At What Price?

Conceptualizing and Measuring Cost-of-Living and Price Indexes

NATIONAL RESEARCH COUNCIL

"My business book of the year."
—Jeff Madrick, *The New York Times*

At What Price?

Conceptualizing and Measuring Cost-of-Living and Price Indexes

Panel on Conceptual, Measurement, and Other Statistical Issues
in Developing Cost-of-Living Indexes

Charles L. Schultze and Christopher Mackie, *Editors*

Committee on National Statistics
Division of Behavioral and Social Sciences and Education
National Research Council

NATIONAL ACADEMY PRESS
Washington, DC

NATIONAL ACADEMY PRESS • 2101 Constitution Avenue, NW • Washington, DC 20418

NOTICE: The project that is the subject of this report was approved by the Governing Board of the National Research Council, whose members are drawn from the councils of the National Academy of Sciences, the National Academy of Engineering, and the Institute of Medicine. The members of the committee responsible for the report were chosen for their special competences and with regard for appropriate balance.

This study was supported by Contract No. J-9-J-8-0039 between the National Academy of Sciences and the Bureau of Labor Statistics. Support of the work of the Committee on National Statistics is provided by a consortium of federal agencies through a grant from the National Science Foundation (Number SBR-9709489). Any opinions, findings, conclusions, or recommendations expressed in this publication are those of the author(s) and do not necessarily reflect the views of the organizations or agencies that provided support for the project.

Library of Congress Cataloging-in-Publication Data

At what price? : conceptualizing and measuring cost-of-living and price
indexes / Panel on Conceptual, Measurement, and Other Statistical Issues
in Developing Cost-of-Living Indexes ; Charles L. Schultze and
Christopher Mackie, editors.
p. cm.
"Committee on National Statistics, Division of Behavioral and Social
Sciences and Education."
Includes bibliographical references and index.
ISBN 0-309-07442-8 (hardcover)
1. Consumer price indexes—United States. 2. Cost and standard of
living—United States—Statistical methods. 3. Prices—United
States—Statistical methods. 4. Economic indicators—United States. 5.
Index numbers (Economics) I. Schultze, Charles L. II. Mackie,
Christopher D. III. National Research Council (U.S.). Panel on
Conceptual, Measurement, and Other Statistical Issues in Developing
Cost-of-Living Indexes.
HB235.U6 A86 2002
338.5′28′0973—dc21

2001007411

Additional copies of this report are available from the National Academy Press, 2101 Constitution Avenue, NW, Lockbox 285, Washington, DC 20055; (800) 624-6242 or (202) 334-3313 (in the Washington metropolitan area); Internet, http://www.nap.edu

Printed in the United States of America

Suggested citation: National Research Council. (2002). *At What Price? Conceptualizing and Measuring Cost-of-Living and Price Indexes.* Panel on Conceptual, Measurement, and Other Statistical Issues in Developing Cost-of-Living Indexes, Charles L. Schultze and Christopher Mackie, Editors. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

THE NATIONAL ACADEMIES

National Academy of Sciences
National Academy of Engineering
Institute of Medicine
National Research Council

The **National Academy of Sciences** is a private, nonprofit, self-perpetuating society of distinguished scholars engaged in scientific and engineering research, dedicated to the furtherance of science and technology and to their use for the general welfare. Upon the authority of the charter granted to it by the Congress in 1863, the Academy has a mandate that requires it to advise the federal government on scientific and technical matters. Dr. Bruce M. Alberts is president of the National Academy of Sciences.

The **National Academy of Engineering** was established in 1964, under the charter of the National Academy of Sciences, as a parallel organization of outstanding engineers. It is autonomous in its administration and in the selection of its members, sharing with the National Academy of Sciences the responsibility for advising the federal government. The National Academy of Engineering also sponsors engineering programs aimed at meeting national needs, encourages education and research, and recognizes the superior achievements of engineers. Dr. Wm. A. Wulf is president of the National Academy of Engineering.

The **Institute of Medicine** was established in 1970 by the National Academy of Sciences to secure the services of eminent members of appropriate professions in the examination of policy matters pertaining to the health of the public. The Institute acts under the responsibility given to the National Academy of Sciences by its congressional charter to be an adviser to the federal government and, upon its own initiative, to identify issues of medical care, research, and education. Dr. Kenneth I. Shine is president of the Institute of Medicine.

The **National Research Council** was organized by the National Academy of Sciences in 1916 to associate the broad community of science and technology with the Academy's purposes of furthering knowledge and advising the federal government. Functioning in accordance with general policies determined by the Academy, the Council has become the principal operating agency of both the National Academy of Sciences and the National Academy of Engineering in providing services to the government, the public, and the scientific and engineering communities. The Council is administered jointly by both Academies and the Institute of Medicine. Dr. Bruce M. Alberts and Dr. Wm. A. Wulf are chairman and vice chairman, respectively, of the National Research Council.

PANEL ON CONCEPTUAL, MEASUREMENT, AND OTHER STATISTICAL ISSUES IN DEVELOPING COST-OF-LIVING INDEXES

CHARLES L. SCHULTZE *(Chair)*, Brookings Institution, Washington, DC
ERNST R. BERNDT, Sloan School of Management, Massachusetts Institute of Technology
ANGUS DEATON, Woodrow Wilson School, Princeton University
ERWIN DIEWERT, Department of Economics, University of British Columbia
CLAUDIA D. GOLDIN, Department of Economics, Harvard University
ZVI GRILICHES,* Department of Economics, Harvard University
CHRISTOPHER JENCKS, John F. Kennedy School of Government, Harvard University
ALBERT MADANSKY, Graduate School of Business, University of Chicago
VAN DOORN OOMS, Committee for Economic Development, Washington, DC
ROBERT A. POLLAK, John M. Olin School of Business, Washington University, St. Louis
RICHARD L. SCHMALENSEE, Sloan School of Management, Massachusetts Institute of Technology
NORBERT SCHWARZ, Institute for Social Research, University of Michigan
KIRK WOLTER, National Opinion Research Center, Chicago

CHRISTOPHER MACKIE, *Study Director*
CONSTANCE F. CITRO, *Senior Program Officer*
MICHAEL SIRI, *Project Assistant*

---

*Died 1999.

COMMITTEE ON NATIONAL STATISTICS
2001

Acknowledgments

The panel's report is the result of the efforts of many people. The project was sponsored by the Bureau of Labor Statistics (BLS) of the U.S. Department of Labor and also, in part, by the many federal agencies that support the Committee on National Statistics through the National Science Foundation. Throughout the study, the panel received unflagging cooperation from the dedicated staff of BLS. In particular Dennis Fixler (now with the Bureau of Economic Analysis, BEA) and John Greenlees gave generously of their time to educate the panel about the methodology and procedures of the Consumer Price Index. Former Commissioner Katharine Abraham, Mary McCarthy, Charles Mason, Ken Dalton, Dan Ginsburg, Walter Lane, Sylvia Leaver, and Paul Liegey attended meetings, gave briefings, or prepared research materials for the panel.

The panel is also indebted to others who made presentations on data and methodological issues, thereby helping us to develop a broad-based understanding of elements that go into the making of price and cost-of-living indexes: David Cutler, Harvard; John Astin, Eurostat; William Nordhaus, Yale; Alan Blinder, Princeton; and Ariel Pakes, Harvard, all served in this capacity. Others, including Kathleen Scholl, U.S. General Accounting Office; Mick Silver, Cardiff University; Tom Palley, AFL-CIO; and Joe Newhouse, Harvard, supplied the panel with helpful information and suggestions. We offer special thanks to Jack Triplett, formerly of BEA and BLS, for attending early meetings and using his long experience and deep knowledge of the major issues to help us get started and plot our course.

No panel with a task as difficult as ours could perform its duties without an excellent and well-managed staff: Andy White, director of the Committee on

National Statistics (CNSTAT) was responsible for seeing that this was the case. Miron Straf, the former director, was instrumental in developing the study. Project assistants Joshua Dick, Julia Kisa, and Michael Siri provided excellent administrative assistance. Constance Citro, senior project officer, provided experienced and wise counsel throughout the life of the panel. The report also benefitted substantially from the work of Eugenia Grohman, associate director for reports in the Division of Behavioral and Social Sciences and Education, who was responsible for editing the report and who also made valuable suggestions about its structure. We thank her for this and also for her role in overseeing the review process.

The panel owes an especial debt of gratitude to Christopher Mackie, CNSTAT study director for the project. He quickly absorbed the relevant professional literature and steeped himself in the issues. He helped organize our work, coordinated and provided insightful comments on papers and drafts covering a wide range of topics, contributed to our substantive discussions, drafted important parts of the report in cooperation with various panel members, and shepherded the report through the review process.

This report has been reviewed in draft form by individuals chosen for their diverse perspectives and technical expertise, in accordance with procedures approved by the Report Review Committee of the National Research Council (NRC). The purpose of this independent review is to provide candid and critical comments that will assist the institution in making the published report as sound as possible and to ensure that the report meets institutional standards for objectivity, evidence, and responsiveness to the study charge. The review comments and draft manuscript remain confidential to protect the integrity of the deliberative process.

We thank the following individuals for their participation in the review of this report: Ellen R. Dulberger, Business Strategy, IBM Global Services, Somers, NY; Charles Engel, Department of Economics, University of Wisconsin; Robert Feenstra, Department of Economics, University of California, Davis; Ariel Pakes, Department of Economics, Harvard University; Marshall Reinsdorf, Bureau of Economic Analysis, U.S. Department of Commerce; Matthew D. Shapiro, Department of Economics and Survey Research Center, University of Michigan; and Richard Valliant, Westat, Rockville, MD.

Although the reviewers listed above have provided many constructive comments and suggestions, they were not asked to endorse the conclusions or recommendations; nor did they see the final draft of the report before its release. The review of this report was overseen by John Geweke, Department of Economics, University of Iowa, and Christopher Sims, Department of Economics, Princeton University. Appointed by the NRC, they were responsible for making certain that an independent examination of this report was carried out in accordance with institutional procedures and that all review comments were carefully considered.

Responsibility for the final content of this report rests entirely with the authoring panel and the institution.

This report reflects the collective expertise and commitment of the individual members of the panel. All participated in the panel's many meetings and discussions and in the drafting and reviewing of sections of the report. Each member brought a critical perspective and I thank them for their hard work and dedication to service for public benefit. Finally, the substance of this report and its many antecedents owe much to Zvi Griliches. He was a prominent contributor to the literature on many of the issues with which the panel wrestled, and he pioneered in exploring ways to deal with the thorniest issue of all, the problem of quality change. Zvi was a member of the panel but died November 4, 1999, shortly after the group began deliberations. I speak for the entire panel in saluting his monumental contributions in this area. We will miss him.

Charles L. Schultze, *Chair*
Panel on Conceptual, Measurement, and
Other Statistical Issues in Developing
Cost-of-Living Indexes

Contents

EXECUTIVE SUMMARY 1
A COGI Versus a COLI, 2
The Scope or Domain of the Index, 3
Aggregation, 4
Accounting for Substitution Behavior, 5
Quality Change, 6
New Goods, 7
Outlets, 8
Pricing Medical Care, 9
Index Design and Index Purpose, 10
Data Collection, 11

1 INTRODUCTION 13
Background, 13
Panel Charge and Report Organization, 17
Alternative Approaches: A COGI Versus a COLI, 18
Domain of the CPI, 19
Substitution, 21
Aggregation, 26
Changes in the Quality of Goods, 27
New Goods, 30
Outlet Substitution, 32
Stocks and Flows, 33
Index Purpose and Design, 35
Data Collection, 36

2 CONCEPTUAL FOUNDATIONS FOR PRICE AND COST-OF-LIVING INDEXES 38
Setting the Stage: What Are Price Indexes?, 41
The Theory of Price Indexes and Its Critics, 43
Two Perspectives, 57
Conclusions, 73
Technical Note: A Mathematical Approach to Price Indexes, 74

3 INDEX DOMAIN 94
An Unconditional COLI: Conceptual Issues, 96
Measurement Problems, 98
Supplemental Indexes and Satellite Accounts, 101
Other Domain Issues, 102
Conclusion and Recommendation, 105

4 EVOLVING MARKET BASKETS: ADJUSTING INDEXES TO ACCOUNT FOR QUALITY CHANGE 106
COLI and COGI Views of the Quality Change Problem, 109
Evidence from the Boskin Commission Report, 112
BLS Approaches to Quality Changes, 114
Hedonic Regression Methods, 122
Cautions and Recommendations, 140
Technical Note 1: Boskin Commission Estimates of Quality Change and New Goods Bias, 146
Technical Note 2: Mathematical Description of Hedonic Methods, 149

5 NEW GOODS AND NEW OUTLETS 155
New Goods, 155
New Outlets, 167

6 THE SPECIAL CASE OF MEDICAL SERVICES 178
Background, 178
Conceptual and Measurement Issues, 181
The Domain of Consumer Health Expenditures: Employers' Health Insurance Payments, 185
Direct Pricing of Health Insurance, 186
Recommendations, 188

7 INDEX DESIGN AND INDEX PURPOSE 191
Indexing Public Transfer Payments, 192
Wage Bargains and Indexed Wages, 207
Indexing Private Contracts, 208
Inflation-Indexed Treasury Securities, 210

Indexing the Federal Income Tax System, 212
Measuring Output Changes, 214
Inflation Indicators for Macroeconomic Policy, 216
Technical Note: The CPI Versus the NIPA Price Index as an Inflation Measure, 219

8 WHOSE INDEX? AGGREGATING ACROSS HOUSEHOLDS 222
Two Kinds of Heterogeneity, 223
Heterogeneity in Prices Paid and in Rates of Inflation, 225
The Consequences of Heterogeneity for Index Construction, 226
Group Indexes: Why the Current Data Collection System Cannot Produce Them, 226
A Price Index for the Elderly?, 228
The Conceptual Basis for Group Indexes, 229
How Might Data for Subgroup Indexes Be Assembled and What Would It Cost?, 232
Suggested Research and Testing, 235
Plutocratic Versus Democratic Weights, 237
Summary and Recommendation, 240
Technical Note 1: Aggregation and the "Representative Consumer," 241
Technical Note 2: Do Inflation Rates Differ by Age or Income Group?, 246

9 DATA COLLECTION FOR CPI CONSTRUCTION 252
The Current Data Collection Process, 253
Alternative Data Collection Approaches, 264
Summary and Recommendations, 274
Technical Note: Additional Description of CPI Data Inputs, 276

APPENDIX: STATISTICAL DEFINITION AND ESTIMATION OF PRICE INDEXES 283

REFERENCES 293

GLOSSARY 307

BIOGRAPHICAL SKETCHES OF PANEL MEMBERS AND STAFF 315

INDEX 319

At What Price?

Conceptualizing and Measuring Cost-of-Living and Price Indexes

Executive Summary

The Consumer Price Index (CPI) is one of the most widely used statistics in the United States. As a measure of inflation it is a key economic indicator. It serves as a guide for the Federal Reserve Board's monetary policy and is an essential tool in calculating changes in the nation's output and living standards. It is used to determine annual cost-of-living allowances for social security retirees and other recipients of federal payments, to index the federal income tax system for inflation, and as the yardstick for U.S. Treasury inflation-indexed bonds.

There has long been both research and policy debate about the appropriate conceptual framework for the CPI and whether it might be overstating changes in consumers' costs of living. Forty years ago the Stigler committee outlined the difference between the CPI and a "true" cost-of-living index and recommended that the Bureau of Labor Statistics (BLS) undertake research to move the CPI closer toward a cost-of-living index. The subject was given new public prominence in the 1990s by increasing congressional concerns over the budget and the role of the CPI in determining social security cost-of-living allowances and tax indexation. A 1996 report by a congressionally appointed committee—known as the Boskin commission after its chair—estimated that the CPI was overstating the rise in the cost of living by about 1.1 percentage points a year and recommended changes in the way the CPI is designed and estimated.

Underlying some of the arguments and questions about the CPI is a fundamental issue of the nature of the index. Traditionally, a consumer price index measures the change in expenditures required by a household to purchase a fixed-weight basket of goods and services when prices change between some initial reference period and a subsequent comparison period. The panel labels this a cost-of-goods index (COGI) (for convenience, we use the term "goods" throughout this report to denote goods and services, unless otherwise specified). In contrast, a cost-of-living index (COLI) measures the change in expenditures a household would have to make in order to maintain a given standard of living.

In 1997 BLS told Congress that it had been using the cost-of-living concept for many years as a framework for making decisions about the CPI and that it accepts the COLI as the measurement objective for the index. Recognizing the many theoretical and measurement issues involved in embodying a cost-of-living concept in an index, BLS asked the Committee on National Statistics of the National Academies to convene a panel of experts "to investigate conceptual, measurement, and other statistical issues in the development of cost-of-living indexes."

A COGI VERSUS A COLI

For dealing with many of the issues considered in this report, there are close parallels between the COGI and COLI approaches. Nevertheless, having a clear conceptual basis for the index is important. It serves as an authority that can be appealed to when making difficult choices among alternative procedures or for accommodating the new developments constantly being generated by a technologically innovative economy.

The cost-of-living approach provides a rationale for taking account of the fact that, when prices change, consumers do not continue to purchase the same fixed basket, but shift their purchases toward goods whose relative prices have fallen. The concept of the COLI explicitly takes into account the effect of this substitution behavior in reducing the expenditure required by a consumer to maintain a given standard of living when prices change.

Probably the single most difficult and important task in index construction is dealing with the ongoing flow of quality changes among consumer goods and services. Many economists consider the economic theory underlying the COLI a helpful way of initially approaching the problem because it prompts the question: "What are the particular attributes of goods that consumers value?" This may provide a way to start, but the panel found that, beyond this point, current techniques for addressing problems associated with changing item quality can be analyzed with minimal use of the theory underlying the COLI and that the techniques could be applied within either a COGI or a COLI framework.

While a COLI framework offers some conceptual advantages, giving up the relative simplicity of the COGI comes at a cost. Conditions that complicate the estimation and cloud the interpretation of COLIs—such as changes in consumer tastes or changes in buying patterns caused by changes in income—may be present in practice. A CPI constructed on cost-of-living principles can, therefore, only be an approximation to the COLI that it seeks to measure. Moreover, restricting the COLI to cover only the universe of private goods and services, as the BLS does and the panel recommends, requires that it be a "conditional" COLI; that is, it should measure changes in consumers' costs of living on the assumption of stability in conditions—such as the weather or the quality of publicly provided goods—that are outside the universe of private goods. But the choice of exactly which outside conditions should be held constant in the conditional COLI is sometimes controversial and cannot itself be derived from the theory underlying the COLI. And the fact that consumers' demands for private goods and services often do change in response to outside conditions provides another reason why there is a range of circumstances under which a CPI constructed on cost-of-living principles can only approximate a COLI.

If asked to assess the relative merits of the two conceptual approaches as a guide for construction of the CPI, various members of the panel would strike the balance differently. All panel members find it difficult to think about the definition of goods and about quality change without considering what it is that consumers value, and agree that it is impossible to think about substitution behavior without the concept of a constant standard of living which allows price changes to be converted into a monetary equivalent. For all these issues, especially the last, the cost-of-living framework is central. However, some panel members are skeptical about our ability to define a constant standard of living in an economy in which the nature of goods and services is constantly changing. They point out that the conceptual framework underlying the COLI is not always well defined in the presence of quality change and, therefore, they conclude it provides, at best, a limited advantage over the COGI approach in handling this most difficult of issues. They are also concerned about the BLS adopting an approach that differs from that of many statistical offices around the world.

Despite these differences, all panel members agree that the COGI and the conditional COLI that the panel recommends share many common aspects. We also concur that neither conceptual approach, viewed in its pure form, can provide the single guide to index construction but that each can make a contribution toward dealing with the various problems that arise in designing the CPI. Taking a pragmatic approach, the panel found that it could come, sometimes by different routes, to unanimous agreement on all of the specific recommendations in this report. But in its inability to achieve unanimity behind a recommendation that the cost-of-living framework be the sole appropriate basis for construction of the CPI, our panel differs from the Stigler committee and the Boskin commission.

THE SCOPE OR DOMAIN OF THE INDEX

For the reasons set forth in Chapters 2 and 3 of this report, we arrived at two general conclusions, largely about the conceptual basis for price and cost-of-living indexes, which serve to guide our more detailed conclusions and recommendations that appear later in the report.

An unconditional cost-of-living index is an unsuitable conceptual basis for the CPI. While research aimed at better understanding the economic effects related to changes in such matters as life expectancy, crime rates, or the environment would be useful for evaluating various aspects of public policy, the CPI should not change in response to changes in such factors. (Conclusion 2-1)

Within the general conceptual framework of cost-of-living indexes, the appropriate theoretical concept for the CPI is a *conditional* cost-of-living index that is restricted to private goods and services and in which environmental background factors are held constant. (Conclusion 2-2)

The BLS should not conduct research on its own aimed at producing a CPI with a substantially broader domain. That said, the panel encourages the BLS—jointly with other federal statistical agencies, particularly the Bureau of Economic Analysis (BEA)—to undertake or sponsor research aimed at producing, on an experimental basis or in satellite accounts, more comprehensive measures of national output, income, and prices. These accounts would seek to include the effects on output, income, and prices from changes in some of what we have labeled "outside conditions" in those cases where there may be at least some chance of measuring those effects—perhaps, for example, changes in the status of the natural environment.

AGGREGATION

Households differ from one another in their consumption patterns and shopping behavior and often pay different prices for the same goods. Part of this heterogeneity is associated with differences in households' economic and demographic characteristics and in their geographic location. This fact gives rise to two kinds of issues: First, for such purposes as adjusting social security payments and the tax system, and for measuring changes in real income, when can one aggregate the data for the whole population into a single official price index; when are different price indexes needed for specific population subgroups; and how can the data needed to produce such subgroup indexes be collected? Second, when a single overall index is produced, how should the costs of living of individual households be combined into a single national index? Should equal weight be given to each household's cost of living (a "democratic" index) or, as is now the case, should costs of living be weighted by the overall consumption spending of each household (a "plutocratic" index)?

The Consumer Expenditure Survey indicates the extent to which various economic and demographic groups allocate their budgets differently among categories of goods and services. The panel believes, however, that substantial variation may also exist among different groups of households with respect to the particular types and qualities of goods they purchase and the prices they pay within each category. But because the price data used to produce the CPI are collected from retail stores and not directly from households, it is impossible to associate the economic and demographic characteristics of buyers with the items they buy and the prices they pay. As a consequence, it is impossible to investigate satisfactorily the two major aggregation issues: To what extent does inflation or changes in living costs differ among the various economic and demographic groups? And to what extent would a democratic index behave differently from a plutocratic one?

With current survey techniques and methods, collecting price as well as expenditure data from households on a scale sufficient to produce the CPI and an array of group indexes would be extremely expensive and possibly even infeasible; we therefore propose a more modest plan:

> **BLS should pursue an exploratory research program that would, initially only on a small scale, investigate and assess several alternative approaches—including, but not limited to, the use by survey respondents of handheld scanners and computers—for collecting prices in a way that allows them to be associated with household characteristics. A first objective might be the production of indexes for a few commodity categories and several demographic groups. (Recommendation 8-1)**

ACCOUNTING FOR SUBSTITUTION BEHAVIOR

When prices change, consumers tend to shift their purchases toward those goods and services whose relative prices have decreased, thereby reducing any adverse consequences of the price changes on their costs of living. A fixed-basket index does not reflect this substitution effect. The BLS has recently made some changes in the method of constructing price indexes for many categories, or *strata*, of goods (utilizing geometric means of individual price relatives) in an effort to capture within-strata substitution effects. It will shortly begin producing a superlative index to approximate substitution effects among strata. But because some of the data necessary to construct a superlative index will not be available to meet the CPI's publication schedule, the superlative index will be available only after a 2-year lag.

The panel agrees that the BLS should continue to produce, as its main index, a real-time CPI. employing a selective use of geometric means for producing individual strata indexes and Laspeyres weights to combine the strata indexes into the overall CPI. Further research should be conducted on consumer shopping and substitution behavior with an eye to improving knowledge of the appropriate application of geometric means at the lower level of index construction.

The BLS should also proceed as planned to begin publishing a superlative index with a 2-year lag. For purposes of producing a timely index for determining cost-of-living allowances for social security benefits and other indexed programs, we recommend an additional series:

> **The BLS should publish, contemporaneous with the real-time CPI, an advance estimate of the superlative index, utilizing either a constant-elasticity-of-substitution method or some other technique. (Recommendation 7-1)**

QUALITY CHANGE

Dealing with the ever-changing mix and quality of available goods and services poses the most numerous and difficult problems in constructing the CPI. Items constantly disappear from store shelves and are replaced in the index with similar but somewhat different items carrying different prices. The BLS must continually make judgments about how much of a price difference represents "pure" price change and how much represents a quality difference. Increasingly, BLS has been turning to explicit quality adjustment techniques, principally hedonics, in which statistical regressions are used to assign monetary values to differences in the particular characteristics of a type of product and to adjust its reported prices accordingly when the characteristics of the good change.

Hedonic techniques currently offer the most promising approach for explicitly adjusting observed prices to account for changing product quality. But our analysis suggests that there are substantial unresolved econometric, data, and other measurement issues that need further attention. The panel makes a number of recommendations to deal with this set of opportunities and problems:

> **BLS should continue to expand its experimental development and testing of hedonic methods and its support of relevant outside research. This research should not be confined to that relating to price adjustment but should also examine the role of hedonics in statistical audits of the other BLS quality adjustment methods and in the review of replacement item selection procedures and comparability decisions. (Recommendation 4-2)**

The above recommendation does not suggest that BLS should immediately expand the use of hedonics in constructing component indexes for its flagship CPI. In fact, the panel takes the opposite position:

> **Relative to our view on BLS research, we recommend a more cautious integration of hedonically adjusted price change estimates into the CPI. (Recommendation 4-3)**

This recommendation is based on theoretical considerations, not on empirical grounds. As documented in the report, the recent BLS expansion of hedonic price adjustments to appliances and electronics has not had a large impact on those item subindexes. Our conservative view on integrating hedonics techniques has more to do with concern for the perceived credibility of the current models. While there is an established academic literature on estimating hedonic functions, researchers are much less experienced using them across a wide variety of goods in price index construction. Thus, while members of the panel agree that BLS and others should vigorously continue to research the viability of hedonics, the methods may, in their current state of development, only be justifiably applied to a narrow class of goods.

So long as hedonic techniques are restricted to replacements for items that have disappeared from store shelves, as is now predominantly the case, their use will not have a significant impact on index growth. Only if extended on a broader basis (e.g., to items coming into the index through the rotation of the retail store sample) will the use of those techniques make much difference. Such an extension would be unwarranted until the recommended research, development, and testing program makes progress on the measurement issues we have identified. To assist in this task, we recommend the following:

> **An independent advisory panel, consisting of econometricians, statisticians, index experts, marketing specialists, and possibly product engineers should be formed to provide guidance on both conceptual and application issues pertaining to hedonic methods. (Recommendation 4-8)**

BLS, working with the recommended advisory panel, should assess the impact of modeling imperfections on the validity of their hedonic adjustments prior to their introduction into the index. This would provide an analytic basis for proceeding sensibly in the face of external pressures to proceed quickly in this area. The advisory panel should also provide outside review to help guide decisions about potential new applications and about which BLS pilot studies are adequately developed to be incorporated into the index. Together, our recommendations emphasize the high priority that the hedonics research program should receive.

NEW GOODS

Another class of product changes involves the appearance of goods with genuinely new characteristics (such as mobility for phones). These goods are sufficiently unlike existing ones in that they do not enter the CPI as part of the item replacement procedure or even when the sample of retail outlets is rotated. Hedonic techniques do not hold much promise for measuring the effect on the index of the introduction of such goods.

If a new good displays new characteristics, it is likely to become eligible for inclusion in the market basket only when item strata are redefined and upper-level weights reestimated. To the extent that new goods offer previously unavailable benefits to early purchasers and because they typically experience price reductions early on, some declines in the cost of living are missed during the period before the new goods are incorporated into the index. Prominent examples of this phenomenon occurred when mobile phones and VCRs were introduced into the index many years after their appearance on the market.

Some proponents of the COLI approach argue that econometric methods should be used to estimate the "virtual" price reduction that occurs when a new product appears. Those estimates, in turn, could be incorporated into the index. However, the panel had serious doubts about the effectiveness of econometric techniques in this regard, and some members dispute the conceptual validity of treating the benefits from introducing new products as a price decrease:

Virtual price reductions associated with the introduction of new goods should not be imputed for use in the CPI. (Conclusion 5-1)

Members of the panel recognize that, outside of price measurement, there is nowhere in the national accounts for the effect of new products to be included, and real growth in the economy may therefore be understated. Rather than modifying the CPI, the panel suggests that research in this area be directed toward developing a separate experimental COLI that is adjusted, to the extent possible, to account for changes as new products and technologies diffuse throughout the economy.

Additionally, because, once introduced, new goods frequently display very different price trends from established ones, the panel does endorse BLS's recent efforts to update weights every 2 years, to streamline sample rotation, and to perform targeted product introductions, all of which should enhance the probability that new products will enter the CPI basket more quickly than has historically been the case.

OUTLETS

Another potential bias of the CPI, when used as a COLI (or possibly even as a COGI), arises because different stores sell identical items at different prices. If price variation is not proportional to differences in the quality of the retail service offered (as the ongoing trend to lower-price, lower-service outlets might suggest), consumers can lower their living costs by altering their shopping behavior. These types of "price reductions" are not fully captured by the CPI. Currently the underlying conceptual apparatus of the CPI assumes that when lower-price outlets enter the sample, there is no net price reduction, because all of the price difference between the old and the new outlet reflects a difference in the quality of service.

Because current techniques cannot consistently and accurately separate quality changes from the price effects associated with the value of retail service, BLS has little choice but to continue this practice, though the body of this report does discuss a couple of alternatives. However, *in principle*, when outlet rotation results in a change in the observed price of an identical product, an attempt should be made to decompose the difference into quality (or convenience) and pure price components instead of attributing it entirely to the former.

> **With longer-term modifications in mind, the panel recommends pursuing research into price variation across outlets with differing characteristics. (Recommendation 5-2)**

PRICING MEDICAL CARE

Medical care, one of the eight major product groups in the CPI, currently accounts for just less than 6 percent of consumer expenditures included in the index. Total expenditures on health care amount to almost 18 percent of consumption outlays, but the domain of the Medical Care Price Index (MCPI) in the CPI is limited to consumers' out-of-pocket expenditures, thus excluding costs paid by Medicare, Medicaid, and employer-financed health insurance (as well as other smaller items). In the case of health insurance premiums paid by households themselves, the BLS does not price the premium cost of the insurance directly but imputes to it the prices of the underlying medical care services that are purchased with the premium.

Because of the complicated institutional setting in which medical care services are provided and financed, together with the rapid pace of development of new medical technologies, their appropriate pricing probably constitutes the most difficult single task in producing the CPI. The panel makes a number of recommendations in this area:

> **BLS should select between about 15 to 40 diagnoses from the ICD (International Classification of Diseases), chosen randomly in proportion to their direct medical treatment expenditures and use information from retrospective claims databases to identify and quantify the inputs used in their treatment and to estimate their cost. On a monthly basis, the BLS could reprice the current set of specific items (e.g., anesthesia, surgery, medications), keeping quantity weights temporarily fixed. Then, at appropriate intervals, perhaps every year or two, the BLS should reconstruct the medical care index by pricing the treatment episodes of the 15 to 40 diagnoses—including the effects of changed inputs on the overall cost of those treatments. The frequency with which these diagnosis adjustments should be made will depend in part on the cost to BLS of doing so. The resulting MCPI price indexes should initially be published on an experimental basis. The panel also recommends that the BLS appoint a study group to consider, among other things, the possibil-**

ity that the index will "jump" at the linkage points and whether a prospective smoothing technique should be used. (Recommendation 6-1)

Additionally, the panel concluded that a price index including a more broadly based measure of the changing cost of medical care would be valuable for a wide range of policy purposes.

BLS should include the portion of health insurance paid for by employers in one version of the CPI, perhaps calling it an "expanded-scope medical CPI." Because many commonly used income measures exclude employer-provided benefits, and because the Consumer Expenditure Survey is based only on out-of-pocket expenditures, the original conception of the MCPI domain should still be maintained in constructing the traditional (flagship) CPI. The panel also recommends examining the practicality of including other employer-paid employee benefits (e.g., dental and cafeteria plans) in the expanded-scope CPI. (Recommendation 6-2)

To inform public policy discussions and to evaluate the performance of the U.S. health care sector, a medical care price index that encompasses purchases from all payers is needed.

A task force should be convened by the BLS, in collaboration with the Centers for Medicare and Medicaid Services and other appropriate agencies, to implement construction and publication of a total medical care expenditure price index, encompassing purchases from all health care payers—governments, private third-party insurers, and consumers. (Recommendation 6-3)

The most difficult issue in the construction of the MCPI concerns adjustments for quality change. New treatments can yield improved outputs in the form of extended and better quality life. The panel believes that an outcomes-based measure is in principle superior to an input-based measure, but we recognize the formidable measurement challenges and do not know how best to proceed. This area is new and requires a good deal more research, much of it interdisciplinary. After BLS has implemented Recommendation 6-1, it can then consider whether, how, and why the outcomes of the treatments for those diagnoses are changing over time, and finally consider how outcomes changes should best be evaluated in computing a quality-adjusted medical care price index.

INDEX DESIGN AND INDEX PURPOSE

The CPI and its individual components are used for a wide range of sometimes dissimilar purposes. In some cases different uses may call for different index designs. But no statistical index can perfectly match what is desired for a particular purpose, and practical considerations limit the number of indexes that can or should be produced. Chapter 7 evaluates the extent to which the CPI and existing or proposed supplemental indexes meet the needs of various users.

The panel concludes that a superlative index is appropriate for adjusting benefits to keep pace with the cost of living. In this context, the panel suggests the following:

> **It would be feasible and appropriate to calculate cost-of-living allowances provided for social security and other programs from an advance estimate of the BLS published superlative index. Any divergence between that estimate and the superlative that appears 2 years later could be incorporated as a correction to the cost-of-living allowance provided for that year. (Conclusion 7-1)**

A related question is whether social security cost-of-living allowances (COLAs) should be based on a special index for the elderly. Using data from the last several decades, BLS has produced a special index for the elderly (CPI-E) by weighting the price indexes for various categories of goods according to the purchasing patterns of the elderly rather than the general population. This index did not rise at a significantly different rate than the overall CPI. Different groups not only have different overall consumption patterns but face different prices and buy different qualities of goods; the BLS has called attention to this limitation of its experimental index. In the absence of an index that can capture these differences, we see no rationale for basing social security COLAs on the type of indexes constructed in the BLS studies. But the CPI-E should be periodically updated to make sure that no significant differences with the CPI have developed.

Adjusting social security benefits for retirees with a wage index would be an alternative to CPI indexation. The panel was not charged to make recommendations on this issue, but we do spell out the implications of this and other indexing methods for public policy.

DATA COLLECTION

The data inputs used to calculate the CPI subindexes originate from several sample-based sources, most notably the Consumer Expenditure Survey (CEX), the Point of Purchase Survey (POPS), the Commodities and Services Survey, and the CPI Housing Survey. The panel considered two distinct approaches for upgrading this apparatus. One is to assume that the basic data collection structure will remain as is and then to seek ways of improving each of the survey components. Another is to redesign, from scratch, the entire data collection structure so that it reflects advances in data collection technology and so that the data collected are more consonant with the ultimate computation of the CPI.

The panel's foremost concern with the CEX, which is the primary tool for establishing CPI weights at the basic item level, is the extent of biases in household-reported expenditures which, in turn, affects the accuracy of upper-level CPI item category weights.

Before additional resources are directed toward increasing its sample size (beyond the current plan), the accuracy of the CEX should be carefully evaluated. Assessing the net advantages of using the BEA's per-capita personal consumption expenditures (PCE) data to produce the upper-level weights for the national CPI should be part of this evaluation. (Recommendation 9-1)

Comparison of the CEX and PCE estimates suggests that, even allowing for errors in the latter, the CEX generates biased weights for a number of items. Even if the current system is ultimately maintained, the effort will produce additional guidance about how the CEX might be improved.

If categories can be reasonably well matched between the CPI and PCE, so that comparable item strata indexes can be created, a program should be set up to produce an experimental CPI that uses PCE-generated weights at the upper (218 item) level but that is otherwise no different from the CPI. (Recommendation 9-2)

Even if it is confirmed that the CEX is the best choice for establishing upper-level expenditure weights, the panel is hesitant to recommend expensive increases in the sample size. The panel's calculations suggest that, if the goal is only to reduce the standard error of the national-level expenditure weights, resources spent to increase the sample size of the CEX beyond that which is currently planned would be largely wasted.

In considering alternative data collection approaches, the panel suggests that BLS (1) investigate the possibility of combining the POPS and CEX into an integrated survey that obtains expenditure and outlet-use data at detailed product levels, along with household demographic information needed for subgroup indexes and (2) continue its work on increasing the utilization of both store- and household-based scanner data.

1

Introduction

There has been widespread interest in evaluating the adequacy of the Consumer Price Index (CPI) as used for various purposes. Some of that interest reflects a view that the annual inflation rate of the CPI exceeds that of some "true" cost-of-living index and leads to an overstatement of cost-of-living adjustments for social security and other public programs. But there has also been a growing research literature among economists and statisticians, much of it from within the Bureau of Labor Statistics (BLS), critically examining and exploring means for improving the design and estimation techniques underlying the CPI.

BACKGROUND

Price indexes have often been popularly labeled cost-of-living indexes; indeed, until it was renamed the CPI in 1945, the index long published by the BLS had officially been labeled a cost-of-living index. Several decades ago the CPI was widely used in labor contracts to index wages, with the goal of providing automatic adjustments to keep wages fully or partially current with changes in the cost of living. With the same goal in mind—protection against changes in the cost of living—the Congress determined in 1972 that the CPI should be used to make annual "cost-of-living adjustments" to social security benefits, and the practice was subsequently extended to many other public transfer payments. Since 1985 the CPI has been used to index tax brackets, exemptions, and deductions in the tax code so as to "neutralize" the effects of inflation. The annual change in the index is widely used in and outside government as a broad measure of inflation. And its components are the main source of deflating the current dollar value of consumer expenditures as part of the measurement of the nation's gross domestic product (GDP).

The growth rate of traditional price indexes like the CPI, which measure the cost of purchasing a *fixed basket* of goods and services, tends to outpace cost-of-living indexes, which attempt to calculate the change in expenditure needed to maintain living standards. Concerns have long been expressed that the CPI does not adequately take account of improvements in the quality of consumer goods and services in a technologically dynamic economy and thereby overstates the price increases consumers are paying for goods of constant quality. As a consequence, indexing wages, social security benefits, or other payments scaled to the CPI would usually overstate the amount needed to compensate for increases in the cost of living. Forty years ago, the Stigler Committee outlined the conceptual and measurement characteristics of the CPI that distinguished it from a "true cost-of-living index"—or, under alternative committee labels, a "welfare index," or a "constant utility" index (National Bureau of Economic Research, 1961). The principal recommendation of the committee was the establishment of a long-run research program designed to make the CPI a better approximation to a cost-of-living index.

In recent years, as the projected long-term financing deficit in the social security system has grown, the question of whether and to what extent the CPI is biased upward, and therefore "overcompensates" social security beneficiaries, has become a concern among some legislators. In 1995 the Senate Finance Committee appointed an Advisory Committee to Study the Consumer Price Index (widely known as the Boskin commission after its chair, Michael Boskin) to review this issue. In its widely publicized final report of December 1996, the Boskin commission concluded that the CPI was currently overstating the rate of increase in consumers' cost of living by about 1.1 percentage points a year, and it cited estimates from other research pointing to approximately the same result. The commission recommended a number of steps designed to move the CPI away from what was essentially an index of the cost of purchasing a fixed basket of consumer goods toward what would be more nearly a cost-of-living index (COLI).

In 1997 the BLS reported to Congress that it had been using a COLI concept for many years to help make decisions about the CPI and that it accepted a COLI as the measurement objective for the index (Bureau of Labor Statistics, 1997c).[1] The report of the Boskin commission, however, undoubtedly spurred BLS to broaden and make more explicit that commitment, and it only recently began taking steps to modify the fixed-weight structure of the CPI so as to bring it closer to a COLI.

---

[1]The *Handbook of Methods* (see, for example, the Bureau of Labor Statistics 1984 and 1992 versions) notes that "a unifying conceptual framework for dealing with practical questions that arise in construction of the CPI is provided by the concept of the cost-of-living (COL) index."

A fixed-basket, or fixed-weight, price index is essentially just that: it measures changes in the cost of purchasing a fixed basket of goods (and services). For the CPI, price quotes are collected monthly, selected to be representative of the various categories of consumer goods and services. The observed price changes are assigned weights, representing the importance of each category in aggregate consumer expenditures during some base period, then combined into the major CPI subcomponents, such as food, shelter, appliances, and so forth and, subsequently, into an overall national average.

A COLI is more ambitious and correspondingly more difficult to produce in that its objective is to measure changes in living costs. Viewed from the standpoint of an individual household, a COLI seeks to measure the percentage change in expenditures a household would have to make in order to hold constant some specified standard of living or level of material well-being.[2] In an aggregate COLI, price and expenditure data must be combined to produce an estimate that reflects some measure of average change in the cost of living for all (or some subgroup of) households.

In recently reiterating its acceptance of a COLI as the measurement objective for the CPI, the BLS added a number of important cautions: "It [the COLI] is a theoretical concept based on the well-being of the individual consumer, so . . . additional assumptions about how to apply it as a measurement objective for an aggregated set of consumers . . . must be made" (Bureau of Labor Statistics, 1997b: 3). Further: "While the CPI may be described formally in the context of a cost-of-living index, there is no single all-purpose definition of the target." The concept of the standard of living that is to be held constant in a cost-of-living index is far from unambiguous. Various analysts have offered different definitions of what universe it should cover (e.g., the standard of living obtainable from public and private goods or from private goods only), and embedding the concept in a regularly published statistical index raises thorny problems.

The discussion and controversy about the CPI reflect a large number of conceptual and measurement issues: As a guide for the BLS in making decisions about how the index should be designed and measured, what are the advantages and limitations of the concepts that underlie fixed-weight and cost-of-living indexes? For many years attention centered on the "substitution issue": To what extent is it possible to incorporate into the index the tendency of households to shift their purchases toward those goods whose prices have risen the least or fallen the most? But there are other important questions of index design whose

---

[2]See Chapter 2 for a discussion of standard of living in the context of cost-of-living theory. Briefly, consumers think more goods are better than less and can consistently rank alternative bundles of goods in terms of a set of preferences. Constrained by income and prices, each consumer chooses the most preferred bundle of goods. The consumer's "standard of living" (or "material well-being") is a measure of the extent to which preferences are satisfied.

resolution depends in part on whether one evaluates them through the prism of a fixed-weight or a cost-of-living viewpoint and on how the cost-of-living approach is interpreted: How aggressively and comprehensively should the BLS pursue efforts to use econometric techniques to adjust observed prices for the effect of quality improvements? How comprehensive should be the universe of goods covered by the CPI—should it cover private goods only or also encompass public goods? Should the BLS take into account, to the extent measurement is feasible, the effects on living standards—and therefore on living costs—of changes in pollution, crime rates, congestion, and other "environmental" developments? How should the index take account of the effect on living standards of the continual introduction of new goods in our technologically innovative society? And, in either a fixed-weight or a cost-of-living index, how should the experiences of the rich and the poor, the old and the young, be combined into a single index, and should indexes for population subgroups also be published?

The difference between the two approaches to index construction is not fully captured by juxtaposing the terms "fixed-weight index" and "cost-of-living index." The objectives of the two indexes are not the same. The former seeks to measure the effects of price changes on the cost to a household of purchasing a specified basket of goods and services. The latter seeks to measure the effects in terms of the cost of maintaining the household's standard of living at some specified level. The two effects are not usually the same. And, in a world in which consumer tastes change and the qualities of many goods and services are constantly being altered, measuring either type of index is a difficult task. A more appropriate terminology would contrast a "cost-of-goods index" (COGI, where "goods" includes both goods and services) with a "cost-of-living index."[3] Considered from the standpoint of an individual household, a COGI seeks to provide a measure of the percentage change in expenditures the household would require to purchase a basket of goods, given a change in prices between some initial period (usually called the reference period) and some later (comparison) period.[4] As its name implies, it seeks to measure changes in the cost of goods. In principle, for a COGI, specification of the basket of goods may be based on a past period's consumption patterns or current patterns, or even a point in between. A COLI, as noted above, seeks to measure the percentage change in expenditures needed to maintain a household's standard of living at some specified level (typically, but not necessarily, the level it had in the reference period). As its name implies, its objective is to measure changes in the cost of living.

---

[3]Unless otherwise specified, "goods" refers to both goods and services throughout the report.

[4]Technically, the index measures the *ratio* of the expenditures needed in the current period to purchase the same basket of goods as in the base period; the percentage change is simply that ratio minus 1.0 (times 100).

PANEL CHARGE AND REPORT ORGANIZATION

In view of the wide range of important issues that have been raised, the BLS asked the Committee on National Statistics to convene the current panel, which was charged with two primary tasks: "(1) investigating conceptual, measurement, statistical, and data issues in the development of cost-of-living indexes and (2) assessing the appropriate use of such indexes for indexing federal programs and other purposes." The statement of task further notes: "Topics of the assessment would include the required frequency, the technical appropriateness of revisions, and the treatment of quality change and new products. The panel would be asked to make explicit the assumptions and models underlying different approaches and to recommend a program of research and experimental measures." The remainder of this initial chapter provides a brief introduction to the key issues and problems considered by the panel in its effort to contribute to the understanding of price and cost-of-living indexes and their limitations and complexities.

Chapter 2 considers alternative conceptual foundations for the CPI, specifically COGI and COLI approaches. The chapter steps through the key attributes (many of which are taken up in greater detail later in the report) that define price and cost-of-living indexes and examines the relative strengths, weaknesses, limitations, and implications associated with each of these approaches.

Chapter 3 presents the panel's assessment of what goods, broadly defined, are appropriate for inclusion in the scope of CPI coverage. An all-encompassing cost-of-living index would attempt to cover—in addition to private market goods—goods provided by government, environmental amenities, and other nonprivate societal conditions (such as public safety).

Chapter 4 discusses the conceptual rationale, methodology, and limitations of adjusting indexes or observed price quotes to account for changing item quality. The chapter reviews in detail current BLS approaches of price adjustment that come into play when items are replaced in sampled outlets. The panel assesses these methods and advances proposals relating to the use and potential of different quality adjustment methods.

Chapter 5 first discusses two issues related to the introduction of new goods: (1) what criteria should determine when and how new goods are introduced into the index and (2) should estimates of "virtual price" decreases associated with their introduction be made and incorporated into the CPI? The second part of the chapter addresses how changes in the patterns of consumer patronage among different types of retail outlets affect living costs and price indexes. The panel specifically considers what, if anything, BLS could do to identify and estimate quality and pure price components of differences in the observed prices of goods across outlets.

Chapter 6 examines conceptual and measurement issues pertaining to the construction of the complicated medical services component of the CPI. Those complications include high variability of prices paid for equivalent services, defining a medical "good," involvement of insurers and government in transactions, pricing risk, and how adjusting medical care prices to account for the quality of outcomes can lead to strange results. The chapter also discusses "outcomes" and direct insurance pricing options.

Chapter 7 examines the relationship between each of the major purposes for which the CPI is used and the appropriate design of the index. It considers the extent to which different index designs are required for different purposes and when a single design can serve as an acceptable measurement instrument for many purposes. It also spells out the implications for various public policy purposes of choosing one index design over another.

Chapter 8 describes the issues that are confronted when a single index must be produced to represent the changes in prices or living costs faced by a heterogeneous population. It emphasizes problems that arise because different consumers buy different types and qualities of goods and pay different prices for them.

Chapter 9 provides an overview and assessment of the current data structure that underlies the CPI and considers ways that data and survey advances might be coordinated to improve the accuracy of the CPI. It also describes the extent to which different data structures permit flexibility in constructing alternative or supplemental indexes (such as for subpopulations).

ALTERNATIVE APPROACHES: A COGI VERSUS A COLI

Between a policy of continuing a traditional COGI with modest changes and one of attempting the modifications necessary to produce a cost-of-living index that reflects the most comprehensive definition of the "standard of living," there is a wide range of intermediate possibilities. Indeed, starting from either basic approach—the fixed-basket price index or the cost-of-living index—many of the same kinds of questions must be faced.

If one thinks of a "simple" fixed-basket index and a comprehensively defined COLI as opposite ends of a spectrum, it is clear that neither alone provides an operational model with which a CPI can be constructed. For example, in a modern innovative economy, even over a relatively short period of time, the characteristics of a wide range of goods and services are constantly changing. When consumers pay more for a new model of a good, how much of that represents a true price increase and how much a payment for higher quality? Moreover, goods with completely new characteristics, like DVDs, come to the market and gradually take over some or all of the market for other goods. In the case of a COLI, the complexities, lack of data, and deviations from the assumptions of the theory that are sometimes encountered in the real world impose limits on the extent to which its implied objective can be achieved. Attempting to push beyond those limits risks introducing an unacceptable amount of subjectivity and the possibility of significant error into index measurement.

The cost-of-living concept has been used to guide index construction by applying the economic theory of consumption to specific problems. The theory assumes that households act rationally to achieve the highest possible standard of living given their income and the prices they face. They, therefore, allocate their incomes so that, at the margin, goods for which they pay more make a larger contribution to their standard of living. As a result, information about the relative values to consumers of different goods can be inferred from their relative prices. However, since individuals have differing (marginal) evaluations of quality, the conceptual framework for deriving overall quality adjustments from observed differences in market prices raises some difficult issues that have not been fully worked through by the BLS or by academic researchers.

In applying the theory to specific cases, it is essential to examine how closely the underlying assumptions match, or depart from, the actual behavior of consumers and markets in the particular case at hand. To take an important example, the allocation of a consumer's expenditures on medical care is to a major extent determined not by the buyer (the consumer) but by a physician, and those expenditures often come not directly out of the consumer's income, but rather from private or public insurance payments.

DOMAIN OF THE CPI

What goods and service should be covered by the CPI—what should be its domain? In the traditional fixed-weight CPI, and in its counterparts in other countries, the domain is specified by the very definition of the index: it measures changes over time in the cost of purchasing a fixed market basket of goods and services. Its domain, therefore, is the universe of private goods and services. A relatively narrow range of essentially private goods that the government produces and sells in the marketplace, such as entrance fees to national parks or fares on a publicly owned transit system, are included. But no effort is made to estimate a "price" for truly public goods and services, such as national defense or the administration of justice. The adoption of a cost-of-living approach to index construction, however, raises a number of questions about what the index ought to cover beyond what is currently included in the CPI since, in addition to the purchase of private goods and services, a large number of economic, social, and environmental factors clearly have an effect on the standard of living and therefore on the cost of living. (For ease of exposition we use the term "outside conditions" when referring collectively to all of these factors.)

Our panel examined the issue of the appropriate scope or domain of a COLI from several perspectives. If we assume (perhaps unrealistically) that tools could be developed to measure the effects on the cost of living that arise from changes in outside conditions and government actions, we must then ask: Should one include the estimated effects of those conditions on a cost-of-living index that is used for the major purposes served now by the CPI? What role should the BLS play, and what priorities should it give to research devoted to developing experimental measures of the contribution to national output and welfare associated with changes in outside conditions and government programs?

With a very broad definition of what should be included in its domain, a COLI would be adjusted up or down to take account of the positive or negative effects on consumer well-being arising from a wide range of sources outside the marketplace that have not traditionally been considered relevant for inclusion in the CPI. These include, among other elements, the quality of the air, water, and other environmental amenities; the presence or absence of congestion on roads and in neighborhoods; changes in perceptions about personal security associated with trends in the crime rate; the effects of significant climate changes; and increases in longevity arising from broad environmental factors (aside from those associated with specific medical procedures).[5] As noted above, the analytic techniques and statistical tools to measure most of these kinds of effects do not currently exist or, if they have been tried, they are still controversial and speculative.

Public Goods, Publicly Provided Private Goods, and Taxes

The CPI is now based on prices charged for private goods, i.e., goods that are sold to individual households. A few of these goods (e.g., tuition at public colleges, bus fares on city-owned buses, or entrance fees to public parks) are private goods produced by government and sold on an individual basis. The CPI does not include public goods (e.g., national defense, clean air) that are made available freely rather than through individual sales.[6] And yet the increased or decreased availability of those goods does affect living standards. Should the value of some or all types of these public goods be included in the CPI, with the taxes to pay for them treated analogously to prices? When goods made available by government are classified in terms of how similar or different they are from the kinds of private goods currently priced in the CPI, they range across a wide spectrum. At the "nearly private" end are things like the airline ticket taxes charged by the government and used to finance flight control, safety, and other operations of the Federal Aviation Administration. The connection between gasoline taxes and highway construction is somewhat looser, but the taxes do bear some relationship

[5]Increases in longevity associated with specific improvements in medical procedures (e.g., heart bypass surgery) might conceivably be treated as a quality change and reflected as a downward adjustment in the price of medical services. We consider these kinds of situations under "quality change."

[6]Public goods are not defined in terms of who sells them—public or private entities—but by their nature. A public good is one that has two characteristics: if the good is available to one person it is available to all, and one person's consumption of the good does not reduce the amount available to others. Public goods, unlike private ones, cannot be sold to individuals.

to the quantity and quality of the goods delivered. As a general rule, an increase in sales taxes is passed on to consumers in the form of higher prices and shows up as an increase in the CPI. Some economists argue that increases in sales tax rates should be adjusted out of any index whose objective is to measure the cost of living, on grounds that the addition to living costs caused by the higher sales tax rates is offset by the benefits from the additional public goods provided thereby (Nordhaus, 1998).

In the examples noted above, one could argue the existence of a rough connection between what individual consumers pay in taxes and the quantity or quality of the services they receive. But what about pure public goods, such as national defense or law enforcement, the benefits of which are not parceled out to individuals? Every individual, willy-nilly, gets the same "quantity" of national defense. Nevertheless, accepting a broad definition of the standard of living would extend the domain of the index to include the value of public goods with net taxes (i.e., taxes minus transfer payments) treated analogously to prices in a cost-of-living index.

In a similar vein, the enactment of environmental, health, and safety regulations requires businesses to incur extra costs that, when passed on in higher prices, are captured by the CPI. But these regulations reduce environmental damages (broadly defined) and increase consumer welfare. Should an estimate of such additional costs be subtracted from the CPI on grounds that they are balanced by the welfare gains?

Employer-Provided Fringe Benefits

In the United States, employers pay, in part or in full, for a wide range of "in-kind" benefits for their workers, health insurance being the most prominent example. The CPI now excludes from the weights assigned to medical care the value of the health insurance premiums paid by employers. Is that treatment appropriate? More generally, how should the BLS treat in-kind employer fringe benefits in designing the CPI?

SUBSTITUTION

The traditional Laspeyres version of the COGI weights the prices of various items in both the initial (*reference*) and ending (*comparison*) periods by the quantities purchased in the reference period. Considered from the standpoint of an individual household, such an index reflects the percentage increase in expenditures the household would have to incur in order to buy the reference period basket of goods at the new, comparison period prices.[7] But when faced with

[7]Technically, the index is the ratio of the comparison period to the base period expenditures, but the percentage change is simply that ratio minus 1.0 (times 100).

changes in relative prices—pork prices rise a lot and beef prices only a little—consumers try to minimize the effect of the price rise on their living standards: they shift their purchases, buying fewer of the goods whose relative prices have risen and more of those whose relative prices have fallen. This substitution allows them to improve their living standards relative to what would have been the case had they been constrained to maintain their old buying patterns in the face of the price changes. The traditional Laspeyres version of the fixed-basket index takes no account of these possibilities since it simply weights both sets of prices by the reference period quantities. As a consequence, the Laspeyres index tends to *overstate* the rise in the cost of maintaining the reference period's standard of living.

An alternative weighting scheme is the Paasche index, which uses as weights the quantities purchased in the ending, or comparison, period. It measures the percentage difference in expenditures between what it would cost the household to buy the comparison period quantities at the old prices and what it costs at the new ones. But because the comparison period quantities already incorporate the household's substitution in favor of goods whose prices have risen the least, the Paasche index *understates* the cost of maintaining the comparison period's standard of living at the old prices. Equivalently, the Paasche index understates the change in the household's cost of living, as evaluated at the comparison period's standard of living.

If substitution behavior plays a major role in explaining changes in quantities purchased between the reference and comparison periods, goods that have experienced relatively large price increases will tend to receive higher weights in the Laspeyres than in the Paasche index, and the opposite will be true for goods that have experienced relative price decreases. Thus, the Paasche index typically tends to produce a lower estimate of average price increase than the Laspeyres index.

Notice, however, that the Laspeyres index overstates the cost of maintaining the *reference* period's standard of living while the Paasche index understates the cost of maintaining the *comparison* period's standard of living. Under conditions in which those two standards of living are significantly different—due perhaps to the size and pattern of the relative price changes or to changes in the incomes and tastes of consumers between the two periods—it is conceptually possible that the change in quantities is not dominated by the effects of substitution behavior. In that case the Laspeyres might produce a lower estimate of price increase than the Paasche. Nevertheless, empirical studies have shown that when the actual inflation measures produced by the two types of indexes are compared over various historical periods, Laspeyres indexes consistently produce a higher measured rate of inflation than Paasche indexes. This finding is widely interpreted as a testament to the importance of substitution behavior by individual households. Similarly, the magnitude of the differences between the two estimates is usually thought to depend on how much relative prices have changed and how much consumers alter their spending patterns when faced with changes in relative prices.

In the 1920s American economist Irving Fisher proposed what he called an "ideal" index that is formed as the geometric mean (the square root of the product) of Laspeyres and Paasche indexes and thus incorporates information about consumer spending patterns from both the base and comparison periods (see Fisher, 1922). In a 1924 article (not available in English until 1939), Russian economist Alexander Konus formally showed how to construct a cost-of-living index as the ratio of the minimum costs required for a consumer to achieve a given standard of living. He also established the relationships between the Laspeyres index and the cost-of-living index for the reference period's standard of living and between the Paasche index and the cost-of-living index for the comparison period. In 1976 W. Erwin Diewert demonstrated that a class of indexes could be constructed using only information on actual quantities and prices in the two periods that would closely approximate a Konus cost-of-living index (see Chapter 2) for some standard of living intermediate between those in the base and comparison periods and would do so for any pattern of (stable) consumer tastes. He labeled such measures superlative indexes. The Fisher ideal index is one of many possible formulations of a superlative index, all of which involve some form of averaging base period and comparison period weights.

Two Levels of Index Construction Underlying the CPI

Familiarity with several key aspects of the way BLS gathers and combines individual price data into an overall index is necessary for understanding how the substitution issue affects the CPI.[8] The BLS collects roughly 80,000 individual prices each month from over 21,000 retail outlets in various geographic areas around the country. For a few CPI categories, it also collects data from 7,300 housing units. The individual prices are classified into 218 categories (termed *strata*) that represent the various types of goods that consumers buy. From the individual item prices that have been collected, separate price indexes are then computed for each stratum in each area, with weights based on the importance in consumer spending of each of the items included in the stratum.[9] This process is called *lower-level aggregation*. The resulting 218 strata indexes are in turn com-

[8]Though this report details various aspects of CPI construction, we do not provide a unified top-to-bottom description. For this, we recommend the primer in Shapiro and Wilcox (1996:95-102) as well as the documentation on the BLS's CPI web page.

[9]This is an oversimplification. The individual items that are priced in each store are selected by a sampling process designed so that the probability for selection is proportional to the importance of that type of good in overall consumer expenditures. A simple average of the prices in each stratum thus produces what is, in effect, a weighted index for the stratum.

bined into an overall CPI with weights derived from the Consumer Expenditure Survey (CEX), reflecting consumption patterns in the base period (currently 1993-1995). This second stage in the process is referred to as *upper-level aggregation.*

How BLS Deals with Substitution in the CPI

If the goods within a stratum are similar in terms of meeting a particular consumer demand (a characteristic of most but by no means all strata), consumer substitution among individual products is clearly important—for example, Golden Delicious substituted for Gala apples. Beginning in 1999, the BLS replaced arithmetic with geometric averaging ("geomeans") to combine the individual item prices in 129 strata (about 60 percent of the strata in the CPI). Under some rather specific assumptions about the degree of substitution among goods and other matters, the geomeans approach will give the right answer from a cost-of-living standpoint. However, the assumptions about the extent of substitution are unlikely to hold precisely. Moreover, consumer responses to price differences may reflect something other than substitution behavior: for example, a consumer stocks up on particular items when sales occur but does not change the amount of those items purchased per month or per year. Nevertheless, most observers regard the adoption of geomeans as moving the CPI closer to a COLI.

The BLS has announced it will continue to use a Laspeyres approach—base period weights and arithmetic averaging—to combine the individual strata indexes into an overall CPI. In 2002 it will also publish an alternative index that uses a superlative index technique to combine the strata. However, a superlative index requires knowledge of consumer expenditure patterns in real time, and no country's statistical system now produces such data. As a consequence, the superlative indexes that BLS will publish will apply to the period 2 years earlier: the index published in 2002 will measure price changes only through 2000. Recent research studies (e.g., Aizcorbe and Jackman, 1993, and Shapiro and Wilcox, 1997), making comparisons over the period of the mid-1980s to the mid-1990s, suggest that a superlative index would rise at about 0.15 percent a year less than the official CPI using fixed weights at the upper level. Of course, future patterns of inflation may differ, possibly producing a different comparison.

Substantial changes in consumer tastes pose problems for the use and interpretation of either a COGI or a COLI. The weights in a COGI have relevance because the index aims to measure the average price change for the things that people buy. If the pattern of purchases changes substantially, either because of substitution behavior or because tastes have changed, the relevance of a COGI diminishes. Since the composition of people's spending is related to their income, age, and possibly other characteristics, changes in income distribution or demographic balance can also lead to changes in aggregate spending patterns. These sorts of factors may very well contribute to long-term changes in the composition of aggregate purchases though they are unlikely to have large effects over periods of 1 or 2 years. (Below, we outline the aggregation problems that arise in constructing a national price index across individuals and households with differing economic, demographic, and other characteristics.)

A superlative index that, in effect, averages beginning and ending weights has some commonsense appeal in that it takes both states of the world into account. However, the theoretical work that demonstrates that a superlative can provide a close approximation to a measure of the change in the reference period cost of living assumes that changes in purchase patterns stem solely from substitution behavior by households with stable tastes or preferences. To the extent that changes in tastes rather than substitution behavior causes purchase patterns to shift, a superlative index will lose some of its accuracy as a measure of the cost of maintaining the reference period level of living.

In the long run, both consumer tastes and the economic and demographic distributions of households can alter substantially. Comparisons of changes over lengthy historical periods in the price level, and perhaps even more so in the cost of living, are difficult to interpret. In the short to medium run, the issue is whether changes in consumer tastes or substitution behavior tend to be the dominant explanations of changes in household purchasing patterns. As noted above, this is essentially an empirical question, one on which the historical data can shed light, though only inferentially.

Given the time lags required to produce a superlative index, the monthly real-time CPI must instead be calculated from a set number of strata indexes aggregated with fixed weights. Assuming that, due to data constraints, this will be the case for the foreseeable future, what alternatives are available to Congress for making cost-of-living adjustments to social security and other public benefits and for indexing the tax system? Should it continue with the traditional fixed-weight index, recognizing that it is likely to modestly overstate rises in the cost of maintaining the reference period's standard of living? Should it make an initial adjustment based on the fixed-weight index (or, perhaps, on the advance estimate of a superlative index based on historical relationships between a fixed-weight and a superlative index) and then incorporate a correction into the cost-of-living adjustments that are made 2 years later, when the superlative becomes available?

In Chapter 2 the panel discusses the conceptual pros and cons of superlative indexes and other methods of accounting for substitution behavior. We also note alternative approaches to producing a lagged superlative index and techniques for making advance estimates of that index. In Chapter 7 we make recommendations about the use of a superlative index, as well as advance estimates of a superlative, in making cost-of-living adjustments (COLAs) for social security benefits and other public transfer programs.

AGGREGATION

Aggregation Across Consumers

A single price index must somehow represent the average experience of a very heterogeneous population, whose members buy different goods, of different qualities, at different prices, in different kinds of outlets and who exhibit different substitution behavior when relative prices change. If the differences were purely idiosyncratic, so that buying patterns, shopping behavior, and prices paid for the same good did not vary systematically according to whether people are rich or poor, old or young, or by other socioeconomic characteristics, alternative ways of aggregating individuals' varied experiences into a single index would not, for most purposes, pose issues of any great significance. But buying patterns, shopping behavior, and prices paid do vary among different groups by income, age, and possibly other characteristics. And so, during any period in which the prices for the particular kinds and qualities of goods that are especially important to one group rise significantly faster or slower than average, the change in the CPI will under- or overstate the rise in the cost of living for that group.

This heterogeneity raises several issues in the construction of the CPI, which we discuss in Chapter 8. The first of these is peculiar to a COLI and the other two are common to both a COGI and a COLI. First, from the standpoint of measuring consumer substitution behavior, different groups may be more or less inclined to switch their expenditure patterns in the face of changes in relative prices. To the extent that heterogeneity of substitution behavior is systematically related to income and demographic characteristics, the substitution effect incorporated in the overall index can vary with changes in income distribution and demographic balance. (In Chapter 2 we briefly consider the conceptual issues raised by this phenomenon, and in Chapter 8 we provide a fuller exposition.)

The second issue can be framed in the form of several questions: When is a single national index appropriate for the whole population, especially for such purposes as adjusting taxes and social security payments, and when are different indexes for different groups and geographic areas needed? If the latter, how does one collect the kind of data needed to produce subindexes that accurately reflect differences among population groups and locations?

Finally, even if it would be desirable to produce one or more subindexes, a single overall index would still be needed for many purposes—as a measure of national inflation, for example. For purposes of combining the prices of individual goods into the overall national CPI, weights are currently assigned to each good based on aggregate consumer expenditures for the item. Since the spending of a household is positively related to the level of its income, the consumption patterns and prices paid by the rich play a greater role in determining the rate of change in the overall CPI than do those of the poor. Because of their expenditure-based weights, the CPI and the corresponding indexes of virtually every country have been labeled *plutocratic* indexes. In an alternative *democratic* index, the purchasing pattern and prices paid of each household would be given equal weight.

Aggregation and Data Collection

Various empirical studies have combined the basic CPI strata price indexes with expenditure weights reflecting the consumption patterns of a particular group—most notably, the elderly and the poor. In general, these studies have tended to show that the individual group indexes have sometimes risen faster or slower than the overall CPI, but the differences were usually small. Common observation shows, however, that within any category of expenditures (such as a stratum) high-income households buy different items, of different qualities, and often at different stores than do low-income consumers. And the probability that a consumer will purchase a new good during the early period, after which the price often falls, is almost certainly correlated positively with income. Simply reweighting strata prices at the upper level will not show whether price or cost-of-living indexes for the rich, the poor, the elderly, and other subgroups in the population sometimes move differently than the index as a whole or whether a democratic index would behave differently from a plutocratic one.

Testing these possibilities can only occur if data are collected in a way that allows the prices of individual items to be linked to the demographic and other characteristics of those who buy them. But, as explained above, the BLS collects data on price changes for individual items not from households but from retail stores and other sellers. There is no link between the prices of individual items and the economic and demographic characteristics of the consumers who bought them. As a consequence, the current collection system cannot produce the data needed to answer the questions posed above.

CHANGES IN THE QUALITY OF GOODS

Ideally, both a COGI and a COLI ought to be based on changes in the prices of "constant-quality" goods. When a consumer switches to a higher (or lower) quality good, the difference in the price paid for the two goods should be adjusted to remove that part of the difference attributable to the change in quality. If, as is usually the case, the average quality of goods that people purchase improves over time, an index appropriately corrected for quality change will rise more slowly than one measured by the change in unadjusted (nominal) prices. The most frequent criticism of the CPI in recent years, typified by the Boskin commission report, has been that it significantly underestimates the extent of quality improvement in goods and services and therefore overstates the rate of inflation. For many decades, the BLS had been aware of problems posed by goods and services whose quality changed over time and had cautiously extended its use of explicit quality adjustments. More recently, it has begun to move somewhat more aggressively in that direction.

Estimates of "quality bias" in the CPI, such as those of the Boskin commission, that combine specific case studies with subjective extensions to the universe of consumer goods, can contribute to informed discussion about the problem. But, as producers of official statistics, the BLS must walk a difficult line: It must seek to develop and apply techniques for measuring quality change, but it also has to recognize that there are substantial conceptual, statistical, and data availability problems to be solved before it can produce careful and replicable estimates that will be widely accepted.

The adjustment of observed price changes, to eliminate those that reflect changes in the quality of the goods purchased, raises conceptual and measurement issues. Even if there are no measurement problems, one would still have to decide how comprehensively the BLS should pursue the goal of quality adjustment. A frequently cited example arises from improvements in specific medical procedures that reduce mortality. Intuitively, many people think it would be inappropriate to adjust the CPI for such quality improvements and thereby reduce the benefits paid to social security recipients to reflect the estimated monetary value of additional longevity resulting from improved medical procedures (recipients should not be put in the position of living a longer life and enjoying it less). Is there a theoretical foundation for this view? Is it exceptional to this particular issue? Are there grounds for establishing limits on the use of quality adjustments? Are there general principles that can be invoked, or is this the kind of issue that must be settled on a case-by-case basis? In Chapter 2 we consider these issues in general and in Chapter 6 examine the medical care example in more detail.

Measurement

Occasions for quality adjustment continually arise when field agents find that sample items no longer appear on outlet shelves. In these cases, using guidelines established by BLS commodity analysts, the field agent selects a replacement item (which may or may not be new to the market) that is as similar to the old item as possible. About 30 percent of the items being priced disappear each year. In about two-thirds of those cases the field agent can identify a comparable good, which is then treated as if it were the old item. In the other cases the agent identifies a similar but not completely comparable product—e.g., a different version of a dining room table, a lawnmower with a more powerful engine, or a different model of a computer—to price. A quality adjustment to the price of the replacement item must then be made.

Adjustment for differences between the old and the replacement items proceeds along one of two paths. Explicit quality adjustments are currently carried out using either a *cost-based method* or a *hedonic regression technique*. For instance, for three decades the BLS has estimated quality adjustments for the annual model changeover in motor vehicles using the cost-based method. The

BLS collects manufacturers' estimates of the cost increases incurred in adding or changing observable features in the new model. On the assumption that those costs are reflected in reported motor vehicle prices, the prices are adjusted downward to reflect the quality improvements associated with the changed characteristics. This approach is applied not merely to features that are immediately evident, e.g., adding cruise control as a standard feature, but also to more subtle changes, such as the introduction of more corrosion-resistant metals on exposed surfaces. Quality adjustments for changes in the attributes of gasoline are also made with a cost-based technique.

Hedonic techniques offer an alternative method of direct quality adjustment. In this approach, statistical regressions are applied to estimate how much consumers pay for combinations of observable characteristics embodied in goods. Hedonic techniques have long been used in the CPI to make quality adjustments for clothing and rent. In the case of computers, differences in market prices have been linked to differences in speed, memory, reliability, and other performance indicators. These statistical estimates are then used to quality adjust the prices of the newer replacement goods as they are substituted for older items that disappear from the index. As would be expected, the hedonically adjusted price of computers has been falling rapidly for many years. The BLS has recently extended this approach to other goods, including TVs, microwave ovens, audio equipment, and several other types of appliances, and it is experimenting with hedonic techniques for still more products. The recent expansion in hedonic applications has thus far not had a large effect on the CPI. In Chapter 4 we discuss this phenomenon in detail and describe how it and other factors shape our assessment of the potential of hedonic methods in the CPI.

When a noncomparable replacement for an item that has disappeared cannot be explicitly adjusted by one of the methods described above—and this constitutes the large majority of cases—an implicit adjustment is made. In linking the old item price to the new item price (to calculate the price relative for the month in which the replacement takes place), a "pure" price change is imputed for the new replacement item on the basis of the average price change for similar items. Any remaining difference in price between the new and the old items is assumed to represent a quality change and is ignored (i.e., the assumed value of the quality change is adjusted away and that part of the price difference has no effect on the CPI). Subsequently, the new item is priced each month as are all the other items in the index. The large volume of items that are implicitly adjusted each year for quality change suggests the potentially high value of research directed toward developing reliable methods for widening the applicability of explicit quality adjustment techniques.

The rapid growth of research in hedonic techniques (coupled with the lack of research on alternatives) suggests that they may be the most promising approach for exploiting data on differences in the observable characteristics of similar goods to generate measures of quality change. But, in practice, their application is fraught with difficult problems of data availability, concept, and statistical method. In Chapter 4 the panel considers in detail the proper balance between the pursuit of improved and expanded direct quality measurement and a rigorous program of selection, testing, and experimentation before implementation.

Medical Care

Pricing medical care embodies the most difficult quality-related problems associated with constructing a cost-of-living index, and the panel gives it special attention. The health services sector is a highly complex industry characterized by rapid advances in technology and a continuing stream of new techniques, propelled in part by substantial federal support for research. The industry also possesses a number of special attributes that make the quantity and quality of its output difficult to define. Complications that BLS must deal with include the high variation in prices paid by consumers for equivalent services, the definition of the medical "good" being purchased (e.g., is it a procedure or treatment or the medical inputs?), the involvement of insurers and government in transactions, and the pricing of risk.

Some important improvements in accounting for changes in the quality of medical care have recently been introduced by redefining the units of service for which prices are collected in the CPI. For example, to price hospital services, the BLS has begun to collect prices for the treatment of particular diagnoses or illnesses in place of the earlier practice of pricing inputs, such as days in the hospital or operating room charges. This has led to substantial improvement in the ability of the index to capture advances in medical technology that reduce the cost of treating a given illness. Another major improvement has been in the BLS treatment of generic drugs, in effect counting the difference between their price and those of the same-molecule branded drugs as a price reduction.

These improvements, however, do not directly deal with advances in the *quality* of a given treatment or procedure, that is, changes in ultimate outcomes—lower mortality and morbidity, fewer undesirable side effects, less pain or trauma, a better quality of life, and so forth. (We touched on the conceptual aspects of this issue above.) To pursue the cost-of-living concept in the pricing of medical care to its logical conclusion, one would need to put a monetary value on how the outcomes of specific medical procedures affect consumer well-being. This obviously poses enormous conceptual and measurement issues. In Chapter 6 the panel considers these conceptual and measurement issues and makes a number of recommendations about the treatment of medical care in the CPI.

NEW GOODS

In the marketplace, there is rarely a sharp dividing line separating a new good from an existing one whose quality has been improved. As described above, monthly BLS price collection procedures continually lead to the pricing of replacements for discontinued items. These replacements, chosen from the same CPI item category, exhibit only incremental quality changes compared to the old products. Goods also appear that are not replacements for any of the items being priced in the current sample but that can be assigned to existing CPI categories. These items may enter the index when retail outlets (and the goods they sell) are rotated in the BLS sample every few years. In addition, though, new goods and services appear that are different enough from existing ones that there is no place for them in any of the CPI categories. Hence, without an explicit decision to add the new item to the list of categories of goods, their impact on general price trends can go unmeasured: VCRs and cell phones are examples of such items that did not enter the index for many years after they appeared on the market.

There are two ways in which the CPI, when used as a measure of changes in living costs, might be biased by the appearance of new goods. First, many of those who advocate a cost-of-living approach to index construction argue that the consumption welfare effects that accompany the appearance of a new good are missed. Under traditional procedures, new goods, both those that enter through item rotation or after item reclassification, are linked into the CPI in such a way that their introduction, in itself, has no effect on the level of the index. But, for example, there were some consumers who found wireless phones so attractive that, rather than do without, they would have paid a higher price than that prevailing when the phones first entered the index (or even at the time of market introduction). An increase in the standard of living of these consumers was made possible by the introduction of the good, but such effects on the cost of living are not captured by simply linking in the new good without making a specific adjustment for the improvement.

The magnitude of the improvement could, in theory, be estimated using detailed market data on the prices of new goods and the quantities sold to infer how much consumers would have paid for the new good rather than doing without. This estimate would then be used to calculate the relative welfare gain associated with the introduction and subsequent consumption of the good. That gain, in turn, would be incorporated into the index as a price decrease. The conceptual and measurement issues at stake here are addressed in Chapter 5. On a measurement basis, is it possible with available econometric tools to estimate the welfare gain with sufficient accuracy and transparency to warrant its use in the way that would be required? Conceptually, even if reliable measurement should become feasible, should the new good welfare effect be treated as if it were just a price decrease and entered into the consumer price index as such? In Chapter 5 the panel examines both the feasibility and advisability, in constructing a COLI, of employing econometric techniques to estimate and incorporate into the index the “virtual” price reduction that accompanies a new good’s appearance in the market.

The second problem posed by the introduction of new goods revolves around the timing of their incorporation into the CPI. New goods are often introduced in the market at a high price, and they have a low volume of sales. Then for some time the price will tend to fall and the market will expand rapidly, until the new good becomes a mature, established product. For some new goods, however, producers follow a different strategy, introducing the new good at a low price in order to promote high initial sales and make the good more widely known to consumers. In this case, as the market expands, the price rises.

If a new good with an initially falling price and rising sales volume—e.g., cell phones—is incorporated into the index only after a long delay, the period of falling prices will be missed and the overall price index will be biased upward. If, in contrast, the new good is incorporated into a fixed-weight index not long after its introduction, the index will reflect most of the decline that occurs early in its price cycle. Early sales are likely to be modest, and if the weights in the index remain fixed for some time, the declining prices will carry only a very small weight, and again the overall index will be biased upward, although to a smaller degree. If the new good is one whose price initially rises rapidly, the opposite results occur: whether the new good is incorporated late or early, the index will be biased downward. With either rising or falling prices, the faster the price change and the more rapid the sales growth, the larger the bias. Working on the assumption that most new goods experience a period of large price declines, the Boskin commission and other observers have attributed an upward bias to the CPI stemming principally from late introduction of many new goods.

The BLS is making changes that will reduce the magnitude of the problem: much more rapid updating of the index weighting system (every 2 years instead of every 10 years or longer) and a faster turnover of the stores from which it collects prices. But the issue of how best to deal with the introduction of new goods will remain an important problem, and handling it will require tradeoffs among competing objectives. Very early introduction of new goods raises the danger of incorporating "duds" into the index—such as Betamax VCRs or 10-inch video disks. And speeding up the rotation of retail stores (see below) into the sample is quite expensive. Chapter 5 examines and makes recommendations about the procedures and criteria that BLS should use for linking the prices of new goods into the index in a timely and nondisruptive way.

To the extent that new goods are disproportionately purchased by the affluent in the early stages of a product's cycle, distributional consequences will arise as these goods are brought into the index. If the relative prices of new goods fall, the growth of the overall index will, on this account at least, tend to understate the inflation faced by low- and middle-income consumers.

OUTLET SUBSTITUTION

When purchasing a good at a particular store, consumers are buying not only the good itself but also a package that includes the quality of the shopping experience associated with the store—the services provided, its locational convenience (or inconvenience), the variety of products available, its return policy, and so forth.

The BLS gradually rotates the sample of retail outlets from which it collects prices, and over time the new samples capture the changing mix of outlets patronized by the buying public. Under current BLS procedures, when new stores enter the sample, *all* of the difference between an item's price at the old outlet and its price at the new outlet is implicitly assumed to reflect differences in the "quality" of the shopping experience; none of it shows up as a pure price change. This practice can produce a bias if price variation across outlets allows consumers, by altering their shopping behavior, to reduce their consumption costs, adjusted for the quality of the shopping experience, in a way not detected by the CPI.

The clearest evidence that consumers are reducing costs is indicated by the increase, for a number of years now, in the market share of high-volume, low-price retail outlets. The prices paid at these outlets are often substantially lower than those in conventional stores. Under current assumptions, lower prices are being fully offset by a lower quality shopping experience—as represented by "goods," such as convenience or a store's return and exchange policy, that are omitted from the CPI. The fact that the market share of the low-price discounters has been steadily growing, however, implies that even after "quality adjustment" the prices at those stores are lower than elsewhere. As new outlets open, consumers in the area gradually change their shopping behavior and take advantage of the lower quality-adjusted prices. At the same time, a minority of consumers who would have preferred to continue shopping at traditional stores find them to be driven out of business by the new outlets, and those consumers suffer an increase in their cost of living.

From the standpoint of a cost-of-living index, the current procedure for handling sample rotation among outlets misses some of the decline in living costs associated with this ongoing shift in consumer purchasing patterns. The few empirical studies that have been done, however, suggest the resulting effect on the overall CPI is relatively small. Identifying and quantifying "quality" differences in the shopping experience offered by different types of outlets, in order to make a proper adjustment for what is happening, poses measurement difficulties for which satisfactory answers are not currently available. In Chapter 5 the panel makes recommendations to the BLS about what to do in the current absence of methods for making outlet-quality adjustments and suggests what priority it should give to research and development efforts in this area.

STOCKS AND FLOWS

Many longer-lived goods, such as owner-occupied housing, automobiles, and appliances are durable capital goods that gradually yield a flow of consumer services over a period of years. Even some "nondurable" goods (e.g., men's suits) often provide services for some years. While there is no consensus among experts and national statistical agencies, a powerful argument can be made that the CPI in any period should measure the price of consumption in that period. In that case the index ought to include not the price that consumers pay for a durable good at the time of purchase, but the estimated cost of the services rendered during the subsequent periods in which the good is owned. In the case of owner-occupied housing, the BLS has, since 1983, estimated ("imputed") a monthly price of the services rendered by that housing (i.e., shelter services). In all other cases the price of a consumer capital good is entered in the month when it is purchased.

There are two ways of estimating the price of the flow of services from consumer capital goods. The first is to estimate the "user" cost of the service (i.e., the cost to a consumer of buying a good at the beginning of a period, using its services during the period, and selling it at the end). Among many other difficulties, the user cost concept would deduct a capital gain (or add a loss) realized by the owner in the transaction. Housing prices can be volatile, and this method of estimating the cost of shelter services can sometimes produce the paradoxical result that just as housing prices are rising rapidly the user cost estimate will show a decrease in the shelter component of the CPI. An alternative method of estimating the cost of housing services, and the one currently used by the BLS, is to try to find a sample of rental housing that is equivalent in quality to owner-occupied housing and use the change in rents within that sample as a measure of changes in the cost of owner-occupied shelter services. This method also poses measurement difficulties of various kinds, including the difficulty of finding, in each geographical area for which the BLS collects prices, a representative sample of rental housing that is truly equivalent to owner-occupied housing.

While a strong conceptual case can be made for incorporating into any index of the price of consumption an estimate of changes in the monthly cost of the services provided by owner-occupied housing and other long-lived capital goods (see Chapter 2), there are some countervailing considerations. For potential low-income home buyers, who are cash constrained and at the margin of acceptability for mortgage loans, sharp changes in housing prices can have a greater impact on their cost of living than would be reflected in cost estimates of the flow of services. And it has been argued that, since the CPI is used as a guide for monetary policy, it should reflect volatile changes in consumer asset prices, as is the practice in a number of other countries (Goodhart, 2001). (We point out in Chapter 7, however, that governments and central banks have plentiful staff resources to consider the effect of changes in assets prices as they relate to monetary policy, even if they are not included in the official index of consumer prices.)

The practices of other countries with respect to housing vary widely. Because of the conceptual and measurement difficulties, many countries simply exclude owner-occupied housing from their consumer price index altogether. A few use the net acquisition price of owner-occupied housing; a few others, in addition to the United States, use equivalent rent; some use a cash flow approach—the sum of down payments, mortgage principal, and interest payments, or some subset of these items; and still others use the sum of depreciation at replacement cost value and mortgage interest payments as a (very) rough measure of user cost. No other CPIs extend the "flow of service" concept to the pricing of other consumer capital goods, although that course has sometimes been proposed to the BLS for the pricing of automobiles (Boskin et al., 1996:recommendation 8).

This brief sketch gives an indication of the wide range of conceptual and measurement issues that arise in determining how to measure changes in the prices of consumer capital goods and their services. However, given both the importance of covering the many other subjects the panel was asked to address and the constraints of time, we did not examine these issues in any depth or formulate recommendations on them.

INDEX PURPOSE AND DESIGN

It is clear that there are many difficult issues to resolve in designing a consumer price index or indexes. How some of the decisions should be made may be dictated by the purpose for which the index is to be used—the index designer needs to keep the index user in mind. Although no restrictions need be placed on the number or variety of research and experimental indexes, a desire to avoid public confusion may constrain the number of separate official indexes that are published. More importantly, there are inherent limits on the extent to which it is possible to match the design of an official index to a particular purpose, limits that are often dictated by what can reliably be measured. As a consequence, public policy makers and private users of indexes need to be aware of the extent to which a particular price index does not measure exactly what they want measured. In fact, considering its range of applications, it is probably rare when the CPI does measure exactly what is needed. The CPI is currently used in many ways, including:

- as a compensation measure to calculate how much is needed to reimburse recipients of social security and other public transfer payments for changes in the cost of living, and for formal or informal use in wage setting;
- for inflation indexation in private contracts;
- as a measure with which to index the income tax system to keep it inflation neutral;
- as a measure of inflation in inflation-indexed U.S. Treasury bonds;
- as an output deflator for separating changes in GDP and its components into changes in prices and changes in real output; and
- as an inflation yardstick for the Federal Reserve Board and other macroeconomic policy makers.

Throughout the report, we explain how alternative choices in index design could affect use of the index for each of these purposes. Where relevant, we spell out the public policy consequences of using alternative index designs for making cost-of-living adjustments in public transfer programs, indexing the tax system, and for other purposes. And while we make no recommendations on the subject, we explore the public policy implications of using a wage instead of a price index for escalating social security and other benefits.

DATA COLLECTION

The CPI is constructed from several sample-based sources: the Consumer Expenditure Survey (CEX), the Point of Purchase Survey, the Commodities and Services Survey, and the CPI Housing Survey. There are two distinct approaches that could be taken to change the data collection apparatus: the first would be to improve each survey component, assuming that the basic structure will remain in place; the second would involve redesigning, from the ground up, the entire data collection apparatus. The panel considered options under both approaches.

The panel addresses questions about both the accuracy and precision of the CEX, which is the primary tool for establishing the CPI upper-level weights (at the basic, 218-item level). The panel's foremost concern is with the extent of bias in the CEX which, in turn, affects the accuracy of CPI expenditure category weights. In this context, it is worth evaluating the pros and cons of using alternative data sources—such as those used to produce per-capita personal consumption expenditures for the national accounts—for deriving the national CPI upper-level weights.

Assuming that the CEX is the appropriate source for generating CPI weights, there is the question of optimal survey sample size. The report addresses these questions and provides some calculations that indicate the relationship between sample size and the precision (variance) of derived item strata weights. Of course, precision requirements set for the national index will yield very different answers than similar ones for component indexes or if population subindexes are desired.

In addition to questions about sample size and accuracy, there are a number of issues that involve assessing the information content of questionnaires and the general structure of the CEX. There are also questions about how the mode of data collection might be modified to take advantage of new computer-based data collection methods, whether all expenditures for all item categories should be collected from all households surveyed (or just some from each), and what processing system is required for the CEX in order to expedite development of a superlative index. Answers to all these questions hinge on the types of indexes that BLS will be called on to produce.

A second major survey input to the CPI is the Point of Purchase Survey (POPS), which is used to determine which outlets BLS data collectors visit to record price changes of index items. The POPS produces outlet-specific expenditure information for item categories so that a sample of those outlets can be selected with a probability proportional to consumer use. The POPS is needed because the CEX does not ask consumers where they purchased goods. Given that there is some functional overlap between the CEX and POPS, the panel considers, among other things, the possibility of merging or better coordinating these two surveys. The report also reviews the CPI's Commodities and Services Survey, a longitudinal survey that tracks changes in price quotes for most CPI sampled consumer items over time.

Since most options for improving CPI support data are expensive, particularly those involving the household surveys, and because there is methodological inflexibility under the current system, it is also worth considering entirely new data production alternatives. Therefore, in Chapter 9 the panel considers (1) the tradeoffs associated with changing to PCE-based expenditure weighting; (2) the possibility of combining POPS and CEX into an integrated survey that contains expenditure and outlet-use data at detailed product levels, along with household demographic information needed for subgroup indexes; and (3) what might be gained from moving toward scanner-based collection systems, which could be used to improve the existing surveys or as a component of an alternative.

2

Conceptual Foundations for Price and Cost-of-Living Indexes

For much of its life, the Consumer Price Index (CPI) was based on the idea of comparing the costs of a fixed bundle or “basket” of goods; this concept leads to what we call a “basket price index” or “cost-of-goods index” (COGI). However, beginning with the Stigler commission report (1961), there has been an increasing emphasis on thinking about the CPI as a cost-of-living index (COLI). Indeed, the “overarching recommendation” of the Boskin et al. (1996) report was that the Bureau of Labor Statistics (BLS) should try to make the CPI approximate a COLI as closely as possible.

This chapter lays out the theory behind both the COLI and the COGI in a form that will serve as a basis for the discussion of specific topics in the chapters that follow. We start from the underlying ideas in their simplest form and then work through a series of practical and conceptual issues, many of which are covered in detail in subsequent chapters. Consideration of each of these issues serves to sharpen and elaborate the concepts in ways that are necessary if either a COGI or a COLI is to fulfill the many sometimes conflicting demands that are placed on a measure like the CPI.

The basket price and cost-of-living approaches to index construction are conceptually quite different. Nonetheless, for many (perhaps even most) purposes, the distinctions are less important than they might seem. In particular, for most of the issues that we discuss in this chapter and in the report more broadly, there are close parallels between the two approaches. In consequence, the two approaches have always drawn on one another, so that a COGI has often been modified to make it more like a COLI and vice versa. Even when the CPI was defined in terms of the basket approach, the BLS kept the cost-of-living concept in mind when making decisions about index methodology. Similarly, and as we show in this chapter, there are strong arguments for replacing a pure cost-of-living index by what is known as a "conditional" cost-of-living index which, in some respects, brings the COLI concept closer to a basket price index. In consequence of this two-way traffic, sharp differences in operating practices are uncommon. Most practical indexes or procedures for computing the CPI can be justified in terms of both approaches, though the arguments will often differ.

Still, the distinctions are important. Indexes derived from the cost-of-living approach allow for the fact that, when relative prices change, consumers tend to substitute toward the relatively cheaper items. Basket price indexes simply measure the cost of a fixed bundle of goods and are not designed with substitution in mind, notwithstanding the fact that a suitable choice of basket sometimes allows them to be interpreted as cost-of-living index numbers. The language is also important, at least in the eyes of policy makers and the public, even if those who make the index know that the formulas are the same. How the CPI is labeled affects the way that people think about it and may influence the credibility of the measure in the view of those who are affected by it. A useful analogy is perhaps the social security system, whose legitimacy in the eyes of many is enhanced by the perception that it is a fund, "the social security trust fund," out of which they draw during retirement the contributions and interest on savings they made when working. Note too that the words "price index" and "cost of living" do not have the same connotation in common speech. The coincidence of the two ideas is relatively recent even among economists. The relationship between the "cost of things" and the "cost of living" needs to be thought about seriously. The argument that, under some circumstances, they are the same thing needs to be carefully argued and clearly laid out. The possibility that under other circumstances they are not the same thing also needs to be kept in mind.

A clear conception of what one is trying to measure also serves as a touchstone to help resolve the many practical issues of price index construction that come up as an economy changes. Theory is the authority to which index designers appeal when it is hard to choose among alternative practical procedures or deal with new developments. Theory can be thought of as a constitution whose wise (if occasionally rather general or even delphic) principles can be applied to settle questions and disputes. A recent practical example is the adoption by the BLS of a new "geometric means" procedure for combining prices at the most detailed level of commodity disaggregation. It seems unlikely that this change would have been adopted without the shift of conceptual basis toward a COLI that followed the Boskin et al. (1996) commission report. An even more important and more difficult issue, which pervades the present report, is how to allow for quality change in the CPI. Here again, the cost-of-living framework has promise for helping design good practical procedures.

Even so, it is important not to expect too much of any conceptual framework. In the words of Commissioner Abraham (Bureau of Labor Statistics, 1997c), "the cost-of-living index is a theoretical construct, however, not a single or straightforward index formula readily amenable to practical use." The consumer price index means different things to different people, and it is used in many, possibly contradictory, ways. An index that is good for one purpose will not always be good for another. It is a lot to ask of any one measure that it provide a general indicator of the level of prices in the country as a whole, that it yield an accurate measure of how much Congress intended social security recipients to be compensated for price changes, that it should hold constant the "real" rate of income taxation, that it should be an appropriate escalator for the poverty line as well as for a host of government, business, and private contracts involving a wide range of people, and that it should be useful to the Federal Reserve Board for setting monetary policy. Each purpose leads to a somewhat different conceptual framework. And as the panel's own discussions have made clear, some of the most difficult issues, such as what to do about quality change, particularly but not exclusively in the provision of medical care, do not seem to be adequately handled by any of the conceptual frameworks currently available, or at least not in a way that commands widespread assent.

The remainder of the chapter has four major sections and a technical note. The first section provides some preliminary definitions of what is meant by a cost-of-living index and by a basket price index. It also lays out some practical considerations that limit the usefulness of at least some of the concepts that might be attractive in theory. The second major section presents the theory of the cost-of-living index. The COLI is rooted in a simple economic theory of consumer behavior that is the workhorse for much practical economic discussion and policy making. For economists, the discussion in the first part of this section will be familiar, although as became apparent in the public discussions on the CPI, this theory is often only vaguely understood. It is often criticized for defects or praised for charms which it may or may not possess. But even when fully understood, the theory is not immune to criticism of its behavioral assumptions, its empirical predictions, nor its approach to well-being. These criticisms, many of which derive from the literature in psychology, are also reviewed in this section.

The next section presents a discussion of specific topics, such as how to relate price indexes for individuals or groups of people to indexes for the nation, how to choose the prices that are appropriately included in a consumer price index, how to use price indexes to compensate people or groups of people for price change, and how to adjust price indexes for changes in the quality of goods. For each topic, we show how the different conceptual approaches are relevant, and how concrete application leads to sharpening and redefinition of the concepts. Although almost all of the topics are dealt with again in subsequent chapters, they need to be covered here in order to develop the conceptual apparatus that will later be used. We present conclusions in the third major section.

SETTING THE STAGE: WHAT ARE PRICE INDEXES?

In Chapter 1 we identified two distinct conceptual bases for the CPI that have dominated the public discussion and are the most relevant for the work of the panel: the fixed-basket approach, which was long the basis for the CPI in the United States, and the cost-of-living approach, which was strongly recommended in the Boskin report. In somewhat more detail, the two approaches are:

- The fixed-basket approach. A basket of goods is priced in each period and the price index calculated as the cost of the basket in the comparison period divided by the cost of the basket in the reference period. Because the goods in the basket are fixed across the comparison, we call this a "cost-of-goods index" or COGI. The relevant basket for a national CPI is the set of all goods and services bought by consumers in the United States during a base period, and the prices are the market prices paid for those goods and services during the reference and comparison periods. The reference period often coincides with the base period but need not necessarily do so.
- The cost-of-living approach, sometimes referred to as the "economic" approach. Prices in the comparison and reference period are compared using the ratio of the cost of living in the two periods. Instead of comparing the costs of two baskets of goods, the comparison is between the cost of maintaining the same standard of living in the comparison and reference periods. Exactly what is meant by the standard of living and the cost of living are matters that we discuss and, as we show in the next section, accurate evaluation of a COLI requires not only data on quantities and prices but also knowledge of how consumers respond to changes in incomes and prices. In practice, therefore, adopting the COLI as a conceptual basis does not imply using an exact COLI for the CPI but, instead, using one of a number of feasible approximations. The nature of these approximations, as well as their relationship to an exact COLI, is developed in the next section. But as is the case for basket price indexes, the calculation of the price index starts from a basket, or baskets, of goods and from lists of prices in the reference and comparison periods.

The research literature contains a number of other approaches to price indexes. One of the most important is the "test" approach associated with Irving Fisher. According to this, price indexes are judged according to a number of desirable "tests" that price indexes should ideally satisfy. For example, one test that is satisfied by all sensible price indexes is that, if all the prices going into the index are doubled, the index doubles too. Another framework is provided by the stochastic approach, in which it is assumed that there is some underlying but unobservable price level, around which the prices of individual goods and services are randomly distributed. For the purposes of this report and in the current historical situation, the COGI and COLI approaches are the obvious contenders to be *the* conceptual basis. Nevertheless, both the test and stochastic approaches have intuitive appeal, and they are often useful for illuminating the properties of specific indexes or for dealing with technical issues that are not otherwise easily addressed. Given the many purposes to which price indexes are put, it is often helpful to have more than one conceptual framework.

It is useful at the outset to put these concepts in context. Note first that general discussions of price index numbers are often cast in terms of two situations, usually labeled the reference and the comparison. The two situations might be geographical locations—Los Angeles versus New York, or the United States versus India—but in the case of the CPI, the two situations are different time periods, typically a reference period that is held fixed for a number of years, and a series of later periods ending with the "current" period, which in practice is a period in the recent past. The CPI is produced on a schedule that, together with the availability of the underlying data, puts constraints on what is possible. In particular, the BLS is able to collect data on prices with a much shorter lag than is possible for collecting data on the quantities of items purchased. The monthly CPI is published quickly: for example, the October 1999 CPI was published on November 17, 1999. However, the basket that was priced for this CPI came from Consumer Expenditure Surveys that collected data during 1993, 1994, and 1995 and was therefore a little more than 5 years old on average. In the past, the basket had been updated infrequently, only once a decade. Although the BLS has undertaken to shorten the time between updates, baskets available for pricing are always likely to be several years old, at least in the absence of some radical new technology, such as the extensive use of scanner data or automatic computer-based reporting of sales from retailers.

The availability of data places limits on what can be achieved within any given conceptual approach to the CPI. To stay with the current production schedule, a basket price index approach must use a base that is considerably earlier than the comparison period. The BLS can compute basket indexes relative to any base period in which quantity (or expenditure) data are available, but no later. Any methodology that requires a current basket to compute the current price index can generate price indexes only with a lag of 2-3 years. To see the implications of this, think about the two most familiar forms of the basket price index, the Laspeyres price index and the Paasche price index. In the Laspeyres, the base period basket is priced in both base and current periods—the base period is also the reference period—and the price index is the ratio of the basket's cost at current prices and at base period prices. No information is required on the current basket. The Paasche index, by contrast, works with the current basket; it is defined as the ratio of the cost of the current basket at current and reference period prices. Because quantity information comes more slowly than price information, the production of a Paasche price index requires a longer time lag than does the production of a Laspeyres. A Laspeyres index can be thought of as an approximation to a COLI. Better approximations are possible using information on both the reference period and current period baskets. One such approximation is Fisher's "ideal" index, which is the geometric mean of the Paasche and the Laspeyres. But the Fisher index cannot be produced any faster than its least timely component, so its production lag is as great as that of the Paasche index. As we show below, there are other indexes that may be more timely than the Fisher ideal, but that still do a better job of approximating a COLI than does the Laspeyres.

For constructing COLI price indexes, as for other economic statistics, there is a tradeoff between timeliness and accuracy. For some purposes, a longer wait is an acceptable price to pay for greater accuracy and closer conformity to a theoretically desirable concept. Moreover, technical and statistical innovations in data collection—such as scanner data—will likely reduce the lag in the future, at least for some components of the CPI. (Bar codes for rent, cars, haircuts, and medical care are still some way off!) As always, much depends on the purpose to which the CPI is to be put. Policy makers and many others value rapid availability, so the BLS puts a good deal of weight on timely production of the index. An index for compensating social security beneficiaries, or for adjusting income tax brackets, can presumably wait longer, though probably not 3 years.

THE THEORY OF PRICE INDEXES AND ITS CRITICS

There is a large literature in economics on the theory of price indexes. We present no more than is needed for use in this report. Much of the relevant literature makes free use of mathematics. While it is possible to give a useful verbal discussion of the main issues, clarity requires some use of formulas. We provide a verbal discussion in the main text and support the argument with a technical note that contains the most important equations. We begin with the basket price index because the ideas are more straightforward and because Laspeyres and Paasche indexes provide useful starting points for thinking about cost-of-living indexes.

Basket Price Indexes

A price index is needed because there are many goods and services in the economy, each with its own price, and each with its own rate of change in price. If all prices in the economy changed at the same rate, there would be no need to construct an index because the ratio of prices in the two periods would be the same for all goods, and any one would summarize all others. Price indexes are needed because prices do not move at the same rate. Because relative prices change over time, a way must be found to combine (or aggregate) all the changes into a reasonable measure of overall price change. This aggregation needs to take into account how much is spent on each good, so that price changes for goods on which more is spent get greater weight. One simple way to do so is to calculate a basket price index.

Beginning with a list of actual purchases in the base period, the total cost of this basket in the reference period can be calculated, as can its total cost in the current (or comparison) period. The ratio of these two costs is a basket price index or cost-of-goods index. If the basket is the list of goods actually purchased in the base period, this is a Laspeyres price index. If the current basket is used as the base and the price index is the ratio of the current to reference period costs of the current period basket, the result is a Paasche price index. Because the ratio of prices in the comparison to reference period differs from one good to another and because the baskets purchased in the two periods are generally different, the Laspeyres and the Paasche indexes are generally not the same. In principle, one could calculate a price index from any basket—for example, one at any point between the current and the base baskets. The relationship between various indexes cannot be known without information about how the baskets are generated and how quantity is related to price. In particular, it is not true, although it is often so claimed, that the Laspeyres must necessarily be greater than the Paasche, though this is usually the case in practice.

As we have noted, the Laspeyres index has an important practical advantage: once base quantities have been set, a Laspeyres index can be produced on the same schedule as prices are collected. The idea of continuously repricing a fixed basket is easily explained even to nonspecialists and corresponds well to what most people think of as a price index. The Laspeyres price index is the concept that is most frequently used by statistical offices around the world.

When the Laspeyres index is used to calculate a national CPI, the basket to be repriced is usually the total purchases of each good by all consumers in the country during the base period. But it is also possible to think about baskets purchased by various subsets of the population. Groups might be defined by region, to derive a regional price index; by age, to look at a price index for the elderly; or by income levels, to construct separate price indexes for the rich and the poor. Indeed, there is nothing in principle to stop us from thinking about a Laspeyres index for each individual in the economy. Different people spend their money in different ways, so that each is affected differently by changes in prices. For example, those who commute long distances to work are seriously affected by an increase in energy prices, while those who walk are not; smokers are affected by an increase in the price of cigarettes; nonsmokers are not.

Two important issues are raised by thinking about price indexes for groups or for individuals. First, not only do different people buy different baskets of goods, but different people often pay different prices for the same goods. Second, if one constructs (say) a national Laspeyres index and an individual Laspeyres index for each person in the country, how does one relate to the other? In particular, is the national price index an average of the individual price indexes? Both of these issues arise repeatedly throughout the report, so it is useful to discuss both at the outset.

The second issue, the aggregation of individual price indexes to get a national price index, is more easily dealt with if one assumes away the first issue and pretends that everyone in the economy pays the same price for everything. In a well-integrated, low-transport-cost economy like that of the United States, the assumption works well for many consumer goods, but there are obvious exceptions, of which shelter and medical care are almost certainly the most important. Nevertheless, imagine an economy in which everyone faces the same prices, and differs only in the total amount they spend and in the how they divide it among different goods. In the individual Laspeyres indexes, prices are weighted in proportion to individual expenditures, while in the national Laspeyres index, prices are weighted by aggregate expenditures. It is useful to think of the Laspeyres index as a weighted average of the "price relatives," which are the ratios of current to reference prices for each good. The Laspeyres weights are the shares of each good in total expenditure, whether for the individual family or the nation (for the equations, see the "Technical Note" at the end of the chapter). The national Laspeyres then differs from the individual Laspeyres only in the weights used: For the national index, the weights are the shares of each good in national total expenditure; for each individual family's index, the weights are the shares of each good in the family's total expenditure.

Is the national price index an average of the price indexes for each family? Yes, but it is a weighted average, not a simple average. Because the national index uses national expenditures as weights, and because families who spend more contribute more to the national expenditure than do families who spend less, those who spend more get a higher weight in the national index. Indeed, the national Laspeyres price index is a weighted average of the individual families' Laspeyres price indexes, with weights equal to the total expenditure on all goods by each family. This weighting was termed *plutocratic* by Prais (1959); the rich—or at least the rich who consume more—get a higher weight in the price index than do the poor. The obvious alternative, in which each family makes an equal contribution to the index, is called the *democratic* price index and would be calculated from the individual price indexes by simple averaging. In general, the democratic and plutocratic price indexes differ, and they will move differently whenever the prices of goods consumed by different income groups change at different rates. A recent example is the price of cigarettes, which makes up a larger share of the budgets of people with lower incomes. Increases in cigarette prices increase a democratic price index by more than a plutocratic index.

The Laspeyres price indexes produced by statistical offices around the world are always plutocratic, not democratic, indexes. Elsewhere in the report, we argue that, were it possible to calculate a democratic price index at reasonable cost, it should be preferred to a plutocratic index for many purposes, especially those to do with compensation. But we also argue that there are real practical difficulties in constructing the democratic index. Those difficulties help explain the universal reliance on plutocratic indexes.

The relationship between national and individual price indexes is much murkier if one allows for the fact that different people often pay different prices for the same product (price heterogeneity). It is still straightforward to imagine a price index for each family; one could simply take a family's basket in a base year and price it at the reference and current market prices paid by that family. The difficulty arises at the national level; an aggregate national bundle is priced, not at the specific prices that individuals actually pay, but at prices that are averaged over all the prices paid. But such an index is not related to the individual indexes in any predictable way; in particular, the national index is no longer a weighted average of the individual indexes. More generally, it is hard to derive any good rationale for the aggregate index when price heterogeneity is important. As always, one remedy is to assume away the problem, which, in effect, is what the BLS currently does. It is a good solution if price heterogeneity is not very important, except for a few goods such as shelter and medical care, both of which require special treatment in any case. If price heterogeneity is important, or if technical change (such as the Internet) allows even greater possibilities in the future than now for firms to charge different prices to different people, there is no good alternative to working at the individual level, at least conceptually. Price indexes would be calculated for each household, or at least for a random sample of households from the population, and averaged to obtain the national index. This radical departure from current practice has many attractions but is almost certainly not feasible given current technology for data collection. We explore these matters further in Chapter 8 on aggregating across households.

When thinking about aggregation from households to nations, it is also worth giving consideration to the opposite process, that of disaggregating households into their individual members. We have used the terms individuals, families, and households more or less interchangeably, contrasting them with national aggregates. Yet multiperson households are themselves collections of individuals whose interests do not always coincide. In the next subsection we deal with the textbook "consumer," who is assumed to make consistent choices within the available opportunities. If such an account is applied to a family or household, it supposes a unity of purpose and preference that might not be the case in practice. Recent research in economics has gone some way to looking inside the household, thinking about ways to model and to recognize non-unitary behavior. Nevertheless, none of this work has been directed toward the construction of price indexes, and in this report we work within the older tradition of regarding households and families as the basic units of the economy.

Cost-of-Living Indexes

Cost-of-living indexes compare prices, not by looking at the cost of a *basket* at different sets of prices but at the cost of *living* at different sets of prices. Basket price indexes work with the cost of specific goods and services; cost-of-living indexes work with the cost of "living." Measuring the cost of living requires one to compare different baskets of goods and to say when they yield the same "standard of living." This is done by using the economic theory of consumer behavior. Consumers always think that more goods are better (or at least no worse) than less, and they can rank different bundles of goods consistently. Consumers' choices are governed by preferences but constrained by the market prices of goods, as well as by the amount of money they have to spend. Subject to these constraints, each consumer chooses the best (most preferred) basket among all the baskets that are affordable. The standard of living is then a measure of the extent to which preferences are satisfied. Given a set of prices that remain constant over a number of periods, the standard of living can be measured by the amount of money spent or, essentially, by real income. More technically, one can measure living standards by the size of the budget at a reference set of prices. This concept of the standard of living is a narrow one, defined entirely in terms of consumption of goods and services. It makes no claim to capture broader aspects of well-being, such as health or happiness, even though consumer choice is often described, for largely historical reasons, as "maximizing utility" or "maximizing consumer satisfaction."

Consider an individual who is behaving according to the theory. In the reference period, there is a set of (reference) prices, and the individual has a certain amount of money to spend. This, together with the prices of goods, sets her standard of living. Next, consider a new, comparison, situation, when the prices are different. How can we think about a cost-of-living index based on holding constant not the original bundle but the standard of living? Since the standard of living is not observed, one may appear to be facing a difficult, if not impossible, task. But there is one straightforward way to make at least a first approximation, which is to calculate the new cost of the reference period basket of goods. This is, of course, the Laspeyres procedure discussed above. The key insight is that, provided nothing else (such as the quality of goods) has changed, the new cost of the original basket is always sufficient to ensure that the individual can reach the original standard of living. If the consumer buys the same bundle, her standard of living is the same. But because relative prices have changed, there may be other bundles that are just as good for the consumer, that also maintain the original standard of living. At the new prices, some of these bundles may cost less than the original bundle. If so, it will be possible to maintain the original standard of living for an amount of money less than the new cost of the original basket. Since it is always possible to reach the original standard of living by buying the original basket, the Laspeyres price index sets an upper bound on the increase in the cost-of-living index based on the original (or base) standard of living.

The difference between changes in the cost of the base period *basket* and changes in the cost of the base period *level of living* plays an important part in cost-of-living index theory, as well as in this report. The size of this difference depends on the extent to which the consumer is able to rearrange her purchases to take advantage of the fact that some goods have become relatively cheaper and others relatively more expensive. This rearrangement of purchases is referred to as consumer substitution, and this substitution effect is one of the most important differences between basket price and cost-of-living indexes.

An important concept in this discussion is that of compensation. When one thinks about taking someone back to his original standard of living after prices have changed, one is asking how much that person must be compensated to make up for the price change. This compensation is the difference between the cost of obtaining the original standard of living at the old and new prices; it is known in the economics literature as the compensating variation. The cost-of-living index is the ratio of the same two costs. It is this close relationship between the compensating variation and the cost-of-living index that makes the latter a natural candidate for price indexes that are to be used for compensation purposes, such as for maintaining the standard of living of social security recipients. Note that there is nothing to stop the compensation from being negative if the price change reduces the cost of obtaining the original standard of living.

The discussion so far has been in terms of the cost-of-living index associated with the reference period level of living and with the corresponding Laspeyres price index, which uses the reference period basket of purchases as the base. In this case, the COLI holds constant the reference period level of living. But one can also construct a cost-of-living index associated with the comparison period level of living and compare the cost of this level of living at the prices in the reference and comparison periods. In this case, the COLI would use the comparison period level of living as the base. If one follows through exactly the same line of argument as above (or checks the equations in "Technical Note" at the end of this chapter), one finds that this current period cost-of-living index is always at least as large as the Paasche price index comparing the current period basket at the two sets of prices. Stating the two results together, for a consumer who behaves according to the theory, the Laspeyres price index is always at least as large as the cost-of-living index using the reference period level of living, and the cost-of-living index using the comparison period standard of living is at least as large as the Paasche price index. It is important to note that these two cost-of-living indexes, one using the reference period level of living as the base and the other using the comparison period level of living as the base, are conceptually different and will only coincide in very special circumstances. As is the case for basket price indexes for which the choice of basket matters, the choice of the base level of living will also generally matter. In consequence, it is not true, though it is often loosely claimed to be true, that the cost-of-living index lies between the Paasche and the Laspeyres. Indeed, it is perfectly possible, even for a consumer who obeys the theory, for the Paasche to exceed the Laspeyres.

The cost-of-living price index is sometimes referred to as the "true" cost-of-living index, a usage which suggests that it is unique. But as we have seen, this is not generally the case. For a consumer obeying the theory, a COLI using the reference period level of living as its base may differ from a COLI using the comparison level of living as its base, and there are potentially an infinite number of other COLIs, each associated with a different level of living. Just as with basket price indexes for which, in principle, one can think about using any basket as the base, so too can one use any level of living to construct the COLI. This multiplicity of possible COLIs is often inconvenient, so that it is natural to ask in what circumstances the multiple indexes are the same. This turns out to be the case if the consumer behaves in accord with what are known as *homothetic preferences*. This is also the condition that is necessary for the Laspeyres to be at least as large as the Paasche whatever the prices may be, a result first established by Frisch (1936). When preferences are not homothetic, there will always be at least one level of living, somewhere between the reference and comparison levels, for which the COLI lies between the Paasche and the Laspeyres (Konüs, 1924). Homotheticity in preferences implies that the way the consumer ranks different bundles of goods is the same no matter what her level of living so that, for example, the rate at which a person is prepared to trade food for tobacco, or baseball tickets for opera tickets, is the same whether the person is rich or poor. Homotheticity also implies that, as people become better off, they simply scale up their purchases without changing the pattern of consumption. However, such behavior is inconsistent with more than a century of empirical evidence dating back to Engel, who showed that the share of food in the budget diminishes at higher levels of income. Because homothetic preferences are not a reasonable description of reality, one must acknowledge a multiplicity of cost-of-living indexes.

So far, we have introduced the concept of a COLI and presented the classic results about the relationship between the Laspeyres and Paasche indexes and the associated COLIs. By themselves, these arguments are of limited practical application. Although they explain the limits of basket price indexes for thinking about cost-of-living indexes or compensation, they tell nothing about how to calculate a cost-of-living index more accurately. For example, one might argue that compensating social security recipients according to a Laspeyres-based CPI ignores their ability to substitute in response to changes in relative prices and therefore overcompensates them. But, from the discussion so far, it is not clear that it is possible to do better without a direct way of observing the standard of living.

One approach to constructing better cost-of-living indexes is to find out more about how consumers respond to changes in prices and income, something that in principle is directly observable. Between the 1950s and the late 1970s, economists worked out theoretical and empirical procedures for measuring the standard of living, given a knowledge of consumer *demand functions*, the relationships that tell us how purchases depend on prices and income. In particular, if the demand functions are known, cost-of-living indexes can be calculated exactly. Here then is a possible procedure. Econometric methods can be used to estimate the demand functions from market data on each individual's purchases, prices, and income and the results used to calculate any cost-of-living index numbers that one wants. While it is useful to know that this is possible, there are serious drawbacks to recommending such procedures for routine use in national statistical offices. Econometric modeling is often controversial because it relies heavily on judgment, and the assumptions needed to justify a given inference can often be challenged. This would make it difficult for the BLS to defend a CPI whose construction was crucially dependent on this sort of subjective work.

There is a less controversial approach that holds great promise for calculating good approximations to cost-of-living indexes. This uses what are known as *superlative* price indexes, which are better approximations to a COLI than the basket price indexes and can be calculated without knowing demand functions. Consider a concrete example. Using the "test" approach to price index construction, Fisher (1922) recommended what is known as Fisher's ideal index, which is a geometric mean (the square root of the product) of the Paasche and the Laspeyres indexes. Although Fisher's index was not derived from cost-of-living considerations, a natural question is whether it has a cost-of-living interpretation. This would be the case if there were consumer demand functions that led back to the Fisher index. This turns out to be true, as was demonstrated by the Russian mathematician Byushgens and economist Konüs in the 1920s (see Konüs and Byushgens, 1926). Indeed, the demand functions that do the trick are relatively general; although they are homothetic—which, as we have seen, is a considerable disadvantage—and have a specific functional form, they leave a large number of parameters unspecified. Subject to homotheticity, these parameters can be chosen to match *any* pattern of consumer substitution that is consistent with the theory. In consequence, the Fisher ideal index can be interpreted as a cost-of-living index without being specific about exactly how consumers substitute in response to changes in relative prices. Apart from the homotheticity (see below), this result comes close to squaring the circle. The statistical agency does not need to make potentially controversial estimates of demand functions. Instead, it can use the two basket price indexes, Paasche and Laspeyres, to calculate another index, the Fisher ideal index, that does what neither basket index can do by itself, namely, capture substitution behavior in a relatively general way.

In his work on superlative indexes, Diewert (1976) extended these results in important ways. First, he went beyond the Fisher ideal index and defined a whole class of superlative indexes whose members, like the Fisher ideal index, are capable of capturing general substitution responses. All of these, like the Fisher index, can be calculated from the same information that goes into basket price indexes—reference and comparison period prices and quantities. They also all require information on comparison baskets so that, like the Paasche index, they can only be produced as quickly as quantity data can be collected.

Diewert also addressed the homotheticity issue. He showed that when demand functions are not homothetic, so that there are different cost-of-living index numbers at different levels of living—and this is the relevant case in practice—superlative indexes can be interpreted as cost-of-living indexes for some level of living intermediate between those of the reference and current periods. If, over the interval of comparison, changes in the level of living are not very important relative to substitution effects in explaining changes in purchases of goods, then cost-of-living indexes evaluated at different intermediate points between the reference and the comparison periods will not differ very much. In that case, the superlative index is a useful approximation to the change in the cost of living evaluated at the reference period of living. The cost of this extension is some loss of conceptual clarity. A COLI indexed on the standard of living in the reference period and a COLI indexed on the standard of living in the comparison period are two distinct concepts, and the superlative index yields neither one nor the other, but something in between. Superlative indexes must therefore be used with caution in situations in which the change in prices involves a substantial change in the level of living.

The apparatus is now almost complete, at least for the case in which the goods themselves remain constant over time. We have presented a theoretical concept for a cost-of-living index, described the intimate link between cost of living indexes and compensation, identified inequalities that link cost-of-living indexes to the Paasche and the Laspeyres indexes, and introduced a set of practical superlative indexes that can capture the consumer substitution effects missed by basket price indexes. But there remains one important step. Everything in this section has been presented for a *single* consumer, not for an aggregate or group of consumers for which price indexes are normally constructed. This step, from an individual to the aggregate or average, is far from straightforward, if only because the concept of standard of living, on which cost-of-living index numbers are based, has no immediate analog for an economy as a whole, or even for a group of consumers.

In some discussions of cost-of-living indexes, this problem is simply ignored, and all consumers together are treated as if their behavior was generated by a single "representative" consumer. This imaginary person has a living standard that is somehow supposed to represent a national level of living and for which a cost-of-living index number can sensibly be defined. Such fictions can be justified only under extremely implausible conditions (see "Technical Note 1" in Chapter 8). To pretend that the theory of living standards and of behavior makes sense at the national level is to do it such violence as to cast into doubt the value of constructing a theoretical basis in the first place. It is much better to construct a framework in which one can explicitly move from an individual to a group or the nation. To do so requires a conceptual basis for an aggregate cost-of-living index number.

The most frequently used theory was first suggested by Pollak (1980, 1981) and is known as the social cost-of-living index. It works as follows. As always, there is a reference period and a comparison period, each with its own set of prices. For each family or household in the economy one calculates the least amount of money needed in the comparison period to be as well off as it was in the base period. This amount, divided by expenditure in the base period, would give the family's own base period COLI. But instead of doing the division, one adds up all the required amounts over all families in the economy to get the total amount of money that would be needed to keep them all just as well off as before. The ratio of this total to the total amount of money spent in the base period is the social cost-of-living index.

An alternative to the social cost-of-living index would be to take the cost-of-living index number for each family and average those numbers over all families to get a national cost-of-living index. This *democratic* COLI is not the same as the social cost of living defined above, which is in fact the *plutocratic* COLI defined as a total expenditure weighted average of each family's COLI. Indeed, the plutocratic and democratic COLIs bear exactly the same relationship to one another as do the plutocratic and democratic basket price indexes. The aggregation of index numbers over the population, or over groups, is not an issue that separates cost-of-living and basket price indexes.

Not surprisingly, the over- and underestimation results linking COLIs to the Paasche and Laspeyres indexes carry through to the social (and, indeed, to the democratic) cost-of-living index. If the CPI is computed as a plutocratic Laspeyres index, Pollak (1980) showed that the CPI is at least as large as the social cost-of-living index using each family's base level of living. Similarly, if the CPI is a plutocratic Paasche, Diewert (1983) showed that the CPI is no larger than the social cost-of-living index using each family's current cost of living. Once again, the aggregate (plutocratic or democratic) Laspeyres need not be larger than the (plutocratic or democratic) aggregate Paasche. But as Konüs (1924) showed for the individual consumer, there is at least one set of intermediate levels of living for which the (plutocratic or democratic) COLI lies between the (plutocratic or democratic) Paasche and Laspeyres indexes. Also as before, one can calculate aggregate superlative indexes, such as the Fisher ideal index (see Diewert, 2000a, for the precise arguments). These indexes will capture the effects of substitution in the aggregate and will provide closer approximations to one particular social cost-of-living index than either the Paasche or Laspeyres indexes. An aggregate superlative index of this kind is one candidate to supplement the Laspeyres-type CPI in the United States. But a superlative index cannot entirely replace the Laspeyres because it cannot be produced in as timely a manner.

Ultimately, an assessment of the ability of a superlative index to approximate a measure of the ratio of expenditures required to maintain a consumer's base period standard of living depends on a judgment about the extent to which changes in the pattern of quantities purchased are driven by changes in income and tastes or by substitution responses to changes in relative prices.

Criticisms of Cost-of-Living Indexes

One central insight of economics is that people respond to changes in prices by selecting away from relatively expensive goods toward relatively cheaper goods. More simply, demand curves slope down. Cost-of-living theory incorporates that insight into the construction of price indexes. Substitution effects are part of what separate cost-of-living indexes from basket price indexes. Yet it is important to remember that substitution in response to price is only a part of what determines purchasing behavior in the economy; other factors, though important, are sometimes neglected in cost-of-living discussions. The aggregate bundle of goods bought by consumers responds to many forces other than prices and incomes: the demographic composition of the population is constantly changing, by age and by ethnic group; tastes are not constant, nor is the distribution of income; technology changes the nature of goods and the way that people use them. Consequently, when one uses prices together with purchases in the reference and comparison periods to form Paasche, Laspeyres, and superlative indexes, the results may differ from what would be expected if only prices had changed, as in the simplified formulations above. In addition, it is not always clear how to interpret superlative indexes when many different forces affect the pattern of purchases. Diewert (2000b) has shown how superlative indexes can be defined in a way that recognizes changes in the environment. Just as a superlative index applies to levels of living that are intermediate between the base and comparison levels, so can other changes be dealt with, noting that the superlative will apply to intermediate values of environmental variables, such as demographic composition, tastes, or income distribution. Of course, the discussion has now come a long way from the straightforward concepts from which it began.

Theoretically, a COLI seeks to measure the amount of expenditure required for a consumer to be equally satisfied in one time period as in another, or "the minimum expenditure necessary to achieve a base period level of *utility*" (Boskin et al., 1998:5). In theory, conceptualizing a COLI in terms of satisfaction or utility has the potential to avoid many of the conceptual problems addressed in this report, from substitution to taste and quality changes. From this perspective, what exactly a consumer consumes is irrelevant; we are merely interested in the price tag associated with a given base level of satisfaction or utility, irrespective of the products and services from which this utility is derived. At present, such an approach is utopian because there are no appropriate measures of utility (see below). Instead, the economic theory of consumer behavior sidesteps the issue by assuming that consumers maximize utility by making the appropriate choices. Hence, consumers' choices can be taken as indicating utility. One result of this theoretical decision is that the choice-based COLI is more similar to a price index than would be the case for a COLI based on other measures of utility.

Utility and Choice

The term utility was originally introduced by Bentham (1789) to refer to pleasures and pains, the "sovereign masters" that "point out what we ought to do, as well as determine what we shall do." From this perspective, utility is an attribute of momentary experience, and a consumption episode that gives one more pleasure has a higher utility than one that gives less pleasure. In principle, this *experienced utility* (Kahneman, 1999) can be assessed by measuring the degree to which a person is pleased or displeased at the time of the experience. Although common sense suggests that past experiences drive future behavior, the relationship is more complex. Individuals' choices are based on *predicted utility*, that is, expected pleasure or pain, for which people draw on memories of previous experiences, that is, the event's *remembered utility*. Unfortunately, experienced utility and remembered utility are not always closely related.

As an example, consider an experiment by Kahneman et al. (1993), who had participants go through two painful experiences of different duration. In a short trial, participants immersed one hand in painfully cold water (14°C) for 60 seconds. In a long trial, they went through the same experience, but kept their hand in the water for an additional 30 seconds, during which the water's temperature increased from 14°C to 15°C, a temperature that is still within the range of pain, as confirmed by contemporaneous reports. Accordingly, the long trial entailed the same 60 seconds of intense pain as the short trial, plus an additional 30 seconds of less intense pain. Nevertheless, participants retrospectively evaluated the longer trial as less painful, exhibiting a bias known as "duration neglect." This bias refers to the observation that people evaluate extended episodes by drawing primarily on two data points, the peak and the end, and largely neglect the overall duration of the episode. Hence, adding a better ending to the otherwise identical experience made the longer episode seem less unpleasant. In conceptual terms, the remembered (dis)utility (pain) of the longer trial is lower than its experienced (dis)utility. Following both experiences, participants were asked which one they wanted to repeat in a third trial. In contrast to common sense and the predictions of consumer behavior theory, a majority chose the longer trial, voluntarily exposing themselves to 60 seconds of intense pain plus 30 seconds of milder pain, instead of the merely 60 seconds of intense pain of the short trial. Apparently, their reliance on the peak and end of both pain episodes led them to prefer the episode with a milder ending, even though it entailed a longer exposure to painful stimulation. Aside from demonstrating that people learn from memories and not from experiences per se, findings of this type highlight the pitfalls of inferring utility from *choice*.

In contrast to core assumptions of the economic theory of consumer behavior, experimental research in psychology and decision making indicates that choice, or *revealed preference*, is at best an imperfect measure of experienced utility. Choices are often based on erroneous assumptions, always dependent on the given context, and frequently fail to increase experienced utility even when the consumer has abundant experience with the product of choice (for a review, see Loewenstein and Schkade, 1999). Hence, *decision utility*, that is, the weight assigned to the desirability of an outcome in the context of a specific decision, is only weakly related to experienced utility in the Bentham sense.

By adopting the economic theory of consumer behavior as the conceptual framework for a COLI, one endorses decision utility, rather than experienced utility, as the crucial measure of utility. Unfortunately, there is little empirical support for the assumption that decision utility is an appropriate measure of the pleasure and displeasure consumers may derive from their choices. Hence, a choice-based COLI may be unlikely to measure what the theoretical definition promises, namely the price tag of a given level of utility or satisfaction.

Satisfaction

The COLI literature often uses *utility* and *satisfaction* as interchangeable terms. However, much as choice is poorly related to experienced utility, so is satisfaction. The same modest pleasure of the taste buds can leave a person very satisfied when he knew from the outset that he picked a local "greasy spoon," but very dissatisfied when the person expected a fancy French restaurant. In short, satisfaction is a function of experience relative to a standard, not a function of the experience per se. Hence, the same objective situation can result in very different satisfaction judgments, which has important conceptual implications.

Every year hundreds of thousands of survey respondents around the world are asked to report how satisfied they are with their lives as a whole or with specific life domains. The answers they provide are mostly based on global evaluations of their living conditions relative to some standard, such as their own past, their current expectations, or the situation of others. Which aspects of their lives they consider, and against which standard they evaluate them, is highly context dependent (for a review, see Schwarz and Strack, 1999). More important for the present purposes, research into the determinants of life satisfaction provides little support for the assumption that improvements in the standard of living will result in corresponding improvements in subjective satisfaction, at least in industrialized nations.

All industrialized nations, and certainly the United States, have experienced enormous improvements in the objective standard of living over the last five decades. Nevertheless, reported life satisfaction has essentially remained flat. In contrast to the assumptions of the economic theory of consumer behavior, access to more and better goods apparently fails to increase consumers' life satisfaction. Cross-national comparisons suggest a similar conclusion. Although the wealth of nations is strongly related to average life satisfaction at low levels of GDP (gross domestic product) per capita, this relationship levels off once basic needs are met (see, e.g., Easterlin, 1967, 1974; Inglehart, 1997; for a recent review of these literatures, see Diener and Suh, 2000). Consistent with these observations, income is only weakly related to life satisfaction within developed nations, accounting for 1 to 2 percent of the variance in reported satisfaction. Overall, the available data indicate that the objective standard of living plays a more limited role in consumers' subjective satisfaction than the theory of consumer behavior would lead us to expect, a finding that poses severe difficulty for a procedure, such as the COLI, that relies on attaching a price tag to satisfaction. Even in the more supportive case of poor nations, the causal nature of the observed relationship remains ambiguous because the wealth of nations is highly correlated with human rights, democracy, and a predominance of individualistic values.

One interpretation of these findings holds that nothing can keep people happy or unhappy for long. According to variants of this *hedonic treadmill* hypothesis, expectations quickly adapt to new circumstances. If these circumstances are characterized by continuous improvement, ever-increasing amounts of goods are required to maintain the same level of satisfaction (see Brickman and Campbell, 1971; Campbell, 1981). Conversely, deteriorating circumstances would make people unhappy for some time, but only until expectations are back in line with reality, as long as basic needs are met. From this perspective, a satisfaction-based COLI would always show inflation in times of economic improvement because more goods are needed to maintain the base level of satisfaction, and it would show deflation in times of economic hardship once adaptation has set in. More troublesome still, a satisfaction-based COLI may show inflation as well as deflation in the absence of any changes in the price of products, hardly a desirable feature for most practical purposes. Moreover, a satisfaction-based COLI could change in the absence of any actual changes in the standard of living. Given the highly comparative nature of "satisfaction," a satisfaction-based COLI might, for example, indicate housing inflation once a luxurious new mansion is built in the neighborhood, by making the existing houses seem less satisfying by comparison.

An alternative interpretation of the weak relationship between the standard of living and consumers' satisfaction suggests that subjective well-being and life satisfaction may be largely a function of people's temperament and genetic endowment. Twin studies suggest, for example, that "the reported well-being of one's identical twin, either now or 10 years earlier, is a far better predictor of one's self-rated happiness than is one's educational achievement, income, or status" (Lykken and Tellegen, 1996:189). From this perspective, the subjective well-being of consumers is largely independent of the level of material well-being, which is consistent with the available survey data.

However, a word of caution is in order. As noted above, reported satisfaction is not closely related to experienced utility. At present, for example, it is not known if a new mansion next door would actually make the neighbors feel less comfortable in their old homes on a moment-to-moment basis or if it would only reduce their global evaluations of their homes, as expressed in satisfaction judgments (see Kahneman, 1999, for a discussion). It is conceivable that a higher standard of living may actually result in more positive moment-to-moment experiences with life which, however, may not show up in global satisfaction judgments made relative to a higher standard. These issues are the topic of current research on the measurement of experienced utility, which is arguably the most adequate measure for assessing the benefits of the material conditions of life.

As this selective discussion of the complexities of utility and satisfaction illustrates, the conceptualization of a COLI in economic theory is based on a very specific definition of utility, namely the decision utility revealed in choice. Decision utility, however, is a poor measure of utility as pleasure and pain, as conceptualized by Bentham. Yet little may be gained for the purpose of an index system by adopting a broader conceptualization of utility. Although measures of experienced utility would provide the most meaningful assessment of consumers' quality of life, these measures would most likely have properties that make them undesirable for an index system. Most importantly, they may indicate inflation or deflation in the absence of any changes in the price of products.

TWO PERSPECTIVES

Theoretical Requirements

One difference between the COLI and COGI approaches is how much theory is built into each. The fixed-basket index (COGI) uses theory to choose the weights for the price index, but it takes very little theoretical background to explain to an intelligent but untrained bystander that a consumer price index ought to price the things that consumers buy. If one thinks of a COGI as the cost of things, one needs to know which things—to which the answer is, the things that people buy. Of course, this still leaves unanswered the questions of which people, when did they buy them, and what varieties of goods. The "which people" question is about the group for which the index is designed, whether one is concerned with individuals or aggregates, with a democratic or plutocratic measure. The question of when leads to questions of a Paasche (the comparison bundle) or a Laspeyres (the reference bundle) index. As already noted, the Paasche index is not feasible for a real-time CPI. If it is acceptable to produce indexes with a delay of (perhaps) 3 or more years, when data on the comparison period purchases will be available, and if there is little to choose between the two indexes on other grounds, it makes good sense to compute superlative indexes tailored to offer approximations to the COLI concept. The third question, about the varieties of goods, raises issues of quality, of whether a good is a simple irreducible "atom" on its own, or should be thought of as a bundle of characteristics. This question is perhaps the hardest to answer, and attempts to do so require a good deal of theoretical structure, in any framework. Indeed, current BLS practice, within a broad COGI framework, requires continuous judgments, many of which are based on theories of how consumers behave and of the relationship between quality and price.

In contrast to a COGI, a COLI framework builds on the economic theory of consumer behavior. Many would argue that this theory is so simple and obvious that there is little reason not to adopt it, but as we have shown in the previous section, this is not clear to everyone. Still, there are great benefits from using the theory. The idea that, as prices change, people substitute toward relatively cheaper goods strikes most people as reasonable, and the theory provides the apparatus to deal with this idea in a formal way. Given the idea, one has a good basis for the further idea that someone with enough money to buy the original bundle in the new situation is at least as well off as before and may be better off, which is the idea of substitution bias (though there is still the further step, that a measure of compensation is the same as a measure of the change in the price level; see below). Many (although not all) price statisticians around the world have at least muted their original suspicions of the COLI approach in the face of its usefulness for thinking about important practical issues, such as the problems with constructing lower-level price indexes. As Triplett (1999) argues, economists use the theory of consumer behavior for all sorts of purposes, and their very success as public policy practitioners attests to the usefulness of their approach, even if they do not believe its literal truth. However, skeptics might attribute the same success to economists' willingness to rely on deductive reasoning even in the face of contrary evidence (such as that reviewed above).

Public Perception and Understanding

When it comes to public understanding, basket price indexes have an advantage over cost-of-living indexes: they are simple and can be explained in seconds to almost anyone. Against this must be noted that, in practice, the actual (modified) Laspeyres indexes used by the BLS are quite complex, so that much of the clarity is lost in day-to-day practice. However, this is true of almost any complex measure or, indeed, any complex object. One might understand very well in a general way why an airplane flies, and the knowledge probably makes people feel better when flying. But the detailed construction of a modern airliner is certainly unknown to almost everyone.

The cost-of-living index is a good deal harder to explain. The basic concept of comparing how much it costs to live at different prices is relatively straightforward, but making the concept practical or precise is quite difficult, sometimes even for those who support the concept. Part of the problem may be less a lack of understanding than a genuine intellectual resistance to the approach. For example, there is nothing incoherent in a position that accepts the argument about substitution bias and accepts that fixed-bundle compensation is overcompensation, but does not accept that the price level and the level of compensation are the same thing, or that the cost of things is the same as the cost of living. According to this view, the COLI is the right framework for calculating compensation, but not necessarily for calculating the price index.

Nevertheless, there is some disagreement, including among members of the panel, on how difficult it is to understand the COLI concept. Some argue that the basic idea of a cost-of-living index is straightforward and that the difficulties come with the detailed implementation. According to this view, there is no great difference between a COLI and a COGI, since the implementation of the latter also requires much that is complex and difficult to understand. Others challenge the comprehensibility of even the basic concepts underlying the COLI. The idea of holding constant the standard of living requires some notion of what is meant by a standard of living and whether this is "happiness," "satisfaction," "utility," or something else. As we have shown, it is indeed something else, and explication of it takes a good deal of space. Finally, there is room for argument about the importance of public perception and understanding and how much weight it should be given in the construction of an index number for which there will inevitably be a great deal of technical detail.

Substitution Bias

The basket price index is just that, the cost of a basket of goods relative to the reference. The issue of substitution bias does not arise, at least if one is happy with the choice of basket. But if one has several reasonable choices of baskets and if they give different price indexes, one has to choose between them and recognize that at least part of the difference between the indexes, say the Paasche and Laspeyres, comes from consumers substituting in response to changes in relative prices. Once again, though, substitution is not the only reason—nor even necessarily the most important—for changes in the bundle; changes in tastes, in quality, and in the sociodemographic composition of the population also have their effects. If one chooses to maintain the distinction between the cost of living and the cost of things, with the former relevant for compensation and the latter for a price index, then recognizing the existence of substitution does not necessarily involve recognizing the existence of substitution bias. A fixed-basket index is biased as a measure of the cost of living but not necessarily as a measure of the price level itself.

From the COLI point of view, which does see the price level in terms of the cost of living, the Laspeyres index is at best an approximation that overstates the change in the price level between the reference and comparison periods. The degree of overstatement is the substitution bias, and it will tend to be larger when the difference in relative prices is large and when consumers' ability and willingness to substitute one good for another are high. But it is not necessarily true, as is sometimes supposed, that the overstatement of the COLI by the Laspeyres becomes increasingly severe simply as the time between the base period and the (current) comparison period increases (see "Technical Note" at the end of this chapter). Indeed, there is some empirical evidence that, for recent U.S. history, price indexes that allow for substitution do not rise faster than the CPI (see Shapiro and Wilcox, 1997).

Two points are noteworthy in this context. First, even if one is committed to a COGI approach for the CPI, there is nothing to stop one from using a COLI for compensation purposes. Indeed, if Congress mandated that social security benefits be indexed to hold constant the living standards of social security recipients who have no other income, the COLI would certainly be the appropriate index for escalation (at least subject to the issues related to compensation; see below). In practice, this might mean retaining the Laspeyres approach for the CPI, while recognizing that CPI-based compensation is overgenerous, and compensating people by the growth in the CPI less some modest amount in recognition of substitution bias. More sophisticated compensation schemes could be implemented using superlative indexes, albeit with a lag (discussed further in Chapter 7).

One alternative to waiting for superlative indexes is to use other indexes that make a somewhat less exact allowance for substitution but that can be produced on the same schedule as the CPI. There are a number of possibilities. One is a constant-elasticity-of-substitution (CES) price index, suggested by Lloyd (1975) and recently evaluated by Shapiro and Wilcox (1997). A CES index starts from the price relatives for each good, the ratios of the price in the current period to the price in the base period. In the Laspeyres index, these relatives are averaged, using as weights the share of the budget devoted to each good. In a CES index, each price relative is raised to a power (for example, 0.5) before being weighted and added up. The final index is then obtained by raising this weighted sum to the power not of the exponent but of its reciprocal (see "Technical Note"). If the exponent is 0.5, one is weighting together the square roots of the price relatives and squaring the result. If the exponent is 1.0 (unity), one would be reproducing the Laspeyres; at the other extreme, with an exponent of zero, one would have the expenditure-weighted geometric mean of the price relatives. The smaller the exponent, the more goods are substitutable for one another. Indeed, if one subtracts the exponent from unity, the result is the measure known as the elasticity of substitution.

The constant elasticity of substitution index is exactly equal to the cost-of-living index number if preferences are homothetic and if all goods are equally substitutable for one another. In practice, historical data could be used to choose the exponent that brings the CES index as close as possible to some superlative index, such as Fisher's ideal index. And although the assumptions of homotheticity and equal substitution are not realistic, such an index will nevertheless capture substitution bias in a way that the Laspeyres does not, and it will do so without requiring data on current purchases. Thus, if substitution is the main concern, the CES index has attractions as a basis for the CPI.

But there are also arguments against the use of a CES price index. Goods are not equally substitutable for one another. If prices change in a way that highlights the assumption to the contrary, a CES index can be quite misleading. Suppose, for example, that the dollar depreciates relative to the currencies of East Asian electronic and automobile producers, so that the imported versions of these products become more expensive. This will hurt American consumers, though the damage will be offset by substitution away from Asian to European and domestic substitutes. If some such substitution is built into its construction, the CPI will rise by less than the increase in the prices of the imports, which is exactly what ought to happen. Now imagine instead that an oil shock increases the price of gasoline and home heating products. In this case, there is much less scope for substitution, and the increase in the CPI ought to be much closer to the increase in the price of fuels: BLS cannot tell consumers that the CPI has not risen by much because they should drive their cars on milk or on orange juice! A superlative index can capture the difference between the two cases because it uses information on purchases after the price change and is sensitive to the fact that the shock induces a much smaller decline in demand for fuels than for imported electronics. But a CES price index treats both identically and assumes that fuels are just as substitutable as imported electronics. In one case or the other, the CPI will be wrong and possibly quite wrong.

Of course, there is no guarantee that a Laspeyres-based CPI will do better as an approximation to a COLI. Indeed, since a Laspeyres is itself a CES with an exponent of unity, a good choice of exponent will certainly lead to an index that does at least as well as the Laspeyres. It is also conceivable that more elaborate CES indexes—such as a two-stage CES, which has the same substitution elasticity between broad groups of goods, with different substitution elasticities within each group—could do even better and remedy the equal substitution problem of a simple CES.

A CES index has an advantage over a superlative index in its timeliness, but it is otherwise inferior. For example, there would be no point in using a CES index instead of a superlative for looking at long-run trends in inflation or for other historical analyses. With some improvement in data collection, much of which is already under way, a superlative index could be produced with a delay of only 1-2 years. Given data up to that time on both a Laspeyres-based CPI and the superlative, it would be possible to make an informed estimate of what the superlative is likely to be, even in advance of its calculation. In this situation, it is not clear that a real-time CES index adds very much. Compensation, such as social security compensation, could ultimately be tied to a superlative index. Interim payments could be made from a forecast of the superlative, with forecast errors rolled into subsequent cost-of-living adjustments (see Chapter 7).

One area in which a COLI concept has already entered BLS practice is the treatment of lower-level price aggregation. This is the procedure whereby the BLS combines prices of the most finely defined goods, such as different varieties of apples or different brands of VCRs. Common sense suggests that consumers are much more willing and able to substitute between two kinds of apples than between an apple and an orange or between two kinds of VCRs than between a VCR and a stereo system. Some lower-level categories like medical supplies or sporting equipment are clearly different: a left-leg prosthesis is not a very good substitute for a right-leg prosthesis, nor a golf club for a tennis racket! Conversely, some goods within an item category may be more substitutable for each other than is explicitly assumed by the BLS procedure for combining prices at the lower level.

At this disaggregated level, the COLI perspective has a distinct advantage because substitution is clearly important and because the COLI recognizes it. It is probably easy to explain to consumers that an increase in the price of a Gala apple is not so serious as (i.e., requires less compensation than) an increase in the price of all apples, even if they usually buy the Galas. And the COLI approach gets this right. In response to this sort of argument and after publication of the Boskin commission's report, the BLS switched from its previous Laspeyres approach to a weighted geometric means procedure (though not for all categories, including artificial limbs). This geomean method will give the right answer under rather specific assumptions about the degree of substitution between goods. Although the specific assumptions are unlikely to be exactly true, most observers regard the change in procedure as an improvement. The change also marks the BLS's own change of perspective from a COGI conceptual basis (informed by COLI considerations) to an explicit COLI basis. It is probably the BLS's first attempt to build substitution effects into the CPI itself, rather than into an experimental index.

We also note that, even prior to the introduction of the geomean procedure, the BLS used a seriously modified Laspeyres index that incorporated a procedure called seasoning, by which the weights come from a different period than either the reference or the comparison periods. The seasoned Laspeyres (including other modifications) used by the BLS was a long way from the simple fixed-basket approach, so that much of the original simplicity of the concept had already been lost. Analysis of the effects of seasoning (for example, by Shapiro and Wilcox, 1996, following the analysis of Reinsdorf and Moulton, 1995) shows that a seasoned index has different biases than an "unseasoned" index, so that it is not clear that the geomean index will always be practically superior to the seasoned Laspeyres (see "Technical Note" for more discussion).

Quality Adjustment

Quality adjustment is possibly the area in which the COLI has the greatest advantage over the COGI approach. For the COGI, there are both practical and conceptual difficulties in trying to work with a fixed basket of goods when the functions and even definitions of goods change. When qualities are changing rapidly, it can be difficult even to find the original basket. Where would one find a 16 MHz computer, let alone an electric calculating machine? And even if one could do so, the results would provide no clue to the effective reduction in price that comes from quality improvements. At worst, a fixed-basket methodology is impossible for such cases, and at best it is irrelevant. The COLI approach is more helpful: when the qualities and even definitions of goods are changing, it makes more sense to try to work with a constant standard of living than with a constant bundle of goods. With the COLI approach, one has conceptual clarity—one can compare the cost of living in two situations in which both prices and qualities of goods are different—which is not possible in the COGI framework.

As we have already seen in other contexts, however, the differences between basket and cost-of-living approaches are rarely so clear cut, and the same is true for quality. Basket approaches can sometimes be modified in sensible ways to deal with quality change, and the COLI approach, while conceptually clear, can sometimes fail to give solid practical guidance. We referred above to the results in the economics literature that show how to construct cost-of-living indexes from demand functions linking consumer purchases to prices and incomes, but these results were derived for situations in which quality is constant. There is no comparable body of theory that allows construction of cost-of-living indexes when prices, income, and quality are all changing. To make progress, one must know more about how quality is changing; and indeed there is a growing, albeit experimental, literature looking at ways of modeling quality change in theory and in specific empirical situations (see Feenstra, 1994, 1995; Berry et al., 1995). But as we argue in Chapter 4, knowledge of quality change can often also be used to adjust the definition of goods in a basket price index approach. Thus, once again, the distinctions between the two approaches are blurred. Indeed, in the judgment of the panel, the central issue in constructing quality-corrected price indexes is not the distinction between COLI and COGI approaches, but the measurement of quality change itself.

In some cases, it is relatively straightforward to see how quality should be adjusted. For example, when coffee now comes in a 12-ounce pack instead of a 16-ounce pack, an obvious (if not necessarily precisely correct) procedure would be to add a third to the price of the new pack before comparing it with the price of the old pack. Such adjustments are routinely made in repricing a basket of goods. In some cases, quality can be thought of as a special form of repackaging. Razor blades might give 10 shaves instead of 5, or a gasoline with an additive might give 25 miles per gallon instead of 20. In both cases, it would make sense to price the cost per shave or the cost per mile. None of this differentiates between basket and cost-of-living approaches, though the latter would make an allowance for substitution toward the improved commodity when cost per shave or cost per mile fell.

Unfortunately, quality changes are seldom easily converted to changes in effective quantity. A computer may run at 500 MHz instead of 200 but have the same infuriatingly slow connection to the Internet. A quality correction requires some way of knowing how to value an improvement in one characteristic while another is held constant. Some new goods, such as cellular telephones, home video machines, or digital cameras, mix characteristics of old goods with wholly new characteristics. The valuation of new goods with entirely new characteristics poses even more difficult problems than does the valuation of characteristics that previously existed but are brought to market in different combinations.

Once again, econometric analysis can provide some insights into the behavior of consumers and producers, including how consumers respond to new goods and how they value the underlying characteristics of goods. Hausman's (1997) work on breakfast cereals provides a widely cited example of an econometric analysis of a new product. But the identification of such relationships is rarely uncontroversial, and we believe it would be unwise for statistical agencies to condition important data on the validity of specific econometric models. Pricing the underlying characteristics of goods is the aim of the hedonic technique whereby the market prices of goods are related to the amounts of each characteristic that they contain. The method requires prior knowledge of what characteristics consumers value, and its application often raises questions of interpretation and econometric technique. Even so, hedonic methods are probably the best hope for improving the way in which quality adjustments are made.

There are fundamental questions as to whether it is possible, even in principle, to measure certain kinds of quality change. In particular, taste change is sometimes indistinguishable from quality change. For example, becoming a vegetarian allows one to obtain the same nutrition from less food expenditure, just as would an improvement in the quality of fruits and vegetables. If all goods suddenly became twice as good, everyone would be better off, but there would not necessarily be any changes in consumers' purchases. One might imagine an economy in which everyone agrees that the "goodness" of goods has increased but where the proponents of the "new goodness economy" say it has increased twofold, while skeptics say it increased by only 50 percent. There is no way of inferring from consumers' behavior which is right, nor of making the corrections to their cost of living that such a change would presumably require. Unless a great deal is known about the nature of the quality change—for example, what the goodness of a good is and how much of it there is, perhaps from the manufacturer's engineering specifications—it is generally not possible to infer quality from examining what has happened to consumer purchases. Yet if someone becomes better at using a good and so gets more out of the same purchase (say, a golf club)—something that is a taste change, not a quality improvement—the associated behavior will look the same as if the manufacturer had increased quality by improving, or putting more goodness into, the good.

The Domain of the Index: Conditional and Unconditional Cost-of-Living Indexes

Any price index, whether derived from a COGI or COLI approach, needs a list of goods that are covered and a list of goods that are not. In its discussions the panel came to call this the domain issue. There is general agreement, within the panel if not universally, that the domain should follow current practice, including market goods and excluding nonmarket goods (e.g., public goods, the environment, crime, life expectancy). In addition, only current goods and services should be covered, not leisure, nor goods and services in past or future periods, even though consumers currently get part of their well-being from consuming time, as well as from the contemplation of past and future purchases. These conclusions are consistent with either a basket or cost-of-living approach to index number construction, though the arguments are different. As we shall see, the definition of a cost-of-living index needs to be modified to become a *conditional* cost-of-living index (for more technical discussions, see Caves et al., 1982; Pollak, 1989; Diewert 2000a). This modification is somewhat controversial, and it has important implications for the application of the cost-of-living framework in other contexts.

For a COGI, the domain can be anything that is thought to be suitable. For example, one can select what people think ought to be in a price index, recognizing that they will certainly need some guidance on how to handle such matters as interest rates or durable goods. Such a procedure would almost certainly lead to the inclusion of the prices of market goods and services. People recognize that it is a good thing when life expectancy goes up, when crime goes down, or when a new product (cell phones or Viagra) makes life more enjoyable, but they seldom think that such improvements reduce the level of prices.

The COLI approach can get to the same place but requires more steps, some of which would be resisted by those who take a comprehensive approach to cost-of-living indexes. A good place to start is with the example of a local government raising sales taxes to build a bridge. Some local taxpayers would prefer to keep their money while others would prefer the bridge, so that the tax to fund the bridge will make some people better off and some worse off. Suppose that, on average and taking into account the taxes, people are about as well off after completion of the bridge as they were before. What has happened to the cost of living? According to the comprehensive approach, nothing. Although prices of goods are higher, the bridge brings benefits which, by assumption, exactly offset the increased cost of goods. So consumers need no compensation, and the cost of living has not changed. By contrast, the COGI approach says that prices have gone up, which means that the CPI has gone up. It is not that the bridge is irrelevant to people's welfare or is not worth anything, but simply that the existence of the bridge seems irrelevant to the measurement of the price level. This seems like an excellent example of a case in which the price index and the cost of living are different things in practice, not just in theory. And, indeed, many of those who argue for the adoption of a comprehensive COLI approach see its treatment of such cases, not how it handles substitution, as the main advantage of COLI over COGI approaches.

One can, however, make the COLI give the same answer as the COGI by constructing a "conditional" cost-of-living measure, defined as the minimum expenditure on market goods needed to attain a given standard of living when the provision of nonmarket goods is at some specified level. In this way, the conditional COLI changes only when prices change. Without changes in prices, the conditional COLI is constant: it cannot be altered by changes in nonmarket goods or changes in the environment (such as the provision of the bridge) or by an increase in life expectancy.

The conditional COLI can be used to hold constant not just the provision of public goods but anything that one does not want to affect the price index. A good example is temperature: in unusually cold winters or hot summers, families have to spend more money to attain the same level of comfort (the same temperature in their homes). Should the CPI rise because the winter is unusually cold, even if the price of heating fuels remain constant? For the panel, the answer is no. The cost of living has gone up, but prices have not. We prefer a price index that does not change in response to temperature changes alone. For this reason, the preferred choice for a cost-of-living index is not the comprehensive or unconditional cost-of-living index but a conditional cost-of-living index that holds constant all environmental nonprice factors that affect people's well-being.

The conditional cost-of-living index can exclude those things that people believe should be excluded—such as fluctuations in winter temperature—leaving it somewhat more like a price index and somewhat less like a cost-of-living index. For most purposes, a conditional COLI is arguably the right concept. It responds to price changes as a price index should, and it takes into account consumer substitution. Nevertheless, a conditional COLI has problems in dealing with some issues and arguably gives the wrong answer in some of them. Some economists object to almost any exclusions. For them a conditional cost-of-living index is no longer a cost-of-living index. Thus, in the case of a sales tax for a bridge, they think the value of the bridge should be taken into account or, if that is impractical, the increase in sales tax should be excluded from the price index. Similarly, the price index should be decreased for an increase in health status or a reduction in the crime rate because both reduce the amount of money required to reach a given standard of living. Although the Boskin report does not formally recommend such a position, it contains a number of statements that are sympathetic to such a treatment.

Even if such arguments are not rejected in principle, there are practical examples for which the case for a conditional COLI is unpersuasive. One example is the construction of regional or city price indexes. Nothing in cost-of-living theory says the base and comparison situations cannot be different *places*, rather than different *times*, and there are many situations in which such cross-place comparisons are needed; for example, the State Department needs to make cost-of-living adjustments for employees living in foreign countries or when an employer wants to adjust salaries for the cost of living in different U.S. cities. A cost-of-living index that compares Phoenix with San Francisco would surely recognize that homes in Phoenix require more air-conditioning than homes in San Francisco. It is hard to see the purpose of a conditional cost of living computed under the assumption that Phoenix has San Francisco's climate, or vice versa. Yet it is precisely this assumption that is needed to prevent a COLI from changing with climate fluctuations over time.

A conditional COLI can also limit our ability to handle quality change. Most people would probably agree that general increases in life expectancy that are not caused by changes in medical care or other market goods should not reduce the CPI, but a conditional COLI should take account of changes in life expectancy due to improvements in the quality of medical care, such as better treatment of heart attacks. Yet it is not clear where to draw the line between general increases in life expectancy and more specific quality improvements, for example, in the treatment of depression or heart attacks or in cataract surgery. When new drugs make it easier to ameliorate an episode of depression or when new techniques reduce the cost of cataract surgery, most would probably want the change to be reflected in a cost-of-living index, and perhaps even in the price index. Some help comes from an appropriate redefinition of commodities, for example as the treatment for an illness, rather than the drugs and medical services themselves. But if quality improvement comes through new technology and if a conditional COLI treats technology as an environmental variable to be held constant, the contribution of quality change to effective price reduction may be ignored or at least understated. Indeed, if one thinks of a conditional COLI as designed to prevent changes in the index level when prices are constant, then it would seem to rule out quality adjustments to price. To capture the contribution of technological change to effective price reduction in the price index, one must remove technology from the conditioning variables in a conditional COLI. Yet as the example of life expectancy shows, such "unconditioning" must be selective; one must hold some technologies constant while others are allowed to change (see "Technical Note" for a more formal discussion of this point). The difficulty of deciding on what to condition is further aggravated by the difficulty in practice of separating changes in technology from changes in tastes, as when the BLS counted as a quality improvement not only the switch from cotton to synthetic shirts but also the subsequent switch in the opposite direction.

One more example is worth thinking about. The availability of a new drug like Viagra certainly makes many people better off. There is no obvious way of redefining one or more commodities so that this shows up as a price decrease. Indeed, many people—including most members of the panel—are quite uncomfortable with the idea that the introduction of Viagra should reduce the CPI. (Here we are considering specifically the "new goods" effect. If, hypothetically, there had been a CPI stratum "treatment for impotence" and the introduction of Viagra had lowered the price of that treatment, then perhaps a price decline for the good, so defined, could be justified. See Chapter 6 for more discussion of this issue.) Yet, on average, consumers are better off than before, even if their incomes and the prices of all other goods are the same. If the price level is not adjusted, the benefits of the technological innovation are missed and show up nowhere in the accounting system. Many economists are concerned that these phenomena are pervasive in modern economies, where the growth in quality has replaced the growth in quantity as the main engine of increased well-being, and where the production of new commodities is as important for economic growth as is the more efficient production of old ones. Neither Laspeyres price indexes nor conditional cost-of-living indexes are likely to capture this progress; it would likely be better captured by an unconditional COLI.

At present, there is probably no alternative to a selective treatment of whether or not the state of technology should be a conditioning variable for a COLI. Such selective treatment of the same variable, technological advance, it not a very comfortable position. It leaves much room for discretion of a kind that will undoubtedly be a source of debate for years to come. More generally, the unacceptability of the unconditional cost-of-living index, together with the apparent impossibility of devising a general way of conditioning the conditional cost-of-living index, has brought several members of the panel to the view that there has probably been too little research on other conceptual approaches, such as the test or stochastic view of index numbers.

There is another somewhat more technical issue concerning conditional cost-of-living indexes. The conditional cost of living itself, and thus a conditional COLI associated with it, depends not just on prices and the level of living but also on the levels chosen for nonmarket goods and other conditioning variables. If the government provides public schools, a family's cost of living is different than when it does not, and the way in which the family's COLI responds to price changes—the price of books or the price of tutoring—will depend on what the public school provides. Similarly, the conditional COLI of someone with a house to heat may be unaffected by changes in the price of fuel when the average temperature is 70°F, but very sensitive to fuel prices when the average temperature is 50°F. A conditional COLI will be independent of the environment only when preferences for market goods are "separable" from nonmarket goods or from characteristics of the environment. Separability requires that the way a family spends its money on market goods must be independent of the provision of nonmarket goods or that a family's choice between food and fuel is independent of the outside temperature. Such conditions are unlikely to hold. Indeed, some public goods may altogether supplant some private goods from a family's budget. When separability does not hold, the conditional cost-of-living index will be different depending on the level at which one chooses to hold constant the provision of nonmarket goods. It is hard to imagine such effects being taken into account in any practical CPI.

Changes in Tastes

A cost-of-living index compares the costs of equivalent standards of living under two different sets of prices. If someone becomes a vegetarian or decides that she prefers not to smoke, the cost of any given level of living will change, even with no change in prices. This can perhaps be dealt with in the same way as a change in the environment, regarding vegetarianism or nonvegetarianism as background variables that are subject to change. In this way, one can think about an unconditional cost-of-living index that calculates the change in the cost of living that comes about from both price and taste changes—becoming a vegetarian reduces the cost of living. In contrast, a conditional cost-of-living index calculates the change in costs, holding the original tastes fixed. It is not entirely clear how much of the original COLI concept is retained under either of these devices. The unconditional COLI essentially attaches values to different systems of tastes, something that most economists prefer to avoid. The conditional COLI is evaluating price changes according to tastes that are no longer valid, so that if tastes change rapidly and if the base period is held fixed for a long period of time, the conditional comparisons will become less and less relevant to consumers. But this is conceptually no different from the usual problems with selecting any base, and it is merely an argument for frequent updating of the base.

More serious are the practical questions, in particular, how to recognize taste change when it has taken place and, having done so, how to correct for it. Taste change is conceptually closely related to quality change. One is a change in the nature of goods, the other is a change in how goods are perceived. In general, it is not possible to distinguish one from the other by watching how consumers behave. Quality change is often directly observable from examination of the goods; observing taste change is much more difficult. Thus, one has little choice but to accept the conditional approach and to assume that tastes are constant.

In a COGI approach, where tastes are not mentioned, it might at first seem that taste change is not an issue. But the choice of a sensible basket is almost impossible if tastes are radically different in the two periods. This issue comes up forcefully when computing price indexes that compare price levels between two diverse countries such as the United States and India. Because the nature and pattern of consumer expenditures are so different in the two countries, with many goods that are bought in one not bought in the other, there is no comparable basket to price. Attempts to use one basket or the other can give absurd results if a staple in one country is not available in the other or is available only occasionally at an extremely high price.

Neither the COGI nor the COLI approach (nor any other we know of) is likely to do a very good job of constructing a CPI when there is a great deal of taste change. In this context, one might be seriously concerned about some of the psychological phenomena discussed above, that nothing makes people happy for long or the hedonic treadmill, which condemns a consumer to ever-increasing expenditures to maintain a constant level of satisfaction. It is not clear what to make of COLIs in such an environment, though one line of approach is through consideration of the literature on habit formation, according to which the cost of living increases in response to previously increased consumption because of the "needs" induced by the earlier consumption, so that the unconditional cost of living will drift upward relative to the conditional cost of living. Once again, a conditional COLI seems to be the appropriate concept for measuring the price level.

Using Indexes for Compensation

Since a COLI is calculated by measuring compensation, it is the natural index to use for compensation and indexation purposes, though one might want different COLIs for different people and circumstances. But a COGI could be adjusted to make it more like a COLI, for example, by making an allowance for substitution bias. Similarly, one could adjust a COLI to make it more like a price index by narrowing the domain (conditioning). However, there are practical and conceptual issues that arise in cost-of-living adjustments when people are sellers as well as consumers of goods, as well as when people have incomes other than those which the index is intended to compensate.

In an industrialized economy, and putting aside the supply of labor which is outside the domain of our index, the most important group of consumers who supply goods are homeowners, who sell housing services to themselves. The issue can perhaps be most clearly seen if one considers the example of a farmer who grows beans for market and uses the proceeds to buy as much as he can afford of a fixed bundle of goods. Sales of beans are his only source of income, and he saves nothing. Suppose that all prices increase by 10 percent. The farmer needs no compensation: the cost of his bundle of goods has gone up by 10 percent, but so has his income. At the conceptual level, a COLI can once again be made to give the right answer but only if one separates the COLI from its basis as compensation. This can be done by adopting the (perhaps strained) device of separating the farmer into his production and consumption selves, so that the latter can be said to have suffered a 10 percent increase in prices. The farmer as producer gets the profits and is better off, while the farmer as consumer pays more for his consumption and is worse off. The "integrated" farmer is both 10 percent better off and 10 percent worse off; he has had no net change in real income. The compensation required by the farmer for the price increase is zero, though he would receive money from a social security or other benefits system that was indexed to a price index of goods. Such a system would therefore fail to hold the farmer at the same level of living as before the prices rose.

Exactly the same issue arise for homeowners, though because they sell only to themselves they cannot be made better off by an increase in the cost of housing. Yet an increase in the price index driven by an increase in rental costs has no effect on their real incomes or their cost of living, and they need no compensation for it. Yet prices have risen, and a price index based on standard COGI or COLI procedures would recognize the fact. In such circumstances, though, compensation by such a COLI would not hold constant the level of living of homeowners whenever the rate of change of the price of housing is different from the rate of change of other prices. A COLI might still be useful as a price index in other contexts, and one might decide on other grounds not to treat homeowners differently from anyone else (or asset holders, for example), but the COLI would no longer be the correct cost-of-living index for homeowners.

Different problems arise when one seeks to compensate people for only part of their income. This issue arises most immediately for social security benefits. Many social security retirees have other income, in some cases substantially exceeding their social security benefits. When Congress legislated to protect these benefits, its intent was to protect the benefits themselves, not the total income of the recipients. It is not immediately obvious how to design a COLI for this purpose. In particular, one does not want an index that holds constant the standards of living of social security recipients supported by more than social security benefits. There are a number of possible approaches to this issue; perhaps the simplest is to define the COLI in terms of the costs of maintaining living standards for those who have no income other than social security benefits.

It seems quite unlikely that it would be worthwhile in practice to try to design price indexes that deal with homeowners separately from renters, or separately for social security recipients with or without other income. Nevertheless, this discussion highlights the fact that, even in the area for which it seems best suited—compensation—the cost-of-living index is not as obvious a choice as at first appears.

Stocks and Flows

Both basket and cost-of-living indexes are constructed from purchases and prices of goods. As we discussed above, the definition of goods cannot be taken for granted in a world of quality change. One property of a good, which can be thought of as an aspect of quality, is the length of time it lasts. For many goods, it is reasonable to use the convenient fiction that consumption happens at the moment of purchase. But for long-lived items like automobiles or houses, consumption is typically spread over several or many years. When computing a price index, it makes no sense to add together prices of durable and nondurable goods. Thus, one must use not the purchase price but the consumption price. For nondurable goods, they are the same thing, but for durable goods they differ. For durables, one needs an estimate of the cost of consuming the good for the same length of time for which one is looking at the consumption of nondurable goods. This concept is known as *user cost*. If the costs of buying and selling (the transactions costs) are ignored, it is calculated by finding out how much it would cost for someone to buy the good, use it for a year (or whatever is the specified period), and then sell it. This cost has three components: the interest foregone on the purchase price, the depreciation due to the physical wear and tear on the good, and any capital loss or gain other than wear and tear. At the time of purchase, the last two components are not known, so that user cost necessarily contains a speculative or expectational element.

The prices of durable goods should be converted to user cost before being aggregated into a price index, whether a basket price index or a COLI. The quantity that is used for pricing, directly in the Laspeyres or through the various approximations for a COLI, should be the *stock* of the good, because user cost is the cost of holding that stock for the year. Note that user cost introduces the (nominal) rate of interest into the consumer price index; user costs are higher at higher interest rates, as are both the cost-of-living and (properly computed) basket price indexes. (If higher nominal interest rates are a product of general price inflation, they will be largely offset in user cost by the expected price increase of the durable good over the holding period.)

There are a number of practical issues to do with user cost. One is whether to calculate it directly and, if so, what interest rate to use and how to proxy the expected capital loss or gain. If a car rental company faces the same costs of owning the car as does a consumer, so that competition sets price equal to cost, car rental rates ought to be close to user costs, and the "rental equivalent" can be used to construct the price index. If transactions costs are important, or are different for people and for rental companies, or if renters treat rental cars differently than their own cars, the two measures will not be equivalent. Currently, BLS treats cars as nondurable and works with user cost only for the "price" of owner-occupied housing. (Nonowners pay rents, which can be used directly.) If user costs are to be constructed directly, there are a number of practical issues associated with the treatment of capital gains. One possibility is to ignore them, but this causes problems when inflation rates are high. Another possibility is to use actual ex-post price movements, but this runs the risk of incorporating volatility into the index, and indeed of imputing *negative* user costs, and potentially even a negative CPI. Using ex-post capital gains also induces a perverse inverse relationship between the price of the durable and its user cost, at least in the short run. A CPI index that falls when complaints about the unaffordability of housing are the loudest would have difficulty gaining public acceptance. It would also be possible to forecast capital gains, or more simply just smooth the ex-post price changes. As is the current practice with housing, we believe that using rental rates is probably the best option.

It is possible to imagine moving to a user cost basis, not only for housing and cars but for other durable goods, such as household appliances and furnishings, electronic equipment, and even clothing. The whole concept of user cost ignores the fact that, if some people cannot get loans, not everyone has access to these goods by paying the user costs. How far to extend the user cost approach remains an important issue for BLS.

CONCLUSIONS

The arguments of this chapter reappear in subsequent chapters of this report as we deal with specific topics. On the basis of our discussion in this chapter, we present two general conclusions, largely about the conceptual basis for price and cost-of-living indexes, which serve to guide our more detailed conclusions and recommendations in the rest of this report.

> **Conclusion 2-1: An unconditional cost-of-living index is an unsuitable conceptual basis for the CPI. While research aimed at better understanding the economic effects related to changes in such matters as life expectancy, crime rates, or the environment would be useful for evaluating various aspects of public policy, the CPI should not change in response to changes in such factors.**

> **Conclusion 2-2: Within the general conceptual framework of cost-of-living indexes, the appropriate theoretical concept for the CPI is a *conditional* cost-of-living index that is restricted to private goods and services and in which environmental background factors are held constant.**

On the broader issue of assessing the relative merits of COGI and COLI conceptual approaches as a guide for construction of the CPI, various members of the panel strike the balance differently. All panel members find it difficult to think about the definition of goods and about quality change without considering what it is that consumers value, and we agree that it is impossible to think about substitution behavior without the concept of a constant standard of living that allows price changes to be converted into a monetary equivalent. For all these issues, especially the last, the cost-of-living framework is central. However, some panel members are skeptical about our ability to define a constant standard of living in an economy in which the nature of goods and services is constantly changing. They are therefore concerned about BLS adopting a conceptual framework that is not always well defined in the presence of quality change. They are also concerned about the BLS adopting an approach that differs from that of many statistical agencies around the world. All panel members do agree that the COGI and the conditional COLI that the panel recommends share many common aspects. We also concur that neither conceptual approach, viewed in its pure form, can provide the single guide to index construction. Rather, each of them can make a contribution toward dealing with the various problems that arise in designing the CPI. Taking a pragmatic approach, the panel found that it could come, sometimes by different routes, to unanimous agreement on all of the specific recommendations in this report. But in its inability to achieve unanimity behind a recommendation that the cost-of-living framework be the sole appropriate basis for construction of the CPI, our panel differs from the Stigler committee and Boskin commission.

TECHNICAL NOTE: A MATHEMATICAL APPROACH TO PRICE INDEXES

Notation, Laspeyres, and Paasche Indexes

We start by introducing some notation for the variables that most concern us, prices and quantities. In each period t, there are N goods, each of which has a price, p_n, and a quantity, q_n, with the subscript n labeling the good and running from 1 to N. We shall also need to refer to these prices and quantities in different periods, typically a *base or reference period*, denoted 0, and a later *comparison* or *current* period, denoted t. (We will occasionally separate base and reference periods later.) Superscripts refer to these time periods, so that q_n^t is the purchase of good n in period t. Sometimes we need to distinguish between purchases by different people, in which case we add another superscript h, for household. It is also occasionally useful to use vector notation, in which case subscripts are dropped; hence q is the (column) vector of N quantities and p the corresponding vector of N prices. Associated with the vector notation is the "dot" or inner product, $p \cdot q$, which denotes the sum of the element by element product of the vectors, in this case the total amount of money spent on q when it is bought at prices p.

Armed with only this notation, we can introduce the two most important *fixed-basket price indexes* or *cost-of-goods indexes*, or COGIs. For the Laspeyres price index, there is a base set of quantities, which we can denote q^0, which is repriced in successive periods. Hence, the *Laspeyres price index for period t*, which we denote P_L^t is defined by the equation

$$P_L^t = \frac{\sum_{n=1}^{N} q_n^0 p_n^t}{\sum_{n=1}^{N} q_n^0 p_n^0} = \frac{q^0 \cdot p^t}{q^0 \cdot p^0}. \tag{1}$$

In equation (1), the two sets of prices p^t and p^0 are compared using the base period quantities, q^0, as weights. Note that the numerator and denominator of (1) are identical, except that the prices in the numerator are current prices p^t, while those in the denominator are base period prices p^0. A useful alternative way to write the Laspeyres index is to define a price *relative* for each good. We write for good n

$$r_n^t = \frac{p_n^t}{p_n^0}, \tag{2}$$

which can be used to rewrite equation (1) in the form

$$P_L^t = \frac{\sum_{n=1}^{N} q_n^0 p_n^0 r_n^t}{\sum_{n=1}^{N} q_n^0 p_n^0} = \sum_{n=1}^{N} \frac{q_n^0 p_n^0}{x^0} r_n^t = \sum_{n=1}^{N} s_n^0 r_n^t, \tag{3}$$

where $x^0 = p^0 \cdot q^0$ is the total amount of money spent on all goods in period 0 and s_n^0 is the *expenditure share of commodity n in period 0*. According to (3), the Laspeyres can be thought of as a weighted sum of the price relatives, where the weights are the shares of the base period budget devoted to each of the goods. This way of thinking about the price index is useful because it shows so clearly how the Laspeyres "solves" the problem of making a single index in a situation where the price of each good has changed in a different way. Each of the N goods has its own rate of inflation, represented by its price relative. The Laspeyres averages these price relatives, each weighted according to the good's importance in the base period.

The *period t Paasche price index* is constructed in the same way as the Laspeyres but with the current basket replacing the base basket. Hence, replacing the base period quantities in (1) with the current period quantities, we have

$$P_P^t = \frac{\sum_{n=1}^{N} q_n^t p_n^t}{\sum_{n=1}^{N} q_n^t p_n^0} = \frac{q^t \cdot p^t}{q^t \cdot p^0}. \tag{4}$$

The Paasche index can also be written in terms of the price relatives and the budget shares, though the formulas are not quite so intuitive. Nevertheless it is easy to show that, instead of (3), we have

$$(P_p^t)^{-1} = \sum_{n=1}^{N} \frac{q_n^t p_n^t}{x^t} \left(r_n^t\right)^{-1} = \sum_{n=1}^{N} s_n^t \left(r_n^t\right)^{-1}, \tag{5}$$

so that the reciprocal of the Paasche is the current budget share weighted average of the reciprocals of the price relatives. If we take reciprocals of both sides of (5), we see that the Paasche index is a *weighted harmonic mean* of N price relatives, as opposed to the Laspeyres index in (3), which is a *weighted arithmetic mean* of the price relatives. If the price relatives are not all equal to one another, and under the special assumption that the expenditure shares in periods 0 and t are equal to one another, then a theorem of Hardy et al. (1934:26) implies that the Paasche index is strictly less than the Laspeyres.

The Paasche and Laspeyres price indexes are the two most familiar fixed-basket price indexes that can be used to measure price change going from period 0 to t. As we have presented them, there is no strong reason to prefer one over the other. However, from the viewpoint of statistical agency practice, there are strong reasons for preferring the Laspeyres. Both indexes require information on the price relatives, r_n^t, and statistical agencies are quite successful in collecting information on prices in a timely manner. However, while the Laspeyres price index requires information on base period expenditure shares, the s_n^0, the Paasche index requires information on current period expenditure shares, the s_n^t. With present methods of data collection, it is not possible to have accurate information on current period expenditure shares in a timely manner. *Thus, from a practical point of view, the preferred fixed-basket price index is the Laspeyres price index since it can be evaluated in a timely manner.* We note another advantage of the Laspeyres price index over its Paasche counterpart in the context of indexation of incomes below.

Averages of Fixed-Price Indexes

After a lag of about 2 years, it becomes feasible to evaluate the Paasche price index.[1] Hence, in the context of making price comparisons over the long run, we have (at least) two different measures of price change between periods 0 and t: the Laspeyres estimate of price change, P_p^t, and the Paasche estimate of price change, P_p^t. These are conceptually different, because they price different bundles over time, and in some cases the distinction may be important, and statistical agencies might wish to make both of these indexes available to the public. However, suppose that for practical or political reasons we need a single estimate of price change between periods 0 and t; is there a "best" such estimate? Obviously, there are many possible approaches to answering this question. We consider two simple and intuitive approaches.

The first way of combining the Paasche and Laspeyres measures of price change is to take some sort of an average, which we write in the form $m(P_L^t, P_P^t)$ so that we can write the new index as

$$P(p^t, p^0, q^t, q^0) = m(P_L^t, P_p^t). \tag{6}$$

We want this average to treat both price indexes symmetrically, to be positive, to be linearly homogeneous in both price indexes, and to be equal to either one when they are the same. In addition, we would like our new index to satisfy the *time reversal test*, which says that a price index from 0 to t should be the reciprocal of the price index from t to 0, so that

$$P(p^t, p^0, q^t, q^0) = 1 / P(p^0, p^t, q^0, q^t). \tag{7}$$

[1]We are assuming here that either the consumer expenditure survey is conducted on a more or less continuous basis or national accounts data, in conjunction with periodic consumer expenditure surveys can be massaged to obtain continuous consumer expenditure weights.

Equation (7) means that it does not matter which period we regard as the base period; we obtain essentially the same answer either way. Since the choice of which period to regard as the base is essentially arbitrary, other things being equal, we would like our price index to satisfy the time reversal test. (It is worth noting that neither the Laspeyres nor the Paasche price index satisfies the time reversal test.)

Diewert (1997:138) showed that only one average satisfies all the properties listed. This is the geometric mean (the square root of the product) of the Paasche and the Laspeyres

$$P_F^t = (P_L^t P_P^t)^{1/2}. \tag{8}$$

The price index defined by (8) is known as the *Fisher (1922) ideal price index*. The foregoing argument provides one justification for thinking of the Fisher price index as a "best" estimator of price change between periods 0 and t.

An alternative approach to combining the Paasche and the Laspeyres is to average not the indexes themselves but the two different baskets that go into them, an approach that was originated by Walsh (1901, 1921) and Knibbs (1924). If we use a geometric mean of the two baskets, we obtain the *Walsh price index*, P_W, written as

$$P_W^t = \frac{\sum_{n=1}^{N} (q_n^t q_n^0)^{1/2} p_n^t}{\sum_{n=1}^{N} (q_n^t q_n^0)^{1/2} p_n^0}. \tag{9}$$

If we replace the geometric mean in (9) with the simple arithmetic mean, we reach yet another index in the Walsh-Knibbs family, known as the *Marshall Edgeworth price index* (Marshall, 1887; Edgeworth, 1925).

Aggregation: Democratic and Plutocratic Indexes

We have been careful so far not to distinguish individual from aggregate quantities. Paasche and Laspeyres indexes can be equally well constructed using individual baskets or aggregate (or average) baskets. In this section, we consider the relationships between these various types of Laspeyres and Paasche indexes under the assumption that each household faces the *same* vector of prices in each period.

Suppose that there are H households in the economy. Household h's period t Laspeyres index can be written following (3) but with the household superscript h in the form

$$P_L^{th} = \sum_{n=1}^{N} \frac{q_n^{0h} p_n^0}{x^{0h}} r_n^t = \sum_{n=1}^{N} S_n^{0h} r_n^t, \tag{10}$$

where q^{0h} is the vector of purchases for household h in the base year 0, x^{0h} is its total expenditure in period 0, and s_n^{0h} is its share of total expenditures on good n in period 0. If, by contrast, we evaluate the national Laspeyres index using the aggregate bundle for all households, we would have

$$P_L^{tA} = \sum_{n=1}^{N} \frac{Q_n^0 p_n^0}{X^0} r_n^t = \sum_{n=1}^{N} S_n^0 r_n^t, \tag{11}$$

where the superscript A denotes "aggregate," Q_n^0 is the aggregate quantity defined as the sum of the individual quantities, X^0 is aggregate expenditures on all goods and services, again the sum of the individual x^{0h}, and S_n^0 is the share of aggregate expenditures on good n.

Both individual and aggregate Laspeyres indexes are weighted averages of the same price relatives, and the formulas (10) and (11) differ only in the weights. The aggregate index (11) uses the shares in the national budget, while the individual index (10) uses the shares in the household's budget. The two sets of weights can be related to one another by noting that

$$S_n^0 = \frac{p_n^0 Q_n^0}{X^0} = \frac{\sum_h p_n^0 q_n^{0h}}{X^0} = \sum_h \frac{x^{0h}}{X^0} \frac{p_n^0 q_n^{0h}}{x^{0h}} = \sum_h \frac{x^{0h}}{X^0} s_n^0, \tag{12}$$

so that the shares in the national budget are the weighted average of the shares in each household's budget, where the weights are each household's total expenditure as a share of national total expenditure. People who spend a lot count more in the national weights than do people who spend a little. Given (12), the individual and national Laspeyres indexes are related by

$$P_L^{tA} = \sum_h \frac{x^{0h}}{X^0} P_L^{th}. \tag{13}$$

Equation (13) is the reason why the aggregate Laspeyres is referred to as a *plutocratic* index; each household's individual Laspeyres price index is weighted by the total amount of money that it spends in period 0. This is in contrast to a *democratic* Laspeyres index in which each household's index is averaged to obtain the national index

$$P_L^{tD} = \frac{1}{H} \sum_h P_L^{th}. \tag{14}$$

Note that the democratic and plutocratic Laspeyres indexes will coincide if everyone has the same income, or if everyone spends their money in the same proportions over the different goods, or if all the price relatives are equal.

Note finally that, if we combine (13) and (14), we can write

$$P_L^{tD} = \frac{1}{H} \sum_h \sum_{n=1}^{N} s_n^{0h} r_n^t = \sum_{n=1}^{N} \bar{s}_n^0 r_n^t, \tag{15}$$

where $\bar{s}_n^{\,0}$ is the simple average over households of the budget shares in the base period. Equation (15) shows that the democratic Laspeyres can be estimated if we can calculate, in addition to the price relatives, the population average of household budget shares, something that can be estimated from a consumer expenditure survey, such as the Consumer Expenditure Survey (CEX). We note that the above average can be repeated for the period t Paasche index, though the calculations are not so straightforward. The national Paasche index, which uses the national aggregate bundle at time t to compare prices at 0 and t, turns out to be a weighted harmonic mean of the individual Paasche indexes. Parallel to the Laspeyres, the weights are plutocratic weights, now the ratios x^{th}/X^t, the shares of each household in aggregate national expenditure on all commodities in the domain of the index but now in period t. We can also define a democratic Paasche index as the simple average of the individual Paasche indexes. However, there is no formula corresponding to (15) for the democratic Paasche index. In consequence, it cannot be calculated by weighting the price relatives by an average of the expenditure shares; instead, it must be calculated directly by averaging the individual Paasche indexes.

What are the merits and demerits of the plutocratic versus democratic price indexes? The democratic indexes, which give each individual an equal weight in the overall index, are the natural indexes for the analysis of welfare when we want each person to count the same rather than in proportion to his expenditure. By contrast, when we want every dollar to count the same, as, for example, when we are calculating the national accounts, the plutocratic indexes are the natural choice. Each family of indexes has its own justification.

Note finally that the arguments in the second section can be repeated in the present context leading, for example, to the use of the plutocratic Fisher ideal index as a good candidate for combining the information in the plutocratic Paasche and Laspeyres indexes into a single measure of the change in prices from 0 to t.

Cost-of-Living Indexes

In the economic theory of consumer behavior, each household (or person) is assumed to spend their money so as to be as well off as possible. The way this is formalized is by writing down a *utility function* whereby the level of utility (or level of living) is determined by the vector of quantities consumed

$$u = f(q). \tag{16}$$

The main role of the utility function is to codify consumer preferences; by inserting any quantity vector q into (16) we can test whether it is better than, the same as, or worse than any other quantity vector, and this ranking tells us the consumer's preferences over goods. The value assigned to u itself is of no significance; provided higher u means a better bundle, it does not matter what particular values are assigned to u. More important is the concept of an *indifference curve* or *indifference surface*; this is a collection of q's all of which yield the same value of the utility function. They are therefore bundles between which the consumer is indifferent. Higher indifference curves are those with a higher value of u and correspond to a higher standard of living.

The most useful concept for cost-of-living theory is the *cost* or *expenditure function*, which measures the least amount of money that the consumer would have to pay at specified prices to reach a specified indifference curve. We write this function as $c(u,p)$ where, as before, p is the vector of prices, and u is some arbitrary label that identifies the indifference curve. Given that the consumer has a total x to spend, and given the assumption that she spends that money to do as well as possible, we can write

$$x = c(u, p). \tag{17}$$

Note that this function also can be thought of as defining u, the standard of living, in terms of the prices p, and total expenditure x.

Cost-of-living index numbers are defined directly from the cost function. Suppose that the base period level of living is u^0. The cost-of-living index number using base period level of living is the ratio of the costs of reaching the indifference curve u^0 at the two sets of prices, p^0 and p^1. Hence,

$$p_{C0}^{t} = \frac{c(u^0, p^t)}{c(u^0, p^0)} \tag{18}$$

is the cost-of-living analog to the Laspeyres index (1). Both indexes compare the current prices in the numerator with the base prices in the denominator. Because $c(u^0,p^0) = x^0 = p^0 \cdot q^0$, the denominators of (1) and (18) are the same. However, the numerator of the Laspeyres is the cost of the base basket q^0 evaluated at period t prices p^t, while the numerator of the cost-of-living index is the minimum cost of obtaining the base period indifference curve at prices p^t. If instead of the base indifference curve in (18), we use the current indifference curve, we get the cost-of-living index corresponding to the Paasche index, which is

$$P_{Ct}^{t} = \frac{c(u^t, p^t)}{c(u^t, p^0)}. \tag{19}$$

Each of the two cost-of-living indexes (18) and (19) involves a *counterfactual* cost; in (18) it is the minimum cost of reaching u^0 at prices p^t, while in (19) it is the minimum cost of reaching u^t at p^0. Although we do not immediately know what these counterfactuals are, we can set limits on them. In particular, since *one* way of reaching u^0 is to buy the original bundle q^0, the minimum cost of reaching u^0 at p^t can be no larger than the cost of that bundle at the current prices, which is $q^0 \cdot p^t$. Similarly, one way of reaching u^t at the original prices is to buy the bundle q^t, so that the minimum cost of u^t at p^0 can be no larger than $q^t \cdot p^0$. Hence, if we go back to the definition of the base period cost-of-living index (18) and note that the minimum cost of u^0 at prices p^0 is the actual expenditure $q^0 \cdot p^0$, we have

$$P_{C0}^{t} = \frac{c(u^0, p^t)}{c(u^0, p^0)} \geq \frac{q^0 \cdot p^t}{q^0 \cdot p^0} = P_L^t, \tag{20}$$

so that the base period cost-of-living index is always no larger than the Laspeyres. For the current period true cost-of-living index, the hypothetical cost is in the denominator, so that replacing it by something larger will make the result smaller. Hence, using (20) and this inequality, we have

$$P_{Ct}^{t} = \frac{c(u^t, p^t)}{c(u^t, p^0)} \geq \frac{q^t \cdot p^t}{q^t \cdot p^0} = P_p^t, \tag{21}$$

so that the current period cost-of-living index P_{Ct}^{t} is always at least as large as the Paasche price index P_p^t.

There is an immediate link between each of these cost-of-living indexes and a measure of compensation. The amount of money that the consumer needs to reach the base level of living at the current prices is simply $c(u^0,p^t)$ so that the (possibly negative) compensation that the consumer requires to make up for the price change from p^0 to p^t is given by

$$CV = c(u^0, p^t) - c(u^0, p^0). \tag{22}$$

This quantity is known as the *compensating variation*. It is the difference between the same two costs whose ratio is the cost-of-living index for the base level of living P_{C0}^{t}. We can also construct the *equivalent variation*, defined as the maximum amount of money that the consumer would have been prepared to pay in the base situation to avoid the price change from p^0 to p^t. It is

$$EV = c(u^t, p^t) - c(u^t, p^0) \tag{23}$$

and bears the same relationship to the cost-of-living index for the current period level of living u^t as does the compensating variation to the cost-of-living index for the base period level of living u^0. To illustrate how these measures work with the cost-of-living indexes, suppose that a consumer's base level of total expenditures is $x^0 = c(u^0,p^0)$ and that we escalate this by the base period cost-of-living index (20). The new escalated total expenditure will be $c(u^0,p^t)$, so that the escalation pays the compensating variation (22) and exactly compensates the consumer for the change in prices. If in the absence of the cost-of-living index, we escalate by the Laspeyres price index P_L^t, the consumer will have *at least* as much as needed to remain as well off. If the object of policy is to ensure that compensation is adequate, and if it is better to compensate too much than to compensate too little, this would be an argument for the use of the Laspeyres price index for escalation.

In principle, we can construct a cost-of-living index around *any* level of living. We might write this arbitrarily based cost-of-living index in the form

$$P_{Cu}^{t} = \frac{c(u, p^{t})}{c(u, p^{0})} \tag{24}$$

for some indifference curve u. From inspection of (24), it is clear that if we want cost-of-living indexes to be the same whatever the choice of u, the cost function must factor into two components, one containing only u and one containing only the prices p, so that we can write

$$c(u, p) = \theta(u)\gamma(p). \tag{25}$$

This condition, known as *homothetic preferences*, implies that the pattern of demand, the way the budget is spread over goods, is the same at all levels of living, something that is not in accord with the empirical evidence. In general then, we can expect cost-of-living indexes to depend on the level of living on which they are based, so that the base period cost-of-living index will be different from the current period cost-of-living index, and cost-of-living indexes will be different for the poor and for the rich. It is only under homotheticity that it is possible to talk about the "true" cost-of-living index, since it is only then that it will be unique, and it is only then that it can be correctly asserted that the "true" cost-of-living index always lies between the Paasche and the Laspeyres (or even that the Laspeyres is always greater than the Paasche).

When preferences are not homothetic, we cannot calculate observable bounds for the two COLIs (20) and (21). However, Konüs (1924:20) proved that there exists a utility level u^*, intermediate between u^0 and u^1, whose cost-of-living index (24) lies between the Paasche and the Laspeyres indexes. Hence, if the Paasche and Laspeyres indexes are not far apart, an average of them, such as the Fisher ideal index (8), is likely to be a good approximation to a COLI such as (24), whose reference standard of living is between the base and current period standards of living.

How can we recover the cost function and the associated cost-of-living index numbers from observable behavior in the market? What we typically see is the relationship between each period's quantities purchased, the q's, their prices p, and the incomes (or total outlays) of consumers. Suppose that we can do so for a single consumer and that we can recover, by experimentation or econometric analysis, the n functions, one for each good

$$q_n = g_n(x, p). \tag{26}$$

According to (26), each of the N purchases is a function of total expenditure x and the vector of N prices p. The cost function is also directly linked to the quantities purchased, and the crucial result here, known as *Shephard's Lemma* (Shephard, 1953:11), states that the quantities are the partial derivatives of the cost function with respect to prices

$$q_n = \frac{\partial c(u,p)}{\partial p_n}. \tag{27}$$

Intuitively, for small changes in price, the effect of the cost of living of a price increase is equal to the amount of the good purchased; if one buys a hamburger every day, an increase of a cent in the price of a hamburger raises one's weekly cost of living by seven cents (provided that we do not substitute hot dogs for hamburgers, something that will not be important for sufficiently small changes in price).

Comparing the demand functions (27) with Shephard's Lemma (26), and noting that expenditure is equal to the cost of living, equation (13), we can write

$$\frac{\partial c(u,p)}{\partial p_n} = g_n[c(u,p),p]. \tag{28}$$

Equations (28) is a set of n partial differential equations whose solution, given knowledge of the observable functions g_n, gives the cost function which, in turn, can be used to construct the cost-of-living indexes. While not all such systems of partial differential equations have a solution at all, (28) will always have a solution if the demand functions from which we begin, equations (26), come from a consumer who is obeying the theory of consumer behavior.

Practical algorithms for calculating the cost function have been worked out in the literature; a simple example is given in Hausman (1981), while a more comprehensive treatment can be found in Vartia (1983). However, these methods cannot be recommended as a practical method for statistical agencies to construct cost-of-living index numbers. The functions (26) must be estimated, which involves estimating the derivatives of each demand function with respect to total expenditures and the prices of all goods, not to mention the other factors that condition consumer behavior. In practical price indexes, there are a large number of goods, so this is a formidable undertaking. Although the theory of consumer behavior provides some help in this task, estimation is not possible without a host of additional assumptions about the structure of preferences, as well as about econometric identification, many of which are not easy to defend. There is therefore a considerable payoff to any method that avoids altogether the need to obtain demand functions.

Superlative Indexes

It is useful to start by recalling the Fisher ideal index, P_F^t, defined by equation (8). Fisher proposed his index because it passes a number of desirable tests not rooted in cost-of-living theory. But it turns out that the Fisher index is a cost-of-living index for a specific utility function, and its associated cost and demand functions. In particular, if the utility function takes the form

$$f(q) = \left(\sum_n^N \sum_m^M a_{nm} q_n q_m \right)^{1/2} \tag{29}$$

for some matrix $A \equiv [a_{mn}]$ (which must be symmetric and have a single positive eigenvalue), then the Fisher ideal index (8) is *exact* in the sense that if we calculated the cost function associated with (29) and used it to calculate the COLIs (18), (19) or (24), we would obtain (8) (Byushgens, 1925). If the matrix A has an inverse B, say, the cost function associated with (29) takes the form

$$c(u, p) = u \left(\sum_n^N \sum_m^M a_{nm} q_n q_m \right)^{1/2}. \tag{30}$$

(If A is not invertible, (30) will still lead to the Fisher ideal index.) The demand functions associated with equation (30) can be written in the form

$$q_n = \frac{x \sum_k a_{nk} p_k}{\sum_j \sum_k a_{jk} p_j p_k}. \tag{31}$$

The remarkable thing about this result is not that it is possible to find a cost function and a set of demand functions that justify a given price index, but the fact that the result is so general. Although preferences (29) are homothetic—and indeed we can see directly from (30) that the cost function is the product of utility and a function of prices, or from (31) that the shares of the budget $p_n q_n / x$ are independent of x—the matrices A and B are not specified, except that they must be symmetric and have a single nonnegative eigenvalue, a requirement that comes from the general theory of consumer demand and guarantees, among other things, that demand curves slope down. As a result, and always subject to homotheticity, the demand functions (31) allow the consumer to respond to price changes in a general way; the price elasticities of demand from (31) are unrestricted, except by the general restrictions of consumer theory. The Fisher ideal index is therefore *exact* for a set of preferences and demand functions that do not restrict *substitution* behavior in ways beyond that required for the theory. It therefore permits a way of computing a general cost-of-living index without having to estimate the demand functions.

Diewert (1976) extended and generalized these results. A particular specification of preferences, or of the cost function, is said to be a *second-order flexible functional form* if the utility (or cost) function can provide a second-order approximation to an arbitrary utility (or cost) function. A *superlative* price index is then one that is *exact* for some second-order flexible functional form for either the cost or utility function but with preferences restricted to be homothetic. Diewert showed that the utility and cost functions (29) and (30) are flexible for homothetic preferences, so that the Fisher ideal index is an example of a superlative index.

There are many other superlative indexes, for example, the Törnqvist index P_T^t defined by

$$\ln P_T^t = \sum_{n=1}^{N} \frac{1}{2}\left(s_n^0 + s_n^t\right)\ln\left(\frac{p_n^t}{p_n^0}\right), \tag{32}$$

which is exact for the *translog* cost function, in which the logarithm of costs is a quadratic form in the logarithms of prices. The Walsh price index (9) is exact for a utility function that is a quadratic form in the square roots of the quantities; it too is therefore a superlative index. Diewert (1978) shows that these three superlative price indexes approximate one another to the second order around any given price-quantity combination, so that the choice between them is unlikely to matter much in practice.

The Fisher ideal index is computed from both the Paasche and Laspeyres, and thus requires information on both base period and current baskets. The (logarithm of the) Törnqvist index (31) is a weighted average of logarithmic price relatives, with weights that are the average of current and base period patterns of demand. Indeed, superlative indexes *always* require both current and base period quantity information. Intuitively, their ability to capture the substitution effects of prices has to be based on observation of the effects of the price change, which requires data on demand both before and after the change.

The analysis so far has been entirely within the framework of homothetic preferences, something that is unattractive in practice. It is possible to accommodate nonhomotheticity at the price of interpreting the superlative index as the cost-of-living index for some specific intermediate level of living. For example, Diewert (1976:122) showed that the Törnqvist price index is exact at the level of utility that is the geometric mean of the utility in periods 0 and t.

Aggregation of Cost-of-Living Indexes

The analysis of the passage from individual to aggregate indexes is essentially identical to the same analysis for the basket price indexes in the second section of these notes. Nevertheless, it is worth defining Pollak's (1980, 1981) social cost-of-living index which is the ratio of the *aggregate* cost of obtaining the base levels of living at current prices to the *aggregate* cost of obtaining the base levels of living at the base period prices. Hence, adding superscripts h to denote individual households

$$P_{S0}^t = \frac{\sum_h c^h(u^{0h}, p^t)}{\sum_h c^h(u^{0h}, p^0)}, \tag{33}$$

where $c^h(u^h,p)$ is the cost function of household h—note that there is no requirement that different households have the same preferences—and u^{0h} is the label for household's h's indifference curve in the base period. Following through the earlier analysis, it is easily seen that the social cost-of-living index (33) is a weighted average of the individual (base period) cost-of-living index numbers, with each household weighted by its total expenditure on goods and services:

$$P_{S0}^{t} = \sum_{h} \frac{x^{0h}}{X^{0}} P_{C0}^{th}. \tag{34}$$

The social cost-of-living index, like the aggregate Laspeyres, is a plutocratic index.

We will not work through the results here, but it is intuitively clear—and true—that we can define a social cost of living around current living standards, and that this too is a plutocratic average of the individual current period cost-of-living indexes. The inequalities between the Paasche and Laspeyres and their corresponding cost-of-living indexes all carry through to the corresponding aggregate and social cost-of-living indexes. We can also define superlative indexes from the social aggregate indexes, such as an aggregate Fisher ideal index, and show that they are exact for social cost-of-living indexes when individual consumers have preferences that are second-order flexible functional forms. For formal demonstrations of this material, see Diewert (2000a). Finally, the whole process can be repeated using democratic instead of plutocratic indexes.

Conditional COLIs, Quality Change, and Health

As we emphasize in the main text, the use of COLIs as price indexes often requires us to ensure that a COLI changes only when prices change, and not when there are changes in the myriad other factors that affect the cost of living. In the text, this is what we refer to as the "domain" issue, that the COLI be a function of the prices of the goods and services that people buy, and not change with such things as the provision of public goods, people's tastes, their family composition, the crime rate, the ambient temperature, or the number of years that they can be expected to live. Yet all of these things affect people's well-being, so that we must formally modify the theoretical framework to allow for their existence. We capture those nonmarket influences on living standards through a vector of "environmental" factors, labeled e, which differs from household to household, and we recognize their effect on utility by writing the utility function in the form $u_h = f^h(q^h,e^h)$. The dependence on e carries through to the cost function, which becomes $c^h(u^h,p,e^h)$. We can then follow the example of Caves, Christensen, and Diewert (1982) and Pollak (1989) and define household h's *conditional cost-of-living index* between periods 0 and t as

$$P_{Cue} = \frac{c^h(u,e,p^t)}{c^h(u,e,p^0)}. \tag{35}$$

The important thing to note here is that, not only is the level of utility held constant between the numerator and denominator of (35), but so also is the level of the environmental variables e. As a result, changes in e from 0 to t do not affect the index. For example, if the winter is colder in t than in 0, so that more fuel must be bought to keep living standards the same, (35) will not show an increase in the cost of living unless prices change. It is a price index that is *conditional* on the temperature or other environmental factors. If prices remain the same in the two periods, so that $p^t = p^0$, the price index will be equal to unity. As discussed in the text, these properties are just what we want in a *price index*; whether they are appropriate for a cost-of-living index is a more controversial question.

Two special cases of (35) are of particular interest: the Laspeyres-type conditional COLI, in which u and e are replaced by u^0 and e^0, and the Paasche-type conditional COLI, in which u and e are replaced by u^t and e^t. It is a routine exercise to check that all of the results and apparatus developed so far apply to these concepts, including the bounding relationships, the construction of superlative indexes, and the aggregation of price indexes to the national level. The results that involve a utility level intermediate between u^0 and u^t, for example, for superlative indexes in the nonhomothetic case, now involve intermediate levels of *both* e and u.

One important use of a conditional COLI is to help us think about the difficult issue of quality change. For example, if a computer costs the same today as it did yesterday but works faster and has more features, a price index that did not control for quality would not capture the effective fall in price. By contrast, a conditional COLI, which treated quality as one of the environmental goods and held it constant from 0 to t, would give a better answer. As will be argued in Chapter 4, using a conditional COLI in this way is straightforward when we know what quality change is and can measure it. Matters become more complicated when quality is not readily observed, or when we do not know the source of quality improvement. In the rest of this section, we provide an example from the important case of health care. This example illustrates how conditional COLIs work in a concrete case, as well as showing that getting the adjustment right can be very difficult in practice.

We start from a utility function in which "health" h is one argument and the vector of other goods q is another, so that the (unconditional) utility function can be written

$$u = f[h, q]. \tag{36}$$

where u denotes utility including health, not just the well-being from goods and services. The quantity h is a latent variable "health status," which determines life, death, and morbidity. More of it is better. Consumers have budget x which has to cover health (or medical) purchases m at price p_m as well as the vector of other goods q at price p. The budget constraint is then

$$x = p \cdot q + p_m m. \tag{37}$$

Health is getting better over time in some disembodied way and is also improved by purchases of health goods m. We assume that the effectiveness of health goods in improving health also changes over time through an efficiency parameter θ. Taking these together, we can write health status at time t as

$$h_t = \delta_t + \theta_t m \tag{38}$$

where δ_t is the cumulated effects up to the beginning of t of the disembodied health progress, and θ_t is the efficiency of health goods and services m in producing health. Examples of δ would be improvements achieved through better childhood nutrition, lower pollution, or reductions in smoking. Combining (37) and (38), we can rewrite the budget constraint as

$$x + \frac{p_m \delta_t}{\theta_t} = p \cdot q + \left(\frac{p_m}{\theta_t} \right) h_t \tag{39}$$

so that the disembodied technical progress δ_t acts like a gift of income (though because it works by reducing the need to purchase health care, its value is reduced the cheaper or more efficient health care is), and the "effective" price of health care is its quality-adjusted price p_m/θ_t. In this set-up, the disembodied improvements in health status increase utility at any given set of prices and thus reduce the (unconditional) cost of living. Writing the budget constraint in the form of (39) allows us to see the consumer's problem as a standard one; utility (36) is defined over q and h, and (39) gives their effective prices, p and p_m/θ_t, as well as the effective budget available to fund them, $x + p_m\delta_t/\theta_t$. Given this, we can immediately see that the unconditional cost function—the minimum cost of reaching u (including both health status and consumption) at prices p_m and p can be written in the form

$$c(u, p_m, p) = \tilde{c}\left(u, \frac{p_m}{\theta_t}, p \right) - \frac{p_m \delta_t}{\theta_t}. \tag{40}$$

From (40) we see that (a) p_m always appears deflated by the efficiency parameter θ_t, so that only the effective price matters, and (b) an increase in disembodied technical progress δ_t decreases the cost of living. The efficiency parameter reduces the price of health care, while the disembodied parameter effectively generates additional income.

Suppose that, in line with our discussion of the domain issue in the main text, we decide that the COLI price index *should not* fall in response to disembodied improvements in health status but *should* fall when new medical procedures or drugs mean that a given episode of illness can be treated at less cost. In the framework here, this decision can be implemented by including δ_t among the environmental variables, e, and holding it constant in cost-of-living comparisons while allowing θ_t to change in comparisons from 0 to t, so that we compare, not the prices p_m^0 and p_m^t, but the quality-corrected prices, p_m^0/θ_0 and p_m^t/θ_t. The conditional cost function that we need to make this work is (40) with δ_t held constant,

$$c(u, p_m^t, p^t) = \tilde{c}\left(U, \frac{p_m^t}{\theta_t}, p^t\right) - \frac{p_m^t \delta_0}{\theta_t}, \tag{41}$$

which no longer changes unless there is a change in price or, more precisely, a change in an "effective" or quality-adjusted price. Although it might seem odd to treat the two sources of technical progress asymmetrically, it can readily be defended as making the distinction between a price change and an income change. In our usual income accounting, we regularly treat income increases differently from price reductions, and that is exactly what is happening here. The part of technical progress that makes health care more efficient is properly counted as a price reduction, while the part that rains down from heaven (or at least is unconnected with current health care provision) is an increase in income; see again (39). Equation (41) is the conditional cost function that would be used to calculate the conditional COLI price index, by insertion into equation (35).

The problem with this approach is an empirical one, that it is very difficult to separate out the two kinds of technical progress. More people are surviving heart disease, and mortality rates are falling rapidly among the age groups most at risk. This outcome could result from better treatment, which is an increase in efficiency and which should rightly be counted as an increase in θ and as a decrease in the effective price of treating heart disease. But it could also be that people are surviving heart disease more frequently because of improvements in some background factor (e.g., they are smoking less or were better nourished in utero), without any increase in efficiency of care, even though its cost is increasing. The argument about causation, between background social factors on the one hand and technical change on the other, has been inconclusively debated in the literature for at least the past 30 years, so it is difficult to think that we can get the assignment right. If we get it wrong and attribute the effects of background factors to medical improvements, we will understate the increase in the price level. And because health status is not included in the National Income and Products Accounts, there is no offsetting effect in the underestimation of income. It is not hard to imagine a situation in which the costs of health care services are rising rapidly, driven by the introduction of new technologies and new drugs. And even if the innovations were not effective, mortality might be falling for other reasons, like the cessation of smoking or improvements in nutrition a long time ago. In this case, the price increase in medical care is real and quality correction would be the wrong thing to do, masking or eliminating the true increase in the price of health care.

The situation is complicated further by the fact that people rarely choose the quantity of their health care, setting price proportional to marginal benefits, but usually have it chosen for them, by a physician or by the combination of an insurer and a physician. Abstracting from the personal contribution to health status, through behavioral choices, we can imagine that health status is set at some level $\bar{h}$ different from what would have been chosen by weighting price against benefit. In this case, the budget constraint (39) becomes

$$x + \frac{p_m \delta_t}{\theta_t} - \left(\frac{p_m}{\theta_t} \right) \bar{h} = p \cdot q \tag{42}$$

so that the fixed amount of health care is simply a charge on the budget for other goods. The conditional cost function with preset health care, sometimes referred to as the *rationed* cost function, written $\bar{C}(U, p_m, p, \bar{h})$, can be linked to the conditional cost function with chosen health care by a linear approximation around the free choice,

$$\bar{c}(u, p_m, p, \bar{h}) \approx c(u, p_m, p) + \frac{(p_m - p_m^*)}{\theta_t}(\bar{h} - h^*), \tag{43}$$

where h^* is the optimal health status for a consumer who is taking price into account and choosing for him- or herself, and p_h^* is the shadow price (willingness to pay) for health care at the margin. When $\bar{h} = h^*$, the shadow and actual prices coincide, but when more health care is provided than would have been chosen in the market, the shadow price is below the market price, so that the last term on the right-hand side of (43) is positive. According to this there is an additional element to the cost of living associated with "overconsumption" of health goods, for example, through point-of-purchase price being low or other considerations. This term is also not taken into account under any of the proposals we are considering and, if present, would further exacerbate the understatement of the cost of living through the sort of effects discussed in the previous paragraph.

Taylor Series Approximations to Cost-of-Living Indexes

Although superlative indexes are better approximations, the Laspeyres index is often itself a useful approximation to the base period cost-of-living index. This depends on a result that we already have, Shephard's Lemma, that the derivatives of the cost function are the quantities, as well as on a result on substitution that we introduce here. If we differentiate Shephard's Lemma (27) for good i with respect to the price of good j, we obtain

$$\frac{\partial q_i(u, p)}{\partial p_j} = \frac{\partial^2 c(u, p)}{\partial p_i \partial p_j} \equiv s_{ij}(u, p). \tag{44}$$

The $N \times N$ matrix of these $s_{ij}(u,p)$ is denoted by S and is called the consumer's substitution matrix (sometimes called the Slutsky matrix) and (44) shows that its i,jth element is equal to the derivative of the demand for good i with respect to the jth price when the consumer is held on the same indifference curve. Such price derivatives are called substitution effects and abstract from the income effects also associated with price changes. They are the key to the substitution behavior that differentiates between basket and cost-of-living price indexes. In what follows, we will be evaluating S at the base level of utility and prices, u^0 and p^0; when we do so, we use the notation S^0 to denote $S(u^0,p^0)$.

The base period cost-of-living index number (18) uses the counterfactual cost of attaining the base period indifference curve at current prices, $c(u^0,p^t)$. One way to approximate it is to take a second-order Taylor series approximation around the point u^0,p^0. Using Shephard's Lemma (27) and (44), we can write this approximation as

$$c(u^0,p^t) \approx c(u^0,p^0) + \sum_{n=1}^{N} q_n^0 (p_n^t - p_n^0) + \frac{1}{2}\sum_i \sum_j s_{ij}^0 (p_i^t - p_i^0)(p_j^t - p_j^0). \quad (45)$$

Recall that $c(u^0,p^0) = p^0 \cdot q^0$ so that the first term on the right-hand side cancels with the second term in the first bracket so that, if we divide through both sides of (45) by $c(u^0,p^0)$, we get the following approximate relationship between the base period COLI, p_{C0}^t, and the Laspeyres index

$$P_{C0}^t - P_L^t \approx \frac{1}{2}\frac{(p^t - p^0)\cdot S^0 \cdot (p^t - p^0)}{p^0 \cdot q^0}. \quad (46)$$

Thus, the difference between the base period cost-of-living index and the Laspeyres price index is zero to the first order so that the Laspeyres is a first-order approximation to the base period cost of living. The approximate difference between them, the right hand-side of (46), depends on how much substitution is possible, which is represented by the matrix S^0 as well as by the size of the difference between the base and current price vectors. Note that, because the Slutsky matrix is a negative semidefinite matrix, the quadratic form on the right-hand side of (46) is nonpositive as we should expect, given that the Laspeyres is an upper bound for the base period cost-of-living index. Note also that p^0 lies in the nullspace of S^0, so that if period t prices are proportional to period 0 prices, the right-hand side of (46) will be zero. More generally, the right-hand side will be larger the more p^t deviates from p^0 in a nonproportionate manner.

We note that a similar approximation analysis can be carried out for the Paasche index and the current period cost-of-living index. We leave the details to the reader.

CES Price Indexes

Suppose that the consumer's cost function takes the form

$$c(u,p) = u\alpha_0 \left(\sum_{n=1}^{N} \alpha_n p_n^{1-\sigma} \right)^{1/(1-\sigma)} \quad (47)$$

when σ is not unity or

$$c(u,p) = u\alpha_0 \prod_{n=1}^{N} p_n^{\alpha_n} \tag{48}$$

when σ = 1. This cost function represents homothetic preference, and the corresponding utility function is the constant elasticity of substitution (CES) utility function introduced into the economics literature by Arrow, Chenery, Minhas, and Solow (1961). The parameter σ is the *elasticity of substitution*; when σ = 0, the unit cost function defined by (47) is linear in prices and hence corresponds to a fixed-coefficients utility function with zero substitutability between all commodities. When σ = 1, equation (48), the corresponding utility function is a Cobb-Douglas function. When σ tends to infinity, the corresponding utility function approaches a linear utility function which exhibits infinite substitutability between all commodities. Even within the class of homothetic preferences, the CES cost function defined by (47) and (48) is *not* a fully flexible functional form (unless the number of commodities is two), but it is more flexible than the zero substitutability utility function that is exact for the Laspeyres and Paasche price indexes.

The base period cost-of-living index associated with (47) takes the form

$$P_{CES}^t = \left(\sum_{n=1}^{N} s_n^0 \left(\frac{p_n^t}{p_n^0} \right)^{(1-\alpha)} \right)^{1/(1-\sigma)} . \tag{49}$$

Note that (49) is itself a CES function of the price relatives; in the mathematical literature, it is also known as the mean of order 1 – σ. When σ takes the value zero, (38) is the Laspeyres index; the Laspeyres is only a COLI when the consumer is unable (or unwilling) to substitute between goods, always consuming them in fixed proportions. As σ tends to unity, (38) tends to the base period expenditure share weighted geometric mean. Provided not all the price relatives are the same, the CES index (49) is monotonically decreasing as the elasticity of substitution increases from 0 to infinity. If some consumers have an extreme aversion to substitution so that their elasticity of substitution is 0, then as relative prices change from period 0 to *t*, they will face a higher cost of living than consumers who substitute toward commodities that have decreased in relative price. Hence, if the elasticity of substitution *s* is positive and prices in period *t* are not proportional to prices in period 0, the Laspeyres price index, P_L^t, will always be *strictly greater* than the corresponding CES price index, P_{CES}^t.

The CES cost-of-living index was first derived from CES preferences by Lloyd (1975), though it was Moulton (1996) who noted its usefulness for statistical agencies. In order to evaluate (50), the only requirements are information on the base period expenditure shares s_n^0, the price relatives p_n^t/p_n^0, and an estimate of the elasticity of substitution σ. The first two requirements are met by the standard information that statistical agencies use to evaluate the Laspeyres price index. Hence, if the statistical agency is somehow able to estimate the elasticity of substitution *s*, the CES price index can be evaluated using the same information used to evaluate the usual Laspeyres index.

How might the statistical agency obtain an estimate for the substitution parameter σ? Shapiro and Wilcox (1997:121-123) provide one method. They calculate superlative Törnqvist indexes for the United States for the years 1986-1995 and then the CES index for the same period using various values of σ. They then chose the value of σ (in this case 0.7) which caused the CES indexes to most closely approximate the corresponding Törnqvist indexes (which could be evaluated on a delayed basis).[2] Assuming that the Törnqvist index is more or less free from substitution bias, it can be seen that the Shapiro and Wilcox procedure will generate a historical time series of CES index values which are largely free of substitution bias. Thus the CES price index, combined with a method for estimating the elasticity of substitution, could be used to provide a timely estimator for a superlative index, which can only be produced on a delayed basis. However, there are some risks associated with this methodology: namely, that past (average) movements in relative prices (which are used in order to obtain an estimator for the elasticity of substitution) are no guarantee for future (or present) movements in relative prices. It is also possible that the historical pattern of demand is determined by other factors not recognized in the analysis, such as changes in incomes, demographic factors, or tastes and technologies. Therefore a risk exists that the CES price index, based on a historical procedure for estimating σ, could generate misleading advance estimates for a superlative index.

[2]Essentially the same methodology was used by Alterman et al. (1999) in their study of U.S. import and export price indexes. For alternative methods for estimating, see Balk (2000).

3

Index Domain

Whether the desired index is a COGI or a COLI, decisions must be made about its "domain," that is, about the universe of the things it is to cover. As we explain in Chapter 2, the conceptual underpinning of a cost-of-goods index (COGI) is sufficiently unstructured that its domain can be straightforwardly defined as whatever is considered appropriate. It is natural to think of the index as the price of what people buy. Throughout its history, the Consumer Price Index (CPI) has been defined to include only the prices paid for private consumer goods and services.[1]

The Bureau of Labor Statistics (BLS) has, in recent decades, increasingly emphasized the concept of the cost-of-living index (COLI) as a framework for making decisions about the CPI and "accepts the COLI as the measurement objective for the index," with the caveat that "while the CPI may be described formally in the context of a cost-of-living index, there is no single all-purpose definition of this target" (Bureau of Labor Statistics, 1997c:3). From the standpoint of an individual consumer, an all-encompassing COLI, defined without limits on its scope or domain, would have to take into account effects on the consumer's standard of living arising from changes in the social and physical environment, such as air quality and the crime rate, and in the provision of public goods furnished by the government. The BLS, however, has continued to define

[1]Some private goods produced and sold by governments to individuals for a price—e.g., entrance fees to national parks and transit fares on publicly owned systems—are included in the domain of the CPI.

the COLI to include only the effects arising from changes in the prices of private goods and services purchased in the market, i.e., the same domain as that specified in the traditional fixed-weight index. More precisely, according to the BLS (1997c:3): "The cost-of-living index approximated by the CPI is a subindex of the all-encompassing cost-of-living concept, specifically a sub-index that is conditional on the excluded factors that affect consumer well-being such as health status and the quantity and quality of government provided goods and services."

A conditional cost-of-living index as defined by the BLS is the minimal expenditure ratio needed to attain a given standard of living in the face of changes in the prices of private goods and services, on the assumption that "outside conditions"—the status of the social and physical environment and the provision of goods by the government—remain at some specified level. Changes in the various conditions outside the universe of the provision of private goods and services can affect a household's cost of living in two ways. They can directly alter its standard of living—the crime rate worsens, additional households become crime victims, and more people begin feeling less secure. But the increase in the crime rate may also change the household's taste for private goods; it may increase its spending on home security systems and cut spending on other goods, especially those, like patronage of downtown restaurants, that it believes will increase exposure to crime. Changes in the relative prices of private goods would then not have the same effect on a household's cost of living as it would with the original crime rate. A COLI conditioned on the stability of environmental conditions would measure the expenditures on private goods that would be needed to maintain a given standard of living for the household in the face of changes in the prices of private goods on the assumption that no change in the crime rate had occurred. It would therefore exclude both the direct and the indirect effects of any changes in outside conditions that did occur.

For most if not all of the purposes for which it would be needed, a conditional cost-of-living index should measure the expenditure ratio needed to maintain the reference period's standard of living, given constancy of the environmental conditions at their reference period status. If changes in environmental conditions do occur between the reference and comparison periods, the superlative index that is used to approximate the conditional COLI would, as desired, exclude their direct effect on the household's standard of living. But if the environmental change alters some of the marginal rates of substitution among private goods—as in the crime rate example above—the superlative index will not provide a first-order approximation to a COLI conditioned on the status of the excluded variables in the reference period.

As we note in Chapter 2, Diewert (2000b) demonstrated that, when external conditions change, a superlative index will approximate a COLI that maintains the household's standard of living at some intermediate level between the reference and comparison periods and under the external conditions that lie between those two periods. Consequently, the conditional COLI that is needed cannot be reproduced by the superlative index. One interpretation of this result is that environmental changes that affect marginal rates of substitution among private goods will reduce the accuracy with which a superlative index measures the conditional cost of living.

The extent to which changes in outside conditions affect the accuracy of a superlative index in meeting its stated objective will depend on how significantly the changes in outside conditions alter the pattern of preferences for private goods between the reference and comparison periods and how those alterations interact with the changes in relative prices that occur over the same period. If the weights used in the index are frequently updated, as the BLS now plans, a slow and gradual drift in such outside conditions as the quality of the physical environment or the crime rate are not likely to have much effect. But an event such as an unusually severe winter might have more noticeable, even if temporary (and reversible), consequences.

AN UNCONDITIONAL COLI: CONCEPTUAL ISSUES

Incorporating into a cost-of-living index the effects of changes in environmental conditions and government-provided public goods would require analytical and measurement techniques that, in most cases, go well beyond the current state of the art. But even if measurement of such effects were feasible, conceptual question arise about whether a cost-of-living-index should take them into account. Put another way, what should be included and what excluded from the variables that are held constant in a conditional COLI? Should the index incorporate the net effects on consumer welfare (not already reflected in the costs of private production) from such public goods as those furnished by the military establishment, the preservation of wilderness areas, or the provision of law and order? Do the effects of changes in the state of national security belong in the index, implying, for example, that a large drop in the cost of living occurred at the end of the Cold War when the threat to national security almost surely declined? What about the motor vehicle accident rate (whose effects would probably be easier to measure than those associated with, say, the crime rate)? Should the index include increases in longevity arising from general improvements in medical knowledge and techniques (as distinguished from those associated with particular medical procedures which, conceptually, should be treated as a quality change in a private good)? How feasible is it to make such a separation? In its discussion of quality-of-life issues the Boskin et al. (1996) report cited as negative factors, presumably tending to drive up the cost of living, "such social issues as divorce, illegitimacy, and the reduced role of the nuclear family." In concept, at least, should the cost of living be defined to include effects of intangible factors such as these?

The panel distinguishes sharply between what is appropriate for inclusion in the official family of consumer price indexes[2] on the one hand and the scope of broader satellite or supplemental measures on the other, which, developed with care and prudence, might provide valuable additional knowledge about the economy and society. Even if acceptably reliable measures were available that could quantify the effects on living standards of changes in some or all environmental conditions, including the provision of public goods, the panel believes those conditions belong on the list of conditioning variables. Their effects should not be incorporated into a measure that is used for such purposes as indexing public and private payments, the income tax code, private contracts, providing an overall measure of inflation for consumer goods, and as an indicator for monetary policy. Within the COLI framework, this requires that the index be defined as a conditional COLI, holding constant the reference period status of environmental and other goods not included in private goods and services.

Congress and the President decide how social security retirees and other transfer recipients should be compensated for the effects of rising prices, and it seems extremely unlikely that any support could be secured for a system that would reduce the purchasing power of social security recipients in response to, say, the end of the Cold War (as national security increased) or to the recent reductions in the crime rate. The concept of inflation would be altered beyond recognition and macroeconomic usefulness if it is defined to include the effect on living standards stemming from a rise in pollution and the crime rate or the effects of a severe winter. For example, in a comprehensive index that included the effects of changes in outside conditions, suppose that the subindex for private goods were stable while deteriorating outside conditions were forcing up the overall index. Surely we would not want the Federal Reserve to pursue a restrictive monetary policy to deflate the prices of private goods (whose stability, after all, is what is relevant for efficient business and consumer planning) in order to avoid an "inflation" in the official index as it responded to increased pollution, higher crime rates, or a colder winter.[3] We discuss these points further in Chapter 7, which examines the relationship between the design of price indexes and the purposes to which they are put.

There are other question about index domain that pose issues that are less clear cut than most of those discussed above. Should the CPI, for example, ultimately aim to include the estimated effect on living standards arising from the introduction of new goods or from the increase in longevity and decrease in

---

[2]We use the term "family of indexes" to reflect the fact that the panel endorses the publication of both a fixed-weight Laspeyres (at the upper level) and a superlative index, as well as indexes for population subgroups when that appears appropriate.

[3]The literature on the macroeconomic difficulties posed by deflation deals with the effect on consumer and business purchases from expectations of falling prices for private goods and services; it nowhere suggests that this problem might somehow be offset by negative effects on living standards flowing from a deterioration in environmental, social, or other outside conditions.

morbidity from technological advances in medicine? Should it seek to incorporate the improvement in living standards that might accompany the introduction and gradual spread of mass merchandising and discount stores, including an explicit adjustment for associated differences in retail service quality and shoppers' time? The conceptual background for considering many of these questions was explored in Chapter 2, and they are discussed in more detail, together with recommendations, in Chapters 5 and 6. In the last part of this chapter we briefly deal with issues related to the treatment of certain goods furnished by government and by private employers (fringe benefits).

MEASUREMENT PROBLEMS

A cost-of-living index that seeks to include the status of the social and physical environment and government-provided public goods must measure the effects of changes in those variables on consumer welfare. That is, it must measure the expenditures necessary to maintain a given standard of living, taking into account the value to consumers—positive or negative—flowing from changes in those variables.

Correspondingly, an expanded measure of changes in national output would include the value of changes in the provision of goods by government and, with the appropriate sign, the value of any changes in the flow of damages from various kinds of environmental "bads," weighted by their implicit prices.[4] Increases in pollution, crime, congestion or other environmental damages, arising chiefly from human activity, would reduce the expanded output measure, while activities (such as pollution controls) that reduced these damages and whose benefits were greater than their costs, would raise the augmented output measure.

Although there are major difficulties that have to be overcome in defining and then measuring the relevant *quantities* of many environmental and government-provided public goods, the problems encountered in assigning *values* to these goods are truly formidable. The fundamental difficulty stems from the fact that not only the goods provided by government but also most environmental goods are public goods. Public goods, such as clean air, have the characteristic that if they are available to one person they are available to all—globally, nationally, or regionally—whether or not particular individuals desire to purchase them. Only to a limited extent can individuals choose how much clean air or safety against crime they want to pay for and consume. They can purchase home security systems to reduce their risk from crime or install water filters to increase the

[4]Changes in government-provided goods and in environmental damage that result in changes in productivity and costs in the production of private goods are already reflected in the conventionally measured gross domestic product (GDP). Damages that show up directly as reduced consumer welfare—e.g., higher rates of respiratory disease from air pollution—would have to be separately identified and accounted for.

purity of their tap water. They can move from a city with a high crime rate to one with a low rate. But in the main they cannot buy a pro rata share of safe streets and ozone reduction or their own personal protection against global warming.[5] This fact has a number of consequences for environmental goods (or any other class of public goods) that make it exceedingly difficult to evaluate their economic worth.

First, by their purchases and the prices they pay, people reveal information about their preferences among private goods. Some of their preferences for environmental and other public goods influence their private decisions—such as in what city to live and in what neighborhood to buy a home—and thereby affect the rents and land prices. But not all environmental and related preferences exert such influences, and when they do the information they convey has to be estimated indirectly, often through a chain of inferences, each step of which entails the likelihood of reducing accuracy. For example, the estimated wage premiums paid to people in high-risk jobs have been used to place a value on the life-threatening effects of air pollution and traffic accidents. The difficulties of valuing public goods have often driven researchers to the use of contingent valuation techniques. Individuals are asked in a survey to indicate what they would pay for a particular public good—e.g., to increase visibility in a specific national park. But these techniques are highly controversial. The hypothetical nature of the question posed, the lack of any budget constraint on the respondent, and the absence of any behavioral observations cloud the results from contingent valuation approaches (see Diamond and Hausman, 1997).

A second characteristic of public goods is that they cannot be bought and sold in the market. Changes in their supply must be decided through political processes, which typically do not lead to an effective balancing of costs and benefits at the margin. People can, by changing their place of residence, make choices among alternative comprehensive packages of public services provided by different local governments. But there are costs of moving, and people cannot make choices among the many individual components of the package of services offered by different local jurisdictions.[6] The choices are still more restricted

[5] Some public goods—e.g., public beaches or highways—are congestible. After some point the consumption of the good by larger numbers of people reduces its value for others. Tolls or fees can be charged, establishing a price, but for reasons of public policy, and politics, fees are usually not charged or are set well below market-clearing levels.

[6] The Tiebout hypothesis proposes that people, by selecting and, if necessary, moving their place of residence sort themselves into heterogenous communities according to their preferences for local public goods. They can effectively, at least to a point, choose the quantity of local public goods to match their preferences. Recently, Rhode and Strumph (2000) conducted an empirical test of the hypothesis, which predicts, they argue, that falling transportation and communication costs should lead to increased heterogeneity of public goods preferences and public goods provision across local communities. Over the past 140 years communication and transportation costs have fallen substantially, but heterogeneity among counties in the United States in terms of outcomes and preferences (as measured by various proxies) has decreased or, in a few cases, remained unchanged.

among public goods furnished by state governments and, for most people, individual choices do not exist for the public goods provided at the national level, such as national security, the federal justice system, or the national air pollution standards. In such cases, households cannot adjust the quantities of public goods they consume in accordance with their individual preferences. The empirical evidence suggests that the marginal costs of providing an extra unit of an environmental good such as clean air through emission abatement measures are not likely to equal the sum of the marginal valuations of the good by the affected households.[7] The demand price is not equal to the supply price and so, in most cases, it cannot be assumed that cost data will be a reasonable proxy for consumer valuations.

Of course, GDP does measure and include the gross amounts spent by federal, state, and local governments to purchase goods and services ($1.7 trillion in 2000). These goods and services include such diverse items as arms purchases and the wages of the armed forces; the wages and salaries of teachers, police forces, and other civil servants; the building of highways and dams; and the maintenance costs of the national parks. But these are not themselves public goods. They represent the costs to the government for the labor force, materials, and capital goods it uses to produce public goods and services for the population—national security, education, law and order, new medical technologies, and so on. And, as we noted above, the political process does not tend to produce public goods in such an amount that, at the margin, the sum of the values attached by the population to an additional unit of public goods typically equals the additional costs of producing it.

In sum, research efforts to develop measures of output and prices that include the effects of changes in some important environmental conditions might well yield useful knowledge. But measures of the effects of changes in such conditions on the welfare of consumers depend on a complex interaction of often controversial assumptions and inferences that are far too speculative to be professionally and publicly accepted as part of an official index used for the important public and private purposes for which the CPI is used.

---

[7]Hahn (1996) assembled estimates of the benefits and costs of 54 major environmental, health, and safety regulations that took effect between 1990 and mid-1995. Aggregate benefits from all the regulations substantially exceeded their costs, but an excess of benefits over costs occurred in only 23 of the 54 cases. By inference, the number of cases in which marginal benefits exceed marginal costs was an even smaller proportion of the total, although some individual regulations showed very large net benefits.

SUPPLEMENTAL INDEXES AND SATELLITE ACCOUNTS

While we recommend that the CPI should continue to be confined to the domain of private goods and services, we also believe that a research program to develop supplemental measures of the economic value of changes in several important environmental conditions is a worthwhile objective. So long as its limitations are made clear, such research should be undertaken in an effort to generate rough-and-ready but still useful knowledge about the effects of some aspects of environmental changes on the material well-being of the population. If successful, it could also aid in the management of the nation's environmental resources. But we do not believe that this is a task that can be efficiently undertaken by the BLS on its own and with the principal objective of producing a more comprehensively defined cost-of-living index. Rather, it can best be pursued in the process of developing an integrated set of expanded output measures and their associated price indexes. Several considerations point in that direction.

In the case of most private goods and services, real output is not measured by collecting data on physical quantities, but by deflating observed nominal expenditure data with an appropriate price index. The output measures are, in effect, derived from the price measures. The prices used are observed market prices, and even when those prices are adjusted for quality change, the adjustments are typically based on estimates themselves derived from market prices (hedonic adjustments) or from cost data. But principally because of their "public goods" characteristics, the measurement of changes in both the "output" and the "prices" of environmental goods have to be derived independently: nominal expenditure data do not exist. Complex measures of physical changes typically have to be estimated, and their implicit prices have to be estimated indirectly.

In the important case of air pollution, for example, ambient air concentrations of individual pollutants (such as carbon monoxide, sulfur dioxide, or particulates) must be measured and weighted in terms of human exposures (or, in some cases, exposures of various building materials). Physical dose-response measures, such as changes in the incidence of respiratory ailments per unit change in ambient concentrations, must then be estimated from clinical or epidemiological studies. In many cases, it is the flows of relatively short-lived pollutants into the environment that cause the damage; in other cases, it is the stocks of pollutants in the atmosphere that produce gradual and long-lasting flows of damages. Once dose-response output effects are identified, a value per unit of response must be constructed by one of a number of methods (noted above).

Some of the damage from air pollutants directly lowers human welfare by increasing morbidity and mortality, and much of this kind of damage is currently unaccounted for in the national economic accounts. Some of the damage, however, lowers the productivity and raises the costs of producing private goods—e.g., the effect of acid rain on building materials or the lost days of production from increased morbidity—and is already included in the current measures of the output and prices of private goods. These latter costs need to be identified, but one must have an accounting system that makes sure that environmental effects are not double counted. While this is an illustration of air pollution effects, there exists, more generally, a complex interweaving of environmental changes, economic production, and consumption.

In summary, estimating the effect of environmental conditions on economic welfare, despite some overlap, will typically involve quite different types of analysis and estimation than that associated with the estimation of a COGI or COLI restricted to the domain of private goods. Most importantly, estimates of the effect of environmental changes on national output, income, and prices must be embedded in a consistent accounting framework that takes account of stocks and flows and outputs and prices and distinguishes the effects of the environment on various categories of human and economic activity.

In recent years the U.S. Department of Commerce's Bureau of Economic Analysis has been engaged in a long-run effort to construct "satellite" accounts that would gradually extend, on an experimental basis, the domain of the government's official measures of national output and income. A recent National Research Council (1999) report provides a highly useful discussion of the challenges, difficulties, and promises of this endeavor. Any research program of the BLS to develop experimental measures that substantially expand the domain of its various price indexes should be carried out in close concert and coordination with the BEA effort.

OTHER DOMAIN ISSUES

Taxes and Government Programs

Governments, particularly state and local governments, use sales taxes and similar levies to finance part of their expenditures. Such taxes tend to be passed along to consumers in the form of higher prices.[8] As a consequence, an increase in the average rate of the sales tax throughout the country tends to raise, and a decrease to lower, the CPI. But as pointed out by Nordhaus (1997), the direct contribution of the education, roads, and other public goods to households' standards of living is not captured by the CPI as an offset against the taxes. As a result, the CPI tends to overstate the rise in the cost of living when sales tax rates

[8]In the national income accounts the category of government revenues labeled "indirect business taxes" includes not only sales taxes and customs duties, which are typically passed along in higher prices, but also state and local property taxes, whose incidence is less clear and that in the short to medium run are more likely to be borne by property owners. In the text we use the term sales taxes to cover indirect business taxes, excluding those levied on (land and structure) property.

are rising and vice versa. Nordhaus also argues that similar anomalies occur in a number of other instances: government environmental, health, and safety regulations require firms to incur costs that are passed on in higher prices for private goods, but the value of the associated benefits is not offset against them. Employers provide fringe benefits to workers, especially pensions and health insurance, whose costs are passed along in higher prices while, again, the benefits are ignored. Ideally, according to Nordhaus, one should estimate the value of the benefits provided and make the necessary adjustments. Since that is not feasible, he recalculates the CPI by subtracting the ratio of these items to aggregate consumption expenditures on the assumption that their cost to consumers is matched by the benefits they finance. He finds that, over the period 1960 to 1995, when both employer fringe benefits and pollution abatement costs were rising relative to aggregate consumption, a CPI stripped of all of those costs, which he calls an augmented COLI (ACOLI), would have risen by 0.4 percent a year slower than the published index (Nordhaus, 1997).

The panel believes that the issues raised by Nordhaus should be explored in the context of research on satellite accounts and the development of expanded and integrated national accounts that we discussed earlier. But we do not believe, nor does Nordhaus recommend, that adjustments of the kind outlined above be incorporated in the flagship CPI. Nor, we conclude, should they be incorporated in any measure used to index public benefits and the tax code.

In the case of sales taxes and mandated pollution control costs, we have discussed the limited or nonexistent mechanisms that exist for equating the marginal tax costs and benefits of public goods, especially at governmental levels higher than local communities. In addition, even if the assumption of cost-benefit equality were assumed, several difficulties remain. Many of the public goods furnished by government are intermediate goods used by business firms in the production of private goods whose benefits are reflected in lower production costs for private goods and therefore already included in the CPI. Also, if we subtract from the index an estimate of the price-raising effects attributable to mandated pollution control activities, must we not also undertake the formidable task of estimating, and adding to the index, the costs directly imposed on consumers from the “free” use of the environment by producing firms, which gave rise to the need for the environmental controls in the first place? It is difficult to justify the first step without the second.

Employer-Paid Benefits

The treatment in the CPI of goods provided in-kind to employees as fringe benefits raises a number of issues. There is a wide, even if not universal, consensus that, apart from some tax advantages and except during short-run transition periods, overall compensation packages are not affected by how they are parceled out into money wages or fringe benefits.[9] In the medium to long run, the greater part of changes in fringe benefits tend to be matched by offsetting changes in wages: workers trade money wages for in-kind benefits. The major categories of fringe benefits are employer contributions to pension plans and health insurance and their payments of social security and Medicare taxes on behalf of employees. Employer-paid pensions are a form of savings (with tax advantages). But the CPI is a "one-period" index. Viewed as a COGI, it measures the average change in the price of currently purchased consumption goods. Viewed as a COLI, it is similarly defined to include only the effect on living costs arising from changes in the prices of current goods.[10] A change in pension saving (i.e., future consumption) ought not to be treated as a change in the price index or cost of living associated with current consumption.

There are several arguments for not adjusting the CPI to take account of the benefits furnished to workers through employers' contributions to the federal social security program. While it has been accumulating surpluses in recent years, the combined social security and Medicare (Part A) system is still essentially a pay-as-you-go system. Because of coming demographic changes, benefit payments under current law will begin to rise faster than revenues under current tax rates. It is unlikely that future changes in the present value of taxes and benefits will be closely linked to each other. Payroll taxes are still to a significant degree a public levy to finance a collective and continuing set of transfers from one generation to another. And, in any event, even if they were fully funded, they would represent another form of "forced saving" whose future consumption benefits are not covered in a one-period CPI.

Employer-provided health insurance benefits raise more difficult questions. Unlike pension contributions and employer-paid social security taxes, they (principally) provide a form of current consumption, namely, medical care plus the provision of risk insurance.

Conceptually, it is not clear whether or not employer-paid health benefits should be considered as an item of private consumption and included in the medical care weights of the CPI. We argued above for excluding public goods from the domain of the index, including locally furnished public goods. And the existence and generosity of health insurance benefits in compensation payments are also to some extent the result of collective decisions and, as such, share some of the characteristics of local public goods. In both cases, decisions are made collectively, but people have some ability to choose individually by changing their residence or their job. By analogy, employer-paid health benefits should be treated like a local public good and excluded from the index.

---

[9]This is an oversimplification. Fringe benefits are tax-free forms of compensation. Hence, by taking a larger fraction of their compensation in the form of fringe benefits, employees can reduce their tax burden.

[10]For a treatment of the conceptual issues involved in a multiperiod cost-of-living index, see Pollak (1989:196-197).

However, it could be argued that health insurance fringe benefits are more nearly private goods, influenced by private choices, than are the goods provided by local government. In choosing among jobs, workers have the possibility of accepting a job with higher money wages but without health benefits, while making their own arrangements for purchasing medical care or self-insuring; or they can choose something in between. But there are many fewer options in the basket of local public goods: one cannot typically buy a bridge single-handedly. Moreover, alternative baskets of compensation items among which workers have to choose will typically contain many fewer goods than the alternative baskets of local public services.[11]

From a practical standpoint, most of the analyses of changes over time in real income are derived by deflating survey data on nominal money incomes by the CPI. In turn, that nominal income data typically exclude the value of employer-financed fringe benefits. As a consequence, in Chapter 6, which deals with the pricing of medical care, the panel pragmatically recommends that the official "flagship" CPI continue to exclude coverage of employer-paid health benefits. On the other hand, while the appropriate conceptual treatment of employer-paid fringe benefits is not an open-and-shut issue, the panel does recommend in Chapter 6 that the BLS also publish a supplemental medical care price index that, among other features, includes health care expenditures financed by employers.

CONCLUSION AND RECOMMENDATION

Drawing from the arguments presented in this chapter, the panel offers the following conclusion and recommendation:

> **Conclusion 3-1: Under either rubric, COGI or COLI, the domain of the "flagship" or official CPI and any subgroup indexes that are produced should be confined to the boundaries that the BLS has adopted—changes in the prices of private goods and services.**
>
> **Recommendation 3-1: The BLS should not, on its own, conduct research aimed at producing a CPI with a substantially expanded domain. But we encourage BLS, jointly with other federal statistical agencies and particularly the Bureau of Economic Analysis (BEA), to undertake research aimed at producing on an experimental basis or in satellite accounts, more comprehensive measures of national output, income, and prices. These would take into account the effects of changes in outside conditions for which there may be at least some chance of reasonably measuring those effects—perhaps, for example, changes in the status of the natural environment.**

---

[11]Just as there are many aspect to a job in addition to the menu of compensation items, there are many aspects to the choice of residential location apart from the basket of local public services.

4

Evolving Market Baskets: Adjusting Indexes to Account for Quality Change

The ever-changing mix and quality of products and services available in the market create difficult problems for price index construction. At a basic conceptual level, the problem is easy to understand: under either of the conceptual frameworks we have discussed, unadjusted price comparisons between an item and a non-identical replacement cannot generally be treated as equivalent to comparisons that involve an unchanged item. However, developing solutions and assessing techniques for correcting the problem is extremely complicated. Quality change has typically been considered the least tractable problem associated with the Consumer Price Index (CPI).

The pervasiveness of item replacement alone makes quality change impossible to ignore. Item replacement refers to the process whereby a Bureau of Labor Statistics (BLS) field agent must select and price a different product because the one previously included in the sample can no longer be found on the store shelf. Moulton and Moses (1997:323) estimate that, based on 1995 data, about 4 percent of price quotations on average every month involve a replacement item. Some items are replaced more than once during a year, and this translates into an annual replacement rate of about 30 percent for CPI items scheduled to remain in the sample. Although a price adjustment is not made in each case, a judgment about quality change is. In about a third of these cases—roughly 10 percent of all CPI items each year—a quality adjustment is deemed necessary. Moulton and Moses also show that, relative to continuously priced items, replacement items have a disproportionately large effect on the rate of change in the CPI.

"Quality change" can take many forms. In the research literature, the distinction is often made between quality change and new goods. Unfortunately, this distinction does not create clear-cut categories that imply specific corrective approaches. For instance, is a cell phone an improved wired phone or an entirely new product? What about a high-definition television, a fuel cell automobile, or on-line stock trading? The line between "new" and "improved" is inevitably arbitrary. The situation is brought into focus somewhat by thinking in slightly different terms, framed by consideration of how CPI product sampling and item identification actually work in real cases. Following Armknecht et al. (1997), three distinct cases can be delineated:

(1) A new item replaces another that has been or soon will be discontinued and that will fall out of the CPI sample. Replacement goods may be substantively similar (in which case there may be no quality issue at all), or they may be improved (or possibly inferior) versions of the discontinued item. These goods replace old goods but fall into familiar CPI categories—e.g., 2001 Fords.

(2) A new "supplemental" good appears that does not replace a specific outgoing good in the CPI, but that does fit appropriately into an existing item strata category—e.g., Honey Nut Cheerios.

(3) A genuinely new item appears that does not fit into an established CPI item or strata—e.g., VCRs or wireless phones.

In some sense, all of the above situations involve new goods; however, the extent of the difference between an old and a new product ranges from close to zero, to run-of-the-mill quality changes that happen on a daily basis, all the way to the appearance of radically new products that reflect what Nordhaus (1998) calls "tectonic shifts in technology."

Over time, BLS confronts situations on all points of this quality change spectrum. On the easy end, a commodity analyst may be forced to compare a 2-pound bag of rice with a 1-pound bag. Perhaps the previous 1-pound bag is out of stock or is not sold much anymore. Something like this happened with butter, which used to be sold in half-pound packets and now is more frequently sold in 1-pound packets. Most economists would simply work with per-pound prices in both cases. Of course, a 1-pound package is not identical to two half-pound packages, since the former requires longer storage, may be more likely to go bad, or may be sold with size discounts, and so on. But in many cases of this sort, per-unit prices seem likely to provide a very good approximation. The BLS apparently agrees. For instance, when the CPI went from pricing 16-ounce cans of tomato sauce to 14.5-ounce cans, all of the difference in price per ounce was attributed to pure price change (Kokoski et al., 2000:2).[1]

[1]This is not to say that the nonlinear pricing issue is unimportant, particularly for large differences in package sizes. As a first step toward estimating its impact on the CPI, BLS could, in a straightforward manner, perform empirical research that examines how unit prices vary with package sizes. Of course, this only applies to products for which a range of sizes is typical.

Some of the distinctions between cases can be clarified by casting them in this repackaging framework (this idea dates back at least to Fisher and Shell, 1968, 1972). This framework deals with situations in which the amount of "good" in the good has changed. It is difficult to think of actual examples other than changes in package size that correspond exactly to this framework, but imagine that gasoline has been improved so that it gives a 25 percent increase in miles per gallon for all vehicles and is otherwise unchanged. Once again, the solution seems fairly clear: the real price has fallen by 20 percent from, say, 5 cents to 4 cents a mile. One useful way of thinking about this is that the good is not gasoline but miles from fuel, and the price of the latter has fallen by 20 percent. Another example might be a new razor that yields more shaves before becoming dull. These cases converge with the butter case when one shifts from thinking of gasoline or razors to thinking of a good that more directly relates to consumer welfare. Once the good is defined appropriately—which is not trivial—and one thinks of the market good as a repackaged real good, the right way to handle quality change becomes transparent.

The basic idea applies to more complicated cases, though the practicalities get harder. In most cases, there is no single obvious quantitative metric (like miles per gallon or number of shaves) to use in redefining the package, which makes it difficult to identify a simple one-to-one relationship between the real goods and the market goods. Economists and marketing specialists often think of situations of this sort in terms of characteristics, with market goods consisting of various combinations (packages) of several characteristics. Since one often does not really know the characteristics—because each good has many and because there is often no nonarbitrary way of defining them—things are rarely as simple as in the gasoline case, let alone the butter case. It is conceptually useful though to think of approaches such as hedonic techniques, which we discuss in detail below, as an attempt to redefine goods so that, by repackaging, one can factor out quality change.

The really hard cases occur when a new good introduces new characteristics, in which case the repackaging idea cannot help with measuring quality change. But it is unclear that any practical technique can help in these cases or, indeed, whether radically different goods can even be appropriately discussed in the context of price measurement. For instance, in no clear sense did the introduction of cellular telephones reduce the general price level. Yet that new product did increase the well-being achievable by a subset of the population for a fixed money outlay and, in that sense, reduced the cost-of-living.

Our coverage of the quality change/new goods problem follows the taxonomy outlined by Armknecht et al. (1997). First, we contrast the nature of the problem as it arises in the COGI and COLI contexts. In the next three sections, we sort though the gradations of quality change that occur along the repackaging spectrum. This discussion includes a brief review of the evidence of CPI bias presented by the Boskin commission (Boskin et al., 1996) as well as a discussion of BLS item replacement procedures and their associated biases. Chapter 5 considers separately the case when goods appear that do not fall into existing product categories. The second half of this chapter assesses the role of hedonic regression techniques in quality adjustment. We offer specific recommendations about the applicability of hedonics for adjusting observed prices or for directly constructing indexes and about approaches to selecting items for quality adjustment.[2]

COLI AND COGI VIEWS OF THE QUALITY CHANGE PROBLEM

The general problem of changing quality can be illustrated by simple example. Consider a price index for automobiles for which, in the reference period, the dominant type of automobile has a steel dashboard and no seat belts and is a gas guzzler. Now suppose that, in the comparison period, the dominant type of automobile has leather appointments, airbags, and efficient fuel economy. Direct comparison of the nominal prices of these cars will yield little meaningful information. What does it tell us if the price of a 2002 Camry is 10 times that of a 1965 Rambler? Similarly, if this year's computer model costs the same as last year's but does more and does it faster, what does the observed price constancy really tell us? Nordhaus (1998:59-60) points out that a fundamental problem associated with quality change is raised by these types of comparisons because "conventional price indexes measure the prices of commodities that consumers buy rather than the cost of attaining a given level of economic well-being or utility." The manner in which quality change and new goods problems arise depends to some extent on the index's underlying conceptual structure whether a cost-of-goods index (COGI) or a cost-of-living index (COLI) though procedures for dealing with these problems are essentially the same in both cases.

COLI

The COLI requires that prices, or the index itself, be adjusted to account for effects on living standards that accompany changes in the quality of goods and services. For certain commodities, the quantitative adjustment could be straightforward—e.g., the new automobile fuel that increases miles traveled per gallon. But in most cases, the relationship between product or product characteristics (inputs to well-being) and actual well-being created cannot be directly observed.

[2]We bypass the issue of quality change as it affects nonmarket inputs to consumer well-being (things like air quality, traffic congestion, and sense of personal safety) that are not captured in conventional price indexes (see Chapter 2). In addition, while we acknowledge the theoretical validity of the Boskin commission's observation that changes in the variety of available goods and services affect consumer well-being, we know of no useful way to deal with this issue in index construction.

It may be impossible to measure the value, even for just one consumer, created by the change from black and white to color television, by an increase in the user friendliness of computers, or by the addition of antismog devices to automobiles.

Even setting aside the problem that the value attached to changing products may differ widely among consumers, changes in the mix of items sold raise two difficult issues. First, when outgoing items are replaced, COLI calculation requires isolating a pure price component from the observed price difference between the outgoing item and its replacement, which reflects both pure price change and quality change. If the underlying index methodology is unable to disentangle the quality-driven price movement from the "pure" price movement, living standards cannot be held constant. Second, techniques must keep the composite of index items—which is in constant churn—relevant to consumers' material well-being. The addition of new goods into the marketplace generally raises (and the elimination of goods lowers) the welfare of some consumers; until the new good is represented, this welfare change is not reflected in the price index.[3]

The cost-of-living approach provides a theoretical framework for thinking about problems associated with the changing nature of goods and services available in the market. If a rational consumer buys two varieties of some product—apples, for instance—in some (relatively short) period, economic theory asserts that the ratio of their prices measures their relative qualities, at least at the margin.[4] The next logical step is to assume that such price ratios provide meaningful measures of relative qualities even if there are many consumers and some do not purchase both products simultaneously. This assumption may be misleading when notions of quality differ across consumers, since demographic changes may then shift relative prices without quality change. Without this assumption, however, there is no way to use market data to recognize that, for instance, replacing a low-price variety with a high-price variety can make all consumers better off if the new variety is of sufficiently higher quality. More generally, if the quality of goods improves on balance over time, a cost-of-living index will discount some of the nominal price increases that occur, and the overall price index will rise more slowly than the average of the unadjusted prices.

[3]Hausman (1997) has argued that the CPI is also biased as a cost-of-living estimator because, to the extent that consumers value variety, it makes no allowance for increases in the number of choices within index categories. Conceptually, this assertion is hard to dispute—if for no other reason than greater variety permits better matching to individual tastes, which gives some people pleasure directly. On the other hand, the existence of greater variety may, in some cases, be welfare decreasing if it creates increased search costs. There is no known way to capture such effects accurately in regular index production.

[4]In practice, of course, it is often a matter of judgment as to whether one is dealing with two varieties of the same product or two different products that happen to be relatively close substitutes. Also, as the number of varieties multiplies, the act of choosing itself may require more time and effort.

Unfortunately, quality adjustment techniques seem destined always to have an ad hoc element. The aspects or characteristics of goods that determine consumers' perceptions of quality are not consistently observable. Moreover, the quality and taste components of price change are often inexorably intertwined. On the other hand, BLS has to balance on a case-by-case basis the errors that inevitably arise in the adjustment process against the errors that would inevitably arise either from ignoring quality changes or from assuming, as is often done, that all price differences between similar items reflect quality differences. We return to these issues below.

COGI

In a conventional Laspeyres index, the changing set of available goods and good characteristics also creates problems. Once an index item from the reference period is replaced by a different item, a strictly defined Laspeyres index cannot be calculated, since an identical bundle can no longer be priced in the comparison period. Given the pace at which new goods are introduced in a modern economy—ranging from those with slightly modified characteristics to those that are entirely new—it would be highly restrictive to monitor price inflation solely from a bundle chosen for stability. A "Big Mac" index may lead to misleading conclusions about general price movements, particularly since stagnant and dynamic sectors of the economy are likely to display systematically different price trends. Nonetheless, in no small part because of the uses to which they are put, it seems desirable to adjust price indexes to account for changing item quality and to reflect the changing mix of goods over time.

In practice, one need not be methodologically boxed in by this narrow textbook view of a Laspeyres index. And the CPI has in fact been modified—since at least 1967 when BLS began adjusting automobile prices—to address quality issues while, at the same time, maintaining its basic COGI structure. A working definition requires only that a set of market goods and services that are valued by consumers be identified for inclusion in the index. Since purchasing patterns and the set of available products have changed, the basket has been allowed to change over time. The organizing principle is the desire to cover the goods on which people spend most of their money and then to make adjustments to account for quality change. A COGI proponent is likely to argue that quality adjustment is necessary because, when the nature of goods change, prices of like items cannot be compared over time since the original bundle of goods no longer exists. A COLI proponent is likely to add that, since improved products generate higher levels of consumer satisfaction, observed prices must be adjusted to isolate changes in the cost of maintaining living standards. These differences would not affect their evaluations of alternative adjustment mechanisms.

Again, the idea of repackaging helps draw some distinctions between a COGI and a COLI in handling quality change. When two half-pounds of butter are replaced by a 1-pound pack, a strictly defined Laspeyres index is an impossibility. However, it is possible to go forward in this framework if the item for pricing has been defined as "butter" instead of "half-pound packets of butter." In our earlier gasoline example, a Laspeyrian working in this fashion would price "miles from fuel" and not "gasoline." The Laspeyres approach has no difficulty pricing characteristics, provided of course that one has some way of identifying and measuring the relevant ones, a difficulty that is common to all approaches. Any Laspeyres-type approach must, however, begin with a definition of the goods (which may be characteristics) to be priced. Selection must be based on some clear notion (e.g., market share) that the set of goods represents that which people buy and what gives them utility. This is as true in a world of fixed quality as in one with changing quality. It is important not to confuse the issue of the definition of goods with the issue of a COLI versus a COGI.

The arguments and recommendations in this chapter reflect the panel's view that the CPI should be adjusted, for most categories of goods and services, to account for changing quality. In the next two sections, we review evidence on quality change bias. First, we briefly examine the Boskin commission report (Boskin et al., 1996) which focused on factors that are external to the CPI sample. We then review CPI methods for adjusting quality-changed items within its sample as well as potential biases associated with those methods, reserving the special case of hedonic adjustment methods for the following section.

EVIDENCE FROM THE BOSKIN COMMISSION REPORT

In accordance with its congressional charge, the Boskin commission ventured to estimate, by source and by item strata, biases in the U.S. CPI, relative to a hypothetical cost-of-living index. The commission's report (Boskin et al., 1996) estimates quality change and new product bias (which they treat interchangeably) to be 0.612 percentage points per year—the largest component of its overall CPI bias estimate of 1.1 percentage points.[5] The commission's report has received extensive attention in the academic literature; numerous studies (both pre- and post-Boskin) corroborate the general view that quality change bias exists, though there is much debate on the size and sources of the biases. Much of the research has focused on specific index items (e.g., Berndt et al. [1996] on prescription drugs, Cutler et al. [1996] on hospital and physician services, Hausman [1997] on new cereal varieties). Shapiro and Wilcox (1996) did estimate an overall CPI bias, in the range of 0.6 to 1.5 percentage points per year, but it is extrapolated from trends for a limited number of products and not from an evaluation ranging across all CPI item categories. Unfortunately, research on the potential magni-

[5]Though the Boskin commission does not attempt to identify separate quality change and new goods bias estimates, the report does make some descriptive distinctions between the two categories.

tude of the quality change or new goods problem has not revealed broadly applicable techniques for correcting these biases.

In contrast, a set of generally accepted methods has emerged for addressing other perceived index problem areas, most notably substitution bias. Shapiro and Wilcox (1996) describe solutions to the substitution component of the bias problem as harvesting the "low-hanging fruit" of the CPI bias problem. Sticking with the harvesting metaphor, solutions to quality change and new goods bias problems must be the fruit at the top of the tree, the kind that requires expensive tools to reach or that may not be reachable at all.

The theoretical COLI perspective provides a rationale for tracking the value to consumers of new models and commodities and suggests why, for certain purposes, an index should be adjusted to reflect these changes. However, the COLI theory is less illuminating when it comes to directing research toward appropriate corrective techniques. Indeed, finding approaches for accurately dealing with changing goods and new goods is the most difficult obstacle to fulfilling the Boskin commission's prescriptions for BLS to establish a cost-of-living index as its objective in measuring consumer prices. Reflecting the difficulty of the issue, the Boskin commission report did not advance any formal recommendations about how BLS could improve its measurement of quality change.[6] The Boskin commission suggested perhaps that BLS should be doing more of the things it already does to correct for quality change bias, but seemed to concede that it did not have new ideas for approaching the problem. Summarizing the commission's report, Gordon and Griliches (1997:84) write:

> The difficult questions posed by quality change and the continual arrival of new products . . . have been called the "house-to-house combat of price measurement." Because the magnitude of quality-change bias differs so much across product categories, any overall evaluation must be conducted "down in the trenches," taking individual categories of consumer expenditure, assessing quality-change bias for each category, and then aggregating using appropriate weights."

The Conference Board (1999:21) study group concurred: "In an advanced, dynamic economy like ours, there is no alternative to thorough, detailed analyses that slog through the data category by category, item by item. This is difficult, costly work, but no shortcuts are available." Such conclusions reinforce the premise that general solutions, equivalent to the use of superlative indexes or geomeans to address substitution bias, do not exist to correct for quality change

[6]In contrast, 5 of the commission's 17 recommendations deal directly with a form of substitution bias—for which concrete options (superlative and superlative approximation indexes) exist. Individual commission members have elsewhere advocated expanding the use of hedonic regression methods to control for quality change for specific product types (see, for example, Boskin et al., 1998:14).

and new goods bias. Furthermore, there is no guarantee that even such "detailed analyses" will produce results that are suitable for inclusion in the CPI.

While the Boskin commission offered no new remedies, it had much to say regarding the magnitude of quality change and new goods effects on the CPI, producing a comprehensive, categorical point estimate. Of the 27 CPI item categories evaluated for quality change by the commission, 8 were assigned a bias of zero; the other 19 were estimated to impart a positive bias on the index.[7] Estimates for 6 of the 19 positive bias categories were calculated using a combination of results from existing studies of specific items and inferred figures for similar unresearched items in the category. Two upper-level CPI categories assessed in this way—appliances (including electronic) and medical services—contributed more than half the estimated overall quality bias. The commission performed original research or detailed back-of-the-envelope calculations for 4 categories. For the remaining 9 categories, empirical evidence was unavailable, and a descriptive approach discussing possible bias sources coupled with guesswork had to suffice (Moulton and Moses, 1997:310). (See "Technical Note 1" to this chapter for a review of upper-level item categories that the commission identified as contributing significantly to its overall CPI bias estimate.)

BLS APPROACHES TO QUALITY CHANGE

In constructing its CPI, the BLS has implemented a number of techniques to minimize perceived biases associated with its modified Laspeyres approach. For many decades—starting long before the comparatively recent calls for a cost-of-living index—BLS has been aware of problems posed by items whose quality is changing over time. In general, the agency has appealed to the cost-of-living theory in describing its efforts to confront the issue.

BLS readily acknowledges that, relative to some ideal COLI, introduction of new goods and quality change of existing ones may bias the CPI in two different ways.[8] First, there are biases associated with quality changes that are detected in the CPI sample and for which BLS attempts to correct. In this case the question is: "What is the bias, if any, of CPI procedures for handling quality change when quality changes appear on CPI items?" (Triplett, 1997:24). Second, there are

---

[7] See Boskin et al. (1996) or Gordon and Griliches (1997) for the complete list of estimated bias by category; see Moulton and Moses (1997) for a detailed critique of the estimates.

[8] As noted above, the distinction between a "new good" and a new variety or improved-quality good is arbitrary. In terms of CPI construction, we think of a "new good" as one that would require creation of a new item strata (or entry-level item) and that can only enter the index by initiation of a new item classification structure—the VCR is an example. Quality change refers to evolving characteristics of a good or service already included in the index and whose price can be adjusted to reflect the change at any point—a laptop computer with more memory is an example.

factors that go unrecognized with current CPI methods that bias the index as a representation of changes in the price to consumers of attaining a given level of well-being. A COLI in its purest sense would respond even to changes in non-market goods (such as air quality or commuting time). Moreover, even in a conditional COLI, changes occur in the market—most visibly the appearance of new goods and services—that affect well-being but are not accounted for, at least not immediately, in the CPI.

Estimating the extent of the first source of bias requires evaluating internal CPI quality adjustment practices. BLS uses a range of quality adjustment approaches when a new item replaces an old one in the sample; the result may add all, some, or none of an observed price change to the index. Some of these approaches implicitly adjust for quality differences; others produce direct cost-based or hedonically derived comparisons of quality that are used to adjust observed prices explicitly.

The Boskin commission report emphasized the second sort of quality-related biases, those created beyond the CPI sample and outside of CPI methodology. They focused on one subcomponent: underrepresentation of new market goods in the CPI. One way of thinking about new goods in the context of a price index (due to Hicks, 1940) is to imagine that the good was always available but at such a high price that no one would buy it. When the good is introduced, one can calculate the effect on the cost of living by translating the new availability into a price reduction, from the (imaginary) price that choked off demand to the new (lower) price at which it was first sold. The CPI as calculated makes no attempt to capture this "price reduction" associated with the introduction of new goods (see Hausman, 1997). Nor does it attempt to capture the later similar "effective" price reductions that occur as more and more consumers learn about new goods and experience a reduction in the cost of living because of that knowledge. Since the CPI market basket has historically only been revised every 10 years or so, new goods often entered the basket only after a long delay, and early stages of product price cycles were missed. Other sources of index bias may go undetected, such as those associated with gradual change in the quality of services (medicine, education, airline travel) or intangible aspects of quality change, such as improved stereo sound or television picture quality.

Estimates of the magnitude of quality change bias seem to be closely tied to the type of bias researchers emphasize. Triplett (1997) argues that the Boskin commission arrived at a high-end estimate of quality bias partly because it focused primarily on biases generated by new goods (such as VCRs and mobile phones) during the periods when they were outside the CPI sample. He further suggests that current BLS methods for within-sample adjustment—which occur when an old product disappears from a CPI outlet and is replaced by a new noncomparable one—may impart some downward biases (Triplett, 1997:24): "The implications of the methods used in the CPI for handling quality changes are not well understood by economists; the CPI [Boskin] Commission did not discuss them adequately, and some of these methods over-adjust for quality change, so that *improving* quality can generate *downward* bias in the CPI."

Essentially, BLS methods for adjusting observed prices of items that have undergone significant quality change, as judged by a commodity analyst, borrow information about price changes observed for similar items. For example, say a new improved microwave oven replaces the old model at a CPI outlet. Under one method, BLS will assume that the pure price portion of an observed price change between the old and new models is the same, in percentage terms, as that for other microwave ovens. Any remaining price difference is attributed to quality change. Such a method would implicitly overstate the effect of quality (and impart a downward bias on the CPI) if, for instance, manufacturers tend to increase prices (beyond those that cover costs of implementing improvements) when they roll out new models. We lay out BLS quality adjustment methods and examine potential biases in greater detail below.

CPI Item Replacement Methods

As noted at the beginning of this chapter, the manner in which goods (and services) appear and disappear can take a number of forms: old models are replaced by new ones that are nearly identical; new models are introduced that embody clear improvements over their predecessors; models may display qualitative change in existing features or may introduce altogether new features. To accommodate some of these differences and to overcome data and procedural limitations, BLS employs alternative methods, shown in Table 4-1, for treating replacement item price quotes.

For cases in which a sample item is replaced, the observed price change must be (1) considered a pure price change (e.g., simple repackaging), (2) attributed entirely to quality differences, or (3) attributed partly to price change and partly to quality change (Kokoski, 1993). Cases 2 and 3 require adjusting observed prices prior to inclusion in the index; all three require judgments by BLS commodity analysts.

Case 1 results in what BLS calls "direct comparison," which applies when the replaced and replacement items are determined to be comparable by the commodity analyst. A repackaged food item or a new color of shirt are examples. Direct comparison is essentially item replacement for cases in which adjustment to the observed price has been deemed unnecessary. As Table 4-1 indicates, this is the most common finding. According to Moulton and Moses (1997), for 1995 about 65 percent of item replacements were in this category.

With direct comparison, a commodity analyst has determined that it is appropriate to treat the observed price change as pure price change. If any quality change does occur, its effect on the index is not filtered out. The Boskin commission wrote that direct comparison, which it called "comparable substitution," likely imparts an upward bias to the index since "in practice most goods tend to

TABLE 4-1 CPI Item Replacement Methods and Use Rates, 1995

Method	Percentage of Price Quotes for Replacement Items Using Method	Percentage CPI Change Attributable to Replacement Items, Decomposed by Method[a]	Item Categories for Which Frequently Applied
Direct comparison[b]	65	60	All categories
Overlap pricing	1	10	Apparel, medical care
Deletion[c]	15	2	Medical care, food and beverages
Class-mean	8	18	Apparel, transportation
Direct quality adjustment[d]	11	19	Transportation, apparel, computers

[a]The percentage change in CPI item categories affected by item replacement was 2.16 in 1995. Of this, 1.09 percent was attributable to replacement items (leaving 1.07 attributable to continuously priced items). Thus, for instance, 0.6 * 1.09 gives the overall percent change in the CPI that could be attributed to direct comparison price quotes in 1995.

[b]Used for comparable replacements; the rest of the methods listed are applied to noncomparable replacements.

[c]This follows Triplett's terminology; it is typically called the link method in BLS literature.

[d]This category includes both cost-based and hedonic methods.

SOURCE: Data from Moulton and Moses (1997).

undergo steady improvement, and often a better model is introduced with no change in price, causing the quality change to be missed entirely" (Boskin et al., 1996:19).[9] Empirical evidence presented by Moulton and Moses (1997) implies that any upward bias from ignoring quality change under the comparable substitution method is small, tempering the Boskin commission view. Triplett (1997:26) synthesized the empirical evidence: "Note . . . that the average price change [shown using 1995 CPI data] for the direct comparison cases (2.51 percent) is not higher than the quality-adjusted price changes for CPI cases where a direct quality adjustment is made (2.66 percent—Table 7). This suggests that the upward bias from ignoring quality in the direct comparison cases is small." Triplett goes on to explain why this is likely to be the case: "Direct comparison is the sanctioned [here used to mean approved] method for cases where the quality difference between varieties *a* and *b* is small, so it is reasonable that the quality errors are also small (though they might be pervasive)."

[9]The Boskin commission is really criticizing the BLS method for assessing comparability. Under the comparable replacement procedure, the new price is recorded, and the bias will be the same (in absolute terms) whether or not price a change has occurred.

Neither side would disagree with the proposition that, to the extent that undetected quality changes are more often improvements than deterioration, direct "comparable replacement" will bias the CPI upward. However, for the bias to be large, undetected quality change would have to be distributed such that, within the sample, missed quality improvements were either much more pervasive or of much greater magnitude than incidents of missed quality deterioration. This has never been shown empirically, possible because judgments by commodity analysts about comparability are not random. A new model (of, say, a microwave oven) with major changes in characteristics is more likely to be judged noncomparable than is a new shirt that displays only minor changes, such as in styling or color, over its predecessor.

For cases deemed by a BLS commodity analyst to require noncomparable substitution—that is, when there is a perceived quality change—BLS uses one of four options (other than hedonics, discussed separately below):

- overlap pricing,
- explicit cost-based price adjustment,
- a deletion link, or
- a class-mean link.

The overlap pricing option can be used when both old and new models are available in at least one period. If the new version is introduced in period t and the old version is also available in that period, the price change recorded for the period t and period $t + 1$ indexes is determined, respectively, by the price relative of the old item for periods t and $t - 1$ and the price relative of the new item for periods $t + 1$ and t. The method does not require direct price or attribute comparison of the old and the new products. Any difference in price in period t is attributed to item quality differences (Kokoski,1993:35). Because the overlap method is employed infrequently, its use (for within-outlet replacement item pricing) is unlikely to be a major source of quality bias in the CPI.[10]

The second option, explicit cost-based adjustment, is regularly used. Cost-based methods were applied to 11 percent of item replacements in 1995; they can be used when information about production cost differences between the replaced and the new items is available. Under the explicit cost-based method, the per-unit change in production cost, as reported by manufacturers, is subtracted from the

[10]It is fairly clear why the overlap method is used infrequently, given CPI price collection methods. If the regularly priced item is available at time t and it is not known that it will be off the shelf in $t + 1$, there is no reason the commodity analyst would price a potential replacement at that time. Even assuming the eventual replacement was available in time t, without prior knowledge the method would require going back, at $t + 1$, to figure out the shelf price of a replacement at the time of the previous trip to the outlet.

change in the observed price paid by consumers. This method, frequently applied to new automobile models, rests on the strong assumption that the perceived value of improved features or new equipment is equal to the incremental costs incurred by the manufacturer to add them. If producers tend to exaggerate the cost of quality improvements, so that reported cost contains an element of pure price change, then the method imposes a downward bias on the CPI (Kokoski, 1993:35). In such a case, a cost-based adjustment erroneously attributes a portion of the observed price difference to quality change. Until competition catches up, however, one might expect profits to be made from an innovative product improvement. In such a case, the price increase would be larger than the per-unit cost associated with the innovation. To the extent that the higher price equates with added value to consumers, the cost-based method (assuming accurate reporting by the manufacturer) might, by ignoring the profit component, understate the quality component of the observed price change.

The Boskin commission judged that BLS use of manufacturers' cost data tends to underadjust for quality change and, in turn, imparts an upward bias to the CPI. The commission estimated quality change bias for new vehicles to be .59 percentage points per year for the period 1983-1996 (Gordon et al., 1997:86). The commission argued specifically that, in the case of automobiles, cost-based adjustment did not include a number of manufacturers' improvements that increased automobile durability and reduced production defects. However, Triplett argues that the Boskin commission assertions about quality bias for new cars were poorly informed (Triplett, 1997:27):

> Bureau of Labor Statistics (1997) listed changes, such as increased use of corrosion-resistant metal, for which cost-based quality adjustments for automotive durability have been made in the CPI. Reduced defects must also have come about from changes made by the car makers. In my experience in the BLS, the auto manufacturers never overlooked quality changes when they submitted costs to the BLS. Rather, manufacturers tried to attribute too much price change to quality improvements. . . . The Commission's idea that quality adjustments are systematically overlooked by the manufacturers when they make reports to the BLS is inconsistent with my experience with these data and also inconsistent with alternative evidence.

Triplett adds that published hedonic studies of new automobiles have produced indexes that rise more rapidly than does the CPI cost-based adjusted index. Griliches (1971:11) also warned that "basing such adjustments largely on data furnished by manufacturers and on 'producer costs' may wind up overestimating 'quality change,' accepting as improvements expenditures which consumers may not interpret as such."

Deletion (and also class-mean, which is a more targeted variant of the deletion method) is used when the replacement and replaced items are judged noncomparable and when neither overlapped prices nor producer cost information is available. Excluding sample rotation, when new independent product samples are drawn, deletion is the most prevalent method used by BLS to address non-identical item replacement in the index (Kokoski et al., 2000:3). If the outgoing item is last seen in period $t-1$ and the new item appears in period t, the old item will be used in index adjustments up to the one made from period $t-2$ to period $t-1$. The new item is used for the adjustment from period t to period $t+1$ and thereafter. For period $t-1$ to period t, change in the index component is proxied by the observed price change of other goods in the same CPI item stratum.

The traditional (non-class-mean) deletion method assumes that the pure price change from the replaced to the replacement item is the same as that for the composite of all other goods in the class. Any remaining portion of the price change observed for the good in question is attributed to quality factors. The direction and extent of residual quality biases associated with this method are disputed, but they essentially hinge on whether the "true" quality-adjusted price change for the item that changed is greater than or less than the measured price changes of the same-class items that were used in the imputation (Triplett, 1997:29).

Triplett and some BLS researchers have argued that deletion can inappropriately attribute a portion of price changes to quality change and, therefore, lead to overadjustment (downward) of price quotes. Their argument is based on the observation that manufacturers are, at least in some sectors, more likely to change prices when a new model is introduced.[11] In an extreme case, if prices changed only when models did, a deletion-based index would pick up no price change at all. Triplett's suspicion of a downward bias is corroborated by Moulton and Moses (1997), who demonstrate that a disproportionate number of price quote changes do in fact occur when new models or varieties of certain goods are introduced. It follows that, if the prices of unchanged models are the only ones that count, as is the case with the standard deletion, the method would impart a bias. Comparative hedonic studies have also indicated a downward bias associated with deletion (e.g., Liegey, 1993).

Prior to the Moulton and Moses (1997) work, the Boskin commission arrived at a different conclusion—that the bias is likely to be upward—stating that the deletion method "bases price change on models that are unchanged in quality and may be further along in the price cycle (Boskin et al., 1996)."[12] It is worth noting that nothing precludes the coexistence of both the type of bias that Moulton/Moses detected and the type that Boskin hypothesized.

Research performed by the BLS indicates that producers frequently take the

---

[11]BLS research showing that price increases tend to coincide with the roll-out of new models is best documented for the apparel and upkeep strata; see Armknecht and Weyback (1989), Liegey (1994), and Reinsdorf et al. (1996).

[12]The underlying assumption here is that product prices drop, or rise less rapidly, immediately after a new product's introduction into the market. This assumption has undoubtedly been true for computers and electronic devices in general but is less clear for other categories of goods and services.

opportunity afforded by introduction of a new model to piggyback price increases (U.S. General Accounting Office, 1999:77). The class-mean method was developed and instituted by BLS to address the problem posed by this pattern. Like deletion, the class-mean method is used to impute the price of a changed item, but it does so from a set of similar goods further limited to those (1) classified as comparable replacements or (2) that could be explicitly quality adjusted by a hedonic or direct cost method. The underlying assumption is that price inflation is different for items that undergo replacement than it is for models that do not change. Looking at this restricted class of goods allows the price trend of a replacement item to be imputed from price quotes observed for other models that have turned over; but by limiting the set to items deemed comparable, it is hoped that quality-related elements are not a major factor. The method has been used in the new cars index since 1989 (Moulton and Moses, 1997:327).[13]

Assessment of Views on Within-Sample Quality Adjustment

As noted above, the Boskin commission report did not seek to identify bias specifically associated with CPI procedures for handling quality change. This was largely a by-product of the commission's decision to estimate new goods and quality change bias together, by CPI category, using independent evidence on quality-adjusted price changes. The work by Triplett (1997) and by Moulton and Moses (1997) indicates that full assessment of quality change should also include an examination of potential biases associated with BLS adjustment procedures. Summarizing the impact of quality adjustments applied by BLS item replacement methods, Moulton and Moses (1997:348-349) conclude:

> For certain important categories of items considered by the Advisory [Boskin] Commission, it would be difficult to argue that the CPI does not overstate the rate of price change. In other cases, however, any bias seems likely to be considerably smaller than the advisory board has estimated and, in certain cases, it could even be negative. . . . Our measurements of quality effects . . . show that any quality bias could go in either direction, either through inadequate quality adjustment (as emphasized by the advisory commission) or through excessive quality adjustment by the application of the link method to items with rising prices.

In addition, Moulton and Moses demonstrate that quality adjustments by the BLS do have a significant effect on measured price change. Examination of BLS methods calls into question the Boskin commission's view that price growth associated specifically with CPI sample items is biased upward.

The debate over the extent to which the treatment of quality improvements produces upward bias in the CPI has been, to a substantial degree, a conflict

[13]We discuss the class-mean method in more detail in the section on hedonics (since class mean is usually what hedonic methods have replaced).

supported by anecdotes. It is not the purpose of this panel to provide new estimates of bias or a detailed technical assessment of BLS procedures. However, the panel has identified several broad issues relating to quality adjustment that deserve attention. First, how can the BLS best assure that the process whereby it identifies and measures quality changes is as objective as possible is not driven by highly subjective assessments of where the problems are likely to be and pays appropriate attention to areas where quality deterioration may have occurred? Second, while adjusting for quality change can in some cases be relatively straightforward, it usually involves product characteristics that are difficult to quantify. Airline deregulation, for example, led to generally lower airfares, but the low fares produced more crowded planes, more cancellations, and more frequent and longer delays (quality deterioration from the standpoint of travelers), as well as an increase in the frequency of flights between many pairs of cities (a quality improvement). How can these serious measurement problems be addressed so that the value of these kinds of quality changes is reflected in the CPI?

Many of the quality change examples used in recent critiques that find an upward bias in the CPI strongly suggest that quality improvements are overlooked. But many of the examples have been chosen from visible product classes presumed to bias the CPI upward. Furthermore, it is difficult to know where BLS should draw the line between adjustments that are sufficiently replicable to be used for producing a publishable index and adjustments that ought to be part of an ongoing research program but are not yet (and may never be) suitable for publication by a federal statistical agency.

HEDONIC REGRESSION METHODS

Hedonics currently offers the most promising technique for explicitly adjusting observed prices to account for changing product quality.[14] Hedonic regressions are used to estimate the value of specific bundles of individual characteristics that, when packaged together, form goods. The principle underlying hedonics is that, if consumers face observable relationships between goods' characteristics and their prices, one should be able to use these relationships to disentangle pure price changes from quality changes. Hedonics essentially replaces the price of goods with the price of bundles of characteristics.[15]

---

[14]This sentiment dates back as far as the Stigler commission report (1961), and is reflected in recent work by Triplett (1990), Kokoski (1993), Boskin et al. (1996), Fixler et al. (1999), and many others.

[15]This basic idea is useful in a variety of other contexts. Particularly when considering product design, marketers routinely treat products as bundles of characteristics; see Green and Krieger (1985). And hedonic regression is routinely used in real estate appraisal and assessment: equations relating sales prices to the characteristics of properties sold during a particular period are widely used to predict the "missing" sales prices of properties that did not change hands; see, e.g., Kang and Reichert (1991).

It is essential to note that hedonic techniques expose a purely empirical relationship between prices and variation among different models of a good. The results of hedonic regressions can be used in either a COGI or a COLI framework. Zvi Griliches, who helped pioneer the application of hedonic methods to price index construction, commented in 1976—and cited the comment approvingly almost 15 year later (Griliches, 1990:189):

> What the hedonic approach attempted was to provide a tool for estimating "missing" prices, prices of particular bundles not observed in the original or later periods. It did not pretend to dispose of the question of whether the various observed differentials are demand or supply oriented, how the observed variety of models in the market is generated, and whether the resulting indexes have an unambiguous welfare interpretation. . . . Its goals were modest. It offered the tool of econometrics, with all its attendant problems, as a help to the solution of the first two issues, the detection of the relevant characteristics of a commodity and the estimation of their marginal market valuation.

One potential advantage of hedonics is that a market may offer products that display a constant set of characteristics over time, even though specific models (and the corresponding characteristic bundles) change. Moreover, in some cases the link between what consumers ultimately value and product characteristics may be more intuitive than the link to a product itself. To this point, Griliches (1990:191) wrote: "Buried within the hedonic idea was the germ of Becker's (1965) 'household production function' and the notion that one should look at the relevant activity as a whole, at its 'ultimate' product in terms of utility or productivity, and not just at the individual components."

A hedonic function relates the price, p_{it}, of variety or model i of some product in some period t, to a vector of its relevant characteristics, $z_{it} : p_{it} = h_t(z_{it})$.[16] In the examples of butter and gasoline discussed at the start of this chapter, z consists of a single variable (ounces of butter and miles of driving, respectively), and there was an implicit presumption that h should be simply proportional to that variable. In more realistic cases, there are multiple relevant characteristics, and h is generally not a linear function of their values. In a typical hedonic regression, price, or the logarithm of price, is the dependent variable, and identifiable and quantifiable product characteristics serve as the explanatory variables.[17] In a well-specified equation, coefficients on the explanatory variables reveal the marginal relationship between the product characteristics and price at

[16]Econometric estimation of hedonic functions dates back at least to the work of Waugh (1928) and Court (1939). This approach received considerable impetus from the seminal work of Griliches (1961).

[17]Interaction terms and nonlinear transformations are also sometimes employed. Some models call for additional explanatory variables such as time period indicators, outlet type, or brand name that may not always be directly indicative of product quality. The implications of the latter additions are discussed below.

various values of z. The basic idea behind hedonic techniques is that one can use a hedonic equation to calculate the expected price of a particular variety—which may not in fact be offered for sale in the period being considered—based on its observable characteristics. Then, as long as the set of observable characteristics includes all characteristics that matter to consumers and the equation is properly specified, these results can be used to correct for product quality change.

To estimate hedonic equations, variation (either cross-sectional or longitudinal, depending on model specification) in the measurable quantity of an attribute is needed to produce coefficient estimates. Categories of goods for which quality change is frequent but incremental, and for which characteristic changes are easy to measure, are considered the best candidates for hedonic analysis. Most obviously, products must have characteristics that are clearly identifiable as valued by consumers. For computers, a relatively easy case, these might include processing speed, hard drive space, memory, and monitor size. However, in most instances, quality characteristics are more difficult to identify, let alone quantify. For example, measuring the performance of cars is highly subjective, as is quantification of their handling, comfort, or safety. Apparel is even more difficult, since consumers' valuations may change over time with fashions. Identifying the characteristics of services that consumers' value can also be very difficult.

The successful use of hedonic methods rests on a modeler's ability to identify and measure quality-determining characteristics and specify an equation that effectively links them to the prices of different models or varieties. It also depends on the availability of good data. In order to produce meaningful results, one generally needs data on more product models than are represented in a typical price index's sample of items. In addition, the reliability of regression coefficients depends directly on the amount of variation (both in terms of presence of indicator variables and magnitude of continuous variables) in the set of characteristics specified in the equation.

In theory, quality adjustments to observed model prices (or of a product-specific index covering all models) can be estimated directly from the hedonic regression. In practice, the critical question is whether one can reliably estimate functions that capture the relationship between market price and characteristics that confront individual consumers. Here, the issue of consumer heterogeneity (see Chapter 8) arises again in a way that affects the index's distributional properties. First, without heterogeneity there would be no hedonic surface in the first place, since identical individuals would choose the same variety (bundle of characteristics) and pay the same price. But because individuals value product characteristics differently at the margin, quality adjustments can alter the relevance of an index as a representation of price changes faced by specific groups or individuals. For instance, people who do not use cell phones do not care about their characteristics, and even the preferences among those who do use them vary greatly. Thus, when prices of cell phones are adjusted to compensate for quality change associated with model turnover, the overall index only becomes more accurate (as a quality-adjusted measure of price change) for one segment of the population. It may become less accurate for those who do not use cell phones.

Moreover, even if everyone faced the same cell phone prices and everyone used a cell phone, the fact that different people choose cell phones with different bundles of characteristics means (because the hedonic function h is not generally linear) that they generally face different marginal prices for characteristics. These issues become thorny when specific groups—such as the poor, the elderly, or those living in a certain area—consume "bias-corrected" products at significantly different rates or in significantly different varieties than do others.[18] Note, though, that these problems are caused by consumer heterogeneity, not the use of hedonic techniques, though they surely complicate our understanding of what is being done with hedonic methods. The hedonic conceptual framework brings to light several difficult issues that have not been fully worked through either by BLS or by academic researchers. In the next section we compare competing hedonic approaches.

Alternative Hedonic Methods

Two basic hedonic adjustment techniques have been developed. For the most part, BLS has pursued an "indirect" approach, designed essentially to supplement price-linking methods when the quality (defined by observable characteristics) of outgoing and incoming models cannot be matched. The method uses a single estimated reference period hedonic function to adjust price quotes—for replacement items that appear at sample outlets—prior to their integration into normal index calculation. The indirect hedonic method is viewed by BLS as an alternative to deletion or cost-based methods for adjusting prices and to the judgments of commodity analysts for assessing comparability of specifications (as such the merits of hedonic methods should be judged against these alternatives). Academic economists, in contrast, have devoted more attention to "direct" techniques, in which indexes are produced directly from estimated hedonic functions based on data for both base and comparison periods.

Indirect Methods

BLS uses the term "indirect" in reference to a specific way of using hedonic functions to deal with situations in which one variety of a good tracked in the CPI

---

[18]As the economy moves toward greater product heterogeneity, this aggregation problem may potentially increase with or without hedonics. Noncomparable item replacement procedures generally attribute price differences (or portions thereof) to quality difference, though many people would not be willing to pay that difference. Independent of the mix of quality adjustment techniques used by BLS, there is a positive correlation between the extent of changes in product characteristics and the magnitude of the aggregation problem. The magnitude of the problem, even without hedonics, is not necessarily an argument for ignoring changes in product characteristics.

system—with a specific vector of characteristics z_1 and price p_1—disappears and is replaced by another—with "similar" characteristics z_2 and price p_2. This method is indirect because it involves adjusting, post hoc, the observed price difference between the outgoing and the replacement items based on the portion of the price change attributable to quality change. The magnitude of the adjustment is determined by the estimated hedonic function and the differences between the characteristics bundles supplied by the old and new items. It is possible and, for reasons of data availability, often necessary to base these adjustments on a hedonic function that is estimated with data from a period well before the substitution occurs.

Though the difficult econometric problems that plague all hedonic analysis—e.g., identifying appropriate functional form and relevant product characteristics—complicate the indirect method, it has considerable commonsense appeal, at least relative to the alternatives: using the price relative (p_2/p_1) with no adjustment for quality change, assuming that the observed price change is due entirely to quality change, or adjusting for quality using one of the standard replacement methods. The fact that, among hedonics approaches, the indirect method is the least demanding in terms of data and procedure adds to its practicality. It simply requires using cross-sectional price and model characteristics data (similar to that which BLS already tracks) to estimate hedonic functions periodically.[19] This function can then be used to estimate the price that would have been charged in the period studied for new models (with the same characteristics but different quantities of them) that are to be brought into the index (see "Technical Note 2" below). In contrasting alternative hedonics approaches, it is imperative to understand that the indirect method is applied by BLS in a comparatively narrow manner—to adjust price quotes, gathered under normal procedures, of replacements for items that have permanently disappeared from a sample outlet. For most products that are now hedonically adjusted (all by the indirect method), the monthly number of quotes adjusted is quite small, as is the effect on the monthly index for the relevant stratum.

Direct Methods

Two distinct direct hedonic adjustment approaches have been developed: the direct time dummy method and the direct characteristics method. In the direct time dummy method, data from multiple periods are used to estimate coefficients of a function relating the logarithm of price to a set of product characteristics and a set of 0-1 dummy variables for the periods covered.[20] As discussed below, this

---

[19]Though the type of data required are similar, BLS typically has needed to expand its sample, or purchase commercial data, in order to generate a sufficient number of price points to estimate the hedonic models recently introduced into the CPI process.

[20]Work on the time dummy method has mainly been developed in the academic literature. Key studies include Griliches (1990), Triplett (1990), Berndt et al. (1995), and Arguea et al. (1994).

procedure assumes that for any two periods, t and u, $h_t(z) = K_{tu}h_u(z)$ for all characteristics bundles z, where K_{tu} depends on t and u but not on z. That is, between any two periods the prices of all models (actual and potential) are assumed to differ by the same percentage. If this assumption is correct and the hedonic function is correctly specified, the characteristics variables pick up all price changes driven by quality changes in the menu of varieties on the market and coefficients on the time dummies pick up the residual pure price change. The index—interpreted as the price ratio net of the quality component captured by the characteristics variables—is produced directly from the difference in the time dummy coefficients from period to period. If the dummy variable for the base period is omitted, as is standard, the antilogarithm of the time dummy coefficient for any other period t gives the ratio of the price(s) of the good in question in period t to the price(s) in the base period.[21] Similarly, the antilog of the difference between the time dummy coefficients for any two periods gives the price relative between those periods.

Under the time dummy method, a single regression covering all periods must be run each time the index is produced. Since regression coefficients involving the characteristics are held constant across periods, changes in marginal cost ratios or in consumer demand patterns are assumed to be negligible. Thus, the basic relationship between product characteristics and relative prices (as well as the mix of characteristics available at market) must be stable in order to accurately isolate the price component associated with quality change over successive periods. This stability is what allows time dummy coefficients to be interpreted as the pure price effect.[22]

The key problem with the time dummy approach is that, for product areas in which quality change bias is likely to be an issue, the relationship between price and characteristics often changes rapidly. As an example, it is unlikely that consumers value, on the margin, a 10 percent increase in computer hard drive memory the same now as a year or two ago. If regression coefficients assumed to be

---

[21]Triplett (2001b:6-7) notes that the dummy variable method, when specified in a double-log or semilog functional form, produces a price index based on the geometric mean formula. Since statistical agencies have begun moving toward using the geometric mean formula to construct elementary item indexes (for other reasons), time dummy approaches have become more consistent with the prevailing methodology.

[22]The problem of obsolete regression coefficients on characteristics is not unique to the time dummy approach. Certainly, the coefficients produced by the indirect approach, if not updated, are also susceptible to the same problem. However, the magnitude of the effect that the changing "true" relationship between characteristics and price can have on the index is more limited for the indirect approach. An index that is adjusted with the indirect hedonic approach will typically be less volatile because it is only affected by those variables representing characteristics whose values have changed from one period to the next. By contrast, all product characteristic variables that experience a divergences between their estimated and "true" relationship to price affect the time dummy coefficients and, in turn, any index derived from them.

constant over time are in fact not constant, the estimated time dummies will reflect a mixture of pure price changes and quality changes, and the resulting index will be biased. More generally, there is neither theoretical support nor much empirical evidence for the assumption that prices of all varieties of particular products generally move proportionately over time.

The second direct approach, variants of which have been suggested by Pakes and Levinsohn (1993), Feenstra (1995), Diewert (2001), and others, is what we call the direct characteristics method. Under this approach one estimates a separate hedonic function for each period and computes price relatives for the product under study by, in effect, comparing the functions for the periods involved. The idea is not to require that all estimated nondummy coefficients differ between periods; it is rather to impose between-period coefficient equality only when that hypothesis withstands statistical scrutiny. To the extent that data from multiple periods can be pooled, estimation efficiency (always a concern in these studies) can be enhanced.[23] In contrast to the time dummy approach, the direct characteristics index is—as its name denotes—constructed from the characteristics coefficients, which are in general allowed to vary over time. The method also offers an advantage over the BLS's deletion or indirect hedonics methods in that it allows for correction of any sample selection bias that may be created because price changes are only sampled from the set of goods or services that remain unchanged from period to period.

However, despite its conceptual appeal, there are reasons that, given the current state of the art, the direct characteristics approach does not have broad applicability across CPI categories. One issue, which applies to all direct methods, involves the general problem of price data that reflect nonobservable seller attributes. Outlet bias (discussed in detail in Chapter 5), for instance, is difficult to control for in an index produced from a time dummy regression or by relating hedonic functions for successive periods. In contrast to other methods in which prices for replacement items are quoted from the same outlet, product price and characteristics data are combined from multiple sources to estimate direct hedonic indexes (Triplett, 2001b:3).

The most obvious obstacle to widespread use of direct hedonic methods, though, involves the data requirements and the operational difficulty of producing characteristics-based indexes on a high-frequency, up-to-date schedule. To produce such an index, routine data collection and processing procedures would need to be directed toward monthly pricing of a comprehensive set of varieties, chosen to represent a population's consumption, rather than a limited sample. Most importantly, it would be necessary to gather data on the sales of all impor-

[23]Two related direct approaches, both of which give the same result as the direct time dummy method when its assumption of stability of nondummy coefficients (and thus of proportional shifts in prices of all varieties) is satisfied are discussed below in "Technical Note 2."

tant varieties with at most a one-period lag (see "Technical Note 2" below). Given current technology, estimating hedonic surfaces for, say, September 2000, in time for the release of the corresponding monthly CPI is infeasible for most goods and services.

Due to the narrow range of products for which data requirements can be satisfied, even proponents of the direct characteristics method acknowledge its current limitations. Pakes recommends starting with computers and moving slowly into other areas. Writing about the applicability of hedonics to index construction, he cautions (Pakes, 1997:9):

> There are, of course, several detailed decisions which will have to be made before the statistical agencies could produce hedonic adjustments for a set of subindices (among them a decision on the instances in which the hedonic bound is likely to be suspect). Moreover, any shift to hedonics will require prior experimentation by commodity group, and will generate adjustment costs.

These warnings aside, moving in this direction would not require a complete overhaul of BLS pricing methods. In fact, data requirements for the direct characteristics method fit in fairly well with current collection procedures since characteristics must now be tracked in order to judge comparability for replacement situations. Even to improve the CPI under current methodology, better quantity and characteristics data are needed, which is what would also be needed here. In addition, the current requirement that identical products be found at outlets by BLS field agents over index periods might be relaxed since only characteristics—which might be found on a number of similar products—need to be tracked.

Silver (1999:19) suggests that agency data collection needs for hedonic indexes might be met in the future with panels of consumer data for frequently purchased items and scanner data for durables. Paasche, Laspeyres, and superlative formulations could be produced, assuming that the time needed to process comprehensive product scanner data is short enough to allow for base and current period weights to be constructed.

BLS Application of Hedonic Methods

The Boskin commission attributed more than half of its estimated 1.1 percentage point CPI bias to a failure of the index to fully account for changing product quality and the appearance of new goods. The BLS has responded to recommendations by the Boskin commission and others (both before and after Boskin) to address this perceived flaw by expanding its use of hedonic quality adjustment—specifically, the indirect method—in the CPI. Kokoski et al. (2000:3) characterize the hedonic method, or class of methods, as the "currently preferred method of quality adjustment." The BLS position is that hedonic analysis provides meaningful information for inferring the value consumers place on quality change and that hedonic function estimates based on regression analysis can be reliably used for certain items to make quality adjustments to indexes.

Hedonic methods were first applied in the CPI during the early 1990s to apparel categories. Initially, the technique was used to develop criteria for identifying appropriate comparable replacements for disappearing items. Shortly thereafter, it evolved to its current use as a tool to filter out the quality-driven component of observed price changes associated with item replacements. For apparel and most other items, hedonics is typically used to make one-time quality adjustments (using the indirect method) when new items replace outgoing ones.[24]

Since January 1999, hedonic analysis of computers and televisions has been incorporated into the official index. The TV index was considered a good candidate since models undergo frequent but nonradical quality change and also because TVs occupy an entire item stratum. The latter feature is convenient because it allows the elementary index to be constructed without combining separate substratum item indexes calculated using different methods. For TVs, a semilog functional form is used to adjust observed prices to account for variable quality characteristics, such as screen size, wide screen, display, projection, and surround sound. Equations must be respecified as new features appear, which become new explanatory variables (Fixler et al., 1999:10). An indirect approach, as described above, is used.

BLS has also tapped into research done for the producer price index (PPI) to develop hedonic regressions for large-scale and network as well as desktop computers. Hedonic adjustments have been used in the PPI since 1990, and application to the desktop computers stratum of the CPI was incorporated in January 1999. Explanatory variables include such characteristics as chip speed, system memory, and monitor type and size (Fixler et al., 1999:11). Since computer technology changes rapidly and the relationship between computer features and value appears unstable, regressions have been respecified frequently (every 3 to 12 months). For 1998, the first year that hedonic regressions were used to adjust computer prices, the price index for personal computer and peripheral equipment was reduced by 6.5 percent, relative to what it would have been had the new method not been implemented.

A hedonics technique has also been applied to the "rent of primary residence" and "owners' equivalence" components of the index since 1988. A more restrictive type of the indirect hedonic method is used to estimate only the effect of aging on the value of housing units. In the application, rents are adjusted for age using a nonlinear (age and age-squared variables) specification.

Items Targeted by the Recent CPI Hedonics Initiative

The BLS is currently conducting research that will extend the use of hedonic regression models to additional CPI items. Kokoski (1993:12) states that "for

---

[24]Interestingly, over the last six months of 1991, using hedonics raised the rate of price change for the apparel category by 0.4 percent per year (Liegey, 1994).

many CPI components, a hedonic approach will likely be adopted before the next scheduled revision in 2002." In fact, the move to expand the hedonics program has already accelerated in response to the fiscal year 1999 CPI improvement initiative.[25] The initiative has added hedonic price adjustment to the following item groups:

- microwave ovens (effective July 2000)
- refrigerators (effective July 2000)
- camcorders (effective January 2000)
- VCRs (effective April 2000)
- DVDs (effective April 2000)
- audio products (12 products, effective January 2000)
- college textbooks (effective July 2000)
- washers and dryers (effective October 2000)

Specific CPI strata are chosen for the expanded hedonics program using the following criteria (Fixler et al., 1999:13):

- There is a perception that the standard procedures inadequately account for quality change (it is unclear if BLS is focusing on items for which quality is perceived to be changing or specifically increasing).
- There is a belief that hedonic models could be developed for at least a subset of items in a stratum (presumably beliefs must be supported by both data availability and theoretical considerations).
- A significant percentage of price quotes exist for substitute items relative to the total number of price quotes for the stratum.

Fixler also notes that the list of candidates provides nice contrast in terms of placement within item life cycles. At one end, DVDs are new and have recently undergone rapid technological development while, on the other, refrigerators and microwaves are well into the product cycle and technology is comparatively stable. It is not surprising that the goods included in the initiative are from the "appliances including consumer electronics" category since it is a product area that has undergone highly visible change. (This category accounted for the largest share of the Boskin commission's estimated 0.6 percent unmeasured quality and new goods bias.)

Much of the ongoing and proposed hedonics-related research must be supported by additional data collection, since routine CPI sampling procedures often

---

[25]The hedonics project was one component of the 1999 CPI improvement initiative, which also designated funds to expand the sample size of the Consumer Expenditure Survey (CEX), to quicken introduction of new products and CPI market updating, and to develop new superlative indexes (Liegey and Shepler, 1999:34).

yield an insufficient number of models to permit reliable estimation of realistically complex hedonic functions. The initiative specifically provided funding to collect additional price observations from current CPI outlets (2,500 quotes distributed among eight items). For some experiments, BLS field agents are also collecting observations from added outlets (as is the case for camcorders); for others (audio products), market data have been purchased from vendors such as A.C. Nielsen or NPD.[26] It is important that BLS continue to examine the implications of using non-uniform data for estimating hedonic regressions and for index construction generally.

The CPI Hedonics Model

Most of the recent BLS work uses the indirect adjustment method. Price adjustments are calculated with an equation in which the (logarithm of) price is estimated as a function of product characteristics. The portion of the observed price difference between a replaced item and a substitute item assigned to differences in quality is determined by the differences in characteristics variables and the associated coefficient values. The process of specifying the model typically involves researching consumer magazines and manufacturer and industry information to develop a sense about which characteristics are related to price. Several specifications may be experimented with prior to final model selection. In the case of VCRs, BLS's final specification consisted of all dummy variables on the right-hand side, each indicating the presence of a particular feature (number of video heads, auto rewind, hi-fi stereo, etc.). Liegey and Shepler (1999:27) write: "The specification for the final VCR regression model was deemed satisfactory, primarily because the magnitude and direction of the parameter estimates matched *a priori* expectations. The high R-squared value further validates the model." For several of the applications, the dependent variable is the list price, not the transaction price. When the model is estimated using retail list price as the dependent variable, a dummy variable indicating that the item was sold at a sale price is included in the model to capture the (negative) effect on actual price in the data.[27]

In addition to tangible characteristics, brand dummies are often included as explanatory variables. Inclusion of brand names in the equations has been de-

---

[26]The audio project, which relies on purchased point-of-sale data, and the video project, which relies on conventional in-house surveys, may provide useful contrasts. The audio data include price and units sold but have limited information about attributes, which forces BLS to supplement the data with manufacturer specifications. Typically, collecting vendor data is more expensive than collecting a sample internally, but the data are available with greater frequency.

[27]This is often done because of data constraints. Earlier studies, such as Liegey and Shepler (1999) on VCRs, used *Consumer Reports* and not CPI price data. Mary Kokoski has questioned this practice commenting that, "since no one really pays full prices, do they (the results) really reflect the equilibrium assumptions that underlie the hedonic method?" (Liegey and Shepler, 1999:32). The panel shares Kokoski's concerns.

fended on the grounds that coefficients for the variables were robust, that it increased the explanatory power of the models, and that is does not create multicollinearity problems.[28] Nonetheless, including brand names is controversial. It is reasonable to worry that the brand variables may "steal" effects from other characteristics—both those that are and that could be included in the model—and thereby bias the estimated effects of characteristics on price. If one assumes that brand, in itself, does not lead to higher valuation by consumers, one must believe that it is an acceptable proxy for unmeasured quality characteristics. Incidence of repairs might be one such example. However, the BLS study on microwave ovens found that the brand names most valued by consumers were actually those with the highest incidence of repair (Liegey, 2000:5). Moreover, brands are repositioned in terms of relative quality from time to time, and reputations sometimes change in response to advertising campaigns, so that brand dummy coefficients may be inherently unstable. Given the difficulty of interpreting coefficients on brand variables, it would be instructive if researchers documented their results with and without brand variables and provided a hypothesis as to what aspects of product value the brand variables are capturing.

The BLS hedonics research program has helped reveal that, in practice, applying hedonic methods to price indexes involves confronting very tough issues. Characteristics cannot be chosen in a formulaic manner—lots of ad hoc judgments are inevitable—and, once chosen, estimated coefficients may exhibit implausible signs.[29] Furthermore, models need to be regularly updated because the relationships between characteristics and price are not stable for long periods. For instance, Liegey and Shepler (1999:27) show that new features on VCRs have a large predictive effect on price but, as they become common, their impact quickly recedes. This is a good example of the kind of work BLS must continue to undertake to support expansion of its hedonic program. Investigations into model stability for different product areas are much needed to improve judgments about the frequency with which hedonic regressions should be reestimated.

While the panel believes that the BLS research program is essential to improving understanding of the theoretical uncertainties about hedonic methods, our concerns have not been allayed by what has actually been done so far. Given these ongoing concerns, we are still quite uncomfortable with extending the

[28]Moulton et al. (1998) include indicator variables for brands in their study of televisions. They argue that brand name is important since "a set with the same screen size and other observable characteristics with a premium brand name, such as Sony, may sell for as much as 50 percent more than a similar television from a less prestigious brand" (p. 9). The authors acknowledge that, if additional characteristics could be added to the regression equation, the effect of brand variables might be reduced.

[29]Pakes (2001) has argued that, given rapid entry and exit and great product differentiation in technologically innovative markets, it may not always be clear what the "right" sign on a characteristics variable should be.

application of hedonic models, in their current state of development, to additional index categories for use in the CPI. Yet the panel is not convinced that anyone could have done this work any better—or is better equipped to continue work in this area—than BLS.

Hedonics Use and the CPI

To make clear the implications that the shift from implicit quality adjustment to hedonics may have on the CPI, we must first describe the BLS in-store item replacement procedure.[30] When a routinely priced item becomes permanently unavailable, BLS field agents are instructed to choose the most similar alternative available at the outlet on the basis of a hierarchical list of characteristics specifications. As explained above, the pure price change for a replacement item in the month of its introduction in the index is measured as the average price change that month among similar items (selected according to one of several different methods). Any remaining difference between the price of the replacement item and the old one is assumed to reflect quality change. The practice of choosing the "most similar" item as the replacement means that the potential quality difference between outgoing and incoming items is smaller than if the practice were to select the most advanced, the newest, or the most frequently purchased product within the same class. Also, this practice increases the number of substitutions that are deemed comparable and that do not require quality adjustment and reduces the magnitude of noncomparability for items that are price adjusted. In 1997, 58 percent of the almost 29,000 nonrent substitutions were judged comparable by commodity analysts (U.S. General Accounting Office, 1999:13). For the subset of substitute items that are deemed noncomparable, BLS then attempts a direct quality adjustment, using hedonics or cost-based calculations or a traditional indirect adjustment method.

Other than in its application to personal computers, hedonic adjustments are producing little if any effect on the CPI.[31] The effect of increased use of hedonics is limited by:

- its narrow application to noncomparable substitute price quotes,
- the nature of CPI item substitution itself, and

---

[30]U.S. General Accounting Office (1999) provides a detailed explanation of the rules that guide item replacement by BLS commodity analysts. The report also describes how the class-mean and other link methods of adjustment work.

[31]It should be noted that, even with hedonic adjustment, the rate of price decline for personal computers in the CPI or PPI is generally much smaller than that estimated by outside researchers. The research in this area is quite controversial.

- the fact that hedonics methods are being applied only to items that were previously quality adjusted by other BLS techniques.

CPI hedonics models are only used to adjust a subset of *substitute* price quotes: that is, to control for quality change when a product that has disappeared from the shelf must be replaced by a noncomparable one. Hedonic adjustments have not been used to offset quality differences associated with item turnover generated by outlet rotation or with respecification of the CPI basket. This is an important distinction because implicit forms of quality adjustment (such as deletion) are already a feature of the in-store item substitution process. As currently used, hedonics is simply replacing another method of quality adjustment, and the item-by-item effect on index growth has been minimal and its direction ambiguous. In some cases, BLS's hedonic models implied price adjustments that would have been larger than the standard (deletion) adjustment actually used; in others the adjustment would have been smaller.[32]

A more broadly based application of hedonic techniques—one that extended beyond routine item replacement cases caused by sample attrition to one that, for example, was also applied to price changes associated with new models appearing as a result of sample rotation—would be expected to have a larger effect on the index. For example, Moulton et al. (1998) simulated an index for televisions that did include hedonic comparisons of items that entered the CPI through sample rotation. Their analysis, which tracked product characteristics over a 5-year period, resulted in a much larger downward adjustment than a simulated index that applied hedonics only to in-store item replacements. The authors argue that sample rotation may be particularly important for TVs since, unlike computers, when models with new features appear on the market, older sets of characteristics (models) remain available for a long time. This means comparable replacements can continue to be found, and commodity analysts need not turn to more radically changed models. Cutting-edge models, even those that quickly gain in market share, seem more likely to enter the CPI when outlet and item samples are rotated (Moulton et al., 1998:12).

Moulton et al. (1998) recommend developing hedonic adjustment and data collection techniques that would make it possible to apply hedonics methods when new products enter the sample during outlet rotation. They also suggest changing the item replacement rule to have field agents select items that are more

[32]This evidence does not speak to questions about the extent to which adjustments are becoming more accurate; matched model (replacement or class-mean) adjustments may conceivably overstate or understate the impact of quality change on price and, while hedonic regressions produce additional evidence about this, the value of the additional information is dependent on the validity of the model and the quality of underlying data used in its estimation.

representative of prevailing quality choices in current consumption, then adjusting the selected substitutes with the standard indirect hedonic models.[33]

Since hedonics currently only replaces other procedures, its adoption by BLS has not led to more widespread adjustment for quality change in the CPI. Hedonic applications have most frequently taken the place of the class-mean method. As discussed above, the class-mean approach, like the deletion method, infers quality differences by comparing the observed price change of the replaced and replacement items to the price change of other goods. However, compared with deletion, the pure price change is imputed from a smaller set of quotes. Quotes are still drawn from a specific entry-level item (ELI), index area cell, but only those for items that are *comparable* replacements or directly quality adjusted are included to calculate average price change for the class. The idea is that the price change of items that are comparable to what they have replaced will reflect only pure price change. This approach is designed to recognize the fact that sellers often use the occasion of introducing model changes to raise prices. The estimated monetized value of quality change is the residual of the observed price change of the substitute item after the average price change of similar but comparable replacement items has been accounted for.

The class-mean method was instituted during the late 1980s partly to capture price increases that accompany introduction of new models. It is the designated method for item strata that experience frequent model and product line turnover. The deletion method misses these price increases because it only follows price movement associated with unchanged models. For this reason, Triplett (1997), Shapiro and Wilcox (1996), and BLS researchers have suggested that the method can overstate quality change and bias the index downward. The class-mean method, by imputing price change only from comparable replacements, attempts to account for the possibility that inflation is different for models that are changing in comparison with models that are stable. But if producers are systematically more likely to include "pure" price increases (or decreases) for new models that are substantively different from old ones than they are for those with only a change in model number (as is the case with 60 percent of comparable VCR substitutions), then this correction may still understate pure price change. In this case, the class-mean method may still lead to a larger quality adjustment than a

[33]The issue of whether or not the current item replacement rule—choosing the closest comparable item—makes sense is a separate but important one. In quickly evolving technology areas, if the replaced item became obsolete, it is likely that the closest substitute is also near obsolescence. If, instead, BLS agents selected the newest model or the one with the highest sales, the frequency with which item substitutions must be made could possibly be reduced. However, such a change in procedure would require making larger quality adjustments, which might pose other problems, particularly if one is not confident that currently available methods can really disentangle pure price and quality contributions to the observed price. A problem with using the newest model is that it will sometimes pick up features that do not last—not all innovations survive in the market.

perfectly specified hedonic adjustment. But hedonics equations cannot be perfectly specified, and BLS research indicates that, in practice, the adjustments fall on both sides of the class-mean imputation. We look at some of these research results next.

The Recent BLS Hedonics Initiative

BLS's hedonics studies are designed to produce equations that can be used to adjust noncomparable replacement item price quotes. Several of the studies compare the performance of a hedonically adjusted index against one simulated with the same methods as the published index. In as many cases as not, the hedonically adjusted price index increases at a faster rate than does the published (class-mean based) version. In other words, the hedonic quality change price adjustment is often smaller than the conventionally used implicit quality adjustment. For instance, for the period June 1999 through December 1999, substitute VCR price quotes used in the published index decreased on average by 4.3 percent (or, annualized, by 7.4 percent). For the same 7-month period, the hedonically adjusted price for substitute quotes decreased by an annualized 2.2 percent (Thompson, 2000:6). This translates into 13.2 and 11.5 percent annualized decreases, respectively, for the published versus hedonically adjusted indexes for the other video equipment strata, of which VCRs are one subcategory. The study of VCRs by Liegey et al. (1999) showed similar results using 1997 data: hedonically adjusted price quote substitutions also grew less rapidly than the published indexes for refrigerators and audio equipment.

In contrast, some of the experimental hedonics applications have slowed index growth by more than the implicit adjustment methods. The TV index falls more rapidly with hedonic adjustment. Moulton et al. (1998) produced a hedonically adjusted index (of the type that has been adopted for other items in the CPI) for the period 1993-1997 that grew 1.4 percent less than the actual CPI for televisions that used the linking methods (p. 11). Table 4.2 summarizes the effects on the CPI from a selection of recent hedonic applications.

For many CPI items, the number of substitute quotes that are available to quality adjust is not large enough for the hedonic adjustment to seriously affect the strata index, much less the overall index. Moulton et al. (1998) note that confining hedonic adjustments to cases of noncomparable substitutions for anything other than very high turnover products like computers will not produce many significant effects on the CPI component indexes.

Another factor that may affect index growth is the frequency with which items are deemed noncomparable and, hence, eligible for hedonic or class-mean adjustment. In each of the BLS studies, the breakdown of substitute items into comparable and noncomparable categories changed with adjustment mode. For instance, in the microwave oven study, moving to hedonics resulted in an increase in the number of noncomparable substitutes increasing from 5 (of 39

TABLE 4-2 Major Effects on CPI Indexes from Five BLS Hedonic Studies

Product	Stratum	Change in Stratum Index		Average Price Change for Product Substitutions		Number of Monthly Substitutions in Product Group
		Published	Hedonically Adjusted	Published	Hedonically Adjusted	
VCR	Video excluding TVs	–13.2	–11.5	–7.4	–2.2	19
Audio products	Audio products[a]	–7.4	–6.0	n.a.	n.a.	n.a.
Refrigerators	Major appliances	1.2	1.2	2.8	3.9	5
Microwaves	Major appliances	1.1	0.8	5.4	1.7	4
DVDs[b]	Video excluding TV	n.a.	n.a.	n.a.	n.a.	1

NOTE: Index and price changes shown in percent and at annual rates.

[a]Not a CPI stratum.
[b]Too few price quotes to estimate an adjustment to the overall stratum index.

quotes) to 21. For VCRs the noncomparables decreased from 63 (out of 130) to 47. BLS explains why using a hedonics system might identify different noncomparable cases from a given set of substitute quotes (Thompson, 2000:5):

> For the purpose of calculating the quality adjusted index, all the substitutions were reevaluated. One of the benefits of using a hedonic model in evaluating substitutions is that the analyst has an opportunity to review price data and item characteristics with a statistical tool, thus enabling him/her to render judgments based on statistics rather than expert judgment alone.

When BLS uses hedonics, comparability of substitute quotes is judged not in terms of an a priori determination about the amount of quality change, but by the extent to which price change is predicted by the regression equation.[34] For substitute price quotes, "differences in the specification or characteristics data of the old and new items were identified to see if the parameter estimates in the hedonic model could be utilized to quality adjust the official price change" (Thompson,

[34]The "standard" comparability decision in forced when a commodity analyst must add to the sample a replacement item that does not match the detailed description of the old one. The analyst judges comparability on the basis of an examination of any differences revealed by the checklists for the old and new versions of an item.

2000:10). Of course, the way a model is specified affects where the line of comparability will be drawn, and it has not been established that the statistical procedure is an improvement over the standard process, particularly if the hedonic regressions are estimated by analysts who know little about the products being adjusted. It is essential that the expert judgment of commodity specialists be retained and perhaps supplemented with that of other product and marketing specialists.

The fact that hedonic methods have produced item comparability judgments that are different from those made in conjunction with deletion methods is a concern (relevant to either method) for several reasons. First and most obviously, different judgments about comparability lead to different rates of quality adjustment, which has obvious implications for index performance—more price adjustments translate into a decreased rate of index growth. Second, the empirical work produced from the recent initiative indicates that price change associated with the comparable substitute quotes can also be significantly affected by the choice of quality adjustment technique. For instance, the 67 directly compared (non-adjusted) VCR price substitutions produced for the published class-mean adjusted index decreased by 3.72 percent; the 83 directly compared price substitutions left over after hedonic adjustments were added decreased by only 1 percent. It is not intuitively obvious why the choice of noncomparable quality adjustment method should have such a large effect on the price change of the subset of comparable substitute quotes. Documentation supporting BLS research does not adequately explain the effect on comparables that results from switching quality adjustment methods.

These criticisms aside, the move toward supplementing judgment—both formal and informal—with replicable, systematic methods of comparing non-identical items is a move in the right direction. Econometric analyses of data that indicate how characteristics are correlated with price change have the potential to improve the ability of BLS analysts to determine what is and is not a comparable substitution. Even if a particular hedonic study is not convincing enough to be used for quality adjustment, it may still offer insights that improve analysts' informal decision making. Statistical audits provide evidence about the variance that arises when different researchers, using the same data, attempt to replicate quality adjustments (Triplett, 2001b:9).

Summary

The incorporation of recent BLS hedonics research into the CPI has not produced evidence for the conclusions offered by the Boskin commission and others about the extent to which quality change biases the CPI when used as an approximation to a COLI. The research indicates, at least, that the commission underestimated the effect on the index of implicit quality adjustment measures already in place. Moreover, even a substantial expansion of hedonics, used as the

BLS now does in the item replacement process, would not be likely to have a big effect on the CPI. Hedonic adjustments tend to wash out relative to those produced by the implicit adjustments that they replace (the computer index is the exception). Confining hedonic adjustment to cases of noncomparable substitutions for anything other than very high turnover products is unlikely to significantly affect CPI component indexes. Also, current BLS rules for replacing disappearing products further minimize quality differences between outgoing and incoming products which, in turn, lessens the importance of which type of quality adjustment is ultimately selected. However, as the Moulton et al. (1998) TV study suggests, the application of hedonic adjustments in a different way and on a larger scale might produce more significant downward adjustments. The panel believes that the BLS should proceed cautiously in its efforts to integrate hedonics into the CPI. Further research, testing, and evaluation of hedonic methodology and specific applications should precede expansion of its use, such as to sample rotation—something that the panel is not in principle opposed to—where the impact on index growth would likely be more significant.

CAUTIONS AND RECOMMENDATIONS

Hedonic methods are not a cure-all for indexing problems related to quality change. Regression techniques do not deal with increases in product variety (e.g., of fruits and vegetables during the winter); nor do they help much with the problem of truly new goods (e.g., cellular phones). The main thing to be said for hedonic methods is that there is nothing better for dealing with certain aspects of the quality change problem. This is not an elegant defense, but it is a powerful one. To a large extent, this reality shapes our recommendations in this area.

BLS should systematically investigate quality change across CPI components.

> **Recommendation 4-1: In addition to its targeted intuitive approach (in which BLS selects for adjustment items thought a priori to have undergone quality change), BLS should pursue experiments to analyze quality change in randomly selected items in order to increase the probability that within-sample quality change biases—both upward and downward—will be identified. Currently, hedonic regression analysis is the leading candidate to serve as the main analytical tool in such experiments.**

One issue that will have to be addressed in such a program is the level of detail that is used in the item selection process. Selection could be randomized across the broad 211 item strata, at more detailed ELI levels, or somewhere in between. Whatever level of disaggregation is chosen, it is logical that selection probability should be proportional to expenditure (perhaps adjusted to account for the rapidity of item replacement) and not random over items independent of weight.

The issue of how to assess service expenditure categories will also pose special problems. One can imagine that the quality of various consumer services changes substantially over time, and certainly not always for the better—think of airline travel for instance. In principle, methods such as hedonics that are used to identify and adjust prices of quality-changed goods can also be applied to services; in practice, for many services, the problem of how to define output appropriately must first be solved. The above recommendation identifies one element in what should be a broad-based hedonics research program.

Recommendation 4-2: BLS should continue to expand its experimental development and testing of hedonic methods and its support of relevant outside research. This research should not be confined to that relating to price adjustment but should also examine the role of hedonics in statistical audits of the other BLS quality adjustment methods and in the review of replacement item selection procedures and comparability decisions.

The above recommendations do not suggest that BLS should immediately expand the use of hedonics in constructing component indexes for its flagship CPI. In fact, the panel takes the opposite position.

Recommendation 4-3: Relative to our view on BLS research, we recommend a more cautious integration of hedonically adjusted price change estimates into the CPI.

This recommendation is based on theoretical considerations, not on empirical grounds. As documented above, the recent BLS expansion of hedonic price adjustments to appliances and electronics has not had a large impact on those item subindexes. The current hedonics program, which only replaces other quality adjustment techniques, actually has an ambiguous effect on index growth. Thus, for practical purposes, the apparent rapid expansion of the use of hedonics is not a pressing empirical concern for those interested only in the accuracy of the final CPI numbers.

Our conservative view on integrating hedonics techniques has more to do with concern for the perceived credibility of the current models. While there is an established academic literature on estimating hedonic functions, researchers are much less experienced using them across a wide variety of goods in price index construction. Thus, while members of the panel agree that BLS and others should continue to research the viability of hedonics, the methods may, in their current state of development, only be justifiably applied to a narrow class of goods. The list of unresolved econometric specification and data issues that may inhibit fully informed use of hedonic quality adjustment is a long one.

For many classes of goods—and perhaps especially services—it can be extremely difficult to identify which characteristics are actually associated with price. Despite the early success of hedonics to move quality adjustment in the CPI toward a statistically based approach, considerable judgment by researchers is still required. For instance, early introduction of video memory as an explanatory characteristic in regressions for PCs yielded "implausibly high coefficient values," so the variable was left out of initial specifications. Later the values "settled down and behaved much more reasonably" and the feature was included in the specification (Fixler et al., 1999:11). Likewise, in the hedonic model for TVs developed by Moulton et al. (1998:10) the "stereo sound" indicator was dropped because it predicted a negative (though insignificant) impact on price. Given the short history of this type of research at BLS, it is not clear what the benchmark should be for assessing what is or is not reasonable. Strange-looking variable coefficients could be indicative of larger problems—including omission of key value indicators, characteristic mismeasurement, and functional form issues.

Whether for a standard comparability decision or for hedonic modeling, identifying and quantifying relevant characteristics is tricky when quality is tied to consumer perceptions that may not be constant over time. Also, it is next to impossible to collect full and timely information on certain types of product characteristics. Quality of fabric in clothing, for example, is determined by a complicated combination of characteristics—not simply by material type, but also by threads per inch, type of weaving, quality of dye, etc. Given the changing nature of fashion, a characteristic may be viewed as a negative quality at one time and as a positive quality at another. For instance, the original move from cotton to synthetic shirts was considered a quality improvement—but so too was the move back to cotton.

Once identified, it is not necessarily any simpler to measure the characteristics thought likely to affect price: consider stylishness in clothing or handling in cars. Even for the best candidates, such as computers, attribute measurement can be problematic. For instance, how does one quantify the user friendliness of hardware or software? For most products, certain elements that contribute to its value will always be difficult to measure consistently.

Theory provides little guidance to help determine the appropriate functional form for hedonic equations. Experience suggests that characteristics often interact in complex ways to affect value. When characteristics work in combination, nonlinear functional forms, perhaps involving interaction variables, must be used to produce reasonably robust results.[35] Furthermore, when one product works as

[35]Curry et al. (2001) summarize some of the advantages of flexible functional forms (and even neural networks) in the context of hedonic modeling applied to consumer goods. The authors use detailed scanner data to estimate and test hedonic models with interaction effects for the U.K. television market.

a complement with another (e.g., hardware and software), it is conceptually unclear how to quality adjust each in isolation. Nor does existing theory say much about the importance, or lack thereof, of explanatory power. It is hard to know when a hedonic function is good enough for CPI work: the absence of coefficients with the "wrong" sign may be necessary, but it is surely not sufficient.

When product and process innovations occur, tastes change, or input prices shift, hedonic surfaces may change rapidly. The ability of the BLS or any other agency to capture those changes in real time is, at best, doubtful. It is unclear whether usable estimates of hedonic surfaces can be routinely and rapidly computed for a wide variety of goods. For many goods, the relationship between characteristics and observed price may not be stable, and the best-fitting functional approximation may change across products or time, particularly when technological change is rapid. Research into the stability of coefficients for different product groups is essential for making informed decisions about how often to reestimate hedonic functions. Without this information, reestimation schedules may be dictated by budget or other factors, which might result in outdated adjustments and be worse than doing nothing.

If the hedonic functions were known in every period, some variant of the direct characteristics method would be the best way to derive price ratios. Sometimes this would reduce to the direct time dummy method, but there is no reason to think this would occur frequently. Since the time dummy method has similar data requirements as the direct characteristics method but rests on much stronger assumptions that lack theoretical support—most notably, stable marginal impact of characteristics on price over time—the former has little to recommend it in principle. The time dummy method seems particularly unsuited to index use in rapidly changing product areas for which, presumably, quality adjustment is most warranted.

> **Recommendation 4-4: BLS should not allocate resources to the direct time dummy method (unless work on other hedonic methods generates empirical evidence that characteristic parameter stability exists for some products).**

The biggest obstacle inhibiting use of the direct characteristics approach is that the data and analysis requirements are daunting. However, the payoff from using this approach could be substantial. The methodology can, in principle, produce quality-adjusted indexes that take into account changing marginal relationships between characteristics, weighted by expenditure shares, and price. And, relative to the indirect method that adjusts an observed price change on the basis of individual coefficients, directly produced hedonic indexes are based on the entire hedonic surface which, in theory, generates more robust and precise estimates over different specifications.

Recommendation 4-5: BLS should experiment with the direct characteristics method, beginning with a few, carefully selected goods. The timely availability of relevant data should be a key selection criterion.

Though its statistical properties require more in-depth study, the indirect method seems at this time the most broadly applicable hedonic approach for use in the CPI.

Recommendation 4-6: BLS should continue to study the value of the indirect method for a wide range of goods.

A large part of its promise rests on the comparatively modest maintenance and data requirements relative to the direct methods. Different adjustment methods imply different updating intensities. Under direct methods, the hedonic regression is part of index construction and must be rerun each time the index is recalculated. Thus, data on prices, characteristics, and purchase shares of a large set of varieties are required in each period and such data must be in hand for the current period before the hedonic function can be estimated and the index computed.

Under the indirect approach, results can be obtained with only periodic reestimation. Only past period estimates are required, so there is less time pressure on data collection and analysis. However, when the relationship between characteristics and price moves quickly, even the indirectly used hedonic functions must be reestimated and, when characteristic sets change, they have to be respecified if they are to remain accurate. Ideally, regression equations would be updated every month. Practical considerations all but eliminate this possibility; BLS is not equipped or adequately funded to do this on a large scale. Rerunning current models with new data may not be overly burdensome, but respecifying models is highly labor intensive. Given that data collection and model estimation requirements may impose more than a 1-month lag in many cases, it may be necessary to figure out how best to use an estimated surface based on 6-month-old data to compute hedonic functions for the most recent monthly index. There are also basic questions regarding which price data to use when estimating hedonic models. For instance, should data be collected on transaction prices or list prices? Although transaction prices seem preferable due to seasonal selling patterns, BLS has used regular list prices in hedonic modeling of apparel—the idea being to avoid looking at different points in a product's life price cycle.

The long list of unresolved issues discussed in this chapter explains why even some proponents of hedonics advocate a less aggressive expansion of its use in the CPI than BLS appears to be taking. It is important that the BLS position on hedonics be shaped by scientific corroboration of the validity of broadly applying the method across index items and not be adopted as the default method to correct for quality bias in an attempt to move the CPI closer toward a COLI ideal. There is certainly no guarantee that hedonic methods always improve accuracy relative to alternative approaches.

The data and specification problems discussed in this chapter are serious, and we believe that the value of hedonic methods, and of alternatives, must be determined over time on an item-by-item basis. This represents a major undertaking.

Recommendation 4-7: Congress should continue to provide the BLS incremental resources to permit it to conduct in-depth and systematic analysis of quality changes across a broad range of goods and services covered by the CPI.

In designing its hedonics research program, BLS should seek to develop tools for dealing with the data and specification problems discussed above. Extending the CPI improvement initiative will allow BLS to continue its experimental research into scanner data; to assess the impact of hedonics on item comparability decisions and on index performance; and to investigate the replicability of competing techniques, perhaps using outside researchers to review and attempt to reproduce BLS results.

Recommendation 4-8: An independent advisory panel, consisting of econometricians, statisticians, index experts, marketing specialists, and possibly product engineers, should be formed to provide guidance on both conceptual and application issues pertaining to hedonic methods.

BLS, working with the advisory panel, should assess the impact of modeling imperfections on the validity of its hedonic adjustments prior to their introduction into the index. This would provide an analytic basis for proceeding sensibly in the face of external pressures to ameliorate the perception that the CPI fails to capture improvements in rapidly evolving sectors and to proceed quickly in this area simply because it is viewed as the only option available. In addition to attempting to advance understanding of the econometric methodology underlying the estimation of hedonic functions, the proposed advisory panel should provide outside review to help guide decisions about potential new applications and about which BLS pilot studies are adequately developed to be incorporated into the index. The hedonic results should always be evaluated against BLS's currently used alternatives (generally those associated with implicit quality adjustment techniques), as opposed to some idealized flawless solution.

To improve its effectiveness, the proposed advisory panel might be charged with helping to promote a major academic research effort to address issues (like the validity of using brand-specific dummy variables in the regressions) that are suspect but are not currently being discussed in the literature. The initiative should aim to increase collaboration between BLS and outside researchers on both theoretical work and practical construction issues. The tendency to emphasize what can be most easily measured, rather than to focus on learning what characteristics are important to consumers, should be resisted. No research program can identify a universal set of criteria against which the BLS can validate its econometric procedures—there will always be a role for detailed case-by-case study. But precisely because so much judgment and knowledge of the product is involved, it makes sense to have outside review before new hedonic applications are brought into the CPI.

TECHNICAL NOTE 1: BOSKIN COMMISSION ESTIMATES OF QUALITY CHANGE AND NEW GOODS BIAS

In this note we briefly review the items, grouped into upper-level categories, that the Boskin commission identified as contributing significantly to its overall CPI bias estimate. We also make note of criticisms of commission methods by Moulton and Moses (1997) to illustrate the lack of consensus that exists regarding the magnitude of quality change and new goods biases—particularly at the level of disaggregated CPI component indexes.

Food and Beverages The estimated bias associated with CPI pricing of fresh fruit and vegetables was the largest among components of the food and beverages category and was attributed by the commission primarily to the value to consumers of increased seasonal availability and variety. Limited by the dearth of published evidence on items in the food category, the commission was forced to lean heavily on Hausman's (1997) work that calculated consumer surplus for a new variety of breakfast cereal as a means to quantitatively estimate the value consumers place on increased product variety. Citing data showing increased total consumption of products within the category, which they linked to increased variety and convenience, the commission arrived at an annual bias estimate of 0.6 percent for fresh fruit and vegetables. Moulton and Moses (1997) challenged this figure, showing that most of the increase in consumption over the period 1972-1995 occurred after 1985, while most of the increase in availability occurred before 1985: "Part of the increase appears to have been driven by shifts in preferences, perhaps as a response to improved knowledge about the health benefits of fresh vegetables" (Moulton and Moses, 1997:314).

Shelter The Boskin commission produced detailed back-of-the-envelope calculations, based on assumptions about rental unit quality and size, to estimate a 0.25 percent annual bias for the shelter cost index. The commission's position that CPI quality adjustments have been inadequate for shelter was deduced from the premise that newer apartments have increased significantly in quality (as reflected by improved amenities, such as central air conditioning) and in size (a dimension of quality). They interpreted housing survey data as indicating that apartments increased in size by 20 percent between 1976 and 1993. Moulton and Moses (1997) countered, arguing that (1) rents generally do not increase proportionately with apartment size and (2) more importantly, that careful examination of data from the American Housing Survey and elsewhere suggests that the Boskin commission overstated historical increases in apartment sizes by perhaps a factor of three.

Appliances and Electronics The commission's bias estimates for this category are the largest—3.6 percent per year for the period 1973-1994 and 5.6 percent per year for 1994-1996. Due to the identifiable and quantifiable nature of appliance characteristics, and probably also to a priori notions about advances in the sector, research into this category of consumer spending is more extensive than for any other. Thus, the commission was able to access direct evidence, and the overall category estimate was extrapolated from items for which studies have been produced. The body of evidence included research by commission member Gordon (1990, cited in Boskin et al., 1996) of model-by-model comparisons from *Consumer Reports*. Moulton and Moses acknowledge that bias estimates for this category were probably the best documented by the Boskin commission: the report cites a number of academic and government studies that "develop hedonic adjustment models and find upward bias for personal computers, television, video equipment, and other items in this category" (Moulton and Moses, 1997:317).

Apparel The Boskin commission used a "conservative reestimation" of figures from Gordon's Sears catalog index, which rose less rapidly than the CPI subindex, to arrive at a 1 percent annual bias for the category. The main shortcoming of the experiment, according to Moulton and Moses (1997), is that Gordon measured year-to-year price changes only for the subset of apparel items that remained identical. The methodology links out—or deletes—the price increases associated with new product lines; the entire observed price change is assumed to reflect quality change. This approach produces misleading estimates if manufacturers are most likely to hike prices when new lines and varieties are introduced, as suggested by BLS studies. Also, apparel prices are known to be affected by lower-level substitution bias because of cross-outlet and seasonal volatility that allows consumers to find similar items at very different prices, depending on the store and on shopping times. Because methods to minimize substitution bias have been applied by BLS to apparel items, Moulton and Moses (1997:318) note that "it is unclear whether the Advisory Commission avoided double counting when sorting through these various sources of bias to produce its estimate of quality bias."

Transportation (New and Used Vehicles/Motor Fuel) On the basis of studies showing increased quality and increased service lifetime, the Boskin commission estimated an annual bias of 0.59 percent for automobiles. The estimate was based on back-of-the-envelope calculations on the effect of increased longevity and, in turn, reduced depreciation rates, on annual operation costs. Triplett (1997), as well as Moulton and Moses, argues that the commission did not have accurate information about measures that BLS has implemented to take into account improved automobile quality. The Boskin commission also estimated a 0.25 percent annual bias associated with CPI pricing of motor fuel, which was attributed to failure of the CPI to capture convenience and time savings associated with automatic credit card readers at gas stations. Moulton and Moses offer their own back-of-the-envelope calculations, based on assumptions about the value of consumers' time, time savings created by the machines, and average purchase size and find a bias about half as large.

Medical Care The Boskin commission's estimate of bias in the medical services index, 3.0 percent for both professional medical services and hospital and related services, is imputed largely from two empirical studies—Shapiro and Wilcox (1996) on treatment of cataracts and Cutler et al. (1996) on treatment for heart attacks. Thus, though Moulton and Moses agree that there is upward bias in the medical index, the validity of the commission's estimate ultimately depends not only on the accuracy of these specific results but also on the extent to which the studied services are representative of the sector. Work by Berndt et al. (1996) and Griliches and Cockburn (1996) for prescription pharmaceuticals—for which the Boskin commission estimated a 2.0 percent per year bias—led BLS, in 1995, to change its method of pricing prescription drugs when generic versions become available. Also, beginning in January 1997, BLS adopted the PPI (Producer Price Index) method of pricing treatment-based bundles of hospital services. Both these measures reduced biases associated with measurement of medical service categories, although it likely did not eliminate them.

Other Goods and Services The estimated biases associated with items other than those noted above were generally minor in terms of their impact on the all-item CPI. The Boskin commission suggested a 2.0 percent bias in sporting equipment and toys; small appliances such as hair dryers were assigned the same bias as large appliances, 3.0 percent per year. Personal financial services, a category for which output is extremely difficult to measure and rapid technological change (e.g., proliferation of ATMs and on-line account management) has occurred, the commission "conservatively" estimated an annual bias of 2.0 percent. The commission also discussed cellular phones but, as Moulton and Moses (1997:321) point out, it is not completely clear whether or not they included this in their estimated 1.0 percent bias for the "other utilities, including telephone" category.

TECHNICAL NOTE 2: MATHEMATICAL DESCRIPTION OF HEDONIC METHODS

In the index number context, the hedonic function $p_{i,t} = h_t(z_i)$ for a product with multiple varieties—where $p_{i,t}$ is the price of the *i*th variety in period t and z_i is a vector of the *i*th variety's characteristics or attributes—plays the same conceptual role as the (scalar) price plays for an undifferentiated good. In the present context, the hedonic function can be viewed as a menu from which individual consumers make choices.

A typical hedonic specification for econometric estimation uses the natural logarithm of an item's price as the dependent variable and several characteristics as the explanatory variables. The model may contain discrete variables, indicating whether or not a model has a feature, such as a CD drive on a computer, as well as continuous variables, such as the thread count of a fabric. Control variables, such as purchase location or outlet type, may also be included. When, as is typically the case, the explanatory variables are included linearly (rather than, say, logarithmically), the coefficients can be interpreted as giving proportional changes in price associated with a one-unit change in the quality characteristic or from a switch in the dichotomous variable. If explanatory variables enter nonlinearly, these proportional changes depend on the values of the explanatory variables.

There is a large theoretical literature on the properties of observed hedonic functions (see, e.g., Rosen, 1974; Muelbauer, 1974; Feenstra, 1995; Barry et al., 1995; Diewert, 2001). Much of this literature is concerned with the extent to which h_t provides information on producers' costs and consumers' preferences under various assumptions about the nature of competition. This is not our concern here: in general, hedonic functions are reduced-form reflections of details of tastes, technologies, endowments, and strategic behavior in differentiated product markets. In particular, when competition is imperfect, it is generally not possible to infer marginal costs from the observed hedonic functions. We follow most of the theoretical literature and assume what Pollak (1983) calls "Houthakker's 'heterogeneous' or 'H-characteristics'" approach, which fits products for which consumers purchase one and only one variety. (The alternative, "Lancaster's 'linear and additive' or 'L-characteristics'" approach, applies when consumers purchase multiple varieties and care about the total amount of each characteristic supplied by all.)

The use of hedonics in the index number context rests on being able to interpret the h_t functions as summarizing the menu of alternatives faced by consumers in period *t*. This raises the general problem that different consumers in fact face different prices and have different stocks of information about their alternatives. Moreover, when price is not linear in the values of characteristics about which consumers care (see Muelbauer, 1974, for some relevant theory), which most hedonic studies seem to find, it follows that, even if h_t is a smooth function, the marginal cost to consumers of any particular characteristic varies with z. Thus, consumers who choose different varieties of some product because of differences in incomes or tastes (or both) face different "characteristics prices" at the margin, and the "characteristics prices" faced by nonbuyers are clearly not well defined.

Finally, most uses of hedonics involve using an estimate of h_t to, in effect, predict the price that would have prevailed in period t for a variety or model not actually offered for sale in that period. While this seems sensible, it is problematic at the theoretical level: under imperfect competition, if an additional variety or model had actually been offered for sale, the prices of other products might also have changed. In addition, smoothness and functional form assumptions are important in these exercises, and, particularly when consumers are heterogeneous, theory provides relatively little guidance regarding such assumptions (see Diewert, 2001, for a useful discussion).

The Indirect Method

As discussed in the body of the chapter, the indirect method is used to handle situations in which one variety of a good tracked in the CPI system—with a specific vector of characteristics z_1 and price $p_{1,t}$, say—disappears after period t and is replaced by another—with characteristics z_2 and price $p_{2,t+1}$, beginning in period $t + 1$. There are two basic types of indirect methods. If the hedonic function, $h_t(z)$, for period t is available, the simplest form of the *forward-looking indirect method* involves using $p_{2,t+1}/h_t(z_2)$ as the estimated "pure" price relative. The denominator of this ratio is an estimate of what a good with characteristics z_2 would have cost if it had been available in period t, based on the empirical relation between price and characteristics in that period. If the hedonic function, $h_{t+1}(z)$, for period $t + 1$ is available, the simplest *backward-looking indirect method* involves using $h_{t+1}(z_1)/p_{1t}$ as the estimated price relative. Because it uses a bundle (of characteristics) purchased in period $t + 1$, the forward-looking method is Paasche-like; similarly, the backward-looking method is Laspeyres-like.

Hedonic functions are typically refit only periodically, so neither the current period nor the prior period function is usually available. Thus, the backward-looking method is rarely feasible. To see how this affects the calculations under the forward-looking method, suppose the hedonic function was last estimated in period 0, with $h_0(z)$ the estimated function, and suppose one wants to calculate the "pure" price relative between periods t and $t + 1$. Clearly, $p_{2,t+1}/h_0(z_2)$ is a forward-looking estimate of the price relative between periods 0 and $t + 1$ for the bundle z_2, while $p_{1,t}/h_0(z_1)$ gives a similar estimate of the price relative between periods 0 and t for bundle z_1. If z_1 and z_2 were the same bundle, the ratio of these quantities

$$R_{t,t+1} = [p_{2,t+1} / p_{1,t}][h_0(z_1) / h_0(z_2)], \tag{1}$$

would give the price relative between periods 1 and 2 for that bundle. Since product 2 is being treated as a replacement for product 1, z_1 and z_2 must be close in some relevant sense. In any case, the BLS proceeds as if they were equal and employs $R_{t,t+1}$ as the price relative.

Another way to look at (1) is that the actual price of product 2 in period $t + 1$ is being compared with the adjusted price of product 1 in period t—adjusted for the quality difference between products 1 and 2 using $h_0(z)$:

$$\text{adjusted } \mathrm{p}_{1,t} = p_{1,t}[h_0(z_2) / h_0(z_1)]. \tag{2}$$

It is easy to show that these methods automatically take into account some forms of unobservable outlet-specific price differences that, along with other factors, prevent hedonic functions from fitting perfectly. Suppose, for instance, that $h_0(z)$ is the estimated marketwide base period hedonic function, as above, but prices of all varieties in some particular outlet exceed marketwide averages by a constant multiple θ. Then $p_2/\theta h_0(z_2)$ is the natural estimate of the price relative between periods 0 and 2 for the bundle z_2, while $p_1/\theta h_0(z_1)$ is the natural estimate of the price relative between periods 0 and 1 for bundle z_1. Neither of these is observable if θ is unknown, but their ratio, which is the quantity of interest, is given simply by equation 1, above.

The Direct Time Dummy Method

This method involves estimating hedonic functions of the following form:

$$\log(p_{i,t}) = h(z_{i,t}) + \sum_{\tau=1}^{T} \beta_\tau \delta(t,\tau), \text{ for } t = 0, \ . \ . \ . \ , T; \text{ all } i, \tag{3}$$

where the subscript i denotes varieties or models and, as above, the β_τ are constants, and $\delta(t,\tau)$ equals 1, if $t = \tau$ and 0 otherwise. Note that there is no time dummy for period 0, the base period; we have arbitrarily normalized at $\beta_0 = 0$ to identify the rest of the model.

Specification (2) implies that in any period t the ratio of the prices of models with, say, characteristics bundles z_1 and z_2, $p_{1,t}/p_{2,t}$, is equal to antilog$[h(z_1) - h(z_2)]$, which does not depend on time. It is thus being assumed that the prices of all (actual and potential) varieties change proportionately over time. (In light of Zvi Griliches's seminal contributions to the theory and practice of hedonic methods, the panel believes it would be appropriate to label this the case of *Griliches neutrality*.) Neither theory nor empirical research provides much support for this assumption, however, particularly in industries experiencing rapid technological change.

If prices of all varieties *do* change proportionally, though, it is simple to use the function above to produce a "pure" price relative for the product under study. For any variety i with a constant characteristic vector z_i, the equation above immediately implies that for any two time periods t and u

$$p_{it} / p_{iu} = \text{antilog}[\log(p_{it}) - \log(p_{iu})] = \text{antilog}[\beta_t - \beta_u], \tag{4}$$

for all i. Thus the expression on the right gives the price relative between periods u and t for the product under study.

The Direct Characteristics Method

Let $h_t(z)$ be the hedonic function in period t, C_t be the set of varieties available—with characteristic vectors $z_{i,t}$, (average) prices $p_{i,t}$, and quantities sold $q_{i,t}$. The direct characteristics method computes price relatives using these data without necessarily imposing the assumption (which underlies the time dummy method) that ratios of hedonic functions are independent of the point in characteristics space at which they are evaluated. If some coefficients of the hedonic function are constant over time, of course, estimation efficiency can be improved by imposing constancy and using data from multiple periods in estimation. Alternatively, if the assumption that all slope coefficients are stable over time (i.e., the assumption of Griliches neutrality that underlies the time dummy method) is rejected by statistical test, some use of some version of the direct characteristics method would seem to be in order.

As noted in the text, the natural way to use the hedonic functions to compute a single price relative in, say, periods 1 and 2, with different sets of products available in each, is to use the hedonic functions to price constant bundles of characteristics over time. The literature suggests two ways of doing this. The first follows Diewert (2001) and uses the average bundles consumed as reference characteristics vectors:

$$z_t^* = \Sigma_i q_{it} z_{it} / \Sigma_i q_{it}, \; t = 1, \; 2. \tag{5}$$

Then Laspeyres-, Paasche-, and Fisher-type indexes, which give alternative measures of price relative between periods 1 and 2, can be defined, respectively, as follows:

$$L_{12} = h_2(z_1^*) / h_1(z_1^*), \tag{6a}$$

$$P_{12} = h_2(z_2^*) / h_1(z_2^*), \tag{6b}$$

$$F_{12} = [L_{12} P_{12}]^{1/2}. \tag{6c}$$

Note that (6a) requires only lagged quantity weights, while both (6b) and (6c) require current quantity data. Note also that all these measures are equal, and all equal the results of the time dummy method, if the ratio $h_2(z)/h_1(z)$ is independent of z.

The second approach follows Feenstra (1995), with some modifications by Diewert (2001). Let C^* be the set of varieties available in both periods, and let C_t' be the set of varieties that are available only in period t. One can use the period 1 hedonic function to "predict" the period 1 prices of those varieties available only in period 2, and one can use the period 2 hedonic function similarly:

$$p_{i1}{}' = h_1(z_{i2}), \qquad z_{i2} \in C_2{}', \tag{7a}$$

$$p_{i2}{}' = h_2(z_{i1}), \qquad z_{i1} \in C_1{}'. \tag{7b}$$

One can then compute a Laspeyres-like measure by taking a weighted average, using period 1 sales shares as weights, of the (actual and "predicted") price ratios of the varieties available in period 1:

$$\begin{aligned} L_{12} &= \Sigma_{C*}\, w_i(p_{i2}/p_{i1}) + \Sigma_{C1'}\, w_i(p_{i2}{}'/p_{i1}) = \\ &[\Sigma_{C*}\, q_{i1}p_{i2} + \Sigma_{C1'}\, q_{i1}p_{i2}{}']/\Sigma_{C1}\, q_{i1}p_{i1}, \end{aligned} \tag{8a}$$

where, as usual, $w_i = q_{i1}p_{i1}/\Sigma_{C1}\, q_{i1}p_{i1}$.

Similarly, using period 2 sales shares of the various varieties as weights and "predicting" the period 1 prices of varieties available only in period 2 yields a Paasche-like measure:

$$\begin{aligned} P_{12} &= \Sigma_{C*}\, w_i(p_{i2}/p_{i1}) + \Sigma_{C2'}\, w_i(p_{i2}/p_{i1}{}') = \\ &[\Sigma_{C2}\, q_{i2}p_{i2}]/[\Sigma_{C*}\, q_{i2}p_{i1} + \Sigma_{C2'}\, q_{i2}p_{i1}{}'], \end{aligned} \tag{8b}$$

where $w_i = q_{i2}p_{i1}/[\Sigma_{C*}\, q_{i2}p_{i1} + \Sigma_{C2'}\, q_{i2}p_{i1}{}']$ for $z_{i2} \in C^*$, and $w_i = q_{i2}p_{i1}{}'/[S_{C*}\, q_{i2}p_{i1} + S_{C2'}\, q_{i2}p_{i1}{}']$ for z_{i2}Œ$C_2{}'$. One can combine these, as in (6c), to obtain a Fisher-like measure of the price relative for this product. Note again that if price ratios for all varieties are the same, as assumed by the time dummy method, all of these measures are equal.

To see the sense in which these two approaches give Laspeyres-like and Paasche-like measures, it is instructive to follow Pakes (2001) and consider a single consumer with income y in periods 1 and 2, with prices the same in both periods for all goods but widgets. In period 1, the consumer has available a set of varieties C_1, the prices of which are given by the known hedonic function $h_1(z)$, and she purchases variety z_1. In period 2, the consumer faces choice set C_2 and known hedonic function $h_2(z)$, and she chooses variety z_2.

Suppose this consumer is given $h_2(z_1) - h_1(z_1)$ additional income in period 2. Is this greater or less than the *compensating variation*, the period 2 income increase that would leave her exactly as well off as in period 1? If $z_1 \in C_2$, buying variety z_1 in period 2 would leave her with $y + h_2(z_1) - h_1(z_1) - h_2(z_1) = y - h_1(z_1)$ to spend on other goods, exactly as in period 1. So $h_2(z_1) - h_1(z_1)$ is at least equal to the compensating variation. But because the two hedonic functions are different, it may be possible for the consumer to do even better by choosing some $z_2{}'$ π z_1 in C_2. Thus $h_2(z_1) - h_1(z_1)$ is greater than or equal to the compensating variation, depending on whether such a $z_1{}'$ exists or not.

Similarly, suppose instead that the consumer's period 1 income is reduced by $h_2(x_2) - h_1(x_2)$. Is this greater or less than the *equivalent variation*, the period 1 income reduction that would leave her exactly as well off as in period 2? If $z_2 \in C_1$, buying variety z_2 in period 1 would leave her with $y + h_1(z_2) - h_2(z_2) - h_1(z_2) = y - h_2(z_2)$ to spend on other goods, exactly as in period 1. So this income reduction will leave the consumer no worse off. She will be better off in the (new) first period if she can afford some $z_1' \in C_1$ that she prefers to z_2. Thus $h_2(z_2) - h_1(z_2)$ is less than or equal to the equivalent variation, depending on whether such a z_1' exists or not.

For our single consumer, the price relative could naturally be computed as either $h_2(z_1)/h_1(z_1)$ or $h_2(z_2)/h_1(z_2)$. The former is a Laspeyres approach and, as above, relates to the compensating variation. The latter is a Paasche approach and relates to the equivalent variation. In the usual sense, and with all the usual caveats plus the requirement that z_1 and z_2 be available in both periods, in this simple case these two measures bound the true, preference-dependent, change in the cost of living.

5

New Goods and New Outlets

This chapter is essentially a continuation of the quality change discussion. In the first half of the chapter we consider the case of new goods that do not fall into existing Consumer Price Index (CPI) item categories. This case presents problems of estimating newly created value, sample rotation frequency, item reclassification, and weight updating. The second half of the chapter addresses the indexing problem that arises when consumer shopping patterns are shifting. The panel considers what, if anything, the Bureau of Labor Statistics (BLS) could do to identify and estimate quality and price components of observed differences in the prices of goods across outlets.

NEW GOODS

The term "new good" is not precise. Routine price collection procedures continually lead to instances in which BLS has to find replacements for items that have disappeared. Similarly, when BLS rotates its sample of retail outlets, it picks up products that are different from those it had been pricing in the old stores.[1] Products also appear that are novel to the point that there is no place in the CPI

[1]Unlike the process in which items are substituted in stores, outlet rotation can introduce "supplemental" goods, such as a new cereal or a generic drug, which do not replace any particular item but do fit into an existing entry-level item (ELI) category. A process also exists whereby supplemental goods can be brought into the index by "directed reinitiation," in which BLS responds (outside of normal rotation) to changing market conditions and subjectively redirects a portion of the sample to cover a new product (Armknecht et al., 1997:377).

item structure to accommodate them: cell phones, home computers, and VCRs are examples.[2] These are products whose characteristics would be difficult to "repackage" (in the sense discussed in Chapter 4) into existing goods and services no matter how broadly definitions are drawn. Without an explicit decision to change the list of goods to be priced, standard indexing procedures will not pick up any of the effect of such newly introduced items on consumers' living standards or costs.[3]

These contrasts notwithstanding, no sharp dividing line separates a new good from a quality improved product. What can be cleanly distinguished are situations that lead to within-sample item replacement and those involving a good or service that has entered the market but would never be brought into the index as part of the in-store pricing process. New items falling into this second category include (1) those that might be picked up during sample rotation (in which case items enter using overlap pricing, where there is no comparison to a previously priced good and, hence, no quality adjustment) and (2) those that can only be brought into the index when item strata are redefined and the sample reset. The previous chapter primarily addressed shortcomings in the process for dealing with quality adjustment of replacement items. However, failure to capture price (or cost-of-living) effects associated with new nonreplacement products may, depending on the objective of the index, cause what Triplett (2001b) has termed "new introduction bias." This failure is not a quality adjustment problem but a sampling one—a case in which rapid product turnover, caused by technological or other changes, leaves the item sample no longer representative of what people buy (Triplett, 2001b:19-20).

The appearance of products that can only enter the index after item reclassification (and, to some extent, those that enter during sample rotation) raises two issues beyond those associated with routine item replacement. The first is what to do to account for price effects that occur during the period in which a new product appears on the market. Specifically, should a price or cost-of-living index reflect the fact that new goods typically enter the market at a price that is below that which would have reduced demand in the period prior to its introduction to

---

[2]There are gradations of incompatibility with the CPI item structure. Some new products might fit into an existing item strata but not into any of the more specific ELI definitions. In this case, a new ELI can be created, and the new product brought in gradually through sample rotation. Other products are so different that they can only be incorporated into the index by revising the CPI item classification structure.

[3]Recent technological innovation has introduced some goods that are, relative to the VCR-type examples, even more difficult to assess. For instance, e-mail has certainly affected people's communication behavior, but it is hard to place a value on it or ascertain its price (even for a single household). What percentage, if any, of Internet access and provider fees, or even of a computer purchase, would be assigned to the "price" of e-mail.

zero or, conversely, for an item that has disappeared, the increase in price that would have driven quantity demanded to zero? Failure to capture this "price reduction" could be argued to cause a new good introduction bias in the index. The second issue is what to do in subsequent periods after the appearance of the new good. Specifically, how and when should these items be brought into the market basket tracked by the index?

Valuation at the Point of Introduction

As new products penetrate the market, the item coverage of a fixed-basket index becomes less and less representative of the things that people are currently buying. This is why various techniques of "unfixing" the market basket, including item replacement and sample rotation, are now regular features of the CPI. However, even with the modifications that these techniques allow, potential indexing problems remain. Identifying one of these problems, Hausman (1997: 209) argues that welfare effects associated with the introduction of new goods should be estimated and used to adjust the CPI:

> The CPI serves as an approximation of an ideal cost-of-living (COL) index. In turn, the COL index answers the question of how much more (or less) income a consumer requires to be as well-off in period 1 as in period 0 given changes in prices, changes in the quality of goods, and the introduction of new goods. . . . The CPI as currently estimated by the Bureau of Labor Statistics (BLS) does a reasonable job of accounting for price changes and has begun to attempt to include quality changes. However, the BLS has not attempted to estimate the effect of the introduction of new goods, despite the recognition of the potential importance of new goods on both a COL index and the CPI.

In this subsection we briefly review the mechanics of how, in theory, price indexes could take into account this new goods effect. We then consider a counterargument to Hausman's (and, implicitly, the Boskin commission's) recommendation for BLS to do so.

The relevant difference between a new good and an established one is that, for the former, the price in previous periods associated with the realized sales levels (zero) cannot be observed, while the price for the established good can be. In theory, a virtual price exists in each prior period that would have been just high enough to drive the quantity demanded of the new good to zero.[4] There are consumers who would have purchased the good at various prices between that virtual price and the lower price at which the good sells when it appears in

[4]The term "virtual price" is synonymous with the terms "choke price" and "reservation price," often used in the literature.

markets.[5] The effective decrease from the virtual to the introductory price of a new good is not captured in the CPI, even in instances when new goods are brought into the market basket very rapidly. The introduction of a new good, and its later diffusion to its ultimate customer base as more consumers learn about it, may be thought of as a series of price reductions. A demand curve that traces the "virtual" prices that some consumers would have been willing to pay for the good can, in theory, be econometrically estimated.[6]

If significant numbers of new goods are continually invented and successfully marketed, an upward bias will be imparted to the overall price index, relative to an unqualified COLI (though this effect may be partially offset by a downward bias created by the disappearance of goods). There is a component of this bias that can occur even if new goods are linked into the index quickly and expenditure weights are updated frequently. A priori, one might expect that only new goods that provide radically improved capabilities or that are sold at reduced prices relative to predecessors would capture market share quickly enough to generate significant point-of-introduction bias. After all, if the new good offers only minor new capabilities relative to existing goods, the virtual price driving

[5]The relationship between price and market sales is not only a matter of different consumers being willing to pay different prices; it also involves, for many goods, the quantity of the good that any one consumer might purchase, depending on the price—e.g., the initial introduction of varieties of fresh vegetables during the winter, made possible by improvements in the speed and efficiency of transportation networks, is an example. Also, consumers will not all learn about a new product at the same moment in time, which means that the virtual prices should be constructed not only for the period just prior to the introduction of the new commodity into the local marketplace but also for subsequent periods, as more consumers learn about the new product.

[6]The conceptual basis for estimating how virtual demand for new commodities could be introduced into a price index is attributable to Hicks (1940). The idea requires first assuming that a single consumer's preferences over new and old goods available to the consumer in period $t + 1$ also apply in period t and earlier. This allows one to look for the lowest price for the new commodity in period t that would cause the consumer to demand zero units in period t: this is the virtual (or shadow) price. With this new price and quantity information for period t, one can proceed to construct a new fixed-weight index.

The new Laspeyres index with period t as the base turns out to equal the initial fixed-base Laspeyres index that ignored the existence of the new commodity, since the new commodity has a zero weight in that period. A Laspeyres index with period $t + 1$ as the base period, however, would show a period $t + 1$ gain from the introduction of the new commodity, since the relevant shadow price for the new commodity must be used in periods t and earlier, and the new product has a positive weight in period $t + 1$. This qualification is critical. Thus, every new product that enters the index requires restating all past values of the index.

A new Paasche index would also now be different from the initial Paasche index. Under the condition that the new good turns out to be a success, the price ratio for the new good will generally be lower than the price ratio for old goods, so the new commodity can steal market share from the old commodities. Thus a new Paasche index will generally show a lower rate of price increase than the old Paasche index.

consumer demand for it to zero would not be far above that for close substitutes.[7] Conceivably, BLS could attempt to identify and estimate demand curves for radically different goods experiencing rapidly growing consumer acceptance. The price or cost-of-living index would include weighted estimates of the difference between the virtual price (in period $t - 1$) and the introductory price (in period t). To be fully consistent, it would also be necessary to identify goods that were forced off the market by newly introduced competitors, the unavailability of which created virtual price increases.

Hausman (1997) argued that it is important to consider new brand introduction in the calculation of economic welfare and consumer price indexes and then proceeded to estimate the demand curve for a new variety of breakfast cereal—Apple-Cinnamon Cheerios—to illustrate that it could be done. His article concludes that introduction of the new cereal variety—hardly a radical expansion of consumer opportunities—created substantial consumer benefits (defined by the difference between the virtual price for the cereal minus the introductory price), sufficient on its own to have reduced the average rate of price increase in the CPI component for breakfast cereals significantly (Hausman, 1997:229, 234):

> . . . to the extent that about 25 percent of cereal demand was from new brands over the past ten years, and under the (perhaps unrealistic) assumptions that the new brands sell for about the same average price as existing brands and that the estimate here would generalize to a reservation price of about two times the actual price, the overall price index for cereals which excludes the effects of new brands would be too high by about the overall share of new brands—25 percent. . . . The introduction of imperfect competition would reduce the overstatement of the cereal CPI to about 20 percent.

If this kind of differentiation produces a significant number of products that enter the market priced well below that which would choke quantity demanded to zero, then our economy, marked as it is by increasing proliferation of product varieties, must be producing a substantial stream of new consumer welfare. This would strengthen the case for making price imputations to account for "price" effects attributable to the introduction of new commodities to the marketplace.

On the other hand, there are weaknesses in the case. Hicks (1940) defined market demand as it relates to new goods, and Hausman demonstrated that a choke price could be estimated for a specific new good. However, there is no clearly acceptable technique for consistently estimating demand curves for new goods or services in such a way that choke prices can be confidently ascertained. Several panel members are also, independent of estimation problems, hesitant

---

[7] In contrast, the existence of huge volumes of demand (and large amounts of unsatisfied demand) for a few hot brands of children's toys in the weeks before Christmas suggests that successful advertising and accompanying peer pressure can create temporary fads that, measured by the techniques outlined in the report, would presumably generate high consumer demand.

about the advisability of doing so on conceptual grounds (these concerns are discussed in the final section of this chapter).

The practical problem facing statistical agencies is exactly how the virtual price of a new commodity should be estimated. Many aspects of Hausman's analysis are highly controversial, even in the context of microeconomic research that is not directly tied to policy. Hausman's findings have been disputed on the grounds that questionable assumptions were built into his econometric specifications, which led to substantial overstatement of the prices that consumers would have paid for the new cereal brand. In a response to Hausman (1997), Bresnahan (1997) argued that Hausman's model requirements—specifically that there be no demand shocks that cause consumers to shift purchases and that shocks are not reflected in prices because they are unanticipated—are inappropriate. The assumptions required about functional form (the shape of the demand curve) and for system identification simply introduce too much uncertainty to be used as an input into a statistic that must be produced in a replicable fashion on a regular basis. Calculating the price that drives demand for a product to zero requires extrapolation outside the range of price and quantity observations. Hence, it must rely more heavily than conventional demand estimation on untestable assumptions about functional form. Bresnahan (1997:237, 246) concluded that the question of how important new goods are in terms of their contribution to social welfare remains unanswered.[8] Procedures for estimating virtual prices would require extensive refinement before they could even be considered for adoption into a national price or cost-of-living index.

Research into welfare and price effects associated with new goods is important and deserves attention, but it is unlikely that such a program will produce a consensus methodology in the near future. Given the level of uncertainty among economists about the accuracy and replicability of current econometric techniques for estimating virtual demand, it would be imprudent for BLS to attempt to adjust the CPI to account for increased welfare that occurs at the point when new products are introduced.[9]

Conclusion 5-1: Virtual price reductions associated with the introduction of new goods should not be imputed for use in the CPI.

Several members of the panel—particularly those advocating separate price and cost-of-living indexes—are unconvinced that adjusting the CPI to account

[8]It is also worth noting that wealthy consumers are likely to place a higher value on the introduction of (at least some classes of) new goods than poor consumers. Hence, incorporating virtual price reductions into a plutocratic index may have a greater effect than they would if incorporated into a democratic index. An index so adjusted could conceivably become less relevant for low- and middle-income consumers.

[9]As far as we know, BLS has no plans to do so.

for point-of-introduction bias would be a good idea even if the practical estimation problems could be solved. Proponents of more traditional price index methodologies argue that it is a perversion of the language to argue that the effect of, say, the introduction of cell phones or the birth control pill is to reduce the price level, a result that comes from confusing the concept of a price level with that of the cost of living. Their position is tempered somewhat by the realization that, outside of price measurement, there is nowhere else in the national accounts for such product quality improvements to be included and, as Nordhaus (1998) and others have argued, real growth in the economy is thereby understated. Additionally, modern economic growth appears to be more quality intensive than quantity intensive, and the statistical system is not keeping up with the change. However, the panel as a whole agrees that adjusting the CPI is not the way to correct the situation. Rather, research in this area should be directed toward developing a separate experimental COLI that is adjusted, to the extent possible, to account for changes as new products and technologies diffuse throughout the economy.

Criteria for Introducing New Goods

Under traditional procedures, a new (nonreplacement) good is linked into the CPI in such a way that its introduction, in and of itself, has no effect on the level of the index. Once in the index, price change of the new item affects index growth in the normal fashion. However, in addition to the point-of-introduction price reductions discussed above, price trends over the interim period between product appearance and introduction to the index also go uncaptured. Thus, a second problem—that of how quickly new goods are brought into the index—exists if early price-cycle patterns are consistently different from general price trends. If a new commodity is a reasonably close substitute for an existing one and is likely to replace it in the marketplace, then instead of explicitly revising the base market basket, one could think of simply replacing the old commodity in the index with the new commodity, after some adjustment for quality change. This is not very different from the within-sample replacement that occurs when an outlet sample item disappears.

For more novel introductions, a new commodity must be brought into the index as part of a revision to the market basket; that is, when the statistical agency switches from the old fixed-basket Laspeyres index to a new fixed-basket Laspeyres index that has a more recent period as its base and includes the commodity. There may still be a problem with use of the latter in comparison with a superlative index because of the properties of new product price cycles. A Laspeyres index that has period 1 as its base will weight the long-term price relative for the new good by its period 1 market share, which will often be much smaller than its period t market share for $t > 1$. Note that this period t market share appears for the period *t* Paasche index. Thus, under the above conditions, a Paasche index will often be considerably lower than its Laspeyres counterpart.[10]

Typically, at least in high-tech sectors, a new good does come to the market at a relatively high price and initially has a small volume of sales. High prices may reflect both production costs that have yet to be reduced by learning and process innovation and seller attempts to maximize profits by first selling to customers for whom the new product is especially valuable. Subsequently, prices often fall.

Armknecht et al. (1997) describe the price cycle observed after the introduction of VCRs to the market. Early in the product cycle, in 1978, approximately 400,000 units were sold at an average price of more than $1200. By 1987, when VCRs were introduced into the CPI, annual sales had increased to almost 12 million units, and the average price was less than $500. Over this period, prices decreased by 60 percent, non-quality adjusted, and, by the time of introduction to the index, sales accounted for about 0.1 percent of consumer expenditures (Armknecht et al., 1997:388). Dulberger (1993) showed how, for such products, more frequent replenishing of item samples can have a large effect on measured price change. Her analysis of semiconductors produced a chained Fisher index that decreased by 29.2 percent per year when new chips were introduced into the index with only a 1-year delay after appearing on the market. However, the index only decreased by 20.1 percent when there was a 3-year delay, and the index showed virtually no price decrease when the lag was 5 years (Dulberger, 1993:Tables 3.7, 3.8).

Observed price patterns such as these have led to charges, by the Boskin commission and others, that delayed introduction of new goods systematically omits product-cycle dynamics that impart an upward bias on price indexes. More frequent updating of the item classification structure and of the sample (which, in turn, would require more frequent index chaining) would have allowed a greater portion of these early product price trends to be captured and led to a more accurate plutocratic index. Each case, individually, would not have had a large cumulative effect on the overall CPI (for VCRs, well under one-tenth of 1 percent). However, in the modern economy, a large number of new goods are introduced each year, each having some effect. It is important to note, though, that not all goods follow this kind of pricing path during their life cycle. Pakes (1997), for instance, has stressed producer efforts to penetrate markets with low introductory

[10] As discussed in Chapter 2, the Paasche and Laspeyres indexes are both valid measures of price change between periods. If only one measure of overall price change is required, it can be argued that an average of these two indexes, such as the Fisher ideal index, which takes the geometric mean of the two indexes, offers a sensible approach. Under the conditions outlined, the Paasche and Fisher indexes will give a lower measure of price change. Thus, one is again led to a strong argument for the production of a superlative index in addition to the present real-time CPI.

prices.[11] Since pricing early in goods' life cycles may be atypical in different ways, the effects of the loss of information on pricing during an inevitable period of delay are, at least in principle, indeterminate.[12]

At a cost, criticisms by the Boskin commission and others could certainly be addressed. New goods could be introduced into the index earlier so as to catch a larger portion of postintroduction price trends; growing sales can be captured by more frequent weight updating. This panel agrees that, other things being equal, more information is better than less. Hence, new commodities should in principle be introduced into the CPI as soon as they become significant in the marketplace. Such an approach would require frequent sample rotation to capture new supplemental goods and more frequent revision of the item classification structure to capture radically different goods.

Unfortunately, other things would not be equal. Survey-based updates of expenditure weights and product samples are expensive to produce and require time to compile. When a new commodity is introduced into the base without new household expenditure information, other complications arise: Should all the old expenditures simply be scaled down proportionally to make room for the new expenditure share? Or should the expenditure shares of commodities that are apparently the closest substitutes for the new commodity be reduced somehow to make room for the new expenditure share? Again, this depends in part on how the new good enters the index.

As noted above, some items, such as VCRs, only enter during item reclassification or, like Viagra, from an ad hoc targeted initiation. However, the majority of supplemental and new items are identified in the Telephone Point of Purchase Survey (TPOPS, the 1998 revision of the POPS) and enter the CPI during sample

---

[11]Economic theory can illustrate cases for which "profit skimming" may make sense and others where penetration pricing does, Kalish (1988). Berndt et al. (1993) observed that the product life-cycle pricing path for prescription drugs typically shows relative declines for about 6 years but, after that point, prices significantly increase, such that the prices of older drugs increased more rapidly than the prices of newer drugs.

[12]The research literature does not offer much empirical evidence on new product price trends, specifically whether falling prices are characteristic of products in the early stages of their life cycle. Curry and Riesz examined *Consumer Reports* test studies that covered five or more brands during the period 1961 through 1980. They were able to gather data for 62 "product forms" covering 4,000 brands from 264 comparative products. The mean price for each product group was traced over time. They found, among other things, that 85 percent of the product prices exhibited negative trends. Tested at the 10 percent level, 56 percent of the products had significant negative trends and only 2 percent had positive trends. But (1) all of the products were goods and, at least from 1958 through 1980, goods, prices fell relative to service prices, and therefore relative to the CPI; and (2) the *Consumer Report* studies were not confined to "new" products. The authors concentrated on branded products and had a heavy preponderance of appliances and traditional electronics (radios, TVs, audio speakers, etc.) for which one might expect theory to apply more than, say, for new medical procedures.

rotation. New item and outlet samples are drawn annually for a subset of the 218 defined TPOPS item categories, so the entire sample turns over periodically (the previous procedure rotated geographic areas rather than item categories). Because of the importance of this process in getting new products into the CPI basket, the Conference Board (1999:25) recommended that BLS speed up sample rotation. Citing how quickly new products enter the marketplace in the modern economy, the group recommended an eventual 2-year rotation schedule for most categories and an annual rotation for categories subject to frequent change.

It is important to point out that the BLS has made significant strides in improving its survey structure to decrease time between outlet and item rotation. Part of the CPI Improvement Initiative was used to begin data collection procedures designed to incorporate new goods into the index more quickly. In addition to moving from an area-based to an item-based outlet rotation process, TPOPS has shortened the amount of time necessary to draw an outlet sample. Instead of one-fifth, one-fourth of outlet samples (and contemporaneously item samples) will soon be rotated each year, which decreases the amount of time needed for full rotation from 5 years to 4. Lane (2000:8) notes that, "by 2003, when the CPI has initiated an entire cycle of outlets based on TPOPS, the outlet samples will be significantly more current than they were before 1999." More frequent rotation might also be complemented to some degree by expanded use of targeted and directed replacement procedures with which BLS is currently experimenting. Targeted outlet item rotation would allow TPOPS categories associated with quickly changing markets to be rotated on a fast-track basis; targeted item rotation involves increasing, ad hoc, the probability of selecting specific items. BLS is also looking at methods for decoupling item and outlet rotation so that items could be rotated (within the current sample of outlets) without waiting for outlet rotation. Targeted replacement is suggested since outlet rotation is a particularly time-consuming and expensive aspect of CPI sampling. Rotation only at the item level may offer a way around this practical constraint, at least for items that enter the index through stores (or types of stores) already represented in the sample.

Sampling is done for 300-400 entry-level items (ELIs), like oranges, which are more finely specified examples of the 218-item strata, such as citrus fruits. Item strata correspond roughly to the TPOPS categories. If new items encountered in the sample rotation process fit existing ELI definitions, they can be readily brought into the CPI through overlap pricing, since new and old items are both available in at least one period (in which case the base period price difference between the new and old items is implicitly treated as completely due to quality differences).

From time to time, new items appear in the system that do not fit existing ELI definitions but are close substitutes for items that do. The example cited in Armknecht et al. (1997) is CD players, which were clearly substitutes for phonographs and tape players. Such items emerge in Consumer Expenditure Survey (CEX) interviews and are coded separately so that expenditure data can be entered; Armknecht et al. (1997) indicate that existing CPI procedures can accommodate these situations. However, "because of the time lapse associated with the CEX and TPOPS surveys, unless special efforts are made to reinitiate an item stratum, three to four years elapse from the time the product appears in the marketplace to the time the product appears in the CPI" (Armknecht et al., 1997:387). BLS accounts of the 1998 revised CPI methodology, however, suggest that the new ELI structure and the change to category rotation under TPOPS will provide more sampling flexibility and bring new items more quickly into the index (see Greenlees and Mason, 1996).

If costs linked to these processes were not a consideration, the panel would likely agree, without qualification, with the Conference Board's suggestion that sample rotation schedules should be further sped up for most goods. Cost is a big factor though and, given current evidence, it not clear that a 2-year rotation plan would yield commensurate benefits in terms of index accuracy.

> **Recommendation 5-1: Until it can be shown that further compression of the sample rotation schedule would create significantly different rates of change in the CPI, the panel is satisfied with the current BLS plans. We do recommend that BLS undertake research designed to assess the impact that moving from a 4-year to a 2-year rotation cycle would have on the rate of index growth.**

This will entail analyzing a broad set of items to simulate the effect that more frequent rotation would have. BLS should specifically investigate the statistical disadvantages of frequent sample rotation. New samples enter the CPI by chaining, and these transient disturbances can cause an index to display higher variances and a tendency to drift upward. Only after these issues are given further attention can recommendations regarding the optimal sample rotation frequency be advanced in a fully informed manner.

As noted above, BLS must also deal with the class of product introductions that are not picked up during normal sample rotation. If a new item is encountered that does not fit an existing ELI definition and is not obviously a close substitute for one that does but does fit within an existing item stratum, the remedy is to define a new ELI. "This process could take five to seven years for full implementation" through the normal sample rotation process (Armknecht et al., 1997:387).

If a new item does not fit any existing item stratum definition (e.g., cell phones, VCRs), it normally does not enter the CPI until item strata are redefined. Until recently, this happened only when major revisions of the CPI were introduced—about every 10 years.[13] Historically, the more novel a new item, the

[13]BLS (Greenlees and Mason, 1996:3, 4) explains that the "the most fundamental and visible activity in each of these CPI revisions is the introduction of a new "market basket," or set of

longer it has taken to appear in the CPI. Home computers and VCRs, for instance, were not introduced into the CPI classification until 1987.

BLS now recognizes that the past delays between the introduction of entirely new and important goods and their appearance in the CPI are no longer acceptable. The most obvious way that this recognition has changed the CPI production schedule is in the updating of upper-level expenditure weights. Weights will be updated more frequently, and they will be based on a shorter span of expenditure data. Beginning in 2002, weights will be updated every 2 years, with a 2-year lag (Bureau of Labor Statistics, 1998a):

> Thus, for example, CPI expenditure weights will be updated to the 2001-02 period effective with release of CPI data for January 2004. As a result of this change, expenditure weight data will be, on average, "two years old" when introduced into the CPI, and four years old when replaced. By contrast, the most recent set of CPI expenditure weights—based on 1993-95 CEX data—were on average, 3 years old when first used in January 1998, and they replaced weights that were about 15 years old.

BLS has also begun researching the advisability of adopting special targeted procedures to quickly bring new products (Viagra was one) into the index (for a full description, see Lane, 2000). BLS's work in this area is commendable, and BLS should continue to develop changes in its procedures designed to reduce those delays substantially.[14] When visible items, such as home computers and VCRs, that achieve significant expenditure shares can be brought into the item samples rapidly, public and policy maker confidence in the CPI and the BLS can only be improved. While the effect of earlier inclusion of any one product is

---

expenditure weights attached to the categories of goods and services comprising the CPI." The year of a revision is identified when the new market basket introduction occurs—as such, the most recent "major revision" is usually identified as the 1998 CPI revision. The authors add:

> The projects and changes encompassed in the current [1998] revision—the sixth major revision in the CPI's history—range from the reselection and reclassification of areas, items, and outlets, to the development of new systems for data collection and processing. . . . It is important to note, however, that numerous methodological improvements in the CPI have taken place outside the revision framework. Among the most prominent examples of these are the annual adjustment for changes in the quality of new cars after model changeovers were introduced in 1967, the shift to flow-of-services measures of the cost of owner-occupied housing in the early 1980s, and the implementation of regression-based methods for quality adjustment of apparel prices starting in 1991.

See Greenlees and Mason (1996) for a full itemization of methodological changes associated with both major and interim CPI revisions.

[14]The 1999 CPI Improvement Initiative provided funds for study and data collection aimed at incorporating new goods into the CPI more quickly. Lane (2000) describes numerous methods, both currently used and proposed, for bringing new items into CPI samples more quickly. Most of the proposed new methods entail expanding current item and outlet rotation, along with augmentation and replacement methods.

likely to be minor, systematic early inclusion of new goods broadly could have a significant effect on index growth.

There are limits on the extent to which speeding up product introduction will relieve the problem, however, since the BLS may not be able to justify—for both analytic consistency and budgetary reasons—revolving a large number of new goods on an extremely short-cycle schedule. Analytically, more rapid rotation and more frequent rebasing require proportionately more chaining of indexes with non-identical components, which can exacerbate index drift (see Glossary). On the cost side, each rotation creates inefficiency because another period must elapse (to produce a price change) before quotes on new items can be used. As rotation frequency increases, the amount of information used in the index relative to the total amount of information collected is decreased.

Finally, though it is important that new items be introduced into the index once they are commonly consumed, they need to be entered with a correct expenditure weight. Since only price data are compiled on a monthly basis, it is not easy to estimate a weight immediately. Thus, there may be a tradeoff between timeliness and accuracy of item weights. For many cases, it may only be practical to introduce new items into the sample rotation after a significant and estimable market share emerges.

In light of these considerations, two approaches to the item introduction problem seem potentially worth considering by the BLS. First, broader ELI and item strata definitions or definitions couched in terms of function instead of product (e.g., audio reproduction instead of phonographs and tape players) might reduce substantially the number of new items that must be excluded for long periods because they do not fit existing definitions.[15] We recognize that there is likely a tradeoff between breadth and clarity of boundaries, but that does not establish that more breadth would not be better. Second, the Conference Board suggests expanding the BLS's small program of special sampling to "arrange for regular consultations with panels of experts . . . persons who are likely to know when important new consumer products have recently or soon will reach the market." This, too, seems sensible and worth serious consideration.

NEW OUTLETS

Current Practice

BLS rotates a portion of its sample of retail stores and business establishments each year. The probability of an outlet being selected is proportional to store-by-store expenditures reported by consumers in TPOPS. Outlet rotation is

[15]The 1998 revisions did, of course, make such adjustments to broaden the scope of video and audio items.

designed to make the index reflect changing consumer shopping patterns—that is, to collect prices from the places where consumers shop most. Item rotation and outlet rotation occur simultaneously; as a new outlet rotates into the sample, so too do many new (to the index) products. A new product is introduced to cover the same market basket category (for instance "records and tapes") as that which was sampled at the outgoing outlet, but whether or not the same specific item type (say, a CD) that was previously priced remains in the sample depends on expenditure shares at the new store in conjunction with the random selection component of the rotation procedure.

In the previous section we addressed difficulties that arise when sales-based outlet and item selection require BLS to compare newly sampled, non-identical items with those that have been replaced from the outgoing outlets. However, a different issue arises when, after outlet rotation, a specific product from the old outlet sample also appears in the new outlet sample but at a different price. Because of the expenditure-based sampling process, products with high market share and high sales volume are particularly likely to be reselected through successive outlet rotations.[16]

Outlet substitution expands the scope of index coverage beyond that represented by specific items, incorporating the notion that, in acquiring goods and services, aspects other than price may affect a consumer's cost of living.[17] When consumers purchase a good at a particular store they are buying a package. The package includes not only the specific item but also the quality of the shopping experience—the services provided, the store's locational convenience, its return and exchange policy, and the variety of products available. In this context, the issue of the value of time naturally arises. As it relates to consumer shopping, and specifically outlet-use patterns, valuing nonmonetary benefits associated with time savings, improved convenience, or better service is central to the concept of a COLI. Treatment of the issue should distinguish between time as (1) a variable that (perhaps combined with transportation costs) might be used to help explain differential outlet quality and (2) any explicit imputations of the value of time spent shopping (perhaps corrected for any entertainment component of the activ-

[16]The outlet effects discussed in this chapter may also be relevant when sample rotation leads to the selection of a non-identical item. However, it is instructive to begin with the easier "all-else-equal" case when prices of identical items are compared, because it allows one to abstract from the item quality problem. Conceptually though, any remedies to outlet bias might apply to the non-identical items case.

[17]A textbook Laspeyres index might avoid this issue since, strictly speaking, pricing a fixed market basket requires selecting the same goods from the same outlets over comparison periods. In a rapidly evolving economy, it is worthwhile to sidestep such a restrictive practice. Failure to rotate outlets allows the index to drift further and further away from reflecting trends in transactions prices—the prices people actually pay at the outlets where they typically shop. That is why the BLS and virtually all statistical agencies in other countries choose to modify their Laspeyres indexes to allow for periodic updating of the sample of outlets from which prices are collected.

ity), derived from wage, survey, or other data, which might be added to the cost of obtaining a given level of material well-being.

Variation in prices charged for the same goods at different kinds of outlets can be substantial. For example, it is not uncommon for a specific brand and size of some good—a laundry detergent or motor oil—to cost 50 to 100 percent more at a 24-hour convenience store than at a large discount outlet. As Pollak (1998) notes, consumers face a distribution of prices for many goods and services, rather than a single price, and at least some of them find it worthwhile to search for lower prices.[18] Consumers benefit if they can reduce costs—and if these gains are not fully offset by inferior service or greater item acquisition costs—by substituting purchases from a high-price seller to a low-price competitor. But because of the way in which new samples are linked to old ones, this type of consumer benefit is not always picked up correctly by CPI procedures.

When a new set of stores enter the sample and prices for various categories of goods are collected, *all* of the difference between the old outlet price and the new outlet price is linked out. For cases in which an identical item (such as a 12-ounce tube of Crest toothpaste) is priced and that price is different, BLS attributes the entire difference to outlet-related quality variation. This could be correct in some cases; however, for all cases, any change in the price recorded for the item has no effect on the CPI.[19] Application of this procedure implies acceptance of the assumption that markets are in equilibrium, so that differences in price are exactly offset by differences in retail service. The practice also means that the BLS approach to cross-outlet price linking allows factors other than observed market price to affect the index. The opposite approach would assign all of any change in the price of an item from one outlet to another to real price change.

If changes in consumer outlet choices exhibited no clear trends or if price differences simply reflected the fact that discount stores are achieving lower costs by cutting the quality of services, there would be little reason for concern. However, the last few decades have produced clear patterns of change in shopping

---

[18]Pollak (1998) notes that this aspect of consumer behavior has been largely ignored in the research, which almost always proceeds as if the "law of one price" holds.

[19]Moreover, even when the prices at the old and new outlets are the same, an assumption about outlet quality is still being made. In this case, though, the assumption is that the outlets offer the same nonprice benefits and costs to consumers. To be fully consistent, BLS would need to extend the implicit quality-based price adjustments made for non-identical pricing when linking identical prices. To illustrate, consider a situation in which rotation from outlet A to outlet B yields a lower observed price for a 12-ounce tube of Crest toothpaste but the same price for a 48-ounce container of Tide laundry detergent. If the lower price of toothpaste is attributed to inferior outlet quality (which requires implicitly adjusting the new price upward to match that from the old outlet), the price of the laundry detergent should also be adjusted upward, meaning that the laundry detergent is actually more expensive at the new outlet. A complicating factor is that quality differences across outlets may be commodity specific; for some goods, care and service could be much more important than for others.

choices. It is well documented that general shifts in shopping habits—specifically, the move from higher-priced full-service outlets to discounters—have occurred. For instance low-price and expanded-format food stores grew from a 31 percent to a 50 percent market share between 1979 and 1988 (Reinsdorf, 1993: 228), and discount outlets have permeated well beyond food sectors, into electronics, computer equipment, home improvement, and others. Under the prevailing methodology, a large shift in the types of outlets that consumers choose to patronize does not directly affect the index trend, since all price change arising from rotation is linked out. This potential bias affects major CPI item categories, including food and beverages and apparel, although other categories—most notably housing—are not affected by outlet substitution.

The fact that the market share of low-price discounters has been steadily growing implies that, even after quality adjustment, prices at those stores are lower than elsewhere. If through economies of scale and other means, the large-volume retailers have been able to provide lower prices to a growing number of consumers for the same quality-adjusted goods, the current procedures bias the CPI upward (as a cost-of-living indicator). At the same time, a minority of consumers who would have preferred to continue shopping at traditional stores found them driven out of business by the new outlets; those consumers experience an increase in their living costs.

New outlet bias is lessened to the extent that low pricing at new outlets forces established ones to follow suit while they are still in the CPI sample. Depending on the timing of store pricing responses and outlet rotation, the CPI may capture such a price decline. If the now relatively less frequented outlets that BLS used to sample began lowering prices before they were rotated out, the price pressures created by the new outlets would be captured.[20] In fact, within the economic model of perfect competition, lower prices would, in equilibrium, be balanced by poorer quality. If equilibrium were always maintained, there would be no potential for this type of bias. However, it appears that the shift to new outlet types has been an ongoing process that is still continuing. As new outlets open, consumers in the area gradually change their shopping behavior and take advantage of the lower quality-adjusted prices.

The recent emergence of e-commerce (the business-to-consumer component) has the potential to create another disequilibrium situation. Expenditure and sales data indicate that consumers are purchasing a small but rapidly increasing share of goods and services through Internet retailers. BLS is planning to rotate these outlets into its sample more or less according to standard protocol. Current CPI procedures for determining where consumers shop should capture increased

[20]Likewise, as Shapiro and Wilcox (1996) point out, the potential for upward outlet bias is also reduced when established outlets respond to price competition by reducing the quality of their service.

patronage of Internet outlets. Even so, emergence of a new mode of shopping highlights time use, item acquisition, and other considerations that must be thought through in order to make pricing of purchases consistent across outlet types.

To illustrate one dimension of the story—item acquisition—consider the example in which, in period one, a compact disk costs $16 at the local record store. In period two, after outlet rotation, a Tower Records superstore that sells the same compact disk for $15 replaces the local store in the sample. Under current procedures, the drop in price does not figure into the CPI; BLS chalks up this price change to a type of quality difference, implicitly assuming that consumers must bear other costs to obtain the item at the lower price. In this case the new costs may derive from things like longer travel time and increased fuel use to get to the superstore or from sales clerks who cannot provide informed answers about music. Since linking essentially equates the old outlet price with the new, some costs beyond the price paid at the register are implicitly part of the adjustment. BLS assumptions about cross-outlet price variation therefore rely to some degree on the idea that an index should reflect full consumer costs, rather than simple transaction price.

For Internet (or catalog) purchases, BLS includes shipping costs in the price. Extending the example from above, say the compact disk in question costs $14 at Tunes.com but requires an additional $2 shipping fee. The e-purchase will be recorded as $16 for index calculation purposes. If a second outlet rotation pushes Tower out of the sample and brings Tunes.com in, the change in price from $15 back to $16 does not affect the index—it is linked out. The higher Internet purchase price is attributed to a difference in quality between the Tower and Tunes.com outlets and contains no pure price component. As discussed above, it may be questionable to assume that the full price difference is equal to the difference in the value of outlet service perceived by consumers, but the method is consistently applied.

On the other hand, the validity of any price *level* comparison would be questionable (since this comparison is avoided, this does not initially create any problems). From consumers' perspectives, the cost of procuring the CD at Tunes.com (shipping) may be viewed as a substitute for time spent (an opportunity cost) and gas used (an explicit cost) traveling to the store. Thus, there may be an element of acquisition cost that is included in the e-purchase that is not in the bricks-and-mortar purchase; the recorded $15 price at Tower does not include the cost to the consumer of driving to the store to buy the CD, while the $16 e-purchase price includes delivery.[21]

[21]Although some share of the computer purchase price and time spent at the web site are additional procurement costs, they are likely to be trivial.

In this example, the value of time, which varies from one consumer to another and which must enter a full-blown COLI concept, becomes a relevant issue since choices relating to it can affect the index. Consider, for instance, that different consumers pay different shipping rates for the same product from the same outlet. Some customers choose to pay an extra \$3 to have one CD delivered overnight; others pay only 10 cents for shipping because they order 20 CDs at the same time and choose 2-day ground delivery. Are these consumers paying different prices for the product, or are they paying for time savings or for additional services that contribute in combination with the primary item to their utility? The customer selecting overnight delivery reveals not that he would rather pay \$19 than \$16 for a CD, but that there is another element besides eventual ownership that affects his well-being. By extension, if everyone changes from buying CDs at the superstore to buying them from the Internet store (and sample rotation reflects this), the price index should not automatically rise (and it wouldn't under the current linking procedure). However, problems could arise if there is a shift in available delivery options or consumer choices. Say, for example, that BLS originally prices CDs delivered by 2-day mail. Now assume that the Tunes.com upgrades its delivery service by replacing 2-day delivery with overnight delivery without raising prices. If no quality adjustment is made when CD purchases under the new setup are priced, the index would miss a true price decrease. This would clearly be incorrect from a COLI perspective (and, depending on how the good is defined, probably also from a cost-of-goods index [COGI] perspective). The problem could arise if item acquisition cost is not treated in a consistent manner—that is, if in some cases it is left out and in others it (or part of it) is included.

The big question is whether or not any of these outlet effects are quantitatively important to index performance. At present, it seems unlikely, but conditions are changing. For instance, market boundaries are rapidly expanding, which might minimize (or exacerbate) the price dispersion problem. By reducing information and search costs, the Internet may one day make the law of one price assumption less unrealistic. The Internet allows consumers access to a vast amount of product information that enables them to easily shop for the lowest price; this visibility also puts pressure on retailers to match competitors' prices. Fuller information and global access mean that indexes for, say, San Francisco and Milwaukee might converge toward the national average. Consumers now have fuller access to product specifications and worldwide price information than ever before. Even consumers who ultimately patronize only brick-and-mortar outlets can save time gathering price and product information. The cost of making informed decisions about purchases has decreased; the chance of purchasing at a noncompetitive price has been reduced. The former effect (cost of information) is not captured in current CPI methods; the latter effect (lower price) probably is.

To the extent that e-commerce forces competition, differential price trends for specific geographic areas may be minimized. As such, the increase in e-market penetration may eventually reduce the relevance of geographic (or demographic group) indexes. In addition, the relevant region of any transaction is becoming more difficult to define. A consumer in Chicago may buy an item produced overseas, sold by a dotcom operating in San Jose, and shipped from a warehouse in Alabama. Theory suggests that elimination of information and geographic market barriers will force retailers toward operating at uniform profit margins. But some unconventional pricing practices within the Internet retailing sector have also emerged: companies that are surviving on market capitalization may operate with large accounting losses, generally in an effort to maximize not short-run profit but market share, though this phenomenon is rapidly perishing with the widespread failure of dotcom start-ups. However, one should not overstate the case for price convergence, given that the evidence so far is surprisingly weak. Part of the reason may be that e-retailers (like catalog merchants) have the ability to price discriminate on the basis of a consumer's past purchasing behavior or, perhaps, information obtained about the purchaser from other vendors. Furthermore, major CPI components, such as housing and utility services, will always be affected by local market conditions and institutional factors.

Evidence of Outlet Bias

The research attempting to estimate the extent of outlet substitution bias is thin.[22] Reinsdorf's 1993 study, probably the most cited on the topic, formally outlines the underlying theory and offers empirical evidence that outlet price differentials are at least partly real rather than merely reflective of quality differences. Reinsdorf—whose research served as the basis for the Boskin commission's outlet substitution bias estimates—compared food and motor fuel prices from outgoing and incoming samples during a 2-year overlap period when samples were being rotated. New sample prices were on average about 1.25 percent lower. Given that sample rotation occurs every 5 years—and given the rather strong assumption that the lower prices were not accompanied by a deterioration in service or other outlet-related quality elements—this implies a 0.25 percent annual bias in the relevant components of the index. Reinsdorf provided a second estimate by tracking changes of CPI components against unlinked average price-paid data (also published by BLS). For foods the average price indexes rose 2 percent more slowly than did linked CPI subindexes, and for unleaded gasoline 0.9 percent slower. Since quality change is not controlled for, Reinsdorf asserted that these estimates should be thought of as an upper bound of sorts for the outlet substitution bias.

Lebow et al. (1994) adjusted the Reinsdorf estimate to reflect that only a subset of CPI goods are affected by outlet substitution. The authors determined

[22]The literature does not appear to go much beyond Pollak (1998), Anglin and Baye (1987), Baye (1985), and Reinsdorf (1993).

that weights associated with the relevant set of categories account for about 40 percent of the CPI, which yields an overall index bias of 0.1 percentage points (0.4×0.25) per year. The Boskin commission adopted this estimate in its categorical reporting of CPI outlet bias. Given the absence of alternative evidence, it is hard to fault this choice, but the quantitative effect of outlet substitution remains unclear. Fixler (1993) notes that comparisons between the movement of average prices (as used by Reinsdorf) and the CPI strata counterpart do not provide direct evidence of the outlet effect because "differences in index formula, in treatment of product quality change, and in coverage of average prices and CPI strata indexes" may also play a part in the divergence of the two series (p. 8).

MacDonald and Nelson (1991) also produced a rough estimate of the bias created by the market shift to discount stores by combining information on prices across outlet types and market shares. At the time of their study, data published by the trade journal *Progressive Grocer* indicated that prices at warehouse food stores were 13.4 percent lower that at traditional outlets. The lower prices, along with a 0.7 percent annual growth of warehouse store market share, imply a non-quality-adjusted 0.1 percent per year index bias. Any quality adjustments in favor of traditional outlets would reduce the bias. Also, bias estimates implied from market share information may be overstated if consumers who preferred traditional outlets are forced, because of the outlets' extinction, to patronize the less desirable superstores. A full estimate of outlet bias would have to consider the increased quality-adjusted price that traditionalists must now pay.

Estimating the Real Component of Price Differences Across Outlets

To accurately remove outlet substitution bias from a COLI, an index producer must abandon the assumption embedded in either extreme position—that any observed price difference of an identical item at two outlets (1) is wholly attributed to outlet-related quality differences valued by consumers or (2) contains no quality component and therefore reflects pure price variation. In order to escape these assumptions, methods would have to be developed that isolate and quantify the value to consumers of the service, time, and other quality dimensions that differ by outlet type so that the pure price component could be identified. That is, differences in observed price changes associated with outlet-rotated items must be broken down into price and quality components, as is done for items whose embodied characteristics change.[23]

---

[23]Dennis Fixler (who, at the time, was at BLS) informed the panel that the producer price index (PPI) program has begun investigating approaches that would treat retail services in the grocery store component explicitly. Fixler's assessment is that "to date, they have not been successful in linking changes in outlet characteristics such as number of cash registers or the number of stock keeping units (as a proxy for the scope of products available) to changes in the retail margin—the measure of price for retail services." He notes that work is also under way to examine links between CPI-collected prices and sets of outlet characteristics (personal communication).

This prescription certainly suggests a role for hedonic analysis, in this case based on outlet (as opposed to product) characteristics. However, work in this direction is embryonic and applicable results appear to be a long way off. This type of hedonics may be even more complex than in the standard quality adjustment context since such an analysis would need to "allow for the existence of temporary market disequilibria and a distribution of preferences across consumers" (Reinsdorf, 1993:250).[24] And the problem of identifying outlet characteristics that are tied to consumer valuations of service is certainly conceptually no simpler than in the parallel item quality case.[25]

Because any attempt to control for price changes attributable to item and outlet characteristics would require collecting detailed data, scanner technology may offer some hope for advancing research on outlet substitution bias. If outlet identifiers could be incorporated into sales data generated by stores, scanner data could reduce the time needed to track which outlet types are experiencing growing or shrinking market shares. Electronic data sources may also facilitate systematic unit-price calculation to account for different packaging that characterizes small and large shops and for varying promotional sale and coupon use patterns. Scanner data may promote more accurate tracking of actual transaction prices, which would be essential to research on price variation across outlets.

Though current evidence suggests outlet bias is significant enough that index producers should be concerned about it, there is a real question as to whether research can generate sensible, reproducible price estimates that neutralize quality in across-outlet product comparisons. Given the complicated nature of any such estimation, the Conference Board (1999:23) suggested, as an interim solution, splitting the difference between the two extremes—that either all or none of observed price differences are due to quality variation—on the basis that it is no more arbitrary than the "all or nothing" assumption and that is likely to be closer to the "truth." The board added that such a practice should only be adopted if solid analysis of specific items suggests that the approximation is reasonable. After all, it is certainly possible that one of the two extreme assumptions is in fact empirically closer to the truth than is the half-and-half solution.

[24]When shopping patterns are not changing over time, it is correct to attribute price differences across outlets to quality variation. Reinsdorf suggest that this is not the case, however: "Large shifts in market share in favor of discounters indicate that the inframarginal consumers making such outlet substitutions experience increased consumer surplus. The average value of this increased consumer surplus depends on the distribution of preferences across consumers, which could be estimated if data providing equilibrium market shares at various price differentials between outlet types were available" (Reinsdorf, 1993:250).

[25]White (1999), using Statistics Canada data from the period 1990-1996, found that outlet type is a relevant price-determining characteristic in terms of *both* price level and price movement differences. He also found that outlet bias in the Canadian CPI is due more to unrepresentative outlet sampling and less to linking.

Conclusions and Recommendations

BLS treatment of items that enter the CPI through outlet rotation is conceptually parallel to its treatment of items that replace disappearing ones (except that, for outlet substitution, overlap pricing can nearly always be used to impute price change for the transition month). In the case of new outlets, the continuing (for now) shift of purchases toward large discounters suggests that the price differences are not, on average, proportional to perceived quality variation of shopping experiences.

On the assumption that the COGI concept does not call for fixed "weights" among types of outlets then, under either a COGI or a COLI approach, an explicit decision is needed about how to make item price quotes comparable as new outlets replace old ones in the CPI sample. For the foreseeable future, the BLS will not have the tools to explicitly adjust observed prices to account for changing outlet quality characteristics. Thus, the range of short-term recommendations from which the panel may select is limited. BLS could:

- continue the current treatment of outlet replacement—in which case all of any price difference for a specific item from one outlet to another is assumed to be equal to the difference in the quality of the shopping experience;
- treat the price difference as a "true" price change—in which case zero net quality difference is assumed to exist between outlet types—though this is a poor option because prices at the newly selected outlet were, in most cases, lower before it was rotated into the sample, so even if a (quality-adjusted) price decrease took place, the timing of its inclusion into the index would be wrong; or
- split the difference—as recommended by the Conference Board, but there is little reason to believe that splitting the difference is more accurate than the results of the current practice (and such a recommendation might be faulted for creating the precedent of solving a difficult problem without quantitative evidence).

Conclusion 5-2: Given the available options and given that current techniques cannot consistently and accurately separate quality and price effects associated with the value of retail service, BLS has little choice but to continue its current practice.

However, *in principle*, when outlet rotation results in a change in the observed price of an identical product, an attempt should be made to decompose the difference into quality (or convenience) and pure price components, instead of attributing it, in its entirety, only to the former.

Recommendation 5-2: With longer-term modifications in mind, the panel recommends pursuing research into price variation across outlets with differing characteristics.

As with product-based hedonic techniques, substantial methodological and data advances would be needed before any such changes to the CPI could be put into practice. Undertaking this approach would require BLS to intensify research on consumer search costs, time use, service valuation, hedonic methods applied to outlet characteristics, and, more generally, how to adapt sampling methods to facilitate more extensive quality adjustment across item rotations. Since it seems unlikely that efforts in these areas could have a large effect on the CPI and given their conceptual difficulty, the panel would assign a low priority to this research.

6

The Special Case of Medical Services

The conceptual and measurement issues that underlie a cost-of-living index (COLI) are perhaps more difficult when one attempts to construct an index for medical care than for any other component of the Consumer Price Index (CPI). Nowhere is the term "cost of living" more literally applicable than for medical care. This chapter begins by briefly outlining current procedures for calculating the Medical Care Price Index (MCPI) and then discusses a number of issues that have been raised concerning MCPI reliability and accuracy. The chapter concludes with recommendations.

BACKGROUND

Medical care is one of the eight major product groups in the current CPI. For each of these major product groups, the domain of the CPI is currently limited to prices for goods and services on which the consumer makes a direct out-of-pocket outlay. This limited domain is of particular importance in the case of medical care expenditures.

BLS decomposes consumers' direct out-of-pocket payments for medical care into five categories:

- prescription drugs;
- over-the-counter drugs and medical supplies;
- services from physicians, dentists, and other medical professionals;
- hospitals and related services; and
- health insurance.

When the revised CPI weights were introduced in January 1998, these five medical care categories together accounted for about 5.6 percent of total consumer expenditures.

The consumers' out-of-pocket health insurance component represents only that portion of total health insurance premiums paid for directly by the consumer. This fact partly explains why medical care is a much smaller share of total expenditure in the CPI (5.6%) than it is in the national income and product accounts. In 1999 national health expenditure was 13 percent of gross domestic product and 17.6 percent of personal consumption expenditure (see Heffler et al., 2001; U.S. Department of Commerce, 2000:Table B.4, p. D-31). The health insurance component of the CPI excludes all employers' contributions to health insurance. To avoid double counting, it also nets out any medical expenditures for which the consumer is subsequently reimbursed by an insurer. This direct out-of-pocket health insurance expenditure is by far the largest of the five MCPI components—in 1995 it comprised 49.6 percent of all out-of-pocket medical expenditures.

If, instead, BLS constructed a price index for health insurance (distinct from health provider services), a number of very significant conceptual and measurement issues would need to be resolved. Although BLS is currently reassessing the feasibility of proceeding with direct measurement of prices of health insurance policies, it has not taken that route (see Greenlees and Fixler, 2000; Bureau of Labor Statistics, 2001). Rather than trying to price health insurance, the BLS distributes out-of-pocket expenditures for private health insurance (fees for service commercial carriers, Blue Cross/Blue Shield, preferred provider health plans, and health maintenance organizations) to the five MCPI subcomponents listed above.

In 1995, consumer expenditures for fee-for-service commercial carrier health insurance were reallocated as follows (Ford, 1995):

- hospital services, 39.7 percent;
- physician and dental services, 34.1 percent;
- other medical professional services (such as home health care), 6.5 percent;
- prescription drugs, 6.2 percent;
- nursing homes, 0.6 percent; and
- pure insurance, 12.9 percent.

By pure insurance the BLS means the services that insurers provide, such as processing claims, not payments to claimants. Thus, consumer expenditures on pure insurance services are computed essentially as premium revenues minus claims paid. Out-of-pocket expenditures to other private health insurers (e.g., Blue Cross/Blue Shield, preferred provider health plans, and health maintenance organizations) are reallocated back to the five MCPI subcomponents in a similar manner.

Consumers also make direct out-of-pocket payments for health insurance to the federal government, which underwrites Medicare Part B (consisting of voluntary payments, primarily by the elderly and disabled, for insurance benefits supplementary to Medicare Part A compulsory hospital insurance).[1] Consumers' out-of-pocket Medicare Part B expenditures for health insurance are also reallocated by the BLS back to the MCPI components; Ford (1995) calculated the following allocation for 1995:

- outpatient hospital services, 27.2 percent;
- physicians' and dental services, 56.8 percent;
- other professional services, 9.2 percent; and
- supplies and durable medical equipment, 6.8 percent.

Notice that BLS specifies there is no pure insurance component for Medicare Part B.

Finally, to obtain the total weights by subcomponent within the MCPI, the BLS sums up, for each category, consumers' direct out-of-pocket expenditures (except health insurance), plus the reallocated private insurance (for commercial carriers fee for service, plus similar reallocations for Blue Cross/Blue Shield, preferred provider health plans, and health maintenance organizations), plus the reallocated federal government insurance. The resulting direct consumer payments plus reallocated health insurance payment expenditure weights were as follows for January 1998 (Ford and Ginsburg, 2001):

- prescription drugs, 14.9 percent;
- over-the-counter drugs and medical supplies, 7.6 percent;
- physician, dental, and other medical professional services, 49.5 percent;
- hospital and related services, 22.9 percent; and
- pure health insurance services, 5.0 percent.

There is considerable price discrimination in medical markets (the law of one price does not hold). Prices paid directly by patients typically exceed those paid by large insurers. Medical care providers and insurers are often reluctant to disclose proprietary price information (indeed, they are sometimes contractually prohibited from doing so). As a result, BLS faces significant operational issues in

[1]Medicare Part B insurance covers physicians' services, outpatient hospital care, and medical equipment prescribed by the physician for use in the home. A deductible applies. Generally, Medicare Part B does not cover prescription drugs, except in a few special cases. See *http://www.medicare.gov/* ("Glossary" and "Frequently Asked Questions").

obtaining appropriate and reliable transaction price quotes separately for the direct and indirect MCPI components.

CONCEPTUAL AND MEASUREMENT ISSUES

Measuring the prices of medical care services presents many challenges, both conceptual and operational. The medical care sector has undergone, and is continuing to undergo, considerable technological progress and institutional changes, resulting in changing quality of care. As discussed in Chapter 4, the prices paid by patients and insurers for medical care goods and services should, in principle, be adjusted for some, perhaps all, such quality changes in medical care. Consumers' health status, however, depends not just on their physicians and the medicines they are prescribed but also on their own behavior and life-styles and on the environment in which they live. Thus, one cannot automatically equate changes in health status with changes in the quality of medical services. There are, in addition, profound ambiguities concerning the identity of the consumer and the individual making the consumer's choice: Is it the patient? The patient's family or other caregivers? The physician acting as an agent for the patient?

There are also ambiguities about exactly what kinds of services the health care sector provides and hence what outputs should be priced. Diagnostic services that lead to the identification and successful treatment of a symptom can, for example, be included as part of the cost of treating a specific condition and generating a specific output. New diagnostic services that mainly allow a physician or a patient to reject an unlikely diagnosis are more difficult to classify and assess, since they may mainly deliver peace of mind rather than health. Alternatively, such diagnostic services may make patients worry about possibilities that they did not consider before, as can occur from false positives yielded by the prostate-specific antigen (PSA) test for prostate cancer or the pap smear test for cervical cancer.

Ideally, the BLS should not be alone in the world trying to answer these questions. But in most other developed countries, medical services are paid for primarily by governments or government-mandated insurance funds. As a result, medical services are outside the domain of their CPIs. In large part, therefore, BLS is going it alone in addressing these difficult conceptual and measurement issues for the construction of an MCPI.

Input Substitution and Pricing Episodes of Medical Treatment

One of the most significant issues facing BLS is whether it should price medical inputs or medical outputs (outcomes). In years past, BLS has priced a fixed bundle of discrete inputs, such as a day in the hospital, a visit to a gastroenterologist, or a serum laboratory test. This procedure was often criticized, even as early as the 1960s (see Scitovsky, 1967). Among other problems, it overlooks substitution possibilities among medical inputs for treating a particular condition. More specifically, the BLS practice of separately pricing and weighting distinct medical item strata neglected the medical care sector's substitution across various strata (e.g., physician services, prescription pharmaceuticals, laboratory tests), thereby overstating price increases.

A well-known example of this problem involves inpatient and outpatient hospitalizations. One way in which managed care has reduced overall hospitalization costs is to shift many surgeries from inpatient to outpatient environments. A consequence of this substitution is that both the severity of illnesses and the complexities of surgery for the average patient have increased over time for both inpatient and outpatient procedures, resulting in an increase in per diem costs in both settings. Because BLS priced these inpatient and outpatient procedures separately and then used constant weights over time to aggregate them, their weighted sum increased over time, even though providers' and insurers' total hospitalization costs (inpatient plus outpatient) for these surgical cases declined. Since the 1998 revisions, the BLS has treated inpatient and outpatient hospitalizations as a combined bundle (stratum), although the frequency with which the substrata weights will be updated has not been made clear to the panel.

Another example of substitution across strata involves the treatment of a mental disorder such as depression (see Frank et al., 1998; Berndt et al., 2001). The clinical literature has demonstrated that a number of alternative treatments for depression involving various combinations of psychotherapy and antidepressant drugs have, on average, equal expected outcomes. In the MCPI, the services of psychotherapists are classified in a separate stratum from antidepressant pharmaceuticals. Over the last decade psychotherapy-intensive treatments for depression have been reduced, and they have been replaced by either a combination of psychotherapy and pharmacological treatments or by pharmacological treatments alone. Although the movement away from psychotherapy-intensive procedures has, in many cases, reduced total costs per treatment episode, this cost reduction is not captured by current BLS procedures because the inputs come from distinct strata, each of which is priced separately and reweighted infrequently. Moreover, this problem would not be resolved by treating inpatient and outpatient hospitalizations as a combined stratum.

BLS has been moving toward the pricing of an episode of an illness, rather than pricing medical inputs. For example, for those hospitals and physicians billing for services provided to Medicare beneficiaries, BLS is now collecting selected price quotes based on Medicare's Diagnostic Related Group (DRG) classification scheme, a system that to some extent encompasses episodes of treatment.

A major set of issues now facing BLS is how to broaden the new approach to encompass the entire treatment of medical conditions, not just hospital stays. One possibility is the following. First, the BLS could choose a subset of diagnoses or illnesses (perhaps between 15 and 40, depending in part on the costs involved in analyzing data) from the *International Classification of Diseases* (*ICD*) (U.S. Department of Health and Human Services, 1980). Once the sample of diagnoses is identified, they can be matched with existing medical treatment data from large retrospective medical claims databases (e.g., those of medical insurers or payers) to monitor changes over time in the cost of a treatment episode. Retrospective medical claims data generally provide detailed information, by individual or family unit, including date of visit with the physician or other medical care professional, medical procedures provided, primary and secondary diagnoses, hospitalizations, dispensed medications, and the patient's and insurer's payments for each of these services.

The diagnoses selected for pricing by the BLS could be chosen on the basis of the relative amounts spent on the disease so that, for example, heart attacks would be far more likely than conjunctivitis to be chosen, even though conjunctivitis has more entries than acute myocardial infarction in the *ICD*. When an individual receives treatments for several distinct but co-occurring illnesses, health services researchers frequently either assign each encounter entirely to the primary diagnosis for that encounter or split the utilization and costs equally among the various diagnoses for that encounter.

Given the present state of information technology and medical claims processing, such a procedure could not be implemented in real time for an MCPI that must be published within a month of data collection. But currently it is feasible with about a 1-year lag, allowing the BLS to publish an experimental index on an annual basis.

Because the BLS would probably find it too costly to produce this type of medical care price index each month, we recommend it consider producing an index based on a pricing of the sample of treatment episodes of distinct diagnoses. These diagnoses have two components—quantity weights and prices, just as with any other price index. For the monthly medical care index, the BLS could reprice the current set of specific items (e.g., anesthesia, surgery, medications), keeping quantity weights temporarily fixed. Then, at appropriate intervals, perhaps every year or two, the BLS should reconstruct the medical care index by pricing the treatment episodes of the 15 to 40 diagnoses in the manner described above.

The index, therefore, might have a break when the set of diagnoses are repriced. There is the possibility that it will jump (up or down), particularly for medical conditions whose treatments have undergone rapid technological change. The panel recommends that BLS explore the possibility that these breaks may be large. A research program could retrospectively estimate the magnitude of such changes and determine what should be done (for example, smoothing the quantity changes based on past trends) if the breaks are expected to be large. At this point, we are agnostic concerning the most appropriate procedures and recommend that the BLS form a study group to examine these issues.

One point on which the panel is clear is that the Medical Care Price Index should not be linked in the usual way when the treatment episodes of diagnoses are repriced. That is, the index that results from the repricing of the 15 to 40 diagnoses should be linked to its previous value in levels, not in changes. The pricing of the set of diagnoses *is* the index. In other words, we advise that the index be the pricing of the treatments of the 15 to 40 diagnoses and that the index that results from the repricing of the components, holding the quantity weights constant, be used only for the high-frequency (monthly) publication requirement.

Thus, we advise that the BLS price treatment episodes for selected diagnoses, independent of the actual treatment components. This method generalizes what the BLS is already doing in the context of using DRG reimbursements in hospitals for pricing hospital services for Medicare patients.

Evaluating Quality Change

A closely related but conceptually distinct issue facing the BLS is how to deal with variation in medical outcomes, which are due in part—but only in part—to changes in the quality of medical care goods and services delivered. (This issue is a component of the more general problem of how to adjust prices for quality change, discussed in Chapter 4.) Some changes in medical treatment outcomes are more easily measured than others.

Consider changing mortality in the treatment of heart attacks. Cutler et al. (1998) have quantified mortality changes in the last few decades attributable to advances in the treatment of acute myocardial infarctions (heart attacks). Even if the entire change were due to higher-quality medical care, the measured improvement in life expectancy following heart attack treatment is not by itself sufficient to construct an outcomes-adjusted price index. Specifically, to calculate consumers' willingness to pay for these changed outcomes and thus to evaluate consumers' relative costs of reaching a given level of satisfaction over time, the change in life expectancy must be assigned some economic value. The marketplace does not provide direct evidence on the value that consumers place on improved life expectancy. Private- and public-sector decisions involving safety and compensation routinely place an implicit value on extending an individual's life by an additional year. But an official medical price index that required BLS to assign an economic value to a year of additional life would likely provoke considerable controversy, in part because it would highlight consumer heterogeneity, raising profound issues of how a price index should deal with heterogeneous consumers' tastes and incomes (see Chapter 8).

When change in mortality is not the most appropriate indicator of changes in medical outcomes, the conceptual and measurement issues for the construction of an MCPI become even more complex. For the treatment of cataracts, Shapiro et al. (2001) document enormous cost-reducing productivity improvements over time from the surgical use of lasers (see also Shapiro et al., 1999). Here the changed outcome is better eyesight, not extended mortality. But how does one place an economic value on additional years of better eyesight?

In other cases, medical advances have no significant effect on medical outcomes for the medical problem being treated but are superior from the patient's viewpoint because they involve fewer side effects. The introduction of nonsedating antihistamines has been significant, for example, not because the new generation of antihistamines is demonstrably more efficacious in reducing symptoms from allergies but because patients receiving the new antihistamines no longer have to face side effects from sedation—side effects that are important in an individual's ability to function and perform normal activities of daily living. In these cases, narrowly defined medical outcomes might not be changed at all, but the quality of life for patients who receive the new treatment is much improved. How is BLS to evaluate such quality-of-life improvements from new medications?

Health services researchers are currently attempting to measure the value of quality-of-life advances. They have constructed metrics such as QALYs (quality-adjusted life years) and DALYs (disability-adjusted life years). Although the conceptual foundations underlying such measures have been criticized, they are increasingly used within the health policy community. While the panel does not believe that BLS should use outcomes-adjusted price indexes based on QALY or DALY measures in its official price indexes in the near future, high priority should be given to research examining the feasibility and sensitivity of alternative ways of incorporating quality-of-life aspects of medical treatments, both at the BLS and elsewhere.

THE DOMAIN OF CONSUMER HEALTH EXPENDITURES: EMPLOYERS' HEALTH INSURANCE PAYMENTS

The CPI is designed to price goods and services for which consumers make direct out-of pocket-expenditures. Because this definition does not include the employee health insurance premiums paid by employers or government payments for medical services (Medicare Part A and Medicaid), the weight given medical care items in the CPI (slightly less than 6%) is much smaller than the share of medical care expenditures in total personal consumption expenditures (about 17.6%). Under certain circumstances, these weighting procedures can result in biases. One possible bias will occur if the health insurance payment part of the MCPI moves differently from other components of the index (e.g., if prices of insurance-covered services, such as treatments for cancer, moved differently from uncovered services, such as cosmetic surgeries or nonreimbursed purchases of nonprescription drugs). The likely direction of such a bias is a priori unclear.

Much more importantly, the panel concluded, is that including only patients' out-of-pocket outlays inappropriately understates the medical care portion in the total CPI. Arguably, this biases the CPI as a whole. In the last 75 years, the MCPI has generally risen more rapidly than the overall CPI. Thus, underweighting the share of MCPI in the total CPI has resulted in an understatement of the aggregate CPI, other things being equal (see Berndt et al., 2000).

The argument for including both components of health care expenditures—those paid by employees and those paid by the employer—in the MCPI (and the CPI)—rests on the assumption that employees, if given the after-tax cash equivalent of employers' contributions, would still spend most if not all of it on health insurance, paying a group rate for the insurance. Employers are generally indifferent as to how total labor compensation is split among wages and benefits, both of which are treated as (tax-deductible) expenses, although employers are obviously aware of the tax advantages for employees of employer-provided health benefits. Since the employer's portion of health care insurance is a benefit provided to employees, and since employees can, to some extent, choose their employers on the basis of the full compensation package (wages, salaries, and health insurance benefits), it makes sense to incorporate the employer portion of health insurance in the CPI and MCPI weights, rather than treating it as a business expense unrelated to employee compensation or consumers' expenditures (for further discussion, see Pauly, 1997; Summers, 1989; Gruber, 1994).

However, if employer health insurance payments were included in the MCPI and CPI, some current uses of the CPI (e.g., deflating earnings, measuring total compensation, escalating wages, escalating social security benefits) would need to be reexamined carefully. Even if employer health insurance payments were not incorporated into the MCPI and CPI, the usefulness and appropriateness of the CPI for measuring movements in the cost of living for subpopulations could be substantially affected by major policy changes, such as a prescription drug benefit for elderly Medicare beneficiaries. The problem of matching CPI design to particular uses is discussed in greater detail in Chapter 7.

Finally, because the CPI currently excludes from its domain government-provided goods and services, no weight is assigned for Medicaid or Medicare Part A services (hospitalizations and related physician services). An important consequence of this omission is that there is no overall price index for medical care—consisting of medical care expenditures from patients' out-of-pocket payments, private insurers, and government insurers (Medicare and Medicaid). To inform public policy discussion and to evaluate the performance of the U.S. medical care sector, a comprehensive MCPI is needed that encompasses purchases from all payers—governments, private third-party insurers, and consumers.

DIRECT PRICING OF HEALTH INSURANCE

As outlined above, BLS currently reallocates consumers' payments for private health insurance and Medicare Part B expenditures back to the five MCPI components. In this indirect pricing of health insurance, the BLS encounters significant difficulty in obtaining reliable separate price quotes for, say, physicians' services and prescription drugs when they are purchased directly by consumers as opposed to when they are paid by third-party insurers. In the medical care sector, the "law of one price" does not hold, making the BLS's task of obtaining reliable distinct price quotes a challenging one.

Recent improvements in information technology and medical care claims processing have greatly expanded the scope of calculations that can be carried out by health insurance actuaries. Stimulated by these information technology and actuarial developments, as well as by the practical difficulties in obtaining accurate and reliable price quotes from third-party payers, the BLS is now examining whether the indirect method can be replaced by direct pricing of health insurance. Such a change would eliminate the need for price quotes from third-party payers for their purchases of physician services, hospital services, and prescription drugs. Implicitly, such direct pricing of health insurance policies would be based on the reasonable assumption that prices paid by insurers for physician services, hospital services, and prescription drugs are reflected in the health insurance premiums they charge. Although the panel had not initially intended to consider direct pricing of medical insurance policies, the BLS raised this issue in its discussions with panel members, and so we address it here in a preliminary way.

Although direct pricing of health insurance could in principle obviate the need for obtaining hospital, physician, and prescription drug price quote information from insurers, the difficult issues of adjusting properly for changing quality and health outcomes remain. Direct pricing of health insurance policies does not address issues of adjusting medical care prices for quality changes. Those problems will be present whether direct or indirect pricing of health insurance takes place.

One major reason for year-to-year movement in health insurance premiums is that the mix of covered risks changes. This mix of covered risks reflects variation in the incidence of illnesses (e.g., AIDS or a flu epidemic), the age-sex mix of the insured population, and changes in the selectivity of coverage (i.e., the mix of individuals who have chosen a particular health insurance policy). The health economics literature is rife with examples of how moral hazard, adverse selection, and adverse retention affect costs (see, e.g., Altman et al., 1998). The presence of these phenomena significantly complicates the task of pricing health insurance.

In theory one could use hedonic pricing techniques (discussed in Chapter 4) to control for variations in health insurance benefits offered and for the age-sex mix of the covered population. In practice, it would be difficult if not impossible to adjust health insurance premium variations for changes in the health status of people who have selected to enroll in that plan. Whether actuarial procedures have improved sufficiently to be able to generate accurate quotes for a frozen or fixed population is unclear to the panel. Moreover, since the Laspeyres index methodology involves fixing product bundles at base-year levels, with base years changing say every 3-5 years, actuaries will likely find it very difficult to generate hypothetical quotes for old policies with obsolete fixed populations, given that insurers' policies and enrollees have changed. Even more problematic is the issue of how actuarial quotes would take into account the premium effects of variations in the health status of those selecting into (and leaving) a particular policy's enrollee population. Such data are typically not observable.

The panel appreciates the difficulties faced by the BLS in obtaining third-party price quotes of insurers' reimbursements for physician and hospital services and prescription drugs and in creating diagnostic-specific episode treatment price indexes. Replacing indirect pricing with the direct pricing of health insurance policies offers very attractive possibilities, and we believe that this direct pricing alternative merits close scrutiny. But we also strongly recommend that no change from indirect to direct pricing be made without extensive experimentation and reliability assessment that includes consultation with leading health economists, actuaries, clinicians, and health insurance specialists.

RECOMMENDATIONS

Pricing Diagnostic Treatments

The changes that the BLS implemented in 1998 regarding the MCPI reduce the bias caused by the aggregation of goods and services to levels that do not allow for substitution among alternative medical treatments. BLS should continue its efforts to eliminate this bias. The major change BLS implemented in 1998 concerned aggregating inpatient and outpatient hospital services into a single stratum, to allow for substitution between them, but that is only one area of potential bias from input substitution.

The panel favors the use of diagnosis-based rather than input-based measures wherever this is feasible. The advantages of diagnosis-based methods have been highlighted in a number of research papers, of which the treatment for depression is only one example. Thus the panel recommends using a disease- or diagnosis-based unit for pricing rather than the current "industry" or medical care strata. This recommendation does *not* imply (or require) outcome-based measures. It implies only that inputs be priced and aggregated by the changing treatment mix for a particular diagnosis, rather than by the traditional BLS medical care strata.

Recommendation 6-1: BLS should select about 15 to 40 diagnoses from the *ICD (International Classification of Diseases)*, chosen randomly in proportion to their direct medical treatment expenditures and use information from retrospective claims databases to identify and quantify the inputs used in their treatment and to estimate their cost. On a monthly basis, the BLS could reprice the current set of

specific items (e.g., anesthesia, surgery, medications), keeping quantity weights temporarily fixed. Then, at appropriate intervals, perhaps every year or two, the BLS should reconstruct the medical care index by pricing the treatment episodes of the 15 to 40 diagnoses—including the effects of changed inputs on the overall cost of those treatments. The frequency with which these diagnosis adjustments should be made will depend in part on the cost to BLS of doing so. The resulting MCPI price indexes should initially be published on an experimental basis. The panel also recommends that the BLS appoint a study group to consider, among other things, the possibility that the index will "jump" at the linkage points and whether a prospective smoothing technique should be used.

Including Employers' Health Insurance Payments

Currently, CPI weights only reflect consumers' out-of-pocket expenditures on medical care, including the cost of insurance. As discussed above, this leads to an underestimation of the medical care category of the CPI. Given the uses to which the flagship CPI is put, it is not obvious that expanding the medical care category to include more types of expenditures is advisable. However, a more broadly based measure of the changing cost of medical would be valuable for a wide range of policy purposes.

Recommendation 6-2: BLS should include the portion of health insurance paid for by employers in one version of the CPI, perhaps calling it an "expanded-scope medical CPI." Because many commonly used income measures exclude employer-provided benefits and because the Consumer Expenditure Survey is based only on out-of-pocket expenditures, the original conception of the MCPI domain should still be maintained in constructing the traditional (flagship) CPI. The panel also recommends examining the practicality of including other employer-paid employee benefits (e.g., dental and life insurance and cafeteria plans) in the expanded-scope CPI.

Expanding the Domain of the MCPI

To inform public policy discussions and to evaluate the performance of the U.S. health care sector, a medical care price index that encompasses purchases from all payers—governments, private third-party insurers, and consumers—is needed. Whether such an index is best constructed by the BLS, the Bureau of Economic Analysis, some other government agency such as the Centers for Medicare and Medicaid Services, or a nongovernmental organization is unclear. What is clear is that a medical care price index encompassing *all* payers for medical services should be computed and published for public consumption.

> **Recommendation 6-3: A task force should be convened by the BLS, in collaboration with the Centers for Medicare and Medicaid Services and other appropriate agencies, to implement construction and publication of a total medical care expenditure price index, encompassing purchases from all health care payers—governments, private third-party insurers, and consumers.**

Pricing Outcomes Rather than Inputs

The most difficult issue in the construction of the MCPI concerns adjustments for quality change. New treatments extend life and make the quality of life better. The panel believes that an outcomes-based measure is in principle superior to an input-based measure, but we recognize the formidable measurement challenges and do not know how best to proceed. This area is new and requires considerably more research, much of it interdisciplinary. After BLS has produced the experimental expanded-scope MCPI recommended above, BLS can consider whether, how, and why the outcomes of the treatments for those diagnoses are changing over time and finally consider how outcome changes should best be evaluated in computing a quality-adjusted medical care price index.

7

Index Design and Index Purpose

As we have seen, a host of difficult questions must be answered in designing a price index. How some of these should be resolved may depend on the particular uses to which an index is put. There is a tension between the goals of providing indexes tailored to specific purposes of public policy or for particular kinds of economic information and avoiding public confusion that might result from too many indexes.

Subject to considerations of costs, feasibility, and reliability, the publication of several overall indexes may well be warranted. For instance, we concur with the decision by the Bureau of Labor Statistics (BLS) to produce a lagged superlative index in addition to the "flagship" Consumer Price Index (CPI); we also suggest publication of an advance forecast of the superlative for compensation purposes (see below and Chapter 2). Yet there are inherent limitations in trying to produce an exact match between the design of official price indexes and the particular purposes for which they are used. Moreover, the media will inevitably emphasize one or, at most, a few indexes.[1] But whether there are few or many indexes available, it is important that public policy makers and private users understand the relationships between the ways in which price indexes are designed and how they serve—or fail to serve—the various purposes for which they are employed. In this chapter we attempt to clarify some of those relationships.

[1]Research and experimental indexes, explicitly labeled as such, have been and can continue to be useful, need not be limited in number, and are less likely to create problems of public perception.

There are many different uses for aggregate indexes of prices and the cost of living, most prominently:

- as a *compensation* measure to calculate how much is needed to reimburse recipients of social security and other public transfer payments against changes in the cost of living and for formal or informal use in wage setting;
- for inflation indexation in *private contracts*;
- as a measure of inflation for *inflation-indexed Treasury bonds*;
- as a measure with which to *index the income tax system* to keep it inflation neutral;
- as an *output deflator* for separating changes in gross domestic product (GDP) and its components into changes in prices and changes in real output; and
- as an *inflation yardstick* for the Federal Reserve and other macroeconomic policy makers.

This chapter examines the application of indexes in each of these contexts.

INDEXING PUBLIC TRANSFER PAYMENTS

The CPI is widely used within government and among private parties as a means of maintaining the purchasing power of a flow of transfer payments in the face of changes in prices, sometimes specifically identified as changes in the cost of living. In a similar vein, the CPI is used to adjust eligibility limits for certain kinds of payments, usually to the poor, that were initially set in nominal dollar terms.

One overarching conceptual issue that arises when cost-of-living adjustments are provided in public transfer payments is whether the adjustments should compensate recipients only for changes in the overall cost of living for the nation as a whole or should take account of any significant differences among particular groups and individuals in society. Even if it were possible to calculate a separate index for every individual, public policy would surely not seek to provide adjustments tailored to each. Indeed, taken literally, this approach would provide incentives for individuals not to substitute away from goods whose prices had risen the most (the government transfer payment would provide the means for an individual to maintain his or her consumption of expensive wines, for example, even if their prices skyrocketed).[2] However, if the goal of public policy is to ensure recipients of various public transfer programs—e.g., the poor and the elderly—against changes in the cost of living, and if cost-of-living indexes for the affected group differ systematically or frequently from the aggregate CPI, then Congress

[2]Of course, this does not preclude the use of democratic cost-of-living indexes, which are averages of the indexes of individual consumers or households.

might wish to consider using a separate index for the affected group as the measure of compensation.

There is no abstract criterion that can be used to determine the extent to which compensation should be tailored to particular subcategories of individuals within each broad group of targeted transfer recipients. In general, the panel believes that, if a particular category of individuals is not itself the target of a transfer program, special indexes based on detailed distinctions within the targeted groups are not a suitable basis for making cost-of-living adjustments. In the sections below, we discuss the potential use of special indexes for making cost-of-living adjustments for the elderly and the poor.

Adjusting Social Security Benefits

The most prominent public policy use of the CPI is for indexing benefits paid to social security retirees. Prior to 1972 Congress had periodically legislated increases in social security benefits, usually by more than enough to cover changes in inflation (as measured by the CPI) since the last increase. In 1969 the Nixon Administration announced its support for automatic indexing of benefits, declaring: "The way to prevent future unfairness is to attach the benefit schedule to the cost of living" (cited in Berkowitz, 1986:48). In 1972 this recommendation became law (along with a 20 percent one-time increase in benefits). Congress explicitly provided for annual cost-of-living-increases for people receiving benefit payments, based on changes in the CPI-W.[3] The CPI is also used for the same objective in a number of other federal programs that provide transfer payments: for the military and civil service retirement systems, the railroad retirement system, veterans' pensions, and Supplemental Security Income.

When it established an indexing procedure, Congress stipulated that benefits be adjusted to offset changes in the cost of living (rather than simply to maintain the purchasing power over a fixed basket of goods) and specified that this should be done through the use of the CPI. At the time, members of Congress appeared to have accepted the widely held presumption that the CPI measures the cost of living. There is no reason to believe they explicitly considered any distinction between a fixed-weight and a cost-of-living index.

The panel was not charged with recommending to Congress what specific objectives it ought to pursue in indexing social security and other benefits. We were charged to make clear the implications for public policy that flow from choosing one or another scheme of indexing and to spell out the consequences for public policy from alternative choices of index design. Moreover, even assuming

---

[3]Various modifications have been made in the timing and details of the adjustment, but the cost-of-living terminology remains. No provision is made for decreases in benefits when the CPI declines (P.L. 92-336).

that public policy seeks to tie benefits to changes in the cost of living, a number of choices still remain about whether the present CPI (together with changes that are planned for the near future) or some alternative version is the appropriate one—for example, a separate index for the elderly.

Using a Lagged Superlative Index for Escalation

We conclude that, for adjusting benefits to keep pace with the cost of living, the superlative index, which the BLS will begin to publish in 2002, will be the appropriate one to use. But that index will be available only after a 2-year delay. One way to deal with this lag would be to pay an initial cost-of-living adjustment based on the change in the fixed-weight CPI and then incorporate a correction 2 years later based on the lagged superlative. Thus, the cost-of-living adjustment (COLA) for 2004 would increase the benefit payment by the change in the regular fixed-weight CPI for the past year minus a correction for the difference between the regular CPI for 2002 and the superlative index for that year, which would just have been published. If recent history is a guide, the superlative index will show an average increase of about 0.1 to 0.2 percent a year less than the real-time CPI, with a range of 0.0 to 0.5 percentage points (Aizcorbe and Jackman, 1993; Shapiro and Wilcox, 1997). A small initial "claw back" would be required; thereafter, the COLA would tend to be very close to the change shown by the current year's real-time CPI minus an adjustment that typically fluctuated within a narrow range.

There is an alternative that may offer some advantages and which the recommendation below would facilitate:

> **Recommendation 7-1: The BLS should publish, contemporaneous with the real-time CPI, an advance estimate of the superlative index, utilizing either a constant-elasticity-of-substitution method or some other technique.**

There are several possible ways to construct that estimate. Individual researchers have constructed estimates of the superlative using only reference-period weights and a constant-elasticity-of-substitution formula, whose major parameter was estimated from a comparison of the real-time CPI with the superlative in prior years. Such estimates, when tested over short periods of time, have very closely tracked the superlative that later became available, although particular patterns of substantial changes in relative prices could produce larger divergences. Alternatively, the BLS could utilize other techniques for making an advance estimate, perhaps taking advantage of the latest information on relative price changes in the real-time CPI.

For purposes of escalation, the panel arrived at the following:

> **Conclusion 7-1: It would be feasible and appropriate to calculate cost-of-living allowances provided for social security and other pro-**

grams from an advance estimate of the BLS's soon-to-be-published superlative index. Any divergence between that estimate and the superlative that appears 2 years later could be incorporated as a correction to the cost-of-living allowance provided for that year.

For example, the 2004 COLA would be based on the advance estimate of the superlative for the prior year plus or minus the difference between that estimate of the superlative for 2002 and the actual value of the superlative for that year. Independent of how the advance estimate of the superlative is arrived at, the panel supports its use.

This alternative offers both an advantage and a disadvantage when compared with the 2-year delay approach. There is a high probability that the later corrections to the initial COLA adjustment would be a good bit smaller if the advance estimate is used. But it would add an additional and hard-to-explain complexity to the index used for the initial adjustment.

Conclusion 7-2: On balance, the advantage of having much smaller corrections outweighs the disadvantage of the additional complexity.

Compensating Beneficiaries Who Have Other Income

Many social security retirees have other income, in some cases substantially exceeding their social security benefits. The broad objective of Congress in providing a cost-of-living adjustment was and is to protect the social security income of beneficiaries, and not their other income, against the consequences of price changes. In Chapter 2 we point out that it is not obvious how to design an index that holds constant the standard of living of social security recipients who have other income. Perhaps the simplest way of dealing with the problem is to define the index as one that provides the compensation needed to maintain living standards for those whose only income is their social security benefit.

The Role of Taxes

In its most usual formulation, a cost-of-living index provides a measure of the percentage change in expenditures a consumer would have to make to maintain a specified standard of living in the face of changes in the prices paid for goods and services. This is an expenditure COLI. An alternative approach is to measure the percentage change in the income a consumer would need to maintain that same standard of living as prices and income and payroll tax rates change. (Indirect taxes, such as sales and value-added taxes, are already included in the prices of private goods and services.) Such an index has sometimes been labeled a tax and price index (TPI; see Gillingham and Greenlees, 1987, 1990). Simply providing an additional amount of income sufficient to pay the higher prices, as would be the case under an expenditure COLI, would not be enough if that income were subject to higher tax rates. A TPI would include the effect of higher income and payroll tax rates and, if used for compensation purposes, would therefore provide enough compensation to cover both the higher taxes and the higher prices. (The domain of a TPI is restricted to private goods and services; it does not impute to the consumer's living standard any value from public goods financed by taxes.)

What are the implications of tying social security payments to a tax-and-price index rather than the current expenditure-based index, assuming that a single overall CPI would continue to be used for indexing purposes? First, social security retirees pay no payroll taxes; and in the federal and many state income tax systems social security benefits are more lightly taxed than other forms of income, at least for low- and lower-middle-income taxpayers. It is likely that legislative changes in tax rates would retain the same sort of preferences for the elderly. A TPI would, therefore, be likely to overadjust social security benefits when tax rates generally are raised and have the opposite results when taxes are cut. If Congress wishes to change the after-tax benefits for social security recipients, a much fairer and more effective way to do so is through explicit changes in the tax code or benefits formula.

Second, the use of an index that reflects changes in tax rates would be inequitable among social security recipients themselves. Legislated changes in income tax rates often vary among people with different incomes and in different economic circumstances, while an overall income COLI would reflect only an average of the rate changes: some beneficiaries would be overcompensated and others undercompensated.

Conclusion 7-3: For purposes of indexing social security and other benefits, shifting from the current expenditure-based CPI to a tax-and-price index that reflected changes in income and payroll tax rates would pose some difficult measurement problems and create unintended distributional inequities.

A Separate Index for the Elderly?

To the extent that prices for goods and services paid by the elderly rise at a different rate than those paid by the population generally, Congress might consider tying retirement benefits to a special index for the elderly. At the request of Congress, BLS developed a special experimental index for the elderly (CPI-E) in which the prices of the 200-plus categories of goods and services in the regularly published CPI-U were reweighted to reflect the consumption patterns of the elderly. From 1984 to 1995 the experimental index rose by an average of 0.4 percent per year faster than the CPI-W, which is used to index social security benefits, and by 0.3 percent per year faster than the CPI-U. Several outside studies covering earlier periods found little difference in the cost-of-living changes faced by the elderly and the general population (see Boskin and Hurd, 1985; Jorgenson and Slesnick, 1983).

These results however do not necessarily mean that the elderly faced only slightly more rapid increases in living costs than the population as a whole during those years. Then, or at other times, the "true" differences could be larger and might be either positive or negative. In the BLS studies, for example, the heavier weight attached to out-of-pocket medical care expenses among the elderly accounts for the majority of the difference between the CPI-E and the CPI-U or CPI-W. It is widely believed that, quite apart from controversial issues of quality adjustment, the measurement of medical care prices in the CPI during the periods studied overstated their rise (Newhouse, 2001). BLS has recently improved the definition of the medical care services it prices, and a recent study has shown that changes of this nature significantly reduce the rise of medical care prices (Newhouse, 2001). In addition, the CPI-E did not capture any possible influence coming from changes in the magnitude and scope of senior citizen discounts.

The major problem, however, lies in the fact that the elderly may, on average, buy different varieties of goods within many CPI strata, face different prices, and shop at different outlets than younger consumers. Unfortunately, as we explain in Chapter 8, there is no database that allows adequate exploration of this facet of behavior and its consequences, for the elderly as well as for the poor or other demographic groups. The special CPI-E index and the other studies cited all used the same prices for the same goods at the same outlets as were priced for the CPI-U and simply reweighted them to reflect the budget allocations of the elderly among large expenditure categories. That is one of the reasons we recommend a long-range BLS research program that would explore the use of innovative techniques (e.g., scanner technology integrated with consumer surveys) to examine this issue.

Conclusion 7-4: In the absence of an index that can capture the differences in the prices or qualities of goods purchased by the elderly, we see no rationale for switching to an index along the lines of the CPI-E for purposes of indexation. However, BLS should periodically update the CPI-E to make sure that significant differences are not developing between it and the CPI-W and the CPI-U.

The CPI-U Versus the CPI-W for Indexing Transfer Payments

In 1978 the CPI was revised in a major way, including an expansion of its coverage from "urban wage earners and clerical workers" (one-third of the population) to "all urban consumers" (four-fifths of the population). The new index was christened the CPI-U, which is now widely accepted as the main or flagship index of consumer prices. The original CPI was renamed the CPI-W and continues to be used as the index for determining annual cost-of-living adjustments in social security and other programs. The weights of the two indexes differ, but only modestly. And to the extent the weighting structures do differ, the CPI-U weighting structure is closer to that of the elderly.[4]

Over the past 20 years, the average difference between the growth of the two indexes has been small—the CPI-U grew 0.13 percent faster, although there was substantially greater divergence over shorter time periods. Over the past 5 years, however, the difference between the growth of the two indexes averaged only 0.04 percent. The differences do not seem to be large enough to warrant switching the index used for indexing social security benefits to the CPI-U. However, if the instrument used for adjusting social security benefits is changed to the superlative index, as we recommend, that index should be based on CPI-U rather than CPI-W weights.

Plutocratic Versus Democratic Indexes

The current CPI is a plutocratic index. In constructing the national index each individual good is assigned a weight equal to overall consumer expenditures on that item. This procedure assigns to the spending pattern of each individual household an importance in the overall index proportional to its total consumption expenditures; the spending patterns and preferences of the rich count more than those of the poor. A democratic CPI would be one in which each household's spending pattern received equal weight.

Arguably, a plutocratic index may be the appropriate choice for an overall indicator of inflation in consumer goods. And since, in the construction of measures of national output, the individual strata indexes of the CPI are used to deflate most of the components of consumption expenditures, a plutocratic version of those individual indexes is needed. But for purposes of indexing social security benefits and other public transfer payments and for dealing with economic welfare considerations generally, a democratically constructed index seems clearly preferable since it assigns the preferences of each household equal importance.

[4]When the goods in the CPI are grouped into nine broad categories, the CPI-U comes closer to the CPI-E in five cases, the CPI-W is closer in one case, and three are about equal. The median income of the population covered by the CPI-W should be closer to that of the median elderly household than to the median for the population covered by the CPI-U. If the individual items and shopping outlets whose prices enter the CPI-W were specifically chosen to reflect the purchases and shopping patterns of the wage-earning population, that index might be superior to the CPI-U for indexing social security benefits. But as we have repeatedly pointed out, the same selection of individual items and outlets goes into all the price indexes—CPI-U as well as CPI-W. The indexes differ only through the variations in their upper-level weights.

In Chapter 8 we note that simply switching to a democratic set of weights at the upper level of CPI aggregation would probably produce only modest differences in the behavior of the index. That was the result of a research index for the elderly constructed along those lines by the BLS for the years 1989 to 1997, as well as of other studies for various earlier periods (see Technical Note 2 in Chapter 8 for a brief summary of these studies). To the extent that systematic and important differences do exist, they must arise at the very detailed level of price collection. This possibility provides yet another reason for the research project we have suggested for investigating differences among households at that level.

Conclusion 7-5: To the extent that the evidence ultimately suggests significant differences between democratic and plutocratic indexes and demonstrates the feasibility of producing demographic indexes, the case for switching to such an index for compensation purposes is a persuasive one.

Indexing Social Security Benefits to Wages Rather than Prices

Some people have argued that social security benefits should be tied to a wage index (see, for example, Griliches, 1995). To the extent that the CPI is a reasonably good approximation to a cost-of-living index, a monthly retirement benefit adjusted for CPI inflation will "buy," over the remaining life of the retiree, the same standard of living as it did on the date of retirement. The retiree receives protection against inflation and is insured against the economic vicissitudes that can erode the real wages and living standards of the working population—such as a large surge in energy prices or a major depreciation of the real exchange rate. With wage indexing, a worker or retiree gives up those protections but gains the advantage of sharing in the fruits of future national productivity growth and any other economic developments that improve real wages.

The choice between wage and price indexing thus involves issues of broad public policy with respect to the distribution of income between social security beneficiaries and the rest of the population. The panel was not charged with providing advice or recommendations on this distributional issue, but we were explicitly asked to assess "the appropriate uses of [cost-of-living] indexes for indexing federal programs and other purposes." While a wage index is not a COLI, the panel believes that a comparison of wage and price indexation for social security recipients, in the light of various criteria of interest to policy makers, can significantly help illuminate some of the public policy choices implicit in the selection of an indexing instrument.

Alternative Wage Indexes

There are many possible wage indexes that might be employed for indexing purposes. They do not all move parallel with each other, even over long periods of time. Which one to adopt is far from obvious, and the choice hinges on important conceptual issues. The specific consequences of adopting wage indexing, both for retirees and the federal budget, would depend on the nature of the wage measure that was chosen and how that choice interacted with future economic developments. The initial choice of a wage index could itself be quite controversial, and the debate might recur in light of subsequent developments. However, subject to one qualification we discuss later, the wage measure could be estimated and social security benefits indexed without having to deal with vexing and often contentious issues involving the effects of substitution among goods and outlets, pervasive quality changes, and the introduction of new goods with which we wrestle throughout this report. We first describe the major conceptual alternatives for a wage index and then briefly summarize the pros and cons and the implications of using each of them.

The social security system already employs an aggregate national wage index as part of the process of calculating each retiree's initial retirement benefit. That benefit is based on the annual earnings of the retiree in each past working year, indexed up to the date of retirement using an economy-wide "average annual wage."[5] The average is the mean of the annual money wages of employees as reported by all employers on their W-2 forms. It includes wages in excess of the social security taxable maximum but excludes nonmonetary fringe benefits.[6] This average is also used to index the bend points in the benefits formula and the maximum earnings subject to tax. In implementing wage indexation, Congress might well decide that the same measure used to index preretirement wages for calculating the initial benefit ought simply to be extended through the retirement years, as the index for maintaining the relationship between postretirement benefits and the real wages of the working population.

However, this choice is not the only one for which a reasonable argument might be advanced. If the broad policy objective is to have the postretirement benefit rise or fall in line with the economic fortunes of the working population, a number of issues arise about the nature of the wage index that would do that appropriately. Perhaps the most important type of choice among alternative wage indexes involves an aggregation question that bears some resemblance to the issue of plutocratic versus democratic prices indexes: Should the index reflect the change in the mean wage or some other point in the wage distribution, the most likely candidate being the median wage? From the late 1970s to the mid-1990s,

[5]Thus a worker's earnings relative to other workers throughout her working life will determine the relative wage that enters into calculating the initial retirement benefit, but the absolute size of the benefit will also depend upon how fast average social security earnings have grown over time. (The formula that is applied to the resultant earnings measure to calculate benefits is itself highly redistributive.)

[6]Some workers are not covered by the social security system, but an estimate of their wages is included in calculating the average social security wage.

TABLE 7-1 Average Annual Real "Wage" Increases, 1980-2000 (percent)

	Means			
Period	ECI Hourly Compensation	ECI Hourly Wage	Social Security Annual Wage[a]	Median CPS Wage
1960-1980	n.a.	n.a.	1.0	1.5
1960-1973	n.a.	n.a.	2.0	2.5
1973-1980	n.a.	n.a.	–0.7	–0.3
1980-2000	0.9	0.6	1.2	0.3

NOTE: Nominal indexes deflated by the CPI-U (research series).

[a]Data through 1999.

SOURCE: ECI data from BLS web site. Social security annual wages from U.S. Social Security Administration (2000). For CPS hourly wage, see text footnote 8.

wage inequality in the United States increased substantially. As a consequence, median hourly wages rose much more slowly than their mean.

A second question concerns the scope of a wage index: Should it measure changes in the overall compensation of workers, including not only wages but fringe benefits, or should it cover wages only? Over most of the past 50 years, fringe benefits—chiefly, employer-paid pensions and health insurance costs—rose more rapidly than wages, so that the growth in real compensation per hour exceeded the growth in wages by about 0.25 percent a year. Since 1980, the BLS has published a quarterly Employment Cost Index (ECI) for total hourly compensation (including fringe benefits) and one for wages alone.[7] Between 1994 and 1999 the excess growth of fringe benefits relative to wages was reversed as growth in the cost of medical care and the generosity of employer health care plans were reduced. But in 2000 fringe benefits once again began to grow more rapidly. The future relationship between the two components of employee compensation is likely to depend importantly on what happens to health care costs.

Table 7-1 shows for various periods the average annual growth in the real wage (or compensation) for each of the four concepts described above: the two published ECIs for hourly compensation and for wages (which are not available for the early years); the average annual social security wage; and the median

[7]Here the term "wages" includes wages and salaries. The ECI defines wages as straight-time wages per hour worked. Paid leave and premium pay are included in fringe benefits. The ECI collects data from employers on a probability sample of compensation components for specific occupations. The data are combined into an overall index with fixed employment weights by occupation and industry.

wage derived from the Current Population Survey (CPS).[8] Alternative choices among these indexes would have produced substantially different outcomes for social security recipients over the different periods. During the period of substantial productivity growth from 1960 through 1973, both the social security and the CPS wage indexes rose strongly, followed by an actual reduction in the seven years following the 1973 growth slowdown. Over the past twenty years rising earnings inequality kept median wage growth close to zero. And, on average over those years, ECI compensation growth exceeded ECI wage growth as fringe benefits rose more rapidly than wages during most of the period.

Some of the apparent anomalies in the data can be explained by differences in their construction. From 1973 on the CPS median wage refers to hourly not annual wages (see footnote 8). During the 1973-80 period, annual hours of work declined, explaining at least in part why the social security wage fell more than the CPS wage. Over the 1980-2000 period, compensation growth exceeded wage growth, but the social security wage, which excludes fringe benefits, rose faster than ECI compensation, which includes them. Hours of work were roughly stable over this period. Some of the difference apparently stems from the fact that the average social security wage is an unweighted measure whose change from year to year includes not only changes in wage rates but also the gradual upgrading of the workforce into higher-skilled and higher-paid occupations. The ECI uses fixed occupation weights to combine individual wage rates and so therefore does not reflect this aspect of the upgrading phenomenon.[9]

The choice among alternative index concepts involves matters of broad public policy. Conceptually, a median wage index might be seen as more suitable than a mean index in the context of the social security system. Social security is essentially a safety net for low- and middle-income workers. Its distribution formula is highly redistributive. Under wage indexing, therefore, it would seem natural to allow beneficiaries to share in the gains and risks of economic develop-

---

[8]Estimates of the median CPS hourly wage are contained in Mishel et al. (2001); *http://www.epinet.org/datazone/dznational.html* for the underlying data. For a description of the methodology to derive the series, see Mishel et al. (2001:App.B). In Table 7.1 (above), the nominal wage data have been deflated by the CPI-U-RS rather than the CPI-U-X1, which was used for the published estimates. The Mishel data were only available since 1973 and were extended backward with an average of the median annual earnings of male and female year-round workers from published CPS data.

[9]There are other compensation and wage indexes available. One of them, the monthly "average hourly earnings index for production and non-supervisory workers," is an outlier. Even though it is a mean not a median index, it rose over the 1980-2000 period by even less than the median CPS—by an average of only 0.2 percent when deflated by the CPI-U (research series). But it is a flawed measure. According to one study that compared alternative wage indexes, "among measures of wages and compensation the hourly earnings index should be dismissed because it is based on sampling techniques that are not properly benchmarked." Indeed, it is not benchmarked at all (see Bosworth and Perry, 1994).

ments as they are reflected in the real earnings of the median employed worker. Retirees would not, in this case, benefit from the above-average gains of upper-income workers during a period of increasing income inequality.[10] Yet the social security benefit formula has always indexed retirees' past wages on the annual social security wage, which is a mean not a median.

The choice of whether the universe of the index should include fringe benefits would make a difference only to the extent that fringe benefits rose faster or slower than money wages. The principal components of fringe benefits are employer contributions for worker retirement and health insurance. Current retirees now have access to their pension benefits, and all are covered by Medicare (some also have employer-paid Medigap policies). Thus, it might seem desirable to index their benefits to the change in the cash wages of current workers, excluding fringe benefits. But the issue is not open and shut.

Most of the rise in fringe benefits relative to wages in the decade prior to 1994, and the decline in that ratio from 1995 to 1999, stemmed from changes in employer contributions for health insurance. Despite the fact that the elderly are covered by Medicare, their out-of-pocket costs for medical care equal, and possibly exceed, the sum of employer contributions and out-of-pocket costs for the rest of the population.[11] Workers receive some of the fruits of national productivity growth in the form of increased employer payments for medical care. If the intent of wage indexing is to have retirees match the economic gains (and share the risks) of workers, it is not obvious that those particular payments should be excluded from the indexing instrument just because they were dedicated to paying for medical care, given the fact that the elderly, despite Medicare, still face large out-of-pocket medical expenditures.

For the other major component of fringe benefits—employer-financed pensions—there are two issues. Fringe benefits, as measured in the CPI and the National Income and Product Accounts (NIPA), include not only employer' payments for private pensions but also their federal payroll taxes. With respect to payroll taxes, social security has always been, and still is, mainly a pay-as-you-go

---

[10] In addition, if the indexing instrument is a mean wage measure, any additional benefits that retirees received through the effect of rising inequality in raising the mean relative to the median wage would sooner or later—in a pay-as-you-go system—have to be paid for by payroll taxes, which, taken by themselves, bear disproportionately on lower-income workers. Indexing social security benefits to a measure of mean wages during periods of rising inequality would, albeit in a modest way, accentuate the effects of the rising wage inequality.

[11] In connection with its estimation of the CPI-E, BLS provided data on the relative value weights for out-of-pocket medical care expenditures for the CPI-U, the CPI-W, and the CPI-E. A rough estimate of the relative value of out-of-pocket expenditures by the non-elderly population was estimated on the basis of the ratio of elderly to non-elderly consumer units from the Consumer Expenditure Survey. An estimate of the weight for employer-paid health premiums was obtained from data from the National Income and Product Account as the ratio of employer payments for health insurance divided by total consumer expenditures.

system, although it is now running a modest surplus. There seems to be no justification for using an increase in the payroll tax rate, most of which goes to supporting the level of benefits for current retirees, as a basis to raise those benefits.

Private employer retirement systems, in contrast, are heavily if not fully funded. Changes in compensation received in the form of employer-paid retirement contributions change the lifetime standard of living of currently employed workers. One could argue that if retirees are to share in the gains of productivity growth accruing to the current workforce, they should not be deprived of part of those gains simply because workers are taking some of the gains in the form of an increase in their own future welfare. In that view, the fact that today's retirees now have access to employer pension contributions made in prior years seems irrelevant.

Whatever the conceptual arguments about fringe benefits, the U.S. social security system has for many years based both the payroll taxes that support it and the calculation of initial benefits on money wages, excluding the value of fringe benefits. It would seem anomalous to leave these aspects of the system unchanged while tying benefits in the years after retirement to a wage index that includes fringe benefits.

Another question is whether the index should reflect changes over time in the average "quality" of the labor force, principally evidenced by changes in average educational attainment and years of experience. Without some explicit adjustment, both a median wage index and a mean wage index would reflect not only changes in the real wages of workers with given educational accomplishments and given experience but also the effects of a gradually changing composition of the workforce. If the real wages of workers with given personal characteristics remain unchanged, an unadjusted wage index would nevertheless rise or fall as the net effect of the outflow of retiring workers and the inflow of new workers gradually changed the quality of the labor force. The answer to this question depends on how one specifies the objective of the indexing: keeping the retiree population abreast of the living standards of the working population as a whole or abreast of the living standards of workers with a fixed amount of education and experience. Adjusting for quality change could make a significant difference. In its ongoing program of estimating multifactor productivity growth, BLS calculates an index of the change in labor quality resulting from shifts in the mix of the labor force according to education, age (a proxy for work experience), and gender. This index shows that changes in workforce composition have produced, over the past 40 years, an average growth in quality amounting to 0.3 percent a year.

As an indexing instrument the average social security wage would have the advantage that it is already used to index a number of aspects of both social security taxes and benefits, including the determination of the initial benefit and the maximum wage subject to tax. And since it is directly calculated from the aggregate payrolls on which the tax is based, indexing postretirement benefits to this measure would mean that the growth of payroll tax receipts would rise or fall in parallel with variations in the growth of those benefits, providing an element of automatic self-financing (in a pay-as-you-go system). It has a number of other characteristics that might or might not be judged desirable: it is a mean index; it excludes fringe benefits; as a measure of annual not hourly wages, it would exclude any gains from productivity growth taken in the form of increased leisure time; and it would include in the wage index the quality component of any change in real earnings.

Making recommendations about the issues involved in choosing the appropriate measure to use for wage indexing is not within the charge of the panel. We have discussed the subject only to demonstrate the kind of decisions that would have to be made in selecting an appropriate wage index.

In sum, from the standpoint of beneficiaries' welfare, wage indexing would have quite different properties than CPI indexing. The latter, albeit imperfectly, tends to fix the purchasing power of a benefit (or its contribution to a recipient's standard of living) as it existed at the time of retirement and insures beneficiaries against the consequences of the ups and downs of macroeconomic supply shocks. Wage indexing, in contrast, has the potential for letting retirees share in the fruits of productivity growth in the nation's consumer goods industries. The actual outcome would depend chiefly on the course of economic developments and to some extent on the interaction between those developments and the choice of the particular wage measure to be used for indexing purposes.

The potential macroeconomic implications of wage indexing should be mentioned. Wage indexing would have the advantage of automatically relating the path of postretirement benefits to the inflow of the payroll taxes that financed them (in a pay-as-you-go system). As we noted, if the average social security wage were used as the indexing instrument, there would be a close correspondence between the rise in the benefits and the rise in the tax base.[12] Yet to the extent that over the long run real wages tended to rise, the overall level of benefits would increase relative to what would be generated by the current benefit formula together with price indexing. Should long-term productivity growth and the trend toward wage inequality revert to 1973-1995 patterns while a median wage measure was used for indexation, the switch to wage indexing would provide benefit adjustments that, on average over the years, would be unlikely to differ very much from those derived by CPI indexation. However, while far from certain, the odds seem a good bit better than even that over the long term the real wages of the median American worker will see some rise. The 2001 report of the social security trustees assumes a 1.0 percent rise over the long term in the real

[12]This takes into account the fact that the size of the initial benefits is also determined by the growth of the average social security wage.

social security covered wage, virtually equal to the average increase over the past 40 years (Board of Trustees of the Federal Old Age Survivors and Disability Trust Funds, 2001:82).[13]

A switch to wage indexing would increase the long-term deficit in the social security system above the large amount that is already projected. This increased cost could be neutralized by reducing the initial replacement rate by an amount that would maintain the projected present value of benefits that would have occurred under price indexing assuming, for example, the long-term growth in real wages projected by the Trustees at the time the switch to wage indexing was made. The path of benefits would be "tilted," with a smaller initial benefit but one that grew more rapidly as retirees aged. An alternative approach would be to subtract from the annual growth of the wage index a legislatively fixed amount that would keep the present value of the time path of benefits equal to what it would have been under price indexing, again using the real-wage growth projected in the Trustees' report at the time of the switch. This approach would preserve many of the basic characteristics of wage indexing without adding to the projected cost of the system.

The issues that would be involved in the use of wage indexing for social security COLAs touch on matters of broad national policy with respect to income distribution and financing, as well as technical issues related to the choice of an appropriate wage index. Given the long-term financial prospects for the social security system, some important changes and reforms in its financing and structure will become inevitable. In the process of considering those changes, one of the options that could be evaluated and considered is a switch to wage indexing for postretirement benefits.

Indexing Other Federal Programs

The Supplemental Security Income (SSI) program provides benefits to the poor elderly, blind, and disabled and indexes them with the CPI-W. A number of other federal programs that provide in-kind transfer payments to low-income households also use the CPI to index the income eligibility limits for program participation. Many of the same issues raised in the discussion of retirement pensions are relevant for these programs, and a number of them were explicitly addressed in a National Research Council (1995) report on measuring poverty. Several of the most important issues are worth mentioning here.

Does Congress want eligibility for poverty-related programs to be based on some absolute level of economic well-being? If so, a cost-of-living index is

[13]The Trustees' report used the CPI-W to deflate nominal wages. Had the estimate of methodological adjustments incorporated in the CPI-U-RS (research series) been used to adjust the CPI-W, the estimated 40-year real-wage growth would have been 0.35 percent higher.

indeed appropriate. However, if policy makers want SSI benefits and eligibility limits in other programs for the poor to be determined on a relative basis, that objective could be achieved by setting the limit at a fixed ratio to the median income or consumption of the working population. No cost-of-living index would be needed. The National Research Council report on measuring poverty recommended a quasi-relative approach, in which poverty thresholds would be indexed by a fixed ratio to the median consumption of basic necessities (food, clothing, and shelter).

Finally, as in the case of the elderly, a number of studies suggest that a special CPI for low-income households which simply reweighted the 200-plus individual strata indexes to reflect differences in the expenditure patterns of the poor would not be likely to differ significantly, over the medium to long run, from the CPI-W or CPI-U. But, again, reweighting at the upper level might not capture the differences that exist at the more detailed level of index construction between the cost of living faced by poor households and more affluent ones.

WAGE BARGAINS AND INDEXED WAGES

The extent of cost-of-living escalators in union wage contracts has fallen dramatically over the past 25 years. In 1976 more than one-fifth of wage and salary workers were union members, and 61 percent of union wage contracts included cost-of-living escalators, typically tied to the CPI-W. In 1995, the last year in which BLS published data on the subject, only 22 percent had such escalator clauses. By 2000, union membership had fallen to a little more than 13 percent of wage and salary workers. If no more than one-fifth of them had escalator clauses in their contracts, the number of American workers now covered by formal escalator clauses could not have been much more than 3 percent of the labor force.

One recent empirical study of the decline of escalator clauses ascribed a large influence to the decline since the late 1970s in uncertainty about inflation (Ragan and Bratsberg, 2000). The authors estimated that a return to the level of inflationary uncertainty that existed in the 1970s would raise the fraction of union contracts covered by formal escalator clauses by 10 percentage points. Should multiyear wage contracts with escalator clauses become more prevalent, giving substantial advance notice of methodological changes and, where feasible, publishing the "old" indexes during a transition period would take on added importance.

On an informal basis, the rate of inflation as measured by the CPI is one factor among many influencing the setting of nominal wages, through its effect on some combination of adaptive and "rational" expectations about the short-term prospects for changes in prices and the cost of living. The current growth of the CPI is about 0.5 percent less than would have been shown by an index based on pre-1995 methodology, and the superlative index, when published, will yield an additional small reduction. A natural question is whether revisions of this nature affect wage growth.

Except for transitory effects, changes in the growth of the CPI stemming from methodological revisions are most unlikely to affect the "true" rate of real wage growth. That is, the long-run growth of *measured* real wages would be expected to change by the same amount (but opposite sign) as the effect of the revisions. Whether, in the long run, methodological revisions affect *nominal* wage growth would depend importantly on how the central bank adapts to the change. If, for example, it changes its formal or informal inflation target by the same amount as the effect of the revision and successfully meets the target, the long-term course of nominal wages would also change in the same direction.[14] In the short run, however, too little is known to draw any conclusion about the extent to which worker and employer perceptions about inflation might differ from what is shown by the published CPI and how that difference might affect nominal wage bargains.[15]

INDEXING PRIVATE CONTRACTS

Parties to private contracts commonly undertake to protect themselves against various contingencies, including changes in prices. Such contracts may be conventional business contracts between buyers and sellers of intermediate or final goods or services, or they may be contracts pertaining to personal or family relationships, such as agreements covering child support, medical support, prenuptial arrangements, or alimony. (We consider wage contracts below in the next section.)

Official indexes of general price inflation may not be needed or desired to provide protection against price changes in many private contracts, especially those between large organizations or for sophisticated transactors. Such protections often will be tailor-made to fit the needs and circumstances of the parties involved, using the prices of specific goods or services or narrowly defined indexes. Nevertheless, for small businesses and private individuals, indexes of general inflation, or subindexes covering particular categories of goods or services, may provide useful and low-cost measures of price change for protective arrangements.

Provisions in contracts that protect against price changes are a form of insurance. (For simplicity we assume here that the insurance protects the seller, or the

[14]See below for a discussion of the effect of CPI methodological revisions on macroeconomic policy.

[15]If workers and employers should continue to set nominal wages as if downward revisions had not occurred while the central bank lowered the target inflation rate, the rate of unemployment consistent with full employment could rise.

payee of support or other payments, since most postwar experience has been with inflation, not deflation.) The greater the protection against unforeseen price changes afforded to the seller or payee by the protective clause, the greater is its cost as measured by what the seller or payee gives up somewhere else in the contract (although such tradeoffs may be less significant in the case of court-ordered agreements, such as child support). In the case of commercial contracts, the highest form of protection might be based on some index of the market prices of raw materials, and possibly the labor services, that the seller purchases. An alternative slightly less costly form of protection might be a price index of materials purchased or goods sold by the relevant industry. The most general, and least risk-free, form of protection is an index of some general measure of inflation. In that case the seller bears the "basis" risk of unforeseen relative price changes in the particular goods that are relevant to his costs but is protected against the common component of price changes manifest in inflation.

For purposes of inflation protection, the usefulness of the CPI, which covers only final consumption goods and services, has to be evaluated against that of available alternatives: the GDP price index, which covers all final goods and services, and the finished goods producer price index (PPI), or one of the PPI stage-of-processing components. Individual and family-related payments, such as child support, have some of the character of compensation, like the public benefit payments discussed above. In this case, the CPI may be the index of choice in protecting household consumption expenditures. However, for commercial contracts, the GDP price index may be superior if the broadest inflation protection is desired, while the narrower coverage of the PPI or one of its components may be preferable in other circumstances.

Most of the design issues that we have been considering in this report apply to the use of the CPI as an inflation index in individual and family compensation-type payments, and we do not discuss them further here. These design issues—for either the CPI or alternative indexes—are not critical for its use as an inflation index in business contracts, except perhaps during a transition period following the date when technical changes are incorporated in the index. Index design or measurement techniques may be periodically changed, with the result that the new index tends systematically to grow at a different rate relative to the prices of interest to a seller than it did in the past. An increased application of quality adjustment techniques in the CPI, for example, might lower its future rate of increase relative to that of the set of prices relevant to a seller. Such differences will eventually be taken into account by the parties to new contracts and reflected in the way contracts are written or in the cost of the insurance.

From the standpoint of sellers' calculations, the relevant prices of purchased inputs (included in some PPI components but not in the CPI) are presumably themselves quality adjusted. Conceivably, the increased use of quality adjustments in the indexes might make the trend growth of the indexes track quality-adjusted input prices more closely than in the past. Other changes in the design of the indexes might work in the opposite direction. In either case, the earlier relationship and the nature of the insurance provided by indexing will change. Between 1995 and 2000, BLS introduced methodological changes in the index that are estimated to have reduced the reported annual rate of CPI inflation by 0.5 percent (Council of Economic Advisors, 1999:94).[16] If recent history is a guide, the introduction of a (lagged) superlative index in the same year will produce a 0.1 to 0.2 percent reduction in the growth rate, which would be approximately reflected in any advance estimate of a superlative index that the BLS decides to produce.

Changes in statistical methodology, when they occur during the life of a contract, can cause difficulties for the parties to indexed contracts. There are a number of ways for statistical agencies to reduce those difficulties. Each time a significant technical change or a series of changes is made, a "research" series that compares the index constructed under both old and new methods for a number of years can be constructed. Both the Bureau of Economic Analysis (BEA) and BLS have been following this practice after recent revisions. To the extent that private parties to indexed contracts have provided for arbitration or other procedures to resolve disputes when significant revisions occur in index design, the availability of such research series could aid the parties in making appropriate adjustments.[17]

If the research that precedes a change in index measurement makes it possible to produce estimates of the effect of the change *prior* to its introduction into the index, this information can give both parties about to enter into an indexed contract notice of what is likely to occur. Along this line, BLS made available the results of its experimental geomeans research before the technique was introduced into the CPI, and the BEA published its Fisher chain price indexes for 5 years before substituting them for the old fixed-weight indexes. Finally, subject to considerations of cost and feasibility, it would be helpful to continue publishing "old-style" indexes for several years after major revisions have been introduced.

INFLATION-INDEXED TREASURY SECURITIES

The uses of price indexes considered thus far have primarily been concerned with adjusting the value of flows of payments of various kinds. However, such indexes can also be used to adjust asset values to reflect price changes over time.

---

[16]This estimate excludes the effect of the 1998 updating of the market basket, measured as the difference between the published index, with its 1982-1984 bases, and the results of a rolling biennial update (a practice which the BLS plans to introduce in 2002).

[17]In its instructions, "How to Use the Consumer Price Index for Escalation," BLS urges the incorporation in the contract of a built-in method for handling such situations (*http://stats.bls.gov/cpifact3.htm*).

The best-known example of such adjustment is the use of the CPI-U to index the value of certain U.S. Treasury securities.

In 1997 the U.S. Treasury, responding to the demand for an asset that combined the safety of Treasury securities with inflation protection, began to issue marketable 5- and 10-year notes and 30-year bonds called Treasury Inflation-Protected Securities (TIPS). By the end of 2000, $121 billion of such securities had been issued, which was 4 percent of the Treasury's total marketable securities. Such securities are held by investors both directly and through such intermediaries as mutual funds and pension plans.[18]

The TIPS are sold, and their interest rates determined, at a single-price auction of the kind used for Treasury's other marketable securities. The principal is indexed to the seasonally unadjusted CPI-U (lagged approximately 3 months), and at maturity the securities are redeemed at the greater of their inflation-adjusted principal or their original par value. The interest rate is fixed for the term of the security, and interest is paid semiannually on the inflation-adjusted principal. Both interest payments and any adjustment to the principal are subject to federal income tax in the year they occur, but they are exempt from state and local taxes.

Since individuals, organizations, and institutions save for a wide variety of purposes, the use of a broad index of general inflation for indexation seems appropriate. Arguably, the GDP price index, which covers all final goods and services, would provide the broadest protection against price changes. However, since the projected need to maintain household consumption expenditures in retirement is a major motivation for individual savings, the CPI is a quite logical alternative. In addition—and perhaps more important—the CPI's familiarity to the public, monthly availability, and freedom from revision provide a strong rationale for its use. Obviously, a broad index like the CPI cannot fully meet the indexing needs of savers with specific and narrower savings goals, such as the college education of children or the purchase of a retirement home. In such cases, investment vehicles that more nearly track the costs of the goods in question, such as real estate investment trusts or state or college prepaid tuition programs, might be more appropriate.

The large investors and institutions that play a major role in the market for Treasury securities will be aware of methodological revisions in the CPI and build them into their expectations about the future rate of measured inflation. Given no change in the rate of inflation expected under the unrevised methodology, the revisions should lead to different auction prices and effective interest yields for TIPS relative to those on other securities than would otherwise have prevailed. An investor will receive a different initial coupon yield and an offsetting difference in the inflation adjustment. However, revisions in CPI methodol-

[18]Inflation-adjusted U.S. savings bonds (Series I) are also available.

ogy that change the subsequently reported rate of inflation relative to what would have been reported under the old methodology will alter the future stream of interest payments to those who purchased TIPS prior to the revisions. The expectations about reported inflation, on the basis of which the auction prices were determined, would have been different had knowledge of the revisions then been available. For a 10- or 30-year investment, modest annual revisions occurring early in the life of the security could add up to a substantial sum and would obviously affect the market value of the TIPS in the intervening years.

On the assumption that an investor in TIPS is seeking to buy a stream of returns insured against perceived changes in purchasing power or the cost of living (as contrasted with hedging liabilities themselves indexed to the CPI), the potential occurrence of future CPI revisions introduces a risk of variance uniquely related to TIPS. Potential investors must face a "revision risk" whose presence will tend to lower the average auction price and raise the cost of borrowing to the Treasury (Emmons, 2000).[19] This premium, however, is likely to be a good bit smaller than the inflation premium associated with nominal yields. Conceptually, after every revision the BLS could continue to publish an unrevised index, which could be used for purposes of indexation for TIPS purchased in the year(s) prior to the revision, thereby eliminating the "revision premium." But in some cases maintaining two sets of indexes could be very expensive, and if there are frequent revisions the number and cost of special indexes could grow rapidly.

Revisions in the CPI are presumably made only after a reasoned determination that they improve the ability of the index to meet its objective, however formulated. Any unavoidable costs associated with creation of a risk premium in TIPS are likely to be much smaller than the benefits to society from an improved index in its many uses.

INDEXING THE FEDERAL INCOME TAX SYSTEM

In the federal income tax system, marginal tax rates rise with nominal income. A taxpaying unit whose real income is unchanged can be pushed into a higher tax bracket by inflation. In 1981 Congress provided for indexation of the individual income tax with the objective of preventing inflation from raising the burden of federal income taxes. Since 1985 (when this tax provision became fully effective) the bracket widths, the personal exemptions, and the standard deduction have been automatically indexed annually with the CPI to ensure that inflation would not result in a higher effective tax rate for any taxpayer. A 1987 paper (Gillingham and Greenlees) showed analytically and demonstrated empirically

[19]Although TIPS constitute only about 5 percent of outstanding marketable Treasury notes and bonds, that percentage is increasing since about one-third to one-half of new Treasury issues are TIPS.

that, under most circumstances, an expenditure-based index like the CPI would not accurately achieve this objective and would tend to under-index the tax system. From a cost-of-living standpoint, CPI indexation will not maintain a specified ratio of federal income taxes to the consumption spending that would yield a constant standard of living (what Gillingham and Greenlees label "direct consumption costs").[20]

There are a number of reasons for this result. Some items of consumption are tax deductible: if their prices rise faster or slower than those of nondeductible consumption, a tax system indexed by the CPI will not be inflation neutral. Some items of expenditure, such as medical care, give rise to tax deductions, and some items of income, such as the imputed rent from owner-occupied housing, are tax exempt. Most importantly, during a period in which federal payroll tax rates increase, the income necessary to maintain any given standard of living in the face of price changes rises by more than the CPI. Hence indexation of bracket limits and other tax parameters by the CPI does not keep federal income taxes at a fixed ratio to direct consumption costs.

Even after a much-simplified representation of the tax code and other simplifying assumptions, neutralizing the effect of inflation requires a complex indexation formula, labeled an "exact indexation measure" (Gillingham and Greenlees, 1987). It takes into account the effects on the ratio of taxes to direct consumption costs arising from the factors listed above, the most important of which have been the indirect effects of changes in federal payroll tax rates. The formula also requires the construction of individual indexes for every consumer in a subsample of the 1973 Consumer Expenditure Survey in order to mimic the effect of the progressive federal income tax system.

Over the 1967-1985 period, the CPI (for the sample of households used in the simulation) rose by 197 percent and the exact index by 241 percent. A simulation of a simplified version of the statutory tax structure of 1973, with prices rising at the rate of growth of the CPI that actually occurred between 1967 and 1985, showed that the average ratio of federal income taxes to direct consumption costs would have risen by 192 percent over the period. CPI indexation would have reduced that to a rise of 13 percent, while exact indexation would have reduced it to zero. (In fact, marginal tax rates were adjusted periodically and other changes were made over the period, so that the actual ratio of taxes to gross consumption costs, without indexation, rose by 40 percent.)

The CPI is clearly an imperfect tool for indexing the income tax system. Yet despite the high rates of inflation during the period studied, CPI indexation, while not rendering the income tax system completely neutral, would have eliminated much of the bracket creep. The Gillingham-Greenlees exact index requires a

[20]Gillingham and Greenlees (1987) describe the base-weighted CPI as an approximation to a conditional COLI.

complex formula even for a stripped-down representation of the tax code. Constructing a more precise index would substantially add to that complexity and still require the use of some judgmental simplifications. The result would be an opaque indexing measure that would be very difficult to explain and justify to taxpayers. Therefore, we conclude the following:

> **Conclusion 7-6: Despite the imperfections of the CPI, it should continue to be the basis for tax indexation and the tax law should not be changed to require the construction of an "exact" index.**

MEASURING OUTPUT CHANGES

In the national income accounts of all countries, changes in national output are estimated by using price indexes to "deflate" the changes in the components of current dollar expenditures and then combining them into aggregate indexes of output (quantity) or, correspondingly, constant dollar output measures. In this context an aggregate index of inflation ought to be evaluated in terms of its ability to partition expenditures into two symmetric components—an index of inflation and an index of output change, which when multiplied together produce the observed change in current dollar expenditures.[21]

In the United States, as in most other industrial countries, the overall CPI or its equivalent is not used as a deflator for aggregate consumption expenditures, but its individual components are the deflators for most of the individual categories of consumption expenditures. Up-to-date estimates of consumer expenditures in current dollars are made quarterly by BEA in the U.S. Department of Commerce to produce the national income and product accounts. The expenditure estimates are based on sales and other data, collected mainly from sellers rather than households. For most categories of consumption goods, the NIPA estimates of expenditures are substantially higher than those derived from the CEX survey.[22] While the NIPA estimates undoubtedly pick up a large volume of expenditures that are missed in the CEX interviews and diaries, for many expenditure categories (e.g., automobiles and computers) one must cull from NIPA the sales made to business firms or to individuals for business use, which introduces a potential source of error. Consumer expenditures constitute about two-thirds of GDP. If the NIPA estimates of aggregate consumer spending are seriously over-

---

[21] See Diewert (2000a:sections 2, 3) for a treatment of this topic.

[22] There are some differences in scope between the NIPA and CPI universes of consumption goods, dictated by the structure of the national income accounts. For example, the NIPA classifies the value of in-kind transfers to consumers as consumption, rather than government expenditures. And in the NIPA price index for personal consumption expenditures, the weight assigned to medical services includes Medicare- and Medicaid-financed outlays; in the CPI the medical service weights reflect only out-of-pocket consumer expenditures.

stating their true magnitude, the statistical discrepancy (which measures the excess of the independently collected gross domestic income relative to gross domestic expenditures) would tend to be a substantially larger (positive) number than it typically is.

Most of the NIPA consumer expenditure data are grouped into detailed categories that can be associated with one or a combination of several CPI strata price indexes. These expenditure categories are then deflated, principally with the use of the CPI price indexes, to produce quantity indexes of real consumption outlays for each category. The detailed price and quantity indexes are aggregated into annual and quarterly Fisher indexes for total consumption and its major subcategories.[23] Fisher quantity and price indexes have the desirable property that an index of current dollar expenditures divided by a Fisher price index produces the corresponding Fisher quantity index.

The quarterly data are available shortly after the quarter ends. Inflation and output changes for each year or quarter reflect the weights of the beginning and ending periods in the comparison. The Fisher price and quantity indexes are superlative indexes and take approximate account of the effect of consumer substitution among the individual strata categories of goods in response to changes in relative prices. For this as well as other reasons, the measured inflation rate is a little lower and the output increase a little higher than would be the case with Laspeyres weighting.[24]

Even though the overall CPI is not itself used as an output deflator for consumption expenditures, its individual components are the critical elements (along with the estimates of current dollar expenditures) in measuring what is happening to the level and structure of national output. Decisions about how to deal with such problems as quality change and new goods thus feed into the aggregate measures through the individual price indexes.

The estimation of national output data through the deflation of current dollar expenditures for the individual components of consumption requires either that the individual strata indexes used for deflators be themselves aggregated from individual prices with expenditure weights or that the individual prices be selected with the sampling procedure based on expenditure weights. That is, they must be plutocratic indexes. If the research into alternative data collection tech-

[23] The annual Fisher indexes are based on weights in adjacent years and (in a very recent change in methodology) the quarterly indexes on weights in adjacent quarters. The quarterly indexes through the most recent completed year are adjusted so that their annual average corresponds to the annual indexes.

[24] BLS recently switched to geometric averaging of individual prices for calculating the individual indexes for strata that account for about 61 percent of the weight in the CPI. Analysis led the agency to conclude that in these strata substitution among goods in response to price changes was large enough to warrant that geometric weights would be superior to arithmetic weights for approximating the effects of substitution behavior on the cost of living.

niques that we recommend (see Chapter 8) ultimately leads to the production of democratically weighted CPI indexes at the lower (strata) level, it will be necessary to retain a parallel system of expenditure-weighted individual strata indexes for use in the NIPA.

Since the detailed components of the CPI are the building blocks of the NIPA as output deflators and inflation measures for consumer goods, which make up two-thirds of GDP, decisions about conceptual and measurement issues in its construction will not only determine what is known about the rate of inflation but will also strongly influence information about the rate at which the U.S. economy grows and the nature of its structural changes over time.

The most important issues in this regard are decisions about how to deal with quality change and the introduction of new goods. One would surely, for example, not consider it inflationary (and thus calling for restrictive monetary policies) if households consumed a large fraction of their growing real incomes upgrading the quality of autos, housing services, and other goods and, in the process, paid higher nominal prices. But the quality issue is not only a matter of doing a better job of capturing product improvements. Some analysts have argued, and provided indirect evidence, that sellers often use the occasion of changes in models and styles to raise prices. As a consequence, the current BLS procedures for pricing the new models may attribute too much of that markup to quality change and too little to "pure" price change (see Chapter 5).

For purposes of well-informed macroeconomic policy making, measures of national inflation, and corresponding measures of national output, ought to incorporate the prices and output of new goods as soon as practicable and also reflect changes in the quality of goods and services to the extent that they can be reliably measured, subject to the conceptual limitations discussed in Chapters 2 and 5. Pursuit of this objective must fully recognize the difficult problems that surround the use of statistical estimation techniques to produce measures of quality change and observe the cautions we express in Chapter 4 about introducing quality adjustments into the index before sufficient preparatory research is done. These difficulties underline the potential value of a continuing research and operational effort directed toward further improving and monitoring the ability of the BLS to deal with changes in quality, the timely introduction of new goods, and related issues.

INFLATION INDICATORS FOR MACROECONOMIC POLICY

The central reason that inflation of any significant magnitude is considered undesirable and economically destructive is that it creates obstacles and uncertainties for business firms and individuals in planning and making commitments for the future. In theory, if the rate of future inflation were known with certainty, then with minor inconvenience one could operate with equal efficiency under a high or a low rate of inflation. But the future rate of inflation is never known with certainty. Moreover, inflation tends to be more variable and relative prices subject to larger changes when overall inflation is high than when it is low. To the extent that inefficiencies in business planning are a major concern, an index that somehow combined both input and output prices would be desirable conceptually, perhaps supplemented by some measure of the variation in relative prices. But since the cross-sectional and time-series variations in inflation tend to be positively correlated with the overall inflation rate, some broad measure of inflation—the CPI or the price index for GDP—is a suitable indicator for purposes of monetary policy. The Technical Note at the end of this chapter compares the GDP price index, the CPI, and the NIPA price index for consumer expenditures as measures of inflation.

Over the last 10 to 15 years, central banks of the economically advanced countries have increasingly come to define their principal, if not always sole, objective as the pursuit of low and stable inflation rates. In some countries these are embedded in numerical targets set by law. The U.S. Federal Reserve has no statutory numerical target, but it does seek a low inflation rate as a top-priority objective.[25] Even if modest changes in inflation may not be harmful to macroeconomic performance or consumer welfare, charging the central bank with achieving sustained low inflation is likely to be desirable. It may, for instance, reduce the temptation to adjust macroeconomic policy to avoid politically difficult decisions by allowing inflation to inch persistently upward.

When inflation strongly overshoots a central bank's targets, "errors" of 0.5 or 1 percent a year in measuring inflation clearly do not matter very much—the steps the central bank ought to take are obvious. But when inflation is in the general neighborhood of the target, small measurement differences can loom large in policy making (which would necessarily be the case if the inflation target were set in law). In this view, the monetary authorities would *mistakenly* restrict the economy if inflation were above the target due to measurement error, whereas such restrictive policy would be unnecessary if the inflation measure were "corrected" to remove the source of error. Because the transition costs of pushing the inflation rate down by, say, a further 1 percent are very substantial, the magnitude of any upward bias in the inflation index can be a critical question of monetary policy. It is not at all clear, however, that this is a correct view of how monetary authorities should react to a change in index measurement techniques.

[25]The Federal Reserve Act directs the Federal Reserve to seek "to promote effectively the goals of maximum employment, stable prices, and moderate long term interest rates." But as the Federal Reserve Board has stated, "A stable level of prices appears to be the condition most conducive to maximum sustained output and employment and to moderate long-term interest rates" (Federal Reserve Board, 2000). History makes it clear that the Federal Reserve considers an inflation rate persistently in the very low single digits to be consistent with the statutory objective of "stable prices."

Assume that the actual rate of inflation is a little above the target and that new statistical procedures are introduced into the index which lower the reported inflation rate by *x* percent, where *x* is a small number. Two alternative policy responses could occur. First, the central bank finds that reported inflation now meets its target and relaxes its efforts to drive inflation lower. Alternatively, it could reduce the target inflation rate by *x* percent. Since nothing in the real economy has changed by virtue of the statistical correction, the former response represents an implicit judgment that (unbeknownst to policy makers) the old target was in fact too high and was unduly restrictive for the economy. But mere changes in measurement can hardly be taken as evidence regarding the appropriateness of the original inflation targets. In the absence of any reason to believe that the old targets were inappropriate relative to the way inflation was reported at the time, the logical reaction to a change in inflation stemming from revised measurement is to alter the target correspondingly.

Because the CPI is a focal point for many transactions, a statistically induced change in the index may, during some transition period, have important real effects—for example, on wage bargains and therefore on actual inflation. The central bank would have to take those changes into account in formulating its operating policy; but it would seem logical to adjust the target itself to reflect the statistical change.

Since the CPI is used for indexing public transfer payments, the income tax code, and private contracts, even those methodological changes that produce relatively small differences in the annual rate of inflation can have significant consequences for government budgets and the welfare of individuals as their effects cumulate over time. When used as an output deflator, even small changes in the GDP price index can cumulate into larger differences in reported output gains and have at least modest effects on the assessment of real economic performance. But as an inflation indicator for the central bank and other policy makers, small differences in reported inflation rates resulting from statistical changes in the CPI or design differences between different price indexes are probably not very significant. Thus, the difference between the use of a Laspeyres or a superlative index (at the upper level) represents a choice of at least some importance for most uses of the CPI, but not for its use by the central bank as an inflation indicator. Even if such small differences should matter, they will likely have been incorporated into the decision making of these sophisticated users of price indexes.

In addition to the use of an inflation index as a measure of their policy target, central banks and other policy makers use various derivative measures of current inflation as devices to help them to filter out the substantial "noise" in the month-to-month changes in inflation so as to detect significant changes in its underlying level. Many different kinds of measures can be used, such as the "core" rate of inflation (excluding such volatile items as food and energy); a "stripped" rate of inflation, removing (for example) the 10 percent of items with the largest positive and negative changes; and the weighted median rate of change of the individual items in the index (see Cecchetti, 1997). Policy makers also typically want to examine current trends in producer prices for finished, intermediate, and crude goods, as well as import prices and the employment cost index. However, the central banks and governments of advanced countries have sizable economic and statistical staffs who can manipulate individual components of the CPI, the PPI, and other indexes to produce these sorts of analytical tools. Except perhaps by generating a demand for additional or different subaggregations of goods, these are not matters that involve the basic design of the indexes.

TECHNICAL NOTE: THE CPI VERSUS THE NIPA PRICE INDEX AS AN INFLATION MEASURE

As measures of aggregate inflation, the NIPA indexes have some advantages over the CPI. First, an overall measure of inflation ought to include the prices of all privately produced goods and not only consumer goods. Some have argued that if an index of inflation for the goods purchased by consumers remains stable, changes in the prices of investment goods can be ignored, since satisfying consumer needs is the objective of the economy. But that is only true in a steady state and is not a very helpful assumption when analyzing economic conditions, which often change rapidly and unexpectedly. Aggregate NIPA inflation indexes are available quarterly for total GDP and for gross domestic purchases (including imports, by all domestic users) and for a fairly detailed set of components, including, of course, many categories of consumer goods.[26]

The fact that an index that includes capital goods is a useful inflation indicator does not imply that indexes of *consumer prices*, CPI or NIPA, should abandon the current practice of pricing the service flows from the housing stock and return (in the case of the CPI) to pricing house purchases. The broader price index, congruent with the definition of GDP, should (and does) price both the flow of goods and services for consumption and the stream of outlays on all capital goods, including housing.

One characteristic of the NIPA inflation indexes is that they provide a Fisher-type index on a real-time basis, whereas such an index for the CPI will only be made available after a 2-year lag. To estimate a real-time Fisher index for the CPI, it would be necessary to rely on a forecast approximation. In a comparison made for recent years, the real-time CPI, which is Laspeyres at the upper

---

[26]Publishing a measure of inflation for the private-sector GDP might be helpful for some individual users. It would exclude the output of government employees (in which, by assumption, productivity gains are zero and price changes equal wage changes) and would save users the effort of calculating it from the published data.

level, has averaged 0.15 percent a year higher growth than the Fisher, although the differences varied from year to year, reaching 0.5 percent in one year.

The domain of the NIPA index includes all medical consumption expenditures, including those paid for by Medicare, Medicaid, and employer-paid health insurance, which currently account for 17.5 percent of the weight in the index. The CPI only includes consumer out-of-pocket costs, which represent about 6.0 percent of the weight in that index. In most (but not all) years of the past several decades, the price of medical care has been rising faster than the average for other consumer goods and services, raising the NIPA inflation rate above the CPI rate. However, the NIPA (atypically) uses not the CPI but the PPI index for hospital service inflation, which is differently constructed and shows a lower rate of increase. On balance, the treatment of medical care costs in the NIPA index raises its measure of consumer goods inflation slightly relative to the CPI rate in most years. As a measure of inflation (as opposed to an index used for compensation purposes) the NIPA estimate, which comes closer to covering all medical service prices, is the more relevant one.

Table 7-2 compares the annual inflation rates of three alternative indexes since 1991: the NIPA GDP and personal consumption expenditure (PCE) price indexes and the BLS research index (CPI-U-RS) that extends backward in time the major revisions in CPI measurement techniques that have been introduced over the last several decades. (Many of these changes had been incorporated earlier in the NIPA indexes, and several others were extended back to 1978 in the latest NIPA revisions.)

With the exception of 1996, the differences between the two measures of consumer goods inflation are small. There are a number of conceptual differences

TABLE 7-2 Comparison of NIPA and CPI Indexes, 1991-2000 (percent change)

Year	NIPA, GDP	NIPA, PCE	CPI-U-RS
1991	3.6	3.8	3.7
1992	2.4	3.1	2.7
1993	2.4	2.4	2.6
1994	2.1	2.0	2.2
1995	2.2	2.3	2.5
1996	1.9	2.1	2.7
1997	1.9	1.9	2.2
1998	1.2	1.1	1.4
1999	1.4	1.6	2.0
2000	2.3	2.7	3.4

NOTES: NIPA, National Income Product Accounts; GDP, gross domestic product; PCE, personal consumption expenditures; CPI-U-RS, research series using the urban consumer price index.

between the two indexes, including the treatment of medical services; the fact that the PCE has updated weights every 4 or 5 years whereas the CPI had 1982-1984 weights until 1998; and the use of a Fisher index for aggregation in the PCE versus a fixed-weight Laspeyres for the CPI. On balance, these differences tend to lower the PCE slightly relative to the CPI: over the 10 years shown in the table, the average annual increase in the PCE was 0.2 percent less than in the CPI. Overall, however, the two indexes move very closely together, exhibiting much the same downward trend in inflation over the period and, with the exception of 1996, roughly the same pattern of small fluctuations around the trend. Except for a 1-year deviation in 1992, the price index for GDP moves closely with the consumption price indexes—not too surprisingly since consumption is two-thirds of GDP.

There is not much to choose among the indexes as an indicator of general inflation. The NIPA indexes allow an analyst to keep track of indexes for investment goods, exports, and consumer goods estimated on a basis consistent with the index for total GDP. It has up-to-date weights and a small advantage in being aggregated as a Fisher index. The CPI, in contrast, is produced monthly—although this advantage is attenuated by the degree of noise that accompanies monthly data. Luckily the Federal Reserve, the executive branch, and the Congressional Budget Office have enough analysts to use and compare both sets of indexes and their various components.

8

Whose Index? Aggregating Across Households

This chapter deals with the consequences for price index construction of the tremendous heterogeneity in the purchasing patterns and shopping behavior of consumers. An important part of these differences among consumers is associated with differences in their economic and demographic characteristics and their geographic location.

The existence of heterogeneity raises two important questions: First, are the rates of inflation experienced by different groups or by people in different geographic locations sufficiently different so that separate indexes should be constructed for each group or location? This issue is particularly important when indexes are used for adjusting taxes, social security benefits, and other public transfer payments: Should they be tailored to the groups to be compensated? If so, how might the data be collected that would allow us to investigate the extent to which inflation rates for particular groups differ and, if they do, to construct separate indexes?

Second, even if households face differing rates of price change, a single national index would still be needed for many purposes—for example, to provide an overall measure of inflation. But construction of a single index requires some method of averaging or aggregating across people, and there are different ways of doing that. Even with many subgroup indexes, there is still enough heterogeneity within those subgroups that the averaging issue would still have to be faced. As noted throughout this report, the Consumer Price Index (CPI) is now a plutocratic index: the weight of each household's consumption pattern in the overall index is proportional to that household's total consumption expenditures. Since consumer expenditures rise with income, this approach gives more influence in the construction of national and regional indexes to the consumption patterns and prices paid by the rich than to those of the poor.

The alternative approach, for which it would be much more difficult to collect the necessary data, is a democratic index: Under this approach, we would construct individual price (or cost-of-living) indexes for a representative sample of the whole population and then average them, assigning the same weight to each person, regardless of the magnitude of their total consumption expenditures. Which kind of index is appropriate for each of the major purposes that the CPI serve? And, in practice, how would one construct the democratic counterpart of the current plutocratic CPI? Such questions arise in much the same form whether one is working with fixed-basket or cost-of-living indexes.

This chapter pays particular attention to the fact that the data collection system underlying the CPI, and those employed to produce price indexes in other countries, cannot now provide important elements of the information needed to explore consequences of consumer heterogeneity and, specifically, to determine whether inflation rates do in fact differ among various groups within the population. As we pointed out in the introductory chapter, the key constraint is that information about consumers and how they budget their income is collected from a household survey, while price information is collected from retail stores; thus we cannot link the characteristics of purchasers with the prices they pay. This chapter outlines the kind of surveys that would be needed to collect price data directly from households. It explains why, using current survey techniques, acquiring such information would be extremely expensive and perhaps impossible. It suggests research options for exploring the feasibility and costs of alternative and more technologically intensive survey methods that might help solve this problem and, in the process, produce information about the inflation experience of particular groups such as the poor or the elderly.

TWO KINDS OF HETEROGENEITY

From the standpoint of constructing a price index, heterogeneity shows up at two stages of the process. First, people allocate their consumption budgets differently across categories of goods such as food, shelter, entertainment, and travel. Some of these differences are idiosyncratic among individuals—vegetarians and meat eaters, book lovers and sports enthusiasts, travelers and homebodies. But many of the differences are systematically related to the economic, demographic, and locational characteristics of households. The poor spend a higher fraction of their income on food and clothing than do the rich and a smaller fraction on travel and entertainment. The elderly tend to devote a smaller fraction of their budgets to durable goods and clothing and a larger fraction to travel and medical care than do non-elderly. People who live in the South spend less on heating fuel and more on air conditioning than those in the North. As we explained earlier, the Bureau of Labor Statistics (BLS) distinguishes some 218 different strata or categories of goods; we label the variation in how consumers allocate their funds among these different categories as "across-stratum" heterogeneity.

Consumer behavior exhibits a second kind of heterogeneity. Within any given stratum of goods, different people buy widely different qualities and brands of goods, often shop at different kinds of retail outlets, and pay different prices for the same product. The price of housing varies a great deal from one part of the United States to another, as well as between the city and the countryside. Different patients pay widely different prices for the same medical treatment. Also, shopping outlets offer goods and services at non-identical prices; catalog stores may tailor their prices to the zip code of the purchaser. The spread of Internet shopping may bring prices closer together, because it makes arbitrage easier, or it may drive them further apart by allowing retailers to set prices more nearly in accordance with the characteristics of the shopper. These within-stratum differences, too—like those across strata—arise not only from idiosyncratic heterogeneity of tastes, but also from differences in age, income, family composition, geographical location, and other factors that have important social implications. Within almost every category of goods the poor choose less expensive and lower-quality brands, often shop in different stores, eat at different types of restaurants, and may pay different prices for the same good. Retirees are more likely to travel on group tours than take bicycling excursions and, within the BLS stratum "sporting equipment," to buy golf clubs instead of soccer balls or skateboards. And they may get senior citizens' discounts on many items, whose scope and value can change from time to time.

It is also quite possible that the ability or willingness of an individual to substitute in response to changes in relative prices differs depending on the level of that individual's income. In consequence, the magnitude of the substitution effect built into a cost-of-living index (COLI) may differ from one person to another, depending upon their income levels. The research on this topic contains many presumptions on how substitution varies with income, but remarkably little evidence. Some researchers argue that, precisely because of limited resources, the poor are more careful with their budgets, hunting out bargains, substituting to the maximum extent possible. Others point out that the poor buy a smaller range of goods—a phenomenon that *is* well documented empirically—and, therefore, have less scope for substitution among goods. Moreover, the poor consume more necessities, formally defined as goods whose demand rises less than in proportion to income, and the term suggests that at least some necessities are hard to substitute. Certainly, for such items as medical care or home heating oil, it is hard to substitute one item for another, especially in the short run. Conceivably, both of these kinds of forces could influence different aspects of the shopping behavior of people with low incomes, but there is nothing in the theory of consumer behavior that creates a presumption that the balance tips one way or the other.

HETEROGENEITY IN PRICES PAID AND IN RATES OF INFLATION

The fact that individuals and groups differ one from the other in their broad consumption patterns, in the kinds and qualities of goods they buy, and in the prices they pay is not itself enough to produce differences in the rate of inflation or the rise in the cost of living they experience. To the extent that relative prices do not change very much over time, the rates of inflation for individuals and groups cannot, of course, differ very much from each other. But relative prices are continually changing. There is some evidence that the higher the overall rate of inflation and the greater the degree of economic disruption, the larger is the variation in relative prices. Moreover, there are periods when inflation is dominated by the rapid rise in the price of one or a few commodities—oil and other energy prices in the 1970s and early1980s and, to a lesser extent, health care costs in the late 1980s and early 1990s. Yet the existence of significant alterations in relative prices will not itself produce important differences in the rate of inflation faced by different groups unless, on average, the collection of prices for the kinds and qualities of goods typically purchased by one group rises faster or slower than the collection of prices of the kinds and qualities of goods purchased by other groups.

Such systematic differences do exist. Lower-income consumers in the northern states are particularly affected in the winter months by increases in the relative price of heating oil; the elderly, even with Medicare, tend to be hit hardest by above-average increases in health care costs. Unless these increases are fortuitously offset by relative price declines in other goods of which they are also especially heavy consumers, the rate of inflation they face will exceed that for the population as a whole. Some observers point out that quality improvements and technological advances, which present one of the most difficult issues in price index calculation, are more prevalent among the types of goods purchased by rich than by poor households and that the rich are the first to acquire such goods. As a consequence, an overall price index corrected for quality change runs the risk of understating inflation for the poor, even if it accurately incorporates the effect of quality change for those who can afford to buy new and improved goods. Others argue that this phenomenon may be partly offset by a "trickle-down" effect of quality change. To the extent that technology produces goods with new characteristics and lower-quality-adjusted prices, the prices of older models also cascade downward, benefitting groups with less income.

In summary, different groups of consumers will experience significantly different rates of inflation whenever three conditions all occur: There are substantial differences in what consumers buy and the prices they pay; there are significant changes in the relative prices of goods; and the distribution of those changes is such as, on average, to raise or lower the prices of the collection of goods typically bought by some groups of consumers relative to the prices of those bought by others. The first two conditions almost certainly occur most of the time; there is little evidence about the third.

THE CONSEQUENCES OF HETEROGENEITY FOR INDEX CONSTRUCTION

If inflation rates differ among people—occasionally or frequently—an overall national index, however averaged over the population, could misstate the rate of inflation for the poor, the elderly, other demographic groups, or people living in different regions of the country. An important aspect of public dialogue about national policy revolves around the changing economic fortunes of different income, age, and racial groups and different regions of the country. We typically measure these fortunes by adjusting changes in nominal incomes or consumption expenditures by a single national index of inflation, which is fine if group differences in inflation are small but misleading if they are not. Even more importantly, if inflation rates differ significantly among groups, the use of an overall CPI to index social security pensions, other public benefits, and the income tax system may overcompensate some groups in the population and undercompensate others in ways that most people would deem unfair or unjust.

As a first step to address the issue, one would like to know the extent to which, in the past, a single national index would have been a reasonably close measure of inflation rates for different groups, and how often and under what economic conditions inflation rates for one or more groups differed significantly from the overall index. While the past need not repeat itself, this kind of information would certainly be valuable for making a reasoned judgment about whether a single national index is sufficient for the major uses to which it is put or whether it needs to be supplemented with one or more indexes representing the different experiences of particular groups within the population. Unfortunately, the current data collection system underlying the CPI is generated is such a way that it is impossible to produce indexes for subgroups in the population that capture the heterogeneity in the qualities of goods purchased and the prices paid among those subgroups.

GROUP INDEXES: WHY THE CURRENT DATA COLLECTION SYSTEM CANNOT PRODUCE THEM

Chapter 1 describes how BLS produces the CPI in two stages. At the first stage, known as the lower level, BLS collects data on monthly price changes for individual items—not from individual households, but from a sample of retail outlets throughout the nation. It groups those items into some 218 categories or strata.[1] Within each stratum, BLS combines into a single-stratum index the

[1] Data on housing rents are collected in a special survey.

monthly price changes for all the individual items in that stratum. For example, within the "new passenger cars" stratum, the monthly price changes for Mercedes, Buicks, Chevys, and Hondas are all averaged together, as are those for soccer balls, hockey helmets, golf clubs, and other items in the "sporting goods" stratum. The price changes for men's clothing purchased at Brooks Brothers and at Walmart are similarly combined in the "men's suits" stratum index. In this process, all within-stratum heterogeneity is lost.[2] And since the price changes are collected from retail stores, there is no way to assemble the data so as to make a direct link between the particular price, quality, and brand of items purchased and the economic or demographic characteristics of those who purchased them.[3] At the second or upper-level stage of estimation an overall CPI is calculated as an average of the 218 stratum indexes, with each index assigned a weight equal to the proportion of total consumer expenditures devoted to purchases of the goods in that stratum, estimated from the Consumer Expenditure Survey (CEX).

BLS and individual researchers have, on occasion, produced indexes for subgroups in the population—for the elderly, the poor, or, in a recent BLS report, the individual quintiles of the income distribution—by reweighting the stratum indexes with expenditure weights that represent the budget allocations of the particular demographic subgroup as determined from the CEX. Although the across-stratum weights are different in each subgroup index, the individual stratum price indexes are the same in all of them. Generally, the subgroup indexes that have been produced have not risen at a substantially different rate than the overall CPI, although at times there have been exceptions (see the second technical note to this chapter for a summary of such comparisons). But such indexes distinguish one subgroup of households from another solely by the differences in the way each one allocates its expenditures among expenditure categories. Only the *across-stratum* heterogeneity is accounted for. The individual stratum price indexes are averages and so do not capture *within-stratum* heterogeneity—the fact that those at the upper end of the scale typically buy the higher-quality items within any given category of goods are the first to acquire many types of new products, shop in high-end grocery and apparel stores, live in areas with high rents, are far more likely to be covered by medical insurance and to fly business

---

[2]This is a highly truncated description of a more complicated process, but it accurately depicts the essence of the procedure.

[3]From a TPOPS-type survey augmented with additional data on household characteristics, it would be possible to determine the average economic and demographic profile of the people who patronized each retail outlet in the BLS sample. One could then observe any differences in the extent to which price inflation varied by category of goods among stores whose patrons had different characteristics. This would provide useful information about the consequences of heterogeneity of prices paid, but in the absence of a direct link between prices paid and individual purchasers, the association between economic and demographic characteristics and prices paid would be quite indirect and partial (see below for a brief discussion of this approach).

or first class on airlines even for personal trips, and so on down a long list of differences with low-income groups.

A PRICE INDEX FOR THE ELDERLY?

An example of the inability of the present data system to answer important questions frequently surfaces in discussions of whether the cost of living for the elderly rises at a faster or slower rate than the CPI as a whole. During the 1995 Senate Finance Committee hearings on the CPI, in response to a question from Senator Kent Conrad about a separate price index for the elderly, former BLS Commissioner Janet Norwood (1995:80) said: "The real point is that we do not know. And we do not know because we do not have prices that are collected for items that are purchased by the elderly." The Boskin commission (Boskin et al., 1996:72), in its discussion of a separate price index for the elderly, points out that an index for the elderly calculated by using CPI prices and reweighting to match the expenditure patterns of the elderly does not differ substantially from the index for the non-elderly but recognizes that "the prices actually paid, not just expenditure shares, may differ." Actually, for the period December 1990 to December 1995, the experimental CPI-E rose by 15.9 percent, somewhat more than the CPI-U and CPI-W, which rose by 14.7 and 14.1 percent, respectively (see *www.bls.gov/news.release/cpi.br12396.a06.htm*). Most of the difference can be explained by the larger expenditure share on the CPI medical care component (which increased faster than the average of other prices); a small portion of this effect was offset by a lower share by the elderly on "other goods and services," a major expenditure group that also showed higher-than-average price growth.

The Boskin commission concluded its discussion by acknowledging that "work on this subject remains to be done" (p. 72). In an article discussing the Boskin commission report, after quoting both Norwood's testimony and the Boskin commission report, Pollak (1998:71) writes:

> Mention of "items that are purchased by the elderly" and "prices actually paid" turns the discussion of group indexes and representative consumers toward the items and qualities priced for the index and the outlets in which they are priced. The literature on group indexes has treated the construction of household indexes as a distinct, prior task and focused on the problem of aggregating household indexes into a group index. In practice, however, because we do not first construct household indexes and then aggregate them, our definition of the group index has implications for the items and qualities we price and the outlets in which we price them.

Would a price index for the elderly behave differently than the overall CPI if data were collected on items and qualities consumed by the elderly and on the prices paid in outlets where the elderly shop? To have a definitive answer to this question, or even relevant evidence instead of speculation and conjecture, an index for the elderly would have to be constructed to reflect "items that are purchased" and "prices actually paid." By comparing such an index with one constructed by applying CPI strata prices to the expenditure pattern of the elderly, one could see whether, given the behavior of prices in a particular historical period, it would have been different.

THE CONCEPTUAL BASIS FOR GROUP INDEXES

One approach to the problem of how to aggregate across heterogeneous individuals starts from the concept that there exists a cost-of-goods index (COGI) and a COLI for each household, based on the prices it pays and the quantities of each good it buys. In a footnote at the beginning of its report, the Boskin commission wrote (Boskin et al., 1996:5): "In principle, if not in practice, a separate cost of living index could be developed for each and every household based on their actual consumption basket and prices paid." Those individual indexes could be combined or averaged into many possible combinations—a single national average for all households and separate indexes for various population subgroups and geographic areas. To produce an overall national index, the indexes for the individual households could be averaged, giving equal importance to each (a democratic index) or be weighted in accordance with each household's total expenditures on consumer goods (a plutocratic index).

Currently, BLS collects monthly price data from retail outlets and other sellers and combines them with information on consumer expenditure patterns derived from separate surveys of households. But to produce individual price or cost-of-living indexes and then combine them into indexes for demographic subgroups, it would be necessary to combine, for each household, monthly information on the prices it paid, the amount expended on each item, and its basic demographic characteristics. Since retailers cannot provide information on either their patrons' demographic characteristics or overall expenditure patterns, it would be necessary to collect the monthly price data, as well as expenditure patterns and demographic information, directly from consumers.

As soon as one begins to think through the implications of collecting such data, however, it becomes increasingly difficult to support the proposition that a monthly (or even annual) price or cost-of-living index could be constructed for individual households. There are many types of goods that an individual consumer buys only at infrequent intervals, and among the goods that a consumer does buy frequently, purchases often vary among different qualities and brands of those goods. Consumers may make rental and utility payments and buy some categories of goods (e.g., food or beverages) on a monthly or more frequent basis, but a major fraction of their purchases occur at longer—in many cases much longer—intervals: How often do people buy a winter suit, a bottle of aspirin, a lawnmower, a resort vacation, a television set, or a refrigerator? For medical care, it is unlikely most people will experience the same major procedure more than once a lifetime.[4]

Even for those categories of goods that are purchased frequently, many households may switch purchases among products that possess somewhat different qualities and are available at different prices, occasionally or often buying different kinds of green vegetables, meat, fish, shampoo, cosmetics, beer, and the like from week to week and month to month (quite apart from substitutions among products driven solely by changes in relative prices). But construction of the current CPI involves the measurement of period-to-period changes in the prices of goods of the same quality purchased in stores that provide the same services to shoppers. BLS goes to great pains to price the identical item each month in each retail store from which it is collecting prices. When that item disappears from a store and BLS must substitute another similar item to price, it devotes substantial resources in an effort to separate the difference in price between the old item and the new item into a component that represents quality changes, which it does not include in the index, and a component of "pure" price change, which it does include (see Chapter 4). And when new outlets are introduced into the sample of stores from which it collects prices the BLS "links" them in, so that any difference in prices is attributed to differences in the quality of service and does not cause a change in the price index. All in all, it would be impossible—even through observing a high-frequency series of prices paid by an individual household (monthly or even annually)—to calculate an index that measures the rate of inflation or the rise in the cost of living that the household has experienced.

An Alternative Approach

A less ambitious but more feasible approach to the aggregation problem would be to exploit the fact that an important part of the heterogeneity among households in consumption patterns and prices paid is systematically related to differences in their economic and demographic characteristics and in their geographic location. The nearest approximation to a homogenous unit that could form the building block for purposes of aggregation across households might, therefore, be an index of the prices of specific goods actually paid by a group of

---

[4]In theory one might treat all goods which satisfied a consumer's wants for any period longer than, say, a month as a durable good. Durable goods so defined might then account for 90 percent or more of aggregate consumer purchases. But that wouldn't simplify the problem. Simply to calculate for a single individual household the user cost or opportunity cost of the monthly flow of services, for all of the services consumed by that household from the "durable" goods it owned, one would have to collect monthly data on prices paid from a group of households sufficiently large to furnish a statistically valid sample of price quotes for each of those durable goods.

households ideally defined in terms of a number of characteristics—age, race, income, and family composition, as well as geographic location—with the prices weighted by the group's budget shares allocated to the purchase of each of those goods. The sample of households within each subgroup would have to be large enough to ensure that a continuing series of price observations could be collected on individual items, including long-lived items only occasionally purchased by individual households. Moreover, the collection system would have to identify the attributes of the items purchased with sufficient detail to allow BLS to determine—as it now does—whether the particular items that are priced month to month were comparable and to make appropriate substitutions when they were not.

No individual household, during the time interval covered, would itself have bought more than a fraction of the type and quality of goods whose prices are included in a subgroup index. But the index would reflect the distribution of the relevant qualities and prices of goods that were available to individual households in the subgroup, given their income, location, and other characteristics, and the conversion of that opportunity set into a distribution of prices actually paid by households who followed the average search strategy and shopping behavior of the group.

Subgroup indexes stratified by income, by age, or by other characteristics would incorporate not only differences among population groups based on the allocation of their budgets among broad expenditure categories but also differences in the prices and qualities of items purchased and in the kind of outlets at which they were purchased. Indexes could then be calculated for groups classified by income level (e.g., income deciles or quintiles) and combined to give equal importance to each income so as to approximate a democratic index. And, of course, if individual group indexes, say for the elderly or the poor, frequently moved differently from an overall national index, they could be used for indexing public transfer payments going to those groups.

Although the sorting of households into separate subgroups by income, or age, or location is likely to remove a good bit of the heterogeneity—and especially the kinds of heterogeneity that have the most social significance—each index for a group classified by one or two characteristics is still an average across individuals who have differences associated with other characteristics and with idiosyncratic tastes. An index for the elderly would combine rich and poor people, and the index for the poor would combine the old and the young. And, in any classification, the weights implicitly assigned to the prices paid by smokers and nonsmokers, vegetarians and meat eaters, represent an average across the remaining heterogeneity. Yet if data were collected in a way which linked prices paid to the characteristics of individuals, it would be possible to produce special indexes for groups with observable differences in tastes—e.g., those for whom a succession of monthly reports shows no purchases of cigarettes or meat—with the data possibly cross-classified by income group.

Our ability to reduce the heterogeneity problem through the use of group indexes is unfortunately limited by the practical problem of costs. To produce true subgroup indexes, one would have to collect both price and quantity data directly from individuals so that prices and individual characteristics could be linked. Once income data are collected from individuals, it is relatively easy to add simple information on a number of other characteristics, such as age, race, and family composition. But, as we explain below, the size of the sample from which data have to be collected rises rapidly as one increases the number of characteristics that demarcate subgroups. A very large and expensive survey would be necessary in order to produce subgroup indexes cross-classified by many characteristics—e.g., elderly, Hispanic, New England households in the third income quintile, although there may be ways to reduce the size of the required sample (see below).

HOW MIGHT DATA FOR SUBGROUP INDEXES BE ASSEMBLED AND WHAT WOULD IT COST?

Currently, data on prices are collected by BLS in a separate operation from the survey that yields the expenditure patterns used to provide the weights for combining strata indexes into the overall CPI. At the lower (within-stratum) level, the separate Telephone Point-of-Purchase Survey (TPOPS) provides data on the distribution of consumer purchases across retail outlets and on consumer purchases for each of more than 200 relatively detailed categories of goods (see Chapters 1 and 9). But as we have repeatedly stressed, the stratum price index (e.g., for "men's suits and sports coats") that goes forward to the upper-stage indexes is based on price data furnished by retail stores, from Neiman-Marcus to J.C. Penney. In order to construct indexes that reflect individual circumstances and shopping behavior, specific households must be tied to specific items, prices, and outlets. To do so, the current system would have to be radically revised so that data on prices and expenditures for specific identifiable items are principally collected not from sellers but from individual households, so that their demographic and locational characteristics can be linked to the prices they pay.

The first problem to be faced in producing these kinds of subgroup indexes lies in the feasibility of collecting monthly data on prices paid from a panel of households. For certain kinds of purchases, such as utilities, panel reporting should be feasible. BLS already conducts a housing survey to obtain rental prices, which might relatively easily collect periodic economic and demographic household information. But with current interview, telephone, or diary survey techniques, the burden of continuous reporting over a period of months and associated problems of reliability and product identification may well be such that for many categories of goods and services it would prove infeasible. However, there are various, more technologically advanced methods, some of which have been used by private survey firms, that could be investigated to determine whether they could ease the burden and increase the reliability of household price reporting (see below).

The second problem in constructing subgroup indexes relates to the size and cost of the survey(s) that would be required. The size of a monthly panel survey needed to collect price data from individual households that could be cross-classified simply by income, age, and region would be unprecedented. For example, simply distinguishing 5 income and 5 age groups, with no regional classification, would require that prices be collected for 25 separate groups. To keep the burden of monthly reporting within reason, the number of categories of goods on which a household could be asked to report would have to be limited to only a fraction of the 218 CPI strata, the number depending on the frequency with which items within the category were typically purchased.[5] If, on average, each household were limited to reporting on, say, 15 categories, the number of demographic/expenditure category cells would exceed 300. Incorporating a geographic classification would expand this number many-fold. To ensure a continuous supply of price quotes in each stratum for a sufficiently large sample of identical or closely comparable goods, it would be necessary to have a substantial number of households in each cell, since in most strata individual households would not be purchasing an identical item month after month. Without research and testing, the required size of the overall sample can only be guessed, but it would undoubtedly be very large.[6]

There are ways, however, in which the size of the needed sample might be significantly reduced. For circumstances in which a household does not purchase a good in a particular month, an item price might be imputed from a household with partially matching characteristics from an adjacent cell, with only a small loss in precision. Moreover, when research and experimentation identify strata for which the variation in prices paid by households across adjacent and nearby demographic groups is small, the relevant demographic cells might be combined, thereby further reducing sample size requirements. To the extent that the use of handheld scanners and technological aids can be implemented to reduce reporting burden, households could report prices and expenditure data on a larger number of strata, also leading to a reduction in the overall sample size.[7]

---

[5]The TPOPS survey, which does not require price reporting, limits the number of categories assigned to a household to somewhere between 10 and 16. The diary survey of the CEX solicits weekly data from each household on a large number of food and other frequently purchased categories of goods, but only for a 2-week period.

[6]The appropriate sample size would be determined in part by the variance of the prices paid within each cell for the items in particular strata; see Chapter 9.

[7]The number of cells could be modestly reduced if all households within a given demographic group were asked to report on very infrequently purchased items, such as automobiles or major appliances; and for some categories like utilities and public transportation, a common price could be assumed for all demographic groups within a given area (although expenditure data by subgroup would be needed).

The per-household cost of the current CEX survey is about five times greater than that of the monthly Current Population Survey (CPS). So even with a cost-saving sample design, such a collection system would be very expensive. Yet there would be cost offsets. The new system could supplant most or all of the current CEX and TPOPS surveys, and it would provide valuable information useful for other statistical purposes so that not all of the extra costs would have to be charged to the construction of the CPI.

An alternative approach exists for associating the prices paid for specific items with the demographic characteristics of the purchasers.[8] Specifically, a household survey could be periodically used (say every 2 to 3 years) to collect a baseline sample of specific items that were purchased in each ELI or POPS category, with an identification of the outlets from which they were purchased; scanners might be used to get detailed product specifications. The survey itself would secure income and demographic data from each household. BLS field agents would then proceed to collect monthly prices on these items from the identified outlets. The item prices could be assigned back to the appropriate demographic subgroup with the appropriate weights. (Implicit in this scheme is the idea that any given item priced at each outlet might end up being attributed to several or many demographic subgroups but presumably with different relative weights within each.)

The sample would still have to be very large to possess the appropriate number of cells. A rotating sample would have to cover purchases in every month of the year to avoid seasonal bias. And using scanners to enter the product specifications without prices would be just about as demanding as entering them together with prices. But a continuous reporting of monthly prices by households would not be necessary.

While the resulting subgroup indexes of strata prices would reflect the specific kinds and qualities of items purchased and the specific outlets patronized by each subgroup, this data collection system would be unable to take into account differences which might exist among subgroups if they differ in the extent to which they concentrate their purchases at times and in outlets where sales occur. The presence or absence of this kind of shopping behavior may or may not turn out to be an important factor affecting the average prices paid by one group relative to another.

If it turns out to be feasible, the collection of data that tie individual prices to household characteristics would make it possible to determine whether or not the cost of living faced by particular subgroups tends to change at different rates,

[8]The core of this approach was suggested by one of the reviewers of an earlier draft of the panel's report.

sporadically or systematically.[9] As answers to these questions gradually emerged, they could provide important information for researchers and have significant ramifications for government indexing policies, either confirming the validity of using an overall index or suggesting the desirability of using subgroup indexes. Even the collection of data for selected expenditure categories within a few demographic groups, undertaken periodically, could usefully inform public dialogue about social issues.

SUGGESTED RESEARCH AND TESTING

Before even assessing the feasibility of collecting the kind of data we have described, a substantial amount of preliminary research would need to be done. Fortunately, that research itself is likely to produce valuable information about the extent to which rates of inflation in some major categories of goods differ among some subgroups of the population.

A necessary prerequisite to collecting usable price data from individual households is the ability of the collection system to provide the product identification sufficient to enable BLS to match identical items whose purchases are reported by different households and to make appropriate substitutions when items disappear. Thus, examining what is already known about the use by survey respondents of handheld scanners and testing their use and that of other information technology ought to have a high priority. Equally important is exploring the willingness of respondents to record, with a reasonable degree of accuracy, a fairly large volume of information over a sustained period of time, especially when the use of scanners may not be feasible. We have not attempted to outline a formal program of research and testing. But we offer alternative possibilities for collecting preliminary information:

(1) An early project might concentrate on a cross-sectional—rather than longitudinal—study to determine the extent to which, within a selected set of strata, individual households pay different prices for the same items and how those differences are related to age, income, and perhaps other household characteristics. As noted above, differences among households in the *level* of prices paid need not be accompanied by significant differences in the rate of price changes over time. But documenting in some depth the existence of substantial differences in prices paid, systematically related to income and other demo-

[9]Within each population subgroup the aggregation of individual item prices into strata indexes would be done with "plutocratic" expenditure weights, but in each subgroup the strata indexes could be aggregated into a subgroup average with democratic weights, and in turn the overall CPI could be aggregated from the subgroup indexes with democratic weights.

graphic characteristics, would help justify further work and provide clues as to where subsequent effort ought to be concentrated.

(2) Several private marketing firms have established panels of consumer households who use scanners to report prices and expenditures on certain classes of goods, generally those purchased from supermarkets, drugstores, and other mass merchandisers. BLS could work with these firms to investigate the potentialities and limitations of these kind of data for meeting its needs. For example, is the product identification sufficiently precise to track identical items over time and to make and monitor substitution decisions? What is the attrition rate among the panels? How comprehensively can purchases be reported? Cooperative arrangements with these private firms might be helpful in proceeding with the study of price-level differences suggested above.[10]

(3) In what, if any, categories of goods could "unit value pricing" be used as a way of tracking the prices through time? Experiments could be conducted that compare time series of various strata or entry-level item (ELI) indexes already calculated by the BLS with those that would result from unit value type calculations. To the extent that unit value indexes do closely and consistently track previously estimated BLS indexes, the net effect of explicit and implicit quality adjustments has presumably been negligible. If a number of categories do lend themselves to unit value calculation, the sample size and respondent burden of the household survey outlined above could be significantly reduced.

(4) BLS might select a limited number of categories of goods that could not be identified through handheld scanners and construct its own identification dictionary and product codes. Either as a separate survey or as part of the regular diary survey within the CEX, several panels of households drawn from different income groups could be furnished with handheld computers and asked to record, over a period of some months, the prices, quantities, and product codes of items they purchased. This experiment could shed light on several important questions, such as how reliably product identification can be reported and, for particular strata, what sample size would be needed to generate a sufficiently large set of matched price quotes each month. This experiment might be conducted from groups selected within the existing CEX survey.[11]

The results from one or more of these investigations would provide information that would help in deciding whether to proceed further in the direction of a more ambitious pilot project to collect price and expenditure data for one or

---

[10]A recent NBER/CRIW conference considered many of these questions (see Richardson, 2000; Feenstra and Shapiro, 2001; Hawkes and Piotrowski, 2000). For an overview of papers presented at the conference, go to *http://www.nber.org/reporter/fall00/conferences/CRIW.html*.

[11]Independent information about the extent of the longer-term variation in the trend of housing costs among subgroups of households, cross-classified in various ways, could be obtained from the Census Bureau's biennial American Housing Survey.

several expenditure categories for a limited set of demographic population groups. If the decision is to go ahead, further research would be required: to estimate the sample sizes required for various categories of goods in order to yield a continuous time series of item prices; to explore methodologies for handling the disappearance and substitution of items and for constructing strata indexes out of the raw price data furnished by household surveys; and to project the costs of the pilot program.

We have stressed throughout this discussion the difficulties and challenges that would have to be overcome in order to construct subgroup indexes that reflect more than simply different expenditure shares. Yet the arguments for at least examining the feasibility of moving in this direction seem to us to be strong ones. It may well be that the existence of differences in prices paid for the same types of goods by households with different economic and other characteristics does not typically produce significant differences in rates of inflation or changes in living costs. But no one knows. Regional price indexes are of perennial interest and importance. The lack of knowledge about the elderly has been noted throughout this report, and the same arguments can be made with respect to the poor and the rich. We think that, at least, it is worth a modest investment in exploratory research to determine the feasibility and costs of generating the information needed to fill these gaps in our knowledge.

PLUTOCRATIC VERSUS DEMOCRATIC WEIGHTS

As we have described, the sample selection technique used by the BLS, at the lower-level stage of constructing the CPI, implicitly assigns to every price change within a commodity stratum a weight that is proportionate to total consumer expenditures on that item in the base period. It averages these weighted individual price changes to produce a separate price index for each stratum. At the upper level of index production, each of the 218 strata indexes is also assigned a weight—one that is proportional to total consumer expenditures on the types of goods included in the stratum. With this weighting scheme, the purchasing pattern of each household in the nation is implicitly assigned an importance in the overall index that is proportional to its total expenditures on consumer goods. As we noted at the beginning of this chapter, the resulting CPI reflects the consumption patterns of upper-income households to a greater degree than those with low incomes. For this reason the current CPI has been described as a *plutocratic* index.[12]

[12]The procedures used by every other national statistical agency in the world to produce their equivalent of the CPI are all plutocratic in nature. This fact is a testament to the very real difficulties that would be faced in collecting the data required for true subgroup indexes.

If one could construct subgroup indexes classified by income levels along the lines described above—say, income quintiles—a simple average of those subgroup indexes would generate an overall CPI in which the inflation experience of households at each income level was assigned equal importance. This in turn would produce a close approximation to a *democratic* index. Even under the best of circumstances, however, it will be a long time before the research program we have suggested above could tell us whether the production of subgroup indexes is feasible with acceptable costs and reliability. But by using data that are already available, it is possible to go part way toward producing an index with democratic weights.

The CEX, from which the information about household expenditure is collected, combines data from two different survey components. One uses diaries, kept by households, to collect data on categories of goods that are purchased on a day-to-day basis, such as food, household supplies, toiletries, and the like. Another group of consumers is surveyed through interviews to collect expenditures on the types of items that are less frequently purchased—rent, appliances, medical care, clothing, etc. Rather detailed demographic information, including income, is collected for each household in the survey. The information from the two surveys can be combined to calculate for each household an allocation of consumer expenditures among some 146 categories of goods over a period of several years.[13] Then an overall index can be computed for each household, weighting the individual stratum price indexes by the budget share devoted to that stratum by each household.[14]

Those indexes could then be combined into an overall CPI, giving equal weight to each. This procedure would go part of the way toward producing a democratic index. It would be partly democratic because the upper-level weights used to combine the strata indexes give equal importance to the budget allocations of each income group. But it would not be a full democratic index, because at the lower level of index construction the individual item prices would not be linked with the income levels of the households who purchased them and would continue to be combined into strata indexes using weights proportional to total expenditures for those items by all consumers. To look at it another way, the stratum prices indexes would be the same for everyone, while the budget shares

[13]For technical reasons, the individual strata within which expenditure data are collected in the diary survey have to be collapsed into fewer categories than the overall CPI, which has 218.

[14]Individual household indexes can be computed at the upper level only on the assumption that all households face the same prices and differ solely in the allocation of their budgets among the different strata. Given that assumption, the same stratum price index, based on prices collected from retail outlets, can be assigned to every household. Constructing such an index is not inconsistent with the proposition put forward earlier in the chapter that, under current data collection procedures, a fully democratic index—which reflects the fact that different households face different prices for the same goods—cannot be produced.

used to combine the indexes into an overall CPI would give equal importance to the spending patterns of each individual—despite the fact that upper-income households contribute a substantially larger fraction to overall consumption than those with lower incomes.

Over the past several decades, a number of studies on inflation rates calculated with expenditure weights for different income groups over varying time periods have found that the differences are slight (Michael, 1979; Hagemann, 1982; Blank and Blinder, 1986; Kokoski, 1987; Garner et al., 1996), although a few earlier studies suggested that larger differences may sometimes arise.[15] A more recent study, by BLS economist Mary Kokoski (2000), estimated an index for the years 1987 through 1997 along the lines outlined here and compared the results to the regularly published CPI-U. As is very often the case when comparing aggregate indexes that differ only in upper-level weighting patterns, there was little difference between the average rate of change in the two indexes. Over the 10-year period, the democratic index rose at an annual rate 0.05 percent faster than the plutocratic version. But for various subintervals the differences were larger: the democratic index rose 0.5 and 0.3 percent faster from 1988 to 1990 and from 1995 to 1997, but 0.2 percent slower from 1990 to 1995.

The Kokoski article also estimates indexes for each quintile of the income distribution.[16] Differences in the performance of the democratic and plutocratic indexes during the various sub-intervals reported above were associated with large differences in the rate of inflation faced by households in the top and the bottom quintiles. From 1988 to 1990 and 1995 to 1997, the index for the bottom quintile rose 1.6 and 0.7 percent a year faster than the index for the top quintile, while in the 1990-95 period the experience was reversed, with the index for the bottom quintile rising 0.9 percent a year slower. As would be expected, deviations from the change in the overall democratic index during these intervals were almost always a good bit smaller among the middle three quintiles than in the top and bottom ones.

While they lasted, the differences in the change between the upper- and lower-quintile indexes were quite large, especially since the indexes captured only the difference in budget shares and not any differences in prices paid. The fact that the differences reversal themselves several times within the period does

---

[15]Deaton and Muellbauer (1980) found a 2 percent greater-than-average inflation for the poor in Britain over the high inflation period of 1975-1976. Kuznets (1966) found that over the long term, as income grows, food prices, which loom large in the budgets of the poor, rise faster than prices of manufactured goods. But some members of the panel believe that the modern speed of innovation in product design, and the likelihood that the rich can take greater advantage of product improvements, would tend, in a heavily quality adjusted index, to show an index for the rich that rises less rapidly than one for the poor.

[16]Kokoski estimates a democratic and plutocratic index for each quintile, but as might be expected when households are segregated by income levels, they exhibit only very small differences.

not guarantee that divergent price behavior is always likely to be so short-lived. We suspect it would not be very expensive to produce quintile indexes on a regular basis—annually, if not monthly—which might prove to be very useful for public policy purposes and would alert us when significant differences reappear. And while the overall quasi-democratic CPI did not depart in any substantial way from the regular plutocratic CPI, the same process that produces the quintile indexes would provide an ongoing quasi-democratic index that would indicate if differences did emerge.

Choosing Between the Indexes: Does It Matter?

There are uses for the CPI or its components in which plutocratic weighting is called for—the component indexes of the CPI are used in deflating current dollar consumer expenditures as part of producing measures of real gross domestic product (GDP). And it is probable that a plutocratic index would come closer than a democratic one to the weights appropriate for indexing the tax system. But for most purposes a democratic index would be preferable. For analysis of economic welfare—e.g., measuring changes in real median incomes—a democratic index would clearly be superior. And that is equally true for the index used to determine cost-of-living allowances in social security and other public transfer programs. Rough calculations cited by Deaton (1998) suggest that the household "represented" by the plutocratic CPI is around the 75th percentile of the income distribution. And it is hard to imagine that anyone would deliberately make decisions about public pensions by tracking households at the 75th percentile of the income distribution.

The fact that, in the past, indexes weighted democratically at the upper level have not tended, over any substantial time period, to move differently from the plutocratic CPI is no guarantee that the future will always produce the same result. The Kokoski article shows that households at the opposite ends of the income spectrum have, at least over short time periods, experienced significantly different rates of inflation simply due to the different allocations of their budgets. On the assumption that, aside from the initial set-up expenses, the costs of maintaining the production of such indexes would not be large, continuing production of such supplemental indexes seems a worthwhile task.

SUMMARY AND RECOMMENDATION

Households differ from one another in their consumption patterns and shopping behavior and often pay different prices for the same goods. Part of this heterogeneity is associated with differences in household economic and demographic characteristics and in their geographic location. This fact gives rise to two kinds of issues: (1) For adjusting social security payments and the tax system or for measuring changes in real income, when can data for the whole population be aggregated into a single official price index? When are different price indexes for specific population subgroups needed? And how should data to produce such subgroup indexes be collected? (2) When a single overall index is produced, how should the costs of living of individual households be combined into a single national index? Should equal weight be given to each household's cost of living or, as is now the case, should the individual costs of living be weighted by the overall consumption spending of each household?

The Consumer Expenditure Survey indicates the extent to which various economic and demographic groups allocate their budgets differently among categories of goods and services. However, substantial variation may also exist among different groups of households with respect to the particular types and qualities of goods they purchase and the prices they pay within each category. But because the price data used to produce the CPI are collected from retail stores and not directly from households, it is impossible to associate the economic and demographic characteristics of buyers with the items they buy and the prices they pay. As a consequence, it is impossible to investigate satisfactorily the two major aggregation issues we identified: To what extent does inflation or changes in living costs differ among the various economic and demographic groups? And to what extent would a democratic index behave differently from a plutocratic one?

With current survey techniques and methods, collecting price as well as expenditure data from households on a scale sufficient to produce the CPI and an array of group indexes would be extremely expensive and possibly even infeasible. We therefore propose something more modest:

Recommendation 8-1: BLS should pursue an exploratory research program that would, initially only on a small scale, investigate and assess several alternative approaches—including, but not limited to, the use by survey respondents of handheld scanners and computers—for collecting prices in a way that allows them to be associated with household characteristics. A first objective might be the production of indexes for a few commodity categories and several demographic groups.

TECHNICAL NOTE 1: AGGREGATION AND THE "REPRESENTATIVE CONSUMER"

The concept of the "representative consumer" frequently comes up in discussions of COLIs and of price indexes more generally. Indeed, it is often difficult to discuss COLIs with non-economists, policy makers, or the public at large without some sort of appeal to the concept. Sometimes the use is ambiguous or implicit: For example, a COLI might be presented in terms of the amount of money needed to keep consumers, or even "the consumer," as well off as before the price change. Or it might appear in thinking about the change in expenditure that would be necessary to offset the effects of inflation on "consumer living standards." Similar phrases are often used to describe substitution effects in response to price changes. Sometimes the language refers explicitly to the representative consumer, sometimes to a "typical" or "average" consumer.

Pollak (1981), in a paper on social cost-of-living index numbers, showed how to define an aggregate or average COLI without any appeal to the existence of a representative agent. For example, there is a well-defined COLI for each person or family, and one can average them. This would be a democratic COLI and could be approximated by a democratic Laspeyres index (for example). Or one could work with a plutocratic COLI in which the amounts of money needed to hold living standards constant are added up over all consumers and compared with the sum of actual expenditures for all consumers. Neither of these constructions involves representative consumers of any sort, and neither is very difficult to understand, certainly not the former.

Nevertheless, the idea of a representative agent is often appealed to, though we have tried hard to avoid it in this report. One danger of the usage is that it is easy to fall into the trap of thinking of the welfare of the representative agent as representing the welfare of everyone. For example, when one talks about the COLI as being calculated (and perhaps paid, as in social security) so as to keep "consumer living standards" constant, it might be taken to mean that everyone's living standards are being held constant. Instead, the best one can hope for is that some average of living standards is being held constant, with some people gaining and some losing. These distribution effects of price changes can sometimes be important.

Another danger of the idea of a representative consumer is that it distracts attention from the need to think explicitly about how to aggregate over different people and families. The economic theory of consumer behavior is a theory of individuals, not of groups, and the analytical results that come with it are results about individual behavior. The theory provides many insights about such topics as substitution effects, the cost of living, and welfare. It provides an apparatus to think about substitution effects and why, when price goes up, demand goes down, as well as some less obvious results, such as the equality between good i's substitution response to the price of good j and good j's substitution response to the price of good i. But these results are for individual consumers, not for the aggregate or average of consumers. As we discuss in Chapter 2, the theory can be used to think about cost-of-living index numbers for groups or nations, but the transition from the individual to the group is not straightforward, and it requires a good deal of explanation. So it is sometimes tempting to avoid the complications, and to apply the theory to average or aggregate behavior, thinking about the country as a whole as a "representative consumer."

In this technical note we discuss two issues. First, what has to be true for the representative consumer to exist, in the sense that the analytic fiction will give the same answers as working with the underlying individuals and thinking about the aggregation explicitly? Second, if the conditions are satisfied, and one computes a COLI for a representative agent, what relationship does this COLI bear to the COLIs of the underlying consumers?

Conditions for the Existence of a Representative Consumer

We begin with a definition. The representative consumer of consumer theory is not defined to be representative of tastes or of levels of living but of behavior. The economic theory of consumer behavior works with a single person who is assumed to be greedy (always wants more), to make consistent choices, and to do the best to satisfy his or her desires within a fixed budget at a fixed set of prices. As a result of doing the best he or she can, each person will have demand functions that relate the amount demanded of each good to income and to the prices of all the goods. Because different people have different tastes and live in different socioeconomic and physical environments, these demand functions will generally differ from person to person, even if they have the same level of income and face the same prices. For the economy as a whole, one can sum (or average) the demand functions of each person to get the aggregate (or average) demand for each good in the economy as a function of prices and the incomes of each person in society. The representative consumer exists if this aggregate can be thought of as coming from a single consumer whose behavior replicates the aggregate of all consumers. Note again that this definition is in terms of behavior, not tastes or welfare.

In general, this construction cannot work, and for a very obvious reason. According to the theory of individual behavior, demand is a function of income and all the prices. But aggregate demand is a function of *all* the incomes and all the prices. The distribution of income between people matters for the aggregate but has no part in the theory of individual behavior. The representative consumer demands goods according to her representative income and the prices, and there is generally no way in which all the possible effects of the distribution of income can be captured through a single representative income. So the representative consumer cannot exist in general.

There are special cases where the distribution of income has no effect on aggregate demands: when each consumer spends his or her marginal dollar in exactly the same way. If so, taking a dollar from A will shrink A's demands for goods in exactly the same way as giving the dollar to B will expand B's demands for goods. As a result, the total demand for each good in the economy depends only on total income, not on who owns that income. Total income will work as representative income, and the representative agent exists.

How realistic is this condition? For it to hold exactly is clearly absurd; no one would seriously claim that everyone in the economy spends an additional dollar in exactly the same way so that, at the margin, all consumers are identical. A more serious question is whether, at the level of aggregation in the CPI (about

200 goods), the approximation is good enough. However, the equal spending condition has (at least) one unpalatable consequence. It can only be satisfied if everyone buys every good. If A does not smoke, and B does and would smoke more if he had more money, then redistributing money from A to B will change the demand for tobacco, and such behavior cannot be accommodated with only a representative consumer. It does not require any sort of econometric analysis, or appeal to the data in the CEX, to know that people consume different subsets of goods (even at the 200-plus commodity level), so that aggregate demand must depend on the distribution of income. If one insists on the fiction of the representative agent, one will be blind to changes in the social cost of living that are brought about by changes in the distribution of income. Conversely, one will also be blind to changes in the distribution of income that are brought about by changes in relative prices. An increases in the price of tobacco redistributes real income from (relatively poor) smokers to (relatively rich) nonsmokers. The representative consumer approach does not recognize such a possibility.

Whose Cost of Living Does the Representative Consumer Represent?

Suppose, contrary to the argument above, that the conditions hold that allow one to construct the analytic fiction of a representative consumer. One then constructs a COLI for this fictitious person to compare prices in period 1 with prices in period 0. Because the representative consumer's behavior is the average of the behavior of each consumer in the economy, one might hope that the representative agent's COLI is the average of the COLIs for the individual consumers. Note that nothing in the construction guarantees this. The representative consumer was constructed to represent average behavior, not the average cost-of-living index. And, in fact, the result is not true. The COLI for the representative agent is the *plutocratic* COLI obtained by averaging the individual COLIs with weights proportional to individual incomes. That this should be the case is intuitively clear from the fact that the representative consumer is constructed to represent average demand and that the rich contribute more to the average than do the poor, simply because they are richer and so spend more. Even though the representative consumer's purchases of each good is a simple average of individual purchases, the representative agent's COLI is not the simple average of the individual COLIs. Consequently, the use of a representative consumer framework in the context of plutocratic weights assigns more importance in the overall index to changes in the cost of living facing the rich than to those facing the poor.

There are alternative definitions of the representative consumer that get around the plutocratic bias. For example, one could average, not the quantities purchased by each consumer, but their budget shares, defined as the fraction of their budget allocated to each good. One could then ask whether it is possible to construct a representative agent with a representative level of income whose budget shares are always equal to the average of the budget shares for each consumer in the economy. The existence of such a representative agents has been investigated by Muellbauer (1975, 1976). In many ways, the conditions to make this story work are less restrictive than those for the original representative consumer, and indeed Muellbauer's representative consumer has a representative income that depends on the distribution of income as well as on its mean. However, it is unclear whether the additional complexity of these formulations would commend them to those seeking straightforward interpretations of the COLI concept of a price index.

Technical Derivation of the Representative Agent COLI

Unless we place restrictions on the distribution of income, the existence of a representative agent requires that individual h has preferences that can be represented by cost functions of the Gorman "polar form" (Gorman, 1959):

$$c^h(u^h, p) = a^h(p) + u^h b(p), \tag{1}$$

where u^h is utility, p is a vector of prices, and $a^h(p)$ and $b(p)$ are nonnegative linearly homogeneous and quasi-concave functions of p. Taste variation is permitted in the function $a^h(p)$ but not in $b(p)$. The representative agent has a cost function that is the average of (1), which is

$$c(u, p) = \bar{a}(p) + ub(p). \tag{2}$$

Denote by x^h the total expenditure of h.

Suppose that the two price vectors to be compared in the COLI are p^1 and p^0. The base-period COLI for h, P^h_{10}, is written

$$P^h_{10} = \frac{a^h(p^1) - a^h(p^0)b(p^1)}{x^{0h}} + \frac{b(p^1)}{b(p^0)}, \tag{3}$$

while the representative consumer's COLI is

$$P_{10} = \frac{\bar{a}(p^1) - \bar{a}(p^0)b(p^1)}{\bar{x}^0} + \frac{b(p^1)}{b(p^0)}, \tag{4}$$

where $\bar{x}^0$ is the population mean of x^{h0}. Straightforward computation then confirms that

$$\sum_h x^{0h}(P^h_{10} - P_{10}) = 0, \tag{5}$$

so that the representative consumer's COLI is the plutocratic average of the individual COLIs,

$$P_{10} = \frac{\sum_h x^{0h} P^h_{10}}{\sum_h x^{0h}}. \tag{6}$$

If we had started from the current period COLI, holding utility at period 1's utility, instead of using base utility from period 0, the relationship between the representative agent's COLI and the individual COLIs is given by

$$\tilde{P}_{10}^{-1} = \frac{\sum_h x^{1h} (\tilde{P}_{10}^{h})^{-1}}{\sum_h x^{1h}}, \tag{7}$$

where the tilde denotes a current base COLI. Note that (6) and (7) also hold for both Laspeyres and Paasche indexes. In (6), if the indexes on the right-hand side are replaced by the individual Laspeyres indexes, the index on the left is the plutocratic Laspeyres, and in (7) the same holds true for the individual and plutocratic Paasche indexes. Indeed, these indexes would be the obvious choices to approximate the "true" COLI concepts if one wanted to measure them.

TECHNICAL NOTE 2: DO INFLATION RATES DIFFER BY AGE OR INCOME GROUP?

Since price indexes are used to adjust benefits paid to well-defined demographic groups, such as the elderly or the poor, it is important to consider the extent to which inflation rates for individuals in these categories differ from those faced by the general population. If purchasing patterns diverge widely *and* if the prices of goods and services that mark this divergence change at significantly different rates, the idea of creating group-specific subindexes becomes compelling. If consumption bundles are proportionally similar or if price changes across group-differentiated bundles consistently balance out, index disaggregation may be superfluous.

Because of the obvious policy implications, age and income-specific subindexes have been given the most attention. This emphasis is reflected in both the academic literature and BLS policy research. BLS produces an experimental index for the elderly as a means to assess the validity of using the CPI-W to index social security benefits.[17] Attention has also been given to separate indexes for the poor and, more generally, to price index variation by income group. Empirical studies of the income-price inflation relationship often bear, at least indirectly, on the closely related issue of plutocratic versus democratic indexes discussed above. This technical note reviews empirical literature that assesses price variation across subpopulation groups.

[17] It is worth noting that not all elderly citizens receive social security, while many non-elderly do receive benefits. Nonetheless, it is curious that social security is indexed to CPI-W, which specifically excludes households whose primary source of income is from retirement and pension accounts (the CPI-U, in contrast, includes all urban consumers, including those who are retired).

Subindexes for the Poor

Government poverty programs and guidelines are regularly adjusted for inflation. The Census Bureau poverty thresholds and the Department of Health and Human Services poverty guidelines, food stamp programs, low-income housing, and home energy assistance programs are all adjusted using the CPI-U. However, because the CPI is plutocratic, the representative household is upper middle class, which means that price changes—as captured by the CPI—are potentially very different than price changes faced by "average" low-income households.

Early on, in work for the Joint Economic Committee, Arrow (1958) pointed out that separate subindexes for different income groups might be appropriate for certain policy applications. He reasoned that observed consumption patterns, most notably the proportion of necessities to luxury goods, are likely to be quite different for low- versus high-income households. Subsequent research has been directed toward generating empirical evidence to ascertain whether or not divergent group consumption patterns do translate into significantly different group inflation rates.

Snyder (1961) pioneered work contrasting the growth rates of experimental indexes for high- and low-income groups. For the period 1936-1955, she estimated Laspeyres indexes for food items—categorized by income and income-food commodity elasticity—purchased by population subgroups. Price growth for low-income (and low-income elasticity) items was generally greater than price growth of middle- or high-income items. However, she also constructed a Paasche index series from 1955 Department of Agriculture food expenditure data that revealed no significant variation across income groups.

Snyder showed that, during the period's recessions, prices of items that constituted high-expenditure shares for the poor declined more slowly than did prices of goods in general. During expansions or periods characterized by high inflation, the price growth of low-income items outpaced the price growth of items purchased proportionately more by middle- or higher-income households. Kuznets (1962) corroborated a specific component of this relationship, documenting a time trend indicating that, as income rises, food prices rise relatively faster than prices of manufactured goods. Deaton and Muellbauer (1980) estimated that in Britain during the high-inflation period 1975-1976, the inflation rate was 2 percentage points higher for the poor than for the general population.

More recently, BLS has tracked price inflation for the poor using item share weight-adjusted indexes. Garner et al. (1996) report results derived from the BLS experimental price index. The stated goal of the program is "to determine whether such an index would be lower than, higher than, or equal to the current CPI-U" (p. 32). In constructing the index, CEX data are used to calculate item category expenditure weights that reflect consumption patterns of the poor. The poor are defined three ways: by program participation, by household expenditure levels, and by income. The authors compute weights using each definition and then compare trends for experimental Laspeyres, Paasche, and Fisher price indexes for 1984-1994.

Garner et al. (1996) found only slight differences between price trends produced by the experimental price index for the poor and the full sample CPI-U. Using 1984 as the base year, the 1994 all-consumer unit Laspeyres is 141.1; the reweighted indexes for the poor range from 139.8 to 140.7, depending on which definition is used. The Paasche and Fisher indexes vary from the all-consumer index by a similar magnitude.[18] The authors conclude that "the poor and the general population have faced similar trends in relative prices over the last several years" (p. 41). In addition to being time specific, they further qualify the results to acknowledge data limitations—e.g., the large share of rural poor do not figure into the calculations and the possible existence of asymmetric substitution opportunities in poor versus high-income consumption bundles.

Michael (1979) empirically examined the effect of demographic factors on price indexes and estimated the statistical significance of the correlation between the two. Using individual-level records from the 1960-1961 CEX, he regressed Laspeyres index values against household demographic characteristics. The equation produced a number of significant coefficients, but no clearly discernable variation between the inflation rates faced by specific income groups and the population sample as a whole. There was no obvious relationship between household income and relative position in the distribution of index values over time. In a similar study, Hagemann (1982) looked at group variation in Laspeyres indexes. The study generated some evidence indicating slightly higher inflation for poorer households, but, again, the results were generally statistically insignificant.

In additional BLS research, Kokoski (1987) constructed a superlative Tornqvist index in order to examine income-specific effects for the period 1972-1980. Differences across groups were generally small and insignificant. Blank and Blinder (1986) round out the available evidence. As part of their investigation into income distribution and poverty and commodity purchase patterns by the poor, they conclude that price inflation faced by the poor was similar to that faced by the general population over the period 1947-1982.

The balance of the evidence, then, points to either modest or no variation in inflation rates faced by different income groups, particularly for more recent periods. Consumption patterns—specifically the relative weights of necessity versus luxury items—may be different, but the differences do not translate into consistently bifurcated subindex growth rates. Even if there is wide-ranging price

---

[18] The paper also includes estimates of substitution bias, computed as the difference between the Fisher ideal and Laspeyres indexes. Substitution bias was estimated to be 1.99 percent for all consumer units, 1.75 percent for the income poor, 2.01 percent for the expenditure poor, and 0.25 percent for the program participation poor.

inflation for item categories that are weighted very differently across subindexes, there may be no clear pattern in which specific price changes would tend to cancel one another out. It is also possible that the available household data are simply inadequate to tease out a significant income-inflation rate relationship. Of course, the bulk of the research identified here is empirical; there is no obvious theoretical basis to assume that the relationship between inflation rates and income group will diverge more or less in the future. Future research may be productively directed toward examining the extent to which suspected index biases correlate to household income. For instance, economists have long argued that quality change and, hence, quality change bias may be more prominent among luxury goods (which would presumably give CPI-type indexes an upward bias for high-income groups). Boskin et al. (1998) challenge the notion that benefits from quality improvements and new products accrue disproportionately to the wealthy; however, there is little empirical documentation to forcefully support either assertion.

Subindexes for the Elderly

Social security is by far the largest government outlay directly adjusted using the CPI. This, along with the perception that the elderly are more vulnerable to adverse affects associated with price inflation, has stimulated research emphasizing this group. Also, the CPI for medical care, a comparatively important component of elderly expenditures, has increased more rapidly that has the overall CPI in recent years; however, measuring medical care costs is extremely complicated and it is hard to assess the accuracy of this CPI component.

The most systematic evidence on inflation faced by the elderly has evolved from a 1987 congressional directive to BLS to develop an experimental index for the population over age 62. In testimony to Congress, Mason (1988) reported the first set of results calculated under the program. For the period 1982-1987, the CPI-U, which captures spending habits of approximately four-fifths of the U.S. population, rose 18.2 percent; the CPI-W, which captures a subset of about one-third of the population rose 16.5 percent; the experimental index for the elderly (CPI-E) rose a slightly higher 19.5% (Amble and Stewart, 1994).

Amble and Stewart updated the results for the ongoing indexing program. For the period 1987-1993, the CPI-U rose 26.3 percent, the CPI-W rose 25.5 percent, and the experimental index for the elderly rose slightly more, 28.7 percent. Stewart and Pavalone (1996) completed the series through 1995, producing similar results. For the period 1990-1995, the CPI-U rose 14.7 percent, the CPI-W rose 14.1 percent, and the CPI-E rose 15.9 percent.

BLS's experimental index consistently produced slightly higher inflation rates for the elderly during the 1980s and 1990s. However, this does not necessarily mean that the elderly have truly faced more rapid increases in living costs. To understand potential inaccuracies of the CPI-E as a true cost-of-living index for the elderly, one must review the BLS index construction method. For the CPI-E, BLS identifies expenditure patterns for the sample of elderly from CEX data. The standard modified Laspeyres index is calculated using a reweighted consumption basket that reflects those patterns. However, as Amble and Stewart (1994:141) report: "The experimental price index for older consumers is a weighted average of price changes for the same set of item strata and [is] collected from the same sample of urban areas used in calculating the CPI-U and CPI-W." Thus, the selection of outlets, as well as the selection of specific item categories to price, may not be representative of those used by the urban population age 62 and over.[19]

The BLS reports also note that, relative to the CPI-U, the CPI-E has a higher sampling error since it is constructed from a smaller sample. Also, the CPI-E does not capture the effect of nonfixed percentage senior citizen price discounts. Nor does it account for higher rates of home ownership among the elderly. Boskin et al. (1998) argue that, because of the rental equivalency indexing method, homeowners are, in effect, getting compensated for capital gains on their homes.

Out-of-pocket medical care expenses account for the majority of the growth rate differences between the CPI-E and the CPI-U; and many economists believe that the medical care component is among the most biased item categories (if the goal is a cost-of-living indicator), due to omitted quality effects and output definition problems. The Boskin commission argued that widespread and systematic quality improvements in the health care sector are not captured by the CPI, creating a significant upward bias in the medical care component—about 3 percent per year when weighted by out-of-pocket expenditures. In short, though the CPI-E has risen more rapidly than the CPI-U, one still cannot estimate relative cost-of-living trends.

To summarize, BLS research shows that the CPI-E series rose slightly faster than the general CPI. However, the CPI-E is computed using a comparatively small CEX sample, and the differences are generally not statistically significant. Also, the growth differential between the CPI-E and CPI-U is attributable to increased weighting of a few item categories, most notably medical services, an item category economists agree has poorly captured improved quality and new item effects.

The non-BLS literature generally concludes that there is a lack of measurable divergence between elderly and general population price inflation trends. Using the reweighted Laspeyres index approach, Boskin and Hurd (1985) found little difference in the cost of living faced by the elderly and the general population during the early 1980s. Jorgenson and Slesnick (1983) arrived at a similar conclu-

[19]Of course, this says nothing of the more general problem that, to the extent that price changes faced by the rural elderly are different than those faced by the urban elderly, inaccuracies are compounded.

sion using a method that attempts to estimate changes in cost of living directly from sets of demand curves representing different demographic groups.[20] Berndt et al. (1998) looked at actual transactions data involving purchases of pharmaceuticals (antidepressants, calcium channel blockers, and antibiotics). For the period 1990-1996, the authors showed that, relative to younger age groups, the elderly faced rates of price inflation that were slightly higher for antibiotics, slightly lower for antidepressants, and about the same for calcium blockers.

Conclusions

To date, researchers have been unable to compellingly support claims that age- and income-defined population subgroups face significantly different rates of price inflation relative to the general population. On the contrary, during the short period for which reliable data exist, little divergence has been found. However, there is no theoretical rationale to assume that these trends must remain constant over time.

It is important to note that, for the most part, data have been available for research that tracks group indexes differentiated only by item category weights. The Jorgenson and Slesnick article, which estimates separate systems of demand equations for different demographic groups, is an exception. On balance, little is known about exactly how quality and substitution biases in current measures may affect subindexes differently. Thus, it is difficult to assess group-specific cost-of-living trends using currently available experimental index measures.

[20]This approach is unfortunately prohibitive at any detailed level of disaggregation since the number of parameters in the system to be estimated rises proportionately with the square of the number of items in the index. Another method for estimating substitution bias that allows for detailed item disaggregation is used by Garner et al. (1996). The method entails comparing Laspeyres indexes, which assume no substitution bias, with Paasche indexes, which weight item categories using the most recent period's consumption and probably overstate substitution.

9

Data Collection for CPI Construction

The data used to calculate most Consumer Price Index (CPI) subindexes originate from three different, though interrelated, sample-based sources: the Consumer Expenditure Survey (CEX), the Point of Purchase Survey (POPS), and the Commodities and Services (C&S) Survey.[1] These surveys have evolved through time, and one of them, the CEX, has multiple uses in the government statistical system. Moreover, the data collection structure itself influences what indexes can and cannot be produced. For example, the current data system does not allow for production of non-urban-area indexes or regional price-level comparisons; nor does it support accurate price indexes for subpopulations such as the elderly, minorities, or the poor, particularly at subnational levels. Also, in order to reduce respondent burden, households are only asked about a portion of CPI item categories, which also inhibits the construction of some potentially useful, alternatively weighted (e.g., democratic) indexes.

There are two distinct approaches one can take when considering how the data underlying CPI computation might be upgraded. One is to assume that the basic data collection structure will remain as it is and then to seek ways of improving each of the survey components. Another is to redesign the entire data collection structure so that it reflects advances in data collection technology and so that the data collected are more consonant with the ultimate computation of the CPI. This second option would require a transition plan that takes the data system

[1] In addition, the CPI housing survey is used to calculate changes in rent of primary residence and owners' equivalent rent (the two largest components of shelter).

from where it is today to where we aspire it to be. In this chapter, we discuss these two options.

THE CURRENT DATA COLLECTION PROCESS

The Consumer Expenditure Survey

The CEX is the primary tool for establishing CPI weights at the basic (218) item level. It is the most comprehensive source of combined household income and expenditure data produced by the statistical system; it is also very expensive to conduct. Nonetheless, a growing consensus is emerging among policy researchers that improvements should be made to the CEX. Probably the most frequently voiced criticism has been that the sample size is too small for the survey to be used for the range of applications to which it is currently put. However, another shortcoming of equal or greater importance—at least in the context of CPI construction—stems from nonsampling-related inaccuracies, such as survey response bias. Suggestions for improving the survey's questionnaire design and substantive scope can be found in the research literature; in this section we review recent recommendations for upgrading the CEX, after first posing several questions that must be answered before a fully informed decision to change the survey can be made.

Accuracy

The panel's foremost concern is with the extent of bias in the CEX which, in turn, affects the accuracy of CPI expenditure category weights. A starting point for evaluating household expenditure allocations estimated by the CEX is to compare them against weights generated by other sources. The Bureau of Economic Analysis (BEA) produces the most obvious alternative, the per-capita personal consumption expenditures (PCE) data, as part of the national income and product accounts (NIPA). During its postsurvey evaluation program, designed to identify areas in which the CEX could be improved, the Bureau of Labor Statistics (BLS) does compare the expenditure pattern of the CEX with that shown in the PCE component of the NIPA (Branch, 1994). Such comparisons might, depending on the outcome, raise a second question: Why not use, for the national CPI, upper-level weights derived from aggregate-level data, such as the PCE?[2]

[2]We note that a price index is already constructed using the PCE—the chain price index for personal consumption expenditures, or PCEPI. See Clark (1999) for a description of the differences between the PCEPI and the CPI in terms of index formula, scope of goods and services covered, underlying price information, and index performance.

In considering such an option, one must (1) judge whether or not the PCE weights are really superior for this application and (2) if they are, determine whether the CEX would still be needed for the CPI program. The answer to the second question depends in part on the value placed on area and group indexes, which could not be constructed using NIPA data. Budgetary considerations aside, there is no inherent reason why BLS could not produce a flagship CPI using NIPA-based upper-level weights, while producing other indexes based on the CEX.

Let us first address the accuracy issue. Branch (1994) provides a comparison of CEX and PCE expenditure categories for the period 1992-1995. The comparison is limited to the universe of categories that are comparably defined; this leaves out two major ones—owner-occupied housing and health care. For a few categories, such as rental rates, utilities, and vehicle purchases, the correspondence ratio (CEX weight divided by PCE weight) is near 1.00, which is what one would generally hope for. For equivalent rent of owner-occupied dwellings, the CEX expenditure weight is much larger than shown in the PCE data, by a ratio of almost 2 to 1. For all other categories, though, the CEX expenditure weight is much smaller, and many are in the 0.4-0.6 range. This discrepancy calls into question the accuracy of the CEX weights. There is also a problem (documented in Triplett, 1997) that the total expenditure of households implied by CEX and PCE weights is drifting further apart, perhaps by as much as 1 percent a year. One should not jump to the conclusion that these differentials imply an accurate PCE and an inaccurate CEX, but the wide discrepancies clearly warrant further investigation since both sets of expenditure weights cannot be correct.

What is known about the relative strengths of the PCE and the CEX data? For certain types of expenditure categories, well-documented sources of household response error damage the credibility of CEX weights. Triplett (1997:15) states that "reporting biases are known to be serious in some consumer expenditure components." For instance, households may underreport "vice" products such as alcohol or tobacco—for 1995, the ratio of CEX to PCE expenditure shares on alcoholic beverages was a dismal 0.34. In addition, survey respondents often fail to accurately recall the volume or timing of some frequently purchased items: for example, "other entertainment" has a correspondence ratio of 0.37, and "miscellaneous" has a ratio of 0.24. For a number of other categories, such as furniture or appliances, it is less clear why expenditure weights differ as sharply as they do between the PCE and CEX. (The ratio of CEX to PCE weights for the "household furnishings and equipment" expenditure category was 0.65-0.66 for the 1992-1995 period.) Here the problem may involve the PCE as much as the CEX. Businesses and governments buy furniture and appliances, but do not necessarily report (or categorize) these purchases in their accounting systems in a consistent way that allows them to be accurately identified and reported. The fact that other items (e.g., books, televisions, sound equipment) that are purchased broadly by both households and businesses—and for which it makes no obvious sense for households to underreport—show large weight differentials that might support this notion (Triplett, 1997). For other components, such as rent or auto purchases, for which reporting rates are known to be high, it is encouraging that the ratios of CEX to PCE weights are close to 1.

There are ways in which the PCE data system appears more developed. The PCE has the advantage that it is based on large surveys of businesses (the most prominent being the Census Bureau's Retail Trade Surveys) that generally keep careful records and that rely less on respondent memory than does the CEX. Triplett (1997:16) does note a "birth bias" in the establishment surveys that arises because there is no mechanism for bringing new businesses into the sampling frame quickly. However, data from the censuses of manufacturers, retail trade, and service industries allow PCE component weights to be revised periodically and benchmarked every 5 years, which surely corrects some reporting and other biases. The benchmarking resets the allocation of purchases by commodity among business, government, and households and updates commodity lists. Furthermore, the BEA methodology for keeping track of inputs and outputs includes cross-checks that impose consistency on the data.

A major advantage of the CEX weights is that they are derived *directly* from reported household expenditures. One benefit of this direct reporting is that it allows household characteristics to be linked to expenditure information and, in turn, subpopulation indexes such as the CPI-E and CPI-W to be calculated. To produce the PCE weights, business and government spending must be subtracted out of sales data. Thus, the PCE is an *indirect* measure, calculated residually as final goods and services minus purchases made by nonconsumer sectors. Triplett (1997:16) notes that it is especially difficult to calculate consumption shares at more refined item levels because sales to consumers are not always distinguishable from sales to businesses and government: "The finer the level of detail, the more likely that the long chain of computations necessary to reach the CPE's indirect estimate of consumer spending will have cumulative errors that affect the totals." Even so, it seems implausible that estimates of business purchases of consumers goods could be off by enough to generate the kind of ratios between NIPA and CEX weights that are now produced.

Difficulties associated with separating business from consumer purchases are compounded by the fact that the PCE covers a wider scope of goods and services than does the CEX. For instance, PCE coverage includes elements of government consumption, such as Medicare and Medicaid, the employer-paid portion of medical insurance, financial services, expenditures by nonprofit institutions, and the value of certain goods and services received in kind by households (Clark, 1999). As discussed throughout this report, the CPI currently covers only out-of-pocket expenditures by urban households. All told, about 25 percent of PCE spending is not reflected in the CPI. This, in itself, redistributes expenditure shares substantially. For instance, the medical care category (since it is not limited to out-of-pocket expenditures) gets a much higher weight in the PCE—

17.6 percent for 1998—than it does in the CPI—5.6 percent for 1998. Also, not all items are defined comparably: in the CEX, for example, expenditures on new cars net out any amount paid for a trade-in vehicle; the PCE tracks the gross amount paid for vehicles, and trade-ins are not taken into account (Clark, 1999).

This imperfectly matched expenditure classification creates a major hurdle to producing a PCE-weighted CPI, though Branch (1994) was able to make adjustments to reduce the noncomparability. For instance, utilities in the PCE can be combined with rent in order to match rent as it is defined in the CEX. But the following categories could not be reconciled for purposes of comparison: homeowner shelter (owners' equivalent rent, as we noted earlier, has recently accounted for around 20 percent of expenditures in the CEX and only around 11 percent in the PCE), capital improvements, health care, insurance, and finance charges (Branch, 1994). Assuming that the basic CPI item structure will remain as is, it is not clear how this problem should be resolved. To maintain the current CPI scope, the additional PCE entries would need to be backed out. Furthermore, BEA actually uses CEX data to estimate expenditure for a small number of commodities—personal computers, vehicle rentals, day care (Triplett, 1997)—which is another reason why moving to PCE weights might not allow the CEX to be eliminated.

On the basis of available evidence, it is unclear whether PCE or CEX weights are superior. What is clear, though, is that for some components the two systems produce very different results. The major hurdle inhibiting comparisons among indexes weighted using alternative source data is the lack of uniformity in the scope and definition of goods and services covered. It is an open question as to how accurately expenditure categories can be mapped from the PCE to the CEX. We are not in a position to advocate one set of weights over the other, but the question certainly warrants further investigation—and this is what we recommend in the final section of this chapter.

Frequency

The CEX is used by BLS to determine the base period household expenditure shares for each of the 11,772 basic CPI strata. The CPI has traditionally determined these quantities from a 3-year span of CEX data; current weights reflect expenditure shares calculated from the 1993-1995 surveys, with immediately prior weights based on the 1982-1984 surveys. In 1998 BLS announced that it would update and apply 1999-2000 expenditure weights effective January 2002 and revise these weights every 2 years, instead of roughly every 10, as has been its prior practice (see the "Technical Notes" at the end of the chapter for additional details about the CEX). To accomplish this objective—which necessitates combining only 2 years of survey data instead of 3 and increasing the per-year number of basic CPI strata for which quantity information is obtained—and to maintain roughly the current level of statistical accuracy, the sample must be increased by 50 percent. The recently requested increase in the CEX sample size from an effective annual sample size of 5,870 to about 7,500 per year approximately does this.

The decision to update CPI-U and CPI-W expenditure weights every 2 years beginning in 2002 was based on a tradeoff between timeliness and concern about "chain drift," which can occur when the price indexes of non-identical items must be linked.[3] BLS agreed with critics (such as Boskin et al., 1996) that the weights should be updated more often than every decade or so as in the past, but little theory or empirical evidence existed to provide guidance on the optimal frequency of updates. BLS chose to move to the more frequent end of the spectrum, every 2 years. There were some operational issues that argued for not updating every year. For instance, BLS reports that there is an advantage to having "off years" in which changes in CEX forms can be implemented without the time pressure of employing the data in the CPI. The main reason, however, was simply that the approach of updating weights every year, which would require overlapping 2-year CE weights, was untested and its statistical properties were uninvestigated. BLS noted that, in its experience, changing index formulas can produce unexpected and undesirable results, so it decided to err on the side of caution by not going to annual updating.

Sample Size

The CEX targeted sample sizes are 6,160 per quarter for the Quarterly Interview Panel Survey and 5,870 per year for the Diary Survey. Because an increased sample size will produce an increase in the precision of an unbiased estimate, recommendations to increase the CEX sample size (primarily directed at the Diary Survey) have tended to be of the "more is better" variety. However if, as we have pointed out, the weights from the CEX are not unbiased, a decrease in sampling variability might actually increase mean squared error, which is what we ultimately care about.

A recent report from the Conference Board recommended increasing the annual sample size of the CEX "perhaps initially to 30,000 households." The

[3]Index (chain) drift refers to the possible bias that can arise when separate price indexes are linked. For example, suppose there are three periods, 0, 1, and 2. A price index could be computed for period 2 relative to period 0 in one step using fixed weights, or a "chain index" could be computed by multiplying the price index from 0 to 1 by the price index from 1 to 2. If each price is stochastic but stationary around a fixed level, or all prices are stationary around the same trend, so that relative prices vary in the short run but not in the long run, chain indexes are likely to be biased in comparison with fixed-base indexes. If relative prices return to their period 0 value in period 2, the chained index will generally differ from unity; this difference is the chain drift.

recommendation was supported with the following comments (Conference Board, 1999:18):

> This [the currently proposed increase in the sample from 11,000 to 15,500] seems inadequate. The comparable Canadian survey, to cover an economy only one-tenth our size, is reported to cover 16,000 households. Sampling theory shows that optimum sample size is *not* proportional to the overall size of the economy. Nevertheless, we feel that a larger U.S. Consumer Expenditure Survey will achieve a worthwhile gain in the accuracy of the weights used in the CPI. A larger [CEX] sample will also improve other important statistics, such as poverty data.

While the intuition of the members of the Conference Board group might be right, their recommendation would have been weightier had they been able to offer a statistically based explanation of how they concluded that the benefit of increasing the sample to 30,000 consumer units would be "worthwhile" and that the value of gains in accuracy from increasing the sample would likely offset the cost of expanding the CEX.[4] Likewise, no indication is given as to why and by what measures, in the group's estimation, the planned 15,000-15,500 household sample is unacceptably inaccurate. The Conference Board did note two factors underlying its recommendation that the CEX should be larger: that it would be valuable for improving other national statistics (besides the CPI) that are based in whole or in part on the CEX (e.g., poverty and savings rates) and that a larger national sample is needed to calculate subgroup indexes.

In a somewhat more careful statement (but one lacking a specific recommendation), Triplett (1997:15) writes:

> . . . The [CEX] sample size (5,000 consumer units) is certainty too small for almost any use for which one wants consumption data. . . . The recently announced increase from 5,000 to 7,500 [CEX] consumer units is a positive, but grossly insufficient, step. . . . The [CEX] is the federal government's only general purpose survey of consumer expenditure. . . . For comparison, the Canadian consumer expenditure survey will soon have a sample size of 36,000.

He adds:

> . . . The [CEX]'s small sample size and lack of a benchmarking statistic means that its estimates for smaller components (e.g., household textiles) particularly are not as reliable as one would want for serious research on consumption. Also, the weights for the *individual* 207 basic components of the CPI are not

[4]The report is a bit unclear in its references to sample size. One can infer though that the 30,000 figure refers to the desired sample size to be used by BLS when it uses 2 years of CEX data to establish expenditure-base quantities; if so, the Conference Board is really recommending an increase from the current annual effective CEX sample size of 5,870 to double that of the proposed 7,500, to 15,000 per year.

determined accurately from a [CEX] of only 5,000 consuming units, although it may also be true that the variance imparted into the *overall* CPI may be small.

Increasing the CEX sample size would also enhance its ability to support other potential uses. For instance, the current sample is not large enough to verify trends among population subgroups needed for a consumption-based poverty measure—especially for specific regions. Also, because it is the only U.S. survey that generates income and consumption microdata together, it is important for research on household savings behavior and on how that behavior varies along with age, income, and other factors. Other agencies (the Congressional Budget Office and the Department of the Treasury) use the data for modeling tax revenues and other research purposes. Better data would certainly improve research prospects in these areas as well.

Finally, in defense of BLS's request for the approximately 50 percent increase in the CEX sample size, Commissioner Abraham testified (Bureau of Labor Statistics, 1998) that the increase "would let us produce superlative measures to a degree of precision comparable to the precision of the current CPI. . . . We currently use three years' worth of data in producing the market basket weights for the CPI. For the superlative measures, you use two years' worth of data, so with a 50 percent increase in sample size, we would have about the same precision in the weights." But there is no reference to the targeted level of accuracy, nor of the impact of the increase in sample size on the precision of the *current* CPI computation. The commissioner's statement merely says that the increase in sample size will enable BLS to estimate a CPI with the similar variance characteristics as those of the current CPI computation.

Given the current state of assessment, it is difficult to offer recommendations about the sample size of the CEX. The most pressing practical issues require weighing the cost of expansion against the advantages that changing the survey sample size would have on the accuracy of expenditure weights and, in turn, the relationship between weight accuracy and index variance. The panel carried out some analysis on this issue. The variance of the CPI, reflecting both the variance of the aggregation weights from the CEX and the price relatives from the C&S survey, is referred to by the BLS as the unconditional variance. The variance of the CPI, reflecting only the variance of the price relatives from the C&S survey (and treating the aggregation weights from the CEX as constant), is referred to by the BLS as the conditional variance. The ratio of the unconditional to the conditional variance is of the form

$$1+\frac{quadratic\ function\ of\ aggregation\ weight\ cov\ mx}{conditional\ variance\ of\ index}=1+\frac{q}{c}.$$

Thus, the impact of increasing the sample size of the CEX by a factor of f will result in a change in the ratio of unconditional variance to conditional variance to

$$1+\frac{q/f}{c}.$$

Let t be an observed ratio of unconditional to conditional standard error. Then the impact of increasing the sample size of the CEX by a factor of f on t will be to reduce t to

$$\sqrt{1+\frac{t^2-1}{f}}.$$

Leaver and Valliant (1995:Table 28.2) provide ratios of the median unconditional to median conditional standard errors across time for major item groups for the period January 1987 through December 1991. Our panel received an update of these ratios for 1998 and 1999 for the "all items" and eight major item group series (Leaver, private communication). These ratios range from essentially 1.00 to a high of 1.23 (for the 1998 apparel index based on a 12-month price change). The all-items ratio ranged from 1.03 for an index based on a 1-month price change to 1.09 for an index based on a 12-month price change.

To see the impact of an increase in the sample size of the CEX, consider an extreme case—in 1998 the 12-month price change in the apparel CPI had a conditional standard error of 0.00811844 and an unconditional standard error of 0.00997372 for a ratio of 1.22853. Doubling the CEX sample size would have reduced this ratio to 1.120107. The apparel CPI went from 131.6 in December 1997 to 130.7 in December 1998. The 95 percent confidence interval for the December 1998 apparel CPI, based on the 12-month change from December 1997, is (128.1, 133.3). The comparable confidence interval, based on a doubled sample size, would be (128.3, 133.1). The panel therefore concludes that there is little evidence to support the recommendation to double the sample size of the CEX.

All this speaks to the index as a whole. One might also want to study the effect of increasing the CEX sample size on the variances at the basic CPI strata level. Acceptable (and optimal) error and variance levels must be defined specifically for the types of indexes that are desired; only then can they be evaluated against the cost of expanding the survey.[5] In other words, one needs to determine the appropriate level of disaggregation at which to assess the effect of a change in sample size of the CEX.

[5]Specifically, one needs to know the range of sampling and nonsampling errors for different index components. Nonsampling errors are caused by failure of respondents to understand survey definitions, their unwillingness to provide correct information, collection and response errors, and a number of other sources. Presumably, sampling errors can be reduced to a much greater extent by increasing sample size than can nonsampling errors.

The following list summarizes research that should be taken into account before BLS statisticians can definitively target an efficient sample size for the CEX:

- Accuracy. As discussed in the previous subsection, it makes little sense to pursue more precise estimates of a biased measure. Differences in expenditure shares estimated by the CEX and PCE must be better understood and (at least partially) reconciled.
- Precision level. If it is established that CEX is the best option for setting expenditure weights, BLS should establish precision requirements for the expenditure weights. The requirements must be informed by an understanding of how precise the CPI needs to be in terms of estimating the level and trend of the index. A primary driver for the sample size would be the extent to which population subindexes are desired. Precision requirements must also be established for other important uses of CEX as well, which may also have demographic or geographic dimensions.
- Cost of expanding the survey sample. The cost of CEX operations should be examined in relation to survey size and design characteristics. BLS and the Census Bureau have a fairly accurate idea of how much it costs to expand the sample size (they now have the experience of the 50 percent increase). In addition to sample size, there may be a considerable clustering effect (both in terms of statistical performance and cost) in the CEX. What is the optimal scheme for clustering surveyed households and designating sampling units? Also, since the survey has many uses other than for BLS weighting, evaluations should consider whether BLS should bear the full budget burden of future changes to the CEX; a cooperative effort shared by the Office of Management and Budget, BEA, the Federal Reserve Board, etc., may be more appropriate.
- Value of CEX. To redesign the CEX, or to expand its sample size, one needs to place a value on the inventory of all of its key uses. We know the CEX data are fundamental to the CPI. All uses, including the CPI, must be considered in making recommendations about the design or size of the CEX.

Other Issues

In addition to questions of frequency, sample size, and accuracy, there are a number of additional issues that involve assessing the information content of questionnaires and the general structure of the CEX. Many of the issues have already been addressed to varying degrees by the BLS and others. Improving the CEX will involve continued assessment of the effectiveness of the interview and diary survey approaches, what methodologies minimize underreporting of purchases or attrition from a diary panel, the appropriate universe of households and goods and services to be covered, and the role of incentives programs in increasing survey accuracy and reducing nonresponse. It will also require answers to questions about how the mode of data collection might be modified to take advantage of new computer-based data collection methods, whether all expenditures for all item categories should be collected from all households surveyed or just some from each, and what processing system is required for the CEX in order to expedite development of a superlative index.

Answers to all three types of questions hinge on the types of indexes that BLS will be called on to produce. For instance, there are increasing demands for subpopulation and geographic (both price level and price change) indexes. Recommendations for modifying the CEX can only be reasonably determined after the BLS and policy makers decide the importance and value of calculating these special-purpose indexes. Assuming different expenditure weights apply to each, a much larger CEX sample will be required.

The Point of Purchase Survey

A second major survey input to the CPI is the POPS, which is used to determine which outlets BLS data collectors will visit in the C&S survey to record actual prices.[6] The POPS produces outlet-specific expenditure information for item categories so that a sample of those outlets can be selected with a probability proportional to consumer use. The POPS is needed because the CEX does not ask consumers where they purchased goods. In addition to its role in selecting outlets to which BLS agents go to price specific items, POPS expenditure data are also used to implicitly assign quantity weights to all items priced within a single item strata (see Cage, 1996:fn. 14 for details). Within the current data support system, the POPS data have been improved in terms of their effectiveness at identifying outlets where households shop and as an input for averaging price quotes within CPI item cells.

The entry-level items (ELIs) in the CEX are not isomorphic with the POPS categories. Thus, some concordance and other adjustments are necessary to match the quantities from the CEX with the prices and price relatives determined from the C&S survey driven by POPS. In a newly designed data system, it seems likely that this mismatch could be eliminated.

There is a substantial overlap between POPS and CEX. If the CEX had no use other than to provide upper-level weights for the CPI, it would make sense to redesign POPS so that it would be the survey vehicle to perform this function as well. This change would then allow for greater index design flexibility, but it would probably increase the sample size required in POPS and also increase the response burden for each participating household.

[6]The POPS provides sample outlets covering items that account for about 72.5 percent of the CPI (as measured in expenditure shares). A housing survey is used for shelter components of rent and owners' equivalent rent, and other sources are used for a few other commodities and services (see the "Technical Notes" at the end of the chapter for additional information about the POPS).

The Commodities and Services Survey: Outlet Pricing

The number of price quotes that are collected is determined at the ELI and index-area level in a process called sample allocation. The stated objective of sample allocation is to produce the most accurate national-level all-item index possible, given the budget constraints. Through this process, item strata in each area are assigned a minimum number of price observations. In practice, this means that sampling rates are dictated, and will be higher, for ELIs that represent a large expenditure weight or display high price variability, as is the case with such items as apples and bananas (Lane, 1996).

The CPI's C&S is a longitudinal survey that tracks changes in price quotes for most CPI-sampled consumer items over time.[7] A few price quotes come from other sources: for instance, the CPI housing survey performs the same function for the shelter category. As described in Chapter 5, the specific items for which (and outlets from which) the C&S samples price quotes are rotated simultaneously. The POPS provides the sampling frames for outlets by producing estimates of expenditures for items in specific POPS categories (corresponding roughly with strata) at specific outlets. Based on POPS results, specific ELIs are assigned to each sample outlet. Each ELI has a checklist of product specifications so that a BLS field agent can identify specific items from the ELI category that are sold at the selected sample outlet. Field agents select a unique item (from within the preselected ELI category) for pricing based on a probability distribution of sales, with high-expenditure items (within that outlet) being more likely to be selected than low-expenditure items. The process whereby an agent narrows down the list of potential items from the ELI group to a specific item is called disaggregation (see Lane, 2000:9, for details). After a unique item is selected, the agent returns to the same outlet every month (or, in some cases, every 2 months) to record the price change. This process is repeated as long as the outlet continues to sell the item or until the outlet is rotated out of the sample. If the item is permanently discontinued, the agent consults a "characteristics" checklist and determines the most comparable replacement to price.

As discussed elsewhere in the report, problems may arise with this pricing system—for instance, if an item is first priced when it is on a special sale or when a specific item remains on store shelves long after a large reduction in its market share. BLS continues to explore methods for improving the quality of price data. The most visible experimental activities involve expanding the use of electronic data, which may offer such advantages as larger samples, reduced variances, more accurate determination of in-store sales shares, more timely publication of superlative indexes, and the potential to use unit pricing.

[7] This paragraph summarizes the description of the C&S survey from Lane (2000).

ALTERNATIVE DATA COLLECTION APPROACHES

Since most options for improving CPI input data, particularly those involving the household surveys, are expensive, and because there is methodological inflexibility under the current system, it is worth considering entirely new data alternatives. Of course, any net benefit of these alternatives hinges on exactly what types of indexes are desired—COLI or fixed-basket, national or regional, plutocratic or democratic, aggregate or subgroup. Other than the PCE-based expenditure weighting possibility, the two most obvious options for breaking from the current data system involve (1) combining POPS and CEX into an integrated survey that contains expenditure and outlet-use data at detailed product levels, along with household demographic information needed for subgroup indexes; and (2) moving toward scanner-based collection systems, which could be used to improve the existing surveys or as a component of an alternative. Current experimentation by BLS using scanner data illustrates its potential within the existing framework. Integrating scanner data as part of a POPS/CEX combined survey, or into a comprehensive household-based pricing system, would entail more radical shifts in CPI methodology.

One advantage of restructuring the entire data support apparatus would be that it could be designed to fulfill current indexing needs. However, as the environment and uses of the index change, even such an optimal data system moves toward obsolescence unless it is much more flexible than current systems. In this section, we examine some approaches to improving the data support system under the assumption that radical changes are one option.

An Integrated CEX/POPS Survey

The CEX and POPS were introduced at different times and evolved out of different needs in an uncoordinated way. The CEX was developed to provide detailed data on household-level expenditure patterns. BLS has been producing expenditure surveys in one form or another since the late nineteenth century; however their production was sporadic (usually not more often than every 10-20 years) in the early part of the century and was motivated by a range of different needs. The 1960-1961 survey was constructed with the primary purpose of revising weights for the CPI and was not limited to urban wage earners, as had typically been the case with previous surveys. The 1972-1973 survey was the first to use the modern interview and diary components, and the sample was selected on a probability basis (Jacobs and Shipp, 1990). The POPS was introduced to provide information about where consumers shop—information not provided in the CEX nor from existing sources of business sales-level data. Also, existing lists were typically based on the Standard Industrial Classification (SIC) system, which is not concordant with BLS-defined ELIs.

Because both CEX and POPS are household-based surveys, it is natural to consider the possibility of merging the two into a single survey. Intuitively, it seems there should be economies of scale in combining them, as well as advantages to having more complete records (both expenditure and shopping pattern data) for each household. While we do think this possibility is worth investigating, there are many complicating factors. To begin with, the reference periods are now different for the two surveys. The quantity weights from the CEX require updating over a longer periodic cycle—formerly every 10 years, but now moving to every 2 years (without necessarily implying a change in the item structure every 2 years); outlet rotation weighting, based on POPS, is done every 4-5 years on average and, since POPS is a continuously rotating survey, a subset of items and areas is considered for change every quarter. Whether or not these are optimal frequencies has not been determined. It is possible that adequate rotation and weighting schemes could be produced from a single survey, but at present the issue remains largely unexplored.

The level of item detail needed to obtain CPI item strata weights and to select outlets and ELI samples is also different in the two surveys. Since POPS asks about product expenditure in greater categorical detail, it is generally believed that it requires a larger sample size to produce accurate probability schedules. It is possible that a unified survey could partition respondents into two or more groups, with some being asked more detail than others (something akin to the census short and long forms). Respondent burden could also be reduced if each household continues to be asked only about a subset of CPI items.

Defenders of the current system could also point out that a combined survey that generates expenditure, demographic, and outlet information concentrates respondent burden unnecessarily. Detailed demographic information is missing from the current POPS; outlet usage information and adequate sample size are missing from the CEX. A combined survey would likely entail greater demands on any given respondent, and the CEX is already considered one of the most burdensome government surveys. There is also a range of data quality issues that would require investigation. The CEX sample may be more representative of the population since it is based on samples drawn from census household files, not on random digital telephone sampling as is the POPS. Each CEX household also reports on a larger share of total household expenditures than does a POPS respondent. Further complicating the issue is the fact that the CEX is used for research and policy purposes other than the CPI.

The most obvious advantage of the multisurvey data system now in place is that—relative to the size of expensive consumer surveys—a large number of price quotes can be generated (and at a reasonable cost) for each specific item that is ultimately tracked by the CPI. This is because price data are not linked to specific households. Households provide just enough information for BLS to assign weights to broad item categories and to identify high-use outlets. If prices had to be gathered from households in the manner laid out in Chapter 8, the survey would presumably have to be much larger (than either the current POPS or

CEX) to ensure an adequate sample of prices for each ELI area cell for the CPI. Yet the real advantage of a survey that links prices paid for specific items to the purchasing households is that, in principle, from such data one could calculate average prices paid for specific items by different household types. The big question is what size household sample would be required to support such an index or, more realistically, how big a sample would be needed to make an experimental pilot project work. This question is discussed in Chapter 8.

Scanner Data

In this section we outline how scanner data work and identify some potential operational and measurement benefits that may be gained by increasing their use; we also point out limitations. However, reflecting the panel's charge, the primary emphasis is on how the use of scanner data (and electronic data in general) might allow greater conceptual flexibility when constructing price or cost-of-living indexes. The discussion comments on the extent to which current BLS research and experimental programs may affect CPI pricing procedures. The panel also assesses the value of incorporating scanner-based pricing methods within the context of its more general recommendations concerning the feasibility and advisability of pursuing a COLI approach. We first look at the potential of point-of-sale scanner data and how it could be used to improve data accuracy and price collection procedures. We then look at the more futuristic idea of household-based scanner data.

Point-of-Sale Scanner Data

The most obvious way in which scanner data could be used to support the CPI would be as a replacement for or supplement to the C&S survey of outlets. Scanners in retail outlet checkout counters record Universal Product Codes that identify specific products and their manufacturers. These data are collected, collated, and sold by two major producers of scanner data: ACNielsen and Information Resources, Inc. (IRI).

A growing literature on the topic is beginning to provide an indication of the feasibility, as well as the benefits and drawbacks, of using scanner data in the production of price indexes. While academic researchers in both the United States and Europe have begun exploring how scanner data could be used to improve the statistical properties of price indexes, BLS has moved to the forefront on work in the area.[8] Reinsdorf (1996) successfully constructed a basic item-level index for coffee using scanner data. Currently, the BLS's ScanData initiative is producing

---

[8]See, for example, Richardson (2000), Bradley et al. (1997), and Reinsdorf (1996).

indexes for breakfast cereal in the New York City area from data provided by Nielsen. To date, Laspeyres, Tornqvist, and geomean indexes have been produced; Paasche and Fisher indexes are under consideration. The BLS team is moving to construct the index for broader geographic areas as well. As additional areas are added, they will use current CPI aggregation weights and Laspeyres formula (Richardson, 2000:11).

Scanner data offer several potential advantages. First, such data could streamline item pricing procedures. Using computer-captured scanner data could reduce the number of manual steps in the C&S survey required to produce subaggregate indexes. Scanner price data may replace or reduce the need to visit stores to price items.

Second, scanner data could generate a more representative selection of items for pricing. Scanner data include the universe of products sold (at outlets that have scanner technology), whereas the current quote sampling method only records prices for a small fraction of items on store shelves. CPI price quotes are drawn from items at outlets made eligible by selection in the most recent POPS sample. Scanner quotes are available if the item has been sold during the pricing period. For the CPI, BLS collects prices for selected items whether or not they have been sold at the POPS-identified outlet. In contrast, transactions scanner data pick up volume of sales. Some stores also maintain files that drive the price identification system and indicate the shelf price of all items for some period, such as a week, whether or not they were sold. CPI outlet and item samples are rotated periodically, every 4 years under current practice. In contrast, since scanner data can include the universe of transacted prices at covered outlets, samples are refreshed continuously and new items appear in the data much more frequently. For the BLS's ScanData geomean and Laspeyres test indexes, weights and item samples are updated each year on the basis of the previous year's expenditure patterns (Richardson, 2000).

Third, scanner data could improve sampling accuracy. Scanner data have introduced new capacity to calculate highly accurate average prices for specific commodities. The large number of outlets and item price points associated with scanner data offer the potential to greatly decrease sample variance and improve data precision. As pointed out by both the Boskin commission (Boskin et al., 1996) and the Conference Board (1999), the high volume of scanner data would allow for production of indexes at finer levels of product detail. Additionally, scanners record actual transaction prices, not shelf prices at which transactions may or may not have taken place for the relevant period. These features may help certain data users, particularly those that perform industry studies or types of analyses where average price movement over fairly short periods is more relevant than shelf price at a given point in time. The tentative result of BLS's ScanData New York experiment—which provides some evidence as to how far these scanner data may improve underlying data quality—have been quite promising. The indexes produced from scanner data have displayed less variability than the CPI sample price counterpart. For cereal in New York, the sample size of price quotes is more than 1,400 times the number in the traditional CPI data. However this translates into a reduction of standard errors, such an increase should create greater index precision. Though it was surmised (Richardson, 2000) that this would reduce the standard error of the cereal CPI by a factor of about 38, a careful study of these data by Leaver and Larson (2001) shows that the reduction in the standard error was by a factor of about 6.

Fourth, scanner data could expand geographic coverage for the CPI. Nielsen compiles scanner data from all states and regions (except for Alaska). Data from nonmetropolitan-area outlets are also available. In contrast, the CPI uses data from only 87 metropolitan areas.

Fifth, scanner data may allow more systematic data-cleaning procedures. Scanner data are more uniform and may be simpler to process for index use. However, data-cleaning rules used by ACNielsen or IRI are different from those at the BLS, particularly in how missing or erroneous prices are imputed. This would become an issue for any index that uses scanner data only for a subset of item categories, while traditional methods continue to be used for the remaining item categories.[9]

Finally, with scanner data, it will be possible to produce price averages (or unit valuations). Scanner data allow transaction prices to be averaged over the relevant period. Unlike BLS pricing methods, scanner datasets are typically produced using aggregated unit values—a quantity-weighted average price of an item. The simplest version is calculated as sales revenue divided by number of units sold. Unit values are used in most basic item indexes in the world; however, this is not the case with the CPI, since a weight is assigned to each price quote (Richardson, 2000).

This last issue, unit pricing, requires further discussion, since it is not transparent that it is conceptually superior to the current practice of pricing items on store shelves at a point in time. The main criticism of unit pricing is that it produces a price at which no single item may actually have been sold.[10] On the other hand, the ScanData team argues that "the unit value index more accurately

[9]See Richardson (2000) for a summary of how scanner data were cleaned for use in the ScanData indexes.

[10]This is the case because stores sell the same item at different prices, which then are averaged. Unit values may be the average of prices over a time period, across some set of outlets (like an outlet chain), or even across product codes that have only minor differences in characteristics. Multistore unit pricing implicitly accepts the assumption that consumers switch easily between outlets in response to price changes. One practical advantage is that chain- (in contrast to individual store-) level data are less expensive to produce. Commenters (such as Diewert, 1995) have expressed reservations about this pricing approach. In the ScanData cereal experiment, outlet- and "organizational"- level data have been very similar.

reflects the preferences of the shopper who searches out the lowest prices each week, and also the consumer who stockpiles during a particularly good special, but then purchases nothing until the next special" (Richardson, 2000:12). In some instances, few consumers purchase at the shelf price that the BLS agent happens to observe. How many people buy Chicken-of-the-Sea tuna fish when the Bumblebee next to it is on sale for half price? Feenstra and Shapiro (2001) cite marketing literature indicating that there is substantial consumer substitution across weeks in response to price changes and advertising. Also, their own data on canned tuna show a high degree of price variation and substantial response of consumer demand to that variation (Feenstra and Shapiro, 2001). Using shelf prices assumes rigidity in consumer shopping behavior, since items in each week of pricing are treated independently and that elasticity of substitution among them is zero (Richardson, 2000). Proponents of unit value pricing argue that is it better to consider purchases in different weeks of a month as purchases of the same good in the context of consumers' utility. It is certainly worth noting as well that, at some level, price averaging *must* take place to construct any price index.

Whatever the outcome of these specific questions, it is clear that scanner data allow researchers to look at all sorts of interesting things. They facilitate comparisons of series that combine price data in different ways, including alternative index formulas, such as short time-lag superlatives. The ScanData team, for instance, was able to compute several indexes contemporaneously (using a Paasche construction as the lower bound with which to test other indexes). Additionally, the sheer volume and detail of scanner data also facilitate hedonic analyses of quality change (such as Ioannidis and Silver, 1999). Even when scanner data are ultimately not used to construct an index, availability of the data can only advance the pace of research that leads to improvements in the index generally.

Early results for the ScanData cereal test indicate that introduction of scanner data may have a significant effect on index performance. For the February 1998 through June 2000 period, cereal inflation for the New York metropolitan statistical area, as measured by the CPI, rose from (a re-based) 100 to 101.1. The geomean scanner index completed the series at 104.9. This 3.8 percent difference may have been attributable to several factors. First, the universe of outlets for the two indexes was not identical; ScanData was missing data from a wholesale club. There was also a sharp decrease in the regular CPI for cereal in October 1999 that did not appear in the scanner data and is difficult to explain. Also, the Tornqvist index rose more rapidly than did the geomean, indicating that, at least for cereal in New York, elasticity of substitution is less than 1.0, as assumed under the geomean method (Richardson, 2000).

It is also important to assess the extent of practical advantages of scanner data that might add to the viability of its regular use. The ScanData experiment has produced favorable results in a number of areas showing, specifically, that:

- Scanner indexes can be produced on the current CPI schedule. Regarding quote timing, CPI and scanner data cover similar periods within the month; scanner data have the advantage of covering weekends and holidays, which CPI data do not.
- For many cases, scanner data cover the entire domain of products within any given item strata and area cell, which is important for methodological consistency.
- The scanner indexes can be produced in a manner generally consistent with BLS sampling procedures.
- The sample is rotated and can be refreshed at least as often as under current CPI practices.
- Indexes work with both standard geomean and superlative formulas.

The cost implications of introducing scanner data and reducing field price observations have yet to be fully evaluated by BLS.

Limitations of Store-Based Scanner Data

Despite the numerous potential advantages described above, issues remain to be sorted out before BLS can proceed toward systematic integration of point-of-sale scanner data into the CPI; these issues relate to pricing, coverage (both geographic and item-specific), cost, integration of scanner data with other data sources, and reliance on private-sector data.

In addition to unit valuation (already discussed), pricing issues include treatment of taxes and comparability between private-sector scanner data and Census Bureau/BLS data. The CPI collects prices without sales taxes; then a calculated tax is applied separately using secondary data. Scanner data also do not include taxes. However, since ACNielsen does not disclose the exact location of outlets, it is not always clear what tax rate should be added to item prices. For the cereal experiment, it was not a problem since New York has no tax on most food items. However, in general, a solution to this problem needs to be found by vendors or BLS. One possibility would be to calculate a population-weighted average sales tax each month for each item based on the outlet usage patterns of consumers in each geostrata (Richardson, 2000).

Coverage issues include geographic definitions and saturation of scanner equipment. Geographic-area definitions for the CPI and for currently produced scanner datasets do not match. Scanner data markets are generally smaller than the census-defined metropolitan areas on which the CPI is based. ACNielsen is currently working to map most of the United States into CPI geographic areas, though when the project is complete, there will still be some gaps (e.g., ACNielsen does not cover Anchorage). Even for the covered areas, scanner price data are not available for all outlets at which items from any given CPI strata are sold. In the cereal experiment, there were CPI quotes that were not included in the scanner universe (in this case, they were from mass merchandisers). Small mom-and-pop stores also frequently do not use scanner technology. Efforts are currently under way at ACNielsen to expand the depth of outlets covered in its datasets. Also, "migrating" quotes come into play when purchases are made across CPI areas. The POPS sample covers purchases in adjacent areas, but these patterns cannot be inferred from scanner data. In other words, the POPS covers purchases of consumers from a certain area while the scanner datasets cover purchases made by any household in a particular area, which is not the CPI objective. It may be possible to construct a scanner index as a weighted index from the areas in which consumers of a given area shop (Richardson, 2000), but this certainly adds complication back into the system.

Scanner data coverage is most broad based for items sold in supermarket outlets, while there is virtually no coverage in service sectors. Hawkes and Piotrowski (2000:1) of ACNielsen report that 43 of the 211 CPI item categories can "in large measure, be represented through scanning data obtained from Supermarkets, Mass Merchandisers, and Drug Stores." These categories account for about 10 percent of all consumer expenditures and about 24.2 percent of expenditures for goods (excluding services such as rent). Item coverage constraints alone severely limit the impact that use of store-based scanners can have on the overall CPI.

In terms of cost, the budget tradeoff between purchasing data from private vendors and traditional price data collection must be evaluated, as BLS is in the process of doing. Another issue concerns integration of scanner-based subindexes (possibly superlative) with traditional sampling-based item indexes: What are the statistical and index performance ramifications when subindexes are compiled using different types of data?

Finally, BLS currently does not have to rely on private outside sources for fundamental pricing data. The ramifications on CPI production of changing this must be explored. For instance, ACNielsen and IRI buy their data from chains, and at times chains decide to no longer sell these data. This means that, while a given store has a positive probability of being in the traditional CPI sample, its probability of being in the scanner dataset is zero. Thus, problems of continuity with the scanner data universe could arise.

Household-Based Scanner Technology

Household scanner technology could be adopted in one of three ways: it could be used to improve the accuracy and coverage of the current household surveys, particularly the CEX; it could also be used in a combined CEX/POPS survey; or, more ambitiously, it could be the technical centerpiece of a household-based panel survey that produces both expenditure share and price information that would be used to produce household or subgroup indexes. Any plan to augment the CEX would require members of sampled households to use handheld scanners to report UPC items and quantities. These data could be enhanced by having the household members key-enter prices as well. BLS could develop scannable menu codes for non-UPC items, which the sampled household could then use to help enter quantities and prices. In addition to this information, household members could be asked to report the store name and address associated with each purchase. One might bypass some of the household recording of prices if the reported store is one from which prices for UPC items can be obtained directly.

Potential to Enhance Accuracy of the CEX and POPS

Even before considering price issues, household scanner devices could increase the quality of current surveys by improving the accuracy of households' documentation of purchases. It could produce more accurate and detailed weighting from the item strata to sub-ELI levels. Household scanning technology could help reduce errors associated with improper identification of products and prices and reduce recall and incomplete information (about location of purchases, for instance) biases.

The technology creates greater breadth and depth of information by tracking product and buyer characteristics and offering more uniform geographic coverage (rural areas, all age groups, etc.), thereby expanding the potential to develop subgroup indexes. It could cover purchases made at outlets that do not use point-of-sale scanning, and even in *sectors* that do not, if supplemental code sheets could be developed for respondents to scan. Lastly, household scanner technology may reduce respondent burden.

The possibility also exists that new types of errors (e.g., keying) could be introduced; this possibility would have to be examined in pilot projects. Pilot projects would also be important for determining whether introduction of this technology into the survey affects the demographic composition of the sample (e.g., bias it away from inclusion of the elderly).

Scanner Technology as a Tool for Moving the CPI Toward a COLI

Independent of whether or not the CPI should be based on a COLI framework, scanner data may be used to help overcome a few of the obstacles that now preclude calculation of anything like a COLI. By providing simultaneous information on prices and quantities, scanner data may reduce the lag in the production of superlative indexes and also enable Paasche indexes to be produced. Under current practices, price and quantity data are produced from different samples and at different frequencies.

Much caution is in order here, though. Feenstra and Shapiro (2001:21) found, in their construction of superlatives using scanner data for tuna fish that "the calculation of conventional price indexes. . . shows substantial pitfalls of mechanically applying price indexes to such data." The superlative index is intended to capture reductions in the cost of living as consumers substitute goods that have decreased in price for those that have increased. However, the superlative index calculated by the authors fails to produce this result (the superlative index grew faster). Feenstra and Shapiro (2001:22) concluded:

> The consumer behavior that generates these data cannot correspond to the static utility maximization that provides the foundation for superlative index numbers. Our tabulations suggest that the index numbers do not properly account for consumer behavior in response to sales. In particular, the chained Tornqvist gives too much weight to price increases that follow the end of sales.

The authors go on to explain that their findings reflect purchases made for storage rather than immediate consumption. In other words, purchases and consumption do not track in a parallel fashion, particularly for items that can be stored. As such, the consumer does not face as much an increase in price (after sales) as the raw data imply. In addition, advertising contributes to the breakdown of the law of demand that is assumed under the superlative index approach: "If advertisements cause consumers to purchase a [larger] quantity than would be consistent with static maximization of a time-invariant utility function, superlative index numbers will not accurately measure the cost of living" (Feenstra and Shapiro, 2001:22). On the basis of their findings, they conclude that unit values might provide a good approximation for construction of a COLI but should be adjusted to reflect consumption and should be adjusted to account for storage costs.[11]

Many of the general advantages of scanner data noted above may also help to address other CPI biases. For instance, scanner data allow for quicker and more accurate identification of both new goods and item attrition (and, as such, could have the capability to reduce new goods bias), as well as of outlet substitution patterns. Furthermore, scanner technology generates more detailed data for hedonic regression and other quality adjustment methods (although quality change bias is probably less of an issue for food items—the potential may be greater in areas such as consumer electronics) and also produces empirical evidence that may allow researchers to estimate the impact of quantity (and other types of) discount pricing on index growth.

[11]Triplett (1998) provides a simple demonstration of several other problems with using high-frequency data to produce a chained superlative index.

SUMMARY AND RECOMMENDATIONS

Without the benefit of extensive research on each of the areas raised in this chapter, the panel cannot make many definitive recommendations with respect to the data inputs to the CPI. We recognize that the BLS has undertaken research projects in these areas, and so BLS's inclusion in our discussion should not be taken as an indication that it has been negligent in its research efforts. It merely means that the panel recognizes the importance of these areas of research and hopes that they will continue systematically and thoroughly.

Research into the accuracy and sample size of the CEX should be a high priority among topic areas relating to the data collection process for the CPI. The panel concluded that it is likely that CEX estimates of consumer expenditure shares are biased, perhaps seriously. There is no obvious benefit to increasing the survey sample size if nonsampling error dominates sampling error—one would simply be achieving more precise estimates of the wrong thing.

Recommendation 9-1: Before additional resources are directed toward increasing its sample size (beyond the current plan), the accuracy of the CEX should be carefully evaluated. Assessing the net advantages of using the BEA's PCE to produce the upper-level weights for the national CPI should be part of this evaluation.

At the very least, research by BLS (and BEA) into the sources of divergence between PCE- and CEX-derived expenditure weights needs to be extended so that these differences can be more fully understood. Even if the current system is ultimately maintained, the effort will produce additional guidance about how the CEX might be improved.

Recommendation 9-2: If categories can be reasonably well matched between the CPI and PCE, so that comparable item strata indexes can be created, a program should be set up to produce an experimental CPI that uses PCE-generated weights at the upper (218 item) level but that is otherwise no different from the CPI.

If full item-by-item mapping turns out to be too problematic, it might still be possible to use PCE estimates for major item categories where the PCE and CEX have comparable coverage. For such categories, estimated totals from the CEX could be forced to equal the PCE estimates, which might allow the PCE to correct for undercoverage in the CEX in much the way that demographic projections are used to correct for undercoverage in household surveys such as the Current Population Survey. The distribution among lower-level aggregations would be determined by the CEX distribution. Investigating how well such experimental indexes perform seems especially sensible given the high cost of revamping the CEX survey or increasing its sample size. We would very much like to see a thorough defense of the choice of CEX-generated upper-level weights, relative to the alternatives, for the national-level CPI.

The CPI data collection process would also benefit from research in several other key areas discussed in this chapter:

- *Frequency of the CEX*: a combined theoretical-empirical study of the impact on the CPI of the frequency of updating weights.
- *Sample size for the CEX*: a combined theoretical-empirical study of the effect of CEX sample size on the variance of the CPI and on any subindexes that are desired.
- *CEX and POPS survey design*: a comprehensive reexamination of the design of each of these surveys.
- *Integration of CEX and POPS*: a study of the feasibility and requirements of a for-CPI-use-only single survey encompassing both the CEX and POPS.

Recognizing that scanner technology has the potential to improve the entire process of data collection for the CPI computation, the panel also identified the following key study areas:

- *Point-of-sale scanner data and item selection*: continuation of research on how these data can be used both to select items for pricing and to replace the C&S Survey and a quantification of the improvement in the CPI based on their use.[12]
- *Point-of-sale data and outlet selection*: initiation of research on how to use store sales information based on scanning to determine the stores to be sampled in the C&S.
- *Household scanner data*: initiation of research on the use of handheld scanners to record UPC items and quantities along with key-entering prices and/or store names and addresses.
- *Integration of UPCs into BLS ELI framework*: development of a concordance between UPCs and the ELIs.
- *Integration of non-UPCs into BLS ELI framework*: development of BLS assignments of UPCs for items which otherwise do not have UPCs for use in household handheld scanning.

[12] Assessment of current BLS scanner data experiments (ScanData cereal index for New York, next year's expansion throughout New England; use of scanner data/hedonics for audio components (using NPD data computers, and consumer electronics) to test impact on statistical properties of price data. We note that in 2002 BLS will consider ScanData recommendations about solving geography issues and about funding requests needed to expand the project and incorporate scanner-based subindexes into the CPI.

• *Experimental development of subgroup indexes*: performance of the household-based price data experiment, likely involving household scanner technology, to produce subgroup indexes that capture variation in both expenditure weights and prices paid.

TECHNICAL NOTE: ADDITIONAL DESCRIPTION OF CPI DATA INPUTS

The Modified Laspeyres CPI

The "Technical Note" to Chapter 2 sets forth the mathematical derivations underlying the development of the recommendations in this report. Equation (1) of that section sets forth the Laspeyres price index $P_L^{\,t}$, namely

$$P_L^t = \frac{\sum_{n=1}^{N} q_n^0 p_n^t}{\sum_{n=1}^{N} q_n^0 p_n^0},$$

relating base period quantities $q_n^{\,0}$, base period prices $p_n^{\,0}$, and current period prices $p_n^{\,t}$, for each of N goods (where the superscripts 0 and t refer to the base and current periods). The actual BLS-reported CPI differs from this in a few respects. First, the index is reported relative to a period in which it was set equal to 100. This period has, since 1987, been July-August 1983; prior to that, it was January 1967.

Second, and of more critical importance, the above equation is based on the assumption that both prices and quantities are collected simultaneously in the base period, but this is not the case for the BLS-reported CPI. For the CPI, the base period quantities are based on data from a household expenditure survey, while the base period prices are based on data from the monthly pricing surveys. Since the quantity data take longer to compile than do the price data, what is instead calculated is a "modified Laspeyres index," namely

$$P_L^{t,a} = \frac{\sum_{n=1}^{N} q_n^0 p_n^t}{\sum_{n=1}^{N} q_n^0 p_n^a} = \frac{\sum_{n=1}^{N} q_n^0 p_n^a [p_n^t / p_n^a]}{\sum_{n=1}^{N} q_n^0 p_n^a},$$

where, as before, n indexes the N goods and the superscript t denotes the current period, but where the superscript 0 refers to the *quantity-base period* (sometimes called the expenditure-base period) and the superscript a refers to the *price-reference period*. Since January 1998, the quantity-base period has been 1993-

1995; prior to that it was 1982-1984.[13] It is planned that, as of January 2002, the quantity-base period will be 1999-2000, and that it will be updated at 2-year intervals subsequently using information from the Consumer Expenditure Survey (CEX) ending 2 years prior to the update.

Finally, the q_n^0 are themselves not directly observed in the household expenditure survey. Rather, the survey provides quantity-base period expenditures e_n^0 for item n, and the quantities q_n^0 are calculated by dividing e_n^0 by p_n^0, where the quantity-base period prices are obtained from the monthly pricing survey.

The CPI can be expressed as a multiple of a Laspeyres index and the reciprocal of a modified Laspeyres index based on the quantity-base period and price-reference period, namely

$$P_L^{t,s} = \frac{\sum_{n=1}^{N} q_n^0 p_n^t}{\sum_{n=1}^{N} q_n^0 p_n^s} = \frac{\sum_{n=1}^{N} q_n^0 p_n^t}{\sum_{n=1}^{N} q_n^0 p_n^0} \frac{\sum_{n=1}^{N} q_n^0 p_n^0}{\sum_{n=1}^{N} q_n^0 p_n^s} = \frac{P_L^t}{P_L^{s,0}}.$$

As seen above, $P_L^{s,0}$ is a constant that relates the quantity-base period to the price-reference period. The critical element of the index is indeed P_L^t, which can be rewritten as

$$P_L^t = \frac{\sum_{n=1}^{N} q_n^0 p_n^t}{\sum_{n=1}^{N} q_n^0 p_n^0} = \frac{\sum_{n=1}^{N} q_n^0 p_n^{t-1}[p_n^t / p_n^{t-1}]}{\sum_{n=1}^{N} q_n^0 p_n^0}$$

$$= P_L^{t-1} \frac{\sum_{n=1}^{N} q_n^0 p_n^{t-1}[p_n^t / p_n^{t-1}]}{\sum_{n=1}^{N} q_n^0 p_n^{t-1}}.$$

This index can be characterized as a "chained" index, where the previous period's index P_L^{t-1} is multiplied by a dollar-weighted average of price relatives, with the dollar expenditure weights being those of the quantity-base period quantities priced at the previous period's prices and the price relative taken with respect to the price in the previous period. One should note that what is reported monthly by BLS is the period-to-period index, namely P_L^t / P_L^{t-1}.

[13]The quantity-base period differs across items so that, strictly speaking, 0 should be subscripted as 0_n, with the specific month for item n depending on the month 0_n in which the q_n^0 are determined from the Consumer Expenditure Survey.

Elements of the Index and Subindexes

The "goods" used in the CPI are organized into expenditure classes (ECs); as of 1999, there were 68 ECs. These are in turn are subdivided into *item strata*; as of 1998, there were 218 strata. Finally, the item strata are subdivided into entry-level items (ELIs); as of 2000, there were 282 ELIs.[14] The following is an example of this hierarchy of goods (Bureau of Labor Statistics, 1997a):

Expenditure Class 24: Maintenance and repair commodities
  Item stratum 2401: Materials, supplies, equipment for home repairs
    Entry-level items:
    24011: Paint, wallpaper, and supplies
    24012: Tools and equipment for painting
    24013: Lumber, paneling, wall and ceiling tile; awnings, glass
    24014: Blacktop and masonry materials
    24015: Plumbing supplies and equipment
    24016: Electrical supplies, heating and cooling equipment.
  Item stratum 2404: Other property maintenance commodities
    Entry-level items:
    24041: Miscellaneous supplies and equipment
    24042: Hard surface floor covering
    24043: Landscaping items

Subsequently, the ECs and their components were redesignated; in the 26 March 1999 list of ELIs, EC24 has been restructured as:

Expenditure Class HM: Tools, hardware, outdoor equipment and supplies
  Item stratum HM01: Tools, hardware, and supplies
    Entry-level items:
    HM011: Paint, wallpaper tools, and supplies
    HM012: Power tools
    HM013: Miscellaneous hardware, supplies, and equipment
    HM014: Nonpower hand yools
  Item stratum HM02: Outdoor equipment and supplies
    Entry-level items:
    HM021: Powered lawn and garden equipment and other outdoor items
    HM022: Lawn and garden supplies and insecticides

The data used in the CPI are collected in 87 primary sampling units (PSUs; see Williams, 1996). The data are aggregated into 54 *basic areas*—34 self-representing areas (e.g., Kansas City, MO-KS) and 20 region- and population-size

[14]Data from BLS dictionaries and Dennis Fixler (BLS, personal communication, 2000).

cross classifications (e.g., Midwest Size A).[15] The basic areas and item strata combine to form (218 × 54) = 11,772 basic CPI strata. Note that each of these basic CPI strata may be comprised of more than one ELI and more than one PSU.

Let h index the basic areas (h = 1, . . . ,54) and z index the item strata (z = 1, ... , 218). Until January 1999, BLS calculated R^t_{hz}—an estimate of the relative price change in basic area h, item stratum z, from period $t - 1$ to period t—using the formula when the samples of items within the item strata are selected with each unit having a probability proportional to quantity, or the formula

$$R^t_{hz} = \frac{\sum_{i\varepsilon z} w_{hi} p^t_{hi}}{\sum_{i\varepsilon z} w_{hi} p^{t-1}_{hi}}$$

$$R^t_{hz} = \frac{\sum_{i\varepsilon z} w_{hi} p^t_{hi} / p^a_{hi}}{\sum_{i\varepsilon z} w_{hi} p^{t-1}_{hi} / p^a_{hi}}$$

when the samples of items within the item strata are selected with each unit having a probability proportional to expenditure. In both forms the weights w_{hi} reflect the probability that item i in item stratum z is selected to be priced in basic area h—in the first of these the weights w_{hi} are essentially q^a_{hi} / π_c; in the second the weights w_{hi} are essentially $p^a_{hi}\, q^a_{hi} / \pi_{hi}$, where π_{hi} is the probability that item i in item stratum z is selected to be priced in basic area h. Since January 1999, they have replaced this computation for most indexes (the housing index being the most notable exception) with a weighted geometric mean, namely

$$R^t_{hz} = \prod_{i\varepsilon z} \left(p^t_{hi} / p^{t-1}_{hi} \right)^{w_{hi}} .$$

When one can obtain prices in basic area h for the universe of items in item stratum z, for both time periods $t - 1$ and t, then R^t_{hz} is given by the weighted average

$$R^t_{hz} = \sum_{i\varepsilon z} w_{hi} \left(p^t_{hi} / p^{t-1}_{hi} \right)$$

or, if the geometric mean computation is used, is given by

$$R^t_{hz} = \prod_{i\varepsilon z} \left(p^t_{hi} / p^{t-1}_{hi} \right)^{w_{hi}},$$

where w_{hi} is the ratio of the expenditure in basic area h on item i of item stratum z to the expenditure in basic area h on all items of item stratum z. Since a census of the prices for the universe of items in item stratum z is impractical, BLS

[15]This count is based on the 1998 CPI item strata spreadsheet provided by Dennis Fixler (personal communication, 2000).

estimates the R^t_{hz}. An oversimplified version of the BLS procedure is the following: Let a sample of N items be drawn from the universe of items in item stratum z (with replacement), with the probability of selection of item i equal to w_{hi}. Then

$$\overline{R}^t_{hz} = \frac{1}{N}\sum_{j=1}^{N}(p^t_{hj}/p^{t-1}_{hj})$$

is an unbiased estimate of the weighted average version of R^t_{hz}, and

$$\overline{R}^t_{hz} = \prod_{j=1}^{N}(p^t_{hj}/p^{t-1}_{hj})^{1/N}$$

is a consistent estimator of the geometric mean version of R^t_{hz}.

The BLS then updates its index P^t_{hz} for this basic stratum by the chaining formula described earlier, namely,

$$I^t_{hz} = I^{t-1}_{hz}R^t_{hz}.$$

These indexes are aggregated to form indexes for aggregate areas (e.g., U.S. cities), aggregate items (e.g., expenditure classes), or both. Let H denote the aggregate area and Z the aggregate item for which an index is to be formed. The index for this aggregate area and item is calculated as

$$I^t_{HZ} = \frac{\sum_{h\varepsilon H}\sum_{z\varepsilon Z} A_{hz}I^t_{hz}}{A_{HZ}},$$

where

$$A_{hz} = \sum_{j\varepsilon z} p^0_{hj}q^a_{hj}/I^0_{hz}$$

and

$$A_{HZ} = \sum_{h\varepsilon H}\sum_{z\varepsilon Z}\sum_{j\varepsilon z} p^0_{hj}q^a_{hj}/I^0_{hz},$$

where $j\varepsilon z$ denotes the items drawn from the universe of items in item stratum j.

Consumer Expenditure Survey

The CEX, sponsored by BLS and conducted by the Bureau of the Census, is a national probability sample of household units. It is comprised of two parts, a Quarterly Interview Panel Survey and a Diary Survey. Each "consumer unit" in the household selected for the Quarterly Interview Panel Survey is interviewed for 5 consecutive quarters about relatively large expenditure items (e.g., major appliances) and expenditures that occur at regular intervals (e.g., utility bills). A sample of 8,910 addresses are contacted for the Quarterly Interview Panel Survey in each of the calendar quarters, and the number of completed interviews per quarter is targeted at 6,160. Each consumer unit selected for the Diary Survey completes a diary on expenditure information on frequently purchased items and relatively small expenditure items for 2 consecutive weeks. A sample of 8,020 addresses are contacted each year to participate in the Diary Survey, so the effective annual sample size participating in this survey is 5,870 households, spaced across the 52 weeks in the year. The CEX has many uses in the governmental statistical framework. Its primary use in the CPI computation is to construct the quantities q_{hz}^0 which underlie the CPI computation. It has also been used "to select new market baskets of goods and services for the index, to determine the relative importance of components, and to derive new cost weights for the baskets" (U.S. Department of Labor, 2000).

Point of Purchase Survey and Commodities and Services Survey

The goal of the Point of Purchase Survey (POPS) is to determine the prices to be used in the CPI computation. The first stage of this survey is a national probability sample of household units, conducted by the Census Bureau, whose primary aim is to define the outlets to be sampled to obtain price data. The survey began in 1978 as a personal interview (and was referred to as CPOPS, for Continuing Point of Purchase Survey). In 1999 BLS revised this survey as a telephone interview, referred to as TPOPS (for Telephone Point of Purchase Survey). CPOPS was conducted annually over a period of 4 to 6 weeks, usually beginning in April; TPOPS interviews households every quarter. In CPOPS approximately one-fifth of the PSUs were sampled each year; the goal in TPOPS is to increase this sampling rate so that one-fourth of the PSUs are sampled each year. All consumer units in the selected household are asked to recall whether or not they purchased categories of goods and services within a specified recall period (varying from 1 week to 5 years, depending on the purchase cycle of the category) and, if so, the expenditure amounts and the names and locations of all places of purchase. Based on the responses to this survey of household units, a frame of outlets is defined for outlet selection. Since approximately one-fourth of the PSUs are currently sampled each year, after the survey of household units the frame of outlets determined by the survey is unchanged for 4 years.

The commodities and services are grouped into POPS categories, consisting of combinations of some of the ELIs; there were 174 POPS categories in 1997 (Bureau of Labor Statistics, 1997a). For example, POPS category 127, materials and supplies for major home repairs, consists of two of the ELIs of item stratum 2401, ELIs 24013 and 24014. POPS category 129, hardware items, hand tools, and other materials for minor home repairs, contains the other four ELIs of item stratum 2401—24011, 24012, 24015, and 24016; it also contains ELI 24041, miscellaneous supplies and equipment; ELI 32043, other hardware; and ELI 32044, nonpowered hand tools.

For the purpose of outlet selection, the BLS has aggregated the POPS categories into eight categories and the PSUs into ten groups (see Bureau of Labor Statistics, 1997a:). After a PSU group has been surveyed, the ELIs to be priced for the C&S Survey are selected with a systematic sampling procedure, with probability of selection proportional to the amount of expenditure in that PSU group and its item stratum. This systematic sampling procedure guarantees that over the 4-year period each of the ELIs will be selected for pricing. The outlets actually sampled from each frame are selected independently for each PSU group and POPS category, with probability of selection proportional to the amount of expenditure in that PSU group and POPS category. To give readers a sense of the number of outlets selected, the largest number is nine, in the POPS foods and beverages category, PSU group Philadelphia.

At a selected outlet a BLS field representative uses a multistage probability selection procedure for selecting the specific item to be priced among those that the outlet sells that fall within the designated-for-pricing ELI definition. The probability of selection is, if the information is available, proportional to the sales of the items in the ELI groups. Otherwise it is either based on the proportion of shelf space or, as a last resort, assigning equal probability to each item. Once the item is selected, its price is recorded. These are the prices that are weighted and used in the computation of the R_{hz}^{t} used in the CPI computation.

Appendix

Statistical Definition and Estimation of Price Indexes

This report addresses foundational economic concepts for cost-of-living or price indexes. In the panel's view, the concepts must reflect the reality of the marketplace; they must capture the change in real prices paid by real consumers. The concepts must be measurable in the context of a system of surveys and other data collection activities that the Bureau of Labor Statistics (BLS) can feasibly implement.

An important step in assessing the measurability and reality of a particular price index concept is to express the concept statistically in the form of a population parameter to be estimated. If one can write down the parameter, one can examine the feasibility of surveys and other data collection activities necessary to produce accurate statistical estimators of the parameters. One can also examine whether the parameter is defined in terms of the prices actually paid by consumers.

In what follows, we translate our concepts into explicit population parameters. We define the price indexes motivated by our concepts and demonstrate briefly the survey data required to estimate the indexes.

To begin, consider a simple world in which there is only one good and two time periods, base and comparison and a static universe of households (HH), denoted by the set H. For cases in which it would be better to work in terms of subgroups within HHs called consumer units (CU), let H denote the universe of CUs.

Next, let us introduce the bulk of the requisite notation. Let

i	signify the HH ($i = 1, \ldots N$),
j	the purchase occasion,
J_{i0}	the set of purchase occasions by the ith HH in base period 0,
J_{it}	the set of purchase occasions by the ith HH in comparison period t,
Q_{gij0}	the number of units of good g purchased by the ith HH, jth purchase occasion, during base period 0,
Q_{gijt}	the number of units of good g purchased by the ith HH, jth purchase occasion, during comparison period t,
N	the number of households in the universe
p_{gij0}	price per unit (of good g) paid by the ith HH, jth purchase occasion, during base period 0, and
p_{gijt}	price per unit (of good g) paid by the ith HH, jth purchase occasion, during comparison period t.

In these definitions, we use the convention

$$\sum_{j \in J_{i0}} = 0$$

for nonbuyers in the base period and

$$\sum_{j \in J_{it}} = 0$$

for nonbuyers in the comparison period. We assume there is at least one buyer, $Q_0 > 0$ and $Q_t > 0$, in each period.

Average unit volumes, $\overline{Q}_0$ and $\overline{Q}_t$, and average prices per unit, $\overline{p}_0$ and $\overline{p}_t$, are defined in the obvious way. The decomposition of the period-to-period trend in total dollar volume is now given by

$$\begin{aligned} T_y &= \frac{Y_t}{Y_0} \\ &= \frac{N_t \overline{Q}_t \overline{p}_t}{N_0 \overline{Q}_0 \overline{p}_0} \\ &= T_N T_q T_p, \end{aligned}$$

where T_N is the trend in the total HH count, T_q is again the trend in average units per HH, and T_p is again the trend in average price per unit. As above, T_p may be called the price index and T_q the unit volume index.

We can next further extend the work to a still more realistic world in which a static set of goods is available in the market at both time periods. Let subscript g signify a good, and to simplify the notation let G represent both the set and the number of goods. Total dollar volumes are now defined by

$$Y_0 = \sum_{g\in G}\sum_{i\in H_0}\sum_{j\in J_{gi0}} Q_{gij0}P_{gij0}$$

$$Y_t = \sum_{g\in G}\sum_{i\in H_t}\sum_{j\in J_{git}} Q_{gijt}P_{gijt}$$

for base and comparison periods, respectively. Average units volumes, $\overline{Q}_{g0}$ and $\overline{Q}_{gt}$, and average prices per unit, $\overline{p}_{g0}$ and $\overline{p}_{gt}$, are defined in the obvious way. Also, define the $G \times 1$ vectors of average unit volumes and average prices per unit

$$\overline{Q}_0 = (\overline{Q}_{10}, Q_{20}, \dots, \overline{Q}_{G0})',$$
$$\overline{Q}_t = (\overline{Q}_{1t}, \overline{Q}_{2t}, \dots, \overline{Q}_{Gt})',$$
$$\overline{p}_t = (\overline{p}_{10}, \overline{p}_{20}, \dots, \overline{p}_{G0})',$$

and

$$\overline{p}_t = (\overline{p}_{it}, \overline{p}_{2t}, \dots, \overline{p}_{Gt})'.$$

Then, the period-to-period trend in total dollar volume is given by

$$\begin{aligned} T_y &= \frac{Y_t}{Y_0} \\ &= \frac{N_t\overline{Q}_t'\overline{p}_t}{N_0\overline{Q}_0'\overline{p}_0} \\ &= \frac{N_t}{N_0}\left(\frac{\overline{Q}_t'\overline{p}_t}{\overline{Q}_0'\overline{p}_t}\right)\left(\frac{\overline{Q}_0'\overline{p}_t}{\overline{Q}_0'\overline{p}_0}\right) \\ &= T_N T_{Pq} T_{Lp}, \end{aligned}$$

where T_N is again the trend in the total HH count, T_{Pq} is the trend in average units per HH, and T_{Lp} is the trend in average price per unit. The trend in average units is weighted by comparison prices, and thus one might view T_{Pq} as a Paasche index of unit volume. Since the trend in average prices is weighted by base units volume, one might thus view T_{Lp} as a Laspeyres price index.

An alternative decomposition of the trend is

$$\begin{aligned} T_y &= \frac{N_t}{N_0}\left(\frac{\overline{Q}_t'\overline{p}_0}{\overline{Q}_0'\overline{p}_0}\right)\left(\frac{\overline{Q}_t'\overline{p}_t}{\overline{Q}_t'\overline{p}_0}\right) \\ &= T_N T_{Lq} T_{p_p}, \end{aligned}$$

where T_{Lq} is a Laspeyres index of units volume and T_{Pp} is a Paasche price index. A second alternative decomposition of the trend is

$$T_y = T_N (T_{Lq} T_{p_q})^{1/2} (T_{p_p} T_{L_p})^{1/2}$$
$$= T_N T_{Fq} T_{F_p}$$

where T_{Fq} and T_{Fp} are Fisher indexes of unit volume and prices, respectively.

Finally, we reach the real world in which both the sets of goods marketed, G_0 and G_t, and the sets of households, H_0 and H_t, vary by period. Partition the set of goods marketed at the base period by

$$G_0 = G_{0Q} \cup G$$

and partition the set of goods marketed at the comparison period by

$$G_t = G \cup G_{tE},$$

where G_{0Q} denotes exiting goods, G_{tE} denotes entering goods, and G denotes both continuing and linkable goods.

Continuing goods are marketed in both time periods, while exiting goods appear in the base period but not in the comparison, and entering goods appear in the comparison but not in the base. There is gray area we have called linkable goods. These are goods for which there is no exact match between the predecessor good and the successor good, but for which economic theory nevertheless accepts the link for purposes of index number construction. BLS has some linkage rules or criteria which it uses currently in producing the monthly CPI.

Period-to-period trend in total dollar volume is now

$$T_y = \frac{Y_t}{Y_0}$$
$$= \frac{\sum_{g \in G_t} N_t \overline{Q}_{gt} \overline{p}_{gt}}{\sum_{g \in G_0} N_0 \overline{Q}_{g0} \overline{p}_{g0}}$$
$$= \frac{\sum_{g \in G_t} N_t \overline{Q}_{gt} \overline{p}_{gt}}{\sum_{g \in G_0} N_0 \overline{Q}_{g0} \overline{p}_{g0}} \frac{R_0}{R_t},$$

where

$$R_0 = \frac{\sum_{g \in G} N_0 \overline{Q}_{g0} \overline{p}_{g0}}{Y_0}$$

is the continuing and linkable volume as a proportion of the total base volume, and

$$R_t = \frac{\sum_{g \in G} N_t \overline{Q}_{gt} \overline{p}_{gt}}{Y_t}$$

is the continuing and linkable volume as a proportion of the total comparison volume.

Let

$$T_R = \frac{R_t}{R_0}$$

be the trend in the proportion continuing or linkable. Then building on the above, the trend in total dollar volume can be decomposed as

$$\begin{aligned} T_y &= T_N T_{Pq} T_{Lp} T_R^{-1} \\ &= T_N T_{Lq} T_{Pp} T_R^{-1} \\ &= T_N T_{Fq} T_{Fp} T_R^{-1} . \end{aligned}$$

Alternative price indexes based on continuing and linkable goods are given byT_{Lp} is Laspeyres price index, T_{Pp} is the Paasche price index, and T_{Fp} is the Fisher price index. All price indexes discussed here extend naturally to a time series of comparison periods.

In the first two formulations, we faced a simple world with only one good. In this world, the price index

$$T_p = \frac{\overline{p}_t}{\overline{p}_0}$$

is both plutocratic and democratic.

In the third formulation, we faced a limited world in which a static set of goods is available in the marketplace. In this world, the plutocratic Laspeyres price index can be rewritten as

$$T_{Lp} = \sum_{g \in G} \frac{\overline{p}_{gt}}{\overline{p}_{g0}} S_{g0},$$

where the plutocratic weight applied to the simple trend in average price

$$S_{g0} = \frac{\sum_{i \in H_0} \sum_{j \in J_{gi0}} Q_{gij0} p_{gij0}}{\sum_{i \in H_0} \sum_{g' \in G} \sum_{j \in J_{g'i0}} Q_{g'ij0} P_{g'ij0}}$$

is simply market share expressed in dollars calculated across all HHs in the base period population with respect to the total market basket, G. Given the same assumptions, the democratic Laspeyres price index is defined by

$$I_{Lp} = \sum_{g \in G} \frac{\bar{p}_{gt}}{\bar{p}_{g0}} \bar{D}_{g+0},$$

where the democratic weight applied to the simple trend in average price

$$\bar{D}_{g+0} = \frac{1}{N_0} \sum_{i \in H_0} \frac{\sum_{j \in J_{gi0}} Q_{gij0} p_{gij0}}{\sum_{g' \in G} \sum_{j \in J_{gio}} Q_{g'ij0} p_{g'ij0}}$$

is the unweighted population mean across all HHs in the base period population of the HH specific market shares. Thus, plutocratic weights are ratios of means and democratic weights are means of ratios. Similar weighting yields I_{Pp}, a democratic Paasche price index, and I_{Fp}, a democratic Fisher price index. It is straightforward to establish the following relationship between plutocratic and democratic weights:

$$\begin{aligned} S_{g0} &= \bar{D}_{g+0} + \frac{\mathrm{Cov}\left\{D_{gi0}, Y_{+i0}\right\}}{\bar{Y}_{+0}} \\ &= \bar{D}_{g+0} + \left(1 + \frac{\mathrm{Cov}\left\{D_{gi0}, Y_{+i0}\right\}}{\bar{D}_{g+0} \bar{Y}_{++0}}\right), \end{aligned}$$

where

$$D_{gi0} = \frac{\sum_{j \in J_{gi0}} Q_{gij0} p_{gij0}}{\sum_{g' \in G} \sum_{j \in J_{g'i0}} Q_{g'ij0} p_{g'ij0}}$$

is market share within the ith household,

$$Y_{+i0} = \sum_{g' \in G} \sum_{j \in J_{g'i0}} Q_{g'ij0} p_{g'ij0}$$

is total consumption volume by the ith HH in the base period, and

$$\bar{Y}_{++0} = \frac{1}{N_0} \sum_{i \in H_0} Y_{+i0}$$

is the population mean per HH of total consumption volume. Thus, the plutocratic weight exceeds (is exceeded by) the democratic weight for any good that displays a positive (negative) correlation between total HH consumption and HH market share. The weights are equal in the event of zero correlation. For example, let g = automobiles. If there is positive correlation between total HH consumption and the share of HH consumption on automobiles, the plutocratic weight will exceed the democratic weight.

Across all goods, one can now conclude the following relationship between price indexes:

$$T_{Lp} = I_{Lp} + \sum_{g \in G} \frac{\bar{p}_{gt}}{\bar{p}_{g0}} \frac{\text{Cov}\{D_{gi0}, Y_{+i0}\}}{\bar{Y}_{++0}}.$$

The difference between the price indexes is determined by the pattern of covariances and price trends across goods. If goods for which the covariance is positive experience relatively large increases in average price, plutocratic price indexes may exceed their democratic counterparts. In general, however, the direction of the difference between the price indexes is far from certain for a given comparison period, t, let alone across periods. This matter is ripe for empirical investigation.

The democratic price indexes, and the relationship between plutocratic and democratic price indexes just discussed, extend naturally to the real-world situation described above where the domain of goods varies by period.

There are at least two approaches to estimating the price indexes: household (HH) survey data and store survey data. In this section, we explore the first approach; in the next section we look at the second.

Let s_0 and s_t denote probability samples drawn from the universe of HHs at times 0 and t, respectively. At each time period, assume that BLS collects unit volume and prices for all buying occasions for all goods from each HH, i, in the sample. Comprehensive data of this kind are not currently collected by any BLS survey. It might be feasible—using scanning technology or other approaches—to design surveys to collect such data.

Let W_{it} and W_{i0} denote survey weights such that

$$\hat{Q}_{gt} = \sum_{i \in s_t} W_{it} \sum_{j \in J_{git}} Q_{gijt}$$

and

$$\hat{Q}_{g0} = \sum_{i \in s_0} W_{i0} \sum_{j \in J_{gi0}} Q_{gij0}$$

are essentially design-unbiased estimators of the totals Q_{gt} and Q_{g0}, respectively. Similarly, define the estimated totals

$$\hat{N}_t = \sum_{i \in s_t} W_{it},$$

$$\hat{N}_0 = \sum_{i \in s_0} W_{i0},$$

$$\hat{Y}_{g+t} = \sum_{i \in s_t} W_{it} \sum_{j \in J_{git}} Q_{gijt} p_{gijt},$$

$$\hat{Y}_{g+0} = \sum_{i \in s_0} W_{i0} \sum_{j \in J_{gi0}} Q_{gij0} p_{gij0},$$

$$\hat{Y}_t = \sum_{g \in G_t} \hat{Y}_{g+t},$$

$$\hat{Y}_0 = \sum_{g \in G_0} \hat{Y}_{g+0},$$

$$\hat{Y}_t^G = \sum_{g \in G} \hat{Y}_{g+t},$$

and

$$\hat{Y}_0^G = \sum_{g \in G} \hat{Y}_{g+0}.$$

The latter two estimators reflect total dollar volume across all continuing and linkable goods. From these basic estimated totals, one can consistently estimate the ratios

$$\hat{R}_t = \frac{\hat{Y}_t^G}{\hat{Y}_t},$$

$$\hat{R}_0 = \frac{\hat{Y}_0^G}{\hat{Y}_0},$$

$$\hat{\bar{P}}_{gt} = \frac{\hat{Y}_{g+t}}{\hat{Q}_{g+t}},$$

$$\hat{\bar{P}}_{g0} = \frac{\hat{Y}_{g+0}}{\hat{Q}_{g+0}},$$

$$\hat{S}_{g0} = \frac{\hat{Y}_{g+0}}{\hat{Y}_0},$$

and

$$\hat{\bar{D}}_{g+0} = \frac{\sum_{i \in s_0} W_{i0} D_{gi0}}{\hat{N}_0}.$$

One can estimate the price indexes:

$$\hat{T}_{Lp} = \frac{\sum_{g \in G} \frac{\hat{Q}_{g0}}{\hat{N}_0} \hat{\bar{p}}_{gt}}{\sum_{g \in G} \frac{\hat{Q}_{g0}}{\hat{N}_0} \hat{\bar{p}}_{g0}} = \sum_{g \in G} \frac{\hat{\bar{p}}_{gt}}{\hat{\bar{p}}_{g0}} \hat{S}_{g0},$$

and

$$\hat{I}_{Lp} = \sum_{g \in G} \frac{\hat{\bar{p}}_{gt}}{\hat{\bar{p}}_{g0}} \hat{\bar{D}}_{g+0}.$$

Estimators for the other trends and indexes—$\hat{T}_y$, $\hat{T}_N$, $\hat{T}_{Pp}$, $\hat{T}_{Fp}$, $\hat{T}_{Lq}$, $\hat{T}_{Pq}$, and $\hat{T}_{Fq}$—are defined in the obvious way.

Next, consider the possibility of estimating the price indexes exclusively using store-level data. Let s_0 and s_t denote probability samples of stores, let the subscript k index the store, and let W_{k0} and W_{kt} denote survey weights corresponding to the unbiased estimator of a population total at times 0 and t, respectively. It is easy to imagine estimators

$$\hat{Y}_{gt} = \sum_{k \in s_t} W_{kt} Y_{gkt}$$

$$\hat{Y}_{g0} = \sum_{k \in s_0} W_{k0} Y_{gk0}$$

of total dollar volume and

$$\hat{Q}_{gt} = \sum_{k \in s_t} W_{kt} Y_{gkt}$$

$$\hat{Q}_{g0} = \sum_{k \in s_0} W_{k0} Y_{gk0}$$

of total unit volume. These estimators obviously require data on prices and unit volume by good at the store level. Current BLS surveys do not collect such data, but surveys based upon scanning technology could produce these data, at least for a subset of goods in a subdomain of stores. Given $\hat{Y}_{gt}$, $\hat{Y}_{g0}$, $\hat{Q}_{gt}$, $\hat{Q}_{g0}$, $\hat{N}_t$, and $\hat{N}_0$, it is possible to estimate the plutocratic price indexes.

The question is whether the price indexes estimated on the basis of store data really estimate T_{Lp}, T_{Pp}, and T_{Fp}. One would anticipate some biases due to such factors as

- goods purchased from stores for business use, not for home consumption;

- shrinkage due to breakage and pilferage (this component of bias would depend on the mode of data collection); and
- coverage errors in the store sampling frame (i.e., missing stores in which consumers shop and including stores in which they do not).

Regrettably, it is not possible to estimate the democratic index I_{Lp} exclusively from store-level data, at least not without additional assumptions. The democratic weights, $\overline{D}_{g+0}$, are population means per HH, and HH data are necessary to estimate the means unbiasedly; such data are not usually available from stores (some store chains have adopted ID card programs that allow tracking of purchases by consumer).

It may be possible to approximate the democratic index from store-level data with periodic adjustment of the weights. This possibility exploits the relationship between plutocratic and democratic weights set forth above. From store-level data, one can construct an estimator of the plutocratic weights

$$\hat{S}_{g0} = \frac{\hat{Y}_{g+0}}{\hat{Y}_0}.$$

Then we define the estimator of the democratic weights as

$$\hat{\overline{D}}_{g+0} = \hat{S}_{g0} - \frac{c(D_{gi0,Y_{+i0}})}{\hat{\overline{Y}}_{++0}},$$

where the adjustment factor is the second term on the right side, developed from an independent HH survey, such as the Consumer Expenditure Survey. In this factor, $c(D_{gi0},Y_{+i0})$ is an estimator of the covariance between HH share and total HH consumption, and $\hat{\overline{Y}}_{++0}$ is an estimator of mean total consumption per HH in the base period. It does not seem necessary to estimate the adjustment factor for each time period (month) the price index is produced. Perhaps it might be acceptable to maintain the adjustment factor only on an infrequent basis.

Without question, one can imagine other hybrid schemes for estimating plutocratic or democratic price indexes. BLS's current method is an outstanding example, with quantity weights coming from one survey and monthly prices from another.

References

Aizcorbe, A.M., and P.C. Jackman

1993 The commodity substitution effect in CPI data, 1982-1991. *Monthly Labor Review* 116(12) (December):25-33.

Alterman, W.F., W.E. Diewert, and R.C. Feenstra

1999 *International Trade Price Indexes and Seasonal Commodities*. Bureau of Labor Statistics, Washington, DC: U.S. Department of Labor.

Altman, Daniel, David M. Cutler, and Richard N. Zeckhauser

1998 Adverse selection and adverse retention. *American Economic Review* 88(May):122-126.

Amble, Nathan, and Ken Stewart

1994 Experimental price index for elderly consumers. *Monthly Labor Review* 117(May):11-16.

Anglin, Paul M., and Michael R. Baye

1987 Information, multiperiod search, and cost-of-living index theory. *Journal of Political Economy* 95(December):1179-1195.

Arguea, N.M., C. Hsiao, and G.A. Taylor

1994 Estimating consumer preferences using market data—an application using U.S. automobile demand. *Journal of Applied Econometrics* 9:1-18.

Armknecht, P.A., Walter F. Lane, and Kenneth J. Stewart

1997 New products and the U.S. consumer price index. In *The Economics of New Goods*, Timothy F. Bresnahan and Robert J. Gordon, eds. NBER Studies in Income and Wealth, Volume 58. Chicago: University of Chicago Press.

Armknecht, Paul, and Donald Weyback

1989 Adjustments for quality change in the U.S. consumer price index. *Journal of Official Statistics* 5(2):107-123.

Arrow, Kenneth J.

1958 The measurement of price changes. In *The Relationship of Prices to Economic Stability and Growth*. 85th Congress, 2nd Session, Joint Economic Committee Print. Washington, DC: U.S. Government Printing Office.

Arrow, K.J., H.B. Chenery, B.S. Minhas, and R.M. Solow

1961 Capital-labor substitution and economic efficiency. *Review of Economics and Statistics* 63:225-250.

Balk, B.M.

2000 On Curing the CPI's Substitution and New Goods Bias. Research Paper 0005, Department of Statistical Methods, Statistics Netherlands, P.O. Box 4000, 2270 JM Voorburg, The Netherlands.

Bentham, J.

1789 Introduction to the Principles of Morals and Legislation. Reprinted in Smith, J.M. and E. Sosa, eds. *Mills Utilitarianism: Text and Criticism*, 1969. Belmont, CA: Wodsworth.

Berkowitz, E.D.

1986 *Social Security After Fifty: Successes and Failures.* Westport, CT: Greenwood Press.

Berndt, E.R., Susan H. Busch, and Richard G. Frank

2001 Treatment price indexes for acute phase major depression. Pp. 463-505 in *Medical Care Output and Productivity*, David M. Cutler and Ernst R. Berndt, eds. Chicago: University of Chicago Press.

Berndt E.R., and I. Cockburn

1998 Is price inflation different for the elderly? An empirical analysis of prescription drugs. Pp. 33-75 in *Frontiers in Health Policy Research*, Alan Garber, ed. Cambridge, MA: MIT Press for the National Bureau of Economic Research.

Berndt, E.R., Iain Cockburn, and Zvi Griliches

1996 Pharmaceutical innovations and market dynamics: Tracking effects on price indexes for anti-depressant drugs. *Brookings Papers on Economic Activity*. Washington, DC: Brookings Institution.

Berndt, E.R., D.M. Cutler, R.G. Frank, Z. Griliches, and J.E. Triplett

2000 Medical care prices and output. Pp. 120-180 in *Handbook of Health Economics*, A.J. Culyer and J.P. Newhouse, eds. Amsterdam: Elsevier Science B.V.

Berndt, E.R., Zvi Griliches, and N.J. Rappaport

1995 Econometric estimates of price indexes for personal computers in the 1990s. *Journal of Econometrics* 68:243-268.

Berndt, E.R., Zvi Griliches, and Joshua G. Rosett

1993 Auditing the Producer Price Index: Micro Evidence from Prescription Pharmaceutical Preparations. *Journal of Business and Economic Statistics* 11(3): 51-264.

Berry, Steven, J. Levinson, and A. Pakes

1995 Automobile prices in market equilibrium. *Econometrica* 63(July):841-890.

Blank, Rebecca M., and Alan S. Blinder

1986 Macroeconomics, income distribution, and poverty. In *Fighting Poverty*, Sheldon Danzinger and Daniel Weinberg, eds. Cambridge, MA: Harvard University Press.

Board of Trustees of the Federal Old Age Survivors and Disability Trust Funds

2001 *Annual Report of the Board of Trustees of the Federal Old Age Survivors and Disability Trust Funds* (March 19). Washington, DC: U.S. Social Security Administration. Available: www.sg.gov/OACT/TR/TR01.

Boskin, M., and M. Hurd

1985 Indexing social security benefits: a separate price index for the elderly. *Public Finance Quarterly* 13(October):436-449.

Boskin, M.J., E. Dulberger, R.J. Gordon, Z. Griliches, and D.W. Jorgenson

1996 *Final Report of the Advisory Commission to Study the Consumer Price Index*. Washington, DC: U.S. Government Printing Office.

1998 Consumer Prices, the consumer price index, and the cost of living. *Journal of Economic Perspectives* 12(1):2-26.

Bosworth, Barry, and George Perry

1994 Productivity and real wages: Is there a puzzle? Pp. 317-335 in *Brookings Papers on Economic Activity*. Washington, DC: Brookings Institution.

Bradley, R.

1997 Potential Benefits from the use of scanner data in the construction of the CPI. *1996 ASA Proceedings*. Available: http://www.science.gmu.edu/gss/bradley.pdf.

Bradley, Ralph, Bill Cook, Sylvia G. Leaver, and Brent Moulton

1997 An Overview of Research on Potential Uses of Scanner Data in the U.S. CPI. Working Paper, Bureau of Labor Statistics, U.S. Department of Labor.

Branch, E. Raphael

1994 The consumer expenditure survey: A comparative analysis. *Monthly Labor Review* (December):47-55.

Bresnahan, Timothy F.

1997 Comment (pp. 237-247) in *The Economic of New Goods,* T.F. Bresnahan and R.J. Gordon, eds. NBER Studies in Income and Wealth, Volume 58. Chicago: University of Chicago Press.

Brickman, P., and D.T. Campbell

1971 Hedonic relativism and planning the good society. In *Adaptation-Level Theory*, M.H. Appley, ed. New York: Academic Press.

Bureau of Labor Statistics

1997a *BLS Handbook of Methods,* Bulletin 2490, April 1997, Chapter 17, The Consumer Price Index, Appendix 4. Washington, DC: U.S. Department of Labor.

1997b Letter from the Commissioner of the Bureau of Labor Statistics to the Chairman of the Joint Economic Committee, U.S. Congress, January 28.

1997c Measurement Issues in the Consumer Price Index. Response to a letter from the Joint Economic Committee of the U.S. Congress. (June). Available: http://www.bls.gov/cpigm697.htm.

1998a Memo, Future Schedule for Expenditure Weight Updates in the Consumer Price Index. (December). Available: http://www.bls.gov/cpiupdt.htm.

1998b Updated Response to the Recommendations of the Advisory Committee to Study the Consumer Price Index, Prepared in response to a request from the Chairman of the Subcommittee of Human Resources of the House Committee on Government Reform and Oversight, June, Section on Short-Run Recommendations. U.S. Department of Labor. Available: http://www.bls.gov/cpi/cpi0698a.htm.

2001 Measuring Price Change for Medical Care in the CPI. Available: http://stats.bls.gov./cpifact4.htm.

Byushgens (Buscheguennce), S.S.

1925 Sur une classe des hypersurfaces: A propos de 'l'index ideal' de M. Irving Fischer. *Mathematischeskii Sbornik* 32:625-631.

Cage, Robert

1996 New methodology for selecting CPI outlet samples. *Monthly Labor Review* (December): 49-61.

Campbell, A.

1981 *The Sense of Well-Being in America*. New York: McGraw-Hill.

Caves, D.W., L.R. Christensen, and W.E. Diewert

1982 The economic theory of index numbers and the measurement of input, output and productivity. *Econometrica* 50:1393-1414.

Cecchetti, Stephen G.

1997 Measuring short-run inflation for central bankers. *Federal Reserve Bank of St. Louis. Review* 79(3):143-155.

Clark, Todd E.
1999 A comparison of the CPI and the PCE price index. *Federal Reserve Bank of Kansas City Review* 84(3):15-29.

The Conference Board
1999 *Measuring Prices in a Dynamic Economy: Re-Examining the CPI*. New York: The Conference Board, Inc.

Council of Economic Advisers
1999 *Annual Report of the Council of Economic Advisers*. Washington, DC: U.S. Government Printing Office. Available: http://w3.access.gpo.gov/eop/.

Court, A.T.
1939 Hedonic price indices with automotive examples. Pp. 99-117 in *The Dynamics of Automobile Demand*. New York: General Motors Corporation.

Curry, Bruce, Peter Morgan, and Mick Silver
2001 Hedonic regressions: Mis-specification and neural networks. *Applied Economics* 33:659-671.

Cutler, David M., et al.
1996 Are Medical Prices Declining? Working Paper 5750. (September). National Bureau of Economic Research, Cambridge, MA.

Cutler, David M., Mark B. McClellan, Joseph P. Newhouse, and Dahlia Remler
1998 Are medical prices declining? Evidence from heart attacks. *Quarterly Journal of Economics* 113(December):991-1024.

Deaton, Angus
1998 Getting prices right: what should be done? *Journal of Economic Perspectives* 12(1):37-46.

Deaton, Angus, and John Muellbauer
1980 *Economics and Consumer Behavior*. New York: Cambridge University Press.

Diamond, Peter, and Jerry Hausman
1997 Contingent valuation: Is some number better than no number? *Journal of Economic Literature* 8(4):45-64.

Diener, E., and E. Suh
2000 National differences in subjective well-being. In *Well-Being: The Foundations of Hedonic Psychology*, D. Kahneman, E. Diener, and N.P. Schwarz, eds. New York: Russell-Sage.

Diewert, W.E.
1976 Exact and superlative index numbers. *Journal of Econometrics* 4:114-145.
1978 Superlative index numbers and consistency in aggregation. *Econometrica* 46:883-900.
1983 The theory of the cost of living index and the measurement of welfare change. Pp. 163-233 in *Price Level Measurement*, W.E. Diewert and C. Montmarquette, eds. Ottawa: Statistics Canada. [Reprinted as pp. 79-147 in *Price Level Measurement*, W.E. Diewert, ed. Amsterdam: North-Holland, 1990].
1987 Index numbers. Pp. 767-780 in *The New Palgrave: A Dictionary of Economics*, *Volume 2*, J. Eatwell, M. Milgate and P. Newman, eds. London: Macmillan. [Reprinted as pp. 71-104 in *Essays in Index Number Theory*, Volume 1, W.E. Diewert and A.O. Nakamura, eds. Amsterdam: North-Holland, 1993].
1993 The early history of price index research. Pp. 33-65 in *Essays in Index Number Theory*, *Volume 1*, W.E. Diewert and A.O. Nakamura, eds. Amsterdam: North-Holland.
1995 Axiomatic and Economic Approaches to Elementary Prize Indexes. Discussion Paper No. 95-01, Department of Economics, University of British Columbia.
1996 Comment on CPI biases. *Business Economics* 31(2):30-35.
1997 Commentary. *Federal Reserve Bank of St. Louis Review* 79(3):127-138.
2000a The Consumer Price Index and Index Number Purpose. Discussion Paper No. 00-02, Department of Economics, University of British Columbia,.

2000b Notes on Producing an Annual Superlative Index Using Monthly Data. Discussion Paper 00-08, Department of Economics, University of British Columbia,.

2001 Hedonic Regressions: A Consumer Theory Approach. Discussion Paper No. 01-12. Department of Economics, University of British Columbia.

Dulberger, Ellen

1993 Sources of price decline in computer processors: Selected electronic components. Pp. 103-124 in *Price Measurements and Their Uses*, Murray F. Foss, Marilyn E. Manser, and Allen Young, eds. Conference on Research in Income and Wealth. Studies in Income and Wealth. Chicago: University of Chicago Press for the National Bureau of Economic Research.

Easterlin, R.A.

1967 *Nations and Households in Economic Growth.* Stanford: University Press.

1974 Does economic growth improve the human lot? Some empirical evidence. In *Nations and Households in Economic Growth*, P.A. David and M.W. Reder, eds. New York: Academic Press.

Edgeworth, F.Y.

1925 *Papers Relating to Political Economy*, *Volume 1*. New York: Burt Franklin.

Emmons, W.R.

2000 The information content of treasury inflation-indexed securities. *Federal Reserve Bank of St. Louis Review* (November/December):33.

Epple, Dennis

1987 Hedonic prices and implicit markets: Estimating demand and supply functions for differentiated products. *Journal of Political Economy* 95(November):59-80.

Feenstra, Robert C.

1994 New product varieties and the measurement of international prices. *American Economic Review* 84(1):157-177.

1995 Exact hedonic price indices. *Review of Economics and Statistics* 77(November):634-653.

Feenstra, Robert C., and Matthew D. Shapiro

2001 High-Frequency Substitution and the Measurement of Price Indexes. National Bureau of Economic Research Working Paper 8176. Available: http//papers.nber.org/papers/W8176.

Fisher, Franklin M., and Karl Shell

1968 Taste and quality change in the pure theory of the true cost-of-living index. Pp. 97-139 in *Value, Capital, and Growth: Papers in Honour of Sir John Hicks*, J.N. Wolfe, ed. Edinburgh: University of Edinburgh Press.

1972 *The Economic Theory of Price Indices: Two Essays on the Effects of Taste, Quality, and Technological Change.* New York: Academic Press.

Fisher, I.

1922 *The Making of Index Numbers.* Boston: Houghton-Mifflin.

Fixler, Dennis

1993 The Consumer Price Index: Underlying concepts and caveats. *Monthly Labor Review* 116(December):3-12.

Fixler, Dennis, C. Fortuna, J. Greenlees, and W. Lane

1999 The Use of Hedonic Regressions to Handle Quality Change: The Experience in the U.S. CPI. Paper presented at the fifth meeting of the International Working Group on Price Indexes. Reykjavik, Iceland. Available: http://www.statice.is/ottawa/bls.pdf.

Ford, Ina Kay

1995 Health Insurance Allocations. Memorandum to Mary Lynn Schmidt, June 22. Bureau of Labor Statistics, U.S. Department of Labor, Washington, DC.

Ford, Ina Kay, and Daniel H. Ginsburg
2001 Medical care in the Consumer Price Index. Pp. 203-219 in *Medical Care Output and Productivity*, David M. Cutler and Ernst R. Berndt, eds. Chicago: University of Chicago Press for the National Bureau of Economic Research.
Frank, Richard G., Susan H. Busch, and Ernst R. Berndt
1998 Measuring prices and quantities of treatment for depression. *American Economic Review* 88(May):106-111.
Frisch, R.
1936 Annual survey of general economic theory: The problem of index numbers. *Econometrica* 4:1-38.
Garner, Thesia I., David S. Johnson, and Mary F. Kokoski
1996 An experimental consumer price index for the poor. *Monthly Labor Review* 119(September):32-42.
Gillingham, Robert, and John S. Greenlees
1987 The impact of direct taxes on the cost of living. *Journal of Political Economy* 95(August): 775-796.
1990 Indexing the federal tax system: A cost-of-living approach. *Journal of Business and Economic Statistics* 8(October):455-473.
Goodhart, Charles
2001 What weight should be given to asset prices in the measurement of inflation? *The Economics Journal* 111(472):335-356.
Gordon, Robert J., and Zvi Griliches
1997 Quality change and new products. *AEA Papers and Proceedings* (May):84-88.
Gorman, W.M.
1959 Separable utility and aggregation. *Econometrica* 27:469-481.
Green, Paul E., and A.M. Krieger
1985 Models and heuristics for product line selection. *Marketing Science* 4:1-19.
Greenlees, John S., and Dennis Fixler
2000 Direct Pricing of Health Insurance in the Consumer Price Index. Paper presented at the fall meeting of the Federal Economic Statistics Advisory Committee, Washington, DC. Bureau of Labor Statistics, U.S. Department of Labor (December 15).
Greenlees, John S., and Charles C. Mason
1996 Overview of the 1998 revision of the Consumer Price Index. *Monthly Labor Review* (December):3-9.
Griliches, Zvi
1961 Hedonic price indexes for automobiles: An econometric analysis of quality change. In *The Price Statistics of the Federal Government*. New York: National Bureau of Economic Research.
1971 *Price Indexes and Quality Change: Studies in New Methods of Measurement*. Cambridge: Harvard University Press.
1990 Hedonic price indexes and the measurement of capital and productivity: Some historical reflections. *Fifty Years of Economic Measurement: The Jubilee Conference of Research in Income and Wealth*, in E.R. Berndt and J.E. Triplett, eds. NBER Studies in Income and Wealth, Vol. 54. Chicago: University of Chicago Press.
1995 Testimony before the U.S. Congress, Senate Finance Committee, June 5.
Griliches, Zvi, and Iain Cockburn
1996 Generics and the producer price index for pharmaceuticals. Pp. 19-34 in *Competitive Strategies in the Pharmaceutical Industry*, Robert Helms, ed. Washington, DC: AEI Press.
Gruber, Jonathan
1994 The incidence of mandated maternity benefits. *American Economic Review* 84(June): 622-641.

Hagemann, Robert
1982 The variability of inflation rates across household types. *Journal of Money, Credit, and Banking* 14(November):494-510.
Hahn, Robert
1996 Regulatory reform: What do the government's numbers tell us? In *Risks, Costs, and Lives Saved*, R. Hahn, ed. Oxford, England: Oxford University Press, AEI Press.
Hardy, G.H., J.E. Littlewood, and G. Pólya
1934 *Inequalities.* Cambridge, England: Cambridge University Press.
Hausman, Jerry A.
1981 Exact consumers' surplus and deadweight loss. *American Economic Review* 71(4):662-676.
1997 Valuation of new goods under perfect and imperfect competition. Pp. 209-237 in *The Economics of New Goods*, T.F. Bresnahan and R.J. Gordon, eds. NBER Studies in Income and Wealth, Volume 58. Chicago: University of Chicago Press.
Hawkes, William J., and Frank W. Piotrowski
2000 Using scanner data to improve the quality of measurement and the measurement of quality in the Consumer Price Index. Prepared for the CRIW conference "Scanner Data and Price Indexes," September 15-16. AC Neilson. Available: http://www.nber.org/books/criw00/quality12-28-00.pdf.
Heffler, Stephen, Katharine Levit, Sheila Smith, Cynthia Smith, Cathy Cowan, Helen Lazenby, and Mark Frieland
2001 Health spending growth up in 1999: Faster growth expected in the future. *Health Affairs* 20(2):193-203.
Hicks, J.R.
1940 The valuation of the social income. *Economica* 7:105-140.
Inglehart, R.
1997 *Modernization and Postmodernization: Cultural, Economic, and Political Change in 43 Societies.* Princeton, NJ: Princeton University Press.
Ioannidis, C., and M. Silver
1999 Estimating hedonic indexes: An application to UK television sets. *Journal of Economics* 69(1):71-94.
Jacobs, Eva, and Stephanie Shipp
1990 How family spending has changed in the U.S. *Monthly Labor Review* (March):20-27.
Jorgenson, Dale, and Dan Slesnick
1983 Individual and social cost-of-living indexes. Pp. 241-323 in *Price Level Measurement*, W.E. Diewert and C. Montmarquette, eds. Ottawa: Statistics Canada.
Kahneman, D.
1999 Objective happiness. In *Well-Being: The Foundations of Hedonic Psychology*, D. Kahneman, E. Diener, and N. Schwarz, eds. New York: Russell-Sage.
Kahneman, D., B.L. Fredrickson, C.A. Schreiber, and D.A. Redelmeier
1993 When more pain is preferred to less: Adding a better end. *Psychological Science* 4:401-405.
Kalish, S.
1988 Pricing new products from birth to decline: An expository review. In *Issues in Pricing: Theory and Research*, T.F. Devinney, ed. Lexington, MA: Lexington Books.
Kang, Han-Bin, and Alan K. Reichert
1991 An empirical analysis of hedonic regression and grid-adjustment techniques in real estate appraisal. *American Real Estate and Urban Economics Association Journal* 19(Spring):70-91.

Knibbs, Sir G.H.

1924 The nature of an unequivocal price index and quantity index. *Journal of the American Statistical Association* 19:42-60, 196-205.

Kokoski, Mary

1987 Consumer Price Indices by Demographic Group. Working Paper No. 167, Bureau of Labor Statistics, U.S. Department of Labor. (April).

1993 Quality adjustment of price indexes. *Monthly Labor Review* 116(December):34-46.

2000 Alternative CPI aggregations: Two approaches. *Monthly Labor Review* (November): 31-39.

Kokoski, Mary, K. Waehrer, and P. Rozaklis

2000 Using Hedonic Methods for Quality Adjustment in the CPI: The Consumer Audio Products Component. BLS working paper. U.S. Department of Labor. Available: wwwlbls.gov/cpi/cpiaudio.htm.

Konüs, A.A.

1924 The problem of the true index of the cost of living. Translated in *Econometrica* 7(1939): 10-29.

Konüs, A.A., and S.S. Byushgens

1926 K probleme pokupatelnoi cili deneg, (On the problem of the purchasing power of money). *Voprosi Konyunkturi* II (1) (supplement to the Economic Bulletin of the Conjuncture Institute) 151-172.

Kuznets, S.

1962 Quantitative aspects of the economic growth of nations: VII: The share and structure of consumption. *Economic Development and Cultural Change* 10(2):1-92.

1996 Modern Economic Growth. New Haven, CT: Yale University Press.

Lane, Walter

1996 The changing item structure of the CPI. *Monthly Labor Review* (December):18-25.

2000 Addressing the New Goods Problem in the Consumer Price Index. Paper presented at the Issues in Measuring Price Change and Consumption Conference, Washington, DC, June 5-8, 2000. Bureau of Labor Statistics, U.S. Department of Labor.

Leaver, Sylvia, and William E. Larson

2001 Estimating variances for a scanner-based Consumer Price Index. Draft Bureau of Labor Statistics paper. Paper presented at 2001 meetings of the American Statistical Association. Bureau of Labor Statistics, U.S. Department of Labor.

Leaver, Sylvia, and Richard Valliant

1995 Statistical problems in estimating the U.S. Consumer Price Index. Chapter 28 in *Business Survey Methods*, B.G. Cox, D.A. Binder, B.N. Chinnappa, A. Christianson, M.J. Colledge, and P.S. Kott, eds. New York: John Wiley & Sons.

Lebow, David E., John M. Roberts, and David J. Stockton

1994 Monetary Policy and the Price Level. Unpublished paper. Federal Reserve Board, Washington, DC.

Levinsohn, Berry, and A. Pakes

1995 Automobile prices in market equilibrium. *Econometrica* (July):841-890.

Liegey, Paul R.

1993 Adjusting apparel indexes in the CPI for quality differences. Pp. 209-226 in *Price Measurements and Their Uses*, M. Foss, M. Manser, and A. Young, eds. National Bureau of Economic Research Studies in Income and Wealth. Chicago: University of Chicago Press.

1994 Apparel price indexes: Effects of hedonic adjustment. *Monthly Labor Review* 117(5):38-45.

2000 Hedonic quality adjustment methods for microwave ovens in the U.S. CPI. Unpublished paper. Available: http://www.bls.gov/cpimwo.htm.

Liegey, Paul, and Nicole Shepler
1999 Using hedonic methods to quality adjust VCR prices: Plucking a piece of the U.S. CPI's "low hanging fruit." *Monthly Labor Review* (September):22-37.
Lloyd, P.J.
1975 Substitution effects and biases in nontrue price indices. *American Economic Review* 65:301-313.
Loewenstein, G., and D. Schkade
1999 Wouldn't it be nice? Predicting future feelings. In *Well-Being: The Foundations of Hedonic Psychology*, D. Kahneman, E. Diener, and N. Schwarz, eds. New York: Russell-Sage.
Lykken, D., and A. Tellegen
1996 Happiness is a stochastic phenomenon. *Psychological Science* 7:186-189.
MacDonald, James M., and Paul E. Nelson, Jr.
1991 Do the poor still pay more? Food price variations in large metropolitan areas. *Journal of Urban Economics* 30(3):344-349.
Marshall, A.
1887 Remedies for fluctuations of general prices. *Contemporary Review* 51:355-375.
Mason, Charles C.
1988 An Analysis of the Rates of Inflation Affecting Older Americans Based on an Experimental Reweighted Consumer Price Index. Report presented to Congress, June. Bureau of Labor Statistics, U.S. Department of Labor.
Michael, Robert
1979 Variation across households in the rate of inflation. *Journal of Money, Credit, and Banking* (February):32-46.
Mishel, Lawrence, Jared Bernstein, and John Schmitt
2001 *The State of Working America 2000-2001*. Ithaca, NY: Cornell University Press.
Moulton, Brent R.
1996 Bias in the Consumer Price Index: What is the evidence? *Journal of Economic Perspectives* 10(4):159-177.
Moulton, Brent R., T.J. LaFleur, and K.E. Moses
1998 Research on improving quality adjustment in the CPI: The case of televisions. Proceedings of the fourth meeting of the International Working Group on Price Indices. U.S. Department of Labor. (April). U.S. Bureau of Labor Statistics.
Moulton, Brent R., and K.E. Moses
1997 Addressing the Quality Change Issue in the Consumer Price Index. *Brookings Papers on Economic Activity* (1):305-49.
Muellbauer, John
1974 Household production theory, quality, and the "hedonic technique." *American Economic Review* (64):977-994.
1975 Aggregation, income distribution and consumer demand. *Review of Economic Studies* 62:525-543.
1976 Community Preferences and the Representative Consumer. *Econometrica* 44:979-999.
National Bureau of Economic Research
1961 *The Price Statistics of the Federal Government.* The Price Statistics Review, Committee of the National Bureau of Economic Research, NBER General Series Number 73. New York: National Bureau of Economic Research.
National Research Council
1995 *Measuring Poverty: A New Approach.* Panel on Poverty and Family Assistance, Constance F. Citro and Robert Michael, eds. Committee on National Statistics, Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

1999 *Nature's Numbers: Expanding the U.S. National Economic Accounts to Include the Environment*. Panel on Integrated Environmental and Economic Accounting, W.D. Nordhaus and E.C. Kokkelenberg, eds. Committee on National Statistics, Commission on Behavioral and Social Sciences and Education. Washington, DC: National Academy Press.

Newhouse, J.P.

2001 Medical care price indices: Problems and opportunities. *Academia Economic Papers* 29(March):1-65.

Nordhaus, William D.

1997 Beyond the CPI: An augmented cost of living index. *Business Economics* (July):48-53.

1998 Quality Change in Price Indexes. *Journal of Economic Perspectives*, 12(1): 59-68.

Norwood, Janet

1995 Testimony. In *Consumer Price Index: Hearings Before the Committee on Finance, United States Senate*. Senate Hearing 104-69. Washington, DC: U.S. Government Printing Office.

Pakes, Ariel

1997 New Goods, Characteristic Based Models, and the Role of Hedonics. Unpublished paper, Department of Economics, Harvard University.

2001 New Goods, Hedonics and Price Indices, with an Application to PCs. Unpublished paper, Department of Economics, Harvard University.

Pakes, A., S. Berry, and J. Levinsohn

1993 Applications and limitations of some recent advances in empirical industrial organization: Price indexes and the analysis of environmental change. *American Economic Review, Papers and Proceedings* 83:240-246.

Pauly, Mark V.

1997 *Health Benefits at Work: An Economic and Political Analysis of Employment-Based Health Insurance*. Ann Arbor, MI: University of Michigan Press.

Pollak, Robert A.

1980 Group cost-of-living indexes. *American Economic Review* 70:273-278.

1981 The social cost-of-living index. *Journal of Public Economics* 15:311-336.

1983a The theory of the cost-of-living index. Pp. 87-161 in *Price Level Measurement*, W.E. Diewert and C. Montmarquette, eds. Ottawa, Canada: Statistics Canada.

1983b Treatment of 'quality' in the cost-of-living index. *Journal of Public Economics* (February):25-53.

1989 The treatment of the environment in the cost of living index. Pp. 181-185 in *The Theory of the Cost of Living Index*, R.A. Pollak, ed. Oxford, England: Oxford University Press.

1998 The Consumer Price Index: A research agenda and three proposals, *Journal of Economic Perspectives* 12(1)(Winter):69-78.

Prais, S.J.

1959 Whose cost of living? *The Review of Economic Studies* 26:126-134.

Ragan, J.F., Jr., and B. Bratsberg

2000 UNCOLA: Why have cost-of-living clauses disappeared from union contracts? *Southern Economic Journal* 67(2):304-324.

Reinsdorf, M.B.

1993 The effect of outlet price differentials in the U.S. Consumer Price Index. Pp. 227-254 in *Price Measurements and Their Use*, M.F. Foss, M.E. Manser, and A.H. Young, eds. Chicago: University of Chicago Press.

1996 Constructing Basic Component Indexes for the U.S. CPI from Scanner Data: A Test Using Data on Coffee. Working Paper 277, Bureau of Labor Statistics. (April). U.S. Department of Labor.

Reinsdorf, M.B., P. Liegey, and K. Stewart

1996 New Ways of Handling Quality Change in the U.S. Consumer Price Index. Working Paper No. 276, Bureau of Labor Statistics, U.S. Department of Labor.

Reinsdorf, M.B., and B.R. Moulton

1995 The Construction of Basic Components of Cost-of-Living Indexes. Working Paper 261, Bureau of Labor Statistics, U.S. Department of Labor.

1997 The construction of basic components of cost-of-living indexes. In *The Economics of New Goods*, T.F. Bresnahan and R.J. Gordon, eds. Chicago: University of Chicago Press.

Rhode, Paul W., and K.S. Strumph

2000 A Historical Test of the Tiebout Hypothesis: Local Heterogeneity from 1850 to 1990. Working Paper No. 7946, National Bureau of Economic Research. (October).

Richardson, David H.

2000 Scanner Indexes for the CPI. Paper presented at the Conference on Research on Income and Wealth Scanner Data and Price Indexes, September 15-16. Bureau of Labor Statistics, U.S. Department of Labor. Available: http://www.nber.org/books/criw00/richardson 12-21-00.pdf.

Rosen, Sherwin

1974 Hedonic prices and hedonic markets: Product differentiation in pure competition. *Journal of Political Economy* (April):34-55.

Schwarz, N., and F. Strack

1999 Reports of subjective well-being: Judgmental processes and their methodological implications. In *Well-Being: The Foundations of Hedonic Psychology*, D. Kahneman, E. Diener, and N. Schwarz, eds. New York: Russell-Sage.

Scitovsky, Anne A.

1967 Changes in the costs of treatment of selected illnesses, 1951-65. *American Economic Review* 57(5):1182-1195.

Shapiro, Irving, M.D. Shapiro, and D.W. Wilcox

1999 Quality improvement in health care: A framework for price and output measurement. *American Economic Review* 89(2):333-337.

2001 Measuring the value of cataract surgery. Chapter 10 in *Medical Care Output and Productivity*, D.M. Cutler and E.R. Berndt, eds. Chicago: University of Chicago Press.

Shapiro, M.D., and D.W. Wilcox

1995 Elementary aggregates, micro-indices and scanner data: Some issues in the compilation of consumer price indices. *Review of Income and Wealth* 41:427-438.

1996 Mismeasurement in the Consumer Price Index: An evaluation. In *NBER Macroeconomics Annual*, B.S. Bernanke and J.J. Rotenberg, eds. Cambridge, MA: MIT Press.

1997 Alternative strategies for aggregating prices in the CPI. *Federal Reserve Bank of St. Louis Review* 79(3):113-125.

Shephard, R.W.

1953 *Cost and Production Functions*. Princeton, NJ: Princeton University Press.

Shepler, Nicole

2001 Developing a Hedonic Regression Model for Camcorders in the U.S. CPI. Unpublished paper, Bureau of Labor Statistics, U.S. Department of Labor.

Silver, Mick

1998 Bias in the Compilation of Consumer Price Indices when Different Models of an Item Coexist. Paper presented at the Ottawa conference, Washington, DC, April. Bureau of Labor Statistics, U.S. Department of Labor.

1999 An evaluation of the use of hedonic regressions for basic components of consumer price indices. *Review of Income and Wealth* 45(1):41-56.

Snyder, Eleanor M.

1961 Cost of living indexes for special classes of consumers. In *The Price Statistics of the Federal Government*. Staff Report No. 7, General Series, No. 73. New York: National Bureau of Economic Review.

Stewart, Kenneth, and Joseph Pavalone

1996 Experimental consumer price index, Americans 62 years of age and older. *CPI Detailed Report, Attachment F*. Bureau of Labor Statistics (April). Washingington, DC: U.S. Department of Labor.

Stewart, K.J., and S.B. Reed

1999 Consumer price index research using current methods. *Monthly Labor Review* (June): 29-38.

Stigler, George

1961 The price statistics of the federal government. In *Report to the Office of Statistical Standards, Bureau of the Budget*. New York: National Bureau of Economic Research.

Summers, Lawrence H.

1989 Some simple economics of mandated benefits. *American Economic Review* 79(2):177-183.

2000 Personal consumption expenditures by type of expenditure, *Survey of Current Business* 80(12).

Thompson, William

2000 Developing a Hedonic Regression Model for VCRs in the U.S. CPI. Unpublished paper. Available: http://www.bls. gov/cpivcrp.htm.

Triplett, J.E.

1986 The economic interpretation of hedonic methods. *Survey of Current Business* (January): 36-40.

1987 Hedonic functions and hedonic indexes. Pp. 630-634 in *The New Palgrave: A Dictionary of Economics*, M.M. Eatwell and P. Newman, eds. London: MacMillan.

1990 Hedonic methods in statistical agency environments: An intellectual biopsy. In *Fifty Years of Economic Measurement: The Jubilee Conference on Research in Income and Wealth*, E.R. Berndt and J.E. Triplett, eds. Chicago: University of Chicago Press.

1997 Measuring consumption: The post-1973 slowdown and the research issues. *Federal Reserve Bank of St. Louis Review* 79(3):9-42.

1998 Elementary indexes for a Consumer Price Index. In *Proceedings of the Fourth Meeting of the International Working Group on Price Indices,* Walter Lane, ed. Washington, DC: U.S. Department of Labor.

2001a Should the cost-of-living index provide the conceptual framework for a consumer price index? *Economic Journal* 111(June):311-334.

2001b *Handbook on Quality Adjustment of Price Indexes for Information and Communication Technology Products*. Paris: Organizaton for Economic Cooperation and Development.

U.S. Department of Commerce

2000 *Survey of Current Business*. Bureau of Economic Analysis, Washington, DC: U.S. Department of Commerce. Available: www.bea.doc.gov.

U.S. Department of Health and Human Services

1980 *International Classification of Diseases, Second Edition*. Washington, DC: U.S. Department of Health and Human Services.

U.S. Department of Labor

2000 *GILS Record: Consumer Expenditure Study*. Washington, DC: U.S. Department of Labor. Available: www.dol.gov/bls/public/gils/records/00166.htm

U.S. General Accounting Office

1999 *Consumer Price Index: Impact of Commodity Analysts' Decisionmaking Needs to Be Assessed.* Report GGD-99-84. Washington, DC: U.S. Government Printing Office.

U.S. Social Security Administration

2000 *Annual Statistical Supplement, 2000*. Washington, DC: U.S. Social Security Administration. Available: http://www.ssa.gov/statistics/Supplement/2000/.

Vartia, Y.Y.

1983 Efficient methods of measuring welfare change and compensated income in terms of ordinary demand functions. *Econometrica* 51(1):79-98.

Walsh, C.M.

1901 *The Measurement of General Exchange Value*. New York: Macmillan.

1921 *The Problem of Estimation*. London: P.S. King & Son.

Waugh, Frederick V.

1928 Quality factors influencing vegetable prices. *Journal of Farm Economics* 10(2):185-196.

White, Alan G.

1999 Measurement biases in consumer price indexes. *International Statistical Review* 67(3): 301-325.

2001 Outlet types and the Canadian consumer price index. *Canadian Journal of Economics* 33(2):488-505.

Williams, Janet L.

1996 The redesign of the CPI geographic sample, *Monthly Labor Review* (December):10-17.

Glossary

Words in ***boldface italic*** appear separately in the glossary.

Aggregation The process whereby prices of individual goods are combined to produce price (or cost-of-living) indexes. For the Consumer Price Index (CPI), some 80,000 price quotes are collected each month from 87 urban areas (primary sampling units) throughout the United States. See ***lower-level aggregation, upper-level aggregation***.

Base period The period in which the market basket of goods and services and the expenditure shares for those goods and services are set. See ***Laspeyres index, reference period.***

Boskin commission The Advisory Committee to Study the Consumer Price Index, formed in 1995 by the Senate Finance Committee and chaired by Michael Boskin of Stanford University.

Chained index A multiperiod index that links indexes for shorter periods that may have different item expenditure weights.

Class-mean method One of the methods used during CPI ***item replacement***. Like ***deletion***, the method involves imputing the price of a changed item, but it does so from a set of similar goods further limited to those that are (1) classified as comparable replacements or (2) that could be explicitly quality adjusted by a hedonic or direct cost method.

Commodities and Services (C&S) Survey A longitudinal survey in which field agents of the Bureau of Labor Statistics (BLS) collect price quotes from sample outlets for items currently included in the CPI.

Commodity analyst A member of BLS's CPI team who determines the comparability of disappearing and replacement items that arise in the CPI sample.

Comparison period The second (often the current) period in the price or cost-of-living comparison.

Conditional cost-of-living index (CCOLI) The minimal expenditure ratio needed to maintain a given standard of living in the face of changes in the prices of private goods and services, on the assumption that "outside conditions"—the status of the social and physical environment and the provision of goods by the government—remain unchanged.

Consumer demand function A construct that indicates the relationship between prices and income on one hand and, on the other, the quantity of a good demanded.

Consumer Expenditure Study (CEX) An annual survey, designed to produce household-level data on both income and consumption, from which the BLS establishes item category (upper-level) expenditure weights.

Consumer substitution The purchase by consumers of more of one good (that has become relatively cheaper) and less of another (that has become relatively more expensive) in response to a relative price change. See ***lower-level substitution, upper-level substitution.***

Cost-of-goods index (COGI) An index in which a fixed basket of goods is priced each period, calculated as the cost of the basket in the ***comparison period*** divided by the cost of the basket in the ***reference period.***

Cost-of-living index (COLI) An index based on the minimal expenditure ratio needed to maintain a given standard of living (however determined) in the ***comparison period*** and the ***reference period.***

CPI The Consumer Price Index, a ***modified Laspeyres index*** which tracks the price of 218 categories of goods across 87 primary sampling units in urban areas of the United States.

CPI-E An experimental version of the CPI in which upper-level weights are based on a subset of the CEX sample aged 62 and over.

CPI Housing Survey A survey conducted by the BLS to track changes in the contract rents (a major component of the shelter item category) paid by a sample of renters.

CPI-U The CPI series that is weighted by the full CEX urban sample. It is the most frequently used version of the CPI.

CPI-U-RS The CPI research series that recalculates the CPI-U from 1978 forward with methods currently used to produce the CPI.

CPI-W A version of the CPI in which upper-level weights are based on a subset of the CEX sample that are wage earners.

Deletion method One of the methods used in CPI ***item replacement*** when the replacement and replaced items are judged noncomparable and when neither overlapped prices nor producer cost information is available. During the item transition period, change in the index component is proxied by the observed price change of other goods in the same CPI item stratum.

Direct comparison In ***item replacement***, a method that is applied when the replaced and replacement items are determined to be comparable by the commodity analyst.

Direct hedonic method A ***hedonic quality adjustment*** technique that produces a price index for a good directly from hedonic regression coefficients.

Democratic index An index in which each household's index is given equal weight.

Entry-level item (ELI) A finely specified item within an ***item strata*** (e.g., oranges is an ELI in the citrus fruit strata).

Expenditure weight The portion of consumer expenditures assigned to a specific good or service category.

Explicit cost-based adjustment An ***item replacement*** method used when information about production cost differences between the replaced item and the new item is available under which the per-unit change in production cost, as reported by manufacturers, is subtracted from the change in the observed price paid by consumers.

Fisher ideal index A ***superlative index*** derived as the geometric mean (the square root of the product) of the ***Paasche index*** and the ***Laspeyres index***.

Fixed-weight index An index constructed as the ratio of the cost of purchasing a basket of goods and services at the prices of one period (or in one location) to the cost of purchasing that same basket at the prices of a subsequent period (or different location). The set of weights assigned to the prices of items in the basket remains the same in each period (or location).

Geometric mean formula A method for combining price quotes, currently used in the CPI for about 61 percent (by weight) of basic item indexes, that uses a set of fixed-expenditure (as opposed to quantity) proportions as weights for averaging individual prices within a basic index. Fixing relative expenditure proportions implies that customers substitute among specific items (in response to changing prices) in such a way that the share of expenditure on each item category remains constant.

Hedonic quality adjustment A method that uses regression technique to separate out the effect of changed item quality on its observed price by establishing a relationship between a good's characteristics and its price. See ***direct hedonic method, indirect hedonic method.***

Homothetic preferences A theoretical assumption positing that a consumer ranks different bundles of goods the same no matter what her level of living so that the rate at which a person is prepared to trade one good for another is independent of whether the person is rich or poor; it also implies that, as people become better off, they simply scale up their purchases without changing the pattern of consumption.

Index (chain) drift The divergence between a ***chained index*** and a ***fixed-weight index*** caused by the linking of indexes with different strata weights.

Index formula A formula that dictates the exact way in which prices and expenditure shares are combined to calculate a price index.

Indirect hedonic method In ***hedonic quality adjustment,*** a technique that involves adjusting, post hoc, the observed price difference between an outgoing item and a replacement item based on the portion of the price change attributable to a changed characteristic. The magnitude of the adjustment is determined by the hedonic function and the differences between the characteristics bundles supplied by the old and new items.

Item replacement The selection by a BLS field agent of a new (previously not priced) item because the item previously priced is no longer available at a particular CPI sample outlet. BLS has implemented a set of methods designed for judging pure price change when prices of two non-identical items must be compared. See ***class-mean method, deletion method, direct comparison, explicit cost-based adjustment, overlap pricing.***

Item strata The 218 categories of goods and services that represent the lowest level of disaggregation at which CPI subindexes are calculated.

Item weighting (upper level) The expenditure share assigned to each category of goods and services in an index.

Laspeyres index A price index measured as a weighted average of the ratios of individual prices in two periods. Item weights are fixed to reflect expenditure shares in the ***base period*** (which for a Laspeyres is the same as the ***reference period***). The denominator is the priced-out fixed bundle established in the base period; the numerator is the same bundle, priced out for a later period.

Link period The period in which new item weights are introduced in an index.

Lower-level aggregation The manner in which individual price quotes are combined by area (or area groups) into separate indexes for some 218 categories of goods, called item strata, the basic building blocks for the CPI.

Lower-level substitution A change in the proportions of expenditure on goods within a CPI entry-level item category (e.g., Gala apples for Golden Delicious). See ***consumer substitution.***

Medical Care Price Index (MCPI) An index designated in scope to track changes in the price of (or cost of living associated with) goods and services related to households' medical care.

Modified Laspeyres index As used for the CPI, an index in which the fixed bundle (strata weights) is established in a base period that does not typically coincide with the reference period. Information from three (not two) periods enters index calculation.

Outlet rotation The process whereby BLS reselects the outlets from which it prices index items.

Outlet substitution The phenomenon in which consumers switch outlet patronage for price or cost-of-living reasons; typically used to describe situations in which consumers switch from purchasing a good at a high-price seller to a lower-priced competitor.

Overlap pricing In ***item replacement*** a method that can be used when both old and new models are available in at least one period.

Paasche index A price index that weights expenditure shares by the ***comparison period*** (typically the most current) consumption patterns. It measures the percentage difference in expenditures between what it would cost the household to buy the comparison period quantities at the old prices and what it costs at the new ones.

Personal consumption expenditure (PCE) data A measure of consumer expenditure shares, produced by the Bureau of Economic Analysis (BEA), based on aggregate sales data.

Personal consumption expenditure (PCE) deflator A chain price index (that uses a ***Fisher ideal index***) for personal consumption expenditures based (mainly) on CPI-collected prices and PCE weights.

Plutocratic index A price index in which all dollars of expenditure (or income) are treated equally so that, implicitly, each household is weighted proportional to its expenditures.

Point of Purchase Survey (POPS) A survey of households, conducted by the Census Bureau, used in the CPI to determine the distribution of households' expenditures across specific outlets.

Private good A good or service sold in markets, whose benefits can be limited to those who pay for it.

PSU One of the 87 primary sampling units that are the geographic delineations from which prices are surveyed.

Public good A good that (1) is available to all if it is available to one and (2) whose consumption by one person does not reduce the amount available to others.

Reference period The first period in a price or cost-of-living comparison.

Standard of living Narrowly defined (entirely in terms of consumption of goods and services), a measure of the extent to which preferences are satisfied which, given a set of prices that remain constant over a number of periods, can be measured by the amount of money spent. More generally, a measure that also captures broader aspects of well-being, such as health or happiness.

Stigler committee The Price Statistics Review Committee of the National Bureau of Economic Research, chaired by George Stigler.

Superlative index A price index that weights expenditure patterns from both the current (***comparison period***) and past (***base or reference period***) in the price ratio calculation.

Tornqvist index A ***superlative index*** in which the rate of growth of individual prices is weighted by their share in the budget averaged over the ***base period*** and the current period.

Upper-level aggregation The process whereby area/***item strata*** indexes are combined to form the national CPI.

Upper-level substitution A change in the proportions of expenditure on goods between item categories (e.g., chicken for beef). See ***consumer substitution.***

Utility The satisfaction derived from the consumption of goods or services.

Virtual price The price of a new good that would have been just high enough, during a period prior to its appearance in the market, to drive quantity demand to zero.

Biographical Sketches of Panel Members and Staff

Charles L. Schultze (*chair*) is a Senior Fellow Emeritus at the Brookings Institution. In addition to holding academic positions at Indiana University and the University of Maryland, he served as chairman of the Council of Economic Advisers and director of the U.S. Bureau of the Budget. He was president of the American Economic Association in 1984, and he served as chair, and of the Panel on Census Requirements in the Year 2000 and Beyond of the National Research Council.

Ernst R. Berndt is the Louis B. Seley Professor of Applied Economics in the Alfred P. Sloan School of Management at the Massachusetts Institute of Technology, where he has taught since 1980. His research focuses on sources of productivity growth, on adjusting computer price measures to account for quality change, and on price and productivity measurement in health care industries. He has written on the measurement of pharmaceutical price changes, on measuring prices of medical treatments that account for changes in outcomes, and has provided testimony before Congress about international pharmaceutical price comparisons. He is an elected Fellow of the Econometric Society, served as member and cochair of the American Economics Association Advisory Committee to the U.S. Census Bureau, is currently the first chairperson of the Federal Economic Statistics Advisory Committee, and also serves as director of the National Bureau of Economic Research Program on Technological Change and Productivity Measurement.

Angus Deaton is Dwight D. Eisenhower Professor of Economics and International Affairs at the Woodrow Wilson School and Economics Department of Princeton University. His areas of expertise include applied econometrics and the analysis of household behavior. He is a Fellow of the British Academy, the American Academy of Arts and Sciences and of the Econometric Society and was the first recipient of the Society's Frisch Medal for applied econometrics. He served as a member of the panel on poverty measurement of the National Research Council.

Erwin Diewert is a professor in the Department of Economics at the University of British Columbia. He is a member of the Statistics Canada Services Advisory Committee, a research associate of the National Bureau of Economic Research, a fellow of the Royal Society of Canada, a fellow of the Econometric Society, and chairman of the Statistics Canada Advisory Committee on Prices. His fields of interest include microeconomic theory, mathematical economics, international trade, public finance, measurement economics and accounting, nonparametric estimation and smoothing methods. He has published extensively on topics relating to theory and development of price indexes.

Claudia Goldin is Henry Lee Professor of Economics at Harvard University and a research associate at the National Bureau of Economic Research, where she is director of the Development of the American Economy Program. She is currently an associate editor of the *Quarterly Journal of Economics*. She is a Fellow of the American Academy of Sciences and of the Econometric Society. Her research has focused on historical analyses of the female labor force, slavery, immigration, education, and economic inequality. She is best known for *Understanding the Gender Gap: An Economic History of American Women* and is the editor of numerous volumes including *The Defining Moment: The Great Depression and the American Economy in the Twentieth Century*.

Zvi Griliches was Paul M. Warburg Professor of Economics at Harvard University. His fields of interest included the economics of research and technological change, econometric methods, rates of return to investment in education, the measurement of quality bias in price indexes, and patent statistics. He was a member of the National Academy of Sciences and served as a member of the Committee on National Statistics. He also served on the Advisory Commission to Study the Consumer Price Index, established by the Senate Finance Committee.

Christopher Jencks is Malcolm Wiener Professor of Social Policy at the Kennedy School at Harvard University. His recent research has dealt with changes in the material standard of living over the past generation, homelessness, welfare reform, the black-white test score gap, and poverty measurement. His book credits include *The Academic Revolution* (with David Riesman), *Inequality, Who*

Gets Ahead? The Urban Underclass, Rethinking Social Policy, *The Homeless*, and most recently, *The Black-White Test Score Gap* (with Meredith Phillips). He is a member the National Academy of Sciences and the American Academy of Arts and Sciences, and serves on the editorial board of *The American Prospect*.

Christopher Mackie is a study director with the Committee on National Statistics (CNSTAT). In addition to working with this panel, he is working on projects on nonmarket economic accounting and data access and confidentiality. Prior to joining CNSTAT, he was a senior economist with SAG Corporation, where he conducted a variety of econometric studies in the areas of labor and personnel economics, primarily for federal agencies. He completed his Ph.D. in economics at the University of North Carolina and, while a graduate student, held teaching positions at the University of North Carolina, North Carolina State University, and Tulane University. He is author of *Canonizing Economic Theory*.

Albert Madansky is the H.G.B. Alexander Professor Emeritus of Business Administration in the Graduate School of Business of the University of Chicago and editor-in-chief of the *Journal of Business*. He has been a mathematician at the RAND corporation, vice president of the Interpublic Group of Companies, president of a computer software and data processing firm, chairman of the computer sciences department at City College of New York, and a Fellow of the Center for Advanced Study in the Behavioral Sciences. He is a Fellow of the American Statistical Association, the Institute of Mathematical Statistics, and the Econometric Society and an elected member of the International Statistical Institute. His primary research is in the area multivariate statistical analysis. He is the author of *Foundations of Econometrics* and of *Prescriptions for Working Statisticians*.

Van Doorn Ooms is Senior Vice President and Director of Research at the Committee for Economic Development, where he has supervised and conducted research in a large number of public policy areas including education, social security and pensions, labor markets, welfare and poverty, international trade and finance, and campaign finance. His primary specialties are in macroeconomics and public finance, with emphasis on the political economy of the federal budget. He was the chief economist of (successively) the Senate Budget Committee, The Office of Management and Budget, and the House Budget Committee. Previously, he was a professor of economics at Swarthmore College.

Robert A. Pollak is the Hernreich Distinguished Professor of Economics at the College of Arts and Sciences and the John M. Olin School of Business at Washington University in St. Louis. Prior to his current appointment, he was a member of the University of Pennsylvania's faculty for 26 years. His areas of expertise include environmental economics and policy, consumer behavior, labor economics, demography, and human capital theory. He is a member of Phi Beta Kappa, a fellow of the Econometric Society, and a fellow of the American Association for the Advancement of Science.

Richard Schmalensee is the Gordon Y. Billard Professor of Economics and Management at the Massachusetts Institute of Technology. Concurrently, he is dean of the Sloan School of Management, director of the Center for Energy and Environmental Policy Research, and a research associate of the National Bureau of Economic Research. He was a member of the Executive Committee of the American Economic Association and, from 1989-1991, served on the President's Council of Economic Advisers. He is also a member of the Committee on National Statistics or the National Research Council.

Norbert Schwarz is a professor of psychology at the University of Michigan and a senior research scientist at Michigan's Institute for Social Research. Prior to joining the University of Michigan, he taught psychology at the University of Heidelberg, Germany. His research focuses on social cognition, the interplay of feeling and thinking, conversational influences on judgment and reasoning, and applications of cognitive psychology to methodological issues of survey research. His research has advanced creative methods in the areas of survey elicitation and on perceptions of well-being and quality of life. He is a former fellow of the Center for Advanced Study in the Behavioral Sciences and a recipient of the Heinz Maier-Leibnitz Preis of the German Department of Science and Education.

Kirk Wolter is professor in the Department of Statistics at the University of Chicago and senior vice-president of statistics and methodology at the National Opinion Research Center. He specializes in sample survey design, federal statistical policy, international statistics, marketing research, census requirements and methods, and sampling/survey training research. Previously he was chief of the Statistical Research Division at the U.S. Bureau of the Census and Vice-President Worldwide for the A.C. Nielson Company. He is a fellow of the American Statistical Association and the International Statistical Institute. He is also President Elect of the International Association of Survey Statisticians.

Index

A

ACNielsen, 266, 270-271
Aggregation. *See also* Subgroup indexes
- across consumers, 26-27
- across households, 4, 45, 79, 223-226, 230-231, 240-246
- for COGIs, 43, 77-79
- for COLIs, 4, 15, 51-52, 53, 61-62, 85-86, 242
- conceptual issues, 39, 51-52
- in CPI, 15, 23-24, 243-244, 305, 308, 311
- data collection issues, 27
- defined, 305
- demand, 243, 244
- democratic indexes, 4, 5, 26-27, 45, 77-79, 86, 235 n.9
- expenditure-weighted averaging, 26, 52, 60
- geometric means formula, 5, 23, 24, 39, 50, 60, 77, 127, 210, 215 n.24, 279-280, 308
- heterogeneity issue, 4, 26, 125 n.18, 223-226, 229, 240-241
- of individual price indexes, 44-45, 215, 235 n.9
- lower-level, 23, 61-62, 309
- mathematical approach, 77-79
- in MCPI, 182, 188
- national indexes, 4, 25, 26, 46, 87, 214, 219-221
- plutocratic indexes, 4, 5, 26-27, 45, 77-79, 86, 235 n.9
- population average of budget shares, 79
- quality change adjustments and, 125 n.18
- recommendations, 240-241
- representative consumer concept, 241-246
- strata indexes, 235 n.9
- substitution effects and, 24-25, 26, 52
- in superlative indexes, 52, 53, 215
- upper-level, 24, 199, 311
- user costs and, 34, 72
- in wage indexes, 200-201

American Housing Survey, 11, 36, 147
Apparel, 29, 67, 72, 117, 118, 120 nn.11&12, 124, 130, 142, 147, 227
Appliances and electronics, 6-7, 72, 114, 131, 138, 141, 147, 148, 163 n.12, 233 n.7. *See also individual items*
Area strata, 311
Automobiles, 28-29, 33, 35, 72, 107, 109, 110, 111, 119, 121, 124, 142, 147-148, 227, 233 n.7

B

Base period, 41, 42, 69, 305, 309, 311
Basket price index. *See* Cost-of-goods index
Bias
in CEX, 253, 254-255
duration neglect, 54
in hedonic quality adjustment methods, 128, 129, 133, 139, 140, 249
in inflation indexes, 217-218
in item replacement methods, 112-114, 121-122, 136
MCPI, 148, 185-186, 188
new goods, 31, 32, 107-108, 112-121, 146-148, 156-158, 160-161, 162, 273
outlet, 8, 33, 128, 168 n.16, 170, 172, 173-174
point-of-introduction, 158-159, 160-161
reporting, 254-255
sample selection, 128, 156-157
seasonal, 234
substitution, 8, 33, 59-62, 93, 113, 128, 147, 168 n.16, 170, 172, 173-174, 248 n.18
within-sample, 140
Boskin commission, 38, 39
on conceptual basis for CPI, 3, 14, 41, 73
criticisms of CPI, 1, 14, 27, 62
defined, 305
on domain of CPI, 66, 96
new goods bias estimates, 32, 157, 162, 163
outlet substitution bias estimates, 173-174
quality adjustment bias estimates, 27, 108, 109 n.2, 112-114, 115, 116-117, 119, 121, 131, 139, 146-148
on subgroup indexes, 228, 229, 250
on theoretical validity of CPI, 109 n.2
Bureau of Economic Analysis (BEA), 4, 102, 105, 189, 210, 214, 253, 274, 310
Bureau of Labor Statistics, 306
charge to committee, 17
commodity analysts, 107, 306
quality adjustment approaches, 39, 73, 107-108, 113 n.6, 114-122, 129-140
ScanData initiative, 266-267, 269

C

Capital gains and losses, 72
Cellular phones, 8, 31, 32, 107, 124-125, 140, 148, 156, 161, 165
Census Bureau, 255, 310
Centers for Medical and Medicaid Services, 10, 189, 190
Chain drift, 167, 257, 308
Chained index, 162, 167, 210, 253 n.2, 277, 305, 308, 310
Choke price. *See* Virtual price
Class-mean method, 117, 118, 119, 121, 134 n.30, 136-137, 139, 305-306
Climate change, 20
Commodities and Services (C&S) Survey, 11, 36, 37, 263, 281-282, 306
Commodity analysts (BLS), 107, 116, 118, 125, 134, 138-139, 306
Comparison situation
conceptual issues, 41, 42, 48-49, 51
period-based, 41, 306, 310, 311
place-based, 66-67, 69-70
Compensating variation, 48, 81, 153
Compensation issues
child support, 209
COGI and, 70, 71
COLI and, 48, 51, 58, 61, 70-71, 81
conceptual bases for indexes, 48, 51, 58, 61, 70-71, 81
domain considerations, 21, 70, 103-105
employer-paid benefits, 21, 98, 103-105
homeowners, 70-71, 72
MCPI components and, 186
social security recipients, 40, 43, 48
and standard of living, 48
superlative indexes, 60, 61, 191
Computers, 29, 109, 110, 117, 120 n.12, 124, 127, 129, 130, 134, 135, 140, 142, 147, 156, 166
Conceptual bases for indexes
aggregation of prices, 39, 51-52
availability of data and, 15, 42-43, 50
basket price (cost-of-goods) approach, 2-3, 14, 16, 18-19, 38-39, 41, 42-46, 47, 48-49, 50, 57, 74-79, 94
COGI vs. COLI, 57-72
comparison situation, 41, 42, 48-49, 51
compensation issues, 48, 51, 58, 61, 70-71, 81
consumer demand functions, 49-50
cost-of-living approach, 2-3, 14, 15, 16, 18-19, 38-39, 41, 43, 46-53, 57, 58, 65-68, 69, 79-93, 283-292
CPI, 1, 2-3, 8, 14, 18-19, 38, 39, 41, 42, 73
criticisms of, 40, 52-53

design considerations, 11, 40, 42, 43, 97, 191-192
domain-related, 3, 65-68, 94, 96-98
homotheticity in preferences, 49, 50-51
income changes, 2, 90
mathematical approach, 74-93
MCPI, 181-185
public perceptions and understanding of, 39, 58-59
purpose of index and, 40, 42, 43
quality adjustment, 2, 3, 14, 19, 28, 30, 39, 40, 57, 59, 62-64, 97-98, 106
reference period, 41, 42, 47-49, 51
satisfaction, 47, 53, 55-57
stochastic approach, 41-42, 68
stocks and flows, 33-35, 71-72
subpopulations and, 229-232
substitution effect, 2, 3, 39, 47-48, 50-51, 52-53, 58, 59-62
superlative indexes, 50-51, 53, 83-85
taste changes, 2, 53, 59, 64, 67, 69-70
test approach, 41-42, 50, 68
theoretical, 39, 40, 43-58, 73, 74-93
utility and choice, 47, 53-55
virtual demand for new commodities, 158 n.6

Conditional cost-of-living index (CCOLI), 39, 65-68, 69, 70, 86-90, 306
Congressional Budget Office, 259
Conrad, Kent, 228
Constant-elasticity-of-substitution (CES) price index, 6, 60-61, 91-93, 194
Constant utility index, 14
Consumer demand functions, 49-50, 306
Consumer Expenditure Survey (CEX), 42, 79, 213
accuracy, 12, 36, 253-256, 261
cost, 234, 253, 261
defined, 306, 307
description of, 189, 238, 244, 280-281
Diary Survey, 257, 280-281
expenditure weights, 11-12, 36, 253-256, 261, 274-275
frequency, 164-165, 256-257, 275
NIPA expenditure estimates compared, 214
POPS integrated with, 37, 264-266, 275
Quarterly Interview Panel Survey, 257, 280
questionnaire and structural issues, 36, 261-262
research needs, 11-12, 261, 274-275
sample size, 12, 36, 131 n.25, 257-261
scanner technology and, 271-272
subgroup data, 4, 203 n.11, 227, 234, 236, 247, 248, 255, 259
substitution with new goods, 164-165
uses, 252, 261, 277, 281

Consumer Price Index (CPI)
aggregation in, 15, 23-24, 243-244, 305, 308, 311
biases in, 112-122, 129, 250
COLI-based, 2-3, 4, 14, 15, 18-19, 24, 94, 104, 110 n.3, 112, 144-145, 272-273
comparison period, 42
conceptual basis, 1-2, 8, 15-16, 18-19, 38, 39, 41, 42, 73, 94, 221
criticisms of, 1, 13, 14, 27, 39
defined, 306
domain, 3-4, 16, 17, 19-21, 73, 94, 105
experimental version (CPI-E), 11, 197, 203 n.11, 228, 246-250, 307
fixed-weight method, 14-15
growth of, 207-208
hedonic quality adjustment, 6-7, 132-137, 138, 139-140, 141
Housing Survey, 307
improvement initiative, 131 n.25, 145, 164, 166 n.14
introduction of new goods, 157, 161-167
item replacement, 134, 305-306, 307-308
item strata, 23, 130, 131, 156, 163, 307, 309
modified Laspeyres-based, 24, 49, 52, 61, 62, 97 n.2, 104, 111, 114, 221, 276-277, 306, 309
NIPA compared, 217, 219-221
product sampling and item identification, 107-108, 250
proxy prices, 72, 120, 307
PSU, 278-279, 310
quality adjustments, 14, 16, 27, 39, 106, 107, 110 n.3, 111, 112-122, 116-121, 130-137, 209-210
quantity-base period, 276-277
recalculation (CPI-U-RS), 202 n.8, 220, 307
reference period, 42, 276
research series, 210
revision frequency, 115, 166 n.13, 221
structure and elements, 23-24, 156, 163, 278-280, 307, 309
substitution behavior and, 15-16, 23-25, 26, 169-170, 113
superlative index, 24-25, 97, 191, 194-195, 207-208, 210

theoretical basis, 73
urban sample (CPI-U), 197-198, 203 n.11, 207, 211, 228, 247, 249, 250, 257, 307
uses, 1, 13-14, 35, 192, 218
wage-earner sample (CPI-W), 197-198, 203 n.11, 207, 228, 246, 250, 257, 307
weights/weighting, 11-12, 14, 15, 26, 104, 163, 179, 180, 198, 227, 307
Consumer substitution
COGI and, 5, 21-23, 25, 59, 70
COLI and, 2, 6, 25, 39, 47-48, 50-51, 52-53, 57, 58, 59-62, 66, 90
conceptual issues, 2, 3, 39, 47-48, 50-51, 52-53, 58, 59-62, 242
defined, 306
and domain, 66, 95
full-service retail to discount outlets, 169-170, 173
heterogeneity of, 26, 224
standard of living and, 73, 215 n.24
superlative index and, 5-6, 23, 25, 60, 61, 84, 85, 93, 96, 113 n.6, 215
Contingent valuation, 98
Contract rents, 307
Cost functions, 80, 82, 83-84, 85, 86, 91-92
conditional, 88-90
Gorman polar form, 245
rationed, 90
Cost-of-goods index (COGI). *See also* Laspeyres index; Paasche indexes
aggregation, 43, 77-79
augmented, 103
averages of, 76-77, 82
base period, 41, 42, 74, 77
bias, 59
COLI compared, 2-3, 16, 18-19, 24, 38-39, 57-72
comparison period, 16, 21-23, 41, 42, 74
compensation issues, 70, 71
conceptual bases, 2-3, 14, 16, 18-19, 38-39, 41, 42-46, 47, 48-49, 50, 57, 74-79
defined, 1-2, 44, 306
domain, 19, 65-66, 94, 95
item replacement, 111-112
mathematical approach, 74-79
objectives, 16
outlet substitution bias, 8, 172, 176
public understanding of, 58
quality adjustment, 62-64, 111-112, 123
rationale for, 43
reference period, 16, 21-23, 41, 74
and substitution, 5, 21-23, 59, 70, 224
and taste change, 24, 69
theory, 43-46
time reversal test, 76-77
Cost-of-living index (COLI)
aggregation, 4, 15, 51-52, 53, 61-62, 85-86, 242
base period, 80-81, 82, 91, 92
calculation, 43-44, 49-50, 172
CES, 91-93
choice-based, 53-55, 57
COGI compared, 2-3, 16, 18-19, 24, 38-39, 57-72
comparison period, 41, 48-49, 51, 53
compensation calculation, 48, 51, 58, 61, 70-71, 81, 153
conceptual bases, 2-3, 14, 15, 16, 18-19, 38-39, 41, 43, 46-53, 57, 58, 65-68, 69, 79-93, 283-292
conditional, 3, 4, 39, 65-68, 69, 70, 73, 86-90, 94-96, 115, 306
cost/expenditure function, 80, 82, 83-84, 85, 86, 89-90
counterfactual cost, 80, 91
criticisms and controversies, 15-16, 52-53
data collection, 289-292
defined, 2, 15, 95, 306
demand functions, 49-50, 63, 83, 84
democratic, 52, 242, 288, 292
domain, 3-4, 17, 19-21, 39, 65-68, 69, 70, 73, 86-90, 94-100, 104, 306
equivalent variation, 81, 153
expenditure, 195
Fisher ideal index as, 50, 52, 60, 286, 287, 288
geometric means formula, 5, 23, 24, 39, 50, 60, 77, 127, 210, 215 n.24, 279-280, 308
homotheticity of preferences, 49, 50-51, 60, 82, 84, 308
income effects, 2, 90
index numbers, 80, 91
indifference curve, 80, 86, 90, 91
item replacement methods, 110
Laspeyres–Paasche index relationship in, 23, 47, 48-49, 51, 80-82, 87, 90-91, 285, 287-289
mathematical approach, 79-83, 90-91
medical care, 19, 86-90
multiperiod, 104 n.10
new goods bias, 8, 31, 160, 161

objectives, 15, 16
outlet substitution bias, 8, 168, 172, 174, 176
plutocratic, 52, 242, 244, 287-288, 292
public understanding of, 58-59
quality adjustment, 2, 3, 18-19, 39, 57, 62-64, 86-90, 109-114, 115, 123, 168
reference period, 41, 47-49, 51-52, 53, 82
representative consumer, 241-246
satisfaction-based, 55-57
Slutsky matrix, 90, 91
social, 51-52, 85-86
standard-of-living measurement, 3, 15 n.2, 19, 41, 46-47, 48, 49, 55-56, 59, 63, 69, 80, 82, 95
statistical definition and estimation of, 283-291
substitution effects, 3, 6, 15-16, 39, 47-48, 50-51, 52-53, 57, 58, 59-62, 66, 90, 224
superlative indexes as, 50-51, 57
and taste change, 2, 16, 24, 69-70
Taylor series approximation, 90-91
theoretical bases, 2, 4, 15, 19, 40, 46-53, 58, 79, 83, 110, 113
"true," 1, 13, 14, 48, 81, 82
unconditional, 4, 17, 66, 68, 69, 70, 73, 96-98, 115
utility function, 53-55, 79-80, 82, 83-84, 85, 86-87
Counterfactual cost, 80, 91
Crime rates, 4, 20, 73, 95, 97, 98
Cross-outlet price linking, 169, 171
Current Population Survey (CPS), 201-202, 234, 274

D

Data collection for CPI, 306
C&S Survey, 11, 36, 37, 252, 263, 281-282
CEX, 11-12, 24, 36, 213, 227, 234, 236, 241, 244, 252-263, 271-272, 274, 280-281
and conceptual bases for indexes, 15, 42-43, 50, 197
costs, 232, 234, 271
for democratic indexes, 215-216, 223
for hedonic quality change adjustments, 124, 126 n.19, 128-129, 131-132, 135, 143, 144
heterogeneity of consumers and, 27, 223, 231, 241
integrated CEX/POPS, 37, 264-266, 271, 275
lagged, 6, 15, 25, 31-32, 42-43, 50, 57, 61, 76, 129, 144, 165, 183, 219-220, 256-257
from manufacturers/vendors, 132 n.26
MCPI, 9, 183
modified Laspeyres, 276-277
new goods, 30-32, 163-165
nonsampling errors, 260 n.5
POPS/TPOPS, 11, 36-37, 163-164, 167-168, 227 n.3, 232, 233 n.5, 234, 262, 267, 281-282, 310
recommendations, 5, 11-12, 241, 261, 274-276
sample rotation, 8, 31, 33, 119-120, 140, 156, 157, 163, 164, 165, 167, 172
sampling errors, 260 n.5
with scanners, 5, 43, 45, 129, 142 n.35, 197, 233, 234, 235, 236, 241, 266-273, 275, 289, 291
scope of, 23, 27, 226-227
Standard Industrial Classification system, 264
for subgroup indexes, 4, 5, 27, 197, 226-228, 229, 231-235, 241
Deletion method, 117, 118, 119-120, 125, 128, 135, 136-137, 139, 305, 307
Demand
aggregation, 243, 244
functions, 49-50, 63, 83, 84
for new commodities, 158 n.6, 159-160
Democratic indexes
aggregation, 4, 5, 26-27, 45, 77-79, 86, 235 n.9
data collection for, 215-216, 223, 252, 270-271
defined, 52, 307
new goods in, 160 n.8, 162
for public transfer program indexing, 198-199
for subgroups, 5, 26-27, 229, 237-240
Demographic issues, and index construction, 24-25, 26, 32, 53, 110, 172-173
Design of indexes. *See also* Conceptual bases for indexes; Domain of indexes; Public transfer payment indexing; Social security benefits
and conceptual framework, 11, 40, 42, 43, 97, 191-192
inflation-indexed treasury securities, 192, 210-212

inflation indicators, 216-221
limits on, 35
for monetary policy setting, 35, 40, 192, 216-221
as output deflator, 35, 214-216, 218
for poverty line, 40
for private contracts, 35, 192, 208-210, 218
for social security compensation, 11, 35, 40, 43, 48, 192
tax-related, 35, 40, 43, 192, 212-214
wage-related, 11, 35, 199-206, 207-208
Diewert, W. Erwin, 23
Direct characteristics method, 126, 128-129, 133, 142, 143-144, 152-154
Direct comparison method, 109, 116-118, 139, 307
Direct consumption costs, 213
Direct time dummy method, 126-128, 143, 145-146, 151-152, 153
Directed reinitiation, 155 n.1
Disaggregation of households, 45
Domain of indexes. *See also* Design of indexes
COGI, 19, 65-66, 94, 95
compensation issues, 21, 70, 103-105
conceptual issues, 3, 65-68, 94, 96-98
conditional COLI, 4, 39, 65-68, 69, 70, 73, 86-90, 95-96, 306
CPI, 3-4, 16, 17, 19-21, 73, 94, 105
cross-place comparisons, 66-67
current practice, 65
and economic indicators, 97
employer-paid benefits, 21, 98, 103-105
environmental variables, 4, 16, 19-20, 66-67, 73, 94-102, 105
MCPI, 9, 10, 95, 97-98, 178-179, 185-186, 189-190, 220
measurement problems, 20, 21, 97, 98-102
NIPA, 220
public goods, 4, 19-21, 94, 96, 98-102, 104
and public transfer programs, 97
quality changes and, 67, 95, 97-98, 101
recommendations, 4, 105
separability issue, 68
substitution and, 66, 95
superlative indexes, 95-96
supplemental indexes and satellite accounts, 4, 101-102, 103, 105
taxes and, 20-21, 97, 102-103
technological changes and, 67-68, 88-89, 97
unconditional COLI, 4, 17, 66, 68, 69, 70, 96-98
Durable goods, 71-72, 230 n.4
DVDs, 131, 138

E

e-Commerce, 170-173
Econometric modeling
of demand functions, 49-50, 63
of new goods, 160
of quality adjustments, 63, 64, 139, 145-146
Economic theory of consumer behavior, 2, 19, 40, 46-53, 58, 79, 83, 110, 242, 243
Elderly, price indexes for, 11, 196-197, 199, 207, 228-229, 249-251
Employer-paid benefits, 9, 10, 21, 98, 103-105, 179, 201, 220
Employment Cost Index (ECI), 201
Entry-level items (ELI), 136, 140, 155 n.1, 164, 165, 167, 234, 236, 262, 307
Environmental (outside) variables
benefit—cost assessment, 100 n.7
conceptual issues, 66-67, 69, 94-98
CPI domain, 4, 19, 73, 105
measurement, 16, 19, 20-21, 98-102
superlative indexes and, 53
taxes and, 20-21, 103
theoretical framework, 86-87, 88
Equivalent variation, 81, 153
Expenditure function, 80
Expenditure weight, 26-27, 52, 86, 166, 167, 307
Expenditure-weighted averaging, 26, 52, 60

F

Fads, 159 n.7
Federal income tax system. *See* Taxes
Fisher ideal index
for aggregation, 52, 86, 215, 221, 248
COLI approximation, 82, 83-84
construction approach, 42-43, 50
defined, 77, 308, 310
introduction of new goods, 162 n.10
mathematical approach, 77, 83-84, 85
plutocratic, 79
production lag, 42-43
publication of research series, 210
quality change adjustment, 152, 153
real-time, 219-220
and substitution bias, 23, 52, 60, 248 n.18

Fisher, Irving, 23, 41
Fixed-weight index, 308
Food and beverages, 64, 83, 112, 117, 146, 159, 160, 170, 173, 174

G

Geometric means formula, 5, 23, 24, 39, 50, 60, 77, 127, 210, 215 n.24, 308
Gillingham-Greenlees exact index, 213-214
"Goodness" of goods, 64
Griliches neutrality, 151, 152
Gross domestic product (GDP), 13-14, 98 n.4, 100, 179, 209, 211, 216, 217, 218, 220, 221, 240
Group indexes. *See* Subgroup indexes

H

Health insurance, 9, 10, 21, 104-105, 179-180, 185-188, 189, 203, 220
Health status, as utility function, 87-90, 95, 187
Hedonic functions, 123, 149
Hedonic quality adjustment, 64
  advantages, 123
  assessment of methods, 7, 145-146
  audit of other methods with, 6, 120, 128
  backward-looking method, 150
  bias in, 128, 129, 133, 139, 140
  BLS application of, 108, 113 n.6, 129-140
  brand-specific, 132-133, 145-146
  candidates for, 28-29, 119, 124, 130-132, 142
  characteristics method, 126, 128-129, 133, 142, 143-144, 152-154
  comparability decisions for, 6, 28, 29, 134-139, 140, 142, 145
  complementary products, 142-143
  consumer heterogeneity and, 124-125
  CPI model, 6-7, 132-137, 138, 139-140, 141
  data requirements, 124, 126 n.19, 128-129, 131-132, 135, 143, 144, 269
  defined, 101
  direct methods, 29, 121, 126-129, 134, 143-144, 151-154, 306, 307
  effect on CPI, 134-137, 138
  forward-looking method, 150
  indirect methods, 29, 125-126, 127 n.22, 129, 130, 132, 134, 135-136, 144, 150-151, 308-309
  limitations of methods, 140
  list price vs. transaction price, 132
  mathematical description of, 123-124, 149-154
  in MCPI, 187-188
  new goods and, 7
  outlet substitution and, 174, 176
  panel cautions, 6-7, 133-134, 139-140, 141, 144-145
  principle, 6, 122
  recommendations, 6-7, 140-146
  reliability, 7, 124, 129-130
  time dummy method, 126-128, 143, 145-146, 151-152, 153
  updating intensities, 144
  uses, 29, 122, 123, 124, 130, 135
Hedonic treadmill hypothesis, 56, 69-70
Heterogeneity. *See also* Subgroup indexes
  across-stratum, 223-224, 227, 229
  and aggregation, 4, 26, 125 n.18, 223-226, 229, 240-241
  and construction of indexes, 18, 26, 226, 240-241
  and data collection methods, 223, 241
  and hedonic quality adjustment methods, 124-125
  and inflation rates, 222, 225-226
  and outcomes-adjusted medical index, 184
  in prices paid, 46, 225-226, 227 n.3
  public goods preferences, 99 n.6
  in substitution behavior, 26, 224
  within-stratum, 223, 227-228
Household
  aggregation across, 4, 223-226, 240-246
  disaggregation into individual members, 45
  furnishings and equipment, 72, 254
  modeling non-unitary behavior, 45
  production function, 123 n.16
  scanner technology, 271-272
Housing. *See* Shelter

I

Imputation of prices, 121, 136, 159, 305
Income distribution, 24-25, 26, 53, 199, 206, 227, 239-240, 243
Income effects, 2, 90, 224
Index drift, 167, 168 n.17, 257, 308
Indifference curve, 80, 86, 90, 91
Individual price indexes, 44-46, 78-79

Inflation rates
age stratification, 246, 249-251
core, 218
effects of bias in indexes, 217-218
heterogeneity and, 222, 225-226
income stratification, 239-240, 247-249
long-run, 61
macroeconomic policy indicators, 1, 97, 198, 216-221
NIPA vs. CPI, 217
quality change adjustments and, 27, 225
and relative price variation, 225
stripped, 218-219
subgroup differences, 222, 239-240, 246-251
tax-rate indexing, 13, 40, 43, 212-214, 222
Treasury securities indexed to, 1, 192, 210-212
Information Resources, Inc., 266, 271
Input substitution, 10, 181-184, 188-189
Interest rates (nominal), 72
Internet, 45, 170-173
Item reclassification, 31, 155, 162, 163, 165-166, 167
Item replacement. *See also* Substitution
class-mean method, 117, 118, 119, 121, 134 n.30, 136-137, 139, 305-306
COLI calculations, 110
comparability decisions by analysts, 6, 28, 29, 116-117, 118, 134-139, 140, 142, 145
defined, 106, 309
deletion method, 117, 118, 119-120, 125, 128, 135, 136-137, 139, 305, 307
direct comparison method, 109, 116-118, 139, 307
explicit cost-based adjustment, 28-29, 118-119, 125, 134, 307-308
in-store procedure, 125, 126, 134, 156
overlap pricing, 117, 118, 119, 310
quality adjustments, 28-29, 31, 106, 110, 116-121, 125, 130-132, 135, 156
research recommendations, 141
targeted, 8, 130-132, 164, 166
within-sample, 156, 161
Item rotation, 31, 164, 168
Item strata, 23, 156, 162, 164, 167, 307, 309, 311
Item weighting, 247-248, 309

K

Konus, Alexander, 23

L

Laspeyres index, 308
advantages, 44, 60, 75-76
aggregation, 77-79, 86
and COLI, 47, 48-49, 50, 51, 53, 59, 80-82, 87, 90-91
complexity, 58
conceptual basis, 42
defined, 42, 44, 309
democratic, 242
domain, 68
escalation with, 81, 215
as inflation indicator, 218, 247, 248
introduction of new goods, 158 n.6, 161-162
mathematical approach, 74-79
modified for CPI, 25, 62, 97 n.2, 111, 114, 168 n.17, 272-273, 306, 309
national vs. individual, 44-46, 78
outlet rotation and, 168 n.17
quality adjustments, 111-112, 129, 152, 153, 154
quantity used for pricing, 72, 78, 82-83
representative agent, 246, 247, 248
seasoned, 62
substitution bias, 5, 21-23, 24, 59, 60, 62, 251 n.20
utility function, 92
weights, 45, 75
Law of one price, 169 n.18, 172, 180, 187
Life expectancy, 4, 20, 28, 67, 73, 96, 97-98, 184
Link period, 309

M

Macroeconomic policy, inflation indicators for, 1, 35, 40, 97, 192, 216-221
Marshall Edgeworth price index, 77
Medical Care Price Index (MCPI)
aggregation, 182, 188
bias in, 148, 185-186, 188, 250-251
conceptual issues, 17-18, 19, 30, 96, 181-185

consumer heterogeneity and, 224, 228
current procedures, 178-181
data collection, 9, 183, 255-256
defined, 309
diagnosis-based pricing, 9, 182-184, 188-189
domain, 9, 10, 20 n.5, 86-90, 96, 97-98, 105, 178-179, 185-186, 189-190, 220
for the elderly, 197, 203, 249, 250-251
health insurance considerations, 9, 10, 21, 104-105, 179-180, 185-188, 189, 220, 255
hospital and physician services, 30, 112, 148, 181
input substitution, 10, 181-182, 188-189
measurement issues, 17-18, 181-185
outcomes-adjusted, 10, 181, 184, 187, 190
prescription drugs, 30, 112, 148, 163, 186
quality adjustments, 10, 19, 20 n.5, 28, 30, 40, 67, 86-90, 112, 114, 117, 148, 181, 184-185, 187-188, 190
recommendations, 9-10, 105, 188-190
weights, 9, 179, 180, 182, 183, 184, 185, 186, 189, 203, 214 n.22, 220, 255
Medicare and Medicaid, 9, 180, 182, 185, 203, 214 n.22, 220, 225, 255
Microwave ovens, 29, 116, 118, 131, 133, 137-138
Motor vehicle accident rates, 96

N

National Bureau of Economic Research, 311
National economic accounts, 4, 20, 101-102, 105
National Income and Products Accounts, 89, 102 n.8, 179, 203, 214-215, 216, 217, 219-221, 253, 254
National price indexes, 4, 8, 25, 26, 44-46, 78-79, 87, 160, 214, 219-221
National security, 21, 96, 97
New goods
bias, 31, 32, 107-108, 112-121, 146-148, 156-158, 160-161, 162, 273
data collection issues, 30-32, 163-165, 273
defined, 7, 155-156
demand for, 158 n.6, 159-160
effect, 31-32, 67-68
improved goods contrasted, 106-107
introduction into index, 7-8, 31-32, 97-98, 111, 112, 120 n.12, 131 n.25, 135, 155 n.1, 161-167
item reclassification, 31, 155, 162, 163, 165-166, 167
measurement issues, 7, 31-32
outlet rotation and, 7, 31, 32, 135, 155, 164, 168
in output-change indexes, 215, 216
plutocratic vs. democratic index, 160 n.8, 162
price cycle, 32, 161-163
quality-improved items, 31, 156, 161
recommended adjustments for, 8, 160-161, 165
sample rotation frequency, 8, 31, 163-164, 165, 167
supplemental, 107, 155 n.1, 163
valuation of, 3, 18, 64, 73, 110, 155, 157-161
variety considerations, 110 n.3, 114, 146
virtual price vs. introductory price, 8, 31, 157-160, 311
weight updating, 32, 155, 163, 166, 167
welfare effects, 31, 157, 159, 160
New outlets. *See* Outlet rotation; Outlet substitution
Nonlinear pricing, 107
Nonsampling errors, 260 n.5
Norwood, Janet, 228

O

Outcome pricing, 190
Output deflation, 1, 13-14, 35, 98, 101-102, 214-216, 218, 240
Outlet rotation
cross-outlet price linking, 169, 171
current practice, 167-173
defined, 309
e-commerce and, 170-173
hedonic techniques applied to, 174, 176
market shift to discount stores and, 8, 33, 169-170, 173, 176
new-product introduction, 7, 31, 32, 135, 155, 164, 168
pricing components, 9, 171, 172, 174-175, 176, 263
probability of selection, 167-168
purpose, 167-168
quality-related issues, 8-9, 32-33, 168-170, 171, 172, 173, 174-175, 176
recommendations, 9, 176-177
sampling frequency, 33, 164, 173
substitution bias, 8, 33, 128, 168 n.16, 170, 172, 173-174

Outlet substitution, 8, 128, 168-170, 172, 173-174, 176, 224, 230, 273, 310
Overlap pricing, 117, 118, 119, 176, 310

P

Paasche indexes, 308
  aggregation, 77-79
  averaging, 79
  COLI and, 48-49, 50, 51, 53, 80-82, 87, 91
  conceptual basis, 42
  defined, 42, 44, 310
  democratic, 79
  introduction of new goods, 158 n.6, 162
  mathematical approach, 75-79, 91
  national vs. individual, 79
  quality adjustments, 129, 152, 153, 154
  representative agent, 246, 247, 248
  substitution bias, 22-23, 251 n.20
  usefulness, 57
  utility function, 92
Penetration pricing, 163 n.11
Pension plans, employer-provided, 104, 203-204
Personal consumption expenditure (PCE) data, 12, 37, 253-256, 274-275, 310
Personal consumption expenditure (PCE) deflator, 220, 221, 310
Plutocratic indexes
  aggregation, 4, 5, 26-27, 45, 77-79, 86, 235 n.9
  bias, 244
  COLI, 52, 242, 244
  defined, 52, 222, 310
  introduction of new goods, 160 n.8, 162
  output-change measurement, 215
  for public transfer program indexing, 198-199
  for subgroups, 5, 26-27, 229, 237-240
Point of Purchase Survey (POPS/TPOPS), 11, 36-37, 163, 167-168, 227 n.3, 232, 233 n.5, 234, 262, 264-266, 272, 281-282, 310
Pollution, 4, 97, 98, 101-102, 103
Preferences
  and COLI, 47, 49, 50, 54, 79-80
  flexible functional form, 84, 86, 142 n.35
  homothetic, 49, 50-51, 60, 82, 84, 92, 308
  revealed, 54, 99
  separability of, 68
Price heterogeneity, 46, 225-226
Private contracts, indexing, 35, 97, 192, 208-210, 218
Private goods
  as bundles of characteristics, 122-123
  characteristics, 108, 111, 112
  defined, 20, 310
  domain of CPI, 4
  government-produced and -sold, 19, 20, 94 n.1
  separability from public goods, 68
Producer price index (PPI), 130, 148, 174 n.23, 209, 219, 220
PSU, 278-279, 281-282, 310
Public goods
  defined, 20, 310
  domain considerations, 19, 20-21, 94, 96, 98-100, 102-103, 104
  pure, 21
  sales taxes and, 21, 102-103
  separability from private goods, 68, 103
  valuation problems, 19, 98-100
Public perceptions
  of indexes, 39, 58-59
  of quality, 111
Public transfer payment indexing, 13
  beneficiaries with other income, 195-197
  consumer heterogeneity and, 226, 231
  CPI-U vs. CPI-W, 193, 197-198
  criticisms of CPI, 13
  domain issues, 97
  for elderly people, 196-197, 207
  lagged superlative index for escalation, 24, 25, 191, 194-195
  plutocratic vs. democratic indexes, 198-199, 240
  poverty-related programs, 206-207, 247
  recommendations, 194-195
  social security adjustments, 13, 25, 193-199, 200, 204-206
  SSI, 193, 206
  for subgroups, 192-193, 206, 231
  tax considerations, 195-196
  wage indexes, 199-206
Pure price changes, 29, 127, 216, 230

Q

Quality change adjustments
  and aggregation, 125 n.18
  apparel, 29, 117, 118, 120, 130, 147

appliances and electronics, 29, 107, 109, 110, 114, 118, 120 n.12, 130, 131, 132, 133, 147
bias in, 27-28, 112-114, 121-122, 129, 133, 139, 140, 144, 249
BLS approaches, 39, 73, 114-122, 129-140
Boskin Commission criticisms, 27-28, 109 n.2, 112-114, 115, 116-117, 119, 120, 146-148
COGI view, 2, 62-64, 111-112
COLI view, 2, 3, 18-19, 39, 57, 62-64, 86-90, 109-111, 112-114, 115
conceptual issues, 14, 19, 28, 30, 39, 40, 57, 59, 62-64, 98, 106
cost-based, 28-29, 118-119, 125, 134, 307-308
in CPI, 14, 16, 18-19, 27, 39, 106, 111, 112-114, 115, 116-121, 130-137, 209-210
defined, 114
domain issues, 67, 95, 97-98, 101, 115
durability of goods and, 71-72
food and beverages, 64, 108, 112, 117, 123, 146
forms of, 106-107
and inflation rates, 27, 225
item replacement methods, 27, 28-29, 31, 106, 110, 111, 116-121, 134, 140-141
items targeted for, 130-132
measurement/modeling, 28-30, 63-64, 108, 122; s*ee also* Hedonic quality adjustment
medical care, 10, 19, 20 n.5, 28, 30, 40, 67, 86-90, 96, 112, 114, 117, 148, 184-185, 187-188, 190
new goods bias, 31, 107-108, 112-114, 115, 146-148, 156, 161
outlet rotation and, 8-9, 168-170, 171, 172, 173, 174-175, 176
per unit approximations, 107
private contracts, 209-210
rates, 106, 137, 139
recommendations and cautions, 140-146
repackaging framework, 107-108, 111-112, 116-117, 156
for services, 115, 122, 141-142, 148, 168
shelter, 146-147
subgroup indexes and, 230
taste change and, 16, 64, 69, 95, 111
transportation, 28-29, 109, 110, 111, 117, 119, 123, 147-148
valuation of new goods, 3, 18, 64, 73, 156
variety of goods and, 110 n.3, 114 n.8
wage indexes, 204, 209-210
within-sample, 115, 121-122, 140, 156
Quality-of-life metrics, 185

R

Reference period, 306, 309, 310, 311
conceptual issues, 41, 42, 47-49, 51
hedonic function, 125
Regional/city price indexes, 66-67, 237, 252
Representative consumer, 241-243
COLI for, 244-246
conditions for existence of, 243-244
Retail Trade Survey, 255

S

Sample rotation, 8, 31, 33, 119-120, 140, 156, 157, 163-164, 165, 167, 172
Sampling errors, 260 n.5
Satellite accounts, 4, 101-102, 103, 105
Satisfaction. *See also* Utility
concept, 47, 53, 55-57
hedonic treadmill hypothesis, 56, 69-70
measures of, 55-57
and utility, 55-57
Scanner data
and COLI-based CPI, 272-273, 275
and hedonic modeling, 142 n.35
household-based technology, 271-272, 275
limitations, 270-271
point-of-sale, 266-271, 275
and quality of CEX and POPS, 272
Seasoning procedure, 62
Semiconductors, 162
Separability of preferences, 68
Services
expenditure categories, 115, 122, 141, 148
flows from consumer capital goods, 33-35, 71-72, 102
pricing, 163 n.12, 172
Shadow price, 90, 158
Shelter
compensation issues, 70-71
consumer heterogeneity and, 224, 236 n.11
data collection, 254, 256, 262 n.6, 263
homeowners vs. renters, 70-71, 72
item category, 307

outlet substitution and, 170
price indexes, 45, 72
quality adjustments, 29, 130, 146-147
service flows from housing, 33-35
Shephard's Lemma, 82-83, 90, 91
Slutsky matrix, 90, 91
Social cost-of-living index, 51-52, 85-86, 242
Social issues, 96
Social security benefits
aggregation issues, 4, 200-201
bias in CPI, 14
consumer heterogeneity and, 226
cost-of-living adjustments, 1, 11, 13, 25, 40, 48, 61, 193-195, 198, 240
CPI-U tied to, 198
democratic vs. plutocratic weights, 240
design of indexes, 11, 35, 40, 43, 48, 192
domain issues, 97, 104, 196
employers' contributions to, 104
initial retirement benefit, 200
other income with, 71, 195-197
rate of increase, 205-206
superlative index for escalation, 11, 6, 25, 61, 198
tax-and-price index tied to, 195-196
timeliness of data collection, 6, 43
wage-based indexing, 11, 199, 200-201, 204-206, 246 n.17
Standard of living
compensation issues, 48, 71
defined, 311
environmental considerations, 19, 95, 96-97
measurement for COLI, 15 n.2, 41, 46-47, 48, 49, 55-56, 59, 63, 69, 80
public goods and, 98
quality of goods and, 63
satisfaction and, 55-56
social security recipients, 71, 97
substitution behavior and, 73
Stigler committee, 1, 3, 14, 38, 73, 122 n.14, 311
Stochastic approach, 41-42, 68
Stocks and flows, 33-35, 71-72, 102
Subgroup indexes. *See also* Aggregation; Heterogeneity
aggregation issues, 4, 5, 26-27, 229
assembly of data for, 232-235
barriers to production, 226-228
biases, 250, 251
conceptual bases, 229-232
costs, 232-235
data collection, 4, 5, 27, 197, 226-228, 229, 231-235, 252, 276
for elderly people, 11, 27, 125, 196-197, 199, 227, 228-229, 231, 249-251
income-stratified, 27, 125, 206-207, 227, 231, 247-249
inflation rates and, 246-251
plutocratic vs. democratic weights, 5, 26-27, 229, 235 n.9, 237-240
and quality changes, 125, 230, 251
recommendations, 5, 97 n.2
with reference-period weights, 194
research and testing suggestions, 235-237
sample size, 233, 234
of strata prices, 234, 238-239
uses, 231
Substitution. *See also* Consumer substitution
and aggregation, 24-25, 26, 52
bias, 59-62, 93, 113, 128, 147, 168 n.16, 170, 172, 173-174, 248 n.18
CPI, 15-16, 23-25, 113, 224, 230
data collection methods and, 269
elasticity of, 6, 60-61, 91-93, 194, 269
input, 10, 181-184, 188-189
lower-level, 147, 309
new goods, 164-165
outlet, 118, 128, 135, 168 n.16, 170, 172, 173-174, 310
upper-level, 311
within-strata effect, 5
Superlative indexes. *See also* Fisher ideal index; Törnqvist index; Walsh price index
advance estimate of, 195
aggregation, 52, 53, 215
as COLI approximation, 50-51, 57, 95-96
compensation calculations, 60, 61, 194-195
conceptual bases, 50-51, 53, 83-85
data collection for, 259, 272-273
defined, 308, 311
domain, 95-96
endorsement for CPI, 97 n.2
environmental variables, 53, 87, 95-96
as inflation indicator, 218, 248
introduction of new goods, 162 n.10
lagged for escalation, 24, 25, 191, 194-195, 207-208, 210
mathematical approach, 83-85
quality adjustments, 129, 131

from social aggregate indexes, 86
for social security cost-of-living adjustments, 11, 6, 25, 61, 198
substitution effect, 5, 6, 23, 24, 25, 60, 61, 84, 85, 93, 96, 113 n.6, 215
taste changes and, 25
Supplemental goods, 107, 155 n.1, 163
Supplemental Security Income (SSI), 206-207

T

Taste changes
COGI vs. COLI, 2, 16, 24-25, 69-70
conceptual issues, 53, 59, 64, 67, 69-70
domain issues, 67, 69, 95
habit formation, 70
and quality changes, 16, 64, 69, 110 n.3
and substitution effect, 25
Tax-and-price index, 195-196
Taxes
aggregation across households, 4
consumer heterogeneity and, 226
data collection issues, 270
democratic vs. plutocratic weights, 240
design of indexes, 35, 40, 43, 192, 212-214
direct consumption costs, 213
domain issues, 20-21, 97, 102-103
exact indexation measure, 213-214
indirect, 102 n.8, 195
public goods offset to, 20-21, 102-103
and public transfer program indexing, 97, 195-196
rate-related indexing, 13, 40, 43, 212-214
Taylor series approximation, 90-91
Technological change, 45, 53
domain issues, 67-68, 88-89, 97
"new goods" effect, 67-68, 162
Televisions, 29, 110, 130, 133, 135, 137, 142, 147
Test approach, 41-42, 50, 68, 267
Theoretical bases of indexes
basket-price, 43-46
and conceptual basis, 39, 40, 43-58, 73, 74-93
cost-of-living, 2, 15, 40, 46-53, 58, 79, 83, 86, 110, 113
CPI, 4, 73
criticisms of, 40, 52-53
satisfaction and, 47, 55-57
utility and choice and, 47, 53-55, 110 n.3
Tiebout hypothesis, 99 n.6
Time reversal test, 76-77
Törnqvist index, 85, 93, 248, 311
Transaction costs, 71, 72
Transfer payments. *See* Public transfer payment indexing; Social security benefits; *specific programs*
Transportation, 72, 109, 112, 117, 122, 147-148, 173, 233 n.7
Treasury Department, 259
Treasury Inflation-Protected Securities (TIPS), 1, 192, 210-212

U

Union wage contracts, 13, 207-208, 218
Unit pricing, 268-269
Unit value indexes, 236
User costs, 34, 71-72
Utility
and choice, 47, 53-55, 110 n.3
decision, 54-55, 57
defined, 53-54, 311
duration neglect bias, 54
environmental factors and, 86-87
experienced, 54, 55, 56-57
function, 79-80, 83-84, 85, 86-90, 92
measures of, 53, 54, 55, 57
predicted, 54
remembered, 54
satisfaction and, 55-57

V

Variety of goods, 110 n.3, 114, 146
VCRs, 8, 31, 115, 131, 132, 133, 137, 138, 139, 156, 162, 163, 165, 166
Virtual price, 8, 31, 157-160, 311

W

Wage indexes
aggregation issues, 200-201
alternatives, 199-206
design issues, 11, 35, 199-206, 207-208
employer-paid benefits in, 201-204
escalators in union contracts, 13, 207-208
hourly earnings, 202 n.9
macroeconomic implications, 205, 218

mean wage measure, 202, 203 n.10
median wage measure, 202, 205
quality adjustments in, 204
social security benefits tied to, 11, 199, 200, 204-206
Walsh price index, 77, 85
Weighted arithmetic mean, 75, 215 n.24
Weighted harmonic mean, 75, 79
Weights/weighting
across-stratum, 227
democratic, 45, 78, 237-240
expenditure, 26, 45, 60, 76 n.1, 86, 166, 167, 198, 227, 307
fixed-weight index, 308
item (upper-level), 97 n.2, 309
Laspeyres index, 45, 75
MCPI, 9, 179, 180, 182, 183, 184, 185, 186, 189, 214 n.22, 220
for new products, 8, 32, 155, 163, 166, 167
plutocratic, 45, 78, 79, 237-240
updating, 8, 32, 96, 155, 163, 166, 167
Welfare effects, 31, 157, 159, 160, 242
Welfare index, 14
Well-being, inputs to, 109-110, 115
Within-sample
item replacement, 156, 161
quality adjustments, 115, 121-122