Volume Title: Price Index Concepts and Measurement

Volume Author/Editor: W. Erwin Diewert, John S. Greenlees and Charles R. Hulten, editors

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-14855-6

Volume URL: http://www.nber.org/books/diew08-1

Conference Date: June 28-29, 2004

Publication Date: December 2009

Chapter Title: A Review of Reviews: Ninety Years of Professional Thinking About the Consumer Price Index

Chapter Author: Marshall Reinsdorf, Jack E. Triplett

Chapter URL: http://www.nber.org/chapters/c5068

Chapter pages in book: (17 - 83)

1

# A Review of Reviews
## Ninety Years of Professional Thinking About the Consumer Price Index

Marshall Reinsdorf and Jack E. Triplett

> It is not often that a price index, a tool of statisticians, becomes an object of political debate.
> —Ostrander 1944, 849

> The theory of price indexes is usually left to specialists, but when a suspicion that something has gone wrong is coupled with the possibility of large political and fiscal benefits from fixing it, the topic can move into the limelight.
> —Deaton 1998, 37

### 1.1 Introduction

The U.S. Bureau of Labor Statistics (BLS) first published an index of consumer prices for food at home in 1903, with continuous publication uninterrupted by budget shortfalls beginning in 1911 (Goldberg and Moye 1985, 37). The next milestone in the development of the index now known as the U.S. Consumer Price Index (CPI) came in 1914, when methods were improved and the index basket was expanded to include cloth and

Marshall Reinsdorf is Chief, Economic Analysis and Research Group, National Economic Accounts, at the U.S. Department of Commerce, Bureau of Economic Analysis. Jack E. Triplett is a nonresident senior fellow at the Brookings Institution.

We appreciate comments from the editors, John Greenlees and Charles Hulten, and also from Bert Balk, Michael Boskin, Angus Deaton, Erwin Diewert, and Charles Schultze. The usual disclaimer disassociating some or all of those who made comments from the content of the chapter applies. The views expressed should not be attributed to the Bureau of Economic Analysis, or to the Brookings Institution, as they are solely those of the authors. This original version of the chapter was presented at the NBER-CRIW conference on price indexes, Vancouver, B.C., June 2004.

clothing. The result was the earliest version of the CPI that is still available from BLS.

A decision during World War I by the Shipbuilding Labor Adjustment Board to escalate wages by a price index led to the development of BLS's first comprehensive index of consumer prices. The existing index, composed almost entirely of fast-rising food prices (which were strongly affected by the war), was obviously unrepresentative of consumer prices in general, so a consumer expenditure survey was conducted to develop a broader index basket. Initially this survey included just the shipbuilding cities, but in June of 1918 its scope was expanded to include other cities. Publication of a national index of consumer prices with weights reflecting survey data on purchasing patterns began in 1919 (Goldberg and Moye 1985, 105).

The U.K. Retail Price Index began, similarly, as a wage escalator during the same war (Roe and Fenwick 2004), as did the Canadian CPI (Statistics Canada 1995, 9–10; Urquhart and Buckley 1965, 287–89). The Swedish index began a little earlier (Dalen 2001). Indexes in all four countries were originally called Cost-of-Living Indexes, but new names were introduced in all four cases after World War II.

The history of economists' analysis of BLS price indexes begins with the assistance of Irving Fisher and Wesley Mitchell, with the food index improvements of 1914 (Goldberg and Moye 1985, 91–92). Here Fisher's hand seems evident in the abandonment by BLS of a method that he often criticized, the averaging of price relatives. A year later, Mitchell increased awareness in the United States of the distinction between consumer prices and what we now call the Producer Price Index (PPI) in a review of what was then called the Wholesale Price Index, or WPI, (Mitchell [1915] 1921; summarized by Banzhaf [2001]). Distinguishing consumer prices from wholesale or producer prices may seem obvious now. But as we point out in section 1.2, Mitchell's distinction was not considered in most index number writing before that time.

Over the years, reviews by committees and panels have critically influenced the development of the U.S. CPI. Moreover, the long record of the debate over CPI concepts and methods preserved in their reports provides important background for understanding the current state of the discussion of CPI methods. This chapter examines the treatment of two questions, one conceptual and one methodological, in officially sanctioned reviews of the CPI.

The conceptual question is, what measurement objective is the appropriate one for the CPI? A modern statement (Schultze and Mackie 2002) presents the alternatives as COLI versus COGI (Cost-of-Living Index versus Cost-of-Goods Index). The conceptual question concerns not just the index number "formula," as it is often called, but essential questions about what components are included in the index, and how the components are to be measured. This question is the subject of our section 1.2.

The methods question is, what sampling procedures and formulas should be used to construct the lowest level, detailed component indexes, or elementary aggregates of the CPI? We discuss the methods question in section 1.3.

Though our topic begins in 1914, the formal reviews that we consider begin with the World War II era reviews of what was then called the Cost-of-Living Index.[1] Four reviews were produced as part of the same investigation; this set of reviews marks a major milestone in the history of thinking about CPI measurement issues.

The second formal review is the 1961 report by the "Price Statistics Review Committee" chaired by George Stigler (Stigler et al. 1961). The World War II criticisms of the procedures used to construct the detailed component indexes of the CPI laid the foundation for the reforms in sampling procedures recommended by the Stigler Committee and later adopted by BLS, and also for the Stigler Committee's concern for the problems posed by quality change. In turn, the Stigler Committee's report influenced the third formal review by the widely-discussed Boskin Commission. Finally, the BLS funded the fourth formal report by a panel selected by the Committee on National Statistics (CNSTAT), largely as a response to the Boskin Commission review. Continuity in the reviewers' recommendations is echoed in continuity among the reviewers, as the Stigler Committee and Boskin Commission included participants in the preceding reviews.

In the following sections, we review each of our topics chronologically.

## 1.2 Recommendations Concerning the Measurement Concept for the Index

The first BLS measure of consumer prices was influenced by Mitchell's review of the Wholesale Price Index, which he called "The Making and Use of Index Numbers" (Mitchell 1915).[2] As the title suggests, this report was more than just a review of the WPI.

1. We restrict our attention to reviews of the U.S. CPI, for lack of resources to conduct a broader investigation. A noteworthy wartime review of the Swedish CPI is described in Dalen (2001), as are several subsequent external reviews. The equivalent index in the U.K., known historically as the Retail Price Index, has been reviewed a number of times by committees established by the relevant Parliamentary Secretary; Roe and Fenwick (2004, appendix A) provide a list of these reviews and a summary of their recommendations. Melser and Hill (2005) present excerpts from Revision Advisory Committees on the New Zealand CPI. In Canada, a Price Measurement Advisory Committee to Statistics Canada meets on a regular schedule. Other similar bodies no doubt exist.

2. After a long period of neglect, Mitchell's contributions to price index history have recently been resurrected in an insightful paper by Banzhaf (2001). Mitchell made the connection between the design of an index number and its purpose so much a part of his approach to price indexes that it was repeated as a mantra by others years after. He was also perhaps the first to focus on measuring the components of the CPI, as opposed to the methods for aggregating them. This focus may partly explain the neglect of Mitchell: index number researchers are often more interested in the problems of higher-level aggregation than in the methods used for the lowest-level aggregates, even though that latter are probably more important empirically.

At the time of this report, consumer price indexes were often constructed by weighting WPI components by family expenditure data. Mitchell excoriated this practice: "To pretend that wholesale price index numbers when weighted on the basis of family expenditures show fluctuation in the cost of living is to overtax the credulity of those who know and to abuse the confidence of those who do not" (Mitchell [1915] 1921, 63). Partly as a result of Mitchell's review, when the comprehensive household inflation measure was launched in 1919, it was based on retail prices. The 1919 index was named the "Cost-of-Living Index" because it covered consumers' entire budgets with weights that reflected their purchasing patterns.[3]

So far as we know, the first professional review of the U.S. price index for households was conducted in 1933 and 1934 by an Advisory Committee appointed by the American Statistical Association. Its recommendations were practical, not conceptual, and concerned surveys to improve the weights for combining items and cities in the "all-items" index, data collection, and the use of imputations for items not priced directly (Hogg 1934; BLS 1966). The Committee suggested that even the cities in the index were not representative, since shipbuilding cities remained overrepresented.

Yet the interwar period was not devoid of thinking about conceptual questions related to the CPI—quite the contrary. The appendix reviews interwar price index research as background for the remainder of the chapter; in the next section, we provide a summary.

### 1.2.1 Interwar Price Index Research

Major developments in the 1920s influenced professional thinking about CPIs and were absorbed into the intellectual tradition that influenced subsequent reviews of the CPI. The first, the "test approach," had its origins in nineteenth century discussions of index number properties and in a 1901 book by Correa Walsh, but its major development came in the work of Irving Fisher. Fisher's (1911) exposition of this approach was focused on finding the $P$ and $T$ terms of the equation of exchange $MV = PT$. In Fisher (1922), the approach was applied to all kinds of index numbers. There is no evidence that Fisher thought much about a specific application to the measurement of consumer inflation.[4]

The second development was Konüs' (1924) theory of the Cost-of-Living Index (COLI). In contrast with Fisher and his forebears, Konüs' contribution was uniquely a contribution to the measurement of consumer prices, inspired, no doubt, by the new "Cost-of-Living Indexes" in the United States, Canada, the United Kingdom, and other countries.[5]

3. Meeker (1919). By 1943, its full name was: "Index of the Cost of Living of Wage Earners and Lower-Salaried Workers in Large Cities." See Ostrander (1944).

4. Diewert (1993, 34) cites Joseph Lowe, who wrote much earlier, as the "father of the consumer price index." Lowe explored a number of problems, including indexes for different demographic groups.

5. See Diewert (1983). A third development from this period, the Divisia index, is discussed in Balk (2005), and used by Hulten (1973).

The appendix sets out our position on test (or axiomatic) and COLI (or economic) approaches to the CPI. In summary, we emphasize three points:

1. It is sometimes said that test and economic approaches converge because each leads to one of the superlative indexes, especially the Fisher index. However, as applied by the advocates of this approach, the tests are an arbitrarily chosen set of index number properties. *The test approach does not yield the Fisher index, unless the tests are chosen to get this result.*
2. The favored index number properties include some that are objectionable on economic grounds. Moreover, advocates of the test approach invariably exclude some other properties (for example, additivity and consistency in aggregation—see the appendix) that are desired by some index number users. The arbitrariness of the set of index number properties traditionally included in the tests means that some other set of properties would yield, not the Fisher index, but some other index. Moreover, the absence of criteria for selecting the tests is matched by the absence of criteria for ranking the importance of the various tests, except by arbitrarily ranking all tests equally (contending, for example, that the Fisher index, or some other index, passes more tests than other indexes). The value of any index number property depends on the index number purpose, so no system of discriminating among tests, including equal weighting, has universal applicability.
3. The economic approach to index numbers is much more than a framework for determining how index number components should be aggregated (the index number formula). It provides a framework that can be used to analyze the domain of the index (the components that are included) and to analyze how index components should be measured. The test approach is completely silent on these essential matters.

Although the test approach was not used in some of the professional reviews of the CPI, we include it in this section and in the appendix because it has come back into vogue in recent years and because we think its limitations are not always understood.

### 1.2.2 Wartime Committees: Clarifying the Meaning of the BLS "Cost-of-Living" Index

In 1942, the National War Labor Board, in what was known as the "Little Steel Agreement," permitted wage increases that matched increases in the BLS Cost-of-Living Index, as it was still then called (citation to the decision is given in Ostrander [1944, 850]). The Labor Board's action repeated the escalation use of a BLS index in the previous war. Very soon thereafter, the labor unions attacked the index as a flawed measure—indeed, an understated measure—of the change in cost of the workers' living standard.

The unions' complaints ultimately led to the preparation of four different reviews of the index. The Labor Department initiated the first, conducted by a special committee of the American Statistical Association chaired by

Frederick Mills, but this review did little to assuage the unions.[6] Almost immediately thereafter (within a month), President Roosevelt set up a "tripartite" (that is, having labor, management, and government members) Presidential Committee on the Cost of Living, chaired by War Labor Board chairman William Davis, to investigate the matter. The Davis Committee also failed to bring about consensus. Instead, it generated a minority report issued by its labor members, George Meany of the American Federation of Labor (AFL) and R. J. Thomas of the Congress of Industrial Organizations (CIO) (the Meany-Thomas report), a business report issued by the National Industrial Conference Board,[7] and finally, a report by a "Technical Committee," which was chaired by Wesley Mitchell. The staff and members of the Mills and Mitchell Committees make up a veritable list of prominent economists and statisticians of the time.[8]

The wartime dispute is unique among political discussions of consumer prices in three respects: (a) the topic was alleged understatement of inflation, not overstatement, as was true of all subsequent reviews; (b) the unions appeared (though the language is not precise) to support the concept of the COLI, unlike their position subsequently; (c) the professional reviews chaired by Mills and Mitchell supported (sometimes, we suspect, too uncritically) BLS methodology—subsequent professional reviews of the CPI range from mildly to overwhelmingly critical.

Examining the period from January 1941 to December 1943, Meany and Thomas (1944) estimated that the true rise in the cost of living was 43.5 percent, compared with only 23.4 percent reported by the BLS—see figure 1.1. The unions gathered some of their own data; the report's empirical sections are considered in part 1.3.

Among many alleged sources of downward bias in the Cost-of-Living Index that the Meany-Thomas report identified, some reflect the authors' views on what the index should have measured. In particular, Meany and Thomas contended that consumers were often forced to substitute more expensive varieties or goods for ones that had disappeared from the marketplace because of wartime shortages or "product line upgrading." They also alleged that consumers were often forced into more expensive dwellings than

6. *Washington Post* editorial, 14 February, 1946. See also Goldberg and Moye (1985, 154), Banzhaf (2001, 354), Mills et al. (1943), and Ostrander (1944).

7. Now the Conference Board. At the time and long thereafter, the organization published its own Cost-of-Living Index.

8. In addition to its chairman, the Mills Committee consisted of E. Wight Bakke, Reavis Cox, Margaret Reid, Theodore W. Schultz, and Samuel Stratton, with staff consisting of Dorothy Brady and Solomon Fabricant. The Mitchell Committee consisted of Mitchell, Reid, and Simon Kuznets, with Fabricant again on the staff. Reid, a prominent academic researcher on the subject of consumer behavior, was at the time on the staff of the Office of Statistical Standards in what is now the Office of Management and Budget (OMB), and Brady, also a prominent researcher on the same topic, was a member of the BLS staff in the 1930s, and again after the war. Reid subsequently returned to the University of Chicago and Brady joined the faculty of the University of Pennsylvania.

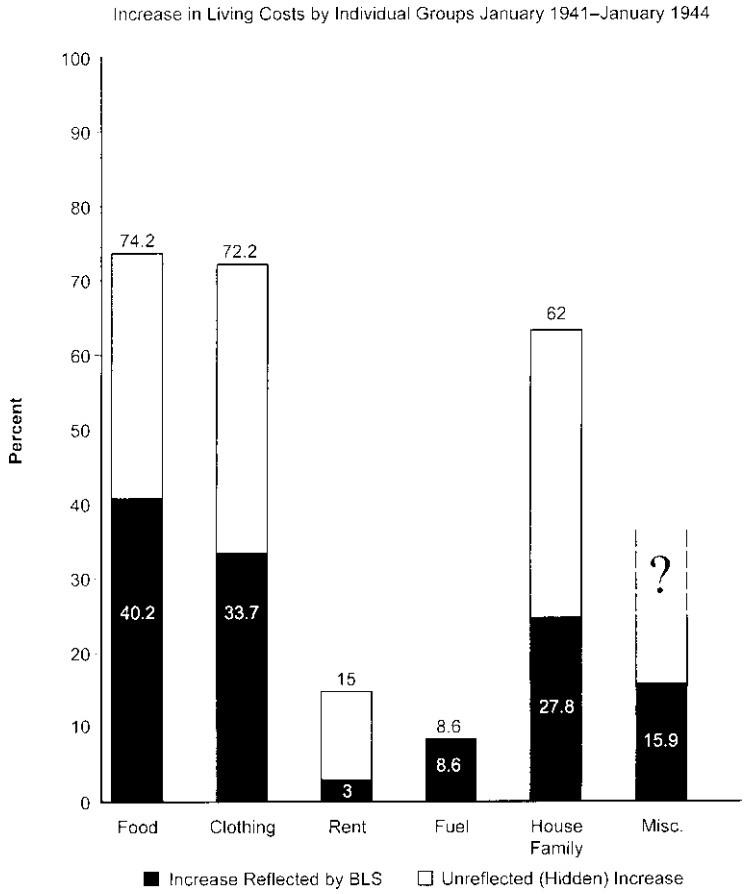


**Fig. 1.1 Illustration from the Meany-Thomas Report**

they wanted because of shortages of affordable housing. These involuntary substitutions raised consumers' cost of living in ways that the BLS index missed. In addition, lower-quality varieties often replaced higher-quality ones in the marketplace, which occurred when manufacturers relabeled a lower grade item as a higher grade one. The index, they contended, took no account of the quality decline. Finally, Meany and Thomas argued for the inclusion of extra expenses necessitated by lifestyle changes, such as increased consumption of restaurant meals due to meat rationing and the entrance of women into the labor force, and extra costs from migration between cities to fill wartime jobs.[9] In their report they did not treat effects

9. Accounting for the value of home production that was lost was overlooked: their only concern was the increased (monetary) costs of substituting market commodities for home production.

of rising standards of living as additional costs to be included in the Cost-of-Living Index, but—judging from rebuttals by their opponents—they were so interpreted.

The BLS vigorously disputed the Meany-Thomas report's contention that the Cost-of-Living Index understated inflation by almost half. The BLS's positions were largely supported by the outside experts on the Mitchell Committee.

An important part of the Mitchell Committee's response to the Meany-Thomas report was a clarification of the conceptual goal of the BLS Cost-of-Living Index. Though the concept of the COLI was known among economists (see section 1.2.1), the term "Cost-of-Living Index" did not always have the same meaning in 1944 that it does today. As Banzhaf (2001) points out, the term was then used, or interpreted, in at least three different ways.

- The now-standard meaning—a price index that holds constant the standard of living. The Mitchell Committee referred to this as a measure of the "real price."
- A fixed basket index that covered the entire family budget, which is what the BLS was in fact producing under the "Cost-of-Living Index" name.
- The cost of attaining the standard of living deemed appropriate, compared to the cost of a possibly lower standard of living in some previous period.

In its discussion of item substitutions involving a change in quality, the Mitchell Committee introduced consumer theory in a limited way as a basis for thinking about the design of the Cost-of-Living Index. When an item disappearance was thought to reflect voluntary substitution behavior by consumers, BLS usually introduced a replacement item into the index via overlap price linking. The Mitchell Committee argued that this procedure was justifiable under the assumption that the relative utility of different qualities varied directly with the ratio of their prices, which requires consumers to be informed about quality and supplies to be freely available (Mitchell, Kuznets, and Reid 1944, 11–12). If, for example, the replacement variety had a lower price, the larger quantity that could now be purchased with the same expenditure might be expected to yield the same utility as the smaller quantity of the higher-priced variety, whose quality was presumably higher.

However, in a passage with a slightly different problem in mind (forced substitution to lower qualities), the Mitchell Committee seemed skeptical about the existence of "a satisfactory way of measuring changes in 'real' prices—that is the price of a given quantity of utility, usefulness or service . . . when poorer qualities are priced" (19). Indeed, the Committee wrote: "To mix in additional factors with price changes would make the meaning of [BLS's] index even harder to determine than war conditions have already made it" (14). The Meany-Thomas "life style changes" were not to be con-

sidered. In the Committee's view, the Cost-of-Living Index ought to measure only the influence of prices on the cost of living, not the influence of other factors such as lifestyle changes, changes in taxes and government-provided services, or obviously, consumption increases that were in response to rising income.[10]

To clarify that nonprice influences on welfare were out of scope, and to avoid confusing the BLS index with one that included some changes in the standard of living, the Mitchell Committee recommended that BLS change the name of its index (20).[11] This recommendation was not intended to mark a change in the measurement goal of the index, which index experts had always understood to be limited to direct effects of prices. In particular, at the time of the name change, the term "Cost-of-Living Index" lacked the economic connotations that it now has, so the change in name should be interpreted as a statement about the domain of the CPI, and about public confusion between a price index and a standard of living index, not as a statement about today's debate over the Cost-of-Living Index concept.[12]

With respect to the other issues, the Mitchell Committee (Mitchell, Kuznets, and Reid 1944) concluded that the effects cited in the Meany-Thomas report were much smaller than claimed or that they were absent. The Committee anticipated the Boskin Commission of fifty years later in performing a "guesstimate" of the probable size of CPI error. It said that the combined effect of the all sources of bias mentioned by Meany and Thomas might be 3 to 5 percentage points over the three year long period, with an additional one-half point possible from the omission of smaller cities from the index, which implies an average rate of roughly 1.0 to 1.8 index points per year.[13] The largest part of the estimate concerned undetected, negative quality changes, set at 2 to 3 points. Scant attention was paid to possible upward biases that might offset the sources of downward bias, presumably because quality improvements were not thought to be of much importance in wartime, and because the Committee thought it infeasible to estimate substitution bias.

Though Meany and Thomas were not disinterested price collectors and their evidence was anecdotal, some of their exhibits were intriguing. For ex-

10. Some of the commentary on the debate considered whether the evidence suggested that workers' real consumption levels had declined. See Ostrander (1944).

11. The BLS did so in September of 1945 following the departure of Labor Secretary Frances Perkins, who had been opposed. The press release stated that the name change to "Consumers' Price Index for Moderate Income Families in Large Cities . . . should end the confusion and controversy caused by the misunderstanding of what the index is designed to measure and by the use of the index for purposes for which it is not adapted."

12. In the much later debate over the Boskin Commission (Boskin et al. 1996) recommendations, certain BLS statements about the 1946 change in name revealed misunderstanding of the episode. The inadequate discussion in Goldberg and Moye (1985) may have contributed to this confusion.

13. Ostrander (1944, 854) says that the 3 to 5 point range was subsequently lowered to 3 to 4.5 points.

ample, they presented menus from cafeterias showing increases in standard meals, which seemed substantially greater than the comparable restaurant component of the index. They documented (graphically, with a drawing) the deterioration in materials and workmanship in shoes, along with an estimate of the shortened lifetime of the shoes that would result. They pointed out that the items in the BLS sample (judgmental, in that era) were disproportionately the ones under price controls or subsidies, and that similar items with higher price changes were omitted from the index. From the vantage point of sixty years later, it is hard to avoid some sense that their evidence was dismissed too readily.[14] Indeed, the unions' complaints retained sufficient credence so that, when the Korean War broke out, BLS felt compelled to rush into production an "interim" revised CPI to avoid a repeat of some of the World War II criticisms (Goldberg and Moye 1985, 193). The CPI was nevertheless attacked in 1951 by the Union of Electrical Radio and Machine Workers (known as "UE," a union that was expelled from the CIO for being Communist dominated) and by the Soviet delegation to the UN Economic and Social Council in Geneva.

### 1.2.3 The Stigler Committee: A Welfare Index as the Measurement Concept for the CPI

In 1957 the Joint Economic Committee (JEC) undertook an investigation of "employment, growth and price levels," which inevitably raised questions about the price-making process and the measurement of prices. The need for reliable price statistics emerged as a minor theme in the subsequent hearings. Notable was a paper by Kenneth Arrow (1958), who argued for a Cost-of-Living Index objective for the CPI because of the importance of commodity substitution behavior (in U.S. Congress Joint Economic Committee [1958, 11]).

Subsequently, the U.S. Bureau of the Budget (now Office of Management and Budget) contracted with the National Bureau of Economic Research, which appointed a "Price Statistics Review Committee" chaired by George Stigler. The Stigler Committee included Dorothy Brady, who had participated in the Mitchell Committee investigation fifteen years earlier.[15] The Committee's report (Stigler et al. [1961]; hereafter, "Stigler Commit-

14. The *American Economic Review* article on the matter describes the Meany-Thomas report in language that is sometimes disparaging: "On the basis of the Meany-Thomas report, *assuming there were any substance to it,* the increased consumer expenditures by the end of 1943 would have been more than absorbed by price increases . . ." (Ostrander 1944, 853; emphasis added).

15. The other members of the Committee were: Edward Denison, Irving Kravis, Albert Rees, Richard Ruggles, Boris Swerling, and Philip McCarthy. Authors of staff papers included with the report were: Philip McCarthy, Victor Zarnowitz, Harry McAllister, Eleanor Snyder, John Flueck, Peter Steiner, Albert Rees, Zvi Griliches, Walter Oi, Geoffrey Shepherd, Earl Swanson, and Reuben Kessel.

tee report"), accompanied by twelve "staff papers," was transmitted to the Bureau of the Budget late in 1960.[16]

The Stigler Committee report made no explicit mention of the Mitchell Committee, but it did refer to the consensus by the participants in the World War II era debate that the CPI ought, in principle, to reflect the effects of substitution:

> In periods of wartime . . . price quotations on the virtually unobtainable commodities may not show much increase, or even be rigidly fixed by price controls. Consumers are driven to available substitutes, which are more expensive relative to desired performance (forced uptrading) or rise rapidly in response to expanding demands. Few economists or consumers come to the defense of the rigidly fixed market basket approach under these circumstances. This suggests strongly that what is in fact being measured is not the cost of a fixed set of consumer goods and services, but rather the cost of maintaining a constant level of utility. (51)

However, the Stigler Committee went beyond the Mitchell Committee in stating unequivocally that the measurement concept for the CPI ought to be the cost of staying on an indifference curve: "A constant-utility index is the appropriate index for the main purposes for which the CPI is used" (Stigler et al. 1961, 52).[17] Furthermore, whereas neither the Meany and Thomas report nor the Mitchell Committee report discussed *voluntary* substitutions by consumers as a source of bias in the official Cost-of-Living Index, the Stigler Committee (52) wrote:

> Since consumers will substitute those goods whose prices rise less or fall more for those whose prices rise more or fall less—and within limits they can do this without reducing their levels of real consumption—the fixed-weight base CPI overstates rises in the cost of equivalent market baskets.

What had changed since the Mitchell review? Partly, professional economists had grown more accepting of economic theory as a guide to practical economics. The contrast between the careers of the two chairmen is illustrative (though both were associated with the National Bureau of Economic Research).

16. The Stigler Committee report can be hard to locate. The report and staff papers were published in January 1961 as Part I of the record of the JEC hearings (Stigler et al. 1961). Subsequent hearings (U.S. Congress, Joint Economic Committee 1961, Part II) elaborated on a number of aspects of the report's findings, including comments by BLS as well as others, and are an important part of the record of the Committee's work, its recommendations, and its impact. The report and staff papers (but not the hearings) were also published by the NBER (Price Statistics Review Committee, NBER 1961).

17. The Stigler Committee also used the term "welfare index," defined as an index that tracks the cost of maintaining a constant level of utility. Interestingly, it never used the term "Cost-of-Living Index," perhaps in response to the wartime confusion over that name. However, in response to a question at the JEC hearing, a member of the Stigler Committee referred to "what Commissioner Clague calls a Cost-of-Living Index."

Secondly, regarding the aggregation part of COLI theory, no proposal for estimating commodity substitution behavior existed in 1944, as the Mitchell Committee noted. Indeed, in a 1942 study on "The Empirical Derivation of Indifference Functions," W. Allen Wallis and Milton Friedman concluded, "We doubt that [the indifference function] has any material value for the organization of empirical data."[18]

But shortly after the Mitchell Committee report, an empirical estimate of bounds for substitution bias was published. In ten comparisons of Laspeyres and Paasche indexes, Ulmer (1946) found that all the differences were less than 1.5 percentage points. He argued that the maximum of the observed Laspeyres-Paasche differentials estimated the maximum possible value for the substitution bias of the Laspeyres index.[19]

Ulmer's bounds study was followed by a proposal for estimating consumer demand functions from which commodity substitution effects in a COLI could be derived—Klein and Rubin's (1947–1948) "linear expenditure system." By 1960, the linear expenditure system (also known as the "Stone-Geary" system) had been employed for empirical work on consumer demand (see Stone [1954]), though no actual COLI estimates existed. In contrast to the situation in 1945, by the time the Stigler Committee wrote its report estimates of substitution effects needed for a Konüs index seemed within reach.

Nevertheless, the Stigler Committee refrained from recommending that the BLS estimate a COLI econometrically. It recommended only that the BLS periodically estimate a Paasche index version of the CPI to gauge the potential size of the bias from substitution, and "possibly" update the weights more frequently to reduce the size of the bias.[20] It did not consider changing the Laspeyres index formula—though at the hearings on the report, Senator Douglas asked BLS to experiment with the use of the Fisher formula (U.S. Congress 1961, 566).

Commodity substitution was not the only question that the Stigler Committee viewed in the COLI framework. As we noted in section 1.2.1, the COLI framework influences many parts of the price index and many decisions that must be made in compiling it. Notable among the Stigler Committee's examples of COLI applications were the effects of changes in quality,

18. Quoted by BLS in a submission to the Joint Economic Committee Hearings; see U.S. Congress (1961, 580).

19. We now know that these are bounds for a common COLI only in the homothetic case; otherwise, they are bounds for two different COLI estimates, one using the initial indifference curve (Laspeyres perspective COLI), the other using the comparison period indifference curve (Paasche perspective COLI). It is thus possible in principle for the substitution bias in the Laspeyres index to be larger than the Paasche-Laspeyres difference, but all empirical COLI estimates lie between Paasche and Laspeyres indexes.

20. In 1960, the BLS was still using 1951 to 1952 weights and those were from an "interim" revision of the weights from the consumer expenditure survey of 1948. The Committee's recommendation seems timid, from today's perspective.

the treatment of consumer durables, and the effects of new products. For example, consumer preferences were the key to evaluating the effects of new products:

> If these new commodities are additional options open to the consumer, he will adopt them only if he prefers them (at their current prices) to goods previously available. (52)

To minimize the bias from new products, the Committee recommended their early introduction into the CPI with weights adjusted to reflect growth in their sales. The Committee thought that "typically" a successful new good enters with a high price and a low quantity sold, but then experiences a rapid decline in price and a rapid growth in quantity. By introducing the product early in its life cycle, tracing out the price decline that accompanied the rise in demand as the price fell, at least part of the welfare gain from the new good would be incorporated into the index.[21]

The problem of quality changes in existing goods had no simple, general solution. The Stigler Committee observed that if a quality increase was accompanied by a decline in price, the CPI should at least reflect the decline in price as a reduction in the cost of living.[22] In response to a question from Senator Douglas, Stigler suggested that instead of pricing the cost of a hospital room and physician's services, the CPI might take account of the more rapid recovery and shorter hospital stay required to treat a condition such as appendicitis (U.S. Congress 1961, 533).[23] Also, one of the report's staff papers, by Griliches (1961), investigated the use of hedonic functions for quality adjustment purposes, a method the Committee viewed as quite promising. Griliches' paper proved to be the most widely cited contribution of the report.

Last, for durable goods, the Stigler Committee noted that consumers' welfare depends on the flow of services from the durable, not on its value at the time of acquisition. Therefore, the cost of the use of the good is theoretically the correct concept for consumer price measurement. As a practical measure, the Stigler Committee recommended that BLS investigate the rental equivalence approach for measuring shelter costs for homeowners.

When asked by Senator Douglas what the Committee's proposal to move the CPI "toward" a Cost-of-Living Index entailed, Stigler gave the following list:

21. The objection that this stereotypical product cycle might not portray the pattern for all goods was raised at the time. On this, see Pakes (2003), who points to low "introductory" prices for new goods, followed by rising prices as knowledge spreads about them and demand increases. Very little empirical information exists on which pattern of price changes predominates for new varieties of goods.

22. The BLS procedures, at least in recent years, are consistent with this recommendation (see Triplett and Bosworth 2004) and were probably consistent in 1960.

23. Stigler noted a study, though he cited no source. Evidently, the Committee was apprised of a preliminary version of the work of Scitovsky (1964).

- "More objective" procedures for handling quality change.
- The same for new products.
- Treating durable goods as the consumption of the flow of services they provided.
- Substituting the average mortgage interest rate for the current rate in the owner-occupied housing measure.[24]
- "Perhaps" more frequent weight changes in the CPI.

Stigler was asked in the JEC hearings to make an estimate of the amount by which the CPI differed from a COLI, but, explaining that the Committee did not know enough to make such an estimate, he refused. In this, the Stigler Committee differed from the Mitchell Committee and the later Boskin Commission, both of which made guesstimates of CPI bias. Richard Ruggles, however, inserted a footnote into the report, which implied an annual bias estimate of 3 percentage points, but this was neither endorsed by the rest of the Committee, nor by Stigler himself.[25]

The initial reaction of the BLS to the Stigler Committee's COLI recommendation was quite negative, based partly on the lack of research showing how to estimate a COLI, and partly on doubts about the suitability of the COLI for the purposes of the CPI.[26] Commissioner Ewan Clague testified to a subcommittee of the Joint Economic Committee in 1961 that:

> There is one very important recommendation [in the report] with which the Bureau of Labor Statistics cannot agree, even with modifications. This is the recommendation that the Consumer Price Index be reoriented gradually toward a "welfare" or "constant utility" index. We would see some value in having a "true cost-of-living" or constant utility index if techniques can be developed for defining such an index, and then for compiling it objectively . . . We must emphasize, however, that this is a long-range goal that is now unattainable, may always be unattainable, and at best could be fully attained only after considerable further theoretical and statistical exploration. (U.S. Congress 1961, 560)

Later in the same hearings, BLS elaborated its position:

> Any partial movement toward a cost-of-living index, before the theoretical frame and operational structure is fully developed, could only lead to ambiguity and subjectivity. Based partly on price index principles and partly on cost-of-living index principles, such an index could not evoke the confidence of its users in its objectivity. (U.S. Congress 1961, 582)

24. A curious recommendation, since the Committee favored the rental equivalence measure of owner-occupied housing that BLS subsequently (in 1983) adopted. Stigler's testimony is also more guarded than the Committee report itself on increasing the frequency of weight updating.

25. Private communication by one of us with Stigler, who wrote that he had never quoted any estimate of the bias.

26. Some of the following discussion parallels Greenlees (2001).

Comments by others at the hearings on the Stigler Committee's report were also skeptical. Even Senator Douglas (a coinventor, after all, of the Cobb-Douglas model of producer substitution behavior, which Klein and Rubin [1947–1948] extended to the consumer case) expressed doubt about the feasibility of valuing quality changes by the amount that would hold utility constant. "I remember those lines of Browning," he remarked at the hearings: "'all, the world's coarse thumb and finger fail to plumb.' I always thought that there was always a large part of satisfaction that could not be plumbed by figures" (U.S. Congress 1961, 572).[27]

Despite BLS's initial opposition to the COLI concept, the tide changed in the 1960s. Part of the reason for the change in BLS's position was a change in its organization and staff. The Stigler Committee recommended that BLS set up a research unit within the BLS Office of Prices and Living Conditions. Funding for that was provided, and after some initial false starts a real research unit was established. It led to a more favorable view of the Stigler Committee's recommendations. As early as 1966, Commissioner Arthur Ross described to the Joint Economic Committee many restrictions on the applicability of economic theory for cost-of-living measurement but noted, "It is the only theory available, and if used with a proper understanding of its limitations does provide some guidance in the operation of a consumer price index" (Ross 1966).

Later statements about the CPI by BLS officials continued to combine references to the COLI measurement objective with caveats about the obstacles to achieving that objective. Statements that the CPI was not a COLI appeared in the *BLS Handbook of Methods* until its 1974 edition. However, a 1974 paper by BLS economist Robert Gillingham laid out the conceptual framework that was adopted for the 1978 revision to the CPI.[28] That paper, which focused on Pollak's concepts of partial and conditional cost-of-living subindexes, states that BLS "assumed that the primary purpose of the [CPI] is to approximate changes in the cost of living of consumers" (Gillingham 1974, 246). Moreover, using language that has been virtually unchanged since at least 1984, the 1997 *BLS Handbook of Methods* states:[29]

27. The Browning line is from "Rabbi Ben Ezra," in *Dramatis Personae.* As set by the Government Printing Office, a crucial comma is omitted, corrected in the above. The sense of the lines preceding the one quoted (stanza 24) is that things "that took the eye and had the price" are readily valued, "But all" (that is, the total of one's work in the eyes of God) is not. We thank Pimone Triplett for providing this reference.

28. Gillingham was a member of the CPI revision staff, and also of the Research Division. The manager of the 1978 CPI revision was John Layng, who became head of the BLS Office of Prices and Living Conditions. Administratively, then, the line officials in charge were closely identified with accepting the COLI framework. Indeed, little or no opposition to this framework was heard inside the BLS from around the mid-1970s until the mid-1990s (see the subsequent section on the Boskin Commission). See also Greenlees (2001) for a similar account.

29. Bureau of Labor Statistics (1997b, 170).

> A unifying 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 [ . . . ] However, the concept is difficult to implement operationally because it holds the standard of living constant, and the living standard must be estimated in some way.
>
> The CPI uses a fixed market basket to hold the base-period living standard constant [ . . . ] The CPI provides an approximation to a COL index as a measure of consumption costs.

Thus, despite its initial skeptical reaction to the Stigler Committee's recommendation on the COLI, the BLS position eventually became (a) the COLI objective provides the framework for the CPI, but (b) the CPI cannot be called a COLI because of limitations of scope, failure to reflect all consumer substitution, and other problems.

The BLS implemented many recommendations of the Stigler Committee. Most notably, it instituted a system of probability sampling and, after a long public battle, it changed the index for owner-occupied housing to a rental equivalence measure.

Yet when the subsequent Boskin Commission was appointed twenty-five years later, BLS had still not implemented other Stigler Committee recommendations aimed at bringing the CPI into closer alignment with a COLI, or enabling the CPI to be compared to a COLI retrospectively. For example, the CPI weights were not updated frequently, new goods that did not fit into the existing item structure of the CPI were not introduced early, and a retrospective index that provided direct evidence on substitution bias in the CPI was not published until 1993. In addition, use of hedonic indexes was limited; for example, as late as 1991, when BLS began working on price indexes for computer equipment, its work plan specified that hedonic methods would be used only if all other methods failed. Nevertheless, considerable research on hedonic indexes was conducted within BLS starting in the late 1960s, which continues to this day. (Triplett [1990] reviews the early BLS hedonic research, and Fixler et al. [1999] review the research in the 1990s on the use of hedonics in the CPI.)

### 1.2.4 BLS Empirical Research on Estimating Substitution Bias

The BLS objected to the Stigler Committee's COLI recommendation in part because little research existed on estimating a COLI. Even on the relatively tractable problem of substitution effects in the estimation of a COLI, the Stigler Committee's report contained little beyond calculating a Paasche-type index to obtain bounds. None of the report's twelve staff papers estimated substitution bias or suggested how to do it, and Klein and Rubin (1947–1948) was not in the references. For the BLS staff, it must have seemed as if they were being told to do the impossible.

Subsequently, however, BLS undertook an extensive program of theoreti-

cal and empirical work on substitution bias under the auspices of the price research unit that the Stigler Committee had recommended. This unit became fully operational in 1968 under the direction of Joel Popkin. Part of its budget was used to fund academic visitors such as Robert Pollak. Pollak's papers for BLS (which form the chapters in Pollak [1989]) resulted in the major advances in cost of living index theory of the 1970s, along with Erwin Diewert's (1976) paper, and the work of Franklin Fisher and Karl Shell (1972).

*Substitution Bias Estimates*

For empirical work, BLS research strategy followed the Klein-Rubin (1948) lead. Improved computer capability finally made estimates of the Klein-Rubin system practical, one of the earliest estimates being Goldberger and Gameletsos (1970).

The BLS developed or adapted improved specifications for systems of consumer demand functions that were less restrictive than the Klein-Rubin system (Brown and Heien 1972; Heien 1973; Christensen and Manser 1976; Braithwait 1980), and that could be used to explore the sensitivity of estimated COLIs to the demand system specification. The BLS empirical estimates of the substitution bias using systems of demand equations include: Christensen and Manser (1976), Manser (1975, 1976)—who studied detailed food categories, and Braithwait (1980), who treated food as an aggregate but included fifty-two other commodity groups, using data from the National Accounts. Taken together, these studies covered consumers' budgets at approximately a sixty-three commodity level of detail. This is a far greater level of detail than was found in previous research: BLS researchers estimated substitution among categories such as "beef, poultry, pork and fish," whereas the previous studies used aggregates like durables, nondurables, and services. Of course, even sixty-three commodity groupings may be too coarse to capture some important substitution behavior, and the use of aggregate data ignores the problem of aggregating over households, which amounts to assuming away one of the major problems with COLI theory.

Nevertheless, the early BLS studies produced two remarkable results. First, substitution bias, at around 0.1 to 0.2 points per year, was discovered to be much smaller than most economists had expected; guesstimates of upwards of 1 percentage point per year, or even more, prevailed in textbooks and informal discussions among economists. Substitution among CPI categories was indeed substantial, as was expected (Braithwait's unpublished BLS working paper has the most exhaustive elasticity estimates), but it did not result in the expected large bias in the index.[30]

Second, BLS researchers found that econometric estimates of substitu-

30. These studies did not consider substitution within CPI basic components; they estimated bias at the CPI level at which weights were held fixed in the Laspeyres formula, or actually, at a somewhat higher level. The question of bias within components emerged later on—see section 1.3.

tion bias in the price index were remarkably robust across different specifications of consumer demand systems. This finding of insensitivity of substitution bias estimates to utility function specifications was surprising because estimates of demand elasticities *are* sensitive to demand specifications. It is true that the roughly eight to ten utility specifications employed in the BLS studies covered only a limited range of possibilities (the studies employed all of the major demand systems that existed at the time, with one major exception that was developed too late for inclusion in the research design).[31] But their results are consistent enough to conclude that substitution bias estimates are empirically robust to model specification choices, contrary to the presumption made in some subsequent discussions of substitution bias.

In the middle of the BLS research project on estimating substitution bias, another development fundamentally changed researchers' perspectives. Diewert (1976), who drew on Byushgens (1925) and Konüs and Byushgens (1926), showed that one could obtain a good approximation to the substitution bias using "superlative" index numbers. Specifically, superlative indexes provide exact COLIs for "flexible" (homothetic) indirect utility functions that have enough free parameters to provide a second-order approximation for an arbitrary twice-differentiable indirect utility function (the "true" function). The approximating function is thus homothetic, but the function being approximated need not be. Superlative indexes include the Fisher index and the Törnqvist index.

One of the flexible functions considered by Diewert, the homothetic translog indirect utility function of Christensen, Jorgenson, and Lau (1975), had already been estimated by Christensen and Manser (1976). Diewert showed that the homothetic translog estimate could be expected to be close to the COLI estimates from any utility function, and Christensen and Manser showed that this was true, empirically, for the ones they estimated. Further experimentation with alternative consumer demand systems seemed unnecessary. Indeed, econometric estimation of a demand system was itself unnecessary, because the COLI could be estimated directly using one of the superlative indexes.

Manser and McDonald (1988) used Diewert's superlative indexes for point estimates of the substitution bias in a Laspeyres index of consumer prices. They supplemented the point estimates with nonparametric bounds derived from revealed preference theory by Afriat (1967) and exposited and advanced by Diewert (1973) and by Varian (1982, 1983). For the period that they studied, the Afriat-Varian method bounds showed that the approximation was very good. The substitution bias estimate averaged 0.18 percent per year.[32]

31. The exception was the "Almost Ideal Demand System." See Deaton and Muellbauer (1980).

32. Blow and Crawford (2001) used an extended version of the procedures in Manser and McDonald to estimate substitution bias in the U.K. index. Also, Balk (1990) estimated the substitution error in the Dutch index.

Finally, Aizcorbe and Jackman (1993) estimated substitution bias using superlative index numbers at a greater level of commodity detail than had been done previously, and unlike earlier studies their database *was* from the CPI. Their estimate, at roughly 0.15 index points per year, was consistent with earlier research. Because it was done on a larger number of commodities, it supplemented the earlier research, which mostly was conducted at a higher level of commodity detail (sixty-three commodities for the combined demand system results of Braithwait and of Manser, and roughly twice that number for Manser and McDonald).

Recently, BLS has begun to publish a "Chained CPI" that uses the Törnqvist index. As Boskin (2005) and Gordon (2006) note, this index differs from the official Laspeyres CPI by more than the 0.1 to 0.2 estimates of earlier studies. (The initial year difference was 0.8 percent, but this has come down to 0.3 percent per year in recent years.) The bias discussed by the Stigler Commission from failure to bring in new products promptly is part of the explanation for the unexpectedly large divergence between the Chained CPI and the ordinary CPI: Shoemaker (2004) traces some of it to higher weights for cell phone service in the Chained CPI. Cage, Greenlees, and Jackman (2003) also point to larger changes in relative prices in the period covered by the Chained CPI than some in earlier years, giving as specific examples rising natural gas prices and falling computer prices. The kind of behavioral response that would convert these relative price changes into significant substitution bias in a Laspeyres index seems more plausible in the case of computers than in the case of the inelastically-demanded natural gas.

*The Empirical Effects of Assuming Homotheticity on COLI Estimates*

Homotheticity is the bête noire of demand analysis. Homotheticity is the condition that Engle curves are straight lines that pass through the origin, so that for any set of relative prices households at all income levels consume commodities in the same proportions. It is well-known that empirical demand curves that are derived from homothetic utility functions are not only unrealistic, but can lead to biased estimates of demand elasticities.

From what is known from the analysis of consumer behavior, use of a homothetic function to estimate a COLI is on its face suspect. The CNSTAT report, discussed following, emphasizes this point. However, the question is not the validity of homotheticity as a specification of consumer behavior, nor is it the impact of maintaining homotheticity on estimates of demand elasticities; the question is the empirical impact of maintaining homotheticity on the size of the estimated substitution bias.

In a notable finding, BLS researchers discovered that an assumption of homotheticity, though a poor specification of consumer demand behavior, does not much influence econometric estimates of substitution bias. Christensen and Manser's (1976, 434–35) estimates of "branch" COLI's for meat are an example of this. They estimated five different nonhomothetic utility functions, which gave "branch" COLI estimates for 1971 (using 1958 prefer-

ences as the base and 1958 = 1.000) that ranged from 1.228 and 1.231; their three homothetic functions produced estimates between 1.228 and 1.233. Thus, over the thirteen-year period 1958 to 1971 as a whole, the maximum difference among all the estimates was only 0.005 index points, and the mean difference between homothetic and nonhomothetic estimates was nil. Using the whole period covered by their data (1947 to 1971), the results were similar: with 1947 = 1.000, the indexes for nonhomothetic forms ranged from 1.437 to 1.440 in 1971, homothetic ones from 1.441 to 1.444. In this case, the homothetic estimates are outside the range of the nonhomothetic ones, but the difference in average annual rates of change of the homothetic and nonhomothetic indexes is again exceedingly small. Yet Christensen and Manser's estimated demand elasticities were sensitive to whether a homothetic system was estimated, in line with previous research.

How could it be true that demand estimates are sensitive to maintaining homotheticity, yet COLI estimates are not? Consider figure 1.2, which shows a nonhomothetic indifference map. For simplicity in drawing the diagram (*only*) the true (nonhomothetic) Engel curve is drawn linear (labeled "A"), and a (counterfactual) homothetic Engel curve is labeled "B."

It is well-known that demand has both income and substitution effects; as figure 1.2 suggests, both would be misestimated if homotheticity were imposed on the data. But for estimating the substitution bias for a COLI, only the substitution term is relevant, and only around the initial (base) period's tangency and the comparison period's tangency. We may thus ask: how much error will be introduced from incorrectly maintaining homotheticity, in effect estimating the substitution term around the tangencies represented by BB, instead of along AA?

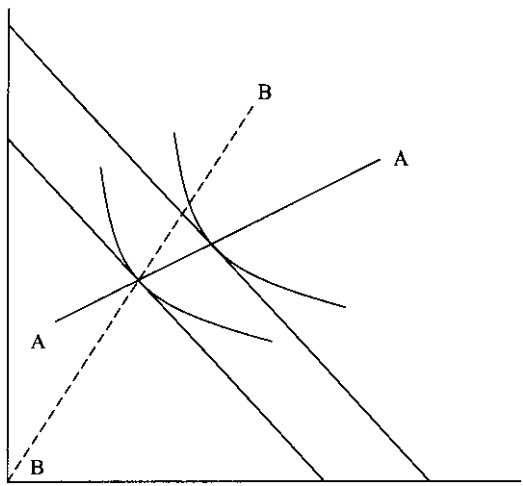


**Fig. 1.2 A change in income with nonhomothetic indifference curves**

The curvature of the implied indifference curves along BB and along AA may well differ. However, for small differences in income levels and barring extreme curvatures of the indifference curves, the curvatures along AA and BB will differ only slightly, so the *estimated* substitution biases can be similar. The Christensen-Manser substitution bias estimates cited in the previous paragraph suggest that something like figure 1.2 represents the data. Estimates of the substitution bias when homotheticity is imposed do indeed differ from substitution bias estimates when it is not, but not by much. Manser-McDonald's bounds on the approximating error of the superlative index indicate the same thing: their assumption of homotheticity does not keep superlative indexes from providing a close bound on the substitution bias, just as the use of homothetic demand systems does not create much of an error in econometric COLI estimates.

The CNSTAT Panel argued that Diewert's superlative index method implied homotheticity, which is an unsupportable assumption about consumer preferences.[33] The panel made a valid, albeit well-known, theoretical point. Yet it overlooked the empirical literature that would have provided perspective on the empirical importance of their criticism—the empirical estimates that exist, cited in the previous paragraphs, show that the effect of imposing homotheticity in COLI estimation is negligible. Further discussion of this question follows in the section on the panel's report.

### 1.2.5 The Boskin Commission: Further Developments in the Cost-of-Living Index Approach

At the 1993 Allied Social Science Associations (ASSA) meetings in Anaheim, Reinsdorf presented a paper documenting upward bias in CPI component indexes for food and gasoline (published in revised form as Reinsdorf [1998]). Partly as a result of the attention drawn to Reinsdorf's results by Erwin Diewert (who, as the paper's discussant, called it "the measurement paper of the decade") and by Robert Gordon, CPI bias became a topic of much discussion. After a remark on upward bias in the CPI in testimony by Federal Reserve Chairman Alan Greenspan (Berry 1995), the Senate Finance Committee appointed an "Advisory Commission to Study the Consumer Price Index." The Commission became known as the Boskin Commission, after its chair, Michael Boskin. Like earlier review panels, the Boskin Commission had one carryover from the Stigler Committee: Zvi Griliches (who had contributed a staff paper on hedonic indexes).

Like the Stigler Committee, the Boskin Commission recommended that BLS adopt the Cost-of-Living Index as the measurement concept for the CPI (Boskin et al. 1995). Indeed, the Boskin Commission took the appro-

33. Actually Diewert's (1976) results included a demonstration that the Törnqvist index is exact for a nonhomothetic translog model if the COLI is evaluated at an intermediate level of utility. This is a less satisfactory measurement concept than either a COLI evaluated at the initial level of utility or a COLI evaluated at the final level of utility.

priateness of the COLI objective for the CPI as almost self-evident, presenting no alternatives and laying out no specifics as to how the COLI should be defined. The COLI factors identified by the Stigler Committee, including substitution induced by changes in relative prices, new products, and changes in quality of existing products, figured prominently in the Boskin Commission's report.

As we noted in an earlier section, BLS had initially been disinclined to accept the COLI recommendation of the Stigler Committee. Even in the run up to the naming of the Boskin Commission, statements from BLS that the CPI is not a COLI had occasionally seemed to disclaim the COLI even as an objective for the CPI.[34] In light of BLS's long-standing acceptance of the COLI concept, these BLS statements were surprising, and they contributed to a perception that the Boskin Commission's COLI recommendation was one that BLS professional staff opposed. In turn, this COLI misunderstanding (if that is what it was) inflamed contentions over the rest of the Commission's findings, some of which were indeed flawed. All manner of political speculations clouded the debate over the Commission's report, ranging from the Senate Committee's motivation in selecting the panel members (reducing the rate of increase in the CPI would reduce expenditure growth on government indexed programs) to whether the independence of a statistical agency was threatened.

Along with its call to embrace the COLI as the measurement objective, the Boskin Commission urged BLS to use superlative or similar index number formulas to approximate a COLI, and here its recommendation constituted a genuine change. Specifically, the Boskin Commission urged BLS to use a superlative index number formula for a retrospective annual index, and to use a geometric mean formula (which is not superlative, but would not require unavailable data on current expenditure patterns) for the monthly CPI. Research results that had become available since the time of the Stigler Committee made these recommendations possible, as noted in section 1.2.4.

The Boskin Commission discussed sources of CPI-COLI difference beyond those that Stigler Committee had considered. One of these was "lower-level substitution bias." The estimates described in section 1.2.4 pertain to what the Boskin Commission called "upper level substitution bias," the bias that occurs *among* the basic components of the index, the levels at which the weights are held fixed (roughly 200 commodity groups in the index structure that was in place at the time). Lower level substitution bias—to be discussed more fully in section 1.3 of this chapter—occurs among the detailed items

34. Baker (1998, 131) interprets the 1995 BLS report to the House Budget Committee as taking this position, and others formed the same interpretation. Moreover, based on personal conversations one of us had with Boskin Commission members at the time, we believe that they also interpreted BLS statements as indicating that the BLS was opposed to the COLI concept.

and varieties that are aggregated to construct the 207 item group indexes in the CPI (that is, *within* the basic components). The problem of aggregating price quotations into basic components had been discussed previously (Carruthers, Sellwood, and Ward 1960; Szulc 1983). However, its importance was little appreciated until the empirical results on the topic that appeared in the early 1990s showed the surprising magnitude of these aggregation effects (Schultz 1994; Reinsdorf 1993; Moulton 1993; Reinsdorf and Moulton 1995; Diewert 1995; Reinsdorf 1998).

The Boskin commission also called attention to the related problem of the unmeasured reduction in the cost of living from substitution to outlets offering lower prices, such as Wal-Mart. We discuss this problem in section 1.3 because it affects the construction of the detailed component indexes.

Like the Mitchell Committee, but unlike the Stigler Committee, the Boskin Commission estimated biases in the CPI. Their 0.15 percentage points per year from upper level substitution was based on the empirical estimates discussed in section 1.2.4, but likely more on the latest of those estimates (Aizcorbe and Jackman 1993) than the earlier ones. The Commission also estimated 0.25 percentage points per year from lower level substitution (this was based on Reinsdorf and Moulton [1995]; Reinsdorf [1998]), 0.6 points from new products and quality change in existing products, and 0.1 points from outlet substitution. The Boskin Commission acknowledged that in some cases the available evidence to make these estimates was not strong, and some reviewers chided the Commission for its lack of objectiveness in indulging in guestimates. Some of the quality change estimates were marred by faulty understanding.[35]

The Boskin Commission also revisited a question that had received little attention since the discussion in the Mitchell Committee's 1944 report: to what extent should developments beyond market price changes that affect the cost of living be reflected in the CPI? Here the Boskin Commission raised some worthwhile questions concerning inconsistencies that can arise from consideration of price effects in isolation from nonprice effects. For example, is it sensible to show a decline in the COLI when the cost of treating AIDS drops because of medical advances if no increase in the COLI was shown for the appearance of this previously unknown disease?

35. For example, its guestimate of upward bias in rental housing (which was based on introspective estimates about the extent that quality of housing had improved) ignored results in Randolph (1988) that showed that the BLS linking procedure for bringing improved housing units into the sample imparted a downward—not upward—bias. An additional downward bias, caused by censored reporting of rents during vacancies, was discovered by Crone, Nakamura, and Voith (2004). These downward biases were replicated by Gordon and van Goethem (2003). Randolph's result was replicated by Gordon and van Goethem (2007) over a longer historical period, and Gordon (2006) revised the Commission's guestimate on housing. Additional discussion of these matters (which are beyond the scope of the present chapter) are in a 2006 symposium on the Boskin Commission report. See, in addition to Gordon (2006), Berndt (2006), Baily (2006), Greenlees (2006), and Triplett (2006).

However, in its discussion of the effects of broader changes in the environment on the cost of living, the Commission failed to make the distinction between the "conditional" COLI and the "unconditional" COLI. Gillingham (1974) and Pollak (1989, chapter 2) contended that the CPI can be interpreted as an approximation to a conditional COLI, for which environmental effects—though appropriate for some broader measures—are out of scope. By discussing environmental effects without providing the appropriate context, the Boskin Commission report implicitly suggested that they might be in scope, creating confusion.[36] This contrasts with the Mitchell Committee's consideration of the need to restrict the domain of the index in its response to the Meany-Thomas report, as well as with the extensive discussion of the domain question in the subsequent CNSTAT report.

The BLS agreed with some, but not all, of the Boskin Commission's recommendations and findings. Regarding its recommendation that the CPI establish a COLI objective, the BLS paper, "Measurement Issues in the Consumer Price Index," indicated:

> The BLS has for many years used the concept of the cost-of-living index as a framework for making decisions about the CPI and accepts the COLI as the measurement objective for the index . . . The cost-of-living index approximated by the CPI is a subindex of the all-encompassing cost-of-living concept. (Bureau of Labor Statistics [1997a, 4]; see also Abraham, Greenlees, and Moulton [1998, 27].)

The Boskin Commission report has been very influential, not just in the U.S., but also worldwide (Diewert 1998). But the basis for its influence, ironically, was its estimates of CPI bias, not its endorsement of the concept of COLI. Indeed, many countries reject the COLI framework, and a number of them initially contended that the Boskin Commission's bias estimates did not apply to their CPI's because their indexes were not intended as approximations to a COLI. Subsequently, however, the discussions the Commission generated resulted in increased efforts to improve CPI measurements, worldwide (Ducharme [1997]; Abraham [2003]; Hausman [2003]; Triplett [2006] and the items cited there).

### 1.2.6 The CNSTAT Panel: Second Thoughts about the Cost-of-Living Index

The Boskin Commission had limited resources and it operated in a politically charged environment. Critics of the Commission's report argued that its treatment of CPI biases lacked balance, and that the membership of the Commission excluded alternative points of view. Robert Pollak—echoing a suggestion made earlier by Boskin Commission member Zvi Griliches—therefore recommended that the technical issues related to the

36. However, two members of the Boskin Commission stated elsewhere that these nonprice effects do not belong in the CPI—see Gordon and Griliches (1997, 87).

CPI be examined by a committee of technical experts under the auspices of an organization such as the National Academy of Sciences (Pollak 1998, 76). Consistent with this proposal, BLS asked the Committee on National Statistics of the National Academy of Sciences to convene a panel to investigate "issues in the development of cost-of-living indexes" (Schultze and Mackie 2002, 17).

In contrast to all previous review panels, the CNSTAT Panel was not composed primarily of scholars with expertise on the CPI: despite the technical nature of the question that the panel was charged with investigating, the goal seemed to be to make sure that a broad range of perspectives was represented. The panel included economists whose primary expertise was in other areas of economics, as well as representatives of disciplines such as statistics, psychology, and marketing. Among the economists on the panel, some had not participated in the debate over the issues discussed by the Boskin Commission, and some were known to differ with the views from the Boskin Commission. The only carryover was Griliches, who had also provided the bridge from Stigler to Boskin, but he died before the panel got very far into its work.[37]

If the goal was to obtain a diversity of perspectives, that is certainly what BLS got. In particular, the view that the COLI is the appropriate measurement concept for the CPI, which had broad support from most economists for half a century,[38] proved so contentious that the panel was unable to reach a consensus on this fundamental issue.

*The COLI versus the COGI*

Rather than measuring the cost of maintaining a constant standard of living, some members of the CNSTAT Panel believed that the conceptual objective of the CPI ought to be measurement of the change in the cost of purchasing a fixed basket of goods and services, or what the panel called a "cost of goods index" or COGI (Schultze and Mackie 2002, 15). In principle, the fixed basket could be based on the initial (or "reference") period consumption pattern, the final (or "comparison") period's consumption pattern, or even some point in between (Schultze and Mackie 2002, 16). However, the report's subsequent discussion of the COGI often suggested its

37. The other members of the CNSTAT panel were its chairman Charles Schultze, Ernst Berndt, Angus Deaton, Erwin Diewert, Claudia Goldin, Christopher Jencks, Albert Madansky, Van Doorn Ooms, Robert Pollak, Richard Schmalensee, Nobert Schwartz, and Kirk Wolter.

38. For example, panel member Robert Pollak (1989, vii) wrote that at BLS in 1968 to 1969 he approached problems in the CPI "with the conviction that a well-developed theory of the cost-of-living index could provide practical solutions." In addition, substitution bias in the CPI had long been a staple of textbooks for principles of economics classes, which is again evidence of a wide consensus that the COLI was the correct concept for the CPI. On the other hand, the economists who were named to the CNSTAT panel were divided on the applicability of the COLI.

equivalence to a Laspeyres index, which uses the initial period's expenditure pattern (see, for example, the report's page 112).

The proponents of the COGI supported their view with three contentions. First, they argued that the COLI concept was unsuitable for the CPI. Second, they pointed to problems in the theory of the COLI. Third, they cited the arbitrariness of any definition for the domain of the conditional COLI, which they thought implied that the concept was unworkable as a guide to index procedures.

*COLI Unsuitable for the CPI*

The COGI proponents who found the COLI concept unsuitable for the CPI accepted that a fixed-basket index overcompensates for price increases because of substitution bias. However, they believed that the CPI should measure the "price level" rather than the level of compensation needed to hold welfare constant (Schultze and Mackie 2002, 58). The COGI advocates apparently viewed the price index in a way that coincides with the views of Fisher—some price level exists, like the level of water in a lake, and the index is a device to find it.

Mitchell's (1921) comment that the notion of the price level is inherently ill-defined (see the appendix) still seems apt in considering this aspect of the CNSTAT 2002 Panel's report. The charge that the domain of the "price level" is undefined might be answered by COGI advocates by saying that Mitchell's critique has less force in the case of consumer prices because the content of consumer expenditures defines the domain. On the other hand, the CNSTAT panel thought that the domain should be defined on the services of durable goods, not on the purchases of durables themselves (thus agreeing with the COLI point of view). In this case, the price index domain is not defined by the observable content of consumer expenditures, but rather by recourse to the theory of consumption.

Leaving aside the flow of services problem, the concept of a "price level" suffers from a more fundamental problem. Given the domain, in principle one could observe all the transactions that take place at a given date and compute the average level of their prices. Different commodities would be weighted, in this system, by their transactions. In this sense, a price level might exist, for a point in time.

Taking another date, the computation could be repeated. The transactions on the second date would differ from those on the first, and again, a price level exists in this sense for the second period. Taking the ratio of the two price levels amounts to a unit value index, whose weights differ between the numerator and denominator. The *average* price of a transaction on each date depends arbitrarily on such factors as the lumpiness of transactions, and the commodity and quality mix. In some sense, this change is a change in the price level (in the sense that, for example, it influences the demand for money). But it is hardly what one wants from a price index.

Even though price indexes are often said to measure the change in the price level, this is just careless or imprecise language, for they do not actually do that at all. As price indexes are actually computed, they are based on samples chosen to measure price change, not the price level. Consider all the apparatus that is put in place in the CPI to hold constant the retail outlet, product characteristics, and so forth. This apparatus is intended to assure that price change is measured without contamination, not that the price level is measured accurately. Though Fisher spoke of "the price level," and of a price index as a device for measuring it (comparable to measuring the level of water in a lake), he also discussed the need to make the price index independent of the units of measurement (his "commensurability axiom"), which ruled out the averaging of prices and the use of unit value indexes.[39]

Rather than addressing specifically what the seemingly vague price level notion means, the report provides only an example of what it does not mean—the treatment of new goods. Recall that the Stigler Committee included the welfare gains associated with new goods in the concept that the CPI should try to measure, and hence advised bringing new goods in early to minimize the bias. Hausman (1997) proposed a method to include all the welfare gains in the CPI by extrapolating the demand curve for a new good out to a Hicksian virtual price (Hicks 1940), at which the good's sales would fall to zero (the virtual price provides the price for the period before the new product was introduced). Some members of the CNSTAT Panel argued against the Hicks-Hausman technique because they contended that gains from the invention or introduction of new goods were not part of the change in the price level (Schultze and Mackie 2002, 160–1). Without a precise definition of the price level, whether their contention is correct is unclear.

*Problems with COLI*

The second argument for the COGI approach was a negative case against the COLI, built from problems with the concept of the COLI, rather than a positive case for the COGI. The "not COLI" position pointed to the need for unrealistic assumptions in some forms of the theory of the COLI (the panel dwelt especially on homotheticity, constant tastes, constant environmental conditions, and the difficulty of aggregating over individual consumer's behavioral functions). Some members also contended that the standard of living (held constant in the COLI) cannot be defined, which provides a counterpart to the assertion of COGI detractors that the price level is not definable.

39. In his (1911) discussion of the equation of exchange $MV = PT$, he defined the quantity units in $T$ as "a dollar's worth in the base period," and found that this implied that $P$ was a Paasche price index, not a ratio of averages.

The COLI for the base-period standard of living is bounded from above by the Laspeyres index and the COLI for the comparison period standard of living is bounded from below by the Paasche index, but these are different COLI indexes if preferences are nonhomothetic. Similarly, the conventional assumption that the COLI of a "representative consumer" who has the aggregate demand patterns summarizes (or "averages") the COLIs of individual households is not justified without unrealistic assumptions about the nature of preferences. Therefore, estimation of an aggregate COLI involves difficult distributional issues, as Pollak (1989, chapters 6 and 7) showed. On the other hand, similar taste and distributional issues affect the COGI (e.g., the Laspeyres COGI will differ from the Paasche COGI), but this seemingly was not regarded as a problem.

*Domain of the Conditional COLI*

Within the context of the COLI approach, the panel agreed that the appropriate version of the COLI for the CPI was a *conditional* COLI that covers private market goods and services and holds the broader environmental factors constant (Schultze and Mackie 2002, 73). The immediate intellectual roots of the concept of the conditional COLI are found in Pollak (1989, chapter 2), though the domain question can also be linked historically to the position that the Mitchell Committee took nearly sixty years earlier, when it recommended that BLS change the name of its "Cost-of-Living Index" to avoid confusion about the breadth of effects that the index covered.

Some members of the CNSTAT panel viewed the need to restrict the domain of the COLI as an argument for the COGI approach because in deciding on how to condition the COLI, arbitrariness and inconsistency can be hard to avoid. One must condition the COLI on some new technologies, the panel asserts, but not on others. For example, the CNSTAT Panel favored diagnostic-based measures of medical care rather than input-based measures whenever possible (188), which implies that the CPI should decline when a new pharmaceutical makes treatment of a diagnosis cheaper. Nevertheless, most of the panelists did not think that the invention of Viagra should cause a decline in the CPI (Schultze and Mackie 2002, 67).[40]

Of course, the COGI must also have a domain. Choosing the domain of the COGI appears to present the same kind of problem as choosing the domain of the conditional COLI. As Schultze (2003, 11) put it in summarizing the debate within the panel: "According to the supporters of the cost-of-living concept, the fact that the basis for . . . domain decisions cannot be provided from within the general theory underlying that concept is not a

40. This appears to be an arbitrary judgment about what medical conditions are appropriately included in the treatments priced for the CPI. This kind of arbitrariness could be considered a powerful argument against the COGI approach, because it lacks any theoretical framework for determining what goods are to be included and how they are to be measured.

reason to preclude using the conditional cost-of-living index as the framework for the design and construction of the Consumer Price Index."

### 1.2.7 Assessment of the Debate over the Conceptual Basis for the CPI

At one level, the COLI-COGI debate involves competing charges of "immeasurable" or "undefined" hurled like spears at the opponents' concept. Detractors of COGI point out that the price level is undefined, as is the COGI domain. The COGI advocates answer that the key concepts of the utility level and the standard of living in the COLI approach are undefined or immeasurable. Observers who are users of the CPI but belong to neither camp might view the competing charges as a draw, and decide that what matters is staying out of spear range.

A more specific criticism of the COLI in the CNSTAT report is the one that has been current in statistical agency circles seemingly forever: the theory rests on questionable behavioral assumptions that do not describe how consumers actually behave and (at a somewhat more sophisticated level) the topics on which the theory is inchoate dominate the topics on which firm conclusions can be drawn. Deaton (1998, 37–38 and 42) summarizes this criticism with his usual eloquence:

> That the Bureau of Labor Statistics should establish a cost-of-living index as its objective in measuring consumer prices, taken by them [the Boskin Commission] as essentially obvious, is a contentious proposition that requires serious argument. In fact, it is unclear that a quality-corrected cost-of-living index in a world with many heterogeneous agents is an operational concept.
>
> We know rather little about whether consumers maximize utility at all, let alone whether they do so instantaneously or take time to adapt to price changes. We do know that there are many consumers, not one, and that, even if each behaves as we like to suppose, we cannot represent their behavior or their welfare by that of a single representative agent.

We emphasize that we do not dispute Deaton's critique of the COLI, nor the elaboration of parallel matters in the CNSTAT report's sections ("The Theory of Price Indexes and Its Critics" and "Two Perspectives," pages 43–72). These shortcomings of COLI theory have long been recognized. (Indeed, Jorgenson [1990] and Jorgenson and Slesnick [1983] proposed a generalization of COLI theory that avoids the use of the single representative agent to represent a society of heterogeneous consumers.) What separates adherents of the COLI view (where we place ourselves) and those of the COGI *is not so much disagreement about the shortcomings or ambiguities of the theory on which the COLI is based as it is disagreement about the shortcomings of the alternatives to COLI, such as COGI.*

A conceptual framework for the CPI must address a broader range of questions than just the scheme for aggregating the basic component indexes. Passages in the CNSTAT Panel report suggest agreement: "Quality adjust-

ment is possibly the area in which the COLI has the greatest advantage over the COGI approach" (Schultze and Mackie 2002, 62).[41] Overall, however, when the panel discusses these "other" questions, the overwhelming weight of its discussion falls on problems, or supposed problems, with the COLI. The COGI "wins" this debate by default.[42]

In constructing the CPI, hundreds of decisions must be made, and many of them involve the question, "What do we want to measure?" (Triplett 2001, 315). An explicit underlying conceptual motivation is needed to provide a unifying framework for consistent decision making in index design and to provide a clear interpretation of the index. A conditional COLI has long served this purpose in the U.S. CPI, and satisfactory alternatives have not been developed.

A fixed basket index as the objective leaves the question of the underlying conceptual motivation largely inchoate. The motivations offered for the COGI—to track "the prices of the things that people buy" or "the price level"—provide little analytic insight into questions such as what items belong in the basket,[43] what price concept to use in cases where the definition of the price is ambiguous, what to do about quality change, how to treat voluntary or involuntary substitution, how to treat product introductions and disappearances, and so forth.

In contrast, the COLI approach provides guidance about how to handle problems such as what to do when items in the index disappear, or what to do when new goods or new quality levels appear, or (when an immediate conclusion is not obvious) it provides a vehicle for thinking about them in a consistent and coherent manner. It can also help to resolve methodological ambiguities and it can aid in the discovery of improvements by making the

41. On the other hand, speaking of the COLI framework, some members of the Panel expressed themselves as "concerned about the BLS adopting a conceptual framework that is not always well defined in the presence of quality change" (Schultze and Mackie 2002, 73). By default, this means adopting the COGI framework (which one can contend is even less adequate), though the Panel also expressed the opinion that with respect to quality change "the distinctions between the two approaches are blurred" (63).

42. One example of this fault is the Panel's discussion of "Using Indexes for Compensation," which is a subsection of its "Two Perspectives." The section contains an excellent discussion of how escalating a *portion* of income with a COLI does not leave the recipient at the same standard of living (as pointed out in Triplett [1983]; see also Triplett [2001]), but it concludes, "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." Because the section in which this conclusion is found is one that purports to be comparing the advantages of COLI and COGI, one would not be amiss in interpreting the Panel as saying that the COGI is immune from the fault, and for this reason should be preferred. The Panel missed the chance to say something meaningful about escalation practices, choosing instead to look for arguments against the COLI. In this, the Panel's report lacks balance.

43. The "domain" of the index. "Prices of things that people buy" provides no guidance: People buy apartment buildings and other investments. The theory of consumption guides, explicitly or implicitly, the domain of the CPI. Some proposals for including some kinds of investments in the CPI appeal to COLI theory for justification (for example, Goodhart [2001]).

index's shortcomings identifiable.[44] An underlying conceptual objective such as a COLI can provide an analytic framework to reason about such basic questions of what price concept should be used in situations where this is not obvious; for example, what to do when a two-part or multipart tariff replaces a simple price, or how to measure complex commodities like insurance.

Using insurance as an example, Beelen (2004, 6) discusses the price concept for insurance in the CPI. When a household purchases an insurance policy, is the service the absorption of risk by the insurance company? If so, the price concept is the risk-adjusted premium. Or is it the administration of the insurance pool on the behalf of the policyholders? In this case, the service is a management service by the insurance company, and the price is the price of the management service (usually measured by premiums minus claims). The European Harmonized Indexes of Consumer Prices (HICP), which operate under a "not COLI" concept, have adopted the latter.[45]

How does one resolve such questions? Though one might contend that COLI theory by itself does not fully answer them without more development, the COLI framework permits development of the theory, just as any problem in economics is typically addressed by developing the relevant theory. The alternative is an ambiguous framework where the decision itself *and the framework for discussing it* are both arbitrary.

Finally, as Blow and Crawford (2001, F359) point out, many uses of the CPI require an interpretation as an approximation to a COLI. As illustrated by the union complaints in the World War II era, users' complaints about the CPI have often reflected a desire for a Cost-of-Living Index, and political debates regarding indexation often refer to adjusting benefits for the "cost of living." The CNSTAT Panel believed that the concepts of the COLI are difficult for noneconomists to understand; newspaper discussions of the Boskin Commission report attest to the contrary—many of them were quite clear and perceptive about the COLI and its objectives, and how the COLI differed from the CPI in its treatment of substitution. In any event, most applied economic research on topics such as tax analysis or benefit-cost analysis also use the same of kind of theoretical measures as the COLI, so an estimate of a COLI is usable in a great many economic analyses with no conceptual inconsistency. Nothing comparable can be said about a "cost of goods" framework.

We also believe that the theoretical problems of the COLI are of less practical importance than is commonly supposed. The CNSTAT report makes much of the implausibility of the assumptions needed for a single

44. Greenlees (2001, 12–14) provides examples of design improvements in the CPI guided by the COLI concept.

45. The HICP strove for compatibility with national accounts, so its decision on insurance was influence by the 1993 System of National Accounts. On measuring insurance in national accounts and CPI, see chapter 6 in Triplett and Bosworth (2004).

representative agent to represent the welfare of a group of households. Yet as the CNSTAT 2002 report (51–52) also observes, these extreme theoretical assumptions are not needed for the aggregate Laspeyres index to be an upper bound for a social Cost-of-Living Index concept known as the "Scitovsky-Laspeyres" index. This index concept, introduced by Pollak (1980, 1981), tracks the cost of keeping every household at its initial standard of living, holding constant its tastes and endowment of environmental factors. A parallel result from Diewert (2001, 172–3) shows that the aggregate Paasche index is a lower bound for an analogous "Scitovsky-Paasche" index, which tracks the cost of keeping every household at its final standard of living, given its final endowment of tastes and environmental factors under similarly weak assumptions.

The empirical literature on estimated COLIs shows that some other criticisms of the COLI are theoretical niceties with less empirical importance. Chapter 2 of the CNSTAT report explains the theoretical significance of homotheticity to consumer demand analysis (not in doubt) and demonstrates that it *might* have an impact on empirical estimates of a COLI (again, not in doubt).

Yet surely the issue for the panel's report is, how much does homotheticity matter empirically for COLI estimates? The CNSTAT Panel made no attempt to distill the empirical literature on estimated COLIs—or any of the empirical evidence in economics—that had a bearing on its COLI-COGI deliberations.

We reviewed the existing empirical COLI research in section 1.2.4. We noted that estimated COLIs based on homothetic indirect utility functions differ trivially from COLIs computed from nonhomothetic indirect utility functions. Empirically, homotheticity matters far less than the CNSTAT panel report implied. One might contend that the existing empirical COLI research amounts to a small number of studies, using a limited number of demand specifications. Nevertheless, a professional review panel has an obligation to consider all the relevant research.

Another way to assess homotheticity is to evaluate the effects of changing the reference base. It is well-known that the value of a COLI is independent of the reference base (the base indifference curve) only with homotheticity. Christensen and Manser (1976, 434–7) found that changing the reference standard of living from the 1947 to the 1967 level (that is, moving it by twenty years) raised their *nonhomothetic* cost-of-living subindexes for meat over this period by about 2 percentage points and lowered their cost-of-living subindexes for produce by about 1.5 percentage points over the whole period—trivial effects given the large change in the reference utility level (see the additional discussion in section 1.2.4), and more trivial still for a normal measurement time horizon. The problem that so occupied the CNSTAT panel—the dependence of the COLI on the reference indifference curve—is correct in theory, but does not have much empirical significance.

Furthermore, in many studies Laspeyres indexes have been found to be above Paasche indexes by an amount that could be expected from substitution effects (the earliest estimate of this kind that we know is Ulmer [1946]). If the other kinds of effects emphasized by the panel had a major role, we would presumably observe many instances where the Paasche index lies above the Laspeyres index or where Paasche-Laspeyres index differences exhibit wide swings.[46]

These empirical results do not, of course, imply that violations of COLI assumptions can never cause distortions large enough to be a practical concern. For example, in a country that is experiencing a severe economic crisis, changes in the standard of living would presumably be large enough for nonhomotheticity to matter, and even in the United States this may occur over long intervals of time. Similarly, a substantial distortion from changing tastes is possible in a long-run index for a good that requires repeated quality adjustments. If each adjustment reflects the tastes of the time when it is made, the cumulative value of many adjustments may not be right for any single configuration of tastes.[47] In the special cases where the potential for important effects from violations of COLI assumptions is high, this possibility should be taken into account. But the solution is not to discard the entire COLI approach, especially when the potential shortcomings of the COGI alternative have not been adequately weighed.

## 1.3 Recommendations Concerning Detailed Component Indexes and Sampling

Sampling theory was not well developed when BLS first began to estimate a price index for consumers. Mitchell recommended judgmental sampling, which the BLS largely followed until 1978. Some improvements in sampling procedures were recommended by an ASA committee in 1933 to 1934 (Hogg 1934).

### 1.3.1 Wartime Committees: The Problem of Unrepresentative Samples of Varieties and Forced Substitution

Among the biases in the BLS Cost-of-Living Index discussed in the Meany-Thomas Report were three that occurred in constructing its detailed component indexes. One of these was caused by failure to collect prices on weekends, when sales were common. The Mitchell Committee conceded

46. The panel noted that "an assessment of the ability of a superlative index to approximate a measure of the [COLI] 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 relative prices" (Schultze and Mackie 2002). This is an empirical proposition that is readily explored (and has been).

47. This is a special application of the well-known principle that a chain index may not give the same measure after several links as an index that directly compares first and last periods.

that this bias was present because weekend sales had become less prevalent, forcing some consumers to pay higher prices. However, their guesstimate for the size of this bias was only half a percent over three years.

A second source of bias identified by Meany and Thomas arose from forced substitution to more expensive varieties or outlets. Inexpensive varieties tended to disappear or to be differentially affected by shortages, resulting in forced "trading up." The Mitchell Committee observed that this likely affected the poorest families more than it did the "average" family (wage earners and salaried clerical workers) tracked by BLS's index. In addition, the Committee noted that some shift to higher-priced outlets (from chain stores to higher-priced independents) could be expected because long hours of work, gasoline rationing, and reduced car ownership made visiting lower-priced outlets inconvenient (II-4). Finally, because of quality deterioration in product lines, consumers were also forced to buy more expensive varieties simply to keep quality constant.[48] After noting the difficulty in valuing quality deterioration ("who is to say how much the real price of shorts has gone up because they have ties rather than elastic sides?") the Mitchell Committee made an educated guess that the downward bias between January 1941 and December 1943 from forced trading up and quality deterioration was 1 to 3 percent for the food index, 4 to 5 percent for the clothing index, and 8 to 11 percent for the house furnishings index.[49]

The third problem identified by Meany and Thomas highlights the vulnerability of judgmental sampling to error or manipulation that can damage the credibility of detailed component indexes. Meany and Thomas argued that samples were unrepresentative and that the varieties (or varieties priced for the detailed component indexes) were causing a large downward bias in the Cost-of-Living Index.[50] Usually one or two tightly specified varieties ("items" in BLS terminology) were selected to represent a commodity category, facilitating use of "specification pricing" as had been recommended many years earlier in Mitchell's report on the Bureau of Labor's Wholesale Price Index. Since the same variety or quality level was priced in many outlets, city level averages for the price could be meaningfully calculated. Thus, the first step in calculating the CPI was to calculate the average price of, say, refrigerators or white bread, in each of the CPI cities, where the averages were always computed for matched samples of varieties and outlets. These city averages could be compared over time to form basic component indexes for commodities or narrowly defined groups of commodities. This

48. Meany and Thomas also alleged that unmeasured quality deterioration was an important cause of downward bias in the Cost-of-Living Index.

49. Note the interesting parallel with the Boskin Commission, which also prepared guestimates of bias on a component-by-component basis.

50. An unrepresentative geographic sample and an unrepresentative market basket of goods and services were also sources of bias alleged by Meany and Thomas. We do not discuss these because our focus is on problems in the construction of the basic component indexes.

simplicity came at a cost, however. The risk that the small, judgmentally selected samples of varieties would fail to represent their commodity group was high.[51]

The errors introduced by the sampling of varieties might, of course, average out to about zero, implying very little bias at the aggregate level. Meany and Thomas claimed, however, that the sampling errors in the component indexes were systematic, biasing the index downward. According to tables 3 and 4 in their report, 77.3 percent of the food index was based on varieties with price subsidies, but of thirty-seven varieties with rising price control ceilings, only eleven were in BLS's index. To illustrate the problem, Meany and Thomas discussed many cases of the inclusion in the CPI of particular varieties subject to subsidies or price rollbacks. For example, subsidized apples represented deciduous fruits; other deciduous fruits not in the sample were unsubsidized (86). Oranges, which had a 25 percent rollback in their price control ceiling, represented citrus fruits; other citrus fruits (not in the CPI) had no price rollback (86). Shortening was used to represent fats and oils; subsidies were about to expire for all fats and oils except shortening (87), so even though most fats and oils prices would rise, the index would not.

To respond to Meany and Thomas's charge that foods not priced by BLS went up twice as fast as those priced, BLS collected a special sample of foods not in the official index (apparently, a retrospective price collection). It found that their weighted average inflation rate between August 1939 and January 1944 was 33.4 percent, compared with 37 percent for the foods used to represent them in the official index (A-9 to A-12). On balance, therefore, the Mitchell Committee concluded that no bias had been caused by the selection of varieties for use in the index. The Committee did, however, recommend pricing broader ranges of items in the future (II-39).

Although the BLS evidence did carry significant weight, it is hard not to be impressed that Meany and Thomas produced actual data to support their position. They were not disinterested samplers, it is true. Although the Mitchell Committee essentially opted for a "not proven" verdict on the points that Meany and Thomas raised, a potential for bias clearly did exist.

### 1.3.2 The Stigler Committee: Representative Samples for Detailed Component Indexes

The Stigler Committee recommended the use of probability sampling to estimate the detailed component indexes used to calculate the CPI.[52] Though the Committee cited technical papers by K. S. Banerjee (1959) and Irma

51. Additionally, such data are subject to error from differences over space in store amenities and services, so interarea indexes might not be so simple to compute as supposed.

52. It also recommended that the CPI be extended to cover all households. This was done in the 1978 Revision (CPI-U), but an index with weights corresponding to the old household definition was published separately (CPI-W).

Adelman (1958) as background, the wartime controversy was implicitly also part of the background of this proposal.[53] Of the Committee's recommendations, this was the one that had the most important effect on the CPI.

In addition to avoiding risk of a repeat of the wartime charges of manipulation of BLS's samples, probability sampling had a number of advantages. It made possible reliable estimates of index variances via the technique of comparing index values from replicated samples calculated using well-defined procedures to ensure consistency of methods (Stigler et al. 1961, 40).[54] Another benefit identified by the Committee was that the attempt to make sampling conform to a probability model would force the index designers to think explicitly about problems of definition and estimation that are easily ignored with judgmental procedures (42). Finally, probability sampling was the only way to guard against biases due to an unrepresentative selection of outlets and varieties (42). For example, types of retail establishments or outlets of growing importance were underrepresented, but use of "sampling frames showing the distribution of consumer expenditures for particular goods and services by market area and type of retail establishment" seemed a promising solution (58).

At the time, the specifications of the items to be priced were still determined centrally. The Stigler Committee thought that centralized specification of what was to be priced risked instructing field agents to price items that were unrepresentative of the sales in the particular retail outlets chosen, or even unavailable in those outlets. Probability selection of items in retail outlets meant that the items selected could indeed be found there, and assured that the full range of product specifications could be represented in the index in proportion to their sales.

One of the "Staff Papers" included with the Stigler Committee's report elaborated on the technical problems involved in probability sampling for estimation of the basic component indexes of the CPI. In this paper McCarthy (1961) discussed the sampling of commodities under the assumptions that the base year weights, base year prices, and comparison year prices were known without error for specified-in-detail commodities, and that the goal was to estimate a Laspeyres index (McCarthy 1961, 209). Following Adelman (1958), within any stratum, price relatives could be sampled with probability proportion to their base period expenditure. In an ideal situation, the sample estimator of the Laspeyres index for the stratum could then be calculated as a simple average of the sampled price relatives (McCarthy

53. Though the Stigler Committee report makes no explicit connection between probability sampling and the wartime controversy, John Marcoot, who was at BLS at the time of the Stigler Committee, cited in discussions with one of the authors the unions' criticisms as a reason for the adoption of probability sampling of outlets and varieties.

54. Assume, for example, availability of two index estimates for a commodity in a city based on two independent samples, selected by identical procedures. The difference between the estimates can be squared and divided by four to obtain an unbiased estimate of the variance of a "best" index estimate, based on a pooled sample.

1961, 213). McCarthy recommended selection by probability sampling of smaller cities, of detailed items within 150 commodity categories, and of outlets (227–9).

A witness at the hearings held to discuss the Stigler Commission's report made a noteworthy observation concerning the detailed component indexes that was not in the Committee's report. Arant (1961, 696) claimed that "[m]ost of the reductions in consumer prices brought about by the growth of mass distribution . . . have not been measured by the Consumer Price Index." Arant argued that the linking procedure used to bring new outlets into the CPI removed the effects of the lower price levels at chain supermarkets compared with the small independent stores that they were replacing.

Hoover and Stotz (1964) investigated Arant's claim and found that had BLS not linked new outlets into the CPI, the food component of the CPI would have dropped by 0.7 percentage points more than it did (their estimate was based on the difference between the weights for chains and independents in 1948 and weights implied by 1958 data). Note, however, that BLS did not ignore the growing importance of chain stores; it linked in interim adjustments to reflect this growing importance on four occasions in between the benchmarking of the weights to the 1948 and 1958 Censuses. Reinsdorf (1998, 184) found that after one of these interim adjustments, which reflected seven years of change in purchasing patterns, the average price series for foods dropped 0.7 percent compared with the official CPI.

Seemingly nothing has been done about the Hoover-Stotz finding, even though Hoover went on to be head of the CPI and Stotz went on to head the PPI. The issue lay quiet until the study by Reinsdorf (1993).

### 1.3.3 The Boskin Commission: Geometric Means as Basic Component Indexes

The Boskin Commission identified four sources of bias in the basic component indexes of the CPI. We leave the discussion of one of these, quality change, to a later paper. Here we will consider formula bias, lower level substitution bias, and outlet substitution bias.[55]

Formula bias was a focus of the Boskin Commission's "interim report," dated September 1995, because it had a substantial effect on the CPI and, unlike most index problems, which are far easier to recognize than to resolve, it was amenable to a quick solution. The Boskin Commission's final estimate of the size of this effect was about 0.5 percent per year. This estimate is consistent with research in Reinsdorf (1998, 185), indicating that formula bias in the commodities and services portion of the CPI may have had an effect on the all-items CPI of around 0.4 percent per year, and that a similar bias in the

55. The Boskin Commission's discussion of bias from new products is not included here because new products are discussed in the sections on the measurement concept. Also, new products are not specifically a problem affecting detailed component indexes since they may not fit into an existing item stratum.

owners' equivalent rent component of the CPI may have had an additional effect of around 0.1 percent per year.[56] The effect of formula bias on the CPI, was, therefore, more than twice as large as the Manser-McDonald estimate of commodity substitution bias of under 0.2 percent per year. Moreover, because important segments of the CPI, such as tenant-occupied shelter, were unaffected, a bias of 0.5 percent per year in the aggregate index implies quite a large bias for many of the affected individual components. For example, for two very homogeneous CPI components indexes—fresh whole chicken and bananas—a reasonable estimate of the combined effect of formula bias and outlet substitution bias can be made by comparing the growth rate of the CPI component index with the growth rate of a weighted average price in the CPI sample. Over an interval from 1980 to 1992, appropriately weighted averages of the prices in the CPI sample grew 1.1 percent per year faster than the CPI component indexes calculated from virtually the same set of prices (Reinsdorf 1998, 192). Comparisons of price levels between outlet samples in Reinsdorf (1993) suggested a typical value of about 0.25 percent per year for outlet substitution bias, leaving a residual not far below 1 percent per year for formula bias.

Formula bias was caused by procedures adopted in the 1978 revision of the CPI to implement probability sampling for selecting varieties and outlets as recommended by the Stigler Committee.[57] Notwithstanding this problem, probability sampling was an important step forward with many benefits. The CPI was less susceptible to bias from unrepresentative samples of varieties or qualities or from manipulation, as Meany and Thomas had alleged. Furthermore, BLS could estimate realistic variances for the CPI, something not possible when a single variety and quality level was chosen judgmentally to represent an entire commodity class.[58] Finally, estimates of variance components could be used to design efficient samples aimed at maximizing the precision of the CPI subject to the budget constraint.

Although one member of the Boskin Commission was heard to remark that the formula bias problem showed that BLS did not understand logarithms, the procedures in question came from papers by distinguished academics (Adelman 1958; McCarthy 1961) and from the Stigler Committee's report. To understand the genesis of the problem, note that random sampling of varieties and quality levels precluded continued use of averages of

56. Subsequently, experimental CPI's showed an average effect on the all-item CPI of about 0.24 percent per year, not counting the effect from owner's equivalent rent or the effect of lower-level substitution bias as measured by geometric mean indexes. See the *Economic Report of the President* (President's Council of Economic Advisors 1999, 94).

57. Probability sampling of outlets, but not varieties, began with the 1964 revision of the CPI with the assistance of Phillip McCarthy. (See U.S. BLS 1966, 26).

58. However, BLS publishes CPI variances only in places like the *CPI Detailed Report* and ASA Proceedings volumes, but not, as yet, on its website. Thus, an important benefit of the substantial expense of probability sampling is not easily available to users of the index.

prices, since prices of dissimilar items cannot meaningfully be averaged.[59] Constructing the refrigerator price index for city A as the ratio of average prices for average quality refrigerators sold in city A in two periods had been sensible, but trying to calculate an average price for musical instruments in a sample that contained guitar picks and pianos (an example used by Moses and Moulton [1997]) would not be. This suggested a change in computation for the basic component indexes from ratios of average prices (used before 1978) to averages of price ratios (used thereafter). The simple average of price ratios discussed by McCarthy was an unbiased sample estimator of a Laspeyres price index *if* the specific items priced were selected with probabilities proportional to expenditures in the base period and price collection began in the base period.

Neither McCarthy nor Adelman (nor Westat, the statistical consulting firm brought in to design the BLS move to probability sampling) considered a problem that arose in practical application. The BLS estimated sampling probabilities as well as possible (a new "Point of Purchase survey" was instituted to collect information on outlets, and item selection was done by probability methods within each retail outlet). Nevertheless, the perfect measurement of base period expenditures and prices assumed by McCarthy is far from achievable in practice. The base period for measuring expenditures was usually long enough to encompass one or more price changes, and additional price changes were likely to occur in the interval between the period of measurement of expenditures and the initial collection of price data. In addition, the estimates of expenditure shares of outlets and varieties within outlets were subject to sampling error.[60] Even if Cobb-Douglas substitution behavior by consumers kept expenditure shares invariant over time, the desired quantities for the Laspeyres index were from the time period when the expenditure shares were measured, and prices in that time period would not be the same as the prices in the period when price sampling began. Consequently, ratios of sampling probabilities to prices were not proportional to the desired quantities for a Laspeyres index.

59. Before 1978, the (unweighted) formula for a CPI basic component was: $(\Sigma p_{i,t+1}/n)/(\Sigma p_{it}/n)$, where $t$ is the last price observation (month for monthly pricing) and $t + 1$ the current month, and the calculation was done separately for each city. In words, then, the initial calculation for, say, refrigerators in the San Francisco area was the change in the average price of a matched sample of refrigerators in this city in the two months, and this calculation became the basic component for that item in the city index for San Francisco.

60. The estimates of expenditures for probability-proportional-to-size sampling of outlets came from surveys of consumers. Once the sample of outlets was known, to select detailed varieties and quality levels within sampled outlets BLS usually calculated sample selection probabilities based on approximate revenue breakdowns furnished by store managers or estimated from a proxy for revenue, such as shelf space. The time period for the outlet expenditures was earlier than the time period for the expenditures on varieties and quality levels within outlets, which was, in turn, earlier than the time of initial collection of price data in a newly selected sample.

Of course, by redefining the desired base period for the Laspeyres index to the time of initial price collection, one could show that the procedure yielded an unbiased estimator of a Laspeyres index objective given a Cobb-Douglas behavioral assumption, but, at the same time, this behavioral assumption would imply that the Laspeyres index with time of initial price data collection as its base period is a particularly poor objective. In comparison with a COLI-like objective, Cobb-Douglas behavior would imply great susceptibility to substitution bias.

What is more, given the "price bouncing" behavior that is typically observed at the lowest level of aggregation, Cobb-Douglas behavior would equally imply upward bias in a statistical sense in comparison with the objective that the index designers had in mind, a Laspeyres index with the time period furnishing the expenditure shares as its base. This bias is present because a downward transitory shock to an initial period price simultaneously implies an upward error in its implicit quantity weight and an upward shock to the change in the price. Since transitory price shocks are common for many types of items, the errors in the weights were positively correlated with the price changes, resulting in an upward bias.

For example, let item 1 and item 2 be represented in a component index, and let the prices for these two items in any month be either ($1, $2) or ($2, $1), with either configuration equally likely. With probabilities of selection based on historical averages of expenditure shares, the probability of selection of the item priced at $1 would be one-half, resulting in an expected value for the estimator of the component index used by BLS of (1/2)($2/$1) + (1/2)($1/$2) = 1.25.

Does any simple set of assumptions exist that would justify this estimator for the index? Two possibilities are shown in table 1.1. Suppose, first, that consumers always buy equal quantities of the items, a Leontief behavioral assumption. Equal quantities imply an expenditure share of one-third if the price is $1, so the item with an initial price of $1 should have a one-third probability of selection for the index sample. The sample estimator implied by this assumption therefore has an expected value of (1/3)($2/$1) + (2/3)($1/$2) = 1, which agrees with the value of 1 for the theoretical index shown at the bottom of the Leontief assumption portion of table 1.1.

Next, consider the assumption of Cobb-Douglas behavior, with consumers always spending equal amounts on the two items. Now the expenditure shares will indeed equal one-half, but the prices observed in the initial or reference period for the index will be of no value in estimating the quantities consumed in the earlier base period for the Laspeyres index basket. The expected values of the reference period prices are uniform. As is shown in the last row of table 1.1, if we estimate the base period quantities as inversely proportional to these uniform expected prices, the implied Laspeyres index again equals 1.

**Table 1.1 Component index objectives under two possible assumptions**

| Time period | Price of item 1 and item 2 | Base period quantities of items 1 and 2 | Cost of base period quantity of item 1 | Cost of base period quantity of item 2 |
|---|---|---|---|---|
| | | *Leontief behavioral assumption* | | |
| Reference period for index | (1,2) | (1,1) | 1 | 2 |
| Comparison period for index | (2,1) | (1,1) | 2 | 1 |
| Base period for index basket | (1,2) or (2,1) equally likely | (1,1) | (1,2) or (2,1) equally likely | |
| | | *Cobb-Douglas behavioral assumption* | | |
| Reference period for index | (1,2) | (1,2) or (2,1) equally likely | 1 or 2 equally likely | 4 or 2 equally likely |
| Comparison period for index | (2,1) | (1,2) or (2,1) equally likely | 2 or 4 equally likely | 2 or 1 equally likely |
| Base period for index basket | (1,2) or (2,1) equally likely | (2,1) or (1,2) | 2 | 2 |

Implied indexes

| Index description | Behavioral assumption | Index value |
|---|---|---|
| True Laspeyres | Leontief | (1 + 2)/(2 + 1) |
| True Laspeyres, based on realized market basket | Cobb-Douglas | (4 + 1)/(2 + 2) or (2 + 2)/(4 + 1) |
| True Laspeyres, based on expected value market basket | Cobb-Douglas | (1.5 + 3)/(3 + 1.5) |
| Estimator adopted in 1978 | n.a. | (1/2)(2/1) + (1/2)(1/2) |

*Note:* n.a. = not available.

Finally, continuing to assume Cobb-Douglas behavior, we could fill in the missing prices from the base period for the Laspeyres index with the two equally likely cases of (a) ($1, $2) and (b) ($2, $1). In case (a), applying the formula for the updating of expenditure weights (needed to calculate the change in a Laspeyres index starting in a period later than its base period; BLS calls these adjusted expenditure shares "relative importances") yields an adjusted expenditure weight (and sampling probability) of 4/5 for item 1 whenever its initial price in the index is $2, and an unchanged expenditure weight of 1/2 whenever its price is back at its base period value of $1. In case (b), the relative importance of item 1 equals 1/5 when its price is $1 and 1/2 when its price is back at the base period value of $2. The two possible outcomes for the change in the correctly based Laspeyres index are therefore (1/2)2 + (1/2)0.5 = 5/4 and (1/5)2 + (4/5)0.5 = 4/5. Thus, under this approach to defining the objective for estimation, the expected value for

the Laspeyres index objective, given the information that is available, again approximately equals 1.[61]

In January 1995, shortly before the naming of the Boskin Commission, BLS introduced the method of "seasoning" as a way to remove the correlation between the price changes and the measurement errors in the item weights that caused the problem of formula bias in the component index estimator adopted in 1978. The initial price data from a new sample was used to calculate the implicit quantity weights, then the sample was allowed to season for several months before its price changes were used in index calculations. Doing this converted the new average-of-ratios formula back into a ratio-of-averages formula; for example, a seasoned index from June to July might be an average of ratios of July prices to January prices divided by an average of ratios of June prices to January prices. Prompted by the interim report of the Boskin Commission, in June 1996 BLS extended the method of "seasoning" to nearly all items other than shelter.

The Boskin Commission was not satisfied with this solution. It regarded seasoning as only a partial cure for the formula-related problems of the basic component indexes, so its final report identified a remaining problem of "lower level substitution bias." This bias could be avoided by using geometric mean of price relatives as the formula for most component indexes in the CPI.

Whereas formula bias exists even with Leontief behavior, a substitution bias is, by definition, caused by substitution behavior. With some exceptions, consumers are likely to regard the varieties and outlets in a component index as highly substitutable. The Boskin Commission (19), citing a rationale in Shapiro and Wilcox (1996, 1997) (for which those authors had, in turn, cited a draft of Reinsdorf [1998]), argued that use of geometric means to average price relatives would result in unbiased estimation of basic component indexes for the COLI objective. Based on estimates of the effect on the CPI of the use of geometric means for basic component indexes, the Boskin Commission therefore estimated that the lower-level substitution bias was one-quarter percent per year.

Finally, the Boskin Commission discussed the effect of "new outlet substitution bias" on the basic component indexes of the CPI. This bias occurred when the entry of outlets offering lower prices, such as Wal-Mart, allowed consumers to save money by changing where they shopped. The effect on the cost of living of substitution between outlets with different price levels had, of course, been identified as an important problem in the Meany-Thomas report—though in the opposite context of involuntary substitu-

61. The reason the expected value is not exactly equal to 1 has to do with the behavior of expected values of nonlinear transformations, such as the division operation in calculating a price index. For larger sample sizes, the expected value would differ from 1 by a negligible amount. This exposition oversimplifies the BLS procedures to make the exposition of the root problem clearer.

tions to higher prices!—and BLS had acknowledged the possibility that price differentials between outlets could cause bias in its index as early as 1964.[62] Nevertheless, the issue was largely forgotten until a paper by Reinsdorf (1993). Multiplying the estimated 40 percent share of the CPI subject to this effect by Reinsdorf's (1993) upper bound estimate of 0.25 percent per year for the food and gasoline components of the CPI, the Boskin Commission estimated this bias at 0.1 percent per year. Subsequent research has shown similar outlet substitution problems in CPIs elsewhere—see, for example, the study by Saglio (1994) on chocolate bars in the French CPI.

### 1.3.4 The CNSTAT Panel: Guarded Agreement with the Boskin Commission

The BLS adopted geometric means for most basic component indexes in the CPI in 1999. This change made seasoning unnecessary, because the geometric mean formula is not subject to the kind of bias that seasoning corrects. The CNSTAT Panel gave qualified support to the change to geometric means. It observed that high substitutability between the product varieties in component indexes was generally plausible, with exceptions for some items such as prescription drugs. Hence, even though the specific behavioral assumption underlying the geometric mean formula—a unitary elasticity of substitution—was unlikely to be exactly true, the formula change was probably an improvement. Nevertheless, it was unclear to the panel that the geometric mean index would always be superior to the seasoned Laspeyres index. The CNSTAT Panel also remarked that this change marked "BLS's first attempt to build substitution effects into the CPI itself," and a "change in perspective from a COGI conceptual basis (informed by COLI considerations) to an explicit COLI basis" (Schultze and Mackie 2002, 62).

On the question of new outlet substitution bias, the CNSTAT Panel noted that the empirical evidence available to the Boskin Commission for its estimate of 0.1 percent per year was limited, and that Reinsdorf had viewed his estimates as upper bounds because price reductions might be accompanied by quality reductions (Schultze and Mackie 2002, 173). The panel's review of the available evidence suggested that outlet substitution bias was significant enough to be a matter of concern, but they doubted whether researchers would be able to produce sensible, reproducible estimates for adjusting for quality differences between outlets (175). They therefore concluded that BLS had little choice but to continue its present practice of linking in new outlets (176). They did, however, recommend continued research on the effects of outlet characteristics on prices.

62. Hoover and Stotz (1964) investigated outlet substitution effects. Consistent with their findings, Jaffee (1984) wrote: "This procedure [fixed samples of outlets] may in fact result in failure of the index to reflect real change in the prices paid by consumers which result from new outlets" (923).

### 1.3.5 Assessment of the Debate over Component Indexes

We have nothing to add to the CNSTAT Panel's treatment of new outlet substitution bias. However, the discussions of the geometric mean indexes in the reports of the Boskin Commission and the CNSTAT Panel raise two important issues.

First, in many cases the standard model of commodity substitution (which justifies the geometric mean, given Cobb-Douglas behavior) is not the appropriate model of the consumer choice process; rather, in cases where the varieties and outlets in a component index are virtually interchangeable, a model of consumer search with costly and imperfect information would be more appropriate, as Pollak (1998), Feenstra and Shapiro (2003), and Triplett (2003) point out. The CNSTAT Panel recognized this issue, and recommended further research on it.

Second, even granting the assumptions that (a) the standard commodity substitution model applies to component indexes and (b) the elasticities of substitution are near unity, the panels' discussion of substitution bias as the reason to adopt geometric means fails to consider the implications of lack of information on expenditure patterns within a component index. If the weights on the items in the component index reflect the purchasing patterns at the initial prices, geometric means are indeed a good way to adjust the weights for the changes in purchasing patterns brought about by price changes. But often what is needed is a theory applicable to situations where the weights do not necessarily reflect initial purchasing patterns. The solution in many of these situations is, again, a geometric mean index, but the logic that supports that solution is not the same as the logic used by the panel.

*Shopping Behavior versus Substitution Behavior*

Economists have sometimes interpreted the difference between arithmetic mean (Laspeyres-type) and geometric mean aggregators for basic components as just the classic substitution bias paradigm drawn from Konüs (1924), only applied one level down. This seems to be the interpretation of the Boskin Commission, which followed Diewert (1995).

Commodity substitution behavior is clearly a relevant concern for basic components. Part of the shift in the composition of musical instrument expenditures is accounted for by changes in relative prices (though it is also hard to maintain the constant tastes assumption for empirical work for such a category), and many other index components are made up from samples of substitutable commodities.

A theory of basic components, however, must be applicable to all basic components. It must explain differences between arithmetic and geometric means for components such as the CPI banana price index, which is as close to a homogeneous product (in the United States, at any rate) as can be found.

An even more challenging example is Schultz (1994), who found enormous formula differences in Canadian CPI data for a single size bottle of a single brand of soft drink—surely there is no room for commodity substitution within a single size and brand of one product.

To apply to index number formulas at this level, a theory of consumer behavior must model consumers' choices across sellers of a homogeneous commodity, as well as choice behavior across different (substitutable) commodities. Pollak (1998) put it well:

> I argue against the view of the Boskin Commission and Diewert (1995) that the "elementary aggregate" problem, which the Commission calls "lower level substitution bias," is primarily a problem of choosing an appropriate formula for combining the prices of items (71).
>
> At least when discussing price indexes . . . economists almost always proceed as if the "law of one price" holds so that the price distribution facing the consumer collapses to a point. With very few exceptions—the published literature appears to consist of three papers: Baye (1985), Anglin and Baye (1987) and Reinsdorf (1994)—economists have ignored the implications of price dispersion and search for the cost-of-living index (73).

Or as Pollak (1998) also put it, in a heading: "Why Shop?" The theory that is relevant to the basic component problem includes consumer shopping behavior, search behavior, inventory, and storage behavior. When soft drinks go on sale, consumers do not necessarily consume more of them (as the theory of commodity substitution has it); they stock up and store the soft drinks. Search, storage, and so forth are not necessarily modeled adequately at all by simply switching to a superlative index or a geometric index, since the theory that lies behind those indexes is not the theory that explains the consumer behavior that motivates consumer purchases.

Indeed, Triplett (2003, 156) presents a simple numerical example to show that with an imputation for search costs, no standard formula applied to prices collected from matched retail outlets will measure the COLI of households who shop. Feenstra and Shapiro (2003) develop a model for analyzing storage behavior (their application was canned tuna, using scanner data), and suggest that when this type of behavior is important the definition of a time period in the index should be as long as the intervals over which consumers plan their purchases. Hendel and Nevo (2002) show that neglect of consumer storage and shopping behavior results in an overestimate of ordinary demand elasticities, surely a fatal problem if one proposes to model index number substitution bias at the component level with a simple Konüs system.

In some cases characterized by shopping behavior, use of a unit value index might be justifiable. The usual objection to this simple method is that the unit value average of prices may change solely because of a change in

the distribution of quantities over sellers, so that the unit value index may be outside not only the Laspeyres-Paasche bounds, but bounds equal to the maximum and minimum price relatives. However, if the average price paid drops because information has become easier to obtain, so that consumers are better able to find the lowest prices, the COLI approach indicates that a drop in the price index is acceptable even though no price has changed.[63] The Laspeyres and Paasche indexes of the prices *from different sellers* are not necessarily bounds on the index needed for a COLI in the presence of costly information.[64]

The CNSTAT Panel (5, 24) addressed the question of consumer search: "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" (5). Also: "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" (24).

We agree that more research on these "nonstandard" problems is the proper future direction for understanding how to measure basic components in price indexes. Attempts to fit the basic component problem into the standard Konüs commodity substitution model lack insight into the nature of the problem and risk yielding misleading conclusions.

*Why Use Geometric Means?*

The property of simultaneously being an average of price relatives and ratio of average price levels makes the geometric mean index well-suited for handling the heterogeneity in varieties and qualities found in probability samples. However, the CNSTAT Panel report implies that the main purpose of the geometric mean formula is to account for substitution effects in component indexes that contain closely substitutable varieties, or closely substitutable outlets.

Assume—as is sometimes the case—that the standard model of commodity substitution is applicable to the items in the component index. Under the assumption that the expenditure shares are measured correctly, the geomet-

63. A BLS memorandum from 1963 summarizing expert advice from Edward Denison stated that the average price paid should ideally be used in the index, although Denison acknowledged the impracticality of collection of data on prices paid from households.

64. Note that costly information is probably immaterial for modeling choices over commodities even if it is important for modeling choices among competing sellers. Sellers' prices for the same commodity are likely to be characterized by high-frequency noise around a common trend, which is harder to learn about than the divergent trends and seasonal cycles that are likely to characterize different commodity prices. Also, as averages, the component price indexes for commodities (or commodity groups) and the weights used to aggregate these indexes into the all-items CPI are likely to be affected only slightly by the randomness associated with costly information. In contrast, the constituents of a single component index are individual prices, not averages.

ric mean index is indeed the formula for a COLI if the elasticity of substitution equals 1 (i.e., inverse movements of relative prices and relative quantities keep expenditure shares constant). Yet the relevant question is: what to do if the expenditure shares are unknown or measured poorly?

Consider first the case when the expenditure shares are unknown. In this case, we cannot measure a COLI with any precision, and to pretend otherwise is to deceive ourselves. With no knowledge of expenditures, the principle of symmetric treatment for items about which one has identical information implies an assumption of uniform expenditures in both the initial and the final period if the index includes dissimilar items. This assumption implies a geometric mean index.[65] Or, to assume explicitly that the elasticity of substitution is zero while avoiding the assumption of uniform quantities because the items are far apart in value, two alternative versions of the Leontief assumption must be treated as equally likely. One is that initial period expenditures are uniform while final period expenditures are directly proportional to the price relatives; the other is that final period expenditures are uniform while initial period expenditures are inversely proportional to the price relatives. An average of the two Laspeyres indexes implied by these equally plausible assumptions virtually equals a geometric mean index. (The equality is exact if the component index contains two items and the Laspeyres indexes are averaged geometrically.) Therefore, if expenditure patterns are unknown, it is a fallacy to infer from the fact that the items in the index are not substitutable that a geometric mean index will be downward biased.

The CNSTAT Panel's presumption of a link between the geometric mean formula and Cobb-Douglas substitution behavior led them to recommend against its use in compiling component indexes for items that are not substitutable (e.g., prescription drugs that treat different conditions). However, few indexes exclusively contain nonsubstitutable items (even a prescription drug index may have some substitutable items, such as generics from different makers or identical drugs from different outlets), so the right question is how to handle a mixture of substitutable and nonsubstitutable items.

Usually the information on expenditures shares of the outlets and varieties covered by a component index is somewhere in between perfect measurement and perfect ignorance, so a plausible assumption is that the expenditure shares reflected in the component index sample approximately equal the shares at the start of price data collection. When substitutes are available for the items with unusually high or low price relatives and the weights reflected in the index approximate the true weights, the geometric mean index for-

65. If the index covers homogeneous items, the symmetric assumption is one of uniform quantities in both time periods. This assumption implies a ratio of average prices for the index formula. But when pianos and guitar picks are in the same component index (an example from Moulton and Moses [1997]), a uniform distribution of quantities implies wildly unequal expenditures.

mula provides a good measure of the component index for the COLI. In the short run, these conditions are plausible, even for a component index containing many nonsubstitutable items. Yet in the long run, a geometric mean component index containing nonsubstitutable items may tend to be biased downward because prices of the nonsubstitutable items in the index often follow divergent trends, but not the substitutable items. If every item that could be substituted follows the same price trend, but items that cannot be substituted follow divergent trends, opportunities for substitution will be limited in the long run. This provides some support for the CNSTAT Panel's recommendation to avoid the geometric mean for component indexes covering nonsubstitutable items.

## 1.4 Conclusion

The history of reviews of BLS price index programs begins almost as early as the impressively long history of those programs. These reviews, and the parallel professional literature on price measurement that influenced them, provide a record of how the interplay between intellectual progress in the disciplines of economics and statistics on the one hand, and the need to address public and professional perceptions of shortcomings in existing procedures on the other, shaped the evolution of the methods used to measure the CPI.

Two of the most distinctive features of the U.S. CPI compared to its counterparts in many other countries are its use of probability sampling to select outlets and product varieties for the basic component indexes and the use of the Konüs Cost-of-Living Index as the measurement concept. Both these features can be traced to the Stigler Committee, whose recommendations were influenced by the World War II era controversy over the charges contained in the Meany-Thomas report. In turn, one of them (the Cost-of-Living Index) played a key role in the subsequent review of the CPI by the Boskin Commission, and a memorable role in the recent deliberations of the CNSTAT Panel.

We also link the Boskin Commission with the component index recommendation. The original implementation of probability sampling gave rise to formula bias in the basic component indexes. The emerging evidence of this problem—which was discovered by BLS researchers—played a key role in the events leading up to the naming of the Boskin Commission. The recommendation of the Boskin Commission that had the greatest practical effect on the CPI was to construct most basic component indexes using a geometric mean index formula that is well-suited for probability samples of outlets and varieties.

The CNSTAT review of the CPI is most memorable for its partial retreat from the Stigler Committee's recommendation of the use of the COLI as the measurement concept for the CPI. Some of the skepticism about the

COLI is the result of people on both sides of the debate taking language too literally—statements about substituting commodities in ways that yield equal amounts of "satisfaction" or "utility" are, at bottom, statements about economic value, not psychological states of mind. The criticisms of the COLI measurement concept in CNSTAT Panel's report provide valuable illumination of the limited domain of this theory, and a valuable reminder that the effects of violations of the underlying assumptions can be important enough to be a practical concern in some situations, such as attempts to adjust the CPI for the effect on the COLI of new kinds of goods.

Yet despite its defects, the COLI offers some unique and critical advantages. No other approach provides an identifiable abstract objective to guide our thinking and to unify our treatment of the broad range of questions that arise in designing and using the CPI. Moreover, many of the assumptions that affect the COLI also affect the alternatives to the COLI, but only in the COLI framework are we able to rigorously analyze their implications for the measure of price change.

## Appendix

### *Economic and Test Approaches to Index Numbers*

Three major developments in the 1920s influenced professional thinking about CPIs and were therefore absorbed into the intellectual tradition that influenced subsequent reviews of the CPI.[66] In the first, Fisher (1911, 1922) elaborated the "test approach" and extended a way of thinking about index numbers that carried over from the nineteenth century (see also Walsh [1901]). Fisher's approach was not specific to consumer price indexes; it applied to all types of index numbers.

The second development was Konüs' (1924) theory of the Cost-of-Living Index (COLI). In contrast with Fisher and his forebears, Konüs' contribution was uniquely a contribution to the measurement of consumer prices, inspired, probably, by the new "Cost-of-Living Indexes" in the United States, Canada, the United Kingdom, and other countries. Konüs' COLI concept resembles one introduced earlier by Pigou (1920). Pigou was interested primarily in what is now called the standard of living index, a quantity index that represents movement from one indifference curve to another. However, he recognized that a price index was needed as a deflator.[67] The

66. This appendix is based largely on Reinsdorf (2007) and unpublished materials by Triplett. It also draws on Balk (1995, 1996), whose views among the contributors to this topic are closest to our own, and on Diewert (1992, 1993).

67. Whether Pigou should get the credit for the invention of the Cost-of-Living Index has been debated. Pigou (1912; 1920; 1932, 59–65) showed that the Laspeyres and Paasche indexes

third development, the Divisia index, is unrelated to current questions about measurement of the CPI, so we ignore it here.[68]

Test and COLI approaches have been reviewed previously. Because ours is not always the standard interpretation, a clarification of our interpretation is important background for the chapter. And because this is an appendix to an already long chapter, it is appropriate to present at the outset how we come out.

We favor the COLI approach over the test approach. It is easy to point to problems with the application of COLI theory to actual index number problems (Schultze and Mackie [2002] present a most comprehensive treatment of that set of problems). However, the problems surrounding the application of the test approach, *as it has been presented in the literature,* are far more daunting, and are less well documented and therefore less well appreciated.

We do think there is some role for pragmatic review of index number properties in particular settings. However, we relegate that role primarily to choosing among index numbers with equally good economic properties.

As an example, a GDP system requires a price index and a quantity index that together exhaust the change in current price GDP between two periods, what Stone and Prais (1952) called a "compatible index number system." A Fisher index number system for national accounts has the convenient property that the product of a Fisher price index and a Fisher quantity index equals the change in current price expenditure between the two periods; thus, both the "change in real GDP" and the associated implicit price deflator can be written as explicit and straightforward index numbers.[69] This convenient property of Fisher indexes led the U.S. Bureau of Economic Analysis to choose, several years ago, a Fisher index number system for both its measures of price change (the implicit deflator, a price index) and of constant price GDP (a quantity index).

The BEA chose the Fisher system only after the options were narrowed to index numbers with equally-appealing economic properties. Diewert (1976)

---

were upper and lower bounds for his index concept, so Staehle (1934) contended that Pigou's theory contained the essential elements of Konüs' contribution. Frisch (1936, 22), however, disagreed, arguing that only Konüs had really used indifference concepts to define the index. Other early discussions of Laspeyres-Paasche bounds were in von Bortkiewicz (1923), Gini (1924), Haberler (1927), Bowley (1928), and Keynes (1930). Even earlier discussions that did not include the bounds were an 1898 discussion by Wicksell of an index number concept that held well-being constant (see Frisch 1936, 11) and a 1707 discussion of an index number objective of "same Ease and Comfort" used by the Bishop of Ely in 1707 (see Samuelson and Swamy 1974, 567).

68. On the connection between the Divisia index and the COLI, see Hulten (1973) and Balk (2005).

69. The deflator corresponding to a Tornqvist quantity index is the change in expenditure divided by the Tornqvist quantity index. It cannot be reduced to a simpler expression that can be easily manipulated and analyzed. Hence, using a Tornqvist index to measure constant price GDP would make the analysis of inflation awkward and difficult.

showed that such indexes include the Törnqvist and the Walsh index numbers, as well as the Fisher. The Fisher index was not chosen from among all the candidate index numbers by applying tests.

Indeed, as we explain in this appendix, the test approach does not yield the Fisher index as the "best" index number, *unless one selects the tests to yield the Fisher index as a result,* which Fisher (and others who have followed him) have done. The test approach, as it has been applied in the literature from Fisher (1922) to the present, includes some properties that are questionable. It also arbitrarily excludes appealing (to some users) properties that would have yielded a different index number as "best." Additivity (of the numerator) and consistency in aggregation are the most notable properties that are omitted in the traditional test literature. Consideration of these properties, particularly if high weight is placed on them, would have tilted the test results toward the Laspeyres index.

In short, the properties that have traditionally been included (and excluded) in the test approach, as it has appeared in its own literature, are disputable. The approach contains no criteria to explain why properties are included or excluded. Lacking an analytical framework, it substitutes instead notions of "appealing" index number properties. Yet different index number users have different views about which index number properties are appealing (perhaps because of different index number settings). To those index number users who endorse a different set of properties, a user of the test approach can only say: "I do not share your views."

Without ignoring its problems (though we think they have often been exaggerated), the COLI approach offers an analytical framework for thinking about index number construction. The COLI's analytical framework extends far beyond the simple problem of choosing an index number formula (the only problem for which the test approach has been developed). It provides a framework for reasoning about the myriad other problems that arise in practical index number construction.

**Fisher and Konüs**

For Fisher (1911), the purpose of price index measurement was to obtain a $P$ for the quantity theory expression $MV = PT$, the equation of exchange. The quantity theory was, at the time, the only well-specified economic theory for which a price index was relevant—though it is the price *level* that is directly relevant.

Mitchell (1921, 1923) disagreed with Fisher. Mitchell criticized the equation of exchange as a foundation for a price index. "To 'measure variations in the exchange value or purchasing power of money' is not a clearly defined aim. . . . What does 'the purchasing power of money' include? Merely the standardized wares of the wholesale markets which are sampled with varying thoroughness in the current index numbers? Or does it include also com-

modities at retail, stocks, bonds, labor of all sorts, farm lands and town lots, loans, transportation, insurance, advertising space, and all the other classes of goods that are bought and sold? [ . . . ] To insist that this problem has but one meaning and therefore one 'best' solution [he refers here to Fisher and to Walsh] obstructs progress" (Mitchell 1921, 23).[70] In this passage, Mitchell raises, possibly for the first time, the question of "domain" of the index number, and the basis for choosing the domain.

As the quantity theory implies, Fisher regarded the price level as an entity in itself. Banzhaf (2001) points to Fisher's use of physical analogies: to measure the level of water in a lake, one ignores the ripples and waves; alternatively, the change in the price level is like an exploding shell where one ignores the paths of the fragments to obtain the true trajectory. The best form of the index number is determined by "tests" on the "reasonableness" of the measured index, which Fisher conceived as a problem that is parallel with measuring the level of the lake or the trajectory of the shell. References to such physical analogies show that Fisher was influenced by the older literature on the stochastic approach to index numbers, which contained many analogies of this type.[71]

Fisher's view of price indexes encompassed neither an economic concept nor an explicit theory of aggregation over commodities.[72] In contrast, Konüs' COLI theory is not only an economic concept, it is also *explicitly a theory of aggregation.* In COLI theory, the index tracks changes in the cost of attaining a specified utility level. To accomplish this requires an aggregating function that is derived from the form of the utility function (strictly, the form of the indirect utility function). Pollak (1983, 1989) and Diewert (1983) provide surveys.

Unlike Fisher, for whom the movements of the individual prices indexes were simply noise (the fragments of a bursting shell or the ripples in the pond), in Konüs' approach the deviations of the individual basic components—in particular, the relative prices—were the essence of the measurement. This roots the theory in the individual prices. Individual prices are real and observable. The aggregate index, on the other hand, is an economic

70. The inclusion of asset prices in the Consumer Price Index has a modern counterpart, but in the COLI context. See Goodhart (2001) and Cecchetti et al. (2000).

71. The stochastic approach sought to estimate statistically the central tendency of the price changes in the economy and generally made no attempt to weight items in the index to reflect expenditure patterns. It was used by Jevons, Edgeworth, and others, but abandoned, at least in its original form, following Keynes' (1930, 85) criticism of it as "root-and-branch erroneous." Keynes credits Fisher, Mitchell, and Walsh for avoiding this mistaken approach.

72. Dimand (2005) pointed out that Fisher's ignoring the potential of consumer theory for a consumer price index (as he did as well for some of his other research) was peculiar, in view of his pioneering nineteenth century work on utility theory. He goes on to attribute this to Fisher's contracting tuberculosis. After his illness, he seems to have belittled pleasure as a guideline for social welfare, and perhaps as well had become agnostic about the rationality of individual choice (Fisher devoted a great amount of energy after his illness toward changing behavior in more healthy directions). If the latter is true, he was a precusor to recent reconsideration by economists of their models of individual and household behavior.

abstraction, useful for economic analysis but hardly analogous to a directly observable quantity like the level of water in a lake. The Konüs view of things implies that the price index is not some physical or metaphysical entity that is "out there," waiting to be captured by an appropriate measuring rod. Instead, it is an invention by economists and statisticians for the purposes of economic analysis.

Fisher thought his ideal index (the geometric mean of Paasche and Laspeyres indexes) provided the true measure of price change between the Laspeyres-Paasche bounds introduced by Pigou for his welfare index deflator. Keynes (1930, 112–113) found Fisher's logic unconvincing:

> Then, as we have seen above, the true measure of comparison—assuming that tastes, etc. are constant and that only relative prices are changed—between the price levels in the two positions necessarily lies somewhere *between p* [the Laspeyres index] and *q* [the Paasche index]. Professor Fisher (amongst others) concludes from this that there must be some mathematical function of *p* and *q* which will afford us the best possible estimate of whereabouts between *p* and *q* the true value lies. Setting out on these lines, he has proposed and examined a great variety of formulae with the object of getting the best possible approximation to the true intermediate position. [. . .] I see no real substance in Professor Fisher's long discussion, by which, after examining a vast number of formulae, he arrives at the conclusion that $\sqrt{pq}$ . . . is theoretically ideal—if, that is to say, he means by this that it is likely to be arithmetically nearer to the truth than other formulae. This conclusion is the result of applying a number of tests. [ . . . ] All these tests, however, are directed to showing not that it is correct in itself, but that it is open to fewer objections than the alternative *a priori* formulae. They do not prove that any one of the formulae has a leg to stand on, regarded as a probable approximation.

Ironically, a demonstration of the approximation property of the Fisher index that Keynes found lacking in Fisher already existed when Keynes wrote these words.

In papers written in Russian that were little known until Diewert (1976) revived them, Byushgens (1925) and Konüs and Byushgens (1926) showed that the homogeneous quadratic indirect utility function yields a COLI that corresponds to Fisher's "ideal" index.[73] Later, Diewert (1976) showed that this utility function provides a second order approximation to the unknown true function; thus, the Fisher index can be used to approximate the COLI index. The exposition of this point in the Committee on National Statistics Panel report (Schultze and Mackie 2002, 84) is worth quoting:

> The remarkable thing about this [Byushgens (1925)] 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. [ . . . ] The Fisher ideal index is therefore *exact* for a set of preferences and demand

73. Because neither of us reads Russian, we follow Diewert's 1976 exposition of these two articles, supplemented by his personal communication on our chapter.

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.

## Comparison of COLI and Test Approaches

Both Fisher (1922) and Konüs and Byushgens (1926) justified the same formula. For this reason, it is sometimes said that test and COLI approaches converge: COLI theory shows that the Fisher index is a good approximation to the true index, which merges with Fisher's conclusions on the basis of his tests. However, the conclusion is too superficial, for two reasons.

First, it implies that the Fisher index clearly emerges from the test approach. In fact, *results from the test approach depend on the subjective process of selecting the tests.* We elaborate on this in the following numbered section, "Selection of Tests."

Second, the notion that test and economic approaches converge presumes that the only thing that matters is the index number formula, and that the implications of the COLI approach are limited to the choice of an aggregator. In fact, the COLI framework is more than a guide to aggregation (that is, to the choice of index number formula). *It is also a way to think about what should be included in the price index and how to measure the price index components.* The test approach is completely silent on these matters, leaving vital portions of CPI construction without a guiding framework.

Again, Mitchell was insightful: "The first step, framing a clear idea of the ultimate use of the results, is most important, since it affords the clue to guide the compiler through the labyrinth of subsequent choices. It is, however, the step most frequently omitted" (Mitchell 1921, 23). He points to Fisher and Walsh as examples of researchers who ignored this principle. It has also been ignored by nearly all of the subsequent followers of Fisher and the test approach.

## 1. Selection of Tests

Fisher discussed numerous index number tests, though he explicitly applied the term "test" only to those he found especially noteworthy. Some of these he dismissed (the circularity test, which involves a kind of transitivity property, came in for particular criticism). Selection among the tests was necessary because, as Frisch (1930) later showed, no index number formula satisfies all reasonable tests.

Fisher treated two tests as particularly important. One of these was the time reversal test, which requires that the index number for the change from period 0 to period $t$, $P_{0t}$, equal the reciprocal of $P_{t0}$. The other was the "factor reversal" test, which requires that the price index and the quantity index have the same functional form. He identified as superior the formula that is

now known as the "Fisher Ideal" index because it satisfied these two tests, and because it also had some additional, less critical, advantages.

Fisher's test approach is subject to three criticisms: First, the index properties included in the tests are arbitrary; little basis is given in the test literature for why the properties that are included are there. Second, as suggested by the first criticism, the approach invariably omits index number properties that some users of National Accounts and the CPI contend are critically important—notably additivity and consistency in aggregation. Third, though some index number properties included in the tests must be more important than others for particular uses, advocates of the test approach usually favor a metric where each test is of equal importance.

### Properties Included

Fisher presented no theory to explain why the particular price index properties he selected provide the relevant tests. Indeed, on economic grounds some authors have rejected some of these tests.

Notably, the factor reversal test was dismissed by Samuelson and Swamy (1974, 575) with the colorful remark, "A man and his wife should be properly matched, but that does not mean I should marry my identical twin!" Because the form of the COLI depends on the form of the indirect utility function, and the form of the associated standard of living index depends on the direct utility function,[74] Fisher's factor reversal test thus implies that direct and indirect utility functions should have the same form. This property is known as "self-dual" in the demand literature. Most demand specifications used empirically do not have this characteristic (for example, Deaton and Muellbauer's [1980] "Almost Ideal Demand System" is not self-dual). There is no reason to specify the self-dual property as some sort of theoretical ideal.

The time reversal test can also be criticized as inconsistent with the economic approach to index numbers. If the indifference curve used to evaluate the COLI is predetermined, the COLI indeed has the time-reversal property. However, the usual convention is to use either the starting or the ending period to define the reference indifference curve. That is, one can frame the cost-of-living question as: what is the minimum expenditure necessary currently to achieve the base period's standard of living?[75] Alternatively: what is the expenditure necessary in the base period to consume the current period's standard of living? Adopting either of these conventions makes the COLI fail the time reversal test, except in the special case of homotheticity. Beginning at one point, passing to another, and then returning to the first (the essence of the time reversal test) involves passing through two utility levels; it is not the same thing at all as remaining at the initial point. The two

74. See Samuelson and Swamy (1974) and Pollak (1989).

75. The index uses the answer to the question as the numerator and the actual base period's expenditure as the denominator.

utility levels have equal influence on the (three-period) result and with nonhomotheticity, the time-reversal test will (*appropriately*) fail. Fisher's time reversal test is seemingly plausible, but he and others who have endorsed it failed to recognize the equally-important role of the two consumption levels that are involved in it.

Thus, Fisher's tests include some that are insupportable from the standpoints of economic theory and empirical work on consumer demand. This is our first criticism of the test approach.

### Properties Omitted

Many users of national accounts (in the United States and elsewhere) have maintained that additivity of the index number is an essential property because it enables the users of the accounts to add up the component measures of constant price output (real output, in U.S. macroeconomics usage) to get, for example, constant price investment expenditure, or the level of constant price GDP. Users desire that GDP = $C + I + G$, in real or constant price terms, as well as in current price terms, and that one can add constant price expenditures on structures, equipment, and inventories to get constant price investment. A similar contention has been made for the CPI. The Laspeyres index is additive in the sense desired, nearly all other index numbers are not.[76]

The test approach provides no basis for evaluating the position that additivity is or is not an important "axiom" for an index number, so one must either arbitrarily accept it or arbitrarily reject it. The Fisher index number, which is among the preferred index numbers under the economic approach and also the index number that Fisher thought the best, fails the additivity test. If one regards additivity as important—even more if additivity is thought to be the most important property of index numbers—one might contend that the Fisher index has worse test properties than index formulas (such as the Laspeyres index, or the Edgeworth-Marshall index) that do satisfy this test. (This, for example, is Baldwin's [2009] conclusion.[77])

Consistency in aggregation is another desirable property of index numbers for many purposes. This property means that the index can be calculated in stages, with the intermediate stage aggregates treated as if they were price relatives at the next higher stage of aggregation. For example, the aggregate Laspeyres CPI can be calculated from its component indexes for food, clothing, housing, appliances, services, and so forth.[78] The Fisher index, as one

76. To avoid confusion, what is wanted is additivity in the levels (the components of the numerator of the index number), not additivity in the index changes. The Walsh index is also additive in the levels under certain circumstances. An index that is additive in the levels will also be consistent in aggregation.

77. To be clear, we are not advocating additivity as a property of either the CPI or of national accounts. But that is because we accept the economic approach to index numbers, which does not support imposing additivity.

78. Balk (1996) presents a more precise definition of consistency in aggregation and a discussion of a family of indexes that have this property.

example, is not consistent in aggregation—a Fisher index of Fisher indexes is not a Fisher index (indeed, the Bureau of Economic Analysis publishes estimates of the error from combining component Fisher indexes).[79] An index number that is additive is also consistent in aggregation, but nonadditive indexes may also be consistent in aggregation.

When BEA adopted Fisher indexes to measure price and quantity change in the U.S. national accounts, users complained that Fisher indexes were not additive and not consistent in aggregation. Yet the literature on the test approach is almost entirely silent about these two desirable properties (the previously cited items by Balk are exceptions). In consequence, the supposedly pragmatic test approach to index numbers has nothing to say about some of the major—and most controversial—pragmatic issues of index number construction that have arisen in recent years.

### No Rationale for Determining Which Properties Are Included in the Tests

No index number formula satisfies all of Fisher's tests, so a decision must be made about which to select as valid. Some tests appear trivial, others as we noted previously, objectionable. Often the tests that are selected are justified as "reasonable." Reasonableness, however, is defined differently by different people (as experience richly shows), and reasonableness may also depend on the purpose of the index number.[80] The economic approach has been used to pare down the list of valid tests, but not in a thorough and comprehensive way.

### No Rationale for Weighting the Index Properties

In the most common application of the test approach, one simply counts the number of the arbitrary tests that are passed by a particular index number formula. The formula that passes the most tests "wins." This amounts to weighting all tests equally. Why should they get equal weight? And, most importantly, why should all the tests get the same weighting, equal or not, for every index number problem?

### Summary

The particular tests that Fisher (and others who have followed him) chose included some that are debatable and excluded others that some index users regard as essential. Had different index number properties been included in the tests, a different index number formula would have emerged as "best,"

79. The error of approximation is fairly small and so empirically supports Diewert's (1978) analytical result that all superlative indexes are approximately consistent in aggregation. Nevertheless, critics have not been satisfied with approximate consistency when other index number systems (Laspeyres) give *exact* consistency in aggregation.

80. We are aware of the counterargument that says that if the test approach rests on "axioms," the COLI approach also rests on an "axiom" (utility maximization). We regard this point as semantic manipulation. The "axiom" of utility maximization is a testable hypothesis, on which a great amount of economic literature exists. It is hardly comparable to index number "tests" such as time reversal and so forth, which are simple index number properties.

especially if one gives up the notion that all tests should be weighted equally. In an extreme case, perhaps, some users of index numbers have contended that additivity and consistency in aggregation should receive all the weight (they prefer, naturally, the Laspeyres index number); the test approach itself has no basis for excluding this view other than arbitrary rejection of it. Accordingly, the test approach does not yield the Fisher index number, *unless one selects the tests in order to get it.* Index number properties have a role for choosing among index number formulas, but that role is primarily to choose among index numbers with good economic properties.

## 2. Measurement of Index Components

The economic and the test approaches to index numbers differ in a more fundamental dimension because the theory of the COLI has two parts. One part, as noted, is a theory of aggregation—that is, selecting the index number "formula." The second part concerns (a) which components are to go into the index—the domain of the index—and (b) how components of the index are to be measured.

The COLI provides the answer to a question, for example: "What is the cost at today's prices of obtaining the standard of living of the base period?" This is not, we emphasize, the same thing as the question: "What index number formula do we choose?" The COLI question is much broader and much more comprehensive. It directs attention to the standard of living, and invites us to ask how the standard of living should be measured.

As an example, faced by at least one of the reviews covered in this chapter, a forced substitution (for example, a regulatory change that removes some variety of a product from the market) and a voluntary substitution induced by changes in relative prices are both changes in the "market basket." Should these be handled in the same way in the price index? Or not? The COLI theory forces us to confront those questions, and suggests some answers, at least in qualitative terms. The test approach is wholly silent.

Another example comes from a current controversy in CPI measurement around the world: is the standard of living defined on the consumption of housing services? Or is it defined on the purchase of houses? Does the consumption of housing or the purchase of houses belong in the domain of the CPI?

The COLI approach has universally been interpreted as implying that housing services are the appropriate concept for the index.[81] On the other

81. The CNSTAT report (Schultze and Mackie 2002) points out that one could use the flow of services approach to owner-occupied housing (which is derived from the theory of consumption) in an index that does not use the theory of consumption for aggregation. That is, the flow of services from durables could be combined with what the CNSTAT committee called the COGI approach. Though the position is logically sustainable, around the world

hand, no "test" on the reasonableness of an index supports any particular treatment of housing in the CPI. Consequently, a great variety of approaches to owner-occupied housing (Diewert [2004] reviews them) have emerged in countries that do not subscribe to the COLI framework. Various strategies (an "acquisitions" index, a "payments" index, and so forth) have been devised to try to provide a non-COLI framework for designing CPI components. None, in our view, succeeds.[82]

## Conclusion

We have contended that the test approach to index numbers cannot be used for guidance about aggregating the basic components (see paragraph section 1). The tests included and excluded from the tests are arbitrary and the index number that emerges as "best" is very sensitive to the arbitrary choice of tests. In particular, the test approach does not, despite so many assurances to the contrary over many years, favor the Fisher index, unless the set of tests is chosen to produce the Fisher.

Secondly, the test approach has nothing to say about which components should be included in a consumption index; it cannot be used to define the index domain. It also says nothing about how those components should be measured. Those are essential points in constructing an index—likely more essential than the choice of index number formula, important as that is—and they are points for which the economic approach has much to say.

Significantly, no U.S. review, with the exception of some passages in the CNSTAT Panel's report, puts any weight on test approaches. This was undoubtedly because, as Reinsdorf (2006) remarks, the economic approach to index numbers dominated the test approach through most of the twentieth century. More recently, the test approach has become popular in Europe, especially in academic circles, while the economic approach remains popular in the United States—a curious trans-Atlantic reversal of roles since the early developers of the economic approach were all European, while main developers of the test approach were the Americans Fisher and Walsh.

no one inhabits this particular "intellectual house." When statistical agencies reject the flow of services approach to owner-occupied housing, they also reject the COLI approach as the conceptual guide for their index, and they frequently reject the COLI *because* they do not wish to implement a flow of services approach to owner-occupied housing.

82. Melser and Hill (2005) suggest that the "use" approach sometimes proposed in these non-COLI frameworks is compatible with the economic approach to index numbers.

## References

Abraham, K. G. 2003. Toward a cost-of-living index: Progress and prospects. *Journal of Economic Perspectives* 17 (11): 45–58.

Abraham, K. G., J. S. Greenlees, and B. R. Moulton. 1998. Working to improve the Consumer Price Index. *Journal of Economic Perspectives* 12 (1): 27–36.

Adelman, I. 1958. A new approach to the construction of index numbers. *Review of Economics and Statistics* 40 (3): 240–9.

Advisory Commission to Study the Consumer Price Index. 1996. See Boskin et al. 1996.

Afriat, S. N. 1967. The construction of utility functions from expenditure data. *International Economic Review* 8 (February): 67–77.

Aizcorbe, A. M., and P. C. Jackman. 1993. The commodity substitution effect in CPI data, 1982–91: Anatomy of price change. *Monthly Labor Review* 116 (12): 25–33.

Anglin, P. M., and M. R. Baye. 1987. Information, multiperiod search, and cost-of-living index theory. *Journal of Political Economy* 95 (6): 1179–95.

Arant, W. D. 1961. Statement. In *U.S. Congress, Joint Economic Committee (1961): Government price statistics, hearings, part 1, May 1, 2, 3, 4, and 5, 1961.* Washington, DC: GPO.

Arrow, K. J. 1958. The measurement of price changes. In *The relationship of prices to economic stability and growth: Compendium of papers submitted by panelists appearing before the Joint Economic Committee.* Joint Economic Committee Print, March 31. Washington, DC: GPO.

Baily, M. N. 2006. Implications of the Boskin Commission Report. *International Productivity Monitor* 12 (Spring): 74–83.

Baker, D. 1998. *Getting prices right.* Washington, DC: M. E. Sharpe.

Baldwin, A. 2009. Chain price and volume aggregates for the System of National Accounts. *Price and Productivity Measurement* 6:241–78.

Balk, B. M. 1990. On calculating cost-of-living index numbers for arbitrary income levels. *Econometrica* 58 (1): 75–92.

———. 1995. Axiomatic price index theory: A survey. *International Statistical Review* 63 (1): 69–93.

———. 1996. Consistency-in-aggregation and Stuvel indices. *Review of Income and Wealth* Series 42 3 (September): 353–63.

———. 2005. Divisia price and quantity indices: 80 years after. *Statistica Neerlandica* 59 (2): 119–58.

Banerjee, K. S. 1959. Precision in the construction of cost-of-living index numbers. *Sankhya* 21 (3/4): 393–400.

Banzhaf, H. S. 2001. Quantifying the qualitative: Quality-adjusted price indexes in the United States, 1915–61. *History of Political Economy* 33 (suppl. 1): 345–70.

Baye, M. R. 1985. Price dispersion and functional price indices. *Econometrica* 53 (1): 213–23.

Beelen, G. 2004. Vehicle, homeowners', and tenants' insurance in the Canadian CPI. Paper presented to the Statistics Canada Price Advisory Committee. April, Tunney's Pasture, Ottawa, Canada.

Berndt, E. R. 2006. The Boskin Commission report after a decade: After-life or requiem? *International Productivity Monitor* 12 (Spring): 61–73.

Berry, J. M. 1995. Agencies debate changes to Consumer Price Index: Billions in cost-of-living changes at stake. *Washington Post,* January 27.

Blow, L., and I. Crawford. 2001. Should the cost-of-living index provide the concep-

tual framework for the Consumer Price Index? *The Economic Journal* 111 (472): F357–F382.
Boskin Commission. 1996. See Boskin et al. 1996.
Boskin, M. J. 2005. Causes and consequences of bias in the Consumer Price Index as a measure of the cost of living. *Atlantic Economic Journal* 33 (1): 1–13.
Boskin, M. J., E. R. Dulberger, R. J. Gordon, Z. Griliches, and D. Jorgenson. 1995. Toward a more accurate measure of the cost of living: Interim report to the Senate Finance Committee from the Advisory Commission to Study the Consumer Price Index. Washington, DC: Senate Finance Committee.
Boskin, M. J., E. R. Dulberger, R. J. Gordon, Z. Griliches, and D. Jorgenson. 1996. Toward a more accurate measure of the cost of living: Final Report to the U.S. Senate Finance Committee. In *U.S. Senate Finance Committee. Senate Print 104-72, December, 104th Congress, 2nd Session.* Washington, DC: GPO.
Bowley, A. L. 1928. Notes on index numbers. *The Economic Journal* 38 (150): 216–37.
Braithwait, S. D. 1980. The substitution bias of the Laspeyres Price Index: An analysis using estimated cost-of-living indexes. *The American Economic Review* 70 (1): 64–77.
Brown, M., and D. Heien. 1972. The s-branch utility tree: A generalization of the linear expenditure system. *Econometrica* 40 (4): 737–47.
Buyshgens, S. S. 1925. Sur une classe des hypersurfaces: A propos de l'index ideal de M. Irving Fischer. *Mathematischkii Sbornik* 32:625–31.
Cage, R., J. Greenlees, and P. Jackman. 2003. Introducing the chained Consumer Price Index. Paper presented at the Seventh Meeting of the International Working Group on Price Indices. May, Paris, France.
Carruthers, A. G., D. J. Sellwood, and P. W. Ward. 1980. Recent developments in the retail prices index. *The Statistician* 29 (1): 1–32.
Cecchetti, S. G., H. Genburg, J. Lipsky, and S. Wadhwani. 2000. Asset prices and Central Bank Policy. Paper presented to ICMB International Center for Monetary and Banking Studies and Centre for Economic Policy Research. July, Geneva.
Christensen, L. R., D. W. Jorgenson, and L. J. Lau. 1975. Transcendental logarithmic utility functions. *The American Economic Review* 65 (3): 367–83.
Christensen, L. R., and M. E. Manser. 1976. Cost-of-living indexes and price indexes for U.S. meat and produce. In *Household production and consumption,* studies in income and wealth vol. 40, ed. N. Terleckyj, 399–446. New York: Columbia University Press.
Committee on National Statistics. 2002. See Schultze and Mackie, eds.
Crone, T. M., L. I. Nakamura, and R. Voith. 2004. Measuring American rents: Regression-based estimates. Federal Reserve Bank of Philadelphia Working Paper no. 04-22.
Dalen, J. 2001. *The Swedish Consumer Price Index: A handbook of methods.* Stockholm: Statistics Sweden.
Deaton, A. 1998. Getting prices right: What should be done? *Journal of Economic Perspectives* 12 (1): 37–46.
Deaton, A., and J. Muellbauer. 1980. *Economics and consumer behaviour.* Cambridge: Cambridge University Press.
Diewert, W. E. 1973. Afriat and revealed preference theory. *Review of Economic Studies* 40 (3): 419–26.
———. 1976. Exact and superlative index numbers. *Journal of Econometrics* 4 (2): 115–45.
———. 1978. Superlative index numbers and consistency in aggregation. *Econometrica* 46 (4): 883–900.

———. 1983. The theory of the cost-of-living index and the measurement of welfare changes. In *Price level measurement,* ed. W. E. Diewert and C. Montmarquette, 163–233. Ottawa, Canada: Minister of Supply and Services.

———. 1992. Fisher ideal output, input and productivity indexes revisited. *Journal of Productivity Analysis* 3 (3): 211–48.

———. 1993. The early history of price index research. In *Essays in index number theory,* vol. 1, ed. W. E. Diewert and A. O. Nakamura, 33–66. New York: North Holland.

———. 1995. Axiomatic and economic approaches to elementary price indices. Discussion Paper no. 95-01. Vancouver, Canada: University of British Columbia, Department of Economics.

———. 1998. Index number issues in the Consumer Price Index. *Journal of Economic Perspectives* 12 (1): 47–58.

———. 2001. The Consumer Price Index and index number purpose. *Journal of Economic and Social Measurement* 27: 167–248.

———. 2004. Index number theory and measurement economics. University of British Columbia. Unpublished Manuscript.

Dimand, R. W. 2005. Comments on William D. Nordhaus's "Irving Fisher and the contribution of improved longevity to living standards." *American Journal of Economics and Sociology* 64 (1): 393–97. January.

Ducharme, L. M. 1997. *Bias in the CPI: Experiences from five OECD countries.* Statistics Canada, Prices Division, Analytic Series No. 10. (catalogue No. 62F0014MPB). Ottawa, Canada: Statistics Canada.

Feenstra, R. C., and M. D. Shapiro. 2003. High-frequency substitution and the measurement of price indexes. In *Scanner data and price indexes,* National Bureau of Economic Research studies in income and wealth vol. 64, ed. R. C. Feenstra and M. D. Shapiro, 123–46. Chicago: The University of Chicago Press.

Fisher, I. 1911. *The purchasing power of money.* New York: MacMillan.

———. 1922. *The making of index numbers.* Boston: Houghton Mifflin.

Fisher, F., and K. Shell. 1972. *The economic theory of price indices: Two essays on the effects of taste, quality, and technological change.* New York: Academic Press.

Fixler, D., C. Fortuna, J. Greenlees, and W. Lane. 1999. The use of hedonic regressions to handle quality change: The experience in the US CPI. Paper presented at the fifth meeting of the International Working Group on Price Indices. 25–27 August, Reykjavik, Iceland. Available at: http://www.ottawagroup.org/pdf/bls.pdf.

Frisch, R. 1930. Necessary and sufficient conditions regarding the form of an index number which shall meet certain of Fisher's tests. *Journal of the American Statistical Association* 25 (December): 397–406.

———. 1936. Annual survey of general economic theory: The problem of index numbers. *Econometrica* 4 (January): 1–39.

Gillingham, R. 1974. A conceptual framework for the CPI. In *Proceedings of the American Statistical Association, Proceedings of the Survey Research Methods System.* Alexandria, VA: American Statistical Association.

Gini, C. 1924. Quelques considérations au sujet de la construction des nombres indices des prix et des questions analogues. *Metron* 4 (July): 3–162.

Goldberg, J. P., and W. T. Moye. 1985. *The first hundred years of the Bureau of Labor Statistics.* Washington, DC: GPO.

Goldberger, A. S., and T. Gamelestos. 1970. A cross-country comparison of consumer expenditure patterns. *European Economic Review* 1 (3): 357–400.

Goodhart, C. A. E. 2001. What weight should be given to asset prices in the measurement of inflation? *The Economic Journal* 111 (472): F335–F356.

Gordon, R. J. 2006. The Boskin Commission report: A retrospective one decade later. *International Productivity Monitor* 12 (Spring): 7–22.
Gordon, R. J., and T. van Goethem. 2007. Downward bias in the most important CPI component: The case of rental shelter, 1914–2003. In *Hard-to-measure goods and services: Essays in Honor of Zvi Griliches,* ed. E. R. Berndt and C. R. Hulten, 153–95. Chicago: University of Chicago Press.
Gordon, R. J., and Z. Griliches. 1997. Quality change and new products. *The American Economic Review* 87 (2): 84–88.
Greenlees, J. S. 2001. The U.S. CPI and the cost-of-living objective. Paper presented at the Conference of European Statisticians Joint ECE/ILO Meeting on Consumer Price Indices. 2 November, Washington, DC.
———. 2006. The BLS response to the Boskin Commission report. *International Productivity Monitor* 12 (Spring): 23–41.
Griliches, Z. 1961. Hedonic price indexes for automobiles: An econometric analysis of quality change. Staff Paper no. 3 in U.S. Congress, Joint Economic Committee. *Government Price Statistics, Hearings, Part 2, May 1, 2, 3, 4, and 5, 1961.* Washington, DC: GPO.
Hausman, J. A. 1997. Valuation of new goods under perfect and imperfect competition. In *The economics of new goods,* studies in income and wealth vol. 58, ed. T. F. Bresnahan and R. J. Gordon, 209–48. Chicago: University of Chicago Press.
———. 2003. Sources of bias and solutions to bias in the Consumer Price Index. *Journal of Economic Perspectives* 17 (1): 23–44.
Heien, D. M. 1973. Some further results on the estimation of the s-branch utility tree. Working Paper no. 10. Washington, DC: Bureau of Labor Statistics.
Hendel, I., and A. Nevo. 2002. Sales and consumer inventory. NBER Working Paper no. 9048. Cambridge, MA: National Bureau of Economic Research, July.
Hicks, J. R. 1940. The valuation of the social income. *Economica* 7 (May): 105–24.
Hogg, M. H. 1934. Revising the wage earners' cost-of-living index. *Journal of the American Statistical Association* 29 (March): 120–4.
Hoover, E. D., and M. S. Stotz. 1964. Food distribution changes and the CPI. *Monthly Labor Review* 87 (January): 58–64.
Hulten, C. R. 1973. Divisia index numbers. *Econometrica* 41 (6): 1017–25.
Jaffee, S. A. 1984. The statistical structure of the revised CPI. *Monthly Labor Review* 87 (August): 916–24.
Jorgenson, D. W. 1990. Aggregate consumer behavior and the measurement of social behavior. *Econometrica* 58 (5): 1007–40.
Jorgenson, D. W., and D. T. Slesnick. 1983. Individual and cost-of-living indexes. In *Price level measurement: Proceedings from a conference sponsored by Statistics Canada,* ed. W. E. Diewert and C. Montmarquette, 864–75. Ottawa: Minister of Supply and Services Canada.
Keynes, J. M. 1930. *A treatise on money.* New York: Harcourt, Brace.
Klein, L. R., and H. Rubin. 1947–1948. A constant utility index of the cost-of-living. *Review of Economic Studies* 15 (2): 84–87.
Konüs, A. A. 1924. The problem of the true index of the cost of living. Published in translation in *Econometrica* 7 (January, 1939): 10–29.
Konüs, A. A., and S. S. Byushgens. 1926. On the problem of the purchasing power of money (in Russian). *Voprosi Konyunkturi* II (1): 151–72.
Manser, M. E. 1975. A note on cost-of-living indexes for U.S. food consumption. BLS Working Paper no. 57. Washington, DC: Bureau of Labor Statistics.
———. 1976. Elasticities of demand for food: an analysis using non-additive utility functions allowing for habit formation. *Southern Economic Journal* 43 (July): 879–91.

Manser, M., and R. MacDonald. 1988. An analysis of substitution bias in measuring inflation, 1959–85. *Econometrica* 46 (4): 909–30.

McCarthy, P. J. 1961. Sampling considerations in the construction of price indexes with particular reference to the United States Consumer Price Index. Staff paper no. 4 in U.S. Congress, Joint Economic Committee. *Government Price Statistics, Hearings, Part 2, May 1, 2, 3, 4, and 5, 1961.* Washington, DC: GPO.

Meany, G., and R. J. Thomas, 1944. U.S. Commerce Department Library. *Recommended report for the Presidential Committee on the cost of living index.* Washington, DC: CIO.

Meeker, R. 1919. The possibility of compiling an index of the cost of living. *American Economic Review,* Papers and Proceedings of the Thirty-First Annual Meeting of the American Economic Association 9 (1): 108–17.

Melser, D., and R. Hill. 2005. Developing a methodology for constructing spatial cost of living indexes: A report prepared for Statistics New Zealand. University of New South Wales, Sydney, Australia. Unpublished Manuscript, August.

Mills, F. C., E. W. Bakke, R. Cox, M. G. Reid, T. W. Schultz, and S. S. Stratton. (Special Committee of the American Statistical Association.) 1943. An appraisal of the U.S. Bureau of Labor Statistics cost of living index. *Journal of the American Statistical Association* 38 (December): 387–405.

Mitchell, W. C. 1915. The making and using of index numbers. Part I of *Index numbers of wholesale prices in the United States and foreign countries,* Bulletin of the U.S. Bureau of Labor Statistics, no. 173 (July): 5–114. Washington, DC: GPO.

———. [1915] 1921. The making and using of index numbers. Part I of *Index numbers of wholesale prices in the United States and foreign countries,* Bulletin of the U.S. Bureau of Labor Statistics, no. 284: 7–114. (Revised ed.) Washington, DC: GPO.

Mitchell, W. C., S. Kuznets, and M. G. Reid. 1944. U.S. Commerce Department Library. Prices and the cost of living in wartime—An appraisal of the Bureau of Labor Statistics index of the cost of living in 1941–1944: Report of the technical committee appointed by the Chairman of the President's Committee on the Cost of Living, June 15. Reprinted in *Report of the President's Committee on the Cost of Living.* Washington, DC: GPO.

Moses, K., and B. Moulton. 1997. Addressing the quality change issue in the consumer price index. *Brookings Papers on Economic Activity,* Issue no. 1: 305–49. Washington, DC: Brookings Institution.

Moulton, B. R. 1993. Basic components of the CPI: Estimation of price changes. *Monthly Labor Review* 116 (12): 13–24.

National Research Council. 2002. See Committee on National Statistics 2002.

Ostrander, F. T. 1944. The Mitchell Committee's report on the cost-of-living index: Comments. *The American Economic Review* 34 (4): 849–56.

Pakes, A. 2003. A reconsideration of hedonic price indexes with an application to PCs. *American Economic Review* 93 (5): 1578–96.

Pigou, A. C. 1912. *Wealth and welfare.* London: Macmillan.

———. 1932. *The economics of welfare,* 4th edition. (1st edition: 1920). London: Macmillan.

Pollak, R. A. 1980. Group cost-of-living indexes. *American Economic Review* 70 (2): 273–8.

———. 1981. The social cost-of-living index. *Journal of Public Economics* 15 (2): 311–36.

———. 1983. The theory of the cost-of-living index. In *Price level measurement: Proceedings from a conference sponsored by Statistics Canada,* ed. W. E. Dei-

wert and C. Montmarquette, 87–162. Ottawa: Minister of Supply and Services Canada.
———. 1989. *The theory of the cost-of-living index.* New York: Oxford University Press.
———. 1998. The Consumer Price Index: A research agenda and three proposals. *Journal of Economic Perspectives* 12 (1): 69–78.
President's Council of Economic Advisors. 1999. *Economic Report of the President.* Washington, DC: GPO.
Price Statistics Review Committee. 1961. See Stigler et al. 1961.
Randolph, W. C. 1988. Housing depreciation and aging bias in the Consumer Price Index. *Journal of Business and Economic Statistics* 6 (3): 359–72.
Reinsdorf, M. B. 1993. The effect of outlet price differentials on the U.S. Consumer Price Index. In *Price measurements and their uses,* studies in income and wealth vol. 57, ed. M. Foss, M. Manser, and A. Young, 227–60. Chicago: University of Chicago Press.
———. 1994. The effect of price dispersion on cost of living indexes. *International Economic Review* 35 (1): 137–49.
———. 1998. Formula bias and within-stratum substitution bias in the U.S. CPI. *Review of Economics and Statistics* 80 (2): 175–87.
———. 2006. Axiomatic price index theory. Washington D.C., Bureau of Economic Analysis. Unpublished Manuscript, May.
———. 2007. Axiomatic price index theory. In *Measurement in economics: A handbook,* ed. M. Boumans, 153–88. Amsterdam: Elsevier.
Reinsdorf, M. B., and B. R. Moulton. 1995. The construction of basic components of cost of living indexes. In *The economics of new goods,* ed. T. F. Bresnahan and R. J. Gordon, 397–423. Chicago: University of Chicago Press.
Roe, D., and D. Fenwick. 2004. The new inflation target: The statistical perspective. *Economic Trends* (January): 24–46.
Ross, A. M. 1966. Prepared statement for hearings of the Joint Economic Committee. In U.S. Congress, Joint Economic Committee. Hearings before the Subcommittee on Economic Statistics of the Joint Economic Committee, Congress of the United States. Eighty-Ninth Congress, second session. May 24, 25, and 25. Washington, DC: GPO.
Saglio, A. 1994. Comparative changes in average price and a price index: Two case studies. In *Papers and final report of the first meeting of the International Working Group on Price Indices,* ed. L. M. Ducharme, 197–270. Ottawa: Statistics Canada.
Samuelson, P. A., and S. Swamy. 1974. Invariant economic index numbers and canonical duality: Survey and synthesis. *The American Economic Review* 64 (4): 566–93.
Schultz (Szulc), B. 1994. Choice of price index formula at the micro-aggregation level: Issue, and a Canadian empirical evidence. In *Papers and final report of the first meeting of the International Working Group on Price Indices,* ed. L. M. Ducharme, 93–128. Ottawa: Statistics Canada.
Schultze, C. 2003. The Consumer Price Index: Conceptual issues and practical suggestions. *Journal of economic perspectives* 17 (1): 3–22.
Schultze, C., and C. Mackie, eds. 2002. *At what price? Conceptualizing and measuring cost-of-living and price indexes.* Washington, DC: Panel on Conceptual, Measurement, and Other Statistical Issues in Developing Cost-of-Living Indexes.
Scitovsky, A. A. 1964. An index of the cost of medical care—a proposed new approach. In *The economics of health and medical care,* ed. S. J. Axelrod: 128–42. Ann Arbor: Bureau of Public Health Economics, University of Michigan.

Shapiro, M., and D. Wilcox. 1996. Mismeasurement in the Consumer Price Index: An evaluation. *NBER Macroeconomics Annual,* ed. B. S. Bernanke and J. Rotemberg, 93. Cambridge, MA: MIT Press.

———. 1997. Alternative strategies for aggregating prices in the CPI. *The Federal Reserve Bank of St. Louis Review* 79 (3): 113–25.

Shoemaker, O. J. 2004. Analysis of divergence between chained CPI-U and regular CPI-U for the all-US all-items indexes (2000–2002). BLS Statistical Survey Working Paper. U.S. Department of Labor, Bureau of Labor Statistics.

Staehle, H. 1934. *International comparisons of food costs.* International Labour Office Report, Series N, no. 20.

Statistics Canada. 1995. *The Consumer Price Index reference paper.* Ottawa: Minister of Industry.

Stigler, G., D. S. Brady, E. Denison, I. B. Kravis, P. J. McCarthy, A. Rees, R. Ruggles, and B. C. Swerling. 1961. *The price statistics of the federal government.* In U.S. Congress, Joint Economic Committee. *Government Price Statistics, Hearings, Part 1, May 1, 2, 3, 4, and 5, 1961.* Washington, DC: GPO.

Stone, J. R. N. 1954. *The measurement of consumers' expenditure and behaviour in the United Kingdom, 1920–1938.* London: Cambridge University Press.

Stone, R., and S. J. Prais. 1952. Systems of aggregative index numbers and their compatibility. *The Economic Journal* 62 (247): 565–83.

Szulc [Schultz], B. J. 1983. Linking price index numbers. In *Price level measurement: Proceedings sponsored by Statistics Canada,* ed. W. E. Diewert and C. Montmarquette, 537–66. Ottawa: Minister of Supply and Services Canada.

Triplett, J. E. 1983. Escalation measures: What is the answer? What is the question? In *Price level measurement: Proceedings from a conference sponsored by Statistics Canada,* ed. W. E. Diewert and C. Montmarquette, 457–82. Ottawa: Minister of Supply and Services Canada.

———. 1990. Hedonic methods in statistical agency environments: An intellectual biopsy. In *Fifty years of economic measurement: The jubilee of the Conference on Research in Income and Wealth,* studies in income and wealth vol. 54, ed. E. R. Berndt and J. E. Triplett, 207–33. Chicago: University of Chicago Press.

———. 2001. Should the cost-of-living index provide the conceptual framework for a Consumer Price Index? *The Economic Journal* 111 (472): F311–F334.

———. 2003. Using scanner data in the Consumer Price Indexes: Some neglected conceptual considerations. In *Scanner data and price indexes,* studies in income and wealth vol. 64, ed. R. C. Feenstra and M. D. Shapiro, 151–62. Chicago: University of Chicago Press.

———. 2006. The Boskin Commission report after a decade. *International Productivity Monitor* 12 (Spring): 42–60.

Triplett, J. E., and B. P. Bosworth. 2004. *Services productivity in the United States: New sources of economic growth.* Washington, DC: Brookings Institution Press.

Ulmer, M. J. 1946. On the economic theory of Cost-of-Living Index numbers. *Journal of the American Statistical Association* (December): 530–42.

Urquhart, M. C., and K. A. H. Buckley. 1965. *Historical indexes of Canada.* Toronto: McMillan.

U.S. Congress, Joint Economic Committee. 1958. *The relationship of prices to economic stability and growth: Compendium of papers submitted by panelists appearing before the Joint Economic Committee.* Joint Committee Print, March 31. Washington, DC: GPO.

———. 1961. *Government price statistics: Hearings before the Joint Economic Committee, Congress of the United States.* Eighty-seventh Congress, first session pursu-

ant to Section 5(a) of Public Law 304 (709th Congress). Part 2. Washington, DC: United States Congress (May 1, 2, 3, 4, and 5).
U.S. Department of Labor, Bureau of Labor Statistics. 1966. The Consumer Price Index: History and techniques. Bulletin no. 1517. Washington, DC: GPO.
———. 1997a. Measurement Issues in the Consumer Price Index. *Response to the U.S. Congress,* Joint Economic Committee, June. Washington, DC: GPO.
———. 1997b. The Consumer Price Index. *BLS Handbook for Methods for Surveys and Studies* (April): 167–202.
Varian, H. R. 1982. The nonparametric approach to demand analysis. *Econometrica* 50 (4): 945–74.
———. 1983. Non-parametric tests of consumer behaviour. *The Review of Economic Studies* 50 (1): 99–110.
von Bortkiewicz, L. 1923. Zweck und Struktur einer preisindexzahl. *Nordisk Statistiek Tidschrift* 2:3–4.
von Haberler, G. 1927. *The meaning of index numbers.* Munich: J. C. B. Mohr.
Walsh, C. M. 1901. *The measurement of general exchange value.* New York: Macmillan.
*Washington Post.* 1946. Editorial. February 14.