

DRAFT—Work in progress (this draft: June 25, 2004)

(Section III to be added)

**A Review of Reviews:
Ninety Years of Professional Thinking About the Consumer Price
Index and How to Measure It**

by

Marshall Reinsdorf
US Bureau Of Economic Analysis
1441 L St, NW, Washington DC 20230
202-606-9665
marshall.reinsdorf@bea.gov

and

Jack E. Triplett
Brookings Institution
Washington, D.C. 20038
JTRIPLETT@brookings.edu

June 2004

The views expressed should not be attributed to the Bureau of Economic Analysis, nor to the Brookings Institution, as they are solely those of the authors.

“It is not often that a price index, a tool of statisticians, becomes an object of political debate.”
Ostrander (*American Economic Review*, 1944, page 849).

Introduction

The U.S. Consumer Price Index was first compiled and published in 1919 (though it was not then called the CPI). It originated in the Shipbuilding Labor Adjustment Board’s decision during World War I to escalate wages by a price index (**BLS, xxxx**; Goldberg and Moye, 1985). The U.K. Retail Price Index, similarly, began as a wage escalator during the same war (**ref**), as did the Canadian CPI (like the U.S. one, originally called the Cost of Living Index—see Statistics Canada, 1995, pp. 9-10; Urquhart and Buckley, 1965, pp. 287-89).

The history of economists’ formal analysis of CPI measurement issues extends as far as their published history. Indeed, awareness in the U.S. of the distinction between consumer prices and what we now call the Producer Price Index seemingly arose from a review of the PPI (then called the Wholesale Price Index, or WPI) by Wesley Mitchell in 1915 (Mitchell, 1915, 1921). As summarized by Banzhaf (2001), Mitchell pointed out that the WPI was not the appropriate index to “measure the general purchasing power of money for consumers.” Distinguishing consumer prices may seem obvious now. But as we point out in section I, Mitchell’s distinction was not made in most of the history of index numbers before that time, partly because their implicit framework revolved around the old quantity theory equation, $MV=PT$, in which the “price level” for the economy was the objective, not the purchasing power of consumers, whose transactions, after all, amounted to only a portion of the vector T in the quantity theory equation.

Over the years, reviews by committees and panels have critically influenced the development of the CPI, and the long record of the debate over CPI concepts and methods preserved in their reports provides an important background for understanding the current state

of the discussion. This paper examines the treatment of three questions, one conceptual, one methodological, and one that is both, in officially sanctioned reviews of the CPI. The conceptual question of interest is: What measurement objective is the appropriate one for the CPI? 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. 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? Finally, we consider a major question that involves both concepts and methods: What are the criteria and appropriate techniques for quality adjustment?

Each section of the paper deals with one of the three questions and proceeds chronologically. After the review of the history of the debate on each question, in the last part of each section we provide our own assessments.

Though our topic begins with Mitchell (1915) and extends through thinking about CPIs in the 1920s and 1930s, our reviews begin with the World War II era reviews on BLS’s Cost of Living Index, which were major milestones in the history of thinking about CPI measurement issues. There were actually four reviews, though all were part of the same investigation. The wartime debate resulted in a change of name: The “Cost of Living Index” became the “Consumer Price Index” in 1945, but the name change has frequently been mis-interpreted. It was not at all rejection of the modern concept of the cost-of-living index (COLI), but rather an attempt to overcome another kind of confusion about index objectives.

The wartime criticisms of the procedures used to construct the detailed component indexes of the CPI laid the foundation for the reforms in sampling procedures adopted by BLS in its implementation of the recommendations of a 1961 report by a committee chaired by George

Stigler (Price Statistics Review Committee, 1961), and also for the Stigler Committee's heightened concern for the problems posed by quality change. In turn, the Stigler Committee's report influenced the report by what became known as the Boskin Commission, 1996 (Advisory Commission to Study the Consumer Price Index), and the Committee on National Statistics (hereafter, CNSTAT—Schultze and Mackie, eds, 2002) report was funded by BLS in response to the Boskin Commission review. The continuity in the history of the reviews is echoed in continuity among the reviewers, for typically one or more of the participants in one review also participated in the one that followed.

We put more emphasis on the role of Wesley Mitchell than has been true of index number history in the past. Mitchell's contributions to price index history have greatly been neglected, and only recently resurrected in the insightful paper by Banzhaf (2001). For example, Diewert (1993) in his history of price index research does not even mention Mitchell. Mitchell's role in focusing attention on the problem of estimating consumer prices has already been mentioned,¹ and 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, sometimes apparently without a clear sense of what was meant. Mitchell was the chair of the main professional wartime committee on the CPI, which reflects his reputation at the time on the topic of price indexes.

Additionally, Banzhaf (2001) discussed Mitchell's correspondence with Irving Fisher on the subject of price indexes. This discussion made us aware of how much the procedures followed by BLS well into the 1960's and 1970's were based on Mitchell's writings on price

¹ Though Mitchell was not the first to recognize that the design of an index number depends on its purpose. Diewert (1993, page 34) cites Joseph Lowe as the "father of the consumer price index." Lowe explored a number of problems, including indexes for different demographic groups.

indexes. Mitchell was perhaps the first authority on price indexes to give prominence to measuring the components, as opposed merely to the methods for aggregating them. Most of our own price index research centers on these same matters, so we are indebted to Banzhaf for pointing us to Mitchell's contribution.

We suspect that the main reason for Mitchell's neglect is the over-emphasis placed by so many index number authorities on determining the functional form of the index number, and on commodity substitution bias—that is, the problem of how one should *aggregate* indexes for apples, oranges, and bananas, and so forth, once these “basic components” are estimated. Though we agree that aggregation and commodity substitution bias are important matters, they are by no means the only important empirical matters for price measurement, and we contend very likely not the most important ones, empirically. Hausman (2003) remarks that aggregate commodity substitution bias is a “second order” problem in measuring price indexes and that the first order problem is measuring the impacts of new goods and quality change, to which we would add the matter of estimating the lower-order aggregates, or “basic components” of the price index. We think Mitchell would have agreed with Hausman, as do we.

I. Recommendations Concerning the Measurement Concept for the Index

“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.” — Stigler Committee, 1961, p. 51.

As noted earlier, the U.S. CPI was influenced by Mitchell’s review of the Wholesale Price Index, which he called “The Making and Use of Index Numbers” (Mitchell, 1915). The title suggests that this WPI review was more than just a review of the WPI. When a household inflation measure was launched, in 1919, it was named the “Cost of Living Index.”

The first professional reviews of the U.S. household inflation index occurred during the Second World War. We discuss them in section I.B. However, the interwar period was not devoid of thinking about the CPI, quite the contrary. Much of the modern understanding of CPI conceptual questions and aggregation was developed in the 1920’s and 1930’s—we have in mind the theory of the COL index, the “test approach” to index number formulas, as well as the Divisia approach, all of which still occupy a great amount of attention from index number theorists. These subjects are well known and have been reviewed many times previously, and hardly merit reviewing again. But in some cases their interpretations are still both controversial and important in assessing subsequent formal CPI reviews, hence, we present a brief overview in section A. Our approach is not always the standard interpretation, and we use our interpretation to review the reviews, so stating our interpretation at the outset is essential to what follows.

A. Price Index Research in the Interwar Period.

Three major developments in the 1920s influenced professional thinking about CPIs and were therefore absorbed into the intellectual tradition that was incorporated into subsequent reviews of the CPI. The first, Fisher's development of what has come to be known as the "test approach," originated earlier than the 1920s (indeed it has a long set of antecedents), but came to fruition in that decade. Nothing in Fisher's (1922) approach was specific to consumer price indexes, it applied to all price indexes. So far as we know, there is no evidence that Fisher thought much about specific applications, such as to consumer inflation.

The second development was Konus (1925) development of the theory of the cost of living index (COLI). So far as we know, Konus was totally original and without precedent.² In contrast with Fisher and his forebears (including Walsh?), Konus' contribution was uniquely a contribution to the measurement of consumer prices, and it was undoubtedly a response to those new "cost of living indexes" in the U.S., Canada, U.K. (and other countries? Sweden). The third development was the Divisia index, which contributes little if anything to questions about measuring the CPI.³

The test (Fisher) and COLI (Konus) approaches to index numbers are very different in their economics and in their implications about what is being measured. For Fisher, the price level was 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, or the price level is like an exploding shell where one ignores the paths of the fragments to obtain the true trajectory. The form of the index number is determined by tests on the reasonableness of the measured P, which Fisher conceived as a problem that is parallel with measuring the level of the lake or the trajectory of the shell.

² Some others may have invented it independently, see Bowley, Staele, Allen, Haberler.

³ On the Divisia index, see Hulten (1973) and Reinsdorf (200x).

The purpose of the price index measurement was, for Fisher, to obtain a P for the quantity theory expression $MV=PT$. The quantity theory was at the time the only systematic economic theory for which a price *index* was relevant. In the quantity theory, of course, PT must refer to all of the transactions in the economy. PT for consumer prices is only relevant as a portion of the total PT that is wanted for that purpose, in the framework of the time. Moreover, the M and V on the left-hand side and the T on the right condition one to think of the P, analogously, as a scalar, as Fisher's physical analogies also suggested. In Fisher's view, aggregation never came into the matter.

In contrast, the COLI theory of Konus is a theory of aggregation. Indeed, Konus is rightly seen as the originator of the economic theory of aggregation, not just as the originator of the economic theory of price indexes. The theory yields the conclusion that aggregating individual price indexes (index basic components) requires an aggregator function that is derived from the form of the utility function (strictly, the form of the indirect utility function). Pollak (1983, 1989) and also Diewert (1993) provide surveys.

Of course, the form of utility function is not generally known. But Byushgens (1925) showed that the indirect utility function that corresponds to Fisher's "ideal" index (the geometric mean of the Laspeyres and Paasche indexes) provides a second order approximation to the unknown true utility function and thus can be used to approximate the COLI index. The very clear exposition of this point in Schultze and Mackie, eds. 2002, page 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 with 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 the demand functions that do not restrict *substitution* behavior in ways beyond that required for the theory. It therefore permits a way of computing a general cost-of-living index without having to estimate the demand functions."

As noted by Schultze and Mackie, eds. (2002), Diewert (1976) generalized Byushgens' result by showing that there are other utility functions, and therefore other index numbers, that are also second order approximations. Before Diewert, these results were completely unknown among researchers and index number theorists.

It has sometimes been said that test and COLI approaches converge: COLI theory shows that the Fisher index is a good approximation to the true aggregator function, which merges with Fisher's conclusions on the basis of his tests. The conclusion, however, is too superficial and too narrow, for two reasons.

First, Fisher's own tests were biased. He excluded possible tests that many index users deem important and that would have overturned his result.

Second, the COLI framework shows not just the way to aggregate (that is, the 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 that will be aggregated with the index number formula. The test approach is completely silent on these matters.

Selection of Tests. Fisher proposed 12 index number tests. He designated what has come to be known as the "Fisher ideal" index as the superior index because the Fisher index satisfied the largest number of tests.

Frisch (1936) pointed out that no index number would satisfy all tests. A far more important criticism is that there is no economic theory to explain why this particular series of 12 tests are the relevant ones, and no theory that says that the tests are to be weighted equally.⁴

This last point is particularly damaging to the test approach since Fisher's tests do not include some index number properties—notably additivity in National Accounts and in CPI's—

⁴ In practice, adherents to the test approach have sometimes made judgments about which of the tests are important or not important.

that some users contend are the most important properties of index numbers. In other words, the Fisher tests are biased.

If one thought that additivity in National Accounts is the most important property of index numbers (in National Accounts it is the numerator of the real output index where additivity is desired), then the Laspeyres index satisfies the relevant test and the Fisher index does not. Indeed, the Fisher index is not only not additive it is not consistent in aggregation—a Fisher index of component Fisher indexes is not an aggregate Fisher index. Neither additivity nor consistency in aggregation are among Fisher’s tests (no doubt because he did not even think of an index number as an aggregation). Since there is no theory for defining the tests and no theory for evaluating them, the tests in Fisher (1922) are arbitrary. No basis exists for excluding additivity and consistency in aggregation from the “reasonableness” tests that index numbers should meet.

Moreover, simply inserting additivity or consistency in aggregation into the list of tests solves nothing. Many users of national accounts (in the U.S. and elsewhere) have maintained that additivity is the most essential property of the index number for their purposes. Equally weighting the tests and simply counting the number of tests “passed” is as fallacious, from their point of view, as omitting their favorite test in the first place. If users insist that additivity is the most important “axiom” for an index number, there is nothing in the test approach for evaluating their position, except to reject it arbitrarily. Fisher’s tests are determined by what he deemed “reasonable,” but reasonableness is defined differently by different people.

For the test approach to have validity, the tests themselves and their weighting must be derived from some sort of theory. They are not. Under this critique, both the test approach itself and its conclusions that the Fisher index is the preferred index number cannot be supported.

Once one asks whether tests can be supported by economic theory, the whole approach breaks down. For example, some tests are inconsistent with economic theory. The Fisher “time reversal” test requires that the index number be invariant to the choice of base. It is well known from consumer demand theory that whenever the utility function is not homothetic, income and substitution effects calculated with the initial indifference curve as the starting point will not be equal to income and substitution effects calculated the other way. The substitution effect, of course, is the same substitution effect that is at the center of COLI index aggregation arguments. The difference between equalizing and compensating variations in public finance and tax policy is exactly equivalent to the difference between cost of living indexes and standard of living indexes calculated from the base period indifference curve or the comparison period indifference curve. Samuelson and Swamy (1974) called the search for a cost of living index that was invariant to the choice of indifference curves as a quest for something that was not obtainable even in “Plato’s Heaven.”⁵

The test approach, accordingly, does not yield the Fisher index number, *unless one selects the tests in order to get it*. The test approach does not yield the same superlative index result as the economic approach to index numbers.

Significantly, no U.S. review (with the exception of brief passages in the CNSTAT Panel’s report) puts any weight on test approaches to determining the index number formula. Test approaches remain more popular in Europe.

Theory for Measuring Index Components. The economic and the test approaches to index numbers yield different results in a more fundamental dimension. The theory of the COLI

⁵ Diewert, as we understand it, has contended that one can solve this problem by choosing an intermediate point midway between the initial indifference curve and the end period indifference curve. Logically, the intermediate point is irrelevant to the question: What is the cost in today’s prices of achieving the base period’s standard of living? We agree with Samuelson and Swamy on this point.

has two parts. One part, as noted, is a theory of aggregation—that is, selecting the index number “formula.” The second part of the COLI approach concerns what is to go into the index and how those individual components of the index are to be measured.

The cost of living index provides the answer to the question: What is the cost in today’s prices of obtaining the standard of living consumed in 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 points to the standard of living, and invites us to ask how that should be measured.

To take an example that is current 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? As another example, faced by at least one of the reviews covered in this paper, forced substitutions (as, for example, a regulatory change that removes some variety of product from the market) and voluntary substitutions are all changes in the “market basket;” should these be handled in the same way in the price index? Or not? COLI theory forces us to confront those questions, and suggests some answers, at least in quantitative terms. The test approach is completely silent on them.

It is striking that in the price index literature up to around 1970, the Fisher test approach was probably invoked more often than the COLI approach. Allen (1975) takes the Laspeyres index as the basic starting point (similar to the CNSTAT Panel’s COGI), though perhaps for expositional reasons, and adds the COLI approach later. Yet, no CPI review committee during this period took the Fisher index as relevant. No doubt one reason was the fact that the Fisher index requires current period weights, which are generally not available for CPI purposes, at least in the U.S. Another is that it was not generally known, until Diewert (1976), that the Fisher

index had a relation to a COLI, and it was not unusual during this interval to read some statement that the Fisher index was theoretically suspect.

B. Wartime Committees: Clarifying the Scope of a 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, page 850). 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. Their position ultimately led to the preparation of four different reviews of the index: The Labor Department initiated the first, 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 AFL and R.J. Thomas of the CIO (the Meany-Thomas report), a business report issued by the National Industrial Conference Board⁷, and finally, a report by a “Technical Committee,” which was chaired by Mitchell. The staff and members of the Mills and Mitchell committees is a veritable list of prominent economists and statisticians of the time.⁸

⁶ *Washington Post* editorial, February 14, 1946 (in US Commerce Department Library, 1944.) See also Goldberg and Moye (1985, p. 154), Banzhaf (2001, p. 354), and Mills, *et al.*, (1943.)

⁷ Now the Conference Board. At the time and for years 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 Brady and Fabricant again on the staff [check: someone else?]. Brady and Reid, both of whom later became prominent academic researchers on the subject of consumer behavior, had both been members of the BLS staff in the 1930’s [check dates, and confirm Reid??].

The wartime dispute is unique among political discussions of consumer prices in three aspects. (1) The topic was alleged understatement of inflation, not overstatement, as was true of all subsequent reviews. (2) The unions appeared (though the language is not entirely precise) to support the concept of the COL index, unlike their position subsequently. (3) The professional reviews (the Mills and Mitchell reports) supported, perhaps too uncritically it seems to us, BLS methodology—subsequent professional reviews of the CPI range from mildly to overwhelmingly critical.

The Meany-Thomas report set forth the specifics of the unions' criticisms of the index. Examining the period from January 1941 to December 1943, they estimated that the true rise in the cost of living was 43.5 percent, compared with only 23.4 percent reported by the BLS (Meany and Thomas, 1944). The unions gathered some of their own data; the report's empirical sections are considered in later sections, particular in section II.

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 to more expensive dwellings than they wanted because of shortages of affordable housing. These involuntary substitutions raised consumers' cost of living in ways that BLS's index missed. In addition, lower-quality varieties were often substituted for higher-quality ones that they had replaced 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 rationing⁹ and the entrance of women into the labor force,¹⁰ and extra costs from migration

⁹ We presume the assertion is correct, though it implies that rationing could be evaded by eating out, an option that must certainly have been more accessible to higher income families than to those of workers.

between cities to fill wartime jobs. They did not include effects of rising income or standards of living as additional costs to be included in the cost of living index in their report, but — judging from the rebuttals by their opponents — statements they made elsewhere gave this impression.

Both the BLS and outside experts vigorously disputed the Meany-Thomas report's contention that the Cost of Living Index understated inflation by almost half. 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 COL index was known among economists (see section A), the term "cost of living index" did not necessarily have the same meaning, generally, in 1944 that it has assumed in later years. As Banzhaf (2001) points out, the term was then used, or interpreted, in at least three different ways.

- (1) The now-standard meaning—a price index that holds constant the standard of living. The committee referred to this as a measure of the "real price."
- (2) A fixed basket index, which is what the BLS was in fact producing under the "cost of living index" name.
- (3) The cost of attaining a standard of living deemed appropriate, compared to the cost of a possibly lower appropriate standard of living in some previous period.

The Mitchell committee introduced utility theory in a limited way as a basis for thinking about the design of the Cost of Living Index in its discussion of item substitutions involving a change in quality. When item disappearances were thought to reflect voluntary substitution behavior by consumers, BLS usually introduced substituted items into the index via overlap price linking. The Mitchell committee argued that this procedure is justifiable under the assumption that the relative utility of different qualities varies directly with the ratio of their prices, which requires consumers to be informed about quality and supplies to be freely available (Mitchell et al, 1944, p. II-12). If, for example, the substituted variety had a lower price, the larger quantity

¹⁰ A bit of conflict here, perhaps. Accounting for the value of home production that was lost does not seem to be the issue, only the increased (monetary) costs of substituting market purchased commodities for home production.

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” (p. 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” (p. 14). The Meany-Thomas “life style changes” were not to be considered. 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.¹¹

To clarify that non-price influences on welfare were out of scope, and—possibly more importantly, considering the confusion of the time—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 (p. 20.)¹² 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 now it 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.¹³

¹¹ Some of the commentary on the debate considered whether the evidence suggested that worker real consumption levels had declines. See Ostrander (1944).

¹² It 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.”

¹³ In the debate over the later Boskin Commission (1996) recommendations, certain BLS statements about the 1946 change in name revealed confusion about the episode. The inadequate discussion in Goldberg and Moye (1985) is no doubt part the source of this confusion.

The Mitchell committee (Mitchell et al., 1944) found 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 50 years later in performing a “guestimate” 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 ½ point possible from the omission of smaller cities from the index. 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.

Though Meany and Thomas were not uninterested price collectors and their evidence was anecdotal, some of their exhibits were intriguing. For example, they presented menus from cafeterias showing increases in standard meals, which seemed greater than the comparable component of the index, and they showed (with a picture) 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) contained disproportionately items that were under price controls, when similar items that experienced higher price changes were omitted from the index. It is hard to avoid, from the vantage point of 60 years, some sense that their evidence was dismissed too readily. Whatever one’s judgment about the quantitative significance of their evidence, their examples certainly pointed to measurement problems in the CPI that would be taken seriously in the future.

C. The Stigler Committee: A Welfare Index as the Measurement Concept for the CPI

In 1957 the Joint Economic Committee 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 theme in the subsequent hearings. Notable was a paper by Kenneth Arrow arguing for a cost of living index objective for the CPI because of the importance of commodity substitution behavior (published in: U.S. Congress, Joint Economic Committee, 1958, p. XI).

Subsequently, the 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 [and someone else?], who had participated in the Mitchell Committee investigation 15 years before.¹⁴ The committee’s report (Stigler, et al. 1961, hereafter: “Stigler committee report”), accompanied by twelve “staff papers,” was transmitted to the Bureau of the Budget in 1960.¹⁵

Unlike the Mitchell committee, the Stigler committee unequivocally stated 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, p. 52). The committee noted that in wartime, when consumers were driven to substitutes that were more expensive or that were rising rapidly in price, few came to the defense of the rigidly fixed market basket, which suggested the cost of maintaining a constant level of utility was the desired measure. The Stigler committee also used the term “welfare index” and “welfare oriented index [check],” defined again as an index that tracks the cost of maintaining a constant level of utility, though interestingly it never used the term “cost-of-living index,” perhaps in response to the wartime confusion over that name.¹⁶

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).

Secondly, however, if one thinks about the aggregation part of COL index theory, no proposal for estimating commodity substitution behavior existed in 1944; the Mitchell committee

¹⁴ List members, and some? Staff paper authors.

¹⁵ The Stigler Committee report can be hard to access. The report and staff papers were published as part of the record of subcommittee hearings held in January 1961 (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, and included 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 were also published by the NBER (Price Statistics Review Committee, NBER, 1961).

¹⁶ At one point, in response to a question at the JEC, a member of the Stigler committee referred to “what Commissioner Clague calls a cost-of-living index.”

noted the [quote]. By 1960 there was at least a proposal for how one might actually estimate commodity substitution. Four years after the Mitchell committee report, Klein and Rubin (1947-48) published what came to be the “linear expenditure system,” a method for estimating systems of consumer demand functions. [description and implications to be added] They proposed explicitly to use their system to estimate a COLI. Thus, that part of the Konus agenda seemed now a practical research issue, where in 1944 its practicality was inconceivable.

Whereas neither the Meany and Thomas report nor the Mitchell committee viewed voluntary substitutions by consumers as a factor causing bias in the official Cost of Living index, the Stigler committee 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.” (p. 52)

To gauge the potential size of the bias from substitution, the Stigler committee recommended that the BLS periodically estimate a Paasche index version of the CPI. In addition, to reduce this bias, the committee recommended “possibly” more frequent updating of the weights.¹⁷ It did not recommend, nor did it discuss, changing the index formula from its historical Laspeyres.

The Stigler committee viewed the COL index framework as pertaining to far more than just commodity substitution and index aggregation formulas. The committee discussed a number of other implications of the COLI as the measurement concept for the CPI. Though some of these properly belong in sections II and III, they also illustrate the point we made in section A: The COLI framework influences many parts of the price index and many decisions that must be made in compiling it.

¹⁷ In 1960, the BLS was still using 1951-52 weights and those were from an “interim” revision of the weights from the consumer expenditure survey of 1948 [check dates]. The committee’s recommendation seems timid, from today’s perspective.

Notable among the Stigler committee's examples were the effects of new products, the effects of changes in quality, and the treatment of consumer durables. It viewed consumer preferences as key to evaluating the effect 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.” (p. 52).

To minimize the bias from neglect of gains 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 has 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.¹⁸

The problem of changes in the quality of 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. Similarly, in response to a question from Senator Paul 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 medical ailment such as appendicitis (p. 533).¹⁹ A staff paper by Griliches included with the report investigated the use of hedonic regression for quality adjustment purposes, a method the committee viewed as quite promising. Griliches' paper proved to be one of the most widely cited contributions of the report.

Last, for durable goods, the Stigler Committee noted that consumers' welfare depends on the flow of services from the durable, not its value at the time of acquisition. Therefore, the cost

¹⁸ 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 “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 for new varieties of goods predominates, empirically.

¹⁹ Stigler noted a study, though he cited no source. Evidently, the committee was apprised of a preliminary version of the work of Scitovsky (1964).

of the use of the good is theoretically the correct concept for 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:

- "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.²⁰
- "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 he refused, saying that the committee did not know enough to make an estimate. In this the Stigler Committee differed from the Mitchell Committee and the following Boskin Commission, both of which made guestimates of CPI bias, though of course the Mitchell Committee's was not necessarily an estimate of bias with respect to a COLI. Richard Ruggles, however, inserted a footnote which implied an estimate of three percentage points annual bias, but this was never endorsed by the rest of the committee, nor by Stigler himself.

The initial reaction of the BLS to the Committee's COLI recommendation was a very negative one, based partly on the difficulty of estimating a COLI, but probably the BLS would have opposed the recommendation even if it were thought to entail changes that were easy to make.²¹ Commissioner Ewan Clague testified to a subcommittee of the Joint Economic Committee in 1961 that:²²

²⁰ 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 updating of weights.

²¹ Some of the following discussion parallels Greenlees (2001).

²² U.S. Congress (1961), p. 560.

“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.”

Other comments at the Hearings on the Committee’s report were also negative. Even Douglas (a co-inventor, after all, of the Cobb-Douglas model of producer substitution behavior that was extended by Klein and Rubin to the consumer case) seemed skeptical. “I remember those words of Browning” he remarked at the hearings: “[complete quote]” And the BLS position in 1961 on the COLI is not so different, in some respects, from Chapter 2 of Schultze and Mackie, eds (2002).

Despite BLS’s initial opposition to the COL index concept, the tide changed in the 1960’s. Part of the reason for change was change in people. One of the Stigler Committee’s recommendations was the proposal to 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 JEC many restrictions on the applicability of economic theory for cost-of-living measurement but noted²³ “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.”

Later statements about the CPI by BLS officials continued to combine sometimes oblique references to the COLI measurement objective with caveats noting the obstacles to achieving that objective. Statements that appeared in the *BLS Handbook of Methods* until its 197? edition said that the CPI was not a COLI. However, a 1974 paper by BLS economist Robert Gillingham

²³ Ross (1966).

laid out the conceptual framework that was adopted for the 1978 revision to the CPI.²⁴ That paper, which focuses on Pollak's concepts of partial and conditional COLI subindexes, states that BLS "... assumed that the primary purpose of the [CPI] is to approximate changes in the cost of living of consumers."²⁵ Moreover, using language that has been virtually unchanged since 1992 (1984?), the current *BLS Handbook of Methods* states, more positively:²⁶

"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 soon became that (i) the COLI objective provides the framework for the CPI, but (ii) 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 key recommendations of the Stigler Committee. Most notably, it instituted a system of probability sampling and it eventually (after a long public battle) changed the measure of owner-occupied housing to a rental equivalence measure. Yet progress on other recommendations aimed primarily at bringing the CPI into closer alignment with a COLI or enabling the CPI to be compared to a COLI retrospectively was still not far along at the time of the subsequent Boskin commission. 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, a retrospective index that provided direct evidence on substitution bias in the

²⁴ The manager of the 1978 revision was John Layng, who later became head of the BLS Office of Prices. 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 next section). See also Greenlees (2001) for a similar account.

²⁵ Gillingham (1974), p. 246.

²⁶ Bureau of Labor Statistics (1997a), p. 170.

CPI was not published until 1993, and use of hedonic regression results was limited to aiding in the identification of substitutions that could be treated as comparable. Even 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 cost of living indexes and on hedonic indexes was conducted within BLS starting in the late 1960s and continuing to this day (the hedonics research was reviewed in Triplett, 1990.)

BLS Empirical Research on Estimating Substitution Bias. The BLS was openly hostile to the Stigler Committee's recommendation on the COLI and to its suggestion that substitution bias should be considered seriously among price index issues. In fairness to BLS, the Stigler Committee's report and the subsequent hearings contained little on how one might estimate substitution bias. Klein and Rubin (1948) was not in the references, for example, and estimating substitution bias was the subject of none of the report's 12 staff papers. Indeed, though the Klein-Rubin article was concerned with estimating a COLI empirically and estimating the substitution bias, in 1961 no empirical estimate of substitution bias using the demand systems approach had been made. For the BLS staff, it must have seemed as if they were being told to do what most of them thought was impossible. In recognition of this, no doubt, the committee's only concrete recommendation on substitution bias was to ask that BLS calculate a Paasche Index to provide a rough estimate of the size of the bias.

Though with a lag of nearly ten years, BLS did respond to the Stigler Committee recommendations with an extensive research program on the COLI, beginning about 1970. The Stigler Committee, as noted earlier, had recommended that the BLS set up a price research unit, and this became fully operational in 1968, under the direction of Joel Popkin.

Part of the research money was used to fund academic visitors. The most valuable contributor was Robert Pollak, who summarized the existing theory of the cost of living index and extended that in a number of directions. These contributions formed chapters of Pollak (1989). In addition to the work of Franklin Fisher and Karl Shell (1972), Pollak's work for the BLS was the major theoretical work on COLI during the period of the 1970s and 1980s.

For empirical work, BLS research strategy followed the Klien-Rubin (1948) lead. Improved computer capability finally made estimates of the Klien-Rubin system practical, one of the earliest estimates (outside BLS) being Goldberger and Gameletsos (1970). BLS work concentrated on finding improved specifications for systems of consumer demand functions (Brown and Heien, 1972; Heien, 1973; Christensen and Manser, 1976; Braithwait, 1980). Another thrust was finding ways to increase the number of demand relations that were included in the system, and to explore issues such as homotheticity and other specification issues.

BLS empirical estimates of the substitution bias using systems of demand equations include Christensen and Manser (1976), and Manser (1975;1976), who studied detailed food categories, and Braithwait (1980), who treated food as an aggregate but included 52 other commodity groups, using data from National Accounts. Taken together, the Christensen-Manser, Manser and Braithwait studies covered the entire consumers budget at approximately a 63-commodity level of detail. This is a far greater of detail that had been estimated by others (earlier studies, for example, used broad aggregations, durables, non-durables and services, where the BLS estimated substitution among categories such as "beef, poultry, poultry, pork and fish").

Taken together, the BLS studies suggested that substitution bias was in the order of 0.1 index points per year. This was a substantially lower number than economists had typically speculated, earlier.

The research was subject to three criticisms: First, even 63 commodity groupings may be too coarse to capture the real substitution behavior that needs to be modeled. Second, though a range of consumer demand systems were employed in the research (each study estimated substitution bias with several demand systems, not just one), one could still speculate that some other consumer demand system would yield a greater estimate of substitution bias. The only answer to that criticism appeared to be to try more systems, but even with improved computer capability, these estimates were neither easy to carry out, nor quick to complete. Third, aggregate demand studies imply that the problem of aggregating over households is solved, so assuming away one of the major problems with COLI theory.

However, these BLS studies also produced a remarkable result: Estimates of substitution bias were remarkably robust across different specifications of consumer demand systems. Put in another way, the empirical estimates were not very sensitive to the underlying utility functions assumed in these econometric estimations. It is true that the roughly 8-10 utility specifications employed in the BLS studies covered perhaps only a limited range of possibilities (they employed all of the major ones that existed at the time, with one major exception that was developed too late for inclusion in the research design). But still, this empirical result was surprising and seems not to have informed some subsequent writing about substitution bias.

In the middle of the BLS research on substitution bias, the influential paper by Diewert (1976) appeared. Diewert showed, following Byushgens (1925), that one could obtain a second order approximation to the substitution bias using what he called “superlative” index numbers,

two of which are the Fisher index and the Tornqvist. The Tornqvist index is an exact COLI for the homothetic translog indirect utility function, which was included among the utility functions for which COLI estimates were prepared by Christensen and Manser (1976).

Combining results from Diewert (1976) and Christensen and Manser (1976) has an implication that has perhaps not been emphasized sufficiently: If the homothetic translog indirect utility function, and its associated Tornqvist COLI, gives a good approximation to the unknown true utility function (Diewert result), and if the COLI from the homothetic translog is empirically close to the COLI estimates from other utility functions (as it was found to be), then further exploration of modified consumer demand systems is not likely to be fruitful empirically.

There is a converse implication that pertains directly to the CNSTAT Panel's concerns: Both the Tornqvist and the Fisher indexes imply that Diewert's second order approximation is a homothetic approximation. The BLS empirical research suggests that homotheticity, though a very poor specification of consumer demand behavior, does not much influence econometric estimates of the substitution bias. Christensen and Manser's (1976) estimated "branch" COLI's for meat provide an example. Five different nonhomothetic utility functions gave COLI estimates between 1.229 and 1.231, over the thirteen year period 1958-1971 (using 1958 preferences as the base, and 1958 = 1.000); three homothetic functions produced estimates between 1.228 and 1.233, virtually indistinguishable on an annual rate of change basis. Using their whole period (1947-1971), the results were comparable: Over that period, indexes for nonhomothetic forms advanced in the range of 1.437 to 1.440, homothetic ones from 1.441 to 1.444. Again, imposition of homotheticity produces some impact on the COLI estimates, but it is tiny.

From this, the criticism of the Diewert-Byushgens superlative index number approximation (that they imply homothetic utility functions) is not, empirically, a compelling criticism if the purpose is to estimate substitution bias.

Diewert's superlative index number results were used in another substitution bias estimate by Manser and McDonald (1989). Manser and McDonald also pursued an alternative approach using a revealed preference method derived from work by Afriat (1967) and Varian (1982, 1983). They used the superlative index to estimate substitution bias and the Afriat-Varian method to calculate a bound on the approximation error inherent in the second order approximation provided by the superlative index. For the period they studied the approximation was very good, and again the substitution bias estimate was on the order of earlier studies.²⁷ They also, as in earlier studies, evaluated homotheticity.

Finally, Aizcorbe and Jackman (1993) estimated substitution bias using superlative index numbers at a far 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 (63 commodities for the combined demand system results, roughly twice that number for Manser and McDonald).

In short, a substantial amount of research was carried out by BLS on the question of substitution bias across index basic components. BLS combines basic components with a Laspeyres index; these studies show that a COLI computed from an approximation to the true utility function suggests a small amount of substitution bias, on the order of 0.1-0.2 points per

²⁷Blow and Crawford (2001) used the Manser and McDonald procedures to estimate substitution bias in the U.K. index.

year. Almost no estimates outside the BLS were produced in the entire period between the early 1970s and the time of the Boston Commission review. These estimates informed the Boston Commission, and indeed the Commission stated that what it called the “upper level substitution bias” was among the firmest empirical estimates the Commission had available.

D. The Boskin Commission: Further Developments in the Cost of Living Index Approach

Prompted in part by a comment from Federal Reserve Chairman Alan Greenspan about upward bias in the CPI (Greenspan, 1995), in 1995 the Senate Finance Committee appointed an “Advisory Commission to Study the Consumer Price Index” chaired by Michael Boskin. [something about the rationale, social security, “deficits as far as the eye can see” etc]] Like earlier reviews, the Boskin Commission had one carryover from Stigler Committee: Zvi Griliches, who contributed its staff paper number 3.

Like the Stigler committee, the Boskin Commission recommended that BLS adopt a cost of living index as the measurement concept for the CPI. Indeed, the Boskin Commission took the appropriateness 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. Furthermore, problems in estimating a COLI identified by the Stigler committee, such as 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 noted in section C, 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.²⁸ These statements were surprising in light

²⁸ Baker (1998, p. 131) interprets the 1995 BLS report the House Budget Committee as taking this position, and others formed the same interpretation. Moreover, based on personal conversations with Boskin Commission members at the time, we believe that they interpreted BLS statements in the same way as did Baker.

of BLS's longstanding acceptance of the COLI concept, and naturally resulted in a perception that the Boskin Commission was making a recommendation that BLS professional staff opposed. In turn, this dispute inflamed the contentions over the rest of the Commission's findings (some of which were indeed controversial) because they raised all manner of political questions from the Senate Committee's motivation in selecting the panel (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 employ formulas to approximate a COLI as closely as possible, and here the recommendations constituted genuine changes. The Boskin commission also added some elements to the discussion of the COLI not present in the Stigler committee's report. As noted in section D, Diewert (1976) showed that the substitution bias part of the COLI is a more practical concept than was surmised at the time of Stigler committee, as commodity substitution bias is easily approximated by means of "superlative" index number formulas such as the Törnqvist index or the Fisher index. As also noted in section D, BLS had already estimated the substitution bias, using several methods, including superlative indexes, so a quantitative estimate was at hand, unlike the situation in the Stigler Committee days, when no information about the magnitude of substitution bias existed.

Second, the Boskin committee discussed some sources of difference between the CPI and a COLI that Stigler committee had not considered. Obviously, commodity substitution is substitution, no matter the level at which it takes place. But the estimates referred to in section D are estimates of "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).

What the Commission called "lower level substitution bias"—to be discussed in the next section—occurs between the detailed items and varieties that are aggregated to construct the 207 item group indexes in the CPI (that is, *within* the basic components). Though the problems

of aggregating price quotations into basic components had been discussed previously (Carruthers, Sellwood and Ward, 1960; Szulc, 1983; Schultz, 1994), awareness of the problem for the U.S. CPI entered the literature with research papers that had appeared in the early 1990s (Reinsdorf, 1993, and Reinsdorf and Moulton, 1995).

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, again a problem whose importance for the U.S. was first demonstrated by Reinsdorf (1993). Related research shows similar problems in CPIs elsewhere (Saglio, 1994). The Stigler Committee did distinguish a related problem, in their fear that product specifications set in Washington would bias the index toward unrepresentative items, but the index form for the lower level aggregate was different in 1960 from what it was in 1990. [this probably belongs in section II]

Like the Mitchell Committee, but unlike the Stigler Committee, the Boskin Commission estimated biases in the CPI compared with a COLI objective. Their 0.15 percentage points per year from upper level substitution was based on the empirical estimates discussed in section B, 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, 1993, 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 charged the commission with lack of objectiveness for its indulgence in guestimates. Some (particularly the quality change estimates) were marred by faulty reasoning or understanding (see part III).

The Boskin commission also revisited a question that had received little attention since the discussion in the Mitchell committee's report. This question is: To what extent should developments beyond price changes that affect the cost of living be reflected in BLS's index? Here the Boskin commission raised some worthwhile questions concerning inconsistencies that

can arise from consideration of price effects in isolation from non-price 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?

However, in its discussion of the effects of broader changes in the environment on the cost of living, it failed to make the distinction between the “conditional” COLI and the “unconditional” COLI, a distinction that has been in the price index literature at least since Pollak (1989, chapter z). In the context of a report on the CPI, which following Gillingham (1974) and Pollak (1989) is best interpreted as an approximation to a conditional COLI, these effects are out of scope for the CPI. Discussing these effects without providing the appropriate context implicitly suggests that they might be in scope, creating confusion. Environmental deterioration might well be considered in a broader measure, but the narrower one is not irrelevant.²⁹ On this matter, note that the Mitchell committee opted for a different resolution in its evaluation of the Meany-Thomas report, though in 1944 neither the distinction between conditional and nonconditional COLI’s was known (the distinction is Pollak’s) and the analytic framework for considering this class of problems was not that advanced at the time.

BLS’s response to the Boskin Commission’s report was mixed. It agreed with some but not all of the Boskin Commission’s recommendations and finding. Regarding its recommendation that the CPI establish a COLI objective, Commissioner Katharine Abraham responded by saying “This seems basically right to me. Indeed, the BLS long has said that it operates within a cost-of-living framework in producing the CPI.”³⁰ Similarly, the BLS paper “Measurement Issues in the Consumer Price Index,” for example, 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-

²⁹ However, two of the members of the Boskin commission did state elsewhere that these non-price effects do not belong in the CPI — see Gordon and Griliches (1997, p. 87.)

³⁰ See the testimony to a Congressional subcommittee in Abraham, 1997, p. 13.

encompassing cost-of-living concept ...” (Bureau of Labor Statistics, 1997b, p. 4; see also Abraham *et al.*, 1998, p. 27).

The Boskin Commission report has been very influential, not just in the U.S., but worldwide. But the basis for its influence, ironically, was its point 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.

E. The CNSTAT Panel: Second Thoughts about the Cost of Living Index Approach

The Boskin commission had operated in a politically charged environment,³¹ and it had very limited resources. Furthermore, critics of the Boskin commission report argued that imbalances existed in its treatment of CPI biases and that the membership of the commission excluded alternative points of view. Robert Pollak — echoing a suggestion made beforehand by Boskin commission member Zvi Griliches — therefore recommended that the technical questions be examined by a committee of technical experts under the auspices of an organization such as the National Academy of Sciences to do a thorough study of the technical issues related to the CPI. (Pollak, 1998, p. 76.)

Consistent with this proposal, BLS asked the Committee on National Statistics of the National Academy of Sciences to convene a panel (hereafter “CNSTAT Panel”) to investigate “issues in the development of cost-of-living indexes” (National Research Council [NRC], 2002, p. 17). In contrast to all previous review panels, the CNSTAT Panel was not composed primarily of economists with expertise on the CPI; the membership included economists whose

³¹ One disadvantage of this type of environment was that it made the Boskin commission reluctant to circulate a draft version of the final report to BLS or outside reviewers for comments. However Gordon (1997, p. 350) complimented BLS’s openness and candor in assisting the Boskin commission in its work. [[This was pro forma, neither he nor the Committee really felt that way.]]

primary expertise was in other areas and representatives of disciplines such as statistics, psychology, and marketing. Thus, the Panel was chosen for the range of the backgrounds they brought to the work. Among the economists on the panel, some had not participated in the debate over the issues raised by the Boskin commission and some were known to differ from the views expressed by the Boskin commission. The goal seemed to be to make sure that a broad range of perspectives was represented. 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.

If the goal was to obtain a diversity of perspectives, that is certainly what BLS got. In particular, the view that the cost of living index is the appropriate measurement concept for the CPI, which had broad support from economists for half a century,³² proved so contentious that the panel was unable to reach a consensus on this fundamental issue.

Rather than measuring the cost of maintaining a constant standard of living, some members of the Panel believed that the conceptual objective of the CPI ought to be measurement of changes in the cost of purchasing a fixed basket of goods and services, what the Panel called a “cost of goods index” or COGI (NRC, p. 15). In principle, this fixed basket could be based on the initial (or “reference”) period consumption patterns, the final (or “comparison”) period’s consumption patterns, or even some point in between (NRC, p. 16). However, the CNSTAT Panel often discussed the COGI in ways that suggest a Laspeyres index, which uses the initial period’s expenditure patterns (for example, see, p. 112.)

The proponents of the COGI justified their view in three ways. Some found the COLI concept unsuitable for the CPI. In particular, these proponents of the COGI accepted that a fixed-basket index over-compensates for price increases because of substitution bias, but believed that the CPI should measure the “price level” rather than the level of compensation that is needed to hold welfare constant (NRC, p. 58). The distinction between measuring the price

³² For example, panel member Robert Pollak (1989, p. vii) writes that at BLS in 1968-9 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.

level and measuring compensation is illustrated by the debate among members of the Schultze panel on the treatment of new goods. Recall that the Stigler committee included all 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 practical method to include all the gains in the CPI by extrapolating the demand curve for a new good out to a Hicksian virtual price, at which the good's sales would fall to zero. The members of the CNSTAT Panel who favored the COGI recommended against use of this technique in the CPI. In their view, gains from the invention of new goods are not a change in the price level, and hence not part of the concept that the CPI should measure (NRC, 2002, pp. 160-1). They did not discuss what the concept of the price level meant. Apparently, their view of the price index coincides with that of Fisher—some price level exists, like the level of water in a lake, and the index is a device to find it. Aggregation over component price indexes is not part of this view of consumer prices.

The second argument for the COGI approach is less charitable to the COLI alternative. Members contended that many uses of COLI theory depended on unrealistic assumptions, namely homotheticity, constant tastes and constant environmental conditions. They also questioned the validity of the representative consumer, when the COLI is formed at the economy-wide level rather than at the level of an individual consumer, where aggregation across households is implied (one can think of the aggregate CPI as the aggregation of individual CPI's for each household).

Commodity substitution due to changes in relative prices is not the only cause of changes in purchasing patterns; changes in the standard of living, changes in tastes, and changes in the broader environment are also important. Hence, the changes in consumption that are measured in empirical estimates of a COLI may reflect factors other than changes in relative prices.³³ If

³³ This point seems mostly directed toward estimation of substitution bias through the Diewert (1976) approach of using superlative index numbers. It is not a valid objection to COLI estimated by demand system estimates, in which income effects are included in the estimate.

purchasing patterns depend on the standard of living (i.e. if preferences are non-homothetic), a different choice for the standard of living held constant would imply a different value for the COLI. In particular, as is well known, the COLI for the base-period standard of living is bounded from above by the Laspeyres fixed-basket 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 (Pollak, 1989). Similarly, without homotheticity, 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.³⁴

Of course, the criticism that unrealistic assumptions are needed for some common uses of the COLI does not imply that such assumptions are needed for the COLI to be usable at all. Violations of the questionable assumptions merely make its implications less sharp and less tidy; in particular, they call into question direct use of Diewert’s (1976) elegant result that the Fisher index (along with some other “superlative” indexes) is exact for a flexible expenditure function and hence able to do a good job of measuring substitution effects under a wide range of conditions. They do not invalidate the basic result from COLI theory that price-induced substitution makes the Laspeyres index an upper bound to a COLI and the Paasche index a lower bound, though they do mean that this result must be stated in a careful way. In particular, the Laspeyres index for a single household is an upper bound to its COLI evaluated at the initial standard of living, tastes and environmental conditions. The aggregate Laspeyres index is an upper bound for Pollak’s (1981) Scitovsky-Laspeyres index, which tracks the cost of keeping every household at its initial standard of living with its initial endowment of tastes and environmental factors. Similarly, the aggregate Paasche index is a lower bound for a “Scitovsky-Paasche” index that tracks the cost of keeping every household at its final standard of living with its final endowment of tastes and environmental factors. The COLI theory’s objective of

³⁴ It is often asserted that preferences must be identical, in addition to being homothetic, to justify use of the representative consumer, but if we are prepared to assume no redistribution of income, only homotheticity is necessary.

accounting for substitution effects can therefore still be accomplished by the averaging upper and lower bounds on COLIs. This average is the aggregate Fisher index.³⁵

The meaningfulness of this average might, of course, be doubted if the COLI concept bounded by the Laspeyres index were numerically distant from the concept bounded by the Paasche index. Admittedly, changes in the standard of living, tastes and the environment are likely to have important effects on purchasing patterns over the very long run, but these effects — which create the theoretical complications with cost of living index concept — are usually relatively unimportant compared with substitution behavior over the shorter intervals of interest for most uses of the CPI. In practice, Laspeyres indexes are typically found to be above Paasche indexes by an amount that could be expected from substitution effects, whereas a major role for the other kinds of effects would presumably cause many instances of the Paasche index being either above the Laspeyres index or further below it than could plausibly be attributed to substitution effects alone.

Indeed, the empirical evidence in Manser and MacDonald's study of substitution effects was that aggregate expenditure patterns are consistent with homotheticity (1988, p. 919). Similarly, Christensen and Manser (1976, pp. 434-7) found that changing the reference standard of living from the 1947 to the 1967 level raised their nonhomothetic cost of living subindexes for meat by about 2 percentage points and lowered their cost of living subindexes for produce by about 1½ percentage points over a 24 year period — trivial effects given the large change in the reference utility level.

A third approach to justifying the COGI hinges on the observation that to use the COLI for the CPI, its domain must be restricted to exclude non-market goods, making it a *conditional* COLI. The need to decide on how to condition the COLI is a problem, because the decisions can be difficult and inconsistencies can be hard to avoid. Where a new technology fits in the scheme

³⁵ A Törnqvist index can be supported as a measure of an aggregate cost of living index using a different approach. If each household's share of aggregate expenditure is constant, then the aggregate log Törnqvist index is a weighted average of household-specific log Törnqvist indexes, which are themselves superlative measures of household level cost of living indexes. If household's shares of aggregate expenditure vary, the aggregate log Törnqvist will still be closely approximated by this weighted average of the household level indexes in most cases.

defined by these principles may be unclear, and we must condition the COLI on some new technologies, the Panel asserts, but not on others (NRC, p. 67). For example, the CNSTAT Panel favored diagnosis-based measures of medical care rather than input-based measures whenever possible (p. 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 (p. 67).³⁶

Although the debate over whether to use the COLI or the COGI is probably the most noteworthy feature of the CNSTAT Panel report, another of its contributions to the debate on the fundamental measurement concept for the CPI is also important. This is its guidance on how best to implement the COLI—offered, perhaps, in tacit recognition that use of the COLI would continue. Within the context of the COLI approach, the panel agreed that the appropriate version of the COLI for the CPI was a *conditional* COLI covering private goods and services and holding the broader environmental factors constant (NRC, 2002, p. 73). This is reminiscent of the position that the Mitchell committee had taken nearly 60 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. (Of course, the effects considered in the CNSTAT Panel report are more broad-ranging than was contemplated at the time of the Mitchell committee.) The Panel asserted that choosing the domain of the conditional COLI had some elements of arbitrariness to it, which is clearly correct. But it did not recognize or emphasize the arbitrariness inherent in COGI.

³⁶ This appears to be an arbitrary judgment about what medical conditions are appropriately included in the treatments priced for the CPI. It is not clear whether the Panel would make the same judgment now that Viagra has become, by news accounts, a recreational drug. In either case, is arbitrariness of this kind not a powerful argument against the COGI approach, and its lack of a theoretical framework for analyzing what goods are to be included and how they are to be measured?

F. Assessment of the Debate over the Conceptual Basis for the CPI

The conditional COLI provides a route to bridge much of the practical gap between the COLI and the COGI, at least if the former is defined narrowly and latter is defined broadly. In practice, estimation of a conditional COLI whose domain is limited to prices of continuously existing private goods and services requires an index formula that accounts for substitution caused by changes in relative prices. The Fisher index is one such formula. The Fisher formula also seems consistent with the remark in the CNSTAT report that the fixed basket used for the COGI can be from either the initial period or the final period. Averaging the possible COGIs based on the initial and final fixed baskets, which yields a Fisher price index, would resolve the ambiguity over which of them to use. Alternatively, if a “pure” COGI—meaning one that has an identifiable fixed basket—is required, we can average the two possible baskets instead of the two possible indexes. To avoid inconsistencies, when we do this we must deflate the quantities in the final period basket by the quantity index before averaging them with the initial period basket, and we must deflate the final period prices by the price index before averaging them with the initial period prices to obtain the average prices for the quantity index. The price and quantity indexes that solve this pair of simultaneous equations are Fisher indexes (Reinsdorf, Diewert and Ehemann, 2002, based on van IJzeren, 1952.) Hence, a Fisher index seems to be a good candidate for a precisely specified COGI concept. Although some passages in the CNSTAT Panel report implicitly define the COGI as a Laspeyres index, to use a Laspeyres index as the measurement goal requires an answer to the question of how often to link in a new basket. Finding a justification for an optimal degree of obsolescence that can serve as a conceptual objective seems impossible unless the answer is “none”, implying a Fisher or similar index.

In addition to helping to bridge the gap with the COGI, a narrow domain for the conditional COLI—which excludes, for example, the welfare gains from the invention of novel products and effects tangential to the market prices paid by consumers—has the advantage of bringing the conceptual objective into alignment with what is practically achievable without

resorting to abstract or subjective adjustments that could put the credibility of the CPI at risk. Admittedly, an income rising in line with such a conditional COLI would likely correspond to a rising standard of living, but uncontroversial exclusions from the conditional COLI, such as government services, already rule out a truly constant standard of living.

Yet even though the COGI can be specified in way that is largely consistent with a narrowly defined conditional COLI, the need to account for substitution effects is not the COLI's only implication. 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, p. 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, and satisfactory alternatives do not come readily to mind. An objective of a fixed basket index leaves the question of the underlying motivation largely unanswered. The motivations offered for the COGI — to track "the prices of the things that people buy" or "the price level"— provide inadequate guidance on questions such as what time period or combination of time periods the market basket should reflect, 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, and how to treat product introductions and disappearances. Indeed, the lack of an explicitly stated abstract motivation could make the CPI hard to interpret or leave it with no clear interpretation because of inconsistent decisions in its design.

In addition to guiding the choice of technique, an explicit abstract objective for measurement such as the COLI can aid in the discovery of improvements by making shortcomings identifiable. It can also help to resolve methodological ambiguities.³⁷ For example, the COLI approach provides more useful guidance than the COGI approach about how to handle problems such as what to do when items in the index disappear, or what to do when

³⁷ Greenlees (2001, pp. 12-14) provides examples of design improvements in the CPI guided by the COLI concept.

new goods or new quality levels appear. In addition, an underlying conceptual purpose 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.

An example is Beelen's (2004, p. 6.) discussion of the treatment of 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 policy-holders? In this case, the service is a management service by the insurance company, and the price is the price of the management service. The European Harmonized Indexes of Consumer Prices (HICP), which operate under a "not COLI" concept, advocate the latter. How does one resolve such questions? Though one might well contend, presumably with the Panel, that COLI theory by itself does not fully answer this question without more development, the alternative is an arbitrary choice where both the decision itself and the framework for discussing it are arbitrary.³⁸

Finally, as Blow and Crawford (2001, p. 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." 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.

³⁸ On measuring insurance in national accounts and CPI, see also Chapter 6 in Triplett and Bosworth (2004).

II. Recommendations Concerning Detailed Component Indexes and Sampling

A. Wartime Committees: The Problem of Unrepresentative Samples of Varieties and Forced Substitution

Among the many biases in BLS's 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 BLS's failure to collect prices on weekends, when sales were common. The Mitchell committee conceded that this bias was present because weekend sales had become less prevalent, forcing some consumers to pay higher prices. However their estimate 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 qualities or outlets.³⁹ Inexpensive qualities tended to be differentially affected by shortages, resulting in forced "trading up." The Mitchell committee observed that this likely affected the poorest families the more than it did the average family tracked by BLS's index (p. 18).⁴⁰ In addition, the Mitchell commission noted some evidence of substitution from chain stores to higher-priced independents. Some shift to higher-priced outlets 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 because. 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

³⁹ Meany and Thomas also alleged that unmeasured quality deterioration was an important cause of downward bias in the Cost of Living Index. We discuss problems posed by quality change in the section on hedonics.

⁴⁰ The Stigler Committee recommended that the CPI be extended to cover all family groups, which was done in the 1978 Revision (CPI-U), but an index with weights corresponding to the old family definition was published separately (CPI-W).

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.⁴¹

The third problem identified by Meany and Thomas highlights the vulnerability to error or manipulation that can undermine the credibility of detailed component indexes not based on scientific sampling. Meany and Thomas argued that unrepresentative samples of varieties or qualities priced for the detailed component indexes were causing a large downward bias in the Cost of Living Index.⁴² Usually one or two tightly specified varieties or qualities 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 and quality was priced everywhere, city level averages for the price of a variety could be calculated without fear of distortion from uncontrolled variation in quality. These averages could be compared over time to form basic component indexes for commodities or narrowly defined groups of commodities, and also they could be compared across cities to investigate inter-area price level differences. This 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.⁴³

The errors introduced by the sampling of varieties or qualities might, of course, have an average near zero, implying very little bias at the aggregate level. Meany and Thomas claimed, however, that the sampling errors in the component indexes were not only high, but also systematic. In particular, 77.3 percent of the food index was based on varieties with price subsidies, but of 37 varieties with rising price control ceilings, only 11 were in BLS’s index (see Tables 3 and 4 in the Meany-Thomas report). To illustrate the problem, Meany and Thomas discussed many cases of the use in the CPI of the particular varieties subject to subsidies or price

⁴¹ Note the interesting parallel with the Boskin Commission, which also prepared guestimates of bias on a component by component basis.

⁴² 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.

⁴³ 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.

rollbacks. For example, subsidized apples represented deciduous fruits, which were otherwise unsubsidized (p. 86). Oranges, which had a 25 percent rollback in their price control ceiling, represented other citrus fruits, which had no price rollback (p. 86). Shortening was used to represent fats and oils, but the subsidies were about to expire for all fats and oils except shortening (p. 87.)

The Mitchell commission considered Meany and Thomas's charge that foods not priced by BLS went up twice as fast as those priced (pp. A-9 to A-12). To respond to Meany and Thomas, BLS had 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 (p. A-10). On balance, therefore, the Mitchell committee's educated guess was that no bias had been caused by BLS's selections of varieties or qualities. The committee did, however, recommend pricing broader ranges of qualities in the future (II-39).

Although the BLS evidence ought and did carry significant weight, it is hard not to be impressed that M-T produced actual data to support their position. They are not disinterested samplers, it is true. But one can also get the feeling that they were subjected to a bit of "Who are you to challenge the experts?" treatment. The language used in some accounts suggests bias. Ostrander (1944, page 853) seemed unduly dismissive for a contribution to a professional journal, for example, in the passage that reads (of the Meany-Thomas report): "assuming there were any substance to it..." One suspects that similar evidence from another source might have been treated differently.

B. The Stigler Committee: Representative Samples for Detailed Component Indexes

The single recommendation of the Stigler committee with the greatest effect on the design of the CPI was to use probability sampling to estimate the detailed component indexes used to calculate the CPI. With such sampling, estimates of index variances could be obtained by

comparing index values calculated from replicated samples using BLS's established policies for adjustment;⁴⁴ the committee found no merit in the argument that the adjustment procedures for handling disappearances of the originally sampled items were so ill-defined that replicated samples were impractical. (Stigler et al., 1961, p. 40.) Another benefit of the recommended approach 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 (p. 42.) Finally, probability sampling was the only way to guard against biases due to an unrepresentative selection of outlets and varieties (p. 42). For example, types of retail establishments or outlets of growing importance were under-represented, 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 (p. 58.)

A paper by committee member Phillip McCarthy, which was 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 his discussion of the sampling of commodities, McCarthy assumed that 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, p. 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, 1961, p. 213.) McCarthy recommended selection by probability sampling of smaller cities, of detailed items within 150 commodity categories, and of outlets (pp. 227-9.)

⁴⁴ 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 estimator of the variance of a "best" index estimate based on a pooled sample.

Another noteworthy observation concerning the detailed component indexes comes from a witness at the hearings held to discuss the Stigler commission's report rather than from the Stigler commission itself. In particular, Arant (1961, p. 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 linking might have prevented a drop in the food component of the CPI of 0.7 percentage point from updating the weights for chains and independents from 1948 values to ones based on 1958 data.⁴⁵

Despite the Hoover-Stotz finding, nothing apparently was done about it (Hoover was at one time head of the CPI, Stotz of the PPI). The issue lay quiet until the study by Reinsdorf (1993).

C. 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. One of these, quality change, is discussed in the next section. Here we will consider formula bias, lower level substitution bias and outlet substitution bias.⁴⁶

Formula bias in the CPI was a focus of the Boskin Commission's "interim report," dated September 1995. 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. Notwithstanding this problem, the benefits of probability sampling were, of

⁴⁵ Note, however, that this estimate may represent the effect of changes in purchasing patterns over fewer than ten years. The official CPI linked in interim adjustments to reflect the growing importance of chain stores on four occasions in between the benchmarking of the weights to the 1948 and 1958 Censuses. Reinsdorf (1998, p. 184) found that at the time of 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.

⁴⁶ 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 and on adjusting for quality change. New products are not specifically a problem affecting detailed component indexes since they may not fit into an existing item stratum.

course, considerable. The CPI was less susceptible to bias from unrepresentative samples of varieties or qualities or from manipulation, as Meany and Thomas had alledged. 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. 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, and 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 prices, since prices of dissimilar items cannot meaningfully be averaged.⁴⁷ This suggested a change in approach for the basic component indexes from taking ratios of average prices (used before 1978) to averaging ratios of prices (used after). The simple average of price ratios discussed by McCarthy constituted 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.

BLS's estimated sampling probabilities as well as possible, but 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

⁴⁷ Before 1978, the formula for an (unweighted) CPI basic component was: $\sum (p_{i, t+1} / n) / \sum (p_{it} / n)$, 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.

error. Consequently, ratios of sampling probabilities to prices were not proportional to the desired quantities for a Laspeyres index.⁴⁸ What is more, any downward transitory shock to an initial period price simultaneously implied an upward shock to its implicit quantity weight and to its change. 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.⁴⁹ Seasoning removed this nuisance correlation by using the initial price data from a new sample solely to set weights, then allowing several months to elapse before using data from the new sample in index calculations.

However, in June 1996, BLS extended the method of “seasoning” (first introduced in January 1995 for food at home) to all commodities and services, a broad category of items other than shelter. Seasoning was intended to eliminate the formula bias problem. In the Boskin commission’s final report, the discussion of formula bias therefore focused on how large this effect had been (Advisory Commission to Study the CPI, 1996, p. 42.) The Boskin commission’s estimate of about 0.5 percent per year is consistent with research by BLS 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 a similar bias in the owners’ equivalent rent component of the CPI may have had an effect of around 0.1 percent per year.⁵⁰ Formula bias, was, therefore, twice as large as the Manser-McDonald estimate of commodity substitution bias of around 0.2 percent year.

⁴⁸ 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.

⁴⁹ For example, suppose that consumers always buy equal amounts of good 1 and good 2 and that their prices bounce back and forth between 1 dollar and 2 dollars. To avoid bias, an initial price of 1 should have a one-third probability of selection. Yet with probabilities of selection based on historical averages of expenditure shares, its probability of selection would likely be one-half, resulting in an expected value for the index of 1.25. For a detailed explanation, see Reinsdorf (1998.)

⁵⁰ See Reinsdorf, 1998, p. 185. Experimental CPI’s available from BLS at <http://www.bls.gov/cpi/cpigm00.htm> show an average effect on the all-item CPI of about 0.3 percent per year not counting the effect from owner’s equivalent rent.

The Boskin commission, however, regarded seasoning as only a partial cure for the formula-related problems of the basic component indexes. In its final report the Boskin commission identified a remaining problem of “lower level substitution bias.” The solution to this problem was to use a geometric mean of price relatives as the formula for most basic component indexes in the CPI.

Whereas consumer behavior plays no role in formula bias,⁵¹ a substitution bias is 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, referring to evidence in Shapiro and Wilcox (1996), argued that use of geometric means to average price relatives would result in unbiased estimation of basic component indexes for the COLI objective (Advisory Commission, 1996, p. 42.) 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. 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.⁵²

D. The CNSTAT Panel: Guarded Agreement with the Boskin Commission

BLS adopted geometric means for most basic component indexes in the CPI in 1999. The CNSTAT Panel gave qualified support to this change. It observed that high substitutability between the product varieties in component indexes was generally plausible, with exceptions for

⁵¹That is, no assumption about behavior is sufficient for formula bias to vanish — see Reinsdorf, 1998, p. 178.

⁵²BLS acknowledged the possibility that price differentials between outlets could cause bias in its index as early as 1964. Consistent with the findings of Hoover and Stotz (1964), Jaffee 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.” (1964, p. 923.)

some items such as medical supplies. 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” (NRC, 2002, p. 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 (NRC, p. 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 (p. 175). They therefore concluded that BLS had little choice but to continue its present practice of linking in new outlets (p. 176.) They did, however, recommend continued research on the effects of outlet characteristics on prices.

E. Assessment of the Debate over Component Indexes

We have little 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 leave an important matter in need of clarification. They imply that the main purpose of the geometric mean formula is to account for substitution effects in component indexes that contain closely substitutable varieties of a good or service or closely substitutable outlets selling the same good or service. Indeed, the geometric mean index is the formula for a COLI if the elasticity of substitution equals 1 (so that inverse movements of relative prices and

relative quantities keep expenditure shares constant) **and** the expenditure shares are measured correctly. The question is, What 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 attempts to do so are likely to be misleading. 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, or an assumption of uniform quantities if the index covers homogeneous items. If implication of substitution behavior implicit in constant expenditure shares is thought to be objectionable, two alternative assumptions based on Leontief behavior must be treated as equally likely to hold. One is that initial period expenditures are uniform while final period expenditures are directly proportional to the price relatives, and the second is that final period expenditures are uniform while initial period expenditures are inversely proportional to the price relatives. Giving equal weight to both these possibilities implies a Fisher index that equals, or virtually equals, a geometric mean index. Hence, making an assumption of uniform expenditures in a logically consistent way leads to a geometric mean index.⁵³ If expenditure patterns are unknown, the inference that a simple average of price ratios provides a better estimate of a COLI than a geometric average when the items in the index are not substitutable because the geometric average assumes substitution is, therefore, a fallacy.

In practice, the information on expenditures shares of outlets and varieties covered by the component indexes of the CPI is usually somewhere in between perfect measurement and perfect ignorance. If prices have persistent divergent trends and consumers do not treat the items covered by the component index as substitutable, the estimates of the expenditure shares from earlier periods may be closer to the true expenditure shares in the index's initial period than in its final period. In this case, the geometric mean index will still provide a good measure of short-term price change, but it may be biased downward over the long run as the influence of the

⁵³ An alternative assumption of identical quantities in both periods implies a ratio of average prices formula for the index.

divergent price trends becomes more pronounced. This provides some legitimate justification for the recommendation of the CNSTAT Panel to avoid the geometric mean formula for component indexes for items that appear non-substitutable. Nevertheless, the need to account for substitution behavior is not the only justification for the geometric mean index: its property of simultaneously being an average of price relatives and ratio of average price levels make the only simple formula that is well-suited for handling the heterogeneity in varieties and qualities that comes from probability sampling. Furthermore, in the 1998 revision of the CPI, resampling of outlets and of varieties within outlets became more frequent, particularly for items with changing expenditure patterns. Chaining in updated indexes more frequently increases the comparative advantage of the geometric mean index formula over more complicated alternatives such as seasoned Laspeyres indexes.

A second issue is more far-reaching and likely of more empirical importance. Economists have interpreted the difference between arithmetic mean (Laspeyres-type) and geometric mean aggregators for basic components as just the classic substitution bias paradigm drawn from Konus (1925), only applied one level down. Substitution behavior is clearly a relevant concern for basic components: When pianos and guitar picks are in the same component (an example used by Moulton and Moses, 1997), certainly 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 as category).

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 banana price index, which is as close to a homogeneous product (in the U.S., at any rate) as can be found. An even stronger 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 explain differences in index number formulas at this level with a theory of consumer behavior requires that the theory model consumer behavior toward homogeneous commodities, as well as behavior across different (substitutable) commodities.

Pollak (1978) 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" (page 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" (page 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, page 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) propose a model for analyzing purchases where storage is intended (their application was canned tuna, using scanner data).

The CNSTAT Panel (pages 5 and 24) addressed this matter: “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” (page 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” (page 24).

We think that more research on these problems is the proper future direction for understanding how to measure basic components in price indexes, much more so than attempts to crush the basic component problem to fit into the standard Konus commodity substitution model. Until we have a more satisfactory model of the consumer behavior that is relevant to basic components, our theoretical grounds for concluding that geometric means are superior to arithmetic means as aggregators for basic components will be weak.⁵⁴

⁵⁴ Both of us think geometric means are better. Our point is that our reasons for believing this are on very shaky theoretical foundations.

References (incomplete)

- Adelman, Irma. 1958. "A New Approach to the Construction of Index Numbers," *Review of Economics and Statistics* 40, 240-9.
- Afriat, Sydney N. 1967. "The Construction of Utility Functions from Expenditure Data." *International Economic Review* 8 (February): 67-77.
- Aizcorbe, Ana M. and Patrick C. Jackman. 1993. "The Commodity Substitution Effect in CPI Data, 1982-91: Anatomy of Price Change." *Monthly Labor Review* 116:12 (December): 25-33.
- Allen, R.G.D. 1975. *Index Numbers in Theory and Practice*. London: MacMillan.
- Anglin, Paul M., and Michael R. Baye. 1987. "Information, Multiperiod Search, and Cost-of-Living Index Theory." *Journal of Political Economy*, 95(6) (December), pp. 1179-95.
- Arrow, Kenneth J. 1958. "The Measurement of Price Changes." in U.S. Congress, Joint Economic Committee. *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: US Government Printing Office.
- Arant, Willard 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, D.C.: Government Printing Office.
- Baker, Dean 1998. *Getting Prices Right*. Washington: M.E. Sharpe.
- Banzhaf, H. Spencer. 2001. Quantifying the Qualitative: Quality-Adjusted Price Indexes in the United States, 1915-61. In *The Age of Economic Measurement*, edited by Judy L. Klein and Mary S. Morgan. Durham, NC: Duke University Press.
- Banzhaf, H. Spencer. Forthcoming. The Form and Function of Price Indexes: An Historical Accounting. To appear in *History of Political Economy*.
- Baye, Michael R. 1985. "Price Dispersion and Functional Price Indices." *Econometrica* 53:1 (January), pp. 213-223.
- Beelen, George. 2004. "Vehicle, Homeowners' and Tenants' Insurance in the Canadian CPI." Presented to the Statistics Canada Price Advisory Committee, April.
- Blow, Laura and Crawford, Ian. 2001. "Should the Cost-of-Living Index Provide the Conceptual Framework for a Consumer Price Index?" *The Economic Journal* 111, no. 472, June, F357-F382.

- Boskin Commission. 1996. See Boskin et al., 1996.
- Boskin, Michael J., Ellen R. Dulberger, Robert J. Gordon, Zvi. Griliches, and Dale Jorgenson. 1995. *Toward a More Accurate Measure of the Cost of Living*. Interim Report to the US Senate Finance Committee.
- Boskin, Michael J., Ellen R. Dulberger, Robert J. Gordon, Zvi. Griliches, and Dale Jorgenson. 1996. *Toward a More Accurate Measure of the Cost of Living*. Final Report to the US Senate Finance Committee, December. In US Senate Finance Committee. Senate Print 104-72, 104th Congress, 2nd Session, Washington: US GPO.
- Braithwait, Steven D. 1980. "The Substitution Bias of the Laspeyres Price Index: An Analysis Using Estimated Cost-of-Living Indexes". *The American Economic Review*, Vol. 70, No. 1., March, 64-77.
- Brown, Murray and Dale Heien. 1972. "The S-Branch Utility Tree: A Generalization of the Linear Expenditure System" *Econometrica*, Vol. 40, No. 4. (Jul., 1972), pp. 737-747.
- Buyshgens. 1925. "Sur une classe des hypersurfaces: A propos de 'l'index ideal' de M. Irving Fischer." *Mathematischkii Sbornik* 32:625-631.
- Carruthers, A.G., D.J. Sellwood, and P.W. Ward. 1980. "Recent Developments in the Retail Prices Index." *The Statistician* Vol. 29 No. 1: 1-32.
- Christensen, Laurits R. and Manser, Marilyn E. 1976. "Cost-of-Living Indexes and Price Indexes for US Meat and Produce." in *Household Production and Consumption*, Nestor Terleckyj, ed. Studies in Income and Wealth Vol. 40, New York: Columbia University Press (for NBER): 399-446.
- Diewert, W. Erwin. 1976. "Exact and Superlative Index Numbers." *Journal of Econometrics* 4(2): 115-45, May.
- Diewert, W. E. 1993. "The Early History of Price Index Research." In Diewert, W. Erwin and Alice O. Nakamura, eds., *Essays in Index Number Theory*, Volume 1, chapter 2, pp. 33-66. New York: North Holland.
- Diewert, W. Erwin. 1995. "Axiomatic and Economic Approaches to Elementary Price Indexes." Discussion paper No. 95-01. University of British Columbia. Vancouver, Canada.
- Diewert, W. E. 2000. "The Consumer Price Index and Index Number Purpose", Discussion Paper No. 00-02, Department of Economics, University of British Columbia, Vancouver, Canada, V6T 1Z1.
- Feenstra, Robert C. and Matthew D. Shapiro. 2003. "High-Frequency Substitution and the Measurement of Price Indexes." In Robert C. Feenstra and Matthew D. Shapiro, eds., *Scanner Data and Price Indexes*, National Bureau of Economic Research Studies in Income and Wealth, Vol. 64. Chicago and London: The University of Chicago Press, pp. 123-46.

- Fisher, Irving. 1921. *The Making of Index Numbers*. Boston, MA: Houghton Mifflin Co.
- Fisher, F. and Shell, K. 1972. *The Economic Theory of Price Indices: Two Essays on the Effects of Taste, Quality, and Technological Change*. New York: Academic Press.
- Frisch, R. 1936. "Annual Survey of General Economic Theory: The Problem of Index Numbers." *Econometrica* 4: 1-39.
- Gillingham, Robert. 1974. "A Conceptual Framework for the CPI" in *Proceedings of the American Statistical Association, Business and Economics Section*.
- Goldberg, Joseph P and Moye, William T. 1985. *The First Hundred Years of the Bureau of Labor Statistics*. Washington: US Government Printing Office.
- Goldberger, Arthur S., and Theodore Gameletsos. 1970. "A Cross-Country Comparison of Consumer Expenditure Patterns." *European Economic Review* 1: 357-400.
- Gordon, Robert J. 1997. "Comment" *Brookings Papers on Economic Activity* 1, 350-7.
- Gordon, Robert J. and Zvi Griliches, Zvi. 1997. "Quality Change and New Products." *The American Economic Review*, Vol. 87, No. 2, Papers and Proceedings of the Hundred and Fourth Annual Meeting of the American Economic Association, May: 84-8.
- Greenlees, John S. 2001. *The U.S. CPI and the Cost-of-Living Objective*. Presented at the Conference of European Statisticians Joint ECE/ILO Meeting on Consumer Price Indices, November 2, 2001. Washington, D.C.: US Bureau of Labor Statistics.
- Greenspan, A. 1995. Testimony to the Committee on Finance, U. S. Senate, March 13.
- Griliches, Zvi. "Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change." Staff Paper no. 3 in U.S. Congress, Joint Economic Committee (1961): *Government Price Statistics, Hearings, Part 2, May 1, 2, 3, 4, and 5, 1961*. Washington, D.C.: Government Printing Office, 173-96.
- Griliches, Zvi and Cockburn, Ian. 1995. "Generics and New Goods in Pharmaceutical Price Indexes." *The American Economic Review*, Vol. 84, No. 5. December: 1213-32.
- Hausman, Jerry A. 1997. Valuation of New Goods Under Perfect and Imperfect Competition. In *The Economics of New Goods* T.F. Bresnahan and R.J. Gordon, eds., NBER Studies in Income and Wealth, Vol. 58. Chicago: University of Chicago Press.
- Hausman, Jerry A. 2003. "Sources of Bias and Solutions to Bias in the Consumer Price Index." *Journal of Economic Perspectives*, 17(1) (Winter), pp. 23-44.
- Heien, Dale M. 1973. "Some Further Results on the Estimation of the S-Branch Utility Tree." *BLS Working Paper* 10. Washington, D.C.: U.S. Bureau of Labor Statistics.

- Hoover, Ethel D. and Stotz, Margaret S. 1964. "Food Distribution Changes and the CPI." *Monthly Labor Review* 87 (January): 58-64.
- Hulten, Charles R. 1973. "Divisia Index Numbers." *Econometrica*, Vol. 41, No. 6. (Nov., 1973), pp. 1017-1025.
- van IJzeren, J. 1952. J. Over de Plausibiliteit van Fisher's Ideale Indices, (On the Plausibility of Fisher's Ideal Indices), *Statistische en Econometrische Onderzoekingen*, Centraal Bureau voor de Statistiek, *Nieuwe Reeks* 7 (no. 2), 104-115.
- Jaffee, Sidney A. "The Statistical Structure of the Revised CPI" *Monthly Labor Review* 87 Aug 1966, pp. 916-24.
- Klein, Lawrence R. and H. Rubin. 1947-48. "A Constant Utility Index of the Cost-of-Living." *Review of Economic Studies* 15, pp. 84-87.
- Konus, A.A. 1925. Translated into English and published as "The Problem of the True Index of the Cost of Living." *Econometrica* 7 (January) 1935, pp.10-29.
- Manser, Marilyn E. 1975. "A Note on Cost-of-Living Indexes and Price Indexes for U.S. Food Consumption." *BLS Working Paper* 57. Washington, D.C.: U.S. Bureau of Labor Statistics.
- Manser, Marilyn E. 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, Marilyn and MacDonald, Richard. 1988. "An Analysis of Substitution Bias in Measuring Inflation, 1959-85." *Econometrica* 46 (no. 4), July: 909-30.
- McCarthy, Phillip 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 (1961): *Government Price Statistics, Hearings, Part 2, May 1, 2, 3, 4, and 5, 1961*. Washington, D.C.: Government Printing Office, 197-232.
- Meany, George, and Thomas, R.J. 1944. *Recommended Report for the Presidential Committee on the Cost of Living Index*.
- Mills, Frederick C., E. Wight Bakke, Reavis Cox, Margaret G. Reid, Theodore W. Schultz, and Samuel S. Stratton. (Special Committee of the American Statistical Association.) 1943. An Appraisal of the US Bureau of Labor Statistics Cost of Living Index. *Journal of the American Statistical Association* 38: December: 387-405.
- Mitchell, Wesley Clair. 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 US Bureau of Labor Statistics, No 173 (July): 5-114. Washington, D.C.: U.S. Government Printing Office.

- Mitchell, Wesley Clair. [1915] 1921. "The Making and Using of Index Numbers." Part 1 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, D.C.: U.S. Government Printing Office.
- Mitchell, Wesley C., Simon Kuznets, and Margaret G. Reid. 1944. 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: Government Printing Office.
- Moses and Moulton. 1997. "Addressing the Quality Change Issue in the Consumer Price Index." *Brookings Papers on Economic Activity* 1: 305-49.
- National Research Council. 2002. *At What Price? Conceptualizing and Measuring Cost-of-Living and Price Indexes*. Schultze, Charles, and Christopher Mackie, eds. Washington, DC: National Academy Press.
- Ostrander, F. Taylor. 1944. "The Mitchell Committee's Report on the Cost-of-Living Index: Comments (in Communications)" *The American Economic Review*, Vol. 34, No. 4. (Dec., 1944), pp. 849-856.
- Pakes, Ariel. 2003. "A Reconsideration of Hedonic Price Indexes with an Application to PC's." *American Economic Review*, 93(5) (December), pp. 1578-96.
- Pollak, Robert A. 1981. "The Social Cost-of-Living Index." *Journal of Public Economics* 15, No. 2, June, 311-36.
- Pollak, Robert A. 1983. "The Theory of the Cost-of-Living Index" In W. Erwin Diewert and Claude Montmarquette, eds. *Price Level Measurement: Proceedings from a Conference Sponsored by Statistics Canada*. Ottawa: Minister of Supply and Services Canada, pp. 87-162.
- Pollak, Robert A. 1989. *The Theory of the Cost-of-Living Index*. New York: Oxford University Press.
- Pollak, Robert A. 1998. "The Consumer Price Index: A Research Agenda and Three Proposals." *Journal of Economic Perspectives* Vol. 12 No. 1 (Winter): 69-78.
- Price Statistics Review Committee. 1961. See Stigler et al. 1961.
- Price Statistics Review Committee, National Bureau of Economic Research. 1961. *The Price Statistics of the Federal Government: Review, Appraisal, and Recommendations*. A Report to the Office of Statistical Standards, Bureau of the Budget. General Series No. 73. New York, NY: National Bureau of Economic Research.

- Reinsdorf, Marshall B. 1993. "The Effect of Outlet Price Differentials on the US Consumer Price Index." In Foss, Murray, Manser, Marilyn and Young, Allan, eds. *Price Measurements and Their Uses*. NBER Studies in Income and Wealth, Vol 57. Chicago: Univ. of Chicago Press, 227-60.
- Reinsdorf, Marshall B. 1998. "Formula Bias and Within-Stratum Substitution Bias in the US CPI." *Review of Economics and Statistics* 80, no. 2., May, pp. 175-87.
- Reinsdorf, Marshall B., Diewert W. Erwin and Ehemann, Christian. 2002. "Additive Decompositions for Fisher, Törnqvist and Geometric Mean Indexes". *Journal of Economic and Social Measurement* 28, 51-61.
- Reinsdorf and Moulton, 1995. "The Construction of Basic Components of Cost of Living Indexes." In *The Economics of New Goods*, Timothy F. Bresnahan and Robert J. Gordon, eds. University of Chicago Press, 1997, pp. 397-423.
- Ross, 1966. (citation to be supplied by JET.)
- 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*. Sponsored by Statistics Canada, Ottawa, November.
- Samuelson, P. A. and S. Swamy. 1974. "Invariant Economic Index Numbers and Canonical Duality: Survey and Synthesis" *The American Economic Review*, Vol. 64, No. 4. (Sep., 1974), pp. 566-593.
- Schultze, Charles and Christopher Mackie, eds. 2002. *At What Price? Conceptualizing and Measuring Cost-of-Living and Price Indexes*. Panel on Conceptual, Measurement, and Other Statistical Issues in Developing Cost-of-Living Indexes. Committee on National Statistics, National Research Council. Washington, D.C.: National Academy Press.
- Scitovsky, Anne A. 1964. "An Index of the Cost of Medical Care--A Proposed New Approach." In Solomon J. Axelrod, ed., *The Economics of Health and Medical Care*. Ann Arbor: Bureau of Public Health Economics, University of Michigan.
- Schultz, Bohdan. 1994. "Choice of Price Index Formula at the Micro-aggregation Level: Issue, and a Canadian Empirical Evidence." In *Papers and Final Report*, International Conference on Price Indices, First Meeting of the International Working Group on Price Indices, Ottawa, Canada, November, pp. 93-128.
- Stigler, George, Dorothy S. Brady, Edward Denison, Irving B. Kravis, Philip J. McCarthy, Albert Rees, Richard Ruggles, and Boris C. Swerling. 1961. *The Price Statistics of the Federal Government*. In U.S. Congress, Joint Economic Committee (1961): *Government Price Statistics, Hearings, Part 1, May 1, 2, 3, 4, and 5, 1961*. Washington, D.C.: Government Printing Office.
- Statistics Canada. 1995. *The Consumer Price Index Reference Paper*. Ottawa, Canada: Minister of Industry, July.

- Szulc, Bohdan J. 1983. "Linking Price Index Numbers" In W. Erwin Diewert and Claude Montmarquette, eds. *Price Level Measurement: Proceedings from a Conference Sponsored by Statistics Canada*. Ottawa: Minister of Supply and Services Canada, pp. 537-566.
- Triplett, Jack E. 1990. "Hedonic Methods in Statistical Agency Environments: An Intellectual Biopsy." In Ernst R. Berndt and Jack E. Triplett, eds., *Fifty Years of Economic Measurement: The Jubilee of the Conference on Research in Income and Wealth*, National Bureau of Economic Research Studies in Income and Wealth, Vol. 54. Chicago, IL: University of Chicago Press, pp. 207-33.
- Triplett, Jack E. 2001. "Should the Cost-of-Living Index Provide the Conceptual Framework for a Consumer Price Index?" *The Economic Journal* 111, no. 472, June, F311-F334.
- Triplett, Jack E. 2003. "Using Scanner Data in the Consumer Price Indexes: Some Neglected Conceptual Considerations." In Robert C. Feenstra and Matthew D. Shapiro, eds., *Scanner Data and Price Indexes*, National Bureau of Economic Research Studies in Income and Wealth, Vol. 64. Chicago and London: The University of Chicago Press, pp. 151-162
- Triplett, Jack E. and Barry P. Bosworth. 2004. *Services Productivity in the United States: New Sources of Economic Growth*. Washington, D.C.: Brookings Institution Press.
- Urquhart, M.C. and K.A.H. Buckley. 1965. *Historical Indexes of Canada*. Toronto, Canada: McMillan.
- U.S. Bureau of Labor Statistics. 1997a. "The Consumer Price Index." In *BLS Handbook of Methods for Surveys and Studies* (April): 167-202.
- U.S. Bureau of Labor Statistics. 1997b. "Measurement Issues in the Consumer Price Index." In *Response to the U.S. Congress*, Joint Economic Committee, June.
- U.S. Bureau of Labor Statistics. 2002. "Discussion of the Behavior of the Chained Consumer Price Index for All Urban Consumers", available at www.bls.gov/cpi/superhist.htm.
- 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: US Government Printing Office.
- U.S. Congress, Joint Economic Committee. 1961. *Government Price Statistics: Hearings Before the Joint Economic Committee, Congress of the United States*. Eighty-seventh Congress, first session pursuant to Section 5(a) of Public Law 304 (709th Congress). Part 2. Washington, D.C.: United States Congress (May 1, 2, 3, 4 and 5).
- U.S. Commerce Department Library. 1944. *Cost of Living Index: Analysis Reports by Various Committees*. (Compilation of reports, press releases and clippings.)

Varian, Hal R. 1982. "The Nonparametric Approach to Demand Analysis." *Econometrica*, Vol. 50, No. 4. (Jul., 1982), pp. 945-974.

Varian, Hal R. 1983 "Non-Parametric Tests of Consumer Behaviour." *The Review of Economic Studies*, Vol. 50, No. 1. (Jan., 1983), pp. 99-110.

Abraham et al. 1998

Abraham 199x

Abraham 1997

BLS, xxxx

U.K. Retail Price Index