The Dema
MON

THE DEMAND FOR MONEY

Theoretical and Empirical Approaches
Second Edition

THE DEMAND FOR MONEY

Theoretical and Empirical Approaches
Second Edition

Apostolos Serletis
University of Calgary
Canada

Springer

Apostolos Serletis
University of Calgary
Canada

Library of Congress Control Number: 2007924968

ISBN 978-0-387-71726-5 e-ISBN 978-0-387-71727-2

Printed on acid-free paper.

Printed in the United States of America.

9 8 7 6 5 4 3 2 1

springer.com

To Anna and Demitre

Contents

Part 2: Dynamic Monetary Macroeconomics

Part 3: Theoretical Approaches to the Demand for Money

Part 4: Empirical Approaches to the Demand for Money

Part 5: Microfoundations and Monetary Aggregation

Part 6: Microfoundations and the Demand for Money

Part 7: Microeconometrics and the Demand for Money

Acknowledgments

There are many people to thank, and I would like to start with the pioneers who have contributed to the development of the discipline. I hope that their ideas have not been too distorted by compression in the pages of this book.

I am deeply grateful to William Barnett, Douglas Fisher, and Melvin Hinich who provided great advice and comments on critical issues over the years. They influenced the way I have come to think about monetary economics and therefore tangibly shaped the contents of this textbook. I am also grateful to Jim Sowfford and Asghar Shahmoradi for their encouragement and support.

Students at the University of Calgary provided important feedback on draft material. Victor Chwee, Guohua Feng, Periklis Gogas, Terence Molik, Ricardo Rangel-Ruiz, Sajjadur Rahman, Aryeh Rosenberg, and Jason Vaccaro deserve special mention. They were particularly helpful, and this is the spot where I would like to offer my thanks. I also thank Joanne Longworth for a thoroughly professional job of preparing the camera-ready copy of this book. It is no exaggeration to say that this work would not have been possible without her help.

Finally, I want to thank my wife Aglaia, my son Demitre, and my daughter Anna, for their love and support, and my parents, who long ago put me on the 'high road' that led to this book.

Apostolos Serletis
Calgary, Winter 2007

Foreword

Almost half a century has elapsed since the demand for money began to attract widespread attention from economists and econometricians, and it has been a topic of ongoing controversy and research ever since. Interest in the topic stemmed from three principal sources.

First of all, there was the matter of the internal dynamics of macroeconomics, to which Harry Johnson drew attention in his 1971 Ely Lecture on "The Keynesian Revolution and the Monetarist Counter-Revolution," *American Economic Review* 61 (May 1971). The main lesson about money that had been drawn from the so-called "Keynesian Revolution" was — rightly or wrongly — that it didn't matter all that much. The inherited wisdom that undergraduates absorbed in the 1950s was that macroeconomics was above all about the determination of income and employment, that the critical factors here were saving and investment decisions, and that monetary factors, to the extent that they mattered at all, only had an influence on these all important variables through a rather narrow range of market interest rates. Conventional wisdom never goes unchallenged in economics, except where its creators manage to control access to graduate schools and the journals, and it is with no cynical intent that I confirm Johnson's suggestion that those of us who embarked on academic careers in the '60s found in this wisdom a ready-made target. University faculties were expanding at that time, so rewards for hitting that target cleanly were both visible before the event, and quickly available after it, particularly when the weapons employed were those provided by then rapidly developing computer technology. Seldom can a novel hypothesis — in this case that the demand for money is a stable function of a few arguments — have been better calculated simultaneously to undermine established beliefs and to exploit newly available technology. Small wonder that studies of the demand for money flourished in the academic journals.

But second, in the late '50s - early '60s there was a new audience for monetary research outside of the academic community. The relaxation of war-time and post-war controls on economies in the 1950s, and the high employment levels achieved during that decade, began to expose monetary policy to new scrutiny. In Britain, the Radcliffe Committee, in the United States, the Commission on Money and Credit, and in Canada, the Porter Commission, all undertook wide-ranging investigations of the scope and strength of monetary policy at around this time. Though the just beginning academic controversy about the interaction of the supply and demand for money and its effects on output and prices, in which demand for money studies stood at the very centre, was low on the agenda of all of these bodies, they could not ignore it, and their work helped to draw public attention to it.

Third and finally, though it was not readily apparent at the time, the mid-1960s saw the onset of a great and more-or-less worldwide inflation which was to last for the next quarter of a century. This would in due course destroy the international monetary system created for the post-war world at Bretton Woods, force a system of flexible exchange rates on the western economies, and provide them with their principal macroeconomic policy problem of the 1970s and '80s. The fact that this inflation, like all the others that had gone before it, turned out to be largely monetary in nature was the final factor ensuring the demand for money function a place of lasting importance in macroeconomic research.

For a while indeed, in the early 1970s, it almost looked as if a new conventional wisdom might impose itself on macroeconomics, in which the central issue was the behaviour of prices, and the influence of the quantity of money thereon was the crucial factor. Since the latter influence was thought to be transmitted through a mechanism in which the demand for money function was a critical link, the theoretical underpinnings of, and empirical support for, this relationship inevitably attracted considerable attention. And of course the same internal dynamic described by Johnson that had undermined Keynesian macroeconomics in the 1960s soon got under way. Critical attention is the natural consequence of success for any economic doctrine, and a fresh cohort of academics, looking for ways to make their own mark turned their attention to the above-mentioned story, and it proved to be all too easy to find weaknesses in it.

Crucially, careful empirical scrutiny showed that the relationship was not quite so well supported by the data as it had initially seemed to be. In particular, the very progress of the inflation that had lent so much importance to the demand for money function, and policies that relied on it, provoked institutional developments that undermined its stability. Even to give precise empirical content to that deceptively simple word "money" turned out to be extremely difficult and controversial. And as new econometric techniques developed, old empirical truths did not always survive their application. But, the implications of all this were not that money, after all, did not matter. Rather they were seen to be, first that its role in the economy was a great deal more complicated than had previously been thought, and second that this role was also likely to be subject to the effects of ongoing institutional change that would need continuous monitoring by anyone interested in the functioning of the system.

There is no need to go into more detail about all this in this brief *Foreword*. In what follows, Apostolos Serletis has provided his readers with a comprehensive account, not just of the current state of play in the field, but also with a sense of how it got there, and where it is likely to go next. He begins with a brief exposition of macroeconomic theories, showing how the demand for money function fits in not only to the old-fashioned short-run IS-LM model that underlay early work on the topic but into more modern models, both long and short run as well. He then describes the theoretical literature on the demand for money function, beginning with its origins in the pre-Keynesian literature and proceeding to the formulations used in the latest theoretical models. He goes on to provide a wide-ranging survey of the principle econometric techniques that have been and are being used to bring empirical discipline to the area. And, as the final chapter shows, he treats the whole area as a field in which research is in progress, rather than as a collection of established truths.

Here, then, is a book which will be valuable to anyone wishing to get up to date with the state of play in this area, and, more important, to anyone looking for a starting point for further work of their own.

David Laidler,
Bank of Montreal Professor
University of Western Ontario

Introduction

The purpose of the second edition of *The Demand for Money: Theoretical and Empirical Approaches*, is the same as that of the first edition. That is, to provide an account of the existing literature on the demand for money, to show how the money demand function fits into static and dynamic macroeconomic analyses, and to discuss the problem of the definition (aggregation) of money. In doing so, it shows how the successful use in recent years of the simple representative consumer paradigm in monetary economics has opened the door to the succeeding introduction into monetary economics of the entire microfoundations, aggregation theory, and micro-econometrics literatures.

A stable demand function for money is a necessary condition for money to exert a predictable influence on the economy so that control of the money supply can be a useful instrument of economic policy. As such, the notion of a stable money demand function appears to require that money holdings, as observed in the real world, should be predictably related to a small set of variables representing significant links to spending and economic activity in the real sector of the economy.

Prior to 1973, both the theoretical derivation and the econometric form of the money demand function were considered settled, and the evidence was interpreted as showing that the money demand function was stable. This evidence, occurring as it did in a climate of worsening inflation, convinced the Federal Reserve to give emphasis to monetary aggregates targeting. After 1973, however, the standard money demand formulation performed poorly, showing inaccurate forecasting ability and parameter instability — both of which remain largely unexplained today despite extensive research devoted to determining the reasons for this poor performance.

In trying to explain what happened, economists in addition to reopening the pre-1973 agenda of empirical issues (mainly concerned with the inappropriate specification of the original function and the choice of dependent and explanatory variables), pointed to financial innovations (and to a lesser extent regulatory changes) which have led to the emergence of new assets and the changing of the relative degrees of 'moneyness' possessed by the various assets. A review of the vast literature devoted to these issues [see Edgar Feige and Douglas Pearce (1977) and John Judd and John Scadding (1992)] reveals that these studies were largely unsuccessful in explaining the instability in money demand after 1973.

There is another problem with this literature, and this is that the many studies of the demand for money (and of the influence of money on the economy in general) are based on official simple-sum monetary aggregates. There are conditions under which such aggregates are appropriate, but if the relative prices of the financial components that constitute the aggregates fluctuate over time (as the evidence suggests) then simple-sum aggregation will produce theoretically unsatisfactory definitions of money. The problem is the incorrect accounting for substitution effects that simple-sum aggregation entails, and the result is a set of monetary aggregates that do not accurately measure the actual quantities of the monetary products that optimizing economic agents select (in the aggregate).

Recently, attention has been focused on the gains that can be achieved by a vigorous use of microeconomic- and aggregation-theoretic foundations in the construction of monetary aggregates. This new approach to monetary aggregation was advocated by William Barnett (1980) and has led to the construction of monetary aggregates based on Erwin Diewert's (1976) class of superlative quantity index numbers — the most recent example is Richard Anderson, Barry Jones, and Travis Nesmith (1997). The new aggregates are Barnett's monetary services indexes (also known as Divisia aggregates), and Julio Rotemberg, John Driscoll, and James Poterba's (1995) currency equivalent (CE) indexes. These aggregates represent a viable and theoretically appropriate alternative to the simple-sum aggregates still in use both by central banks and researchers in the field.

This new literature is actually an ongoing one that has only just begun to produce empirical results worthy of the effort required to understand it. The main research lies in two areas — the construction of monetary aggregates that conform to the specifications of systems of demand theory and the estimation of systems of monetary asset-demand equations in which the restrictions of demand theory are incorporated in such a manner as to assure consistency with the optimizing behavior of economic agents. I think that this new literature suggests answers to a number of problems raised over previous studies of the demand for money. Most important, I think, is the idea that traditional measures of money and log-linear money demand functions are simply unbelievable in the volatile financial environment in which we find ourselves.

My aim in this textbook is to discuss the problem of the definition (aggregation) of money and to show how the successful use in recent years of the simple representative consumer paradigm in monetary economics has opened the door to the succeeding introduction into monetary economics of the entire microfoundations, aggregation theory, and microeconometrics literatures. In particular, the book will illustrate how a simultaneous-equations monetary assets structure both fits neatly into the new microeconomic- and aggregation-theoretic approach to the definition of money and provides a structure that can be used to measure income and interest rate elasticities as well as the important elasticities of substitution among financial entities.

Although this text has undergone a major revision, it retains the basic hallmarks that have made it the best book on money demand:

- A microeconomic- and aggregation-theoretic approach to the demand for money
- Focus on issues pertaining to the idea that traditional measures of money and log-linear money demand functions are inappropriate for monetary policy purposes
- The presentation of empirical evidence using state-of-the-art econometric methodology
- The recognizing of the existence of unsolved problems and the need for further developments

In addition to the expected updating of all data used in the text, there is major new material in every part of the text. Moreover new material to this edition is:

- a new chapter (Chapter 3) on rational expectations macroeconomics and issues such as the Lucas critique, rules versus discretion in monetary policymaking, and time inconsistency
- a new chapter (Chapter 6) on money demand issues and estimation of the welfare cost of inflation using tools from public finance and applied microeconomics
- increased coverage of the univariate and multivariate properties of the money demand variables, nonlinear chaotic dynamics, and self-organized criticality (see Chapter 11)

- increased coverage of theoretical and empirical approaches to the demand for money, including a new chapter (Chapter 14) on cross-country evidence
- revised coverage of monetary asset demand systems based on locally flexible functional forms such as the translog, generalized Leontief, almost ideal demand system, Minflex Laurent, and the Normalized Quadratic reciprocal indirect utility function (see Chapter 20)
- revised coverage of monetary asset demand systems based on globally flexible functional forms such as the Fourier and the Asymptotically Ideal Model (see Chapter 21)
- increased coverage of the econometrics of demand systems highlighting the challenge inherent with achieving both economic and econometric regularity (see Chapter 22)

The Demand for Money is primarily aimed at upper-level undergraduate and graduate students. The emphasis is on theoretical and empirical approaches to the demand for money and the empirical analysis of data sets. Although the book uses data from the United States economy, it is indended to be used internationally as the main text in one-semester courses in Monetary Economics and as a supplement in a wide range of courses in Macroeconomics, Applied Microeconomics and Applied Econometrics. I hope that those interested in various aspects of the demand for money will find this book valuable.

Apostolos Serletis

Part 1: Static Monetary Macroeconomics

Overview of Part 1

Chapters 1 and 2 concern macroeconomic analysis with a strong emphasis on monetary aspects, in the context of static 'classical' and 'Keynesian' models. These models, as Bennett McCallum (1989, p. 13) puts it, "have been extremely important in macroeconomic analysis and teaching over the last 40 years."

An important feature of these models is that they each incorporate a demand for money function, but make different assumptions about the flexibility of some prices. Our purpose, then, is to investigate the implications for monetary macroeconomics of different assumptions about the money demand function, in different economic environments.

1

Classical Macroeconomic Theory

We begin with an issue described by David Laidler in the (last) 1993 edition of his book, *The Demand for Money: Theories, Evidence, and Problems*, as follows

> "Macroeconomics is controversial. There is no single model upon whose validity all practitioners agree. One area of disagreement of particular importance is the behavior of money wages and money prices. If these are extremely flexible in their response to shocks to the economy, then so will be the general price level. If they are not, then the price level will be slow moving, or 'sticky.' This matters because the general price level is one of the key variables upon which the demand for money depends. If the price level is flexible, then it is free to move to absorb the consequences of shifts in exogenous factors such as the supply of money, and their effects on other variables, notably real income and employment, will be relatively muted. If the price level is sticky, those consequences will spill over onto real income and employment and cause them to fluctuate relatively more." (p. 8)

The above quotation shows that the assumptions we make about the flexibility of prices (and wages) matter. In this chapter, we address the issue by discussing a model that has been extremely important in monetary macroeconomics — the classical model. One important feature of this model is its assumption that prices and nominal wages are fully flexible, in the sense that they continuously adjust to clear markets; the implications of introducing some inflexibility of prices are discussed in the next chapter.

Another important feature of the classical model is that it incorporates a money demand function, a function that explains people's willingness to hold money. Our task is to make as clear as possible what the implications are for monetary macroeconomics of different assumptions about the demand for money function. In describing the model, we follow Chapter 1 of Thomas Sargent's 1979 book, *Macroeconomic Theory.*

1.1 The Complete Classical Model

The classical model can be summarized as consisting of the following seven equations, potentially able to determine seven endogenous variables,

$$\frac{w}{P} = F_L; \tag{1.1}$$

$$L = L\left(\frac{w}{P}\right); \tag{1.2}$$

$$Y = F(K, L); \tag{1.3}$$

$$C = C(R - \pi^e); \tag{1.4}$$

$$I = I\Big(q(K, L, R - \pi^e, \delta)\Big); \tag{1.5}$$

$$Y = C + I + G; \tag{1.6}$$

$$\frac{M}{P} = \Phi(Y, R). \tag{1.7}$$

Equation (1.1) is the demand function for labor, derived by maximizing economy wide profits with respect to employment. The basic hypothesis is that firms maximize profits (that is, gross revenue less factor costs). Formally, the firms' problem is

$$\max_L \Big\{PF(K, L) - wL - (R - \pi^e + \delta)PK\Big\},$$

where P the price of the economy's single good, w the money wage rate, and $(R - \pi^e + \delta)$ the cost of capital. The reader should note that R is the nominal interest rate on bonds, π^e the expected inflation rate, and δ the rate of depreciation of capital. Taking the stock of capital as given, the first-order condition for profit maximization with respect to L is

$$PF_L - w = 0,$$

which can be rewritten as in equation (1.1), and states that firms maximize profits by equating the marginal product of labor, F_L, to the real wage rate, w/P.

Equation (1.2) is the labor supply function and describes the labor-leisure preferences of workers. It is assumed that the supply of labor is an increasing function of w/P (that is, $L' > 0$) and that the labor market is in equilibrium (that is, actual employment, L, equals labor supply, L^s). Equation (1.3) is the aggregate production function where Y is output of the economy's single good, with K and L denoting capital and labor inputs. We assume that both marginal products are positive but diminishing, that is,

$$F_L > 0, \;\; F_K > 0, \;\; F_{LL} < 0, \;\; F_{KK} < 0,$$

where subscripts stand for partial derivatives. We also assume that capital and labor are complements, that is,

$$F_{LK} = F_{KL} > 0.$$

Equation (1.4) is the consumption function relating real consumption spending, C, to the real interest rate on bonds, $R - \pi^e$, which is the difference between the nominal interest rate, R, and the expected inflation rate, π^e. It is assumed that $C' < 0$, because of the intertemporal substitution effect arising from changes in the rate of interest — see Robert Barro (1997, Chapter 3). Equation (1.5) is the investment function that relates real investment spending by firms, I, to the relative price q, defined by

$$q(K, L, R - \pi^e, \delta) = \frac{F_K - \delta}{R - \pi^e}.$$

The assumption is that investment demand is a function of the gap between the real rate of return to physical capital, $F_K - \delta$, and the real rate of return to financial capital, $R - \pi^e$. In particular, investment demand is higher the higher is the marginal product of capital and the lower is the real interest rate, $R - \pi^e$; that is, $I' > 0$ — see Barro

(1997, Chapter 9) for more details regarding a theoretical analysis of investment. Notice that the derivatives of q with respect to K, L, and $R - \pi^e$ are

$$q_K = \frac{F_{KK}}{R - \pi^e} < 0;$$

$$q_L = \frac{F_{KL}}{R - \pi^e} > 0;$$

$$q_{R-\pi^e} = -\frac{q}{R - \pi^e} < 0,$$

so q is an increasing function of L and a decreasing function of K and $R - \pi^e$.

Equation (1.6) is the national income identity linking aggregate real output, Y, and its components — real consumption, C, real investment, I, and real government purchases, G. Finally, equation (1.7) characterizes portfolio equilibrium by equating the real money supply, M/P — which is the ratio of the nominal money supply, M, to the price level, P — and the real money demand, $\Phi(Y, R)$. Notice that real output, Y, enters the $\Phi(\cdot)$ function as a proxy for the rate of transactions in the economy and also that the nominal interest rate, R, enters the $\Phi(\cdot)$ function as a proxy for the opportunity cost of holding money — which is the real interest rate on bonds, $R - \pi^e$, less the real interest rate on money, $-\pi^e$. We assume that

$$\Phi_Y > 0;$$

$$\Phi_R < 0,$$

that is, the demand for money depends positively on real income and negatively on the nominal interest rate.

Assuming that at any moment the stock of capital is fixed, equations (1.1)-(1.7) determine seven endogenous variables:

$$L,\ \frac{w}{P},\ Y,\ C,\ I,\ R,\ \text{and}\ P.$$

The exogenous variables are:

$$G,\ K,\ M,\ \text{and}\ \pi^e.$$

The parameters of the model,

$$F_K,\ F_{KK},\ F_L,\ F_{LL},\ F_{KL},\ L',\ C',\ I',$$

$$q_L,\ q_K,\ q_{R-\pi^e},\ \Phi_Y,\ \Phi_R,\ \text{and}\ \delta,$$

determine the shapes of the underlying functions. Notice that we assume that the expected inflation rate, π^e, is exogenously determined.

1.2 The Classical Dichotomy

The hallmark of classical macroeconomic theory is its separation of real and nominal variables, known as the *classical dichotomy.* This classical dichotomy arises because in the classical model changes in the money supply do not influence real variables and allows us to study first how the values of the real variables are determined in isolation. Given the equilibrium values of the real variables, the equilibrium in the money market then determines the price level and, as a result, all other nominal variables.

It is easy to verify that the classical model we have been studying dichotomizes. Consider the model formed by equations (1.1)-(1.7) and assume that an initial equilibrium exists. Write the model in change form to obtain the following linear system (assuming, for simplicity, that δ is always constant, so $d\delta = 0$)

$$d\,(w/P) = F_{LL}dL + F_{LK}dK; \tag{1.8}$$

$$dL = L'd\,(w/P)\,; \tag{1.9}$$

$$dY = F_K dK + F_L dL; \tag{1.10}$$

$$dC = C'\,(dR - d\pi^e)\,; \tag{1.11}$$

$$dI = I'q_K dK + I'q_L dL + I'q_{R-\pi^e}\,(dR - d\pi^e)\,; \tag{1.12}$$

$$dY = dC + dI + dG; \tag{1.13}$$

$$\frac{dM}{P} - \frac{M}{P}\frac{dP}{P} = \Phi_Y dY + \Phi_R dR. \tag{1.14}$$

Notice that this system is not fully simultaneous. In particular, only two endogenous variables, $d(w/P)$ and dL, appear in the first two equations, implying that these two equations form an independent subset that can determine employment and the real wage rate. Similarly, only three endogenous variables, $d(w/P)$, dL, and dY, appear in the first three equations. As a consequence, these equations form an independent subset that determines employment, the real wage, and output. This very important property of the classical model is known as *block* *recursiveness* and is what yields the dichotomy. That is, the key real variables (output and employment) are determined solely in a subsystem involving only production considerations, and are independent of the level of the money supply and the general price level. In such a system *money is a veil*.

1.3 The Classical AD-AS Model

In order to solve the classical model, we utilize the aggregate demand (AD)-aggregate supply (AS) apparatus. That is, we collapse equations (1.8)-(1.14) into a system of two equations in dR and dY. This is accomplished by eliminating $d(w/P)$, dL, dC, dI, and dP by substitution.

First we obtain the total differential of the aggregate supply schedule. Substituting (1.9) into (1.8) to eliminate dL yields

$$d\left(\frac{w}{P}\right) = \frac{F_{LK}}{1 - F_{LL}L'}dK, \tag{1.15}$$

which implies that an increase in the capital stock increases the real wage, since

$$F_{LK} > 0, \; F_{LL} < 0, \; L' > 0,$$

and hence

$$\frac{F_{LK}}{(1 - F_{LL}L')} > 0.$$

Also, substituting (1.15) into (1.9), to eliminate $d(w/P)$, yields

$$dL = L'\frac{F_{LK}}{1 - F_{LL}L'}dK, \tag{1.16}$$

which implies that an increase in the capital stock also increases employment.

The total differential of the aggregate supply curve can be obtained by substituting (1.16) into (1.10), to eliminate dL

$$dY = \left(F_K + \frac{F_L L' F_{LK}}{1 - F_{LL}L'}\right) dK. \tag{1.17}$$

Equation (1.17) implies that an increase in the capital stock would increase output. In fact, the increase in capital increases output, both because the marginal product of capital is positive as well as because the increase in capital increases the marginal product of labor.

Clearly, equations (1.15)-(1.17) completely determine the values of the only three endogenous variables involved and show that K is the only exogenous variable that enters into the determination of Y, L, and w/P — in this model, there is no interaction with other variables. Thus output, Y, is determined independently of the price level. In what follows we assume that capital can be accumulated only by investing, thus ruling out once-and-for-all changes in the stock of capital. This implies that, at a point in time, output, employment, and the real wage are constants, independent of fiscal and monetary variables and the public's expectations.

We now turn our attention to deriving the total differential of the aggregate demand schedule. Assuming that $dK = 0$ [which implies, from solving (1.8)-(1.10), that $dY = dL = 0$] and substituting (1.11) and (1.12) into (1.13) yields the total differential of the AD schedule or, equivalently, the total differential of the reduced form of R (after solving for dR)

$$dR = -\frac{1}{C' + I' q_{R-\pi^e}} dG + d\pi^e, \tag{1.18}$$

where $C' + I' q_{R-\pi^e}$ — the total derivative of aggregate demand with respect to the interest rate — is negative since $C' < 0$, $I' > 0$, and $q_{R-\pi^e} < 0$.

Manipulation of the reduced form for R, equation (1.18), implies that

$$\frac{\partial R}{\partial G} = -\frac{1}{C' + I' q_{R-\pi^e}} > 0;$$

$$\frac{\partial R}{\partial \pi^e} = 1.$$

Thus the nominal interest rate rises in response to an increase in government spending. Also, a change in π^e produces an equivalent change in R, with no change in $R - \pi^e$ — the *Fisher effect*. Notice that in this version of the model, the interest rate bears the entire burden of adjusting the level of aggregate demand, so that it equals the level of aggregate supply determined by equations (1.1)-(1.3), given the capital stock.

To determine the effect of changes in G and π^e on consumption and net investment we substitute (1.18) into (1.11) and (1.12), respectively, and solve for the reduced form partial derivatives with respect to these exogenous variables, keeping $dK = dL = dY = 0$. The effects on consumption and investment are

$$\frac{\partial C}{\partial G} = C' \frac{\partial R}{\partial G} < 0;$$

$$\frac{\partial I}{\partial G} = I' q_{R-\pi^e} \frac{\partial R}{\partial G} < 0;$$

$$\frac{\partial C}{\partial \pi^e} = C' \frac{\partial R}{\partial \pi^e} - C' = 0;$$

$$\frac{\partial I}{\partial \pi^e} = I' q_{R-\pi^e} \frac{\partial R}{\partial \pi^e} - I' q_{R-\pi^e} = 0.$$

Thus, an increase in government expenditures tends to increase the interest rate, which in turn, through equations (1.11) and (1.12) induces changes in consumption and rates of capital accumulation. In fact, the rise in the interest rate *crowds out* both forms of private spending, C and I. However, changes in π^e do not affect consumption and investment. This is so because of the Fisherian link, according to which a change in π^e leads to an equivalent change in R, leaving $R - \pi^e$ unchanged.

Once the differentials for R and Y are determined, equation (1.14) has only one free variable in it — the differential of the price level, dP. In fact, the role of equation (1.14) is to determine dP/P to equate the nominal demand for money to the given nominal quantity of money.

1.4 The Neutrality of Money

So far we have dealt with changes in the demand for money while holding fixed the aggregate supply of nominal money. We have shown that disturbances which end up changing the interest rate or output change the demand for real money balances. With the money stock held constant, the price level then changes to clear the money market. Notice that the price level moves in the direction opposite to changes in the real demand for money. For example, increases in R reduce the demand for real money and drive the price level upward, while increases in Y increase the demand for real money and drive the price level downward.

Although disturbances that end up changing output or the interest rate are possible sources of price level changes, many economists argue that fluctuations in nominal money, M, are the principal source of variations in the price level. In fact, if the supply of money is the only exogenous variable that changes, equation (1.14) implies that only the price level is affected, and it changes proportionately with the money supply — i.e.,

$$\frac{dP}{P} = \frac{dM}{M},$$

since $dY = dR = 0$. This property of the classical model is referred to as the *neutrality of money*, meaning the (null) effect on real variables of a once-and-for-all change in the nominal money supply.

Notice that monetary neutrality and dichotomy are distinct concepts. For example, a system that dichotomizes need not posses the property of neutrality, while a system in which there is neutrality need not dichotomize. See Sargent (1979, p. 47) for an (artificial) example of a system that dichotomizes but in which neutrality fails.

1.5 Conclusion

We began with a representation of the aggregate economy designed to facilitate analysis of the interaction between real and monetary variables, under perfect wage and price flexibility — the classical model. Although the model is essentially static in nature, it is still used by a large part of the economics profession and provides a good introduction to some of the important issues pertaining to the role of money in the economy and the importance of the money demand function.

An important result is that — with wage and price flexibility — the real variables (real wage rate, real interest rate, and the aggregates of output, consumption, investment, and employment) are invariant with variations in the quantity of money. Money is, therefore, neutral in the model — only the general price level and all other nominal variables (nominal output, consumption, investment, and so on) are (equiproportionally) affected by changes in the supply of nominal money balances. In other words, the money market and the money demand function play no crucial role in determining the aggregates of output and employment.

So far we haven't been able to show why the money demand function is an important relationship. In the next chapter, we investigate how inflexibility of some prices, and the resulting imbalance between quantities supplied and demanded, change the nature of our conclusions regarding the importance of the demand for money function.

2

Keynesian Macroeconomic Theory

In Chapter 1 we began our discussion of macroeconomic theory with a view of nominal wages and prices as fully flexible. This approach ensures that markets are always in equilibrium, in the sense that there is continual balance between the quantities demanded and the quantities supplied. The classical model was the dominant macroeconomic theory until the Great Depression in the 1930s. The prolonged unemployment, however, in the United Kingdom and the United States during the 1930s prompted John Maynard Keynes to significantly depart from the classical assumption of perfectly flexible prices and develop models based on the assumption that there are constraints on the flexibility of some prices.

The crucial assumption in the Keynesian models is that some prices are sticky — i.e., do not adjust promptly to ensure continual balance between the quantities supplied and demanded. Hence, unlike the classical model, some markets do not always clear and output and employment typically end up below the optimal amounts. Although Keynes's analysis and some subsequent treatments — see, for example, Don Patinkin (1965, Chapter 13) and Barro and Herschel Grossman (1976) — focused on sticky money wages, the price level is sometimes assumed to be perfectly flexible (leading to the so-called *complete Keynesian model*), but is more often treated also as sticky. Here, we follow Sargent (1979, Chapter 2) and develop the complete Keynesian model.

2.1 The Keynesian Consumption Function

The key assumption in Keynesian analysis is that sticky money wages result in excess supply in the labor market. This has important consequences for employment and consumption. For example, producers do not produce more than is demanded, suggesting that employment is restricted to the minimum necessary to produce the given level of output. Also, consumers find that their income is constrained to be less than it would have been in the absence of excess supply in the output market.

The implication of this is that people's consumption depends on their exogenously given level of real income. Formally, the consumption function now takes the form

$$C = C(Y - T,\ R - \pi^e),$$

relating real consumption spending, C, to real disposable income, $Y-T$ (where T is defined as real tax collections net of transfers), and to the real rate of interest, $R-\pi^e$. It is assumed that $0 < C_1 < 1$ and $C_2 < 0$, where C_1 is the marginal propensity to consume out of real disposable income and C_2 is the interest sensitivity of consumption demand. The above function is known as the *Keynesian consumption function*.

Another argument that has been put forward in deriving the Keynesian consumption function concerns the loan market. In particular, it is argued that many people are *liquidity constrained*, in the sense that they would like to borrow at market rates, but face higher borrowing costs, because they have poor collateral. In this case, people will change their consumption almost one-to-one with changes in their income. Hence, this point of view can also explain why consumption is a function of current income — see Barro (1997, Chapter 20) for a more detailed discussion.

2.2 The Complete Keynesian Model

The Keynesian model's assumption of excess supply in the output market means that employment is determined by labor demand, which defines the *short side* of the market. Therefore, removing the labor supply schedule, $L = L(w/P)$, and the implicit labor-market-clearing condition ($L^D = L^S = L$), making the money wage rate exogenous, and making consumption also a function of real disposable income are the essential changes that must be made in the classical model of Chapter 1 in order to arrive at the Keynesian model.

The resulting system, which constitutes the complete Keynesian model, consists of the following six equations:

$$\frac{\overline{w}}{P} = F_L; \tag{2.1}$$

$$Y = F(K, L); \tag{2.2}$$

$$C = C(Y - T, R - \pi^e); \tag{2.3}$$

$$I = I\left(q(K, L, R - \pi^e, \delta)\right); \tag{2.4}$$

$$Y = C + I + G; \tag{2.5}$$

$$\frac{M}{P} = \Phi(Y, R), \tag{2.6}$$

where $\overline{w}$ denotes the exogenous money wage, determined outside the system. In fact, we treat $\overline{w}$ as a predetermined variable, not as one that is strictly exogenous, and we assume that it changes through time — that is, $d\overline{w}/dt$ can be nonzero. Notice that the other assumptions underlying these equations are all as they were in the classical model. Thus, the Keynesian model consists of the above six equations in the six endogenous variables:

$$L,\ Y,\ C,\ I,\ R,\ \text{and } P.$$

The exogenous variables are:

$$M,\ G,\ T,\ K,\ \pi^e,\ \delta,\ \text{and } \overline{w},$$

and the parameters of the model are:

$$F_K,\ F_{KK},\ F_L,\ F_{LL},\ F_{KL},\ L',\ C_1,\ C_2,\ I',$$

$$q_L,\ q_K,\ q_{R-\pi^e},\ \Phi_Y,\ \Phi_R,\ \text{and } \delta.$$

Clearly, the Keynesian model utilizes the same theory of aggregate demand as the classical model [i.e., equations (2.3)-(2.6) are the same in the two models] but a different theory of aggregate supply. To investigate the implications of the Keynesian version of the aggregate supply function, we totally differentiate the above six equations (assuming that $dK = d\delta = 0$) to obtain the following linear system in the differentials of the six variables:

$$\frac{d\overline{w}}{\overline{w}} - \frac{dP}{P} = \frac{F_{LL}}{F_L} dL; \tag{2.7}$$

$$dY = F_L dL; \tag{2.8}$$

$$dC = C_1 dY - C_1 dT + C_2 (dR - d\pi^e); \tag{2.9}$$

$$dI = I' q_L dL + I' q_{R-\pi^e} \left(dR - d\pi^e\right); \tag{2.10}$$

$$dY = dC + dI + dG; \tag{2.11}$$

$$\frac{dM}{P} - \frac{M}{P}\frac{dP}{P} = \Phi_Y dY + \Phi_R dR. \tag{2.12}$$

Notice that this system, unlike the classical system, is not block recursive, in the sense that it is impossible to find an independent subset of equations that determine a subset of variables. That means, of course, that output is not determined solely on the basis of aggregate supply considerations, as it is in the classical model. In other words, the Keynesian model does not dichotomize.

In what follows, we shall generate the basic Keynesian results in the context of three different versions of the Keynesian model. In particular, the Keynesian-cross, the IS-LM, and the AD-AS Keynesian models are developed.

2.3 The Keynesian-Cross Model

An extreme version of the Keynesian model is the Keynesian-cross. In addition to assuming that there is perpetual excess supply in the goods market, the Keynesian-cross model also assumes that the nominal interest rate is fixed. This allows it to ignore the money market and focus exclusively on the goods market to determine the level of output, which is demand determined.

We can summarize the main aspects of this simple Keynesian model using equations (2.8), (2.9), (2.10), and (2.11) to obtain the total differential of the reduced form of Y (assuming that $dR = d\pi^e = 0$)

$$\left(1 - C_1 - I' \frac{q_L}{F_L}\right) dY = dG - C_1 dT. \tag{2.13}$$

Assuming that the marginal propensity to save out of disposable income, $1 - C_1$, exceeds the marginal propensity to invest out of income,

$I'q_L/F_L$, the coefficient on dY in (2.13) is positive.[1] Then the reduced form partial derivatives of Y with respect to G and T are given by

$$\frac{\partial Y}{\partial G} = \frac{1}{1 - C_1 - I'q_L/F_L};$$

$$\frac{\partial Y}{\partial T} = -\frac{C_1}{1 - C_1 - I'q_L/F_L},$$

where the expression for $\partial Y/\partial G$ is the *government purchases multiplier* — the amount output changes in response to a unit change in government purchases. The expression for $\partial Y/\partial T$ is the *tax multiplier* — the amount output changes in response to a unit change in taxes. Notice that if investment does not respond to changes in income, that is if I' equals zero, the above expressions reduce to the standard simple Keynesian multiplier formulas.

Finally, if $I' = 0$, the effect on output of a change in G matched by an equal change in T is given by

$$\left.\frac{\partial Y}{\partial G}\right|_{dG=dT} = 1,$$

which is the so-called *balanced budget multiplier.*

2.4 The IS-LM Model

The Keynesian-cross model shows how to determine the level of output for a given interest rate. The assumption, however, that the interest rate is given means that the analysis is seriously incomplete. Therefore, we now want to go further to determine simultaneously the interest rate and the level of output. To carry this analysis we use John Hick's (1937) IS-LM curve apparatus. That is, we collapse equations (2.7), (2.8), (2.9), (2.10), (2.11), and (2.12) into a system of two equations in dY and dR, this being accomplished by eliminating the other endogenous variables by substitution.

---

[1] To see that $I'q_L/F_L$ is the marginal propensity to invest out of income, differentiate the investment schedule partially with respect to Y to obtain

$$\frac{\partial I}{\partial Y} = \frac{\partial I}{\partial q}\frac{\partial q}{\partial L}\frac{\partial L}{\partial Y} = I'\frac{q_L}{F_L}.$$

First we obtain the total differential of the IS curve, the locus of the combinations of R and Y that satisfy (2.5), the aggregate demand-aggregate supply equality. Substituting (2.8), (2.9), and (2.10) into (2.11) and rearranging yields the total differential of the IS curve

$$\left(1 - C_1 - I' q_L / F_L\right) dY = -C_1 dT + dG + \left(C_2 + I' q_{R-\pi^e}\right)\left(dR - d\pi^e\right). \tag{2.14}$$

The slope of the IS curve in the $R - Y$ plane is thus given by

$$\frac{dR}{dY} = \frac{1 - C_1 - I' q_L / F_L}{C_2 + I' q_{R-\pi^e}},$$

which is negative, since $C_2 + I' q_{R-\pi^e} < 0$ and $1 - C_1$ has been assumed to be greater than $I' q_L / F_L$. Notice that the smaller the government purchases multiplier and the smaller the sensitivity of aggregate demand to the interest rate, $C_2 + I' q_{R-\pi^e}$, the steeper the IS curve.

To determine how the IS curve shifts when the exogenous variables, T, G, and π^e, change, we can use (2.14) to determine the horizontal shift in the IS curve by evaluating the partial derivatives of Y with respect to each exogenous variable, dR being set equal to zero. Alternatively, we can use (2.14) to determine the vertical shift in the IS curve by evaluating the partial derivatives of R with respect to each exogenous variable, dY being set equal to zero. So we have:

$$\frac{\partial Y}{\partial T} = \frac{-C_1}{1 - C_1 - I' q_L / F_L} < 0;$$

$$\frac{\partial Y}{\partial G} = \frac{1}{1 - C_1 - I' q_L / F_L} > 0;$$

$$\frac{\partial R}{\partial \pi^e} = 1.$$

An increase in government purchases or a decrease in taxes will shift the IS curve out to the right, the extent of the shift depending on the size of the relevant (Keynesian-cross model) multiplier. Also, when the expected inflation rate changes, the IS curve shifts upward by the amount of the increase in π^e.

The IS curve does not determine either R or Y. It only provides the combinations of nominal interest rates and income (output) that clear the goods market. To determine the equilibrium of the economy, we need another relationship between these two variables, to which we now turn.

By using (2.7) and (2.8) to eliminate dP/P from (2.12) yields the total differential of the LM curve, a schedule that shows all combinations of interest rates and levels of income that clear the market for money balances

$$\left(\frac{F_{LL}}{F_L^2}\frac{M}{P} - \Phi_Y\right) dY = -\frac{dM}{P} + \frac{M}{P}\frac{d\overline{w}}{\overline{w}} + \Phi_R dR. \tag{2.15}$$

The slope of the LM curve is

$$\frac{dR}{dY} = \frac{1}{\Phi_R}\left(\frac{F_{LL}}{F_L^2}\frac{M}{P} - \Phi_Y\right) > 0.$$

Notice that the smaller the interest sensitivity and the larger the income sensitivity of the demand for money, the steeper the LM curve. In fact, as $\Phi_R \to 0$, the LM curve approaches a vertical position while as $\Phi_R \to -\infty$, as is supposed in the case of the *liquidity trap*, the LM curve approaches a horizontal position.

To determine how the LM curve shifts when the exogenous variables, M and $\overline{w}$ change, we use equation (2.15) to evaluate the partial derivatives of R with respect to each of the exogenous variables, dY being set equal to zero. Thus,

$$\frac{\partial R}{\partial M} = \frac{1}{\Phi_R P};$$

$$\frac{\partial R}{\partial \overline{w}} = -\frac{M}{\Phi_R P \overline{w}}.$$

The expression $\partial R/\partial M$ is zero when $\Phi_R \to -\infty$ and negative when $\Phi_R > -\infty$. Also, the expression $\partial R/\partial \overline{w}$ is zero when $\Phi_R \to -\infty$ and positive as long as $\Phi_R > -\infty$. Hence, the LM curve shifts down and to the right when the nominal money supply rises or the money wage falls.

We now have all the components of the IS-LM model. Given that the two equations of this model are (2.14) and (2.15) we can solve the system analytically to analyze the (short run) effects of policy changes and other events on national income. Alternatively, using our knowledge of how changes in the various exogenous variables of the model shift the IS and LM curves, we can make use of a graphical device — see, for example, Barro (1997, Chapter 20).

Substituting (2.15) into (2.14) to eliminate dR yields the total differential of the reduced form of Y

$$HdY = -C_1 dT + dG - (C_2 + I' q_{R-\pi^e}) d\pi^e$$

$$+ \frac{C_2 + I' q_{R-\pi^e}}{\Phi_R} \left(\frac{dM}{P} - \frac{M}{P} \frac{d\overline{w}}{\overline{w}} \right),$$

where the coefficient on dY, H, is given by

$$H = 1 - C_1 - I' \frac{q_L}{F_L} - \frac{C_2 + I' q_{R-\pi^e}}{\Phi_R} \left(\frac{F_{LL}}{F_L^2} \frac{M}{P} - \Phi_Y \right).$$

Under the assumption that $1 - C_1$ exceeds $I' q_L / F_L$, H is positive and the reduced form partial derivatives of Y with respect to the exogenous variables of the model are given by

$$\frac{\partial Y}{\partial T} = -\frac{C_1}{H} \leq 0; \quad \frac{\partial Y}{\partial G} = \frac{1}{H} \geq 0;$$

$$\frac{\partial Y}{\partial \pi^e} = -\frac{C_2 + I' q_{R-\pi^e}}{H} \geq 0; \quad \frac{\partial Y}{\partial M} = \frac{C_2 + I' q_{R-\pi^e}}{\Phi_R P H} \geq 0;$$

$$\frac{\partial Y}{\partial \overline{w}} = -\frac{(C_2 + I' q_{R-\pi^e})\, M}{\Phi_R P \overline{w} H} \leq 0.$$

Thus, except in limiting cases, increases in G, π^e, and M and decreases in T and $\overline{w}$ will in general increase the level of real income. Therefore, money is not neutral in this model.

Notice that if money demand is insensitive to the interest rate ($\Phi_R \to 0$ and the LM curve is vertical), $H \to \infty$ and the effect on output from a disturbance that shifts the IS curve is nil, that is, $\partial Y/\partial T$, $\partial Y/\partial G$, and $\partial Y/\partial \pi^e$ all approach zero. Under those circumstances, a fiscal expansion raises the interest rate and crowds out interest sensitive private spending. However, any shift in the (vertical) LM curve has a maximal effect on the level of income.

On the other hand, in the liquidity trap ($\Phi_R \to -\infty$ and the LM curve is horizontal), monetary policy has no impact on the equilibrium of the economy, since $\partial Y/\partial M$ and $\partial Y/\partial \overline{w}$ both approach zero. Fiscal policy, however, has its full multiplier effect on the level of income, since $\partial Y/\partial T$ and $\partial Y/\partial G$ reduce to the tax multiplier and government purchases multiplier, respectively, of the Keynesian cross model.

Finally, if the interest rate has a negligible effect on aggregate demand ($C_2 = I' = 0$), the IS curve is vertical and changes in the money supply and the money wage have no effect on output, that is, $\partial Y/\partial M = \partial Y/\partial \overline{w} = 0$. On the other hand, if aggregate demand is extremely sensitive to the interest rate, the IS curve is very flat and shifts in the LM have a large effect on output.

2.5 The Keynesian AD-AS Model

In the previous section we solved the Keynesian model [equations (2.7), (2.8), (2.9), (2.10), (2.11), and (2.12)] by collapsing it into two equations in a pair of variables, dR and dY. We can also solve the same model by collapsing it into two equations in another pair of variables, dP and dY, thereby obtaining the Keynesian version of the aggregate demand (AD)-aggregate supply (AS) model.

Solving (2.7) and (2.8) for dY yields the total differential of the aggregate supply function in the $P-Y$ plane

$$dY = \frac{F_L^2}{F_{LL}}\frac{d\overline{w}}{\overline{w}} - \frac{F_L^2}{F_{LL}}\frac{dP}{P}. \tag{2.16}$$

Since $F_{LL} < 0$, equation (2.16) implies that aggregate supply increases in response to an increase in the price level or (for a given price level) a decline in the money wage. The slope of the aggregate supply schedule is

$$\frac{dP}{dY} = -\frac{PF_{LL}}{F_L^2} > 0.$$

This expression equals zero if the marginal product of labor is constant — that is, if $F_{LL} = 0$ — as it happens, for example, when capital and labor are combined in fixed proportions.

The total differential of the aggregate demand curve in the $P-Y$ plane, a schedule that, in the present context, represents all those combinations of P and Y that satisfy the demands for goods and assets, comes from equations (2.9), (2.10), (2.11), and (2.12) and is given by

$$\widehat{H}dY = -C_1dT + dG - (C_2 + I'q_{R-\pi^e})d\pi^e$$

$$+ \frac{C_2 + I'q_{R-\pi^e}}{\Phi_R}\frac{dM}{P} - \frac{C_2 + I'q_{R-\pi^e}}{\Phi_R}\frac{M}{P^2}dP, \tag{2.17}$$

where the coefficient on dY, $\widehat{H}$, is

$$\widehat{H} = 1 - C_1 - I' \frac{q_L}{F_L} + \frac{C_2 + I' q_{R-\pi^e}}{\Phi_R} \Phi_Y.$$

The slope of the aggregate demand schedule is thus given (substituting back for $\widehat{H}$) by

$$\frac{dP}{dY} = -\frac{\left[\left(1 - C_1 - I' \frac{q_L}{F_L}\right) \Phi_R + (C_2 + I' q_{R-\pi^e}) \Phi_Y\right] P^2}{(C_2 + I' q_{R-\pi^e}) M},$$

which, under the assumption that $1 - C_1 > I' q_L / F_L$, is negative.

Notice that the AD curve is flatter the smaller the interest sensitivity of the demand for money, Φ_R, the smaller the income sensitivity of money demand, Φ_Y, and the larger the interest sensitivity of aggregate demand, $C_2 + I' q_{R-\pi^e}$. Also, the larger the marginal propensity to consume out of disposable income, C_1, (or, equivalently, the smaller the marginal propensity to save out of disposable income, $1 - C_1$), and the larger the sensitivity of investment demand to income, that is, the larger $I' q_L / F_L$, the flatter the AD curve. It is also interesting to note that as $C_2 + I' q_{R-\pi^e} \to 0$ or $\Phi_R \to -\infty$, the aggregate demand curve becomes vertical in the $P - Y$ plane.

To determine how the aggregate demand curve shifts when the exogenous variables, T, G, π^e, and M change, we use equation (2.17) to evaluate the partial derivatives of Y with respect to each exogenous variable, dP being set equal to zero. Letting $\widehat{H}$ stand for the coefficient on dY in equation (2.17) we obtain

$$\frac{\partial Y}{\partial T} = -\frac{C_1}{\widehat{H}} < 0; \quad \frac{\partial Y}{\partial G} = \frac{1}{\widehat{H}} > 0;$$

$$\frac{\partial Y}{\partial \pi^e} = -\frac{C_2 + I' q_{R-\pi^e}}{\widehat{H}} > 0; \quad \frac{\partial Y}{\partial M} = \frac{C_2 + I' q_{R-\pi^e}}{\widehat{H} \Phi_R P} > 0.$$

Thus, increases in G, π^e, and M and decreases in T will in general shift the aggregate demand curve outward and/or upward in the $P - Y$ plane.

We now have all the components of the Keynesian AD-AS model. It consists of equations (2.16) and (2.17), which we can solve to analyze the effects of policy actions on national income — alternatively, we can make use of a graphical device as, for example, in Barro (1997, Chapter 20). In particular, substituting (2.16) into (2.17) to eliminate dP/P yields the total differential of the reduced form of Y

$$\widetilde{H}dY = -C_1 dT + dG - (C_2 + I' q_{R-\pi^e}) d\pi^e$$

$$+ \frac{C_2 + I' q_{R-\pi^e}}{\Phi_R} \left(\frac{dM}{P} - \frac{M}{P} \frac{d\overline{w}}{\overline{w}} \right),$$

where the coefficient on dY, now $\widetilde{H}$, is given by

$$\widetilde{H} = 1 - C_1 - I' \frac{q_L}{F_L} - \frac{C_2 + I' q_{R-\pi^e}}{\Phi_R} \left(\frac{F_{LL}}{F_L^2} \frac{M}{P} - \Phi_Y \right).$$

Again, under the assumption that $1 - C_1$ exceeds $I' q_L / F_L$, $\widetilde{H}$ is positive and the reduced form partial derivatives of Y with respect to the exogenous variables are:

$$\frac{\partial Y}{\partial T} = -\frac{C_1}{\widetilde{H}} \leq 0; \quad \frac{\partial Y}{\partial G} = \frac{1}{\widetilde{H}} \geq 0;$$

$$\frac{\partial Y}{\partial \pi^e} = -\frac{C_2 + I' q_{R-\pi^e}}{\widetilde{H}} \geq 0; \quad \frac{\partial Y}{\partial M} = \frac{C_2 + I' q_{R-\pi^e}}{\Phi_R P \widetilde{H}} \geq 0;$$

$$\frac{\partial Y}{\partial \overline{w}} = -\frac{(C_2 + I' q_{R-\pi^e}) M}{\Phi_R P \overline{w} \widetilde{H}} \leq 0.$$

Hence, this model produces the same qualitative results as the IS-LM model. The reader should also notice that the AD-AS Keynesian model does not dichotomize as the classical model does, as presented in Chapter 1.

2.6 Conclusion

In this chapter, we have summarized a great deal of traditional Keynesian macroeconomic theory. We have seen that if there are constraints on the flexibility of some prices, then the financial market and the money demand function play a crucial role in determining the effects not only of monetary policy, but also of fiscal policy. In fact, the relationship between the demand for money and the level of real income and the nominal rate of interest is of crucial importance in these Keynesian models.

In particular, with sticky prices knowledge of the various functions and of the values of their parameters is particularly useful in evaluating the effects of policy actions on the macroeconomy. In the case, for example, of the money demand function, if the interest elasticity of the demand for money balances is high, then fluctuations in the level of income are not likely to be caused by variations in the money supply. If it is low, then exactly the converse is true.

Of course, the theories of macroeconomic behavior that we have so far discussed in Chapters 1 and 2 are static in specification. As Bennett McCallum (1989, p. 77-78) puts it

> "one way in which these models are static is that they treat the economy's capital stock — its collection of productive machines, plants, highways, and so on — as *fixed* in quantity. As a result of that simplification, the models are not well designed for the analysis of policy actions or other events that would tend to induce substantial changes in the stock of capital within the relevant time frame."

Although this weakness of the classical and Keynesian models can be remedied, the current fashion is to explore short-run and long-run phenomena in the context of dynamic analyses. Models of this type have displaced the IS-LM and AD-AS frameworks in mainstream macroeconomic theory and dominate current research in almost all areas in economics. In the light of these developments, we now turn to these models.

Part 2:
Dynamic Monetary Macroeconomics

Overview of Part 2

We begin Chapter 3 by introducing the Cagan (1956) money demand model and the *adaptive* and *rational* expectation hypotheses. We introduce random shocks and present models designed to trace out the time paths of the endogenous variables. We proceed under the assumption that expectations are formed rationally and demonstrate how policy analysis is conducted in dynamic stochastic models. We also discuss some important and interesting research developments in the rational expectation macroeconomics literature.

In Chapter 4, we begin developing the framework for dynamic monetary macroeconomics, using the tools of neoclassical growth theory and related dynamical approaches. As Costas Azariadis (1993, p. xii) puts it "dynamical systems have spread so widely into macroeconomics that vector fields and phase diagrams are on the verge of displacing the familiar supply-demand schedules and Hicksian crosses of static macroeconomics."

In Chapter 5, we discuss monetary versions of neoclassical growth theory. Among monetary growth models, three that have seen wide and expanding use in the last two decades are the Tobin model, the Sidrauski model, and the overlapping generations model. Chapter 5 covers the Tobin and Sidrauski models in detail, leaving a discussion of the overlapping generations model for Chapter 9.

Chapter 6 provides a brief summary of the theoretical issues regarding the estimation of the welfare cost of inflation. We note that the welfare cost of inflation question is an outstanding one in macroeconomics and monetary economics.

3

Models with Rational Expectations

We begin this chapter by introducing the Cagan (1956) model and the *adaptive* and *rational* expectation hypotheses. Our analysis in this chapter is more complicated, because we introduce random shocks and use models designed to trace out the time paths of the endogenous variables on a period-by-period basis. Most of our discussion proceeds under the assumption that expectations are formed rationally, and in the context of dynamic, stochastic models, we discuss some important and interesting research developments.

In doing so, we also implicitly demonstrate how policy analysis is conducted in dynamic stochastic models. In particular, in the static models of Chapters 1 and 2 we used comparative static analysis and compared outcomes of changes in policy variables. In the context, however, of dynamic stochastic models, as McCallum (1989, p. 228) puts it,

> "the appropriate comparison pertains to average outcomes resulting, over a large number of periods, from different policy rules when maintained over these periods. Averages are represented analytically by unconditional expectations of relevant random variables."

Let us start with Cagan's (1956) famous study of hyperinflations — Philip Cagan was the 2006 Nobel laureate in economics.

3.1 The Cagan Model

Consider the following aggregate money demand function (to be discussed in more detail in Parts 3 and 4 of this book)

$$m_t - p_t = \alpha_0 + \alpha_1 \log y_t + \alpha_2 R_t + u_t, \tag{3.1}$$

where $m_t = \log M_t$, $p_t = \log P_t$, $R_t = r_t + \pi_t^e$, r_t is the real rate of interest, π_t^e is the expected inflation rate, and u_t is a white noise money demand innovation term.

Since during hyperinflations, movements in M and P are so large so as to dominate movements in real variables, we neglect movements in r_t and y_t and write (3.1) as

$$\begin{aligned} m_t - p_t &= (\alpha_0 + \alpha_1 \log y_t + \alpha_2 r_t) + \alpha_2 \pi_t^e + u_t \\ &= \gamma + \alpha \pi_t^e + u_t, \end{aligned} \tag{3.2}$$

where $\gamma = (\alpha_0 + \alpha_1 \log y_t + \alpha_2 r_t)$ is a composite constant term and $\alpha = \alpha_2 < 0$. Equation (3.2) is the central ingredient of the Cagan model. It involves only two variables, m_t and p_t, since π_t^e is taken as exogenous. Under the additional assumption that m_t is determined exogenously, (3.2) describes the behavior of p_t and can be used as a theory of price level determination during hyperinflations (i.e., extremely severe inflationary episodes).

Cagan wanted to econometrically estimate (3.2) and provide evidence that during hyperinflations the demand for real balances depends negatively on the expected inflation rate, π_t^e. He examined seven hyperinflations, but because he had no data on the expected inflation rate, π_t^e, he came up with the *adaptive expectations hypothesis*, to which we now turn.

3.2 Adaptive Expectations

In terms of our notation, the adaptive expectations model for the unobserved expected inflation rate at time t, π_t^e, can be expressed as

$$\pi_t^e - \pi_{t-1}^e = \theta \left(\Delta p_t - \pi_{t-1}^e \right),$$

where Δp_t $(= \log P_t - \log P_{t-1}) = \pi_t$, and $0 \leq \theta \leq 1$. The adaptive expectations model states that the change in the expected inflation rate from period $t-1$ to the current period t, $\pi_t^e - \pi_{t-1}^e$, is proportional to the forecast error — the discrepancy between the current actual and last period's anticipated inflation rate, $\pi_t - \pi_{t-1}^e$ — with the factor of proportionality being θ.

Clearly, the adaptive expectations model expresses the ability of economic agents to learn from their past mistakes, and this is why it is also known as the *error learning* hypothesis. In particular, if expectations realize (i.e., $\pi_t = \pi_{t-1}^e$), then there will be no revision in expectations ($\pi_t^e = \pi_{t-1}^e$). If, however, the inflation rate turns out to be surprisingly high (i.e., $\pi_t > \pi_{t-1}^e$), then there will be an upward revision in expectations ($\pi_t^e > \pi_{t-1}^e$), and if it turns out to be surprisingly low (i.e., $\pi_t < \pi_{t-1}^e$), then there will be a downward revision in expectations ($\pi_t^e < \pi_{t-1}^e$).

A simple rearrangement of the adaptive expectations model yields

$$\pi_t^e = \theta \Delta p_t + (1 - \theta)\, \pi_{t-1}^e. \tag{3.3}$$

This formulation states that the expected inflation rate at time t is a weighted average of the current actual inflation rate and last period's expected inflation rate, with the weights being the adjustment parameters θ and $1 - \theta$.

3.2.1 Application to the Cagan Model

Using (3.3), the Cagan model, equation (3.2), can be written as

$$m_t - p_t = \gamma + \alpha \left[\theta \Delta p_t + (1 - \theta) \pi_{t-1}^e \right] + u_t \tag{3.4}$$

To get rid of the expectational term, π_{t-1}^e, on the right-hand side of (3.4), we can write (3.2) for period $t-1$ to get an expression for π_{t-1}^e, which when substituted in (3.4) gives

$$m_t - p_t = \gamma\theta + \alpha\theta \Delta p_t + (1 - \theta)\left(m_{t-1} - p_{t-1}\right) + v_t$$

where $v_t = u_t - (1-\theta)u_{t-1}$. Clearly this last equation no longer includes terms involving the unobserved expected inflation rate variable. So, it can be estimated econometrically — see Cagan (1956) or McCallum (1989, Chapter 7) for a discussion and interpretation of Cagan's (1956) estimates for the seven hyperinflations that he studied.

3.3 Rational Expectations

The adaptive expectations hypothesis that we just examined has been faulted on the grounds that it doesn't assume enough rationality on the part of economic agents. In particular, according to the second presentation of the adaptive expectations hypothesis, economic agents use only current and last period's expected inflation rate when formulating expectations for the future. An alternative hypothesis for economic analysis of expectational behavior is John Muth's (1961) *rational expectations* hypothesis.

According to the rational expectations notion, economic agents use all of the available and economically usable information, including relevant economic theory, in the formation of expectations for the future. Formally, let π be the variable that is being forecast, π^e the rational expectation of π, and π^{of} the optimal forecast of π. Then according to the rational expectations hypothesis

$$\pi_t^e = \pi_t^{of} = E\left(\pi_t \,|I_{t-1}\right),$$

where I_{t-1} is the available information set. That is, the agents' subjective expectations are equal to the mathematical conditional expectations, meaning that expectations will not differ from optimal forecasts (i.e., the best guess possible) using all available information. Of course, there are two reasons why expectations may fail to be rational: (i) economic agents might be aware of all available information, but they are not making their expectation the best guess possible, and (ii) agents might be unware of some available relevant information, so that their expectation will not be the best guess possible.

The theory of rational expectations leads to the following two, common-sense implications for the way expectations are formed:

- if there is a change in the way a variable moves, the way expectations of this variable are formed will also change, and
- the expectational forecast error, $\varepsilon_t = \pi_t - E\left(\pi_t \,|I_{t-1}\right)$, will on average be zero and uncorrelated with available information. That is,

$$E\left(\varepsilon_t \,|I_{t-1}\right) = E\left[\left(\pi_t - E\left(\pi_t \,|I_{t-1}\right)\right) |I_{t-1}\right] = 0,$$

and

$$E\left(\varepsilon_t \times I_{t-1} \,|I_{t-1}\right) = 0.$$

If this were not the case it would be possible to improve the forecast by incorporating the available information.

3.3.1 Application to the Cagan Model

The Cagan model with rational expectations can be written as

$$m_t - p_t = \gamma + \alpha\pi_t^e + u_t$$

$$= \gamma + \alpha E_t(p_{t+1} - p_t) + u_t \tag{3.5}$$

where $\alpha < 0$ and $E_t p_{t+1} = E_t(p_{t+1} \,|I_t)$. Solving for p_t we obtain

$$p_t = \frac{m_t - \gamma - \alpha E_t p_{t+1} - u_t}{1 - \alpha}. \tag{3.6}$$

Expression (3.6), however, is not a solution for p_t, because of the expectational variable, $E_t p_{t+1}$, on the right-hand side.

To derive the rational expectation solution for this model's endogenous variable, p_t, we use the 'minimal set of state variables (MSV)' solution procedure — see McCallum (1989, Chapter 8) for more details. In doing so, we first complete the model (of price level determination), by assuming a monetary policy rule, determining the money supply, as follows,

$$m_t = \mu_0 + \mu_1 m_{t-1} + e_t, \tag{3.7}$$

where $|\mu_1| < 1$ and e_t is white noise. According to (3.7), the money supply at time t depends on its value last period and also on the random component, e_t. In (3.7), $\mu_0 + \mu_1 m_{t-1}$ represents the 'systematic' part of monetary policy and e_t the 'unsystematic' part.

Next, using (3.7) in (3.5) yields

$$\gamma + \alpha E_t p_{t+1} + (1 - \alpha)p_t + u_t = \mu_0 + \mu_1 m_{t-1} + e_t, \tag{3.8}$$

which shows that p_t will depend on m_{t-1}, u_t, e_t, and $E_t p_{t+1}$. Hence, we conjecture the following solution

$$p_t = \phi_0 + \phi_1 m_{t-1} + \phi_2 u_t + \phi_3 e_t, \tag{3.9}$$

where ϕ_0, ϕ_1, ϕ_2, and ϕ_3, are constants which we will have to express in terms of the parameters of the model, α, γ, μ_0, and μ_1.

Assuming that our conjecture is true, we can write (3.9) for period $t+1$ to get

$$p_{t+1} = \phi_0 + \phi_1 m_t + \phi_2 u_{t+1} + \phi_3 e_{t+1},$$

and apply the time t expectations operator, E_t, to obtain

$$\begin{aligned} E_t p_{t+1} &= \phi_0 + \phi_1 m_t \\ &= \phi_0 + \phi_1 \left(\mu_0 + \mu_1 m_{t-1} + e_t\right). \end{aligned} \tag{3.10}$$

Substituting (3.9) and (3.10) into (3.8) to eliminate p_t and $E_t p_{t+1}$ yields

$$\begin{aligned} &\Big[\gamma + \alpha\phi_0 + \alpha\phi_1\mu_0 + (1-\alpha)\phi_0\Big] \\ &+ \Big[\alpha\phi_1\mu_1 + (1-\alpha)\phi_1\Big] m_{t-1} \\ &+ \Big[(1-\alpha)\phi_2 + 1\Big] u_t \\ &+ \Big[\alpha\phi_1 + (1-\alpha)\phi_3\Big] e_t \\ &= \mu_0 + \mu_1 m_{t-1} + e_t. \end{aligned} \tag{3.11}$$

The equality in (3.11) can be used to solve for ϕ_0, ϕ_1, ϕ_2, and ϕ_3 as a function of the model's parameters, α, γ, μ_0, and μ_1. In fact, it implies the following restrictions on the parameters:

$$\gamma + \alpha\phi_0 + \alpha\phi_1\mu_0 + (1-\alpha)\phi_0 = \mu_0; \tag{3.12}$$

$$\alpha\phi_1\mu_1 + (1-\alpha)\phi_1 = \mu_1; \tag{3.13}$$

$$(1-\alpha)\phi_2 + 1 = 0; \tag{3.14}$$

$$\alpha\phi_1 + (1-\alpha)\phi_3 = 1. \tag{3.15}$$

These are the conditions that we need to express ϕ_0, ϕ_1, ϕ_2, and ϕ_3 as a function of the model's parameters. In fact, (3.13) and (3.14) imply, respectively,

$$\phi_1 = \frac{\mu_1}{1-\alpha+\alpha\mu_1};$$

$$\phi_2 = -\frac{1}{1-\alpha}.$$

If we substitute the solution for ϕ_1 into (3.15) and (3.12) we get, respectively,

$$\phi_3 = \frac{1}{1-\alpha+\alpha\mu_1};$$

$$\phi_0 = \frac{\mu_0(1-\alpha)}{1-\alpha+\alpha\mu_1} - \gamma.$$

Hence, the rational expectation solution for the price level is

$$p_t = \frac{\mu_0(1-\alpha)}{1-\alpha+\alpha\mu_1} - \gamma + \frac{\mu_1}{1-\alpha+\alpha\mu_1} m_{t-1} - \frac{1}{1-\alpha} u_t + \frac{1}{1-\alpha+\alpha\mu_1} e_t.$$

This equation defines the time path of p_t in terms of the exogenous shocks u_t and e_t and the predetermined variable, m_{t-1} — see McCallum (1989, Chapter 8) regarding the properties of this solution and other examples of rational expectation solutions.

As it happens, the concept of rational expectations has been embraced by the economics profession and the theory has been enhanced by important contributions by Lucas (1972, 1973, 1976), Sargent and Wallace (1975), Barro (1976), and Barro and Gordon (1983). In the rest of this chapter we review a number of results in the rational expectation macroeconomics literature.

3.4 The Lucas Critique

Robert Lucas in his famous (1976) paper, "Econometric Policy Evaluation: A Critique," presented an argument against the use of conventional econometric models as forecasting tools and for policy evaluation. The Lucas cririque is an important insight, and for it, Lucas was awarded the Nobel Prize in 1995.

To illustrate the Lucas critique of conventional policy evaluation, let's use Cagan's money demand function,

$$m_t - p_t = -\beta\left(E_t p_{t+1} - p_t\right) + u_t, \tag{3.16}$$

and assume that u_t is a serially correlated process,

$$u_t = \rho u_{t-1} + \varepsilon_t,$$

with $|\rho| < 1$ and ε_t being white noise. Assume that in the past the money supply was fixed, say $m_t = 0$ (or, equivalently, $M_t = 1$), and that, under this fixed money supply policy, prices were thought to be too volatile. The central bank asks the econometric policy advisor to advice on how m_t can be used in the indefinite future in order to minimize the fluctuations in p_t.

The behavior of p_t during the past is given by the rational expectation solution for p_t. To find the rational expectation solution for p_t, we need to solve the following model for p_t,

$$m_t - p_t = -\beta\left(E_t p_{t+1} - p_t\right) + u_t;$$

$$u_t = \rho u_{t-1} + \varepsilon_t;$$

$$m_t = 0.$$

which can be written as

$$-p_t = -\beta\left(E_t p_{t+1} - p_t\right) + \rho u_{t-1} + \varepsilon_t. \tag{3.17}$$

Using the method of conjectured solutions, we conjecture the solution

$$p_t = \phi_0 + \phi_1 u_{t-1} + \phi_2 \varepsilon_t, \tag{3.18}$$

which implies (after we apply the period t expectations operator, E_t)

$$E_t p_{t+1} = \phi_0 + \phi_1 u_t. \tag{3.19}$$

Substituting (3.18) and (3.19) into (3.17) yields the equality

$$\phi_0 + \phi_1 u_{t-1} + \phi_2 \varepsilon_t = \frac{1}{1+\beta}\left[\beta\phi_0 + (\beta\phi_1 - 1)\rho u_{t-1} + (\beta\phi_1 - 1)\varepsilon_t\right].$$

Hence, the implied conditions on the ϕ's are:

$$\phi_0 = \frac{\beta\phi_0}{1+\beta};$$

$$\phi_1 = \frac{\left(\beta\phi_1 - 1\right)\rho}{1+\beta};$$

$$\phi_2 = \frac{\beta\phi_1 - 1}{1+\beta}.$$

Solving these conditions for ϕ_0, ϕ_1, and ϕ_2 yields,

$$\phi_0 = 0;$$

$$\phi_1 = -\frac{\rho}{1+\beta(1-\rho)};$$

$$\phi_2 = -\frac{1}{1+\beta(1-\rho)}.$$

Hence, the rational expectation soltion for p_t is

$$p_t = -\frac{\rho}{1+\beta(1-\rho)}u_{t-1} - \frac{1}{1+\beta(1-\rho)}\varepsilon_t.$$

In fact, since

$$p_{t-1} = -\frac{\rho}{1+\beta(1-\rho)}u_{t-2} - \frac{1}{1+\beta(1-\rho)}\varepsilon_{t-1}$$

$$= -\frac{1}{1+\beta(1-\rho)}\left(\rho u_{t-2} + \varepsilon_{t-1}\right)$$

$$= -\frac{1}{1+\beta(1-\rho)}u_{t-1},$$

we can express the rational expectation solution for p_t as

$$p_t = \rho p_{t-1} - \frac{\varepsilon_t}{1+\beta(1-\rho)}. \tag{3.20}$$

Conventional policy evaluation might proceed as follows. The econometrician would use time series data to estimate (3.20) and get an estimate of ρ, $\widehat{\rho}$, over the sample period. The estimated model would then serve as a model of expectations to find $E_t p_{t+1} = \widehat{\rho} p_t$ which would be substituted into (3.16) to give

$$m_t - p_t = -\beta\left(\widehat{\rho} p_t - p_t\right) + u_t.$$

The conventional econometrician's model of p_t would then be

$$p_t = \frac{m_t - u_t}{1+\beta(1-\widehat{\rho})}. \tag{3.21}$$

Next, considering a policy rule of the form $m_t = g u_{t-1}$ equation (3.21) implies

$$\mathrm{Var}\,(p_t) = \frac{\sigma_\varepsilon^2 \left(g^2 + 1 - 2g\widehat{\rho}\right)}{\left[1 + \beta(1 - \widehat{\rho})\right]^2 \left(1 - \widehat{\rho}^2\right)}. \tag{3.22}$$

Hence, the g that minimizes $\mathrm{Var}(p_t)$ is the one that solves the equation

$$\frac{\partial \mathrm{Var}\,(p_t)}{\partial g} = 0.$$

Taking the partial derivative of $\mathrm{Var}(p_t)$ with respect to g, setting it equal to zero, and rearranging yields the optimal value of g,

$$g = \widehat{\rho},$$

implying the minimum variance

$$\mathrm{Var}\,(p_t) = \frac{\sigma_\varepsilon^2}{\left[1 + \beta(1 - \widehat{\rho})\right]^2}.$$

But we know that (3.22) is incorrect when $g = \widehat{\rho}$, since (3.22) was derived under the assumption that $g = 0$ (that is, under the money supply rule $m_t = 0$). The error was to assume that $E_t p_{t+1} = \widehat{\rho} p_t$, regardless of the choice of policy. This is exactly the point of the Lucas critique.

In fact, the correct approach would be to solve the following model

$$m_t - p_t = -\beta\left(E_t p_{t+1} - p_t\right) + u_t;$$

$$u_t = \rho u_{t-1} + \varepsilon_t;$$

$$m_t = g u_{t-1},$$

in which case the rational expectation solution for p_t is not (3.20). To find the rational expectation solution for p_t, we write the above model as

$$g u_{t-1} - p_t = -\beta\left(E_t p_{t+1} - p_t\right) + \rho u_{t-1} + \varepsilon_t \tag{3.23}$$

and conjecture the solution

$$p_t = \phi_0 + \phi_1 u_{t-1} + \phi_2 \varepsilon_t, \tag{3.24}$$

which implies

$$E_t p_{t+1} = \phi_0 + \phi_1 u_t. \tag{3.25}$$

Substituting (3.24) and (3.25) into (3.23) yields

$$\phi_0+\phi_1 u_{t-1}+\phi_2\varepsilon_t = \frac{1}{1+\beta}\left[\beta\phi_0 + (\beta\phi_1\rho + g - \rho)\rho u_{t-1} + (\beta\phi_1 - 1)\varepsilon_t\right]$$

The implied conditions on the ϕ's are:

$$\phi_0 = \frac{\beta\phi_0}{1+\beta};$$

$$\phi_1 = \frac{\beta\phi_1\rho + g - \rho}{1+\beta};$$

$$\phi_2 = \frac{\beta\phi_1 - 1}{1+\beta},$$

which, when solved for ϕ_0, ϕ_1, and ϕ_2 yield:

$$\phi_0 = 0;$$

$$\phi_1 = \frac{g-\rho}{1+\beta(1-\rho)};$$

$$\phi_2 = \frac{-1-\beta(1-g)}{(1+\beta)\left[1+\beta(1-\rho)\right]}.$$

Hence, the rational expectation solution for p_t is

$$\begin{aligned} p_t &= \frac{g-\rho}{1+\beta(1-\rho)} u_{t-1} + \frac{-1-\beta(1-g)}{(1+\beta)\left[1+\beta(1-\rho)\right]}\varepsilon_t \\ &= \frac{g-\rho}{1+\beta(1-\rho)} u_{t-1} - \frac{1+\beta(1-g)}{(1+\beta)\left[1+\beta(1-\rho)\right]}\left(u_t - \rho u_{t-1}\right) \\ &= -\frac{1+\beta(1-g)}{(1+\beta)\left[1+\beta(1-\rho)\right]} u_t + \frac{g}{1+\beta} u_{t-1}. \end{aligned} \tag{3.26}$$

Note that p_t in (3.26) depends on the parameters of the policy rule. We can calculate $\text{Var}(p_t)$ and minimize it with respect to g.

The message of the Lucas critique is that the public's expectations about a policy will influence the response to that policy, and that we should avoid using equations that will tend to shift with policy changes. Although the critique has typically being destructive, implicit in the critique is a constructive way to improve on conventional techniques for policy evaluation, by constructing models in terms of 'structural parameters,' that is, parameters that are invariant with respect to policy intervention. Whether, however, a parameter is invariant or not is a matter of judgement. This even applies to the 'deep parameters' of aggregator functions (utility and production functions) that we will deal with in Chapters 4 and 5.

3.5 Rules versus Discretion

The topic of 'rules versus discretion' in the conduct of monetary policy has a long history in macroeconomics — see, for example, Simons (1936). Recently, however, Barro and Gordon (1983), building on work by Kydland and Prescott (1977), study monetary policymaking in a world with rational expectations, taking into consideration the incentives of the monetary authority and the political constraints it may face. As they show, rules are better, because they lead to a lower average inflation rate than discretionary policymaking does.

Let's follow Barro and Gordon (1983) and assume that the expectational Phillips curve captures how the economy works

$$u = u^* - \alpha(\pi - \pi^e), \quad \alpha > 0, \tag{3.27}$$

where u is the unemployment rate, π the inflation rate, u^* the natural rate of unemployment, π^e the expected inflation rate, and α measures the 'marginal benefit of surprise inflation.' Notice that $u < u^*$ when $\pi > \pi^e$ and $u > u^*$ when $\pi < \pi^e$. Also, notice that the slope of the expectational Phillips curve is

$$\frac{\Delta(\pi - \pi^e)}{\Delta u} = -\frac{1}{\alpha},$$

suggesting that an increase in α makes the curve flatter, increasing the policymaker's temptation to create surprise inflation, $\pi - \pi^e$, which now gives a larger reduction in the unemployment rate, u, per point of π.

To determine the monetary authority's policy choice, we need to specify the preferences of the monetary authority. We assume that the monetary authority has a single period loss function (assumed to reflect the preferences of both government and society) quadratic in the actual inflation rate, π, and in the deviation of u from u^*, as follows

$$L = \beta\pi^2 + (u - ku^*)^2, \quad \beta > 0, k < 1, \tag{3.28}$$

where β measures the 'social cost of inflation' and k the strength of the policymaker's incentive to create surprise inflation, $\pi - \pi^e$.

3.5.1 Rules

Rule-like policymaking calls for period-by-period implementation of a formula that has been selected by the monetary authority to be applicable each period and for a large number of periods. It is to be noted that rule-type policy could be either nonactivist (that is, it does not depend on the current state of the economy) as, for example, in the following two monetary policy rules

$$\mu_t = 0.02;$$

$$\mu_t = 0.02 + 0.0001t,$$

where μ_t is the monetary growth rate in period t (assumed to be the monetary authority's instrument) and t is a time trend, or activist (that is, it depends on the current state of the economy), as in the rule

$$\mu_t = 0.02 + 0.5\left(u_{t-1} - 0.05\right),$$

where u_{t-1} is the unemployment rate in period $t-1$. Clearly, under the activist policy rule, this period's setting of the monetary policy instrument, μ_t, depends on last period's unemployment rate, u_{t-1}. For example, when $u_{t-1} = 5\%$, $\mu_t = 2\%$, but $\mu_t > 2\%$ if $u_{t-1} > 5\%$ and $\mu_t < 2\%$ if $u_{t-1} < 5\%$.

Keeping in mind that rule-type policy could be either activist or nonactivist and that the issue of rules versus discretion is separate form the issue of activist versus nonactivist policy behavior, let's examine how the economy performs when monetary policy is conducted under rules. Under rules, the monetary authority makes commitments about future monetary growth and inflation. Assuming that the monetary authority can use its instruments to produce any desired inflation rate, π, consider first the case where the monetary authority is committed to a constant π — this is known as 'fixed (or constant-growth-rate) rule.'

With this commitment, under rational expectations, economic agents will neither overpredict nor underpredict the inflation rate, π. Hence, $\pi = \pi^e$, $\pi - \pi^e = 0$, and [according to equation (3.27)], $u = u^*$. In this case, since $u = u^*$, regardless of π, there is no reason in having any inflation at all, and thus the optimal inflation rate is zero. Hence

$$\pi_p = \pi^e = 0, \tag{3.29}$$

and the value of the loss function is

$$L_p = (1-k)^2 u^{*2},$$

where the subscript p denotes the 'precommitment,' rules-type equilibrium.

3.5.2 Discretion

Discretionary policymaking involves period-by-period reoptimization on the part of the monetary authority. In particular, under discretion, the monetary authority takes the public's inflationary expectations, π^e, as given and minimizes (3.28) subject to (3.27). Hence, by substituting (3.27) into (3.28) to eliminate u, taking the partial derivative with respect to the choice variable, π, and setting it equal to zero, yields

$$\frac{\partial L}{\partial \pi} = -2\alpha \left[u^* - \alpha(\pi - \pi^e) - ku^* \right] + 2\beta\pi = 0,$$

which, when solved for the optimal inflation rate gives

$$\pi = \frac{\alpha\left[(1-k)u^* + \alpha\pi^e\right]}{\alpha^2 + \beta}. \tag{3.30}$$

When expectations are fulfilled (i.e., when $\pi = \pi^e$), the inflation rate will be (put $\pi = \pi^e$ in the above to get)

$$\pi_d = \frac{\alpha}{\beta}(1-k)u^*, \tag{3.31}$$

and the value of the loss function will be

$$L_d = \left[1 + \alpha^2/\beta\right] L_p, \tag{3.32}$$

where d denotes the 'discretionary' equilibrium.

On the basis of these results, Barro and Gordon (1983) argue that rule-type policy is superior to discretionary policy, because it leads to a lower average inflation rate.

3.6 Time Inconsistency

We have shown that rule-type policymaking is superior because it leads to zero inflation while discretionary (period-by-period) policymaking leads to positive inflation, with no additional output in compensation. The rules-type equilibrium, however, is often referred to as the 'optimal, but time-inconsistent solution' — see Kydland and Prescott (1977). The term 'time-inconsistent' refers to the policymaker's incentives to deviate from the rule when economic agents expect it to be followed.

On the other hand, the discretionary equilibrium is referred to as the 'suboptimal, but time-consistent solution.'

In order to discuss the time-inconsistency problem, let's calculate the value of π and L in the fooling solution, in which the public expects zero inflation but the monetary authority instead acts opportunistically and produces surprise inflation in order to reduce u below u^*. With $\pi^e = 0$, equation (3.30) implies that the monetary authority's short-run optimal inflation rate, π, is

$$\pi_f = \frac{\alpha(1-k)u^*}{\alpha^2+\beta},$$

and the value of the loss function is

$$L_f = \frac{(1-k)^2 u^{*2}}{1+\alpha^2/\beta} = \frac{1}{1+\alpha^2/\beta} L_p,$$

where the subscript f represents 'fooling.'

Clearly,

$$L_f < L_p < L_d,$$

demonstrating the benefits of precommitment. Kydland and Prescott (1977) argued that although the precommitted solution is optimal, it is time-inconsistent, in the sense that it lacks credibilty because the monetary authority has an incentive (called 'temptation') to behave inconsistently and inflate more than the public expects to obtain L_f, which is lower than L_p.

In doing so, however, the monetary authority completely ignores the consequences for future expectations and ends up raising the inflation rate, producing the worst possible outcome — the discretionary solution. In other words, the 'reneging' outcome is feasible only in the short-run, because in a rational expectations world economic agents cannot be fooled forever. That is, the rule will not be credible, the reputation of the monetary authority will be damaged, and the discretionary outcome will be realized.

One way to deal with the time-inconsistency problem is to 'constrain' the monetary authority by the rule — as already mentioned, the exact form of the rule is less important than the need to establish a credible commitment to the rule. It is, for example, possible to design an activist policy rule (specifying how the monetary authority will adjust μ in the light of new information about the economy) or a nonactivist policy rule (such as a fixed μ). Another solution is to appoint a central banker with an excessive dislike for inflation; in this case $\beta \to \infty$ and $L_d \to L_p$.

3.7 Inflation Mitigation

Anticipated inflation reduces the welfare of money holders. Here, we follow Fischer and Summers (1989) and examine the welfare consequences of institutional changes (such as the payment of interest on money, the issuance of government indexed bonds, the introduction of mortgage contracts that keep real rather than nominal payments constant, etc.) that reduce the costs of inflation. Such measures are widespread in high inflation countries, but governments in moderate inflation countries, despite the experience of significant inflation, have been reluctant to promote such measures. In fact, in most countries, only social security payments are indexed, but nothing else. Why?

Some economists have argued that the general reluctance of governments in moderate inflation countries to promote reforms to reduce the costs of inflation is due to the 'transitional costs' of moving to an indexed system as well as the 'transactions costs' of operating in such a system. Policymakers, however, argue that inflation-cost mitigation is counterproductive, because it promotes the inflation whose harmful effects seeks to mitigate. To evaluate this argument, let's use the Barro-Gordon (1983) model of the previous section.

We have seen that the discretionary equilibrium is given by (3.31) and (3.32). Clearly, the π_d function is increasing in the marginal benefit of surprise inflation, α. Hence, policy measures that increase α, will increase the inflation rate, π, and reduce social welfare. Hence, wage indexation is good because it reduces α, making the expectational Phillips curve steeper, and increasing economic welfare by reducing the monetary authority's incentive to create surprise inflation.

Let's now think of β as representing the effects of changes in the extent of inflation mitigation on utility — in particular, β falls as inflation mitigation increases. The π_d and L_d functions in (3.31) and (3.32) are decreasing in β, implying that policy measures that reduce β end up increasing the total cost of inflation to society. Hence, inflation mitigation policies, although they reduce the costs associated with a given inflation rate, π, they make things worse by reducing the monetary authority's commitment to low inflation and by causing adjustments of inflationary expectations.

3.8 Conclusion

The traditional models that we discussed in Chapters 1 and 2 and the rational expectation models we discussed in this chapter have been criticized on the grounds that they are not based on sound microeconomic foundations. In particular, it has been argued that the behavioral relationships of a good macro model should be derived from the intertemporal optimization of economic agents.

In the next three chapters we turn to explicit optimization analysis of the choice problems of representative economic agents. In doing so, we review neoclassical growth theory and related dynamical approaches that have widely spread into both macroeconomics and monetary economics and are now routinely used for macroeconomic and monetary analysis.

4

Neoclassical Growth Theory

In Chapters 1-3 we considered the role of the money demand function in comparative static models. These models were the dominant macroeconomic paradigm up until thirty years ago. Recently, however, neoclassical growth theory and related dynamical approaches have widely spread into both macroeconomics and monetary economics and are now routinely used in exploring fiscal and monetary policy issues.

Among dynamic macroeconomic models, three that have seen wide and expanding use in the last twenty years are the neoclassical growth model of Robert Solow (1956), the *optimal growth model* originated by Frank Ramsey (1928), and further developed by David Cass (1965) and Tjalling Koopmans (1965), and the *overlapping generations model* of Peter Diamond (1965). In what follows, we briefly discuss non-monetary versions of neoclassical growth theory, leaving monetary versions of the theory for the next chapter. In doing so, we focus on discrete time systems, given that economic data are available in discrete form.

4.1 The Solow Model

With the publication of Solow's (1956) seminal article on growth theory, entitled "A Contribution to the Theory of Economic Growth," macroeconomics and monetary economics started developing a central theoretical core. The Solow model is the cornerstone of that core.

The model consists of two equations, a production function and a capital accumulation equation. The production function describes how private factor inputs of capital, K_t, and labor, L_t, combine to produce output, Y_t. It takes the form

$$Y_t = F(K_t, L_t),$$

and is assumed to exhibit constant returns to scale, so that

$$F(\psi K_t, \psi L_t) = \psi Y_t,$$

for $\psi > 0$. Choosing $\psi = 1/L_t$ for $L_t > 0$, we can write the production function as

$$Y_t = F(K_t, L_t) = L_t F\left(\frac{K_t}{L_t}, 1\right) = L_t F\left(k_t, 1\right) = L_t f(k_t),$$

or in per capita (that is, per worker) terms,

$$y_t = f(k_t), \tag{4.1}$$

where y_t is output per person, Y_t/L_t, and k_t is capital per person, K_t/L_t. Writing the production function as in equation (4.1), has the advantage of focusing attention on per capita output, y_t, which is a better measure of living standards than total output, Y_t. The production function is also assumed to satisfy the conditions

$$f' > 0,\ f'' < 0,\ f'(0) = \infty,\ \text{and } f'(\infty) = 0.$$

It is assumed that there are many firms in the economy, so that perfect competition prevails, and that the firms are price-takers. Each firm maximizes profits, Π_t, by solving the following problem

$$\max_{K_t, L_t} \ \Pi_t = L_t f(k_t) - w_t L_t - r_t K_t, \tag{4.2}$$

where w_t is the real wage rate, r_t is the real rental price of capital, and $L_t f(k_t) = Y_t$ — the latter obtained by rearranging (4.1). The first-order conditions for profit maximization are

$$f(k_t) - k_t f'(k_t) = w_t; \tag{4.3}$$

$$f'(k_t) = r_t. \tag{4.4}$$

Equation (4.3) states that firms will hire labor until the marginal product of labor, $f(k_t)-k_tf'(k_t)$, equals the real wage rate, w_t, and equation (4.4) states that firms will hire capital until the marginal product of capital, $f'(k_t)$, equals the real rental price of capital, r_t. In other words, under perfect competition and profit maximization, markets clear when the real return to each factor equals its marginal product.

Notice also that, with constant returns to scale, payments to capital and labor sum to equal national income — that is, $r_tK_t + w_tL_t = Y_t$. We can show this (in per capita terms) by combining the results from (4.3) and (4.4) as follows

$$r_tk_t + w_t = f'(k_t)k_t + f(k_t) - k_tf'(k_t)$$

$$= f(k_t),$$

which when multiplied by L_t becomes

$$r_tK_t + w_tL_t = Y_t.$$

The second key equation of the Solow model is the capital accumulation equation. Assuming that capital depreciates at the constant rate $\delta > 0$, the capital accumulation equation is given by

$$K_{t+1} = I_t + (1-\delta)K_t$$

$$= sY_t + (1-\delta)K_t,$$

where I_t is gross investment and s $(0 \leq s \leq 1)$ is the *saving rate* — the fraction of output that is saved and invested.

In what follows we let ν denote the population growth rate and τ the rate of technical change and consider three versions of the model:

- with no population growth ($\nu = 0$) and no technical change ($\tau = 0$)
- with population growth ($\nu \neq 0$) but no technical change ($\tau = 0$), and
- with both population growth ($\nu \neq 0$) and technical change ($\tau \neq 0$)

4.1.1 Steady State ($\nu = \tau = 0$)

Under the assumption that the rate of population growth is zero ($\nu = 0$), $L_{t+1} = L_t$, and we have

$$k_{t+1} = \frac{K_{t+1}}{L_{t+1}} = \frac{sY_t + (1-\delta)K_t}{L_t},$$

which implies the following first-order difference equation

$$k_{t+1} = sf(k_t) + (1 - \delta)k_t. \tag{4.5}$$

Equation (4.5) is the Solow model in discrete time. It says that the amount of capital per worker depends positively on the saving rate, s, and negatively on the depreciation rate, δ.

Using the Solow model, equation (4.5), we can consider *dynamic equilibria,* in which every variable grows at some constant rate. Such equilibria are known as *steady states.* Under our present assumptions (with no population growth and technical change), *steady-state* requires that capital per person is constant over time — that is, $k_{t+1} = k_t$. In this case, (4.5) reduces to the standard steady-state equation

$$sf(k^*) = \delta k^*, \tag{4.6}$$

suggesting that δk^* is the amount of saving (and therefore, investment) per capita, necessary to keep constant the level of capital per worker.

We can also discuss what amount of capital accumulation is optimal. In order to do so, we assume a policymaker whose objective is the same as that of the representative economic agent, and in particular, to maximize steady-state per capita consumption. Such an objective is called the *golden rule* and the equilibrium that maximizes per capita consumption is called the *golden rule level of capital accumulation.*

To find the golden rule level of capital accumulation, we maximize steady-state per capita consumption, c^*,

$$c^* = f(k^*) - sf(k^*)$$

$$= f(k^*) - \delta k^*,$$

with respect to steady-state per capita capital, k^*. The first-order condition implies

$$f'(k^*) - \delta = 0. \tag{4.7}$$

Condition (4.10) is known as the *golden rule of accumulation* and states that (under present assumptions) steady-state per capita consumption is maximized when the marginal product of capital net of the depreciation rate equals zero (or, equivalently, when the marginal product of capital equals the depreciation rate).

4.1.2 Steady State Growth ($\nu \neq 0$ and $\tau = 0$)

Assuming that population grows at the rate ν, we have

$$k_{t+1} = \frac{K_{t+1}}{L_{t+1}} = \frac{sY_t + (1-\delta)K_t}{(1+\nu)L_t},$$

which implies the following first-order difference equation

$$(1+\nu)k_{t+1} = sf(k_t) + (1-\delta)k_t. \tag{4.8}$$

Equation (4.8) is the Solow model in discrete time. It says that the amount of capital per worker depends positively on the saving rate, s, and negatively on the depreciation rate, δ, and the population growth rate, ν.

Under our present assumptions (with constant population growth and no technical change), *steady-state growth* requires that capital per person is constant over time — that is, $k_{t+1} = k_t$. In this case, (4.8) reduces to the standard steady-state equation

$$sf(k^*) = (\nu + \delta)k^*, \tag{4.9}$$

suggesting that $(\nu + \delta)k^*$ is the amount of saving (and therefore, investment) per capita, necessary to keep constant the level of capital per worker. It is to be noted that in a steady state growth position of the economy, L grows at the rate ν, each of K and Y also grows at the same rate ν, so that $k = K/L$ and $y = Y/L$ are constant. In other words, there is growth in the levels of the variables but not in per capita quantities.

To find the golden rule level of capital accumulation, we maximize steady-state per capita consumption, c^*,

$$c^* = f(k^*) - sf(k^*)$$

$$= f(k^*) - (\nu + \delta)k^*,$$

with respect to steady-state per capita capital, k^*. The first-order condition implies

$$f'(k^*) - \delta = \nu. \tag{4.10}$$

and states that (under present assumptions) steady-state per capita consumption is maximized when the marginal product of capital net of the depreciation rate equals the population growth rate.

4.1.3 Steady State per Capita Growth ($\nu \neq 0$ and $\tau \neq 0$)

The Solow model can easily be extended to incorporate exogenous technical change. In that case, the first-order difference equation becomes

$$\Delta\widehat{k} = sf(\widehat{k}) - (\delta + \nu + \tau)\widehat{k}$$

where $\widehat{k} = K/AL$, and A is a factor measuring the efficiency of labor and assumed to grow at the rate τ. In this case, the steady-state equation becomes

$$sf(\widehat{k}^*) = (\nu + \delta + \tau)\widehat{k}^*,$$

It is to be noted that in a steady state, L grows at the rate ν, AL grows at the rate $\nu + \tau$, each of K and Y also grows at the rate $\nu + \tau$, so that $\widehat{k} = K/AL$ is constant. However, $y = Y/L$ and $k = K/L$ each grow at the rate τ. In other words, there is growth in the levels of the variables (at the rate $\nu + \tau$) as well as growth in the per capita quantities (at the rate τ). This is why such a steady state is referred to as a steady state per capita growth position of the economy. Hence, according to the model, technological progress is the only source of rising living standards over time.

Finally, in this case the golden rule of accumulation (4.10) becomes

$$f'(\widehat{k}^*) - \delta = \nu + \tau.$$

Corresponding theoretical discussion can be found in Barro and Xavier Sala-i-Martin (2004), David Romer (2001), and Solow (1999, 2000).

4.2 The Optimal Growth Model

One limitation of the neoclassical growth model of Solow is its *ad hoc* assumption that the saving rate, s, is an exogenous parameter. Although this assumption has allowed us to ignore the consumption-saving decision in order to concentrate our attention on the production side of the economy, as Solow (1999, p. 646) puts it,

> "the current fashion is to derive the consumption-investment decision from the decentralized behavior of intertemporal-utility-maximizing households and perfectly competitive profit-maximizing firms."

The Solow model has been improved upon. Today, there are two different versions of neoclassical growth theory, both explicitly based on maximizing behavior by economic agents — the optimal growth model of Ramsey (1928) and the overlapping generations model of Diamond (1965). In this section we discuss the Ramsey model, leaving the Diamond model for the next section.

Consider an economy populated by a large number of infinite-lived households each of which has preferences (at an arbitrary time, denoted $t = 0$) given by

$$\mathcal{U}(c_0, c_1, c_2, \cdots) = \sum_{t=0}^{\infty} \beta^t u(c_t), \tag{4.11}$$

or, written out in full,

$$\mathcal{U} = u(c_0) + \beta u(c_1) + \beta^2 u(c_2) + \cdots,$$

where c_t is per capita consumption at time t. The discount factor β equals $1/(1+\rho)$, where ρ $(0 < \rho < \infty)$ is a *time preference* parameter. Notice that a positive ρ implies that $\beta < 1$ and therefore a positive time preference — i.e., a preference for current over future consumption. The within-period utility function, $u(c_t)$, satisfies the following conditions

$$u'(c_t) > 0,\ u''(c_t) < 0,\ u'(0) = \infty,\ \text{and } u'(\infty) = 0.$$

We also assume that the household supplies inelastically one unit of labor each period — in other words, leisure is not valued.

The household operates a production function with constant returns to scale in capital and labor, given by (4.1). Assuming that population grows at the rate ν and that there is no technical change $(\tau = 0)$, the national income accounts identity can be written as,

$$f(k_t) = c_t + i_t,$$

and the capital accumulation equation as,

$$K_{t+1} = I_t + (1-\delta)K_t.$$

Dividing the capital accumulation equation by L_t (to express the equation in per capita terms) and rearranging gives

$$(1+\nu)k_{t+1} = i_t + (1-\delta)k_t,$$

where the population growth rate, ν, is defined by $(1+\nu) = L_{t+1}/L_t$. Finally, solving the last expression for i_t and substituting into the national income accounts identity to eliminate i_t, yields the household's budget constraint for period t,

$$f(k_t) = c_t + (1+\nu)k_{t+1} - (1-\delta)k_t. \tag{4.12}$$

As of time 0, the household chooses c_t and k_{t+1} (for $t = 0, 1, 2, ...$) to maximize (4.11) subject to (4.12), taking the initial stock of capital, k_0, as given. Formally, the household's problem in period 0 is

$$\max_{\{c_t,\, k_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t u\left(c_t\right)$$

subject to

$$f(k_t) = c_t + (1+\nu)k_{t+1} - (1-\delta)k_t,$$

for $t = 0, 1, 2, \cdots$, with k_0 given.

4.2.1 The Method of Lagrange Multipliers

Discrete time optimization methods can be used to solve finite as well as infinite horizon problems in both deterministic and stochastic environments. One method that can be used to solve this problem is the method of Lagrange multipliers; another method is 'dynamic programming.'

Let's start by using the method of Lagrange multipliers and write the Lagrangian function as

$$\mathcal{L} = \sum_{t=0}^{\infty} \beta^t u\left(c_t\right) + \sum_{t=0}^{\infty} \beta^t \lambda_t \left[f(k_t) - c_t - (1+\nu)k_{t+1} + (1-\delta)k_t \right].$$

As c_t is subject to the control of the economic agent it is called a *control* variable. k_{t+1} is called a *state* variable. λ is the Lagrange multiplier associated with the household's period t budget constraint.

The first-order conditions necessary for optimality can be obtained by differentiating $\mathcal{L}$ with respect to c_t and k_{t+1}. They are (for all t)

$$u'(c_t) = \lambda_t; \tag{4.13}$$

$$-(1+\nu)\lambda_t + \beta\lambda_{t+1}\left[f'(k_{t+1}) + 1 - \delta\right] = 0. \tag{4.14}$$

Conditions (4.12)-(4.14) are necessary for a maximum. In addition, there is a transversality condition,

$$\lim_{t\to\infty} k_{t+1}\beta^t u'(c_t) = 0, \tag{4.15}$$

stating that the present value of the stock of capital, k_{t+1}, in marginal utility units, must approach zero as $t \to \infty$. Notice that (4.15) does not require that $k_{t+1} \to 0$, since $\beta^t \to 0$ as $t \to \infty$.

In this setting, (4.12)-(4.14) are necessary for a maximum, while (4.12)-(4.15) are jointly sufficient. In other words, if (4.15) is satisfied, the household's choices of c_t and k_{t+1} will be described by (4.12)-(4.14).

Conditions (4.13) and (4.14) can be rearranged to give

$$(1+\nu)\frac{u'(c_t)}{\beta u'(c_{t+1})} = f'(k_{t+1}) + 1 - \delta. \tag{4.16}$$

This condition is known as the *Euler equation* (or *Keynes-Ramsey rule)* and relates the time path of consumption to the marginal product of capital and the rate of time preference. In the special case, for example, when

$$u(c_t) = \frac{c_t^\theta - 1}{\theta}, \quad \text{for} \quad \theta < 1 \quad \text{and} \quad \theta \neq 0,$$

and the marginal utility of consumption is $u'(c_t) = c_t^{\theta-1}$, the Euler equation (4.16) reduces to the following difference equation, describing a necessary condition that needs to be satisfied along an optimal path,

$$\frac{c_{t+1}}{c_t} = \left[\frac{f'(k_{t+1}) + 1 - \delta}{1+\rho}\right]^{1/(1-\theta)}.$$

4.2.2 The Method of Dynamic Programming

As already noted, another method of solving the household's maximization problem is dynamic programming. In the dynamic programming approach, we convert the infinite period problem into a two period problem as follows

$$\max_c \left\{ u(c) + \beta v(k') \right\} \tag{4.17}$$

subject to

$$(1+\nu)k' = f(k) - c + (1-\delta)k \tag{4.18}$$

with k given, where $u(c)$ is utility in the current period, k' denotes next period's value of k (that is $k' = k_{t+1}$), and $v(k')$ is the optimal value of the infinite period problem from period $t+1$ onwards (that is, starting with k_{t+1}).

Substituting (4.18) into (4.17) we get the Bellman equation (often called 'functional equation')

$$v(k) = \max_c \left\{ u(c) + \beta v\left[\left(f(k) - c + (1-\delta)k\right)/(1+\nu)\right] \right\},$$

for all k. For this problem the state variable is k and is given at the start of any period. The state completely summarizes all information from the past that is needed for the forward looking optimization problem. The control variable is the variable that is being chosen. In this case, it is the level of current consumption, c. The dependence of the state tomorrow on the state today and the control today is geven by (4.18).

In this two period problem, instead of choosing a sequence of consumption and capital levels, $\{c_t, k_{t+1}\}_{t=0}^{\infty}$, the agent just chooses current consumption, c, since all future controls and $v(k')$ have already been obtained.

A different version of the functional equation can be obtained by specifying the problem so that instead of choosing today's consumption, c, we choose tomorrow's state, k', as follows

$$v(k) = \max_{k'} \left\{ u\left(f(k) - (1+\nu)k' + (1-\delta)k \right) + \beta v(k') \right\}, \qquad (4.19)$$

for all k. Either specification yields the same result. We will choose this latter approach because it makes the algebra easier. Note that the unknown in the Bellman equation is the value function itself, $v(k')$.

The first order condition for the maximum problem (4.19) is

$$-u'\left(f(k) - (1+\nu)k' + (1-\delta)k \right)(1+\nu) + \beta v'(k') = 0.$$

The Benveniste-Scheinkman formula (envelope condition), saying that the value function is differentiable, is

$$v'(k) = u'\left(f(k) - (1+\nu)k' - (1-\delta)k \right)\left[f'(k) + 1 - \delta \right], \qquad (4.20)$$

which, when combined with the first order condition, gives

$$u'\left(f(k) - (1+\nu)k' + (1-\delta)k \right)(1+\nu) = \beta u'\Big(f(k') - (1+\nu)k''$$

$$+(1-\delta)k' \Big)\left[f'(k') + 1 - \delta \right]$$

where k'' denotes the value of k two periods ahead. The above equation can written as

$$(1+\nu)\frac{u'(c_t)}{\beta u'(c_{t+1})} = f'(k_{t+1}) + 1 - \delta,$$

which is the Euler equation we derived earlier — that is, equation (4.16).

4.2.3 The Euler Equation

The Euler equation (4.16) states that the marginal rate of intertemporal substitution equals the gross marginal product of capital net of the depreciation rate. It shows that when $f'(k_{t+1})$ rises, $u'(c_t)$ must increase relative to $u'(c_{t+1})$, suggesting that c_t must fall relative to c_{t+1}. Alternatively, it shows that when $f'(k_{t+1})$ rises, period $t+1$ consumption becomes relatively less expensive, motivating utility-maximizing households to switch away from current consumption and toward future consumption. An increase in δ has the opposite effect, since it reduces $f'(k_{t+1}) + 1 - \delta$.

Finally, an increase in β (which results from a fall in ρ, that is, a decrease in impatience) reduces c_t relative to c_{t+1}. Thus the effect of a fall in ρ in the Ramsey model is the same as that of a rise in s in the Solow model with a capital stock below the golden-rule level. The only difference between the two models is that in the Solow model s is constant whereas in the Ramsey model the saving rate is not constant during the transition to the new steady state.

4.2.4 The Modified Golden Rule

Equation (4.16) also implies that the competitive equilibrium steady state is characterized by

$$f'(k^*) - \delta = \nu + \rho, \tag{4.21}$$

since in the steady state $c_t = c_{t+1}$ and therefore $u'(c_t) = u'(c_{t+1})$. Equation (4.21) is known as the *modified golden rule of accumulation* and states that the capital stock is reduced below the golden rule level by an amount that depends on the rate of time preference. This is a powerful result, suggesting that the rate of time preference, ρ, determines the marginal product of capital, $f'(k^*)$, and the production function determines the stock of capital consistent with that marginal product of capital.

4.3 The Overlapping Generations Model

The optimal growth model assumes that the economy is populated by identical, infinitely-lived households, each endowed with perfect foresight over the infinite future. This is restrictive. The overlapping generations model avoids some of the restrictiveness by assuming that individuals have finite lives, that they care only about their own consumption, and that they leave no bequests when they die.

The simplest version of the overlapping generations model assumes that individuals live for two periods, t and $t+1$, so that an individual born at time t is young at t and old at $t+1$. Hence, at any time there are two heterogeneous generations (or cohorts) alive — in particular, in period t the young generation overlaps with an older generation and in period $t+1$ with a subsequent younger generation. An individual born at time t consumes c_{1t} in period t and c_{2t+1} in period $t+1$ and derives utility

$$\mathcal{U} = u(c_{1t}) + \beta u(c_{2t+1}),$$

where β is the subjective discount factor and $u'(\cdot) > 0$, $u''(\cdot) < 0$, $u'(0) = \infty$, and $u'(\infty) = 0$. Notice that there are two subscripts on consumption. The first subscript gives the age of the consumer and the second subscript the date (because the economy itself goes on forever).

Individuals work only when they are young, supplying inelastically one unit of labor and earning the going real wage rate of w_t. They decide how much to spend on the single good for first-period consumption, c_{1t}, they save and invest $s_t = w_t - c_{1t}$ at the going interest rate r_{t+1}, and spend all of their wealth — that is, $(1 + r_{t+1})s_t$ — on second-period retirement consumption, leaving nothing behind. The individuals born at time t and working in period $t+1$ are L_t. Population grows at rate ν so that $L_t = (1+\nu)^t L_0$. In what follows, we will examine the optimization problems of individuals and firms and define the competitive equilibrium.

Consider an individual born at time t. Her (constrained) maximization problem is

$$\max_{\{c_{1t}, c_{2t+1}\}} \left\{ u(c_{1t}) + \beta u(c_{2t+1}) \right\}$$

subject to

$$c_{1t} + s_t = w_t$$

$$c_{2t+1} = (1 + r_{t+1})s_t.$$

By substituting the budget constraints into the objective function, the individual's optimization problem can be written (in unconstrained form) as

$$\max_{\{s_t\}} \left\{ u(w_t - s_t) + \beta u\left((1 + r_{t+1})s_t\right) \right\}.$$

The first-order condition for choice of s_t is

$$\frac{u'(c_{1t})}{\beta u'(c_{2t+1})} = (1 + r_{t+1}). \tag{4.22}$$

This is the Euler equation, saying that the marginal rate of intertemporal substitution equals the gross real rate of interest.

Alternatively, we can solve the individual's problem by obtaining the intertemporal budget constraint, by eliminating s_t from the two, one-period budget constraints to get

$$c_{2t+1} - (1 + r_{t+1})(w_t - c_{1t}) = 0.$$

Setting up the Lagrangian

$$\mathcal{L} = u(c_{1t}) + \beta u(c_{2t+1}) + \lambda \left[c_{2t+1} - (1 + r_{t+1})(w_t - c_{1t}) \right],$$

the first-order conditions for choice of c_{1t} and c_{2t+1} are

$$u'(c_{1t}) + \lambda(1 + r_{t+1}) = 0;$$

$$\beta u'(c_{2t+1}) + \lambda = 0,$$

which when combined give the Euler equation (4.22).

Equation (4.22) implies that the quantity saved can be expressed as a function of the wage rate and interest rate,

$$s_t = s(w_t, r_{t+1}).$$

We assume that s is a differentiable function with $0 < s_w < 1$. However, s_r may be positive, negative, or zero, because of the income and intertemporal substitution effects. For example, an increase in the interest rate reduces the price of second period consumption, leading individuals to substitute second- for first-period consumption — this is the *intertemporal substitution effect*. But it also increases the feasible consumption set, making it possible to increase consumption in both periods — this is the *income effect*. The net effect of these substitution and income effects is ambiguous.[1]

[1] Consider, for example, the case where utility is log linear,

$$\mathcal{U} = \log c_{1t} + \beta \log c_{2t+1},$$

in which case the Euler equation (4.22) can be written as

$$\frac{c_{2t+1}}{\beta c_{1t}} = (1 + r_{t+1}).$$

By substituting the intertemporal budget constraint into the Euler equation to eliminate c_{2t+1} we get (after solving for c_{1t})

$$c_{1t} = \frac{w_t}{1+\beta} \quad \text{or} \quad s_t = w_t - c_{1t} = \frac{\beta w_t}{1+\beta}.$$

Turning now to firms, it is assumed that they act competitively using a constant returns to scale production function, (4.1). Each firm is assumed to maximize profits, taking the real wage rate, w_t, and the rental rate on capital, r_t, as given. The representative firm's maximization problem (assuming that $\delta = 0$) is given by (4.2), and profit maximization requires that conditions (4.3) and (4.4) are satisfied.

To derive the market equilibrium we need to find conditions for equilibrium in the goods and factor markets. Regarding equilibrium in the factor markets, the equilibrium conditions are those given by (4.3) and (4.4). Hence, equilibrium in the factor markets obtains when labor is hired to the point where the marginal product of labor equals the real wage rate and capital is rented to the point where the marginal product of capital equals the real rental rate.

Equilibrium in the goods market requires that the demand for goods equals the supply of goods or, equivalently, that investment equals saving

$$K_{t+1} - K_t = L_t s(w_t, r_{t+1}) - K_t. \tag{4.23}$$

In equation (4.23), $K_{t+1} - K_t$ is net investment, $L_t s(w_t, r_{t+1})$ is the saving of the young, and K_t is the dissaving of the old. This equation says that the capital stock increases only if the amount saved by the young, $L_t s(w_t, r_{t+1})$, exceeds the amount set aside last period by the current old, K_t, who withdraw their savings in this period.

Eliminating K_t from both sides of (4.23), we get

$$K_{t+1} = L_t s(w_t, r_{t+1}),$$

which says that the capital stock at time $t+1$ equals the saving of the young people at time t. Dividing both sides of the above by L_t gives the following *capital accumulation equation*

$$(1+\nu)k_{t+1} = s(w_t, r_{t+1}). \tag{4.24}$$

The capital accumulation equation (4.24), together with the factor market equilibrium conditions (4.3) and (4.4), yields the following relationship between k_{t+1} and k_t

$$(1+\nu)\, k_{t+1} = s\left(f(k_t) - k_t f'(k_t),\ f'(k_{t+1})\right). \tag{4.25}$$

Clearly, in this (log linear utility) case, saving does not depend on the interest rate, implying that the income and intertemporal substitution effects offset each other exactly.

We will refer to this equation as the *saving locus*. To study this equation is to study equilibria in the overlapping generations model.[2]

As Olivier Blanchard and Stanley Fischer (1989, p. 95) argue, the properties of the saving locus depend on the derivative

$$\frac{dk_{t+1}}{dk_t} = \frac{-s_w(k_t)k_t f''(k_t)}{(1+\nu) - s_r(k_{t+1})f''(k_{t+1})},$$

and the model does not, without further assumptions about utility and production, guarantee either existence or uniqueness of a steady state equilibrium with a positive capital stock.[3] One way to obtain definite results on the properties of the model is to specify explicit functional forms for the underlying utility and production functions.

As an example, consider the case where utility is log linear, $\mathcal{U} = \log c_{1t} + \beta \log c_{2t+1}$, and the production function is Cobb-Douglas, $y_t = k_t^\alpha$. In this case the capital accumulation equation is

$$k_{t+1} = \frac{(1-\alpha)\beta}{(1+\nu)(1+\beta)} k_t^\alpha,$$

suggesting that the steady state capital stock is

$$k^* = \left[\frac{(1-\alpha)\beta}{(1+\nu)(1+\beta)}\right]^{\dfrac{1}{1-\alpha}}.$$

[2] Let's see how equation (4.25) looks like with logarithmic utility, $\mathcal{U} = \log c_{1t} + \beta \log c_{2t+1}$, and a Cobb-Douglas functional form, $y = k^\alpha$, for the production function. We have

$$s_t = \frac{\beta w_t}{1+\beta}, \quad r_t = \alpha k_t^{\alpha-1}, \quad \text{and} \quad w_t = (1-\alpha)k_t^\alpha$$

so that the capital accumulation equation becomes

$$k_{t+1} = \frac{s_t}{(1+\nu)} = \frac{\beta w_t}{(1+\nu)(1+\beta)} = \frac{(1-\alpha)\beta}{(1+\nu)(1+\beta)} k_t^\alpha,$$

which is a difference equation for k_t. If we can solve this equation then we can have a complete solution, since we can read off w_t, r_t, and s_t (and hence consumption).

[3] If we are willing to assume that a unique equilibrium with positive capital stock exists, then stability requires that dk_{t+1}/dk_t is less than one in absolute value. That is, the stability condition is

$$\left|\frac{-s_w k^* f''(k^*)}{(1+\nu) - s_r f''(k^*)}\right| < 1.$$

In this case, $dk_{t+1}/dk_t > 0$. Stability requires that $|dk_{t+1}/dk_t| < 1$, and we can check if this is true for given values of the technology parameter α and the preference parameter, β. Notice that the properties of the economy, once it has converged to its balanced growth path, are the same as those in the optimal growth model — the saving rate is constant, per capita output is growing at the rate ν, the capital-output ratio is constant, and so on.

To see how the economy responds to shocks, consider a fall in ρ, when the economy is initially on its balanced growth path. The fall in ρ causes the young to save a greater fraction of their labor income, thereby increasing k^*. Thus the effects of a fall in the utility rate of time preference in the Diamond model (in the case we are considering, with logarithmic utility and Cobb-Douglas technology) are similar to the effects of a fall in ρ in the optimal growth model and to the effects of a rise in the saving rate in the Solow model.

The change shifts the paths over time of output and capital per worker permanently up, but it leads only to a temporary increase in the growth rates of these variables. The reader should also notice that in the Diamond model, as in the Ramsey model, the saving rate is not constant during the adjustment process.

4.4 Conclusion

We have reviewed basic, one-sector models of neoclassical growth theory and showed that these models, unlike the static IS-LM and AD-AS models that we discussed in Chapters 1-3, are dynamic structures built on solid microeconomic foundations. These models are also extremely versatile. They can be extended to deal with a number of issues in growth theory such as, for example, increasing returns to scale, human capital, endogenous population growth, and technological progress — see Robert Lucas (1988), Paul Romer (1986, 1990), Barro and Sala-i-Martin (2004), and Solow (1999, 2000) for references.

Neoclassical growth theory, however, also has uses in monetary economics and macroeconomics. As Costas Azariadis (1993, p. xii) puts it, neoclassical growth theory has

> "evolved into a *language* in which many macroeconomists, especially of the younger generation, choose to express their work and communicate their findings."

In fact, mainstream macroeconomic analysis amounts to 'complicating' one of the models discussed in this chapter. Introducing, for example, taxes and government debt we can study the effects of fiscal policy. Introducing money, we can explore the effects of monetary policy. Excellent treatments can be found in Blanchard and Fischer (1989), Athanasios Orphanides and Solow (1990), McCallum (1990), Azariadis (1993), and David Romer (2001).

In the next chapter, we review the ongoing debates about the role of money and the money demand function in neoclassical growth theory.

5

Monetary Growth Theory

The neoclassical growth models that we studied in Chapter 4 are models of a non-monetary economy. In this chapter, we review monetary versions of neoclassical growth theory. This involves putting money in the models of neoclassical growth theory and studying the implications for monetary policy. We begin with James Tobin (1965) who, as Orphanides and Solow (1990, p. 224) put it,

> "asked the question that has mainly preoccupied the literature ever since 1965. Different long-run rates of growth of the money supply will certainly be reflected eventually in different rates of inflation; but will there be any *real* effects in the long-run? Tobin studied this ("superneutrality") question in a simple "descriptive" model with aggregate saving depending only on current income, and seigniorage distributed in such a way as to preclude any distributional effects. He found that faster money growth is associated with higher capital stock and output per person in the steady state."

We also discuss an optimizing framework, originally due to Miguel Sidrauski (1967), that has played an important role in the development of monetary theory and has been used widely to study a variety of issues in monetary economics. In the Sidrauski, infinite-horizon optimization model superneutrality prevails.

We leave a discussion of a monetary version of the overlapping generations model for Chapter 9.

5.1 The Tobin Model

Tobin (1965) in his article, "Money and Economic Growth," developed one of the early dynamic models with money. He did so by introducing a *portfolio decision*, connecting money growth and capital accumulation, into the neoclassical growth model of Solow.

As with the Solow model, Tobin assumes a linearly homogeneous production function, given by $y_t = f(k_t)$, and that real per capita wealth in period t, a_t, is kept in the form of physical capital, k_t, and real money balances, m_t. That is,

$$a_t = k_t + m_t,$$

where m_t is real per capita money balances,

$$m_t = \frac{M_t}{P_t L_t}.$$

Above, M_t is nominal money balances, P_t is the price level, and L_t is population. The intuition of this *asset-allocation decision* is that physical capital and real money balances are substitutes in asset portfolios. In particular, for a given level of real wealth, a decrease in real balances would increase per capita capital and output, whereas an increase in real balances would have the opposite effect.

To introduce a means of changing the stock of money, Tobin assumes a government whose only role in the economy is to make lump-sum, real (per capita) transfers of money, in the amount of v_t. He then defines real aggregate disposable income in period t, Y_t^d, as real output, Y_t, plus the monetary transfers, $v_t L_t$, plus the change in the real value of money holdings arising from changes in the price level from period t to period $t+1$. That is,

$$Y_t^d = Y_t + v_t L_t + \frac{M_t}{P_{t+1}} - \frac{M_t}{P_t}.$$

Assuming that the inflation rate from period t to period $t+1$ is $\pi_t = (P_{t+1} - P_t)/P_t$, real aggregate disposable income can be written as

$$Y_t^d = Y_t + v_t L_t - \frac{\pi_t}{1+\pi_t}\frac{M_t}{P_t},$$

where $-\pi_t/(1+\pi_t)$ is the real rate of return on money.[1] Clearly, inflation (i.e., $\pi_t > 0$) produces a capital loss and therefore reduces real aggregate disposable income, whereas deflation (i.e., $\pi_t < 0$) produces a capital gain and increases it.

As with the Solow model, there is a *saving decision* in the Tobin model, similar to that in the Solow model. In particular, assuming a fixed saving rate, s, (physical and financial) asset accumulation equals the saving rate times real aggregate disposable income, Y_t^d. That is,

$$K_{t+1} - K_t + \frac{M_{t+1}}{P_{t+1}} - \frac{M_t}{P_t} = s\left(Y_t + v_t L_t - \frac{\pi_t}{1+\pi_t}\frac{M_t}{P_t}\right),$$

which implies that real aggregate investment (i.e., investment in physical capital) is given by

$$K_{t+1} - K_t = s\left(Y_t + v_t L_t - \frac{\pi_t}{1+\pi_t}\frac{M_t}{P_t}\right) - \left(\frac{M_{t+1}}{P_{t+1}} - \frac{M_t}{P_t}\right). \qquad (5.1)$$

If we divide both sides of equation (5.1) by L_t, we can write it in per capita terms as follows (the same expression can be obtained by dividing both sides of (5.1) by L_{t+1})[2]

$$(1+\nu_t)k_{t+1} = s\left(y_t + v_t - \frac{\pi_t}{1+\pi_t}m_t\right) - \frac{\mu_t - \pi_t}{1+\pi_t}m_t + k_t, \qquad (5.2)$$

where μ_t denotes the monetary growth rate, $\mu_t = (M_{t+1} - M_t)/M_t$, and ν_t the population growth rate (as before).

[1] To see this, notice that $1/P$ is the purchasing power of money (that is, the value of a unit of money in terms of goods that it buys). Hence, the real rate of return on money from period t to period $t+1$ is

$$\left(\frac{1}{P_{t+1}} - \frac{1}{P_t}\right) \Big/ \frac{1}{P_t} = \frac{P_t}{P_{t+1}} - 1 = -\frac{\pi_t}{1+\pi_t}.$$

[2] In deriving (5.2), we also made use of the fact that

$$\begin{aligned}\left(\frac{M_{t+1}}{P_{t+1}} - \frac{M_t}{P_t}\right)\frac{1}{L_{t+1}} &= \left(\frac{M_{t+1}P_t}{M_t P_{t+1}}\frac{L_t}{L_{t+1}} - \frac{L_t}{L_{t+1}}\right)\frac{M_t}{P_t L_t} \\ &= \left[\frac{1+\mu_t}{(1+\pi_t)(1+\nu_t)} - \frac{1}{1+\nu_t}\right] m_t \\ &= \frac{\mu_t - \pi_t}{(1+\pi_t)(1+\nu_t)} m_t.\end{aligned}$$

Since lump-sum transfers are equal to the real per capita value of the change in nominal balances, or

$$v_t = \frac{M_{t+1} - M_t}{P_{t+1}L_{t+1}} = \frac{\mu_t}{(1+\pi_t)(1+\nu_t)}m_t,$$

equation (5.2) reduces to (under the approximation that $\pi_t \nu_t \to 0$)

$$(1+\nu_t)k_{t+1} = sf(k_t) - \left(1 - \frac{s}{1+\nu_t}\right)\frac{\mu_t - \pi_t}{1+\pi_t}m_t + k_t. \tag{5.3}$$

Equation (5.3) is the Tobin model, clearly showing that anything that increases real per capita money balances will result in a lower level of per capita capital and output. The basic intuition can be stated as follows. People regard the transfer of money as income and therefore raise their total saving, but only by a fraction s of the increase in real money holdings. Thus, they are induced to consume more and hence save less for capital accumulation.

Using the Tobin model, and assuming that the population growth rate and the growth rate of nominal balances are constant, we now consider properties of steady states. Since m_t is real per capita money balances, $m_t = M_t/P_tL_t$, its evolution through time is given by

$$\frac{dm}{dt} = (\mu - \pi - \nu)m,$$

suggesting that in the steady state (where m is constant)

$$\pi = \mu - \nu. \tag{5.4}$$

Condition (5.4) states that steady-state inflation is directly determined by the monetary growth rate, μ.

In the steady state we also have $k_{t+1} = k_t$, so that equation (5.3) reduces to, after using equation (5.4),

$$sf(k) = \left[\left(1 - \frac{s}{1+\nu}\right)\frac{1}{1+\pi}m + k\right]\nu$$

$$= \left[\left(1 - \frac{s}{1+\nu}\right)\frac{1}{1+\pi}\frac{m}{k} + 1\right]\nu k. \tag{5.5}$$

Equation (5.5) is the steady-state equation in the Tobin monetary growth model. For $m = 0$, it reduces to the standard equation of the Solow non-monetary growth model, with $\delta = 0$.

It is immediately clear from (5.5) that monetary neutrality prevails in this model, since changes in the supply of nominal money balances produce proportional changes in the aggregate price level, leaving the real equilibrium unaffected. Notice that this property follows from the fact that $m = M/PL$ in (5.5), implying that proportional changes in M and P do not affect the equilibrium level of per capita capital and output.

To investigate the issue regarding the superneutrality of money, we close the model by assuming that the ratio of money holdings to capital holdings, m/k, is a function of the real rates of return yielded by the two assets. Since r is the real rate of return on capital and $-\pi/(1+\pi)$ the real rate of return on money, m/k depends negatively on r and positively on $-\pi/(1+\pi)$, or, equivalently, negatively on π. Algebraically, we have the following money demand function

$$\frac{m}{k} = \Phi(r, \pi), \tag{5.6}$$

with $\Phi_r < 0$ and $\Phi_\pi < 0$.

With profit maximization and perfect competition in factor and output markets, we know that $r = f'(k)$ and equation (5.6) becomes

$$\frac{m}{k} = \Phi\left(f'(k), \pi\right),$$

and the steady-state equation (5.5) in terms of s, π, and ν becomes

$$sf(k) = \left[\left(1 - \frac{s}{1+\nu}\right)\frac{1}{1+\pi}\Phi\left(f'(k), \pi\right) + 1\right]\nu k. \tag{5.7}$$

Totally differentiating (5.7) with respect to μ yields

$$\frac{dk}{d\mu} = \frac{\left(\Phi_\pi - \dfrac{\Phi(\cdot)}{1+\pi}\right)\vartheta\nu k}{sf'(k) - \nu\left[1 + \vartheta\Phi + \vartheta\Phi_r f''(k)k\right]}\frac{d\pi}{d\mu},$$

where $\vartheta = 1/(1+\pi) - s/(1+\nu)(1+\pi)$. Clearly, $dk/d\mu$ is positive, meaning that the Tobin model does not have the property of superneutrality, in the sense that changes in the monetary growth rate affect real variables. Intuitively, an increase in the growth rate of money raises the steady-state rate of inflation and lowers the real rate of return on money relative to physical capital. This reduces the ratio of money holdings to capital holdings and increases the steady-state level of per capita capital. This *portfolio substitution effect* of money growth on the equilibrium capital intensity of the economy is known as the *Tobin effect.*

5.2 The Sidrauski Model

The descriptive Tobin model, like the Solow model, assumes that the saving rate is an exogenous parameter. Sidrauski (1967), in his paper "Rational Choice and Patterns of Growth in a Monetary Economy," studied the superneutrality of money question in the context of a monetary growth model in an explicitly optimizing framework. In particular, he incorporated money balances into the utility function of the representative economic agent of the Ramsey (1928) optimal growth model, discussed in Chapter 4.

Sidrauski (1967) assumes that the representative household's lifetime utility function is of the form

$$\mathcal{U} = \sum_{t=0}^{\infty} \beta^t u(c_t, m_t), \tag{5.8}$$

where c_t and m_t are per capita consumption and real balances at time t. The within-period utility function, $u(c_t, m_t)$, satisfies the conditions, $u_i(c_t, m_t) > 0$, $u_{ii}(c_t, m_t) < 0$, for $i = 1, 2$, where $u_i(c_t, m_t)$ denotes the partial derivatives of $u(c_t, m_t)$ with respect to the ith argument.

Assuming (for simplicity) that there is no population growth (i.e., $\nu = 0$) and that the inflation rate from period t to period $t+1$ is $\pi_t = (P_{t+1} - P_t)/P_t$, the household's budget constraint (in per capita terms) can be written as

$$\begin{aligned} f(k_t) + v_t &= c_t + i_t \\ &= c_t + k_{t+1} - (1-\delta)k_t + (1+\pi_t)m_{t+1} - m_t, \end{aligned} \tag{5.9}$$

where m_t is real (time t) cash holdings and v_t denotes (lump-sum) real government transfers (net of taxes), received at the start of the period.[3]

The problem is solved by maximizing (5.8) subject to the constraint (5.9), taking k_0 and m_0 as given. The Lagrangian expression is

[3] The term $(1+\pi_t)m_{t+1} - m_t$ on the right-hand side of (5.9) gives the change in real money holdings from period t to period $t+1$. In particular, it is

$$\begin{aligned} \frac{M_{t+1} - M_t}{P_t} = \frac{M_{t+1}}{P_t} - \frac{M_t}{P_t} &= \frac{M_{t+1}}{P_{t+1}} \frac{P_{t+1}}{P_t} - m_t \\ &= (1+\pi_t)m_{t+1} - m_t. \end{aligned}$$

$$\mathcal{L} = \sum_{t=0}^{\infty} \beta^t u\left(c_t, m_t\right) + \sum_{t=0}^{\infty} \beta^t \lambda_t \Big[f(k_t) + v_t$$

$$-c_t - k_{t+1} + (1-\delta)k_t - (1+\pi_t)m_{t+1} + m_t \Big].$$

The necessary first-order conditions for optimality can be obtained by differentiating $\mathcal{L}$ with respect to c_t, m_{t+1}, and k_{t+1}. They are (for all t)

$$u_1(c_t, m_t) - \lambda_t = 0; \tag{5.10}$$

$$\beta u_2(c_t, m_t) - \lambda_t \left(1 + \pi_t\right) + \beta \lambda_{t+1} = 0; \tag{5.11}$$

$$-\lambda_t + \beta \lambda_{t+1} \left[f'(k_{t+1}) + 1 - \delta \right] = 0. \tag{5.12}$$

Conditions (5.9)-(5.12) are necessary for a maximum. In addition, there are two transversality conditions,

$$\lim_{t \to \infty} m_{t+1} \beta^t \lambda_t (1 + \pi_t) = 0; \tag{5.13}$$

$$\lim_{t \to \infty} k_{t+1} \beta^t \lambda_t = 0. \tag{5.14}$$

In this setting, (5.9)-(5.12) are necessary for a maximum, while (5.9)-(5.14) are jointly sufficient. In other words, if (5.13)-(5.14) are satisfied, the household's choices of c_t, m_{t+1}, and k_{t+1} will be described by (5.9)-(5.12).

We can now consider properties of steady states. Under present assumptions (with $\nu = 0$ and no technical change), c_t, k_t, m_t, v_t, and λ_t will be constant over time. With zero growth, conditions (5.10)-(5.12) reduce to

$$\beta u_2(c, m) = (1 + \pi - \beta) u_1(c, m); \tag{5.15}$$

$$\beta \left[f'(k) + 1 - \delta \right] = 1, \tag{5.16}$$

with (5.15) coming from (5.10) and (5.11) and (5.16) from (5.12). As equation (5.16) shows, in the steady state the marginal product of capital, $f'(k)$, is independent of π (and μ). This means that the real rate of interest (which equals the marginal product of capital) is independent of π (and μ) and that the Fisherian link between the nominal interest rate and the inflation rate holds across steady states. Moreover, because of the one-to-one mapping from $f'(k)$ to k, the capital intensity is also independent of π (and μ). Thus, superneutrality prevails and the Tobin effect is invalidated.

The superneutrality of money, however, is not a general result. Even minor modifications of Sidrauski's optimizing framework can lead to quite different results. In what follows, we consider a variation of the Sidrauski model, due to William Brock (1974), in which superneutrality fails.

5.3 A Variation of the Sidrauski Model

The Ramsey and Sidrauski models are both based on the assumption that labor is supplied inelastically. This is an unreasonable assumption and we can drop it, by introducing the amount of work as another decision variable. In doing so, we include labor input as an argument in the utility function of the representative economic agent, as follows

$$\mathcal{U} = \sum_{t=0}^{\infty} \beta^t u(c_t, m_t, n_t),$$

where c_t and m_t are (as before) per capita consumption and real balances at time t and n_t is the amount of work during period t. The within-period utility function, $u(c_t, m_t, n_t)$, satisfies the conditions, $u_i(c_t, m_t, n_t) > 0$, $u_{ii}(c_t, m_t, n_t) < 0$, for $i = 1, 2$, and $u_3(c_t, m_t, n_t) < 0$, $u_{33}(c_t, m_t, n_t) < 0$.

When labor supply is not inelastic, the production function can also be written as $y_t = f(k_t, n_t)$, with $f_i > 0$ and $f_{ii} < 0$, for $i = 1, 2$. The household's budget constraint (in per capita terms) can then be written as

$$\begin{aligned} f(k_t, n_t) + v_t &= c_t + i_t \\ &= c_t + k_{t+1} - (1-\delta)k_t + (1+\pi_t)m_{t+1} - m_t. \end{aligned}$$

Taking a dynamic programming approach, we write the problem as follows

$$\max_{c, m', k', n} \left\{ u(c, m, n) + \beta v(k', m') \right\} \tag{5.17}$$

subject to

$$f(k,n) + v = c + k' - (1-\delta)k + (1+\pi)\, m' - m, \tag{5.18}$$

where $v(k', m')$ is the value function; as in Chapter 4, a prime on a variable denotes next period's value of that variable.

By substitutiong (5.18) in (5.17) to eliminate c, we get the Bellman equation

$$v(k,m) = \max_{k',m',n} \Big\{ u \Big[f(k,n) + v - k' + (1-\delta)k - (1+\pi)\, m' + m, m, n \Big]$$

$$+\beta v(k', m') \Big\}$$

The first order conditions for the maximum problem are

$$u_1\,(c,m,n) = \beta v_1(k', m') \tag{5.19}$$

$$(1+\pi)\, u_1\,(c,m,n) = \beta v_2(k', m') \tag{5.20}$$

$$f_2(k,n) u_1(c,m,n) + u_3(c,m,n) = 0 \tag{5.21}$$

The Benveniste-Scheinkman formulas (envelope conditions) are

$$v_1(k,m) = u_1\,(c,m,n) \Big[f_1(k,n) + 1 - \delta \Big]$$

$$v_2(k,m) = u_1\,(c,m,n) + u_2(c,m,n)$$

which for period $t+1$ can be written as

$$v_1(k', m') = u_1\left(c', m', n'\right) \Big[f_1(k', n') + 1 - \delta \Big] \tag{5.22}$$

$$v_2(k', m') = u_1\left(c', m', n'\right) + u_2(c', m', n') \tag{5.23}$$

Equations (5.19) and (5.22) imply the Euler equation

$$\frac{u_1\,(c,m,n)}{\beta u_1\,(c', m', n')} = f_1(k', n') + 1 - \delta \tag{5.24}$$

and (5.20) and (5.23) yield

$$\frac{u_1(c,m,n)}{\beta\left[u_1(c',m',n')+u_2(c',m',n')\right]}=\frac{1}{1+\pi} \tag{5.25}$$

In the steady state, with $\nu=\tau=0$ (5.24), (5.25), and (5.21) reduce to

$$f_1(k,n)-\delta=\rho \tag{5.26}$$

$$\frac{u_1(c,m,n)}{u_2(c,m,n)}=\frac{\beta}{1+\pi-\beta} \tag{5.27}$$

$$u_3(c,m,n)=-f_2(k,n)u_1(c,m,n) \tag{5.28}$$

Clearly, equations (5.26)-(5.28) do not form a block recursive system of equations. In fact, as equation (5.26) shows, the steady-state marginal product of capital, $f_1(k,n)$, is no longer independent of π (and μ). Hence, this model does not have the property of superneutrality.

In general, if the marginal product of capital depends on other things besides per capita capital, then superneutrality will fail.

5.4 The New Empirics of Monetary Growth

Long-run monetary superneutrality asserts invariance of real variables with respect to inflation rates and monetary growth rates. This invariance derives from invariance in the marginal productivity of capital, to which the real rate of interest is equal in equilibrium. Over the years, these invariances have been investigated in a large number of studies. The evidence, however, has been in a state of flux. In his Nobel lecture, Robert Lucas (1996, p. 661) addresses this issue as follows:

> "the work for which I have received the Nobel Prize was part of an effort to understand how changes in the conduct of monetary policy can influence inflation, employment, and production. So much thought has been devoted to this question and so much evidence is available that one might reasonably assume that it has been solved long ago. But this is not the case: It had not been solved in the 1970s when I began my work on it, and even now this question has not been given anything like a satisfactory answer."

The 1990's, however, have been fruitful in this area of macroeconometrics, with new tests having been devised and executed. For example, Mark Fisher and John Seater (1993) and Robert King and Mark Watson (1997) contribute to the literature on testing long-run neutrality propositions by developing tests using recent advances in the theory of nonstationary regressors, to be briefly discussed later in this book. They show that meaningful long-run monetary neutrality tests can only be constructed if both nominal and real variables satisfy certain nonstationarity conditions and that much of the older literature violates these requirements, and hence has to be disregarded.

In particular, they show that neutrality tests are possible if real and nominal variables are integrated of order one [or I(1) in the terminology of Robert Engle and Clive Granger (1987)] and do not cointegrate; superneutrality tests are possible if the order of integration of the nominal variables is equal to one plus the order of integration of the real variables. Similarly, they show that the Fisherian link between inflation and nominal interest rates can be tested if the inflation and interest rate series are integrated of order one and do not cointegrate. Serletis and Zisimos Koustas (1998) and Koustas and Serletis (1999) provide international evidence on long-run monetary neutrality propositions, based on the King and Watson (1997) methodology.

Overall, recent empirical tests of long-run monetary neutrality propositions do not provide much evidence against the long-run neutrality of money. They provide, however, mixed evidence regarding long-run monetary superneutrality and the Fisherian link between nominal interest rates and inflation rates. See James Bullard (1999) for a review of the recent evidence regarding testing long-run monetary neutrality propositions.

5.5 Conclusion

We developed the descriptive Tobin model and the Sidrauski optimizing model and showed that in the former high inflation is associated with higher levels of per capita capital and output, whereas in the latter the superneutrality of money prevails. However, as we argued here and as Orphanides and Solow (1990) and McCallum (1990) discuss in more detail, these results are not robust to even modest modifications of each of these models. There is undoubtedly some intuition about the superneutrality question, but none of the models is comprehensive enough to provide a good understanding of the long-run effects of monetary growth.

One problem with these models is that they concentrate on the transactions role of money, ignoring the precautionary and speculative motives for holding money. In fact, although these models dominate current research in a large number of areas, they haven't been successfully expanded into the demand for money area in which static models are still the rule. In Chapters 7, 8, and 9 we review such static models of the demand for money.

6

The Welfare Cost of Inflation

The specification of the money demand function is also crucial in the estimation of the welfare cost of inflation. Whether inflation is costly is an important question, especially given the prevalence of inflation in the economic history of many countries around the world. In this chapter we provide a brief summary of the theoretical issues regarding the estimation of the welfare cost of inflation and note that the welfare cost of inflation question is an outstanding one in macroeconomics and monetray economics.

6.1 The Money Demand Function

Consider the following money demand function

$$\frac{M}{P} = \digamma(R, y),$$

where M denotes nominal money balances, P the price level, y real income, and R the nominal rate of interest, all at time t. Assuming that the $\digamma(R, y)$ function takes the form $\digamma(R, y) = \Phi(R)y$, the money demand function can be written as $m = \Phi(R)y$, where m denotes real money balances, M/P. Equivalently, we can write

$$z = \frac{m}{y} = \Phi(R),$$

which gives the demand for real money balances per unit of income as a function of the nominal interest rate R.

The specification of the money demand function is crucial in the estimation of the welfare cost of inflation. As you will see in the next section, Bailey (1956) and Friedman (1969) use a semi-log demand schedule whereas Lucas (2000) uses a double log (constant elasticity) schedule on the grounds that the double log performs better on the U.S. data that does not include regions of hyperinflation or rates of interest approaching zero.

6.2 The Consumer Surplus Approach

Bailey (1956) uses tools from public finance and applied microeconomics to measure the welfare cost of inflation. He argues that the welfare cost of inflation is the area under the inverse money demand schedule — the 'consumer surplus' that can be gained by reducing the nominal interest rate from a positive level of R to the lowest possible level (perhaps zero). In doing so, he implicitly assumes that individuals hold money (thereby sacrificing interest) because of the benefits from the transaction-facilitating services provided by money. These benefits are the reduced time and energy devoted to shopping and for any given change in the level of money holdings, the change in these benefits is represented by the area under the inverse money demand schedule between the initial and final levels of money holdings.

In particular, based on Bailey's consumer surplus approach, we estimate the function $z = \Phi(R)$, calculate its inverse $R = \Psi(z)$, and define the welfare cost function $w(R)$ by

$$w(R) = \int_{\Phi(R)}^{\Phi(0)} \Psi(x)dx = \int_0^R \Phi(x)dx - R\Phi(R). \tag{6.1}$$

Above, $w(R)$ is the welfare cost of inflation expressed as a fraction of income.

Clearly any measure of the welfare cost of inflation depends on the money demand function $\Phi(R)$ that is used. Bailey (1956) and Friedman

(1969) use a semi-log (Cagan-type) functional form for $\Phi(R)$ whereas Lucas (2000) uses a log-log functional form.

In particular, assuming a double-log money demand function

$$\ln \Phi(R) = \ln A + \eta \ln R, \tag{6.2}$$

or, equivalently,[1]

$$\Phi(R) = AR^{\eta}, \tag{6.3}$$

the welfare cost function (6.1) takes the form

$$w(R) = \int_0^R \Phi(x)dx - R\Phi(R)$$

$$= \left[\frac{A}{\eta+1}x^{\eta+1}\right]_0^R - RAR^{\eta} = -A\frac{\eta}{\eta+1}R^{\eta+1}.$$

To calculate $w(R)$, we use an estimate of η and calculate the value of A such that the curve obtained passes through the geometric means of the data.[2]

On the other hand, with a semi-log functional form for $\Phi(R)$,

$$\ln \Phi(R) = \alpha - \xi R, \tag{6.4}$$

or, equivalently,[3]

$$\Phi(R) = Be^{-\xi R}, \tag{6.5}$$

the welfare cost function (6.1) takes the form

[1] Equation (6.3) is obtained from equation (6.2) by writing (6.2) as $e^{\ln \Phi(R)} = e^{\ln A + \eta \ln R}$ which implies $\Phi(R) = e^{\ln A} e^{\ln R^{\eta}} = AR^{\eta}$.

[2] As an example, suppose that $\eta = -0.3$ and the geometric means of z and R are 1 and 0.05, respectively. Then

$$A = \frac{z}{R^{\eta}} = \frac{1}{0.05^{-0.3}} = 0.41.$$

In this case, the welfare cost of a steady state nominal interest rate of 10% relative to a steady state interest rate of 0% is

$$w(0.10) = -A\frac{\eta}{\eta+1}R^{\eta+1} = -(0.41)\frac{-0.3}{0.7}(0.10)^{0.7} = 0.035.$$

That is, the welfare cost of a 10% steady state interest rate is equal to 3.5% of people's income in the steady state.

[3] Equation (6.5) is obtained from equation (6.4) by writing (6.4) as $e^{\ln \Phi(R)} = e^{\alpha - \xi R}$, which implies $\Phi(R) = e^{\alpha} e^{-\xi R} = Be^{-\xi R}$, where $B = e^{\alpha}$.

$$w(R) = \int_0^R \Phi(x)dx - R\Phi(R)$$

$$= \left[\frac{B}{-\xi}e^{-\xi x}\right]_0^R - RBe^{-\xi R} = \frac{B}{\xi}\left[1-(1+\xi R)e^{-\xi R}\right]$$

The difference between the log-log and semi-log money demand functions is that the log-log schedule always produces greater welfare gains since the demand for real balances increases without limit as the nominal interest rate approaches zero. As Marty (1999) argues the log-log form works well in times of moderate inflation, but is not likely to work well in times of hyperinflation or in cases where policy is set according to Friedman's (1969) rule.

6.3 The Compensating Variation Approach

Lucas (2000) takes a 'compensating variation' approach in estimating the welfare cost of inflation. In particular, in the context of general equilibrium Ramsey type models, he calculates the reduction in consumption needed to compensate for (and measure) the gain in utility from a larger stock of money balances. In doing so, he derives an exact measure for the welfare cost of inflation using the Sidrauski (1967) framework and also investigates the robustness of the results to the non-existence of lump sum taxes and to the assumed transactions technology.

Let's consider the Sidrauski (1967) framework and assume an economy populated by a large number of infinite-lived agents each of which has preferences (at an arbitrary time, denoted $t = 0$) given by

$$\mathcal{U} = \sum_{t=0}^{\infty} \beta^t u(c_t, m_t), \tag{6.6}$$

where c_t and m_t are real consumption and money balances at time t. The discount factor β equals $1/(1+\rho)$, where ρ $(0 < \rho < \infty)$ is a time preference parameter. We assume a homothetic current period utility function (with consumption, c, and the ratio of real balances to consumption, m/c, as arguments) of the form

$$u(c,m) = \frac{1}{1-\sigma}\left[cf\left(\frac{m}{c}\right)\right]^{1-\sigma}, \tag{6.7}$$

with $\sigma \neq 1$.[4]

We also assume that each household is endowed with one unit of time which is inelastically supplied to the market and produces $y_t = (1+\gamma)^t y_0$ units of the consumption good in period t. Because the consumption good is nonstorable, one equilibrium condition is

$$c_t = y_t = (1+\gamma)^t y_0, \tag{6.8}$$

where γ is the real growth rate, assumed to be independent of monetary policy.

Households also face a cash flow constraint, which in nominal terms for period t is

$$P_t y_t = P_t c_t + M_{t+1} - M_t + H_t, \tag{6.9}$$

where H_t denotes lump sum taxes (or, if $H_t < 0$, lump sum transfers). Dividing both sides of (6.9) by P_t, we can write the household's cash flow constraint in real terms as

$$y_t = c_t + (1+\pi_{t+1})m_{t+1} - m_t + h_t, \tag{6.10}$$

where $(1+\pi_{t+1}) = P_{t+1}/P_t$ and $h_t = H_t/P_t$.

We assume a balanced growth equilibrium where the money growth rate $\mu_t = (M_{t+1} - M_t)/M_t$ is constant at μ, maintained by a constant ratio of transfers to income, h/y. In this case m/y will be constant and the condition

$$1 + \pi = \frac{1+\mu}{1+\gamma}$$

will also be satisfied.

Using dynamic programming, let $\tilde{v}(m, y)$ be the value of the maximized objective function (6.6) for a household in such an equilibrium that has real balances m when the economy-wide income level has reached y. Then the value function $\tilde{v}(m, y)$ satisfies the Bellman equation

[4] Homotheticity requires that the slope of the indifference curves (that is, the marginal rate of substitution) depend only on the m/c ratio. In the context of equation (6.7), we have

$$\frac{\frac{\partial u(c,m)}{\partial c}}{\frac{\partial u(c,m)}{\partial c}} = \frac{\left[cf\left(\frac{m}{c}\right)\right]^{-\sigma}\left[f\left(\frac{m}{c}\right) - f'\left(\frac{m}{c}\right)\frac{m}{c}\right]}{\left[cf\left(\frac{m}{c}\right)\right]^{-\sigma} f'\left(\frac{m}{c}\right)} = \frac{f\left(\frac{m}{c}\right) - f'\left(\frac{m}{c}\right)\frac{m}{c}}{f'\left(\frac{m}{c}\right)},$$

which depends on the ratio of real balances to consumption m/c.

$$\tilde{v}(m,y) = \max_{c} \left\{ \frac{1}{1-\sigma} \left[cf\left(\frac{m}{c}\right)\right]^{1-\sigma} + \beta\tilde{v}\left(m',(1+\gamma)y\right)\right\}, \quad (6.11)$$

where m' denotes next period's real balances, m_{t+1}, given by equation (6.10).

Under the homotheticity assumption the value function can be simplified to a function of a single state variable as follows. Define a value function $v(z)$ such that

$$\tilde{v}(m,y) = v(z)y^{1-\sigma},$$

where $z = m/y$, and let $\omega = c/y$ be the household's choice variable. Then the function $v(z)$ satisfies

$$v(z) = \max_{\omega} \left\{ \frac{1}{1-\sigma} \left[\omega f\left(\frac{z}{\omega}\right)\right]^{1-\sigma} + \beta(1+\gamma)^{1-\sigma} v(z')\right\}, \quad (6.12)$$

where z' is next period's value of the state variable z, z_{t+1}, defined as

$$z' = \frac{m'}{(1+\gamma)y} = \frac{y-c+m-h}{(1+\pi)(1+\gamma)y} = \frac{1-\omega+z-h/y}{1+\mu}.$$

The first-order and envelope conditions for the problem (6.12), evaluated along any equilibrium path where $c = y$ (and thus $\omega = 1$) are[5]

$$[f(z)]^{-\sigma}\left[f(z) - zf'(z)\right] = \frac{1}{1+R}v'(z'); \quad (6.13)$$

and

$$v'(z) = [f(z)]^{-\sigma} f'(z) + \frac{1}{1+R}v'(z'), \quad (6.14)$$

where the nominal interest rate R is defined by

$$\frac{1}{1+R} = \frac{\beta(1+\gamma)^{1-\sigma}}{1+\mu}. \quad (6.15)$$

Since along the balanced path z is constant, $v'(z) = v'(z')$. Eliminating $v'(z)$ and $v'(z')$ between (6.13) and (6.14) we get

$$R = \frac{f'(z)}{f(z) - zf'(z)}. \quad (6.16)$$

Let $\Phi(R)$ denote the z value that satisfies (6.16). It is this kind of steady state equilibrium relationship that Lucas (2000, p. 256) refers to as 'money demand function.'

[5] The first-order condition is obtained by evaluating $\partial v(z)/\partial w = 0$ at $w = 1$.

Next we define the welfare cost of a nominal interest rate R, $w(R)$, to be the income compensation needed to leave the household indifferent between living in a steady state with an interest rate constant at R and an otherwise identical steady state with an interest rate of zero. Thus, $w(R)$ is the solution to the following equality

$$\mathcal{U}\left[(1+w(R))\,y, \Phi(R)y\right] = \mathcal{U}\left[y, \Phi(0)y\right]. \tag{6.17}$$

With the assumed homothetic utility function (6.7), the equality in (6.17) reduces to[6]

$$(1+w(R))\,f\left(\frac{\Phi(R)}{1+w(R)}\right) = f\left(\Phi(0)\right), \tag{6.18}$$

and the welfare cost $w(R)$ can be calculated using an estimated money demand function $\Phi(R)$.

Suppose for example, that $\Phi(R)$ is given and substitute its inverse, $R = \Psi(z)$, into equation (6.16) to get the differential equation

$$f'(z) = \frac{\Psi(z)}{1+z\Psi(z)} f(z). \tag{6.19}$$

Differentiating (6.18) with respect to R, yields

$$w'(R)f\left(\frac{\Phi(R)}{1+w(R)}\right) + f'\left(\frac{\Phi(R)}{1+w(R)}\right)\left[\Phi'(R) - \frac{w'(R)\Phi(R)}{1+w(R)}\right] = 0. \tag{6.20}$$

Applying equation (6.19) with $z = \Phi(R)/\left(1+w(R)\right)$ to equation (6.20) and rearranging yields the differential equation

$$w'(R) = -\Psi\left(\frac{\Phi(R)}{1+w(R)}\right)\Phi'(R) \tag{6.21}$$

in the welfare cost function w.

6 Equation (6.18) is obtained by writing (6.17) as

$$\frac{1}{1-\sigma}\left[(1+w(R))\,yf\left(\frac{\Phi(R)y}{(1+w(R))\,y}\right)\right]^{1-\sigma} = \frac{1}{1-\sigma}\left[yf\left(\frac{\Phi(0)y}{y}\right)\right]^{1-\sigma},$$

and rearranging.

For any given money demand function, equation (6.21) can be solved numerically for an exact welfare cost function $w(R)$.[7] For example, with the log-log functional form (6.3) for $\Phi(R)$, equation (6.21) can be written as

$$w'(R) = -\eta A R^{\eta} \left(1 + w(R)\right)^{-1/\eta}, \qquad (6.22)$$

with solution

$$w(R) = \exp\left[-\frac{\eta \ln\left(-\frac{1}{A(R\exp(\eta\ln R)) - \frac{\eta}{A(\eta+1)} - \frac{1}{A(\eta+1)}}\right)}{\eta+1}\right] - 1. \qquad (6.23)$$

Thus the welfare cost of inflation is easily obtained using equation (6.23).

Lucas also investigates the robustness of his results to the nonexistence of lump sum taxes and inelastic labor supply, by introducing theoretical modifications to the Sidrauski model as well as a version of the McCallum and Goodfriend (1987) variation of the Sidrauski model to provide another general equilibrium rationale for Bailey's consumer surplus approach.

6.4 Empirical Evidence

Lucas (2000) provides estimates of the welfare cost of inflation in the United States, based on time series for 1900-1994. In doing so, he defines the money supply as simple-sum M1, assumes that money pays no interest, and estimates the welfare cost of inflation using Bailey's (1956) consumer surplus approach as well as the compensating variation approach. Lucas argues that money demand behavior at hyperinflation

[7] It is also possible to solve the differential equation (6.19) for the function f and to reconstruct the utility function. Consider, for example, the money demand function $z = AR^{-1/2}$ and its inverse $R = \Psi(z) = (A/z)^2$. By substituting in (6.19) yields

$$\frac{f'(z)}{f(z)} = \frac{1}{z} - \frac{1}{z + A^2}.$$

Solving this differential equation we get $\ln f(z) = \ln z - \ln\left(z + A^2\right) = \ln\left(z/\left(z + A^2\right)\right)$ or, equivalently, $f(z) = z/\left(z + A^2\right) = \left[\left(z + A^2\right)/z\right]^{-1} = \left(1 + A^2/z\right)^{-1}$, which when substituted in the Sidrauski utility function (6.7) gives

$$u(c, m) = \frac{1}{1-\sigma}\left[c\left(1 + \frac{A^2}{z}\right)^{-1}\right]^{1-\sigma} = \frac{1}{1-\sigma}\left[\frac{1}{c} + \frac{A^2}{m}\right]^{\sigma-1}.$$

or at rates of interest close to zero is crucial for welfare cost calculations; in those cases the semi-log money demand function, used by Cagan (1956) and Bailey (1956), fits the data better and should be used for such calculations. However, the U.S. time series data includes only moderate inflation rates, and Lucas' calculations, based on the double log demand schedule, indicate that reducing the interest rate from 3% to zero yields a benefit equivalent to an increase in real output of about 0.009 (or 0.9%; that is, nine tenths of one percent).

More recently, Serletis and Yavari (2004) calculate the welfare cost of inflation for Canada and the United States, in the post-World War II period, from 1948 to 2001. In doing so, they use the same double log money demand specification used by Lucas (2000), but pay particular attention to the integration and cointegration properties of the money demand variables and use recent advances in the field of applied econometrics (to be discussed in detail in Chapters 11-13) to estimate the interest elasticity of money demand. They conclude that the welfare cost of inflation is significantly lower than Lucas reported. In particular, for the United States, they find that reducing the interest rate from 3% to zero, would yield a benefit equivalent to 0.0018 (less than two tenths of one percent) of real income. This is much smaller than the 0.9% (nine tenths of one percent) figure obtained by Lucas under the assumption that the interest elasticity of money demand is -0.5. Similar welfare cost estimates are also reported by Serletis and Yavari (2005) for Italy, using the low frequency data from Muscatelli and Spinelli (2000) over the 1861 to 1996 period.

Finally, as Lucas (2000, p. 270) puts it in his conclusions, a direction for potentially productive research "is to replace M1 with an aggregate in which different monetary assets are given different weights." Serletis and Virk (2006) have taken up Lucas on this suggestion and provide a comparison among the official simple-sum aggregates, Barnett's (1980) Divisia aggregates, and Rotemberg's (1991) currency equivalent (CE) aggregates, at four different levels of monetary aggregation, to investigate the welfare implications of alternative monetary aggregation procedures — monetary aggregation issues will be discussed in detail in Chapters 15-17 of the book. In doing so, they assume that the different monetary aggregates face the same double log demand function, since their data does not include regions of hyperinflation or rates of interest approaching zero. However, following Serletis and Yavari (2004), they pay particular attention to the integration and cointegration properties of the money demand variables and use the Fisher and Seater (1993) long-horizon regression approach to obtain an estimate of the interest rate elasticity of money demand.

Their results indicate that the choice of monetary aggregation procedure is crucial in evaluating the welfare cost of inflation. In particular, the Divisia monetary aggregates, which (as you will see later in this book) embody differentials in opportunity costs and correctly measure the monetary services furnished by the non-currency components (valued by households), suggest a smaller welfare cost than the simple-sum and currency equivalent aggregates. This result is robust to whether they use the traditional approach developed by Bailey (1956) or the compensating variation approach used by Lucas (2000). Serletis and Virk (2006), however, have also made the bold assumption that money is non-interest bearing and used the 90-day T-bill rate to capture the opportunity cost of holding money. Investigating how much this matters, and also dealing with the issues raised in the last section of Marty (1999), is an area for productive research.

6.5 Conclusion

Of course the issue regarding the welfare cost of inflation is not closed. Recently, for example, Bullard and Russell (2004) use a quantitative-theoretic general equilibrium model of the U.S. economy and report that

> "a permanent, 10-percentage-point increase in the inflation rate — a standard experiment in this literature — imposes an annual welfare loss equivalent to 11.2 percent of output."

This is an estimate that is an order of magnitude larger than those estimates reported by Lucas (2000), Serletis and Yavari (2004, 2005), and Serletis and Virk (2006), suggesting that the welfare cost of inflation question is an outstanding one in macroeconomics and monetary economics.

As already noted, the choice of a money measure and the assumptions that we make about the interest elascticity of money demand are crucial in evaluating the welfare cost of inflation. In fact, the use of monetary aggregates (in various forms and at different levels of aggregation) is also subject to a comment by Prescott (1996, p.114) that (in the case of M1)

> "the theory has households holding non-interest bearing money, while the monetary aggregate used in the demand for money

> function is M1. Most of M1 is not non-interest bearing debt held by households. Only a third is currency and half of that is probably held abroad. Another third is demand deposits held by businesses, which often earn interest *de facto*. Households do not use these demand deposits to economize on shopping time. The final third is demand deposits held by households that, at least in recent years, can pay interest."

Dealing with these issues is an area for potentially productive future research. It should also be kept in mind that much of the welfare cost of inflation is borne by the poor, and thus depends on the income distribution, meaning that aggregate methods of the type discussed in this chapter might not be the most appropriate ones to use.

Part 3: Theoretical Approaches to the Demand for Money

Overview of Part 3

Chapters 7, 8, and 9 deal with conventional theoretical approaches to the demand for money. As in Laidler (1993), I discuss Fisher, Keynes, Friedman, Baumol and Tobin, McCallum, and Sargent and Wallace. Some of this theoretical literature on money demand, unlike the 'micro-foundations' approach to be discussed in Part 5 of the book, contains the result that the demand for money should be linear (or linear in the logs) and should have as arguments a small set of variables, themselves representing significant links to spending and economic activity in the other sectors of the economy.

7

The Classics, Keynes, and Friedman

In this chapter we survey the early theoretical literature on the macroeconomic demand for money. We begin with the classical version of the quantity theory of money, which remains considerably relevant even today. Then we move on to the Keynesian liquidity preference theory and we end with Milton Friedman's modern quantity theory.

A central question in this literature, crucial to how we view money's effects on aggregate economic activity, is whether and to what extent the demand for money is affected by changes in the interest rate. If the demand for money is insensitive to interest rates, the velocity of money is constant and the quantity of money is the primary determinant of nominal aggregate spending. If, however, the demand for money is affected by changes in interest rates, then velocity is not constant and money is not the primary determinant of aggregate spending.

For discussing these theories of the demand for money, the equation of exchange is a useful point of departure.

7.1 The Equation of Exchange

We begin with the *transactions version* of the *equation of exchange*, introduced by Irving Fisher in his 1911 book, *The Purchasing Power of Money*,

$$M^s V = PT,$$

where M^s is the actual stock of money, V its transactions velocity of circulation (or more simply velocity — the average number of times per period that the stock of money changes hands to finance transactions), P is the price level, and T is the volume of transactions. The equation of exchange states that the quantity of money multiplied by the average number of times that it changes hands per period in making transactions (which equals the number of purchases) must equal the number of transactions conducted over the period multiplied by the average price at which they take place (which equals the value of sales).

In the literature one finds a second presentation of the equation of exchange, known as the *income version* of the equation of exchange,

$$M^s V = PY, \tag{7.1}$$

where instead of the volume of transactions, T, real output, Y, appears in the equation and the income velocity (the rate of circulation of money relative to the rate of production of real income) replaces the transactions velocity. Underlying this substitution is the assumption that real income and the volume of transactions are proportionately related. In what follows, we adopt the convention of working with the income version of the equation of exchange.

7.2 The Quantity Theory of Money

Although equation (7.1) is nothing more than an identity, it can be used to develop a theory by postulating certain things about the determinants of the equation of exchange variables. In particular, assuming (as Fisher did) that real activity and money are exogenously determined, that velocity has a constant equilibrium long-run value, and that, within the monetary sector, the price level is the only endogenous variable, the equation of exchange (7.1) can be transformed into a version of the *quantity theory of money*, which can be written as

$$\overline{M}^s \overline{V} = P\overline{Y}, \tag{7.2}$$

with bars over M^s, V, and Y indicating that they are determined independently of the other variables. Equation (7.2) is the quantity theory of money, which states the conditions under which nominal income is determined solely by movements in the quantity of money. Alternatively, equation (7.2) can be viewed as a theory of price level determination, suggesting that the equilibrium price level is strictly proportional to the quantity of money.

7.3 The Quantity Theory Demand for Money

The quantity theory of money becomes a theory of the demand for money once one assumes that the money market is in equilibrium, so that $M^s = M^d = M$. In that case, equation (7.2) becomes (when solved for M^d)

$$M^d = kPY \quad \text{or} \quad \frac{M^d}{P} = kY, \tag{7.3}$$

where $k = 1/V$. Equation (7.3) is the long-run demand for money function, interpreted from the viewpoint of the quantity theory of money. It says that the demand for nominal (real) money is proportional to nominal (real) income.

A convenient linearization of equation (7.3) is achieved if we write it in logarithmic form as

$$\log M - \log P = \alpha + \log Y, \tag{7.4}$$

where $\alpha = \log k$. Equation (7.4) implies that for given values of real income, the demand for real money balances, $\log M - \log P$, is unaffected by exogenous changes in nominal money. In fact equation (7.4) implies that the price level elasticity of the demand for nominal money balances, $\eta(M, P)$, is[1]

---

[1] The reader should note that the above mentioned price level homogeneity condition can easily be tested by reformulating equation (7.4) as

$$\log P = -\alpha - \beta \log Y + \gamma \log M,$$

and testing the hypothesis that $\gamma = 1$. In fact, the above equation can also be written in differenced form as

$$\Delta \log P = -\alpha - \beta \Delta \log Y + \gamma \Delta \log M,$$

where Δ is the difference operator, and be used to test the steady-state hypothesis between the inflation rate, $\Delta \log P$, and the monetary growth rate, $\Delta \log M$, by testing that $\gamma = 1$.

$$\eta\left(M, P\right) = \frac{\mathrm{d}\log M}{\mathrm{d}\log P} = 1,$$

and that the real income elasticity of the demand for real money balances, $\eta(M/P, Y)$, is[2]

$$\eta\left(\frac{M}{P}, Y\right) = \frac{\mathrm{d}\log(M/P)}{\mathrm{d}\log Y} = 1.$$

Equation (7.3) also suggests that the demand for money is purely a function of income and that interest rates have no effect on the demand for money. In other words, the (nominal) interest rate elasticity of the demand for real money balances, $\eta(M/P, R)$, is

$$\eta\left(\frac{M}{P}, R\right) = \frac{\mathrm{d}\log(M/P)}{\mathrm{d}\log R} = 0.$$

7.4 The Cambridge Cash Balance Equation

A somewhat different approach within the quantity theory tradition was taken by the neoclassical economists in Cambridge University, England.[3] In contrast to the classical macroeconomic approach, the Cambridge economists took a microeconomic approach, by asking what determines the amount of money an economic agent would wish to hold. The emphasis was therefore on the choice-making behavior at the microeconomic level rather than on wants at a macroeconomic level.

The Cambridge economists treated money as a durable good yielding a flow of services (such as, according to Pigou for example, 'convenience' and 'security') and they also raised to a position of importance variables such as wealth and interest rates. They argued that total wealth puts an upper bound on money holdings and that money competes with other financial assets, many of which offer advantages relative to money. In this regard they argued that the division of total wealth into money and other assets individuals could hold, is optimal

[2] The implication of a unitary income elasticity can also be tested by reformulating equation (7.4) as

$$\log\left(\frac{M}{P}\right) = \alpha + \beta \log Y,$$

and testing the hypothesis that $\beta = 1$.

[3] By *neoclassical* economics we generally mean the work of Leon Walras (1834-1910), Alfred Marshall (1842-1924), and Arthur C. Pigou (1887-1959).

only if the marginal utility of money equals the marginal utility of an investment in an alternative asset.

The Cambridge economists, however, significantly simplified their formal demand for money relationship, by assuming that — in the short run at least — an economic agent would not alter the relationship between his level of wealth, the volume of transactions, and the level of income. They then argued that money demand will be a constant fraction, k, of income, as follows

$$M^d = kPY \quad \text{or} \quad \frac{M^d}{P} = kY. \tag{7.5}$$

Equation (7.5) is known as the *Cambridge cash balance equation.* It looks similar to equation (7.3), but rests on fundamentally different notions of the role of money in the economy, as we discussed.

The Cambridge economists assumed, in common with Fisher, that the level of real income is exogenous, suggesting that their demand for money is roughly proportional to the general level of prices. Notice that under the additional assumptions that the supply of money is exogenous and that money is willingly held (so that $M^s = M^d$), the Cambridge cash balance equation also implies the quantity theory prediction that nominal income is determined by the quantity of money.

However, unlike the quantity theorists (who assumed that velocity can change with changes in institutional factors, but not with changes in other variables of the economic system), the Cambridge economists allowed for the possibility of interest-rate effects on the demand for money in the short run. They argued that k could fluctuate in the short run with fluctuations in the yields and expected returns on other assets individuals could hold. This was a major departure from the quantity theorists' view, and led Keynes (a later Cambridge economist) to develop a theory of the demand for money that emphasized the importance of interest rates.

7.5 The Keynesian Approach

Although the Cambridge economists raised to a position of importance variables such as interest rates and wealth, they did not explicitly include these variables in their money demand function. It is, however, from this tradition of approaching the subject of money demand that their successor Keynes developed his analysis in his famous 1936 book, *The General Theory of Employment, Interest, and Money.*

Keynes studied both transaction and asset theories of money demand. He called his overall theory of the demand for money the *liquidity preference theory* and distinguished three motives for holding money — a 'transactions motive,' a 'precautionary motive,' and a 'speculative motive' — suggesting that people regard holding of money for one motive, at least in part, as separate from holdings of money for another motive.

In his discussions of the transactions demand for money, Keynes followed closely Fisher and the Cambridge economists and listed the transactions motive as an important (but not the only) motive underlying the demand for money. He postulated that the transactions (or business) demand for money is a stable function of the level of income. In fact, he wrote the transactions demand for money as in equation (7.3). Also, regarding the precautionary motive for holding money, Keynes suggested that the demand for precautionary money balances depends on the level of income and slightly on the interest rate, but for the most part, on the level of uncertainty about the future.

However, the most important innovation in Keynes's analysis of the demand for money is his *speculative demand* for money, or the demand for money as an asset alternative to other interest-yielding assets. The primary result of the Keynesian speculative theory is that the demand for money depends negatively on the interest rate. Keynes derived this result by analyzing only the choice between interest-yielding bonds and money as an issue of liquidity preference. In doing so, however, he raised to a position of importance variables such as interest rates, expectations, and uncertainty, which although considered by the Cambridge economists, were ultimately accorded a secondary role.

We may illustrate the Keynesian speculative theory of money demand by dividing the assets into two broad categories: money and bonds. Assume that the expected return on money is zero, as Keynes did (in his time, unlike today, this was a reasonable assumption, since money was mostly of the outside type). The expected rate of return on bonds is the sum of the current yield and the expected rate of capital gain (or loss).

If people expect interest rates to increase in the future (and therefore bond prices to decline), the expected rate of return on bonds would be less than the current yield, because the expected rate of capital gain is negative — that is, an expected capital loss. In fact, if people expect future interest rates to increase substantially, the expected rate of capital loss might outweigh the current yield, so that the expected rate of return on their bonds would be negative. In this case, they will put all of their liquid wealth into money. On the other hand, if people expect a substantial decline in interest rates (and therefore a significant increase in bond prices), the expected rate of return on bonds will exceed the current yield, because the expected rate of capital gain is positive. In this case, people will hold all bonds and no money for speculative purposes.

The implication of this is that the demand for speculative money balances depends on both the observable market (nominal) interest rate and people's expectation concerning that rate in the future. The decision with respect to holding bonds or money is described in Keynes in terms of some *normal value* that interest rates tend to. If interest rates are above this normal value, people will expect them to fall, bond prices to rise, and capital gains to be realized. As a result, people will be more likely to hold their liquid wealth as bonds rather than money, and the demand for money will be low.

If interest rates are below the normal value, people will expect them to rise, bond prices to fall, and capital losses to be realized. They will be more likely to hold money than bonds and the demand for money will be high. In fact, at some very low interest rate, everyone will expect it to rise and the demand for money in the aggregate will be perfectly elastic with respect to the interest rate — this is known as the *liquidity trap*. Overall, assuming a normal distribution on the population's expectations of the future interest rate, the aggregate demand for money will be negatively related to the level of interest rates.

We have discussed three separate demands for money — the transactions demand, the precautionary demand, and the speculative (or asset) demand for money. Combining these three demands, we get the Keynesian *liquidity preference function*, describing the total demand for money

$$\frac{M^d}{P} = \Phi(R, Y),$$

with $\Phi_1 < 0$ and $\Phi_2 > 0$, where Φ_i denotes the partial derivative of $\Phi(\cdot)$ with respect to its ith argument. That is, the demand for real money balances is negatively related to the nominal interest rate, R, and positively related to real income, Y.

One implication of the Keynesian liquidity preference theory of the demand for money, which contrasts sharply with the classical quantity theory approach, is that velocity is not constant but instead positively related to nominal interest rates. We can see this by writing down the velocity that is implied by the liquidity preference function

$$V = \frac{Y}{M/P} = \frac{Y}{\Phi(R, Y)}.$$

We know that when the interest rate increases the demand for money declines and therefore velocity rises. Hence, in contrast to the quantity theorists' view of a constant velocity, the Keynesian liquidity preference theory implies that velocity is procyclical, since procyclical interest rate movements induce procyclical velocity movements.

7.6 Friedman's Modern Quantity Theory

The Keynesian theory of liquidity preference draws a distinction between transactions, precautionary, and speculative demands for money. Friedman (1956), however, by assuming that money is *abstract purchasing power*, meaning that people hold it with the intention of using it for upcoming purchases of goods and services, integrated an asset theory and a transactions theory of the demand for money within the context of neoclassical microeconomic theory of consumer and producer behavior.

In particular, Friedman did not specify, as Keynes did, any particular motives for holding money. Rather, by taking for granted the fact that people hold money, he viewed money as a durable good (or monetary assets as durable goods) yielding a flow of nonobservable services (proportional to the stock), which enter as arguments in aggregator functions (i.e., utility and production functions). He also assumed that money competes with other assets (such as, for example, bonds, stocks, and physical goods) for a place in individuals' and business firms' portfolios and that the marginal utility of monetary services declines as the quantity of money held increases.

Friedman's theory of the demand for money can be expressed in terms of the following demand function for money for an individual wealth holder

$$\frac{M^d}{P} = \Phi(Y_p, R_b - R_m, R_e - R_m, \pi^e - R_m, \cdots), \tag{7.6}$$

where

$$\begin{aligned} Y_p &= \text{real permanent income} \\ R_b &= \text{expected nominal rate of return on bonds} \\ R_e &= \text{expected nominal rate of return on equities} \\ R_m &= \text{expected nominal rate of return on money, and} \\ \pi^e &= \text{expected inflation rate.} \end{aligned}$$

The dots in equation (7.6) stand for other variables (such as, for example, the ratio of human to nonhuman wealth) that are regarded as relevant but play no essential role in Friedman's theory and have no important implications for monetary policy. The expected nominal rate of return on bonds includes expected capital gains or losses, that on equities includes expected changes in their prices, and π^e is used as a proxy for the expected nominal rate of return on physical assets. It is also assumed the demand for real money balances is positively related to permanent income, Y_p, and negatively related to the yield on other assets.

The first point in an exposition of this theory must be an explanation of the concept of permanent income, which differs from actual measured income. In fact, an individual's permanent income is the hypothetical constant level of income that has the same discounted present value as the expected future income streams. More formally, (real) permanent income, Y_p, can be defined by the condition

$$Y_p + \frac{Y_p}{(1+r)} + \frac{Y_p}{(1+r)^2} + \cdots = \sum_{j=0}^{\infty} \frac{Y_{t+j}}{(1+r)^j}, \tag{7.7}$$

where variables dated after t are anticipated values, viewed from today (period t), and r is the real rate of interest. By algebraic manipulation, the above equation yields[4]

$$Y_p = \frac{r}{1+r} \sum_{j=0}^{\infty} \frac{Y_{t+j}}{(1+r)^j} \tag{7.8}$$

[4] To see this, rewrite equation (7.7) as

$$[1 + z + z^2 + \quad \cdots \quad]Y_p = \sum_{j=0}^{\infty} \frac{Y_{t+j}}{(1+r)^j}$$

in which $z = 1/(1+r)$. Since

$$1 + z + z^2 + \quad \cdots \quad = \sum_{j=0}^{\infty} z^j = \frac{1}{1-z}, \qquad \text{for } -1 < z < 1,$$

the above equation can be written as

$$\frac{1}{1-z} Y_p = \sum_{j=0}^{\infty} \frac{Y_{t+j}}{(1+r)^j}$$

which by substitution (to eliminate z) and suitable algebraic manipulation yields equation (7.8).

Friedman stressed two issues regarding his money demand function that distinguish it from Keynes's liquidity preference theory. First, Friedman did not take the expected rate of return on money to be a constant, as did Keynes, and by assuming that the demand for money depends on the incentives for holding other assets relative to money, he argued that the demand for money is insensitive to interest rates. In particular, in Friedman's view, when interest rates rise in the economy the expected rate of return on money held as bank deposits also rises (along with the rise in the expected rates of return on other assets), so that there is no change in the incentive terms, $R_b - R_m$, $R_e - R_m$, and $\pi^e - R_m$, in the money demand function.

Hence, unlike Keynes's liquidity preference theory in which interest rates are an important determinant of the demand for money, Friedman's theory suggests that although the demand for money is sensitive to changes in the incentives for holding other assets relative to money, these incentives stay relatively constant when interest rates change, implying that the demand for money is insensitive to interest rates. Therefore, Friedman's money demand function, equation (7.6), can be approximated by

$$\frac{M^d}{P} = \Phi(Y_p),$$

which indicates that real permanent income is the only determinant of real money demand.

The second issue Friedman stressed is the stability of the money demand function. In particular, unlike Keynes (who felt that the demand for money is erratic and shifts with changed expectations of the rate of interest), Friedman suggested that the money demand function is highly stable, implying that the quantity of money demanded can be predicted accurately by the money demand function. Also, when combined with his view that the demand for money is insensitive to interest rates, this means that the velocity of money is highly predictable.

The reader should notice that the stability of the money demand function and the consequent predictability of the velocity of money derive from the relationship between current income and permanent income. In particular, according to equation (7.8), a permanent change in income changes permanent income by the same amount, whereas a temporary change in income (as, for example, in a business cycle expansion or recession) changes permanent income by a small amount.[5]

[5] Mathematically, suppose that $r = 10\%$ and consider a temporary change in income, such as $\Delta Y_t = 1$ with $\Delta Y_{t+j} = 0$, $j = 1, 2, \cdots$. According to equation (7.8), we have

We can see this implication of Friedman's theory, by converting the money demand function to the corresponding velocity of money function

$$V = \frac{Y}{M/P} = \frac{Y}{\Phi(Y_p)},$$

which suggests that since the relationship between current income, Y, and permanent income, Y_p, is usually quite predictable, the velocity of money is predictable (although not constant) as well. This means that a given change in the nominal money supply will produce a predictable change in aggregate spending. Therefore, Friedman's theory of the demand for money is indeed a reformulation of the quantity theory of money, because it leads to the quantity theory conclusion that money is the primary determinant of aggregate nominal spending.

Finally, Friedman's money demand formulation can also explain the procyclical movements of velocity we find in the data, by the relationship between the demand for real money balances and permanent income and by the relationship between permanent income and actual measured income. For example, in a business cycle expansion the demand for money rises less than income because the increase in permanent income is small relative to the increase in actual measured income [see equation (7.8)], and velocity rises. Similarly, in a recession, the demand for money falls less than income because the decline in permanent income is small relative to the decline in actual measured income, and velocity falls.

7.7 Conclusion

We have discussed the early theories of money demand and identified similarities and differences that exist among them. According to the classical quantity theory (developed by Fisher and the Cambridge economists), nominal income is determined primarily by the quantity

---

$$\Delta Y_p = \frac{.10}{1+.10}\Delta Y_t = .09,$$

that is, permanent income changes by only .09. A permanent change in income, however, such as $\Delta Y_{t+j} = 1, \quad j = 0, 1, 2, \cdots$, produces an equal change in permanent income, since

$$\Delta Y_p = \frac{r}{1+r}\frac{1+r}{r} = 1.$$

of money. This proposition, however, rests on the classical economists' assumption that velocity could be treated as reasonably constant.

Keynes criticized the quantity theorists for their assumption of a constant velocity and argued that velocity is affected by behavioral economic variables, most importantly by the nominal interest rate. His conclusion that the demand for money is negatively related to the nominal interest rate is a significant departure from the classical quantity theory of money demand. It is, however, less of a departure from the classical Cambridge approach, which did not rule out such a relationship.

Friedman's theory of the demand for money used a similar approach to that of Keynes and the earlier Cambridge economists, but did not deal with the motives for holding money. By using the theory of portfolio choice, Friedman argued that the demand for money depends on permanent income and the incentives for holding other assets relative to money. In contrast to Keynes, however, he concluded that the demand for money is stable and insensitive to interest rates. This implies that velocity is predictable, yielding the quantity theory conclusion that money is the primary determinant of nominal aggregate spending.

8

Transactions Theories of Money Demand

Theories of the demand for money that emphasize money's medium-of-exchange role in the economy are called *transactions theories*. These theories emphasize that money, unlike other assets, is held to make purchases and in general show that the average amount of real money held involves a trade-off between transactions costs (that arise when people economize on their holdings of money) and interest income foregone.

Transactions theories of the demand for money take many different forms, depending on how the process of obtaining money and making transactions is modeled. To see how these theories explain the demand for money, in this chapter we develop explicitly three prominent models of this type.

8.1 The Baumol-Tobin Model

The choice of when and how often to exchange bonds for money is an important margin of choice for individuals and has been analyzed independently by William Baumol (1952) and James Tobin (1956). Both emphasize the costs and benefits of holding money, coming to similar conclusions about the variables that determine the transactions demand for money. It is argued, for example, that the benefit of holding money is convenience and that the cost of this convenience is the interest income foregone by not holding interest-yielding assets, such as bonds.

To see how maximizing economic agents trade off these benefits, we follow Baumol's (slightly simpler) approach and consider an individual agent who plans to spend Y, in real terms, gradually over the course of a year. The agent has a choice of holding his wealth in the form of (non-interest-yielding) money or in the form of interest-yielding bonds — bonds yield an interest rate of R per period, which is assumed constant over the period and reflects the opportunity cost of holding money. In addition, it is also assumed that each exchange of interest-bearing bonds for money involves a lump-sum transactions cost b in real terms — b is what Baumol calls the *brokerage fee.*

In the setting described, assuming that K is the real value of bonds turned into money each time such a transfer takes place, the total cost of making transactions is the sum of the brokerage cost, $b\,(Y/K)$, where (Y/K) is the number of withdrawals, and the foregone interest if money is held instead of bonds, which is $R(K/2)$, where $K/2$ is the average amount of real money holdings $(= M/P)$. Thus, the total cost can be written as

$$\text{Total Cost} = b\frac{Y}{K} + R\frac{K}{2}. \tag{8.1}$$

Clearly, the fewer the withdrawals, Y/K (and as a result the larger the money balances, $K/2$, held by the individual), the lower will be the brokerage cost and the higher the interest cost. In fact the number of withdrawals that minimizes the total cost of making transactions occurs when the increase in brokerage cost as the result of an additional withdrawal is just offset by the reduction in the interest cost as a result of this withdrawal.

By taking the partial derivative of equation (8.1) with respect to K, setting it equal to zero and solving for K we find the optimal value of K — the value that minimizes total cost. Thus

$$\frac{\partial(\text{Total Cost})}{\partial K} = -\frac{bY}{K^2} + \frac{R}{2} = 0,$$

which yields the following *square root* relationship between K and Y, b, and R

$$K = \sqrt{\frac{2bY}{R}}.$$

At this value of K, average money holding in real terms is, as noted earlier

$$\frac{M}{P} = \frac{K}{2} = \frac{1}{2}\sqrt{\frac{2bY}{R}}, \tag{8.2}$$

suggesting that the demand for real (transactions) money balances is proportional to the square root of Y and inversely proportional to the square root of R. Notice that as $b \to 0$, $M/P \to 0$, meaning that without transactions costs there would be no demand for money, since in this case the individual will be synchronizing cash withdrawals with the purchase of goods and services. Hence, transactions costs have an important role in determining average money balances held, suggesting that the demand for money emerges from a trade-off between transactions costs and interest earnings.

The merit of this approach to the demand for money is that it produces testable relationships between the demand for money and its determinants. For example, taking logarithms of equation (8.2), we can express it as

$$\log\left(\frac{M}{P}\right) = \alpha + \frac{1}{2}\log Y - \frac{1}{2}\log R, \tag{8.3}$$

where $\alpha = \log(1/2)\sqrt{2b}$. In the log-linear equation (8.3), the elasticity of M/P with respect to Y is

$$\eta\left(\frac{M}{P}, Y\right) = \frac{d\log(M/P)}{d\log Y} = \frac{1}{2},$$

implying that a rise in real spending leads to a less-than-proportionate increase in the average holding of real money. Economists refer to this result as *economies of scale* in money holding, meaning that individuals with a larger scale of spending hold less money when expressed as a ratio to their expenditures.

Also the elasticity of M/P with respect to the interest rate is

$$\eta\left(\frac{M}{P}, R\right) = \frac{d\log(M/P)}{d\log R} = -\frac{1}{2},$$

and the elasticity of nominal money, M, with respect to the price level is

$$\eta(M, P) = \frac{d\log M}{d\log P} = 1.$$

Clearly, the Baumol-Tobin model represents a significant departure from the classical quantity theory of money, as it implies economies of scale in the demand for money and an interest elasticity away from zero. This conflict between the Baumol-Tobin model and the quantity theory led Karl Brunner and Allan Meltzer (1967) to reformulate the Baumol-Tobin model and show that for large values of Y or small values of b, there will be no economies of scale in the use of money. On the basis of this result, they argue that the Baumol-Tobin model is not an alternative to the quantity theory, but that it implies it.

However, as Syed Ahmad (1977) shows, although the Brunner and Meltzer specification eliminates economies of scale, it implies an interest elasticity of money demand of -2.0, when $Y \to \infty$ or $b \to 0$. In other words, the Brunner and Meltzer formulation, instead of bringing the Baumol-Tobin model closer to the quantity theory, takes it further away from it.

8.2 The Shopping-Time Model

Although the Baumol-Tobin model pays attention to the medium of exchange role of money, it does not explicitly focus on that most obvious distinguishing characteristic of money. More recently, however, McCallum and Marvin Goodfriend (1987) and Kevin Dowd (1990) suggest that we analyze the demand for money by taking explicitly into account the transactions facilitating services provided by money.

They argue that trade with money, unlike trade by barter which is inefficient and time consuming, produces large savings of what is called *shopping time*. Such savings are desirable, because shopping time reduces leisure which, in turn, reduces utility. In what follows a formal model is presented in which this appealing idea is developed, following McCallum and Goodfriend (1987) and McCallum (1989, Chapter 3).[1]

Consider an economy composed of a large number of similar, infinite-lived individuals. The representative person, who can be viewed as the head of the representative extended family, has preferences given by

$$\mathcal{U} = \sum_{t=0}^{\infty} \beta^t u(c_t, \ell_t),$$

where c_t and ℓ_t are the individual's consumption of goods and leisure respectively, during period t. β is the discount factor and the within-period utility function, $u(c_t, \ell_t)$, is assumed to satisfy the conditions, $u_i(c_t, \ell_t) > 0$ and $u_{ii}(c_t, \ell_t) < 0$, for $i = 1, 2$.

To bring in the role of money, it is assumed that the representative agent holds money (even though higher-yielding assets are available)

[1] This is basically Miguel Sidrauski's (1967) model, expanded to include the 'shopping time' specification developed by Thomas Saving (1971).

because it helps to facilitate transactions. In particular, the agent expends time (and energy) in shopping and the amount of time (and energy) devoted to shopping is positively related to consumption but, for a given level of consumption, negatively related to real money holdings. Of course, the greater the time (and energy) spent in shopping, the smaller the amount left over for leisure, which in turn suggests that leisure will be negatively related to consumption and positively related to real money holdings. We can summarize these ideas in the form of a function, ψ, for leisure demanded

$$\ell_t = \psi(c_t, m_t), \quad \psi_1 < 0, \psi_2 > 0,$$

where $m_t = M_t/P_t$, with M_t being nominal money balances held during period t and P_t is the price level.

The household has access to a production function that is homogeneous of degree one in physical capital and labor. Assuming, for simplicity, that labor is supplied inelastically, the production function can be written as $y_t = f(k_t)$. The production function is assumed to satisfy the conditions $f' > 0$, $f'' < 0$, $f'(0) = \infty$, and $f'(\infty) = 0$.

It is also assumed that the household can buy at time t government bonds at a money price of $1/(1 + R_t)$ and redeem them for one unit of money at time $t + 1$. Hence, the nominal rate of return on bonds is R_t. Assuming that the inflation rate from period t to period $t + 1$ is $\pi_t = (P_{t+1} - P_t)/P_t$, the household's budget constraint can be written as

$$f(k_t) + v_t = c_t + k_{t+1} - k_t$$

$$+ (1 + \pi_t)m_{t+1} - m_t + \frac{b_{t+1}}{1 + r_t} - b_t, \tag{8.4}$$

where m_t and b_t are real (time t) cash and bond holdings, and v_t denotes (lump-sum) real government transfers (net of taxes), received at the start of the period.[2]

The Lagrangian associated with this problem can be written as

[2] The term $b_{t+1}/(1 + r_t) - b_t$ on the right-hand side of (8.4) gives the change in real bond holdings from period t to $t + 1$. In particular, it is

$$\frac{B_{t+1}/(1 + R_t) - B_t}{P_t} = \frac{B_{t+1}}{(1 + R_t)P_t} - b_t = \frac{1 + \pi_t}{1 + R_t} b_{t+1} - b_t = \frac{b_{t+1}}{1 + r_t} - b_t.$$

$$\mathcal{L} = \sum_{t=0}^{\infty} \beta^t u\left[c_t, \psi(c_t, m_t)\right] + \sum_{t=0}^{\infty} \beta^t \lambda_t \Bigg[f(k_t) + v_t$$

$$- c_t - k_{t+1} + k_t - (1+\pi_t)m_{t+1} + m_t - \frac{b_{t+1}}{1+r_t} + b_t \Bigg],$$

where $\psi(c_t, m_t)$ is substituted for ℓ_t, and λ_t is the Lagrange multiplier associated with the household's period t budget constraint. The necessary first-order conditions for optimality can be obtained by differentiating $\mathcal{L}$ with respect to c_t, m_{t+1}, k_{t+1}, and b_{t+1}. They are (for all t)

$$u_1(c_t, \ell_t) + u_2(c_t, \ell_t)\psi_1(c_t, m_t) - \lambda_t = 0; \tag{8.5}$$

$$\beta u_2(c_{t+1}, \ell_{t+1})\psi_2(c_{t+1}, m_{t+1}) - \lambda_t(1+\pi_t) + \beta\lambda_{t+1} = 0; \tag{8.6}$$

$$-\lambda_t + \beta\lambda_{t+1}\left[f'(k_{t+1}) + 1\right] = 0; \tag{8.7}$$

$$-\frac{\lambda_t}{1+r_t} + \beta\lambda_{t+1} = 0, \tag{8.8}$$

where $u_i(c_t, \ell_t)$ and $\psi_i(c_t, m_t)$ denote the partial derivatives of $u(c_t, \ell_t)$ and $\psi(c_t, m_t)$ with respect to the ith argument and $f'(k_{t+1})$ is the rate of return on capital between periods t and $t+1$. Conditions (8.4)-(8.8) are necessary for a maximum. In addition, there are three transversality conditions,

$$\lim_{t\to\infty} m_{t+1}\beta^t\lambda_t(1+\pi_t) = 0; \tag{8.9}$$

$$\lim_{t\to\infty} k_{t+1}\beta^t\lambda_t = 0; \tag{8.10}$$

$$\lim_{t\to\infty} b_{t+1}\beta^t\frac{\lambda_t}{1+r_t} = 0. \tag{8.11}$$

In this setting, (8.4)-(8.8) are necessary for a maximum, while (8.4)-(8.11) are jointly sufficient. In other words, if (8.9)-(8.11) are satisfied, the household's choices of c_t, m_{t+1}, k_{t+1}, and b_{t+1} will be described by (8.4)-(8.8).

Eliminating $\beta\lambda_{t+1}$ between (8.6) and (8.8) and λ_t from the resultant equation, by using (8.5), we get the following optimality condition[3]

$$\frac{u_1(c_t, l_t) + u_2(c_t, l_t)\psi_1(c_t, m_t)}{\beta u_2(c_{t+1}, \ell_{t+1})\psi_2(c_{t+1}, m_{t+1})} = \frac{1 + r_t}{R_t}. \tag{8.12}$$

The optimality condition (8.12) involves only three variables: c_t, m_t, and R_t. Assuming that it can be uniquely solved for m_t as a function of c_t and R_t we obtain an exact *equilibrium* relationship,

$$\frac{M_t}{P_t} = \Phi(c_t, R_t), \tag{8.13}$$

relating the household's optimal consumption, his demand for real balances and the nominal interest rate. Since M_t/P_t and c_t are choice variables, equation (8.13) is not a (money) demand function but, instead, an equilibrium condition among choice variables that the demand functions must satisfy. However, the practice of calling relations like (8.13) money demand functions is extremely common, and although improper, we shall use that terminology.

The foregoing theoretical model, although it is explicit and general, is lacking in one way. In particular, it does not imply, as the Baumol (1952) and Tobin (1956) models do, that c_t enters positively and R_t negatively on the right-hand side of (8.13). To complete the model, we assume that $\Phi(\cdot)$ possesses partial derivatives and that $\Phi_1 > 0$ and $\Phi_2 < 0$. It is to be noted, however, that it is not true that those signs are strictly implied for all functions satisfying the assumptions that we placed on $u(c_t, \ell_t)$ and $\psi(c_t, M_t/P_t)$.

8.3 Cash-in-Advance Models

Another popular device for introducing money into macroeconomic equilibrium models is the cash-in-advance constraint, proposed by Robert Clower (1967). This approach captures the role of money as a medium of exchange by requiring that a transaction can take place only if the money needed for the transaction is held in advance. Moreover, it provides an explanation as to why rational economic agents hold

[3] From (8.7) and (8.8) we also get

$$1 + f'(k_{t+1}) = \frac{1 + R_t}{1 + \pi_{t+1}},$$

which is basically the Fisher equation, linking the nominal interest rate to the real interest rate and the inflation rate.

money — an asset that is intrinsically useless and return-dominated by other assets.

The simplest cash-in-advance model parallels the Sidrauski and shopping time models and was introduced by Alan Stockman (1981). Following Stockman (1981), we consider an economy with a representative individual with perfect foresight, solving the following problem

$$\max_{\{c_t,\, k_{t+1},\, M_{t+1}\}_{t=0}^{\infty}} \sum_{t=0}^{\infty} \beta^t u(c_t), \tag{8.14}$$

subject to a series of one period budget constraints (for all t)

$$f(k_t) + v_t = c_t + k_{t+1} - (1-\delta)k_t + \frac{M_{t+1} - M_t}{P_t}, \tag{8.15}$$

and

$$\frac{M_t}{P_t} + v_t \geq c_t. \tag{8.16}$$

The first constraint is the dynamic budget constraint corresponding to the asset accumulation equation of the Sidrauski model. M_t is the amount of nominal money balances carried over from the previous period, period $t-1$, and M_{t+1} is the amount of nominal balances held at the end of period t and carried forward to period $t+1$. The second constraint is the cash-in-advance constraint, according to which real money balances carried into the period plus the government transfer received at the start of the period cannot be less than real consumption spending during the period.

The problem is solved by maximizing (8.14) subject to constraints (8.15) and (8.16), taking k_0 and M_0 as given. The Lagrangian is as follows

$$\mathcal{L} = \sum_{t=0}^{\infty} \beta^t \Big\{ u(c_t)$$

$$+ \lambda_t \left[f(k_t) + v_t - c_t - k_{t+1} + (1-\delta)k_t - \frac{M_{t+1} - M_t}{P_t} \right]$$

$$+ \gamma_t \left[\frac{M_t}{P_t} + v_t - c_t \right] \Big\},$$

where λ and γ are the Lagrange multipliers for the two constraints. The necessary first-order conditions for optimality can be obtained by differentiating $\mathcal{L}$ with respect to c_t, k_{t+1}, and M_{t+1}. They are (for all t)

$$u'(c_t) = \lambda_t + \gamma_t; \tag{8.17}$$

$$\beta\lambda_{t+1}\left[f'(k_{t+1}) + 1 - \delta\right] = \lambda_t; \tag{8.18}$$

$$\beta\lambda_{t+1}\frac{1}{P_{t+1}} + \beta\gamma_{t+1}\frac{1}{P_{t+1}} = \lambda_t\frac{1}{P_t}. \tag{8.19}$$

Equation (8.17) equates the marginal utility of consumption, $u'(c_t)$, to the marginal cost of consumption (which is the marginal utility of having an additional unit of real money balances), $\lambda_t + \gamma_t$. Equation (8.18) equates the marginal value of an additional unit of capital in period $t+1$ to the marginal cost of holding an additional unit of real balances in period t. Finally, equation (8.19) states that the marginal value of an extra unit of nominal balances in period $t+1$, deflated by that period's price level, equals the marginal cost of having that additional unit of money. Conditions (8.15)-(8.19) are necessary for a maximum. In addition, there are two transversality conditions,

$$\lim_{t\to\infty} k_{t+1}\beta^t\lambda_t = 0; \tag{8.20}$$

$$\lim_{t\to\infty} M_{t+1}\beta^t\lambda_t\frac{1}{P_t} = 0. \tag{8.21}$$

In this setting, (8.15)-(8.19) are necessary for a maximum, while (8.15)-(8.21) are jointly sufficient. In other words, if (8.20)-(8.21) are satisfied, the household's choices of c_t, k_{t+1}, and M_{t+1} will be described by (8.15)-(8.19).

We can now turn to steady-state analysis. Consumption and the capital stock are constant in the steady state. This implies that $\lambda + \gamma$ is constant over time, suggesting that λ and γ must each be constant. Hence, in the steady state (8.18) becomes

$$\beta\left[f'(k^*) + 1 - \delta\right] = 1,$$

suggesting that the steady-state capital and real interest rate are independent of the rate of inflation. Clearly, this model provides the same result as the Sidrauski model.

So far, however, we have assumed that the liquidity-cash-in-advance constraint pertains only to the purchases of consumption goods, c_t. We now investigate the robustness of our results regarding the superneutrality of money in cash-in-advance models, by assuming that the cash-in-advance constraint pertains to purchases of consumption as well as capital. Under this assumption, we write the cash-in-advance constraint as

$$\frac{M_t}{P_t} + v_t \geq c_t + k_{t+1} - (1-\delta)k_t, \tag{8.22}$$

where $k_{t+1} - (1-\delta)k_t$ is gross investment. The cash-in-advance constraint (8.22) states that the individual must be able to finance current consumption and gross investment out of money balances carried over from the previous period plus current transfers.

Now the problem of a private agent is to maximize (8.14) subject to (8.15) and (8.22). The Lagrangian associated with this problem is

$$\mathcal{L} = \sum_{t=0}^{\infty} \beta^t \Big\{ u(c_t)$$

$$+ \lambda_t \left[f(k_t) + v_t - c_t - k_{t+1} + (1-\delta)k_t - \frac{M_{t+1} - M_t}{P_t} \right]$$

$$+ \gamma_t \left[\frac{M_t}{P_t} + v_t - c_t - k_{t+1} + (1-\delta)k_t \right] \Big\},$$

and the necessary first-order conditions are

$$u'(c_t) = \lambda_t + \gamma_t; \tag{8.23}$$

$$\beta\lambda_{t+1}\left[f'(k_{t+1}) + 1 - \delta\right] + \beta\gamma_{t+1}(1-\delta) = \lambda_t + \gamma_t; \tag{8.24}$$

$$\beta\lambda_{t+1}\frac{1}{P_{t+1}} + \beta\gamma_{t+1}\frac{1}{P_{t+1}} = \lambda_t \frac{1}{P_t}. \tag{8.25}$$

Using (8.23)-(8.25), we again turn to steady state analysis. As before, the capital stock and consumption are constant in the steady state, implying that $\lambda + \gamma$ is constant over time. Hence, in the steady state (8.25) becomes

$$\beta(\lambda_{t+1} + \gamma_{t+1}) = \lambda_t \frac{P_{t+1}}{P_t},$$

which yields

$$\gamma = \frac{1+\pi}{\beta}\lambda - \lambda,$$

since λ and γ are each constant over time and $P_{t+1}/P_t = 1+\pi$. Substituting the last expression into (8.24) and rearranging yields the steady-state condition

$$f'(k^*) = (1+\pi)\frac{1-(1-\delta)\beta}{\beta^2}.$$

Thus, when both consumption and gross-investment are subject to the liquidity constraint, higher inflation rates are associated with higher steady-state real rates of interest and lower capital stock and money balances.

The reason for this result is the complementarity of money and capital. In particular, as Orphanides and Solow (1990, p. 256) put it

> "[i]nvestment of an additional unit of capital in period $t+1$ requires an additional unit of money holdings in period t. Higher inflation increases the cost of the additional unit of investment by increasing the cost of holding the money necessary for the investment. Thus, it reduces the (net of money holding costs) return on a unit of investment. As a result, the demand for capital is reduced and less money is held."

Finally, since consumption and the capital stock are constant in the steady-state market equilibrium, the cash-in-advance constraint is satisfied as an equality, implying the following money demand function

$$\frac{M_t}{P_t} = f(k),$$

which is similar to the quantity-theoretic money demand function.

We have presented a simple cash-in-advance model and seen that the specification of the transactions subject to the liquidity constraint is important. For other (more recent) cash-in-advance frameworks, see Lars Svennson (1985), Lucas and Nancy Stokey (1987), and Thomas Cooley and Gary Hansen (1989).

8.4 Conclusion

The models that we have discussed in this chapter seek to derive the demand function for money from explicit consideration of the notion that money facilitates transactions. In particular, the Baumol-Tobin model shows how the use of money in completely foreseen transactions implies economies of scale and an interest elasticity of money demand significantly different from zero.

However, although the Baumol-Tobin model pays attention to money's role as a means of exchange in markets for goods and services, it does not focus explicitly on that role. It does not, for example, explain the holding of money in terms of the transactions facilitating services provided by money, but in terms of transactions costs, which influence money demand and consumption decisions. It is only the shopping time and cash-in-advance models that focus explicitly on transactions services and money's role as a medium of exchange.

9

Portfolio Theories of Money Demand

Theories of the demand for money that emphasize the role of money as a store of value are called *asset* or *portfolio theories*. These theories stress that people hold money as part of their portfolio of assets and predict that the demand for money depends on the return and risk offered by money and by other assets that people can hold instead of money.

We have already discussed two asset theories of the demand for money — the Keynesian speculative theory of money demand and Friedman's modern quantity theory. In what follows we will discuss the portfolio theories of the demand for money developed by Tobin (1958) and Thomas Sargent and Neil Wallace (1982).

9.1 Tobin's Theory of Liquidity Preference

We have already discussed Tobin's contribution to the transactions theory of money demand in the Baumol-Tobin model of cash management. Tobin, in his 1958 article, "Liquidity Preference as Behavior Towards Risk," has also reformulated Keynes's speculative theory of money demand. While Keynes derived an inverse aggregate relationship between the demand for money and the interest rate from the assumption of certain expectations that differ among individuals, Tobin (1958) derived this same demand for money relationship for an individual from the assumption of uncertain expectations and risk avoidance — the latter being the basis for his Nobel Prize in economics. He refers to his theory as a theory of *liquidity preference*, following Keynes's terminology.

Tobin assumes that the individual holds a portfolio consisting of a proportion of wealth w_1 in money and w_2 in the risky asset, say perpetual bonds. Notice that $w_1 + w_2 = 1$. Money has a riskless rate of return $E_1 = R_f$ (≥ 0) and therefore a variance of return that is exactly zero, $\sigma_1^2 = 0$. The risky asset has an expected rate of return E_2 ($> R_f$) and a variance of return σ_2^2 ($> \sigma_1^2$).

The expected return on the portfolio, E_p, is simply a weighted average of the expected returns on each of the assets, with the weights being the proportion of wealth invested in each asset,

$$E_p = \sum_{i=1}^{k} w_i E_i,$$

where k is the number of assets in the portfolio and E_i is the expected return on asset i. Since $k = 2$, $E_1 = R_f$, and $w_1 = 1 - w_2$, we have

$$E_p = (1 - w_2)R_f + w_2 E_2. \tag{9.1}$$

The total variance of the portfolio, σ_p^2, is

$$\sigma_p^2 = \sum_{i=1}^{k}\sum_{j=1}^{k} w_i w_j R_{ij} \sigma_i \sigma_j,$$

where R_{ij} is the simple correlation between returns on assets i and j. However, since $k = 2$ and $R_{11} = R_{22} = 1$ we have

$$\sigma_p^2 = w_1^2 \sigma_1^2 + w_2^2 \sigma_2^2 + 2 w_1 w_2 R_{12} \sigma_1 \sigma_2.$$

Finally, since $\sigma_1 = 0$ and the assets are independent (i.e., $R_{12} = 0$), the last equation reduces to

$$\sigma_p^2 = w_2^2 \sigma_2^2. \tag{9.2}$$

Rearranging equation (9.2) yields $w_2 = \sigma_p / \sigma_2$ which, after substituting back into equation (9.1) and rearranging terms, gives

$$E_p = R_f + \left(\frac{E_2 - R_f}{\sigma_2} \right) \sigma_p. \tag{9.3}$$

This equation shows a simple linear relationship between expected portfolio return, E_p, and portfolio risk, σ_p. Specifically, expected portfolio return is the sum of the risk-free rate of return, R_f, and $(E_2 - R_f)/\sigma_2$ times the portfolio risk, σ_p. The slope $(E_2 - R_f)/\sigma_2$ is referred to as the *price of risk*, since it measures how σ_p and E_p can be traded off in making portfolio choices.

Equation (9.3) describes the options available to investors with respect to holding alternative portfolios. To examine, however, how investors determine what portfolios to hold, we must first examine investor preferences over expected portfolio return and portfolio risk. In doing so, we assume that the individual wishes to maximize a utility function depending on E_p and σ_p^2, as follows

$$\mathcal{U} = u(E_p, \sigma_p^2),$$

with $u_1 > 0$, $u_2 < 0$, $u_{11} < 0$, and $u_{22} < 0$. Approximating the explicit utility function by

$$\mathcal{U} = E_p - \frac{\gamma}{2}\sigma_p^2,$$

where γ is a constant representing the degree of risk aversion, and using equations (9.1) and (9.2), the individual's (unconstrained) optimization problem is to maximize

$$\mathcal{U} = (1 - w_2)R_f + w_2 E_2 - \frac{\gamma}{2} w_2^2 \sigma_2^2,$$

with respect to w_2.

The first-order condition for maximization is

$$\frac{\partial \mathcal{U}}{\partial w_2} = -R_f + E_2 - \gamma w_2 \sigma_2^2 = 0,$$

which implies the optimal proportion of the holding of the risky asset, w_2^*,

$$w_2^* = \frac{E_2 - R_f}{\gamma \sigma_2^2}. \tag{9.4}$$

Equation (9.4) is Tobin's (1958) *mean-variance model* of asset demands and holds for any risky portfolio.

Consider now the effects of an increase in the interest rate, with no change in the perceived riskiness of bonds. Clearly, according to Tobin's asset demand model the optimal proportion of the holding of the risky asset will increase and money holdings will decline. That is, the increase in the interest rate reduces the demand for money. Hence, Tobin's model implies a negative interest rate elasticity of the same general form as we saw in the Keynesian and Baumol-Tobin models.

The effect of a change in the perceived riskiness of bonds could also be discussed. In terms of equation (9.4), an increase in the riskiness of bonds, σ_2^2, involves a decline in w_2^*. This causes money holdings to increase, bond holdings and expected return to decrease, and portfolio risk, σ_p^2, may increase or decrease, depending on the magnitude of the proportionate decline in w_2 relative to the magnitude of the proportionate increase in σ_2^2 — see equation (9.2).

Tobin has therefore reformulated the Keynesian asset theory of money demand, largely in terms of portfolio theory. The key characteristic of Tobin's theory is that it explains the speculative demand for money by the assumption of uncertain expectations and the principle of portfolio diversification by individuals, rather than Keynes's assumption of certain expectations that differ among people. As Laidler (1993, p. 85) puts it,

> "this model suggests that some measure of the economy's assessment of the riskiness of assets other than money may be worth including in the demand-for-money function."

Given, however, the menu of assets available in most countries, Tobin's approach actually undermines the speculative demand for money, by requiring that money be an important component of diversified portfolios. The reason is that other risk-free assets (such as, for example, savings deposits and Treasury bills) paying a higher rate of return than money, may displace money from portfolios.

9.2 Money and Overlapping Generations

The demand for money as an asset has also been analyzed by non-Keynesian economists. Recently, Sargent and Wallace, two leading exponents of new classical macroeconomic theory, developed monetary theory based on the overlapping generations model that we discussed in Chapter 4 — see Sargent and Wallace (1982). As we will see, this model also has nothing to do with the means-of-exchange function of money.[1]

Assume that time is discrete (indexed by the subscript $t = 1, \ldots, \infty$), that people live for two periods, and that population grows at the rate ν, so that by appropriate normalization $L_t = (1 + \nu)^t$. People born at time t are young at t and old at $t + 1$. In the first period of life, each

[1] More detailed expositions can be found in Blanchard and Fischer (1989), Laidler (1993), and Bruce Champ and Scott Freeman (1994).

individual is endowed with one unit of the nonstorable consumption good, but receives no endowment when old. To keep things simple we will assume that all agents born at time t have identical preferences and that the representative agent's utility function is given by

$$\mathcal{U} = u(c_{1t}) + \beta u(c_{2t+1}),$$

where (as before) c_{1t} denotes the amount of the good consumed in the first period of life by an individual born in period t and c_{2t+1} denotes the amount that the same individual consumes in the second period of life. In other words, the first subscript gives the age of the consumer and the second one gives the date (because the economy itself goes on forever).

In order to open up an intergenerational trade opportunity, we now introduce money into the economy by assuming that the government gives H perfectly divisible units of fiat money to the old in the initial period. Unlike the consumption good, fiat money can be stored between periods and we assume that fiat money is not valued for its own sake but simply because individuals believe it will have value in the future. In particular, we assume that the old and every subsequent generation believe that money can be exchanged for goods at price P_t at time t — we refer to P_t as the price level.

For fiat money to have value, the economy must go on forever, the supply of fiat money must be limited, and it must be impossible (or very costly) to counterfeit. If, for example, the economy ended at some time T, generation T would have no incentive to buy money (from generation $T-1$) that it could not spend at time $T+1$. This in turn implies that generation $T-1$, knowing this, would not want to buy money from generation $T-2$, and so on. In other words, in such a finite-lived economy, money could not be introduced at all and the economy would remain at the barter equilibrium. Hence, the assumption that the economy goes on forever is a necessary but not a sufficient condition for money to be valued. If, for example, individuals had the ability to print money costlessly, its supply would rapidly approach infinity, driving its value to zero.

Let us now examine how agents will decide how much money to hold. Consider an individual born at t, $t \geq 1$. His problem is to maximize,

$$u(c_{1t}) + \beta u(c_{2t+1}),$$

with respect to c_{1t} and c_{2t+1}, subject to the budget constraints

$$c_{1t} + \frac{M_t^d}{P_t} = 1;$$

$$c_{2t+1} = \frac{M_t^d}{P_{t+1}},$$

where M_t^d is the individual's demand for money at time t, which equals $1 - c_{1t}$. The first equation is the budget constraint facing the individual in the first period of life. The right-hand side represents the individual's total sources of goods (his endowment) and the left-hand side the individual's total uses of goods (consumption and the acquisition of money — $1/P_t$ is the value of one unit of money in terms of goods). The second equation is the budget constraint facing the individual in the second period of life (period $t+1$). This equation makes clear that since the individual receives no endowment when old, he can acquire second-period consumption goods only by spending the money acquired in the previous period.

By substituting the period-$(t+1)$ budget constraint into the period-t budget constraint to eliminate M_t^d we can obtain the individual's lifetime budget constraint

$$c_{1t} + \frac{P_{t+1}}{P_t} c_{2t+1} = 1,$$

which shows combinations of first- and second-period consumption that an individual can afford over his lifetime.

To solve the individual's problem we set up the Lagrangian

$$\mathcal{L} = u(c_{1t}) + \beta u(c_{2t+1}) + \lambda \left(1 - c_{1t} - \frac{P_{t+1}}{P_t} c_{2t+1} \right),$$

where λ is the Lagrange multiplier on the lifetime budget constraint. The first-order conditions are

$$u'(c_{1t}) = \lambda;$$

$$\beta u'(c_{2t+1}) = \lambda \frac{P_{t+1}}{P_t}.$$

These combine to give the Euler equation

$$\frac{u'(c_{1t})}{\beta u'(c_{2t+1})} = \frac{P_t}{P_{t+1}},$$

which states that the marginal rate of substitution between first- and second-period consumption equals the rate of return on money, P_t/P_{t+1}.

The Euler equation implies a money demand function (which is a saving function)

$$\frac{M_t^d}{P_t} = \Phi\left(\frac{P_t}{P_{t+1}}\right).$$

If we define the deflation rate, π_t, by $(1+\pi_t) = P_t/P_{t+1}$, we can write the money demand function as

$$\frac{M_t^d}{P_t} = \Phi(1+\pi_t).$$

Let us now describe equilibrium in the money market and find an equilibrium time path of the value of money. Since the old supply inelastically the money they have (which is H) and the young buy money according to the above equation, the money market — or equivalently, by Walras' law the goods market — will be in equilibrium when

$$(1+\nu)^t M_t^d = H,$$

where ν is the constant population growth rate and $(1+\nu)^t M_t^d$ is the total demand for money by all individuals in the economy at time t. Using the above two equations at time t and $t+1$ we obtain

$$\frac{1+\nu}{1+\pi_t} = \frac{\Phi(1+\pi_t)}{\Phi(1+\pi_{t+1})}.$$

To simplify, we consider a stationary allocation, where the members of every generation have the same lifetime consumption pattern, that is, $c_{1t} = c_1$ and $c_{2t+1} = c_2$ for every period t. This definition implies that the price ratio, P_t/P_{t+1} will also be independent of time and that $\Phi(1+\pi_t) = \Phi(1+\pi_{t+1})$. In turn, our definition of a stationary equilibrium and the last equation imply that the rate of deflation must be equal to the constant population growth rate ($\pi = \nu$).

Since $\nu > 0, \pi = \nu$ means that the price of the consumption good is falling over time, or, equivalently, that the value of money is increasing over time. In other words, in a growing economy with a constant fiat money stock, the price of the consumption good must decrease at a rate such that the supply of real money balances grows at the same rate as the total demand for money, which is itself growing at the population growth rate.

The analysis also applies to a shrinking economy, where $\nu < 0$. In such a case, with a constant fiat money stock, the price level will be rising over time, implying a falling value of money. Finally, in the special case of constant population ($\nu = 0$), the last equation gives that

$$1 + \pi_t = 1 \quad \text{or} \quad \frac{P_t}{P_{t+1}} = 1,$$

implying a constant price level, or, equivalently, a rate of return on fiat money of 1. Since the value of money is the inverse of the price level, it, too, is constant over time.

So far we have shown that fiat money can be valued (in the sense that it can be traded for consumption goods), and that the introduction of positive valued fiat money can lead to a Pareto optimal allocation of resources across generations, assuming that the economy reaches a stationary equilibrium. We have also concentrated on factors that affect the demand for money and we have found that the overlapping generations model explains the demand for money as an asset but not as a means of exchange, as it might appear at first glance — what matters is money's capacity to act as a store of value between periods.

There seem to be problems, however, with some of the results the overlapping generations model generates. For example, in an economy in which money coexists with another asset that yields a real return (like government bonds or land), money will be driven out of the model since it doesn't pay interest and the other asset does. Hence, the overlapping generations model cannot explain the value of the rate of return-dominated money in actual economies, and like the Keynesian speculative theory of the demand for money, implies an infinite elasticity of the demand for money as an asset with respect to the rate of return on other assets.

Regarding rate of return dominance, Sargent and Wallace (1982) rely on *legal restrictions* to explain the coexistence of fiat money and government bonds in an overlapping generations equilibrium. This answer to rate of return dominance has become known as *legal restrictions theory*. It claims that it is only because of institutionally imposed prohibitions, that introduce an element of coercion into the decision to hold money, that people hold rate of return-dominated money.

9.3 Conclusion

The models that we have discussed in this chapter are a sample of those models that seek to derive the demand for money as an asset. These models are theoretically interesting, but certainly unrealistic since they abstract from money's most obvious distinguishing characteristic — its ability to function as a medium of exchange.

These models, however, should not be dismissed out of hand for this reason. As Laidler (1993, p. 90) puts it,

> "models of the demand for money as an asset yield predictions about the nature of the demand-for-money function that arise when the means-of-exchange function is ignored. Hence they enable us to formulate empirical questions whose answers might help us to decide whether or not we need take account of money's peculiar characteristics when we construct theories about it."

From this point of view, models of the demand for money as an asset have important implications for how the macroeconomy functions.

Part 4: Empirical Approaches to the Demand for Money

Overview of Part 4

Chapter 10 deals with the empirical relevance of some of the theories presented in Part 3, taking a conventional approach to estimation and hypothesis testing. In Chapters 11, 12, and 13, we turn to a discussion of the same issues using recent advances in the field for applied econometrics, such as integration and cointegration theory. The approach here is similar to that taken by Hoffman and Rasche (1996). That is, we pay explicit attention to the econometric consequences of nonstationary data and their implications for the study of money demand.

In both Chapters 10 and 13, comparisons are made among simple-sum, Divisia, and CE monetary aggregates (of M1, M2, M3, and MZM). Similar comparisons will be provided in Chapter 16, as one of the objectives of this textbook is to provide empirical evidence regarding the relative merits of alternative monetary aggregation procedures.

Chapter 14 examines money demand issues using cross-country data, for 48 countries over the 1980-1995 period. In particular, we investigate conventional money demand functions, for both narrow and broad monetary aggregates, and the role that institutions, financial structure, and financial development may have in the demand for money.

10

Conventional Demand for Money Functions

In our discussions of theories of macroeconomic behavior, we have talked about the demand for money function. As we saw, this function is a critical component in the formulation of monetary policy. Moreover, it has been argued over the years that a stable demand function for money is a necessary condition for money to exert a predictable influence on the economy so that control of the monetary aggregates can be a useful instrument of economic policy.

Not surprisingly, then, numerous empirical studies have been conducted in many countries to evaluate the determinants and stability of the demand for money. As Stephen Goldfeld and Daniel Sichel (1990, p. 300) put it,

> "the evidence that emerged, at least prior to the mid-1970s, suggested that a few variables (essentially income and interest rates, with appropriate allowance for lags) were capable of providing a plausible and stable explanation of money demand."

In this chapter we look at the factors that have shaped the evolution of the research on modeling and estimating money demand functions. In doing so, we discuss measurement issues on a variable-by-variable basis, and distinguish the long-run and short-run concepts of the demand for money by the absence of adjustment costs in the former and their presence in the latter. We conclude that conventional money demand functions are seriously misspecified, and argue for new modelling approaches.

10.1 The Basic Specification

As we have seen there are different money demand theories, emphasizing different considerations and implying different testable theoretical hypotheses. These theories, however, share common important elements. In particular, most of them suggest a relationship between the quantity of money demanded and a few important variables that represent significant links to the level of economic activity. In general, this theoretical money demand relationship can be written as

$$\frac{M_t}{P_t} = \Phi(R_t, Y_t),$$

where M_t is nominal money balances demanded, P_t is the price index used to convert nominal balances to real balances, Y_t is the *scale* variable relating to activity in the real sector of the economy, and R_t is the *opportunity cost* of holding money.

In what follows, we discuss the choice of variables as suggested by the different theories of the demand for money. For other similar, and perhaps more detailed discussions, see Edgar Feige and Douglas Pearce (1977), Judd and Scadding (1982), Laidler (1993), and Goldfeld and Sichel (1990).

10.1.1 Definition of Money

The first problem in the empirical estimation of money demand functions is the selection of an explicit measure of money. In general, transactions-based theories of the demand for money emphasize narrow definitions of money that include currency and checkable deposits. Once one moves away, however, from a transactions approach, there are problems in determining which monetary assets belong to which monetary aggregate — see, for example, Goldfeld and Sichel (1990) for a discussion of the relevant issues as they pertain to the United States.

In addition to problems of determining the monetary assets over which to aggregate, the monetary aggregates currently in use by most central banks around the world have been criticized for being based on the simple-sum method of aggregation. The essential property of this method of monetary aggregation is its assigning all monetary components a constant and equal (unitary weight). This index is M_t in

$$M_t = \sum_{j=1}^{n} x_{jt}, \tag{10.1}$$

where x_{jt} is one of the n monetary components of the monetary aggregate M_t. This summation index implies that all monetary components contribute equally to the money total and it views all components as dollar for dollar perfect substitutes. Such an index, there is no question, represents an index of the stock of nominal monetary wealth, but cannot, in general, represent a valid structural economic variable for the services of the quantity of money.

Over the years, there has been a steady stream of attempts at properly weighting monetary components within a simple-sum aggregate. With no theory, however, any weighting scheme is questionable. As we will see in later chapters, it was Barnett (1980) who derived the theoretical linkage between monetary theory and aggregation and index number theory. He applied economic aggregation and index number theory and constructed monetary aggregates based upon Erwin Diewert's (1976) class of superlative quantity index numbers, to be discussed in more detail later in this book. The new aggregates are Divisia quantity indexes which are elements of the superlative class. The Divisia index (in discrete time) is defined as

$$\log M_t^D - \log M_{t-1}^D = \sum_{j=1}^{n} w_{jt}^*(\log x_{jt} - \log x_{j,t-1}). \tag{10.2}$$

According to equation (10.2) the growth rate of the aggregate is the weighted average of the growth rates of the component quantities, with the Divisia weights being defined as the expenditure shares averaged over the two periods of the change,

$$w_{jt}^* = (1/2)(w_{jt} + w_{j,t-1}),$$

for $j = 1, \cdots, n$, where

$$w_{jt} = \frac{p_{jt}x_{jt}}{\sum_{k=1}^{n} p_{kt}x_{kt}}$$

is the expenditure share of asset j during period t, and p_{jt} is the *user cost* of asset j, derived in Barnett (1978),

$$p_{jt} = \frac{(R_t - r_{jt})}{(1 + R_t)}, \tag{10.3}$$

which is just the opportunity cost of holding a dollar's worth of the jth asset. In equation (10.3), r_{jt} is the market yield on the jth asset, and R_t is the yield available on a *benchmark* asset that is held only to carry wealth between multiperiods.

More recently, Rotemberg (1991) and Rotemberg, Driscoll, and Poterba (1995) proposed the currency equivalent (CE) index

$$CE_t = \sum_{j=1}^{n} \frac{R_t - r_{jt}}{R_t} x_{jt}. \tag{10.4}$$

In (10.4), as long as currency yields no interest, units of currency are added together with a weight of one. Other assets are added to currency but with a weight that declines toward zero as their return increases toward R_t.

Clearly, the problem of the definition of money is an aggregation problem. We will not get into the specific aggregation issues in this discussion, but will reserve the topic for Part 5 of the book. There we will consider some important theoretical issues and approach the topic by means of a system of demand equations for the various monetary assets, estimating the degree of substitution between monetary assets and testing for weakly separable asset groupings.

10.1.2 Scale Variables

The scale variable in the money demand function is used as a measure of transactions relating to economic activity. As we saw in earlier chapters, transactions theories of money demand emphasize the level of income as the relevant scale variable whereas asset theories place more emphasis on wealth. Wealth, however, is difficult to measure. In fact, only in a handful of countries like the United Kingdom and the United States it is possible to construct long time series on financial wealth. Moreover, these measures are less inclusive than a general measure of wealth that includes the value of human as well as nonhuman capital, as suggested by Friedman's (1956) modern quantity theory discussed in Chapter 7.

To measure this more inclusive concept of wealth, as Laidler (1993, pp. 99-100) put it

> "presents formidable difficulties of its own, and virtually all attempts to come to grips with them have started from the simple

idea that wealth is the discounted present value of expected future income. So long as the rate of discount used can be regarded as constant, wealth varies in exactly the same fashion as expected income. If expected income rises by 10%, so will wealth; if it falls, so will wealth, and so on. One is interested in studying the relationship between *variations* in the level of wealth and *variations* in the demand for money and, because this is the case, it is not important whether wealth is measured directly or whether *expected income*, or, as it is often called, *permanent income*, is used as a proxy for this variable."

One way to measure expected income is to use Cagan's (1956) model of *adaptive expectations* that we discussed in Chapter 3. In terms of our notation, the adaptive expectations model for the unobserved expected level of income at time t, Y_t^e, can be expressed as

$$Y_t^e - Y_{t-1}^e = \theta \left(Y_t - Y_{t-1}^e \right),$$

where $0 \leq \theta \leq 1$. A simple rearrangement of the adaptive expectations model yields

$$Y_t^e = \theta Y_t + (1 - \theta) Y_{t-1}^e.$$

This formulation states that the expected level of income at time t is a weighted average of the current actual level of income and last period's expected value of income, with the weights being the adjustment parameters θ and $1 - \theta$. Finally, through continuous back-substitution, the second presentation of the adaptive expectations model yields

$$Y_t^e = \theta Y_t + \theta (1 - \theta) Y_{t-1} + \theta (1 - \theta)^2 Y_{t-2} + \cdots,$$

according to which the unobserved expected level of income at time t is a weighted average of the current actual level of income and already known income levels of the past, Y_{t-1}, Y_{t-2}, and so on. The weighting scheme, θ, $\theta (1 - \theta)$, $\theta (1 - \theta)^2$, and so on, represents a memory expressing the influence of past income levels on the formation of expectations. If, for example, θ is close to zero, then the weights decline slowly and the economic agent is said to have a 'long memory,' in the sense that information from the distant past significantly influences the formation of expectations. If θ is close to one, then the weights decline quickly and the agent is said to have a 'short memory,' in the sense that only information from the recent past influences the formation of expectations.

The adaptive expectations model, however, has been faulted on the grounds that it doesn't assume enough rationality on the part of economic agents. In particular, according to the third presentation of the adaptive expectations hypothesis, economic agents use only current and past values of the variable in question when formulating expectations for the future. An alternative hypothesis for economic analysis of expectational behavior is John Muth's (1961) *rational expectations* hypothesis. According to the rational expectations notion, economic agents use all of the available and economically usable information, including relevant economic theory, in the formation of expectations for the future.

As it happens, the concept of rational expectations has been embraced by the economics profession and the theory has been enhanced by important contributions by Lucas (1972, 1973), Sargent and Wallace (1975), and Barro (1976). Of course, in order to implement the notion of rational expectations empirically, it is necessary to quantify the concepts of 'available information' and 'relevant economic theory.' Such quantification, although potentially fruitful, is very difficult, since it also requires an explicit treatment of a large number of other issues, such as, for example, structural shifts in the income growth process — for empirical work along these lines, see Barro (1977, 1978).

As an empirical matter, the level of current income is most often used to represent the scale variable in the money demand function. As Laidler (1993, pp. 98-99) put it

> "the measurement of this variable presents little problem because, although gross national product series, net national product series and gross domestic product series have been used to measure it, these variables move rather closely together over time and no important difference in results is obtained by using one or the other."

The level of income, however, is less inclusive than a more comprehensive measure of transactions. For example, gross national product (GNP) excludes transactions in financial assets, sales of intermediate goods, transfers, and purchases of existing goods, all of which are likely to affect the demand for money. For this reason, in recent years research has focused on the construction of scale variables based on more general measures of transactions. It is too early, however, to tell if these new data will yield significant improvements in the explanation of aggregate money demand.

Recent research has also focused on the disaggregation of GNP into several scale variables, reflecting the notion that not all transactions are equally money intensive. For example, Gregory Mankiw and Lawrence Summers (1986) argue that consumption is a more empirically successful scale variable in estimated money demand functions than GNP. It has also been argued that the disaggregation of GNP into components that reflect the nature of international transactions is likely to be important for open economies. However, there is no firm evidence that disaggregation of GNP improves the performance of money demand functions.

10.1.3 Opportunity Costs

For a given definition of money, the opportunity cost of holding money is the difference between the rate of return on assets alternative to money and the own rate on money. Regarding the rate of return on alternative assets, those researchers that adopt a transactions approach and use a narrow definition of money typically use one or more short-term interest rates, such as the Treasury bill rate, the commercial paper rate, or the saving deposit rate. On the other hand, those that adopt an asset approach and use broader definitions of money typically use longer-term rates of interest.

As to the own rate on money, most researchers treat it as zero, implicitly assuming that the explicit rate of return on most forms of money (i.e., currency, demand deposits, etc.) is zero. This is not correct, however, because even when the explicit return is zero, money earns an *implicit rate* of return, in the form of gifts, services, or reduced transactions fees, when deposit holders maintain a minimum level of deposits. The measurement, however, of this implicit rate of return is a difficult matter and it is perhaps for this reason that this issue has generally been ignored — see Benjamin Klein (1974) and Richard Startz (1979) for exceptions.

Of course, there are other variables that may play a role in the money demand function. For a discussion with further references see Goldfeld and Sichel (1990), Laidler (1993), and Subramanian Sriram (1999).

10.2 The Long-Run Function

In general, the starting point in the empirical estimation of money demand functions is the long-run, log linear function of the form

$$\log\left(\frac{M_t^*}{P_t}\right) = \alpha + \beta_1 \log Y_t + \beta_2 R_t + \varepsilon_t, \tag{10.5}$$

where M^* denotes the *desired* stock of nominal money, P is the price index used to convert nominal balances to real balances, Y is the scale variable, and R is the opportunity cost variable.

As an example of estimated long-run money demand functions, in Table 10.1 we report estimation results based on monthly observations for the United States over the 1960:1 to 2006:1 period. In doing so, we use the industrial production index as the scale variable, the 90-day T-bill rate as the opportunity cost variable, and make comparisons between simple-sum, Divisia, and currency equivalent monetary aggregates (of M1, M2, M3, and MZM). The monetary aggregates were obtained from the St. Louis MSI database, maintained by the Federal Reserve Bank of St. Louis as a part of the Bank's Federal Reserve Economic Database (FRED). The three different monetary aggregation procedures will be discussed in detail in Chapters 15-17.

The numbers in parentheses, under the ordinary least squares (OLS) coefficients are p-values. Other notation is: R^2 is the unadjusted squared multiple correlation coefficient and DW is the conventional Durbin-Watson statistic. Q is the Ljung-Box (1978) Q-statistic for testing residual serial correlation, asymptotically distributed as a $\chi^2(36)$ on the null of no autocorrelation; RESET is (an F-version of) Ramsey's (1969) test of functional form (using the square of the fitted values), and has an asymptotic F distribution on the null of no misspecification; J-B is the Jarque-Bera (1980) test for normality of the regression residuals, distributed as a $\chi^2(2)$ under the null hypothesis of normality; ARCH is Robert Engle's (1982) Autoregressive Conditional Heteroskedasticity (ARCH) test, distributed as a $\chi^2(1)$ on the null of no ARCH; CHOW is (an F-version of) Gregory Chow's (1960) test for parameter constancy over the 1960:1 – 1982:10 and 1982:11 – 2006:1 sample periods, and has an asymptotic F distribution on the null of parameter constancy.

The estimates seem reasonable by conventional standards. The coefficient on real income is statistically significant, has the correct sign (i.e., $\beta_1 > 0$), and is of reasonable magnitude. The coefficient on the opportunity cost variable is also as the theory implies (i.e., $\beta_2 < 0$), although it is not always statistically significant. The test statistics, however, give indication of model misspecification. In particular the Q-statistic indicates significant residual serial correlation and this is consistent with the DW statistic. The Chow test statistic reveals parameter non-constancy of the regression model and the RESET and J-B statistics indicate model misspecification.

Table 10.1. Long-Run Money Demand Functions

Aggregate	Coefficients estimates and p-values (in parentheses): Constant	$\log Y$	R	R^2	DW	Tail areas of tests: Q	RESET	J-B	ARCH	CHOW
Sum M1	.299 (.000)	.290 (.000)	−.008 (.000)	.77	.02	.000	.000	.000	.000	.000
Divisia M1	−1.008 (.000)	.351 (.000)	−.008 (.000)	.89	.03	.000	.000	.000	.000	.000
CE M1	−.103 (.000)	.466 (.000)	−.014 (.000)	.90	.09	.000	.030	.000	.000	.000
Sum M2	.020 (.111)	.732 (.000)	−.000 (.032)	.95	.01	.000	.000	.000	.000	.000
Divisia M2	−.521 (.000)	.361 (.000)	−.005 (.000)	.83	.01	.000	.001	.000	.000	.000
CE M2	−.808 (.000)	1.096 (.000)	−.014 (.000)	.90	.07	.000	.000	.000	.000	.000
Sum M3	−.393 (.000)	1.012 (.000)	.000 (.860)	.95	.01	.000	.000	.000	.000	.000
Divisia M3	−.629 (.000)	.473 (.000)	−.004 (.000)	.86	.08	.000	.374	.000	.000	.000
CE M3	−1.083 (.000)	1.298 (.000)	−.018 (.000)	.90	.05	.000	.000	.001	.000	.000
Sum MZM	.188 (.000)	.626 (.000)	−.021 (.000)	.94	.11	.000	.000	.006	.000	.000
Divisia MZM	−.563 (.000)	.351 (.000)	−.018 (.000)	.84	.06	.000	.000	.762	.000	.000
CE MZM	−.634 (.000)	.985 (.000)	−.026 (.000)	.90	.09	.000	.000	.000	.000	.000

NOTES: Sample period, monthly observations, 1960:1 - 2006:1. The numbers in parentheses (next to the OLS coefficients) are p-values. Q is the Ljung-Box (1978) Q-statistic for testing serial correlation (asymptotically) distributed as a $\chi^2(36)$ on the null of no autocorrelation. The RESET statistic is the outcome of (an F-version of) Ramsey's (1969) test of functional form (using the square of the fitted values), and has an asymptotic F distribution on the null of no misspecification. J-B is the Jarque-Bera's (1980) test statistic distributed as a $\chi^2(2)$ under the null hypothesis of normality. ARCH is Engle's (1982) Autoregressive Conditional Heteroskedasticity (ARCH) χ^2 test statistic, distributed as a $\chi^2(1)$ on the null of no ARCH. The CHOW statistic is the outcome of (an F-version of) Chow's (1960) test for parameter constancy over the 1960:1 - 1982:10 and 1982:11 - 2006:1 sample periods, and has an asymptotic F distribution on the null of parameter constancy.

10.3 Money Demand Dynamics

Many of the early studies tended to ignore dynamic aspects of the money demand specification. However, the standard practice of using monthly or quarterly, instead of annual, observations and hence the need to take account of sluggish adjustment by money holders to fluctuations in the determinants of money demand, prompted a number of researchers to address this issue, most frequently by assuming that agents behave as posited by the *partial adjustment* model. This model posits the existence of a desired level of real money balances M^*/P — reflecting what real money demand would be if there were no adjustment costs — and further assumes that the actual level of money balances adjusts in each period only part of the way toward its desired level.

If the adjustment of actual to desired money holdings is in real terms, the adjustment mechanism is

$$\log\left(\frac{M_t}{P_t}\right) - \log\left(\frac{M_{t-1}}{P_{t-1}}\right) = \lambda\left[\log\left(\frac{M_t^*}{P_t}\right) - \log\left(\frac{M_{t-1}}{P_{t-1}}\right)\right], \quad (10.6)$$

where M_t/P_t denotes the actual value of real money balances and λ is a measure of the speed of adjustment, with $0 \leq \lambda \leq 1$; $\lambda = 1$ corresponds to full immediate adjustment while smaller values represent slower, more sluggish, adjustment. Implementation of the real partial adjustment model is achieved by assuming that log (M_t^*/P_t) is given by an equation of the form (10.5) and by substituting equation (10.5) into equation (10.6) to obtain the short-run demand for money function with real [e.g., Chow (1966) and Goldfeld (1973)] partial adjustment

$$\log\left(\frac{M_t}{P_t}\right) = \lambda\alpha + \lambda\beta_1 \log Y_t$$

$$+ \lambda\beta_2 R_t + (1-\lambda)\log\left(\frac{M_{t-1}}{P_{t-1}}\right) + e_t, \quad (10.7)$$

where e_t is a random error term.

The real partial adjustment model of equation (10.6), however, is not without its shortcomings. One aspect of this can be seen by rewriting (10.6) as follows

$$\log M_t - \log M_{t-1} = \lambda\left[\log\left(\frac{M_t^*}{P_t}\right) - \log\left(\frac{M_{t-1}}{P_{t-1}}\right)\right] + \Delta \log P_t.$$

As this equation shows, since the coefficient of $\Delta \log P_t$ is unity, the real partial adjustment specification presumes an immediate adjustment to changes in the price level. As this assumption is unlikely to hold, more recent research has used the so-called nominal adjustment model given by

$$\log M_t - \log M_{t-1} = \lambda \left[\log M_t^* - \log M_{t-1} \right]. \tag{10.8}$$

Implementation of the nominal adjustment model (10.8) is achieved by assuming again that $\log(M_t^*/P_t)$ is given by an equation of the form (10.5) and substituting equation (10.5) into equation (10.8) to obtain the short-run demand for money function with nominal [e.g., Goldfeld (1976)] partial adjustment

$$\log\left(\frac{M_t}{P_t}\right) = \lambda\alpha + \lambda\beta_1 \log Y_t$$

$$+ \lambda\beta_2 R_t + (1-\lambda)\log\left(\frac{M_{t-1}}{P_t}\right) + v_t, \tag{10.9}$$

where v_t is a stochastic disturbance term. A number of empirical tests suggest that the nominal model is to be preferred.

A final attack on the real and nominal partial adjustment models involves a more general reconsideration of the adjustment process. In particular, if we assume that the monetary authorities exogenously fix the nominal money supply, then the desired nominal stock of money must adjust to the given stock, presumably by adjustments in the price level. A particularly simple version of this idea would replace equation (10.8) with an adjustment equation in prices as in

$$\log P_t - \log P_{t-1} = \lambda \left[\log P_t^* - \log P_{t-1} \right]. \tag{10.10}$$

Implementation of (10.10) is achieved by assuming that $\log(M_t^*/P_t)$ is given by an equation of the form (10.5) and by substituting equation (10.10) into equation (10.5) to obtain the short-run demand for money function with price [e.g., Robert Gordon (1984)] adjustment

$$\log\left(\frac{M_t}{P_t}\right) = \lambda\alpha + \lambda\beta_1 \log Y_t$$

$$+ \lambda\beta_2 R_t + (1-\lambda)\log\left(\frac{M_t}{P_{t-1}}\right) + \zeta_t, \tag{10.11}$$

where ζ_t is a stochastic disturbance term.

Equations (10.7), (10.9), and (10.11) differ in the lagged money term. In equation (10.7), which is the real adjustment specification, the lagged dependent variable is M_{t-1}/P_{t-1}, whereas in equation (10.9), which is the nominal adjustment specification, the lagged dependent variable is M_{t-1}/P_t, and in equation (10.11), which is the price adjustment specification, the lagged dependent variable is M_t/P_{t-1}.

Because specifications (10.7), (10.9), and (10.11) are not nested hypotheses, each should be evaluated for its stability and its consistency with the theory, the latter meaning that the coefficients should be correctly signed, statistically significant, and the adjustment coefficient should obey its restriction. Ordinary least squares estimates of the real, nominal, and price adjustment equations (not reported here) indicate that these changes do not repair the money demand function, since the test statistics indicate model misspecification with almost all money measures (irrespective of the method of aggregation) and for all partial adjustment specifications.

10.4 First-Difference Specifications

Another complication is introduced by the widely held belief that equations like (10.7), (10.9), and (10.11) would not remain the same with the passage of time. A simple way to take account of this is to add to each of these equations a trend term, $\lambda\beta_3 t$, where the variable is time itself and $\lambda\beta_3$ is the associated coefficient. Estimates, however, of the formulation just described give rise to highly serially correlated disturbances. One very simple way to take account of that statistical problem is to work with first-differenced data. Thus, if we add $\lambda\beta_3 t$ to each of (10.7), (10.9), and (10.11), use the same equation for period $t-1$, and subtract the latter from the former, we get

$$\Delta \log\left(\frac{M_t}{P_t}\right) = \lambda\beta_1 \Delta \log Y_t + \lambda\beta_2 \Delta R_t + (1-\lambda)\Delta \log\left(\frac{M_{t-1}}{P_{t-1}}\right) + \lambda\beta_3 + \Delta e_t, \tag{10.12}$$

for the real partial adjustment specification,

$$\Delta \log \left(\frac{M_t}{P_t} \right) = \lambda\beta_1 \Delta \log Y_t + \lambda\beta_2 \Delta R_t$$

$$+ (1-\lambda)\Delta \log \left(\frac{M_{t-1}}{P_t} \right) + \lambda\beta_3 + \Delta v_t, \qquad (10.13)$$

for the nominal partial adjustment specification, and

$$\Delta \log \left(\frac{M_t}{P_t} \right) = \lambda\beta_1 \Delta \log Y_t + \lambda\beta_2 \Delta R_t$$

$$+ (1-\lambda)\Delta \log \left(\frac{M_t}{P_{t-1}} \right) + \lambda\beta_3 + \Delta \zeta_t, \qquad (10.14)$$

for the case with price adjustment. In each of (10.12), (10.13), and (10.14), $\lambda\beta_3 = \lambda\beta_3 t - \lambda\beta_3(t-1)$.

Estimates of equations (10.12), (10.13), and (10.14), not reported here, indicate that the first-difference specifications appear to eliminate some of the autocorrelation problems in comparison with the log-levels specifications, suggesting that the econometric estimates obtained with differenced data might be more reliable. However, the first-difference specifications are not consistent with the theory, since the coefficients are not always 'correctly' signed neither are they always statistically significant.

Regarding the levels versus first-difference formulations, many researchers, under the assumption that the (log) levels of the variables are nonstationary, carried out the empirical analysis in terms of first differences of the variables. This practice, however, of first differencing to induce stationarity has recently been questioned by Engle and Clive Granger in their 1987 article "Co-Integration and Error Correction: Representation, Estimation and Testing". They argue that the traditional approach of first differencing to induce stationarity disregards potentially important equilibrium relationships among the levels of the series to which the hypotheses of economic theory are usually taken to apply.

Hence, a strategy that is consistent with recent developments in the theory of nonstationary regressors should be applied to analyze the money demand variables.

10.5 Conclusion

We have looked at the factors that have shaped the evolution of the research on the demand for money function and discussed some specific results on the demand for money in the United States. We have determined that, irrespective of how money is measured, conventional money demand functions are seriously misspecified and that recent developments in the theory of nonstationary regressors should be used to analyze aggregate money demand. These developments are the subject matter of the next two chapters.

11

Modeling Trends in the Variables of the Money Demand Function

As discussed in the previous chapter the issue of whether economic time series are nonstationary or not is important for both estimation and hypothesis testing, both of which rely on asymptotic distribution theory. Moreover, the nature of nonstationarity has important implications for the appropriate transformation to attain a stationary series as well as for the estimation of long-run relationships between nonstationary variables.

In this chapter we explore recent exciting developments in the field of applied econometrics to distinguish between two different types of nonstationary time series — those with a deterministic trend and those with a stochastic trend. We also use tools from dynamical systems theory and statistical physics to distinguish between stochastic and deterministic behavior in the money demand variables. As we shall see, such a distinction is an important part of the analysis of data on the demand for money.

11.1 Deterministic and Stochastic Trends

The money demand variables (and most economic and financial time series in general) are nonstationary and the basic statistical issue is the appropriate representation of the nature of nonstationarity. Nonstationary time series are frequently assumed to be *trend-stationary* (TS) and are detrended in empirical investigations by regressing the series on time or a function of time. However, Charles Nelson and Charles Plosser (1982) show that most economic time series are better characterized as *difference-stationary* (DS) processes rather than TS processes. As a result, differencing rather than detrending is preferable to achieve stationarity.

The issue of whether economic time series are TS or DS has also important implications for the nature and existence of business cycles. For example, according to Nelson and Plosser (1982), trend stationarity in aggregate output would be evidence for traditional (monetary or Keynesian) business cycle models — according to which output fluctuations from a variety of macroeconomic disturbances are temporary deviations from trend. On the other hand, difference stationarity in output would be providing support for the *real business cycle* (RBC) theory of economic fluctuations — according to which most disturbances to output are permanent.

To distinguish between TS and DS processes, let us start with the time series model most commonly used to describe trend stationarity

$$y_t = \mu t + \sum_{j=0}^{\infty} a_j \epsilon_{t-j}, \tag{11.1}$$

where μt describes the trend and ϵ_t is a random disturbance. If a_j approaches zero as $j \to \infty$, $\sum_{j=0}^{\infty} a_j \epsilon_{t-j}$ is a stationary stochastic process. In this case, fluctuations in y_t are temporary and y_t is called trend stationary. As a result, a one-unit shock to y, say in period 1 (i.e., $\epsilon_1 = 1$) with no further shocks (i.e., $\epsilon_t = 0$, for $t > 1$), increases the growth rate of y above its historical average for a few periods, but does not change the long-range forecast of the level of y.

The simplest time series model most commonly used to describe difference stationarity is the 'random walk with drift' which is a first-order autoregressive process with unit coefficient (also known as a *unit root* process)

$$y_t = \mu + y_{t-1} + \epsilon_t, \tag{11.2}$$

where ϵ_t is white noise with zero mean and variance σ_ϵ, and where μ is the (fixed) mean of the first differences, often called the 'drift.' When

$\mu = 0$, equation (11.2) reduces to a 'pure random walk' (i.e., random walk with no drift), $y_t = y_{t-1} + \epsilon_t$.

The conditional mean of y_t is

$$E(y_t | y_0) = y_0 + \mu t,$$

which increases or decreases without limit as t increases. It is for this reason that a random walk with drift model is also known as a model of *stochastic trend*, because the trend is driven by stochastic shocks. Also, the 'conditional variance' of y_t is

$$\begin{aligned} \text{var}(y_t | y_0) &= E\left[y_t - E(y_t | y_0)\right]^2 \\ &= E\left[\epsilon_t + \epsilon_{t-1} + ... + \epsilon_1\right]^2 \\ &= t\sigma_\epsilon^2, \end{aligned}$$

with $\lim_{t\to\infty} t\sigma_\epsilon^2 = \infty$. Hence, the conditional variance of a random walk increases without limit, rather than converging to some finite unconditional variance. In fact, the unconditional mean and variance of a random walk do not exist. Hence, the random walk is also nonstationary.

Fluctuations in a random walk are permanent in the following sense. Accumulating changes in y from some initial value y_0 at time 0, we get from (11.2)

$$y_t = y_0 + \mu t + \sum_{j=1}^{t} \epsilon_j, \tag{11.3}$$

which has the same form as equation (11.1), but is fundamentally different from (11.1). In particular, the intercept in (11.3) is not a fixed parameter but rather depends on the initial value y_0. Also, equation (11.3) implies that a one-unit shock to y, say in period 1 (i.e., $\epsilon_1 = 1$) with no further shocks (i.e., $\epsilon_t = 0$, for $t > 1$), will forever increase y by one unit. Hence fluctuations in a random walk are permanent. Nelson and Plosser (1982) refer to processes like (11.2) as DS processes. Such processes are also known as *integrated* with an 'order of integration' of one, denoted $I(1)$, meaning that they need to be differenced once to yield a stationary series — a stationary series is said to be integrated of order zero, $I(0)$.

We have distinguished between two different types of nonstationary time series — those with a deterministic trend and those with a stochastic trend. We have also argued that time series with a deterministic trend can be transformed into stationary series by removing the deterministic trend and those with a stochastic trend by first-differencing. Obviously, subtracting a deterministic trend from a DS process or first-differencing a TS process will result in serious misspecification errors in applied work. Thus, the issue is how to distinguish between time series with and without a unit root.

11.2 Testing for Unit Roots

The literature on unit root testing is vast — see Francis Diebold and Mark Nerlove (1990), John Campbell and Pierre Perron (1991), and James Stock (1994) for selective surveys, and Walter Enders (2004, Chapter 4) for a textbook treatment. In what follows, we shall only briefly illustrate some of the issues that have arisen in the broader search for unit roots in economic and financial time series.

11.2.1 Dickey-Fuller (DF) Tests

The point of tests for unit roots is to distinguish between TS and DS processes. In the simplest case one starts with a zero-mean AR(1) process

$$y_t = \phi_1 y_{t-1} + e_t, \tag{11.4}$$

where the shock e_t is white noise. By subtracting y_{t-1} from both sides of (11.4) we obtain

$$\Delta y_t = \alpha_1 y_{t-1} + e_t, \tag{11.5}$$

where $\alpha_1 = (\phi_1 - 1)$. Notice that testing the hypothesis $\phi_1 = 1$ is equivalent to testing the hypothesis $\alpha_1 = 0$. With this in mind, in what follows we shall consider regression equations of the form (11.5).

Hence, we can test for the presence of a unit root by estimating (by ordinary least squares) the coefficients in (11.5) and using the standard t-test, labeled $\widehat{\tau}$, to test the null $H_0 : \alpha_1 = 0$. When the null is true, equation (11.5) reduces to

$$\Delta y_t = e_t,$$

so that y_t is a pure random walk and thus nonstationary.

In the context of (11.5) we test the null hypothesis of a pure random walk against the alternative hypothesis of a zero-mean, covariance-stationary AR(1) process. However, given that economic time series rarely have zero mean, we should allow for a nonzero mean, α_0, under the alternative hypothesis. In this case, we can estimate the regression equation

$$\Delta y_t = \alpha_0 + \alpha_1 y_{t-1} + e_t, \tag{11.6}$$

and use the standard t-test, now labeled $\widehat{\tau}_\mu$ (since a nonzero mean is allowed), to test the null $H_0 : \alpha_1 = 0$. Under the null hypothesis of a unit root, equation (11.6) reduces to

$$\Delta y_t = \alpha_0 + e_t,$$

so that y_t is a random walk with drift and thus nonstationary.

Finally, we can also allow for a deterministic trend under the alternative hypothesis, by estimating the regression equation

$$\Delta y_t = \alpha_0 + \alpha_1 y_{t-1} + \alpha_2 t + e_t, \tag{11.7}$$

and using the standard t-test, in this case labeled $\widehat{\tau}_\tau$ (since a linear trend is allowed), we can test the null hypothesis $H_0 : \alpha_1 = 0$.

A key result is that, under the null hypothesis $H_0 : \alpha_1 = 0$, in each of (11.5), (11.6), and (11.7) the y_t sequence is generated by a nonstationary process and the t-statistic for testing $\alpha_1 = 0$, $\widehat{\tau}$, $\widehat{\tau}_\mu$, and $\widehat{\tau}_\tau$, respectively, does not have the usual t-distribution. This problem has been solved by David Dickey and Wayne Fuller (1979, 1981) who devised special distributions, now called the 'Dickey-Fuller distributions' — see Fuller (1976) for the Dickey-Fuller tables.

11.2.2 Augmented Dickey-Fuller (ADF) Tests

The Dickey-Fuller test can also be generalized to allow for higher-order autoregressive dynamics, in case that an AR(1) process is inadequate to render e_t white noise. Consider, for example, the zero-mean AR(p) process

$$y_t = \sum_{j=1}^{p} \alpha_j y_{t-j} + e_t, \tag{11.8}$$

which can be written as[1]

[1] To obtain (11.9), first subtract y_{t-1} from both sides of (11.8) to obtain

$$\Delta y_t = \alpha y_{t-1} + \sum_{j=1}^{k} c_j \Delta y_{t-j} + e_t, \tag{11.9}$$

where $k = p - 1$ and

$$\alpha = -\left(1 - \sum_{j=1}^{p} \alpha_j\right) \quad \text{and} \quad c_j = -\sum_{i=j+1}^{p} \alpha_i. \tag{11.10}$$

Of course, depending on whether a zero mean, a nonzero mean, or a linear trend is allowed under the alternative hypothesis, one can use either (11.9) or

$$\Delta y_t = \alpha_0 + \alpha y_{t-1} + \sum_{j=1}^{k} c_j \Delta y_{t-j} + e_t, \tag{11.11}$$

or

$$\Delta y_t = \alpha_0 + \alpha y_{t-1} + \beta t + \sum_{j=1}^{k} c_j \Delta y_{t-j} + e_t. \tag{11.12}$$

The same $\widehat{\tau}$, $\widehat{\tau}_\mu$, and $\widehat{\tau}_\tau$ Dickey-Fuller statistics are used to test the null that $\alpha = 1$ in each of (11.9), (11.11), and (11.12), respectively. The k extra regressors are added to eliminate possible nuisance parameter dependencies of the test statistic caused by temporal dependencies in the disturbances. The optimal lag length, k, can be chosen using data-dependent methods that have desirable statistical properties when applied to unit root tests.

Based on such ADF unit root tests, Nelson and Plosser (1982) argue that most macroeconomic and financial time series are better characterized as DS processes rather than TS processes. As a result, differencing rather than detrending is usually necessary to achieve stationarity.

$$\Delta y_t = -(1 - \alpha_1) y_{t-1} + \sum_{j=2}^{p} \alpha_j y_{t-j} + e_t.$$

Then add and subtract $\alpha_p y_{t-p+1}$ to obtain

$$\Delta y_t = -(1 - \alpha_1) y_{t-1} + \alpha_2 y_{t-2} + ... + (\alpha_{p-1} + \alpha_p) y_{t-p+1} - \alpha_p \Delta y_{t-p+1} + e_t.$$

Next, add and subtract $(\alpha_{p-1} + \alpha_p) y_{t-p+2}$. Continuing in this fashion, we obtain (11.9), with α and c_j $(j = 1, \cdots, p-1)$ defined as in (11.10).

11.2.3 Breaking Trend Functions

Perron (1989), however, argues that most time series [and in particular those used by Nelson and Plosser (1982)] are trend stationary if one allows for a one-time change in the intercept or in the slope (or both) of the trend function. The postulate is that certain ‘big shocks’ do not represent a realization of the underlying data generation mechanism of the series under consideration and that the null should be tested against the trend-stationary alternative by allowing, under both the null and the alternative hypotheses, for the presence of a one-time break (at a known point in time) in the intercept or in the slope (or both) of the trend function.

In particular, Perron (1989) uses the following modification to the ADF regression

$$y_t = \mu + \theta DU_t + \beta t + \gamma DT_t + \delta D(T_B)_t + \alpha y_{t-1}$$

$$+ \sum_{j=1}^{k} c_j \Delta y_{t-j} + e_t, \tag{11.13}$$

where $DU_t = 1$ and $DT_t = t$ if $t > T_B$ and 0 otherwise, and $D(T_B)_t = 1$ if $t = T_B + 1$ and 0 otherwise. T_B (with $1 < T_B < T$, where T is the sample size) denotes the time at which the change in the trend function occurs.[2] In this framework, testing the null hypothesis of a unit root amounts to comparing the t statistic for testing (taking the break fraction, or break point, $\lambda = T_B/T$, to be exogenous) $\alpha = 1$, $t_\alpha(\lambda)$, with the critical values tabulated by Perron over different values of λ. In particular, reject the null hypothesis of a unit root if $t_\alpha(\lambda) < \tau(\lambda)$, where $\tau(\lambda)$ denotes the critical value from the asymptotic distribution of $t_\alpha(\lambda)$ for a fixed λ.

Perron’s (1989) assumption that the break point is uncorrelated with the data has been criticized, most notably by Lawrence Christiano (1992) who argues that problems associated with ‘pre-testing’ are applicable to Perron’s methodology and that the structural break should instead be treated as being correlated with the data. More recently, Eric Zivot and Donald Andrews (1992), Perron and Timothy Vogelsang (1992a, 1992b), and Anindya Banerjee, Robin Lumsdaine,

[2] Equation (11.13) is Perron’s (1989) regression (14), Model (C). It nests the null (i.e., $y_t = \mu_1 + y_{t-1} + e_t$ for $t \leq T_B$ and $y_t = \mu_2 + \gamma D(T_B)_t + y_{t-1} + e_t$ for $t > T_B$) and the alternative (i.e., $y_t = \mu_1 + \beta_1 t + e_t$ for $t \leq T_B$ and $y_t = \mu_2 + (\beta_1 - \beta_2)T_B + \beta_2 t + e_t$ for $t > T_B$) hypotheses.

and Stock (1992), in the spirit of Christiano (1992), treat the selection of the break point as the outcome of an estimation procedure and transform Perron's (1989) conditional (on structural change at a known point in time) unit root test into an unconditional unit root test.

The idea here is that the choice of a break point should be an explicit part of the estimation procedure, because in practice one never selects a date to test for a break point without prior information about the data. Moreover, endogenizing the break point leads to critical values that are much more conservative that Perron's (1989) ones. For example, the Zivot and Andrews (1992) estimation procedure involves using regression (11.13) without the dummy variable $D(T_B)_t$,

$$y_t = \mu + \theta DU_t + \beta t + \gamma DT_t + \alpha y_{t-1} + \sum_{j=1}^{k} c_j \Delta y_{t-j} + e_t,$$

and choosing λ to minimize the one-sided t-statistic for testing $\alpha = 1$, over all $T - 2$ regressions.

In general, existing empirical evidence indicates that the unit root hypothesis could be rejected if allowance is made for the possibility of a one-time break in the intercept or in the slope (or both) of the trend function, irrespective of whether the break point is estimated or fixed. Hence, whether the unit root model is rejected or not depends on how big shocks are treated. If big shocks are treated like any other shock, then ADF unit root testing procedures are appropriate and the unit root null hypothesis cannot (in general) be rejected. If, however, they are treated differently, then Perron-type procedures are appropriate and the null hypothesis of a unit root will most likely be rejected.

11.3 Testing for Stationarity

It is important to note that in the tests that we have discussed so far the unit root is the null hypothesis to be tested and that the way in which classical hypothesis testing is carried out ensures that the null hypothesis is accepted unless there is strong evidence against it. In fact, Denis Kwiatkowski, Peter Phillips, Peter Schmidt, and Yoncheol Shin (1992) argue that such unit root tests fail to reject a unit root because they have low power against relevant alternatives (meaning that they do not reject the hypothesis $\alpha = 1$, but they do not reject the hypothesis $\alpha = 0.95$ either).

Kwiatkowski *et al.* (1992) propose tests (known as the KPSS tests) of the null hypothesis of stationarity against the alternative of a unit root. They argue that such tests should complement unit root tests and that by testing both the unit root hypothesis and the stationarity hypothesis, one can distinguish series that appear to be stationary, series that appear to be integrated, and series that are not very informative about whether or not they are stationary or have a unit root.

In particular, the null hypothesis of *level stationarity* in y_t is tested by calculating the test statistic

$$\widehat{\eta}_\mu = \frac{1}{T^2} \sum_{t=1}^{T} \frac{S_t^2}{\widehat{\sigma}_k^2},$$

where $S_t = \sum_{i=1}^{t} e_i$, $t = 1, 2, \cdots, T$, e_t is the residuals from the regression of y_t on an intercept, and $\widehat{\sigma}_k$ is a consistent estimate of the long-run variance of y_t calculated, using the Whitney Newey and Kenneth West (1987) method, as

$$\widehat{\sigma}_k = \frac{1}{T} \sum_{t=1}^{T} e_t^2 + \frac{2}{T} \sum_{s=1}^{T} b(s,k) \sum_{t=s+1}^{T} e_t e_{t-s},$$

where T is the number of observations, $b(s,k) = 1 + s/(1+k)$ is a weighing function and k is the lag truncation parameter.

The null hypothesis of *trend stationarity* in y_t can also be tested by defining e_t as the residuals from the regression of y_t on an intercept and time trend (instead of as above) and calculating the $\widehat{\eta}_\tau$ test statistic as above.

11.4 Fractional Integration

Fractional integration is a generalization of integer integration, under which time series are usually presumed to be integrated of order zero or one. For example, an autoregressive moving-average process integrated of order d — denoted ARFIMA(p, d, q) — can be represented as

$$(1 - L)^d \, \theta(L) x(t) = \phi(L) u(t),$$

where $u(t)$ is an i.i.d. random variable with zero mean and variance σ_u^2, L denotes the lag operator, and $\theta(L)$ and $\phi(L)$ denote finite polynomials in the lag operator with roots outside the unit circle. For $d = 0$, the process is stationary, and the effect of a shock to $u(t)$ on $x(t+j)$ decays geometrically as j increases. For $d = 1$, the process is said to have a unit root, and the effect of a shock to $u(t)$ on $x(t+j)$ persists into the infinite future.

Fractional integration defines the polynomial in the lag operator $(1-L)^d$ for non-integer values of d. Following Hosking (1981) and Granger and Joyeaux (1980), a more general definition of $(1-L)^d$ can be derived from a power series expansion as follows

$$(1-L)^d = \sum_{k=0}^{\infty} \binom{d}{k} (-L)^k$$

$$= 1 - dL - \frac{1}{2}d(1-d)L^2 - \frac{1}{6}d(1-d)(2-d)L^3 - \cdots .$$

This power expansion can be re-expressed in terms of the gamma function as

$$(1-L)^d = \sum_{k=0}^{\infty} \frac{\Gamma(k-d)\,L^k}{\Gamma(-d)\,\Gamma(k+1)}.$$

For $-0.5 < d < 0.5$, the process $x(t)$ is stationary and invertible. For such processes, the effect of a shock $u(t)$ on $x(t+j)$ decays as j increases, but much more slowly than for a process with $d = 0$. More precisely, the autocovariance function for a fractionally integrated process decays hyperbolically, while the autocovariance function for zero-integrated processes decays geometrically. In both cases, the sign of the autocovariances has the same sign as d.

Fractionally integrated processes are often distinguished by their properties in the frequency domain, where Fourier analysis is utilized to represent a time series in terms of sine and cosine functions. In particular, for a fractionally integrated process with $0 < d < 0.5$, a large portion of the variance is explained by low frequency components. The extent of the low frequency variation is so great that the spectral density at frequency zero is infinite. Analogously, for fractionally integrated processes with $-0.5 < d < 0$, a large portion of the variance is explained by high frequency components, such that the spectral density at frequency zero is zero.

11.5 Testing for Nonlinearity and Chaos

Most of the tests that we discussed so far are designed to detect *linear* structure in the data. However, as John Campbell, Andrew Lo, and Craig MacKinlay (1997, p. 467) argue,

> "many aspects of economic behavior may not be linear. Experimental evidence and casual introspection suggest that investors' attitudes towards risk and expected return are nonlinear. The terms of many financial contracts such as options and other derivative securities are nonlinear. And the strategic interactions among market participants, the process by which information is incorporated into security prices, and the dynamics of economy-wide fluctuations are all inherently nonlinear.
> This is quite a challenge, since the collection of nonlinear models is much 'larger' than the collection of linear models — after all, everything which is not linear is nonlinear. Moreover, nonlinear models are generally more difficult to analyze than linear ones, rarely producing closed-form expressions which can be easily manipulated and empirically implemented. In some cases, the only mode of analysis is computational, and this is unfamiliar territory to those of us who are accustomed to thinking analytically, intuitively, and linearly."

It is for such reasons that interest in deterministic nonlinear chaotic processes has in the recent past experienced a tremendous rate of development. Besides its obvious intellectual appeal, chaos is interesting because of its ability to generate output that mimics the output of stochastic systems, thereby offering an alternative explanation for the behavior of economic variables. Clearly then, an important research inquiry is to test for chaos in the money demand variables. In other words, we are interested in whether it is possible for the money demand variables to appear to be random but not to be really random.

Sensitive dependence on initial conditions is the most relevant property of chaos to economics and finance and its characterization in terms of Lyapunov exponents is the most satisfactory from a computable (i.e. possible to estimate) perspective. Lyapunov exponents measure average exponential divergence or convergence between trajectories that differ only in having an 'infinitesimally small' difference in their initial conditions and remain well-defined for noisy systems. A bounded system with a positive Lyapunov exponent is one operational definition of chaotic behavior. See Barnett and Serletis (2000) and Kyrtsou and Serletis (2006) for several other univariate statistical tests for independence, nonlinearity and chaos, that have been recently motivated by the mathematics of deterministic nonlinear dynamical systems.

One early method for calculating the dominant Lyapunov exponent is the one proposed by Alan Wolf, Jack Swift, Harry Swinney, and John Vastano (1985). This method, however, requires long data series and is sensitive to dynamic noise, so inflated estimates of the dominant Lyapunov exponent are obtained. Douglas Nychka, Stephen Ellner, Ronald Gallant, and Daniel McCaffrey (1992) have proposed a new method, involving the use of neural network models, to test for positivity of the dominant Lyapunov exponent. The Nychka *et al.* (1992) Lyapunov exponent estimator is a regression (or Jacobian) method, unlike the Wolf *et al.* (1985) direct method which [as William Brock and Chera Sayers (1988) have found] requires long data series and is sensitive to dynamic noise. Another very promising approach to the estimation of Lyapunov exponents [that is similar in some respects to the Nychka *et al.* (1992) approach] has also been proposed by Ramazan Gencay and Davis Dechert (1992). This involves estimating all Lyapunov exponents of an unknown dynamical system. The estimation is carried out, as in Nychka *et al.* (1992), by a multivariate feedforward network estimation technique — see Gencay and Dechert (1992) for more details.

Until recently, however, it was not possible to investigate the statistical significance of the sign of the Lyapunov exponent point estimates. Thus, it was difficult to tell whether the positive Lyapunov exponents were evidence of chaotic behavior. This problem motivated Whang and Linton (1999) and Shintani and Linton (2003, 2004) to construct the standard error for the Nychka *et al.* (1992) dominant Lyapunov exponent and provide a statistical test for chaos — see also Serletis and Shintani (2006) for another application of this approach. In what follows, we discuss the key features of the Whang and Linton (1999) and Shintani and Linton (2003, 2004) approach.

Let $\{X_t\}_{t=1}^{T}$ be a random scalar sequence generated by the following non-linear autoregressive model

$$X_t = \theta(X_{t-1}, \dots, X_{t-m}) + u_t \tag{11.14}$$

where θ: $\mathbb{R}^m \to \mathbb{R}$ is a non-linear dynamic map and $\{u_t\}_{t=1}^{T}$ is a random sequence of iid disturbances with $E(u_t) = 0$ and $E(u_t^2) = \sigma^2 < \infty$. We also assume θ to satisfy a smoothness condition, and $Z_t = (X_t, \dots, X_{t-m+1})' \in \mathbb{R}^m$ to be strictly stationary and to satisfy a class of mixing conditions — see Whang and Linton (1999) and Shintani and Linton (2003, 2004) for details regarding these conditions.

Let us express the model (11.14) in terms of a map

$$F(Z_t) = \left(\theta(X_{t-1}, \dots, X_{t-m}), X_{t-1}, ..., X_{t-m+1}\right)', \tag{11.15}$$

with $U_t = (u_t, 0, ..., 0)'$ such that

$$Z_t = F(Z_{t-1}) + U_t,$$

and let J_t be the Jacobian of the map F in (11.15) evaluated at Z_t. Then the dominant Lyapunov exponent of the system (11.14) is defined by

$$\lambda \equiv \lim_{M\to\infty} \frac{1}{2M} \ln \nu_1 \left(\mathbf{T}'_M \mathbf{T}_M\right), \quad \mathbf{T}_M = \prod_{t=1}^{M} J_{M-t} = J_{M-1} \cdot J_{M-2} \cdot \cdots \cdot J_0, \tag{11.16}$$

where $v_i(A)$ is the i-th largest eigenvalue of a matrix A. Necessary conditions for the existence of the Lyapunov exponent are available in the literature. Usually, if $\max\{\ln \nu_1 (J'_t J_t), 0\}$ has a finite first moment with respect to the distribution of Z_t, then the limit in (11.16) almost surely exists and will be a constant, irrespective of the initial condition.

To obtain the Lyapunov exponent from observational data, Eckmann and Ruelle (1985) and Eckmann *et al.* (1986) proposed a method based on nonparametric regression which is known as the Jacobian method. The basic idea of the Jacobian method is to substitute θ in the Jacobian formula by its nonparametric estimator $\widehat{\theta}$. In other words, it is the sample analogue estimator of (11.16). It should be noted that we distinguish between the 'sample size' T used for estimating the Jacobian $\widehat{J}_t$ and the 'block lenght' M which is the number of evaluation points used for estimating the Lyapunov exponent. Formally, the Lyapunov exponent estimator of λ can be obtained by

$$\widehat{\lambda}_M = \frac{1}{2M} \ln \nu_1 \left(\widehat{\mathbf{T}}'_M \widehat{\mathbf{T}}_M\right), \quad \widehat{\mathbf{T}}_M = \prod_{t=1}^{M} \widehat{J}_{M-t} = \widehat{J}_{M-1} \cdot \widehat{J}_{M-2} \cdot \cdots \cdot \widehat{J}_0, \tag{11.17}$$

where

$$\widehat{J}_t = \frac{\partial \widehat{F}(Z_t)}{\partial Z'} = \begin{bmatrix} \Delta\widehat{\theta}_{1t} & \Delta\widehat{\theta}_{2t} & \cdots & \Delta\widehat{\theta}_{m-1,t} & \Delta\widehat{\theta}_{mt} \\ 1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \end{bmatrix}, \tag{11.18}$$

for $t = 0, 1, ..., M-1$, and $\Delta\widehat{\theta}_{jt} = D^{e_j}\widehat{\theta}(Z_t)$ for $j = 1, ..., m$ in which $e_j = (0, ..., 1, ..., 0)' \in \mathbb{R}^m$ denotes the j-th elementary vector.

In principle, any nonparametric derivative estimator $D^{e_j}\widehat{\theta}$ can be used for the Jacobian method. However, in practice, the Jacobian method based on the neural network estimation first proposed by Nychka *et al.* (1992) and Gençay and Dechert (1992) is the most widely used method in recent empirical analyses in economics. The neural network estimator $\widehat{\theta}$ can be obtained by minimizing the least square criterion

$$S_T(\theta_T) = \frac{1}{T}\sum_{t=1}^{T}\frac{1}{2}\left(X_t - \theta_T(Z_{t-1})\right)^2,$$

where the neural network sieve $\theta_T : \mathbb{R}^m \to \mathbb{R}$ is an approximation function defined by

$$\theta_T(z) = \beta_0 + \sum_{j=1}^{k}\beta_j\psi\left(a_j' z + b_j\right),$$

where ψ is an activation function and k is the number of hidden units. For the neural network estimation, we use the FUNFITS program developed by Nychka *et al.* (1996). As an activation function ψ, this program uses a type of sigmoid function

$$\psi(u) = \frac{u(1+|u/2|)}{2+|u|+u^2/2},$$

which was also employed by Nychka *et al.* (1992). The number of hidden units (k) will be determined by minimizing the generalized cross validation (GCV) criterion defined by

$$GCV(m,k) = \frac{\widehat{\sigma}^2}{\left(1 - \frac{2}{T}\left[1 + k(m+2)\right]\right)^2},$$

where $\widehat{\sigma}^2 = T^{-1}\sum_{t=1}^{T}\left(X_t - \widehat{\theta}(X_{t-1},\ldots,X_{t-m})\right)^2$. Notice that the GCV criterion is closely related to the BIC criterion. This particular type of GCV has been recommended by Nychka *et al.* (1996).

Using the argument in Whang and Linton (1999), Shintani and Linton (2003, 2004) showed that under some reasonable condition, the neural network estimator $\widehat{\lambda}_M$ is asymptotically normal and its standard error can be obtained using

$$\widehat{\Phi} = \sum_{j=-M+1}^{M-1} w(j/S_M)\widehat{\gamma}(j) \text{ and } \widehat{\gamma}(j) = \frac{1}{M}\sum_{t=|j|+1}^{M}\widehat{\eta}_t\widehat{\eta}_{t-|j|},$$

where

$$\widehat{\eta}_t = \widehat{\xi}_t - \widehat{\lambda}_M,$$

with

$$\widehat{\xi}_t = \frac{1}{2} \ln \left(\frac{\nu_1 \left(\widehat{\mathbf{T}}_t' \widehat{\mathbf{T}}_t \right)}{\nu_1 \left(\widehat{\mathbf{T}}_{t-1}' \widehat{\mathbf{T}}_{t-1} \right)} \right) \text{ for } t \geq 2 \text{ and } \widehat{\xi}_1 = \frac{1}{2} \ln \nu_1 \left(\widehat{\mathbf{T}}_1' \widehat{\mathbf{T}}_1 \right).$$

Above, $\omega(\cdot)$ and S_M denote a kernel function and a lag truncation parameter, respectively. Note that the standard error is essentially the heteroskedasticity and autocorrelation covariance estimator of Andrews (1991) applied to $\widehat{\eta}_t$. We employ the QS kernel for $\omega(\cdot)$ with S_M selected by the optimal bandwidth selection method recommended in Andrews (1991).

It should also be noted that before conducting nonlinear analysis the data must be rendered stationary, delinearized (by replacing the stationary data with residuals from an autoregression of the data) and transformed (if necessary). Also, since the interest is in nonlinear dependence, one should remove any linear dependence in the stationary data by fitting the best possible linear model. In particular, one can prefilter the stationary series by the following autoregression

$$z_t = b_0 + \sum_{j=1}^{q} b_j z_{t-j} + \varepsilon_t, \quad \varepsilon_t |\, I_{t-1} \sim N(0, w_0),$$

using for each series the number of lags, q, for which the Ljung-Box (1978) $Q(36)$ statistic, for example, is not significant at the 5% level.

Finally, since the interest is in deterministic nonlinear dependence, one should remove any stochastic nonlinear dependence by fitting a GARCH model with the same AR structure as the one determined above, using the $Q(36)$ statistic. In particular, one can estimate the following GARCH (1,1) model

$$z_t = b_0 + \sum_{j=1}^{q} b_j z_{t-j} + \varepsilon_t, \quad \varepsilon_t |\, I_{t-1} \sim N(0, \sigma_t^2),$$

$$\sigma_t^2 = w_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2,$$

where $N(0, \sigma_t^2)$ represents the normal distribution with mean zero and variance σ_t^2. Lyapunov exponent estimates can then be calculated for the standardized GARCH (1,1) residuals, $\widehat{\varepsilon}_t / \widehat{\sigma}_t$ — see Serletis (1995) or Serletis and Ioannis Andreadis (2000) for more details regarding these issues.

In general, it has been proven difficult to produce reliable evidence regarding the existence of chaotic processes in the money demand variables, or in macroeconomic and financial variables in general. Recently, however, Serletis and Shintani (2006) have looked at Canadian and U.S. simple-sum, Divisia, and currency equivalent money and velocity measures (a total of 54 variables) to investigate their dynamic structure and to address disputes about their chaoticity, using the Whang and Linton (1999) and Shintani and Linton (2003, 2004) approach. They have found statistically significant evidence against low-dimensional chaos. In fact, as Barnett (2006) put it, the Serletis and Shintani (2006) paper

> "is important, since it resolves some of the problems associated with a long standing controversy. In fact the paper is close to being the "last word" on the subject."

11.6 Detecting Signatures of Self-Organization

Another type of nonlinear process is self-organized criticality, recently discovered in physics by Bak *et al.* (1987). Unlike chaos, self-organized criticality is a probabilistic process. It incorporates a dominant long-run trend toward greater sensitivity and a short-run catastrophic element, which is triggered by random shocks within the system. Moreover, self-organized criticality produces fractal patterns that are generically similar, unlike chaotic fractals that differ substantially among systems. While self-organized criticality has not received much attention in economics and finance, it has been extensively investigated in the hard sciences.

As Bak *et al.* (1988, p. 364) put it "the temporal 'fingerprint' of the self-organized critical state is the presence of flicker noise or $1/f$ noise." A statistical physics approach — namely 'detrended fluctuation analysis' (DFA), introduced by Peng *et al.* (1994) — has recently been used by Serletis, Uritskaya, and Uritsky (2007) to investigate the fractal structure of some of the money demand variables. DFA is a sensitive statistical tool for detecting multiscale autocorrelations in various types of data, including financial, geophysical, and physiological signals. The main advantage of this method consists in its ability to distinguish intrinsic autocorrelations associated with memory effects in the underlying dynamical system from those imposed by external nonstationary trends.

Consider the time series $X(t)$ with $t = 1, \cdots, N$. The first step of the DFA technique consists of creating the following integrated signal (a running sum of the $X(t)$ fluctuations)

$$y(k) = \sum_{t=1}^{k} \Big[X(t) - \langle x \rangle \Big],$$

where $\langle x \rangle$ is the average value of the series $x(t)$ and $k = 1, \cdots, N$. If the returns are completely uncorrelated, this integrated signal is a random walk — a self-affine stochastic process described by the coastline fractal dimension of 1.5 — see Mandelbrot (1982). In the presence of long-range correlations, however, this signal should have another fractal dimension or exhibit deviations from fractality.

To investigate this issue, DFA of the integrated signal $y(k)$ enables us to reveal long-range correlations in $x(t)$ by getting rid of trends. In particular, the integrated series $y(k)$ is divided into $M = \lfloor N/n \rfloor$ nonoverlapping boxes (subintervals) of equal length n, where $\lfloor \cdot \rfloor$ is the floor function. The boxes are indexed by $m = 1, \cdots, M$ and their starting times are denoted as k_{nm}. For each m-th box of size n, the least squares line $y_{nm}(k)$ representing a local linear trend in that box is fit to the data. Next, the integrated series $y(k)$ is detrended by subtracting $y_{nm}(k)$, and the root mean square fluctuation of the integrated and detrended series is calculated as follows

$$F(n) = \frac{1}{M} \sum_{m=1}^{M} \sqrt{\frac{1}{N} \sum_{k=k_{nm}}^{k_{nm}+n} \Big[y(k) - y_{nm}(k) \Big]^2}.$$

This computation is repeated over all box sizes in order to characterize the relationship between the average detrended fluctuation $F(n)$ and the time scale n.

Typically, $F(n)$ will increase with the box size. A linear relationship between $F(n)$ and n on a log-log plot indicates the presence of power law (fractal) scaling — power laws indicate that there is 'scale invariance' (or 'self-similarity'), in the sense that fluctuations over small time scales are related to the fluctuations at larger time scales. Under such conditions, the slope of the line relating $\log F(n)$ to $\log n$ determines the 'scaling exponent' (or 'self-similarity parameter') α that can be used to characterize the fluctuations.

The scaling exponent α is related to the slope γ of the $1/f^{\gamma}$ power spectrum of scale-invariant fluctuations by $\gamma = 2\alpha - 1$. In particular, if $\alpha = 0.5$ (and $\gamma = 0$), the time series $x(t)$ is uncorrelated (white noise). If $\alpha = 1$ (and $\gamma = 1$), $x(t)$ corresponds to the $1/f$ noise or flicker noise. If $\alpha = 1.5$ (and $\gamma = 2$), $x(t)$ can be represented as a $1/f^2$ noise —

a random walk series (the best known nonstationary series) is exactly $1/f^2$ noise — see Li (1991).

As already noted, recently, Serletis *et al.* (2007) extend the work in Serletis and Shintani (2006) by using DFA to investigate the dynamic structure in United States money and velocity measures. In doing so, they use monthly data, over the period from 1959:1 to 2006:2, and provide a comparison between sum, Divisia, and CE money and velocity measures at each of the four levels of monetary aggregation, M1, M2, M3, and MZM — see Serletis *et al.* (2007) for more details.

11.7 Conclusion

We have argued that the first step in estimating money demand functions is to test for stochastic trends (unit roots) in the autoregressive representation of each individual time series. Moreover, since the power of unit root tests against alternative hypotheses near the null hypothesis is low, we should use alternative testing procedures (such as, for example, the KPSS level and trend stationarity tests) to deal with anomalies that arise when the data are not very informative about whether or not there is a unit root.

In a univariate time series modeling context, unit root and stationarity tests are useful in regard to the decision of whether to specify models [such as, for example, moving-average (MA) models, autoregressive (AR) models, and autoregressive moving-average (ARMA) models] in levels or first differences. If the series are stationary (i.e., there is no unit root), then it is desirable to work in levels, and if the series are integrated (i.e., there is a unit root), then differencing is appropriate.

At one time, the conventional wisdom was to generalize this idea and difference all integrated variables used in a multivariate context. Recently, however, Engle and Granger (1987) argue that the appropriate way to treat integrated variables is not so straightforward in a regression analysis. It is possible, for example, that the integrated variables cointegrate — in the sense that a linear relationship among the variables is stationary. Differencing such an already stationary relationship entails a misspecification error, which we should avoid. It is to these issues that the next chapter is devoted.

12

Cointegration and the Aggregate Demand for Money Function

We have argued that in a multivariate context with integrated variables it is important to test for cointegration (i.e., long-run equilibrium relationships). If the variables are integrated of the same order, but not cointegrated, ordinary least squares yields misleading results. In fact, Peter Phillips (1987) formally proves that a regression involving integrated variables is spurious in the absence of cointegration. In this case, the only valid relationship that can exist between the variables is in terms of their first differences.

However, if the variables are integrated and cointegrate, then there is a long-run equilibrium relationship between them. Moreover, the dynamics of the variables in the system can be described by an *error correction model* in which the short-run dynamics are influenced by the deviation from the long-run equilibrium. This is known as the 'Granger representation theorem' stating that for any set of integrated variables, cointegration and error correction are equivalent representations.

In this chapter we explore recent exciting developments in the field of applied econometrics, pertaining to the empirical analysis of models characterized by integrated and cointegrated variables. We begin with a brief review of recent theoretical developments and then discuss empirical issues in modeling and estimating aggregate money demand functions.

12.1 Cointegration

Cointegration is a relatively new statistical concept, introduced into the economics literature by Engle and Granger (1987). It is designed to deal explicitly with the analysis of the relationship between integrated series. In particular, it allows individual time series to be integrated, but requires a linear combination of the series to be stationary. Therefore, the basic idea behind cointegration is to search for a linear combination of individually integrated time series that is itself stationary.

Consider the null hypothesis that there is no cointegration between two integrated series, y_t and x_t, or equivalently, there are no shared stochastic trends (i.e., there are two distinct stochastic trends) between these series, in the terminology of Stock and Mark Watson (1988). The alternative hypothesis is that there is cointegration (or equivalently, y_t and x_t share a stochastic trend). Following Engle and Granger (1987), one can estimate the so-called *cointegrating regression* (selecting arbitrarily a normalization)

$$y_t = a + bx_t + \varepsilon_t. \tag{12.1}$$

A test of the null hypothesis of no cointegration (against the alternative of cointegration) is based on testing for a unit root in the ordinary least squares (OLS) regression residuals, $\widehat{\varepsilon}_t$, using the testing procedures discussed in Chapter 11.

12.2 Cointegration and Common Trends

The observation by Stock and Watson (1988) that cointegrated variables share common stochastic trends provides a useful way to understand long-run and/or short-run relationships. Let's follow Stock and Watson (1988) and decompose each of the y_t and x_t variables into a trend, cyclical, and stationary (but not necessarily white-noise) irregular component as follows

$$y_t = \tau_{yt} + c_{yt} + \epsilon_{yt}; \tag{12.2}$$

$$x_t = \tau_{xt} + c_{xt} + \epsilon_{xt}, \tag{12.3}$$

where τ_{jt} is the trend component of variable j at time t, c_{jt} is the cyclical component, and ϵ_{jt} is the noise (or irregular) component. If the individual series have a stochastic trend, we can explore for shared stochastic trends between the series. In particular, if the stochastic trend of x_t is shared with the y_t series (i.e., τ_{xt} is linearly related to τ_{yt}), then we have the following structure

$$y_t = \tau_{yt} + c_{yt} + \epsilon_{yt}; \tag{12.4}$$

$$x_t = \alpha\tau_{yt} + c_{xt} + \epsilon_{xt}, \tag{12.5}$$

where α is the factor of proportionality between the two trends. In this case there is a unique coefficient λ, such that the following linear combination of y_t and x_t

$$z_t = y_t - \lambda x_t$$

is a stationary series — see Engle and Granger (1987). In fact, if there is a shared stochastic trend, the linear combination z_t can be written as

$$z_t = \tau_{yt} + c_{yt} + \epsilon_{yt} - \lambda \left(\alpha\tau_{yt} + c_{xt} + \epsilon_{xt}\right)$$

$$= \tau_{yt} - \lambda\alpha\tau_{yt} + c_{yt} - \lambda c_{xt} + \epsilon_{yt} - \lambda\epsilon_{xt},$$

which for $\lambda = 1/\alpha$ reduces to

$$z_t = c_{yt} - \lambda c_{xt} + \epsilon_{yt} - \lambda\epsilon_{xt}.$$

Of course, λ may not be known a priori. Stock (1987) shows that λ can be consistently estimated using OLS in the following regression

$$y_t = \lambda x_t + z_t$$

The test for a common stochastic trend is therefore a cointegration test.

12.3 Cointegration and Common Cycles

Regarding common cycles, the approach adopted in the business cycle literature is a modern counterpart of the methods developed by Burns and Mitchell (1946). It involves the measurement of the degree of comovement between two series by the magnitude of the correlation coefficient, $\rho(j)$, $j \in \{0, \pm 1, \pm 2, \ldots\}$, between (stationary) cyclical deviations from trends — see Chapter 17 for more details.

Recently, however, Engle and Kozicki (1993) and Vahid and Engle (1993) suggested an alternative and more informative test for common cycles based on an extension of the common trends (cointegration) analysis in a stationary setting. They show that the presence of a cyclical component in the first difference of an integrated of order one variable implies the existence of some feature and that the test for common cycles in a set of I(1) variables is essentially a test for the existence of common features — features are data properties such as, for example, seasonality, heteroscedasticity, autoregressive conditional heteroscedasticity, and serial correlation.

Here, we follow Engle and Kozicki (1993) and consider how to test for a common feature of serial correlation. Therefore, the basic idea behind such a serial correlation (co)feature test is to determine whether a serial correlation feature is present in the first differences of a set of cointegrated I(1) variables and then to examine whether there exists a linear combination of the stationary variables that does not have the serial correlation feature. If the linear combination of the stationary variables eliminates the feature, it means that the feature is common across the stationary variables and that they were generated by similar (stationary) stochastic processes. Evidence to the contrary provides strong empirical support that the series are generated by significantly different (stationary) stochastic processes.

Suppose, for example, that in our bivariate setting the y_t and x_t series are I(1) variables and that each series has been rendered stationary by removing the stochastic trend. We can then write equations (12.2) and (12.3) as

$$\Delta y_t = c_{yt} + \epsilon_{yt}$$

$$\Delta x_t = c_{xt} + \epsilon_{xt}.$$

Assuming that the cyclical component is common across the two series, $c_{xt} = \beta c_{yt}$ where β is the factor of proportionality between the cyclical components, a linear combination between Δy_t and Δx_t can be written as

$$\Delta z_t = c_{yt} + \epsilon_{yt} - \mu \left(\beta c_{yt} + \epsilon_{xt} \right)$$

$$= c_{yt} - \mu\beta c_{yt} + \epsilon_{yt} - \mu\epsilon_{xt},$$

which for $\mu = 1/\beta$ reduces to a series made up of the noise components. The test for a common serial correlation feature is thus a test of whether there is some 'cofeature vector' $[1, \mu]$ for which Δz_t does not have the serial correlation feature.

12.4 Cointegration and Codependent Cycles

In introducing the notion of common features, Engle and Kozicki (1993) expand on the work by Engle and Granger (1987) on common trends and cointegration and provide a test for the existence of common cycles. However, as Ericsson (1993, p. 380) argues, in an early critique of the Engle and Kozicki (1993) methodology, common feature tests have some shortcomings and that

> "detecting the presence of a cofeature depends on the dating of the series. If the relative lag between the series is not correct, a test for a cofeature may fail to find a cofeature when there is one, even asymptotically."

To illustrate, suppose that the Δy_t and Δx_t series have exactly the same serial correlation cofeature but at different lags, as follows,

$$\Delta y_t = c_{yt} + \epsilon_{yt};$$

$$\Delta x_t = \beta c_{yt-k} + \epsilon_{xt}.$$

In this case, a linear combination of Δy_t and Δx_t at time t will not remove the feature even though each of the Δy_t and Δx_t series individually has the same feature. If, however, Δy_t enters the linear combination at lag k, as follows,

$$\Delta z_t = c_{yt-k} + \epsilon_{yt-k} - \mu \left(\beta c_{yt-k} + \epsilon_{xt} \right)$$

$$= c_{yt-k} - \mu\beta c_{yt-k} + \epsilon_{yt-k} - \mu\epsilon_{xt},$$

then for $\mu = 1/\beta$ the serial correlation common feature is eliminated from the Δz_t series. Vahid and Engle (1997) refer to the presence of a lagged serial correlation cofeature of this kind as a 'codependent cycle.'

A codependent cycle is not as strong a form of comovement as a common cycle. It provides, however, a stronger and more informative test of underlying comovements between a group of variables than traditional (lagged) cross-correlation analysis does.

12.5 Cointegration and Error Correction

If a cointegrating relationship is identified, for example $\widehat{\varepsilon}_t$ is integrated of order zero in (12.1), then according to the Engle and Granger (1987) representation theorem there must exist an error correction representation relating current and lagged first differences of y_t and x_t, and at least one lagged value of $\widehat{\varepsilon}_t$. In particular, in the present context of the y_t and x_t variables, the error correction model can be written as

$$\begin{aligned}\Delta y_t = {} & \alpha_1 + \alpha_y \widehat{\varepsilon}_{t-1} \\ & + \sum_{j=1}^{r} \alpha_{11}(j) \Delta y_{t-j} + \sum_{j=1}^{s} \alpha_{12}(j) \Delta x_{t-j} + \varepsilon_{yt};\end{aligned} \tag{12.6}$$

$$\begin{aligned}\Delta x_t = {} & \alpha_2 + \alpha_x \widehat{\varepsilon}_{t-1} \\ & + \sum_{j=1}^{r} \alpha_{21}(j) \Delta y_{t-j} + \sum_{j=1}^{s} \alpha_{22}(j) \Delta x_{t-j} + \varepsilon_{xt},\end{aligned} \tag{12.7}$$

where $\alpha_1, \alpha_2, \alpha_y, \alpha_x, \alpha_{11}(j), \alpha_{12}(j), \alpha_{21}(j)$, and $\alpha_{22}(j)$ are all parameters, ε_{yt} and ε_{xt} are white noise disturbances, and $\widehat{\varepsilon}_{t-1}$ estimates the deviation from long-run equilibrium in period $t-1$.

The purpose of the error correction model is to focus on the short-run dynamics while making them consistent with the long-run equilibrium. In particular, the error correction model shows how y_t and x_t change in response to stochastic shocks, represented by ε_{yt} and ε_{xt}, and to the previous period's deviation from long-run equilibrium, represented by $\widehat{\varepsilon}_{t-1}$. If, for example, $\widehat{\varepsilon}_{t-1}$ is positive (so that $y_t - a - bx_t > 0$), x_t would rise and y_t would fall until long-run equilibrium is attained, when $y_t = a + bx_t$.

Notice that α_y and α_x can be interpreted as *speed of adjustment* parameters. For example, the larger is α_y, the greater the response of y_t to the previous period's deviation from long-run equilibrium. On the other hand, very small values of α_y imply that y_t is relatively unresponsive to last period's equilibrium error. In fact, for y_t to be unaffected by x_t, α_y and all the $\alpha_{12}(j)$ coefficients in (12.6) must be equal to zero. This is the empirical definition of Granger causality in cointegrated systems. In other words, the absence of Granger causality for cointegrated variables requires the additional condition that the speed of adjustment coefficient be equal to zero.

In fact, in the context of two integrated variables, y_t and x_t, that cointegrate, the causal relationship between y_t and x_t can be determined by first fitting equation (12.6) by ordinary least squares and obtaining the unrestricted sum of squared residuals, SSR_u. Then by running another regression equation under the null hypothesis that α_y and all the coefficients of the lagged values of Δx_t are zero, the restricted sum of squared residuals, SSR_r, is obtained. The statistic

$$\frac{(SSR_r - SSR_u)/(s+1)}{SSR_u/(T-r-s-2)},$$

has an asymptotic F-distribution with numerator degrees of freedom $(s+1)$ and denominator degrees of freedom $(T-r-s-2)$. T is the number of observations, r represents the number of lags of Δy_t in equation (12.6), s represents the number of lags for Δx_t, and 2 is subtracted in order to account for the constant term and the error correction term in equation (12.6).

If the null hypothesis cannot be rejected, than the conclusion is that the data do not show causality. If the null hypothesis is rejected, then the conclusion is that the data do show causality. The roles of y_t and x_t are reversed in another F-test [as in equation (12.7)] to see whether there is a feedback relationship among these series.

It is to be noted that if the two integrated variables, y_t and x_t, do not cointegrate, then the causal relationship between them can be tested as above, but with the restrictions $\alpha_y = \alpha_x = 0$ imposed.

12.6 Cointegration and Money Demand

Consider the long-run demand function for real money balances of Chapter 10, which we now write as

$$(m-p)_t = \beta_0 + \beta_1 y_t + \beta_2 r_t + \varepsilon_t, \tag{12.8}$$

where m, p, y, and r respectively denote the logs of nominal money, the price level, aggregate real income, and the nominal interest rate. The behavioral assumptions require that $\beta_1 > 0$, $\beta_2 < 0$, and that the ε_t sequence is stationary, so that any deviations from long-run money market equilibrium are temporary in nature. Hence, the theory requires the existence of a combination of the nonstationary variables $(m-p)_t$, y_t, and r_t, such as, for example,

$$\varepsilon_t = (m-p)_t - \beta_0 - \beta_1 y_t - \beta_2 r_t,$$

that is stationary.

We have argued in Chapter 11 that real money balances, real income, and the nominal interest rate are most likely integrated of order one, so that their changes are stationary. If these variables are each I(1), then it is typically true that the error ε_t will also be I(1). A stochastic trend (i.e., a unit root) in ε_t, would imply that $(m-p)_t$, y_t, and r_t deviate permanently from each other, thus invalidating model (12.8). However, stationarity in ε_t would establish (12.8) as a plausible long-run relationship, with the short-run dynamics incorporated in ε_t, usually referred to as the *equilibrium error*. Then the integrated variables $(m-p)_t$, y_t, and r_t are said to be cointegrated and equation (12.8) is referred to as the cointegrating regression, as in Engle and Granger (1987).

In matrix notation, an equilibrium money demand model requires that

$$\varepsilon_t = \boldsymbol{\beta}'\boldsymbol{X}_t = \begin{bmatrix} 1 & -\beta_0 & -\beta_1 & -\beta_2 \end{bmatrix} \begin{bmatrix} (m-p)_t \\ 1 \\ y_t \\ r_t \end{bmatrix}$$

is stationary. The vector $\boldsymbol{\beta}' = [1 \; -\beta_0 \; -\beta_1 \; -\beta_2]$ is called the *cointegrating vector* for the nonstationary stochastic process $\boldsymbol{X}_t$, corresponding to $[(m-p)_t \; 1 \; y_t \; r_t]'$. This cointegrating vector isolates (in the present context) the stationary linear combination, ε_t.

There are several important points to note about cointegration — see Enders (2004) for more details. First, the cointegrating vector $\boldsymbol{\beta}$ is not unique. In particular, if $\boldsymbol{\beta}$ is a cointegrating vector, then for any $\vartheta \neq 0$, $\vartheta\boldsymbol{\beta}$ is also a cointegrating vector. Also, cointegration requires that the nonstationary variables are integrated of the same order — if the variables are integrated of different orders, they cannot be cointegrated. Moreover, cointegration refers to a 'linear' combination of integrated variables. Theoretically, it is possible that nonlinear long-run relationships exist among a set of integrated variables. The current state of econometric practice, however, cannot test for nonlinear cointegrating long-run relationships.

Finally, if $\boldsymbol{X}_t$ has two components, then there can be at most one independent cointegrating vector. If, however, $\boldsymbol{X}_t$ contains n variables, then there may be as many as $n-1$ linearly independent cointegrating vectors — the number of cointegrating vectors is called the *cointegrating rank* of $\boldsymbol{X}_t$. To see this point, suppose that $\boldsymbol{X}_t$ is $[c_t \; i_t \; (m-p)_t \; 1 \; y_t \; r_t]$, where c_t is logged consumption and i_t is logged investment. From economic theory, we expect three long-run equilibrium relations among these variables, given by

$$\boldsymbol{\beta}'\boldsymbol{X}_t = \begin{bmatrix} 1 & 0 & 0 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 & 0 \\ 0 & 0 & 1 & -\beta_0 & -\beta_1 & -\beta_2 \end{bmatrix} \begin{bmatrix} c_t \\ i_t \\ (m-p)_t \\ 1 \\ y_t \\ r_t \end{bmatrix} = \text{stationary}.$$

The three rows of the matrix $\boldsymbol{\beta}'$ are the cointegrating vectors of the nonstationary stochastic process $\boldsymbol{X}_t$. These linearly independent cointegrating vectors isolate stationary linear combinations of the $\boldsymbol{X}_t$ process corresponding to the logarithms of the balanced-growth consumption to output ratio, investment to output ratio, and the long-run money demand function. As such the cointegrating rank of $\boldsymbol{X}_t$ is 3.

12.7 Testing for Cointegration

Several methods have been proposed in the literature to test for cointegration. For a survey on statistical issues in cointegrated systems, see Campbell and Perron (1991), Engle and Byung Yoo (1987), Jesus Gonzalo (1994), and Watson (1994). Excellent textbook treatments can also be found in Enders (2004) and Dennis Hoffman and Robert Rasche (1996). In what follows, we consider two of the most frequently used cointegration testing approaches — the Engle and Granger (1987) approach and the Soren Johansen (1988) approach.

12.7.1 The Engle-Granger Approach

The Engle and Granger (1987) approach is to select arbitrarily a normalization and regress one variable against the others to obtain the OLS regression residuals. In the context of the money demand function, this involves estimating the long-run equilibrium relationship (12.8) to obtain the residual sequence $\widehat{\varepsilon}_t$. A test of the null hypothesis of no cointegration (against the alternative of cointegration) could then be based on testing for a unit root in $\widehat{\varepsilon}_t$ using a Dickey-Fuller test.

In particular, we could estimate the following autoregression of the residuals [without an intercept and trend term, since the $\widehat{\varepsilon}_t$ sequence is a residual from a regression],

$$\Delta\widehat{\varepsilon}_t = \alpha_1\widehat{\varepsilon}_{t-1} + \xi_t, \tag{12.9}$$

and test the null hypothesis that $\alpha_1 = 0$, using critical values that reflect the fact that the $\widehat{\varepsilon}_t$ sequence is generated from a regression equation — the problem here is that we cannot use the ordinary Dickey-Fuller tables, because $\widehat{\varepsilon}_t$ is not the actual error, but an estimate of the error. If we cannot reject the null hypothesis $\alpha_1 = 0$, we can conclude that the $\widehat{\varepsilon}_t$ sequence contains a unit root, suggesting that the money demand variables are not cointegrated. If, however, we can reject the null hypothesis $\alpha_1 = 0$, we can conclude that the residual series is stationary and that the money demand variables are cointegrated.

If the residual sequence $\widehat{\xi}_t$ of (12.9) does not appear to be white noise, we could perform an augmented Dickey-Fuller test on the sequence $\widehat{\varepsilon}_t$. That is, instead of using (12.9), we could estimate the following autoregression

$$\Delta\widehat{\varepsilon}_t = \alpha_1\widehat{\varepsilon}_{t-1} + \sum_{j=1}^{k} c_j\Delta\widehat{\varepsilon}_{t-j} + \xi_t, \tag{12.10}$$

and test the null hypothesis $\alpha_1 = 0$, using simulated critical values which correctly take into account the number of variables in the cointegrating regression — see Engle and Yoo (1987) for the appropriate tables.

The Engle-Granger procedure has several important defects. One defect is that it is a 'two-stage' estimator. In the first stage we generate the residual sequence $\widehat{\varepsilon}_t$, by estimating the long-run equilibrium relationship (12.8). In the second stage we use the generated sequence $\widehat{\varepsilon}_t$ to test the null hypothesis of a unit root in the context of a regression equation of the form (12.9) or (12.10), depending on whether or not the $\widehat{\xi}_t$ sequence exhibits serial correlation. This is an undesirable feature of the procedure, since any errors introduced in the first stage are carried into the second stage.

Another defect of the Engle-Granger procedure is that (with limited amounts of data typically available in economics) the test for cointegration depends on the arbitrary normalization implicit in the selection of the ‘dependent’ variable in the regression equation. In the context, for example, of the nonstationary stochastic process $\boldsymbol{X}_t = [\log(M/P)_t \; \log Y_t \; \log R_t]'$, the long-run equilibrium regression can be estimated using either $\log(M/P)_t$, $\log Y_t$, or $\log R_t$ as the dependent variable. The problem is that it is possible to find that the variables are cointegrated using one variable as the dependent variable, but are not cointegrated using another variable. This possible ambiguity is a weakness of the test.

Moreover, in tests using three or more variables, the Engle-Granger procedure does not distinguish between the existence of one or more cointegrating vectors. As a consequence, the Engle-Granger approach is well suited for the bivariate case which can have at most one cointegrating vector. All these problems can be avoided by using Johansen’s (1988) maximum likelihood (ML) extension of the Engle and Granger (1987) cointegration approach. This approach is sufficiently flexible to account for long-run properties as well as short-run dynamics, in the context of multivariate vector autoregressive models.

12.7.2 The Johansen ML Approach

The Johansen procedure is a multivariate generalization of the ADF test. Following Johansen and Katarina Juselius (1992), let us consider the following p-dimensional vector autoregressive (VAR) model of order k

$$\boldsymbol{X}_t = \sum_{i=1}^{k} \boldsymbol{A}_i \boldsymbol{X}_{t-i} + \boldsymbol{u}_t, \tag{12.11}$$

where $\boldsymbol{X}_t$ is a $p \times 1$ vector and $\boldsymbol{u}_t$ is an independently and identically distributed p-dimensional vector of innovations with zero mean and variance matrix $\Sigma_{\boldsymbol{u}}$. In the case of the stochastic process $\boldsymbol{X}_t = [\log(M/P)_t \; \log Y_t \; \log R_t]'$, $p = 3$.

The maximum likelihood estimation and likelihood ratio test of this model has been investigated by Johansen (1988), and can be described as follows. First, letting $\Delta = 1 - L$, where L is the lag operator, Johansen and Juselius (1992) suggest writing equation (12.11) as

$$\Delta \boldsymbol{X}_t = \sum_{i=1}^{k-1} \boldsymbol{\Gamma}_i \Delta \boldsymbol{X}_{t-i} + \boldsymbol{\Pi} \boldsymbol{X}_{t-k} + \boldsymbol{u}_t, \tag{12.12}$$

where

$$\boldsymbol{\Gamma}_i = -\left(\boldsymbol{I} - \sum_{j=1}^{i} \boldsymbol{A}_j\right) \quad \text{and} \quad \boldsymbol{\Pi} = -\left(\boldsymbol{I} - \sum_{i=1}^{k} \boldsymbol{A}_i\right), \tag{12.13}$$

with the $p \times p$ 'total impact' matrix $\boldsymbol{\Pi}$ containing information about the long-run relationships between the variables in $\boldsymbol{X}_t$.[1]

In the context of (12.12), the number of distinct cointegrating vectors that exist between the p elements of $\boldsymbol{X}_t$ will be given by the rank of $\boldsymbol{\Pi}$, denoted as r. The rank of a (square) matrix is the number of linearly independent rows (columns) in the matrix and is given by the number of its 'eigenvalues' that are significantly different from zero.

To recall some linear algebra, note that for an $n \times n$ square matrix $\boldsymbol{A}$, a real number λ is an eigenvalue (or 'characteristic root') of $\boldsymbol{A}$ if the system of linear equations

$$\boldsymbol{A}\boldsymbol{z} = \lambda \boldsymbol{z},$$

(where $\boldsymbol{z}$ is an $n \times 1$ vector) has nonzero solutions $\boldsymbol{z}$ (called 'eigenvectors' or 'characteristic vectors'). The condition $\boldsymbol{A}\boldsymbol{z} = \lambda \boldsymbol{z}$ can be written as

$$(\boldsymbol{A} - \lambda \boldsymbol{I})\boldsymbol{z} = \boldsymbol{0},$$

where $\boldsymbol{I}$ is the $n \times n$ identity matrix. Hence, λ is an eigenvalue of $\boldsymbol{A}$ if and only if $\boldsymbol{A} - \lambda \boldsymbol{I}$ is not invertible, which in turn means that the determinant $|\boldsymbol{A} - \lambda \boldsymbol{I}| = \boldsymbol{0}$. Thus, we can find the eigenvalues of $\boldsymbol{A}$ by finding the values of λ that satisfy the 'characteristic equation'

$$|\boldsymbol{A} - \lambda \boldsymbol{I}| = \boldsymbol{0}.$$

In the context of (12.12), if $\boldsymbol{\Pi}$ consists of all zeros, its characteristic equation has solutions $\lambda_1 = \lambda_2 = ... = \lambda_p = 0$, and $\text{rank}(\boldsymbol{\Pi}) = 0$. In

[1] Equation (12.12) can be obtained as follows. Subtract $\boldsymbol{X}_{t-1}$ from both sides of (12.11) to get

$$\Delta \boldsymbol{X}_t = (\boldsymbol{A}_1 - \boldsymbol{I})\boldsymbol{X}_{t-1} + \sum_{i=2}^{k} \boldsymbol{A}_i \boldsymbol{X}_{t-i} + \boldsymbol{u}_t.$$

Now, add and subtract $(\boldsymbol{A}_1 - \boldsymbol{I})\boldsymbol{X}_{t-2}$ to obtain

$$\Delta \boldsymbol{X}_t = (\boldsymbol{A}_1 - \boldsymbol{I})\Delta \boldsymbol{X}_{t-1} + (\boldsymbol{A}_2 + \boldsymbol{A}_1 - \boldsymbol{I})\boldsymbol{X}_{t-2} + \sum_{i=3}^{k} \boldsymbol{A}_i \boldsymbol{X}_{t-i} + \boldsymbol{u}_t.$$

Next add and subtract $(\boldsymbol{A}_2 + \boldsymbol{A}_1 - \boldsymbol{I})\boldsymbol{X}_{t-3}$. Continuing in this fashion, we obtain (12.12).

this case, (12.12) is the usual VAR model in first differences and there are p unit roots and no cointegration — that is, all elements of $\boldsymbol{X}_t$ have unit roots and so do all linear combinations of these elements.

If all rows of $\boldsymbol{\Pi}$ are linearly independent, $\boldsymbol{\Pi}$ has full rank so that $\text{rank}(\boldsymbol{\Pi}) = p$, and the vector process is stationary — that is, all elements of $\boldsymbol{X}_t$ (as well as any linear combination of these elements) will be stationary. In the more interesting case when $0 < \text{rank}(\boldsymbol{\Pi}) = r < p$, there are r cointegrating relations among the elements of $\boldsymbol{X}_t$ and $p - r$ common stochastic trends.

Johansen proposes two tests for the number of distinct cointegrating vectors — the *trace* and *maximum eigenvalue* tests. In the trace test, the null hypothesis that there are at most r cointegrating vectors is tested (against a general alternative) by calculating the test statistic

$$\lambda_{\text{trace}}(r) = -T \sum_{i=r+1}^{p} \log(1 - \widehat{\lambda}_i),$$

where $\widehat{\lambda}_i$ $(i = 1, ..., p)$ are the estimated eigenvalues, obtained from the estimated $\boldsymbol{\Pi}$ matrix. If the variables do not cointegrate, $\text{rank}(\boldsymbol{\Pi}) = 0$, and the characteristic equation of $\boldsymbol{\Pi}$ has solutions $\lambda_1 = \lambda_2 = ... = \lambda_p = 0$. In this case, each $\log(1 - \widehat{\lambda}_i)$ will equal zero (since $\log 1 = 0$), and λ_{trace} equals zero. However, the farther the estimated eigenvalues are from zero, the more negative is each of the expressions $\log(1 - \widehat{\lambda}_i)$, and the larger the λ_{trace} statistic.

In the maximum eigenvalue test, the null hypothesis of r cointegrating vectors is tested against the alternative of $r + 1$ cointegrating vectors by calculating the test statistic

$$\lambda_{\max}(r, r+1) = -T \log(1 - \widehat{\lambda}_{r+1}).$$

Again, if the estimated eigenvalue, $\widehat{\lambda}_{r+1}$, is close to zero, $\lambda_{\max}$ will be small, and the null hypothesis that the number of cointegrating vectors is r will not be rejected.

The cointegration and error-correction frameworks have proved to be successful tools in the identification and estimation of aggregate money demand functions. This type of approach to the demand for money captures the long-run equilibrium relationship between money and its determinants as well as the short-run variation and dynamics. It does so by allowing economic theory to specify the long-term equilibrium while the underlying data-generating process determines the short-term dynamics. It is in this sense that this approach represents a significant improvement over the partial adjustment specifications that we discussed in Chapter 10, which severely restrict the lag structure by relying solely on ad hoc economic theory without examining the actual data.

There is a growing literature on the application of cointegration and error-correction models to the examination of aggregate money demand functions. The earlier applications tended to be based on the Engle and Granger (1987) cointegration approach. Further research, however, suggests that we undertake the identification and estimation of aggregate money demand functions in a multivariate framework, using procedures developed by Johansen (1988) and Johansen and Juselius (1992). For an excellent textbook treatment see Hoffman and Rasche (1996), and for further references regarding the existing empirical literature on the demand for money in different countries (including developing countries), see Sriram (1999).

12.8 A Bounds Testing Approach

Although the demand for money has been investigated in a large number of recent studies taking a cointegration and error-correction approach, this approach requires the researcher to take a stance on a common order of integration for the individual variables in the money demand function. As a result, most of the literature ignores a recent important contribution to this topic by Serena Ng and Perron (1997) who show that we should be very wary of estimation and inference in 'nearly unbalanced,' 'nearly cointegrated' systems.

In this section we discuss a new econometric technique developed by Hashem Pesaran, Yongcheol Shin, and Richard Smith (2001). Their autoregressive distributed lag (ARDL), bounds test approach to testing for the existence of a single long-run relationship among a set of variables is particularly relevant as it does not require that we take a stand on the time series properties of the data. Therefore one is able to test for the existence of a long-run relationship without having to assume that the money demand variables are I(0), I(1), or even integrated of the same order.

Let us consider the existence of a single long-run relationship between the logarithm of real money balances, $(m-p)_t$, which here we will denote by s_t, and $\boldsymbol{x}_t$, where $\boldsymbol{x}_t$ is the vector time series $[\log Y_t \;\; \log R_t]$. In describing the Pesaran *et al.* (2001) methodology, we begin with an unrestricted vector autoregression

$$\boldsymbol{Z}_t = \boldsymbol{\mu}+\boldsymbol{\delta}t+\sum_{j=1}^{p}\boldsymbol{\phi}_j\boldsymbol{Z}_{t-j} + \boldsymbol{\varepsilon}_t \tag{12.14}$$

where $\boldsymbol{Z}_t = [s_t\ \boldsymbol{x}_t]'$, $\boldsymbol{\mu}$ is a vector of constant terms, $\boldsymbol{\mu} = [\mu_s\ \boldsymbol{\mu}_{\boldsymbol{x}}]'$, t is a linear time trend, $\boldsymbol{\delta} = [\delta_s\ \boldsymbol{\delta}_{\boldsymbol{x}}]'$ and $\boldsymbol{\phi}_j$ is a matrix of VAR parameters for lag j. As noted earlier, the money demand variables can be either I(0) or I(1). In this case, equation (12.14) describes a trivariate VAR.

The vector of error terms $\boldsymbol{\varepsilon}_t = [\varepsilon_{s,t}\ \boldsymbol{\varepsilon}_{\boldsymbol{x},t}]' \sim IN(\boldsymbol{0}, \boldsymbol{\Omega})$ where $\boldsymbol{\Omega}$ is positive definite and given by

$$\boldsymbol{\Omega} = \begin{bmatrix} \omega_{ss} & \omega_{s\boldsymbol{x}} \\ \omega_{s\boldsymbol{x}} & \omega_{\boldsymbol{x}\boldsymbol{x}} \end{bmatrix}.$$

Given this, $\varepsilon_{s,t}$ can be expressed in terms of $\boldsymbol{\varepsilon}_{\boldsymbol{x},t}$ as

$$\varepsilon_{s,t} = \omega\boldsymbol{\varepsilon}_{\boldsymbol{x},t} + u_t \tag{12.15}$$

where $\omega = \omega_{s\boldsymbol{x}}/\omega_{\boldsymbol{x}\boldsymbol{x}}$ and $u_t \sim IN(0, \omega_{ss})$.

Manipulation of equation (12.14) allows us to write it as a vector error correction model, as follows

$$\Delta\boldsymbol{Z}_t = \boldsymbol{\mu} + \boldsymbol{\delta}t + \lambda\boldsymbol{Z}_{t-1} + \sum_{j=1}^{p-1}\boldsymbol{\gamma}_j\Delta\boldsymbol{Z}_{t-j} + \boldsymbol{\varepsilon}_t \tag{12.16}$$

where $\Delta = 1 - L$, and

$$\boldsymbol{\gamma}_j = \begin{bmatrix} \gamma_{ss,j} & \gamma_{s\boldsymbol{x},j} \\ \gamma_{s\boldsymbol{x},j} & \gamma_{\boldsymbol{x}\boldsymbol{x},j} \end{bmatrix} = -\sum_{k=j+1}^{p}\boldsymbol{\phi}_k.$$

Here λ is the long-run multiplier matrix and is given by

$$\lambda = \begin{bmatrix} \lambda_{ss} & \lambda_{s\boldsymbol{x}} \\ \lambda_{\boldsymbol{x}s} & \lambda_{\boldsymbol{x}\boldsymbol{x}} \end{bmatrix} = -\left(\boldsymbol{I} - \sum_{j=1}^{p}\boldsymbol{\phi}_j\right),$$

where $\boldsymbol{I}$ is an identity matrix. The diagonal elements of this matrix are left unrestricted. This allows for the possibility that each of the series can be either I(0) or I(1) — for example, $\lambda_{ss} = 0$ implies that the real balances series is I(1) and $\lambda_{ss} < 0$ implies that it is I(0).

This procedure allows for the testing for the existence of a maximum of one long-run relationship that includes both $(m-p)_t$ and $\boldsymbol{x}_t$. This implies that only one of $\lambda_{\boldsymbol{x}s}$ and $\lambda_{s\boldsymbol{x}}$ can be non-zero. As our interest is on the long-run effect of real output and the nominal interest rate on real money balances, one can impose the restriction $\lambda_{\boldsymbol{x}s} = 0$. This implies that real balances have no long-run impact on real output and the nominal interest rate or that the real output and nominal interest rate series are *long-run forcing* for real money balances, in the terminology of Pesaran *et al.* (2001). Note that this does not preclude real money balances being Granger causal for the real output and nominal interest rate series in the short-run. These effects are captured through the short-run response coefficients described by the matrices $\boldsymbol{\phi}_1$ to $\boldsymbol{\phi}_p$.

Under the assumption $\lambda_{\boldsymbol{x}s} = 0$, and using (12.15), the equation for real money balances from (12.16) can be written as

$$\Delta(m-p)_t = \alpha_0 + \alpha_1 t + \varphi s_{t-1} + \boldsymbol{\psi}\boldsymbol{x}_{t-1}$$

$$+ \sum_{j=1}^{p-1} \beta_{s,j}\Delta(m-p)_{t-j}$$

$$+ \sum_{j=1}^{\mathbf{q}-1} \beta_{\boldsymbol{x},j}\Delta\boldsymbol{x}_{t-j} + \omega\Delta\boldsymbol{x}_t + u_t \qquad (12.17)$$

where $\alpha_0 = \mu_s - \omega'\mu_{\boldsymbol{x}}$, $\alpha_1 = \delta_s + \omega'\delta_{\boldsymbol{x}}$, $\varphi = \lambda_{ss}$, $\boldsymbol{\psi} = \lambda_{s\boldsymbol{x}} - \omega'\lambda_{\boldsymbol{xx}}$, $\beta_{s,j} = \gamma_{ss,j} - \omega'\boldsymbol{\gamma}_{\boldsymbol{x}s,j}$ and $\beta_{\boldsymbol{x},j} = \gamma_{s\boldsymbol{x},j} - \omega'\boldsymbol{\gamma}_{\boldsymbol{xx},j}$. This can also be interpreted as an autoregressive distributed lag [ARDL(p,$\boldsymbol{q}$)] model. One can estimate equation (12.17) by ordinary least squares and test the absence of a long-run relationship between s_t and $\boldsymbol{x}_t$, by calculating the F statistic for the null hypothesis of $\boldsymbol{\phi} = \boldsymbol{\psi} = \mathbf{0}$. Under the alternative of interest, $\boldsymbol{\phi} \neq \mathbf{0}$ and $\boldsymbol{\psi} \neq \mathbf{0}$, there is a stable long-run relationship between $(m-p)_t$ and $\boldsymbol{x}_t$, which is described by

$$(m-p)_t = \theta_0 + \theta_1 t + \theta_2\boldsymbol{x}_t + v_t$$

where $\theta_0 = -\alpha_0/\varphi$, $\theta_1 = -a_1/\varphi$, $\theta_2 = \delta/\varphi$ and v_t is a mean zero stationary process.

The distribution of the test statistic under the null depends on the order of integration of $(m-p)_t$ and $\boldsymbol{x}_t$. In the trivariate case where all variables are I(0), and the regression includes an unrestricted intercept and trend, the appropriate 95% asymptotic critical value is 4.87. When all variables are I(1) this critical value is 5.85. For cases in which one series is I(0) and the other is I(1), the 95% asymptotic critical value falls in-between these two bounds — see Pesaran *et al.* (2001, Table C1.v).

12.9 Conclusion

We have argued that cointegration provides a correct method of estimating and testing hypotheses in models characterized by long-run relations between nonstationary time series data. It avoids the spurious regression problem and indicates whether it is possible to model the integrated data in an error correction model. In particular, if the variables are integrated and cointegrate, then there is an error-correction representation that enables the estimation of long-run equilibrium relationships without simultaneously having to take a strong position on how to model short-run dynamics. If, however, the variables are integrated and do not cointegrate, then the only valid relationship that can exist between them is in terms of their first differences.

More detailed discussion of the issues raised in this chapter is best carried on in the context of a specific investigation. Such an investigation forms the subject matter of the next chapter. In particular, we examine the evidence for an equilibrium aggregate money demand function in the United States, using quarterly data and making comparisons among simple-sum, Divisia, and currency equivalent methods of monetary aggregation. We also discuss the implications of such an equilibrium relationship for the sources of shocks in the U.S. economy.

13

Balanced Growth, the Demand for Money, and Monetary Aggregation

In this chapter, building on a previous empirical study by King, Plosser, and Watson (1991), we apply the Johansen (1988) maximum likelihood approach for estimating long-run steady-state relations in multivariate vector autoregressive models, to test the implications of neoclassical stochastic growth theory and traditional money demand theory. As we argued in Chapter 12, the Johansen approach is superior to the Engle and Granger (1987) methodology, because it fully captures the underlying time series properties of the data, provides estimates of all the cointegrating relations among a given set of variables, offers a set of test statistics for the number of cointegrating vectors, and allows direct hypothesis tests on the elements of the cointegrating vectors.

Our objective is to apply the Johansen methodology to U.S. quarterly observations over the 1960:1 to 2005:4 period, and also determine whether the evidence supports certain theoretical claims in the real business cycle literature as well as in the traditional money demand literature. In doing so, we make comparisons among simple-sum, Divisia, and currency equivalent monetary aggregates (of M1, M2, M3, and MZM), to deal with the possible anomalies that arise because of different definitions of money. The monetary aggregates were obtained from the St. Louis MSI database, maintained by the Federal Reserve Bank of St. Louis as a part of the Bank's Federal Reserve Economic Database (FRED). The monetary data and the different monetary aggregation procedures will be discussed in great detail in Chapters 15-17.

13.1 Theoretical Background

Following King *et al.* (1991), let's consider the following simple real business cycle model. The single final good, Y_t, is produced via a constant-returns-to-scale Cobb-Douglas production function,

$$Y_t = \lambda_t K_t^{1-\theta} L_t^{\theta}, \tag{13.1}$$

where K_t is the predetermined capital stock, chosen in period $t-1$, and L_t is labor input in period t. Total factor productivity, λ_t, follows a logarithmic random walk

$$\log(\lambda_t) = \mu_\lambda + \log(\lambda_{t-1}) + \xi_t, \tag{13.2}$$

where μ_λ represents the average productivity growth rate and ξ_t is an independent and identically distributed process with mean zero and variance σ^2. In equation (13.2), $\mu_\lambda + \log(\lambda_{t-1})$ represents the deterministic part of the productivity evolution and ξ_t represents the stochastic innovations (or shocks).

Under the assumption that the intertemporal elasticity of substitution in consumption is constant and independent of the level of consumption, the basic neoclassical growth model with deterministic trends implies that the two great ratios — the log output-consumption ratio and the log output-investment ratio — are constant along the steady-state growth path, since the deterministic model's steady-state common growth rate is μ_λ/θ. With stochastic trends, however, there is a common stochastic trend $\log(\lambda_t)/\theta$ with a growth rate of $(\mu_\lambda + \xi_t)/\theta$, implying that the great ratios, $c_t - y_t$ and $i_t - y_t$ become stationary stochastic processes — see King, Plosser, and Sergio Rebelo (1988) for more details.

As we argued in Chapter 12, these theoretical results can be formulated as testable hypotheses in a cointegration framework. Let $\boldsymbol{X}_t$ be the multivariate stochastic process consisting of the logarithms of real per capita consumption, investment, and output, $\boldsymbol{X}_t = [c_t, i_t, y_t]$. Each component of $\boldsymbol{X}_t$ is integrated of order one [or I(1) in the terminology of Engle and Granger (1987)] — because of the random walk nature of productivity. The balanced growth implication of the neoclassical growth model with stochastic trends is that the differences $c_t - y_t$ and $i_t - y_t$ will be I(0) variables. That is, there should be two cointegrating vectors, [1, 0, −1] and [0, 1, −1].

If $\boldsymbol{X}_t$ is augmented to include real per capita money balances, $(m-p)_t$ and the nominal interest rate, R_t, that is, if $\boldsymbol{X}_t = [c_t,\ i_t,\ (m-p)_t,\ y_t,\ R_t]$, and if $(m-p)_t$ and R_t are each integrated of order one, then according to the theory we would expect to find three cointegrating vectors — the two great ratios, [1, 0, 0, −1, 0] and [0, 1, 0, −1, 0], and the money demand relation, $[0, 0, 1, \beta_y, \beta_R]$. In fact, according to the theory we expect $\beta_y = -1$ and β_R to be small and positive. These coefficients in the cointegrating vector imply a one-to-one positive relation between real money balances and real output and a small but negative relation between real balances and the nominal rate of interest.

13.2 Univariate Tests for Unit Roots

As we argued earlier meaningful cointegration tests can only be conducted if both nominal and real variables are integrated of order one and of the same order of integration. Hence, the first step before conducting Johansen maximum likelihood cointegration tests is to test for stochastic trends (unit roots) in the autoregressive representation of each individual time series. In doing so, in what follows we use four alternative testing procedures to deal with anomalies that arise when the data are not very informative about whether or not there is a unit root.

In the first three columns of panel A of Table 13.1, we report p-values for the augmented Weighted Symmetric (WS) unit root test [see Pantula *et al.* (1994)], the augmented Dickey-Fuller (ADF) test [see Dickey and Fuller (1981)], and the nonparametric, $Z(t_{\widehat{\alpha}})$, test of Phillips (1987) and Phillips and Perron (1988). These p-values (calculated using TSP 4.5) are based on the response surface estimates given by MacKinnon (1994). As discussed in Pantula *et al.* (1994), the WS test dominates the ADF test in terms of power. Also, the $Z(t_{\widehat{\alpha}})$ test is robust to a wide variety of serial correlation and time-dependent heteroskedasticity. For the WS and ADF tests, the optimal lag length was taken to be the order selected by the Akaike information criterion (AIC) plus 2 — see Pantula *et al.* (1992) for details regarding the advantages of this rule for choosing the number of augmenting lags. The $Z(t_{\widehat{\alpha}})$ test is done with the same Dickey-Fuller regression variables, using no augmenting lags. Based on the p-values for the WS, ADF, and $Z(t_{\widehat{\alpha}})$ test statistics reported in panel A of Table 13.1, the null hypothesis of a unit root in levels cannot in general be rejected for each of the variables, except for the CE M3 monetary aggregate and investment.

Given that unit root tests have low power against relevant (trend stationary) alternatives, we also follow Kwiatkowski *et al.* (1992) and test for level and trend stationarity to distinguish between series that appear to be stationary, series that appear to be integrated, and series that are not very informative about whether or not they are stationary or have a unit root. KPSS tests for level and trend stationarity are presented in columns 4 and 5 of panel A of Table 13.1. As can be seen, the t-statistic $\widehat{\eta}_{\mu}$ that tests the null hypothesis of level stationarity is large relative to the 5% critical value of .463 given in Kwiatkowski *et al.* (1992). Also, the t-statistic $\widehat{\eta}_{\tau}$ that tests the null hypothesis of trend stationarity exceeds the 5% critical value of .146 [also given in Kwiatkowski *et al.* (1992)]. Hence, combining the results of our tests of the stationarity hypothesis with the results of our tests of the unit root hypothesis, we conclude that all the series have at least one unit root.

To test the null hypothesis of a second unit root, in panel B of Table 13.1 we test the null hypothesis of a unit root (using the WS, ADF, and $Z(t_{\widehat{\alpha}})$ tests) as well as the null hypotheses of level and trend stationarity in the first (logged) differences of the series. Clearly, all the series appear to be stationary in first differences, since the null hypothesis of a unit root is rejected and the null hypotheses of level and trend stationarity cannot be rejected. The decision of the order of integration of the series is documented in the last column of Table 13.1.

13.3 Testing the c, i, y System

In this section, we apply the Johansen and Juselius (1992) maximum likelihood cointegration tests to test the balanced growth hypothesis in a three-variable model containing the real variables, c, i, and y on a per capita basis. According to the theory, we expect two cointegrating relationships among these three I(1) variables, given by the (log) differences of consumption and output and of investment and output. These are known as the 'great ratios.'

Table 13.2 reports the results of the cointegration tests on a quarterly VAR of length 6 ($k = 6$). Two test statistics are used to test for the number of cointegrating vectors: the maximum eigenvalue ($\lambda_{\max}$)

Table 13.1. Unit Root Test Results

	A. Log Levels					B. First Differences of Log Levels					
	p-values			KPSS		*p*-values			KPSS		
Variable	WS	ADF	$Z(t_{\hat{\alpha}})$	$\hat{n}_\mu$	$\hat{n}_\tau$	WS	ADF	$Z(t_{\hat{\alpha}})$	$\hat{n}_\mu$	$\hat{n}_\tau$	Decision
Sum M1	.999	.085	.432	1.709*	.213*	.005	.016	.000	.076	.073	I(1)
Sum M2	.999	.135	.700	1.887*	.177*	.006	.021	.001	.173	.094	I(1)
Sum M3	.973	.740	.660	1.883*	.190*	.001	.002	.000	.146	.107	I(1)
Sum MZM	.996	.223	.796	1.689*	.370*	.005	.016	.000	.189	.064	I(1)
Divisia M1	.939	.290	.853	1.785*	.246*	.009	.028	.001	.083	.068	I(1)
Divisia M2	.888	.855	.801	1.758*	.178*	.000	.000	.000	.119	.112	I(1)
Divisia M3	.984	.952	.924	1.770*	.176*	.005	.010	.000	.139	.127	I(1)
Divisia MZM	.988	.982	.969	1.446*	.378*	.005	.009	.000	.239	.082	I(1)
CE M1	.611	.620	.835	1.706*	.330*	.000	.000	.000	.087	.062	I(1)
CE M2	.080	.165	.144	1.796*	.179*	.000	.000	.000	.057	.030	I(1)
CE M3	.035	.076	.143	1.784*	.157*	.000	.000	.000	.044	.029	I(1)
CE MZM	.928	.703	.642	1.665*	.368*	.001	.004	.000	.183	.038	I(1)
c	.471	.005	.295	1.919*	.188*	.000	.000	.000	.089	.033	I(1)
i	.007	.070	.052	1.757*	.172*	.000	.000	.000	.038	.031	I(1)
y	.358	.031	.222	1.930*	.079	.000	.000	.000	.044	.033	I(1)
R	.566	.526	.490	.363	.346*	.000	.000	.000	.075	.037	I(1)

Notes: Numbers in the WS, ADF, and $Z(t_{\hat{\alpha}})$ columns are tail areas of unit root tests. An asterisk (next to a t-statistic) indicates significance at the 5 percent level. The 5 percent critical values for the KPSS $\hat{n}_\mu$ and $\hat{n}_\tau$ test statistics [given in Kwiatkowski *et al.* (1992)] are .463 and .146, respectively.

and trace (λ_{trace}) test statistics. As we argued in Chapter 12, in the trace test the null hypothesis that there are at most r cointegrating vectors (where $r = 0, 1, 2$) is tested against a general alternative whereas in the maximum eigenvalue test the alternative is explicit. That is, the null hypothesis $r = 0$ is tested against the alternative $r = 1$, $r = 1$ against the alternative $r = 2$ et cetera. Based on the 90% critical values for the λ_{max} and λ_{trace} test statistics reported in Table 13.2, the hypothesis of zero cointegrating vectors cannot be rejected — notice that in this case the two test statistics give similar results regarding the number of cointegrating relations.

Hence, we conclude that the two great ratios are nonstationary, in contrast to the predictions of the balanced growth literature and the more recent real business cycle literature. These results are also in conflict with the findings by King *et al.* (1991), using quarterly U.S. data over the 1949:1 to 1988:4 period..

Table 13.2
Maximum Likelihood Cointegration
Tests for the c, i, y System

Null hypothesis	λ_{max}	90 % critical value	λ_{trace}	90 % critical value
$r = 0$	18.802	19.020	29.816	28.780
$r \leq 1$	10.981	12.980	11.013	15.750
$r \leq 2$	0.032	6.500	0.032	6.500

Notes: Sample period, quarterly data: 1960:1–2005:4.
An asterisk indicates significance of the 10% level.

13.4 Testing the $m - p$, y, R System

The next system that we test is the $m - p$, y, R system. As we argued earlier, according to theory, in this case we expect to find one cointegrating vector, $[1, \beta_y, \beta_R]$, which corresponds to the long-run money demand function. In fact, according to the theory we expect $\beta_y = -1$ and $\beta_R > 0$. That is, real balances should be positively related to income and negatively related to the nominal rate of interest.

The results of the Johansen maximum likelihood cointegration tests are reported in Table 13.3 for the twelve sum, Divisia, and CE monetary aggregates. Using the 10% critical values reported in the notes to Table 13.3, we see that the $\lambda_{\max}$ and λ_{trace} test statistics give similar results regarding the number of cointegrating relations. In fact, the results indicate that the hypothesis of zero cointegrating vectors cannot be rejected

13.5 Testing the c, i, $m - p$, y, R System

We now turn to the multivariate stochastic process, $\boldsymbol{X}_t = [c_t, i_t, (m - p)_t, y_t, R_t]$ and report results in Table 13.4, in the same fashion as those for the trivariate system, $m - p$, y, R, in Table 13.3. Using the 10% critical values reported in the notes to Table 13.4, we see that the $\lambda_{\max}$ and λ_{trace} test statistics give different results regarding the number of cointegrating relations. According to Johansen (1991) this ambiguity is due to the low power in cases when the cointegration relation is quite close to the nonstationary boundary. However, since the trace test takes account of all of the smallest eigenvalues it tends to have more power than the $\lambda_{\max}$ test.

According to the λ_{trace} test statistic, we cannot reject the null of $r = 1$ in all systems. The next step is to identify the cointegrating vector. Clearly, the evidence in support of one cointegrating relationship does not provide any direction as to which one of the three vectors expected by economic theory is picked up by the Johansen procedure. It is more likely that the one cointegrating vector is the long-run money demand function, since in the trivariate c, i, y system we did not find evidence of cointegration. Identifying the cointegrating vector is beyond the scope of this Chapter — see King *et al.* (1991) for work along these lines

13.6 Conclusion

We have looked at data consisting of the traditional simple-sum monetary aggregates, as published by the Federal Reserve Board, and Divisia and CE monetary aggregates, recently produced by the Federal Reserve Bank of St. Louis, to investigate the univariate time series properties of the different monetary aggregates and to test the predictions of the balanced growth literature and the traditional money demand literature.

Table 13.3

Maximum Likelihood Cointegration Tests for the $m - p$, y, R System

Null hypothesis	λ_{max}	λ_{trace}	λ_{max}	λ_{trace}	λ_{max}	λ_{trace}	λ_{max}	λ_{trace}
	Sum M1		Sum M2		Sum M3		Sum MZM	
$r = 0$	8.881	20.781	11.239	20.072	8.552	15.197	16.461	24.631
$r \leq 1$	8.330	11.900	7.394	8.832	3.810	6.645	8.164	8.170
$r \leq 2$	3.570	3.570	1.438	1.438	2.834	2.834	.006	.006
	Divisia M1		Divisia M2		Divisia M3		Divisia MZM	
$r = 0$	10.579	22.382	8.630	18.611	8.484	17.746	10.718	19.868
$r \leq 1$	8.409	11.803	7.163	9.981	6.616	9.262	9.147	9.149
$r \leq 2$	3.394	3.394	2.817	2.817	2.645	2.645	.002	.002
	CE M1		CE M2		CE M3		CE MZM	
$r = 0$	11.009	21.078	11.555	22.792	11.731	22.961	9.532	17.622
$r \leq 1$	9.181	10.069	9.104	11.237	8.766	11.230	7.734	8.089
$r \leq 2$	.887	.887	2.133	2.133	2.463	2.463	.355	.355

Notes: Sample period, quarterly data: 1960:1–2005:4. An asterisk indicates significance at the 10% level. The 90 percent critical values for the λ_{max} and λ_{trace} test statistics are (for $r = 0$, $r \leq 1$, and $r \leq 2$) 19.020, 12.980, and 6.500 and 28.780, 15.750, and 6.500.

Table 13.4

Maximum Likelihood Cointegration Tests for the c, i, $m-p$, y, R System

Null hypothesis	$\lambda_{\max}$	λ_{trace}	$\lambda_{\max}$	λ_{trace}	$\lambda_{\max}$	λ_{trace}	$\lambda_{\max}$	λ_{trace}
	Sum M1		Sum M2		Sum M3		Sum MZM	
$r=0$	39.098	89.752*	35.824*	86.272*	33.017	82.460*	33.449	100.144*
$r \leq 1$	21.233	50.654	20.707	50.448	22.446	49.442	27.768	66.695
$r \leq 2$	16.254	29.420	15.612	29.740	15.403	26.996	21.965	38.927
$r \leq 3$	8.002	13.165	11.978	14.128	9.652	11.592	11.261	16.962
$r \leq 4$	5.163	5.163	2.149	2.149	1.940	1.940	1.700	1.700
	Divisia M1		Divisia M2		Divisia M3		Divisia MZM	
$r=0$	40.144*	92.679*	36.753*	85.725*	34.953*	81.495*	32.030	83.983*
$r \leq 1$	21.882	52.534	20.895	48.972	20.924	46.541	26.122	51.953
$r \leq 2$	16.236	30.652	16.939	28.076	16.622	25.616	17.356	25.830
$r \leq 3$	8.645	14.415	10.815	11.137	8.878	8.994	8.370	8.474
$r \leq 4$	5.770	5.770	.321	.321	.116	.116	.103	.103
	CE M1		CE M2		CE M3		CE MZM	
$r=0$	31.216	83.909*	35.080*	84.566*	34.049	82.271*	31.962	80.349*
$r \leq 1$	21.843	52.692	23.447	49.486	23.269	48.222	22.003	48.387
$r \leq 2$	15.802	30.849	14.374	26.039	13.433	24.953	17.924	26.383
$r \leq 3$	11.846	15.046	7.070	11.664	7.718	11.520	6.306	8.459
$r \leq 4$	3.200	3.200	4.593	4.593	3.801	3.801	2.153	2.153

Notes: Sample period, quarterly data: 1960:1–2005:4. An asterisk indicates significance at the 10% level. The 90 percent critical values for the $\lambda_{\max}$ and λ_{trace} test statistics are (for $r=0$, $r \leq 1$, $r \leq 2$, $r \leq 3$, and $r \leq 4$) 34.160, 28.320, 22.260, 16.280, and 9.750 and 77.550, 55.010, 36.280, 21.230, and 9.750, respectively.

Our results, although not in line with the simple one-factor neoclassical growth model, are consistent with the evidence reported by Kunst and Neusser (1990) for Austria, Neusser (1991) for Canada, Germany, and Japan, and Serletis (1994) for Canada..

We have also established that the different monetary aggregates have different time series properties. Of course, in such cases our economic intuition is hard-pressed for explanations. In this regard, we think that our results in this chapter suggest answers to a number of questions raised over previous studies of the role of money in the economy. In fact, as Serletis and Koustas (2001, p. 137) put it

> "a meaningful comparison of alternative monetary aggregation procedures requires the discovery of the structure of preferences over monetary assets by testing for weakly separable subgroupings. Leaving aside the method of aggregating over monetary assets (i.e., Divisia as opposed to other possibilities), the problem is the *a priori* assignment of monetary assets to monetary aggregates."

We explore such issues in the rest of this book, starting with the microeconomic- and aggregation-theoretic approach to the definition of money in Part 5 of the book.

14

Cross-Country Evidence on the Demand for Money

Past estimation of money demand functions has primarily been confined to industrialized countries, especially the United States and the United Kingdom — see Goldfeld and Sichel (1990) and Sriram (1999) for surveys on past theoretical and empirical money demand studies. However, the estimates derived from the time-series approach seem to be sensitive to the choice of sample period, functional form, and the univariate and multivariate time series properties of the underlying variables. Thus, it has been difficult to draw broad conclusions about long-run money demand based on only a handful of countries, which can be argued to be similar in nature. For these reasons (among others), Friedman and Kuttner (1992, p. 490) argue that time-series data does not uncover a "close or reliable relationship between money and nonfinancial economic activity."

Recently, however, Kenny (1991), Mulligan and Sala-i-Martin (1992), Fujiki and Mulligan (1996), and Fischer (2005), have opted for an alternative modeling approach, by estimating money demand cross-sectionally. This approach allows researchers to utilize additional conditional variables, which may not be available as a time-series. For example, the Mulligan and Sala-i-Martin (1992) cross-state American study includes state specific variables for population, population density, agricultural sector's share of income, and regional dummies. Also, Fischer (2005) attempts to reconcile parameter biases in the conventional money demand estimates by conditioning on heterogeneous levels of financial sophistication in his cross-regional panel analysis of Switzerland. The findings and conclusions drawn from such studies indicate that supplementary variables can enhance standard inferences regarding money demand.

Motivated by these considerations, in this chapter we examine money demand issues using cross-country data, for 48 countries over the 1980-1995 period. In particular, we investigate conventional money demand functions, for both narrow and broad monetary aggregates, and the role that institutions, financial structure, and financial development may have in the demand for money. As Levine (2002, p. 405) puts it,

> "one advantage of the broad cross-country approach is that it permits a consistent treatment of financial system structures across countries and thereby facilitates international comparisons."

14.1 Cross-Country Data

In order to analyze the possible relationships between real money balances, real GDP, the nominal interest rate, and different institutional, financial structure, and financial development measures, we adopt the common broad cross-country approach, using one observation for each variable under consideration, per country, for 48 countries (over the 1980-1995 period). The countries we consider are the same as those investigated in Levine (2002) and are listed in Table 14.1. The institutional, financial structure, and financial development measures are also from Levine (2002) — see Levine (2002) and Serletis and Vaccaro (2006) for a detailed description of the data.

14.2 Cross-Country Specifications

Following Levine (2002), we argue that different views regarding money demand can be represented as rival predictions on the parameters of a standard money demand equation and consider the following cross-country money demand regression equations:

Table 14.1. Countries

Argentina	Kenya
Australia	Malaysia
Austria	Mexico
Belgium	Netherlands
Brazil	New Zealand
Canada	Norway
Chile	Pakistan
Colombia	Panama
Cyprus	Peru
Denmark	Philippines
Ecuador	Portugal
Egypt	South Africa
Finland	Spain
France	Sri Lanka
Germany	Sweden
Ghana	Switzerland
Greece	Taiwan, China
Honduras	Thailand
India	Trinidad and Tobago
Ireland	Turkey
Israel	Tunisia
Italy	United Kingdom
Jamaica	United States
Japan	Zimbabwe

$$\log\left(\frac{M}{P}\right) = \boldsymbol{a}'\boldsymbol{X} + \varepsilon_1;$$
$$\log\left(\frac{M}{P}\right) = \boldsymbol{a}'\boldsymbol{X} + \boldsymbol{b}'\boldsymbol{I} + \varepsilon_2;$$
$$\log\left(\frac{M}{P}\right) = \boldsymbol{a}'\boldsymbol{X} + cS + \varepsilon_3;$$
$$\log\left(\frac{M}{P}\right) = \boldsymbol{a}'\boldsymbol{X} + dF + \varepsilon_4,$$

where M is the money stock (defined by either a narrow or broad definition) and $\boldsymbol{X}$ represents the standard set of conditioning information — that is, the natural logarithm of real GDP and a short term nominal interest rate.

As in Levine (2002), $\boldsymbol{I}$ represents a vector of institutional variables which measure macroeconomic stability, openness to international trade, and political stability. S gauges financial structure, with larger values suggesting a more market-based economy and smaller values implying a bank-based economy. F measures the degree of financial development; larger measures of F imply an increased development of securities markets, banks, and non-banks. Such measures can also be interpreted as a proxy for financial services. ε_i, with $i = 1, 2, 3, 4$, is the corresponding error term for each of the four equations, respectively. $\boldsymbol{a}$, $\boldsymbol{b}$, c, and d are estimated coefficients (with bold letters indicating vectors of coefficients). $\boldsymbol{I}$, S, and F are the same variables that Levine (2002) considers as possible growth determinants.

The idea is that countries with greater institutional stability should exhibit less uncertainty and therefore display a reduced demand for money. Specifically, the sign of $\boldsymbol{b}$ will depend on each of the institutional variables under consideration. For example, a higher level of average schooling years over the population implies a stronger knowledge of the mechanics of the economy and the money market, suggesting that the demand for money will be lower as the educational index rises. Large black market premium values indicate that the transaction costs incurred while purchasing goods and services are also large, which in turn requires agents to hold more liquid money. There is also a possible relationship between government expenditure and money demand. Theory asserts that private spending and public spending maybe perfect substitutes or complements — see, for example, Barro (1997). If perfect substitutes, then the expenditure on goods and services by the government will reduce expenditure by agents, requiring them to hold less money, ceteris paribus. If complements, then providing additional services will require agents to purchase these services and compel them to retain additional funds, ceteris paribus.

The trade variable attempts to proxy the degree of openness. With enhanced trade comes exposure to different markets, where agents must now consider foreign interest rates and balance of payment issues. As a result, agents will have to divide their monetary holdings between domestic and foreign accounts. Higher degrees of openness would suggest that there would be lower demand for domestic money. Measures of civil liberties, revolutions and coups and political assassinations can be thought of as proxies for political stability. With domestic political instability comes capital flight. The theory is that as the future of the financial system becomes dismal, faith in a paper promise declines and faith in other assets such as gold and tangible goods useful for bartering increases. Kenny (1991) considers a similar approach by trying to control for the type of government by including a dummy variable for dictatorships. Our interpretation differs given that the three political stability variables are not mutually exclusive to countries considered either a dictatorship or democracy. As well, Kenny (1991) emphasizes precautionary motives for his interpretation but neglects speculative motives, which have increasingly dominated financial markets during our sample period.

Bureaucratic efficiency measures the extent of autonomy from political pressures and strength to govern. This is important because it signals a degree of competence within key governmental departments such as finance and the central bank. Given that autonomy and expertise indicate certainty and provide faith in the monetary and political system, the implication is that as the quality of the bureaucracy rises, the demand for money should decline. As with the black market premium measure, corruption can also accordingly be considered a source of raising transaction costs. It is not unreasonable to assume that an increase in corruption would be followed by bribery and possibly influence peddling. Therefore, as we observe an increase in corruption we should also observe an increase in the demand for money.

The addition of financial structure measures allows for investigation into the possible heterogeneity in money demand under diverse financial systems. Specifically, a better understanding of whether money demand is higher or lower in a bank-based or market-based system can be explored. Such analysis and its insights may be useful in formulating monetary policy to remedy a financial crisis or to restructure a command style economy to a more capital driven one, from a policy perspective. Given that, the hypothesis is that under market-based systems firms can easily raise funds in the open market for financing and investment through capital markets, which in turn would broaden loan possibilities. Boot and Thakor (1997) along with Allen and Gale (1999) articulate that competitive capital markets contribute positively in aggregating dispersed information signals and efficiently relay such information to investors, with favorable implications for firm level financing — see also Levine (2002) for a further explanation and other references on the subject matter. In comparison, under a bank-based system, funds would have to be raised through banks, therefore limiting financing possibilities. Bhide (1993) along with Boot and Thakor (1997) argue that banks act as a coordinated coalition of investors which can monitor firms more efficiently to diminish post lending moral hazard issues and a myopic investor climate. Thus, given the possibility of easily attainable funds under a market based system and the possible impediments under a bank-based system, we should observe the demand for money to be lower in economies where there are market-based characteristics and higher in economies where bank-based characteristics are observed. Hence, we should observe $c < 0$.

Financial services, whether provided by banks or capital markets, can also give broad insight into transaction costs. The idea is that financial arrangements such as contracts, markets, and intermediaries alleviate market imperfections. Levine (1997) stresses that this view curtails the significance of the bank-based and market-based discussion. The argument Levine (2002) makes is that financial arrangements (such as contracts, markets, and intermediaries) highlight prospective investment opportunities, promote corporate responsibility, contribute to risk management, develop liquidity, and reduce savings mobilization. With regards to money demand, the issue is whether such arrangements assist in lowering transaction costs or aid in increasing them. Standard economics textbooks describe financial innovations having a negative effect on the demand for real money balances — see, for example, Barro (1997). However, there is not a definitive hypothesis given that reductions in market imperfections come at a price. Ambiguity arises because the derived benefits from financial services may not outweigh the costs and vice versa. As a result, the data will have to dictate which case is more likely. If the benefits offset the costs, transaction costs decline and the implied sign is $d < 0$. Whereas, if the costs overshadow the benefits, transaction costs could rise and the implied sign is $d > 0$. Kenny (1991) presents a similar idea by using population density as a surrogate for bank proximity and their corresponding services.

14.3 Cross-Country Evidence

Table 14.2 presents the initial conventional money demand results using ordinary least squares (OLS) estimation with heteroskedasticity-consistent standard errors. The top panel displays the results for M1 as the dependent variable and the bottom panel those for M2. For both money measures, the estimated income elasticity of the demand for real money balances is highly significant and close to the quantity theory demand for money predictions. Specifically, for both aggregates we tested the null hypothesis that the income elasticity is equal to one, and cannot reject the null at the 5% level. The estimated interest elasticities of the demand for real balances are negative and both significant at the 5% level. Although the interest elasticity estimates are not zero for both aggregates, as predicted by the quantity theory demand for money, they are quite low and statistically different than the implied value of the Baumol-Tobin transactions theory.

Table 14.2. Conventional Money Demand Functions

Explanatory Variable	Coefficient	Standard error	t-statistic	p-value	R^2	RESET F
			M1			
Constant	-6.575	0.943	-6.969	0.000	0.897	1.099
Ln R	-0.108	0.037	-2.899	0.006		
Ln Y	1.012	0.040	25.240	0.000		
			M2			
Constant	-6.615	0.787	-8.396	0.000	0.940	0.110
Ln R	-0.102	0.024	-4.165	0.000		
Ln Y	1.061	0.032	32.468	0.000		

Note: The reported explanatory variables are all included in each of the regressions. The simple information set only includes the logarithm of short term interest rates and the logarithm of real GDP.

Table 14.3 presents the institution results for both money measures. The estimation procedure we opt for is to control sequentially for each institutional variable conditioned on the simple information set. The reasoning stems from issues regarding simultaneity and mutual exclusiveness. In particular, we are concerned with high correlations between the bureaucracy and corruption indexes and the small variance of the political indexes. As well, we are also apprehensive about the validity and consistency of OLS once multiple indexes measured by scale are included concurrently and when numerous degrees of freedom are lost from including multiple explanatory variables in our small sample. Although Kenny (1991) and Levine (2002) do not take the same approach, Beck and Levine (2004) do take a similar approach when investigating associations between stock market and bank development with economic growth. As a result, we are simply interested in the influential direction each of the explanatory variables has on the money measures and caution on interpreting the results as exploitable elasticities.

The results in the top panel of Table 14.3 imply that only the educational variable is significantly related to money demand when considering a narrow measure. The sign of the coefficient also theoretically

Table 14.3. Institutions, Political, Macro-Stability, and Money Demand

Explanatory Variable	Coefficient	Standard error	t-statistic	p-value	R^2	RESET F
			M1			
Ln School80	-0.395	0.161	-2.452	0.018	0.904	1.003
Ln BMP	0.095	0.136	0.698	0.489	0.895	1.027
Ln GOV	-0.086	0.310	-0.279	0.781	0.895	1.093
Ln Trade	-0.422	0.309	-1.363	0.180	0.902	0.289
Civil	0.034	0.058	0.593	0.556	0.895	1.205
REVC	-0.257	0.219	-1.174	0.247	0.896	1.105
ASSASS	-0.081	0.128	-0.632	0.530	0.895	0.892
Bureau	-0.051	0.095	-0.543	0.590	0.895	1.067
Corrupt	-0.045	0.087	-0.512	0.611	0.895	0.934
			M2			
Ln School80	0.056	0.132	0.424	0.673	0.939	0.107
Ln BMP	-0.531	0.131	-4.059	0.000	0.943	0.096
Ln GOV	0.274	0.183	1.490	0.143	0.941	0.151
Ln Trade	-0.046	0.319	-0.145	0.885	0.939	0.169
Civil	0.002	0.036	0.071	0.943	0.939	0.104
REVC	-0.228	0.163	-1.400	0.168	0.940	0.148
ASSASS	-0.178	0.103	-1.724	0.092	0.942	0.354
Bureau	0.028	0.068	0.416	0.679	0.939	0.090
Corrupt	0.061	0.054	1.130	0.264	0.940	0.020

Note: The reported explanatory variables are included one-by-one in each of the regressions. The simple information set only includes the log of the interest rate and the log of real GDP.

conforms because increases in the level of workforce education impact money demand negatively from a narrow perspective. This result is also consistent with Kenny (1991) where he also finds a negative relationship between literacy and M1. None of the other institutional indicators enter the narrow money demand regressions at the 10% level. With regards to the broader aggregate, Table 14.3 shows that the black market premium and assassination variables enter significantly. However, the sign of the black market premium coefficient is incorrect from the theoretical expectation. The negative sign on the assassination coefficient corresponds to our prediction that domestic turmoil would lead to a substitution out of money and into other tangible assets. However, given that it narrowly makes the 10% level we are still aware of potentially making a Type II error. None of the other institutional indicators enter the broad money demand regressions at the 10% level.

The implication of both the narrow and broad money regressions is that conditioning on institutions may not be so informative and unnecessary when investigating money demand issues. This follows from only one out of the nine institutional variables entering the narrow specification significantly and only two out of the nine being significant in the broad specification. As a result, it would be suspect to add any of the institutional variables to the conditioning information set. One interpretation may be that the demand for both aggregates could be stable irrespective of most institutional differences. In fact, in both specifications the elasticities with respect to income and the interest rate remain statistically similar to those in Table 14.2.

Table 14.4 presents the results when controlling for financial structure. The same estimation methods were used as in the institutional specification. Three of the structure measures enter the narrow specification significantly at the 10% level. In particular, the activity, size, and aggregate coefficients are all negative and of similar statistical magnitude, with size having the largest effect. The implication is that some measures of financial structure indicate that money demand is negatively related to market-based economies. This result corresponds to the economic theory outlined in the specification section. However, it also shows that there is some measurement sensitivity to such a conclusion. On the other hand, only the size variable is significant at the 10% level in the broad specification. This result suggests that measures of financial structure are for the most part statistically trivial when investigating money demand from a broad perspective. Again, the elasticities with respect to income and the interest rate remain statistically similar to those in Table 14.2.

Finally, Table 14.5 presents the results when conditioning on the simple information set and controlling for financial development. The elasticities with respect to income and the interest rate again remain statistically similar to those in Table 14.2 for both aggregates. Using the same estimation method as the previous specification for financial structure, the results indicate that measures of financial development do not bring forth additional information regarding narrow money

Table 14.4. Financial Structure and Money Demand

Explanatory Variable	Coefficient	Standard error	t-statistic	p-value	R^2	RESET F
		M1				
Structure-Activity	-0.170	0.091	-1.854	0.070	0.902	1.038
Structure-Size	-0.204	0.107	-1.900	0.064	0.901	1.920
Structure-Efficiency	-0.110	0.094	-1.168	0.249	0.899	1.157
Structure-Aggregate	-0.194	0.105	-1.839	0.073	0.903	1.377
Structure-Regulatory	0.014	0.031	0.461	0.647	0.895	0.929
		M2				
Structure-Activity	-0.002	0.066	-0.031	0.975	0.939	0.113
Structure-Size	-0.132	0.070	-1.872	0.068	0.942	0.117
Structure-Efficiency	0.071	0.075	0.951	0.347	0.941	0.064
Structure-Aggregate	-0.009	0.074	-0.121	0.904	0.939	0.116
Structure-Regulatory	-0.002	0.027	-0.108	0.914	0.939	0.102

Note: The reported explanatory variables are included one-by-one in each of the regressions. The simple information set only includes the log of the interest rate and the log of real GDP.

demand. None of the financial variables enter significantly at the 10% level. Conversely, in the broad specification there are intuitive results. All of the four measures of financial development enter significantly at the 10% level or higher. The sign on all of the coefficients is positive. Recall that the implied sign may be positive or negative. Given the consistent positive sign, we argue that this may suggest possible evidence that although greater financial development would bring forth additional services through financial arrangements, the benefits of such services may be outweighed by the costs and may actually raise transaction costs on a cross country scale. Kenny (1991) also finds a significantly positive estimate on the bank proximity variable in his M2 specification. Such results warrant further analysis before a definitive conclusion can be made.

Table 14.5. Financial Development and Money Demand

Explanatory Variable	Coefficient	Standard error	t-statistic	p-value	R^2	RESET F
		M1				
Finance-Activity	-0.044	0.073	-0.609	0.545	0.896	0.946
Finance-Size	0.033	0.173	0.194	0.847	0.895	1.080
Finance-Efficiency	-0.069	0.064	-1.072	0.289	0.897	0.785
Finance-Aggregate	-0.074	0.143	-0.514	0.609	0.895	0.974
		M2				
Finance-Activity	0.143	0.060	2.356	0.023	0.949	0.009
Finance-Size	0.447	0.127	3.522	0.001		0.952
Finance-Efficiency	0.110	0.057	1.917	0.062	0.944	0.022
Finance-Aggregate	0.304	0.113	2.679	0.010	0.950	0.006

Notes: The reported explanatory variables are included one-by-one in each of the regressions.The simple information set only includes the log of the interest rate and the log of real GDP.

14.4 Robustness

So far, we have followed Kenny (1991), Levine (2002), and Beck and Levine (2004) and treated countries as homogeneous units using the same regression model for all countries in the sample. Recently, Serletis and Vaccaro (2006) explored whether heterogeneity exists in our cross-country database, and in doing so, they provide an approach to overcome it. They used an automatic classification program (AutoClass) for cluster analysis, developed by researchers at the Ames Research Center — for a description of the AutoClass program, see Stutz and Cheeseman (1996) or Serletis (2007) for a recent application in the context of monetary aggregation.

They have shown that the assumption that all of the countries can be treated as a homogeneous unit can cause systematic distortions. Specifically, they used unsupervised Bayesian methods based on finite mixture models and mathematical properties, to cluster the data set into two distinct groups. Regressions based on each of the partitioned data sets displayed heterogeneity with respect to the influence institutions, financial structure, and financial development have on money demand, for each of the two groups. They found that the developing, high-inflation class somewhat dominated the data set and distorted some of the developed, low-inflation class results. In particular, the role that the supplementary variables have in the money demand function depends not only on the specified aggregate, but also on the countries specified in the sample — see Serletis and Vaccaro (2006) for more details.

14.5 Conclusion

In this chapter we used cross-country data (for 48 countries, over the 1980-1995 period) to investigate the long-run relationship between both narrow and broad monetary aggregates and interest rates, real GDP, institutions, financial structure, and financial development. We have shown that the interest and income elasticities of real balances are fairly stable and conform to the theoretical prediction of the quantity theory demand for money. As well, we have found that institutions, financial structure, and development do play a role in the demand for money in an aggregate setting; albeit a limited role.

Part 5: Microfoundations and Monetary Aggregation

Overview of Part 5

Chapters 15 and 16 provide the microeconomic foundations to the problem of monetary aggregation. Most of the material here is by now well established in the monetary literature. In fact, the manifest advantages of the microfoundations approach have been laid out carefully by Barnett, Douglas Fisher, and Serletis (1992) and Barnett and Serletis (2000).

Chapter 17 investigates the cyclical behavior of the monetary variables, using the methodology suggested by Kydland and Prescott (1990). In doing so, comparisons are made among simple-sum, Divisia, and CE monetary aggregates (of M1, M2, M3, and MZM) using the data set that we discussed in Chapters 10 and 13.

15

The Microeconomic Foundations of the Definition of Money

In the discussion to this point, we have used the word 'money' as though it were obvious what it means, but this is not the case. Currently, the common practice among central banks is to construct money measures from a list of possible components by simply adding together those that are considered to be the likely sources of monetary services. These are usually highly liquid financial assets, and the approach is referred to in the literature as that of *simple-sum* aggregation.

In recent years, however, such a monetary aggregation procedure has been questioned and explicit attention has been focused on the rigorous use of microeconomic- and aggregation-theoretic foundations in the construction of monetary aggregates. In this chapter we provide a brief qualitative assessment of the relative merits of the conventional (summation) versus the new approach to monetary aggregation. In doing so, we follow closely the presentation of Barnett, Fisher, and Serletis (1992).

15.1 The Simple-Sum Index

As already suggested, the monetary aggregates currently in use by most central banks around the world are simple-sum indexes in which all monetary components are assigned a constant and equal (unitary) weight. This index is M in

$$M = \sum_{j=1}^{n} x_j$$

where x_j is one of the n monetary components of the monetary aggregate M. This summation index views all components as dollar-for-dollar perfect substitutes. There is no question that such an index represents an index of the stock of nominal monetary wealth, but it is a special case, at best, of the appropriate type of index for monetary services, as we will see.

Friedman and Schwartz (1970, pp. 151–152) dismissed simple-sum monetary aggregates when discussing the potential generalization of the simple-sum aggregates to index numbers

> "this (summation) procedure is a very special case of the more general approach. In brief, the general approach consists of regarding each asset as a joint product having different degrees of *moneyness*, and defining the quantity of money as the weighted sum of the aggregate value of all assets, the weights for individual assets varying from zero to unity with a weight of unity assigned to that asset or assets regarded as having the largest quantity of "moneyness" per dollar of aggregate value. The procedure we have followed implies that all weights are either zero or unity. The more general approach has been suggested frequently but experimented with only occasionally. We conjecture that this approach deserves and will get much more attention than it has so far received."

Over the years, there has been a steady stream of attempts at properly weighting the monetary components within a simple-sum aggregate. Without theory, however, any weighting scheme is questionable. Chetty (1969) appears to have been the first to recognize the direct relevancy of microeconomic aggregation theory to monetary aggregation, since he was the first to produce a structure for monetary aggregation embedded within a constrained optimization problem.

More recently, Barnett (1980) in a challenging paper, "Economic Monetary Aggregates: An Application of Index Number and Aggregation Theory," voiced objections to simple-sum aggregation procedures and argued instead for applying aggregation theory and statistical index number theory to monetary aggregation. As Barnett, Offenbacher, and Spindt (1984, p. 1051) put it,

> "by equally weighting components, aggregation by summation can badly distort an aggregate. For example, if one wished to obtain an aggregate of transportation vehicles, one would never aggregate by summation over the physical units of, say, subway trains and roller skates. Instead one could construct a quantity index (such as the Department of Commerce's indexes) using weights based on the *values* of the different modes of transportation."

Barnett has argued that a more satisfactory approach to monetary aggregation must involve consideration of the utility function underlying the demand for monetary assets. For example, the appropriate form of aggregation (simple-sum as opposed to other possibilities) will be determined by the relationship that monetary assets bear to one another and their contribution to total 'moneyness.' It turns out that simple-sum aggregation is justified, when viewed in this framework, only if the component assets are perfect substitutes.

The case for using microeconomic aggregation theory in monetary economics is now very strong. The theory has two branches, one leading to the construction of index numbers and methods derived from economic theory and one leading to the construction of money-demand functions in the context of a system of equations modeling the wealth holder's allocation of funds between money and nonmoney assets. The two branches are supported by the same structure in that the supporting theory in both cases is that of the constrained maximization of the aggregate consumer's dynamic utility function.

In what follows, we briefly spell out the microtheoretical framework to the aggregation of money, leaving the related discussion of the demand systems approach to modeling the demand for money (and monetary assets) for later chapters.

15.2 The User Cost of Money

The meaning of the *price* of money is not obvious in monetary theory. Usually this price has been viewed as varying inversely to the general price level. In this sense, the price of money is its purchasing power in terms of real goods and services. The price of money has also been viewed as an opportunity cost — the cost of not holding interest-yielding assets. In fact, as we will see later in this book, the usual assumption is that the demand for money depends negatively on the incentives for holding other assets relative to money.

In the recent literature, however, money is treated as a durable good having an infinite life and it is assumed that money retains at least some value beyond the holding period. Under such an assumption, it would be wrong to attribute a price of unity — the full purchase price — to a unit of the stock of money, simply because this one dollar price represents the price of a unit of the stock over an infinite holding period. There is no question that money is a stock (at an instant of time). But money is also an economic good that provides a variety of services (i.e., liquidity, safety, convenience). These services of money are better described in a flow dimension (per period of time).

Donovan (1978) argued that a *user cost* concept, rather than the full purchase price, is more appropriate for pricing money. Barnett (1978) derived the user cost formula in a constrained intertemporal consumer optimization framework — see also Barnett and Serletis (2000, Chapter 1). The user cost is given by

$$p_j = p^* \left(\frac{R - r_j}{1 + R} \right) \tag{15.1}$$

and denotes the discounted interest foregone by holding a dollar's worth of the jth asset. Here, r_j is the yield on the jth asset, R is the yield on the benchmark asset, and p^* is the true cost of living index.

The benchmark asset is specifically assumed to provide no liquidity or other monetary services and is held solely to transfer wealth intertemporally. In theory, R is the maximum expected holding period yield in the economy. It is usually defined in practice in such a way that the user costs for the monetary assets are positive. Note that if p^* is deleted from the user cost formula, the formula produces real rather than nominal user cost. The interest rates are nominal so that inflationary expectations appear here (mainly in the denominator, since the effects in the two rates in the numerator of the formula may well cancel out).

15.3 Microeconomic Foundations

With Barnett's (1978) derivation of the user cost of monetary assets the stage has been set for formulating a representative consumer's decision problem over consumption goods, leisure, and the services of monetary assets. In doing so, we assume that the services of consumption goods, as well as the services of monetary assets and leisure, enter as arguments in the representative agent's utility function

$$u = u(\boldsymbol{c}, \ell, \boldsymbol{x}) \tag{15.2}$$

where

$\boldsymbol{c} =$ a vector of the services of consumption goods
$\ell =$ leisure time, and
$\boldsymbol{x} =$ vector of the services of monetary assets
(assumed to be proportional to the stocks)

The utility function (15.2) is assumed to be maximized subject to a full income constraint

$$\boldsymbol{q}'\boldsymbol{c} + \boldsymbol{p}'\boldsymbol{x} + w\ell = Y,$$

where

$Y =$ full income
$w =$ wage rate
$\boldsymbol{q} =$ a vector of prices of the consumption goods
(with the prime indicating a row vector)
$\boldsymbol{p} =$ a vector of monetary asset user costs (or rental prices),
with the ith component given as above

In order to focus on the details of the demand for monetary services, ignoring other types of goods, a good starting point is the theory of two-stage optimization investigated initially in the context of consumer theory by Strotz (1957, 1959) and Gorman (1959). It refers to a sequential expenditure allocation, where in the first stage (that of *budgeting* or *price aggregation*), the consumer allocates his expenditure among broad categories (consumption goods, leisure, and monetary services in our context) relying on price indexes for these categories, and then in the second state (that of *decentralization*) allocates expenditure within each category.

Decomposition of the consumer choice problem along these lines is possible if and only if the representative individual's utility function is weakly separable, implying a *utility tree*, in the services of monetary assets. That is, it must be possible to write the utility function as

$$u = u\Big(\boldsymbol{c}, \ell, f(\boldsymbol{x})\Big), \tag{15.3}$$

in which $f(\boldsymbol{x})$ is the monetary services aggregator function (quantity index) satisfying a number of economically motivated conditions that will be mentioned later.

As we shall argue later in the book, according to the original definition of separability by Leontief (1947) and Sono (1961) the algebraic requirement of (direct) weak separability in the services of monetary assets is that

$$\frac{\partial}{\partial \zeta}\left(\frac{\dfrac{\partial u}{\partial x_i}}{\dfrac{\partial u}{\partial x_j}}\right) = 0, \quad \zeta = \boldsymbol{c}, \ell,$$

for $i \neq j$. That is, the marginal rate of substitution between any two monetary assets does not depend upon the values of $\boldsymbol{c}$ and ℓ. This means that the demand for monetary services is independent of relative prices outside the monetary group.[1]

Whether or not the utility function (15.2) is weakly separable in monetary services is, of course, an empirical question. Ideally, instead of treating (15.3) as a maintained (untested) hypothesis, as we do here, one could test whether the utility function (15.2) is appropriately separable in monetary services — an assumption implicit in the traditional money-nonmoney dichotomization. This issue remains relatively unexplored.

If one is willing to continue focusing on the details of the demand for services of monetary assets, ignoring other types of goods, the following classical consumer problem can be utilized

$$\max_{\boldsymbol{x}} f(\boldsymbol{x}) \quad \text{subject to} \quad \boldsymbol{p}'\boldsymbol{x} = y, \tag{15.4}$$

in which y is the expenditure on the services of monetary assets (determined in the first stage of the two level optimization problem) and $\boldsymbol{p}$ is as defined above, a vector of monetary asset user costs.

[1] Note that the separability structure is asymmetric. That is, $\boldsymbol{c}$ is not separable from $\boldsymbol{x}$ and ℓ in $u(\cdot)$ unless there exists a function $g(\boldsymbol{c})$ such that

$$u = u(\boldsymbol{c}, \ell, \boldsymbol{x}) = u\Big(g(\boldsymbol{c}), \ell, f(\boldsymbol{x})\Big).$$

For an extensive discussion of separability, see Blackorby, Primont, and Russell (1978).

15.4 Aggregation Theory

In the discussion to this point, we have shown the steps that are normally taken to reduce a very general consumer choice problem to an asset-choice problem. At this point, we are prepared to proceed to results in the aggregation-theoretic literature, in which we are looking for monetary aggregates that are consistent with the optimizing behavior of rational economic agents. We begin with the monetary services utility function, $f(\boldsymbol{x})$, assuming that the utility function (15.2) is weakly separable in monetary services.

Using a specific and differentiable form for $f(\boldsymbol{x})$, and solving decision (15.4), we can derive the demand-function system. Using these derived solution functions and specific monetary data, we then could estimate the parameters and replace the unknown parameters of $f(\boldsymbol{x})$ by their estimates. The resulting estimated function is called an *economic* (or *functional*) monetary index, and its calculated value at any point in time is an economic monetary-quantity index number.

The problem is that the use of a specific function necessarily implies a set of implicit assumptions about the underlying preference structure of the economic agent. For example, the use of a weighted linear aggregator function,

$$f(\boldsymbol{x}) = \sum_{j=1}^{n} a_j x_j,$$

implies perfect substitutability among the n assets and hence should logically lead to specialization in consumption of the least expensive asset.[2] If this is inaccurate, obviously, we commit a specification error by using this functional form.

The use of a Cobb-Douglas functional form,

$$f(\boldsymbol{x}) = \prod_{j=1}^{n} x_j^{a_j},$$

imposes an elasticity of substitution equal to unity ($\sigma = 1$) between every pair of assets and its use implies that each asset always accounts for a constant share of the expenditure. Again, if this proposition is at odds with the facts, as it is likely to be, the use of the Cobb-Douglas seems inappropriate.[3]

[2] The more restrictive unit-weighted ($a_j = 1$, $j = 1, ..., n$) aggregator function implies dollar for dollar perfect substitutability. This is the simple-sum aggregation procedure.

[3] In general, the elasticity of substitution between assets i and j is defined as

As a last example, a constant elasticity of substitution (CES) functional form

$$f(\boldsymbol{x}) = \sum_{j=1}^{n} (a_j x_j^r)^{1/r},$$

where $0 < a_j < 1$, $-\infty < r < 1$, relaxes the unitary elasticity of substitution restriction imposed by the Cobb-Douglas, but imposes the restriction that the elasticity of substitution between any pair of assets is always constant, $\sigma = 1/(1-r)$. Again this seems contrary to fact.

The list of specific functional forms is, of course, boundless, but the defining property of the more popular of these entities is that they imply limitations on the behavior of the consumer that may be incorrect in practice. While the issue of their usefulness is ultimately an empirical question — and we shall treat the issue that way in this book — we feel that most members of this class of functions should be rejected, partly in view of the restrictive nature of their implicit assumptions, and partly because of the existence of attractive alternatives.

Among the alternatives is a member of the class of quadratic utility functions. With a member of the quadratic class, we would be using a *flexible functional form* to approximate the unknown monetary-services aggregator function, $f(\boldsymbol{x})$. Flexible functional forms such as the translog, introduced by Christensen, Jorgenson, and Lau in their 1975 article, "Transcendental Logarithmic Utility Functions,"

$$f(\boldsymbol{x}) = \alpha_0 + \sum_{i=1}^{n} \alpha_i \log x_i + \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_{ij} \log x_i \log x_j, \tag{15.5}$$

can locally approximate to the second order any unknown functional form for the monetary services aggregator function, and even higher

---

$$\sigma_{ij} = \frac{d \log(x_j/x_i)}{d \log\left(f_i(\boldsymbol{x})/f_j(\boldsymbol{x})\right)}.$$

To calculate the elasticity of substitution for the simple, two-asset (i.e., $n = 2$) Cobb-Douglas utility function, we note that the numerator of the above expression is

$$d \log(x_2/x_1) = d \log x_2 - d \log x_1,$$

and that the denominator is

$$d \log\left(f_1(\boldsymbol{x})/f_2(\boldsymbol{x})\right) = d \log\left(\frac{\alpha_1}{\alpha_2} \frac{x_2}{x_1}\right) = d \log x_2 - d \log x_1.$$

Hence, $\sigma = 1$.

quality approximations are available. We will consider the details of such functional forms later in this book.

If one is to do away with the simple-sum method of aggregating money and replace it with a nonlinear aggregator function as suggested, one will be able to deal with less than perfect substitutability and, for that matter, with variations over time in the elasticities of substitution among the components of the monetary aggregates. There is a problem, however, and this is that the functions must be estimated over specific data sets (and re-estimated periodically) with the attendant result that the index becomes dependent upon the specification.

This dependence is particularly troublesome to government agencies that have to justify their procedures to persons untrained in econometrics. This is a reasonable concern — and it is exacerbated by the fact that there are many possible nonlinear models from which to choose. Under these circumstances, government agencies around the world have always viewed aggregation theory as being solely a research tool, and have instead used index number formulas from statistical index number theory, to which we now turn.

15.5 Index Number Theory

Statistical index-number theory provides a class of quantity and price indexes that can be computed from price and quantity data alone, thus eliminating the need to estimate an underlying structure. In fact since the appearance of Fisher's (1922) early classic book on statistical index number theory, nearly all federal government economic data series have been based upon aggregation formulas from that literature. Well known examples are the Consumer Price Index, which is a Laspeyres price index, the Implicit Price Deflator, which is a Paasche price index, and real GNP, which is a Laspeyres quantity index.[4] The simple-sum index,

[4] The Laspeyres quantity index is

$$M_t^L = \sum_{j=1}^{n} w_{j,t-1} \left(\frac{x_{jt}}{x_{j,t-1}} \right),$$

where $w_{jt} = p_{jt}x_{jt} / \sum_{k=1}^{n} p_{kt}x_{kt}$ is the jth asset's share in expenditure on all assets. The Paasche quantity index is

$$M_t^P = \frac{1}{\displaystyle\sum_{j=1}^{n} w_{jt} \left(\frac{x_{j,t-1}}{x_{jt}} \right)}.$$

often used for monetary quantities, is a member of the broad class, but the simple-sum is a special case, since it contains no prices.

Statistical indexes are mainly characterized by their statistical properties. These properties were examined in great detail by Fisher (1922) and serve as tests in assessing the quality of a particular statistical index. They have been named, after Fisher, as *Fisher's system of tests.* Eichhorn (1976, 1978) provides a detailed analysis as well as a comprehensive bibliography of Fisher's *test* (or *axiomatic*) approach to index numbers.

While Fisher found the simple-sum index to be the worst known index number formula, the index that he found to be the best, in the sense of possessing the largest number of appropriate statistical properties, has now become known as the *Fisher ideal* index. Another index found to possess a very large number of such properties is the (Törnqvist) discrete time approximation to the continuous Divisia index. That index commonly is called the Törnqvist index or just the Divisia index (in discrete time). We shall use the latter naming convention.

Let x_{jt} be the quantity of the jth asset during period t, and let p_{jt} be the rental price (i.e., user cost) for that asset during period t. Then, the Fisher ideal index, M_t^F, during period t, is the geometric average of the Laspeyres and Paasche indexes

$$M_t^F = M_{t-1}^F \left[\sum_{j=1}^{n} w_{j,t-1} \left(\frac{x_{jt}}{x_{j,t-1}} \right) \times 1 \Bigg/ \sum_{j=1}^{n} w_{jt} \left(\frac{x_{j,t-1}}{x_{jt}} \right) \right]^{1/2},$$

where

$$w_{jt} = \frac{p_{jt} x_{jt}}{\sum_{k=1}^{n} p_{kt} x_{kt}}$$

is the jth asset's share in expenditure on the total portfolio's service flow.

On the other hand, the discrete time (Törnqvist) Divisia index, M_t^D, during period t, is

$$M_t^D = M_{t-1}^D \prod_{j=1}^{n} \left(\frac{x_{jt}}{x_{j,t-1}} \right)^{(1/2)(w_{jt}+w_{j,t-1})}.$$

It is informative to take the logarithm of each side of the above equation, so that

$$\log M_t^D - \log M_{t-1}^D = \sum_{j=1}^{n} w_{jt}^* (\log x_{jt} - \log x_{j,t-1}), \tag{15.6}$$

where $w^*_{jt} = (1/2)(w_{jt} + w_{j,t-1})$. In this form, it is easy to see that for the Divisia index the growth rate (log change) of the aggregate is the share-weighted average of the growth rates of the component quantities.

The primary advantage of the Fisher ideal index over the Divisia index is that the Fisher ideal index satisfies Fisher's *factor reversal test* — which requires that the product of the price and quantity indexes for an aggregated asset (or good) should equal actual expenditures on the component assets (or goods) — while the Divisia index fails that test. However, the magnitude of the error is very small (third order in the changes), and the Divisia index has the very large advantage of possessing the easily interpreted functional form, given as equation (15.6).

15.6 Diewert's Link

Until relatively recently, the fields of aggregation theory and statistical index number theory developed independently. However, Diewert in his 1976 paper, "Exact and Superlative Index Numbers," provided the link between aggregation theory and statistical index number theory by attaching economic properties to statistical indexes. These properties are defined in terms of the statistical indexes' ability to approximate a particular functional form for the unknown aggregator function, $f(\boldsymbol{x})$ in our case.

For example, for a number of well known statistical indexes Diewert shows that they are equivalent to the use of a particular functional form. Such statistical indexes are called *exact*. Exactness, however is not sufficient for acceptability of a particular statistical index when the functional form for the aggregator function is not known. In this case it seems desirable to choose a statistical index which is exact for a flexible functional form. Diewert termed such statistical indexes *superlative*. Diewert also showed that the Divisia index is exact for the linearly homogeneous translog and is, therefore, superlative.

Following Diewert (1976) we will demonstrate how an exact index for the homogeneous translog functional form can be derived and that the index is the Divisia index. Consider the homogeneous (of degree one) translog functional form given by equation (15.5), with the following (homogeneity) restrictions imposed

$$\sum_{i=1}^{n} \alpha_i = 1 \quad \text{and} \quad \sum_{j=1}^{n} \beta_{ij} = 0, \quad \text{for all } j,$$

and define a quantity index between periods 0 and r, $Q(\boldsymbol{p}^0,\boldsymbol{x}^0,\boldsymbol{p}^r,\boldsymbol{x}^r)$, $r = 1,...,T$, as a function of the n prices in periods 0 and r. To ensure that the statistical index approximates the functional form for the aggregator function, it is required that the following relation is satisfied

$$\frac{f(\boldsymbol{x}^r)}{f(\boldsymbol{x}^0)} = Q(\boldsymbol{p}^0,\boldsymbol{x}^0,\boldsymbol{p}^r,\boldsymbol{x}^r), \quad \text{for } r = 1,...,T \tag{15.7}$$

whenever $\boldsymbol{x}^r > 0$ is the solution to the following aggregator maximization problem

$$\max_{\boldsymbol{x}} \left\{ f(\boldsymbol{x}) : \boldsymbol{p}^r\boldsymbol{x} \leq \boldsymbol{p}^r\boldsymbol{x}^r, \quad \boldsymbol{x} \geq 0 \right\}, \quad \text{for } r = 0,...,T.$$

For a base period normalization $f(\boldsymbol{x}^0) = 1$, equation (15.7) implies that the quantity index at time t equals the aggregator function evaluated at that point. If equation (15.7) is satisfied the quantity index $Q(\boldsymbol{p}^0,\boldsymbol{x}^0,\boldsymbol{p}^r,\boldsymbol{x}^r)$ is said to be exact for the aggregator function $f(\boldsymbol{x})$.

Next, we make use of the quadratic approximation lemma of Theil (1967, p. 222-223)

$$f(\boldsymbol{z}^1) - f(\boldsymbol{z}^0) = \frac{1}{2}\left[\nabla f(\boldsymbol{z}^1) + \nabla f(\boldsymbol{z}^0)\right](\boldsymbol{z}^1 - \boldsymbol{z}^0), \tag{15.8}$$

where $\nabla f(\boldsymbol{z}^r)$ is the gradient vector of $f(\boldsymbol{z})$ evaluated at $\boldsymbol{z}^r$. Now, since for the translog $z_i^r = \log x_i^r$ and $f(\boldsymbol{z}^r) = \log f(\boldsymbol{x}^r)$, for $r = 0,1$ and $i = 1,...,n$, we have

$$\nabla f(\boldsymbol{z}^r) = \frac{\partial \log f(\boldsymbol{x}^r)}{\partial \log \partial \boldsymbol{x}^r} = \frac{\partial f(\boldsymbol{x}^r)}{\partial \boldsymbol{x}^r}\frac{\boldsymbol{x}^r}{f(\boldsymbol{x}^r)} = \widehat{\boldsymbol{x}}^r \frac{\nabla f(\boldsymbol{x}^r)}{f(\boldsymbol{x}^r)}, \tag{15.9}$$

where $\widehat{\boldsymbol{x}}^r, r = 0,1$ is the vector $\boldsymbol{x}$ diagonalized into a matrix. Then if we substitute (15.9) into (15.8), we obtain

$$\log\frac{f(\boldsymbol{x}^1)}{f(\boldsymbol{x}^0)} = \frac{1}{2}\left[\widehat{\boldsymbol{x}}^1\frac{\nabla f(\boldsymbol{x}^1)}{f(\boldsymbol{x}^1)} + \widehat{\boldsymbol{x}}^0\frac{\nabla f(\boldsymbol{x}^0)}{f(\boldsymbol{x}^0)}\right]\left(\log \boldsymbol{x}^1 - \log \boldsymbol{x}^0\right), \tag{15.10}$$

where $\log\boldsymbol{x}^r = (\log x_1^r, \log x_2^r, \cdots, \log x_n^r)$, for $r = 0,1$. Using Wold's theorem (for a linear homogeneous function),

$$\frac{\boldsymbol{p}^r}{\boldsymbol{p}^r\boldsymbol{x}^r} = \frac{\nabla f(\boldsymbol{x}^r)}{f(\boldsymbol{x}^r)},$$

and substituting into (15.10) we obtain

$$\log \frac{f(\boldsymbol{x}^1)}{f(\boldsymbol{x}^0)} = \frac{1}{2}\left[\frac{\widehat{\boldsymbol{x}}^1 \boldsymbol{p}^1}{\boldsymbol{p}^{1T}\boldsymbol{x}^1} + \frac{\widehat{\boldsymbol{x}}^0 \boldsymbol{p}^0}{\boldsymbol{p}^{0T}\boldsymbol{x}^0}\right]\left(\log \boldsymbol{x}^1 - \log \boldsymbol{x}^0\right)$$

$$= \sum_{j=1}^{n} \frac{1}{2}\left(w_j^1 + w_j^0\right)\left(\log x_j^1 - \log x_j^0\right).$$

The right-hand side of the above equation is the same as that in equation (15.6). In fact, it is the Divisia index in growth rate form. Hence, we have shown that the Divisia index is exact for the homogeneous translog. Since the homogeneous translog is a flexible functional form, the Divisia index is a superlative index.

It is obvious that the definition of exact statistical indexes depends upon microeconomic maximizing behavior and is completely independent of the form or properties the aggregator function might have. However, if we do not know the true functional form for the aggregator function (that is, if we do not have *a priori* information about preferences) it would be wise to choose a statistical index that is exact for a flexible functional form. Diewert (1976) also showed that the Fisher ideal index is exact for the square root of a homogeneous quadratic function — see also Lau (1978).

15.7 Conclusion

With Diewert's (1976) successful merging of index number theory with economic aggregation theory and Barnett's (1978) derivation of the user cost of the services of monetary assets, the stage has been set for introducing index number theory into monetary economics. The moral of the story is that the nonlinearity produced by economic theory is important and that the simple-sum index should be abandoned (both as a source of research data and as an intermediate target or indicator for monetary policy).

The most obvious conclusion of this brief theoretical discussion is that the rigorous use of nonlinear microeconomic theory will result in consistent and satisfactory monetary aggregates. In fact, Barnett (1980) originated the Divisia monetary aggregates, which are elements of the superlative class, and which represent a practically viable, and theoretically meaningful, alternative to the inappropriate simple-summation aggregates. We discuss the Divisia monetary aggregates in the next chapter.

16

The New Monetary Aggregates

We have argued in Chapter 15 that the simple-sum method of aggregation makes strong *a priori* assumptions about substitution effects and the result is a set of monetary aggregates that do not accurately measure the actual quantities of the monetary products that optimizing economic agents select (in the aggregate). We also surveyed the microeconomic theory of monetary aggregation, as it has evolved during the past twenty years, using a model of the optimizing behavior of representative economic agents.

Our objective in this chapter is to develop a better understanding of the Divisia monetary aggregates, by presenting the source and the underlying microeconomic theory of the Divisia index. In addition, we provide an empirical assessment of the relative merits of the Divisia versus the simple sum method of monetary aggregation as well as other, recently proposed, aggregation procedures. Our objective is to be able to settle on a satisfactory method of 'measuring' money.

16.1 The Neoclassical Monetary Problem

In the discussion to this point, we have shown that in the second stage of the two-stage maximization problem, with weak separability between monetary assets and consumer goods and leisure, the consumer faces the following problem

$$\max_{\boldsymbol{x}} f(\boldsymbol{x}) \quad \text{subject to} \quad \boldsymbol{p}'\boldsymbol{x} = y, \tag{16.1}$$

or, written out in full,

$$\max_{x_1, x_2, \cdots, x_n} f(x_1, x_2, \cdots, x_n)$$

subject to

$$\sum_{i=1}^{n} p_i x_i = p_1 x_1 + p_2 x_2 + \cdots + p_n x_n = y,$$

where $\boldsymbol{x} = (x_1, x_2, \cdots, x_n)$ is a vector of services from monetary assets, $\boldsymbol{p} = (p_1, p_2, \cdots, p_n)$ is a vector of monetary asset user costs, and y is the expenditure on the services of monetary assets.

The *first order conditions* for a maximum can be found by forming an auxiliary function known as the *Lagrangian*

$$\mathcal{L} = f(\boldsymbol{x}) + \lambda \left(y - \sum_{i=1}^{n} p_i x_i \right),$$

where λ is the *Lagrange multiplier*. By differentiating $\mathcal{L}$ with respect to x_i, and using the budget constraint, we obtain the $(n+1)$ first order conditions

$$\frac{\partial f(\boldsymbol{x})}{\partial x_i} - \lambda p_i = 0, \quad i = 1, ..., n;$$

$$y - \sum_{i=1}^{n} p_i x_i = 0,$$

where the partial derivative $\partial f(\boldsymbol{x})/\partial x_i$ is the *marginal utility* of asset i.

What do these first order conditions tell us about the solution to the utility maximization problem? Notice that the first n conditions can be written as

$$\frac{\partial f(\boldsymbol{x})/\partial x_1}{p_1} = \frac{\partial f(\boldsymbol{x})/\partial x_2}{p_2} = \cdots = \frac{\partial f(\boldsymbol{x})/\partial x_n}{p_n} = \lambda, \tag{16.2}$$

which simply say that, in equilibrium, the ratio of marginal utility to price must be the same for all assets. Alternatively, for any two assets i and j, the above condition can be rewritten as

$$\frac{\partial f(\boldsymbol{x})/\partial x_i}{\partial f(\boldsymbol{x})/\partial x_j} = \frac{p_i}{p_j},$$

which says that, in equilibrium, the ratio of marginal utilities (also known as the *marginal rate of substitution*) must equal the respective price ratio.

Notice that according to equation (16.2), the optimal Lagrange multiplier is utility per unit of asset k divided by the number of dollars per unit of asset k ($k = 1, \cdots, n$), reducing to utility per dollar. By this interpretation, the optimal Lagrange multiplier is also called the *marginal utility of income.*

16.2 Understanding the Divisia Aggregates

Consider the representative consumer's utility function over monetary assets, $f(\boldsymbol{x})$. Take the total differential of $f(\boldsymbol{x})$ to get

$$df(\boldsymbol{x}) = \sum_{i=1}^{n} \left(\frac{\partial f(\boldsymbol{x})}{\partial x_i} \right) dx_i,$$

where $\partial f(\boldsymbol{x})/\partial x_i$ ($i = 1, ..., n$) are marginal utilities containing the unknown parameters of the function $f(\boldsymbol{x})$. From the first-order conditions of the neoclassical monetary problem (discussed in the previous section), we can write the marginal utilities as

$$\frac{\partial f(\boldsymbol{x})}{\partial x_i} = \lambda p_i, \quad i = 1, ..., n,$$

where λ is the Lagrange multiplier and p_i is the user-cost of asset i. This expression can then be substituted into the total differential of $f(\boldsymbol{x})$, to eliminate $\partial f(\boldsymbol{x})/\partial x_i$ ($i = 1, ..., n$), and yield

$$df(\boldsymbol{x}) = \sum_{i=1}^{n} \lambda p_i dx_i, \tag{16.3}$$

which is written not in unknown marginal utilities but in the unknown Lagrange multiplier, user costs, and *changes* in quantities.

In equation (16.3) the Lagrange multiplier is itself a function of unknown tastes and thereby a function of the parameters of the unknown utility function. We would rather not have to estimate it econometrically. In order to get rid of the Lagrange multiplier, we assume that the economic quantity aggregator, $f(\boldsymbol{x})$, is linearly homogeneous in its components — that is, $f(\kappa \boldsymbol{x}) = \kappa f(\boldsymbol{x})$. This is, indeed, a reasonable assumption, since it would be very curious indeed if linear homogeneity of $f(\boldsymbol{x})$ failed — in such a case the growth rate of the aggregate would differ from the growth rates of its components, even if all components were growing at the same rate.

Next define $P(\boldsymbol{p})$ to be the dual price index satisfying Fisher's factor reversal test[1]

$$P(\boldsymbol{p})f(\boldsymbol{x}) = \sum_{i=1}^{n} p_i x_i = y.$$

It can be shown [see Barnett, Fisher, and Serletis (1992, footnote 22)] that $\lambda = 1/P(\boldsymbol{p})$ in which case equation(16.3) reduces to

$$df(\boldsymbol{x}) = \sum_{i=1}^{n} \left(\frac{1}{P(\boldsymbol{p})} \right) p_i dx_i. \tag{16.4}$$

Manipulating equation (16.4) algebraically, to convert to growth rate (log change) form, we find that

$$d \log f(\boldsymbol{x}) = \sum_{i=1}^{n} w_i^* d \log x_i. \tag{16.5}$$

The result is that the log change in the utility level (and therefore in the level of the aggregate) is the weighted average of the log changes of the component levels. Equation (16.5) is the Divisia index in growth rate form, as defined in Chapter 10.

This exercise demonstrates the solid microeconomic foundations of the Divisia index. It is indeed, the logical choice for an index from a theoretical point of view, being the transformed first-order conditions for constrained optimization.

[1] Recall that Fisher's factor reversal test requires that the product of the price and quantity indexes for an aggregated asset (or good) should equal actual expenditure on the component assets (or goods).

16.3 Divisia Second Moments

Henri Theil in his 1967 book, *Economics and Information Theory*, observed that there is an interesting stochastic interpretation of the Divisia index. In particular, he observed that the Divisia weights are nonnegative and sum to 1 in every period. Given this, we can treat the growth rates of the components as drawn randomly from a population such that the right-hand side of equation (16.5) becomes an expectation. Under this interpretation, the left-hand side of equation (16.5) is the mean of the growth rates of the components, $d\log M^D$. What this then suggests is that the Divisia quantity index is a first moment and, by appealing to Theil's sampling analogy, we can define the Divisia second moments. In particular, the Divisia quantity variance (the second moment) is

$$K_t = \sum_{i=1}^{n} w_{it}^{*}(d\log x_{it} - d\log M_t^D)^2.$$

Also, since the Divisia price index (mean or first moment) is

$$d\log P_t = \sum_{i=1}^{n} w_{it}^{*} d\log p_{it},$$

the corresponding Divisia price variance is

$$J_t = \sum_{i=1}^{n} w_{it}^{*}(d\log p_{it} - d\log P_t)^2,$$

and the Divisia price-quantity covariance is

$$\Gamma_t = \sum_{i=1}^{n} w_{it}^{*}(d\log x_{it} - d\log M_t^D)(d\log p_{it} - d\log P_t).$$

Similarly, we can define the Divisia share mean as

$$d\log W_t = \sum_{i=1}^{n} w_{it}^{*} d\log w_{it},$$

and the Divisia share variance as

$$\Psi_t = \sum_{i=1}^{n} w_{it}^{*}(d\log w_{it} - d\log W_t)^2.$$

Note that Theil (1967) has shown that the Divisia second moments are related by the equality

$$K_t = \Psi_t - J_t - 2\Gamma_t.$$

Recently, Barnett and Serletis (1990) applied Theil's stochastic index number theory and tested for aggregation error in the Divisia monetary aggregates. Aggregation errors can be produced when the conditions for exact aggregation are violated. As we have seen, exact aggregation (over monetary assets) requires the existence of a weakly separable and linearly homogeneous aggregator function and removes dependency of market behavior upon distributional effects.

This means that the Divisia second moments would contain no information about the economy, if one already had conditioned upon the information contained in the Divisia mean. Barnett and Serletis (1990), by implicitly assuming that the appropriate distributional variable is the Divisia quantity variance, explored its macroeconomic effects by explicitly introducing it along with the Divisia quantity mean in various tests. They found no evidence of major aggregation error in the Divisia monetary aggregates.

16.4 Measurement Matters

At this stage, it is perhaps worth asking whether 'measurement' matters. To highlight the importance of measurement, let's recall the three-year 'monetarist experiment' of November 1979 to August 1982. During that time the Federal Reserve Board embarked on an experiment in monetarist policy designed to control the money supply and permit interest rates to be determined in the money markets, free of control.

The Fed's views on monetary policy, however, were based on the simple sum monetary aggregates (in particular simple sum M2) and monetary policy during the three-year period was considerably tighter than the Fed thought and led to the recession of 1982. The following quotation, from Barnett (1997, pp. 1174-1175), explains what happened:

> "As I reported in Barnett (1984), the growth rate of simple sum M2 during the period of the 'monetarist experiment' averaged 9.3%, while the growth rate of Divisia M2 during the period averaged 4.5%. Similarly, the growth rate of simple sum M3 during the period averaged 10%, while the growth rate of Divisia M3 during the period averaged 4.8%. This period followed double digit growth rates of all simple sum and Divisia

> monetary aggregates. In short, believers in simple sum monetary aggregation, who had been the advocates of the 'monetarist experiment,' were put in the embarrassing position of witnessing an outcome (the subsequent recession) that was inconsistent with the intent of the prescribed policy and with the behavior of the simple sum aggregates during the period. This unwelcome and unexpected outcome rendered vulnerable those economists who advocated a policy based upon the assumption of a stable simple sum demand for money function.
>
> Friedman's very visible forecast error on 26 September 1983 followed closely on the heels of the end of the monetarist experiment in August 1982 and the recession that it produced. The road buckled and collapsed below the monetarists and those who believed in stable simple sum demand for money functions. Those two associated groups have never recovered. But the recession that followed the monetarist experiment was no surprise to anyone who had followed the Divisia monetary aggregates, since those aggregates indicated that a severe deflationary shock had occurred. To those who were using data based upon valid index number and aggregation theory, rather than the obsolete simple sum monetary aggregates, the road remained smooth — no bumps, no breaks. Nothing unexpected had happened."

The above quotation shows that simple sum and Divisia monetary aggregates tell very different stories. Monetary policy, as indicated by the Divisia monetary aggregates, was tighter than indicated by the simple sum aggregates. That resulted in the severe 1982 recession, despite the fact that the Fed's intention, as indicated by the simple sum monetary aggregates, was to produce a gradual disinflation rather than a severe disinflationary shock. Hence, measurement matters and the failure to use superior aggregates can have big practical consequences.

16.5 The MQ and CE Indexes

While we are on the topic of monetary indexes, we will briefly consider two recent additions to the list of alternative index numbers, with a potential application to monetary aggregation. These are Spindt's (1985) 'monetary quantities' (MQ) index (which is no longer in use in the monetary literature), and the 'currency equivalent' (CE) index more recently introduced by Rotemberg (1991) and Rotemberg , Driscoll, and Poterba (1995).

In the particular form computed by Spindt (1985), MQ is measured as a Fisher ideal index, but with the user costs replaced by monetary-asset turnover rates. That is

$$MQ_t = MQ_{t-1} \left[\sum_{j=1}^{n} w_{j,t-1} \left(\frac{x_{jt}}{x_{j,t-1}} \right) \times 1 \Bigg/ \sum_{j=1}^{n} w_{jt} \left(\frac{x_{j,t-1}}{x_{jt}} \right) \right]^{1/2},$$

where

$$w_{jt} = \frac{v_{jt} x_{jt}}{\sum\limits_{k=1}^{n} v_{kt} x_{kt}},$$

with v_j being the turnover rate of monetary asset j. The problem with this procedure is that the MQ index, unlike the Divisia index, is inconsistent both with existing aggregation theory and index number theory. The relevant foundations (both index-number theoretic and aggregation theoretic) for the Fisher-ideal index require the use of prices and quantities and not turnover rates and quantities. If nothing else, MQ can be said to be no less arbitrary than the official simple-sum aggregates.

A more recent addition to the list of alternative index numbers is the Rotemberg, Driscoll, and Poterba (1995) currency equivalent (CE) index

$$CE = \sum_{j=1}^{n} \frac{R_t - r_{jt}}{R_t} x_{jt}. \tag{16.6}$$

This index is basically the simple sum index with the addition of a simple weighting mechanism. In (16.6), as long as currency gives no interest, units of currency are added together with a weight of one. Other assets are added to currency, but with a weight that declines toward zero as their return increases toward R_t and the assets come to behave more like the benchmark asset (a means to transfer wealth) and less like money.

The difference between Divisia and CE methods of monetary aggregation is that the former measures the flow of monetary services whereas the latter, like simple summation aggregation, measures the stock of monetary assets. There is also a considerably less attractive interpretation of the CE index as a flow index. See Barnett, Hinich, and Yue (2000) regarding the flow interpretation, which requires stronger assumptions than those needed to derive the Divisia flow index.

16.6 Empirical Comparisons

In order to provide a quantitative assessment of the simple sum, Divisia, and currency equivalent monetary aggregation procedures, we employ (seasonally-adjusted) monthly data, from 1959:1 to 2006:2, on United States simple sum, Divisia, and CE indexes. The data were obtained from the St. Louis MSI database, maintained by the Federal Reserve Bank of St. Louis as part of the Federal Reserve Economic Database (FRED) — see Anderson, Jones, and Nesmith (1997) for details regarding the construction of the Divisia and currency equivalent aggregates and related data.

Figures 16.1 to 16.4 provide graphical representations of the four major measures of money (M1, M2, M3, and MZM) under the simple sum, Divisia, and currency equivalent aggregation procedures. As the graphs indicate, the numbers differ considerably across the three monetary aggregation procedures. Even more interesting are the graphs of the (industrial production) velocities for these same aggregates appearing in Figures 16.5 to 16.8. Not only are the fluctuations of the velocity series different at different levels of aggregation, but also across aggregation methods.

Finally, in Table 16.1 we provide summary statistics based upon the first and second order sample moments of the monthly data on annual monetary growth rates.[2] Inspection of the summary statistics suggests that the average growth rate of the Divisia aggregates is, in general, less than that of the simple sum aggregates. In addition, the Divisia aggregates indicate less volatile monetary growth, during our sample period, than is indicated by the official simple sum aggregates.

However, the currency equivalent money measures tell a very different story. In particular, the average growth rate of the CE aggregates is always higher than that of the corresponding simple sum and Divisia aggregates. Also, observe that monetary policy, as measured by the CE aggregates, is much more volatile than is suggested by either the simple sum or Divisia aggregates. These differences reflect the essentially complicated monetary aggregation issues — something we mentioned when we discussed the 'monetarist experiment' of November 1979 to August 1982 and to be kept in mind throughout this book.

[2] The annual growth rates are simple percent monthly changes at an annual rate — that is, $1200 \times \ln(x_t/x_{t-1})$.

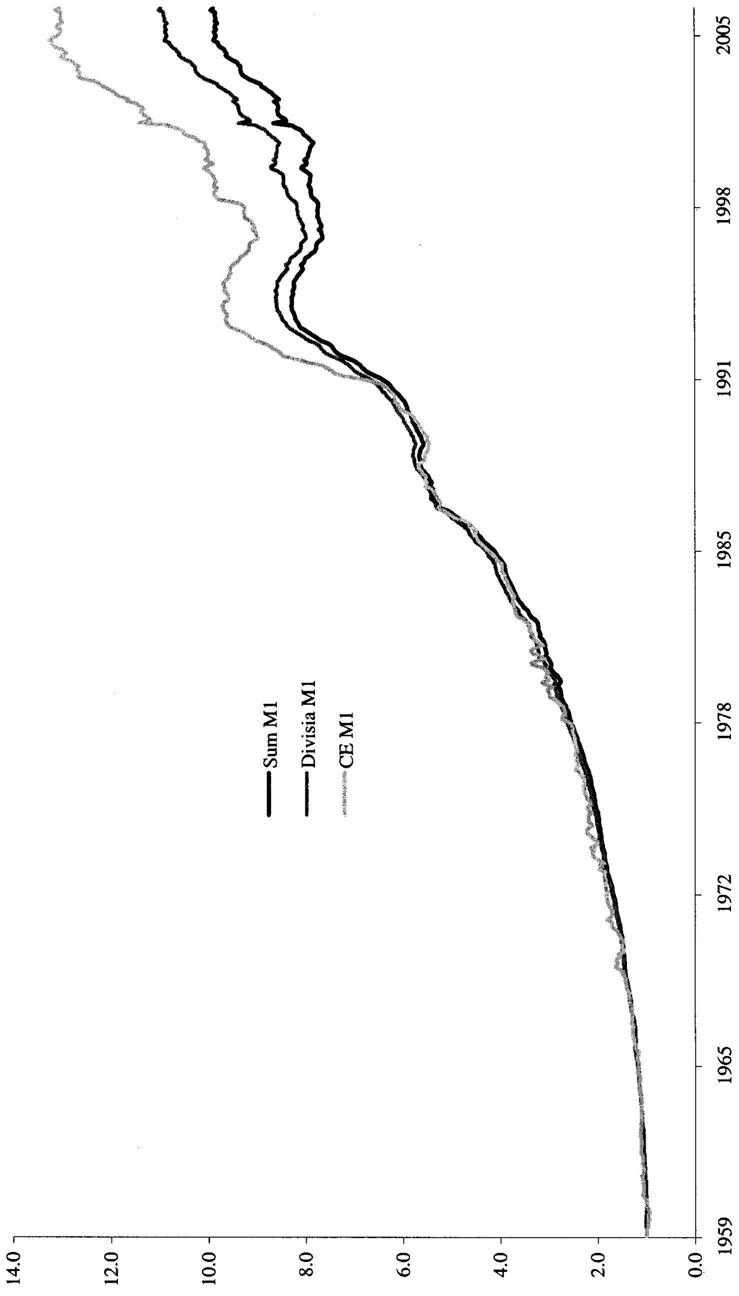


Figure 16.1 . Sum, Divisia, and CE M1 Money Measures

Figure 16.2 . Sum, Divisia, and CE M2 Money Measures

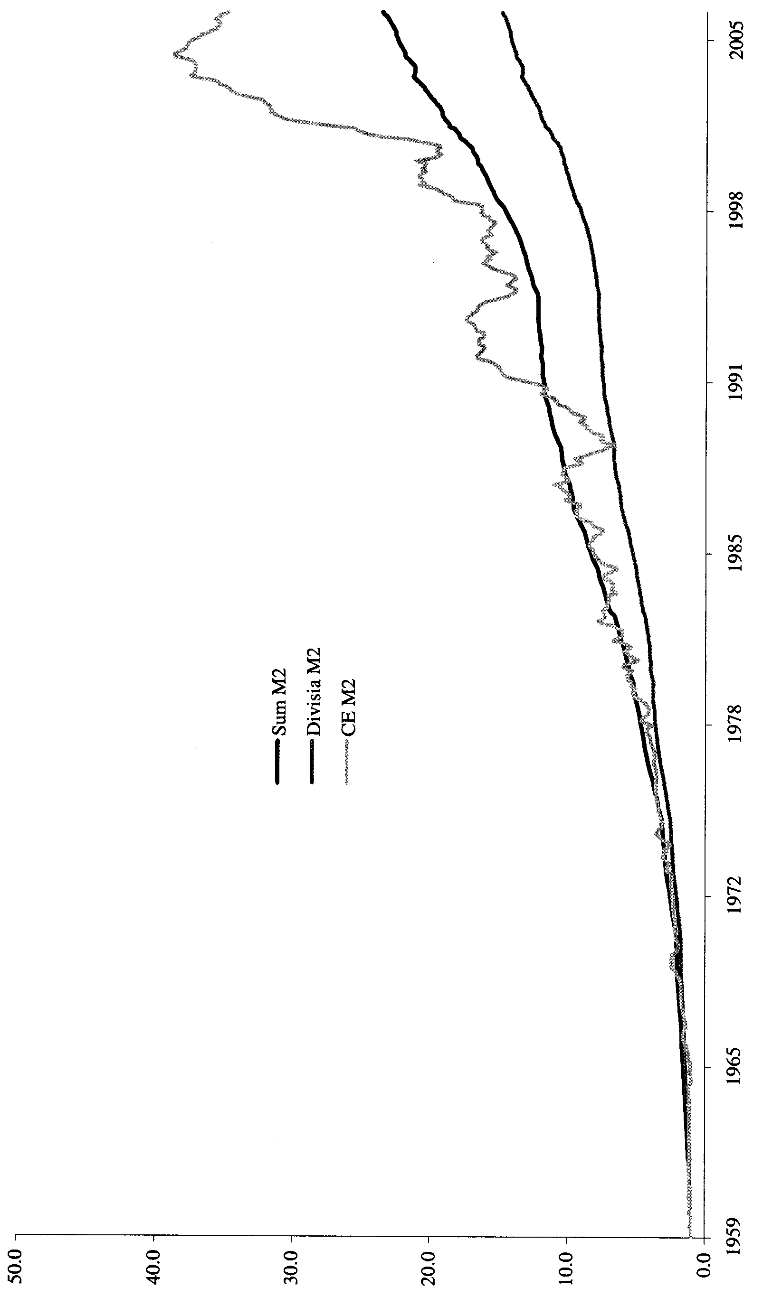


Figure 16.3. Sum, Divisia, and CE M3 Money Measures

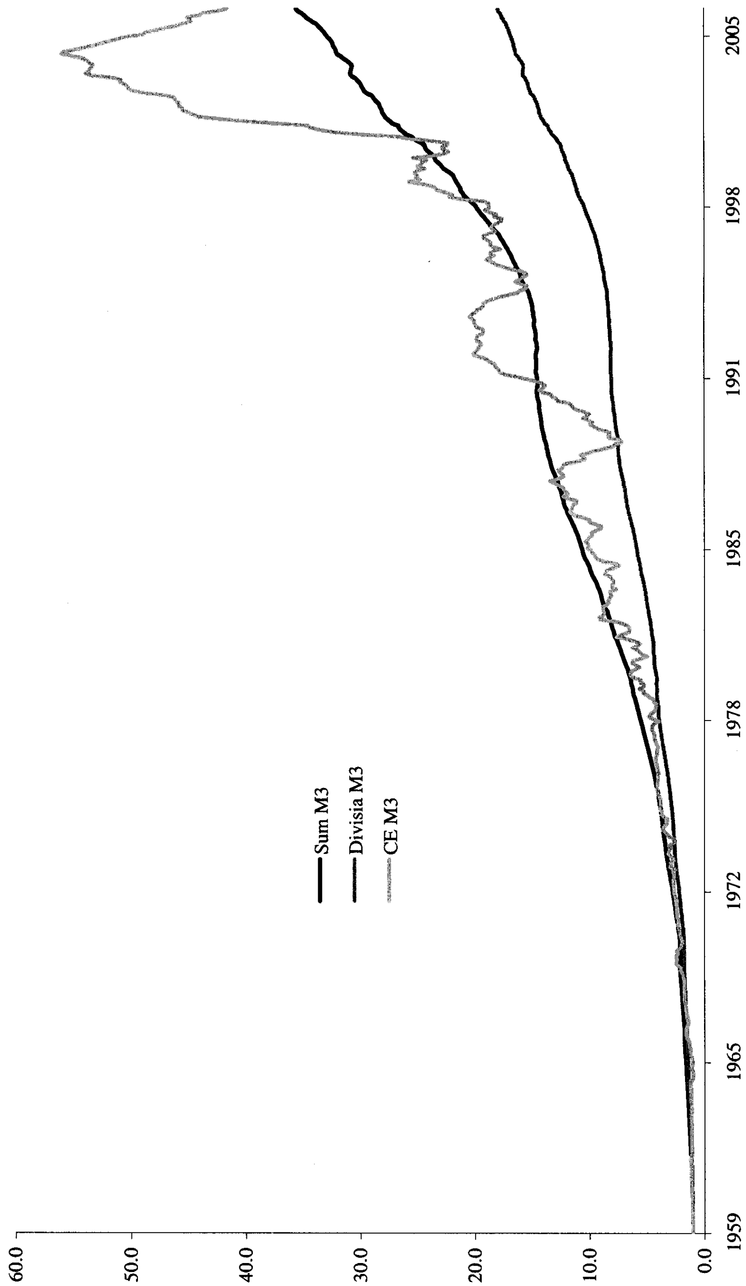


Figure 16.4. Sum, Divisia, and CE MZM Money Measures

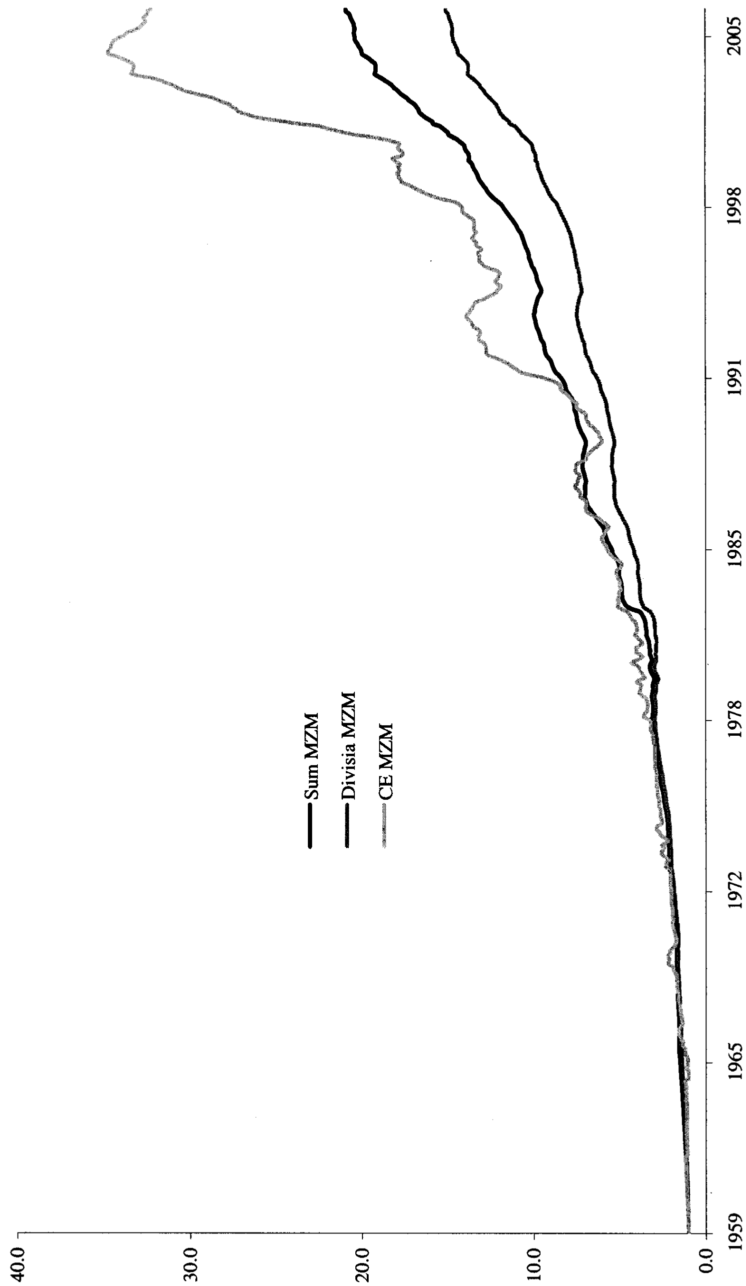


Figure 16.5. Sum, Divisia, and CE M1 Velocity Measures

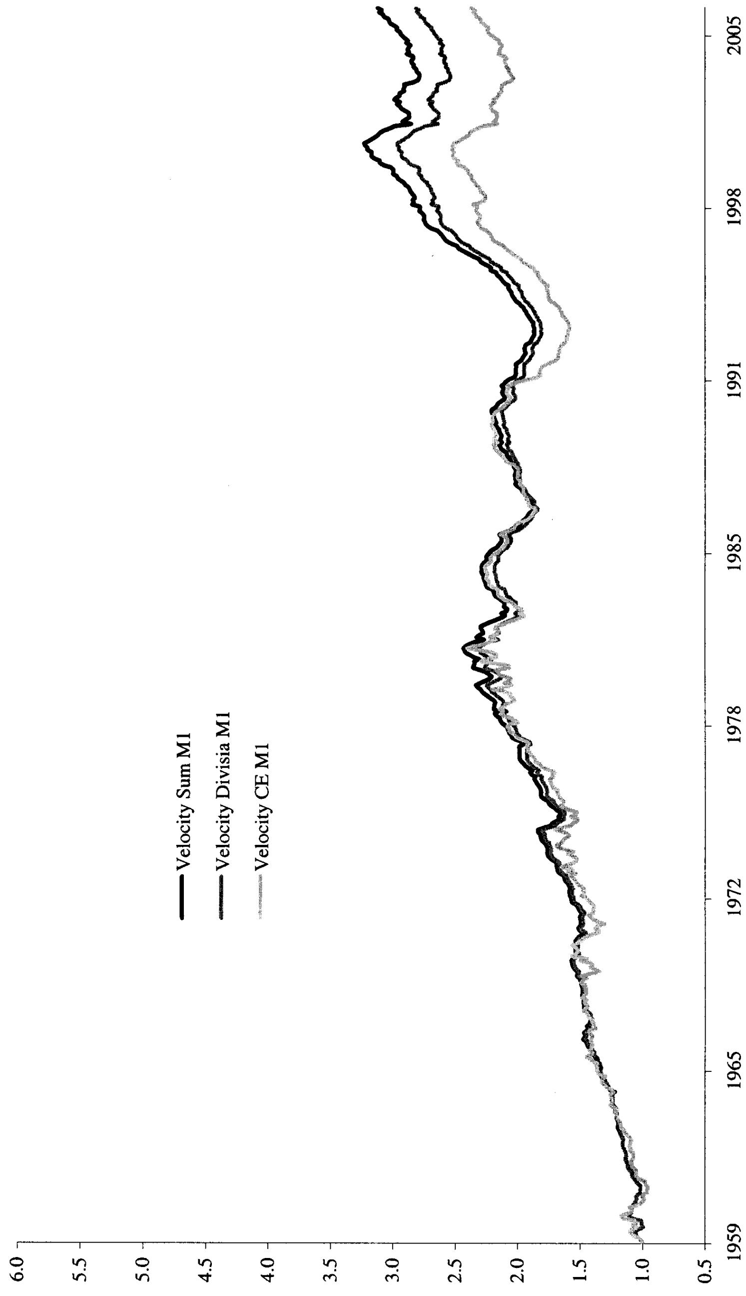


Figure 16.6. Sum, Divisia, and CE M2 Velocity Measures

Figure 16.7 . Sum, Divisia, and CE M3 Velocity Measures

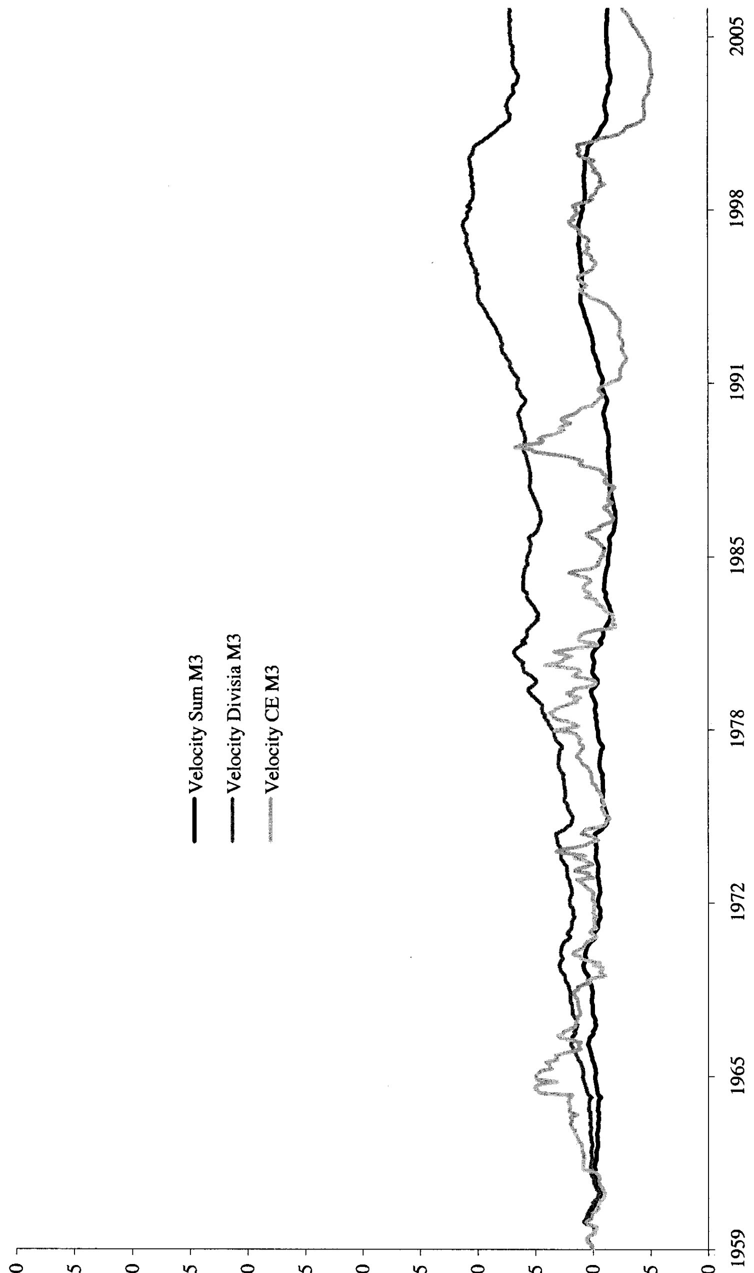


Figure 16.8. Sum, Divisia, and CE MZM Velocity Measures

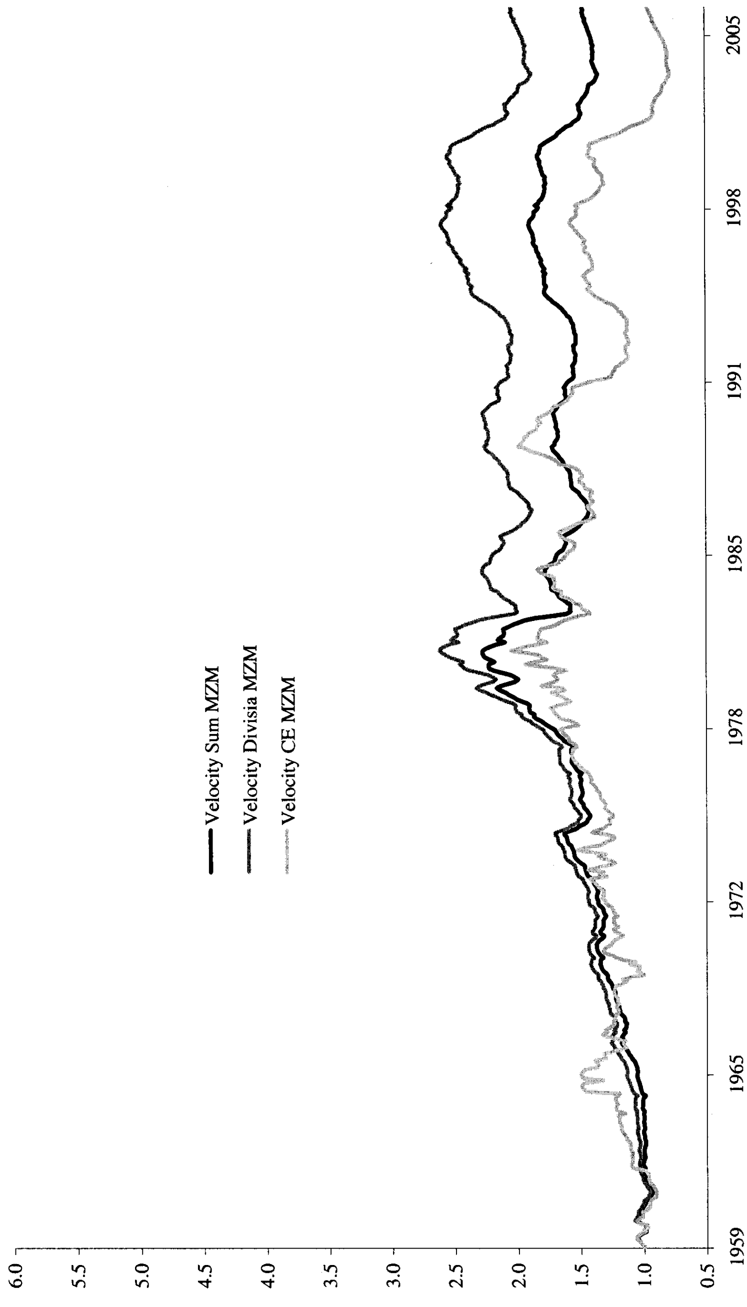


Table 16.1. Sample Moments of Annual Rates of Growth of Monetary Aggregates

Aggregate	Mean	Standard deviation
Sum M1	4.868	6.692
Divisia M1	5.088	5.912
CE M1	5.453	17.667
Sum M2	6.712	4.130
Divisia M2	5.731	4.012
CE M2	7.541	41.041
Sum M3	7.589	4.426
Divisia M3	6.138	4.001
CE M3	7.920	47.384
Sum MZM	6.459	8.728
Divisia MZM	5.766	7.226
CE MZM	7.378	31.371

Note: Monthly data 1959:1-2006:2.

16.7 Conclusion

We have surveyed a growing literature on the importance of the use of microeconomic aggregation theory in monetary aggregation. The issue is of practical importance, because effective conduct of monetary policy presupposes an appropriate monetary aggregate. It is also of academic interest for the insight it provides into the nature of ‘moneyness’. We have argued that the simple sum aggregates are just accounting identities, not economic aggregates. The Divisia aggregates are economic aggregates, and hence are useful indicators (or intermediate targets) for ultimate policy goals.

17

Nominal Stylized Facts

Kydland and Prescott in their 1990 article, "Business Cycles: Real Facts and a Monetary Myth," argue that business cycle research took a wrong turn when it abandoned the effort to account for the cyclical behavior of aggregate data, following Koopmans' (1965) criticism of the methodology developed by Burns and Mitchell (1946), as being 'measurement without theory.' Crediting Lucas (1977) with reviving interest in business cycle research, they initiated a line of research that builds on the growth theory literature and part of it involves an effort to assemble business cycle facts.

Kydland and Prescott report some original evidence for the United States economy, and conclude that several accepted nominal facts, such as the procyclical movements of money and prices, appear to be business cycle myths. In this chapter, we follow Kydland and Prescott (1990) and examine the cyclical behavior of United States money, prices, nominal interest rates, and velocity, using the monthly data that we discussed in the previous chapter.

In doing so, we start by discussing the popular Hodrick-Prescott (1980) filter, and then present Hodrick and Prescott cyclical correlations, making comparisons among simple sum, Divisia, and currency equivalent money and velocity measures. We also discuss the robustness of our results to relevant alternative filtering procedures.

17.1 The Hodrick and Prescott Filter

For a description of the stylized facts, we follow the current practice of detrending the data with the Hodrick-Prescott (HP) filter — see Hodrick and Prescott (1980). For the logarithm of a time series x_t, for $t = 1, 2, \ldots, T$, this procedure defines the (smoothed) *trend* or growth component, denoted τ_t, for $t = 1, 2, \ldots, T$, as the solution to the following minimization problem,

$$\min_{\{\tau_t\}_{t=1}^T} \sum_{t=1}^{T} (x_t - \tau_t)^2,$$

subject to

$$\sum_{t=2}^{T-1} [(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2 \leq \Lambda.$$

That is, the smoothed trend component, $\{\tau_t\}_{t=1}^T$, is obtained by minimizing the sum of squared differences from the data subject to the constraint that the sum of the squared differences be less than an appropriate bound Λ.

The above minimization problem is equivalent to the following unconstrained (minimization) problem,

$$\min_{\{\tau_t\}_{t=1}^T} \sum_{t=1}^{T} (x_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} [(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2,$$

for an appropriate value of the smoothing parameter λ — the Lagrange multiplier. The smaller is λ, the smoother the trend path and when $\lambda = 0$, the linear trend results. In our computations we set $\lambda = 14,400$, as it has been suggested by Kydland and Prescott (1990) for monthly data (with quarterly data, λ is set equal to $1,600$). Notice that $x_t - \tau_t$ is the filtered series.

As noted by Kydland and Prescott (1990), the Hodrick and Prescott filter has several attractive features. In particular, it occupies an intermediate position between the linear filter (which permits most low frequency components to pass through) and the first difference filter (which permits the least). Moreover, the Hodrick and Prescott trend is a linear transformation of the original series and is a smooth curve — like one that one would draw through a plot of the original series.

17.2 The Cyclical Behavior of Money

We describe the empirical regularities of the monetary variables, using the Hodrick and Prescott filter and by investigating whether deviations from their HP trends are correlated — and at what leads and lags — with the cycle. In particular, we measure the degree of comovement of a monetary aggregate with the cycle by the magnitude of the correlation coefficient $\rho(j)$, $j \in \{0, \pm 1, \pm 2, \ldots\}$. All the variables are in logarithms (with the exception of the rate variables) and the statistics discussed pertain to variables that have been processed via the Hodrick and Prescott filter — that is, to stationary HP cyclical deviations.

The contemporaneous correlation coefficient — $\rho(0)$ — gives information on the degree of contemporaneous comovement between the monetary series and the cycle. In particular, if $\rho(0)$ is positive, zero, or negative, we say that the series is *procyclical*, *acyclical*, or *countercyclical*, respectively. In fact, for data samples of this size, it has been suggested [see, for example, Ricardo Fiorito and Tryphon Kollintzas (1994)] that for $0.5 \leq |\rho(0)| < 1$, $0.2 \leq |\rho(0)| < 0.5$, and $0 \leq |\rho(0)| < 0.2$, we say that the series is strongly contemporaneously correlated, weakly contemporaneously correlated, and contemporaneously uncorrelated with the cycle, respectively. Also, $\rho(j)$ $j \in \{\pm 1, \pm 2, \ldots\}$ — the cross correlation coefficient — gives information on the phase shift of the monetary series relative to the cycle. If $|\rho(j)|$ is maximum for a positive, zero, or negative j, we say that the series is *leading* the cycle by j periods, is *synchronous*, or is *lagging* the cycle by j periods, respectively.

In Table 17.1 we report contemporaneous correlations as well as cross correlations (at lags and leads of $3, 6, 9, 12, 18$, and 24 months, given the high frequency nature of the data and the traditional view that there are 'long and variable lags' in the relationship between real and monetary variables) between the cyclical components of money and the cyclical component of industrial production. We see that money, irrespective of how it is measured is acyclical.

To investigate the robustness of this result to changes in the cyclical indicator, we report in Table 17.2 correlations (in the same fashion as in Table 17.1) using the unemployment rate as an indicator of the cycle. The seasonally adjusted unemployment rate includes all workers (including resident armed forces). Of course, since the cyclical components of industrial production and the unemployment rate are negatively correlated, a negative correlation in Table 17.2 indicates

Table 17.1. H-P Cyclical Correlations of Sum, Divisia, and CE Money Measures with Industrial Production

	$\rho(x_t, y_{t+j})$, $j = -24, -18, -12, -9, -6, -3, 0, 3, 6, 9, 12, 18, 24$												
Series	$j=-24$	$j=-18$	$j=-12$	$j=-9$	$j=-6$	$j=-3$	$j=0$	$j=3$	$j=6$	$j=9$	$j=12$	$j=18$	$j=24$
Sum M1	−.005	−.012	−.009	−.005	−.004	.003	.009	.031	.044	.047	.040	.010	−.019
Divisia M1	−.005	−.010	−.005	−.002	−.001	.005	.008	.029	.041	.045	.039	.009	−.019
CE M1	−.004	.000	−.017	−.023	−.036	−.025	−.004	.004	.010	.017	.023	.008	−.000
Sum M2	−.026	−.026	−.016	−.008	−.001	.010	.017	.035	.049	.057	.055	.037	.011
Divisia M2	−.027	−.030	−.019	−.010	−.003	.009	.021	.042	.057	.065	.063	.042	.013
CE M2	.018	.003	−.013	−.019	−.024	−.021	−.012	−.002	.003	.010	.017	.016	.006
Sum M3	−.010	.000	.016	.025	.029	.033	.027	.021	.018	.020	.021	.019	.006
Divisia M3	−.024	−.018	.001	.012	.021	.031	.034	.040	.044	.047	.044	.029	.002
CE M3	.016	−.000	−.019	−.026	−.031	−.030	−.022	−.006	.006	.018	.029	.030	.016
Sum MZM	−.017	−.027	−.031	−.027	−.020	−.008	.006	.032	.050	.060	.060	.040	.013
Divisia MZM	−.016	−.026	−.029	−.024	−.017	−.005	.010	.036	.054	.064	.063	.041	.011
CE MZM	.021	.002	−.017	−.022	−.024	−.017	−.002	.010	.016	.022	.024	.015	.003

Note: Sample period, monthly data: 1960:01 – 2006:01. x_t = Money, y_t = Industrial production.

Table 17.2. H-P Cyclical Correlations of Sum, Divisia, and CE Money Measures with the Unemployment Rate

Series	$\rho(x_t, u_{t+j})$, $j = -24, -18, -12, -9, -6, -3, 0, 3, 6, 9, 12, 18, 24$												
	$j=-24$	$j=-18$	$j=-12$	$j=-9$	$j=-6$	$j=-3$	$j=0$	$j=3$	$j=6$	$j=9$	$j=12$	$j=18$	$j=24$
Sum M1	.095	.128	.137	.118	.082	.026	−.049	−.116	−.162	−.182	−.174	−.113	−.028
Divisia M1	.106	.136	.142	.121	.085	.031	−.037	−.097	−.136	−.153	−.147	−.096	−.025
CE M1	.037	.069	.077	.067	.062	.057	.053	.027	.006	−.021	−.056	−.044	−.034
Sum M2	.107	.172	.181	.159	.119	.057	−.042	−.132	−.196	−.223	−.214	−.143	−.041
Divisia M2	.131	.175	.176	.150	.106	.036	−.072	−.170	−.237	−.269	−.265	−.187	−.059
CE M2	−.021	−.009	.027	.058	.093	.123	.123	.110	.092	.057	.012	−.032	−.039
Sum M3	.048	.039	−.001	−.031	−.061	−.089	−.121	−.132	−.128	−.112	−.089	−.015	.049
Divisia M3	.126	.106	.053	.011	−.039	−.096	−.166	−.218	−.244	−.245	−.221	−.112	.027
CE M3	−.030	.002	.058	.098	.140	.179	.184	.155	.114	.057	−.008	−.075	−.084
Sum MZM	.110	.202	.263	.250	.211	.138	.015	−.104	−.201	−.260	−.274	−.224	−.130
Divisia MZM	.114	.190	.242	.228	.187	.111	−.013	−.134	−.228	−.283	−.293	−.228	−.115
CE MZM	.002	.023	.066	.080	.096	.095	.062	.030	.007	−.017	−.038	−.031	−.027

Note: Sample period, monthly data: 1960:01 – 2006:01. x_t = Money, u_t = Unemployment rate.

procyclical variation and a positive correlation indicates countercyclical variation. Clearly, the results in Table 17.2 strongly confirm those in Table 17.1.

We interpret these results as being generally illustrating no significant differences across simple-sum, Divisia, and currency equivalent monetary aggregates. They also appear to support no monetary effect of money on real output (based on the monthly data that we use).

17.3 Prices, Interest Rates, and Velocity

While we are investigating nominal stylized facts, we also describe the statistical properties of the cyclical components of the price level (measured by the consumer price indexr) and two short-term nominal interest rates (to deal with anomalies that arise because of different ways of measuring financial market price information) — the Treasury bill rate and the federal funds rate. The Treasury bill rate is the interest rate on short-term unsecured borrowing by the U.S. government whereas the fed funds rate is the interest rate on fed funds. Again, with the exception of the rate variables, all the other variables are in logarithms. Table 17.3 reports HP cyclical correlations of prices and short-term nominal interest rates with each of industrial production (panel A) and the unemployment rate (panel B).

Irrespective of the cyclical indicator, we see that the price level is acyclical. This result clearly supports the Kydland and Prescott (1990) claim that the perceived fact of procyclical prices is but a myth. We also see that when industrial production is used as the cyclical indicator, the federal funds and Treasury bill rates are acyclical. However, when the unemployment rate is used as the cyclical indicator, these interest rate series are weakly procyclical and lead the cycle — recall that a variable leads the cycle if its cross-correlations with future industrial production are larger (in absolute value) than the contemporaneous correlation.

In addition to the statistical properties of the cyclical components of money, the price level, and nominal interest rates, we also examine the cyclical behavior of simple-sum, Divisia, and CE velocity (at the M1, M2, M3, and MZM levels of monetary aggregation), using both industrial production as well as the unemployment rate as measures of the cycle. The results reported in Tables 17.4 and 17.5 (in the same fashion as those in Tables 17.1 and 17.2 for the monetary aggregates), indicate that velocity (irrespective of how it is measured) is in general acyclical.

Table 17.3. H-P Cyclical Correlations of the Price Level, Treasury Bill Rate, and Federal Funds Rate with Industrial Production and the Unemployment Rate

Series	$j=-24$	$j=-18$	$j=-12$	$j=-9$	$j=-6$	$j=-3$	$j=0$	$j=3$	$j=6$	$j=9$	$j=12$	$j=18$	$j=24$
A. Cross correlations with industrial production: $\rho(x_t, y_{t+j})$, $j = -24, -18, -12, -9, -6, -3, 0, 3, 6, 9, 12, 18, 24$													
Price level	.042	.038	.022	.003	−.010	−.024	−.037	−.052	−.072	−.074	−.070	−.039	−.013
Treasury bill rate	−.000	.015	.032	.044	.050	.051	.038	.013	−.012	−.032	−.049	−.059	−.043
Federal funds rate	−.006	.016	.042	.054	.056	.052	.041	.013	−.014	−.034	−.049	−.057	−.041
B. Cross correlations with the unemployment rate: $\rho(x_t, u_{t+j})$, $j = -24, -18, -12, -9, -6, -3, 0, 3, 6, 9, 12, 18, 24$													
Price level	−.236	−.241	−.161	−.089	−.010	.076	.184	.287	.356	.386	.382	.284	.139
Treasury bill rate	−.013	−.084	−.151	−.200	−.227	−.259	−.265	−.174	−.058	.056	.160	.267	.255
Federal funds rate	−.003	−.083	−.161	−.219	−.253	−.288	−.277	−.173	−.039	.083	.179	.274	.254

Note: Sample period, monthly data: 1960:01 – 2006:01. $x_t =$ (Price level, Treasury bill rate, Federal funds rate).

Table 17.4. H-P Cyclical Correlations of Velocity of Money Measures with Industrial Production

	$\rho(V_t, y_{t+j})$, $j = -24, -18, -12, -9, -6, -3, 0, 3, 6, 9, 12, 18, 24$												
Velocity Series	$j=-24$	$j=-18$	$j=-12$	$j=-9$	$j=-6$	$j=-3$	$j=0$	$j=3$	$j=6$	$j=9$	$j=12$	$j=18$	$j=24$
Sum M1	.022	.031	.031	.024	.017	.009	−.003	−.024	−.053	−.069	−.072	−.053	−.030
Divisia M1	.023	.031	.030	.023	.016	.008	−.002	−.023	−.052	−.068	−.071	−.054	−.032
CE M1	.024	.027	.030	.030	.026	.023	.001	−.013	−.037	−.050	−.057	−.048	−.032
Sum M2	.029	.035	.034	.026	.017	.008	−.003	−.024	−.051	−.066	−.071	−.060	−.041
Divisia M2	.029	.036	.034	.026	.017	.008	−.004	−.025	−.053	−.068	−.072	−.061	−.041
CE M2	−.003	.004	.017	.022	.023	.022	.012	.000	−.014	−.024	−.032	−.034	−.022
Sum M3	.027	.029	.026	.016	.007	−.000	−.010	−.026	−.047	−.059	−.063	−.056	−.041
Divisia M3	.030	.034	.029	.020	.010	.000	−.010	−.028	−.052	−.065	−.068	−.058	−.038
CE M3	−.005	.005	.020	.027	.029	.029	.022	.009	−.010	−.025	−.038	−.044	−.030
Sum MZM	.024	.033	.038	.038	.026	.016	.002	−.012	−.050	−.067	−.073	−.061	−.038
Divisia MZM	.024	.033	.037	.031	.023	.014	.000	−.023	−.053	−.070	−.075	−.062	−.038
CE MZM	−.002	.007	.022	.027	.024	.020	.004	−.010	−.028	−.039	−.045	−.038	−.021

Note: Sample period, monthly data: 1960:01 – 2006:01.

Table 17.5. H-P Cyclical Correlations of Velocity of Money Measures with the Unemployment Rate

Series	$\rho(V_t, u_{t+j})$, $j = -24, -18, -12, -9, -6, -3, 0, 3, 6, 9, 12, 18, 24$												
	$j=-24$	$j=-18$	$j=-12$	$j=-9$	$j=-6$	$j=-3$	$j=0$	$j=3$	$j=6$	$j=9$	$j=12$	$j=18$	$j=24$
Sum M1	−.138	−.199	−.204	−.184	−.144	−.093	−.015	.089	.198	.270	.311	.296	.215
Divisia M1	−.144	−.204	−.207	−.186	−.145	−.094	−.018	.083	.190	.261	.302	.292	.218
CE M1	−.096	−.168	−.173	−.159	−.127	−.098	−.052	.019	.104	.171	.218	.235	.188
Sum M2	−.128	−.201	−.207	−.185	−.146	−.098	−.023	.082	.192	.267	.308	.297	.221
Divisia M2	−.136	−.201	−.202	−.181	−.141	−.092	−.015	.092	.202	.277	.319	.308	.225
CE M2	−.015	−.040	−.068	−.085	−.103	−.122	−.122	−.086	−.040	.004	.054	.116	.115
Sum M3	−.126	−.177	−.163	−.134	−.091	−.046	.018	.106	.197	.256	.289	.274	.201
Divisia M3	−.145	−.190	−.172	−.143	−.097	−.047	.024	.120	.218	.283	.316	.295	.203
CE M3	.004	−.035	−.080	−.107	−.136	−.165	−.177	−.144	−.087	−.023	.044	.134	.145
Sum MZM	−.125	−.208	−.253	−.246	−.210	−.158	−.070	.057	.184	.275	.327	.319	.238
Divisia MZM	−.130	−.204	−.239	−.229	−.192	−.140	−.051	.075	.200	.289	.338	.325	.235
CE MZM	−.042	−.080	−.114	−.124	−.122	−.118	−.084	−.015	.047	.096	.133	.143	.116

Note: Sample period, monthly data: 1960:01 – 2006:01.

17.4 Robustness

We have characterized the key nominal features of U.S. business cycles using a modern counterpart of the methods developed by Burns and Mitchell (1946) — HP cyclical components. The HP filter is almost universally used in the real business cycle research program and extracts a long-run component from the data. HP filtering, however, has recently been questioned as a unique method of trend elimination. For example, King and Rebelo (1993) argue that HP filtering may seriously change measures of persistence, variability, and comovement. They also give a number of examples that demonstrate that the dynamics of HP filtered data can differ significantly from the dynamics of differenced or detrended data.

Also, Cogley and Nason (1995) in analyzing the effect of HP filtering on trend- and difference-stationary time series, argue that the interpretation of HP stylized facts depends on assumptions about the time series properties of the original data. For example, when the original data are trend stationary, the HP filter operates like a high pass filter (removes the low frequency components and allows the high frequency components to pass through) on deviations from trend. However, when the original data are difference stationary, the HP filter does not operate like a high pass filter. In this case, HP stylized facts about periodicity and comovement are determined primarily by the filter and reveal very little about the dynamic properties of the original data.

More recently, two other filters have been proposed in the literature — the Baxter and King (1999) and the Christiano and Fitzgerald (2003) filters, to which we now briefly turn.

17.4.1 The Baxter and King Filter

The Hodrick and Prescott filter is a high-pass filter that passes through frequencies higher than a chosen cut-off frequency, ω. The Baxter and King (1999) filter improves upon the Hodrick and Prescott filter. It is a band-pass filter, with a gain function that takes the value 1 for all frequencies in the desired band, the interval $[\omega_1, \omega_2]$, and the value 0 for all other frequencies. Although an infinite number of observations is required to construct an ideal filter, Baxter and King (1999) approximate the ideal filter y_t with a finite and symmetric moving average filter over $2K+1$ periods, as follows

$$\widehat{y}_t = \sum_{j=-K}^{K} \alpha_j L^j x_{t-j},$$

where $\widehat{y}_t$ is the filtered series, L^j is the lag operator of order j, α_j are the filter weights, K is the order of the approximation, and the x_t's are the observed data. The filter weights are chosen in the frequency domain, by minimizing the difference between the ideal but unfeasible filter $A(\omega)$ and the proposed feasible filter $B(\omega)$, as follows [see Baxter and King (1999) for more details],

$$\min \frac{1}{2\pi} \int_{-\pi}^{\pi} \left| A(\omega) - B(\omega) \right|^2,$$

with

$$B(\omega) = \sum_{j=-K}^{K} g_j \exp\left(-i\omega j\right).$$

17.4.2 The Christiano and Fitzerland Filter

The Christiano and Fitzgerald (2003) filter is an improved version of the Baxter and King (1999) filter. It approximates the ideal filter y_t by a linear function $\widehat{y}_t$ of the observed data, $x \equiv (x_1, \cdots, x_T)$. In doing so, it chooses the filter weights to make $\widehat{y}_t$ as close as possible to y_t by minimizing the mean criterion

$$E\left[(y_t - \widehat{y}_t)^2 \Big| x \right],$$

where E is evaluated using the time series properties of x. The Christiano and Fitzgerald (2003) recommended solution for the filtered series is given by

$$\widehat{y}_t = \beta_0 x_t + \beta_1 x_{t+1} + \cdots + \beta_{T-1-t} x_{T-1}$$

$$+ \widetilde{\beta}_{T-t} x_T + \beta_1 x_{t-1} + ... + \beta_{t-2} x_2 + \widetilde{\beta}_{t-1} x_1$$

for $t = 3, 4, \cdots, T-2$, where the β_j's are defined as in Baxter and King (1999) and $\widetilde{\beta}_{T-t}$ and $\widetilde{\beta}_{t-1}$ are linear functions of the β_j's — see Christiano and Fitzgerald (2003) for more details.

17.5 Conclusion

We have investigated the cyclical behavior of United States money, prices, short term nominal interest rates, and velocity, using the methodology of Kydland and Prescott (1990). Although we have not investigated the robustness of our results to alternative filtering procedures, we believe that the results reported here, based on the Hodrick-Prescott filter (and our monthly data), are reasonably robust across business cycle filters. In fact, as Baxter and King (1999) argue HP filtering can produce reasonable approximations to an ideal business cycle filter.

Based on monthly data, the results are (in general) robust to alternative measures of the cycle and match recent evidence regarding the cyclical behavior of the price level. We also found that short-term nominal interest rates are weakly procyclical and that money and velocity (however measured) are acyclical. These findings do not support a monetary effect on the cycle and illustrate the importance of constructing theoretically meaningful monetary aggregates — an issue that we will attempt to deal with in more detail in the rest of this book.

Part 6: Microfoundations and the Demand for Money

Overview of Part 6

In Chapters 18 and 19, we move from the theoretical model of utility maximization (presented in Chapters 15 and 16) to the inter-related problems of monetary aggregation and estimation of money demand. To achieve this, we conduct the analysis within a microtheoretical framework, and discuss nonparametric and parametric approaches to demand analysis. The 'Slutsky conditions' that every demand system should satisfy, irrespective of the form of the utility function, are also discussed.

In Chapters 20 and 21, we discuss classes of flexible functional forms for utility and demand systems that play important roles in the parametric approach to empirical demand analysis. In particular, we discuss five locally flexible forms and two asumptotically globally flexible forms. In doing so, we pay explicit attention to the theoretical regularity conditions of positivity, monotonicity, and curvature and argue that much of the older empirical literature ignores economic regularity and hence has to be disregarded.

18

The Nonparametric Approach to Demand Analysis

In Chapter 15 we showed the steps that are normally taken to reduce a very general consumer choice problem to a monetary asset choice problem. At this point, we are prepared to proceed and develop the microeconomic- and aggregation-theoretic literature on the demand for money and monetary assets. This is achieved by conducting the analysis within a microtheoretical framework, making use of a number of theoretical advances in a set of related theories — revealed preference, index numbers, duality, separability, and demand systems.

The standard approach to applied demand analysis is *parametric*, in the sense that it postulates parametric forms for the utility function and fits the derived demand functions to observed data. The estimated demand functions can then be tested for consistency with the utility-maximizing hypothesis underlying the model, used to estimate price and substitution elasticities, or used to forecast behavior for other price configurations. As Varian (1982, p. 945) puts it, this approach "will be satisfactory only when the postulated parametric forms are good approximations to the 'true' demand functions."

An alternative approach to demand analysis is *nonparametric*, in the sense that it requires no specification of the form of the demand functions. This approach, fully developed by Varian (1982) in his article, "The Nonparametric Approach to Demand Analysis," deals with the raw data itself using techniques of finite mathematics. It typically addresses three issues concerning consumer behavior: (i) consistency of observed behavior with the preference maximization model; (ii) the recovering of preferences, given observations on consumer behavior; and (iii) the forecasting of demand for different price configurations. However, there are also advantages and disadvantages of this approach to demand analysis, as we will discuss later in this chapter.

Let us now turn to a detailed discussion of the nonparametric approach to the demand for liquid assets, leaving a discussion of the parametric approach for the next chapter and the rest of this book.

18.1 The Idea of Revealed Preference

Consider the n-vector $\boldsymbol{x}$ of monetary assets and its corresponding n-vector of user costs, $\boldsymbol{p}$. Suppose also that we have T observations on these quantities and user costs. Let $\boldsymbol{x}^i = (x_1^i, \cdots, x_n^i)$ denote the ith observation of $\boldsymbol{x}$ and let $\boldsymbol{p}^i = (p_1^i, \cdots, p_n^i)$ be the associated user costs, $i = 1, \cdots, T$. Let us consider the following definitions from Varian (1982, 1983).

Definition 18.1. *An observation $\boldsymbol{x}^i$ is directly revealed preferred to a bundle $\boldsymbol{x}$, written $\boldsymbol{x}^i R^0 \boldsymbol{x}$, if $\boldsymbol{p}^i\boldsymbol{x}^i \geq \boldsymbol{p}^i\boldsymbol{x}$. An observation $\boldsymbol{x}^i$ is revealed preferred to a bundle $\boldsymbol{x}$, written $\boldsymbol{x}^i R\boldsymbol{x}$, if there is a sequence of observations $\left(\boldsymbol{x}^j, \boldsymbol{x}^k, \cdots, \boldsymbol{x}^l\right)$ such that $\boldsymbol{x}^i R^0 \boldsymbol{x}^j$, $\boldsymbol{x}^j R^0 \boldsymbol{x}^k$, $\cdots$, $\boldsymbol{x}^l R^0 \boldsymbol{x}$.*

Note that revealed preference is a relation that holds between the optimal bundle at some budget and anything else the consumer could have bought at the given budget.

Definition 18.2. *The data satisfies the Generalized Axiom of Revealed Preference (GARP) if $\boldsymbol{x}^i R\boldsymbol{x}^j$ implies $\boldsymbol{p}^j\boldsymbol{x}^j \leq \boldsymbol{p}^j\boldsymbol{x}^i$.*

What this definition tells us is that the set of choices $\boldsymbol{x}^i$ is revealed to be preferred to $\boldsymbol{x}^j$ if the expenditures on $\boldsymbol{x}^i$ exceed or are equal to those on $\boldsymbol{x}^j$ evaluated at the original set of prices, where i and j refer to dates (not necessarily consecutive). Note from the above definition that GARP is a necessary and sufficient condition for observed demand data to be consistent with utility maximization.

18.2 The Maximization Hypothesis

The importance of GARP is that we can use it to see if there is a utility function that has generated a given set of data. As we just mentioned, consistency with GARP is a necessary and sufficient condition for the existence of a well-behaved utility function.

Definition 18.3. *A utility function* $f(\boldsymbol{x})$ *rationalizes the data* $\left(\boldsymbol{p}^i, \boldsymbol{x}^i\right)$, $i = 1, ..., T$, *if* $f(\boldsymbol{x}^i) \geq f(\boldsymbol{x})$ *for all* $\boldsymbol{x}$ *such that* $\boldsymbol{p}^i\boldsymbol{x}^i \geq \boldsymbol{p}^i\boldsymbol{x}$, *for* $i \geq 1, \cdots, T$.

This definition simply means that $f(\cdot)$ is consistent with the data if observed consumption would be optimal under $f(\cdot)$. Varian (1982) developed methods for examining whether any such utility function exists for a given data set based on the following theorem due to Afriat (1967) and Diewert (1973b).

Theorem 18.4. *The following conditions are equivalent: (1) there exists a nonsatiated utility function that rationalizes the data; (2) the data satisfies GARP; (3) there exist numbers* $U^i, \lambda^i > 0$, $i = 1, \cdots, T$ *that satisfy the Afriat inequalities:*

$$U^i \leq U^j + \lambda^j \boldsymbol{p}^j (\boldsymbol{x}^i - \boldsymbol{x}^j),$$

for $i, j = 1, ..., T$; *(4) there exists a concave, monotonic, continuous, nonsatiated utility function that rationalizes the data.*

By parts 1 and 4 of the theorem, if there exists a rationalizing utility function, it will have the properties typically assumed in consumer theory. By parts 1 and 2, some rationalizing utility function exists if and only if the data satisfy GARP. By parts 2 and 3, one can test for GARP by examining if there exist numbers $U^i > 0$ and $\lambda^i > 0$ that satisfy the Afriat inequalities.

18.3 Homotheticity

Varian (1983) also developed the Homothetic Axiom of Revealed Preference (HARP), to test for consistency with homothetic preferences. Let's consider the following definition.

Definition 18.5. *A function $f(\boldsymbol{x}) : \mathbb{R}^n \to \mathbb{R}$ is homothetic if it is a linear monotonic transformation of a linearly homogeneous (i.e., homogeneous of degree 1) function.*

That is, $f(\boldsymbol{x})$ is homothetic if it can be written as $f(\boldsymbol{x}) = g\left(h(\boldsymbol{x})\right)$, where $h(\boldsymbol{x})$ is homogeneous of degree 1 and $g(h)$ is positive monotonic. Normalizing the prices by the level of expenditure at each observation, so that $p^i x^i = y^i = 1$, for $i = 1, \cdots, T$, consider the following theorem from Varian (1983):

Theorem 18.6. *The following conditions are equivalent: (1) there exists a nonsatiated homothetic utility function that rationalizes the data; (2) the data satisfies HARP: for all distinct choices of indexes $(i, j, \cdots, m)$ we have*

$$\left(\boldsymbol{p}^i \boldsymbol{x}^j\right)\left(\boldsymbol{p}^j \boldsymbol{x}^k\right) \cdots \left(\boldsymbol{p}^m \boldsymbol{x}^i\right) \geq 1;$$

(3) there exist numbers $U^i > 0$, $i = 1, \cdots, T$ such that

$$U^i \leq U^j \boldsymbol{p}^j \boldsymbol{x}^i,$$

for $i, j = 1, ..., T$; (4) there exists a concave, monotonic, continuous, nonsatiated, homothetic utility function that rationalizes the data.

18.4 Direct Separability

The notion of separability is of considerable importance, because it provides a means of justifying the use of monetary aggregates. It also resolves the statistical problem caused by the lack of degrees of freedom, since it rationalizes the estimation of a smaller set of demand equations when one takes a parametric approach to demand analysis (as we shall do in the rest of this book). In the context of preference structures there are different separability concepts, giving rise to both different grouping patterns and different behavioral implications — for a good exposition of alternative forms of separability and their behavioral implications, see Pudney (1981). Here we deal with the utility relation expressed in the direct form.

Let $I = (1, 2, \cdots, n)$ be a set of integers that identify the variables over which preferences are defined and consider the partition of I into two subsets

$$I = \{I^c, I^r\}$$

such that $I^c \cup I^r = I$, $I^c \cap I^r = \emptyset$, $I^c \neq \emptyset$, and $I^r \neq \emptyset$. Corresponding to the binary partition I, denote vectors in Ω^n in ways that reflect the partition. In particular, express Ω^n as a Cartesian product of the subspaces

$$\Omega^n = \Omega^{(c)} \times \Omega^{(r)}$$

with the dimensions of $\Omega^{(c)}$ and $\Omega^{(r)}$ given by the cardinalities of I^c and I^r, respectively. An asset vector, $\boldsymbol{x} \in \Omega^n$, can be written as $\boldsymbol{x} = (\boldsymbol{x}^c, \boldsymbol{x}^r)$ and if the ith asset is in the rth category, then x_i is a component of the vector $\boldsymbol{x}^r \in \Omega^{(r)}$. Consider the following definition, adapted from Blackorby, Primont, and Russell (1978) — Blackorby *et al.* (1978) should be consulted as a definite source concerning separability, duality, and functional structure.

Definition 18.7. *I^r is weakly separable in $f(\boldsymbol{x})$ if and only if there exist functions*

$$f^r : \Omega^{(r)} \to \mathbb{R}$$

and

$$\overline{f} : \Omega^{(r)} \times \mathcal{R}(f^r) \to \mathbb{R},$$

where $\mathcal{R}(f^r)$ is the range of f^r, such that

$$f(\boldsymbol{x}) = \overline{f}(\boldsymbol{x}^c, f^r(\boldsymbol{x}^r)),$$

where $\overline{f}$ is strictly inreasing in $f^r(\boldsymbol{x}^r)$.

According to this theorem, $f(\boldsymbol{x})$ is called the *parent function*, $\overline{f}(\boldsymbol{x}^c, f^r(\boldsymbol{x}^r))$ the *macro function*, and $f^r(\boldsymbol{x}^r)$ the *aggregator function.* The requirement of weak separability is that the marginal rate of substitution between any two assets in a separable component group be invariant with respect to any asset outside the group. Algebraically, assets i and j are separable from asset k, if and only if

$$\frac{\partial}{\partial x_k}\left(\frac{f_i(\boldsymbol{x})}{f_j(\boldsymbol{x})}\right) = 0, \quad \forall \boldsymbol{x} \in \Omega^n,$$

where $f_i(\boldsymbol{x})$ is the marginal utility of asset i and $f_j(\boldsymbol{x})$ the marginal utility of asset j. Equivalently, assets i and j are separable from asset k if and only if

$$\frac{f_{ik}(\boldsymbol{x})}{f_i(\boldsymbol{x})} = \frac{f_{jk}(\boldsymbol{x})}{f_j(\boldsymbol{x})},$$

which is the explicit differential statement of the previous equation and $f_{ik}(\boldsymbol{x})$ and $f_{jk}(\boldsymbol{x})$ are the cross partial derivatives. Note that a sufficient condition for weak separability is perfect substitutability. Under perfect substitutability, the ratio of marginal utilities is constant and hence invariant to any asset change.

If the utility function $f(\boldsymbol{x})$ is weakly separable in $\boldsymbol{x}^r$, its value does not depend on the elements of $\boldsymbol{x}^r$ individually, but rather on the quantity index $f^r(\boldsymbol{x}^r)$. As we discussed in Chapter 15, weak separability implies a sequential expenditure allocation, where in the first stage the consumer divides expenditures on $\boldsymbol{x}^c$ and $\boldsymbol{x}^r$, and then decides the optimal allocation of expenditure among the elements of $\boldsymbol{x}^r$ independently of the choices of the elements of $\boldsymbol{x}^c$. In other words, when the function $f(\boldsymbol{x})$ is weakly separable in $\boldsymbol{x}^r$, the econometrician can estimate a separable demand system in the assets of $\boldsymbol{x}^r$, disregarding the assets in $\boldsymbol{x}^c$.

Varian (1982, 1983) provides two tests of weak separability. The first test, which is the weaker of the two, is a test of the necessary conditions for weak separability. It checks if the subdata in group r satisfy GARP. In particular, since each observation in group r must solve the problem

$$\max_{\boldsymbol{x}^r} f^r(\boldsymbol{x}^r) \quad \text{subject to} \quad \boldsymbol{p}^r\boldsymbol{x}^r \geq \boldsymbol{p}^r\boldsymbol{x},$$

it is necessary for separability that Afriat numbers exist for the data in $\boldsymbol{x}^r$. Otherwise the aggregator function $f^r(\boldsymbol{x}^r)$ does not exist. The second, stronger test checks necessary and sufficient conditions for weak separability. It does so, by checking if the Afriat numbers for the aggregator function $f^r(\boldsymbol{x}^r)$ are consistent with those for the parent function $f(\boldsymbol{x})$. The following theorem from Varian (1983) states the relationship precisely.

Theorem 18.8. *The following conditions are equivalent: (1) there exists a weakly separable concave, monotonic, continuous, nonsatiated utility function that rationalizes the data; (2) there exist numbers* $U^i, W^i, \lambda^i > 0, \mu^i > 0$, $i = 1, \cdots, T$ *that satisfy:*

$$U^i \leq U^j + \lambda^j \boldsymbol{p}^j(\boldsymbol{x}^i - \boldsymbol{x}^j) + \frac{\lambda^j}{\mu^j}(W^i - W^j),$$

$$W^i \leq W^j + \mu^j \boldsymbol{q}^j(\boldsymbol{z}^i - \boldsymbol{z}^j),$$

for $i, j = 1, ..., T$; *(3) the data* $(\boldsymbol{q}^j \boldsymbol{z}^j)$ *and* $\left(\boldsymbol{p}^i, 1/\mu^i; \boldsymbol{x}^i, W^i\right)$ *satisfy GARP for some choices of* $\left(W^i, \mu^i\right)$ *that satisfy the Afriat inequalities.*

Note that part 2 of the theorem provides the means for testing the necessary and sufficient conditions for direct weak separability — clearly one must construct two sets of interrelated Afriat numbers.

18.5 Indirect Separability

The utility function $f(\boldsymbol{x})$ introduced above is *direct* — it has the actual quantities of liquid assets used, x_i, $i = 1, ..., n$, as arguments. Alternatively, preferences may be represented by the indirect utility function, which indicates the indifference curve attainable at prices $\boldsymbol{p}$ and total expenditure y. In particular, the utility maximization problem

$$\max_{\boldsymbol{x}} f(\boldsymbol{x}) \quad \text{subject to} \quad \boldsymbol{p}'\boldsymbol{x} = y,$$

can be reformulated equivalently as

$$\max_{\boldsymbol{x}} f(\boldsymbol{x}) \quad \text{subject to} \quad \sum_{i=1}^{n} v_i x_i = 1. \tag{18.1}$$

where $v_i = p_i/y$, $i = 1, ..., n$ denotes 'expenditure-normalized' user costs. In this reformulated version, the maximization problem has two sets of n variables: monetary asset services, with values $\boldsymbol{x} = (x_1, ..., x_n)$, and normalized monetary asset user costs, with values $\boldsymbol{v} = (v_1, ..., v_n)$.

As we shall see in the next chapter, the solution to (18.1) is the system of demand functions

$$x_i = x_i(v_1, ..., v_n), \quad i = 1, ..., n. \tag{18.2}$$

Substituting solution (18.2) into the objective function yields the maximum attainable utility given normalized monetary asset user costs

$$h(\boldsymbol{v}) = h\left(x_1(\boldsymbol{v}), ..., x_n(\boldsymbol{v})\right),$$

where $h(\boldsymbol{v})$ is quasi convex, continuous, and decreasing. The function $h(\boldsymbol{v})$ reflects the fact that utility depends indirectly on prices and income rather than on quantities. For this reason, $h(\boldsymbol{v})$ is called the *indirect* utility function.

The important thing is that the direct utility function and the indirect utility function are equivalent representations of the underlying preference ordering. However, a structural property of the direct utility function does not imply the same property of the indirect utility function. The point to be stressed here is that the behavioral implications of direct separability are different from those of indirect separability. Let's consider the following definition of indirect weak separability.

Definition 18.9. *I^r is weakly separable in $h(\boldsymbol{v})$ from I^c if and only if there exist continuous functions*

$$h^r : \Omega_+^{(r)} \to \mathbb{R}$$

and

$$\overline{h} : \Omega_+^{(c)} \times \mathcal{R}(V^r) \to \mathbb{R},$$

where $\mathcal{R}(h^r)$ is the range of h^r, such that

$$h(\boldsymbol{v}) = \overline{h}(\boldsymbol{v}^c, h^r(\boldsymbol{v}^r)),$$

where $\overline{h}$ is nondecreasing in $h^r(\boldsymbol{v}^r)$.

The algebraic requirement of indirect weak separability is that

$$\frac{\partial}{\partial v_k}\left(\frac{h_i(\boldsymbol{v})}{h_j(\boldsymbol{v})}\right) = 0, \quad \forall \boldsymbol{v} \in \Omega^n,$$

or using Roy's identity (to be discussed in detail in the next chapter)

$$\frac{\partial}{\partial v_k}\left(\frac{x_i(\boldsymbol{v})}{x_j(\boldsymbol{v})}\right) = 0.$$

Also, since $h(\boldsymbol{v}) = h(\boldsymbol{p}, y)$, due to homogeneity of degree zero in $\boldsymbol{p}$ and y, the last equation can be rewritten as

$$\frac{\partial}{\partial p_k}\left(\frac{x_i(\boldsymbol{v})}{x_j(\boldsymbol{v})}\right) = 0,$$

which implies that the optimal asset ratios in I^r are independent of the kth price. Equivalently, the above can be written as

$$x_j(\boldsymbol{v})\left(\frac{\partial x_i(\boldsymbol{v})}{\partial p_k}\right) = x_i(\boldsymbol{v})\left(\frac{\partial x_j(\boldsymbol{v})}{\partial p_k}\right),$$

or

$$\eta_{ik} = \eta_{jk}, \tag{18.3}$$

where η_{ik} and η_{jk} are the cross price elasticities of assets i and j with respect to the kth price, respectively. Hence, indirect weak separability implies an equality restriction on the cross price elasticities.

18.6 Homothetic Separability

Separability is a characteristic of functional structure that does not necessarily carry over from direct to indirect utility functions. In particular, a separable direct utility function implies a different preference ordering than a separable indirect utility function. However, Lau (1970) has shown that if the direct utility function is homothetically separable, then the indirect utility function will have the same structure with respect to normalized prices. Hence, testing for homothetic separability is of considerable practical interest, since homothetic separability is a sufficient condition for simultaneous separability of $f(\boldsymbol{x})$ and $h(\boldsymbol{v})$.

Moreover, direct separability establishes only a necessary condition for aggregation in its simplest form. In particular, if we wish to measure the subaggregate $f^r(\boldsymbol{x}^r)$ using the most elementary method, we would require the additional assumption that $f^r(\boldsymbol{x}^r)$ be homothetic. In fact, homothetic weak separability is necessary and sufficient for the existence of the simplified form of subaggregation. Let us now define homothetic separability of the direct utility function.

Definition 18.10. *I^r is homothetically weakly separable in $f(\boldsymbol{x})$ if and only if there exist functions*

$$f^r : \Omega^{(r)} \to \mathbb{R}$$

and

$$\overline{f} : \Omega^{(r)} \times \mathcal{R}(f^r) \to \mathbb{R},$$

where $\mathcal{R}(f^r)$ is the range of f^r, such that

$$f(\boldsymbol{x}) = \overline{f}(\boldsymbol{x}^c, f^r(\boldsymbol{x}^r)),$$

where $\overline{f}$ is nondecreasing in $f^r(\boldsymbol{x}^r)$ and $f^r(\boldsymbol{x}^r)$ is homothetic.

Varian (1983) provides the following theorem to test for direct homothetic separability, based on the Diewert and Parkan (1978) suggested procedure.

Theorem 18.11. *The following conditions are equivalent: (1) there exists a homothetically weakly separable, concave, monotonic, continuous, nonsatiated utility function that rationalizes the data; (2) the data $\left(\boldsymbol{p}^i, 1/W^i; \boldsymbol{x}^i, W^i\right)$ satisfy GARP for some choices of W^i that satisfies the homotheticity inequalities; (3) there exist numbers $U^i, W^i, \lambda^i > 0$, $i = 1, \cdots, T$ that satisfy*

$$U^i \leq U^j + \lambda^j \boldsymbol{p}^j(\boldsymbol{x}^i - \boldsymbol{x}^j) + \frac{\lambda^j}{W^j}(W^i - W^j),$$

$$W^i \leq W^j \boldsymbol{q}^j \boldsymbol{z}^j,$$

for $i, j = 1, ..., T$.

18.7 NONPAR Tests of Consumer Behavior

GARP, homotheticity, direct separability, and direct homothetic separability are quite simple to test, using Varian's nonparametric (NONPAR) computational package. As we have argued, this approach to applied demand analysis imposes no functional form restrictions and requires only actual market data (quantities and user costs) generated by the consumers of financial services.

For example, in the case of GARP, NONPAR takes advantage of the formulation,

$$\text{if } \boldsymbol{x}^i R \boldsymbol{x}^j \text{ then } \boldsymbol{p}^j \boldsymbol{x}^j \leq \boldsymbol{p}^j \boldsymbol{x}^i,$$

and evaluates all pairs in the data (which are, of course, finite) in order to see if the expenditures on $\boldsymbol{x}^i$, evaluated at $\boldsymbol{p}^j$, are greater than those on $\boldsymbol{x}^j$, evaluated at the same price, for all $i, j = 1, \cdots, T$. In doing so, NONPAR reports the number of violations (reversals of the inequality), which can be considerable if the data set contains assets held by different sorts of economic agents, as possibly by consumers versus business firms.

What this suggests is that we can take a revealed preference approach to monetary aggregation. In particular, we can use the observed monetary data and their respective user costs to find sets of assets that are consistent with preference maximization and then test for direct separability to see which grouping has been used by money-holders in practice. Thus, if the data for $\boldsymbol{x}$ satisfies GARP, then there is a utility function $f(\boldsymbol{x})$ that rationalizes the data. There may also be an aggregator function $f^r(\boldsymbol{x}^r)$ such that $f(\boldsymbol{x})$ is weakly separable. GARP will still be satisfied by $f(\boldsymbol{x})$, but $f^r(\boldsymbol{x}^r)$ not only must satisfy GARP but must also be homothetically weakly separable, in order for a monetary aggregate to behave like an elementary asset.

The nonparametric approach to demand analysis has been used in numerous recent papers, such as, for example, Fleissig, Hall, and Seater (2000), Fisher (1989, Chapter 1), Fisher and Fleissig (1997), Swofford and Whitney (1986, 1987, 1988, 1994), Serletis and Rangel-Ruiz (2001), Fleissig and Whitney (2003, 2005), de Peretti (2005), Jones and de Peretti (2005), and Jones *et al.* (2005). Note, however, that the nonparametric approach to demand analysis is not without problems. As Fleissig, Hall, and Seater (2000, p. 329) put it,

> "[t]he NONPAR tests have advantages and disadvantages. The main advantage is that the tests are non-parametric; one need not specify the form of the utility function. Also, the tests can handle a large number of goods. The main disadvantage is that the tests are non-stochastic. Violations are all or nothing; either there is a utility function that rationalizes the data or there is not. We therefore must be especially careful about the possibility of measurement error. If there were no measurement error, then any observed rejection of GARP would be a genuine rejection. However, with measurement error, false rejections may occur, and without a distribution theory for the tests, we cannot judge the importance of observed rejections by the conventional significance tests."

Thus, establishing consistency with preference maximization and the existence of consistent monetary aggregates, using Varian's (1982, 1983) nonparametric techniques of revealed preference analysis, is a very strong standard, and it is not surprising that most recent studies of the demand for money cannot rationalize a well-behaved utility function over liquid assets. Of course, the nonparametric revealed preference analysis has implications for the parametric analysis, but for these implications to be fully investigated it is necessary that the nonparametric approach rationalizes a well-behaved utility function over monetary assets over long samples, to enable the estimation of large demand systems like the ones used in this paper.

18.8 Conclusion

The nonparametric techniques of revealed preference analysis we have sketched out above suggest one approach to testing for consistency with preference maximization and for the existence of consistent monetary aggregates. As we saw, however, there is a bias toward rejection, since the GARP test is all or nothing — even one inconsistency leads to total rejection of GARP. Hence, passing the GARP test is a very strong standard and it is not surprising that most recent studies of the demand for money cannot rationalize a well-behaved utility function over monetary assets.

It remains now to assess the significance of the parametric approach to demand analysis. This matter is the subject of the rest of this book.

19

The Parametric Approach to the Demand for Monetary Assets

The parametric approach to applied demand analysis involves postulating parametric forms for the utility function and then fitting the resulting demand functions to a finite number of observations on consumer behavior. As we argued earlier, this approach will be satisfactory only when the postulated parametric forms are good approximations to the generating demand functions.

Our approach in this chapter addresses the question of how to derive a set of demand functions for monetary assets from a framework in which the representative asset holder maximizes the monetary services utility function, $f(\boldsymbol{x})$, subject to the budget constraint. We will state the problem first and then show why and how duality theory might be employed explicitly in the rationalization of estimable demand functions.

19.1 The Direct Utility Approach

In chapter 15 we saw that the neoclassical monetary problem is that of choosing a bundle of monetary services, given the utility function and the budget constraint. So the problem is

$$\max_{\boldsymbol{x}} f(\boldsymbol{x}) \quad \text{subject to} \quad \boldsymbol{p}'\boldsymbol{x} = y, \tag{19.1}$$

with the following $(n+1)$ first-order conditions for utility maximization

$$f_i(\boldsymbol{x}) = \lambda p_i, \quad i = 1, ..., n;$$

$$\boldsymbol{p}'\boldsymbol{x} = y,$$

where $f_i(\boldsymbol{x})$ is the marginal utility of asset i, $f_i(\boldsymbol{x}) = \partial f(\boldsymbol{x})/\partial x_i$.

The first order conditions can be solved for the n optimal (i.e., equilibrium) values of x_i

$$x_i = x_i(\boldsymbol{p}, y), \quad i = 1, ..., n, \tag{19.2}$$

and the optimal value of λ

$$\lambda = \lambda(\boldsymbol{p}, y).$$

System (19.2) is the *demand system*, giving the quantity demanded as a function of the prices of all assets and income.

As an example, consider the Cobb-Douglas utility function introduced in Chapter 15, but take logs to obtain

$$\log f(\boldsymbol{x}) = \sum_{i=1}^{n} \alpha_i \log x_i.$$

To find the demand system, we need to maximize the utility function subject to the budget constraint

$$\sum_{i=1}^{n} p_i x_i = y.$$

Forming the Lagrangian, deriving the first order conditions, and rearranging these conditions, we obtain the following demand system for the Cobb-Douglas utility function

$$x_i = \frac{\alpha_i}{\sum_{i=1}^{n} \alpha_i} \frac{y}{p_i}, \quad i = 1, ..., n.$$

Demand systems like (19.2) are the systems whose parameters we will want to estimate and whose properties we will want to analyze, although this is not our intention in this book. As you will see shortly, there are further techniques in the microeconomic literature which can be used on this problem.

We should also note that the form of the demand system is determined by the utility function. For example, the demand system for the two-asset ($n = 2$) constant elasticity of substitution (CES) utility function (introduced in chapter 15),

$$f(\boldsymbol{x}) = \sum_{i=1}^{n}(a_i x_i^r)^{1/r},$$

where $0 < a_i < 1$, $-\infty < r < 1$, with $\alpha_1 = \alpha_2 = 1$ for simplicity, can be shown to have the following form

$$x_1 = \frac{p_1^{1/(r-1)}y}{p_1^{r/(r-1)} + p_2^{r/(r-1)}} \quad \text{and} \quad x_2 = \frac{p_2^{1/(r-1)}y}{p_1^{r/(r-1)} + p_2^{r/(r-1)}}.$$

That is, when the utility function changes (or rather when preferences change) the demand system also changes. Empirical demand analysis is mainly concerned with the identification of preferences and the estimation of the parameters of demand systems such as the above.

19.2 The Indirect Utility Approach

An alternative method of deriving the demand system is from the indirect utility function, defined on prices and income (which we briefly discussed in Chapter 18). As we saw, the utility maximization problem (19.1) can be reformulated equivalently as

$$\max_{\boldsymbol{x}} f(\boldsymbol{x}) \quad \text{subject to} \quad \boldsymbol{v}'\boldsymbol{x} = 1, \tag{19.3}$$

with the following solution

$$x_i = x_i(\boldsymbol{v}), \quad i = 1, ..., n, \tag{19.4}$$

where $v_i = p_i/y$, $i = 1, ..., n$ denotes 'expenditure-normalized' prices (user costs). Substituting solution (19.4) into the objective function yields the maximum attainable utility, given normalized monetary asset user costs,

$$h(\boldsymbol{v}) = h\left(x_1(\boldsymbol{v}), ..., x_n(\boldsymbol{v})\right)$$

$$= \max_{\boldsymbol{x}} f(\boldsymbol{x}) \quad \text{subject to} \quad \boldsymbol{v}'\boldsymbol{x} = 1.$$

The indirect utility function $h(\boldsymbol{v})$ reflects the fact that utility depends indirectly on prices and income.

The important thing is that the direct utility function and the indirect utility function are equivalent representations of the underlying preference ordering. In fact, there is a *duality* relationship between the direct utility function and the indirect utility function, in the sense that maximization of $f(\boldsymbol{x})$ with respect to $\boldsymbol{x}$, with given $\boldsymbol{p}$ and y, and minimization of $h(\boldsymbol{v})$ with respect to $\boldsymbol{v}$, with given $\boldsymbol{x}$, leads to the same demand equations — see, for example, Mas-Colell, Whinston, and Green (1995).

While the direct utility function has greater intuitive appeal than the indirect utility function, being able to represent preferences by an indirect utility function is particularly appealing. This is so, because the indirect utility function has prices exogenous in explaining consumer behavior. Moreover, we can easily derive the demand system by straightforward differentiation, without having to solve a system of simultaneous equations (as is the case with the direct utility function approach).

In particular, a result known as *Roy's identity*

$$x_i = -\frac{\partial h(\boldsymbol{v})/\partial p_i}{\partial h(\boldsymbol{v})/\partial y}, \qquad i = 1, \ldots, n,$$

allows us to derive the demand system, provided, of course, that $p_i > 0$ and $y > 0$. Alternatively, we can apply the 'logarithmic form' of Roy's identity

$$s_i = -\frac{\partial \log h(\boldsymbol{v})/\partial \log p_i}{\partial \log h(\boldsymbol{v})/\partial \log y}, \qquad i = 1, \ldots, n, \tag{19.5}$$

to derive the budget share equations, where $s_i = p_i x_i / y$ is the budget share of the ith asset.

As an example of how we can use the framework to get a set of equations that we can estimate, consider the *homothetic translog* (HTL) flexible functional form — the simplest member of the translog family of flexible functional forms (to be discussed in more detail in Chapter 20 — for the indirect utility function $h(\boldsymbol{v})$. The homothetic translog indirect utility function is given by[1]

$$\log h(\boldsymbol{v}) = a_0 + \sum_{k=1}^{n} a_k \log v_k + \frac{1}{2} \sum_{k=1}^{n} \sum_{j=1}^{n} \beta_{jk} \log v_k \log v_j,$$

with the following restrictions imposed:

$$\beta_{jk} = \beta_{kj}, \quad \text{for all } k, j;$$

$$\sum_{i=1}^{n} \beta_{ik} = 0, \quad \text{for all } k;$$

$$\sum_{k=1}^{n} \alpha_k = 1.$$

This function is a generalization of the Cobb-Douglas function and reduces to it when all β_i are equal to zero. In fact, when all β_i are equal to zero, the homothetic translog decays to

$$\log h(\boldsymbol{v}) = a_0 + \sum_{k=1}^{n} a_k \log v_k,$$

which is the Cobb-Douglas, written in logs.

Application of Roy's identity in share form, equation (19.5) yields a set of share equations for the homothetic translog

$$s_i = \alpha_i + \sum_{k=1}^{n} \beta_{ik} \log v_k, \quad i = 1, ..., n. \tag{19.6}$$

With n assets, the n homothetic translog share equations have $n(n+3)/2$ parameters to be estimated. For example, let us assume that there are only three assets ($n = 3$). In this three-asset case the homothetic translog share equations become

[1] Recall that a homothetic function is a positive monotonic transformation of a linearly homogeneous (i.e., homogeneous of degree 1) function. That is, $f(\boldsymbol{x})$ is homothetic if and only if we can write $f(\boldsymbol{x}) = \varphi\left(g(\boldsymbol{x})\right)$, where $g(\boldsymbol{x})$ is linearly homogeneous and $\varphi(\cdot)$ is a monotonic function. Notice that since any monotonic transformation of a utility function implies the same preference ordering, a homothetic preference ordering can be equivalently represented by a linearly homogeneous utility function.

$$s_1 = \alpha_1 + \beta_{11}\log v_1 + \beta_{12}\log v_2 + \beta_{13}\log v_3;$$

$$s_2 = \alpha_2 + \beta_{12}\log v_1 + \beta_{22}\log v_2 + \beta_{23}\log v_3;$$

$$s_3 = \alpha_1 + \beta_{13}\log v_1 + \beta_{23}\log v_2 + \beta_{33}\log v_3,$$

and have 9 parameters, α_1, α_2, α_3, β_{11}, β_{12}, β_{13}, β_{22}, β_{23}, and β_{33}.

Budget share equation systems such as (19.6), written in matrix notation as

$$\boldsymbol{s}_t = \boldsymbol{\psi}(\boldsymbol{v}_t, \boldsymbol{\vartheta}),$$

where $\boldsymbol{s} = (s_1, \cdots, s_n)'$, $\boldsymbol{\psi}(\boldsymbol{v}, \boldsymbol{\vartheta}) = (\psi_1(\boldsymbol{v}, \boldsymbol{\vartheta}), \cdots, \psi_n(\boldsymbol{v}, \boldsymbol{\vartheta}))'$, $\boldsymbol{\vartheta}$ is the parameter vector to be estimated, and $\psi_i(\boldsymbol{v}, \boldsymbol{\vartheta})$ is given by the right-hand side of (19.6), are what are typically estimated. Once the parameters of the indirect utility function are estimated, we can move directly to calculations of income and price elasticities and the elasticities of substitution. We will consider the empirical implementation of such systems later in this book.

19.3 The Slutsky Conditions

In this section we analyze the properties of demand systems which result from the fact that demand systems are obtained by preference-maximizing behavior. These properties translate into mathematical restrictions on the derivatives of the demand functions and hold whatever the functional form of the utility function. It is for this reason that they are referred to as *general restrictions*. They are also known as *Slutsky conditions*, in the terminology suggested by Barten (1967) in honor of Eugene Slutsky who was the first to state them explicitly.

19.3.1 Homogeneity (of Degree Zero)

Utility maximization implies that demand functions must be homogeneous of degree zero in prices and nominal income, meaning that if all prices and income are multiplied by the same factor κ, the quantities demanded must not change. To demonstrate that homogeneity of degree zero in prices and income holds, let's consider the $(n+1)$ first order conditions for utility maximization (derived earlier):

$$f_i(\boldsymbol{x}) = \lambda p_i, \quad i = 1, ..., n; \tag{19.7}$$

$$\boldsymbol{p}'\boldsymbol{x} = y, \tag{19.8}$$

and rewrite these conditions as

$$\frac{f_i(\boldsymbol{x})}{f_n(\boldsymbol{x})} = \frac{p_i}{p_n}, \quad i = 1, ..., n-1; \tag{19.9}$$

$$\boldsymbol{p}'\boldsymbol{x} = y. \tag{19.10}$$

Multiplying all prices and income by κ, we see that κ drops out from the numerator and denominator of the right hand side of (19.9) and is also eliminated form (19.10). Hence, the first order conditions remain unchanged by proportionate changes in all prices and income. This result is summarized as follows

$$x_i(\kappa\boldsymbol{p}, \kappa y) = \kappa^0 x_i(\boldsymbol{p}, y) = x_i(\boldsymbol{p}, y),$$

for $i = 1, ...n$. That is, the demand functions are homogenous of degree zero in prices and income.

The knowledge that a system of demand equations is homogeneous of degree zero in prices and income is not very useful as such. However, it can be made operational by expressing it in terms of derivatives of the demand functions. In particular, by applying Euler's theorem to $x_i = x_i(\boldsymbol{p}, y)$, we obtain[2]

$$\sum_{j=1}^{n} p_j \frac{\partial x_i}{\partial p_j} + y\frac{\partial x_i}{\partial y} = 0, \quad i = 1, ..., n.$$

Dividing all terms of the above expression by x_i we get the following reformulated restriction, in terms of price and income elasticities

$$\sum_{j=1}^{n} \eta_{ij} = -\eta_{iy}, \quad i = 1, ..., n, \tag{19.11}$$

where η_{ij} is the elasticity of demand of asset i with respect to asset j,

[2] In general, if a function $f(\boldsymbol{x}) = f(x_1, ..., x_n)$ is homogeneous of degree γ, then Euler's theorem implies that

$$\sum_{i=1}^{n} x_i \frac{\partial f(x_1, ..., x_n)}{\partial x_i} = \gamma f(x_1, ..., x_n).$$

In the special case of homogeneity of degree zero ($\gamma = 0$), the above reduces to

$$\sum_{i=1}^{n} x_i \frac{\partial f(x_1, ..., x_n)}{\partial x_i} = 0.$$

$$\eta_{ij} = \frac{p_j}{x_i}\frac{\partial x_i}{\partial p_j},$$

and η_{iy} is the income elasticity of demand for asset i,

$$\eta_{iy} = \frac{y}{x_i}\frac{\partial x_i}{\partial y}.$$

Hence, the homogeneity condition (19.11) states that the sum of the own- and all cross-price elasticities of any asset i has to equal the negative of its income elasticity.

As already stated, homogeneity of degree zero in income and prices has to be exactly satisfied for our mathematical functions to be candidates for qualification as demand functions. Evidence that contradicts the homogeneity restriction has also important implications from a macroeconomics perspective. For example, homogeneity of degree zero with respect to prices and nominal income is an important assumption underlying classical macroeconomic theory, which requires the real side of the economy to be homogeneous of degree zero in the nominal variables.

19.3.2 Adding-Up (Summability)

Since the budget constraint has to be satisfied, the demand equations have to be such that the sum of the estimated expenditures on the different monetary assets equals total monetary asset expenditure in any period. Such a system is called *additive* — not to be confused with the additivity property of utility functions.

The adding-up restriction can also be expressed in terms of elasticities. In particular, partially differentiating (19.10) with respect to y we get

$$\sum_{i=1}^{n} p_i \frac{\partial x_i}{\partial y} = 1,$$

which implies that an increase in total expenditure is completely allocated to all monetary assets. Manipulating the above we obtain

$$\sum_{i=1}^{n} s_i \eta_{iy} = 1, \tag{19.12}$$

where $s_i = p_i x_i / y$ is the budget share of asset i. Thus, the summability condition (19.12), often called the *Engel aggregation* condition in the terminology of Frisch (1959), states that the sum of the income elasticities weighted by their respective expenditure proportions has to equal unity.

19.3.3 Symmetry of the Slutsky Matrix

Total differentiation of the first order conditions for utility maximization, conditions (19.7) and (19.8), gives

$$\begin{bmatrix} \boldsymbol{F} & \boldsymbol{p} \\ \boldsymbol{p}' & \mathbf{0} \end{bmatrix} \begin{bmatrix} d\boldsymbol{x} \\ -d\lambda \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \lambda \mathbf{I} \\ 1 & -\boldsymbol{x}' \end{bmatrix} \begin{bmatrix} dy \\ d\boldsymbol{p} \end{bmatrix}, \tag{19.13}$$

where $\boldsymbol{F}$ is the $n \times n$ Hessian matrix of the utility function,

$$\boldsymbol{F} = \begin{bmatrix} \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_1^2} & \cdots & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_1 \partial x_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_n \partial x_1} & \cdots & \dfrac{\partial^2 f(\boldsymbol{x})}{\partial x_n^2} \end{bmatrix}.$$

As already noted in Chapter 19, the solution to (19.7) and (19.8) is the demand system

$$x_i = x_i(\boldsymbol{p}, y), \quad i = 1, ..., n;$$

$$\lambda = \lambda(\boldsymbol{p}, y).$$

Total differentiation of this demand system yields

$$\begin{bmatrix} d\boldsymbol{x} \\ -d\lambda \end{bmatrix} = \begin{bmatrix} \boldsymbol{x}_y & \mathbf{X}_{\boldsymbol{p}} \\ -\lambda_y & -\lambda_{\boldsymbol{p}}' \end{bmatrix} \begin{bmatrix} dy \\ d\boldsymbol{p} \end{bmatrix}, \tag{19.14}$$

where

$$\lambda_{\boldsymbol{p}} = \begin{bmatrix} \dfrac{\partial \lambda}{\partial p_1} \\ \vdots \\ \dfrac{\partial \lambda}{\partial p_n} \end{bmatrix}, \quad \boldsymbol{x}_y = \begin{bmatrix} \dfrac{\partial x_1}{\partial y} \\ \vdots \\ \dfrac{\partial x_n}{\partial y} \end{bmatrix}, \quad \mathbf{X}_{\boldsymbol{p}} = \begin{bmatrix} \dfrac{\partial x_1}{\partial p_1} & \cdots & \dfrac{\partial x_1}{\partial p_n} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial x_n}{\partial p_1} & \cdots & \dfrac{\partial x_n}{\partial p_n} \end{bmatrix},$$

and $\lambda_y = \partial \lambda / \partial y$.

Substitution of (19.14) into (19.13) leads to

$$\begin{bmatrix} \boldsymbol{F} & \boldsymbol{p} \\ \boldsymbol{p}' & \mathbf{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_y & \mathbf{X}_{\boldsymbol{p}} \\ -\lambda_y & -\lambda_{\boldsymbol{p}}' \end{bmatrix} = \begin{bmatrix} \mathbf{0} & \lambda \mathbf{I} \\ 1 & -\boldsymbol{x}' \end{bmatrix},$$

the solution of which can be written in the form

$$\begin{bmatrix} \boldsymbol{x}_y & \mathbf{X}_{\boldsymbol{p}} \\ -\lambda_y & -\lambda'_{\boldsymbol{p}} \end{bmatrix} = \begin{bmatrix} \boldsymbol{F} & \boldsymbol{p} \\ \boldsymbol{p}' & \mathbf{0} \end{bmatrix}^{-1} \begin{bmatrix} \mathbf{0} & \lambda\mathbf{I} \\ 1 & -\boldsymbol{x}' \end{bmatrix}. \tag{19.15}$$

Equation (19.15) implies [see Barten (1964) or Phlips (1974) for details]

$$\boldsymbol{x}_y = \lambda_y \boldsymbol{F}^{-1}\boldsymbol{p};$$

$$\mathbf{X}_{\boldsymbol{p}} = \lambda \boldsymbol{F}^{-1} - (\lambda/\lambda_y)\boldsymbol{x}_y\boldsymbol{x}'_y - \boldsymbol{x}_y\boldsymbol{x}'. \tag{19.16}$$

Equation (19.16) is known as the *Slutsky equation* — the fundamental equation of value theory. It can be written as

$$\mathbf{X}_{\boldsymbol{p}} = \boldsymbol{K} - \boldsymbol{x}_y\boldsymbol{x}', \tag{19.17}$$

where $\boldsymbol{K} = \lambda\boldsymbol{F}^{-1} - (\lambda/\lambda_y)\ \boldsymbol{x}_y\ \boldsymbol{x}'_y$ is the *substitution matrix* (also known as the *Slutsky matrix*) of income compensated price responses and $\boldsymbol{x}_y\boldsymbol{x}'$ is the *matrix of income effects*. Notice that the i, j element of (19.17) is

$$\frac{\partial x_i}{\partial p_j} = k_{ij} - \frac{\partial x_i}{\partial y}x_j, \tag{19.18}$$

where $\partial x_i/\partial p_j$ is the *total effect* of a price change on demand, k_{ij} (i.e., the i, j element of $\boldsymbol{K}$) is the *substitution effect* of a compensated price change on demand, and $(-\partial x_i/\partial y)x_j$ is the *income effect,* resulting from a change in price (not in income).

Notice that, in the absence of a particular specification of the utility function, economic theory has nothing to say about the sign of the income effect. It is an empirical question to determine the sign of the income effect. In the case, for example, of normal goods (also known as superior goods), $\partial x_i/\partial y > 0$ and the income effect is negative; in the case of inferior goods, $\partial x_i/\partial y < 0$ and the income effect is positive

The Slutsky decomposition is not a restriction in itself. Its importance lies in the fact that the Slutsky matrix $\boldsymbol{K}$ is an $n \times n$ symmetric matrix, since $\lambda\boldsymbol{F}^{-1} - (\lambda/\lambda_y)\ \boldsymbol{x}_y\boldsymbol{x}'_y$ is symmetric. Hence,

$$k_{ij} = k_{ji}.$$

This symmetry restriction of the Slutsky matrix may also be written in elasticity terms, making use of equation (19.18), as follows

$$\frac{\eta_{ij}}{s_j} + \eta_{iy} = \frac{\eta_{ji}}{s_i} + \eta_{jy},$$

where η_{ij} is the elasticity of demand of asset i with respect to the price of asset j, η_{iy} is the income elasticity of demand of asset i, and

$s_j = p_j x_j / y$ is the proportion of total expenditure devoted to asset j. Clearly, the symmetrical terms are the Allen elasticities of substitution

$$\sigma_{ij}^a = \frac{\eta_{ij}}{s_j} + \eta_{iy} = \frac{\eta_{ji}}{s_i} + \eta_{jy} = \sigma_{ji}^a, \tag{19.19}$$

where σ_{ij}^a is the Allen elasticity of substitution between assets i and j. Hence, the Allen elasticities of substitution are equivalent to compensated price elasticities. As with the sign of the income effect, the sign of the cross substitution effect, k_{ij}, is not determined, in the absence of a particular specification of the utility function. This sign is to be determined empirically.

19.3.4 Negativity of the Own Substitution Effect

Finally, the most important restriction of all is the negativity of the own substitution effect [see Phlips (1974, p. 52-53) for a proof of the negativity property]

$$k_{ii} < 0, \quad i = 1, ..., n. \tag{19.20}$$

This restriction establishes the negative relationship between quantity and price (i.e., the negativity of the slope of the demand curves), for those assets that are not inferior (i.e., assets for which the income elasticity is positive or zero).

Equations (19.11), (19.12), (19.19), and (19.20) are the Slutsky conditions that each system of demand equations should satisfy, irrespective of the choice of a particular utility function. These conditions, together with the usual neoclassical monotonicity requirement that the direct utility function should be an increasing function of each good consumed, are also known as the *integrability conditions.* If they hold, then the demand system is integrable in the sense that it can be generated by utility maximization subject to a budget constraint. As Fisher (1989, p. 87) puts it

> "[t]he importance of this is clear: if the integrability conditions are valid, then the theory of individual consumer behavior is applicable to the analysis of aggregate consumer demand functions in *per capita* form ... ."

19.4 Conclusion

We have developed the demand systems approach to the demand for liquid assets, paying explicit attention to the increasingly obvious tradeoff that exists between theoretical purity and econometric simplicity. The reader should also note that the nonparametric and parametric approaches to applied demand analysis are not mutually exclusive. For example, one can test the data with GARP to find subsets that are consistent with utility maximization and then estimate parametric demand functions imposing the utility maximization restrictions.

Clearly, the use in recent years of the simple representative consumer paradigm in monetary economics has opened the door to the succeeding introduction into monetary economics of the entire microfoundations, aggregation theory, and microeconometrics literatures. This new literature is actually an ongoing one and has only just begun to produce empirical results worthy of the effort required to understand it. In the following chapter we emphasize the contribution that can be made by using alternative functional forms for demand systems.

20

Locally Flexible Functional Forms and Demand Systems

For many years, econometricians used *globally regular functional* forms, such as the Cobb-Douglas and the Constant Elasticity of Substitution (CES) forms, to approximate the generating functions (such as, for example, direct utility, indirect utility, production, and cost functions) of neoclassical microeconomic theory. By regularity we mean that the indirect utility function is consistent with rational economic behavior — that is, the economic agent maximises direct utility subject to a budget constraint. In particular, regularity requires that the indirect utility function is homogeneous of degree zero in $\boldsymbol{p}$ and y, non-increasing in $\boldsymbol{p}$, non-decreasing in y, and convex or quasi-convex in $\boldsymbol{p}$. These constraints imply that the Slutsky conditions (discussed in the previous chapter) are satisfied. Regularity is global if it holds for all (positive) $\boldsymbol{p}$ and y values.

Although globally regular functional forms satisfy everywhere the theoretical regularity conditions for rational neoclassical economic behavior, they do not provide the capability to attain arbitrary elasticities of substitution — see, for example, Uzawa (1962). In recent years a number of empirical studies have made use of the *flexible* functional forms method to approximate aggregator functions. A flexible functional form is an approximation to an arbitrary function, with parameters that can be chosen to make the value of the first and second derivatives of the approximation equal to the first and second derivatives of the true function at any point — see Diewert (1973a) for more details. In this and the next chapter we provide a theoretical comparison of a number of popular flexible functional forms by grouping them into three sets that have broadly similar characteristics. These are (i) *locally flexible* forms, (ii) *effectively globally regular* forms, and (iii) *asymptotically globally flexible* forms.

We discuss locally flexible and effectively globally regular functional forms in this chapter, leaving the discussion of the asymptotically globally flexible forms for the next chapter.

20.1 Locally Flexible Forms

We begin with three of the most popular flexible functional forms — the *generalized Leontief*, the *translog*, and the *almost ideal demand system.* These forms provide the capability to approximate systems resulting from a broad class of generating functions and also to attain arbitrary elasticities of substitution — although at only one point. They have, however, very small regions of theoretical regularity, as we shall discuss in the sequel.

20.1.1 The Generalized Leontief

The generalized Leontief (GL) functional form was introduced by Diewert (1973) in the context of cost and profit functions. Diewert (1974) introduced the GL reciprocal indirect utility function

$$h(\boldsymbol{v}) = a_0 + \sum_{i=1}^{n} a_i v_i^{1/2} + \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} \beta_{ij} v_i^{1/2} v_j^{1/2}, \tag{20.1}$$

where $\boldsymbol{v} = [v_1, v_2, \cdots, v_n]$ is a vector of income normalized user costs, with the ith element being $v_i = p_i/y$, where p_i is the user cost of asset i and y is the total expenditure on the n assets. $\boldsymbol{B} = [\beta_{ij}]$ is an $n \times n$ symmetric matrix of parameters and a_0 and a_i are other parameters, for a total of $\left(n^2 + 3n + 2\right)/2$ parameters.

Using Diewert's (1974) modified version of Roy's identity

$$s_i = \frac{v_i \partial h(\boldsymbol{v}) / \partial v_i}{\sum_{j=1}^{n} v_j \partial h(\boldsymbol{v}) / \partial v_j}, \tag{20.2}$$

where $s_i = v_i x_i$ and x_i is the demand for asset i, the GL demand system can be written as

$$s_i = \frac{a_i v_i^{1/2} + \sum_{j=1}^{n} \beta_{ij} v_i^{1/2} v_j^{1/2}}{\sum_{j=1}^{n} a_j v_j^{1/2} + \sum_{k=1}^{n} \sum_{m=1}^{n} \beta_{km} v_k^{1/2} v_m^{1/2}}, \qquad i = 1, \cdots, n. \tag{20.3}$$

Because the share equations are homogenous of degree zero in the parameters, the following normalization — see Barnett and Lee (1985) — could be used in estimation

$$2\sum_{i=1}^{n} a_i + \sum_{i=1}^{n} \sum_{j=1}^{n} \beta_{ij} = 1. \tag{20.4}$$

20.1.2 The Basic Translog

The basic translog (BTL) flexible functional form was introduced by Christensen, Jorgenson, and Lau (1975). The BTL reciprocal indirect utility function can be written as

$$\log h(\boldsymbol{v}) = a_0 + \sum_{k=1}^{n} a_k \log v_k$$

$$+ \frac{1}{2} \sum_{k=1}^{n} \sum_{j=1}^{n} \beta_{jk} \log v_k \log v_j, \tag{20.5}$$

where $\boldsymbol{B} = [\beta_{ij}]$ is an $n \times n$ symmetric matrix of parameters and a_0 and a_i are other parameters, for a total of $\left(n^2 + 3n + 2\right)/2$ parameters.

The share equations, derived using the logarithmic form of Roy's identity,

$$s_i = -\frac{\partial \log h(\boldsymbol{v}) / \partial \log p_i}{\partial \log h(\boldsymbol{v}) / \partial \log y}, \qquad i = 1, \ldots, n,$$

are

$$s_i = \frac{a_i + \sum_{k=1}^{n} \beta_{ik} \log v_k}{\sum_{k=1}^{n} a_k + \sum_{k=1}^{n} \sum_{j=1}^{n} \beta_{jk} \log v_k}, \qquad i = 1, \ldots, n. \tag{20.6}$$

By imposing the restrictions

$$\sum_{i=1}^{n} \beta_{ik} = 0, \quad \text{for all } k,$$

on the BTL, the homothetic translog (HTL) model's share equations can also be obtained. These are (as we saw in Chapter 19)

$$s_i = \frac{a_i + \sum\limits_{k=1}^{n} \beta_{ik} \log v_k}{\sum\limits_{k=1}^{n} a_k}, \qquad i = 1, \dots, n.$$

With n assets, the HTL model's share equations contain $n(n+3)/2$ parameters. Notice that estimation of each model's share equations requires some parameter normalization, as the share equations are homogeneous of degree zero in the a's. Usually the normalization $\sum_{i=1}^{n} a_i = 1$ is used.

20.1.3 The Almost Ideal Demand System

The almost ideal demand system (AIDS) is written is share equation form [see Deaton and Muellbauer (1980) for more details] as

$$s_i = a_i + \sum_{k=1}^{n} \beta_{ik} \log p_k + b_i \left(\log y - \log g(\boldsymbol{p}) \right), \quad i = 1, \dots, n, \tag{20.7}$$

where $\log g(\boldsymbol{p})$ is a translog price index defined by

$$\log g(\boldsymbol{p}) = a_0 + \sum_{k=1}^{n} a_k \log p_k + \frac{1}{2} \sum_{k=1}^{n} \sum_{j=1}^{n} \beta_{kj} \log p_k \log p_j .$$

In equation (20.7), s_i is the ith budget share, y is income, p_k is the kth price, and $(\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{\beta})$ are parameters of the demand system to be estimated. Symmetry ($\beta_{ij} = \beta_{ji}$ for all i, j), adding up ($\sum_{k=1}^{n} a_k = 1$, $\sum_{i=1}^{n} \beta_{ij} = 0$ for all j, and $\sum_{i=1}^{n} b_i = 0$), and homogeneity ($\sum_{j=1}^{n} \beta_{ij} = 0$ for all i) are imposed in estimation. With n assets the AIDS model's share equations contain $(n^2 + 3n - 2)/2$ free parameters.

20.2 Effectively Globally Regular Forms

As argued earlier, models such as the generalized Leontief, translog, and AIDS are locally flexible but may have a relatively small regular region. In fact, Caves and Christensen (1980), Barnett and Lee (1985), and Barnett, Lee, and Wolfe (1985) show that the regularity regions of local flexible functional forms can be relatively small. Furthermore, the Monte Carlo analysis of Guilkey and Lovell (1980) finds that the generalized Leontief and the translog fail to provide a satisfactory approximation to the true data generating process for the moderate and even large elasticities of substitution that often arise in applications.

These problems led to the development of locally flexible functional forms that have larger regularity regions that Cooper and McLaren (1996) classify as 'effectively globally regular.' These functions typically have regular regions that include all data points in the sample. In addition, the regularity region increases as real expenditure levels grow, as is often the case with time series data. Furthermore, these functions provide more general Engel curve approximations, especially when income varies considerably.

Examples of these functions include the Minflex Laurent models introduced by Barnett (1983, 1985), Barnett and Lee (1985), and Barnett, Lee, and Wolfe (1985, 1987), based on the Laurent series expansion; the 'quadratic AIDS' (QUAIDS) model of Banks, Blundell and Lewbel (1996); and the 'general exponential form' (GEF) of Cooper and McLaren (1996). In what follows, we will consider two effectively globally regular flexible functional forms — the Minflex Laurent (ML) model and the normalized quadratic (NQ) reciprocal indirect utility function.

20.2.1 The Minflex Laurent

The Minflex Laurent model, introduced by Barnett (1983) and Barnett and Lee (1985), is a special case of the Full Laurent model also introduced by Barnett (1983). Following Barnett (1983), the Full Laurent reciprocal indirect utility function is

$$h(\boldsymbol{v}) = a_0 + 2\sum_{i=1}^{n} a_i v_i^{1/2} + \sum_{i=1}^{n}\sum_{j=1}^{n} a_{ij} v_i^{1/2} v_j^{1/2}$$

$$- 2\sum_{i=1}^{n} b_i v_i^{-1/2} - \sum_{i=1}^{n}\sum_{j=1}^{n} b_{ij} v_i^{-1/2} v_j^{-1/2}, \tag{20.8}$$

where a_0, a_i, a_{ij}, b_i, and b_{ij} are unknown parameters and v_i denotes the income normalized price, as before.

By assuming that $b_i = 0$, $b_{ii} = 0\ \forall i$, $a_{ij}b_{ij} = 0\ \forall i, j$, and forcing the off diagonal elements of the symmetric matrices $\boldsymbol{A} \equiv [a_{ij}]$ and $\boldsymbol{B} \equiv [b_{ij}]$ to be nonnegative, (20.8) reduces to the ML reciprocal indirect utility function

$$h(\boldsymbol{v}) = a_0 + 2\sum_{i=1}^{n} a_i v_i^{1/2} + \sum_{i=1}^{n} a_{ii} v_i + \sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} a_{ij}^2 v_i^{1/2} v_j^{1/2} - \sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} b_{ij}^2 v_i^{-1/2} v_j^{-1/2}. \tag{20.9}$$

Note that the off diagonal elements of $\boldsymbol{A}$ and $\boldsymbol{B}$ are nonnegative as they are raised to the power of two.

By applying Roy's identity to (20.9), the share equations of the ML demand system are

$$s_i = \frac{a_i v_i^{1/2} + a_{ii} v_i + \sum_{\substack{j=1\\ i\neq j}}^{n} a_{ij}^2 v_i^{1/2} v_j^{1/2} + \sum_{\substack{j=1\\ i\neq j}}^{n} b_{ij}^2 v_i^{-1/2} v_j^{-1/2}}{\sum_{i=1}^{n} a_i v_i^{1/2} + \sum_{i=1}^{n} a_{ii} v_i + \sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} a_{ij}^2 v_i^{1/2} v_j^{1/2} + \sum_{i=1}^{n}\sum_{\substack{j=1\\ i\neq j}}^{n} b_{ij}^2 v_i^{-1/2} v_j^{-1/2}}. \tag{20.10}$$

Since the share equations are homogenous of degree zero in the parameters, one can follow Barnett and Lee (1985) and impose the following normalization in the estimation of (20.10)

$$\sum_{i=1}^{n} a_{ii} + 2\sum_{i=1}^{n} a_i + \sum_{\substack{j=1\\ i\neq j}}^{n} a_{ij}^2 - \sum_{\substack{j=1\\ i\neq j}}^{n} b_{ij}^2 = 1. \tag{20.11}$$

Hence, there are

$$1 + n + \frac{n(n+1)}{2} + \frac{n(n-1)}{2}.$$

parameters in (20.9), but the $n\,(n-1)\,/2$ equality restrictions, $a_{ij}b_{ij} = 0\ \forall i, j$, and the normalization (20.11) reduce the number of parameters in equation (20.10) to $\left(n^2 + 3n\right)/2$.

20.2.2 The NQ Reciprocal Indirect Utility Function

Following Diewert and Wales (1988), the normalized quadratic (NQ) reciprocal indirect utility function is defined as

$$h(\boldsymbol{v}) = b_0 + \sum_{i=1}^{n} b_i v_i + \frac{1}{2} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} \beta_{ij} v_i v_j}{\sum_{i=1}^{n} \alpha_i v_i} + \sum_{i=1}^{n} \theta_i \log v_i, \qquad (20.12)$$

where b_0, $\boldsymbol{b} = [b_1, b_2, \cdots, b_n]$, $\boldsymbol{\theta} = [\theta_1, \theta_2, \cdots, \theta_n]$, and the elements of the $n \times n$ symmetric $\boldsymbol{B} \equiv [\beta_{ij}]$ matrix are the unknown parameters to be estimated. It is important to note that the quadratic term in (20.12) is normalized by dividing through by a linear function

$$\sum_{i=1}^{n} \alpha_i v_i,$$

and that the nonnegative vector of parameters $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_n]$ is assumed to be predetermined.

As in Diewert and Wales (1988), we assume that $\boldsymbol{\alpha}$ satisfies

$$\sum_{j=1}^{n} \alpha_j v_j^* = 1, \;\; \alpha_j \geq 0, \;\; \forall j. \qquad (20.13)$$

Moreover, we pick a reference (or base-period) vector of income normalized prices, $\boldsymbol{v}^* = \mathbf{1}$, and assume that the $\boldsymbol{B}$ matrix satisfies the following n restrictions

$$\sum_{j=1}^{n} \beta_{ij} v_j^* = 0, \;\; \forall i. \qquad (20.14)$$

Using the modified version of Roy's identity (20.2), the NQ demand system can be written as

$$s_i = \frac{v_i\left(b_i - \frac{1}{2}\frac{\alpha_i \sum_{k=1}^{n}\sum_{j=1}^{n}\beta_{kj}v_k v_j}{\left(\sum_{i=1}^{n}\alpha_i v_i\right)^2} + \frac{\sum_{j=1}^{n}\beta_{ij}v_i}{\left(\sum_{i=1}^{n}\alpha_i v_i\right)} + \theta_i\right)}{\sum_{i=1}^{n} b_i v_i + \frac{1}{2}\frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\beta_{ij}v_i v_j}{\left(\sum_{i=1}^{n}\alpha_i v_i\right)} + \sum_{i=1}^{n}\theta_i}. \qquad (20.15)$$

Finally, as the share equations are homogeneous of degree zero in the parameters, one can follow Diewert and Wales (1988) and impose the normalization

$$\sum_{j=1}^{n} b_j = 1. \qquad (20.16)$$

Hence, there are $n\,(n+5)\,/2$ parameters in (20.15), but the imposition of the $(n-1)$ restrictions in (20.14) and (20.16) reduces the number of parameters to be estimated to $\left(n^2+3n-2\right)/2$.

20.3 Imposing Local Curvature

The usefulness of flexible functional forms depends on whether they satisfy the theoretical regularity conditions of positivity, monotonicity, and curvature, and in most of the monetary asset demand literature there has been a tendency to ignore theoretical regularity. In fact, as Barnett (2002, p. 199) put it in his *Journal of Econometrics* Fellow's opinion article, without satisfaction of all three theoretical regularity conditions,

> "$\cdots$ the second-order conditions for optimizing behavior fail, and duality theory fails. The resulting first-order conditions, demand functions, and supply functions become invalid."

Recently, Ryan and Wales (1998) suggest a relatively simple procedure for imposing local curvature conditions. Their procedure applies to those locally flexible demand systems for which, at the point of approximation, the $n \times n$ Slutsky matrix $\boldsymbol{S}$ can be written as

$$\boldsymbol{S} = \boldsymbol{B} + \boldsymbol{C}, \qquad (20.17)$$

where $\boldsymbol{B}$ is an $n \times n$ symmetric matrix, containing the same number of independent elements as the Slutsky matrix, and $\boldsymbol{C}$ is an $n \times n$ matrix whose elements are functions of the other parameters of the system. Curvature requires the Slutsky matrix to be negative semidefinite. Ryan and Wales (1998) draw on related work by Lau (1978) and Diewert and Wales (1987) and impose curvature by replacing $\boldsymbol{S}$ in equation (20.17) with $-\boldsymbol{K}\boldsymbol{K}'$, where $\boldsymbol{K}$ is an $n \times n$ lower triangular matrix, so that $-\boldsymbol{K}\boldsymbol{K}'$ is by construction a negative semidefinite matrix. Then solving explicitly for $\boldsymbol{B}$ in terms of $\boldsymbol{K}$ and $\boldsymbol{C}$ yields

$$\boldsymbol{B} = -\boldsymbol{K}\boldsymbol{K}' - \boldsymbol{C},$$

meaning that the models can be reparameterized by estimating the parameters in $\boldsymbol{K}$ and $\boldsymbol{C}$ instead of the parameters in $\boldsymbol{B}$ and $\boldsymbol{C}$. That is, we can replace the elements of $\boldsymbol{B}$ in the estimating equations by the elements of $\boldsymbol{K}$ and the other parameters of the model, thus ensuring that $\boldsymbol{S}$ is negative semidefinite at the point of approximation, which could be any data point.

Ryan and Wales (1998) applied their procedure to three locally flexible functional forms — the almost ideal demand system, the normalized quadratic, and the linear translog. Moreover, Moschini (1999) suggested a possible reparameterization of the basic translog to overcome some problems noted by Ryan and Wales (1998) and also imposed curvature conditions locally in the basic translog. In this section, we follow Ryan and Wales (1998) and Moschini (1999) and impose curvature conditions locally on the flexible functional forms discussed so far in this chapter.

It should be noted, however, that in general the imposition of curvature does not assure true theoretical regularity as it might produce spurious violations of monotonicity. In this regard, Barnett and Pasupathy (2003, p. 151) argue that

> "imposing curvature without monotonicity, while perhaps to be preferred to the prior common practice of imposing neither, is not adequate without at least reporting data points at which violations of monotonicity occur. Monotonicity is too important to be ignored."

20.3.1 The Generalized Leontief and Local Curvature

Regarding the generalized Leontief model, we follow Serletis and Shahmoradi (2007) who build on Ryan and Wales (1998) and Moschini (1999) and impose curvature conditions locally on the generalized Leontief model by exploiting the Hessian matrix of second order derivatives of the reciprocal indirect utility function, unlike Ryan and Wales (1998) and Moschini (1999) who exploit the Slutsky matrix.

In particular, since curvature of the GL reciprocal indirect utility function requires that the Hessian matrix is negative semidefinite, we impose local curvature (at the reference point) by evaluating the Hessian terms of (20.1) at $\boldsymbol{v}^* = \mathbf{1}$, as follows

$$H_{ij} = -\delta_{ij} \left(a_i + \sum_{j=1, j \neq i}^{n} \beta_{ij} \right) + (1 - \delta_{ij})\beta_{ij},$$

where δ_{ij} is the Kronecker delta (that is, $\delta_{ij} = 1$ when $i = j$ and 0 otherwise). By replacing $\boldsymbol{H}$ by $-\boldsymbol{K}\boldsymbol{K}'$, where $\boldsymbol{K}$ is an $n \times n$ lower triangular matrix and $\boldsymbol{K}'$ its transpose, the above can be written as

$$-\left(KK'\right)_{ij} = -\delta_{ij} \left(a_i + \sum_{j=1, j \neq i}^{n} \beta_{ij} \right) + (1 - \delta_{ij})\beta_{ij}. \tag{20.18}$$

Solving for the a_i and β_{ij} terms as a function of the $(KK')_{ij}$ we can get the restrictions that ensure the negative semidefiniteness of the Hessian matrix [without destroying the flexibility properties of (20.1), since the number of free parameters remains the same]. In particular, when $i \neq j$, equation (20.18) implies

$$\beta_{ij} = -\left(KK'\right)_{ij}, \tag{20.19}$$

and for $i = j$ implies

$$\left(KK'\right)_{ii} = a_i + \sum_{j=1, j \neq i}^{n} \beta_{ij}.$$

Substituting β_{ij} from (20.19) in the above equation we get

$$\left(KK'\right)_{ii} = a_i - \sum_{j=1, j \neq i}^{n} \left(KK'\right)_{ij},$$

or

$$a_i = \left(KK'\right)_{ii} + \sum_{j=1, j \neq i}^{n} \left(KK'\right)_{ij},$$

which after some rearrangement yields

$$a_i = \sum_{j=1}^{n} \left(KK'\right)_{ij}. \tag{20.20}$$

For the case of three assets, conditions (20.19) and (20.20) imply the following six restrictions on (20.3)

$$\begin{aligned}
\beta_{12} &= -k_{11}k_{21}; \\
\beta_{13} &= -k_{11}k_{31}; \\
\beta_{23} &= -\left(k_{21}k_{31} + k_{22}k_{32}\right); \\
a_1 &= k_{11}^2 + k_{11}k_{21} + k_{11}k_{31}; \\
a_2 &= k_{21}^2 + k_{22}^2 + k_{11}k_{21} + k_{21}k_{31} + k_{22}k_{32}; \\
a_3 &= k_{31}^2 + k_{32}^2 + k_{33}^2 + k_{11}k_{31} + k_{21}k_{31} + k_{22}k_{32},
\end{aligned}$$

where the k_{ij} terms are the elements of the $\boldsymbol{K}$ matrix.

20.3.2 The Basic Translog and Local Curvature

Applying the Ryan and Wales (1998) procedure for imposing local curvature to the basic translog, the Slutsky terms of (20.5) can be written as

$$S_{ij} = \beta_{ij} - a_i\delta_{ij} - a_i \sum_{k=1}^{n} \beta_{kj}$$

$$- a_j \sum_{k=1}^{n} \beta_{ik} + a_i a_j \left(1 + \sum_{k=1}^{n}\sum_{m=1}^{n} \beta_{km}\right),$$

for $i, j = 1, ..., n$, where δ_{ij} is the Kronecker delta, as before. Ryan and Wales (1998) argued that in the case of the basic translog replacing $\boldsymbol{S}$ by $-\boldsymbol{K}\boldsymbol{K}'$ is of little help in imposing local curvature because the ijth element of $\boldsymbol{S}$ contains not just β_{ij} but also the terms $\sum_{k=1}^{n} \beta_{kj}$, $\sum_{k=1}^{n} \beta_{ik}$, and $a_i a_j \left(1 + \sum_{k=1}^{n}\sum_{m=1}^{n} \beta_{km}\right)$. As they noted, there are $n(n+1)/2$ independent β_{ij} parameters, but only $n(n-1)/2$ independent elements in $\boldsymbol{S}$, rendering it no longer possible to express the β_{ij} terms in terms of the elements of $\boldsymbol{K}$ and of the other parameters of the model.

However, Moschini (1999) suggested a possible reparameterization of the basic translog to overcome the problem noted by Ryan and Wales (1998) so that we can still use their procedure for imposing local curvature in the BTL demand system. In particular, he showed that by letting $\theta_i = \sum_{j=1}^{n} \beta_{ij}$ we can rewrite (20.6) as

$$s_i = \frac{a_i + \sum_{k=1}^{n-1} \beta_{ik} \log v_k + \theta_i \log v_n}{1 + \sum_{k=1}^{n} \theta_k \log v_k}, \quad i = 1, \ldots, n-1, \tag{20.21}$$

with s_n given by $s_n = 1 - \sum_{i=1}^{n-1} s_i$. With this parameterization, the Slutsky terms can be expressed in terms of a matrix of dimension $(n-1) \times (n-1)$, denoted by $\widetilde{\boldsymbol{S}}$, with the ijth element written as

$$\widetilde{S}_{ij} = \beta_{ij} - a_i \delta_{ij} - a_i \theta_j - a_j \theta_i + a_i a_j \left(1 + \sum_{k=1}^{n} \theta_k \right), \tag{20.22}$$

for $i, j = 1, \cdots, n-1$. Note that now in equation (20.22) there are exactly $n(n-1)/2$ $\widetilde{S}_{ij}$ terms as there are $n(n-1)/2$ β_{ij} terms.

By replacing $\widetilde{\boldsymbol{S}}$ by $-\widetilde{\boldsymbol{K}}\widetilde{\boldsymbol{K}}'$ in (20.22), for $n = 3$ we get the following three restrictions on (20.21)

$$\beta_{11} = -k_{11}^2 + a_1 + 2a_1\theta_1 - a_1^2 \left(1 + \sum_{k=1}^{n} \theta_k \right);$$

$$\beta_{12} = -k_{11}k_{21} + a_1\theta_2 + a_2\theta_1 - a_1 a_2 \left(1 + \sum_{k=1}^{n} \theta_k \right);$$

$$\beta_{22} = -k_{21}^2 - k_{22}^2 + a_2 + 2a_2\theta_2 - a_2^2 \left(1 + \sum_{k=1}^{n} \theta_k \right),$$

where the k_{ij} terms are the elements of the $\widetilde{\boldsymbol{K}}$ matrix.

20.3.3 The Almost Ideal Demand System and Local Curvature

Applying the Ryan and Wales (1998) procedure for imposing local curvature, we write the ijth element of the Slutsky matrix associated with the AIDS demand system, equation (20.7), at the point $y = p_k = 1$ ($\forall$ k) as

$$S_{ij} = \beta_{ij} - (a_i - b_i a_0)\delta_{ij}$$

$$+ (a_j - b_j a_0)(a_i - b_i a_0) - b_i b_j a_0,$$

for $i, j = 1, ..., n$, where $\delta_{ij} = 1$ when $i = j$ and 0 otherwise. Thus, following Ryan and Wales (1998) local curvature can be imposed by replacing the elements of $\boldsymbol{B}$ in the estimating share equations by the elements of $\boldsymbol{K}$ and the other parameters, as follows for the ijth element of $\boldsymbol{B}$

$$\beta_{ij} = \left(-KK'\right)_{ij} + (a_i - b_i a_0)\delta_{ij}$$

$$- (a_j - b_j a_0)(a_i - b_i a_0) + b_i b_j a_0, \tag{20.23}$$

for $i, j = 1, ..., n$.

For $n = 3$, for example, equation (20.23) implies the following three restrictions on (20.7)

$$\beta_{11} = -k_{11}^2 + a_1 - b_1 a_0 - (a_1 - b_1 a_0)^2 + b_1^2 a_0;$$
$$\beta_{12} = -k_{11} k_{21} - (a_2 - b_2 a_0)(a_1 - b_1 a_0) + b_1 b_2 a_0;$$
$$\beta_{22} = -k_{21}^2 - k_{22}^2 + a_2 - b_2 a_0 - (a_2 - b_2 a_0)^2 + b_2^2 a_0,$$

where the k_{ij} terms are the elements of the $\boldsymbol{K}$ matrix.

20.3.4 The Minflex Laurent and Local Curvature

As shown by Barnett (1983, Theorem A.3), (20.9) is globally concave for every $\boldsymbol{v} \geq \boldsymbol{0}$, if all parameters are nonnegative, as in that case (20.9) would be a sum of concave functions. If the initially estimated parameters of the vector $\boldsymbol{a}$ and matrix $\boldsymbol{A}$ are not nonnegative, curvature can be imposed globally by replacing each unsquared parameter by a squared parameter, as in Barnett (1983).

20.3.5 The NQ Reciprocal Indirect Utility Function and Local Curvature

The normalized quadratic reciprocal indirect utility function defined by (20.12), (20.13), and (20.14) will be globally concave over the positive orthant if $\boldsymbol{B}$ is a negative semidefinite matrix and $\theta \geq \boldsymbol{0}$ — see Diewert and Wales (1988, Theorem 3). Although curvature conditions can be imposed globally if the initially estimated $\boldsymbol{B}$ matrix is not negative semidefinite or the initially estimated θ vector is not nonegative, we follow Ryan and Wales (1998) and impose curvature conditions locally.

Using the Ryan and Wales (1998) technique, the Slutsky terms associated with the NQ demand system at the reference point $\boldsymbol{v}^* = \boldsymbol{1}$ can be written as

$$S_{ij} = \beta_{ij} - \theta_i \delta_{ij} + \theta_i b_j + \theta_j b_i + 2\theta_i \theta_j, \tag{20.24}$$

for $i, j = 1, 2, \cdots, n$, where δ_{ij} is the Kronecker delta, as before. As already noted, according to Moschini's (1999) result, a necessary and sufficient condition for $\boldsymbol{S}$ to be negative semidefinite is that $\widetilde{\boldsymbol{S}}$ (obtained by deleting the last row and column of $\boldsymbol{S}$) is also negative semidefinite. Thus, (20.24) can be expressed as

$$\widetilde{S}_{ij} = \beta_{ij} - \theta_i \delta_{ij} + \theta_i b_j + \theta_j b_i + 2\theta_i \theta_j, \tag{20.25}$$

for $i, j = 1, 2, \cdots, n-1$. Hence, local curvature can be imposed (at $\boldsymbol{v}^*$ $= \mathbf{1}$) by setting $\widetilde{\boldsymbol{S}} = -\widetilde{\boldsymbol{K}}\widetilde{\boldsymbol{K}}'$ in (20.25) and then using (20.25) to solve for the β_{ij} elements, as follows

$$\beta_{ij} = -\left(\widetilde{K}\widetilde{K}'\right)_{ij} + \theta_i \delta_{ij} - \theta_i b_j - \theta_j b_i - 2\theta_i \theta_j, \tag{20.26}$$

for $i, j = 1, 2, \cdots, n-1$. It is to be noted that this re-parametrization does not destroy the flexibility of the NQ reciprocal indirect utility function, since the $n(n-1)/2$ elements of $\boldsymbol{B}$ are replaced by the $n(n-1)/2$ elements of $\widetilde{\boldsymbol{K}}$.

For the case of three assets ($n = 3$), (20.26) implies the following restrictions on (20.15)

$$\begin{aligned}
\beta_{11} &= -k_{11}^2 + \theta_1 - 2b_1\theta_1 - 2\theta_1^2; \\
\beta_{12} &= -k_{21}k_{11} - \theta_1 b_2 - \theta_2 b_1 - 2\theta_1\theta_2; \\
\beta_{22} &= -\left(k_{21}^2 + k_{22}^2\right) + \theta_2 - 2\theta_2 b_2 - 2\theta_2^2,
\end{aligned}$$

where the k_{ij} terms are the elements of the $\widetilde{\boldsymbol{K}}$ matrix.

20.4 Conclusion

We have provided a theoretical discussion of a number of different locally flexible functional forms, by grouping them into two groups that have similar characteristics. Of course, there are many other possibilities, but we selected these functional forms because they provide a representation of the two groups of locally flexible functional forms that are in the widest use in applied work based on the demand systems approach.

We want to emphasize that it would be preferable if one could nest at least some flexible functional forms, so that the choice between them could be the subject of a statistical hypothesis test. This is currently possible for those flexible functional forms that have interpretations as Taylor series expansions. For example, Serletis (1988) uses likelihood ratio tests to choose between four nested translog demand systems — the generalized translog, the basic translog, the linear translog, and the homothetic translog. In general, however, given known estimation techniques, it is not possible to nest flexible functional forms with different approximation properties.

21

Globally Flexible Functional Forms and Demand Systems

As already noted in the previous chapter, most locally flexible functional forms provide arbitrary elasticity estimates at the point of approximation and they gain this precision at the expense of giving up global regularity. Barnett (1983, 1985), Barnett and Lee (1985) and Barnett, Lee, and Wolfe (1985, 1987) provided a partial solution to this problem by proposing the minflex Laurent model that is locally flexible and regular over a large region but is still not globally regular.

An innovation in this respect are the semi-nonparametric flexible functional forms that possess global flexibility and in which asymptotic inferences are, potentially, free from any specification error. Seminonparametric functions can provide an asymptotically global approximation to complex economic relationships. These functions provide global approximations to the true data generating process and its partial derivatives. By global approximation, we mean that the flexible functional form is capable, in the limit, of approximating the unknown underlying generating function at all points and thus of producing arbitrarily accurate elasticities at all data points. Two such seminonparametric functions are the Fourier flexible functional form, introduced by Gallant (1981), and the Asymptotically Ideal Model (AIM), introduced by Barnett and Jonas (1983) and employed and explained in Barnett and Yue (1988).

This chapter focuses on these two globally flexible functional forms — the Fourier and the Asymptotically Ideal Model. While there is some comparison implied in our presentation, our purpose in this chapter is basically to make clear the properties of these two models.

21.1 The Fourier Model

We follow the procedure explained in Gallant (1981) for expanding the indirect utility function using the Fourier series,

$$h(\boldsymbol{v}) = u_0 + \boldsymbol{b}'\boldsymbol{v} + \frac{1}{2}\boldsymbol{v}'\mathbf{C}\boldsymbol{v}$$

$$+\sum_{\alpha=1}^{A}\left(u_{0\alpha} + 2\sum_{j=1}^{J}\left[u_{j\alpha}\cos(j\boldsymbol{k}_{\alpha}'\boldsymbol{v}) - w_{j\alpha}\sin(j\boldsymbol{k}_{\alpha}'\boldsymbol{v})\right]\right), \quad (21.1)$$

in which

$$\mathbf{C} = -\sum_{\alpha=1}^{A} u_{0\alpha}\boldsymbol{k}_{\alpha}\boldsymbol{k}_{\alpha}',$$

where $\boldsymbol{v}$ denotes income normalized prices $(=\boldsymbol{p}/y)$, $\boldsymbol{k}_{\alpha}$ is a multi-index — an n-vector with integer components — and u_0, $\{b\}$, $\{u\}$, and $\{w\}$ are parameters to be estimated. As Gallant (1981) shows, the length of a multi-index, denoted as $|\boldsymbol{k}_{\alpha}|^* = \sum_{i=1}^{n}|k_{i\alpha}|$, reduces the complexity of the notation required to denote high-order partial differentiation and multivariate Fourier series expansions. For example, with $n = 3$ in (21.1), the multi-index $\lambda' = (5, 2, 7)$, generates the 14th order partial derivative, as follows — see Gallant (1981) for more details:

$$D^{\lambda}h(\boldsymbol{v}) = \frac{\partial^{|\lambda|^*}}{\partial v_1^{\lambda_1}\partial v_2^{\lambda_2}\partial v_3^{\lambda_3}}h(\boldsymbol{v}) = \frac{\partial^{14}}{\partial v_1^5\partial v_2^2\partial v_3^7}h(\boldsymbol{v}),$$

The parameters A (the number of terms) and J (the degree of the approximation) determine the degree of the Fourier polynomials. The Fourier flexible functional form has the ability of achieving close approximation in Sobolev norm which confers nonparametric properties on the functional form. This is the reason the Fourier flexible form is considered to be a semi-nonparametric functional form.

By applying Roy's modified identity,

$$s_i(\boldsymbol{v}) = \frac{v_i\left(\partial h(\boldsymbol{v})/\partial v_i\right)}{\boldsymbol{v}'\left(\partial h(\boldsymbol{v})/\partial v_i\right)}, \tag{21.2}$$

to (21.1), we obtain the Fourier demand system

$$s_i = \frac{v_i b_i - \displaystyle\sum_{\alpha=1}^{A}\left(u_{0\alpha}\boldsymbol{v}'\boldsymbol{k}_\alpha + 2\sum_{j=1}^{J} j\left[u_{j\alpha}\sin(j\boldsymbol{k}'_\alpha\boldsymbol{v}) + w_{j\alpha}\cos(j\boldsymbol{k}'_\alpha\boldsymbol{v})\right]\right)k_{i\alpha}v_i}{\boldsymbol{b}'\boldsymbol{v} - \displaystyle\sum_{\alpha=1}^{A}\left(u_{0\alpha}\boldsymbol{v}'\boldsymbol{k}_\alpha + 2\sum_{j=1}^{J} j\left[u_{j\alpha}\sin(j\boldsymbol{k}'_\alpha\boldsymbol{v}) + w_{j\alpha}\cos(j\boldsymbol{k}'_\alpha\boldsymbol{v})\right]\right)\boldsymbol{k}'_\alpha\boldsymbol{v}}, \tag{21.3}$$

for $i = 1, 2, 3$ — the time subscript t has been suppressed.

Eastwood and Gallant (1991) show that Fourier functions produce consistent and asymptotically normal parameter estimates when the number of parameters to be estimated equals the number of effective observations raised to the power of 2/3 — this result follows from Huber (1981) and is similar to optimal bandwidth results in many nonparametric models. For example, with $n = 3$ and $T = 134$, the number of effective observations is 268 ($= 2 \times 134$) — since we estimate $(n-1)$ share equations — and we should therefore estimate (approximately) ($268^{2/3} =$) 41 parameters.

As we impose the normalization $b_n = \sum_{i=1}^{n-1} b_i$, the Fourier demand system has $(n-1)$ b, A $u_{0\alpha}$, AJ $u_{j\alpha}$, and AJ $w_{j\alpha}$ parameters to be estimated, for a total of $(n-1) + A(1+2J)$ free parameters. By setting $(n-1) + A(1+2J)$ equal to 41, for the $n = 3$ and $T = 134$ example, we choose the values of A and J to be 13 and 1, respectively. This also determines the elementary multi-indexes, as shown in the following table:

Elementary Multi-indexes $\{k\}_{\alpha=1}^{13}$

α	1	2	3	4	5	6	7	8	9	10	11	12	13
v_1	1	0	0	1	1	0	1	0	0	1	1	2	2
v_2	0	1	0	1	0	1	1	1	2	2	0	1	0
v_3	0	0	1	0	1	1	1	2	1	0	2	0	1
$\lvert k_\alpha \rvert^*$	1	1	1	2	2	2	3	3	3	3	3	3	3

As a Fourier series is a periodic function in its arguments but the indirect utility function is not, the scaling of the data is also important. In empirical applications, to avoid the approximation from diverging from the true indirect utility function the data should be rescaled so that the income normalized prices lie on $0 \leq v_i \leq 2\pi$. The income normalized prices v_i $(i = 1, \cdots, n)$ are typically rescaled as follows $v_i \times [(2\pi - \varepsilon) / \max \{v_i : i = 1, \cdots, n\}]$, with $(2\pi - \varepsilon)$ set equal to 6, as in Gallant (1982). In cases, however, that the income normalized prices v_i $(i = 1, \cdots, n)$ are already between 0 and 2π, such rescaling is not necessary.

21.2 The AIM Model

We follow Barnett and Yue (1988) and use the reciprocal indirect utility function for the asymptotically ideal model for $n = 3$ (as an example):

$$
\begin{aligned}
h(\boldsymbol{v}) = a_0 &+ \sum_{k=1}^{K} \sum_{i=1}^{3} a_{ik} v_i^{\lambda(k)} \\
&+ \sum_{k=1}^{K} \sum_{m=1}^{K} \left[\sum_{i=1}^{3} \sum_{j=1}^{3} a_{ijkm} v_i^{\lambda(k)} v_j^{\lambda(m)} \right] \\
&+ \sum_{k=1}^{K} \sum_{m=1}^{K} \sum_{g=1}^{K} \left[\sum_{i=1}^{3} \sum_{j=1}^{3} \sum_{h=1}^{3} a_{ijhkmg} v_i^{\lambda(k)} v_j^{\lambda(m)} v_h^{\lambda(g)} \right], \qquad (21.4)
\end{aligned}
$$

where $\lambda(z) = 2^{-z}$ for $z = \{k, m, g\}$ is the exponent set and a_{ik}, a_{ijkm}, and a_{ijhkmg}, for all $i, j, h = 1, 2, 3$, are the parameters to be estimated. The number of parameters is reduced by deleting the diagonal elements of the parameter arrays so that $i \neq j$, $j \neq h$ and $i \neq h$. This does not alter the span of the model's approximation.

By applying the modified Roy's identity to (21.4), we obtain the AIM(K) demand system, where $s_i = p_i x_i / \boldsymbol{p}'\boldsymbol{x} = v_i x_i$. With n assets and a degree of approximation of K, the number of parameters to be estimated in the AIM(K) model is given by the following formula:

$$
\frac{nk}{1!} + \frac{n(n-1)k^2}{2!} + \frac{n(n-1)(n-2)k^3}{3!} + \cdots.
$$

In what follows, we briefly present the basic properties of three AIM models — the AIM model for $K = 1$, 2, and 3. While there is some comparison in our presentation in this section, our purpose is basically to make clear the properties and complexities of these models. It is to be noted that for $n = 3$ and $K = 4$ the AIM(4) has 124 parameters to be estimated!

21.2.1 The AIM(1) Model

For $K = 1$ equation (21.4) becomes, since $\lambda(z) = 1/2$ for $z = \{k, m, g\}$,

$$h_{K=1}(\boldsymbol{v}) = a_0 + \sum_{i=1}^{3} a_i v_i^{1/2} + \sum_{i=1}^{3}\sum_{j=1}^{3} a_{ij} v_i^{1/2} v_j^{1/2}$$

$$+ \sum_{i=1}^{3}\sum_{j=1}^{3}\sum_{h=1}^{3} a_{ijh} v_i^{1/2} v_j^{1/2} v_h^{1/2}. \tag{21.5}$$

We delete the diagonal terms (so that $i \neq j$, $j \neq h$ and $i \neq h$) and follow Barnett and Yue (1988) and reparameterize by stacking the coefficients as they appear in (21.5) into a single vector, $\boldsymbol{b} = (b_0, \cdots, b_7)'$ containing the 8 coefficients in (21.5), as follows,

$$h_{K=1}(\boldsymbol{v}) = b_0 + b_1 v_1^{1/2} + b_2 v_2^{1/2} + b_3 v_3^{1/2}$$

$$+ b_4 v_1^{1/2} v_2^{1/2} + b_5 v_1^{1/2} v_3^{1/2}$$

$$+ b_6 v_2^{1/2} v_3^{1/2} + b_7 v_1^{1/2} v_2^{1/2} v_3^{1/2}, \tag{21.6}$$

where

$$\begin{aligned}
b_0 &= a_0; \\
b_1 &= a_1; \\
b_2 &= a_2; \\
b_3 &= a_3; \\
b_4 &= a_{12} + a_{21}; \\
b_5 &= a_{13} + a_{31}; \\
b_6 &= a_{23} + a_{32}; \\
b_7 &= a_{123} + a_{132} + a_{213} + a_{231} + a_{312} + a_{321}.
\end{aligned}$$

Applying the modified Roy's identity (21.2) to (21.6) yields the AIM(1) demand system,

$$s_1 = \left(b_1 v_1^{1/2} + b_4 v_1^{1/2} v_2^{1/2} + b_5 v_1^{1/2} v_3^{1/2} + b_7 v_1^{1/2} v_2^{1/2} v_3^{1/2}\right)/D; \quad (21.7)$$

$$s_2 = \left(b_2 v_2^{1/2} + b_4 v_1^{1/2} v_2^{1/2} + b_6 v_2^{1/2} v_3^{1/2} + b_7 v_1^{1/2} v_2^{1/2} v_3^{1/2}\right)/D; \quad (21.8)$$

$$s_3 = \left(b_3 v_3^{1/2} + b_5 v_1^{1/2} v_3^{1/2} + b_6 v_2^{1/2} v_3^{1/2} + b_7 v_1^{1/2} v_2^{1/2} v_3^{1/2}\right)/D, \quad (21.9)$$

where D is the sum of the numerators in equations (21.7), (21.8), and (21.9).

21.2.2 The AIM(2) Model

For $K = 2$, equation (21.4) becomes:

$$h_{K=2}(\boldsymbol{v}) = a_0 + \sum_{k=1}^{2} \sum_{i=1}^{3} a_{ik} v_i^{\lambda(k)}$$

$$+ \sum_{k=1}^{2} \sum_{m=1}^{2} \left[\sum_{i=1}^{3} \sum_{j=1}^{3} a_{ijkm} v_i^{\lambda(k)} v_j^{\lambda(m)} \right]$$

$$+ \sum_{k=1}^{2} \sum_{m=1}^{2} \sum_{g=1}^{2} \left[\sum_{i=1}^{3} \sum_{j=1}^{3} \sum_{h=1}^{3} a_{ijhkmg} v_i^{\lambda(k)} v_j^{\lambda(m)} v_h^{\lambda(g)} \right]. \quad (21.10)$$

Again, to avoid the extensive multiple subscripting in the coefficients a_{ijhkmg}, we follow Barnett and Yue (1988), and reparameterize by stacking the coefficients as they appear in (21.10) into a single vector of parameters, $\boldsymbol{b} = (b_0, \cdots, b_{26})'$ containing the 27 coefficients in (21.10), as follows [since $z = 1, 2$ so that $\lambda(1) = 1/2$ and $\lambda(2) = 1/4$, for $z = \{k, m, g\}$],

$$\begin{aligned}h_{K=2}(\boldsymbol{v}) = {} & b_0 + b_1 v_1^{1/2} + b_2 v_2^{1/2} + b_3 v_3^{1/2} + b_4 v_1^{1/4} + b_5 v_2^{1/4} + b_6 v_3^{1/4} \\ & + b_7 v_1^{1/2} v_2^{1/2} + b_8 v_1^{1/2} v_2^{1/4} + b_9 v_1^{1/4} v_2^{1/2} + b_{10} v_1^{1/4} v_2^{1/4} \\ & + b_{11} v_1^{1/2} v_3^{1/2} + b_{12} v_1^{1/2} v_3^{1/4} + b_{13} v_1^{1/4} v_3^{1/2} + b_{14} v_1^{1/4} v_3^{1/4} \\ & + b_{15} v_2^{1/2} v_3^{1/2} + b_{16} v_2^{1/2} v_3^{1/4} + b_{17} v_2^{1/4} v_3^{1/2} + b_{18} v_2^{1/4} v_3^{1/4} \\ & + b_{19} v_1^{1/2} v_2^{1/2} v_3^{1/2} + b_{20} v_1^{1/4} v_2^{1/2} v_3^{1/2} + b_{21} v_1^{1/2} v_2^{1/4} v_3^{1/2} \\ & + b_{22} v_1^{1/2} v_2^{1/2} v_3^{1/4} + b_{23} v_1^{1/2} v_2^{1/4} v_3^{1/4} + b_{24} v_1^{1/4} v_2^{1/2} v_3^{1/4} \\ & + b_{25} v_1^{1/4} v_2^{1/4} v_3^{1/2} + b_{26} v_1^{1/4} v_2^{1/4} v_3^{1/4}. \end{aligned} \tag{21.11}$$

Applying the modified version of Roy's identity, (21.2), to (21.11) we obtain the AIM(2) demand system,

$$\begin{aligned}s_1 = {} & \Big(2b_1 v_1^{1/2} + b_4 v_1^{1/4} + 2b_7 v_1^{1/2} v_2^{1/2} + 2b_8 v_1^{1/2} v_2^{1/4} + b_9 v_1^{1/4} v_2^{1/2} \\ & + b_{10} v_1^{1/4} v_2^{1/4} + 2b_{11} v_1^{1/2} v_3^{1/2} + 2b_{12} v_1^{1/2} v_3^{1/4} + b_{13} v_1^{1/4} v_3^{1/2} \\ & + b_{14} v_1^{1/4} v_3^{1/4} + 2b_{19} v_1^{1/2} v_2^{1/2} v_3^{1/2} + b_{20} v_1^{1/4} v_2^{1/2} v_3^{1/2} \\ & + 2b_{21} v_1^{1/2} v_2^{1/4} v_3^{1/2} + 2b_{22} v_1^{1/2} v_2^{1/2} v_3^{1/4} + 2b_{23} v_1^{1/2} v_2^{1/4} v_3^{1/4} \\ & + b_{24} v_1^{1/4} v_2^{1/2} v_3^{1/4} + b_{25} v_1^{1/4} v_2^{1/4} v_3^{1/2} + b_{26} v_1^{1/4} v_2^{1/4} v_3^{1/4}\Big) \Big/ D; \end{aligned} \tag{21.12}$$

$$\begin{aligned}s_2 = {} & \Big(2b_2 v_2^{1/2} + b_5 v_2^{1/4} + 2b_7 v_1^{1/2} v_2^{1/2} + b_8 v_1^{1/2} v_2^{1/4} + 2b_9 v_1^{1/4} v_2^{1/2} \\ & + b_{10} v_1^{1/4} v_2^{1/4} + 2b_{15} v_2^{1/2} v_3^{1/2} + 2b_{16} v_2^{1/2} v_3^{1/4} + b_{17} v_2^{1/4} v_3^{1/2} \\ & + b_{18} v_2^{1/4} v_3^{1/4} + 2b_{19} v_1^{1/2} v_2^{1/2} v_3^{1/2} + 2b_{20} v_1^{1/4} v_2^{1/2} v_3^{1/2} \\ & + b_{21} v_1^{1/2} v_2^{1/4} v_3^{1/2} + 2b_{22} v_1^{1/2} v_2^{1/2} v_3^{1/4} + b_{23} v_1^{1/2} v_2^{1/4} v_3^{1/4} \\ & + 2b_{24} v_1^{1/4} v_2^{1/2} v_3^{1/4} + b_{25} v_1^{1/4} v_2^{1/4} v_3^{1/2} + b_{26} v_1^{1/4} v_2^{1/4} v_3^{1/4}\Big) \Big/ D; \end{aligned} \tag{21.13}$$

$$\begin{aligned}s_3 = {} & \Big(2b_3 v_3^{1/2} + b_6 v_4^{1/4} + 2b_{11} v_1^{1/2} v_3^{1/2} + b_{12} v_1^{1/2} v_3^{1/4} + 2b_{13} v_1^{1/4} v_3^{1/2} \\ & + b_{14} v_1^{1/4} v_2^{1/4} + 2b_{15} v_1^{1/2} v_3^{1/2} + b_{16} v_1^{1/2} v_3^{1/4} + 2b_{17} v_2^{1/4} v_3^{1/2} \\ & + b_{18} v_2^{1/4} v_3^{1/4} + 2b_{19} v_1^{1/2} v_2^{1/2} v_3^{1/2} + 2b_{20} v_1^{1/4} v_2^{1/2} v_3^{1/2} \\ & + 2b_{21} v_1^{1/2} v_2^{1/4} v_3^{1/2} + b_{22} v_1^{1/2} v_2^{1/2} v_3^{1/4} + b_{23} v_1^{1/2} v_2^{1/4} v_3^{1/4} \\ & + b_{24} v_1^{1/4} v_2^{1/2} v_3^{1/4} + 2b_{25} v_1^{1/4} v_2^{1/4} v_3^{1/2} + b_{26} v_1^{1/4} v_2^{1/4} v_3^{1/4}\Big) \Big/ D, \end{aligned} \tag{21.14}$$

where now D is the sum of the numerators in equations (21.12), (21.13), and (21.14).

21.2.3 The AIM(3) Model

For $K = 3$ and $\lambda(z) = 2^{-z}$ for $z = \{k, m, g\}$, equation (21.4) becomes, after reparameterizing by stacking the coefficients as they appear in (21.4) for $K = 3$ into a single vector of parameters $\boldsymbol{b} = (b_0, \cdots, b_{63})'$ containing the 64 coefficients in (21.4) for $K = 3$ and $n = 3$,

$$\begin{aligned}
h_{K=3}(\boldsymbol{v}) = {} & b_0 + b_1 v_1^{1/2} + b_2 v_2^{1/2} + b_3 v_3^{1/2} + b_4 v_1^{1/4} + b_5 v_2^{1/4} \\
& + b_6 v_3^{1/4} + b_7 v_1^{1/8} + b_8 v_2^{1/8} + b_9 v_3^{1/8} + b_{10} v_1^{1/2} v_2^{1/2} \\
& + b_{11} v_1^{1/2} v_2^{1/4} + b_{12} v_1^{1/2} v_2^{1/8} + b_{13} v_1^{1/2} v_3^{1/2} + b_{14} v_1^{1/2} v_3^{1/4} \\
& + b_{15} v_1^{1/2} v_3^{1/8} + b_{16} v_1^{1/4} v_3^{1/2} + b_{17} v_1^{1/4} v_2^{1/4} + b_{18} v_1^{1/4} v_2^{1/8} \\
& + b_{19} v_1^{1/4} v_3^{1/2} + b_{20} v_1^{1/4} v_3^{1/4} + b_{21} v_1^{1/4} v_3^{1/8} + b_{22} v_1^{1/8} v_2^{1/2} \\
& + b_{23} v_1^{1/8} v_2^{1/4} + b_{24} v_1^{1/8} v_2^{1/8} + b_{25} v_1^{1/8} v_3^{1/2} + b_{26} v_1^{1/8} v_3^{1/4} \\
& + b_{27} v_1^{1/8} v_3^{1/8} + b_{28} v_2^{1/2} v_3^{1/2} + b_{29} v_2^{1/2} v_3^{1/4} + b_{30} v_2^{1/2} v_3^{1/8} \\
& + b_{31} v_2^{1/4} v_3^{1/2} + b_{32} v_2^{1/4} v_3^{1/4} + b_{33} v_2^{1/4} v_3^{1/8} + b_{34} v_2^{1/8} v_3^{1/2} \\
& + b_{35} v_2^{1/8} v_3^{1/4} + b_{36} v_2^{1/8} v_3^{1/8} + b_{37} v_1^{1/2} v_2^{1/2} v_3^{1/2} \\
& + b_{38} v_1^{1/2} v_2^{1/4} v_3^{1/2} + b_{39} v_1^{1/2} v_2^{1/8} v_3^{1/2} + b_{40} v_1^{1/2} v_2^{1/2} v_3^{1/4} \\
& + b_{41} v_1^{1/2} v_2^{1/4} v_3^{1/4} + b_{42} v_1^{1/2} v_2^{1/8} v_3^{1/4} + b_{43} v_1^{1/2} v_2^{1/2} v_3^{1/8} \\
& + b_{44} v_1^{1/2} v_2^{1/4} v_3^{1/8} + b_{45} v_1^{1/2} v_2^{1/8} v_3^{1/8} + b_{46} v_1^{1/4} v_2^{1/2} v_3^{1/2} \\
& + b_{47} v_1^{1/4} v_2^{1/4} v_3^{1/2} + b_{48} v_1^{1/4} v_2^{1/8} v_3^{1/2} + b_{49} v_1^{1/4} v_2^{1/2} v_3^{1/4} \\
& + b_{50} v_1^{1/4} v_2^{1/4} v_3^{1/4} + b_{51} v_1^{1/4} v_2^{1/8} v_3^{1/4} + b_{52} v_1^{1/4} v_2^{1/2} v_3^{1/8} \\
& + b_{53} v_1^{1/4} v_2^{1/4} v_3^{1/8} + b_{54} v_1^{1/4} v_2^{1/8} v_3^{1/8} + b_{55} v_1^{1/8} v_2^{1/2} v_3^{1/2} \\
& + b_{56} v_1^{1/8} v_2^{1/4} v_3^{1/2} + b_{57} v_1^{1/8} v_2^{1/8} v_3^{1/2} + b_{58} v_1^{1/8} v_2^{1/2} v_3^{1/4} \\
& + b_{59} v_1^{1/8} v_2^{1/4} v_3^{1/4} + b_{60} v_1^{1/8} v_2^{1/8} v_3^{1/4} + b_{61} v_1^{1/8} v_2^{1/2} v_3^{1/8} \\
& + b_{62} v_1^{1/8} v_2^{1/4} v^{1/8} + b_{63} v_1^{1/8} v_2^{1/8} v_3^{1/8}.
\end{aligned} \tag{21.15}$$

Applying the modified Roy's identity to (21.15), yields the AIM(3) demand system,

$$
\begin{aligned}
s_1 = \Big(& 4b_1 v_1^{1/2} + 2b_4 v_1^{1/4} + b_7 v_1^{1/8} + 4b_{10} v_1^{1/2} v_2^{1/2} + 4b_{11} v_1^{1/2} v_2^{1/4} \\
& + 4b_{12} v_1^{1/2} v_2^{1/8} + 2b_{13} v_1^{1/4} v_2^{1/2} + 2b_{14} v_1^{1/4} v_2^{1/4} + 2b_{15} v_1^{1/4} v_2^{1/8} \\
& + 4b_{16} v_1^{1/2} v_3^{1/2} + 4b_{17} v_1^{1/2} v_3^{1/4} + 4b_{18} v_1^{1/2} v_3^{1/8} + 2b_{19} v_1^{1/4} v_3^{1/2} \\
& + 2b_{20} v_1^{1/4} v_3^{1/4} + 2b_{21} v_1^{1/4} v_3^{1/8} + b_{22} v_1^{1/8} v_2^{1/2} + b_{23} v_1^{1/8} v_2^{1/4} \\
& + b_{24} v_1^{1/8} v_2^{1/8} + b_{25} v_1^{1/8} v_3^{1/2} + b_{26} v_1^{1/8} v_3^{1/4} + b_{27} v_1^{1/8} v_3^{1/8} \\
& + 4b_{37} v_1^{1/2} v_2^{1/2} v_3^{1/2} + 4b_{38} v_1^{1/2} v_2^{1/4} v_3^{1/2} + 4b_{39} v_1^{1/2} v_2^{1/8} v_3^{1/2} \\
& + 4b_{40} v_1^{1/2} v_2^{1/2} v_3^{1/4} + 4b_{41} v_1^{1/2} v_2^{1/4} v_3^{1/4} + 4b_{42} v_1^{1/2} v_2^{1/8} v_3^{1/4} \\
& + 4b_{43} v_1^{1/2} v_2^{1/2} v_3^{1/8} + 4b_{44} v_1^{1/2} v_2^{1/4} v_3^{1/8} + 4b_{45} v_1^{1/2} v_2^{1/8} v_3^{1/8} \\
& + 2b_{46} v_1^{1/4} v_2^{1/2} v_3^{1/2} + 2b_{47} v_1^{1/4} v_2^{1/4} v_3^{1/2} + 2b_{48} v_1^{1/4} v_2^{1/8} v_3^{1/2} \\
& + 2b_{49} v_1^{1/4} v_2^{1/2} v_3^{1/4} + 2b_{50} v_1^{1/4} v_2^{1/4} v_3^{1/4} + 2b_{51} v_1^{1/4} v_2^{1/8} v_3^{1/4} \\
& + 2b_{52} v_1^{1/4} v_2^{1/2} v_3^{1/8} + 2b_{53} v_1^{1/4} v_2^{1/4} v_3^{1/8} + 2b_{54} v_1^{1/4} v_2^{1/8} v_3^{1/8} \\
& + b_{55} v_1^{1/8} v_2^{1/2} v_3^{1/2} + b_{56} v_1^{1/8} v_2^{1/4} v_3^{1/2} + b_{57} v_1^{1/8} v_2^{1/8} v_3^{1/2} \\
& + b_{58} v_1^{1/8} v_2^{1/2} v_3^{1/4} + b_{59} v_1^{1/8} v_2^{1/4} v_3^{1/4} + b_{60} v_1^{1/8} v_2^{1/8} v_3^{1/4} \\
& + b_{61} v_1^{1/8} v_2^{1/2} v_3^{1/8} + b_{62} v_1^{1/8} v_2^{1/4} v^{1/8} + b_{63} v_1^{1/8} v_2^{1/8} v_3^{1/8} \Big) / D; \qquad (21.16)
\end{aligned}
$$

$$
\begin{aligned}
s_2 = \Big(& 4b_2 v_2^{1/2} + 2b_5 v_2^{1/4} + b_8 v_2^{1/8} + 4b_{10} v_1^{1/2} v_2^{1/2} + 2b_{11} v_1^{1/2} v_2^{1/4} + b_{12} v_1^{1/2} v_2^{1/8} \\
& + 4b_{13} v_1^{1/4} v_2^{1/2} + 2b_{14} v_1^{1/4} v_2^{1/4} + b_{15} v_1^{1/4} v_2^{1/8} + 4b_{22} v_1^{1/8} v_2^{1/2} + 2b_{23} v_1^{1/8} v_2^{1/4} \\
& + b_{24} v_1^{1/8} v_2^{1/8} + 4b_{28} v_2^{1/2} v_3^{1/2} + 4b_{29} v_2^{1/2} v_3^{1/4} + 4b_{30} v_2^{1/2} v_3^{1/8} + 2b_{31} v_2^{1/4} v_3^{1/2} \\
& + 2b_{32} v_2^{1/4} v_3^{1/4} + 2b_{33} v_2^{1/4} v_3^{1/8} + b_{34} v_2^{1/8} v_3^{1/2} + b_{35} v_2^{1/8} v_3^{1/4} + b_{36} v_2^{1/8} v_3^{1/8} \\
& + 4b_{37} v_1^{1/2} v_2^{1/2} v_3^{1/2} + 2b_{38} v_1^{1/2} v_2^{1/4} v_3^{1/2} + b_{39} v_1^{1/2} v_2^{1/8} v_3^{1/2} + 4b_{40} v_1^{1/2} v_2^{1/2} v_3^{1/4} \\
& + 2b_{41} v_1^{1/2} v_2^{1/4} v_3^{1/4} + b_{42} v_1^{1/2} v_2^{1/8} v_3^{1/4} + 4b_{43} v_1^{1/2} v_2^{1/2} v_3^{1/8} + 2b_{44} v_1^{1/2} v_2^{1/4} v_3^{1/8} \\
& + b_{45} v_1^{1/2} v_2^{1/8} v_3^{1/8} + 4b_{46} v_1^{1/4} v_2^{1/2} v_3^{1/2} + 2b_{47} v_1^{1/4} v_2^{1/4} v_3^{1/2} + b_{48} v_1^{1/4} v_2^{1/8} v_3^{1/2} \\
& + 4b_{49} v_1^{1/4} v_2^{1/2} v_3^{1/4} + 2b_{50} v_1^{1/4} v_2^{1/4} v_3^{1/4} + b_{51} v_1^{1/4} v_2^{1/8} v_3^{1/4} + 4b_{52} v_1^{1/4} v_2^{1/2} v_3^{1/8} \\
& + 2b_{53} v_1^{1/4} v_2^{1/4} v_3^{1/8} + b_{54} v_1^{1/4} v_2^{1/8} v_3^{1/8} + 4b_{55} v_1^{1/8} v_2^{1/2} v_3^{1/2} + 2b_{56} v_1^{1/8} v_2^{1/4} v_3^{1/2} \\
& + b_{57} v_1^{1/8} v_2^{1/8} v_3^{1/2} + 4b_{58} v_1^{1/8} v_2^{1/2} v_3^{1/4} + 2b_{59} v_1^{1/8} v_2^{1/4} v_3^{1/4} + b_{60} v_1^{1/8} v_2^{1/8} v_3^{1/4} \\
& + 4b_{61} v_1^{1/8} v_2^{1/2} v_3^{1/8} + 2b_{62} v_1^{1/8} v_2^{1/4} v_3^{1/8} + b_{63} v_1^{1/8} v_2^{1/8} v_3^{1/8} \Big) / D; \qquad (21.17)
\end{aligned}
$$

$$
\begin{aligned}
s_3 = \Big(& 4b_3 v_3^{1/2} + 2b_6 v_3^{1/4} + b_9 v_3^{1/8} + 4b_{16} v_1^{1/2} v_3^{1/2} + 2b_{17} v_1^{1/2} v_3^{1/4} \\
& + b_{18} v_1^{1/2} v_3^{1/8} + 4b_{19} v_1^{1/4} v_3^{1/2} + 2b_{20} v_1^{1/4} v_3^{1/4} + 4b_{25} v_1^{1/8} v_3^{1/2} \\
& + 2b_{26} v_1^{1/8} v_3^{1/4} + b_{27} v_1^{1/8} v_3^{1/8} + 4b_{28} v_2^{1/2} v_3^{1/2} + 2b_{29} v_2^{1/2} v_3^{1/4} \\
& + b_{30} v_2^{1/2} v_3^{1/8} + 4b_{31} v_2^{1/4} v_3^{1/2} + 2b_{32} v_2^{1/4} v_3^{1/4} + b_{33} v_2^{1/4} v_3^{1/8} \\
& + 4b_{34} v_2^{1/8} v_3^{1/2} + 2b_{35} v_2^{1/8} v_3^{1/4} + b_{36} v_2^{1/8} v_3^{1/8} + 4b_{37} v_1^{1/2} v_2^{1/2} v_3^{1/2} \\
& + 4b_{38} v_1^{1/2} v_2^{1/4} v_3^{1/2} + 4b_{39} v_1^{1/2} v_2^{1/8} v_3^{1/2} + 2b_{40} v_1^{1/2} v_2^{1/2} v_3^{1/4} \\
& + 2b_{41} v_1^{1/2} v_2^{1/4} v_3^{1/4} + 4b_{42} v_1^{1/2} v_2^{1/8} v_3^{1/4} + b_{43} v_1^{1/2} v_2^{1/2} v_3^{1/8} \\
& + b_{44} v_1^{1/2} v_2^{1/4} v_3^{1/8} + b_{45} v_1^{1/2} v_2^{1/8} v_3^{1/8} + 4b_{46} v_1^{1/4} v_2^{1/2} v_3^{1/2} \\
& + 4b_{47} v_1^{1/4} v_2^{1/4} v_3^{1/2} + 4b_{48} v_1^{1/4} v_2^{1/8} v_3^{1/2} + 2b_{49} v_1^{1/4} v_2^{1/2} v_3^{1/4} \\
& + 2b_{50} v_1^{1/4} v_2^{1/4} v_3^{1/4} + 2b_{51} v_1^{1/4} v_2^{1/8} v_3^{1/4} + b_{52} v_1^{1/4} v_2^{1/2} v_3^{1/8} \\
& + b_{53} v_1^{1/4} v_2^{1/4} v_3^{1/8} + b_{54} v_1^{1/4} v_2^{1/8} v_3^{1/8} + 4b_{55} v_1^{1/8} v_2^{1/2} v_3^{1/2} \\
& + 4b_{56} v_1^{1/8} v_2^{1/4} v_3^{1/2} + 4b_{57} v_1^{1/8} v_2^{1/8} v_3^{1/2} + 2b_{58} v_1^{1/8} v_2^{1/2} v_3^{1/4} \\
& + 2b_{59} v_1^{1/8} v_2^{1/4} v_3^{1/4} + 2b_{60} v_1^{1/8} v_2^{1/8} v_3^{1/4} \\
& + 2b_{61} v_1^{1/8} v_2^{1/2} v_3^{1/8} + b_{62} v_1^{1/8} v_2^{1/4} v_3^{1/8} + b_{63} v_1^{1/8} v_2^{1/8} v_3^{1/8} \Big) / D, \quad (21.18)
\end{aligned}
$$

where now D is the sum of the numerators in equations (21.16), (21.17), and (21.18).

21.3 Computational Considerations

Demand systems (21.3) and (21.7)-(21.9), (21.12)-(21.14), and (21.16)-(21.18) can be written as

$$\boldsymbol{s}_t = \boldsymbol{\psi}(\boldsymbol{v}_t, \boldsymbol{\theta}) + \boldsymbol{\epsilon}_t, \qquad (21.19)$$

with an error term appended. In (21.19), $\boldsymbol{s} = (s_1, \cdots, s_n)'$, $\boldsymbol{\psi}(\boldsymbol{v}, \boldsymbol{\theta}) = (\psi_1(\boldsymbol{v}, \boldsymbol{\theta}), \cdots, \psi_n(\boldsymbol{v}, \boldsymbol{\theta}))'$, and $\psi_i(\boldsymbol{v}, \boldsymbol{\theta})$ is given by the right-hand side of each of (21.3) and (21.7)-(21.9), (21.12)-(21.14), and (21.16)-(21.18).

In our recent work with globally flexible demand systems [see Serletis and Shahmoradi (2005)], we have followed Gallant and Golub (1984, p. 298) who argue that

> "all statistical estimation procedures that are commonly used in econometric research can be formulated as an optimization problem of the following type [Burguete, Gallant and Souza (1982)]:

$$\widehat{\boldsymbol{\theta}} \text{ minimizes } \varphi(\boldsymbol{\theta}) \text{ over } \Theta$$

with $\varphi(\boldsymbol{\theta})$ twice continuously differentiable in $\boldsymbol{\theta}$."

Hence, following Gallant and Golub (1984) we can use Zellner's (1962) seemingly unrelated regression method to estimate $\boldsymbol{\theta}$. Hence, $\varphi(\boldsymbol{\theta})$ has the form

$$\varphi(\boldsymbol{\theta}) = \frac{1}{T}\boldsymbol{\epsilon}_t^{\prime}\boldsymbol{\epsilon}_t = \frac{1}{T}\sum_{t=1}^{T}\left[\boldsymbol{s}_t - \boldsymbol{\psi}(\boldsymbol{v}_t,\boldsymbol{\theta})\right]^{\prime}\widehat{\boldsymbol{\Sigma}}^{-1}\left[\boldsymbol{s}_t - \boldsymbol{\psi}(\boldsymbol{v}_t,\boldsymbol{\theta})\right], \quad (21.20)$$

where T is the number of observations and $\widehat{\boldsymbol{\Sigma}}$ is an estimate of the variance-covariance matrix of (21.19). In minimizing (21.20), we have used the TOMLAB/NPSOL tool box with MATLAB — see http://tomlab.biz/products/npsol. NPSOL uses a sequential quadratic programming algorithm and is suitable for both unconstrained and constrained optimization of smooth (that is, at least twice-continuously differentiable) nonlinear functions.

It is to be noted that as results in nonlinear optimization are sensitive to the initial parameter values, to achieve global convergence, in Serletis and Shahmoradi (2005) we randomly generated 500 sets of initial parameter values and chose the starting $\boldsymbol{\theta}$ that led to the lowest value of the objective function. Also, as in Gallant (1981) and Barnett and Yue (1988) we do not have access to asymptotic standard errors that can be supported by statistical theory.

21.4 Imposing Curvature Restrictions

The indirect utility function should be a quasi-convex function in income normalized prices, v_i $(i = 1, \cdots, n)$ — as already noted, this is the curvature condition. Gallant and Golub (1984), following Diewert, Avriel, and Zang (1977), argue that a necessary and sufficient condition for quasi-convexity of $h(\boldsymbol{v}, \boldsymbol{\theta})$ is

$$g(\boldsymbol{v}, \boldsymbol{\theta}) = \min_{\boldsymbol{z}}\left\{\boldsymbol{z}^{\prime}\nabla^2 h(\boldsymbol{v}, \boldsymbol{\theta})\boldsymbol{z} : \boldsymbol{z}^{\prime}\nabla h(\boldsymbol{v}, \boldsymbol{\theta}) = 0,\ \boldsymbol{z}^{\prime}\boldsymbol{z} = 1\right\}, \quad (21.21)$$

where $\nabla h(\boldsymbol{v}, \boldsymbol{\theta}) = (\partial/\partial\boldsymbol{v})h(\boldsymbol{v}, \boldsymbol{\theta})$ and $\nabla^2 h(\boldsymbol{v}, \boldsymbol{\theta}) = (\partial^2/\partial\boldsymbol{v}\partial\boldsymbol{v}^{\prime})h(\boldsymbol{v}, \boldsymbol{\theta})$, and $g(\boldsymbol{v}, \boldsymbol{\theta})$ is non-negative (that is, zero or positive) when the quasi-convexity (curvature) constraint is satisfied and negative when it is violated. $g(\boldsymbol{v}, \boldsymbol{\theta})$ is refered to as the 'constraint indicator.'

Hence, as in Gallant and Golub (1984), we can impose quasi-convexity by modifying the optimization problem as follows

$$\text{minimize } \varphi(\boldsymbol{\theta}) \text{ subject to } \min_{\boldsymbol{v}\in\Omega} g(\boldsymbol{v}, \boldsymbol{\theta}) \geq 0,$$

where Ω is a finite set with the finite number of elements v_i $(i = 1, \cdots, n)$. Curvature can be imposed at some representative point in the data (that is, locally), over a region of data points, or at every data point in the sample (that is, globally).

Let us briefly describe in more detail the Gallant and Golub (1984) method for imposing curvature restrictions on flexible functional forms. Define a real symmetric $n \times n$ matrix $\boldsymbol{A} = \nabla^2 h(\boldsymbol{v}, \boldsymbol{\theta})$ — note that this is the Hessian matrix of the indirect utility function, $h(\boldsymbol{v}, \boldsymbol{\theta})$ — and an $n \times 1$ vector $\boldsymbol{\alpha} = \nabla h(\boldsymbol{v}, \boldsymbol{\theta})$ as the gradient vector of $h(\boldsymbol{v}, \boldsymbol{\theta})$. The curvature condition (21.21) can be written as

$$g(\boldsymbol{v}, \boldsymbol{\theta}) = \min_{\boldsymbol{z}} \left\{ \boldsymbol{z}'\boldsymbol{A}\boldsymbol{z} : \boldsymbol{z}'\boldsymbol{\alpha} = 0, \ \boldsymbol{z}'\boldsymbol{z} = 1 \right\}.$$

The next step is to partition $\boldsymbol{\alpha}$ as $\boldsymbol{\alpha} = (\alpha_1, \boldsymbol{\alpha}'_{(2)})'$, where α_1 is the first element of $\boldsymbol{\alpha}$ and $\boldsymbol{\alpha}_{(2)}$ is an $(n-1) \times 1$ vector of the remaining elements of $\boldsymbol{\alpha}$, and construct an $n \times 1$ vector $\boldsymbol{u}$

$$\boldsymbol{u} = \begin{pmatrix} \alpha_1 - \|\boldsymbol{\alpha}\| \\ \\ \boldsymbol{\alpha}_{(2)} \end{pmatrix},$$

where $\|\boldsymbol{\alpha}\|$ is the norm of $\boldsymbol{\alpha}$, defined as $\|\boldsymbol{\alpha}\| = \left(\sum_{i=1}^{n} \alpha_i^2\right)^{1/2}$. With this notation we define the following

$$\begin{aligned} \gamma &= -\frac{1}{2}\boldsymbol{u}'\boldsymbol{u}; \\ \boldsymbol{\omega} &= -\gamma^{-1}\boldsymbol{A}\boldsymbol{u}; \\ \Phi &= \left(\gamma^{-2}\boldsymbol{u}'\boldsymbol{A}\boldsymbol{u}\right); \\ \boldsymbol{\phi} &= (\Phi/2)\boldsymbol{u} - \boldsymbol{\omega}, \end{aligned}$$

where γ is a scalar, Φ is an $n \times n$ matrix, and $\boldsymbol{\omega}$ and $\boldsymbol{\phi}$ are $n \times 1$ vectors. The next and final step is to form an $n \times n$ matrix $\boldsymbol{K}$ as follows

$$\boldsymbol{K} = \boldsymbol{A} + \boldsymbol{u}\boldsymbol{\phi}' + \boldsymbol{\phi}\boldsymbol{u}'.$$

Let's delete the first row and column of $\boldsymbol{K}$ and rename the $n-1$ by $n-1$ thereby obtained matrix as $\boldsymbol{K}_{22}$. A necessary and sufficient condition for curvature (or equivalently for the indicator function (21.21) to be non-negative) is that $\boldsymbol{K}_{22}$ should be a positive semidefinite matrix. We can use the 'chol' command in MATLAB to perform a Cholesky factorization of the $\boldsymbol{K}_{22}$ matrix and construct an indicator of whether

$\boldsymbol{K}_{22}$ is positive semidefinite (this indicator is zero when $\boldsymbol{K}_{22}$ is positive semidefinite and a positive integer otherwise). Hence, we run a constrained optimization subject to the constraint that $\boldsymbol{K}_{22}$ is positive semidefinite (in which case curvature is satisfied). As already noted, we can evaluate $\boldsymbol{K}_{22}$ at a single data point, over a region of data points, or at every data point in the sample.

21.5 Conclusion

We have provided a theoretical discussion of two semi-nonparametric flexible functional forms — the Fourier and the AIM. We also addressed computational considerations and discussed how global curvature can be imposed in these models, using methods suggested over 20 years ago by Gallant and Golub (1984). Unlike the locally flexible functional forms we discussed in the previous chapter — the generalized Leontief, translog, AIDS, minflex Laurent, and normalized quadratic — that provide arbitrary elasticity estimates at the point of approximation, semi-nonparametric flexible functional forms are free from specification error and can provide an asymptotically global approximation to complex economic relationships.

Part 7: Microeconometrics and the Demand for Money

Overview of Part 7

In Chapter 22, stochastic specifications of monetary asset budget share equations are discussed as well as income elasticities and own- and cross-price elasticities. These elasticities, along with the (Allen and Morishima) elasticities of substitution, are particularly useful in interpreting demand system parameter estimates. Attention is also focussed on the dynamic context in which policy operates.

Chapter 23 presents an econometrics digression emphasizing the contribution that can be made by using the demand-systems approach to the demand for money and monetary assets. We tackle the problem in two stages: (i) we employ the Divisia index to perform the aggregation over monetary assets and (ii) we estimate a system of share equations based on the basic translog flexible functional form and the AIM(2) globally flexible demand system.

The concluding chapter discusses a number of issues and presents a possible future research agenda that might flow out of this book.

22

The Econometrics of Demand Systems

There are three main purposes to our investigations in the context of demand systems. First and foremost, we want to focus our attention on econometric techniques that can be used to analyze the interrelated demand for money and liquid assets in the context of share equation systems. Secondly, we want to analyze the properties of demand systems which result from the fact that demand systems are obtained by preference-maximizing behavior. Finally, we are interested in the substitutability/complementarity relationship between money and other liquid assets.

In this chapter we outline a standard stochastic specification for demand systems written in share form that forms the basis of our discussion in the rest of this book. Although our primary focus is the estimation of price and substitution elasticities, we also pay explicit attention to recent developments that have increased the usefulness of the demand-systems modeling approach for monetary studies. We attempt to clarify just what these developments are and how they are tending to reorganize a very traditional literature on an important topic.

22.1 Dimension Reduction

There is an immediate problem with the parametric approach to the demand for money and liquid assets. It is the difficult problem of estimating monetary asset demand systems when there is a large number of assets. In particular, if n is large, the estimation of a highly disaggregated demand system encompassing the full range of assets is econometrically intractable, because of computational difficulties in the large parameter space. In the United States, for example, the Federal Reserve Board's M3 monetary aggregate contains 22 monetary assets. If we were to deal with a homothetic translog demand system (that we discussed in Chapter 19) encompassing all 22 liquid assets, the share equations would contain $n(n+3)/2 = 275$ parameters. It is not feasible, for *degrees of freedom* reasons, to estimate that many parameters.

In such cases, the number of variables can be reduced, in a large number of *ad hoc* ways, by assuming separability and using Divisia subaggregate indexes. We can assume, for example, in accordance with the Federal Reserve Board's *a priori* assignment of monetary assets to monetary aggregates, that the monetary services utility function $f(\boldsymbol{x})$ has the strongly recursive separable form

$$f(\boldsymbol{x}) = f^4(\boldsymbol{x}^4, f^3(\boldsymbol{x}^3, f^2(\boldsymbol{x}^2, f^1(\boldsymbol{x}^1)))),$$

where the components of $\boldsymbol{x}^1$ are those that are included in the Fed's M1 monetary aggregate, the components of $\boldsymbol{x}^2$ are those of the Board's MZM aggregate net of $\boldsymbol{x}^1$, the components of $\boldsymbol{x}^3$ are those of the Board's M2 aggregate net of $\boldsymbol{x}^1$ and $\boldsymbol{x}^2$, and the components of $\boldsymbol{x}^4$ are those of the Board's M3 aggregate net of $\boldsymbol{x}^1$, $\boldsymbol{x}^2$, and $\boldsymbol{x}^3$.

Each aggregator function f^r, $r = 1, ..., 4$, has two rather natural (mutually consistent) interpretations. On one hand it can be thought of as a (specific) category utility function; on the other hand, it may be interpreted as a subaggregate measure of monetary services. In the latter case, the aggregator functions f^r, $r = 1, ..., 4$ are the Board's functional monetary aggregates M1, MZM, M2 and M3, respectively. In particular, if Q_1 is the monetary aggregate for the components of M1, Q_2 for MZM, Q_3 for M2, and Q_4 for M3, then it follows that

$$Q_1 = f^1(\boldsymbol{x}^1);$$

$$Q_2 = f^2(\boldsymbol{x}^2, f^1(\boldsymbol{x}^1)) = f^2(\boldsymbol{x}^2, Q_1);$$

$$Q_3 = f^3(\boldsymbol{x}^3, f^2(\boldsymbol{x}^2, f^1(\boldsymbol{x}^1))) = f^3(\boldsymbol{x}^3, Q_2);$$

$$Q_4 = f^4(\boldsymbol{x}^4, f^3(\boldsymbol{x}^3, f^2(\boldsymbol{x}^2, f^1(\boldsymbol{x}^1)))) = f^4(\boldsymbol{x}^4, Q_3).$$

Of course, the actual numbers produced by the official monetary aggregates require the restrictive assumption that f^r, $r = 1, ..., 4$, and hence $f(\boldsymbol{x})$ itself, are all simple summations.

To focus on the details of demand for services of money and liquid assets at different levels of aggregation, one can assume a recursively decentralized decision-making process, reflected in the solution to the following optimizing problems,

$$\max_{\boldsymbol{x}^r, Q_{r-1}} f^r(\boldsymbol{x}^r, Q_{r-1}),$$

subject to

$$\boldsymbol{p}^r \boldsymbol{x}^r + P_{r-1} Q_{r-1} = P_r Q_r,$$

for $r = 4, 3, 2$, where P_r is the Divisia price aggregate corresponding to the Divisia quantity aggregate Q_r. Recall that the Divisia quantity index (in discrete time) is defined as

$$\log M_t^D - \log M_{t-1}^D = \sum_{j=1}^{n} w_{jt}^* (\log x_{jt} - \log x_{j,t-1}),$$

according to which the growth rate of the aggregate M is the weighted average of the growth rates of the component quantities, with the Divisia weights being defined as the expenditure shares averaged over the two periods of the change, $w_{jt}^* = (1/2)(w_{jt} + w_{j,t-1})$ for $j = 1, ..., n$, where $w_{jt} = p_{jt} x_{jt} / \sum p_{kt} x_{kt}$ is the expenditure share of asset j during period t, and p_{jt} is the nominal user cost of asset j, derived in Barnett (1978),

$$p_{jt} = p^* \frac{R_t - r_{jt}}{1 + R_t},$$

which is just the opportunity cost of holding a dollar's worth of the jth asset. Above, p^* is the true-cost of living index, r_{jt} is the market yield on the jth asset, and R_t is the yield available on a 'benchmark' asset that is held only to carry wealth between multiperiods.

Thus the allocation of expenditure between the assets within the rth group and the $(r-1)$th monetary aggregate may be carried out optimally knowing only the prices within the rth group, the price index of the $(r-1)$th monetary aggregate, and the optimal expenditure on Q_r (being passed down recursively from the previous stage constrained maximization). This system of optimization problems reflects a sequential budgeting procedure, similar to the two-stage budgeting procedure discussed in Chapter 15. Although the consumer is making the decentralization decisions from the top of the tree down, one can estimate conditional money demand models at successive levels of aggregation recursively, from the bottom up. This approach to the recursive estimation of utility trees has been developed by Barnett (1977) and Anderson (1979), and has been applied to the demand for money problem by Serletis (1991a).

It should be obvious that strong recursive separability is a 'strong' assumption and that the Fed's (and most other central banks') present practice of having this sort of structure is entirely unrealistic, and therefore not suitable for applied econometric work. In particular, the algebraic requirement of *strong recursive separability* is that

$$\frac{\partial}{\partial x_k}\left(\frac{f_i(\boldsymbol{x})}{f_j(\boldsymbol{x})}\right)=0,$$

for every $i \in I^r,\ j \in I^s,\ k \in I^t,\ \ t > r, s$. That is, strong recursive separability implies that the marginal rate of substitution between, say, an asset in $\boldsymbol{x}^1$ and an asset in $\boldsymbol{x}^2$ is independent of assets in $\boldsymbol{x}^3$ and $\boldsymbol{x}^4$.

In fact, one of the objectives of empirical demand analysis is to discover the structure of preferences. That is, instead of imposing a grouping pattern on the model, as is the case with most central banks' *a priori* assignment of monetary assets to monetary aggregates, the structure of preferences over monetary assets could be discovered by actually testing for weakly separable subgroups. We feel that separability-based modeling of the demand for liquid assets is an area for potentially productive future research. This matter is the subject of Chapter 24.

22.2 Duality and Functional Structure

There is another issue to be discussed, which is ultimately related to the uncertainty about the 'true stucture' of preferences. We have argued that the structure of preferences can be represented by either a direct or an indirect utility function, with the latter being more easily approached because it simplifies the estimation considerably, since it has prices exogenous in explaining consumer behavior. However, a structural property of the direct utility function does not imply the same property on the indirect utility function, and in order to implement a model of demand based on the indirect function that satisfies properties of the direct function, a correspondence between direct and indirect properties is needed.

Although nonhomothetic direct and indirect separable utility functions are distinct structures, as we mentioned in the previous chapter, Lau (1970) showed that if the direct utility function is weakly separable with homothetic aggregator functions then the indirect utility function will have the same structure with respect to (expenditure-normalized) prices. Moreover, Blackorby, Nissen, Primont, and Russell (1974) show that if the direct utility function is strongly recursively separable with homothetic aggregator functions then the indirect utility function will have the same structure with respect to normalized prices. The above results also apply to *quasi-homothetic* (i.e., homothetic to a point other than the origin) preferences — see Gorman (1970) for an extensive treatment of quasi-homotheticity.

The choice of homothetic indirect utility functions in some of the empirical work [such as, for example, Serletis (1991a, 1991b)] has been primarily motivated by these considerations. That is, it is motivated by the need to maintain a correspondence between direct and indirect utility function properties.

22.3 Stochastic Specification

In order to estimate share equation systems such as those discussed in Chapters 20 and 21, a stochastic version must be specified. Since these systems are in share form and only exogenous variables appear on the right-hand side, it seems reasonable to assume that the observed share in the ith equation, $i = 1, ..., n$, deviates from the true share by an additive disturbance term ϵ_i. Furthermore, we assume that the resulting disturbance vector $\boldsymbol{\epsilon} = (\epsilon_1, \epsilon_2, ..., \epsilon_n)$ is a 'classical disturbance' term with the following properties

$$E(\boldsymbol{\epsilon}_t) = 0, \quad E(\boldsymbol{\epsilon}_t, \boldsymbol{\epsilon}_s') = \begin{cases} \boldsymbol{\Omega} & \text{for } s = t \\ \mathbf{0} & \text{for } s \neq t \end{cases} \qquad \text{all } s, t,$$

where $\boldsymbol{\Omega}$ is the $n \times n$ symmetric and positive semidefinite covariance matrix, and $\mathbf{0}$ is a null matrix. With the addition of additive errors, the share equation system can be written in matrix form as

$$\boldsymbol{s}_t = \boldsymbol{\psi}(\boldsymbol{v}_t, \boldsymbol{\theta}) + \boldsymbol{\epsilon}_t, \tag{22.1}$$

where $\boldsymbol{\theta}$ is the coefficients vector.

As an example, consider the three-asset homothetic translog demand system

$$s_{1t} = \alpha_1 + \beta_{11} \log v_{1t} + \beta_{12} \log v_{2t} + \beta_{13} \log v_{3t} + \epsilon_{1t};$$

$$s_{2t} = \alpha_2 + \beta_{21} \log v_{1t} + \beta_{22} \log v_{2t} + \beta_{23} \log v_{3t} + \epsilon_{2t};$$

$$s_{3t} = \alpha_3 + \beta_{31} \log v_{1t} + \beta_{32} \log v_{2t} + \beta_{33} \log v_{3t} + \epsilon_{3t},$$

which we can write in matrix notation as

$$\begin{bmatrix} s_{1t} \\ s_{2t} \\ s_{3t} \end{bmatrix} = \begin{bmatrix} \alpha_1 & \beta_{11} & \beta_{12} & \beta_{13} \\ \alpha_2 & \beta_{21} & \beta_{22} & \beta_{23} \\ \alpha_3 & \beta_{31} & \beta_{32} & \beta_{33} \end{bmatrix} \begin{bmatrix} 1 \\ \log v_{1t} \\ \log v_{2t} \\ \log v_{3t} \end{bmatrix} + \begin{bmatrix} \epsilon_{1t} \\ \epsilon_{2t} \\ \epsilon_{3t} \end{bmatrix},$$

or more compactly,

$$\boldsymbol{s}_t = \boldsymbol{\Pi} \boldsymbol{v}_t + \boldsymbol{\epsilon}_t, \tag{22.2}$$

where $\boldsymbol{s}_t$ is a vector of positive expenditure shares, $\boldsymbol{v}_t$ is a vector of expenditure normalized prices with unity as the first element, $\boldsymbol{\Pi}$ is a matrix of preference parameters, and $\boldsymbol{\epsilon}_t$ is a vector of random disturbances.

Since the s_{it} are budget shares, they satisfy the adding up (singularity) condition, $\boldsymbol{i}'\boldsymbol{s}_t = 1$, for all t, so that we must have $\boldsymbol{i}'\boldsymbol{\Pi} = [1 \ 0 \ 0 \ 0]$ and $\boldsymbol{i}'\boldsymbol{\epsilon}_t = 0$, for all t, where $\boldsymbol{i}$ is an appropriately dimensioned unit vector. Written out in full, the adding up condition for the three asset homothetic translog implies

$$\begin{aligned} \alpha_1 + \alpha_2 + \alpha_3 &= 1; \\ \beta_{11} + \beta_{21} + \beta_{31} &= 0; \\ \beta_{12} + \beta_{22} + \beta_{32} &= 0; \\ \beta_{13} + \beta_{23} + \beta_{33} &= 0; \\ u_{1t} + u_{2t} + u_{3t} &= 0. \end{aligned}$$

In addition to the adding up restrictions on $\boldsymbol{\Pi}$, other restrictions are also imposed. An example of such other restrictions is the symmetry restrictions (to be discussed in more detail in the next chapter),

$$\beta_{12} = \beta_{21};$$
$$\beta_{13} = \beta_{31};$$
$$\beta_{23} = \beta_{32}.$$

The assumption that we have made about $\boldsymbol{\epsilon}_t$ permits correlation among the disturbances at time t but rules out the possibility of autocorrelated disturbances. This assumption and the fact that the shares satisfy an adding up condition (because this is a singular system) imply that the disturbance covariance matrix is also singular.

Another issue concerns our assumption that the error terms are normally distributed. As we are dealing with shares, such that $0 \leq s_i \leq 1$, the error terms cannot be exactly normally distributed and a multivariate logistic distribution might be a better assumption, as in Barnett, Geweke, and Yue (1991). However, as Davidson and MacKinnon (1993) argue, if the sample does not contain observations which are near 0 or 1, one can use the normal distribution as an approximation in the inference process.

If autocorrelation in the disturbances is absent, Barten (1969) has shown that full information maximum likelihood (FIML) estimates of the parameters can be obtained by arbitrarily deleting one equation in such a system, and that the resulting estimates are invariant with respect to the equation deleted. The parameter estimates from the deleted equation can be recovered from the restrictions imposed.

22.4 Autoregressive Disturbances

The assumption of a classical disturbance term permits correlation among the disturbances at time t but rules out the possibility of autocorrelated disturbances. This assumption and the fact that $\boldsymbol{s}_t$ (and therefore the $\boldsymbol{\epsilon}_t$) satisfy the adding up condition imply that the disturbance covariance matrix is also singular. As we argued earlier, if autocorrelation in the disturbances is absent, then FIML estimates of the parameters can be obtained by arbitrarily deleting one equation, with the resulting estimates being invariant with respect to the equation deleted.

However, autocorrelation in money demand systems is a common result and may be caused by institutional constraints which prevent people from adjusting their asset holdings within one period. In cases where the equation-by-equation Durbin-Watson statistics suggest that the disturbances are serially correlated, then usually a first-order autoregressive process is assumed, such that

$$\boldsymbol{\epsilon}_t = \boldsymbol{\rho}\boldsymbol{\epsilon}_{t-1} + \boldsymbol{e}_t,$$

where $\boldsymbol{\rho} = [\rho_{ij}]$ is a matrix of unknown parameters and $\boldsymbol{e}_t$ is a non-autocorrelated vector disturbance term with constant covariance matrix. For example, in the three asset HTL case, we have

$$\boldsymbol{\rho} = \begin{bmatrix} \rho_{11} & \rho_{12} & \rho_{13} \\ \rho_{21} & \rho_{22} & \rho_{23} \\ \rho_{31} & \rho_{32} & \rho_{33} \end{bmatrix} \quad \text{and} \quad \boldsymbol{e}_t = \begin{bmatrix} e_{1t} \\ e_{2t} \\ e_{3t} \end{bmatrix},$$

and adding up of the budget shares implies the following restrictions on $\boldsymbol{\rho}$,

$$\rho_{1i} + \rho_{2i} + \rho_{3i} = \kappa, \quad i = 1, 2, 3.$$

In this case, FIML estimates of the parameters can be obtained by using a result developed by Berndt and Savin (1975). They showed that if one assumes no autocorrelation across equations (i.e., $\boldsymbol{\rho}$ is diagonal), the autocorrelation coefficients for each equation must be identical, $\rho_{11} = \rho_{22} = \rho_{33} = \rho$. Consequently, by writing equation (22.1) for period $t-1$, multiplying by $\boldsymbol{\rho}$, and subtracting from (22.1), we can estimate (using FIML procedures) stochastic budget share equations given by

$$\boldsymbol{s}_t = \boldsymbol{\psi}(\boldsymbol{v}_t, \boldsymbol{\theta}) + \boldsymbol{\rho}\boldsymbol{s}_{t-1} - \boldsymbol{\rho}\boldsymbol{\psi}(\boldsymbol{v}_{t-1}, \boldsymbol{\theta}) + \boldsymbol{e}_t. \tag{22.3}$$

As an example, the homothetic translog version of the problem, with symmetry imposed and the third equation deleted, can be set up as follows

$$\begin{aligned} s_{1t} = {} & \alpha_1 + \beta_{11}\log v_{1t} + \beta_{12}\log v_{2t} + \beta_{13}\log v_{3t} + \rho s_{1t-1} \\ & - \rho\left(\alpha_1 + \beta_{11}\log v_{1t-1} + \beta_{12}\log v_{2t-1} + \beta_{13}\log v_{3t-1}\right) + e_{1t}; \end{aligned}$$

$$\begin{aligned} s_{2t} = {} & \alpha_2 + \beta_{12}\log v_{1t} + \beta_{22}\log v_{2t} + \beta_{23}\log v_{3t} + \rho s_{2t-1} \\ & - \rho\left(\alpha_2 + \beta_{12}\log v_{1t-1} + \beta_{22}\log v_{2t-1} + \beta_{23}\log v_{3t-1}\right) + e_{2t}. \end{aligned}$$

Notice that imposing a common factor across the equations ensures the invariance of the FIML parameter estimates with respect to the equation deleted.

22.5 Theoretical Regularity

As we noted in Chapter 20, the usefulness of flexible functional forms depends on whether they satisfy the theoretical regularity conditions of positivity, monotonicity, and curvature. These conditions can be checked as follows [see, for example, Serletis and Shahmoradi (2005) for more details]:

- Positivity can be checked by direct computation of the values of the estimated budget shares, $\widehat{\boldsymbol{s}}_t$. It is satisfied if $\widehat{\boldsymbol{s}}_t \geq 0$, for all t.
- Monotonicity can be checked by choosing a normalization on the indirect utility function so as to make $h(\boldsymbol{v})$ decreasing in its arguments and by direct computation of the values of the first gradient vector of the estimated indirect utility function. It is satisfied if $\nabla\widehat{h}(\boldsymbol{v}) < 0$, where $\nabla\widehat{h}(\boldsymbol{v}) = (\partial/\partial\boldsymbol{v})\widehat{h}(\boldsymbol{v})$.
- Curvature requires that the Slutsky matrix be negative semidefinite and can be checked by performing a Cholesky factorization of that matrix and checking whether the Cholesky values are nonpositive [since a matrix is negative semidefinite if its Cholesky factors are nonpositive — see Lau (1978, Theorem 3.2)]. Curvature can also be checked by examining the Allen elasticities of substitution matrix provided that the monotonicity condition holds. It requires that this matrix be negative semidefinite. In the case of four assets ($n = 4$), for example, this requires that (i) all own four σ_{ii}^a are negative at each observation, (ii) each of the six possible 2×2 matrices

$$\begin{bmatrix} \sigma_{ii}^a & \sigma_{ij}^a \\ \sigma_{ij}^a & \sigma_{jj}^a \end{bmatrix}$$

for $i, j = 1, 2, 3, 4$ but $i \neq j$, has a positive determinant at every observation, (iii) each of the four possible 3×3 matrices

$$\begin{bmatrix} \sigma_{ii}^a & \sigma_{ij}^a & \sigma_{ik}^a \\ \sigma_{ij}^a & \sigma_{jj}^a & \sigma_{jk}^a \\ \sigma_{ik}^a & \sigma_{jk}^a & \sigma_{kk}^a \end{bmatrix}$$

for $i, j, k = 1, 2, 3, 4$ but $i \neq j$, $i \neq k$, $j \neq k$, has a negative determinant at every observation, and (iv) the 4×4 matrix consisting of all the σ_{ij}^a, $i, j = 1, 2, 3, 4$,

$$\begin{bmatrix} \sigma_{11}^a & \sigma_{12}^a & \sigma_{13}^a & \sigma_{14}^a \\ \sigma_{12}^a & \sigma_{22}^a & \sigma_{23}^a & \sigma_{24}^a \\ \sigma_{13}^a & \sigma_{23}^a & \sigma_{33}^a & \sigma_{34}^a \\ \sigma_{14}^a & \sigma_{24}^a & \sigma_{34}^a & \sigma_{44}^a \end{bmatrix}$$

has a determinant whose value is zero (or near zero).

22.6 Econometric Regularity

We have shown how to estimate money demand functions from aggregate time series data and highlighted the challenge inherent with achieving economic regularity and the need for economic theory to inform econometric research. Incorporating restrictions from economic theory seems to be gaining popularity as there are also numerous recent papers that estimate stochastic dynamic general equilibrium models using economic restrictions — see, for example, Aliprantis *et al.* (2006). With the focus on economic theory, however, we should not be ignoring econometric regularity. In particular, we should not be ignoring unit root and cointegration issues, because the combination of nonstationary data and nonlinear estimation in large models (like the ones discussed in the previous two chapters) is an extremely difficult problem.

In this regard, there is a great deal of consensus in the literature that aggregate budget shares and price and expenditure variables are integrated of order one [or I(1) in the terminology of Engle and Granger (1987)]. It follows then that for demand models to make any sense the variables must be cointegrated in levels; that is, the equation errors must be stationary. If the errors are nonstationary, then there is no theory linking the left hand side to the right hand side variables in equation (22.1) or, equivalently, no evidence for the theoretical models in level form. In such cases, an important nonstationary variable might have been omitted and as a minimal step towards addressing the issue the models should be reestimated with the inclusion of a time trend, which can at least roughly proxy the omitted dynamics, omitted demographic shifts, and deterministic nonstationarity. Allowing for first order serial correlation, as in equation (22.3), is almost the same as taking first differences of the data if the autocorrelation coefficient is close to unity. In that case, the equation errors become stationary, but there is no theory for the models in first differences.

If the errors are stationary, the estimates are super consistent. However, as argued by Attfield (1997) and Ng (1999), standard estimation procedures are inadequate for obtaining correctly estimated standard errors for coefficients in cointegrating equations. If the equations were all linear, the DOLS method of Stock and Watson (1993) or the FM-OLS method of Phillips (1991, 1995) could have been used to obtain correctly estimated standard errors. With nonlinear models, however, some sort of modification of these procedures is called for, but this is a very difficult issue to deal with and beyond the scope of this book.

22.7 Expenditure and Price Elasticities

A system of budget share equations provides a complete characterization of consumer preferences over the services of monetary assets and can be used to estimate the income elasticities as well as the own- and cross-price elasticities. These elasticities are particularly useful in judging the validity of the parameter estimates and can be calculated directly from the estimated budget share equations by writing the left-hand side as

$$x_i = \frac{s_i y}{p_i}, \quad i = 1, \ldots, n.$$

In particular, the income elasticities, η_{iy}, can be calculated as

$$\eta_{iy} = \frac{y}{s_i}\frac{\partial s_i}{\partial y} + 1, \quad i = 1, \ldots, n,$$

and the uncompensated (Cournot) price elasticities, η_{ij}, as

$$\eta_{ij} = \frac{p_j}{s_i}\frac{\partial s_i}{\partial p_j} - \delta_{ij}, \quad i,\ j = 1, \ldots, n,$$

where $\delta_{ij} = 0$ for $i \neq j$ and 1 otherwise. If $\eta_{ij} > 0$ the assets are gross substitutes, if $\eta_{ij} < 0$ they are gross complements, and if $\eta_{ij} = 0$ they are independent.

As an example, the elasticity formulas for the homothetic translog are

$$\eta_{iy} = 1, \quad \eta_{ii} = \frac{\beta_{ii}}{s_i} - 1, \quad \text{and} \quad \eta_{ij} = \frac{\beta_{ij}}{s_i}.$$

22.8 Elasticities of Substitution

We can also interpret the estimated parameter values by computing elasticities of substitution. One reason for our interest in these measures is that the degree of substitutability among monetary assets has been used — explicitly or implicitly — to provide a rationale for the appropriate definition of money, which as we have seen has been the focus of continuing controversy over the years. Moreover, knowledge of the substitutability between monetary assets is essential in order to understand the potential effects of monetary policy actions as well as the effects of the growth of financial intermediation.

There are currently two methods employed for calculating the partial elasticity of substitution between two variables — the Allen and the Morishima. The Allen elasticity of substitution (AES) between two liquid assets i and j, denoted by σ_{ij}^a, can be calculated from the income and price elasticities, using the Slutsky equation (to be discussed in detail in the next section)

$$\sigma_{ij}^a = \eta_{iy} + \frac{\eta_{ij}}{s_j}.$$

It categorizes goods as complements if an increase in the price of asset j causes a decreased consumption of asset i ($\sigma_{ij}^a < 0$). If $\sigma_{ij}^a > 0$, goods are Allen substitutes. Alternatively, following Diewert (1974) and Gallant (1981), the Allen elasticity of substitution can be computed from the estimated indirect utility function as follows

$$\sigma_{ij}^a = \frac{\left[\sum_k \boldsymbol{v}_k V_k\right] V_{ij}}{V_i V_j} - \frac{\sum_k \boldsymbol{v}_k V_{jk}}{V_j} - \frac{\sum_k \boldsymbol{v}_k V_{ik}}{V_i} + \frac{\sum_m \sum_k \boldsymbol{v}_k V_{km} \boldsymbol{v}_m}{\sum_n \boldsymbol{v}_n V_n},$$

where V_i and V_{ij} denote elements of $\partial h(\boldsymbol{v})/\partial \boldsymbol{v}$ and $\partial^2 h(\boldsymbol{v})/\partial \boldsymbol{v} \partial \boldsymbol{v}'$, respectively.

Although the AES has been used widely to study substitution behavior and structural instability, Blackorby and Russell (1989) have shown that the AES is quantitatively and qualitatively uninformative and that the Morishima elasticity of substitution (MES) is the correct measure of the substitution elasticity. The Morishima elasticity of substitution, denoted by σ_{ij}^m, is defined as [see Blackorby and Russell (1989) for more details]

$$\sigma_{ij}^m = s_i(\sigma_{ji}^a - \sigma_{ii}^a),$$

and addresses impacts on the ratios of two goods. In particular, it categorizes goods as complements ($\sigma_{ij}^m < 0$) if an increase in the price of j causes x_i/x_j to decrease. If $\sigma_{ij}^m > 0$, goods are Morishima substitutes.

Comparing the AES and the MES, we see that [since σ_{ii}^a is always positive (given negative own price elasticities)] if two goods are Allen substitutes ($\sigma_{ji}^a > 0$) they must also be Morishima substitutes ($\sigma_{ij}^m > 0$). However, two goods may be Allen complements ($\sigma_{ji}^a < 0$), but Morishima substitutes if $|\sigma_{ii}^a| > |\sigma_{ji}^a|$, suggesting that the AES always overstates the complementarity relationship. Moreover, the AES matrix is symmetric ($\sigma_{ij}^a = \sigma_{ji}^a$), but the MES matrix is not — Blackorby and Russell (1981) show that the MES matrix is symmetric only when the aggregator function is a member of the CES family.

22.9 Conclusion

In this chapter we have illustrated some basic aspects of demand system specification that will provide context and motivation for the discussion in Chapter 23. One of the possible ways of improving on the foregoing is to pay explicit attention to the time series properties of the data. A recent finding in the econometrics literature is that estimation and hypotheses testing critically depend on the integration and cointegration properties of the variables. For example, in the context of linear demand systems such as the homothetic translog and the AIDS, Ng (1995) and Attfield (1997) test the null hypothesis of homogeneity (with respect to prices and nominal income) and show that this cannot be rejected once the time series properties of the data are imposed in estimation.

This implies that testing for cointegration [in the spirit of Engle and Granger (1987)], and constructing a form of the error correction model is appropriate. Most demand systems, however, have share equations that are nonlinear and, as Granger (1995) points out, nonlinear modeling of nonstationary variables is a new, complicated, and largely undeveloped area. We generally ignore this issue in this book, keeping in mind that this is an area for potentially productive future research.

23

Applied Monetary Demand Analysis

In this chapter we consider and illustrate a solution to the interrelated problems of monetary aggregation and estimation of money demand functions. In doing so, we use quarterly U.S. data and take a demand systems approach. We handle the problem in two stages: (i) we aggregate liquid assets using a superlative index — the Divisia index; and (iii) we use a flexible demand system to deal with the problem of money demand.

Our objective is to estimate income, price, and elasticities of substitution, by estimating a system of demand equations derived from the indirect utility function. In order to do so, we need to choose a functional form that will accurately approximate both the true indirect utility function and its partial derivatives. As already noted, parametric functions (such as, for example, the Cobb-Douglas) fail to accurately approximate the data generating function and often restrict the substitutability/complementarity relationship between money and near monies.

23.1 The Monetary Problem

Following Serletis and Shahmoradi (2005, 2007), we assume that the representative money holder faces the following problem

$$\max_{\boldsymbol{x}} f\left(\boldsymbol{x}\right) \quad \text{subject to} \quad \boldsymbol{p}'\boldsymbol{x} = y \tag{23.1}$$

where $\boldsymbol{x} = (x_1, x_2, \cdots, x_8)$ is the vector of monetary asset quantities described in Table 23.1; $\boldsymbol{p} = (p_1, p_2, \cdots, p_8)$ is the corresponding vector of monetary asset user costs; and y is the expenditure on the services of monetary assets. Because demand system estimation requires heavy dimension reduction (as already noted in Chapter 22), we follow Serletis and Shahmoradi (2005, 2007) and separate the group of assets into three collections based on empirical pre-testing. Thus the monetary utility function in (23.1) can be written as

$$f\left(\boldsymbol{x}\right) = f\left(f_A\left(x_1, x_2, x_3, x_4\right),\ f_B\left(x_5, x_6\right),\ f_C\left(x_7, x_8\right)\right)$$

where the subaggregate functions f_i $(i = A, B, C)$ provide subaggregate measures of monetary services.

Table 23.1. Monetary Assets/Components

A	1 Currency + Travelers checks
	2 Demand deposits
	3 Other checkable deposits at banks including Super Now accounts
	4 Other checkable deposits at thrifts including Super Now accounts
B	5 Savings deposits at banks including money market deposit accounts
	6 Savings deposits at thrifts including money market deposit accounts
C	7 Small denomination time deposits at commercial banks
	8 Small denomination time deposits at thrift institutions

Although not the same, this structure of preferences is very similar to the one uncovered by Fisher and Fleissig (1994) and also used by Fleissig and Swofford (1996) and Fisher and Fleissig (1997) when they estimated their money demand models. Fisher and Fleissig (1994) found, using the NONPAR program of Varian (1982, 1983), that these groups of assets satisfy the weak separability condition for several Generalized Axiom of Revealed Preference (GARP) consistent subperiods.

Instead of using the simple-sum index, currently in use by the Federal Reserve and most central banks around the world, to contruct the monetary subbaggregates, f_i $(i = A, B, C)$, we use the Divisia quantity index to allow for less than perfect substitutability among the relevant monetary components. Recall that the Divisia index (in discrete time) is defined as

$$\log M_t^D - \log M_{t-1}^D = \sum_{j=1}^{n} w_{jt}^* (\log x_{jt} - \log x_{j,t-1}),$$

according to which the growth rate of the aggregate is the weighted average of the growth rates of the component quantities, with the Divisia weights being defined as the expenditure shares averaged over the two periods of the change, $w_{jt}^* = (1/2)(w_{jt} + w_{j,t-1})$ for $j = 1, ..., n$, where $w_{jt} = p_{jt}x_{jt} / \sum p_{kt}x_{kt}$ is the expenditure share of asset j during period t, and p_{jt} is the nominal user cost of asset j, derived in Barnett (1978),

$$p_{jt} = p^* \frac{R_t - r_{jt}}{1 + R_t},$$

which is just the opportunity cost of holding a dollar's worth of the jth asset. Above, p^* is the true-cost of living index, r_{jt} is the market yield on the jth asset, and R_t is the yield available on a 'benchmark' asset that is held only to carry wealth between multiperiods.

23.2 The Basic Translog and the Demand for Money

We begin by using the basic translog functional form to approximate the unknown indirect utility function. As we argued in Chapter 20, however, the translog is capable of approximating an arbitrary function only locally (at a point), and that a more constructive approach would be based on the use of flexible functional forms that possess global properties. Thus, the results in this section are not definitive in any sense, but are meant to demonstrate the demand systems methodology described in the discussion thus far.

The basic translog functional form, discussed in Chapter 20, with the symmetry restrictions, $\beta_{ij} = \beta_{ji}$, imposed for the three-asset ($n = 3$) case can be written as

$$\log h(\boldsymbol{v}) = a_0 + \sum_{k=1}^{3} a_k \log v_k + \frac{1}{2} \sum_{k=1}^{3} \sum_{j=1}^{3} \beta_{jk} \log v_k \log v_j.$$

Applying the logarithmic form of Roy's identity allows us to derive the model's share equations,

$$s_i = \frac{a_i + \sum_{k=1}^{3} \beta_{ik} \log v_k}{\sum_{k=1}^{3} a_k + \sum_{k=1}^{3}\sum_{j=1}^{3} \beta_{jk} \log v_k} + \epsilon_i, \qquad i = 1, 2, 3.$$

The disturbance terms e_i $(i = 1, 2, 3)$ have been added to capture deviations of the observed shares from the true shares. As we argued in Chapter 22, the errors are assumed to be additive, jointly normally distributed with zero means, and with constant but unknown variances and covariances. This distributional assumption on the errors is standard and is fundamental in the derivation of the FIML estimator.

Since demand theory provides that the budget shares sum to 1, it follows that the disturbance covariance matrix is singular. If autocorrelation in the disturbances is absent (as assumed here), Barten (1969) showed that FIML estimates of the parameters can be obtained by arbitrarily deleting an equation in such a system and that the estimation results are invariant with respect to the equation deleted. Thus, with three shares one equation must be dropped, and only two equations are estimated. Here, we drop the third equation and estimate the remaining two equations. Notice also that estimation of this system requires some parameter normalization, as the share equations are homogeneous of degree zero in the a's. We use the normalization $a_1 + a_2 + a_3 = 1$, and in terms of the variables defined above, we estimate

$$s_1 = \frac{a_1 + \beta_{11} \log v_1 + \beta_{12} \log v_2 + \beta_{13} \log v_3}{1 + \sum_{j=1}^{3} \sum_{i=1}^{3} \beta_{ji} \log v_j} + \epsilon_1;$$

$$s_2 = \frac{a_2 + \beta_{12} \log v_1 + \beta_{22} \log v_2 + \beta_{23} \log v_3}{1 + \sum_{j=1}^{3} \sum_{i=1}^{3} \beta_{ji} \log v_j} + \epsilon_2.$$

This system has 8 free parameters — that is, 8 parameters estimated directly. These parameters are: a_1, a_2 β_{11} β_{12} β_{13} β_{22} β_{23}, and β_{33}.

23.2.1 Data and Econometric Issues

We use the same quarterly data set (from 1970:1 to 2003:2, a total of 134 observations) as in Serletis and Shahmoradi (2005, 2007). It consists of asset quantities and nominal user costs for the eight items listed in Table 23.1, obtained from the Monetary Services Indices (MSI) project of the Federal Reserve Bank of St. Louis. As we require real per capita asset quantities for our empirical work, we have divided each measure of monetary services by the U.S. CPI (all items) and total U.S. population in each period. The calculation of the user costs, which are the appropriate prices for monetary services, has been explained earlier.

Prior to estimation (and the logarithmic transformation of the data), the income-normalized prices were normalized again by dividing each v_i by its mean in order to place the Taylor's expansion around the point $\boldsymbol{v}^* = 1$. The system was estimated using the FIML regression procedure in TSP International (version 4.5) — convergence is set at 0.00001. As with vector autoregressions and other time series models, there are many parameters to be estimated and it does not matter if all the parameters are statistically significant or not — what is important is for the model to fit the data well.

As results in nonlinear optimization are sensitive to the initial parameter values, to avoid being caught in local minima and in order to achieve global convergence, we randomly generate sets of initial parameter values and choose those parameter estimates that lead to the lowest value of the objective function.

23.2.2 Empirical Evidence

The parameter estimates that minimize the objective function are reported in Table 23.2, with p-values in the last column. We also report the number of positivity, monotonicity, and curvature violations, checked as we discussed in Chapter 22.

Clearly, although the model satisfies positivity and monotonicity at all sample observations, it violates curvature at most observations (in particular, at 65 observations). Because regularity hasn't been attained (by luck), we follow the suggestions by Barnett (2002) and Barnett and Pasupathy (2003) and estimate the model by imposing local curvature. We impose local curvature using the Ryan and Wales (1998) and Moschini (1999) procedures, discussed in detail in Chapter 20. In particular, we impose the following three restrictions [see Chapter 20 for more details]:

Table 23.2. Basic Translog Parameter Estimates

Parameter	Estimate	p-value
a_1	.412	.004
a_2	.290	.003
β_{11}	.770	.114
β_{12}	.117	.143
β_{13}	.461	.126
β_{22}	.249	.088
β_{23}	.228	.111
β_{33}	.511	.112
Positivity violations	0	
Monotonicity violations	0	
Curvature violations	65	

Notes: Quarterly data 1970:1-2003:2 ($T = 134$).

$$\beta_{11} = -k_{11}^2 + a_1 + 2a_1\theta_1 - a_1^2 (1 + \theta_1 + \theta_2 + \theta_3);$$
$$\beta_{12} = -k_{11}k_{21} + a_1\theta_2 + a_2\theta_1 - a_1a_2 (1 + \theta_1 + \theta_2 + \theta_3);$$
$$\beta_{22} = -k_{21}^2 - k_{22}^2 + a_2 + 2a_2\theta_2 - a_2^2 (1 + \theta_1 + \theta_2 + \theta_3),$$

on the share equations.

Also, as noted by Ryan and Wales (1998), the ability of locally flexible models to satisfy curvature at other sample observations other than the point of approximation, depends on the choice of approximation point. Thus, we estimated the model 134 times (a number of times equal to the number of observations) and report results for the best approximation point (best in the sense of satisfying the curvature conditions at the largest number of observations). The best approximation point is 2002:4. The results are reported in Table 23.3 in the same fashion as those in Table 23.2.

Our findings in terms of regularity violations when the curvature conditions are imposed are disappointing. In particular, the imposition of local curvature reduces the number of curvature violations from 65 to 50. This means that based on this model inferences about money demand (including those about income and price elasticities as well as the elasticities of substitution) will not significantly improve our understanding of real world money demand.

Table 23.3. Basic Translog Parameter Estimates With Local Curvature Imposed

Parameter	Estimate	p-value
a_1	.398	.011
a_2	.248	.016
β_{11}	.140	.012
β_{12}	.041	.013
β_{13}	.089	.004
β_{22}	.059	.005
β_{23}	.052	.006
β_{33}	.093	.008
Positivity violations	0	
Monotonicity violations	0	
Curvature violations	50	

Notes: Quarterly data 1970:1-2003:2 ($T = 134$).

23.2.3 Regularity Effects of Serial Correlation Correction

We have used a static model, implicitly assuming that the pattern of demand adjusts to a change in exogenous variables instantaneously. We paid no attention to the dynamic structure of the model used, although many recent studies report results with serially correlated residuals suggesting that the underlying models are dynamically misspecified. Autocorrelation in the disturbances has mostly been dealt with by assuming a first-order autoregressive process (discussed in detail in Chapter 22) — see, for example, Ewiss and Fisher (1984), Serletis and Robb (1986), Serletis (1987, 1988), Fisher and Fleissig (1994, 1997), Fleissig (1997), Fleissig and Swofford (1996, 1997), Fleissig and Serletis (2002), and Drake and Fleissig (2004).

Here we investigate the effects on theoretical regularity of serial correlation corrections by allowing the possibility of a first-order autoregressive process in the error terms, as follows

$$\boldsymbol{\epsilon}_t = \boldsymbol{\rho}\boldsymbol{\epsilon}_{t-1} + \boldsymbol{e}_t,$$

where $\boldsymbol{\rho} = [\rho_{ij}]$ is a matrix of unknown parameters and $\boldsymbol{e}_t$ is a non-autocorrelated vector disturbance term with constant covariance matrix. As already noted in Chapter 22, estimates of the parameters can be obtained by using a result developed by Berndt and Savin (1975). They showed that if one assumes no autocorrelation across equations (i.e., $\boldsymbol{\rho}$ is diagonal), the autocorrelation coefficients for each equation must be identical. Consequently, we can estimate (using FIML procedures) stochastic budget share equations given by

$$\boldsymbol{s}_t = \boldsymbol{\psi}(\boldsymbol{v}_t, \boldsymbol{\theta}) + \boldsymbol{\rho} \boldsymbol{s}_{t-1} - \boldsymbol{\rho} \boldsymbol{\psi}(\boldsymbol{v}_{t-1}, \boldsymbol{\theta}) + \boldsymbol{e}_t.$$

We estimated the above equation for the basic translog and observed that serial correlation correction increases the number of curvature violations (from 65 to 93) and also leads to induced violations of monotonicity (at 40 data points) — see Serletis and Shahmoradi (2007) for more details. It seems that the current practice of correcting for serial correlation without reporting the results of monotonicity checks (even when the curvature conditions are imposed) is not justified. Moreover, allowing for first order serial correlation is almost the same as taking first differences of the data if the autocorrelation coefficient is close to unity. In that case, the equation errors become stationary, but there is no theory for the models in first differences.

We believe that in order to deal with dynamically misspecified models attention should be focused in the development of unrestricted dynamic formulations to accommodate short-run disequilibrium situations as, for example, in Serletis (1991) who builds on the Anderson and Blundell (1982) approach to dynamic specification in the spirit of error correction models. Alternatively, attention should be focused in the development of dynamic generalizations of the traditional static models by considering specific theories of dynamic adjustment.

23.3 The AIM(2) Model and the Demand for Money

Because of the dissapointing results with the basic translog demand system, in this section we build on Serletis and Shahmoradi (2005) and estimate the AIM model using the same optimization procedures as in Gallant and Golub (1984) and Serletis and Shahmoradi (2005), already discussed in Chapter 21 of this book. In particular, we estimate the AIM(2) demand system (reproduced here):

$$\begin{aligned}
s_1 = \Big(& 2b_1 v_1^{1/2} + b_4 v_1^{1/4} + 2b_7 v_1^{1/2} v_2^{1/2} + 2b_8 v_1^{1/2} v_2^{1/4} + b_9 v_1^{1/4} v_2^{1/2} \\
&+ b_{10} v_1^{1/4} v_2^{1/4} + 2b_{11} v_1^{1/2} v_3^{1/2} + 2b_{12} v_1^{1/2} v_3^{1/4} + b_{13} v_1^{1/4} v_3^{1/2} \\
&+ b_{14} v_1^{1/4} v_3^{1/4} + 2b_{19} v_1^{1/2} v_2^{1/2} v_3^{1/2} + b_{20} v_1^{1/4} v_2^{1/2} v_3^{1/2} \\
&+ 2b_{21} v_1^{1/2} v_2^{1/4} v_3^{1/2} + 2b_{22} v_1^{1/2} v_2^{1/2} v_3^{1/4} + 2b_{23} v_1^{1/2} v_2^{1/4} v_3^{1/4} \\
&+ b_{24} v_1^{1/4} v_2^{1/2} v_3^{1/4} + b_{25} v_1^{1/4} v_2^{1/4} v_3^{1/2} + b_{26} v_1^{1/4} v_2^{1/4} v_3^{1/4}\Big) / D \qquad (23.2)
\end{aligned}$$

$$\begin{aligned}
s_2 = \Big(& 2b_2 v_2^{1/2} + b_5 v_2^{1/4} + 2b_7 v_1^{1/2} v_2^{1/2} + b_8 v_1^{1/2} v_2^{1/4} + 2b_9 v_1^{1/4} v_2^{1/2} \\
&+ b_{10} v_1^{1/4} v_2^{1/4} + 2b_{15} v_2^{1/2} v_3^{1/2} + 2b_{16} v_2^{1/2} v_3^{1/4} + b_{17} v_2^{1/4} v_3^{1/2} \\
&+ b_{18} v_2^{1/4} v_3^{1/4} + 2b_{19} v_1^{1/2} v_2^{1/2} v_3^{1/2} + 2b_{20} v_1^{1/4} v_2^{1/2} v_3^{1/2} \\
&+ b_{21} v_1^{1/2} v_2^{1/4} v_3^{1/2} + 2b_{22} v_1^{1/2} v_2^{1/2} v_3^{1/4} + b_{23} v_1^{1/2} v_2^{1/4} v_3^{1/4} \\
&+ 2b_{24} v_1^{1/4} v_2^{1/2} v_3^{1/4} + b_{25} v_1^{1/4} v_2^{1/4} v_3^{1/2} + b_{26} v_1^{1/4} v_2^{1/4} v_3^{1/4}\Big) / D \qquad (23.3)
\end{aligned}$$

$$\begin{aligned}
s_3 = \Big(& 2b_3 v_3^{1/2} + b_6 v_4^{1/4} + 2b_{11} v_1^{1/2} v_3^{1/2} + b_{12} v_1^{1/2} v_3^{1/4} + 2b_{13} v_1^{1/4} v_3^{1/2} \\
&+ b_{14} v_1^{1/4} v_2^{1/4} + 2b_{15} v_1^{1/2} v_3^{1/2} + b_{16} v_1^{1/2} v_3^{1/4} + 2b_{17} v_2^{1/4} v_3^{1/2} \\
&+ b_{18} v_2^{1/4} v_3^{1/4} + 2b_{19} v_1^{1/2} v_2^{1/2} v_3^{1/2} + 2b_{20} v_1^{1/4} v_2^{1/2} v_3^{1/2} \\
&+ 2b_{21} v_1^{1/2} v_2^{1/4} v_3^{1/2} + b_{22} v_1^{1/2} v_2^{1/2} v_3^{1/4} + b_{23} v_1^{1/2} v_2^{1/4} v_3^{1/4} \\
&+ b_{24} v_1^{1/4} v_2^{1/2} v_3^{1/4} + 2b_{25} v_1^{1/4} v_2^{1/4} v_3^{1/2} + b_{26} v_1^{1/4} v_2^{1/4} v_3^{1/4}\Big) / D \qquad (23.4)
\end{aligned}$$

where D is the sum of the numerators in equations (23.2), (23.3), and (23.4).

Since the usefulness of flexible functional forms depends on whether they satisfy the theoretical regularity conditions (of positivity, monotonicity, and curvature), we impose the curvature restriction globally, using the methods that we discussed in Chapter 21. Using NPSOL we performed the computations and report the parameter estimates in Table 23.4, together with the minimized value of the objective function, in the last row of the table.

Table 23.4. AIM(2) Parameter Estimates

Parameter	Estimate	Parameter	Estimate
b_1	-6.926	b_{15}	3.568
b_2	-1.935	b_{16}	-13.164
b_4	-2.977	b_{17}	-4.136
b_5	-14.185	b_{18}	5.035
b_6	-4.432	b_{19}	-7.425
b_7	-3.326	b_{20}	2.626
b_8	-11.115	b_{21}	13.912
b_9	14.818	b_{22}	-0.721
b_{10}	2.416	b_{23}	1.980
b_{11}	-12.377	b_{24}	5.727
b_{12}	12.813	b_{25}	-7.050
b_{13}	-10.896	b_{26}	5.627
b_{14}	4.425		

Value of the objective function: .0236.

Note: Quarterly data 1970:1-2003:2.

The imposition of curvature globally does not produce spurious violations of monotonicity, mentioned by Barnett and Pasupathy (2003), thereby assuring true theoretical regularity. Hence, in what follows we discuss the income and price elasticities as well as the elasticities of substitution based on the AIM(2) model which (with our data set) satisfies both the neoclassical monotonicity and curvature conditions.

23.3.1 AIM(2) Income and Price Elasticities

In the demand systems approach to estimation of economic relationships, the primary interest, especially in policy analysis, is in how the arguments of the underlying functions affect the quantities demanded. This is conventionally and completely expressed in terms of income and price elasticities and in elasticities of substitution. We are interested in the policy issues, but we are also interested in the informational context of the the model.

We begin by presenting the income elasticities (η_{iy}) in Table 23.5, evaluated at the mean of the data, for the three subaggregates, A, B, and C — all elasticities have been acquired using numerical differentiation and the formulas presented in Chapter 22. η_{Ay}, η_{By}, and η_{Cy}

are all positive (suggesting that assets A, B, and C are all normal goods) which is consistent with economic theory. In Table 23.5 we also show the uncompensated (Cournot) own- and cross-price elasticities, evaluated at the mean of the data. The own-price elasticities are all negative, as predicted by the theory. For the cross-price elasticities, economic theory does not predict any signs, but we note that most of the off-diagonal terms are negative, indicating that the assets taken as a whole, are gross complements. This is a frequent finding in the literature.

Table 23.5. Income and Price Elasticities

Asset	η_{iy}	η_{iA}	η_{iB}	η_{iC}
A	.988	−.551	−.225	−.211
B	1.821	−.750	−.751	−.322
C	.115	.025	.130	−.270

Note: Quarterly data 1970:1-2003:2

23.3.2 AIM(2) Elasticities of Substitution

From the point of view of monetary policy, the measurement of the elasticities of substitution among the three monetary assets is of prime importance. The currently popular simple sum approach to monetary aggregation requires that the elasticities of substitution be very high (perhaps infinite) among the components of the monetary aggregates. An additional concern relates to the volatility of the elasticities. Specifically, if there is evidence of significant volatility in the elasticies of substitution (with models such as the AIM(2) that satisfies all three theoretical regularity conditions), the simple sum aggregates will surely be invalid and methods of aggregation that allow for variable elasticities of substitution would be preferable.

In Table 23.6 we show estimates of the Allen elasticities of substitution, evaluated at the means of the data. We expect the three diagonal terms, representing the own-elasticities of substitution for the three assets, to be negative. This expectation is clearly achieved. However, because the Allen elasticity of substitution produces ambiguous results off-diagonal, we will use the Morishima elasticity of substitution to investigate the substitutability/complementarity relation between money and near money.

Table 23.6. Allen Elasticities

Asset	σ^a_{iA}	σ^a_{iB}	σ^a_{iC}
A	−.212	.190	.170
B		−.833	.575
C			−.934

Note: Quarterly data 1970:1-2003:2.

Based on the asymmetrical Morishima elasticities of substitution — the correct measures of substitution — the assets are all Morishima substitutes, as documented in Table 23.7. Moreover, all Morishima elasticities of substitution are less than unity. This clearly indicates difficulties for a simple-sum based monetary policy and helps explain why recent attempts to target and control the money supply (simple sum M2) have been abandoned in favor of interest rate procedures.

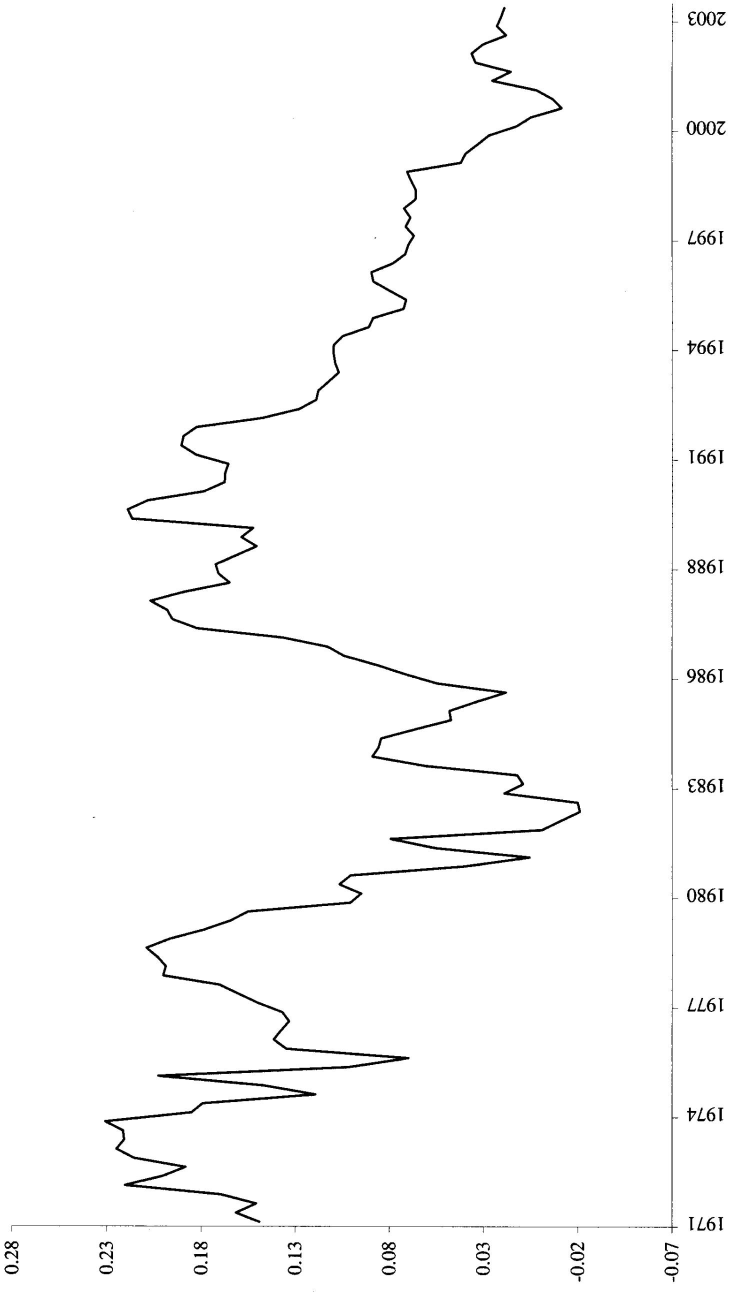
Table 23.7. Morishima Elasticities

Asset	σ^m_{iA}	σ^m_{iB}	σ^m_{iC}
A		.185	.176
B	.289		.427
C	.285	.363	

Note: Quarterly data 1970:1-2003:2.

Because we are providing global approximations, it is also interesting to present graphs for the Morishima elasticities, as those in Figures 23.1 and 23.2. As already noted, the Morishima approach to the calculation of the elasticity of substitution provides a different estimate depending on which asset price is varied (of the two being considered). For example, Figure 23.1 shows the Morishima elasticity between assets A and B with the price of A changing whereas Figure 23.2 shows the same elasticity with the price of B varying, in effect approaching

Figure 23.1 . AIM(2) Morishima Elasticity of Substitution Between A and B with the Price of A Changing



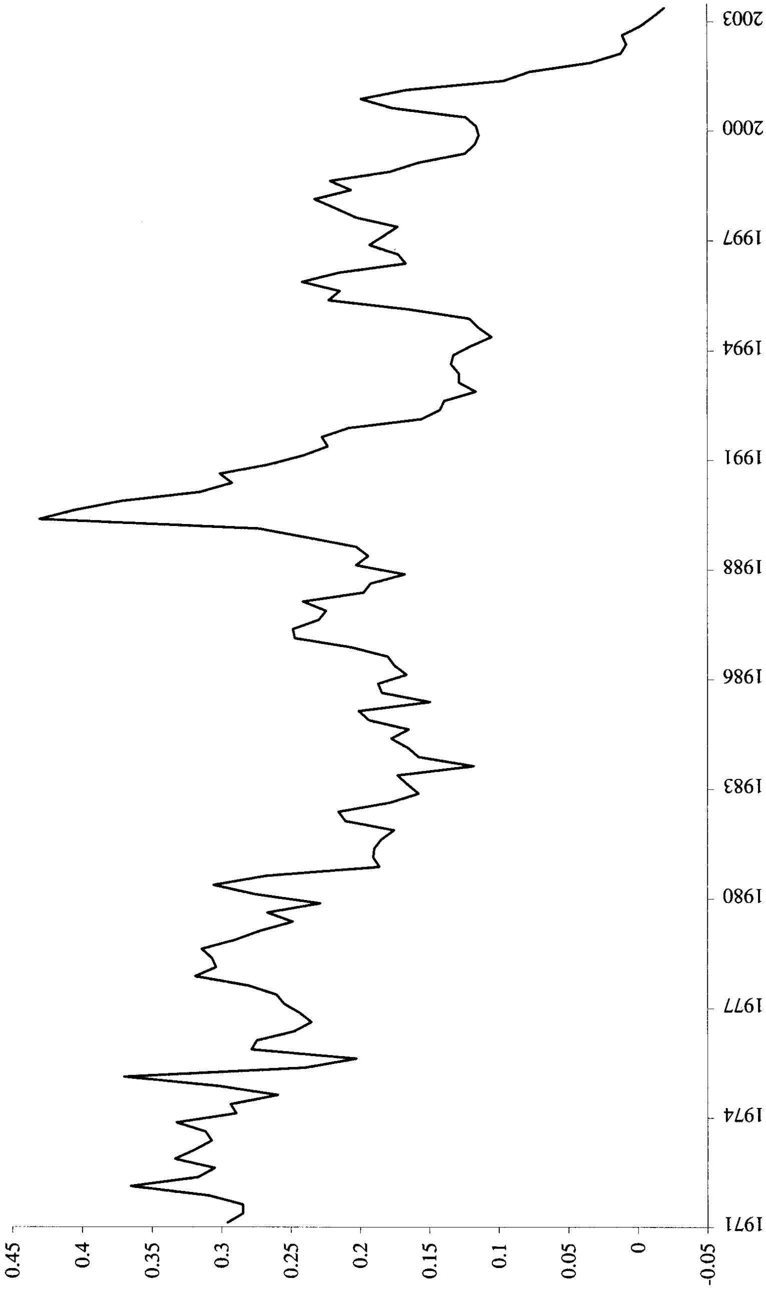


Figure 23.2. AIM(2) Morishima Elasticity of Substitution Between A and B with the Price of B Changing

from a different direction. As expected, there are no inconsistencies in the elasticity calculations and all the estimates are less than unity over the entire sample, showing mild substitutabilty (no matter what price is varied in the Morishima calculation).

23.4 Conclusion

We have investigated the substitutabilty of money and near monies in the United States in the context of the AIM(2) model. The model satisfies the theoretical regularity conditions (of positivity, monotonicity, and curvature) and our parameter and elasticity estimates are therefore consistent with neoclassical microeconomic theory. We have also provided a policy perspective, using parameter estimates that are consistent with global regularity, in that a very strong case can be made for abandoning the simple sum approach to monetary aggregation, on the basis of the low and volatile elasticities of substitution among the components of the popular M2 aggregate of money. This is the same (qualitative) conclusion reached by Serletis and Shahmoradi (2005) using the Fourier and the AIM(2) models and Serletis and Shahmoradi (2007) using a number of locally flexible demand system specifications.

We have thus provided a solution to the inter-related problems of monetary aggregation and estimation of money demand, using demand systems analysis. As Fisher (1989, p. xiii) puts it,

> "[t]his work represents a literature that has developed in the monetary context over the last 15 or so years and one that now offers theoretical and empirical insights on both the nature of what we might call 'the monetary problem' and on most of the major issues involving monetary policy."

We have also addressed several other matters that we haven't simultaneously examined. A discussion of these matters is the subject of the next, final, chapter of this book.

24

Future Research Agenda

Clearly many unsolved problems exist in what Barnett (1997) calls the 'high road' literature regarding the inter-related problems of monetary aggregation and estimation of money demand. These problems continue to be the subject of expanding research. In fact, as Barnett (1997, p. 1182) puts it,

> "the high road builds on the foundation of existing microeconomic theory, including the theory of the firm, consumer theory, and the implied microeconomic aggregation and index number theory. Advances in those areas are accepted and absorbed in a coherent manner as research proceeds up the high road. Constructive criticism of the high road is based upon recognition of the existence of unsolved problems in the supporting areas of economic research and the need for further development in those areas to permit the high road to climb even higher."

With this in mind, in this concluding chapter of the book we address a number of issues and point particular directions (not necessarily in order of importance) that high road research could take.

24.1 Outstanding Credit

The monetary aggregates currently available (either simple-sum, Divisia, or currency equivalent) exclude unused credit, potentially generating spurious instability of money demand. Given that a conceptually more appropriate measure of the quantity of the medium of exchange should assign a nonzero weight to outstanding credit, the current monetary aggregation literature should be extended to include in the definition of the medium of exchange some measure of outstanding credit available to the public.

Outstanding credit is a very useful asset. The problem is how to price assets such as, for example, overdraft facilities of firms and individuals and approved credit lines available to the public, in order to be explicitly included in microeconomic- and aggregation-theoretic monetary aggregates. The gains remain to be seen, of course, but we suggest that many of the anomalies in the literature on monetary topics will be eliminated when theoretically designed index numbers are employed.

24.2 Monetary Policy

An exposition is needed of what the usefulness of the estimated elasticities of substitution and, more importantly, the underlying demand equations would be in practice. That is, to what extent are higher or lower elasticities of substitution important for the conduct of effective monetary policy? To what use would the central bank put this information in its actions? What are the practical implications of using the wrong elasticities?

A breakthrough from the current state of 'interest target' monetary policy back to the correct control of monetary quantities will be through demand systems. The research agenda is clear, and it starts with getting across the procedures and showing that elasticities make sense and that the properties of the models are nicely neoclassical. Since this has been done, the next step is to analyze actual policies, using actual demand system estimates, compared to the other ways of guiding policy.

24.3 Dynamics

We should note that most of the early studies of demand systems directly applied data to static models, implicitly assuming that the pattern of demand adjusts to a change in exogenous variables instantaneously. No attention had been paid to the dynamic structure of the models used, although many studies reported results with serially correlated residuals suggesting that the underlying models are perhaps dynamically misspecified. In fact, most of the early money demand studies tended to ignore dynamic issues, although the issue of dynamic adjustment has been considerably addressed in the 'traditional' log-levels money demand specification by considering different, short-run adjustment processes. In particular, in Chapter 10 we considered three fundamentally different, short-run dynamic adjustment processes in the traditional approach — the 'real' adjustment specification, the 'price' adjustment specification, and the 'nominal' adjustment specification.

Recently, however, a number of demand studies have focussed attention on the development of dynamic generalizations of the traditional static models that allow a test of the static model itself, as well as the theoretical restrictions and simplifications from demand theory. For example, Serletis (1991a) develops microtheoretic dynamic generalizations of four static translog models, by appealing to the habit *hysteresis* theory and assuming that consumers' current preferences depend on their past consumption pattern so that lagged variables will influence current demand. Also, Gordon Anderson and Richard Blundell (1982) motivated from the lack of accord between the postulates of demand theory and empirical static demand functions estimated on time series data, develop an unrestricted dynamic formulation to accommodate short-run disequilibrium situations, by including lagged endogenous and exogenous variables as regressors.

To illustrate the Anderson and Blundell (1982) approach, let's consider the share equation system of the HTL. The dynamics are introduced by replacing the usual static assumption of instantaneous adjustment by a more general one that the static model holds only asymptotically. Under this assumption it is possible to specify the dynamic structure (data generation process) by a general stationary stochastic process. In particular, following Anderson and Blundell (1982), we replace $\boldsymbol{s}$ in equation (22.2) by a vector autoregressive process in $\boldsymbol{s}$ of order 1 and $\boldsymbol{v}$ by a vector autoregressive process in $\boldsymbol{v}$ of order 1. After some manipulation and consideration of the adding up restrictions — see Anderson and Blundell (1982, p. 1560-66) — a general first-order dynamic model may be written as

$$\Delta \boldsymbol{s}_t = \boldsymbol{D}\Delta\widetilde{\boldsymbol{v}}_t - \boldsymbol{A}\left(\boldsymbol{s}^n_{t-1} - \boldsymbol{\Pi}^n \boldsymbol{v}_{t-1}\right) + \boldsymbol{u}_t \qquad (24.1)$$

where Δ represents the first difference operator, $\widetilde{\boldsymbol{v}}$ refers to $\boldsymbol{v}$ with the first element excluded, superscript n on a matrix or a vector denotes the deletion of the nth row and $\boldsymbol{D}$ and $\boldsymbol{A}$ are appropriately dimensioned short-run coefficient matrices. Note that the adding-up restrictions associated with (22.2) require certain additional restrictions on the elements of $\boldsymbol{D}$ and $\boldsymbol{A}$ in (24.1). These imply that the column sums of $\boldsymbol{D}$ and $\boldsymbol{A}$ in (24.1) are all zero.

The advantage of estimating in the context of (24.1) is that equation (24.1) is the alternative hypothesis against which a number of hypotheses can be tested. For example, if $\boldsymbol{D} = \boldsymbol{\Pi}_1$, where $\boldsymbol{\Pi}_1$ denotes $\boldsymbol{\Pi}$ with the first column corresponding to the intercept term deleted, equation (24.1) reduces to the static model with AR(1) error term,

$$\boldsymbol{s}_t = \boldsymbol{\Pi} \boldsymbol{v}_t + \boldsymbol{u}_t,$$

$$\boldsymbol{u}_t = \boldsymbol{\rho} \boldsymbol{u}_{t-1} + \boldsymbol{e}_t.$$

If $\boldsymbol{D} = \boldsymbol{A}\boldsymbol{\Pi}_1$, equation (24.1) reduces to the partial adjustment model considered by Ishag Nadiri and Sherwin Rosen (1969). Finally, if $\boldsymbol{D} = \boldsymbol{\Pi}_1$, and $\boldsymbol{A} = \boldsymbol{I}$, equation (24.1) reduces to the static model (22.1).

Notice that the Anderson and Blundell approach to dynamic specification, adopted in the money demand literature by Serletis (1991b), follows in the spirit of the error correction models and stands in contrast to the theoretical approaches that maintain specific theories of dynamic adjustment. It is, however, intuitively appealing as it seems that no theoretical approach is likely to deal with the actual dynamics of a demand system, which are likely to be a complicated amalgam of effects, including habit persistence, adjustment costs, the formation of expectations, and misinterpretation of real price changes.

More recently, Fisher and Fleissig (1994) produce two versions of the dynamic Fourier demand system — one is a 'dynamic utility function approach,' following the lead of Serletis (1991a), and the other is a 'time series approach,' following the lead of Anderson and Blundell (1982, 1983) and Serletis (1991b). For other interesting follow-up papers that utilize some of the flexible functional forms that we discussed in Chapters 20 and 21, see Fleissig and Swofford (1996, 1997), Fleissig (1997), Fisher and Fleissig (1997), Leigh Drake, Fleissig, and Andy Mullineaux (1999), and Fleissig and Serletis (2001).

Those works are interesting and attractive; they include estimates of the degree of substitutability using some constrained flexible functional forms or unconstrained versions used to test for theoretical and functional form restrictions. Yet Serletis (1991b), in addition to modeling the demand for aggregate Divisia money measures, systematically tests for the appropriateness of the weak separability (aggregation) conditions using flexible functional form interpretations of the translog functional form. All these specifications, however, simply add lagged variables to the system. There is no explicit dynamic optimization framework under consideration. We believe that a particularly constructive approach will be based on the use of dynamic models.

Moreover, high autocorrelation in demand systems estimated with per capita data could also be due to either the effects of nonstationarity of prices — see, for example, Arthur Lewbel (1996) — and/or aggregation across consumers — see, for example, Thomas Stoker (1986). None of these potential biases to the models have been successfully considered in the existing monetary demand literature. Of course, dealing with these issues is not easy. The combination, for example, of nonstationary data and nonlinear estimation in fairly large models is an extremely difficult problem.

24.4 Risk Matters

In this book we haven't dealt with the extension of riskless models to situations where the economic agent makes decisions under uncertainty. In fact, in the theory of microeconomic quantity and price aggregation reviewed in this book, the theoretical existence of exact aggregates is proved through the use of nested two-stage budgeting and duality theory. Under risk, however, two-stage budgeting theorems do not work, and most duality theory does not apply.

Recently, however, Barnett (1995), and Barnett, Yi Liu, and Mark Jensen (1997) have extended aggregation theory to the case of risk and derived the 'generalized Divisia index.' This work is very interesting and innovative and we briefly review it here in the context of an optimal growth model, similar to those we dealt with in Chapters 4 and 5. Our objective is to show the process by which the generalized Divisia index may be derived, its connection to optimal growth theory and capital asset pricing models, and to also discuss a possible future research agenda that might flow out of this work.

Assume an infinitely lived economic agent having expected utility

$$E_0 \sum_{t=0}^{\infty} \beta^t u\left(c_t, \boldsymbol{x}_t\right), \tag{24.2}$$

where E_0 is the mathematical expectation operator conditional on all information at time 0, β is the discount factor, c is aggregate real consumption of goods and services, and $\boldsymbol{x}$ is an n-dimensional vector of real balances of monetary assets. It is assumed that the one-period utility function, $u(\cdot)$ is weakly separable in $\boldsymbol{x}$. That is, it must be possible to write it as $u = u\left(c_t, f\left(\boldsymbol{x}_t\right)\right)$, in which $f(\boldsymbol{x})$ defines the monetary subutility function.

The objective (24.2) is maximized with respect to $\{c_t, \boldsymbol{x}_t, A_t\}_{t=0}^{\infty}$, where A_t denotes holdings of the benchmark asset during period t, subject to a sequence of one period budget constraints

$$q_t c_t = \sum_{i=1}^{n} \left[\left(1 + r_{i,t-1}\right) p_{t-1}^{*} x_{i,t-1} - p_t^{*} x_{it}\right]$$

$$+ \left(1 + R_{t-1}\right) p_{t-1}^{*} A_{t-1} - p_t^{*} A_t + I_t, \tag{24.3}$$

for $t = 0, 1, \cdots$. In equation (24.3), q is the exact price aggregate that is dual to the consumer goods quantity aggregate c; r_i is the nominal holding period yield on assets i; R is the holding-period yield on the benchmark asset; p^* is the true cost of living index, $p^* = p^*(q)$; and I_t is the sum of all other sources of income during period t. To rule out perpetual borrowing on the part of the consumer, we impose the condition

$$\lim_{t \to \infty} d_t A_t = 0, \tag{24.4}$$

where the present value factor d_t is defined by $d_t = d_{t-1} / \left(1 + R_{t-1}\right)$, with $d_{t-1} = 1$.

Solving the difference equation (24.3) forward, using (24.4) as the terminal condition, we obtain the following intertemporal version of the budget constraint

$$\sum_{t=0}^{\infty} d_t q_t c_t = p_{t-1}^{*} A_{t-1} + \sum_{t=0}^{\infty} d_t I_t$$

$$+ \sum_{t=0}^{\infty} \sum_{i=1}^{n} d_t \left[\left(1 + r_{i,t-1}\right) p_{t-1}^{*} x_{i,t-1} - p_t^{*} x_{it}\right]. \tag{24.5}$$

The Euler equations for monetary assets and the consumer goods aggregate are

$$\frac{\partial u}{\partial x_{it}} = \beta E_t \frac{p_t^{*}\left(R_t - r_{it}\right)}{p_{t+1}^{*}} \frac{\partial u}{\partial c_{t+1}}, \quad i = 1, ..., n; \tag{24.6}$$

$$\frac{\partial u}{\partial c_t} = \beta E_t \frac{p_t^{*}\left(1 + R_t\right)}{p_{t+1}^{*}} \frac{\partial u}{\partial c_{t+1}}. \tag{24.7}$$

For notational convenience, we convert the nominal rates of return, R_t and r_{it}, into real rates, R_t^* and r_{it}^*, such that

$$(1 + R_t^*) = p_t^* \frac{(1 + R_t)}{p_{t+1}^*};$$

$$(1 + r_{it}^*) = p_t^* \frac{(1 + r_{it})}{p_{t+1}^*}.$$

Under this new notation (24.6) and (24.7) can be written as

$$\frac{\partial u}{\partial x_{it}} = \beta E_t \left[\left(R_t^* - r_{it}^* \right) \frac{\partial u}{\partial c_{t+1}} \right], \quad i = 1, ..., n; \tag{24.8}$$

$$\frac{\partial u}{\partial c_t} = \beta E_t \left[\left(1 + R_t^* \right) \frac{\partial u}{\partial c_{t+1}} \right]. \tag{24.9}$$

Under risk aversion, the marginal utility of consumption and the interest rates in the expectation on the right-hand side of each of (24.8) and (24.9) are correlated. Hence, these equations can be written as

$$\frac{\partial u}{\partial x_{it}} = \beta E_t \left[\frac{\partial u}{\partial c_{t+1}} \right] \left(E_t R_t^* - E_t r_{it}^* \right) + \beta \text{Cov} \left(R_t^*, \frac{\partial u}{\partial c_{t+1}} \right)$$

$$- \beta \text{Cov} \left(r_{it}^*, \frac{\partial u}{\partial c_{t+1}} \right), \quad \text{for all } i \tag{24.10}$$

$$\frac{\partial u}{\partial c_t} = \beta E_t \left[\frac{\partial u}{\partial c_{t+1}} \right] + \beta E_t \left[R_t^* \right] E_t \left[\frac{\partial u}{\partial c_{t+1}} \right]$$

$$+ \beta \text{Cov} \left(R_t^*, \frac{\partial u}{\partial c_{t+1}} \right). \tag{24.11}$$

Solving (24.11) for $\beta E_t \left[\partial u / \partial c_{t+1} \right]$ and substituting into (24.10) yields

$$\frac{\partial u}{\partial x_{it}} = \left(p_{it} + \psi_{it} \right) \frac{\partial u}{\partial c_t} \quad i = 1, \cdots, n, \tag{24.12}$$

where $(p_{it} + \psi_{it})$ is the risk-adjusted user cost of holding asset i, with

$$p_{it} = \frac{E_t R_t^* - E_t r_{it}^*}{1 + R_t^*},$$

or in nominal terms

$$p_{it} = \frac{E_t R_t - E_t r_{it}}{1 + R_t},$$

and

$$\psi_{it} = \beta \left(1 - p_{it}\right) \frac{\mathrm{Cov}\left(R_t^*, \dfrac{\partial u}{\partial c_{t+1}}\right)}{\dfrac{\partial u}{\partial c_t}} - \beta \frac{\mathrm{Cov}\left(r_{it}^*, \dfrac{\partial u}{\partial c_{t+1}}\right)}{\dfrac{\partial u}{\partial c_t}}.$$

To derive the Divisia index under risk aversion, let us return to the consumer's utility function as defined above, this was $u = u\left(c_t, f(\boldsymbol{x}_t)\right)$. Because of the weak separability assumption of $u(\cdot)$, we have

$$\frac{\partial u}{\partial x_{it}} = \frac{\partial u}{\partial f(\boldsymbol{x}_t)} \frac{\partial f(\boldsymbol{x}_t)}{\partial x_{it}}, \quad i = 1, \cdots, n,$$

which because of (24.12) becomes

$$\frac{\partial f(\boldsymbol{x}_t)}{\partial x_{it}} = \left(p_{it} + \psi_{it}\right) \frac{\partial u / \partial c_t}{\partial u / \partial f(\boldsymbol{x}_t)}. \tag{24.13}$$

Since the total differential of $f(\boldsymbol{x})$ is

$$df(\boldsymbol{x}_t) = \sum_{i=1}^{n} \left(\frac{\partial f(\boldsymbol{x}_t)}{\partial x_{it}}\right) dx_{it}, \tag{24.14}$$

substitution of (24.13) into (24.14) yields

$$df(\boldsymbol{x}_t) = \frac{\partial u / \partial c_t}{\partial u / \partial f(\boldsymbol{x}_t)} \sum_{i=1}^{n} \left(p_{it} + \psi_{it}\right) dx_{it}. \tag{24.15}$$

Also, since linear homogeneity of $f(\boldsymbol{x})$ implies

$$f(\boldsymbol{x}_t) = \sum_{i=1}^{n} \left(\frac{\partial f(\boldsymbol{x}_t)}{\partial x_{it}}\right) x_{it}, \tag{24.16}$$

substitution of (24.13) into (24.16) yields

$$f(\boldsymbol{x}_t) = \frac{\partial u / \partial c_t}{\partial u / \partial f(\boldsymbol{x}_t)} \sum_{i=1}^{n} \left(p_{it} + \psi_{it}\right) x_{it}. \tag{24.17}$$

Finally, dividing (24.15) by (24.17) and rearranging yields the *generalized Divisia index*

$$d \log f(\boldsymbol{x}_t) = \sum_{i=1}^{n} w_{it}^{*} d \log x_{it}, \tag{24.18}$$

where

$$w_{it}^{*} = \frac{(p_{it} + \psi_{it})\, x_{it}}{\sum_{k=1}^{n} (p_{kt} + \psi_{kt})\, x_{kt}},$$

is the risk-adjusted share of asset i. Of course, under risk neutrality, the marginal utility of consumption and the interest rates in the expectations on the right-hand sides of (24.10) and (24.11) are uncorrelated. Hence, ψ_{it} would be zero, for all i, and the generalized Divisia index (24.18) reduces to the Divisia index in the perfect-certainty (or risk neutrality) case that we studied in this book.

More recently, Barnett and Wu (2005) extend the monetary asset user cost risk adjustment of Barnett *et al.* (1997) to the case of multiple risky non-monetary assets and intertemporal non-separability. They show that for any individual monetary asset, the risk adjustment (to its certainty equivalent user cost) can be measured by its beta, which depends on the covariance between the rate of return on the monetary asset and the wealth portfolio of the consumers. This result is analogous to the standard Capital Asset Pricing Model (CAPM). These extensions are especially useful, when own rates of return are subject to exchange rate risk, as in Barnett (2007).

This extension of the theory of microeconomic quantity and price aggregation under perfect certainty to situations under risk is interesting and innovative. However, there is a need for the expected utility approach to be extended to a nonexpected utility approach. As Diewert (2000, p. xxvi) puts it, in his introduction to Barnett and Serletis (2000a),

> "starting with Allais (1953), various researchers, including for example, Machina (1982), Mehra and Prescott (1985), and Chew and Epstein (1989), have noted various paradoxes associated with the use of the expected utility approach. Using the state contingent commodity approach to choice under risk that was pioneered by Blackorby, Davidson, and Donaldson (1977), Diewert (1993) tried to show that the expected utility framework led to a relatively inflexible class of functional forms to model preferences over uncertain alternatives. Diewert showed that a much more flexible class of functional forms can be obtained by

> moving to nonexpected utility models that are counterparts to the choice over lotteries models of the type pioneered by Dekel (1986), Chew (1989), Epstein and Zin (1990, 1991), and Gul (1991). Epstein and Zin (1990), Epstein (1992), and Diewert (1993, 1995) showed that these more flexible models can explain many of the choice under uncertainty paradoxes, including the equity premium puzzle of Mehra and Prescott (1985)."

24.5 Conclusion

Until fairly recently most money demand studies were based on the use of simple-sum measures of money and log-linear demand for money functions. The subject was, as Barnett (1987, p. 1184) puts it,

> "the unstable non-demand function for a non-variable regressed on other non-variables through non-theory."

This book has focused attention on the gains that can be achieved by a rigorous use of microeconomic theory, index number theory, aggregation theory, and related econometric approaches to the study of the inter-related problems of monetary aggregation and estimation of money demand. Of course, many unsolved problems exist in this literature. The investigation of those problems, however, is likely to be significantly useful and productive.

References

1. Afriat, S., 1967. “The Construction of a Utility Function from Expenditure Data,” *International Economic Review* 8, 67-77.
2. Ahmad, S., 1977. “Transactions Demand for Money and the Quantity Theory,” *Quarterly Journal of Economics* 91, 327-335.
3. Allais, M., 1953. “Le comportement de l’homme rationnel devant le risque: critique de postulats et axiomes de l’école américaines,” *Econometrica* 21, 503-546.
4. Allen, F. and D. Gale, 1999. *Comparing Financial Systems,* Cambridge, MA: MIT Press.
5. Aliprantis, C. D., W. A. Barnett, B. Cornet, and S. Durlauf, 2007. “Special Issue Editors’ Introduction: The Interface Between Econometrics and Economic Theory,” *Journal of Econometrics* 136, 324-329.
6. Alston, J. M., K. A. Foster, and R. D. Green, 1994. “Estimating Elasticities with the Linear Approximate Ideal Demand System: Some Monte Carlo Results,” *Review of Economics and Statistics* 76, 351-356.
7. Anderson, G. J. and R. W. Blundell, 1982. “Estimation and Hypothesis Testing in Dynamic Singular Equation Systems,” *Econometrica* 50, 1559-1571.
8. Anderson, G. J. and R. W. Blundell, 1983. “Testing Restrictions in a Flexible Dynamic Demand System: An Application to Consumers’ Expenditure in Canada,” *Review of Economic Studies* 50, 397-410.
9. Anderson, R. G., 2006. “Replicability, Real-Time Data, and the Science of Economic Research: FRED, ALFRED, and VDC,” Federal Reserve Bank of St. Louis *Review* 88, 81-93.
10. Anderson, R. G. and J. Buol, 2005. “Revisions to User Costs for the Federal Reserve Bank of St. Louis Monetary Services Indices,” Federal Reserve Bank of St. Louis *Review* 87, 735-749.

11. Anderson, R. G. and J. G. Thursby, 1986. "Confidence Intervals for Elasticity Estimators in Translog Models," *Review of Economics and Statistics* 68, 647-656.

12. Anderson, R. G., B. E. Jones, and T. D. Nesmith, 1997a. "Monetary Aggregation Theory and Statistical Index Numbers," Federal Reserve Bank of St. Louis *Review* 79, 31-51.

13. Anderson, R. G., B. E. Jones, and T. D. Nesmith, 1997b. "Building New Monetary Services Indexes: Concepts, Data and Methods," The Federal Reserve Bank of St. Louis *Review* 79, 53-82.

14. Anderson, R. W., 1979. "Perfect Price Aggregation and Empirical Demand Analysis," *Econometrica* 47, 1209-1230.

15. Andrews, D. W. K., 1991. "Heteroskedasticity and Autocorrelation Consistent Covariance Matrix Estimation," *Econometrica* 59, 817-858.

16. Attfield, C. L. F., 1997. "Estimating a Cointegrating Demand System," *European Economic Review* 41, 61-73.

17. Azariadis, C., 1993. *Intertemporal Macroeconomics*, Oxford: Blackwell.

18. Backus, D. K. and P. J. Kehoe, 1992. "International Evidence on the Historical Properties of Business Cycles," *American Economic Review* 82, 864-888.

19. Bailey, M. J., 1956. "The Welfare Cost of Inflationary Finance," *Journal of Political Economy* 64, 93-110.

20. Bak, P., C. Tang, and K. Wiesenfeld, 1987. "Self-Organized Criticality: An Explanation of $1/f$Noise," *Physical Review Letters* 59, 381-384.

21. Bak, P., C. Tang, and K. Wiesenfeld, 1988. "Self-Organized Criticality," *Physical Review A* 38, 364-374.

22. Banerjee, A., R. L. Lumsdaine, and J. H. Stock, 1992. "Recursive and Sequential Tests of the Unit-Root and Trend-Break Hypotheses: Theory and International Evidence," *Journal of Business and Economic Statistics* 10, 271-287.

23. Banks, J., R. Blundell, and A. Lewbel, 1997. "Quadratic Engel Curves and Consumer Demand," *Review of Economics and Statistics* 79, 527-539.

24. Barnett, W. A., 1977. "Recursive Subaggregation and a Generalized Hypocycloidal Demand Model," *Econometrica* 45, 1117-1136.

25. Barnett, W. A., 1978. "The User Cost of Money," *Economics Letters* 1, 145-14.

26. Barnett, W. A., 1980. "Economic Monetary Aggregates: An Application of Aggregation and Index Number Theory," *Journal of Econometrics* 14, 11-48.

27. Barnett, W. A., 1982. "The Optimal Level of Monetary Aggregation," *Journal of Money, Credit, and Banking* 14, 687-710.

28. Barnett, W. A., 1983. "New Indices of Money Supply and the Flexible Laurent Demand System," *Journal of Business and Economic Statistics* 1, 7-23.

29. Barnett, W. A., 1984. "Recent Monetary Policy and the Divisia Monetary Aggregates," *American Statistician* 38, 165-172.

30. Barnett, W. A., 1985. "The Minflex Laurent Translog Flexible Functional Form," *Journal of Econometrics* 30, 33-44.

31. Barnett, W. A., 1995. "Exact Aggregation Under Risk." In *Social Choice, Welfare, and Ethics,* W. A. Barnett, H. Moulin, M. Salles, and N. Schofield (eds.), Cambridge: Cambridge University Press, 353-374.

32. Barnett, W. A., 1997. "Which Road Leads to a Stable Money Demand?" *The Economic Journal* 107, 1171-1185.

33. Barnett, W. A., 2002. "Tastes and Technology: Curvature is not Sufficient for Regularity," *Journal of Econometrics* 108, 199-202.

34. Barnett, W. A., 2007. "Multilateral Aggregation-Theoretic Monetary Aggregation Over Heterogeneous Countries," *Journal of Econometrics* 136, 457-482.

35. Barnett, W. A. and P. Chen, 1988. "The Aggregation Theoretic Monetary Aggregates are Chaotic and Have Strange Attractors: An Economic Application of Mathematical Chaos." In *Dynamic Econometric Modeling*, W. A. Barnett, E. Berndt, and H. White (eds.), Cambridge: Cambridge University Press, 199-245.

36. Barnett, W. A. and S. Choi, 1989. "A Monte Carlo Study of Tests of Blockwise Weak Separability," *Journal of Business and Economic Statistics* 7, 363-377.

37. Barnett, W. A. and A. Jonas, 1983. "The Muntz-Szatz Demand System: An Application of a Globally Well Behaved Series Expansion," *Economics Letters* 11, 337-342.

38. Barnett, W. A. and Y. W. Lee, 1985. "The Global Properties of the Minflex Laurent, Generalized Leontief, and Translog Flexible Functional Forms," *Econometrica* 53, 1421-1437.

39. Barnett, W. A. and M. Pasupathy, 2003. "Regularity of the Generalized Quadratic Production Model: A Counterexample," *Econometric Reviews* 22, 135-154.

40. Barnett, W. A. and A. Serletis, 1990. “A Dispersion-Dependency Diagnostic Test for Aggregation Error: With Applications to Monetary Economics and Income Distribution,” *Journal of Econometrics* 43, 5-34.

41. Barnett, W. A. and A. Serletis, 2000a. *The Theory of Monetary Aggregation*, Amsterdam: North-Holland.

42. Barnett, W. A. and A. Serletis, 2000b. “Martingales, Nonlinearity, and Chaos,” *Journal of Economic Dynamics and Control* 24, 703-724.

43. Barnett, W. A. and S. Wu, 2005. “On User Costs of Risky Monetary Assets,” *Annals of Finance* 1, 35-50.

44. Barnett, W. A. and P. Yue, 1988. “Seminonparametric Estimation of the Asymptotically Ideal Model: The AIM Demand System.” In *Nonparametric and Robust Inference: Advances in Econometrics,* Vol. 7, G. F. Rhodes and T. B. Fomby (eds.), Greenwich, CT : JAI Press, 229-251.

45. Barnett, W. A., D. Fisher, and A. Serletis, 1992. “Consumer Theory and the Demand for Money,” *Journal of Economic Literature* 30, 2086-2119.

46. Barnett, W. A., A. J. Geweke, and M. Wolfe, 1991. “Seminonparametric Bayesian Estimation of the Asymptotically Ideal Production Model,” *Journal of Econometrics* 49, 5-50.

47. Barnett, W. A., M. J. Hinich, and P. Yue, 2000. “The Exact Theoretical Rational Expectations Monetary Aggregates,” *Macroeconomic Dynamics* 4, 197-221.

48. Barnett, W. A., Y. W. Lee, and M. D. Wolfe, 1985. “The Three-Dimensional Global Properties of the Minflex Laurent, Generalized Leontief, and Translog Flexible Functional Forms,” *Journal of Econometrics* 30, 3-31.

49. Barnett, W. A., Y. W. Lee, and M. D. Wolfe, 1987. “The Global Properties of the Two Minflex Laurent Flexible Functional Forms,” *Journal of Econometrics* 36, 281-298.

50. Barnett, W. A., Y. Liu, and M. J. Jensen, 1997. “CAPM Risk Adjustment for Exact Aggregation over Financial Assets,” *Macroeconomic Dynamics* 1, 485-512.

51. Barnett, W. A., E. K. Offenbacher, and P. A. Spindt, 1984. “The New Divisia Monetary Aggregates,” *Journal of Political Economy* 92, 1049-1085.

52. Barro, R. J., 1976. “Rational Expectations and the Role of Monetary Policy,” *Journal of Monetary Economics* 2, 1-32.

53. Barro, R. J., 1977. “Unanticipated Money Growth and Unemployment in the United States,” *American Economic Review* 67, 101-115.

54. Barro, R. J., 1978. "Unanticipated Money, Output, and the Price Level in the United States," *Journal of Political Economy* 86, 548-580.

55. Barro, R. J., 1997. *Macroeconomics* (5th edition), Cambridge: The MIT Press.

56. Barro, R. J. and D. B. Gordon, 1983. "A Positive Theory of Monetary Policy in a Natural-Rate Model," *Journal of Political Economy* 91, 589-610.

57. Barro, R. J. and H. I. Grossman, 1976. *Money, Employment, and Inflation*, Cabmridge: Cambridge University Press.

58. Barro, R. J. and X. Sala-i-Martin, 2004. *Economic Growth* (2nd edition), Cambridge, Massachusetts: The MIT Press.

59. Barten, A. P., 1964. "Consumer Demand Functions under Conditions of Almost Additive Preferences," *Econometrica* 32, 1-38.

60. Barten, A. P., 1967. "Evidence on the Slutsky Conditions for Demand Equations," *Review of Economics and Statistics* 49, 77-84.

61. Barten, A. P., 1969. "Maximum Likelihood Estimation of a Complete System of Demand Equations," *European Economic Review* 1, 7-73.

62. Baumol, W. J., 1952. "The Transactions Demand for Cash: An Inventory Theoretic Approach," *Quarterly Journal of Economics* 66, 545-556.

63. Baxter, M. and R. G. King, 1999. "Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series," *Review of Economics and Statistics* 81, 573-593.

64. Beck, T. and R. Levine, 2004. "Stock Markets, Banks and Growth: Panel Evidence," *Journal of Banking and Finance* 28, 423-442.

65. Belongia, M. T., 1996. "Measurement Matters: Recent Results from Monetary Economics Reexamined," *Journal of Political Economy* 104, 1065-1083.

66. Berndt, E. R. and N. E. Savin, 1975. "Estimation and Hypothesis Testing in Singular Equation Systems with Autoregressive Disturbances," *Econometrica* 43, 937-957.

67. Bhide, A., 1993. "The Hidden Costs of Stock Market Liquidity," *Journal of Financial Intermediation* 34, 1-51.

68. Blackorby, C. and R. R. Russell, 1981. "The Morishima Elasticity of Substitution, Symmetry, Constancy, Separability, and its Relationship to the Hicks and Allen Elasticities," *Review of Economic Studies* 48, 147-158.

69. Blackorby, C. and R. R. Russell, 1989. "Will the Real Elasticity of Substitution Please Stand Up?" *American Economic Review* 79, 882-888.

70. Blackorby, C., R. Davidson, and D. Donaldson, 1977. "A Homiletic Exposition of the Expected Utility Hypothesis," *Economica* 44, 351-358.

71. Blackorby, C., D. Primont, and R. R. Russell, 1978. *Duality, Separability, and Functional Structure*, Amsterdam: North-Holland.

72. Blackorby, C., D. Nissen, D. Primont, and R. R. Russell, 1974. "Recursively Decentralized Decision Making," *Econometrica* 42, 487-496.

73. Blanchard, O. J. and S. Fischer, 1989. *Lectures in Macroeconomics*, Cambridge, MA: The MIT Press.

74. Boot, A. and A. V. Thakor, 1997. "Financial System Architecture," *Review of Financial Studies* 10, 693-733.

75. Brock, W. A., 1974. "Money and Growth: The Case of Long-Run Perfect Foresight," *International Economic Review* 15, 750-777.

76. Brock, W. A. and C. Sayers, 1988. "Is the Business Cycle Characterized by Deterministic Chaos?" *Journal of Monetary Economics* 22, 71-90.

77. Brunner, K. and A. H. Meltzer, 1967. "Economies of Scale in Cash Balances Reconsidered," *Quarterly Journal of Economics* 81, 422-436.

78. Buse, A., 1994. "Evaluating the Linearized Almost Ideal Demand System," *American Journal of Agricultural Economics* 76, 781-793.

79. Bullard, J. B., 1999. "Testing Long-Run Monetary Neutrality Propositions: Lessons from the Recent Research," Federal Reserve Bank of St. Louis *Review* 81, 57-77.

80. Bullard, J. B. and S. Russell, 2004. "How Costly Is Sustained Low Inflation for the U.S. Economy," The Federal Reserve Bank of St. Louis *Review* 86, 35-67.

81. Burguete, J. F., A. R. Gallant, and G. Souza, 1982. "On Unification of the Asymptotic Theory of Nonlinear Econometric Models," *Econometric Reviews* 1, 151-190.

82. Burns, A. F. and W. C. Mitchell, 1946. *Measuring Business Cycles*, New York: National Bureau of Economic Research.

83. Cagan, P., 1956. "The Monetary Dynamics of Hyperinflation." In *Studies in the Quantity Theory of Money*, M. Friedman (ed.), Chicago: University of Chicago Press.

84. Campbell, J. Y. and P. Perron, 1991. "Pitfalls and Opportunities: What Macroeconomists Should Know about Unit Roots," *NBER Macroeconomics Annual,* 141-200.

85. Campbell, J. Y., A. W. Lo, and A. C. MacKinlay, 1997. *The Econometrics of Financial Markets*, Princeton, NJ: Princeton University Press.

86. Cass, D., 1965. "Optimum Growth in an Aggregative Model of Capital Accumulation," *Review of Economic Studies* 32, 233-240.

87. Caves, D. W. and L. R. Christensen, 1980. "Global Properties of Flexible Functional Forms," *American Economic Review* 70, 422-432.

88. Chalfant, J. A., 1987. "A Globally Flexible Almost Ideal System," *Journal of Business and Economic Statistics* 5, 233-242.

89. Chalfant, J. A. and A. R. Gallant, 1985. "Estimating Substitution Elasticities with the Fourier Cost Function: Some Monte Carlo Results," *Journal of Econometrics* 28, 205-222.

90. Champ, B. and S. Freeman, 1994. *Modeling Monetary Economies*, New York: John Wiley & Sons, Inc.

91. Chetty, K. V., 1969. "On Measuring the Nearness of Near-Moneys," *American Economic Review* 59, 270-281.

92. Chew, S. H., 1989. "Axiomatic Utility Theories with the Betweenness Property," *Annals of Operations Research* 19, 273-298.

93. Chew, S. H. and L. G. Epstein, 1989. "A Unifying Approach to Axiomatic Non-Expected Utility Theories," *Journal of Economic Theory* 49, 207-240.

94. Chow, G. C., 1960. "Tests of Equality Between Sets of Coefficients in Two Linear Regressions," *Econometrica* 28, 591-605.

95. Chow, G. C., 1966. "On the Long-Run and Short-Run Demand for Money," *Journal of Political Economy* 74, 111-131.

96. Christensen, L. R., D. W. Jorgenson, and L. J. Lau, 1975. "Transcendental Logarithmic Utility Functions," *American Economic Review* 65, 367-383.

97. Christiano, L. J., 1986. "Money and the U.S. Economy in the 1980s: A Break from the Past?" Federal Reserve Bank of Minneapolis *Quarterly Review* 10, 2-13.

98. Christiano, L. J., 1992. "Searching for a Break in GNP," *Journal of Business and Economic Statistics* 10, 237-250.

99. Christiano, L. J. and T. Fitzgerald, 2003. "The Band Pass Filter," *International Economic Review* 44, 435-465.

100. Clower, R. W., 1967. "A Reconsideration of the Microeconomic Foundations of Monetary Theory," *Western Economic Journal* 6, 1-8.

101. Cogley, T. and J. M. Nason, 1995. "Effects of the Hodrick-Prescott Filter on Trend and Difference Stationary Time Series: Implications for Business Cycle Research," *Journal of Economic Dynamics and Control* 19, 253-278.

102. Collins, S. and C. L. Edwards, 1994. "An Alternative Monetary Aggregate: M2 Plus Household Holdings of Bond and Equity Mutual Funds," Federal Reserve Bank of St. Louis *Review* 76, 7-29.

103. Cooley, T. F. and G. D. Hansen, 1989. "The Inflation Tax in a Real Business Cycle Model," *American Economic Review* 79, 733-748.

104. Cooley, T. F. and L. E. Ohanian, 1991. "The Cyclical Behavior of Prices," *Journal of Monetary Economics* 28, 25-60.

105. Cooper, R. J. and K. R. McLaren, 1996. "A System of Demand Equations Satisfying Effectively Global Regularity Conditions," *Review of Economics and Statistics* 78, 359-364.

106. Davidson, R. and J. G. Mackinnon, 1993. *Estimation and Inference in Econometrics,* Oxford: Oxford University Press.

107. Deaton, A. and J. N. Muellbauer, 1980. "An Almost Ideal Demand System," *American Economic Review* 70, 312-326.

108. Dekel, E., 1986. "An Axiomatic Characterization of Preferences Under Uncertainty: Weakening the Independence Axiom," *Journal of Economic Theory* 40, 304-314.

109. de Peretti, P., 2005. "Testing the Significance of the Departures from Utility Maximization," *Macroeconomic Dynamics* 9, 372-397.

110. Diamond, P. A., 1965. "National Debt in a Neoclassical Growth Model," *American Economic Review* 55, 1126-1150.

111. Dickey, D. A. and W. A. Fuller, 1979. "Distribution of the Estimates for Autoregressive Time Series with a Unit Root," *Journal of the American Statistical Association* 74, 427-431.

112. Dickey, D. A. and W. A. Fuller, 1981. "Likelihood Ratio Tests for Autoregressive Time Series with a Unit Root," *Econometrica* 49, 1057-1072.

113. Diebold, F. X. and M. Nerlove, 1990. "Unit Roots in Economic Time Series: A Selective Survey." In *Advances in Econometrics: Co-Integration, Spurious Regressions, and Unit Roots*, Vol. 8, T. B. Fomby and G. F. Rhodes Jr. (eds.), Greenwich, CT: JAI Press, 3-69.

114. Diewert, W. E., 1971. "An Application of the Shephard Duality Theorem: A Generalized Leontief Production Function," *Journal of Political Economy* 79, 481-507.

115. Diewert, W. E., 1973a. "Functional Forms for Profit and Transformation Functions," *Journal of Economic Theory* 6, 284–316.

116. Diewert, W. E., 1973b. "Afriat and Revealed Preference Theory," *Review of Economic Studies* 40, 419-426.

117. Diewert, W. E., 1974. "Applications of Duality Theory." In *Frontiers in Quantitive Economics* Vol 2, M. Intriligator and D. Kendrick (eds.). Amsterdam: North-Holland, p. 106-171.

118. Diewert, W. E., 1976. "Exact and Superlative Index Numbers," *Journal of Econometrics* 4, 115-145.

119. Diewert, W. E., 1993. "Symmetric Means and Choice Under Uncertainty." In *Essays in Index Number Theory,* Vol. 1, W. E. Diewert and A. O. Nakamura (eds.), Amsterdam: North-Holland, p. 355-433.

120. Diewert, W. E., 1995. "Functional Form Problems in Modeling Insurance and Gambling," *Geneva Papers on Risk and Insurance Theory* 20, 135-150.

121. Diewert, W. E., 2000. "Preface to *The Theory of Monetary Aggregation*." In *Theory of Monetary Aggregation*, W. A. Barnett and A. Serletis (eds.), Amsterdam: North-Holland, p. xxiii-xxviii.

122. Diewert, W. E. and C. Parkan, 1978. "Test for Consistency of Consumer Data and Nonparametric Index Numbers," Discussion Paper #78-27. Department of Economics, University of British Columbia.

123. Diewert, W. E. and T. J. Wales, 1987. "Flexible Functional Forms and Global Curvature Conditions," *Econometrica* 55, 43-68.

124. Diewert, W. E. and T. J. Wales, 1988. "Normalized Quadratic Systems of Consumer Demand Functions," *Journal of Business and Economic Statistics* 6, 303-312.

125. Diewert, W. E., M. Avriel, and I. Zang, 1977. "Nine Kinds of Quasiconcavity and Concavity," *Journal of Economic Theory* 25, 397-420.

126. Donovan, D. J., 1978. "Modeling the Demand for Liquid Assets: An Application to Canada," International Monetary Fund *Staff Papers* 25, 676-704.

127. Dowd, K., 1990. "The Value of Time and the Transactions Demand for Money," *Journal Money, Credit, and Banking* 22, 51-64.

128. Drake, L. and A. R. Fleissig, 2004. "Semi-Nonparametric Estimates of Currency Substitution: The Demand for Sterling in Europe," *Review of International Economics* 12, 374-394.

129. Drake, L., A. R. Fleissig, and A. Mullineaux, 1999. "Are 'Risky' Assets Substitutes for 'Monetary Assets'? Evidence from an AIM Demand System," *Economic Inquiry* 37, 510-526.

130. Drake, L., A. R. Fleissig, and J. L. Swofford, 2003. "A Seminonparametric Approach to the Demand for U.K. Monetary Assets," *Economica* 70, 99-120.

131. Eastwood, B. J. and A. R. Gallant, 1991. "Adaptive Rules for Seminonparametric Estimators that Achieve Asymptotic Normality," *Econometric Theory* 7, 307-340.

132. Eckmann, J. P. and D. Ruelle, 1985. "Ergodic Theory of Chaos and Strange Attractors," *Reviews of Modern Physics* 57, 617-656.

133. Eckmann, J. P., S. O. Kamphorst, D. Ruelle, and S. Ciliberto, 1986. "Lyapunov Exponents from Time Series," *Physical Review A* 34, 4971-4979.

134. Eichhorn, W., 1976. "Fisher's Tests Revisited," *Econometrica* 44, 247-256.

135. Eichhorn, W., 1978. *Functional Equations in Economics*, Reading, MA: Addison-Wesley.

136. Enders, W., 2004. *Applied Econometric Time Series* (2nd edition), New York: John Wiley & Sons, Inc.

137. Engle, R. F., 1982. "Autoregressive Conditional Heteroskedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica* 50, 987-1008.

138. Engle, R. F. and C. W. Granger, 1987. "Cointegration and Error Correction: Representation, Estimation and Testing," *Econometrica* 55, 251-276.

139. Engle, R. F. and B. S. Yoo, 1987. "Forecasting and Testing in Cointegrated Systems," *Journal of Econometrics* 35, 143-159.

140. Epstein, L. G., 1992. "Behavior Under Risk: Recent Developments in Theory and Applications." In *Advances in Economic Theory*, J. J. Laffont (ed.), New York: Cambridge University Press.

141. Epstein, L. G. and S. E. Zin, 1990. "First Order Risk Aversion and the Equity Premium Puzzle," *Journal of Monetary Economics* 26, 387-407.

142. Epstein, L. G. and S. E. Zin, 1991. "The Independence Axiom and Asset Returns," NBER Working Paper #109. Cambridge, Mass.

143. Ewiss, N. A. and D. Fisher, 1984. "The Translog Utility Function and the Demand for Money in the United States," *Journal of Money, Credit and Banking* 16, 34-52.

144. Ewiss, N. A. and D. Fisher, 1985. "Toward a Consistent Estimate of the Substitutability between Money and Near Monies: An Application of the Fourier Flexible Form," *Journal of Macroeconomics* 7, 151-174.

145. Feenstra, R. C., 1986. "Functional Equivalence Between Liquidity Costs and the Utility of Money," *Journal of Monetary Economics* 17, 271-291.

146. Feige, E. L. and D. K. Pearce, 1977. "The Substitutability of Money and Near-Monies: A Survey of the Time-Series Evidence," *Journal of Economic Literature* 15, 439-469.

147. Fiorito, R. and T. Kollintzas, 1994. "Stylized Facts of Business Cycles in the G7 from a Real Business Cycles Perspective," *European Economic Review* 38, 235-269.

148. Fischer, A., 2005. "Measuring Income Elasticity for Swiss Money Demand: What Do the Cantons Say About Financial Innovation?" Working Paper 05.01, Swiss National Bank.

149. Fischer, S. and Summers, L. H., 1989. "Should Governments Learn to Live with Inflation?" *American Economic Review* 79, 382-388.

150. Fisher, D., 1989. *Money Demand and Monetary Policy*, Ann Arbor: The University of Michigan Press.

151. Fisher, D. and A. R. Fleissig, 1994. "Money Demand in a Flexible Dynamic Fourier Expenditure System," Federal Reserve Bank of St. Louis *Review* 76, 117-128.

152. Fisher, D. and A. R. Fleissig, 1997. "Monetary Aggregation and the Demand for Assets," *Journal of Money, Credit and Banking* 29, 458-475.

153. Fisher, D. and A. Serletis, 1989. "Velocity and the Growth of Money in the United States, 1970-1985," *Journal of Macroeconomics* 11, 323-332.

154. Fisher, D., A. R. Fleissig, and A. Serletis, 2001. "An Empirical Comparison of Flexible Demand System Functional Forms," *Journal of Applied Econometrics* 16, 59-80.

155. Fisher, M. and J. J. Seater, 1993. "Long-Run Neutrality and Superneutrality in an ARIMA Framework," *American Economic Review* 83, 402-415.

156. Fisher, I., 1911. *The Purchasing Power of Money*, London: Macmillan.

157. Fisher, I., 1922. *The Making of Index Numbers: A Study of Their Varieties, Tests, and Reliability*, Boston: Houghton Mifflin.

158. Fleissig, A. R., 1997. "The Dynamic Laurent Flexible Form and Long-Run Analysis," *Journal of Applied Econometrics* 12, 687-699.

159. Fleissig, A. R. and A. Serletis, 2002. "Seminonparametric Estimates of Substitution for Canadian Monetary Assets," *Canadian Journal of Economics* 35, 78-91.

160. Fleissig, A. R. and J. L. Swofford, 1996. "A Dynamic Asymptotically Ideal Model of Money Demand," *Journal of Monetary Economics* 37, 371-380.

161. Fleissig, A. R. and J. L. Swofford, 1997. "Dynamic Asymptotically Ideal Models and Finite Approximation," *Journal of Business and Economic Statistics* 15, 482-492.

162. Fleissig, A. R. and G. A. Whitney, 2003. "A New PC-Based Test for Varian's Weak Separability Conditions," *Journal of Business and Economic Statistics* 21, 133-144.

163. Fleissig, A. R. and G. A. Whitney, 2005. "Testing for the Significance of Violations of Afriat's Inequalities." *Journal of Business and Economic Statistics* 23, 355-362.

164. Fleissig, A. R., A. R. Hall, and J. J. Seater, 2000. "GARP, Separability, and the Representative Agent," *Macroeconomic Dynamics* 4, 324-342.

165. Fleissig, A. R., T. Kastens, and D. Terrell, 1997. "Semi-Nonparametric Estimates of Substitution Elasticities," *Economics Letters* 54, 209-215.

166. Friedman, B. M. and K. N. Kuttner, 1992. "Money, Income, Prices, and Interest Rates," *American Economic Review* 82, 472-492.

167. Friedman, M., 1956. "The Quantity Theory of Money: A Restatement." In *Studies in the Quantity Theory of Money*, M. Friedman (ed.), Chicago: University of Chicago Press.

168. Friedman, M., 1969. *The Optimum Quantity of Money and Other Essays.* Chicago: University of Chicago Press.

169. Friedman, M. and A. J. Schwartz, 1970. *Monetary Statistics of the United States: Estimates, Sources, Methods, and Data*, New York: Columbia University Press (for the NBER).

170. Frisch, R., 1959. "A Complete Scheme for Computing All Direct and Cross Demand Elasticities in a Model with Many Sectors," *Econometrica* 27, 177-196.

171. Fujiki, H. and C. B. Mulligan, 1996. "A Structural Analysis of Money Demand: Cross-Sectional Evidence from Japan," *Monetary and Economic Studies* 14, 53-78.

172. Fuller, W. A., 1976. *Introduction to Statistical Time Series*, New York: John Wiley & Sons, Inc.

173. Gallant, A. R., 1981. "On the Bias in Flexible Functional Forms and an Essentially Unbiased Form: The Fourier Flexible Form," *Journal of Econometrics* 15, 211-245.

174. Gallant, A. R., 1982. "Unbiased Determination of Production Technologies," *Journal of Econometrics* 20, 285-323.

175. Gallant, A. R. and G. H. Golub, 1984. "Imposing Curvature Restrictions on Flexible Functional Forms," *Journal of Econometrics* 26, 295-321.

176. Gallant, A. R. and H. White, 1992. "On Learning the Derivatives of an Unknown Mapping with Multilayer Feedforward Networks," *Neural Networks* 5, 129-138.

177. Gencay, R. and W. D. Dechert, 1992. "An Algorithm for the n-Lyapunov Exponents of an n-Dimensional Unknown Dynamical System," *Physica D* 59, 142-157.

178. Goldfeld, S. M., 1973. "The Demand for Money Revisited," *Brookings Papers on Economic Activity* 3, 577-638.

179. Goldfeld, S. M., 1976. "The Case of the Missing Money," *Brookings Papers on Economic Activity* 3, 683-730.

180. Goldfeld, S. M. and D. Sichel, 1990. "The Demand for Money." In *Handbook of Monetary Economics*, Vol. 1, B. M. Friedman and F. H. Hahn (eds.), Amsterdam: North-Holland, 299-356.

181. Gonzalo, J., 1994. "Five Alternative Methods of Estimating Long-Run Equilibrium Relationships," *Journal of Econometrics* 60, 203-233.

182. Gordon, R. J., 1984. "The Short-Run Demand for Money: A Reconsideration," *Journal of Money, Credit, and Banking* 16, 403-434.

183. Gorman, W. M., 1959. "Separable Utility and Aggregation," *Econometrica* 27, 469-481.

184. Gorman, W. M., 1970. "Tricks with Utility Functions." In *Essays in Economic Analysis*, M. J. Artis and A. R. Nobay (eds.), Cambridge: Cambridge University Press, 211-243.

185. Granger, C. W. J., 1995. "Modelling Nonlinear Relationships Between Extended-Memory Variables," *Econometrica* 63, 265-279.

186. Granger, C. W. J. and R. Joyeaux, 1980. "An Introduction to Long-Memory Time Series Models and Fractional Differencing," *The Journal of Time Series Analysis* 1, 15-29.

187. Green, H. A. J., 1964. *Aggregation in Economic Analysis: An Introductory Survey*, Princeton, N.J.: Princeton University Press.

188. Green, R., D. Rocke, and W. Hahn, 1987. "Standard Errors for Elasticities: A Comparison of Bootstrap and Asymptotic Standard Errors," *Journal of Business and Economic Statistics* 5, 145-149.

189. Guilkey, D. K. and C. A. K. Lovell, 1980. "On the Flexibility of the Translog Approximation," *International Economic Review* 21, 137-147.

190. Guilkey, D. K., C. A. K. Lovell, and R. C. Sickles, 1983. "A Comparison of the Performance of Three Flexible Functional Forms," *International Economic Review* 24, 591-616.

191. Gul, F., 1991. "A Theory of Disappointment Aversion," *Econometrica* 59, 667-686.

192. Jones, B., and P. de Peretti, 2005. "A Comparison of Two Methods for Testing the Utility Maximization Hypothesis when Quantity Data is Measured with Error," *Macroeconomic Dynamics* 9, 612-629.

193. Jones, B., D. Dutkowsky, and T. Elger, 2005. "Sweep Programs and Optimal Monetary Aggregation," *Journal of Banking and Finance* 29, 483-508.

194. Hendry, D. and G. E. Mizon, 1978. "Serial Correlation as a Convenient Simplification, not a Nuisance: A Comment of a Study of the Demand for Money by the Bank of England," *Economic Journal* 88, 549-563.

195. Hicks, J. R., 1937. "Mr. Keynes and the 'Classics'," *Econometrica* 5, 147-159.

196. Hodrick, R. J. and E. C. Prescott, 1980. "Post War U.S. Business Cycles: An Empirical Investigation," Mimeo. Carnegie-Mellon University, Pittsburgh, PA.

197. Hoffman, D. L. and R. H. Rasche, 1996. *Aggregate Money Demand Functions*, Boston: Kluwer Academic.

198. Hosking, J.R., 1981. "Fractional Differencing," *Biometrika* 68, 165-176.

199. Huber, P. J., 1981. *Robust Statistics*, New York: Wiley.

200. Jarque, C. M. and A. K. Bera, 1980. "Efficient Tests for Normality, Homoscedasticity, and Serial Independence of Regression Residuals," *Economics Letters* 6, 255-259.

201. Johansen, S., 1988. "Statistical Analysis of Cointegrated Vectors," *Journal of Economic Dynamics and Control* 12, 231-254.

202. Johansen, S., 1991. “The Power Function for the Likelihood Ratio Test for Cointegration.” In *Econometric Decision Models: New Methods of Modelling and Applications*, J. Gruber (ed.), New York: Springer Verlag, 323-335.

203. Johansen, S. and K. Juselius, 1992. “Testing Structural Hypotheses in a Multivariate Cointegration Analysis of the PPP and UIP for the UK,” *Journal of Econometrics* 53, 211-244.

204. Jones, B. and P. de Peretti, 2005. “A Comparison of Two Methods for Testing the Utility Maximization Hypothesis when Quantity Data is Measured with Error.” *Macroeconomic Dynamics* 9, 612-629.

205. Jones, B., D. Dutkowsky, and T. Elger, 2005. “Sweep Programs and Optimal Monetary Aggregation.” *Journal of Banking and Finance* 29, 483-508.

206. Judd, J. P. and J. L. Scadding, 1982. “The Search for a Stable Money Demand Function,” *Journal of Economic Literature* 20, 993-1023.

207. Kenny, L. W., 1991. “Cross-Country Estimates of the Demand for Money and its Components,” *Economic Inquiry* 29, 696-705.

208. Keynes, J. M., 1936. *The General Theory of Employment, Interest, and Money*, London and New York: Macmillan.

209. King, R. G. and S. T. Rebelo, 1993. “Low Frequency Filtering and Real Business Cycles,” *Journal of Economic Dynamics and Control* 17, 207-232.

210. King, R. G. and M. W. Watson, 1997. “Testing Long-Run Neutrality,” Federal Reserve Bank of Richmond *Economic Quarterly* 83, 69-101.

211. King, R. G., C. I. Plosser, J. H. Stock, and M. W. Watson, 1991. “Stochastic Trends and Economic Fluctuations,” *American Economic Review* 81, 819-840.

212. King, R. G., C. I. Plosser, and S. T. Rebelo, 1988. “Production, Growth and Business Cycles: I. The Basic Neoclassical Model,” *Journal of Monetary Economics* 21, 195-232.

213. Klein, B., 1974. “Competitive Interest Payments on Bank Deposits and the Long-Run Demand for Money,” *American Economic Review* 64, 931-949.

214. Koopmans, T. C., 1947. “Measurement Without Theory,” *Review of Economics and Statistics* 29, 161-172.

215. Koopmans, T. C., 1965. “On the Concept of Optimal Economic Growth.” In *The Econometric Approach to Development Planning*, Chicago: Rand-McNally.

216. Koustas, Z. and A. Serletis, 1999. “On the Fisher Effect,” *Journal of Monetary Economics* 44, 105-130.

217. Kunst, R. and K. Neusser, 1990. “Cointegration in a Macroeconomic System,” *Journal of Applied Econometrics* 5, 351-365.

218. Kydland, F. E. and E. C. Prescott, 1977. “Rules Rather Than Discretion: The Inconsistency of Optimal Plans,” *Journal of Political Economy* 85, 473-491.

219. Kydland, F. E. and E. C. Prescott, 1990. “Business Cycles: Real Facts and a Monetary Myth,” Federal Reserve Bank of Minneapolis *Quarterly Review*, 3-18.

220. Kyrtsou, C. and A. Serletis, 2006. “Univariate Tests for Nonlinear Structure,” *Journal of Macroeconomics* 28, 154-168.

221. Kwiatkowski, D., P. C. B. Phillips, P. Schmidt, and Y. Shin, 1992. “Testing the Null Hypothesis of Stationarity Against the Alternative of a Unit Root,” *Journal of Econometrics* 54, 159-178.

222. Laidler, D. E. W., 1993. *The Demand for Money: Theories, Evidence, and Problems*, 4th ed., New York: HarperCollins College Publishers.

223. Lau, L. J., 1970. “Duality and the Structure of Utility Functions,” *Journal of Economic Theory* 1, 374-396.

224. Lau, L. J., 1978a. “On Exact Index Numbers,” *Review of Economics and Statistics* 61, 73-82.

225. Lau, L. J., 1978b. “Testing and Imposing Monotonicity, Convexity, and Quasi-Convexity Constraints.” In *Production Economics: A Dual Approach to Theory and Applications* (Vol. 1) M. Fuss and D. McFadden (eds.), Amsterdam: North Holland, 409-453.

226. Leontief, W. W., 1947. “An Introduction to a Theory of the Internal Structure of Functional Relationships,” *Econometrica* 15, 361-373.

227. Levine, R., 1997. “Financial Development and Economic Growth: Views and Agenda,” *Journal of Economic Literature* 35, 688-726.

228. Levine, R., 2002. “Bank-Based or Market-Based Financial Systems: Which is Better?” *Journal of Financial Intermediation* 11, 398-428.

229. Lewbel, A., 1996. “Aggregation without Separability: A Generalized Composite Commodity Theorem,” *American Economic Review* 86, 524-543.

230. Ljung, G. M. and G. E. P. Box, 1978. “On a Measure of Lack of Fit in Time Series Models,” *Biometrica* 65, 297-303.

231. Li W., 1991. "Absence of $1/f$Spectra in Dow Jones Average," *International Journal of Bifurcation and Chaos* 1, 583-597.

232. Lucas, R. E., Jr., 1972. "Expectations and the Neutrality of Money," *Journal of Economic Theory* 4, 103-124.

233. Lucas, R. E., Jr., 1973. "Some International Evidence on Output-Inflation Tradeoffs," *American Economic Review* 63, 326-334.

234. Lucas, R. E., Jr., 1976. "Econometric Policy Evaluation: A Critique." In *The Phillips Curve and Labor Markets*, Carnegie-Rochester Conference Series on Public Policy, Vol. 1, K. Brunner and A. H. Meltzer (eds.), Amsterdam: North-Holland.

235. Lucas, R. E., Jr., 1977. "Understanding Business Cycles." In *Stabilization of the Domestic and International Economy*, K. Brunner and A. H. Meltzer (eds.), Carnegie-Rochester Series on Public Policy 5, 7-29.

236. Lucas, R. E., Jr., 1988. "On the Mechanics of Economic Development," *Journal of Monetary Economics* 22, 3-42.

237. Lucas, R. E., Jr., 1996. "Nobel Lecture: Monetary Neutrality," *Journal of Political Economy* 104, 661-682.

238. Lucas, R. E., Jr., 2000. "Inflation and Welfare," *Econometrica* 68, 247-274.

239. Lucas, R. E., Jr. and N. L. Stockey, 1987. "Money and Interest in a Cash-in-Advance Economy," *Econometrica* 55, 491-513.

240. Machina, M., 1982. "'Expected Utility' Analysis without the Independence Axiom," *Econometrica* 50, 277-323.

241. MacKinnon J. G., 1994. "Approximate Asymptotic Distribution Functions for Unit-Root and Cointegration Tests," *Journal of Business and Economic Statistics* 12, 167-176.

242. Mandelbrot, B. B., 1982. *The Fractal Geometry of Nature*. New York: W. H. Freeman and Co.

243. Mankiw, N. G. and L. H. Summers, 1986. "Money Demand and the Effects of Fiscal Policies," *Journal of Money, Credit, and Banking* 18, 415-429.

244. Marty, A. L., 1999. "The Welfare Cost of Inflation: A Critique of Bailey and Lucas," Federal Reserve Bank of St. Louis *Review* 81, 41-46.

245. Mas-Colell, A., M. D. Whinston, and J. R. Green, 1995. *Microeconomic Theory*, New York: Oxford University Press.

246. McCaffrey, D., S. Ellner, R. Gallant, and D. Nychka, 1992. "Estimating the Lyapunov Exponent of a Chaotic System with Nonparametric Regression," *Journal of the American Statistical Association* 87, 682-695.

247. McCallum, B. T., 1989. *Monetary Economics: Theory and Policy,* New York: Macmillan.

248. McCallum, B. T., 1990. "Inflation: Theory and Evidence." In *Handbook of Monetary Economics*, Vol. 2, B. M. Friedman and F. H. Hahn (eds.), Amsterdam: North-Holland, 963-1012.

249. McCallum, B. T. and M. S. Goodfriend, 1987. "Demand for Money: Theoretical Studies." In *The New Palgrave: A Dictionary of Economics,* J. Eatwell, M. Millgate, and P. Newman (eds.), London: Macmillan, 117-130.

250. Mehra, R. and E. C. Prescott, 1985. "The Equity Premium: A Puzzle," *Journal of Monetary Economics* 15, 145-161.

251. Moschini, G., 1999. "Imposing Local Curvature in Flexible Demand Systems," *Journal of Business and Economic Statistics* 17, 487-490.

252. Mulligan, C. B. and X. Sala-i-Martin, 1992. "U.S Money Demand: Surprising Cross-Sectional Estimates," *Brookings Papers on Economic Activity* 2, 285-343.

253. Muscatelli, V. A. and F. Spinelli, 2000. "The Long-Run Stability of the Demand for Money: Italy 1861-1996," *Journal of Monetary Economics* 45, 717-739.

254. Muth, J. F., 1961. "Rational Expectations and the Theory of Price Movements," *Econometrica* 29, 315-335.

255. Nadiri, I. I. and S. Rosen, 1969. "Interrelated Factor Demand Functions," *American Economic Review* 59, 457-471.

256. Nelson, C. R. and C. I. Plosser, 1982. "Trends and Random Walks in Macroeconomic Time Series," *Journal of Monetary Economics* 10, 139-162.

257. Neusser, K., 1991. "Testing the Long-Run Implications of the Neoclassical Growth Model," *Journal of Monetary Economics* 27, 3-37.

258. Newey, W. K. and K. D. West, 1987. "A Simple, Positive Semi-Definite, Heteroskedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica* 55, 703-708.

259. Ng, S., 1995. "Testing for Homogeneity in Demand Systems when the Regressors are Nonstationary," *Journal of Applied Econometrics* 10, 147-163.

260. Ng, S. and P. Perron, 1997. "Estimation and Inference in Nearly Unbalanced Nearly Cointegrated Systems," *Journal of Econometrics* 79, 53-81.

261. Nychka, D., S. Ellner, R. Gallant, and D. McCaffrey, 1992. "Finding Chaos in Noisy Systems," *Journal of the Royal Statistical Society B* 54, 399-426.

262. Offenbacher, E.K., 1979. *The Substitution of Monetary Assets.* Ph.D. Dissertation, University of Chicago.

263. Orphanides, A. and R. Solow, 1990. "Money, Inflation and Growth." In *Handbook of Monetary Economics*, Vol. 1, B. M. Friedman and F. H. Hahn (eds.), Amsterdam: North-Holland, 223-261.

264. Pantula, S. G., G. Gonsalez-Farias, and W. A. Fuller, 1994. "A Comparison of Unit-Root Test Criteria," *Journal of Business and Economics Statistics* 12, 449-459.

265. Pashardes, P., 1993. "Bias in Estimating the Almost Ideal Demand System with the Stone Index Approximation," *Economic Journal* 103, 908-915.

266. Patinkin, D., 1965. *Money, Interest, and Prices*, 2nd ed., New York: Harper & Row Publishers.

267. Peng C.-K., S. V. Buldyrev, S. Havlin, M. Simmons, H. E. Stanley, and A. L. Goldberger, 1994. "Mosaic Organization of DNA Nucleotides," *Physical Review E* 49, 1685-1689.

268. Perron, P., 1989. "The Great Crash, the Oil Price Shosk, and the Unit Root Hypothesis," *Econometrica* 57, 1361-1401.

269. Perron, P. and T. J. Vogelsang, 1992a. "Nonstationarity and Level Shifts with an Application to Purchasing Power Parity," *Journal of Business and Economic Statistics* 10, 301-320.

270. Perron, P. and T. J. Vogelsang, 1992b. "Testing for a Unit Root in a Time Series with a Changing Mean: Corrections and Extensions," *Journal of Business and Economic Statistics* 10, 467-470.

271. Pesaran, M. H., Y. Shin, and R. J. Smith, 2001. "Bounds Testing Approaches to the Analysis of Long-Run Relationships," *Journal of Applied Econometrics* 16, 289-326.

272. Phlips, L., 1974. *Applied Consumption Analysis*, Amsterdam: North-Holland.

273. Phillips, P. C. B., 1987. "Time Series Regresion with a Unit Root," *Econometrica* 55, 277-301.

274. Phillips, P. C. B. and P. Perron, 1987. "Testing for a Unit Root in Time Series Regression," *Biometrica* 75, 335-346.

275. Pollak, R. A. and T. J. Wales, 1980. "Comparison of the Quadratic Expenditure System and Translog Demand Systems with Alternative Specifications of Demographic Effects," *Econometrica* 48, 595-612.

276. Prescott, E. C., 1996. "Commentary [on King and Wolman, "Inflation Targeting in a St. Louis Model of the 21st Century"]," Federal Reserve Bank of St. Louis *Review* 78, 112-115.

277. Pudney, S. E., 1981. "An Empirical Method of Approximating the Separable Structure of Consumer Preferences," *Review of Economic Studies* 48, 561-577.

278. Ramajo, J., 1994. "Curvature Restrictions on Flexible Functional Forms: An Application of the Minflex Almost Ideal Demand System to the Pattern of Spanish Demand," *Journal of Business and Economic Statistics* 4, 431-436.

279. Ramsey, F., 1928. "A Mathematical Theory of Saving," *Economic Journal* 88, 543-559.

280. Ramsey, J. B., 1969. "Tests for Specification Errors in Classical Linear Least Squares Regression Analysis," *Journal of the Royal Statistical Society* B 31, 350-371.

281. Romer, D., 2001. *Advanced Macroeconomics*, 2nd ed., New York: McGraw-Hill.

282. Romer, P. M., 1986. "Increasing Returns and Long-Run Growth," *Journal of Political Economy* 94, 1002-1037.

283. Romer, P. M., 1990. "Endogenous Technological Change," *Journal of Political Economy* 98, S71-S102.

284. Rotemberg, J. J., 1991. "Commentary: Monetary Aggregates and Their Uses." In *Monetary Policy on the 75th Anniversary of the Federal Reserve System*, M. T. Belongia (ed.), Boston: Kluwer Academic, 223-231.

285. Rotemberg, J. J., J. C. Driscoll, and J. M. Poterba, 1995. "Money, Output, and Prices: Evidence from a New Monetary Aggregate," *Journal of Business and Economic Statistics* 13, 67-83.

286. Ryan, D. L. and T. J. Wales, 1998. "A Simple Method for Imposing Local Curvature in Some Flexible Consumer-Demand Systems," *Journal of Business and Economic Statistics* 16, 331-338.

287. Samuelson, P. A., 1958. "An Exact Consumption-Loan Model of Interest, With or Without the Social Contrivance of Money," *Journal of Political Economy* 66, 467-482.

288. Sargent, T. J., 1979. *Macroeconomic Theory*, 2nd ed., New York: Academic Press.

289. Sargent, T. J., 1987. *Dynamic Macroeconomic Theory*, Cambridge, MA: Harvard University Press.

290. Sargent, T. J. and N. Wallace, 1975. "Rational Expectations, the Optimal Monetary Instrument, and the Optimal Money Supply Rule," *Journal of Political Economy* 83, 241-254.

291. Sargent, T. J. and N. Wallace, 1982. "The Real Bills Doctrine Versus the Quantity Theory: A Reconsideration," *Journal of Political Economy* 90, 1212-1236.

292. Saving, T. R., 1971. "Transactions Costs and the Demand for Money," *American Economic Review* 61, 407-420.

293. Schwartz, G., 1978. "Estimating the Dimension of a Model," *Annals of Statistics* 6, 461-464.

294. Serletis, A., 1987. "The Demand for Divisia M1, M2, and M3 in the United States," *Journal of Macroeconomics* 9, 567-591.

295. Serletis, A., 1988. "Translog Flexible Functional Forms and Substitutability of Monetary Assets," *Journal of Business and Economic Statistics* 6, 59-67.

296. Serletis, A., 1991a. "Modeling the Demand for Consumption Goods and Liquid Assets," *Journal of Macroeconomics* 13, 435-457.

297. Serletis, A., 1991b. "The Demand for Divisia Money in the United States: A Dynamic Flexible Demand System," *Journal of Money, Credit, and Banking* 23, 35-52.

298. Serletis, A., 1994. "Testing the Long-Run Implications of the Neoclassical Growth Model for Canada," *Journal of Macroeconomics* 16, 329-346.

299. Serletis, A., 1995. "Random Walks, Breaking Trend Functions, and the Chaotic Structure of the Velocity of Money," *Journal of Business and Economic Statistics* 13, 453-458.

300. Serletis, A., 2007. "A Bayesian Classification Approach to Monetary Aggregation," Mimeo, Department of Economics, University of Calgary.

301. Serletis, A. and I. Andreadis, 2000. "Chaotic Analysis of U.S. Money and Velocity Measures," *International Journal of Systems Science* 31, 161-169.

302. Serletis, A. and Z. Koustas, 1998. "International Evidence on the Neutrality of Money," *Journal of Money, Credit, and Banking* 30, 1-25.

303. Serletis, A. and Z. Koustas, 2001. "Monetary Aggregation and the Neutrality of Money," *Economic Inquiry* 39, 124-138.

304. Serletis, A. and A. L. Robb, 1986. "Divisia Aggregation and Substitutability among Monetary Assets," *Journal of Money, Credit, and Banking* 18, 430-446.

305. Serletis, A. and A. Shahmoradi, 2005. "Semi-Nonparametric Estimates of the Demand for Money in the United States," *Macroeconomic Dynamics* 9, 542-559.

306. Serletis, A. and A. Shahmoradi, 2007. "Flexible Functional Forms, Curvature Conditions, and the Demand for Assets." *Macroeconomic Dynamics,* forthcoming.

307. Serletis, A. and M. Shintani, 2006. "Chaotic Monetary Dynamics with Confidence," *Journal of Macroeconomics* 28, 228-252.

308. Serletis, A. and K. Yavari, 2004. "The Welfare Cost of Inflation in Canada and the United States," *Economics Letters* 84, 199-204.

309. Serletis, A. and K. Yavari, 2005. "The Welfare Cost of Inflation in Italy," *Applied Economics Letters* 12, 165-168.

310. Serletis, A. and J. Vaccaro, 2006. "Cross-Country Evidence on the Demand for Money," *Economia Internazionale* 59, 199-224.

311. Serletis, A. and J. J. Virk, 2006. "Monetary Aggregation, Inflation, and Welfare," *Applied Financial Economics* 16, 499-512.

312. Shintani, M. and O. Linton, 2003. "Is There Chaos in the World Economy? A Nonparametric Test Using Consistent Standard Errors," *International Economic Review* 44, 331-358.

313. Shintani, M. and O. Linton, 2004. "Nonparametric Neural Network Estimation of Lyapunov Exponents and Direct Tests for Chaos," *Journal of Econometrics* 120, 1-33.

314. Sidrauski, M., 1967. "Rational Choice and Patterns of Growth in a Monetary Economy," *American Economic Review* 57, 533-544.

315. Simons, H., 1936. "Rules Versus Authorities in Monetary Policy," *Journal of Political Economy* 44, 1-30.

316. Solow, R. M., 1956. "A Contibution to the Theory of Economic Growth," *Quarterly Journal of Economics* 70, 65-94.

317. Solow, R. M., 1999. "Neoclassical Growth Theory." In *Handbook of Macroeconomic*s, Vol. 1A, J. B. Taylor and M. Woodford (eds.), Amsterdam: North-Holland, 637-667.

318. Sono, M., 1961. "The Effect of Price Changes on the Demand and Supply of Separable Goods," *International Economic Review* 2, 239-271.

319. Spindt, P. A., 1985. "Money Is What Money Does: Monetary Aggregation and the Equation of Exchange," *Journal of Political Economy* 93, 175-204.

320. Sriram, S., 1999. "Survey of Literature on Demand for Money: Theoretical and Empirical Work with Special Reference to Error-Correction Models," Working Paper #99-64, International Monetary Fund.

321. Startz, R., 1979. "Implicit Interest on Demand Deposits," *Journal of Political Economy* 37, 1190-1291.

322. Stock, J. H., 1994. "Unit Roots, Structural Breaks and Trends." In *Handbook of Econometrics*, Vol. 4, R. F. Engle and D. L. McFadden (eds.), Amsterdam: North-Holland, 2739-2841.

323. Stock, J. H. and M. W. Watson, 1988. "Testing for Common Trends," *Journal of the American Statistical Association* 83, 1097-1107.

324. Stockman, A. C., 1981. "Anticipated Inflation and the Capital Stock in a Cash-in-Advance Economy," *Journal of Monetary Economics* 8, 387-393.

325. Stoker, T. M., 1986. "Simple Tests of Distributional Effects on Macroeconomic Equations," *Journal of Political Economy* 94, 763-795.

326. Strotz, R. H., 1957. "The Empirical Implications of a Utility Tree," *Econometrica* 25, 169-180.

327. Strotz, R. H., 1959. "The Utility Tree - A Correction and Further Appraisal," *Econometrica* 27, 482-488.

328. Stutz, J. and P. Cheeseman, 1996. "AutoClass: A Bayesian Approach to Classification." In *Maximum Entropy and Bayesian Methods*. J. Skilling and S. Sibisi (eds.). Kluwer Academic Publishers.

329. Svensson, L. E. O., 1985. "Money and Asset Prices in a Cash-in-Advance Economy," *Journal of Political Economy* 93, 919-944.

330. Swofford, J. L. and G. A. Whitney, 1986. "Flexible Functional Forms and the Utility Approach to the Demand for Money: A Nonparametric Analysis," *Journal of Money, Credit, and Banking* 18, 383-389.

331. Swofford, J. L. and G. A. Whitney, 1987. "Nonparametric Tests of Utility Maximization and Weak Separability for Consumption, Leisure, and Money," *Review of Economics and Statistics* 69, 458-464.

332. Swofford, J. L. and G. A. Whitney, 1988. "A Comparison of Nonparametric Tests of Weak Separability for Annual and Quarterly Data on Consumption, Leisure, and Money," *Journal of Business and Economic Statistics* 6, 241-246.

333. Swofford, J. L. and G. A. Whitney, 1994. "A Revealed Preference Test for Weakly Separable Utility Maximization with Incomplete Adjustment," *Journal of Econometrics* 60, 234-249.

334. Theil, H., 1967. *Economics and Information Theory*, Amsterdam: North-Holland.

335. Tobin, J., 1956. "The Interest Elasticity of the Transactions Demand for Cash," *Review of Economics and Statistics* 38, 241-247.

336. Tobin, J., 1958. "Liquidity Preference as Behaviour Towards Risk," *Review of Economic Studies* 25, 65-86.

337. Tobin, J., 1965. "Money and Economic Growth," *Econometrica* 33, 671-684.

338. Toers, A. L., 1980. "Approximate Variance Formulas for the Elasticities of Substitution Obtained from Translog Production Functions," *Economics Letters* 5, 155-160.

339. Toers, A. L., 1982. "Approximate Variance Formulas for the Elasticities of Substitution Obtained from Translog Cost Functions," *Economics Letters* 10, 107-113.

340. Uzawa, H., 1962. "Production Functions with Constant Elasticities of Substitution," *Review of Economic Studies* 29, 291-299.

341. Varian, H. R., 1982. "The Nonparametric Approach to Demand Analysis," *Econometrica* 50, 945-973.

342. Varian, H. R., 1983. "Non-Parametric Tests of Consumer Behaviour," *Review of Economic Studies* 50, 99-110.

343. Wales, T. J., 1977. "On the Flexibility of Flexible Functional Forms: An Empirical Approach," *Journal of Econometrics* 5, 183-193.

344. Walsh, C. E., 2003. *Monetary Theory and Policy* (2nd edition), Cambridge and London: The MIT Press.

345. Watson, M. W., 1994. "Vector Autoregressions and Cointegration." In *Handbook of Econometrics,* Vol. 4, R. F. Engle and D. L. McFadden (eds.), Amsterdam: North-Holland, 2843-2915.

346. Whang, Y.-J. and O. Linton, 1999. "The Asymptotic Distribution of Nonparametric Estimates of the Lyapunov Exponent for Stochastic Time Series," *Journal of Econometrics* 91, 1-42.

347. Wolf, A., J. B. Swift, H. L. Swinney, and J. A. Vastano, 1985. "Determining Lyapunov Exponents from a Time Series," *Physica D* 16, 285-317.

348. Zellner, A., 1962. "An Efficient Method of Estimating Seemingly Unrelated Regressions and Tests for Aggregation Bias," *Journal of the American Statistical Association* 57, 348-368.

349. Zivot, E. and D. W. K. Andrews, 1992. "Further Evidence on the Great Crash, the Oil Price Shock and the Unit Root Hypothesis," *Journal of Business and Economic Statistics* 10, 251-270.

Author Index

Subject Index