improving Reforms*

The first stage of our optimal design analysis explores the normative conclusions that may be derived from a Pareto improvement criterion. This exercise is closely related to Werning (2007), which characterized the set of Pareto efficient tax schedules within the Mirrlees (1971) model, and which proposed a simple test for the efficiency of a given tax schedule through the lens of that model.

To explore the efficiency of a given tax and transfer system T_e we first calculate the (incentive compatibility) maximized value of utility under this system for all $(X, \epsilon, \varepsilon)$. With slight abuse of our above notation, we denote these maximized utility levels as $U(T_e, X, \epsilon, \varepsilon)$. We then consider reforms to T_e by constructing the composite schedule $T_p = T_e + T_d$. While T_e accurately reflects the full heterogeneity in the actual system we will restrict ourselves to reforms where T_d belongs to a particular parametric class.

The parameters of T_d are chosen to maximize the revenue of the government $R(T_d)$:

$$R(T_d) = \int_{X,\epsilon}\int_{\varepsilon}\left[T_e(wh^*, h^*; X) + T_d(wh^*, h^*; X)\right]dF(\varepsilon)dG(X,\epsilon) \qquad (2)$$

subject to the requirement that each individual is at least as well off as the actual tax and transfer system T_e:

$$U(T_e + T_d, X, \epsilon, \varepsilon) \geq U(T_e, X, \epsilon, \varepsilon)\,\forall(X, \epsilon, \varepsilon), \qquad (3)$$

and where $U(T_e + T_d, X, \epsilon, \varepsilon)$ denotes maximized utility under the reformed system. If revenue is not maximized under T_e then it cannot be Pareto efficient, since it would be possible to reform the system in a direction which, by raising revenue, allows the welfare of some individuals to be improved without harming others. Note that Pareto improvements in this setting require *reductions* in tax schedules.

13.2.2 *Social Welfare Improving Reforms*

The second stage of our policy analysis maintains the same positive aspects as described above, but introduces an alternative normative framework. It concerns the choice of a tax schedule T in which the government is allocating a fixed amount of revenue R to a specific demographic group in a way which will maximize the social welfare for this group. Such a schedule balances redistributive objectives with efficiency considerations. Redistributive preferences are represented through the social welfare function W, defined as the sum of transformed individual utilities:

$$W(T) = \int_{X,\epsilon}\int_{\varepsilon} \Upsilon(c(h^*; T, X, \epsilon), h^*; X, \epsilon, \varepsilon)dF(\varepsilon)dG(X,\epsilon) \qquad (4)$$

where for a given cardinal representation of U, the utility transformation function Υ determines the governments relative preference for the equality of utilities. This maximization is subject to the incentive compatibility constraint which states that lone mothers choose their hours of work optimally given T (as in Equation 1) and the government resource constraint:

$$\int_{X,\epsilon}\int_{\varepsilon} T(wh^*, h^*; X)dF(\varepsilon)dG(X,\epsilon) \geq \bar{T}(\equiv -R). \qquad (5)$$

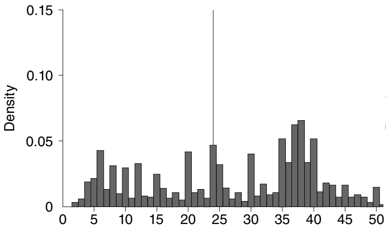
As in our exploration of Pareto improving reforms, we will restrict T to belong to a particular parametric class of tax functions. This is discussed in Section 13.7 when we empirically examine the optimal design of the UK tax and transfer schedule.

13.3 Tax Credit Reform and Low-Income Support

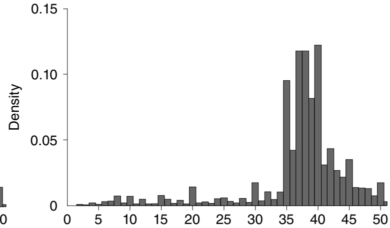
The increasing reliance on tax credit policies during the 1980s and 1990s, especially in the UK and the US, reflected the secular decline in the relative wages of low-skilled workers with low labor market attachment together with the growth in single-parent households (see Blundell 2002, and references therein). The specific policy context for this paper is the Working Families Tax Credit (WFTC) reform which took place in the UK at the end of 1999. A novel feature of the British tax credit system is that it makes use of minimum hours conditions in addition to an employment condition. Specifically, WFTC eligibility required a working parent to record at least 16 hours of work per week. Moreover, there was a further hours contingent bonus for working 30 hours or more.

As in the US, the UK has a long history of in-work benefits, starting with the introduction of Family Income Supplement (FIS) in 1971. In 1988 FIS became Family Credit (FC), and in October 1999 this was replaced by Working Families Tax Credit. While these programs have maintained a similar structure, the reforms have been associated with notable increases in their generosity. As described above, an important feature of British programs of in-work support since their inception – and in contrast with programs such as the US Earned Income Tax Credit – is that awards depend not only on earned and unearned income and family characteristics, but also on a minimum weekly hours of work requirement. Originally set at 24 hours per week, this was reduced to 16 hours per week in April 1992, where it has stayed since (an additional but smaller credit at 30 hours was introduced in 1995). The impact of this reform to FC on single parents' labor supply is ambiguous: those working more than 16 hours a week had an incentive to reduce their weekly hours to (no less than) 16, while those previously working fewer than 16 hours had an incentive to increase their labor supply to (at least) the new cut-off. Figure 13.1 shows that the pattern of observed hours of work over this period strongly reflects these incentives. Single women without children were ineligible.

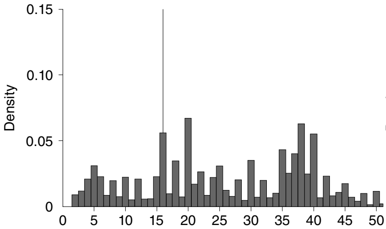
The tax design problem we discuss here draws directly on some of the key features of the WFTC. Indeed, we assess the reliability of our structural labor supply model by its ability to explain behavior before and after the WFTC reform. The WFTC reform increased the attractiveness of working 16 or more hours a week compared to working fewer hours, and the largest potential beneficiaries of WFTC were those families who were just at the end of the FC benefit withdrawal taper. Conditional on working 16 or more hours, the theoretical impact of



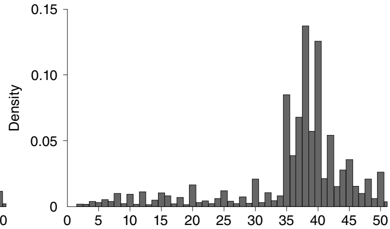
a) Lone Mothers, 1991



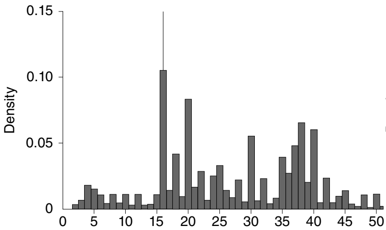
b) Single Women, 1991



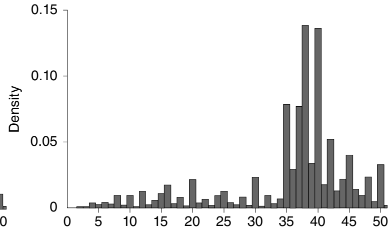
c) Lone Mothers, 1995



d) Single Women, 1995



e) Lone Mothers, 2002



f) Single Women, 2002

Figure 13.1. Female Hours of Work by Survey Year

Notes: Figure shows the distribution of usual hours of work for women by year and presence of children. Sample is restricted to women aged 18–45. Calculated using UK Labour Force Survey data (for 1991) and UK Quarterly Labour Force Survey data (1995 and 2002). Horizontal axes measure weekly hours of work; the vertical line indicates the minimum-hours eligibility.

WFTC is as follows: (i) people receiving the maximum FC award face an income effect reducing labor supply, but not below 16 hours a week; (ii) people working more than 16 hours and not on maximum FC will face an income effect away from work (but not below 16 hours a week), and a substitution effect towards work; (iii) people working more than 16 hours and earning too much to be entitled to FC but not WFTC will face income and substitution effects away from work if they claim WFTC (see Blundell and Hoynes 2004). The main parameters of FC and WFTC are presented in the Supplementary Material for this paper (Blundell and Shephard 2012).

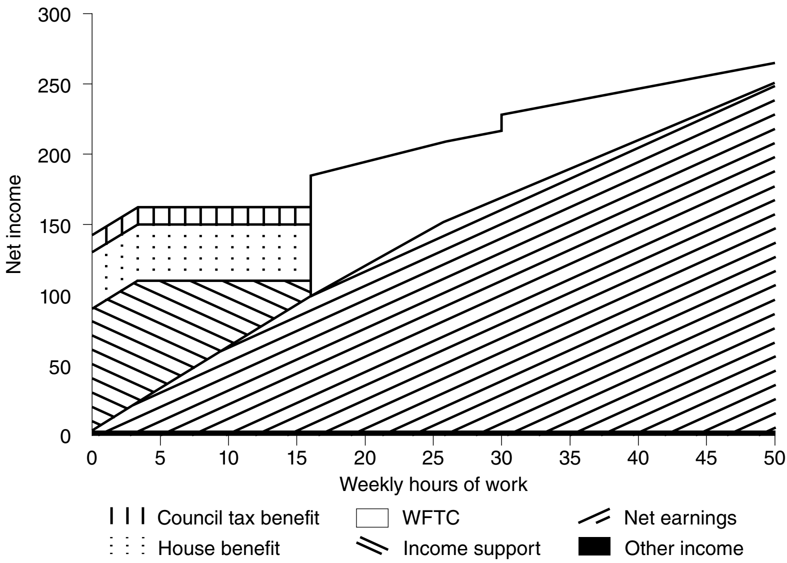


Figure 13.2. Tax and Transfer System Interactions

Note: Figure shows interaction of tax and transfer system under April 2002 system for a lone parent with a single child aged 5, average band C council tax, £40 per week housing costs, £6 gross hourly wage rate, and no childcare costs. All incomes expressed in April 2002 prices. Calculated using FORTAX.

When analyzing low-income support we take an integrated view of the tax system. This is because tax credit awards in the UK are counted as income when calculating entitlements to other benefits, such as Housing Benefit and Council Tax Benefit. Families in receipt of such benefits would gain less from the WFTC reform than otherwise equivalent families not receiving these benefits; Figure 13.2 illustrates how the various policies impact on the budget constraint for a low wage lone parent. Moreover, there were other important changes to the tax system affecting families with children that coincided with the expansion of tax credits, and which make the potential labor supply responses considerably more complex. In particular, there were increases in the generosity of Child Benefit (a cash benefit available to all families with children regardless of income), as well as notable increases in the child additions in Income Support (a welfare benefit for low-income families working less than 16 hours a week).[6]

13.4 A Structural Labor Supply Model

The labor supply specification develops from earlier studies of structural labor supply that use discrete choice techniques and incorporate non-participation in transfer programs, specifically Hoynes (1996) and Keane and Moffitt (1998). Our aim is to construct a credible model of labor supply behavior that adequately allows for individual heterogeneity in preferences and can well describe observed labor market outcomes. As initially discussed in section 2, lone mothers have preferences defined over consumption c and hours of work h. Hours of work h are chosen from some finite set H, which in our main empirical results will correspond to the discrete weekly hours points $H = \{0, 10, 19, 26, 33, 40\}$. These hours points correspond to the empirical hours ranges 0, 1–15, 16–22, 23–29, 30–36 and 37+ respectively. In the supplementary material we discuss the sensitivity of our results to a finer discretisation of weekly hours; our main results appear robust to this.

13.4.1 *Preference Specification*

We augment the framework presented in section 2 to allow the take-up of tax credits to have a direct impact on preferences through the presence of some stigma or hassle cost, and we use P (equal to one if tax credits are received, zero otherwise) to denote the endogenous take-up (program participation) decision. The utility function is now given by:

$$U(c,h,P;X,\epsilon,\varepsilon) = u(c,h,P;X,) + \varepsilon_h,$$

with these preferences allowed to vary with observable demographic characteristics X, and vectors of unobservable (to the econometrician) characteristics ϵ and ε. The state specific errors ε that are attached to each discrete hours point are assumed to follow a Type-I extreme value distribution.

All the estimation and simulation results presented here assume preferences of the form:

$$u(c,h,P;X,\epsilon) = \alpha_y(X,\epsilon)\frac{c^{\theta_y}-1}{\theta_y} + \alpha_1(X)\frac{\left(1-\frac{h}{H}\right)^{\theta_t}}{\theta_t} - P\eta(X,\epsilon) \tag{6}$$

where $H = 168$ denotes the total weekly time endowment, and where the set of functions $\alpha_y(X, \epsilon)$, $\alpha_l(X)$ and $\eta(X, \epsilon)$ capture observed and unobserved preference heterogeneity. The function $\eta(X, \epsilon)$ is included to reflect the possible disutility associated with claiming in-work tax credits ($P = 1$), and its presence allows us to rationalize less then complete take-up of tax credits. In each case we allow observed and unobserved heterogeneity to influence the preference shifter functions through appropriate index restrictions. We parameterize these as:

$$\log \alpha_y(X,\epsilon) = X_y'\beta_y + \epsilon_y$$
$$\log \alpha_l(X) = X_l'\beta_l$$
$$\eta(X,\epsilon) = X_\eta'\beta_\eta + \epsilon_\eta.$$

The sets of included demographics are described in Section 13.5.2.

13.4.2 *Budget Constraint, Fixed Costs of Work and Childcare Costs*

Individuals face a budget constraint, determined by a fixed gross hourly wage rate (generated by a log-linear relationship of the form $\log w = X'_w\beta_w + \epsilon_w$) and the tax and transfer system $T(wh, h, P; X)$. Non-labor income, such as child maintenance payments, enter the budget constraint through the dependence of the tax and transfer schedule T on demographic characteristics X.

We arrive at our measure of consumption c by subtracting both childcare expenditure and fixed work related costs from net income, $wh - T(wh, h, P; X)$.[7] Both of these processes are described in detail below. Essentially, the choice of work hours will affect consumption through three main channels: firstly, through its direct effect on labor market earnings and its interactions with the tax and transfer system; secondly, through fixed working costs which are payable only if hours of work are strictly positive; thirdly, since working mothers may be required to purchase childcare for their children which varies with maternal hours of employment.

13.4.2.1 FIXED COSTS OF WORK

Fixed work-related costs (as in Cogan 1981) help provide a potentially important wedge that separates the intensive and extensive margin. They reflect the actual and psychological costs that an individual has to pay to get to work. We model these work-related costs $\alpha_f(h; X)$ as a fixed, one-off, weekly cost subtracted from net income at positive values of working time. We parameterize fixed costs as:

$$\alpha_f(h; X) = 1(h > 0) \times X_f'\beta_f$$

where $1(\cdot)$ is the indicator function.

13.4.2.2 CHILDCARE EXPENDITURE

Given the rather limited information that our data contains on the types of childcare use, we take a similarly limited approach to modelling, whereby hours of childcare use h_c is essentially viewed as a constraint: working mothers are required to purchase a minimum level of childcare $h_c \geq \alpha_c(h, X, \epsilon)$ which varies stochastically with hours of work and demographic characteristics. Since we observe a mass of working mothers across the hours of work distribution who do not use any childcare, a linear relationship (as in Blundell et al. 2000b) is unlikely to be appropriate. Instead, we assume the presence of some underlying latent variable that governs both the selection mechanism and the value of required childcare itself. More specifically, we assume that the total childcare hours constraint is given by:

$$\alpha_c(h, X, \epsilon) = 1(h > 0) \times 1(\epsilon_c > -\beta_c h - \gamma_c) \times (\gamma_c + \beta_c h + \epsilon_c). \tag{7}$$

In our empirical application we will allow all the parameters of this relationship to vary with the set of observable characteristics X_c. Total weekly childcare expenditure is then given by $p_c h_c$ with p_c denoting the hourly price of childcare. Empirically, we observe a large amount of dispersion in childcare prices, with this distribution varying systematically with the age composition of children. This is modelled by assuming that p_c follows some distribution $p_c \sim F_c(\cdot; X_c)$ which again varies with demographic characteristics.

13.4.3 *Optimal Individual Behavior*

The relationships described above allow us to write consumption at a given hours of work and program participation combination (h, P) as:

$$c(h, P; T, X, \epsilon) = wh - T(wh, h, P; X) - p_c \alpha_c(h, X, \epsilon) - \alpha_f(h; X) \tag{8}$$

which may then be substituted into the utility function in equation (6). Each single mother is assumed to jointly choose her hours of work and program participation decision to maximize her utility. Note that individuals may only be eligible to receive tax credits for some hours choices, and we use $E(h; X, \epsilon)$ to denote such eligibility (equal to one if eligible, zero otherwise). For given hours of work h eligible mothers will elect to receive tax credits if the utility gain from the associated higher consumption level exceeds the utility cost of claiming in-work tax credits. More formally, the optimal program participation decision $P^*(h)$ will be given by:

$$\begin{aligned} P^*(h) = 1 \text{ if } E(h; X, \epsilon) = 1 \text{ and } & u(c(h, P = 1; T, X, \epsilon), h, P = 1; X, \epsilon) \\ & \geq u(c(h, P = 0; T, X, \epsilon), h, P = 0; X, \epsilon) \\ & P^*(h) = 0 \text{ otherwise.} \end{aligned}$$

It then follows that the optimal (incentive compatible) choice of individual work hours $h^* \in H$ solves:

$$\max_{h \in H} U(c(h, P^*(h); T, X, \epsilon), h, P^*(h); X, \epsilon, \varepsilon).$$

13.5 Data and Estimation

13.5.1 *Data*

We use six repeated cross-sections from the Family Resources Survey (FRS), from the financial year 1997/8 through to 2002/3, which covers the introduction and subsequent expansion of WFTC. The FRS is a cross-section household-based survey drawn from postcode records across Great Britain: around 30,000 families with and without children each year are asked detailed questions about earnings, other forms of income and receipt of state benefits.

Our sample is restricted to lone mothers who are aged between 18 and 45 at the interview date, not residing in a multiple tax unit household, and not in receipt of any disability related benefits. Dropping families with missing observations of crucial variables, and those observed during the WFTC phase-in period of October 1999 to March 2000 inclusive, restricts our estimation sample to 7,090 lone mothers.

13.5.2 *Estimation*

The full model (preferences, wages, and childcare) is estimated simultaneously by maximum likelihood; the likelihood function is presented in Appendix A. This simultaneous estimation procedure contrasts with labor supply studies in the UK that have used discrete choice techniques. Perhaps largely owing to the complexity of the UK transfer system, these existing studies (such as Blundell et al. 2000b) typically pre-estimate wages which allows net incomes to be computed prior to the main preference estimation. In addition to the usual efficiency arguments, the simultaneous estimation here imposes internal coherency with regards to the various selection mechanisms. We incorporate highly detailed representations of the tax and transfer system using FORTAX (Shephard 2009). The budget constraint varies with individual circumstances, and reflects the complex interactions between the many components of the tax and transfer system. To facilitate the estimation procedure, the actual tax and transfer schedules are modified slightly to ensure that there are no discontinuities in net income as either the gross wage or childcare expenditure vary for given hours of work. We do not attempt to describe the full UK system here, but the interested reader may consult Adam and Browne (2006) and O'Dea, Phillips and Vink (2007) for recent surveys; see Shephard (2009) for a discussion of the implementation of the UK system in FORTAX.

The set of demographics characteristics contained in both X_y and X_l, and therefore affecting the marginal rate of substitution between consumption and leisure, are age, education (a zero-one dummy equal to one if the individual had completed compulsory schooling), number of children, and a series of dummies for age-of-youngest child (0–4, 5–10, and with 11–18 as the omitted category). Age and education also affect the cost of accessing tax credits through X_η, as does ethnicity and the currently operating tax credit program. We model this by including a zero-one dummy for the entire WFTC period, together with an additional variable to capture possible first year introductory effects. The wage equation regressors X_w comprise the age that education was completed, a polynomial in age, ethnicity, region, and a series of time dummies. Age, education, number of children, age-of-youngest child, ethnicity, and region, are all contained in X_f and so affect the fixed costs of work.

For the purpose of modelling childcare, we define six groups by the age of youngest child (0–4, 5–10, and 11–18) and by the number of children (1 and 2 or more). The stochastic relationship determining hours of required childcare $\alpha_c(h, X, \epsilon)$ varies within each of these groups, as does the childcare price distribution $F_c(\cdot; X_c)$. Using data from the entire sample period, the childcare price distribution is discretized into either four price points (if the youngest child is aged 0–4 or 5–10) or 2 points (if the youngest child is aged 11-18). In each case, the zero price point is included. The positive price points pc are fixed prior to estimation and correspond to the mid-points in equally sized groups amongst those using paid childcare (these values are presented in the supplementary material). The probability that lone mothers face each of these discrete price points is estimated together with the full model.

We impose concavity on the utility function by restricting the power terms θ_l and θ_y to be between 0 and 1. The unobserved wage component ϵ_w and the random preference heterogeneity terms (ϵ_y, ϵ_η, ϵ_c) are assumed to be normally distributed. Given the difficulty in identifying flexible correlation structures from observed outcomes (see Keane 1992), we allow ϵ_y to be correlated with ϵ_w, but otherwise assume that the errors are independent. The integrals over ϵ in the log-likelihood function are approximated using Gaussian quadrature with 11 nodes in each integration dimension. See Appendix A for further details.

13.5.3 *Specification and Structural Parameter Estimates*

The estimates of the parameters of our structural model are presented in the supplementary material. The results show that the age of the youngest child has a significant impact on the estimated fixed costs of work α_f; fixed work related costs are higher by around £16 per week if the youngest child is of pre-school age. The presence of young children also has a significant effect on the linear preference terms α_y (negatively) and α_l (positively). Parents with more children are also estimated to have a higher valuation for leisure, as well as higher fixed costs of work.

Lone mothers who are older are estimated to have a lower preference for both consumption and leisure, but higher costs of claiming in-work support. Meanwhile, the main impact of education comes through the preference for leisure α_l; mothers who have completed compulsory schooling have a lower preference for leisure. Ethnicity enters the model through both fixed costs of work and program participation costs η; we find that program participation costs are significantly higher for non-white lone mothers. These costs are found to fall significantly following the introduction of WFTC, although the reduction in the first year is small (as captured by the inclusion of a first year zero-one dummy variable). In contrast to many theoretical optimal tax studies which assume that preferences are quasi-linear in consumption, our estimate of θ_y places significant curvature on consumption. The estimate of θ_l is equal to the upper bound imposed so that estimated preferences are linear in leisure.

Both the intercept γ_c and slope coefficient β_c in the childcare equation are typically lower for those with older children. This reflects the fact that mothers with older children use childcare less, and that the total childcare required varies less with maternal hours of work. To rationalize the observed distributions, we also require a larger standard deviation σ_c for those with older children. As noted in section 5.2, the price distribution of childcare for each group was discretized in such a way that amongst those mothers using paid childcare, there are equal numbers in each discrete price group. Our estimates attach greater probability on the relatively high childcare prices (and less on zero price) than in our raw data. Individuals who do not work are therefore more likely to face relatively expensive childcare were they to work.

The hourly log-wage equation includes the age at which full-time education was completed (which enters positively), and both age and age squared (potential wages are increasing in age, but at a diminishing rate). Lone others who reside in the Greater London area have significantly higher wages, and the inclusion of time dummies track the general increase in real wages over time. There is considerable dispersion in the unobserved component of log-wages.

The within sample fit of the model is presented in Tables 13.1 and 13.2. The estimated model matches the observed employment states and the take-up rate over the entire sample period very well (see the first two columns of Table 13.1). We slightly under predict the number of lone mothers working 19 hours per week, and slightly over predict the number working either 26 or 33 hours per week, but the difference is not quantitatively large. Similarly, we obtain a very good fit by age of youngest child. The fit to the employment rate is encouraging, and the difference between predicted and empirical hours frequencies never differs by more than around three percentage points and is typically smaller. Furthermore, despite the relatively simple stochastic specification for childcare, our model performs reasonably well in matching both the use of childcare by maternal employment hours (both overall and by age of youngest child), and conditional hours of childcare. Full results are presented in the supplementary material.

Table 13.1. Predicted and Empirical Frequencies by Age of Youngest Child

	All		0–4		5–10		11–18	
	Predicted	Empirical	Predicted	Empirical	Predicted	Empirical	Predicted	Empirical
0 hour	0.549	0.550	0.704	0.708	0.490	0.489	0.319	0.320
	(0.005)	(0.006)	(0.007)	(0.008)	(0.009)	(0.010)	(0.012)	(0.013)
10 hours	0.078	0.068	0.063	0.049	0.090	0.083	0.086	0.081
	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)	(0.005)	(0.006)	(0.007)
19 hours	0.1D5	0.134	0.089	0.108	0.117	0.156	0.117	0.147
	(0.002)	(0.004)	(0.003)	(0.006)	(0.003)	(0.007)	(0.004)	(0.010)
26 hours	0.079	0.057	0.054	0.035	0.090	0.068	0.112	0.082
	(0.002)	(0.003)	(0.002)	(0.003)	(0.002)	(0.005)	(0.003)	(0.007)
33 hours	0.087	0.077	0.048	0.042	0.099	0.086	0.152	0.136
	(0.002)	(0.003)	(0.002)	(0.004)	(0.003)	(0.005)	(0.004)	(0.009)
40 hours	0.103	0.115	0.044	0.058	0.114	0.120	0.214	0.234
	(0.003)	(0.004)	(0.003)	(0.004)	(0.005)	(0.006)	(0.010)	(0.012)
Take-up	0.769	0.764	0.840	0.788	0.768	0.781	0.702	0.715
rate	(0.010)	(0.009)	(0.010)	(0.017)	(0.011)	(0.013)	(0.016)	(0.018)

Notes: Empirical frequencies calculated using FRS data with sample selection as detailed in Section 5.1. The discrete points 0, 10, 19, 26, 33 and 40 correspond to the hours ranges 0, 1–15, 16–22, 23–29, 30–36 and 37+ respectively. Empirical take-up rates calculated using reported receipt of FC/WFTC with entitlement simulated using FORTAX. Predicted frequencies are calculated using FRS data and the maximum likelihood estimates (presented in the Supplementary Material). Standard errors are in parentheses, and calculated for the predicted frequencies by sampling 500 times from the distribution of parameter estimates and conditional on the sample distribution of observables.

The fit of the model over time is presented in Table 13.2. Fitting the model over time is more challenging given that time enters our specification in a very limited manner through the wage equation and via the change in the stigma costs of accessing the tax credit. Despite this we are able to replicate the 9 percentage point increase in employment between 1997/98 and 2002/03 reasonably well with our model, although we do slightly underpredict the growth in part-time employment over this period.

To understand what our parameter estimates mean for labor supply behavior we simulate labor supply elasticities under the actual 2002 tax systems across a range of household types. All elasticities are calculated by simulating a 1% increase in consumption at all positive hours points. The results of this exercise are presented in Table 13.3. Across

Table 13.2. Predicted and Empirical Frequencies: 1997–2002

	1997		2002	
	Predicted	Empirical	Predicted	Empirical
0 hour	0.595	0.600	0.493	0.507
	(0.007)	(0.014)	(0.007)	(0.013)
10 hours	0.079	0.080	0.079	0.062
	(0.003)	(0.008)	(0.003)	(0.006)
19 hours	0.098	0.110	0.116	0.155
	(0.003)	(0.009)	(0.003)	(0.010)
26 hours	0.069	0.043	0.090	0.063
	(0.002)	(0.006)	(0.002)	(0.007)
33 hours	0.072	0.063	0.104	0.093
	(0.002)	(0.007)	(0.002)	(0.008)
40 hours	0.086	0.104	0.119	0.120
	(0.004)	(0.009)	(0.003)	(0.009)
Take-up rate	0.736	0.684	0.808	0.838
	(0.013)	(0.029)	(0.014)	(0.016)

Notes: See notes accompanying Table 13.1.

Table 13.3. Simulated Elasticities

	All		0–4		5–10		11–18	
	Uncomp.	Comp.	Uncomp.	Comp.	Uncomp.	Comp.	Uncomp.	Comp.
Participation	0.770	0.770	0.663	0.663	0.897	0.897	0.745	0.745
Intensive	0.042	0.123	0.032	0.094	0.043	0.128	0.047	0.136
Total hours	1.534	1.616	2.253	2.317	1.590	1.676	1.007	1.097

Notes: All elasticities simulated under actual 2002 tax systems with complete take-up of WFTC. Elasticities are calculated by increasing consumption by 1% at all positive hours choices. Participation elasticities measure the percentage point increase in the employment rate; intensive elasticities measure the percentage increase in hours of work amongst workers in the base system; total hours elasticities measure the percentage increase in total hours.

our sample of single mothers, we obtain an overall participation elasticity of 0.77, with our estimates implying a lower participation elasticity for single mothers whose youngest child is under 4 (an elasticity of 0.66), while they are significantly higher for mothers with school aged children (0.90 if youngest child is aged 5–10; 0.75 if the youngest child is aged 11–18). Intensive elasticities, which here measure the responsiveness of hours worked amongst employed single mothers to changes in in-work consumption, are small and are also increasing for parents with older children. Since mothers with older children also work more hours on average (see Table 13.1), these intensive elasticities also reflect larger increases in absolute hours for these groups. Compensated intensive elasticities are slightly higher. Finally, the total hours elasticities reported in the table combine these intensive and extensive responses.[8] Here, the lower employment rates for single mothers with younger children produces somewhat higher total hours elasticities for these groups.[9]

13.5.4 *Simulating the WFTC Reform*

Before we proceed to consider optimal design problems using our structural model, we first provide an evaluation of the impact of the WFTC reform described in section 3 on single mothers. This exercise considers the impact of replacing the actual 2002 tax systems with the April 1997 tax system on the 2002 population. This exercise is slightly different to simply examining the change in predicted states over this time period as it removes the influence of changing demographic characteristics.

The full results of this policy reform simulation are presented in the supplementary material. Overall we predict that employment increased by 5 percentage points as a result of these reforms, with the increase due to movements into both part-time and full-time employment. Comparing with Table 13.2 we find the reform explains over a half of the rise in employment over this period. The predicted increase in take-up of tax credits is also substantial, with this increase driven both by the changing entitlement and the estimated reduction in program participation costs.

13.6 Pareto Improving Reforms

In this section we use our structural model to examine the efficiency of the actual 2002 tax and transfer system T_e for single mothers with one child (and with complete take-up of tax credits). We will first restrict ourselves to reforms where the change in the tax schedule T_d is a function only of earnings *wh*; later we will also allow it to be a function of partially observed hours of work. To identify the regions where Pareto improvements are attainable we specify T_d as a flexible piecewise linear function of earnings. This schedule is characterized by a uniform change in out-of-work income, together with 21 different marginal tax rates. These marginal tax rates, which are restricted to lie between –100% and 100%, apply to weekly earnings from £0 to £500 in increments of £25, and then all weekly earnings above £500.

As described in Section 13.2.1 we search for the parameters of this schedule which maximize the revenue of the government, subject to the requirement that each individual is at least as well off as under the actual tax and transfer systems T_e. That is, we require that $U(T_e + T_d, X, \epsilon, \varepsilon) \geq U(T, X, \epsilon, \varepsilon)$ for all $(X, \epsilon, \varepsilon)$. Recall that Pareto improvements in this setting require reductions in tax schedules.

13.6.1 *Efficiency Implications for the Tax Schedule*

The results of this exercise are presented in column 2 of Table 13.4. Reductions in the tax schedule are found for weekly earnings between 225 and 400 pounds per week. This is precisely the range where the density of earnings is falling most quickly (see column 1 of the same table). As Werning (2007) notes, reductions in the tax schedule at a point will cause some individuals to reduce their labor supply, and others to increase it. While tax revenue is always lost from the former group, it can be increased for the latter. If the earnings density is falling sufficiently quickly, then the number of individuals who increase their labor supply will be large relative to the number who decrease it, making an increase in tax revenue more likely.

The table also quantifies the inefficiency under the existing system by comparing the actual and maximized revenue levels from this exercise. The same metric was proposed by Werning (2007) but was not quantitatively explored. As a result of this reform, we find that the government expenditure on single mothers is reduced by around 0.1%. Thus, the increase in tax revenue that this particular reform delivers is clearly very small and suggests that the actual system is close to being efficient. Of course, this metric does not quantify any gains that accrue to single mothers as a result of the reductions in the tax schedules that they face.

Before we explore incorporating partial hours observability into T_d we first consider a somewhat more relaxed criterion where we integrate

Table 13.4. Pareto Improving Changes to the Tax Schedule

Weekly earnings	Base density	Conditional on (X, ϵ, ε)		Conditional on (X, ϵ_w)	
		No hours rule	Hours rule	No hours rule	Hours rule
0–25	0.005	0.000	0.000	0.000	0.205
25–50	0.041	0.000	0.000	0.000	–0.297
50–75	0.046	0.000	0.000	0.000	0.243
75–100	0.044	0.000	0.000	0.000	0.194
100–125	0.054	0.000	0.000	0.000	–0.119
125–150	0.048	0.000	0.000	0.000	0.025
150–175	0.049	0.000	0.000	0.000	0.192
175–200	0.042	0.000	0.000	0.000	–0.231
200–225	0.034	0.000	0.000	0.000	–0.075
225–250	0.032	–0.076	–0.076	–0.083	0.167
250–275	0.021	0.076	0.076	0.088	–0.048
275–300	0.020	–0.434	–0.434	–0.456	–0.092
300–325	0.016	0.064	0.064	0.074	–0.107
325–350	0.018	–0.073	–0.073	–0.052	0.072
350–375	0.010	0.273	0.273	0.167	0.074
375–400	0.008	0.170	0.170	0.253	0.193
400–425	0.008	0.000	0.000	0.059	0.224
425–450	0.007	0.000	0.000	0.026	0.107
450–475	0.007	0.000	0.000	–0.030	–0.354
475–500	0.006	0.000	0.000	–0.038	0.178
500+	0.027	0.000	0.000	–0.001	–0.001
Out-of-work income		0.000	0.000	0.000	0.269
Bonus at 16 hours		—	0.000	—	–1.370
Bonus at 30 hours		—	0.000	—	18.616
Change in expenditure		–0.090%	–0.090%	–0.095%	–0.692%

Notes: Table presents changes to the structure of marginal tax rates, out-of-work income, and hours contingent payments that yield Pareto improvements conditional on $(X, \epsilon, \varepsilon)$ and (X, ϵ_w) respectively. The base system refers to the actual 2002 tax and transfer system with complete take-up of tax credits. All incomes are in pounds per week and are expressed in April 2002 prices.

over some dimensions of the unobserved heterogeneity and require that individuals are made no worse off for all (X, ϵ_w). The set of inequality constraints (Equation 3) are then replaced by:

$$\int_{\epsilon_{-w}}\int_{\varepsilon} U(T_e + T_d, X, \epsilon, \varepsilon)dF(\varepsilon)dG(X, \epsilon_{-w} \mid \epsilon_w)$$
$$\geq \int_{\epsilon_{-w}}\int_{\varepsilon} U(T_e, X, \epsilon, \varepsilon)dF(\varepsilon)dG(X, \epsilon_{-w} \mid \epsilon_w)$$

for all (X, ϵ_w). This may be viewed as an appropriate criterion if we think of welfare conditional on characteristics X and idiosyncratic productive capacity ϵ_w. Note that this relaxed criterion does not necessarily require reductions in the tax schedule everywhere. The results are shown in column 4 of Table 13.4, and are extremely similar to those obtained in our initial exercise.

13.6.2 *Incorporating Hours Information*

We now consider the use of hours information to improve efficiency. The hours rules in T_d are restricted to operate at the same location as under the actual system T_e (that is, further payments are received if working at the discrete points corresponding to more than 16 and more than 30 hours per week). Note that if we condition on all the observed and unobserved heterogeneity, then Pareto improvements do not permit any reductions in these hours contingent payments since it would make individuals with a particularly high attachment to a given hours state worse off. This severely limits the potential for reforms to the hours rules to yield Pareto improvements. Indeed, the revenue maximizing tax schedules (column 3) does not alter the hours bonuses, with the reformed schedule the same as reported in column 2 of the same table.

Unsurprisingly, the more relaxed criterion produces quite different results as we are integrating over the unobserved heterogeneity ε that is responsible for this hours attachment. The results from this exercise (see column 5) point to a small increase in out-of-work income, together with a reduction in the size of the part-time hours bonus and a large increase in the full-time hours bonus. There are also pronounced changes to marginal tax rates over the entire distribution of labor earnings. This reform produces larger reductions in government expenditure relative to when we did not adjust the size of the hours bonuses (around 1%). The requirement that no individual is made worse off following a tax reform is a demanding criterion, particularly in the presence of preference heterogeneity. In the supplementary material we quantify the extent to which imposing this requirement may restrict the potential for the type of social welfare improving reforms that we consider in the following section.

13.7 Optimal Design of the Tax and Transfer Schedule

In this section we consider the normative implications when we adopt a social welfare function with a set of subjective social welfare weights. The analysis here shows the key importance of the differences in labor supply responses at the extensive and intensive margin. We also examine the welfare cost from moving to an administratively simpler linear tax system. The variation in response elasticities noted in our discussion of the estimated model above points to potential gains from allowing the optimal schedule to vary with children's age. We investigate such a design.

Given the use of a minimum hours condition for eligibility in the British tax credit system, we also consider the design in the case of a minimum hours rule. We show that if hours of work are partially (but otherwise accurately) observable, then there can be welfare gains from introducing an hours rule for lone mothers. However, accurately observing hours of work is crucial for this result. Our results suggest that if hours of work are subject to random measurement error or direct misreporting then the welfare gains that can be realized may be much reduced. Our analysis here therefore supports the informal discussion regarding the inclusion of hours in the tax base in Banks and Diamond (2010). Before detailing these results, we first turn to the choice of social welfare transformation and the parameterization of the tax and transfer schedule.

13.7.1 *Optimal Tax Specification*

To implement the optimal design analysis described in Section 13.2.2 we approximate the underlying non-parametric optimal schedule by a piecewise linear tax schedule as in Section 13.6. Here the tax schedule will be characterized by a level of out-of-work income (income support), and nine different marginal tax rates.[10] We do not tax any non-labor sources of income, and do not allow childcare usage to interact with tax and transfer schedule unless explicitly stated. When we later allow for partial observability of hours, we introduce additional payments that are received only if the individual fulfills the relevant hours criteria.

In all of these illustrations we continue to condition upon the presence of a single child, and set the value of government expenditure equal to the predicted expenditure on this group within our sample. Conditioning upon this expenditure we numerically solve for the tax and transfer schedule that maximizes social welfare. Throughout this section we adopt the following utility transformation in the social welfare function:

$$Y(U;\,\theta) = \frac{(\exp U)^{\theta} - 1}{\theta} \tag{9}$$

which controls the preference for equality by the parameter θ and also permits negative utilities which is important in our analysis given that the state specific errors ε can span the entire real line. When θ is negative, the function in equation (9) favors the equality of utilities; when θ is positive the reverse is true. By L'Hôpital's rule $\theta = 0$ corresponds to the linear case. Note that $-\theta = -Y''(U;\,\theta)/Y'(U;\,\theta)$ so that $-\theta$ can be interpreted as the coefficient of absolute inequality aversion.

We solve the schedule for a set of parameter values $\theta = \{-0.4, -0.2, 0.0\}$ and then derive the social weights that characterize these redistributive preferences. We do not consider cases where $\theta > 0$. The presence of state specific Type-I extreme value errors, together with our above choice of utility transformation has some particularly convenient properties, as the following proposition now demonstrates.

Proposition 1. Suppose that the utility transformation function is as specified in equation (9). If $\theta = 0$ then conditional on X and ϵ the integral over (Type-I extreme value) state specific errors ε in equation (4) is given by:

$$\log\left(\sum\nolimits_{h\in H}\exp(u(c(h;T,X,\epsilon),h;X,\epsilon))\right)+\gamma$$

where $\gamma \approx 0.57721$ is the Euler-Mascheroni constant. If $\theta < 0$ then conditional on X and the integral over state specific errors is given by:

$$\frac{1}{\theta}\left[\Gamma(1-\theta)\times\left(\sum_{h\in H}(u(c(h;T,X,\epsilon),h;X,\epsilon))\right)^{\theta}-1\right]$$

where Γ is the gamma function.

Proof. The result for $\theta = 0$ follows directly from an application of L'Hôpital's rule, and the well known result for expected utility in the presence of Type-I extreme value errors (see McFadden 1978). See Appendix B for a proof in the case where $\theta < 0$.

This proposition, which essentially generalizes the result of McFadden (1978), facilitates the numerical analysis as the integral over state specific errors does not require simulating. Moreover, the relationship between the utilities in each state, and the contribution to social welfare for given (X, ϵ) is made explicit and transparent.

13.7.2 *Implications for the Tax Schedule*

The underlying properties from the labor supply model, together with the choice of social welfare weights, are the key ingredients in the empirical design problem. As set out in Section 13.2, our model is characterized by both intensive and extensive labor supply responses. The summary labor supply elasticity measures presented in Table 13.3 point to a sizeable extensive elasticity and a relatively small intensive elasticity. As formalized by Saez (2002), whenever extensive labor supply responses are high at low earnings relative to intensive responses, low (or even negative) marginal tax rates are more likely to be optimal.

Starting from an initially high marginal tax rate, the marginal cost associated with a reduction in this tax rate (higher earners reducing their labor supply on the intensive margin), is likely to be dominated by the marginal benefit (by encouraging non-workers to enter work).

The parameter estimates presented in the supplementary material (and discussed in Section 13.5.3) show that both observed and unobserved preference heterogeneity are important determinants of individual utilities. Even when only intensive labor supply responses are permitted, the presence of such multi-dimensional heterogeneity (preferences and ability) exerts an important influence on the structure of tax rates, and can provide another source of departure from the predictions of the standard Mirrlees (1971) model. Choné and Laroque (2010) demonstrate that the optimal schedule depends on the average social weights of individuals conditional on observed earnings. The precise influence of heterogeneity then depends on how its distribution varies with earnings; they also show that there are conditions under which negative marginal tax rates may become optimal in this setting.

We now describe our results. For the choice of utility transformation function in Equation (9) we examine the impact of alternative θ values. In Table 13.5 we present the underlying average social welfare weights evaluated at the optimal schedule (discussed below) according to these alternative θ values. For all three values of θ considered here the weights are broadly downward sloping. For the most part we focus our discussion here on the –0.2 value, although we do provide

Table 13.5. Social Welfare Weights Under Optimal System

Weekly earnings	θ = –0.4		θ = –0.2		θ = 0.0	
	Density	Weight	Density	Weight	Density	Weight
0	0.398	1.378	0.367	1.305	0.281	1.073
0–50	0.055	1.340	0.051	1.218	0.039	0.968
50–100	0.109	1.088	0.104	1.071	0.088	0.935
100–150	0.101	0.907	0.110	0.987	0.123	1.015
150–200	0.100	0.718	0.111	0.855	0.136	1.024
200–250	0.078	0.563	0.087	0.721	0.115	1.021
250–300	0.049	0.457	0.054	0.615	0.071	0.959
300–350	0.043	0.347	0.046	0.504	0.060	0.945
350–400	0.021	0.307	0.023	0.454	0.029	0.880
400+	0.046	0.184	0.047	0.305	0.058	0.806

Notes: Table presents social welfare weights under optimal structure of marginal tax rates and out-of-work income under range of distributional taste parameters θ as presented in Table 13.6. All incomes are in pounds per week and are expressed in April 2002 prices. Welfare weights are obtained by increasing consumption uniformly in the respective earnings range and calculating a derivative; weights are normalized so that the earnings-density-weighted sum under optimal system is equal to unity.

Table 13.6. Optimal Tax Schedules

Weekly Earnings	No hours			19 hours			Optimal hours		
	θ = −0.4	θ = −0.2	0 = 0.0	θ = −0.4	θ = −0.2	θ = 0.0	θ = −0.4	θ = −0.2	θ = 0.0
0–50	0.132	0.144	0.139	0.266	0.280	0.252	0.053	0.056	0.078
	(0.028)	(0.025)	(0.029)	(0.029)	(0.034)	(0.037)	(0.031)	(0.028)	(0.027)
50–100	0.520	0.344	−0.022	0.995	0.899	0.328	0.778	0.646	0.123
	(0.030)	(0.030)	(0.044)	(0.006)	(0.034)	(0.062)	(0.030)	(0.032)	(0.036)
100–150	0.354	0.275	−0.022	0.466	0.355	−0.013	0.535	0.481	0.221
	(0.019)	(0.020)	(0.037)	(0.027)	(0.019)	(0.039)	(0.021)	(0.022)	(0.025)
150–200	0.483	0.414	0.069	0.503	0.440	0.090	0.698	0.650	0.229
	(0.014)	(0.017)	(0.033)	(0.014)	(0.017)	(0.035)	(0.028)	(0.030)	(0 042)
200–250	0.520	0.471	0.167	0.535	0.484	0.173	0.672	0.638	0.483
	(0.015)	(0.017)	(0.038)	(0.015)	(0.017)	(0.039)	(0.030)	(0.032)	(0.071)
250–300	0.540	0.501	0.189	0.551	0.512	0.197	0.659	0.632	0.231
	(0.020)	(0.021)	(0.040)	(0.020)	(0.022)	(0.042)	(0.043)	(0.045)	(0.069)
300–350	0.546	0.514	0.266	0.554	0.521	0.270	0.644	0.613	0.670
	(0.023)	(0.025)	(0.053)	(0.024)	(0.026)	(0.053)	(0.038)	(0.040)	(0.082)
350–400	0.590	0.551	0.286	0.604	0.575	0.293	0.728	0.715	0.284
	(0.019)	(0.020)	(0.040)	(0.019)	(0.021)	(0.042)	(0.029)	(0.031)	(0.045)
400+	0.616	0.599	0.401	0.623	0.607	0.403	0.687	0.676	0.558
	(0.008)	(0.009)	(0.023)	(0.008)	(0.009)	(0.024)	(0.008)	(0.009)	(0.028)
Out-of-work	135.975	131.170	103.651	136.226	131.361	104.407	137.262	132.204	106.072
income	(s1.672)	(s1.680)	(s3.308)	(1.704)	(1.686)	(3 348)	(1.740)	(1.736)	(3.271)
Hours bonus	—	—	—	36.290	38.698	23 231	44.056	48.632	51.702
				(1.670)	(1.357)	(2.944)	(2.037)	(1.540)	(6.136)
Hours point	—	—	—	19	19	19	33	33	40

Notes: Table presents optimal structure of marginal tax rates and out-of-work income under range of distributional taste parameters θ. All incomes are in pounds per week and are expressed in April 2002 prices. Standard errors are in parentheses and are calculated by sampling 500 times from the distribution of parameter estimates and conditional on the sample distribution of observables.

a sensitivity of our results to the choice of θ and find the broad conclusions are robust to this choice.

In the first three columns of Table 13.6 we present the optimal tax and transfer schedules across the alternative θ values; these schedules are also illustrated in Figure 13.3. In the table we present standard errors for the parameters of the optimal tax schedule, which are obtained by sampling 500 times from the distribution of parameter estimates and resolving for the optimal schedule conditional on the sample distribution of covariates. In all the simulations performed here, we obtain a broadly progressive marginal tax rate structure: marginal tax rates are typically much lower in the first tax bracket (earnings up to £50 per-week) than at higher earnings. Apart from the $\theta = 0.0$ case, the calculated marginal tax rates are much higher in the second bracket than the first, but then fall before proceeding to generally increase with labor earnings. As we increase the value of θ (less redistributive concern), we obtain reductions in the value of out-of-work income. This is accompanied by broad decreases in marginal tax rates, except in the first tax bracket where marginal tax rates are largely unchanged. The social welfare weights presented in Table 13.5 reflect these changes.[11]

The results presented in Table 13.6 point towards a non-linear tax schedule over a large range of earnings. For each value of θ considered we quantify the welfare gains from allowing for such non-linearity by

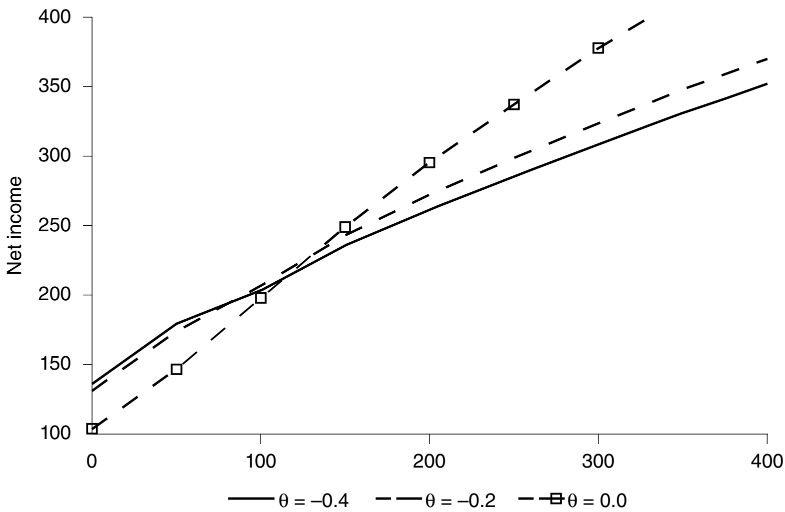


Figure 13.3. Optimal Tax Schedules with Alternative Values of θ

Note: All incomes are measured in April 2002 prices and are expressed in pounds per week.

calculating the increase in government expenditure required such that the value of social welfare under the optimal linear tax system is the same as under the non-linear systems above. This produces optimal constant marginal tax rates of 43.5%, 37.6% and 11.3% (for $\theta = -0.4$, $\theta = -0.2$ and $\theta = 0.0$ respectively). In the illustrations when $\theta = -0.2$, government expenditure would need to increase by 1.5% to achieve the same level of social welfare.

13.7.3 *Tagging by Age of Child*

Before exploring the use of hours contingent payments in the tax schedule we consider how the optimal schedule varies by age of children, should the government decide to condition (or tag) the tax and transfer schedule upon this information. The nature of the optimal income tax schedule in the presence of tagging was explored by Akerlof (1978). Note that WFTC awards depended upon on the age of children (the different rates are presented in the supplementary material) as do other parts of the UK tax and transfer system (including income support, the main transfer available to low-income families working less than 16 hours per week).

Our results in Table 13.2 suggest that labor supply responses differ significantly at the extensive and the intensive margin according to the age of children. Whenever the labor supply of an identified group is more responsive to tax rates than is that for other groups, then this identified group should face lower marginal tax rates. By shifting the tax burden to otherwise equivalent individuals with lower elasticities of labor supply, the tax structure can create lower efficiency costs while holding unchanged the degree of redistribution from rich to poor, see Gordon and Kopczuk (2013), for example.

In our analysis we do not change the resources going to parents we just adjust the payments according to the age of the child. Nonetheless, since our model is static this exercise ignores the dynamics that are introduced by the child ageing process. Clearly, such considerations could be important for the optimal design problem and will be explored in future work. However, this remains an important benchmark case and is likely to still yield important insights, particularly if the population of interest has a sufficiently low discount factor, or is liquidity constrained.

We proceed to solve for the optimal tax schedules for three different groups on the basis of the age of youngest child: under 4, aged 5 to 10 and 11 to 18. Since the childcare requirements of mothers with young children are considerably higher, we also allow for a childcare expenditure subsidy of 70% (which corresponds to the formal childcare subsidy

Table 13.7. Optimal Tax System by age of Child with Childcare Subsidy (Conditional on Group Expenditure)

a) Fixed Expenditure Division

Weekly earnings	0–4			5–10			11–18		
	θ = −0.4	θ = −0.2	θ = 0.0	θ = −0.4	θ = −0.2	θ = 0.0	θ = −0.4	θ = −0.2	θ = 0.0
0–50	0.19B	0.287	0.432	−0.003	0.006	0.085	−0.107	−0.111	−0.009
50–100	0.503	0.344	0.043	0.545	0.370	0.013	0.478	0.279	−0.013
100–150	0.309	0.232	−0.033	0.395	0.320	0.038	0.445	0.343	−0.004
150–200	0.478	0.415	0.151	0.517	0.444	0.085	0.552	0.472	0.086
200–250	0.490	0.442	0.145	0.579	0.537	0.265	0.577	0.510	0.154
250–300	0.557	0.526	0.348	0.532	0.480	0.101	0.674	0.629	0.222
300–350	0.530	0.496	0.220	0.640	0.614	0.449	0.488	0.441	0.160
350–400	0.592	0.563	0.384	0.583	0.540	0.168	0.771	0.734	0.383
400+	0.607	0.590	0.431	0.640	0.622	0.420	0.654	0.631	0.377
Out-of-work income	140.950	139.152	126.405	131.855	125.374	95.572	118.382	106.947	66.850

b) Optimal Expenditure Division

Weekly earnings	0–4			5–10			11–18		
	θ = −0.4	θ = −0.2	θ = 0.0	θ = −0.4	θ = −0.2	θ = 0.0	θ = −0.4	θ = −0.2	θ = 0.0
0–50	0.167	0.265	0.429	−0.002	0.008	0.085	−0.121	−0.115	−0.009
50–100	0.535	0.368	0.047	0.536	0.362	0.016	0.441	0.254	−0.024
100–150	0.316	0.238	−0.028	0.398	0.323	0.041	0.458	0.353	−0.015
150–200	0.473	0.406	0.156	0.519	0.447	0.088	0.564	0.483	0.073
200–250	0.482	0.433	0.153	0.584	0.541	0.268	0.585	0.517	0.146
250–300	0.544	0.513	0.351	0.533	0.482	0.104	0.685	0.640	0.209
300–350	0.523	0.490	0.223	0.643	0.618	0.450	0.495	0.447	0.154
350–400	0.581	0.551	0.387	0.585	0.543	0.171	0.780	0.742	0.372
400+	0.502	0.584	0.433	0.642	0.623	0.422	0.660	0.636	0.370
Out-of-work income	156.618	154.340	123.959	127.071	120.336	93.975	100.615	90.768	71.954

Notes: Table presents optimal structure of marginal tax rates and out-of-work income by age of youngest child under range of distributional taste parameters θ. All schedules calculated with an uncapped childcare subsidy equal to 70%. All incomes are in pounds per week and are expressed in April 2002 prices.

rate under WFTC) to facilitate the comparison of marginal tax rates across these groups. We first solve for these schedules separately when we condition on the predicted expenditure on each of these groups in our sample; we then solve for these schedules jointly allowing the division of overall expenditure to be reoptimized. Full results are presented in Tables 13.7a and 13.7b; Figure 13.4 illustrates these with fixed group expenditure when $\theta = -0.2$.

While the overall structure of the schedules retain many of the features present in our earlier simulations, our optimal tax simulations here reveal some important differences by the age of children. In the case of fixed within group expenditure (see Table 13.7a), marginal tax rates tend to be higher at low earnings for lone mothers with younger children: in the first tax bracket marginal tax rates for the youngest group are around 40 percentage points higher than for the oldest group. Amongst women with children from the oldest group we also obtain negative marginal tax rates. The higher marginal tax rates at low earnings for parents with younger children are also accompanied by higher levels of out-of-work support for these groups.

Conditioning upon within group expenditure levels makes an implicit assumption on the weight that the government attaches on the welfare

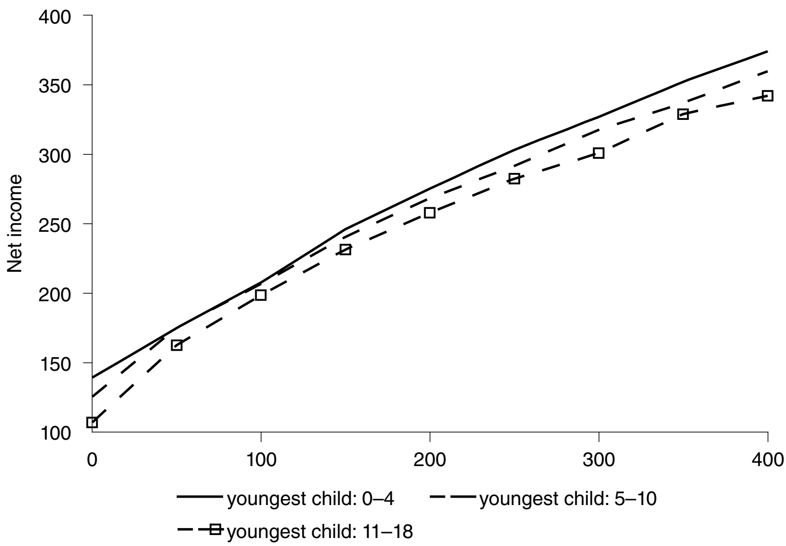


Figure 13.4. Optimal Tax Schedules by Age of Child

Note: All schedules are calculated with fixed expenditure division and with $\theta = -0.2$. All incomes are measured in April 2002 prices and are expressed in pounds per week.

of parents with children of different ages. Under the assumption that the government places equal valuation on these groups we solve for the three schedules jointly (see Table 13.7b). Relative to the previous simulations, this makes the differences across groups more pronounced. In particular, there are notable increases in expenditure (and out-of-work income levels) for lone mothers with younger children. And while there are some changes in the structure of marginal tax rates (due to income effects) these changes are somewhat smaller in magnitude.

The welfare gains from tagging on the basis of age of children can be calculated in much the same way as when comparing a non-linear schedule to one which is linear. The potential welfare gains appear reasonably large: relative to a system where tagging by the age of youngest child is not possible, government expenditure would have to increase by 2.6% (when $\theta = -0.2$) to obtain the same level of social welfare as that achieved when such tagging is possible. These gains are even larger when more redistributive preferences are considered.

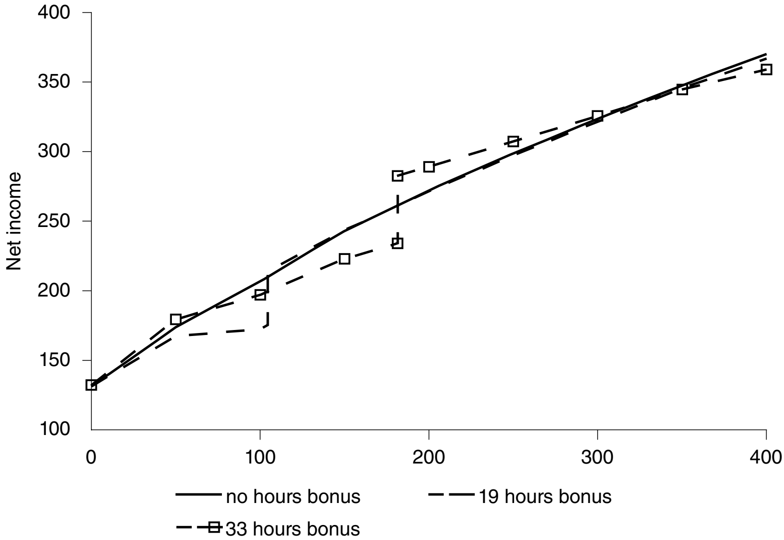
13.7.4 *Introducing an Hours Rule*

For several decades the UK's tax credits and welfare benefits have made use of rules related to weekly hours of work. As discussed in Section 13.3, individuals must work at least 16 hours a week to be eligible for in-work tax credits, and receive a further credit when working 30 or more hours. While many theoretical models rule out the observability of any hours information, this design feature motivates us to explore the optimal structure of the tax and transfer system when hours can be partially observed as set out in Section 13.2. Essentially, observing some hours of work information allows the government to better distinguish between different types of individuals. In the absence of any labor supply participation response, and when the only source of worker heterogeneity is the exogenous wage rate (productive ability), the government is able to redistribute without costs when both hours and earnings are perfectly observable since it can now infer ability. The first best ceases to be attainable once hours of work are only partially observed, but even this information allows the government to better separate types relative to when labor earnings are the only signal.

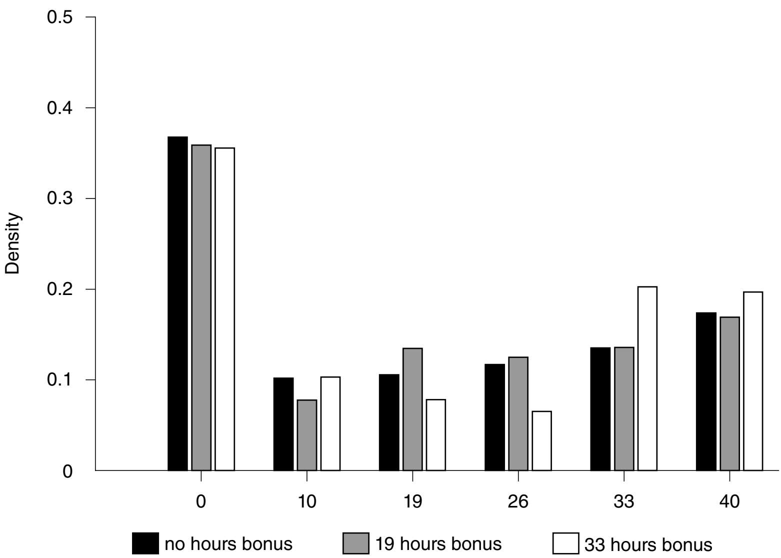
We begin by assuming that the tax authority is able to observe whether individuals are working 19 hours or more, which roughly corresponds to the placement of the main 16 hours condition in the British tax credit system, and for now we do not allow for any form of measurement error. In this case the tax authority is able to condition an additional payment on individuals working such hours. The results of this exercise are presented in columns 4–6 in Table 13.6, and the $\theta = -0.2$ case is also presented in Figure 13.5a. Relative to the optimal system when such a rule is not implementable, the hours bonus increases marginal rates in the part of the earnings distribution where this hours rule would roughly come into effect (particularly in the £50 to £100 earnings bracket) while marginal rates further up the distribution, as well as the level of out-of-work support, are essentially unchanged. As a result, some non-workers with low potential wages may be induced to work part-time, while some low hours individuals will either not work or increase their hours. Similarly, some high earnings individuals reduce their hours to that required for the bonus.[12]

Although there are some notable changes in the structure of the constraint when hours information is partially observable, it does not follow that it necessarily leads to a large improvement in social welfare. Indeed, in the absence of the hours conditioning, there are only a few individuals working less than 19 hours (see Figure 13.5b when $\theta = -0.2$) so the potential that it offers to improve social welfare appear limited. We now provide some guidance concerning the size of the welfare gain from introducing hours rules. The exact experiment we perform is as follows: we calculate the level of social welfare under the optimal schedule with hours contingent payments, and then determine the increase in expenditure that is required to obtain the same level of social welfare in the absence of such hours conditioning; we allow all the parameters of the (earnings) tax schedule to vary so this is obtained at least cost.

Perhaps unsurprisingly, these welfare gains are found to be relatively small. In both the $\theta = -0.4$ and $\theta = -0.2$ cases the expenditure increase required to achieve the level of social welfare obtained under the 19 hours rule is less than 1%. When the least redistributive preferences are considered, this falls to just 0.2%. Even without allowing for any measurement error, it follows that unless the costs of partial hours observability is sufficiently low, it would appear difficult to advocate the use of a 19 hour rule based upon this analysis. This has very important policy implications given that the UK tax credit system makes heavy use of very similar hours conditions. We note that Keane and Moffitt (1998) considered introducing a work subsidy in a model with three employment states (non-workers, part-time and full-time work) and multiple benefit take-up. Even small subsidies were found to increase labor supply and to reduce dependence on welfare benefits, and at reduced cost. In contrast to our application (where we are moving from a base with marginal rates well below 100% at low earnings), their



a) Optimal Tax Schedules



b) Distribution of Work Hours

Figure 13.5. Optimal Tax Schedules with Hours Bonuses and Associated Hours Distribution

Note: All schedules are calculated with $\theta = -0.2$ and assuming a gross hourly wage of £5.50. All incomes are measured in April 2002 prices and are expressed in pounds per week.

simulations considered introducing the subsidy in an environment where many workers faced marginal effective tax rates which often exceeded 100%, and where the receipt of these work subsidies encourages women to exit welfare benefits (and so no longer be affected by the associated stigma costs).

13.7.4.1 AN OPTIMAL HOURS RULE?

The social welfare gains from introducing a 19 hours rule appear to be only very modest in size at best. In this section we explore whether there are potentially larger gains by allowing the choice of the point at which the hours rule becomes effective to be part of the optimal design problem. The parameters of the optimal tax schedules for all θ are presented in columns 7–9 of Table 13.6, while the optimal schedule when $\theta = -0.2$ is also shown in Figure 13.5a. Apart from when considering the least redistributive government preferences, we obtain an optimal hours rule at the fifth (out of six) discrete hours point, which corresponds to 33 hours per week (when $\theta = 0.0$ the optimal placement shifts to 40 hours per-week). We also note that the size of the optimally placed hours bonus always exceeds that calculated when the hours rule became effective at 19 hours per week.

Introducing an hours rule further up the hours distribution allows the government to become more effective in distinguishing between high wage/low effort and high effort/low wage individuals than at 19 hours to the extent that few higher wage individuals would choose to work very few hours. Relative to the schedule when the hours rule is set at around 19 hours, this alternative placement tends to make people with low and high earnings better off, while people in the middle range lose. While we again obtain very small adjustments to the level of out-of-work income, there are much more pronounced changes to the overall structure of marginal rates. In particular, there are large reductions in the marginal tax rate in the first tax bracket, while marginal rates increase at higher earnings. Figure 13.5b shows the resulting impact on the hours distribution when $\theta = -0.2$.

As before, we attempt to quantify the benefits from allowing for hours conditioning. Performing the same experiment as we conducted under the 19 hours rule we find that the required increase in expenditure is considerably larger than that obtained previously. We find that a 2.5% increase in expenditure would be required to achieve the same level of social welfare when $\theta = -0.2$ (with very similar increases for the alternative θ values), which represents a non-trivial welfare gain. In any case, if the government wishes to maintain the use of hours conditional eligibility, the analysis here suggests that it may be able to improve design by shifting towards a system that primarily rewards full-time rather than part-time work.[13]

13.7.4.2 MEASUREMENT ERROR AND HOURS MISREPORTING

The results presented have not allowed for any form of measurement error. While earnings may not always be perfectly measured, it seems likely that there is more scope for mismeasurement of hours as they are conceivably harder to monitor and verify. Indeed, the presence of hours rules in the tax and transfer system presents individuals with an incentive to not truthfully declare whether they satisfy the relevant hours criteria. Relative to when hours are always accurately reported, this would seem to weaken the case for introducing a measure of hours in the tax base as the signal is now less informative about individual type. While we do not explore this issue, we note that the government may be able to improve design by using additional tax instruments that are related to hours of work. An example of such an instrument is childcare expenditure, which may be observed more accurately than self reported hours of work if the tax authorities require expenditure receipts. We quantify the importance of such measurement error by considering two alternative scenarios: firstly, when hours are imperfectly observed due to random measurement error; secondly, when we allow individuals to directly misreport their hours of work to the tax authorities.

In the supplementary material we present results from the first case with random measurement error. We show how both the size of the optimal hours bonus and the associated welfare gains decline as reported hours become less informative. Here we focus upon the arguably more plausible case of systematic hours misreporting. We modify our setup by distinguishing between actual hours of work h, and reported hours of work h_R; actual hours determine both leisure and earnings, while reported hours of work directly affect consumption through the tax schedule, with $T = T(wh, h_R; X)$. If individuals misreport their hours then they must incur a utility cost, which is assumed to be proportional to the distance $h_R - h$. We therefore modify the individual utility function by including $h_R - h$ as an explicit argument, with $U = u(c, h, h_R - h; X, \epsilon) + \varepsilon_h$. This modified utility function is as in equation (6) but now with the additional cost term $b \times (h_R - h)$ subtracted from u whenever $h_R > h$.[14]

Misreporting is only possible when $h > 0$, and we refer to the parameter $b \geq 0$ as the misreporting cost. We do not allow individuals to manipulate their earnings wh.

As before, we consider tax schedules with a single hours eligibility threshold, and denote this hours requirement as h_B. Since misreporting hours is costly, it is only necessary to consider the cases when hours are

Table 13.8. The Effect of Hours Misreporting on the Optimal Hours Bonus

Misreporting cost	θ = −0.4			θ = −0.2			θ = 0.0		
	Bonus	Hours	Welfare (%)	Bonus	Hours	Welfare (%)	Bonus	Hours	Welfare (%)
∞	44.06	33	2.24	48.63	33	2.46	51.70	40	2.44
0.64	44.05	33	2.24	48.62	33	2.46	51.70	40	2.44
0.32	42.71	33	2.21	47.44	33	2.42	51.34	40	2.43
0.16	31.02	33	1.89	33.55	33	2.05	35.99	40	2.10
0.08	21.04	33	1.28	25.08	33	1.44	28.78	40	1.54
0.04	13.26	33	0.83	14.39	33	0.94	16.47	40	1.07
0.02	8.02	33	0.53	9.27	33	0.62	11.25	40	0.78
0.01	5.98	33	0.37	7.17	33	0.47	8.59	40	0.68

Notes: Table shows how the optimal placement and size of hours contingent payments varies with the utility cost of hours misreporting. "Misreporting Cost" refers to the additive utility cost associated with misreporting, and is measured per-hour overstated and relative to standard deviation of the state specific error ε. The columns "welfare" refer to the percentage increase in required expenditure to achieve the same level of social welfare compared to when no hours conditioning is performed. All incomes are in pounds per week and are expressed in April 2002 prices.

truthfully revealed $h_R = h$, or when $h_R = h_B > h$. At a given actual hours of work $h < h_B$ individuals will report their hours as $h_R = h_B$ if and only if the utility gain exceeds the cost. That is:

$$u(c(h,T(wh,h_B;X),X,\epsilon),h,h_B-h;X,\epsilon)$$
$$> u(c(h,T(wh,h;X),X,\epsilon),h,0;X,\epsilon).$$

We present results from this exercise in Table 13.8.[15] The table illustrates that as the utility cost of misreporting becomes very low, the welfare gain from using reported hours of work diminishes (but the optimal placement remains unchanged for all values considered). Note also that when $b = \infty$ misreporting is never optimal. This analysis suggests that the welfare gains from using hours of work information may be small unless the scope for misreporting hours of work is limited.

13.8 Summary and Conclusions

The aim of this paper has been to examine the optimal design of low-income support using a stochastic structural labor supply model. The application focussed on the design of the tax schedule for parents with children, in particular single mothers. The structural labor supply model was shown to be reliable and found to match closely the changes in observed behavior that followed a large reform to the tax credit system in the UK.

The paper has made three contributions to the existing literature on tax design. First, we have taken the structural model of employment and hours of work seriously in designing the optimal schedule of taxes and transfers. To operationalize this we have developed the design problem within an extended Mirrlees framework which has incorporated unobserved heterogeneity, the non-convexities of the tax and transfer system, as well as allowing for childcare costs and fixed costs of work. We first used this model to identify inefficiencies in the actual tax and transfer system and characterized Pareto improving reforms. While this analysis pointed to relatively minor improvements in the UK tax and transfer schedule for lone parents, by imposing a specific social welfare function with reasonable social welfare weights we obtained a reformed non-linear tax schedule with lower tax rates over a large range of earnings for many families, and with tax credits only optimal for low earners.

Tagging has been suggested to improve the trade-off between equality and efficiency. Our second contribution has been to empirically assess the role of tagging taxes by the age of children under a social welfare function. These results highlighted an importance of conditioning effective tax rates on the age of children, with tax credits being found to be most important for low-earning families with school age children. The welfare gains from this age based tagging were also found to be quantitatively significant.

We have noted that hours contingent payments are a key feature in the British tax credit system. Our third contribution was to consider the case where hours of work are partially observable to the tax authorities and to quantify the value of this signal. If the tax authorities are able to choose the lower limit on working hours that trigger eligibility for such families, then we find an empirical case for using a full-time work rule rather than the main part-time rule currently in place for parents in the UK. While this is found to be a more effective instrument, we demonstrate how these welfare gains diminish with both misreporting and measurement error.

Appendix A: Likelihood Function

In what follows let $P_j(X, p_{c_k}, \epsilon) \equiv Pr(h = h_j|X, p\_c\_k, \epsilon)$ denote the probability of choosing hours $h_j \in H$ conditional on demographics X, the childcare price, and the vector of unobserved preference heterogeneity $\epsilon = (\epsilon_w, \epsilon_c, \epsilon_y, \epsilon_\eta)$. Given the presence of state specific Type-I extreme value errors, this choice probability takes the familiar conditional logit form. We also use $\pi k(X) \equiv Pr(p_c = p_{c_k}|X)$ to denote the probability of a lone mother with observable characteristics X facing childcare price p_{c_k}.

In the case of non-workers ($h = h_0$), neither wages nor childcare are observed so that the likelihood contribution is simply given by:

$$\sum_k \pi_k(X) \int_\epsilon P_0(X, p_{c_k}, \epsilon) dG(\epsilon).$$

Now consider the case for workers when both wages and childcare is observed so that h_c is not censored at zero. Using $E_h \equiv E(h; X, p_c, \epsilon)$ to denote eligibility for in-work support we define the indicator $D(e, p) = 1(E_h = e, P = p)$. We also let $\Delta u(h_j \mid p_{c_k}, X, \epsilon_{|\eta=0})$ denote the (possibly negative) utility gain from claiming in-work support at hours h_j, conditional on demographics X, the childcare price p_{c_k}, and the vector of unobserved preference heterogeneity ϵ with $\epsilon_\eta = 0$. Suppressing the explicit conditioning for notational simplicity, the likelihood contribution is given by:

$$\prod_k \pi_k(X)^{1(p_c = p_{c_k})} \int_{\epsilon_y} \left\{ D(1,1) \int_{\epsilon_\eta < \Delta u} \prod_j P_j(X, p_{c_k}, \epsilon)^{1(h=h_j)} \right.$$

$$\left. + D(1,0) \int_{\epsilon_\eta > \Delta u} \prod_j P_j(X, p_{c_k}, \epsilon)^{1(h=h_j)} + D(0,0) \int_{\epsilon_\eta} \prod_j P_j(X, p_{c_k}, \epsilon)^{1(h=h_j)} \right\}$$

$$dG(\epsilon \mid \epsilon_w = \log w - X_w' \beta_w,\ \epsilon_c = h_c - \gamma_c - \beta_c h)$$

$$g_w(\log w - X_w' \beta_w) g_c(h_c - \gamma_c - \beta_c h).$$

If working mothers are not observed using childcare, then h_c is censored at zero and the childcare price also unobserved. Defining $\bar{\epsilon}_c = -\gamma_c - \beta_c h$, the likelihood contribution is given then by:

$$\sum_k \pi_k(X) \int \int_{\epsilon_c < \bar{\epsilon}_c, \epsilon_y} \left\{ D(1,1) \int_{\epsilon_\eta < \Delta u} \prod_j P_j(X, p_{c_k}, \epsilon)^{1(h=h_j)} \right.$$

$$\left. + D(1,0) \int_{\epsilon_\eta > \Delta u} \prod_j P_j(X, p_{c_k}, \epsilon)^{1(h=h_j)} + D(0,0) \int_{\epsilon_\eta} \prod_j P_j(X, p_{c_k}, \epsilon)^{1(h=h_j)} \right\}$$

$$dG(\epsilon \mid \epsilon_w = \log w - X_w' \beta_w) g\_w(\log w - X_w' \beta_w).$$

Our estimation also allows for workers with missing wages. This takes a similar form to the above, except that it is now necessary to also integrate over the unobserved component of wages ϵ_w.

All the integration over ϵ is performed using Gaussian Hermite quadrature with 11 nodes in each integration dimension. When it is unnecessary to integrate over the entire real line in a given dimension, a change of variable is conducted so that integration is performed over $(0, +\infty)$, with appropriate semi-Hermite quadrature formulae then applied.

Appendix B: Proof of Proposition

For notational simplicity we abstract from the explicit conditioning of utility on observed and unobserved preference heterogeneity and let $u(h) \equiv u(c(h), h; X, \epsilon)$. We then define V as the integral of transformed utility over state specific errors conditional on (X, ϵ):

$$V \equiv \textstyle\int_{\varepsilon} Y(\max_{h \in H}[u(h) + \varepsilon_h])dF(e). \qquad (A\text{–}1)$$

To prove this result we first differentiate V with respect to $u(h)$:

$$\begin{aligned}\frac{\partial V}{\partial u(h)} &= \int_{\varepsilon} \left(\frac{\partial Y(\max_{h \in H}[u(h) + \varepsilon_h])}{\partial u(h)} \right) dF(\varepsilon) \\ &= \int_{\varepsilon} Y'(u(h) + \varepsilon_h) \times 1(h = \arg\max_{h \in H}[u(h') + \varepsilon_{h'}])dF(\varepsilon). \qquad (A\text{–}2)\end{aligned}$$

Given our choice of utility transformation function in Equation (9) and our distributional assumptions concerning ε the above becomes:

$$\begin{aligned}\frac{\partial V}{\partial u(h)} &= \int_{\epsilon_c = -\infty}^{\infty} \{e^{(u(h)+c_k)}\}^{\theta} \left(\prod_{h' \neq h} e^{-e^{-\{\varepsilon_h + u(h) - u(h')\}}} \right) \times e^{-\varepsilon_h} e^{-e^{-\varepsilon_h}} d\varepsilon_h \\ &= \{e^{u(h)}\}^{\theta} \int_{\varepsilon_h = -\infty}^{\infty} \{e^{u(h)}\}^{\theta} \times \exp\left(-e^{-\varepsilon_h} \sum_{h' \in H} e^{-(u(h) - u(h'))} \right) e^{-\varepsilon_h} d\varepsilon_h.\end{aligned}$$

We proceed by using the change of variable $t = \exp(-\varepsilon_h)$ so that the above partial derivative becomes:

$$\frac{\partial V}{\partial u(h)} = \{e^{u(h)}\}^{\theta} \int_{t=0}^{\infty} t^{-\theta} \times \exp\left(-t \sum_{h' \in H} e^{-(u(h) - u(h'))} \right) dt.$$

By defining $z \equiv t \times \sum_{h' \in H} e^{-(u(h)-u(h'))}$ we can once again perform a simple change of variable and express the above as:

$$\begin{aligned}\frac{\partial V}{\partial u(h)} &= \{e^{u(h)}\}^{\theta} \left\{ \sum_{h' \in H} e^{-(u(h) - u(h'))} \right\}^{\theta - 1} \int_{z=0}^{\infty} z^{-\theta} e^{-z} dz \\ &= e^{u(h)} \left\{ \sum_{h' \in H} e^{u'(h)} \right\}^{\theta - 1} \int_{z=0}^{\infty} z^{-\theta} e^{-z} dz \\ &= e^{u(h)} \left\{ \sum_{h' \in H} e^{u'(h)} \right\}^{\theta - 1} \Gamma(1 - \theta)\end{aligned}$$

where the third equality follows directly from the definition of the Gamma function $\Gamma(\cdot)$. Note that this integral will always converge given that we are considering cases where $\theta < 0$. Integrating equation (A–2) we obtain:

$$V = \frac{1}{\theta}\left[\Gamma(1-\theta)\times\left(\sum_{h'\in H}\exp\{u(h')\}\right)^{\theta} - 1\right] \qquad (A\text{–}3)$$

where the constant of integration is easily obtained by considering the case of a degenerate choice set and directly integrating (A–1).

Part V
Conclusions and Future Directions

Richard Blundell

This book began with a question (see Chapter 1): how should empirical evidence on labor supply be used in the analysis of tax policy reform? The subsequent chapters have attempted to provide a partial answer, drawing on empirical contributions from my own (limited) publications. This is team research and I thank again my many coauthors in this venture. It is also a field of research that continues to advance, triggered by new policy questions, new sources of data and new models of behavior. Although great progress has been made in understanding the nature of labor supply decisions and in quantifying labor supply responses to tax policy reform, there is much more to be done.

The interplay between evidence and policy design is something that continues to keep research in the field of labor supply and taxation alive. The ideas developed in this volume, and the resulting estimated models, have been used to frame many policy discussions and to analyze many policy reforms over the past two decades. They were particularly influential in the Mirrlees Review and related studies, see Mirrlees et al. (2010; 2011) and references therein. Perhaps not surprisingly though, there still remain many important gaps in our knowledge. Here I focus on three: How should we incorporate *human capital decisions* with labor supply more explicitly – both formal education and work experience – in an uncertain environment? Can we improve our understanding of *decisions within the family* – especially the dynamics of life cycle family labor supply decisions? How should we more carefully incorporate job offers and *restrictions on choices of hours of work* when considering labor supply responses to tax reform?

These three questions tee-up a large and exciting agenda for future research. I conclude therefore by considering how the framework described in this book might be extended to answer each of these key questions. In outlining these directions for future analysis, I draw on more recent, and consequently, more speculative research.

Having mapped out these three directions for the future I finish up the discussion with a brief assessment of other key areas ripe for development. The study of labor supply and taxation is clearly set to remain an active and engaging field of economic research for many years to come!

(i) Incorporating Human Capital

As far as labor supply analysis is concerned there is a natural partition between two forms of human capital investment. First, is the formal education that occurs before an individual enters the labor market. Second, is the work experience and the on-the-job learning that occurs while in the labor market. The former is of key importance for tax and benefit design as changes to the structure of work incentives can naturally change the returns to education. They can change both the expected return and the uncertainty of returns. Perhaps even the preference for work itself, see Blundell, Duncan and Meghir (1998). Nonetheless, labor supply decisions can be analyzed *conditional* on chosen education levels, even though the education level should then be treated as endogenous. The later form of human capital investment – experience capital and on-the-job investment – drives an additional wedge between today's labor supply decision and tomorrow's labor supply choices. In this case past work experience can change the return to today's labor supply through its impact on tomorrow's wages.

The break in intertemporal separability generated by wages depending on past work experience implies that consumption is no longer a sufficient statistic for all external decisions and expectations in respect of within period labor supply decisions. Unlike in Chapter 4 there is now an explicit role for labor market dynamics. This would occur were the break in separability to come about through habits, adjustment costs or experience dynamics.

The standard separable model of Chapter 4 can be adapted to explicitly allow for saving and intertemporal labor supply decisions. Blundell, Meghir and Neves (1993) provide an example of this and allow for selection into work. But this model rules out state dependence in preferences or in the budget constraint. Thereby eliminating a role for habits, adjustment costs and returns to work experience, and consequently maintaining separability.

Experience capital would seem to be a key missing element of the standard labor supply model. There is good evidence that human capital accumulated through work. The simplest mechanism for this is through passive learning by doing. Imai and Keane (2004) make a strong case for including this in the analysis of female labor supply. Relaxing intertemporal separability this way implies that we are bound to explicitly model intertemporal decisions. That is we have to model risk preferences and discounting. We also have to make some assumptions about expectations and the nature of uncertainty. Uncertainty enters explicitly, not only over wages, but also in respect of household composition and tax policy itself.

To provide an insight into how this approach to modelling labor supply might best be done I draw on new research reported in Blundell et al. (2013). This work builds on a long history of dynamic labor supply models: it is related to Heckman and MaCurdy (1980) who developed the life cycle model of female labor supply, to Eckstein and Wolpin (1989) who introduced a dynamic discrete choice model of labor supply, wages and fertility, to Keane and Wolpin (1997) who estimate a dynamic model of education, occupational choice and labor supply and to Shaw (1989), Heckman, Lochner and Taber (1998b) and Imai and Keane (2004) who consider life cycle models of labor supply and consumption with human capital accumulation. The work also builds naturally on the intertemporal framework developed by MaCurdy (1983), Altonji (1986), Blundell and Walker (1986), Blundell, Meghir and Neves (1993), Blundell, Duncan and Meghir (1998) and Attanasio, Low and Sanchez-Marcos (2008).

The key modelling issue in introducing experience capital into the standard labor supply model is in the precise form by which past employment and past hours enter the wage equation. Does experience capital compliment formal education qualifications? How should more distant work experience be depreciated? Should part-time work be treated differently to full-time work? Blundell et al. (2013) address each of these issues and have uncovered some striking initial results. They use a sample of women drawn from the British Household Panel Survey (a household panel much like the PSID in the US) and estimate a structural model of life cycle labor supply and human capital choices incorporating a dynamic wage equation which allows for these different experience on earnings.

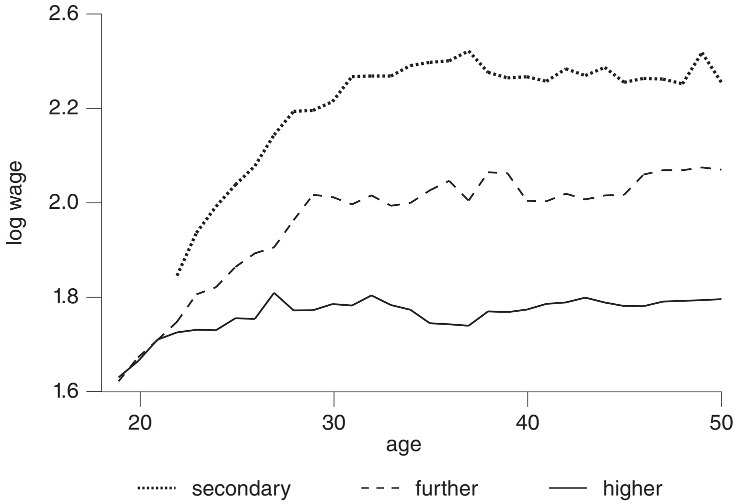


Figure V.1. Log Average Hourly Wages Across the Life Cycle for UK Women Workers by Education Group
Source: BHPS and Blundell et al. (2013).

Figure V.1 shows a clearly increasing life cycle profile of hourly wages with education level. However, this does not imply a significant or stronger experience profile. For this we need to separate persistent shocks from experience profiles as well as adjusting for selection into employment. After carefully allowing for these factors, the results in Blundell et al. (2013) suggest that the returns to experience are significantly more valuable for those with higher education qualifications. They also find little experience pay-off for part-time work and this holds for the educated group. Overall their estimates point to very little value placed on experience in part-time work and for lower-educated women. They suggest that a strong part-time penalty can develop in earnings for women who work for several years in part-time employment.

These results have clear tax policy implications, especially in the UK where, as we saw in Chapters 8 and 9, there have been a series of tax credit reforms that place strong incentives on part-time work for low-wage workers. The longer-term pay-off through wage progression from these incentives for part-time employment is likely to be negligible if the results of this study are confirmed. For those though with higher education investments, full-time work experience is an important driver of future wages. Future analyses of tax policies could usefully put some weight on exploring the implications of such experience capital in deriving the impact and optimality of policy. Indeed, there are important distinctions to be made within the class of on-the-job human investment models. The model I have outlined here is the simplest passive on-the-job learning model. Active investment models and on-the-job training provide important alternatives and will have different tax policy implications.

Perhaps as important for improving our understanding of the longer-term impacts and design of tax policy, as pointed out earlier, are education and occupation decisions. Tax and welfare policy change the relative returns to these choices. Moreover, as Blundell et al (2013) show, the insurance value of tax and welfare reform can substantively change the return to educational investments.

The interaction between labor supply, human capital investment and taxation will surely be of ever increasing importance in refining the design of tax and welfare policy.

(ii) Family Labor Supply

Tax policy design and evaluation has to account for the way families react to tax reform. But decisions taken within families are complicated by the obvious, if not always acknowledged, fact that families consist of a collection of individuals. They may take actions as if they are a single decision maker, they may take decisions collectively attaining the Pareto frontier, or they may take decisions in a non-cooperative way. Distinguishing between these alternative decision making models and their implications remains a key part of modern research on labor supply, see Apps and Rees (2009), Lundberg and Pollak (2008), Chiappori and Donni (2009), Blundell, Chiappori and Meghir (2005) and Blundell et al. (2007). As these papers, and many references therein point out, they also have strong implications for tax policy.

Policies that change relative incentives in the labor market will induce different responses in each of these models. Indeed, a policy that makes ‘households’ slightly better off might do so in a way that makes the poorest individuals much worse off. What I want to consider here is the way labor supply decisions of family members might be used as a way to insure families against adverse shocks. This could be incorporated in any of the family decision-making frameworks above, but to outline the issues involved I will focus here on the more standard, if less convincing(!), unitary model. I leave dynamic insurance considerations in non-unitary models to be taken forward in future research.

Mazzocco (2007) and Mazzocco and Saini (2012) have already made important progress.

The aim is to examine the impact of labor market shocks in families that make decisions in a unitary framework. Many of the key ideas are likely to become even more interesting in alternative family decision-making settings. Among mechanisms families might turn to for dealing with adverse shocks, the labor supply of different members is potentially one of the key ways that families have to smooth their marginal utility over time. For example, Attanasio, Low and Sanchez-Marcos (2005) note the importance female labor supply as insurance to labor market shocks. Other potential mechanisms include credit markets and savings, and gifts and informal contracts. The tax and welfare system will also provide insurance. In designing good tax and benefit policies for families it is important to know how these various alternative mechanisms are used. How important are family labor supply responses? For which type of households do they matter most? The recent paper Blundell, Pistaferri and Saporta-Eksten (2012) attempts to address these questions and here I draw on their results.

There are two aspects to modelling insurance behavior. First, is the nature of the labor market shocks facing the household. In the simplest framework I outline here, these are productivity shocks represented through the dynamic structure of wages of individuals in families. Additionally we might want to allow explicitly for lay-off risk, see Low, Meghir and Pistaferri (2010), for example. Second, is the nature of preferences, in particular the size of intertemporal elasticities and the degree of non-separability between consumption and family labor supply. In terms of elasticities, it is the intertemporal Marshallian elasticities (that measure the impact of an unexpected permanent shock to wages on labor supply) that are key, as this most likely describes the most difficult shock for families to insure. Moreover, this elasticity is also the key component in measuring the labor supply response to an unexpected but permanent shift in tax policy.

Non-separabilities are important because they emphasize that it is marginal utility that is smoothed not consumption itself. For example, families may not wish to smooth consumption if there are consumption costs of work. Consumption in this case will increase when family members enter work reflecting, for example, clothing and childcare costs that are required for work. These consumption items may not, on their own, produce utility! Eventually though we might expect utility constant consumption to fall back if all family members are at work or at school – there is no time for consumption. Income is saved and some consumption is put off to the future!

The recent work in Blundell, Pistaferri and Saporta-Eksten (2012) finds a key role for all these aspects. But it also finds that once taxation, family labor supply and savings are properly accounted for there is little room for other insurance. That is, for most families dealing with adverse shocks is a balance between using their savings, changing labor supply of other members of the household and the tax and welfare system. Figure V.2 displays this most clearly. It shows the average impact of an unexpected 10% permanent reduction in after-tax wages for the male worker in families drawn from a sample of PSID married couples. On average male workers earn around 70% of future discounted earned income and so there is a 7% decline in household consumption if there are no other insurance mechanisms available. Perhaps what is most important to draw from the Figure is the impact of family labor supply. For families with heads in the 30–40 age-range this reduces consumption loss to 4% on average. Indeed, for families in this age range there is little impact on savings. Savings grow in importance as families approach retirement age but still the labor supply of other family members is a key mechanism. These results are generated

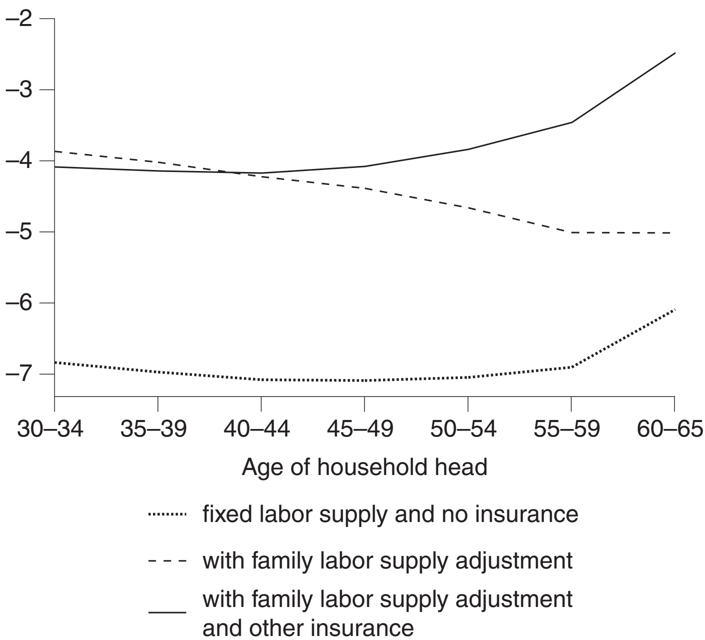


Figure V.2. Mechanisms of Insurance over the Life Cycle

Source: PSID and Blundell, Pistaferri and Saporta-Eksten (2012).

from estimated Marshallian labor supply elasticities of around 0.3 for women and zero for men. Nothing unusual about those (see Blundell and MaCurdy 1999, for example). Of course, there is a distribution of these affects across different types of families depending on their asset levels and their labor supply elasticities.

This work gives a glimpse at the kind of results that can be uncovered by placing family labor supply and savings decisions in an intertemporal model with wage uncertainty. The data requirements are more demanding and this work was made easier by the remarkable improvement in the consumption and assets data collected in the PSID since 1998. The results show clearly the value of family labor supply responses for younger families and especially those with limited access to assets. Of course, this does not come for free, the increased labor supply represents a decline in time for leisure and home production. Tax and welfare policies need to add to and complement these family mechanisms in a way that minimizes deadweight. This represents an exciting agenda for future research.

(iii) Restrictions on Labor Supply Choices

In Chapter 5 we saw how a job offer arrival rate could be superimposed on a life cycle consistent model of labor supply. However, this operated only at the extensive margin. That is, once an individual is in work that model assumed hours of work were chosen optimally. Chapter 3 considered the polar opposite case where, for some individuals in the family, hours of work were fully rationed. Even though in Chapter 7 we saw that some individuals adjust fairly rapidly to incentives in the tax and benefit system, it seems likely that, for some workers at least, hours of work reflect some limited choice. Perhaps the individual had to choose between two or maybe three alternative offers. Observed hours could then deviate quite substantially from the optimal choice according to a neoclassical model of labor supply. Moreover, short-run changes may be a poor guide to the longer run adjustment to tax and welfare reforms.

There is a large literature incorporating hours restrictions and frictions into labor supply models, including, Aaberge and Colombino (2008), Aaberge, Colombino and Strom (1999), Altonji and Paxson (1992), Bloemen (2008), Chetty et al. (2011a), Chetty et al. (2013), Dickens and Lundberg (1993), van Soest, Woittiez and Kapteyn (1990), Ham and Reilly (2002) and Dagsvik and Strom (2006). The idea of the new work I describe here, see Beffy et al. (2014), is to extend the standard structural labor supply model with nonlinear budget sets to a constrained choice setting in which the set of alternative choices on offer is restricted. This 'bounded rationality' framework concerns a general choice setting where the econometrician does not directly observe the choice set from which the individual has made his or her choices.

The underlying structural model is one where agents do not make their choices over the whole set of possible choices, but on a random subset of it. This choice model is placed in a life cycle setting developing the models outlined in Chapters 4–6. Hours of work, employment and savings decisions are made subject to a nonlinear tax system, fixed costs of work and an offer distribution. We draw on the extensive existing literature on labor supply models with nonlinear budget sets (Hausman 1980; Heckman 1974), with fixed costs of work (Heckman 1974; 1979; Cogan 1981), intertemporal choices (Heckman and MaCurdy 1980). The aim is to place these models in an environment in which individuals face constraints on the choice set of hours of work.

As empirical economists we typically do not know the complete set of alternatives available to a specific individual. We are not sure of their choice set. Moreover, different individuals may well face different constraint sets. Building on Dagsvik and Strom (2006) and Aaberge, Colombino and Strom (1999), Beffy et al. (2014) make the first steps to develop and estimate a structural model of labor supply that embeds restrictions on the distribution of available hours offers. Although not fully integrated within a model of job search (see Shephard 2012 for an important development in that respect), the approach does allow observed hours of work to deviate from rational choices in a way that reflects a structural wage offer distribution.

The key question is whether there are plausible assumptions under which the parameters of the underlying preference model are identified even though outcomes only reflect constrained choices. The nonlinearity of budget constraints induced by the tax and benefit system suggests that the offer distribution may indeed be identified without restrictions of preferences. For example, in Figure V.3, an observation on the 'horizontal' part of the constraint must rule out lower hours offers to that individual, as these would strictly dominate. Beffy et al. (2014) formalize this identification idea. They first show that if the offer distribution is known then preferences are non-parametrically identified. They then show the converse, if preferences are known the offer distribution is identified. In some cases, like the one in Figure V.3, the form of the budget constraint restricts possible choices and delivers identification of the offer distribution. In general, the distribution of

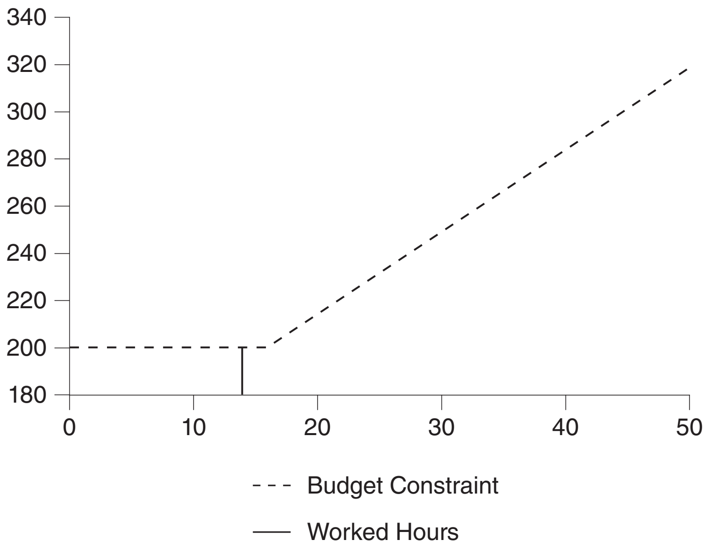


Figure V.3. Non-linear Budget Set Showing Non-Optimal Labor Supply Behavior

Source: Beffy et al. (2014).

offers can be unknown but is typically only fully identified if restricted to be a function of a finite set of unknown parameters.

Beffy et al. (2014) focus attention on developing and estimating a two-offer model in which each individual is assumed to face two independent hours offers – the one at which they are observed to work, if they are working positive hours, and one they turned down. The 'alternative' offer could include the observed hours point in which case the individual would be completely constrained and able to make no other hours choices. The option of not working is always available. As the number of offers increases the specification is shown to approach the neoclassical labor supply model at which observed choices coincide with the fully optimal choice over all hours options. The model is estimated on the same sample of UK lone parents as described in Chapter 14. These are key groups for this type of analysis. They face non-convex budget constraints and it is unlikely their observed hours of work will always reflect an optimal choice. The estimates point to a wage offer distribution with modes at part-time and full-time hours. The preference parameters imply significant 'long-run' labor supply elasticities at both intensive and extensive margins.

There is enormous scope for developing these models of partially restricted choices to incorporate job offer distributions into life cycle models of labor supply in the presence of non-linear budget constraints. They seem essential to fully understand how tax reforms work over the business cycle and in different local labor markets. As pointed out in Chapter 2, together with our understanding of human capital responses, they will surely also be a key part of understanding the relationship between macro and micro responses to taxes, see also Chetty et al. (2011b), Chetty et al. (2013), Keane and Rogerson (2012), and Rogerson and Wallenius (2009), and references therein. Further, these models may help explain some of the differences across countries in labor supply behavior, documented in, for example, Bargain, Orsini and Peichl (2014) and Ohanian, Raffo and Rogerson (2008) and Prescott (2004).

(iv) Further Developments

The directions for future research outlined above which incorporate into labor supply analysis *human capital accumulation*, *decisions in families*, and *restrictions on choices*, are but three exciting areas where we might expect new breakthroughs in our understanding. There are many more, including:

The social stigma of different benefits, the 'salience' of different taxes and the costs of 'take-up'. These are key areas where we need to improve our understanding. The 'single integrated benefit' reform of the Mirrlees Review (2011) was partly motivated by the low take-up and complexity of the interaction of many of the existing benefits in the UK welfare system. The current UK Universal Credit reform made extensive use of this research. Incorporating the salience and stigma of benefits and taxes into labor supply models will be an important advance in the evaluation and design of earnings tax reform.

- Social insurance and the contributory aspects of taxation and welfare is another field that has been rather sidelined recently. Yet it should move more centre stage. The design issues are closely related to new dynamic public finance models and fit very neatly within modern life cycle labor supply models that emphasize the role of uncertainty and credit constraints, see Farhi and Werning (2013), and Golosov, Troshkin, and Tsyvinski (2011), for example.
- The anticipation of tax reforms and the impact of policy announcements would seem another key avenue for research, see Blundell, Francesconi and van der Klaauw (2011), for a review.

Also Pistaferri (2003) for an important early reference on the use of expectations data in modelling dynamic labor supply responses. Tax policy making is a slow process in most modern democracies and reforms come into existence over many months and even years of debate and refinement. How do workers react to anticipations and announcements? How does this change the evaluation of tax reforms?

- There is much we still do not know about the nature and size of taxable income elasticities. The use of 'bunching at kinks' estimators, as championed by Saez (2010), is clearly a very fruitful approach, see also the incorporation of information in Chetty, Friedman and Saez (2013). Understanding taxable income elasticities, especially for top earners, remains a key factor in tax design. It is likely to remain a provocative and stimulating area of research, see for example Blomquist and Selin (2010), and Piketty, Saez and Stantcheva (2014), and references therein.
- The impact of tax and welfare incentives on migration and choice of jurisdiction for tax purposes is a key part of getting the design of the tax and welfare system right. They are a determinant of the taxable income elasticity and have increasingly important implications for tax and welfare design, see some initial discussion in Brewer, Saez and Shephard (2010).

Having put the chapters in this volume together and reflected on a number of new directions for research, it is not difficult to see why research in this area remains active. Indeed, it is these developments and remaining puzzles that continue to make the analysis of labor supply and taxation a compelling and, perhaps surprisingly, engaging area of empirical microeconomic research.

Notes

Introduction by the Editors

1. This concept has been proposed and analyzed by several IZA studies before, see, e.g., Bonin and Schneider (2006).
2. Those written in the 1980s mainly focus on estimations using the continuous labor supply model of Hausman (1981) and provide evidence essentially for individuals in couples (Pencavel 1986, for married men; Killingsworth and Heckman 1986, for married women). More recent surveys incorporate other methods (see Blundell and MaCurdy 1999) including life cycle models (see Meghir and Phillips 2008; Keane 2011; Keane and Rogerson 2012). Evidence on the elasticities of taxable income is surveyed in Meghir and Phillips (2008) and Saez, Slemrod and Giertz (2012). Evers, de Mooij and van Vuuren (2008) suggest a meta-analysis based on estimates for different Western countries, focusing essentially on those obtained with the traditional Hausman approach.

Chapter 1

1. See also Chetty (2009).
2. See Blundell, Bozio and Laroque (2011a) for a more detailed analysis of the composition effects.
3. See Adam, Browne and Heady (2010) for a more detailed overview.
4. Brewer et al. (2006) provide an excellent guide to the issues involved.
5. Background programs and data are available at http://www.ucl.ac.uk/˜uctp39a/lect.html, under Section III, Labour Supply and Tax Policy Simulation.
6. In the Supplementary Material to the Blundell and Shephard paper the sensitivity of the results to a finer discretization of weekly hours are examined; the main results appear robust.
7. Blundell and Shephard (2010) relax this to allow for partial observability of hours to capture the minimum hours conditions in the British tax credit system.
8. In this design problem we assume full take-up and do not include the stigma model used in estimating the preference parameters.

Chapter 3

1. Using a similar data set, but without explicitly modelling female labor supply, Blundell (1980) was unable to reject the separability of goods from male leisure. This indicates the crucial importance of female leisure in the rejection of separability.
2. Energy covers fuel, light and power, clothing includes footwear; transport includes vehicles; we have excluded housing expenditure, implicitly treating it as separable from all other decisions, and we treat expenditure on durables as current consumption. Note that the durables definition includes items which are "time saving" as well as "time using", so that it would be desirable to disaggregate this group.
3. It is simple to demonstrate that if we let the unrationed model have additive homoscedastic disturbances and integrate these back into the cost function, then the corresponding rationed model will have neither additive nor homoscedastic disturbances.

Chapter 4

1. This formulation of the life cycle problem assumes $AL = 0$ and abstracts from any bequest motive. However, the empirical formulation below is also consistent with an intertemporally separable bequest argument in the utility function but is unable to identify its parameters.
2. The functional form of (11) remains the same at a comer solution. However, the relevant price would be replaced by a virtual price, see Browning, Deaton and Irish (1985).
3. Such a form is a generalization of the life cycle model employed by Ashenfelter and Ham (1979).
4. In the context of a life cycle consumption see also Hall (1978) and Muellbauer (1983).
5. In a linear model it is possible, under certain circumstances (see Borjas 1980), to use the actual hourly wage as an instrument for the normal hourly wage providing a consistent estimator under measurement error. For our nonlinear selectivity adjusted model this may also be a useful approach but the properties of such an instrumental variable estimator would clearly need some further investigation.

Chapter 5

1. See also Cogan (1981) and Blank (1985). Heckman (1980) finds that his data reject the restrictions that the Tobit specification places on the constant term of the participation equation, but do not reject the restrictions that the Tobit model places on the slope coefficients in a linear participation equation.
2. Blank (1985) finds that a state unemployment rate is statistically significant in her reduced form hours per week and weeks per year equations.

However, as she notes, this variable may enter these equations simply as a determinant of the wage.

3. We allowed for a selection adjustment term in the wage equation for participants but this was found to be insignificant.
4. Further details of these tests are given in Blundell and Meghir (1986).
5. Omitting the industry dummy leads to an increase in the weight given to industry unemployment, vacancy and redundancy rates but does not affect the overall conclusions from this section. Similarly, the conclusions are little affected by the omission of the services and manual dummies from the labor supply model.
6. A worrying aspect of the results presented in Table 5.2b is the large normality test statistics. As these tests are conditional on independence they could simply be "picking up" the failure of that assumption. In future work we will investigate this issue by deriving appropriate test statistics for non-normality in the model with dependence.
7. In fact it is possible for the derivative (23) to be negative for workers even when labor supply is forward sloping. This could occur as $y + a(\cdot)$ becomes very small and the last term in the expression dominates.

Chapter 6

1. See, for example, Mroz (1987).
2. As a referee pointed out, the business cycle could generate some composition effects for which we do not account due to the conditioning on households with employed men only.
3. The key to the approach is the choice of groups. We discuss this later in detail.
4. Whether this is the case is hard to evaluate and depends heavily on where the woman lives and how good the public childcare provision is in the area.
5. See Heckman (1974), Gronau (1974), and Heckman (1979).
6. This is exactly analogous to conditioning out non-workers, which avoids modelling the participation decision. As discussed in Blundell, Duncan and Meghir (1992), this may also account for the presence of "optimization" errors whose distribution has bounded and limited support.
7. In practice this implies the exclusion of time-group interactions from the labor supply model. A full set of time effects and group effects are included.
8. In principle married and cohabiting couples face the same budget constraint. The FES distinguishes between the two types of households only after 1988.
9. The wage elasticity is the coefficient on the log wage divided by hours of work and the other income elasticity is the coefficient on other income divided by hours of work and multiplied by other income.
10. As a referee pointed out we would ideally have liked to compare the labor market trends of these two groups during a period of a stable wage structure

and the absence of reforms. The UK does not offer us such a chance since during the 70s a number of income policies, designed to compress wages, were implemented, as well as other tax reforms.

Chapter 7

1. See Blundell and MaCurdy (1999) for an overview.
2. Blundell and Hoynes (2004) and Brewer et al. (2006) present a comprehensive review of the evidence.
3. Card (1990) argues that constraints are the result of nonconvexities in the relationship between output and individual hours due to start-up costs or other aspects of the technology used.
4. There has been relatively little analysis of hours constraints in Britain. Two studies that have investigated the extent of constrains on desired hours are Stewart and Swaffield (1997) and Bryan (2007). Using data from the British Household Panel Survey, they both find that a substantial proportion of male workers (Stewart-Swaffield) and male and female workers (Bryan) are not putting in the hours they would like, with most of the dissatisfied workers wishing to work fewer hours per week. Both studies, however, abstract from the way in which job changes are related to hours changes, and, more broadly, from the issue of the path of labor supply adjustment.
5. Hours rules are an important feature of the UK's welfare system more generally. Receipt of the basic safety-net welfare benefit (Income Support or income-related Jobseeker's Allowance) is conditional on both working less than a certain number of hours and having a sufficiently low income. For parents, the hours rules for welfare benefits and in-work benefits are aligned so that families can never be entitled to both.
6. Since 1998, the transfer system affecting lone parents has undergone nearly continuous reform. However, the most important change, in terms of both government expenditure and potential labor supply effects, was the introduction of WFTC. We do not want to claim however that there has been a stable post-reform period since October 1999. On this and other related and concurrent policy initiatives, see the discussion in the next subsection.
7. It is worthwhile noticing that, for all three reforms, work incentives were likely to be dampened for single mothers living in areas with high childcare costs or high house rents (e.g., London and the South East of England). The availability of a more generous childcare tax credit component under WFTC might reduce this problem (Francesconi and van der Klaauw 2007), although high and increasing rents had to be weighed within the trade-off between additional tax credit gains and lower Housing Benefit entitlements (Gregg and Harkness 2007). In Section 7.4.2 we will present and discuss estimation results obtained after stratifying the sample by child's age, housing tenure, and region of residence.
8. In Section 7.4.1 we shall return to the definition of the post-reform periods. Brewer (2001) has a detailed time-line of reforms to in-work benefits

between 1971 and 2000. This does not reflect the reforms in April 2003 which lie outside our sample, and which are described in Brewer (2003).

9. The choice of single women without children as control group in our analysis is somewhat arbitrary. Albeit not eligible to receive FC or WFTC because they do not have children, these women are different from single mothers along a number of observable characteristics (see Section 7.3). Most of the existing studies on the effect of in-work benefits on lone mothers use the same control group as used here, whether they look at the UK (Blundell et al. 2000b; Blundell and Hoynes 2004; Francesconi and van der Klaauw 2007; Gregg and Harkness 2007) or the US experience (Eissa and Liebman 1996; Meyer and Rosenbaum 2001). Blundell and MaCurdy (1999) lay out the identification conditions for such an analysis and their credibility in the context of the analysis of tax reform is further discussed in Heckman's (1996) comment on Eissa (1996). In Section 7.4, however, we perform some sensitivity analysis in which the control group is restricted to single childless women with lower educational attainment.
10. In general, conditioning can be accomplished non-parametrically by combining matching and difference-in-differences. We find that this makes very little difference to our estimates, which condition linearly on covariates.
11. Eligibility to and provisions of the various New Deal schemes have slightly changed over time. In relation to NDLP, since 2002 lone parents are eligible to NDLP not only if they are in receipt of Income Support (as they were in previous years) but also if they receive other benefits (such as Housing Benefit and Council Tax Benefit) and, importantly, WFTC (as well as maternity allowance and statutory maternity pay). Also eligible are lone parents working under 16 hours per week (and thus ineligible to WFTC) who are not claiming any benefits, except child benefit. All these changes, however, were implemented outside our sample period.
12. Compulsory Work Focused Interviews (CWFI) for lone parents claiming Income Support were introduced in April 2001. Under CWFI people of working age seeking to claim Income Support are obliged to participate in a work-focused interview with an advisor at the start of their claim as a condition of receiving the benefit. Kirby and Riley (2004) find little evidence that CWFI increase labor market participation amongst inactive benefit claiming lone mothers.
13. The (levels of) time-varying variables are all measured at $t - 1$.
14. It is worth noting at this point that were firms to adjust their overall shift lengths in response to these changes in desired labor supply, there could be important spillover effects on other workers in our control group. We have not been able to locate any evidence either way on changes in shift length at the time of these reforms but it would clearly be useful to document evidence on this.
15. Women with zero hours at the time of any of the twelve interviews are excluded from the analysis. For further discussion on this point, see Section 7.3.

16. Of the individuals interviewed in 1991, 88 percent were re-interviewed in wave 2 (1992). The wave-on-wave response rates from the third wave onwards have been consistently above 95 percent (that is, 95 percent of the previous-wave respondents get interviewed). Detailed information on the BHPS is presented in Lynn et al. (2006) and can be obtained at http://www.iser.essex.ac.uk/ulsc/bhps/doc/. The households from the European Community Household Panel subsample (followed since the seventh wave in 1997), those from the Scotland and Wales booster subsamples (added to the BHPS in the ninth wave) and those from the Northern Ireland booster subsample (which started in wave 11) are excluded from our analysis.
17. Restricting our analysis to women who work for at least two consecutive periods leads to a sample of women who are more educated and less poor than those who are observed out of the labor market more frequently. But crucially the differences between treatment and control groups in this larger sample are very similar to those found in the more restricted sample used in the paper.
18. For non-British readers, "A (Advanced)-level" corresponds to education beyond high school, but short of a university degree.
19. Our measure of job change does not include internal promotions or job changes within the same firm or establishment, but includes all moves from one firm to another (either through quits or layoffs). Alternative definitions of job change (e.g., dropping laid-off workers from the pool of movers, or dropping promoted workers from the group of the stayers) produce similar results to those reported in this paper. See also Section 7.4.3.
20. To account for potential differential attrition over the panel and individual/item nonresponse in each specific wave, we recomputed group-specific means using weighted data (with either cross-sectional or longitudinal enumerated individual weights). The results (not shown) are very similar to those obtained with unweighted data and presented in Table 7.1, suggesting that the problems induced by panel attrition and changing sample composition are likely to be relatively small in our data. We shall return to some of these issues while performing sensitivity analysis (see Section 7.4.3).
21. For the sake of visual clarity, Figure 7.1 does not show the observations with more than 60 weekly hours. These, however, represent less than one percent of the subsamples in each survey period, and are included in the regression analysis reported below.
22. This contrasts with the estimates reported in Stewart and Swaffield (2004), which provide evidence of female labor supply reduction of 1-2 hours per week as a result of the introduction of the National Minimum Wage in April 1999. Their results are not robust across data sets and specifications, and are obtained from data that stop in September 2000 at the latest (that is, just before the second post-WFTC year in our sample). In addition, as pointed out in the Introduction, the Stewart-Swaffield estimates refer to all women, so we do not know how single women with and without children have been differentially affected by the minimum wage.

23. In parentheses, this and the subsequent tables report the absolute value of t-ratios obtained from standard errors that are adjusted to reflect multiple observations per person (and are robust to arbitrary forms of serial correlation and misspecification). For the sake of brevity, the estimates on the control variables are not reported but are available from the authors.
24. To understand this lack of effect, we estimated variants of equations (1)–(4) with Q interacted with marital status (not shown). Regardless of whether we control for group-specific time trends, changing job is associated with increases of about 1.2 hours per week for single mothers and with reductions of 0.8 hours per week for single childless women.
25. We reach the same conclusion if we keep the entire sample as in Table 7.2 but redefine the post-reform period as either 1999–2000 or 1999–2001. Similarly, redefining the FC period over 1992–93 (rather than 1992–94) does not alter our baseline results.
26. These estimates are also accompanied by a significantly different impact of changing job on hours changes for the two groups of women (α_2). When moving from one job to another, women in the bottom part of the hours distribution faced an average increase of nearly 2 hours per week, while women in the top part of the distribution reduced their labor supply by about 1 hour per week.
27. If a large proportion of better-educated single mothers had not been eligible to WFTC, the effects reported in Table 7.2 should be attributed to shocks other than WFTC. However, using data from the Family Resources Survey, we find that tax credit eligibility has increased proportionally more for more educated lone mothers than for the less educated after the introduction of WFTC (albeit a greater fraction of the less educated are eligible). In particular, between 1995 and 1998, about 26 percent of better educated lone mothers who work 16 or more hours per week were eligible to FC. Between 2000 and 2002, 49 percent were eligible to WFTC (an increase in eligibility rate by 88 percent). For the less educated, the increase in eligibility rate was only 28 percent (from 65 to 83 percent).
28. Stratifying the sample jointly by region and house tenure leads to small subsamples. But when we performed the analysis on the entire sample and included an interaction term between these two variables, the largest increases in worked hours occurred in association with changing job after the introduction of WFTC for single mothers who lived in rented accommodations outside the London/South-East region.
29. Single mothers who were employed in manufacturing industries also showed a significant increase of 3 hours of work per week if they changed job after the introduction of WFTC (panel F). For the same group of women there is also evidence (significant only at the 10 percent level) of positive labor supply adjustments of about 1.6 hours per week if they changed job between 1992 and 1994 (i.e., during the FC regime). This effect involves only 25 percent of the whole sample, and this may be why it does not show up in the baseline estimates of Table 2 for the whole sample. Manufacturing

production is based on technologies that are traditionally less flexible than those used in services, such as batch methods and robotized assembly lines (Goldin and Katz 1998), which may be reflected in a greater rigidity in (downward) adjustments in hours.

30. Over the whole sample period, about 19 percent of lone mothers report being overemployed, 18 percent underemployed, and the remaining 62 percent report being satisfied with their hours of work. The corresponding proportions for single women without children are 28, 11, and 61 percent. Considering all women in the sample, the most mobile are the underemployed (with 27 percent of them changing job in any two consecutive years), while the job changing rates for the overemployed and the other group of workers are lower (19 and 15 percent respectively).
31. The hypothesis that the estimated b and β coefficients are equal can be rejected at the 5 percent level (p-value = 0.027).
32. Notice that "unconstrained" are defined to be those workers who would like to continue to work the same number of hours. This definition may not precisely reflect their entire preference ordering, since they may be constrained in other dimension (e.g., job location and family responsibilities).
33. Following Altonji and Paxson (1992) and Euwals (2001), we also checked whether the hours adjustments estimated in conjunction with the WFTC reform are in line with women's stated preferences. The results (which are not reported for convenience) show that this is the case, especially for underemployed lone mothers. Almost 80 percent of single mothers who wanted to work more did adjust their hours upward by changing job after the 1999 reform as opposed to only 30 percent among those who did not change job. The corresponding downward adjustments for women who wanted to work fewer hours were instead 55 and 18 percent for movers and stayers respectively.
34. We also reestimated the models eliminating laid-off workers from Q, or dropping promoted workers from the group of the stayers. Both these exercises produced results that were virtually identical to those shown in Table 7.2, and are thus not reported.

Introduction to Part IV

1. The suggestion that the longer term pay-off to part-time work may be limited is confirmed in the recent more comprehensive study Blundell et al. (2013).

Chapter 8

1. We choose not to impose an equivalence between hours of work and hours of childcare; empirical evidence in Duncan, Giles and Webb (1995) suggests equivalence to be an unreasonable assumption. For this illustrative

example, and for the simulation exercise that follows, we assume a linear relationship between childcare hours and hours of work. The form of the linear relationship is estimated from observed patterns of childcare use in the FRS for a range of households of different demographic types. Details of these estimated relationships between childcare hours and hours of work are given in Table 8A.2 of the appendix.

2. This is not universally true, however. If the man were to work 30 rather than 40 hours, then examination of Figure 8.5 shows the income differential rising initially as the woman moves into work, before ultimately falling once household income reaches the pre-reform cut-out point for Family Credit.
3. See Duncan, Giles and Webb (1995).
4. A point well documented in Dilnot and Giles (1998), for example.
5. Based on observed patterns of work and childcare use among working women.
6. See Moffitt (1992) for a survey of incentive responses to welfare reform in the US, and Dickert, Hauser and Scholz (1995), Eissa and Hoynes (1998) or Eissa and Liebman (1996) for an evaluation of the labor market responses to changes in the US system of Earned Income Tax Credit (EITC). For the UK, see Dilnot and Duncan (1992) for a simulation study of the 16-hour-rule reform in Family Credit.
7. The discrete approach to estimation has become increasingly popular in recent literature. See Blundell and MaCurdy (1999) for a survey and Bingley and Walker (1997), Duncan and Giles (1998), Hoynes (1996), Keane and Moffitt (1998) and van Soest (1995) for specific applications of discrete choice methods to labor supply analysis.
8. This requires re-sampling techniques, since to operationalize the computations of the new probabilities involves drawing repeated realisations of the stochastic elements of the discrete choice model laid out in earlier sections. See Duncan and Weeks (1998) for a more detailed explanation.
9. This standard approach for dealing with missing wages is discussed in Blundell et al. (2000a).
10. See Blundell and Reed (1999) for a more detailed comparison of the two approaches.
11. Since this fifth scenario requires full re-estimation of the structural model, we restrict the reported comparison to an illustrative sample of single-parent households only.

Chapter 9

1. See Dilnot and Duncan (1992), for example.
2. Gregg et al. (1999) emphasize this point.
3. See Blundell and MaCurdy (1999).
4. Gregg, Johnson and Reed (1999) consider a number of alternative methods for 'imputing' the wage for new entrants. Their preferred method is to use

the wages of recent entrants in the UK Labour Force Survey. As one might expect, these are often well below the wages of similar individuals in work, and suggest that the use of wages of workers without correction could lead to a biased assessment of a policy reform. However, for lone parents at least, they report little difference between entry wages and the average of wages for all those in work. The standard reform simulation approach, as used in the figures reported in Section 9.4 below, uses wages adjusted for selection. These will also typically be lower than the observed wages for those currently in work (see Blundell et al. 2000b, for a discussion of this approach).
5. The discussion in this section draws heavily on the excellent review by Eissa and Hoynes (1998), where further details can be found.
6. See Eissa and Liebman (1996) for a more extensive discussion of EITC rules.
7. See Blundell, Duncan and Meghir (1998) for the precise derivation of these conditions.
8. Liebman (1998) provides an update of this study and shows similar effects through to 1996.
9. Introducing a stigma cost to participation in WFTC allows the simulation model to predict a low probability of take-up among those with low eligibility, something found in earlier studies of welfare program take-up. Moreover, it suggests a higher take-up of WFTC (in contrast to FC) for those whose eligible amount of credit increases as a result of the WFTC reform.
10. It is worth pointing out that the movement into work requires the individual receiving a suitable job offer. Since this is a probability-based model, the impact of a low job offer rate is reflected in a lower participation probability. In the simulation model a job offer probability is included and allowed to depend on regional labor market characteristics and the characteristics of the individual, following the model of search and discouraged workers developed in Blundell, Ham and Meghir (1987).
11. See Blundell et al. (2000b) for details.
12. See, for example, the discussion of the design of the EITC in Meyer and Rosenbaum (2001).
13. See Paull, Walker and Zhu (2000).
14. See Dickert and Houser (1998), Ellwood (2000), and Moffitt (1998).
15. See Dickert and Giles (1996) for a detailed discussion of childcare availability.

Chapter 10

1. LaLonde (1986) is perhaps the most influential paper expressing this view (see Heckman and Hotz 1989 for an early riposte). Recently, Dehejia and Wahba (1998; 1999) have argued that careful matching using propensity score methods can overcome many of the problems with conventional non-

experimental estimators and sought to demonstrate this using LaLonde's original data on the National Supported Work (NSW) program. Smith and Todd (2005), however, showed that such "success" came from discarding a large proportion of the original NSW data and that cross sectional matching estimators remained highly sensitive on the full sample. As with our own results presented here Smith and Todd found difference-in-differences estimators are the most robust.

2. This is the main British form of unemployment insurance (see Appendix D).
3. For more information about training programs in Britain and their effects see, for instance, Dolton, Makepeace and Treble (1992), Blundell, Dearden and Meghir (1996) and Blundell et al. (1997).
4. See Heckman (1979), Heckman and Robb (1986), Blundell, Duncan and Meghir (1998), Bell, Blundell and Van Reenen (1999) and Blundell and Costa Dias (2000) for precise descriptions of these conditions. Davidson and Woodbury (1993) is an example of an attempted calibration of substitution effects using data from the Illinois unemployment insurance (UI) experiments (see also Woodbury and Spiegelman 1987, on this program).
5. See Card and Hyslop (2005) for evidence of the absence of dynamic effects in the Canadian Self-Sufficiency Project.
6. For example, Knab et al. (2000).
7. On job assistance see the survey by Meyer (1995); on wage subsidies see Katz (1998).
8. JSA is the main form of unemployment benefit in the UK. It is essentially a flat rate benefit paid every two week of about £40 ($60) a week. Past work experience is not a condition of receipt of JSA and although there is a requirement to "actively seek employment", it is not time limited. See Appendix D for details.
9. Note that certain groups of especially disadvantaged individuals (e.g. the disabled, ex-convicts, those with basic skills problems, etc.) are allowed to enter the New Deal earlier than six months if they wish. Additionally, in the early stages of the program those individuals on JSA for over six months were only obliged to enter the New Deal as they reached their 12th, 18th, 24th month, etc of JSA (unless they choose to be early entrants). We are careful to control for these "early entrants" in the work below.
10. This is quite generous. Hales et al. (2000) find that the mean starting wage for those on a subsidized job is £3.78 an hour, implying a forty per cent level of subsidy for a thirty-seven hour week.
11. The intention was that the treatments were staged. The employment service would seek to place an individual in an unsubsidized job in the first month of the program, a subsidized job in the second month, in education/training in the third month and the Environmental Taskforce in the fourth month. This guidance was not strictly enforced on the ground, however.
12. See Anderson, Riley and Young (1999).

13. Our data currently ends in July 1999. Individuals entering the Gateway in April 1998 and joining the year-long education and training option after four months will only start job search in August 1999.
14. For example, Heckman, Lochner and Taber (1998a).
15. See Hahn, Todd and Van der Klaauw (2001).
16. The matching method we use smoothes the counterfactual outcomes either with a Kernel based method or with splines (see, Heckman, Ichimura and Todd 1997). We also present results based on the nearest neighbour weighting scheme. These however turn out to be much less precise.
17. We also considered more finely disaggregated age groups – e.g. 24 vs. 25 year olds – which generates similar qualitative results, although with much less precision on account of the smaller sample size.
18. This analysis is also informative on whether the assumptions on the comparability between any two groups being used are valid. In fact, before the introduction of the New Deal the estimated impacts are expected to be zero given the absence of a policy that causes a differential behavior between any two groups being compared. If, however, a large number of point estimates are found to be significantly different from zero, one might suspect that the assumptions on the comparability of the two groups being used are not valid.
19. There is a code in the JUVOS data which purports to have New Deal destinations, but on investigation it proved to be unreliable.
20. One could also worry about 18-22 year olds in college education. There is only a tiny fraction of this group in the unemployed pool, however.
21. Britain had never had a national minimum wage before this date. There was a system of Wages Councils that set minimum wages for certain groups of occupations in low-wage industries. These only covered about two million of the approximately 30 million UK workforce when they were abolished in 1993 (see Dickens, Machin and Manning 1999, for an analysis).
22. All regressions include a set of other covariates, including age (when similar age groups are being compared), marital status, region, sought occupation and labor market history variables. All computations have been performed excluding these covariates as well. Given the similarity of the results, however, we skip their presentation.
23. Appendix A presents some comparisons between treatment and comparison groups with respect with some of the covariates being considered, including a few checks on the quality of the propensity score matching.
24. See Van Reenen (2001) for discussion of Restart and the introduction of JSA.
25. See Eissa and Leibman (1996) for an evaluation.
26. See Blank, Card and Robins (2000) for example.
27. Katz (1998). See also Burtless (1985) and Dubin and Rivers (1993) for evaluations of wage subsidy programs.
28. Other comparisons are available and can be provided under request.

Chapter 11

1. Here we refer explicitly for the New Deal for Young People, directed towards 18-24 year olds with at least six months unemployment. However, there are now similar policies in the UK directed toward those on disability benefit (New Deal for Disabled People), for those aged over twenty five – 25 Plus, for Lone Parents and for older workers – 50 Plus. Although different, each have similar characteristics and are subject to similar design issues.
2. The "scaring" effect of spells of unemployment and welfare is also raised as a further deterrent to work (see Gregg and Wadsworth 1999).
3. See Eissa and Liebman (1996).
4. See Gradus and Jusling (2001), who also review similar schemes and proposals in Germany, the Netherlands, Ireland and Finland.
5. See Holcomb et al. (1998) for a review of these schemes. In particular the Work Mandate designs which are very close to the design of the New Deal.
6. See Card and Robins (1998).
7. See Bos et al. (1999).
8. Quets et al. (1999) provide a careful evaluation of the effect of adding job search services to the SSP. This evidence is used later in our discussion of job search assistance in financial incentive programs.
9. See Miller et al. (1997). Continuation of the MFIP in work is conditional on accredited training for workers who do not have children under one year old and who are in jobs of less than 30 hours per week.
10. Inland Revenue (2001).
11. These are sometimes referred to as the "windfall beneficiaries" of the program.
12. The proportion of NEETs (not in employment, education or training), sometimes referred to as the "idle", was 8.4% in the UK in 1997 compared to 2.3% in 1984. In 1997 the corresponding figure was 5.6% in the US, 4.2% in Germany, 3.3% in France and 9.1% in Italy (see Blanchflower and Freeman 2000).
13. See Nickell (1999) for an extensive review of the British data. Hoynes (2000) also notes a strong degree of sensitivity to the cycle among young welfare recipients and low-skilled workers in the US.
14. These figures are drawn from the repeated cross-sections of the British Family Expenditure Survey. As such they refer to different people over time and will therefore exhibit systematic composition changes according to birth cohort, education and other factors.
15. See Dickens (2000) and the references therein.
16. See Blundell and Preston (1998).
17. This is quite generous. The mean starting wage for those on a subsidized job is £3.78 an hour, implying a 40 per cent level of subsidy for a 37 hour week.
18. Once the option period is over, if the individual has not managed to keep/find a job or leave the claimant count for any other reason, the third stage

of the program is initiated, the Follow Through. This is a process to the Gateway, taking up to 13 weeks, where job search assistance is the main treatment being provided.

19. See Blundell and Hoynes (2004) for a brief historical review.
20. See Brewer (2001).
21. See Dickert-Conlin and Holz-Eakin (2000).
22. See Pavetti et al. (2001).
23. See Dilnot and Giles (1999) for a discussion of the impact of the WFTC reform on implicit tax rates.
24. This is the second "peak" in Figure 9.7.
25. See MaCurdy and McIntyre (2004), for example.
26. As noted already, empirical evidence on wage progression for specific types of workers is sadly lacking. Reliable evidence for low-skilled workers is dogged by selection and attrition problems as highlighted in the study of age growth in the SSP treatment population by Card, Michalopoulos and Robins (2001).
27. See Katz (1998).
28. See Sianesi (2001).
29. As already noted, there is no reason why continuation of the credit could not be conditional on various advancement conditions oriented toward wage progression such as job search and accredited training.
30. Dolton and O'Neill (1996).
31. See Abbring, van Berg and van Ours (1996) for further evidence on the effectiveness of sanctions on transition rates into employment from unemployment.
32. See Abbring, van Berg and van Ours (1996) for further evidence on the effectiveness of sanctions on transition rates into employment from unemployment.
33. Miller et al. (1997), and the discussion in Card and Blank (2000a).
34. See Riccio and Bloom (2001), for example.
35. See also the design of the New Deal for Long Term Unemployed, 25 Plus, analyzed in Lissenburgh (2001).
36. See Blundell, Duncan and Meghir (1998).
37. See Blundell and MaCurdy (1999), for a review.
38. See Blundell and Costa Dias (2000).
39. Bonjour et al. (2001) and Dorsett (2001) provide a detailed description of the option and post-option results.
40. See Van Reenen (2004) for a careful cost-benefit analysis of the New Deal.
41. Blundell et al. (2001) assess the importance of the estimated effects and interpret them in an historical perspective. They provide some lower and upper bounds for the treatment effect by using other pre-program time periods. This can be done for total outflow for all years since 1982.
42. See Adnett and Dawson (1996) for a look at European schemes and Dickert-Conlin and Holz-Eakin (2000) for an extensive review of the issues.

43. All regressions in Table 9.1 and Table 9.2 include a set of other controls, including age (when similar age groups are being compared), marital status, region, sought occupation and labor market history variables. All computations have been performed excluding these covariates as well. Given the similarity of the results, however, we skip their presentation.
44. Blundell et al. (2001) presents some comparisons between treatments and controls with respect to some of the covariates being considered, including a few checks on the quality of the propensity score matching.
45. Bartik (2000) finds relatively small displacement and substitution effects for the recent US welfare reforms, despite the large declines in caseload.
46. See the surveys by Katz (1998) and by Meyer (1995). Also a round up of the evidence on wage subsidy programs in Bell, Blundell and Van Reenen (1999).
47. This work develops earlier structural labor supply simulation models by Hoynes (1996), for example. In particular, it allows for childcare demands to vary with hours worked and it allows for fixed costs of work. It also accounts for take-up by incorporating welfare stigma following on from Keane and Moffitt (1998).
48. Interestingly a similar "unintended" adverse effect on employment rates among married women has been documented for the EITC expansions in the US, see Eissa and Hoynes (1998).
49. Department of Social Security, Client Group Analysis.
50. There has also been a large increase in take-up of the Childcare Tax Credit compared to the childcare disregard under Family Credit. 111,000 families were receiving help with childcare costs in May 2000, a 156% increase over 12 months. The average amount of costs claimed was £32 a week. But although a large increase, this is still only 10% of the total WFTC caseload.
51. See Eissa and Liebman (1996) for use of similar control groups in the evaluation of the EITC reforms. As mentioned above this type of before and after evaluation is often termed a "natural experiment".
52. One caveat to this is the possible impact of childcare credit. Under WFTC this is a generous scheme available only to those in work (requiring both parents in a couple to work at least 16 hours) but it is currently taken up by only a small fraction of WFTC recipients. If participation in this part of the WFTC program was to expand significantly it could further encourage labor supply among those low-income parents who are currently out of work and claiming Income Support.
53. See also the discussion in Dickert-Conlin and Holz-Eakin (2000).
54. Meyer (1995) highlights the potential for cycling effects.
55. The review of such schemes by Meyer (1995) finds a significant (and cost effective) impact of mandatory job search assistance schemes operating in the late 1980s in the US.
56. For two recent important studies see Ermisch and Francesconi (2000) and Morris and Michalopoulos (2000).

Chapter 12

1. Although earned income tax credit policies have been implemented for couples in the UK and in the US, I will focus on the design of such policies for lone parents. The issues are similar but, for couples, we need to consider a model for joint labor supply decisions in families. A number of alternative models exist; see Blundell, Brewer and Shephard (2006) for example. However, the implementation and estimation of collective models for the analysis of tax and welfare reform is still in its infancy, see Bargain et al. (2006).
2. Gosling, Machin and Meghir (2000).
3. Brewer and Gregg (2001).
4. See Eissa and Liebman (1996).
5. See Gradus and Julsing (2001), who also review similar schemes and proposals in Germany, the Netherlands, Ireland and Finland.
6. See Holcomb et al. (1998) for a review of these schemes. In particular the Work Mandate designs which are very close to the design of the New Deal.
7. See Card and Robins (1998).
8. See Bos et al. (1999).
9. Quets et al. (1999) provide a careful evaluation of the effect of adding job search services to the SSP. This evidence is used later in our discussion of job search assistance in financial incentive programs.
10. See Miller et al. (1997). Continuation of the MFIP in work is conditional on accredited training for workers who do not have children under one year old and who are in jobs of less than 30 hours per week.
11. These figures are drawn from the repeated cross-sections of the British Family Expenditure Survey. As such they refer to different people over time and will therefore exhibit systematic composition changes according to birth cohort, education and other factors.
12. An important start on this has been made in Kleven, Kreiner and Saez (2006) and for a discussion of collective models see Bargain et al. (2006).
13. See Blundell (2002) and references therein.
14. See Moffitt (1983).
15. See also Chone and Laroque (2005), Beaudry and Blackorby (2000), Liebman (2002).
16. See Blundell and MaCurdy (1999).
17. These are summarized here, for a full description see Blundell et al. (2000b) and Brewer et al. (2006).
18. Up to £80, between £80 and £140, between £140 and £220, and £220 and above respectively.

Chapter 13

1. The Mirless Review provides a recent overview of UK earnings tax design (Brewer, Saez and Shephard 2010).

2. Hours conditions are used in the tax credit systems in Ireland and New Zealand. They are also proposed in Keane (1995), although not within an optimal tax framework.
3. An alternative model which incorporates constraints on the labor supply choices in an optimal design problem is developed in Aaberge and Colombino (2008).
4. Depending on the size of the interval, this framework nests two important special cases; (i) when hours are perfectly observable $\underline{h} = \overline{h} = h$ for all $h \in H$; (ii) only earnings information is observed $h = H_{++}$ for all $h > 0$. In general this is viewed as a problem of partial observability since actual hours h are always contained in the interval h. In our later analysis in section 7.4 we will explore the effect that both random hours measurement error, and possible hours misreporting have upon the optimal design problem.
5. Conditional on work hours h, consumption will not depend on ε given our assumption that ε enters the utility function additively and is independent of (X, ϵ).
6. For many families with children, these increases in out-of-work income meant that replacement rates remained relatively stable.
7. The potential dependence of childcare expenditure on T has been suppressed for simplicity.
8. The total hours elasticity η_r is related to the intensive and extensive elasticities (respectively η_e and η_i) according to $\eta_t = \eta_i + (Q/P) \times \eta_e$. Here, P denotes the employment rate, and Q is the ratio of average hours of new workers, relative to the initial average hours of existing workers.
9. A large participation (extensive) elasticity and a relatively small intensive elasticity have been reported in other studies, see Blundell and Macurdy (1999). A useful recent reference is Bishop, Heim and Mihaly (2009) who report a (fitted) intensive elasticity of 0.05 in 2003, as well as a (fitted) participation elasticity of 0.25 in the same year, for single mothers in the US.
10. These marginal tax rates are again restricted to lie between –100% and 100%, but now apply to weekly earnings from £0 to £400 in increments of £50, and then all weekly earnings above £400.
11. Comparing actual tax schedules to the optimal schedules from Table 14.6 is complicated as the actual systems vary in multiple dimensions. Broadly speaking, the optimal tax schedule (when $\theta = -0.2$) has higher (lower) values of out-of-work support than the actual April 2002 system for families with low (high) values of housing rent and Council Tax. For low values of earnings we obtain lower marginal tax rates (except at very low earnings due to an income disregard in Income Support). For lone mothers with moderate wages we obtain lower marginal tax rates over a large range of earnings.
12. The hours bonus is sufficiently large that a mother earning the minimum wage would face an effectively zero participation tax rate at 19 hours.
13. The welfare gains from a part-time hours rule are also small if we condition by the age of children as described in Section 14.7.3. And while the welfare

gains from an optimally placed (full-time) hours rule are also small for mothers with pre-school aged children, these gains are found to be much more substantial for parents with school age children. Full results are available upon request.

14. In practice misreporting costs are likely to vary with both observed and unobserved worker characteristics. While it is sufficient to model this as a single cost for the purpose of our discussion and simulations here, our framework can easily be extended to incorporate such heterogeneity.
15. The misreporting cost b is measured relative to standard deviation of the state specific error ε. With an hours bonus payable at 33 hours per week (for example), a value of $b = 0.16$ would mean that the utility cost of reporting 33 hours when actual hours are 26 is equivalent to a $0.16 \times (33 - 26) = 1.12$ standard deviation change in the realization of the state specific error.

References

Introduction by the Editors

Bargain, O., Peichl, A. (2013). Steady-State Labor Supply Elasticities: A Survey, ZEW Discussion Paper No. 13-084.

Bargain, O., Orsini, K., Peichl, A. (2014). Comparing Labor Supply Elasticities in Europe and the US: New Results, in: Journal of Human Resources, 49(3): 723–838.

Blundell, R., Ham, J., Meghir, C. (1987). Unemployment and Female Labour Supply, in: Economic Journal, 338a(97): 44–6.

Blundell, R. (1993). Taxation and Labour Supply Incentives in the UK, in: Atkinson, A. B., Morgensen, G. V. (Eds.), Welfare and Work Incentives: A North European Perspective, Oxford, UK: Clarendon Press, 135–60.

Blundell, R., Chiappori, P.-A., Meghir, C. (2005). Collective Labour Supply with Children, in: Journal of Political Economy, 113(6): 1277–306.

Blundell, R., Chiappori, P.-A., Magnac, T., Meghir, C. (2007). Collective Labour Supply: Heterogeneity and Non-Participation, in: Review of Economic Studies, 74(2): 417–45.

Blundell, R., Duncan, A., McCrae, J., Meghir, C. (2000). The Labour Market Impact of the Working Families' Tax Credit, in: Fiscal Studies, 21(1): 75–103.

Blundell, R., MaCurdy, T. (1999). Labor Supply: A Review of Alternative Approaches, in: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, Vol. 3A, Amsterdam: Elsevier, 1559–695.

Blundell, R., MaCurdy, T., Meghir, C. (2007). Labor Supply Models: Unobserved Heterogeneity, Non-participation and Dynamics, in: Heckman, J., Leamer, E. (Eds.), Handbook of Econometrics Vol. 6A, Amsterdam: Elsevier, 4667–775.

Blundell, R., Meghir, C., Neves, P. (1993). Labour Supply and Intertemporal Substitution, in: Journal of Econometrics, 59(1–2): 137–60.

Blundell, R., Meghir, C., Symons, E., Walker, I. (1988). Labour Supply Specification and the Evaluation of Tax Reforms, in: Journal of Public Economics, 36(1): 23–52.

Blundell, R., Walker, I., Bourguignon, F. (1988). Labour Supply Incentives and the Taxation of Family Income, in: Economic Policy, 3(6): 133–161.

Bonin, H., Schneider, H. (2006). Workfare: Eine wirksame Alternative zum Kombilohn, in: Wirtschaftsdienst, 86(10): 645–50.

Dolls, M., Doorley, K., Schneider, H., Sommer, E. (2013). Structural and Behavioral Dimensions of Labor Supply in Europe in 2030, NEUJOBS Working Paper D10.9.

Eichhorst, W., Hinz, T., Marx, P., Peichl, A., Pestel, N., Siegloch, S., Thode, E., Tobsch, V. (2012). Geringfügige Beschäftigung: Situation und Gestaltungsoptionen, IZA Research Report No. 47.

Eissa, N., Kleven, H. J., Kreiner, C. T. (2008). Evaluation of Four Tax Reforms in the United States: Labor Supply and Welfare Effects for Single Mothers, in: Journal of Public Economics, 92(3–4): 795–816.

Evers, M., de Mooij, R., van Vuuren, D. (2008). The Wage Elasticity of Labour Supply: A Synthesis of Empirical Estimates, in: De Economist, 156(1): 25–43.

Hausman, J. (1981). Labor Supply, in: Aaron, H. J., Pechman, J. A. (Eds.), How Taxes Affect Economic Behavior, Washington D.C.: The Brookings Institution, 27–72.

Heim, B. T. (2007). The Incredible Shrinking Elasticities: Married Female Labor Supply, 1978–2002, in: Journal of Human Resources, 42(4): 881–918.

Hoynes, H. W. (1996). Welfare Transfers in Two-Parent Families: Labor Supply and Welfare Participation under AFDC-UP, in: Econometrica, 64(2): 295–332.

Immervoll, H., Kleven, H. J., Kreiner, C. T., Verdelin, N. (2011). Optimal Tax and Transfer Programs for Couples with Extensive Labor Supply Responses, in: Journal of Public Economics, 95(11): 1485–500.

Keane, M. P. (2011). Labor Supply and Taxes: A Survey, in: Journal of Economic Literature, 49(4): 961–1075.

Keane, M., Rogerson, R. (2012). Micro and Macro Labor Supply Elasticities: A Reassessment of Conventional Wisdom, in: Journal of Economic Literature, 50(2): 464–76.

Keane, M., Moffitt, R. (1998). A Structural Model of Multiple Welfare Program Participation and Labor Supply, in: International Economic Review, 39(3): 553–89.

Killingsworth, M., Heckman, J. J. (1986). Female Labor Supply: A Survey, in: Ashenfelter, O., Layard, R. (Eds.), Handbook of Labor Economics, Vol. 1, Amsterdam: Elsevier, 103–204.

Löffler, M., Peichl, A., Pestel, N., Schneider, H., Siegloch, S. (2012). Effizient, Einfach und Gerecht: Ein Integriertes System zur Reform von Einkommensteuer und Sozialabgaben, in: Perspektiven der Wirtschaftspolitik, 13(3): 196–213.

Löffler, M., Peichl, A., Siegloch, S. (2014). Structural Labor Supply Models and Wage Exogeneity, IZA Discussion Paper No. 8281.

Meghir, C., Phillips, D. (2008). Labour Supply and Taxes, The Institute for Fiscal Studies, Working Paper No. 08/04.

Peichl, A., Schneider, H., Siegloch, S. (2010). Documentation IZAΨMOD: The IZA Policy Simulation Model, IZA Discussion Paper No. 4865.

Peichl, A., Siegloch, S. (2012). Accounting for Labor Demand Effects in Structural Labor Supply Models, in: Labour Economics, 19(1): 129–38.

Pencavel, J. (1986). Labor Supply of Men: A Survey, in: Ashenfelter, O., Layard, R. (Eds.), Handbook of Labor Economics, Vol. 1, Amsterdam: Elsevier, 3–102.

Saez, E. (2001). Using Elasticities To Derive Optimal Income Tax Rates, in: Review of Economic Studies, 68(1): 205–29.

Saez, E., Slemrod, J. B., Giertz, S. H. (2012). The Elasticity of Taxable Income with Respect to Marginal Tax Rates: A Critical Review, in: Journal of Economic Literature, 50(1): 3–50.

van Soest, A. (1995). Structural Models of Family Labor Supply: a Discrete Choice Approach, in: Journal of Human Resources, 30(1): 63–88.

Zimmermann, K. F. (1993). Labour Responses to Taxes and Benefits in Germany, in: Atkinson, A. B., Morgensen, G. V. (Eds.), Welfare and Work Incentives: A North European Perspective, Oxford, UK: Clarendon Press, 135–160.

Bibliography

Aaberge, R., Colombino, U. (2008). Designing Optimal Taxes with a Microeconometric Model of Household Labour Supply, CHILD Working Paper No. 06/08.

Aaberge, R., Colombino, U., Strom, S. (1999). Labour Supply in Italy: An Empirical Analysis of Joint Household Decisions, with Taxes and Quantity Constraints, in: Journal of Applied Econometrics, 14(4): 403–22.

Abbott, M., Ashenfelter, O. (1976). Labour Supply, Commodity Demand and the Allocation of Time, in: Review of Economic Studies, 46(3): 389–412.

Abbring, J., van den Berg, G., van Ours, J. (1996). The Effect of Unemployment Insurance Sanctions on the Transition Rate From Unemployment to Employment, Serie Research Memoranda No. 0038.

Adam, S., Brewer, M. (2005). Take-up of Family Credit and Working Families' Tax Credit: final report, Open Access publications from University College London http://discovery.ucl.ac.uk/18448/.

Adam, S., Browne, J. (2006). A Survey of the UK Tax System, Open Access publications from University College London http://discovery.ucl.ac.uk/14872.

Adam, S., Browne, J., Heady, C. (2010). Taxation in the UK, in: Mirrlees, J., Adam, S., Besley, T., Blundell, R., Bond, S., Chote, R., Gammie, M., Johnson, P., Myles, G., Poterba, J. (Eds.), Dimensions of Tax Design: The Mirrlees Review, New York: Oxford University Press, 1–89.

Adnett, N., Dawson, A. (1996). Wage Subsidies and European Unemployment: Theory and Evidence, in: Economic Issues, 1: 1–21.

Akerlof, G. A. (1978). The Economics of "Tagging" as Applied to the Optimal Income Tax, Welfare Programs, and Manpower Planning, in: American Economic Review, 68: 8–19.

Altonji, J. G. (1986). Intertemporal Substitution in Labour Supply: Evidence from Micro-Data, in: Journal of Political Economy, 94(3): 176–215.

Altonji, J. G., Paxson, C. H. (1988). Labor Supply Preferences, Hours Constraints, and Hours-Wage Tradeoffs, in: Journal of Labor Economics, 6(2): 254–76.

Altonji, J. G., Paxson, C. H. (1992). Labor Supply, Hours Constraints, and Job Mobility, in: Journal of Human Resource, 27(2): 256–78.

Amemiya, T. (1973). Regression Analysis when the Dependent Variable is Truncated Normal, in: Econometrica, 41(6): 997–1017.

Amemiya, T. (1974). Multivariate Regression and Simultaneous Equation Models When the Dependent Variables are Truncated Normal, in: Econometrica, 42: 999–1012.

Anderson, B., Riley, R., Young, G. (1999). The New Deal for Young People: Early Findings from the Pathfinder Areas, Employment Service Research and Development Employment Service Research No. 34.

Anderson, P. (2001). Monitoring and Assisting Active Job Search, in: OECD Proceedings: Labour Market Policies and the Public Employment Service, Paris: OECD.

Anderson, T. W. (1951). The Asymptotic Distribution of Certain Characteristic Roots and Vectors, in: Neyman, J. (Ed.), Proceedings of the 2nd Berkeley Symposium on Mathematical Statistics and Probability, Berkeley: University of California Press, 103–50.

Angrist, J. (1991). Grouped Data Estimation and Testing in Simple Labor Supply Models, in: Journal of Econometrics, 47(2–3): 243–66.

Angrist, J. D., Krueger, A. B. (1999). Empirical Strategies in Labor Economics, in: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, Vol. 3A, Amsterdam: Elsevier, 1277–366.

Apps, P., Rees, R., (2009). Relational Contracts, Taxation and the Household, IZA Discussion Paper No. 4189.

Arellano, M., Meghir, C. (1992). Female Labour Supply and on the Job Search: An Empirical Analysis using Complementary Data Sources, in: Review of Economic Studies, 59(3): 537–59.

Ashenfelter, O. (1978). Estimating the Effect of Training Programs on Earnings, in: Review of Economics and Statistics, 60(1): 47–57.

Ashenfelter, O. (1980). Unemployment as Disequilibrium in a Model of Aggregate Labor Supply, in: Econometrica, 48(3): 547–64.

Ashenfelter, O., Card, D. (1985). Using the Longitudinal Structure of Earnings to Estimate the Effect of Training Programs, in: Review of Economics and Statistics, 67(4): 648–60.

Ashenfelter, O., Ham, J. (1979). Education, Unemployment and Earnings, in: Journal of Political Economy, 87(5): 99–166.

Ashenfelter, O., Plant, M. W. (1990). Non-parametric Estimates of the Labor Supply Effects of Negative Income Tax Programs, in: Journal of Labor Economics, 8(1): 396–415.

Ashenfelter, O. Ashmore, D., Dechenes, O. (1999). Do Unemployment Insurance Recipients Actively Seek Work? Randomized Trials in Four U.S. States, NBER Working Paper No. 6982.

Ashworth, K., Hardman, J., Liu, W.-C., Maguire, S., Middleton, S., Dearden, L., Emmerson, C., Frayne, C., Goodman, A., Ichimura, H., Meghir, C. (2001). Education Maintenance Allowance: The First Year. A Quantitative Evaluation, London: Department for Education and Employment.

Atkinson, A. B., Piketty, T. (2007). Top Incomes over the Twentieth Century: A Contrast between European and English-Speaking Countries, New York: Oxford University Press.

Atkinson, A. B., Gomulka, J., Stern, N. (1984). Household Expenditure on Tobacco 1970–80, London School of Economics ESRC Programme on Taxation, Incentives and the Distribution of Income, Discussion Paper No. 57.

Atkinson, A. B., Micklewright, J., Stern, N. H. (1981). A Comparison of the Family Expenditure Survey and the New Earnings Survey 1971–1977: Part I: Characteristics of the Sample, Social Science Research Council Programme on Taxation, Incentives and the Distribution of Income Working Paper No. 27.

Atkinson, A. B., Micklewright, J., Stern, N. H. (1982). A Comparison of the Family Expenditure Survey and the New Earnings Survey 1971–1977: Part II: Hours and Earnings, Social Science Research Council Programme on Taxation, Incentives and the Distribution of Income Working Paper No. 32.

Atkinson, A. B., Stern, N. H. (1980). On the Switch from Direct to Indirect Taxation, in: Journal of Public Economics, 14(2): 195–224.

Atkinson, A. B., Stiglitz, J. E. (1972). The Structure of Indirect Taxation and Economic Efficiency, in: Journal of Public Economics, 1(1): 97–119.

Attanasio, O., Low, H., Sánchez-Marcos, V. (2008). Explaining Changes in Female Labor Supply in a Life-Cycle Model, in: American Economic Review, 98(4): 1517–52.

Attanasio, O., Low, H., Sánchez-Marcos, V. (2005). Female Labor Supply As Insurance Against Idiosyncratic Risk, in: Journal of the European Economic Association, 3(2–3): 755–64.

Banks, J., Diamond, P. (2010): The Base for Direct Taxation, in: Mirrlees, J., Adam, S., Besley, T., Blundell, R., Bond, S., Chote, R., Gammie, M., Johnson, P., Myles, G., Poterba, J. (Eds.), Dimensions of Tax Design: The Mirrlees Review, New York: Oxford University Press, 548–674.

Bargain, O., Beblo, M., Beninger, D., Blundell, R., Carrasco, R., Chiuri, M.-C., Laisney, F., Lechene, V., Moreau, N., Myck, M., Ruiz-Castillo, J., Vermeulen, F. (2006). Does the Representation of Household Behavior Matter for Welfare Analysis of Tax-benefit Policies? An Introduction, in: Review of Economics of the Household, 4(2): 99–111.

Bargain, O., Orsini, K., Peichl, A. (2014). Labor Supply Elasticities in Europe and the U.S, Journal of Human Resources, 49(3): 723–838.

Barnett, W. A. (1979). The Joint Allocation of Leisure and Goods Expenditure, in: Econometrica, 47(3): 539–63.

Bartik, T. J. (2000). Displacement and Wage Effects of Welfare Reform, in: Card, D., Blank, R. (Eds.), Finding Jobs: Work and Welfare Reform, New York: Russell Sage Foundation, 72–124.

Barton, M., Layard, R., Zabalza, A. (1980). Married Women's Participation and Hours, in: Economica, 47(185): 51–72.

Barzel, Y. (1973). The Determination of Daily Hours and Wages, in: Quarterly Journal of Economics, 87(2): 220–38.

Bassi, L. (1983). The Effect of CETA on the Post-Program Earnings of Participants, in: Journal of Human Resources, 18(4): 539–56.

Bassi, L. (1984). Estimating the Effects of Training Programs with Non-Random Selection, in: Review of Economics and Statistics, 66(1): 36–43.

Beaudry, P., Blackorby, C. (2000). Taxes and Employment Subsidies in Optimal Redistribution Programs, UBC Discussion Paper No. 22.

Beffy, M., Blundell, R., Bozio, A., Laroque, G. (2014). Labour Supply and Taxation with Restricted Choices, IFS Working Paper W14/04.

Bell, B., Blundell, R., Van Reenen, J. (1999). Getting the Unemployed Back to Work: The Role of Targeted Wage Subsidies, IFS Working Paper No. W99/12.

Bera, A. K., Jarque, C. M., Lee, L. F. (1984). Testing for the Normality Assumption in Limited Dependent Variable Models, in: International Economic Review, 25(3): 563–78.

Besley, T., Coate, S. (1992). Workfare versus Welfare: Incentive Arguments for Work Requirements in Poverty Alleviation Schemes, in: American Economic Review, 82(1): 249–61.

Biddle, J. E., Zarkin, G. A. (1989). Choice Among Wage-Hours Packages: An Empirical Investigation of Male Labor Supply, in: Journal of Labor Economics, 7(4): 415–37.

Bingley, P., Walker, I. (1997). The Labour Supply, Unemployment and Participation of Lone Mothers in In-Work Transfer Programmes, in: Economic Journal, 107(444): 1375–90.

Bishop, J., Montgomery, M. (1993). Does the Targeted Jobs Tax Credit Create Jobs at Participating Firms?, in: Industrial Relations, 32(3): 289–306.

Bishop, K., Heim, B., Mihaly, K. (2009). Single Women's Labor Supply Elasticities: Trends and Policy Implications, in: Industrial and Labor Relations Review, 63(1):146–68.

Black, D., Smith, J., Berger, M., Noel, B. (2003). Is the Threat of Reemployment Services More Effective than the Services Themselves? Evidence from Random Assignment in the UI System, in: American Economic Review, 93(4): 1313–27.

Blackorby, C., Boyce, R., Russell, R. R. (1978). Estimation of Demand Systems Generated by the Gorman Polar Form: A Generalization of the S-Branch Utility Tree, in: Econometrica, 46(2): 345–63.

Blackorby, C., Primont, D. Russell, R. R. (1978). Duality, Separability and Functional Structure: Theory and Economic Applications, New York: Elsevier.

Blanchflower, D., Freeman, R. (2000). The Declining Economic Status of Young Workers in OECD Countries, in: Blanchflower, D., Freeman, R. B. (Eds.), Youth Unemployment and Joblessness in Advanced Countries, Chicago: University of Chicago Press, 19–56.

Blank, R. (1985). Simultaneously Modelling the Supply of Weeks and Hours of Work Among Female Household Heads, Industrial Relations Section, Princeton University Working Paper No. 197.

Blank, R. (2000). Declining Caseloads/Increased Work: What Can We Conclude about the Effects of Welfare Reform, Prepared for Conference Welfare Reform Four Years Later, Federal Reserve Bank of New York.

Blank, R., Card, D., Robbins, P. (2000). Financial Incentives for Increasing Work and Income Among Low Income Families, in: Card, D., Blank, R. (Eds.), Finding Jobs: Work and Welfare Reform, New York: Russell Sage Foundation, 373–419.

Blau, F., Kahn, L. (2007). Changes in the Labor Supply Behavior of Married Women: 1980–2000, in: Journal of Labor Economics, 25: 393–438.

Bloemen, H. G. (2008). Job Search, Hours Restrictions, and Desired Hours of Work, in: Journal of Labor Economics, 26: 137–179.

Blomquist, S., Selin, H. (2010). Hourly Wage Rate and Taxable Labor Income Responsiveness to Changes in Marginal Tax Rates, in: Journal of Public Economics, 94(11–12): 878–89.

Bloom, D., Michalopoulos, C. (2001). How Welfare to Work Programs Affect Employment and Income: A Synthesis of Research, Washington, D.C.: mdrc.

Blundell, R. (1980). Estimating Continuous Consumer Equivalence Scales in an Expenditure Model with Labour Supply, in: European Economic Review, 14(2): 145–57.

Blundell, R. (2001). Welfare Reform for Low Income Workers, in: Oxford Economic Papers, 53(2): 189–214.

Blundell, R. (2002). Welfare-to-work: Which Policies Work and Why?, in: Proceedings of the British Academy, 117: 477–524.

Blundell, R., Bozio, A., Laroque, G. (2011a). Labour Supply and the Extensive Margin, in: American Economic Review, 101(3): 482–486.

Blundell, R., Bozio, A., Laroque, G. (2011b). Extensive and Intensive Margins of Labour Supply: Working Hours in the US, UK and France, IZA Discussion Paper No. 6051.

Blundell, R., Brewer, M. (2000). Tax and Benefit Reform in the UK: Making Work Pay', in: Peer Review of UK Tax and Benefit Reform: Making Work Pay, European Commission - DG EML, Peer Review Programme 2000.

Blundell, R., Brewer, M., Francesconi, M. (2008). Job Changes, Hours Changes and the Path of Labour Supply Adjustment, in: Journal of Labor Economics, 26(3): 421–53.

Blundell, R., Brewer, M., Shephard, A. (2005). Evaluating the Labour Market Impact of Working Families' Tax Credit Using Difference-In-Differences, HM Customs and Revenue, http://www.hmrc.gov.uk/research/ifs-did.pdf.

Blundell, R., Brewer, M., Shephard, A. (2006). Earned Income Tax Credit Policies: Impact and Optimality: The Adam Smith Lecture, 2005, in: Labour Economics, 13(4): 423–43.

Blundell, R., Browning, M., Meghir, C. (1994). Consumer Demand and the Life-Cycle Allocation of Household Expenditures, in: Review of Economic Studies, 61(1): 57–80.

Blundell, R., Chiappori, P.-A., Magnac, T., Meghir, C. (2007). Collective Labor Supply: Heterogeneity and Non-participation, in: Review of Economic Studies 74(2): 417–45.

Blundell, R., Chiappori, P.-A., Meghir, C. (2005). Collective Labour Supply with Children, in: Journal of Political Economy, 113(6): 1277–1306.

Blundell, R., Costa Dias, M. (2000). Evaluation Methods for Non-Experimental Data, in: Fiscal Studies, 21(4): 427–68.

Blundell, R., Costa Dias, M. (2002). Alternative Approaches to Evaluation in Empirical Microeconomics, CeMMAP Working Paper CWP10/02, http://cemmap.ifs.org.uk/working_papers.shtml.

Blundell, R., Costa Dias, M. (2009). Alternative Approaches to Evaluation in Empirical Microeconomics, in: Journal of Human Resources, 44(3): 565–640.

Blundell, R., Costa Dias, M., Meghir, C. (2003). The Overall Impact of Wage Subsidies Under Idiosyncratic Uncertainty, Institute for Fiscal Studies, http://editorialexpress.com/cgi-bin/conference/download.cgi?db_name=NASM2003&paper_id=567.

Blundell, R., Costa Dias, M., Meghir, C., Shaw, J. (2013). Female Labour Supply, Human Capital and Welfare Reform, IZA Discussion Paper No. 7375.

Blundell, R., Costa Dias, M., Meghir, C., Van Reenen, J. (2004). Evaluating the Employment Impact of a Mandatory Job Search Assistance Program, in: Journal of the European Economic Association 2(4): 569–606.

Blundell, R., Dearden, L., Goodman, A., Reed, H. (1997). Higher Education, Employment and Earnings in Britain, London: Institute for Fiscal Studies.

Blundell, R., Dearden, L., Meghir, C. (1996). The Determinants and Effects of Work Related Training in Britain, IFS Commentary, London.

Blundell, R., Dearden, L., Sianesi, B. (2003). Evaluating the Impact of Education on Earnings in the UK: Models, Methods and Results from the NCDS, IFS Working Paper No. W03/20.

Blundell, R., Duncan, A., McCrae, J., Meghir, C. (2000a). Evaluating In-Work Benefit Reform: The Working Families Tax Credit in the UK, JCPR Working Paper No. 160.

Blundell, R., Duncan, A., McCrae, J., Meghir, C. (2000b). The Labour Market Impact of the Working Families Tax Credit, in: Fiscal Studies, 21(1): 75–103.

Blundell, R., Duncan, A., Meghir, C. (1992). Taxation in Empirical Labour Supply Models: Lone Mothers in the U.K., in: Economic Journal, 102(411): 265–78.

Blundell, R., Duncan, A., Meghir, C. (1998). Estimating Labour Supply Responses Using Tax Policy Reforms, in: Econometrica, 66(4): 827–62.

Blundell, R., Francesconi, M., van der Klaauw, W. (2011). Anatomy of Welfare Reform Evaluation: Announcement and Implementation Effects, IZA Discussion Paper No. 6050.

Blundell, R., Ham, J., Meghir, C. (1987). Unemployment and Female Labour Supply, in: Economic Journal, 97(388a): 44–64.

Blundell, R., Hoynes, H. W. (2004). Has "In-Work" Benefit Reform Helped the Labor Market?, in: Card, D., Blundell, R., Freeman, R. B. (Eds.), Seeking a Premier Economy: The Economic Effects of British Economic Reforms, 1980–2000, Chicago: University of Chicago Press, 411–59.

Blundell, R., Johnson, P. (1998). Pensions and Labor-Market Participation in the United Kingdom, in: American Economic Review, 88(2): 168–72.

Blundell, R., MaCurdy, T. (1999). Labor Supply: A Review of Alternative Approaches, in: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, Vol. 3A, Amsterdam: Elsevier, 1559–695.

Blundell, R., Meghir, C. (1986). Selection Criteria for a Microeconometric Model of Labour Supply, in: Journal of Applied Econometrics, 1(1): 55–80.

Blundell, R., Meghir, C. (2002). Active Labour Market Policy vs Employment Tax Credits: Lessons from Recent UK Reforms, IFAU Working Paper No. 2002:1.

Blundell, R., Meghir, C., Neves, P. (1993). Labour Supply and Intertemporal Substitution, in: Journal of Econometrics, 59(1–2): 137–60.

Blundell, R., Meghir, C., Smith, S. (2002). Pension Incentives and the Pattern Of Early Retirement, in: Economic Journal 112(478): C153–70.

Blundell, R., Meghir, C., Symons, E., Walker, I. (1986). A Labour Supply Model for the Simulation of Tax and Benefit Reforms, in: Blundell, R., Walker, I. (Eds.), Unemployment, Search and Labour Supply, Cambridge: Cambridge University Press, 267–93.

Blundell, R., Meghir, C., Symons, E., Walker, I. (1988). Labour Supply Specification and the Evaluation of Tax Reforms, in: Journal of Public Economics, 36(1): 23–52.

Blundell, R., Pistaferri, L., Saporta-Eksten, I. (2012). Consumption Inequality, Consumption Smoothing and Family Labor Supply, IZA Discussion Paper No. 6900.

Blundell, R., Preston, I. (1998). Consumption Inequality and Income Uncertainty, in: Quarterly Journal of Economics, 113(2): 603–40.

Blundell, R., Reed, H. (1999). The Employment Effects of the Working Families' Tax Credit, IFS Briefing Note 6/99.

Blundell, R., Shephard, A. (2010). Employment, Hours of Work and the Optimal Taxation of Low Income Families, IFS Working Paper No. W08/01.

Blundell, R., Shephard, A. (2012). Supplement to 'Employment, Hours of Work and the Optimal Taxation of Low Income Families', in: Review of Economic Studies, 79(2): 481–510.

Blundell, R., Smith, R. (1986). An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply, in: Econometrica, 54(3): 679–85.

Blundell, R., Smith, R. (1994). Coherency and Estimation in Simultaneous Models With Censored or Qualitative Dependent Variables, in: Journal of Econometrics, 64(1–2): 355–73.

Blundell, R., Walker, I. (1982). Modelling the Joint Determination of Household Labour Supplies and Commodity Demands, in: Economic Journal, 92(366): 351–64.

Blundell, R., Walker, I. (1983). Limited Dependent Variables in Demand Analysis: An Application to Modelling Family Labour Supply and Commodity Demand Behaviour, Discussion Paper in Econometrics No. ES 126.

Blundell, R., Walker, I. (1984). A Household Production Specification of Demographic Variables in Demand Analysis, in: Economic Journal, 94(376a): 59–68.

Blundell, R., Walker, I. (1986). A Life-Cycle Consistent Empirical Model of Family Labour Supply Using Cross-Section Data, in: Review of Economic Studies, 53(4): 539–58.

Bonjour, D., Dorsett, R., Knight, G., Lissenburgh, S., Mukherjee, A., Payne, J., Range, M., Urwin, P., White, M. (2001). New Deal for Young People: National Survey of Participants: Stage 2, Employment Service Report 67.

Borjas, G. J. (1980). The Relationship Between Wages and Weekly Hours of Work: The Role of Division Bias, in: Journal of Human Resources, 15(3): 409–421.

Bos, J., Houston, A., Granger, R., Duncan, G., Brook, T., McLoyd, V. (1999). New Hope for People with Low Income, Washington, DC: mdrc.

Bound, J., Jaeger, D. A., Baker, R. M. (1995). Problems with Instrumental Variables Estimation when the Correlation between the Instrument and the Endogenous Explanatory Variable is Weak, in: Journal of the American Statistical Association, 90(430): 443–50.

Brewer, M. (2001). Comparing In-work Benefits and the Reward to Work for Low-Income Families with Children in the US and UK, in: Fiscal Studies, 22(1): 41–77.

Brewer, M. (2003). The New Tax Credits, London: IFS.

Brewer, M., Browne, J. (2009). Can More Revenue be Raised by Increasing Income Tax Rates for the Very Rich?, London: IFS.

Brewer, M., Clark, T., Myck, M. (2001). Credit Where It's Due? An Assessment of the New Tax Credits, London: IFS.

Brewer, M., Duncan, A., Shephard, A., Suárez, M.-J. (2006). Did Working Families' Tax Credit Work? The Impact of In-Work Support on Labour Supply in Great Britain, in: Labour Economics, 13(6): 699–720.

Brewer, M., Gregg, P. (2001). Eradicating Child Poverty in Britain: Welfare Reform and Child Poverty since 1997, IFS Working Paper No. W01/08.

Brewer, M., Myck, M., Reed, H. (2001). Financial Support for Families with Children: Options for the New Integrated Child Credit, London: IFS.

Brewer, M., Saez, E., Shephard, A. (2010). Means-testing and Tax Rates on Earnings, in: Mirrlees, J., Adam, S., Besley, T., Blundell, R., Bond, S., Chote, R., Grammie, M., Johnson, P., Myles, G., Poterba, J. (Eds.), Dimensions of Tax Design: The Mirrlees Review, New York: Oxford University Press, 90–201.

Browning, M. J., Deaton, A. S., Irish, M. (1985). A Profitable Approach to Labour Supply and Commodity Demands over the Life-Cycle, in: Econometrica, 53(3): 503–43.

Bryan, M. L. (2007). Free to Choose? Differences in the Hours Determination of Constrained and Unconstrained Workers, in: Oxford Economic Papers, 59(2): 226–52.

Bryson, A., Knight, G., White, M. (2000). New Deal for Young People: National Survey of Participants: Stage 1, Employment Service Report 44.

Bryson, A., Marsh, A. (1996). Leaving Family Credit, London: Department of Social Security.

Burtless, G. (1985). Are Targeted Wage Subsidies Harmful? Evidence from a Wage Voucher Experiment, in: Industrial and Labor Relations Review, 39(1): 105–14.

Burtless, G. (1995). The Case for Randomized Field Trials in Economic and Policy Research, in: Journal of Economic Perspectives 9(2): 63–84.

Card, D. (1990). Labor Supply with a Minimum Hours Threshold, in: Carnegie-Rochester Conference Series on Public Policy, 33(1): 137–68.

Card, D., Blank, R. (Eds.) (2000a). Finding Jobs: Work and Welfare Reform, New York: Russell Sage Foundation.

Card, D., Blank, R. (Eds.) (2000b). Labor Markets and Less-Skilled Workers, New York: Russell Sage Foundation.

Card, D., DellaVigna, S., Malmendier, U. (2011). The Role of Theory in Field Experiments, in: Journal of Economic Perspectives, 25(3), 39–62.

Card, D., Hyslop, D. (2005). Estimating the Effects of a Time-Limited Earnings Subsidy for Welfare-Leavers, in: Econometrica, 73(6): 1723–70.

Card, D., Michalopoulos, C., Robins, P. K. (2001). The Limits to Wage Growth: Measuring the Growth Rate of Wages for Recent Welfare Leavers, NBER Working Paper No. 8444.

Card, D., Robins, P. K. (1998). Do Financial Incentives Encourage Welfare Recipients To Work?, in: Research in Labor Economics, 17: 1–56.

Chesher, A., Irish, M. (1984). Residuals and Diagnostics for Probit, Tobit and Related Models, University of Bristol Economics Discussion Paper.

Chetty, R. (2008). Sufficient Statistics for Welfare Analysis: A Bridge between Structural and Reduced-form Methods, NBER Working Paper No. 14399.

Chetty, R. (2009). Bounds on Elasticities With Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply, NBER Working Paper No. 15616.

Chetty, R. (2012). Bounds on Elasticities With Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply, in: Econometrica 80(3): 968–1018.

Chetty, R., Friedman, J., Olsen, T., Pistaferri, L. (2011a). Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records, in: Quarterly Journal of Economics 126(2): 749–804.

Chetty, R., Friedman, J. N., Olsen, T., Pistaferri, L., Frederiksen, A. (2009). The Effect of Adjustment Costs and Institutional Constraints on Labor Supply Elasticities: Evidence from Denmark, Harvard University Working Paper.

Chetty, R., Friedman, J., Saez, E. (2013). Using Differences in Knowledge Across Neighborhoods to Uncover the Impacts of the EITC on Earnings, in: American Economic Review, 103(7): 2683–721.

Chetty, R., Guren, A., Manoli, D., Weber, A. (2013). Does Indivisible Labor Explain the Difference between Micro and Macro Elasticities? A Meta-Analysis of Extensive Margin Elasticities, in: NBER Macroeconomics Annual, 27(1): 1–56.

Chetty, R., Manoli, D., Guren, A., Weber, A. (2011b). Are Micro and Macro Labor Supply Elasticities Consistent? A Review of Evidence on the Intensive and Extensive Margins, in: American Economic Review Papers and Proceedings, 101(3): 471–5.

Chiappori, P. A., Donni, O. (2009). Non-unitary Models of Household Behavior: A Survey of the Literature, IZA Discussion Paper No. 4603.

Choné, P., Laroque, G. (2010). Negative Marginal Tax Rates and Heterogeneity, in: American Economic Review, 100(5): 2532–47.

Choné, P., Laroque, G. (2005). Optimal Incentives for Labor Force Participation, in: Journal of Public Economics, 89(2–3): 395–425.

Cogan, J. F. (1981). Fixed Costs and Labor Supply, in: Econometrica, 49(4): 945–63.

Cossa, R, Heckman, J., Lochner, L. (1999). Wage Subsidies and Skill Formation: A Study of the Earned Income Tax Credit, Chicago: Russel Sage Foundation.

Cossa, R., Heckman, J., Lochner, L. (2002). Learning-By-Doing vs. On-the-Job Training: Using Variation Induced by the EITC to Distinguish Between Models of Skill Formation, NBER Working Paper No. 9083.

Cragg, J. G. (1971). Some Statistical Models for Limited Dependent Variables with Application to the Demand for Durable Goods, in: Econometrica, 39(5): 829–44.

Crawford, I., Keen, M., Smith, S. (2010). Indirect Taxation and VAT, in: Mirrlees, J., Adam, S., Besley, T., Blundell, R., Bond, S., Chote, R., Gammie, M., Johnson, P., Myles, G., Poterba, J. (Eds.), Dimensions of Tax Design: The Mirrlees Review, New York: Oxford University Press, 275–422.

Dagsvik, J. K., Strom, S. (2006). Sectoral Labor Supply, Choice Restrictions and Functional Form, in: Journal of Applied Econometrics, 21(6): 803–26.

Danziger, S. (2000). Approaching the Limit: Early Lessons from Welfare Reform, Center on Poverty, Risk and Mental Health, University of Michigan, mimeo.

Davidson, C., Woodbury, S. A. (1993). The Displacement Effect of Reemployment Bonus Programs, in: Journal of Labor Economics, 11(4): 575–605.

Davis, S., Henrekson, M. (2004). Tax Effects on Work Activity, Industry Mix and Shadow Economy Size: Evidence from Rich-Country Comparisons, NBER Working Paper No. 10509.

Deaton, A. (1974). Reconsideration of the Empirical Implications of Additive Preferences, in: Economic Journal, 84(334): 338–48.

Deaton, A., Irish, M. (1984). Statistical Models for Zero Expenditures in Household Budgets, in: Journal of Public Economics, 23(1–2): 59–80.

Deaton, A., Muellbauer, J. (1980). Economics and Consumer Behaviour, Cambridge, UK: Cambridge University Press.

Deaton, A., Muellbauer, J. (1981). Functional Forms for Labor Supply and Commodity Demands with and without Quantity Restrictions, in: Econometrica, 49(6): 1521–32.

Deheijia, R., Wahba, S. (1998). Propensity Score Matching Methods for Non-Experimental Causal Studies, NBER Working Paper No. 6829.

Dehejia, R., Wahba, S. (1999). Causal Effects in Non-Experimental Studies: Reevaluating the Evaluation of Training Programs, in: Journal of the American Statistical Association, 94(448): 1053–62.

DfEE (1997). Design of the New Deal for 18–24 Year Olds, London, UK: Department for Education and Employment.

Diamond, P. (1980). Income Taxation with Fixed Hours of Work, in: Journal of Public Economics, 13(1): 101–10.

Dickens, R. (2000): The Evolution of Individual Male Earnings in Britain: 1975–1995, in: Economic Journal, 110(460): 27–49.

Dickens, R., Ellwood, D. (2001). Whither Poverty in Great Britain and the United States? The Determinants of Changing Poverty and Whether Work Will Work, in: Card, D., Blundell, R., Freeman, R. B. (Eds.), Seeking a Premier Economy: The Economic Effects of British Economic Reforms, 1980–2000, Chicago: University of Chicago Press, 313–370.

Dickens, R., Machin, S., Manning, A. (1999). The Effects of Minimum Wages on Employment: Theory and Evidence from Britain, in: Journal of Labor Economics, 17(1): 1–22.

Dickens, R., Manning, A. (2004). Has the National Minimum Wage Reduced UK Wage Inequality?, in: Journal of the Royal Statistical Society, Series A 167(4): 613–26.

Dickens, W. T., Lundberg, S. J. (1993). Hours Restrictions and Labor Supply, in: International Economic Review, 34(1): 169–92.

Dickert, S., Giles, C. (1996). Should We Subsidise Childcare, and If So How?, in: Fiscal Studies, 17(3): 39–62.

Dickert, S., Houser, S. (1998). Taxes and Transfers: A New Look at the Marriage Penalty, in: National Tax Journal, 51(2): 175–217.

Dickert, S., Houser, S., Scholtz, J. K. (1995). The Earned Income Tax Credit and Transfer program: A Study of Labor Market and Program Participation, in: Poterba, J. (Ed.), Tax Policy and the Economy, Vol. 9, Cambridge, MA: MIT Press, 1–50.

Dickert-Conlin, S., Holtz-Eakin, D. (2000). Employee-based versus Employer Based Subsidies to Low Wage Workers: A Public Finance Perspective, in: Card, D., Blank, R. (Eds.), Finding Jobs: Work and Welfare Reform, New York: Russell Sage Foundation, 262–96.

Dilnot, A., Duncan, A. (1992). Lone Mothers, Family Credit and Paid Work, in: Fiscal Studies, 13(1): 1–21.

Dilnot, A., Giles, C. (1998). The IFS Green Budget: January 1998, London: Institute for Fiscal Studies.

Dolton, P., Makepeace, G., Treble, J. (1992). Public- and Private-Sector Training of Young People in Britain, in: Lynch, L. M. (Ed.), Training and the Public Sector, Chicago: University of Chicago Press, 261–82.

Dolton, P., O'Niell, D. (1996). Unemployment Duration and the Restart Effect: Some Experimental Evidence, in: Economic Journal, 106(435): 387–400.

Dorsett, R. (2001). The New Deal for Young People: Relative Effectiveness of the Options in Reducing Male Unemployment, London: PSI.

Dubin, J., Rivers, D. (1993). Experimental Estimates of the Impact of Wage Subsidies, in: Journal of Econometrics, 56(1–2): 219–42.

Duncan, A., Giles, C. (1998). The Labour Market Impact of the Working Families' Tax Credit, paper presented at the International Institute for Public Finance conference, Cordoba.

Duncan, A., Giles, C., Webb, S. (1995). The Impact of Subsidising Childcare, Equal Opportunities Commission, Research Discussion Series No. 13.

Duncan, A., Weeks, M. (1998). Simulating Transitions Using Discrete Choice Models, in: Papers and Proceedings of the American Statistical Association, 106: 151–6.

Dustmann, C., Meghir, C. (2001). Wages, Experience and Seniority, IFS Working Paper No. W01/01.

Eckstein, Z., Wolpin, K. (1989). Dynamic Labour Force Participation of Married Women and Endogenous Work Experience, in: Review of Economic Studies, 56(3): 375–390.

Eissa, N. (1995). Taxation and Labor Supply of Married Women: The Tax Reform Act of 1986 as a Natural Experiment, NBER Working Paper No. 5023.

Eissa, N. (1996). Labor Supply and the Economic Recovery Tax Act of 1981, in: Feldstein, M., Poterba, J. (Eds.), Empirical Foundations of Household Taxation, Chicago: University of Chicago Press, 5–38.

Eissa, N., Hoynes, H. (1998). The Earned Income Tax Credit and the Labor Supply of Married Couples, NBER Working Paper No. 6856.

Eissa, N., Hoynes, H. (2004). Taxes and the Labor Market Participation of Married Couples: The Earned Income Tax Credit, in: Journal of Public Economics, 88(9–10): 1931–58.

Eissa, N., Kleven, H. J., Kreiner, C. T. (2004). Evaluation of Four Tax Reforms in the United States: Labor Supply and Welfare Effects for Single Mothers, NBER Working Paper No. 10935.

Eissa, N., Liebman, J. (1996). Labor Supply Response to the Earned Income Tax Credit, in: Quarterly Journal of Economics, 111(2): 605–37.

Ellwood, D., Welty, E. (2000). Public Service Employment and Mandatory Work: A Policy Whose Time Has Come and Gone and Come Again?, in: Card, D., Blank, R. (Eds.), Finding Jobs: Work and Welfare Reform, New York: Russell Sage Foundation, 299–372.

Ellwood, D. T. (2000). The Impact of Earned Income Tax Credit and Social Policy Reforms on Work, Marriage and Living Arrangements, in: National Tax Journal, 53(4): 1063–105.

Ermisch, J., Francesconi, M. (2000). The Effect of Parents' Employment on Children's Educational Attainment, ISER Working Paper No. 2000–31.

Euwals, R. (2001). Female Labour Supply, Flexibility of Working Hours, and Job Mobility, in: Economic Journal, 111(471): C120-C134.

Farhi, E., Werning, I. (2013). Insurance and Taxation over the Life Cycle, in: Review of Economic Studies, 80(2): 596–635.

Feldstein, M. S. (1995). The Effect of Marginal Tax Rates on Taxable Income: A Panel Study of the 1986 Tax Reform Act, in: Journal of Political Economy, 103(3): 551–72.

Feldstein, M. S. (1999). Tax Avoidance and the Deadweight Loss of the Income Tax, in: Review of Economics and Statistics, 81(4): 674–80.

Fraker, T., Maynard, R. (1987). The Adequacy of Comparison Group Designs for Evaluations of Employment Related Programs, in: Journal of Human Resources, 22(2): 194–227.

Francesconi, M., van der Klaauw, W. (2007). The Socioeconomic Consequences of "In-Work" Benefit Reform for British Lone Mothers, in: Journal of Human Resources, 42(1): 1–31.

Ghez, G., Becker, G. S. (1975). The Allocation of Time and Goods Over the Life Cycle, New York: Columbia University Press.

Giles, C., Johnson, P., McCrae, J. (1997). Housing Benefit and the Financial Returns to Employment for Tenants in the Social Sector, in: Fiscal Studies, 18(1): 49–72.

Gill, P. E., Murray, W. (1972). Quasi-Newton Methods for Unconstrained Optimization, in: Journal of the Institute of Mathematics and its Applications, 9: 91–108.

Gladden, T., Taber, C. (2000). Wage Progression Among Less Skilled Workers, in: Card, D., Blank, R. (Eds.), Finding Jobs: Work and Welfare Reform, New York: Russell Sage Foundation, 160–192.

Goldin, C., Katz, L. F. (1998). The Origins of Technology Skills Complementarity, in: Quarterly Journal of Economics, 113(3): 693–732.

Goldman, S. M., Uzawa, H. (1964). A Note on Separability and Demand Analysis, in: Econometrica, 32(3): 387–98.

Golosov, M., Troshkin, M., Tsyvinski, A. (2011). Optimal Taxation: Merging Micro and Macro Approaches, Journal of Money, Credit and Banking, Supplement to 43(5), 147–174.

Gordon, R., Kopczuk, W. (2013). The Choice of the Personal Income Tax Base, New York.

Gorman, W. M. (1968). Conditions for Additive Separability, in: Econometrica, 36(3–4): 605–9.

Gorman, W. M. (1976). Tricks with Utility Functions, in: Artis, M. J., Nobay, A. R. (Eds.), Essays in Economic Analysis: The Proceedings of the Association of University Teachers of Economics, Cambridge: Cambridge University Press, 211–44.

Gosling, A., Machin, S., Meghir, C. (2000). The Changing Distribution of Male Wages in the U.K, in: Review of Economic Studies, 67(4): 635–66.

Gourieroux, C., Laffont, J., Montfort, A. (1980). Coherency Conditions in Simultaneous Linear Equation Models with Endogenous Switching Regimes, in: Econometrica, 48(3): 675–95.

Gourieroux, C., Monfort, A., Renault, E., Trognon, A. (1984). Résidus généralisés ou interprétations linéaires de l'économétrie non linéaire, INSEE Document de Travail No. 8410.

Gradus, R. H., Julsing, J. M. (2001). Comparing Different European Income Tax Policies: Making Work Pay, OCFEB Research Memorandum 0101.

Greenhalgh, C. (1977). A Labour Supply Function for Married Women in Great Britain, in: Economica, 44(175): 249–65.

Greenwood, J., Voyer, J.-P. (2000). Experimental Evidence on the Use of Earnings Supplements as a Strategy to "Make Work Pay", OECD Economic Studies No.31.

Gregg, P., Hansen, K., Wadsworth, J. (1999). The Rise of the Workless Household, in: Gregg, P., Wadsworth, J. (Eds.), The State of Working Britain, Manchester: Manchester University Press, 75–89.

Gregg, P., Harkness, S. (2007). Welfare Reform and Lone Parents in the UK, CMPO Working Paper No. 07/182.

Gregg, P., Johnson, P., Reed, H. (1999). Entering Work and the British Tax and Benefit System, London: Institute for Fiscal Studies.

Gregg, P., Wadsworth, J. (1999). The State of Working Britain, Manchester: Manchester University Press.

Grogger, J. (2000). Time Limits and Welfare Use, NBER Working Paper No. 7709.

Grogger, J., Michalopoulos, C. (1999). Welfare Dynamics Under Time Limits, NBER Working Paper No. 7353.

Gronau, R. (1974). Wage Comparisons – A Selectivity Bias, in: Journal of Political Economy, 82(6): 1119–43.

Gruber, J., Saez, E. (2002). The Elasticity of Taxable Income: Evidence and Implications, in: Journal of Public Economics, 84(1): 1–32.

Gruber, J., Wise, D. (1999). Social Security and Retirement around the World, Chicago: The University of Chicago Press.

Hahn, J., Todd, P., Van der Klaauw, W. (2001). Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design, in: Econometrica, 69(1): 201–9.

Hales, J., Collins, D., Hasluck, C., Woodland, S. (2000). New Deals for Young People and for Long-Term Unemployed: Survey of Employers, Employment Service Research and Development Report No. 58.

Hall, R. E. (1978). Stochastic Implications of the Life Cycle-Permanent Income Hypothesis: Theory and Evidence, in: Journal of Political Economy, 86(6): 971–88.

Ham, J. (1982). Estimation of a Labour Supply Model with Censoring Due to Unemployment and Underemployment, in: Review of Economic Studies, 49(3): 335–54.

Ham, J. (1986). Testing Whether Unemployment Represents Intertemporal Labour Supply Behaviour, in: Review of Economic Studies, 53(4): 559–78.

Ham, J., Reilly, K. T. (2002). Testing Intertemporal Substitution, Implicit Contracts, and Hours Restriction Models of the Labor Market Using Micro Data, in: American Economic Review, 92(4): 905–27.

Hanoch, G. (1980). A Multivariate Model of Labour Supply: Methodology and Estimation, in: Smith, J. P. (Ed.), Female Labour Supply: Theory and Estimation, Princeton: Princeton University Press, 249–326.

Hansen, L. P., Singleton, K. J. (1982). Generalized Instrumental Variables Estimation of Non-linear Rational Expectations Models, in: Econometrica, 50(5): 1269–86.

Hausman, J., Wise, D. (1985). Social Experimentation, Chicago: University of Chicago Press.

Hausman, J. A. (1980). The Effect of Wages, Taxes and Fixed Costs on Women's Labor Force Participation, in: Journal of Public Economics, 14(2): 161–94.

Hausman, J. A. (1981). Labor Supply, in: Aaron, H., Pechman, J. (Eds.), How Taxes Affect Economic Behaviour, Washington, DC: Brookings Institution, 27–72.

Hausman, J. A., Ruud, P. (1984). Family Labor Supply with Taxes, in: American Economic Review, 74(2): 242–8.

Hausman, J. A., Wise, D. A. (1977). Social Experimentation, Truncated Distributions, and Efficient Estimation, in: Econometrica, 45(4): 913–38.

Heckman, J. J. (1974). Shadow Prices, Market Wages, and Labor Supply, in: Econometrica, 42(4): 679–94.

Heckman, J. J. (1974a). Life Cycle Consumption and Labour Supply: An Explanation of the Relationship Between Income and Consumption Over the Life Cycle, in: American Economic Review, 64(1): 188–94.

Heckman, J. J. (1974b). The Effect of Child Care Programs on Women's Work Effort, in: Journal of Political Economy 82(2): S136-S163.

Heckman, J. J. (1978). Dummy Endogenous Variables in a Simultaneous Equation System, in: Econometrica 46(4): 931–59.

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error, in: Econometrica, 47(1): 153–61.

Heckman, J. J. (1980). Sample Selection Bias as a Specification Error, in: Smith, J. (Ed.), Female Labor Supply: Theory and Estimation, Princeton: Princeton University Press, 206–48.

Heckman, J. J. (1993). What Has Been Learned About Labor Supply in the Past Twenty Years?, in: American Economic Review 83(2): 116–21.

Heckman, J. J. (1996). Comment, in: Feldstein, M., Poterba, J. (Eds.), Empirical Foundations of Household Taxation, Chicago: University of Chicago Press, 32–38.

Heckman, J. J., Hotz, V. (1989). Choosing Among Alternative Non-experimental Methods for Estimating the Impact of Social Programs: The Case of Manpower Training, in: Journal of the American Statistical Association, 84(408): 862–74.

Heckman, J. J., Ichimura, H., Smith, J., Todd, P. (1998). Characterizing Selection Bias Using Experimental Data, in: Econometrica, 66(5): 1017–98.

Heckman, J. J., Ichimura, H., Todd, P. (1997). Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program, in: Review of Economic Studies, 64(4): 605–54.

Heckman, J. J., Lalonde, R., Smith, J. (2000). The Economics and Econometrics of Active Labor Market Policies, in: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, Vol. 3, Amsterdam: Elsevier, 1865–2097.

Heckman, J. J., Lochner, L., Cossa, R. (2003). Learning-By-Doing vs. On-the- Job Training: Using Variation Induced by the EITC to Distinguish Between Models of Skill Formation, in: Phelps, E. (Ed.), Designing Inclusion: Tools to Raise Low-End Pay and Employment in Private Enterprise, Cambridge: Cambridge University Press, 74–130.

Heckman, J. J., Lochner, L., Taber, C. (1998a). Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Earnings, in: Review of Economic Dynamics, 1(1): 1–58.

Heckman, J. J., Lochner, L., Taber, C. (1998b). General-Equilibrium Treatment Effects: A Study of Tuition Policy, in: The American Economic Review 88(2): 381–6.

Heckman, J. J., Macurdy, T. E. (1980). A Life Cycle Model of Female Labour Supply, in: Review of Economic Studies, 47(1): 47–74.

Heckman, J. J., Robb, R. (1985). Alternative Methods for Evaluating the Impact of Interventions, in: Heckman, J., Singer, B. (Eds.), Longitudinal Analysis of Labor Market Data, New York: Cambridge University Press, 156–246.

Heckman, J. J., Robb, R. (1986). Alternative Methods for Solving the Problem of Selection Bias in Evaluating the Impact of Treatments on Outcomes, in: Wainer, H. (Ed.), Drawing Inferences from Self-Selected Samples, Berlin: Springer Verlag, 63–107.

Heckman, J. J., Smith, J. A. (1995). Assessing the Case for Social Experiments, in: Journal of Economic Perspectives, 9(2): 85–110.

Heckman, J. J., Smith, J., Clements, N. (1997). Making the Most out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts, in: Review of Economic Studies, 64(4): 487–535.

Heim, B. T., Meyer, B. (2004). Work Costs and Non-convex Preferences in the Estimation of Labor Supply Models, in: Journal of Public Economics, 88(11): 2323–38.

Holcomb, P. A., Pavetti, L., Ratcliffe, C., Riedinger, S. (1998). Building an Employment Focused Welfare System. Work First and Other Work-Oriented Strategies in Five States, Washington, DC: Urban Institute, http://www.urban.org/publications/407477.html

Hollenbeck, K., Willke, R. (1991). The Employment and Earnings Impact of the Targeted Job Tax Credit, Upjohn Working Paper No. 91–07.

Hoynes, H. (1996). Welfare Transfers in Two-Parent Families: Labor Supply and Welfare Participation under AFDC-UP, in: Econometrica, 64(2): 295–332.

Hoynes, H. (1997a). Does Welfare Play Any Role in Female Headship Decisions?, in: Journal of Public Economics, 65(2): 89–117.

Hoynes, H. (1997b). Work, Welfare, and Family Structure: What Have We Learned?, in: Auerbach, A. (Ed.), Fiscal Policy: Lessons From Economic Research, Cambridge, MA: MIT Press, 101–146.

Hoynes, H. (2000). The Employment, Earnings and Income of Less Skilled Workers over the Business Cycle, in: Card, D., Blank, R. (Eds.), Finding Jobs: Work and Welfare Reform, New York: Russell Sage Foundation, 23–71.

Hurd, M. (1976). The Estimation of Non-linear Labor Supply Functions with Taxes from a Truncated Sample, Stanford Research Institute Center for the Study of Welfare Policy Research Memorandum No. 36.

Imai, S., Keane, M. (2004). Intertemporal Labor Supply and Human Capital Accumulation, in: International Economic Review, 45(2): 601–41.

Immervol, H., Kleven, H., Kreiner, C., Saez, E. (2007). Welfare Reform in European Countries: A Microsimulation Analysis, in: Economic Journal 117(516): 1–44.

Inland Revenue (2001). New Tax Credits: Supporting Families, Making Work Pay and Tackling Poverty, London: Inland Revenue.

Inland Revenue (2003). Working Families' Tax Credit Statistics Quarterly Enquiry, London: Inland Revenue.

Inland Revenue (2005). Working Families' Tax Credit Estimates of Take-Up Rates in 2002–03, London: Inland Revenue.

Inland Revenue (various years). Working Families' Tax Credit Quarterly Enquiry, London: Government Statistical Service.

Johnson, T. R., Pencavel, J. H. (1980). Utility-Based Hours Of Work Functions for Husbands, Wives, and Single Females Estimated from Seattle-Denver Experimental Data, Menlo Park: SRI International.

Kahn, S., Lang, K. (1991). The Effect of Hours Constraints on Labor Supply Estimates, in: Review of Economics and Statistics, 73(4): 605–11.

Katz, L. (1998). Wage Subsidies for the Disadvantaged, in: Freeman, R., Gottschalk, P. (Eds.), Generating Jobs. How to Increase Demand for Less-Skilled Workers, New York, NY: Russell Sage Foundation, 21–53.

Katz, L., Meyer, B. (1990). Unemployment Insurance, Recall Expectations and Unemployment Outcomes, in: Quarterly Journal of Economics 105(4): 973–1002.

Katz, L. F., Autor, D. H. (1999). Changes in the Wage Structure and Earnings Inequality, in: Ashenfelter, O., Card, D. (Eds.), Handbook of Labor Economics, Vol. 3A, Amsterdam: Elsevier, 3–26.

Keane, M. P. (1992). A Note on Identification in the Multinomial Probit Model, in: Journal of Business and Economic Statistics, 10(2): 193–200.

Keane, M. P. (1995). A New Idea for Welfare Reform, in: Quarterly Review, 19: 2–28.

Keane, M. P., Moffitt, R. (1998). A Structural Model of Multiple Welfare Program Participation and Labor Supply, in: International Economic Review, 39(3): 553–89.

Keane, M. P., Rogerson, R. (2012). Micro and Macro Labor Supply Elasticities: A Reassessment of the Conventional Wisdom, in: Journal of Economic Literature 50(2): 464–76.

Keane, M. P., Wolpin, K. I. (1997). The Career Decisions of Young Men, in: Journal of Political Economy, 105(3): 473–522.

Keane, M. P., Wolpin, K. I. (2010). The Role Of Labor And Marriage Markets, Preference Heterogeneity, and the Welfare System in the Life Cycle Decisions of Black, Hispanic, and White Women, in: International Economic Review, 51(3): 851–92.

Kemsley, W. F. F. (1977). Family Expenditure Survey. A Study of Differential Response Based on a Comparison of the 1971 Sample with the Census, in: Statistical News, 31: 16–22.

Kirby, S., Riley, R. (2004). Compulsory Work-Focused Interviews for Inactive Benefit Claimants: An Evaluation of the British ONE Pilots, in: Labour Economics, 11(4): 415–29.

Kleven, H. J., Kreiner, C. T., Saez, E. (2006). The Optimal Income Taxation of Couples, NBER Working Paper No. 12685.

Knab, J., Bos, J., Friedlander, D., Weissman, J. (2000). Do Mandates Matter? The Effects of a Mandate to Enter a Welfare-to-Work Program, New York: mdrc.

Kopczuk, W. (2005). Tax Bases, Tax Rates and the Elasticity of Reported Income, in: Journal of Public Economics, 89(11–12): 2093–119.

Lalonde, R. (1986). Evaluating the Econometric Evaluations of Training Programs with Experimental Data, in: American Economic Review, 76(4): 604–20.

Laroque, G. (2005). Income Maintenance and Labor Force Participation, in: Econometrica, 73(2): 341–76.

Latham, R. W. (1980). Quantity Constrained Demand Functions, in: Econometrica, 48(2): 307–13.

Layard, R. (2000). Welfare to Work and the New Deal, in: The Business Economist, 31(3): 28–40.

Layard, R., Barton, M., Zabalza, A. (1980). Married Women's Participation and Hours, in: Economica, 47(185): 51–72.

Lee, L.-F. (1984). Tests for the Bivariate Normal Distribution in Econometric Models with Selectivity, in: Econometrica, 52(4): 843–63.

Lee, L.-F., Maddala, G. S., Trost, R. P. (1980). Asymptotic Covariance Matrices of Two-Stage Probit and Two-Stage Tobit Methods for Simultaneous Equations Models with Selectivity, in: Econometrica, 48(2): 491–503.

Liebman, J. B. (1998). The Impact of the EITC on Incentives and Income Distribution, in: Poterba, J. (Ed.), Tax Policy and the Economy, Vol. 12, Cambridge, MA: MIT Press, 83–120.

Liebman, J. B. (2002). The Optimal Design of the Earned Income Tax Credit, in: Meyer, B., Holtz-Eakin, D. (Eds.), Making Work Pay: The Earned Income Tax Credit and Its Impact on American Families, New York: Russell Sage Foundation, 196–234.

Lissenburgh, S. (2001). New Deal for the Long Term Unemployed: A Comparison of Provision in Pilot and National Areas, Employment Service Report 69.

Low, H., Meghir, C., Pistaferri, L. (2010). Wage Risk and Employment Risk over the Life Cycle, in: American Economic Review, 100(4): 1432–67.

Lundberg, S. (1985). Tied Wage-Hours Offers and the Endogeneity of Wages, in: Review of Economics and Statistics, 67(3): 405–10.

Lundberg S., Pollak, R. A. (2008). Family Decision-Making, in: Durlauf, S. N., Blume, L. E. (Eds.), The New Palgrave Dictionary of Economics, 2nd edition, London: Palgrave Macmillan.

Lynn, P., Buck, N., Burton, J., Laurie, H., Urhig, S. C. N. (2006). Quality Profile: British Household Panel Survey Discussion Paper, Institute for Social and Economic Research, University of Essex, https://www.iser.essex.ac.uk/files/bhps/quality-profiles/BHPS-QP-01-03-06-v2.pdf.

Macurdy, T. E. (1981). An Empirical Model of Labor Supply in a Life-Cycle Setting, in: Journal of Political Economy, 89(6): 1059–85.

Macurdy, T. E. (1983). A Simple Scheme for Estimating an Intertemporal Model of Labor Supply and Consumption in the Presence of Taxes and Uncertainty, in: International Economic Review, 24(2): 265–89.

Macurdy, T. E. (1985). Interpreting Empirical Models of Labor Supply in an Intertemporal Framework with Uncertainty, in: Macurdy, T. E., Longitudinal Analysis of Labor Market Data, Cambridge: Cambridge University Press.

Macurdy, T. E., Green, D., Paarsch, H. (1990). Assessing Empirical Approaches for Analyzing Taxes and Labor Supply, in: Journal of Human Resources, 25(3): 415–90.

MaCurdy, T. E., McIntyre, F. (2004). Helping Working-Poor Families. Advantages of Wage-Based Tax Credits Over the EITC and Minimum Wages, Washington, DC: Employment Policies Institute.

Manning, A. (2003). Monopsony in Motion: Imperfect Competition in Labor Markets, Princeton, NJ: Princeton University Press.

Martin, J. (1998). What Works Among Active Labour Market Policies: Evidence From OECD Countries' Experiences, OECD Labour Market and Social Policy Occasional Papers No. 35.

Martin, J., Grubb, D. (2001). What Works and For Whom: A Review of OECD Countries' Experience with Active Labour Market Policies, IFAU Working Paper No. 2001:14.

Mazzocco, M. (2007). Household Intertemporal Behavior: A Collective Characterization and a Test of Commitment, in: Review of Economic Studies, 74(3): 857–95.

Mazzocco, M., Saini, S. (2012). Testing Efficient Risk Sharing with Heterogeneous Risk Preferences, in: American Economic Review, 102(1): 428–68.

McFadden, D. (1978). Modeling the Choice of Residential Location, in: Karlqvist, A., Lundqvist, L., Snickars, F., Weibull, J. (Eds.), Spatial Interaction Theory and Planning Models, Amsterdam: North Holland, 75–96.

Meade, J. (1978). The Structure and Reform of Direct Taxation: Report of a Committee chaired by Professor J. E. Meade, London: George Allen and Unwin, http://www.ifs.org.uk/docs/meade.pdf.

Meghir, C. (1985). The Specification of Labour Supply Models and their Use in the Simulation of Tax and Benefit Reforms, Ph.D. Thesis, University of Manchester.

Meghir, C., Palme, M. (2001). The Effects of a Social Experiment in Education, IFS working Paper No. 01/11.

Meghir, C., Phillips, D. (2010). Labour Supply and Taxes, in: Mirrlees, J., Adam, S., Besley, T., Blundell, R., Bond, S., Chote, R., Gammie, M., Johnson, P., Myles, G., Poterba, J. (Eds.), Dimensions of Tax Design: The Mirrlees Review, New York: Oxford University Press, 202–74.

Meghir, C., Whitehouse, E. (1996). The Evolution of Wages in the United Kingdom: Evidence from Micro Data, in: Journal of Labor Economics, 14(1): 1–25.

Meyer, B. D. (1990). Unemployment Insurance and Unemployment Spells, in: Econometrica, 58(4): 757–82.

Meyer, B. D. (1995). Lessons from the U.S. Unemployment Insurance Experiments, in: Journal of Economic Literature 33(1): 91–131.

Meyer, B. D., Rosenbaum, D. T. (2001). Welfare, the Earned Income Tax Credit, and the Labor Supply of Single Mothers, in: The Quarterly Journal of Economics, 116(3): 1063–114.

Miller, C., Knox, V., Auspos, P., Hunter, J. A., Orenstein, A. (1997). Making Welfare Work and Work Pay: Implementation and 18-Month Impacts of the Minnesota Family Investment Program, New York: mdrc.

Mirrlees, J. A. (1971). An Exploration in the Theory of Optimum Income Taxation, in: Review of Economic Studies, 38(114): 175–208.

Mirrlees, J., Adam, S., Besley, T., Blundell, R., Bond, S., Chote, R., Gammie, M., Johnson, P., Myles, G., Poterba, J. (Eds.) (2010). Dimensions of Tax Design: The Mirrlees Review, New York: Oxford University Press.

Mirrlees, J., Adam, S., Besley, T., Blundell, R., Bond, S., Chote, R., Gammie, M., Johnson, P., Myles, G., Poterba, J. (Eds.) (2011). Tax by Design: the Mirrlees Review, New York: Oxford University Press.

Moffitt, R. (1983). An Economic Model of Welfare Stigma, in: American Economic Review, 73(5): 1023–35.

Moffitt, R. (1984). The Estimation of a Joint Wage-Hours Labor Supply Model, in: Journal of Labor Economics, 2(4): 550–66.

Moffitt, R. (1992). Incentive Effects of the U.S. Welfare System: A Review, in: Journal of Economic Literature, 30(1): 1–61.

Moffitt, R. (1998). The Effect of Welfare on Marriage and Fertility, in: R. Moffitt (Ed.), Welfare, the Family and Reproductive Behavior. Research Perspectives, Washington, DC: National Academy Press, 50–97.

Moffitt, R. (2005). Welfare Work Requirements with Paternalistic Government Preferences, Johns Hopkins University Working Paper No. 534.

Moffit, R., Pavetti, L. A. (2000). Time Limits, in: Card, D., Blank, R. (Eds.), Finding Jobs: Work and Welfare Reform, New York: Russell Sage Foundation, 507–36.

Morris, P., Michalopoulos, C. (2000). The Self-Sufficiency Project at 36 Months: Effects on Children of a Program that Increased Parental Employment and Income, Ottawa/Ontario: SDRC.

Moulton, B. R. (1986). Random Group Effects and the Precision of Regression Estimates, in: Journal of Econometrics, 32(3): 385–97.

Mroz, T. A (1987). The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions, in: Econometrica, 55(4): 765–800.

Muellbauer, J. (1974). Household Composition, Engel Curves and Welfare Comparisons between Households, in: European Economic Review 5(2): 103–22.

Muellbauer, J. (1977). Testing the Barten Model of Household Composition Effects and the Cost of Children, in: The Economic Journal 87(347): 460–87.

Muellbauer, J. (1981). Linear Aggregation in Neoclassical Labour Supply, in: Review of Economic Studies, 48(1): 21–36.

Muellbauer, J. (1983). Surprises in the Consumption Function, in: Economic Journal, 93(369a): 34–50.

Myck, M. (2000). Fiscal Reforms since May 1997, IFS Briefing Notes, London: IFS.

Neary, J. P., Roberts, K. W. S. (1980). The Theory of Household Behaviour under Rationing, in: European Economic Review, 13(1): 25–42.

Nickell, S. J. (1997). Unemployment and Labour Market Rigidities: Europe versus North America, in: Journal of Economic Perspectives, 11(3): 55–74.

Nickell, S. J. (1999). Unemployment in Britain, in: Gregg, P., Wadsworth, J. (Ed.), The State of Working Britain, Manchester: Manchester University Press, 7–28.

Ohanian, L., Raffo, A., Rogerson, R. (2008). Long-Term Changes in Labor Supply and Taxes: Evidence from OECD Countries, 1956–2004, in: Journal of Monetary Economics, 55(8): 1353–62.

Organisation for Economic Co-operation and Development (1994). The OECD Jobs Study. Facts, Analysis, Strategies, Paris: OECD.

Organisation for Economic Co-operation and Development (2000). Make Work Pay: Special Issue, OECD Economic Studies No. 31 2000/II.

O'Dea, C., Phillips, D., Vink, A. (2007). A Survey of the UK Benefit System, IFS Briefing Note 13.

O'Neill, D. (2000). Evaluation of Labour Market Interventions, in: Journal of the Statistical and Social Inquiry Society of Ireland, 29: 177–214.

Pagan, A. (1986). Two Stage and Related Estimators and Their Applications, in: Review of Economic Studies, 53(4): 517–38.

Paull, G., Walker, I., Zhu, Y. (2000). Child Support Reform: Some Analysis of the 1999 White Paper, in: Fiscal Studies, 21(1): 105–40.

Phelps, E. S. (1994). Raising the Employment and Pay of the Working Poor: Low wage Employment Subsidies vs the Welfare State, in: AER Papers and Proceedings, 84(2): 54–8.

Phlips, L. (1978). The Demand for Leisure and Money, in: Econometrica, 46(5): 1025–43.

Piketty, T., Saez, E., Stantcheva, S. (2014). Optimal Taxation of Top Labor Incomes: A Tale of Three Elasticities, in: American Economic Journal: Economic Policy, American Economic Association, 6(1), 230–71.

Pistaferri, L. (2003). Anticipated and Unanticipated Wage Changes, Wage Risk, and Intertemporal Labor Supply, in: Journal of Labor Economics, 21(3): 729–54.

Pollak, R. A., Wales, T. J. (1978). Estimation of Complete Demand Systems from Household Budget Data: The Linear and Quadratic Expenditure Systems, in: American Economic Review, 68(3): 348–59.

Pollak, R. A., Wales, T. J. (1979). Welfare Comparisons and Equivalence Scales, in: American Economic Review, 69(2): 216–21.

Prescott, E. (2004). Why Do Americans Work So Much More than Europeans?, in: Quarterly Review, 28(1): 2–13.

Quets, G., Robins, P. K., Pan, E. P., Michalopoulos, C., Card, D. (1999). Does SSP Plus Increase Employment? The Effect of Adding Services to the Self Sufficiency Project's Financial Incentives, Ottawa/Ontario: SDRC.

Riccio, J. A., Bloom, H. S. (2001). Extending the Reach of Randomized Experiments: New Directions in Evaluations of American Welfare to Work and Employment Initiative, MDRC Working Paper No. 10/2001.

Riley, R., Young, G. (2001). The Macroeconomic Impact of the New Deal for Young People, NIESR Discussion Paper No. 184.

Robin, J.-M., Smith, R. J. (2000). Test of Rank, in: Econometric Theory, 16(2): 151–75.

Rogerson, R., Wallenius, J. (2009). Micro and Macro Elasticities in a Life Cycle Model With Taxes, in: Journal of Economic Theory, 144(6): 2277–92.

Rosen, H. S. (1976). Taxes in a Labor Supply Model with Joint Wage-Hours Determination, in: Econometrica, 44(3): 485–507.

Saez, E. (2000). Optimal Income and Transfer Programs: Intensive versus Extensive Labor Supply Responses, NBER Working Paper No. 7708.

Saez, E. (2001). Using Elasticities to Derive Optimal Income Tax Rates, in: Review of Economic Studies 68(1): 205–29.

Saez, E. (2002). Optimal Income Transfer Programs: Intensive versus Extensive Labor Supply Responses, in: The Quarterly Journal of Economics, 117(3): 1039–73.

Saez, E. (2010). Do Taxpayers Bunch at Kink Points?, in: American Economic Journal: Economic Policy, 2(3): 180–212.

Sandmo, A. (1976). Optimal Taxation: An Introduction to the Literature, in: Journal of Public Economics, 6(1–2): 37–54.

Schmitt, J. (1993). The Changing Structure of Male Earnings in Britain, CEP Discussion Papers No. dp0122.

Schmitt, J. (1995). The Changing Structure of Male Earnings in Britain, 1974–1988, in: Freeman, R. B., Katz, L. F. (Eds), Differences and Changes in Wage Structures, Chicago: University of Chicago Press, 177–204.

Schoeni, R. F., Blank, R. M. (2000). What Has Welfare Reform Accomplished? Impacts on Welfare Participation, Employment, Income, Poverty, and Family Structure, NBER Working Papers No. 7627.

Scholz, J. K. (1996). In-Work Benefits in the United States: The Earned Income Tax Credit, in: The Economic Journal 106(434): 156–69.

Shaw, K. L. (1989). Life-Cycle Labor Supply with Human Capital Accumulation, in: International Economic Review 30(2): 431–456.

Shephard, A. (2009). FORTAX: Reference Manual, University College London, Manuscript.

Shephard, A. (2012). Equilibrium Search and Tax Credit Reform, CEPS Research Paper.

Sianesi, B. (2001). An Evaluation of the Active Labour Market Programmes in Sweden, IFAU Working Paper No. 2001:5, http://swopec.hhs.se/ifauwp/abs/ifauwp2001_005.htm

Slemrod, J., Kopczuk, W. (2002). The Optimal Elasticity of Taxable Income, in: Journal of Public Economics, 84(1): 91–112.

Smith, J. A., Todd, P. (2005). Does Matching Overcome LaLonde's Critique of Non-experimental Estimators?, in: Journal of Econometrics 125(1–2): 305–53.

Smith, J. P. (1977). Family Labour Supply over the Life Cycle, in: NBER (Ed.), Explorations in Economic Research, Vol. 4, Cambridge, MA: NBER, 1–72.

Smith, R. J., Blundell, R. (1983). An Exogeneity Test for the Simultaneous Equation Tobit Model with an Application to Labour Supply, Queen's University Working Paper No. 546.

Smith, R. J., Blundell, R. (1986). An Exogeneity Test for a Simultaneous Equation Tobit Model with an Application to Labor Supply, in: Econometrica 54(3): 679–85.

Snower, D. (1994). Converting Unemployment Benefits into Employment Subsidies, in: American Economic Review 84(2): 65–70.

Staiger, D., Stock, J. H. (1997). Instrumental Variables Regression with Weak Instruments, in: Econometrica 65(3): 557–86.

Stern, N. (1986). On the Specification of Labour Supply Functions, in: Blundell, R., Walker, I. (Eds.), Unemployment, Search and Labour Supply, Cambridge, UK: Cambridge University Press, 143–89.

Stewart, M. B. (2004). The Employment Effects of the National Minimum Wage, in: Economic Journal 114(494), C110-C116.

Stewart, M. B., Swaffield, J. K. (1997). Constraints on the Desired Hours of Work of British Men, in: Economic Journal, 107(441): 520–35.

Stewart, M. B., Swaffield, J. K. (2006). The Other Margin: Do Minimum Wages Cause Working Hours Adjustments for Low-Wage Workers?, The Warwick Economics Research Paper No. 746.

Tallis, G. M. (196I). The Moment Generating Function of the Truncated Multinormal Distribution, in: Journal of the Royal Statistical Society 23(1): 223–9.

Tobin, J. (1958). Estimation of Relationships for Limited Dependent Variables, in: Econometrica, 26(1): 24–36.

Van der Berg, G., van der Klaauw, B., van Ours, J. (2004). Punitive Sanctions and the Transition Rate from Welfare to Work, in: Journal of Labor Economics 22(1): 211–41.

Van Reenen, J. (2001). No More Skivvy Schemes? Active Labour Market Policies and the British New Deal for the Young Unemployed in Context, IFS Working Paper No. WP01/09.

Van Reenen, J. (2004). Active Labour Market Policies and the British New Deal for the Young Unemployed in Context, in: Card, D., Blundell, R., Freeman, R. B. (Eds.), Seeking a Premier Economy: The Economic Effects of British Economic Reforms, 1980–2000, Chicago: University of Chicago Press, 461–96.

van Soest, A. (1995). Structural Models of Family Labor Supply: A Discrete Choice Approach, in: Journal of Human Resources 30(1): 63–88.

van Soest, A., Das, M., Gong, X. (2002). A Structural Labour Supply Model with Flexible Preferences, in: Journal of Econometrics, 107(1–2): 345–74.

van Soest, A., Woittiez, I., Kapteyn, A. (1990). Labor Supply, Income Taxes, and Hours Restrictions in the Netherlands, in: Journal of Human Resources 25(3): 517–58.

Wales, T. J., Woodland, A. D. (1980). Sample Selectivity and the Estimation of Labor Supply Functions, in: International Economic Review, 21(2): 437–68.

Walker, R., Wiseman, I. (1997). The Possibility of a British Earned Income Tax Credit, in: Fiscal Studies 18(4): 401–25.

Werning, I. (2007). Pareto Efficient Income Taxation, MIT Working Paper.

West, R. W. (1979). The Effects of the Seattle and Denver Income Maintenance Experiments on the Labor Supply of Young Nonheads." SRI International Research Memorandum 60.

White, H. (1982). Instrumental Variables Regression with Independent Observations, in: Econometrica 50(2): 483–99.

White, M., Lakey, J. (1992). The Restart Effect: Does Active Labour Market Policy Reduce Unemployment?, London: Policy Studies Institute.

Woodbury, S., Spiegelman, R. (1987). Bonuses to Workers and Employers to Reduce Unemployment: Randomized Trials in Illinois, in: American Economic Review 77(4): 513–30.

Wymer, C. R. (1978). Computer programs: RESIMUL manual, Washington D.C.: International Monetary Fund.

Index

About the Author...

Richard Blundell holds the David Ricardo Chair of Political Economy at University College London where he was appointed Professor of Economics in 1984. Since 1986 he has been Research Director of the Institute for Fiscal Studies, where he is also Director of the ESRC Centre for the Microeconomic Analysis of Public Policy. A graduate of University of Bristol and the LSE, he holds Honorary Doctorates from the University of St. Gallen, the Norwegian School of Economics, and the University of Mannheim. He has held visiting professor positions at UBC, MIT and Berkeley.

In 1995 he was awarded the Yrjö Jahnsson Prize, given every two years to the best young economist in Europe, for his work in microeconometrics and the analysis of labour supply, welfare reform and consumer behaviour. In 2000 he was awarded the Econometric Society Frisch Prize Medal for his paper 'Estimating Labour Supply Responses using Tax Reforms'. He was awarded the Commander of the British Empire in 2006 for his services to Economics and Social Science. In 2008 he was the recipient of the Jean-Jacques Laffont Prize given to an economist whose research combines both the theoretical and applied aspects of economics. He was awarded the CES-Ifo Prize in 2010 and the Sandmo Prize in 2011.

He served as President of the European Economics Association (2004) and of the Econometric Society (2006). He was President of the Society of Labor Economics (2010) and of the Royal Economic Society (2011–2013). He has been on the editorial board of many academic journals and was co-editor of Econometrica from 1997–2001 and co-editor of the Journal of Econometrics from 1992–1997. His papers have appeared in academic journals including Econometrica, Review of Economic Studies, American Economic Review, Journal of Political Economy, Quarterly Journal of Economics, Journal of the Royal Statistical Society, Journal of Econometrics, Journal of Applied Econometrics and Economic Journal. He has also published a number of edited volumes and monographs and is one of the editors of the Mirrlees Review of Tax Reform.

Richard Blundell received the 2012 IZA Prize in Labor Economics.

...and the Editors

Andreas Peichl has been Professor of Public Economics at the University of Mannheim and head of the research group "International Distribution and Redistribution" at the Centre for European Economic Research (ZEW) since October 2013. From 2008 until 2013 he was Senior Research Associate at IZA. He received his PhD from the University of Cologne in 2008. Prior to his doctoral studies, he studied economics at the Universities of Marburg and Cologne where he graduated in 2004. He was a visiting scholar at ISER, University of Essex, in 2008 and at UC Berkeley in 2012. His research has been published in academic journals such as the Economic Journal, the Journal of Public Economics and the Journal of Human Resources, among others. His current research interests include public, labor and welfare economics with particular reference to labor supply, optimal taxation, and the evaluation of tax reforms.

Klaus F. Zimmermann has been Full Professor of Economics at the University of Bonn and Director of the Institute for the Study of Labor (IZA Bonn) since 1998. From 2000 until 2011 he was President of the German Institute for Economic Research (DIW Berlin). Zimmermann is Honorary Professor of Economics at the Free University of Berlin (since 2001) and Honorary Professor at the Renmin University of China (since 2006), and was Chairman of the Society of the German Economic Research Institutes (ARGE) from 2005 until 2011. Klaus F. Zimmermann is author or editor of more than 50 books, more than 140 papers in refereed research journals, and almost 150 chapters in collected volumes. He is Editor-in-Chief of the Journal of Population Economics, acts as the Chair of the Economics Section and a Council member of the Academia Europaea, and is a member of the German Academy of Sciences Leopoldina. Zimmermann regularly advises the World Bank, the European Commission, and various national governments. His main research interests are in migration, labor economics, and population economics. In 2013 he received the Outstanding Contribution Award of the European Investment Bank (EIB Prize).