The Boskin Commission Report After a Decade

Jack E. Triplett[1]
Brookings Institution

THE ADVISORY COMMISSION TO Study the Consumer Price Index, known as the Boskin Commission after its chairman, Michael Boskin, published its final report at the end of 1996 (Boskin et al., 1996, hereafter Commission Report). The Commission Report has been exceptionally influential, not only in the United States but worldwide. Its chairman noted recently that "statistical agencies around the world have been using the Commission report as a major input to their own agendas for improvement" (Boskin, 2005:2).

To review the Commission Report and its impact, I have selected six topics. Other papers in this symposium elaborate the report's influence on U.S. data, so I specialize by discussing the report's world-wide influence (Section one). Section two concerns one of its less salutatory legacies, the increased popularity of "guestimates."

Two parts of the Commission's report have not stood up to the test of time. In its analysis of quality change bias (the largest part of its famous 1.1 percentage point CPI upward bias estimate), the Commission ignored the possibility that quality improvements could nevertheless produce a net downward bias to CPI components because the implicit quality adjustments inherent in BLS procedures might over-adjust (Section three). The Commission's analysis of CPI basic components likewise has been overtaken by more recent analysis, which has focused on consumer search behavior; the research agenda for measuring CPI basic components does not now seem to be solely Konüs-type commodity substitution, one level down, as the Commission supposed (Section four).

In Section five, I consider the Commission's neglected recommendation on CPI classifications, which I contend ought to be implemented by BLS. The concluding section of the paper includes some remarks on the politics of the Commission Report, and the lost opportunity for distinguishing more clearly the difference between improving the CPI and articulating improved bases for escalating payments to Social Security participants.

I neglect the Commission's perhaps major recommendation — that the CPI should be an approximation to a cost of living index (COLI) — because I thoroughly agree with it. My position is presented in Triplett (2001) and in Reinsdorf and Triplett (2004).

The Report's Influence World-Wide

I begin with an anecdote. The "Ottawa Group" is a recurring international conference on price indexes that is attended by many statistical agencies of the countries of the Organization for Economic Cooperation and

1 The author is Visiting Fellow, Brookings Institution. This paper was originally presented at the session "The Boskin Commission Report After a Decade: Is the CPI Still Biased?," held at the annual meeting of the American Economic Association, January 6-8, 2006, Boston, Mass. B.K. Atrostic's perceptive comments improved this paper substantially, as did suggestions from the editor. Email: jtriplett@brookings.edu

Development (OECD). At the 1994 meeting of the Ottawa Group (Statistics Canada, 1994), I introduced a discussion of the survey paper by Wynne and Sigalla (1993), which was one of the first attempts to pull together research findings in order to quantify bias in the CPI. After the discussion, representatives of several countries said: It is very interesting that in the U.S. you have these public criticisms of your CPI. That could never happen in our country.

At the time (1994), they were more or less right. Most countries had no tradition of outside reviews and criticisms by economists and statisticians of the statistical outputs produced by government agencies, and no organizations (such as the Conference on Research in Income and Wealth) to facilitate reviews and exchanges of views between analytic users and statistical agency staffs.

The Boskin Commission changed all that. After publication of the Commission Report — indeed after the Commission's "Interim Report" was published in 1995 — many countries' statistical agencies found, to their surprise, that the questions that the Commission asked were for the first time being asked of their CPIs as well. Though they had not been very vocal in the past, economists in other countries shared the Commission's conception of the measurement issues that surround the CPI, and were persuaded by the force of the Commission's analysis that CPI bias was a topic to be taken seriously.

The Commission Report, and other major U.S. studies such as Wynne and Sigalla (1993), Lebow, Roberts, and Stockton (1994), and Shapiro and Wilcox (1996), were followed by studies similar in intent, such as Oulton (1995) for the U.K., Hoffmann (1998) for Germany, and Shiratsuka (1999) for Japan. But these published or formal studies are just the visible tip of the ferment, for more commonly central banks and treasury or finance ministries posed questions to their statistical agencies that were conditioned by the Commission Report, or by the controversy in the U.S. that originally spawned the Commission. Some of these queries led statistical agencies to publish their own self-evaluations, Lequillier (1997) for France being an example (Wynne and Rodriguez-Palenzuela, 2001, mention some others).

Many statistical agencies outside the U.S. initially responded to the Boskin Commission's bias estimates by claiming that they did not apply to their indexes. Eurostat, the statistical agency of the European community, summarized this point of view in its response to an OECD country survey in 1997 (OECD, 1997:27): "Most member states feel that Boskin biases apply to the United States's CPI and that the United States aims at a COLI while that is not the case in the EU. There is no bias that could arise due to the use of the HICP as a basis for measuring inflation instead of any sort of COLI." [The HICP are the harmonized indexes of consumer prices that Eurostat designed, with input from member states, for all of the European Union countries.] That is, because the HICP were not intended to be COLIs, Eurostat was contending that they were not biased measures of inflation.[2]

International response to the Commission's detailed bias estimates typically took the following forms.

Upper Level Substitution Bias

Upper level substitution bias is the classic Konüs (1925) type substitution in response to relative price change among the components of the CPI. The Commission's point estimate of bias from upper level substitution was 0.15 index points per year.

As noted above, many countries claimed that their indexes were not intended to be cost of

2 Eurostat and some other countries' statistical agencies insisted that inflation was not measured by a COLI. A paper by Peter Hill conveys the joint position (Hill, 1997). I did not find it convincing, for the reasons in Triplett (2001).

living indexes (COLI). In any case, they also pointed out that they updated their weights more frequently than did the BLS. At the time the Commission wrote, weights in the U.S. CPI were about 13 years old, which was indeed older than CPI weights for any other major OECD country. Eurostat's standard for HICP indexes was one to two years, ideally, and no more than five years maximum. Thus, their assertions had some plausibility, though no other country outside the U.S. had an actual estimate of substitution bias at that time.[3]

Lower Level Substitution Bias

Substitution within a CPI detailed component (examples of detailed components are bananas in the fruit category and musical instruments in the entertainment category) the Commission called lower level substitution bias. The Commission set lower level substitution bias at 0.25 points, or more than upper-level substitution bias, and suggested that the BLS move to the unweighted geometric mean formula, in place of the arithmetic average of price relatives (APR) that had been in place since 1978.[4]

European countries were quick to point out that Eurostat had already endorsed the geometric mean formula for basic components in HICPs, and indeed had banned the BLS APR method. Accordingly, they believed their indexes were immune to lower level substitution bias (even if they were considered as COLIs, which they were not). Again, this European contention had merit, though some countries were subsequently discovered still using arithmetic means.

Outlet Substitution Bias

This bias arises from consumers shifting to lower cost outlets. Outlet substitution behavior is not recorded in the CPI because the CPI is constructed from a matched outlet sample. The Commission put outlet substitution bias at 0.10 index points annually, but it did not recommend any CPI improvements that would reduce the bias, other than perhaps bringing new outlets more quickly into the sample.

Many countries claimed that outlet substitution was no problem in their CPIs because their retail sectors were less dynamic than United States retailing. It is true that strong multi-factor productivity (MFP) growth in the U.S. retail sector in the 1990s (Triplett and Bosworth, 2004) seems without a parallel in Europe at the same time (Timmer and Inklaar, 2005), which would be evidence in favor of the European position.

On the other hand, at the time of the Commission Report the only existing European study relating to outlet effects in the CPI (Saglio, 1994, on chocolate bars in France) found substantial shifts from traditional small French retail shops to hyper-markets and so forth. Covas and Silva (1999) found outlet bias for Portugal that varied by year and by methodology; it averaged a little under 0.1 point per year in one method (which the authors believed understated substitution) and 0.49 in the other, probably an overstatement. These estimates, then, are not lower than the Commission's estimate for the U.S.

Another factor bearing on this whole matter is sampling. Few countries outside the U.S. employ much of a probability sample for items and outlets in their CPIs. The U.K. and Sweden choose outlets by probability sample, but they are exceptions (an international comparison of CPI methods is OECD, 1997). Outlet bias might well be larger in a non-probability sample than in a probability sample (the Stigler Committee thought so — see Price Statistics Review Committee, 1961), though this might not be outlet *substitution* bias.

3 This has now changed. For a sophisticated substitution bias estimate for the U.K., see Blow and Crawford (2001), who implemented the procedures used for the U.S. by Manser and McDonald (1988).

4 In Section four, I suggest that the Commission and the BLS were too quick to judgment on the lower level substitution bias question.

On balance, it is not clear that other countries' protestations — that their CPIs were not subject to outlet substitution bias — were correct.

Quality Change Error

The Commission put error arising from inadequate adjustments for quality change at +0.60 index points per year. As suggested in Section three, this estimate is very shaky. Surprisingly, the Commission did not recommend that BLS use hedonic methods to produce better quality adjusted indexes.

Some countries contended that their quality change error was lower than in the U.S. CPI. Some of their comments revealed lack of knowledge about U.S. quality change procedures, and also, more surprisingly, lack of understanding of the biases in the quality adjustment procedures that many of them were using.[5] The Eurostat statement mentioned earlier maintained that in the CPI the "main bias is agreed to come from quality changes. Since there is no agreement on what could or should be done, there can be no agreement on what biases exist" (quoted in OECD, 1997:27).

Quality change is a major problem in all countries' CPIs, and little reason exists for believing that European countries have devised better methods than those used in North America. Dalén (2002) and Ribe (2002) provide revealing analyses of European country practices on quality change. Hoffman (1998) showed that differences in treating quality change in the German CPI created enormous differences in the price indexes for household appliances,[6] and Silver and his colleagues found substantial quality change error in U.K. price indexes (for example, Silver and Heravi, 2001). No OECD country, with the exceptions of France and Sweden, had any hedonic price indexes for computers before the Commission wrote (of course, neither did the U.S. CPI), and in most of them the environment was very hostile toward hedonic indexes. By roughly the turn of the century, however, countries outside North America became more receptive to new methods, perhaps for more reasons than the ferment caused by the Commission Report, but the report certainly contributed to their changed veiwpoints.

Overall Impact

The Commission Report had an extremely salutatory effect on international price statistics. Countries differ greatly in the quality of their statistics, in their agencies' receptiveness to new methods, and in their lines of communications with analytic users. Accordingly, generalizations do not fit individual country experiences. Yet, several generalizations apply across many countries.

- The report brought into open discussion a set of price measurement issues that are properly concerns in every country, not just in the U.S.
- The report, and the controversy that surrounded it, engendered a more extensive dialogue between statistical agencies and their analytic users than had been the norm before.
- The report, and the world-wide discussion of it, jarred some agencies out of their complacency about the quality of the numbers they were producing.
- The report also encouraged price index research among economists and statisticians in countries outside of North America, where it had been difficult to find before. For example, a 2001 Workshop in the Measurement of Price Indices sponsored by the European Central Bank and the Centre for Economic Policy Research contained

5 This is a more complex matter than is sometimes appreciated. An analysis of all methods for handling quality change in price indexes, and the probable errors in each of them, is found in Chapter 2 of Triplett (2004).

6 Production of the German CPI is not centralized, unlike the situations in the U.S., Canada, and most other OECD countries.

contributions on all the price measurement topics addressed by the Commission and estimates of price index bias that applied to a variety of components and countries (Camba-Mendez, Gaspar and Wynne, 2001). Wynne and Rodriguez-Palenzuela (2001) pointed to much recent research but even so, found insufficient research in Europe to permit an evaluation of measurement bias in HICP indexes.

Guestimates

Ironically, what I liked least about the Commission Report was exactly what made it so influential — its guestimate of 1.1 percentage points of bias.

A guestimate is a number produced when one does not have research results. The Commission (and others who have followed its lead) used ad hoc reasoning to come up with a number. I did not think that we economists knew enough to estimate the overall CPI bias when the Commission wrote its report, and I still think we do not know enough.

But this seemingly so precise 1.1 number caught the eyes not only of the press and the politicians, but also of economists in the U.S. and in other countries. Jacob Ryten wrote just after the release of the Commission's interim report: "...for the first time ever, a blue ribbon commission dared give a number for the estimate of total bias and detailed each of the contributing factors" (Ryten, in Ducharme, 1997:3). Without the guestimates, the Commission Report was just another dry, academic study to be perused by professionals (and perhaps only those professionals who were previously interested in price index research). Without the guestimates, the report would likely have had minimal impact.

Conversations with Commission members suggest that some, at least, were ill at ease themselves with guestimates, but I take it that they felt that their mandate from the Senate Finance Committee compelled them to brew one. The report notes the shortage of research findings on which to base a better and more scientific estimate. Morever, the Commission presented a "plausible range" (also a guestimate, of course) of 0.80 to 1.60 index points, so they themselves were not asserting the degree of precision that some of their readers took away from the report.

The Commission's 1.1 point guestimate, made to carry out their Senate Finance Committee mandate, is one thing. My personal preference is to resist the seductive blandishments of politics and politicians, but not everyone feels that way. But the Commission's 1.1 point guestimate seems to have legitimized the making of guestimates by professionals who are not under the sway of the Senate Finance Committee, and who seem eager to charge into what apparently the Commission members entered into with some reluctance.

I should specify more precisely what I mean by guestimate, since they have now been so much used that the distinction I am drawing may no longer be clear to some economists. By guestimate, I do not mean a compilation of research results. The Commission's estimate of substitution bias (0.15 points) was not a guestimate. It was a professional distillation of research results that encompassed a number of different studies (most of them conducted within the BLS) that used a variety of methods and reached closely similar quantitative conclusions on the size of the bias.[7]

Little controversy exists about the Commission's substitution bias estimate because the Commission could cite the research studies on which its estimate was based. Of course,

7 The major studies are Christensen and Manser (1976), Manser (1975), Braithwait (1980), Manser and McDonald (1988), and Aizcorbe and Jackman (1993). BLS research on substitution bias is summarized in Reinsdorf and Triplett (2004). Recent BLS publications suggest a somewhat larger substitution bias estimate, but those BLS numbers are suspect because they commingle commodity substitution with area effects (substitution of cabbage in New York for carrots in Los Angeles).

judgment is involved in integrating these studies. I believe that any competent group of professionals, viewing the same evidence, would come up with a similar judgment.

However, for its measure of the impact of quality change (at 0.6 points, much larger than its substitution bias estimate), the Commission had no comparable body of research on which to rely. Instead, the commission entered into a variety of more or less ad hoc reasoning exercises.

In some cases the Commission's indirect logic seems persuasive. But persuasive or not, its conclusions were not backed up by firm empirical estimates of the size of the bias. There is no contradiction in saying at the same time that the Commission's guestimates were better than some previous ones (partly because they derived them on a component-by-component basis, which imposes a certain consistency and check on the total) and in saying that they were nevertheless inadequate. They are inadequate in terms of what one should demand in a scientific estimate — that it is evidenced-based and reproducible. Evidence-based, in this case, means a comparison of actual CPI indexes with others computed on a different basis or with different data, such as comparisons involving conventional methods for handling quality change and hedonic indexes, or else an empirical analysis of the impacts of applying alternative methods.

The guestimates made parts of the report's 1.1 point bias estimate unreliable, especially its 0.6 point estimate for quality change and to a lesser degree its 0.1 point estimate for what the Commission called lower level substitution bias. More regrettably, the Commission's guestimates have spawned a flurry of other guestimates in the U.S. and in other countries. Recent ones are Lebow and Rudd (2003) for the U.S. Federal Reserve Board and Rossiter (2005) for the Bank of Canada. Especially in their estimates of quality change error, these new guestimates are flawed in the same way that the Commission's guestimates were flawed, and for less reason. In the next section, I give some concrete reasons why guesstimates fail.

In my view, the economics profession would be better served if the resources put into producing new guestimates were instead devoted to price index research. New research would not only improve our information on the accuracy of the CPI, but perhaps as well improve the index. [8]

The Boskin Commission on Quality Change

Quality change is the biggest problem in obtaining accurate price indexes. The Commission was surely right that estimating the bias caused by quality change — in fact, doing any meaningful research on quality change — must proceed on a component by component basis. To illustrate the principle, it is now well established that the rate of quality-corrected price decline differs substantially across different categories of computer and communications equipment, and across different categories of semiconductors (see the review of studies by Mark Doms, Ana Aizcorbe, and collaborators, in Triplett and

8 In an interesting parallel, the Mitchell Committee that investigated alleged bias in the CPI during World War II guestimated that the downward bias (not upward bias) in the index from forced trading up and quality deterioration was 1-3 per cent for food, 4-5 per cent for clothing, and 8-11 per cent for house furnishings. These numbers are all several multiples of the comparable component estimates in Boskin et al. (1996, Table 2), and of course are of opposite sign because of wartime conditions. George Stigler, head of the 1961 Stigler Committee on price indexes, declined to make an estimate when asked by a Congressional committee to do so, saying that the committee did not know enough. However, one of the Committee's members, Richard Ruggles, inserted a footnote suggesting 3 percentage points of bias, a guestimate nearly 3 times that of the Boskin Commission. The Mitchell and Stigler reports are reviewed in Reinsdorf and Triplett (2004). I suspect that the clear downward trend in the magnitudes of professional guestimates about CPI bias does not reflect entirely a consensus that the BLS is doing a better job. Rather, more recent professional judgments draw on more empirical research. Speculative judgments that are not informed by research seem invariably to produce larger numbers than are documented in research findings.

Bosworth, 2004, chapter 10). Each component needs its own study. One cannot just presume that "electronics" price indexes fall equally. The same is true of components in other product groupings. There is no substitute for what Shapiro and Wilcox (1996) termed the "house to house combat" approach to quality change.

The Commission examined research estimates for those cases where research existed. Some research missed their net, and in one or two cases, regrettably, they substituted their own judgments for research findings. However, as the Commission also observed, in many categories "There is little if any published evidence that allows us to reach a determination" (Boskin et al., 1996:41). For those cases, they used their intuitions and a variety of indirect data and hypotheses — guestimates.

What is logically wrong with guestimates of quality change error?

Most economists, I suspect, believe that our dynamic and competitive economy generates an overall improvement in the quality and variety of goods and services that are available for consumption. I share that presumption. Counter examples can be cited — the most recent versions of some software that I use may not be better than older versions, quality deterioration in some services is often alleged, and some traditional manufactured products (desk telephone sets and toasters) seem less reliable than their ancestors. But overall, the direction of quality change is undoubtedly positive, not negative.

From the presumption that quality is on balance improved, economists have often inferred that quality change must bias price indexes upward, that $\Delta Ql > 0$ implies $\beta > 0$ (where I use ΔQl to represent quality change, or its value, and β to indicate the price index bias). The converse expectation is thought for those cases when $\Delta Ql < 0$. Accordingly, an additional inference goes, if one can form some estimate of the direction and size of ΔQl, that information can be used to estimate the price index bias. Even if the price index agency makes some adjustments, these economists infer that information on ΔQl will at least provide a bound on β.

All three of these inferences are wrong. Information on ΔQl, even if it were reliable, does not determine the magnitude of β, *or even its sign*.

The quality error that is incorporated into the index must be a product of:

a) the type and extent of quality change that takes place and
b) what the price index agency did about it.

If the price index agency did nothing, of course, either ignored the quality change or did not notice it, then the full amount of the quality change passes through into a bias in the price index. In this case *only*, quality improvements cause upward bias to price indexes — that is, $\Delta Ql > 0$ implies $\beta > 0$ — and quality deteriorations results in downward biased price indexes.

Obviously, if the price index agency does something about the quality change that it observes, then the full amount of the quality error does not pass through untouched into index number bias. That is probably well understood.

Less well understood is that most of the procedures used by price index agencies create *implicit* quality adjustments when quality change is encountered. Those implicit adjustments can be either too small or too large, and a fair amount of evidence has accumulated that the implicit adjustments are frequently too large. *When the implicit adjustments are too large, quality change can create an error of the opposite sign from the sign of the quality change itself.*

Why is the sign of the error not determined by the sign of the quality change?

I first examine the Canadian CPI because its method is somewhat easier to analyze. Statistics

Canada frequently uses the "link-to-show-no-change" method when quality changes are encountered (see Statistics Canada, 1995). Under this method, if a new model has higher quality than the model that it replaces and its price is also higher, the price difference between the new model and the old one is taken as the value of the quality change, so: $\Delta P \approx$ est ΔQl (the estimated quality adjustment). Note that the new and old typically are not observed in the same month, so $\Delta P = P_{nt} / P_{o,t-1}$, where n and o designate new and old models. This implies that when the new model first enters the sample, *it can generate no price change in the index*, by construction.

Suppose that the seller takes the opportunity of the new model to introduce a price increase, above the value of the quality change the new model represents. In this case, $\Delta P > \Delta Ql$. The implicit quality adjustment is too high, relative to the true quality change, because est ΔQl $(\equiv \Delta P) > \Delta Ql$; the price index bias is negative, $\beta<0$, even though the quality change is positive.

Conversely, suppose that $\Delta P < \Delta Ql$ — either the price increase is less than the value of the quality change or some price decline accompanies the improved new model (frequently the case with electronics, for example). Then, the implicit quality adjustment given by ΔP is too low; any price decline is missed or adjusted out of the index, and $\beta > 0$, the bias is positive.

Even when quality is improving, therefore, *the quality change bias could go either way*. The bias depends on the ratio of $\Delta P / \Delta Ql$, and whether that ratio is greater or lesser than unity; it does not depend on whether $\Delta Ql > 0$ or $0 < \Delta Ql$, that is, whether quality is improving or deteriorating.

Note that the quality change bias is more nearly a function of the direction of price change than of the direction of quality change. When prices are rising, one can infer that the index will be biased downward, it will show too little inflation because the method tends to remove price increases from the index. The proposition is completely symmetric: If prices are falling (true of electronic goods), the bias is upward because the method picks up too little of the price decline, which is one reason that hedonic indexes for electronic goods so frequently fall more than indexes constructed with the usual statistical agency methods (a review of studies is in Triplett, 2004, chapter 4). It has been shown repeatedly that in many CPI components price changes occur most frequently when quality changes occur (Moulton and Moses (1997), Armknecht and Weyback (1989), and the discussion of Armknecht's findings in Shapiro and Wilcox (1996)).

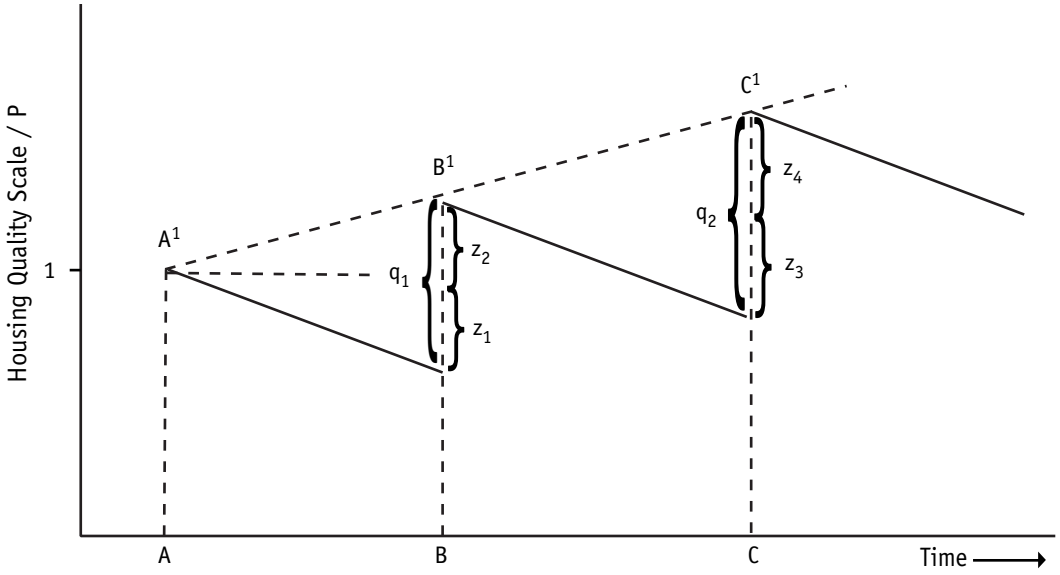
In cases where quality is deteriorating, a parallel analysis applies, which need not be detailed here.

The U.S. does not use Statistics Canada's link-to-show-no-change method in the CPI, but rather another linking method often termed "deletion." In the deletion linking method, the prices of the item whose quality changed (both the old item and its replacement) are deleted from the index for the adjacent periods in question. The *price change* for the missing item (not its quality change, and not its price level, as sometimes said) is imputed from price movements of other items that have not changed in quality. Again, this method creates an implicit quality adjustment. [9]

Although the U.S. method is different from the Statistics Canada method, the analysis is similar: The quality bias in the index depends on the size of the implicit quality adjustment and whether the implicit quality adjustment is too high or too low. The BLS found a number of years ago that the deletion method is biased (though not so much as the Canadian method), so it now uses a modification of the deletion method developed by Armknecht (Armknecht and Weyback, 1989), in an attempt to reduce the

9 The deletion method, its probable bias, and the algebra that provides its analysis are presented in Triplett (2004, Chapter two).

Chart 1



bias. They call the modification the "class mean" method. The best empirical evaluation of the implications of the deletion method is Moulton and Moses (1997), whose results suggest that it is on balance biased toward overadjusting for quality improvements.

No guestimate of quality change bias has paid much attention to where the bias comes from and how its origin influences one's expectations about it. The guestimating methods that have been used to estimate CPI quality change bias (β) are predominantly guestimates about ΔQl. But the sign of the quality error in the index is not determined by whether quality is improving or deteriorating, and the amount of the error or bias is not determined by the amount of quality change. Accordingly, the guestimates are flawed, possibly even with respect to their sign. The following example illustrates.

An Example: Housing

The Commission made a guestimate for how much it thought the quality of housing had improved and used that guestimate as the basis for its CPI housing bias estimate. Studies have shown (the major one is Randolph, 1988) that even though housing quality is steadily improving, the quality bias is negative in the CPI rent and homeowners equivalent rent housing components, unless a correction is made.

In Chart 1, the vertical axis measures simultaneously a housing quality scale and the price of rental housing (I assume for simplicity in drawing the diagram that competitive markets operate to bring about this equality instantaneously). At point A, a new rental housing unit comes into the sample. I normalize its price and quality at 1.0, and assume that no housing inflation is taking place (again for simplicity in drawing the diagram).

In use, the rental unit begins a slow, probably imperceptible, process of deterioration depicted in Chart 1 as a decline in the unit's quality (and price) over time. The monthly deterioration would not normally be detected when BLS collects the price for this rental unit for the CPI. The price index is accordingly biased downward (the term in the literature is "aging bias") over the interval from A to B because the price/quality decline shown from A to B should have been adjusted out of the index.

At point "B" on Chart 1, the rental unit is renovated. It may not only be painted, repaired, and

so forth, but it may also receive new and better appliances, perhaps it is centrally air conditioned where it was not before, and so forth. Thus, the renovation may create a rental unit of higher quality than it was originally, by the amount z_2 in Chart 1. The amount z_2 is the correct measure of improvement in housing quality, and if the Commission (or other guestimators) arrived at the right number, they would have estimated improved housing quality in the amount of z_2.

However, the price index agency would never compare the deteriorated unit with the renovated unit, they differ too much to make a valid price index comparison. If the agency used Statistics Canada's method (the simplest to show on the diagram), it would apply the quality adjustment q_1, that is, the sum of z_1 and z_2, as the quality adjustment. This quality adjustment is too large because z_1 is just the amount that restores the rental unit's quality to its as-new level. Put another way, when quality deterioration occurred with aging, no adjustment was made because the monthly changes were too small to be observed, so the index was biased downward. Then when the quality was restored, there is so much difference in the unit that it does not meet the standards for a the month-to-month "match," so an implicit quality adjustment would be made — the change in rent associated with the renovation is linked out. That means, however, not only that the negative quality error from aging bias remains in the index, but also that the value of its corresponding renovation is linked out of the index (the adjustment for z_2 is not at issue). Overall, the index is biased downward, even if on average housing quality is improving, as shown in Chart 1.

The U.S. CPI contains an aging correction, based on Randolph (1988). Aging bias has become very well known in the housing literature — see, for recent examples, McCarthy and Peach (2002) and Krone, Nakamura, and Voith (2002). Aging bias was rejected by the Commission, (but see Gordon's contribution to this symposium), and it is apparently not known at all by those who have followed the Commission by trying to estimate, in effect, z_2 and adjusting it out of the index.

Wrap Up

The implications of the linking and deletion methods have been in the price index literature for a long time (at least since Triplett, 1971). The Commission ignored the matter, and though I think that was a flaw in their report, perhaps at the time the Commission wrote, it was still regarded as controversial. But since then, Moulton and Moses (1997) produced for the first time estimates of the aggregate impact of the method, and the Committee on National Statistics Panel report (Schultze and Mackie, 2002) spelled out clearly and plainly the price index implications of this form of handling quality change. Gordon's (2003, 2004) studies on clothing and housing imply that the same negative bias from linking in the matched model method infests historical price measures (that is, he confirms with historical data the earlier BLS methodological studies by Liegey (1993) on clothing and Randoph (1988) on housing). Yet, the implications of deletion, and other methods that the agencies use for handling quality change, are still ignored in the most recent U.S. guestimate (Lebow and Rudd, 2003) and the implications of Statistics Canada's somewhat different methods are similarly overlooked by Rossiter (2005).

Boskin et al. (1996:32) contend that generating a positive number as the quality change bias estimate is better than accepting zero, because zero "represents an extreme one-sided answer to the question as to whether the components of the CPI subject to relatively little research are biased." The analysis in this section, however, shows that when quality is improving, the lower bound on the quality error that is incorporated into the index is not zero. Conversely, when

_quality is deteriorating, the upper bound of the quality change error is not zero. When one has no empirical information on the magnitude of the quality change error, one also does not know its sign. That is, $\Delta Ql > 0$ may imply $\beta > 0$ or $\beta < 0$, and similarly $\Delta Ql < 0$ may imply $\beta > 0$ or $\beta < 0$. Accordingly, zero may be the best point estimate.

Extrapolations are not always invalid. In some circumstances, one has information on related products or related sampling conditions that may facilitate estimating a bound. For example, studies of electronics have overwhelmingly shown that prices of those products are falling. One might well extrapolate falling price indexes to electronic products that have not been studied, on the grounds that existing studies have also shown the reasons why the conventional linking techniques have missed price decline. Or, one might be willing to extrapolate results from one study to a class of related products that are thought to exhibit similar behavior, or similar pricing and measurement problems.

The point is: One needs a study to extrapolate from. Without some empirical work to base them on, guestimates have little value.[10]

What the price index agencies do to adjust for the quality changes they encounter ("inside the sample" quality change) is but one aspect of the more general problem of quality change. For example, bias may occur because the sampling methodology systematically misses the varieties of products on which quality change occurs ("outside the sample" price and quality change, such as in Berndt, Griliches and Rosset (1993), and other research questions exist. But guestimates do not address them adequately, either.

The Commission on CPI Basic Components[11]

Examples of CPI basic components include musical instruments in the entertainment category and bananas in the food indexes. CPI basic components are formed from microdata collected from matched retail outlets. Two measurement problems have dominated recent discussion in the U.S. — what the Commission called "lower level substitution bias" and the related question that the BLS called "formula bias."

Formula bias was caused by procedures adopted in the 1978 revision of the CPI to implement probability sampling for varieties and outlets, as recommended by the Stigler committee.[12] Before 1978, the (unweighted) formula for a CPI basic component was: $(\Sigma\, p_{i,\,t+1} / n) / (\Sigma\, p_{it} / n)$, where t is the last price observation (month for monthly pricing), t+1 the current month, and the calculation was done separately for each city. In words, then, the basic component index 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. In the index number literature, this is known as the ratio of average prices method, or RAP. Note that it is a form of arithmetic mean.

After 1978, BLS estimated outlet sampling probabilities from a new "Point of Purchase survey" and item selection was done by probability methods within each retail outlet. However, random sampling of varieties and quality levels precluded continued use of averages of prices. If a piano and a guitar pick are probability selec-

10 Lebow and Rudd (2003) present a table that lists CPI components, their estimated biases, and the sources for the authors' bias estimates. Components for which they believe they have good research estimates amount to only 7 per cent of the CPI and contribute only 0.02 points to their total bias estimate. Components for which they have no research whatever account for over half the CPI, and they themselves mark these estimates as "almost entirely subjective." They remark that this table is "sobering;" it ought to have been discouraging.

11 Portions of this section are drawn from Section II.C in Reinsdorf and Triplett (2004), which also reviews related topics in reports of the Mitchell, Stigler and Committe on National Statistics panels, in 1945, 1961, and 2002, respectively.

12 Probability sampling of outlets, but not varieties, began with the 1964 revision of the CPI.

tions for musical instruments, an example presented by Moulton and Moses (1997), the average price movement would be dominated by the piano. This suggested changing the computation for basic component indexes from ratios of average prices (or RAP), used before 1978, to averages of price ratios (APR), used thereafter. The simple APR discussed by McCarthy (1961) was an unbiased sample estimator of a Laspeyres price index **if** the specific items priced were selected with probabilities proportional to expenditures in the base period and price collection began in the base period.

However, neither McCarthy (1961), nor Adelman (1958), nor Westat (the statistical consulting firm brought in to design the BLS move to probability sampling) considered problems that arose in practical application. The BLS produced some long and complicated analyses that focused on the expectation of the APR sample estimator with respect to a Laspeyres population parameter, and how the particular implementation of that estimator was biased because the perfect measurement of base period expenditures and prices assumed by McCarthy is far from achievable in practice (Reinsdorf, 1998; Reinsdorf and Moulton, 1997). BLS called the resulting bias "formula bias," and proposed a solution (they called it "seasoning") that attenuated the bias.

Additionally, Reinsdorf (1993) showed that use of a geometric mean as the basic component aggregator in the CPI gave lower rates of price change than the ARP that the BLS had been using. As soon as "geometric-arithmetic" appeared in the discussion, some economists seized on the apparent parallel between a Laspeyres index (weighted arithmetic mean) and a superlative index (weighted geometric mean or a geometric combination of arithmetic means) to raise the well-known Konüs substitution bias as an explanation for the geo-arithmetic mean difference at the basic component level. Before the dust had settled, the Commission appeared on the scene and took up this "lower level substitution bias" explanation for the difference between APR and geometric mean indexes. The Commission proposed that the BLS use the geometric mean to bring basic components more closely in line with COLI theory, a recommendation that the BLS accepted for most components of the CPI.

What Economic Behavior Applies?

Economists have sometimes interpreted the difference between arithmetic mean and geometric mean aggregators for basic components as just the classic substitution bias paradigm drawn from Konüs (1925), only applied one level down. In this context, much (probably too much) has been made of the fact that Cobb-Douglas behavior justifies a (weighted) geometric mean price index — that is, a weighted geometric index is a COLI if the elasticity of substitution is everywhere equal to unity. The question concerns unweighted means, not weighted indexes.

It is a mere mechanical fact that an unweighted arithmetic mean of positive quantities will be greater than an unweighted geometric mean. The difference between the two is not evidence of substitution bias. No inference about lower level substitution can be drawn from the fact that the geometric mean basic component gives a lower estimate of price change than the arithmetic mean, since that will always be the case, substitution or no. The fixed-weight Laspeyres index, on the other hand, only exceeds the COLI when commodity substitution takes place and will equal the COLI when substitution is zero.

Empirically, geo-arithmetic mean divergences have usually been associated with "price bouncing," the periodic if not monthly sales that dominate some sections of retailing. Schultz (1994),[13] though not the first study of

13 An earlier paper by Carruthers, Sellwood and Ward (1980) influenced Europeans more than North Americans, one example of numerous trans-Atlantic divides in price statistics.

its kind, brought the formula for basic components to attention in North America. He reported huge variances from different formulas applied to microdata from Statistics Canada's CPI. His subject was a single brand and size of soft drink in a single Canadian city — clearly not a *commodity* substitution story of the classic Konüs kind. Prices for this soft drink fluctuated, often dramatically, from month to month in a single retail outlet. When prices returned to normal from the sales prices, the increases dominated the movements in an arithmetic mean but less so in a geometric mean, even though sales weights were not available,.

This example and others suggest that the arithmetic mean-geometric mean differences that dominated the discussion at the time the Commission was preparing its report were not, as the Commission supposed, evidence of commodity substitution at the lower level (Granny Smith for American Delicious apples was a favorite example). To be sure, substitution must take place there. But the ubiquity of the difference even when substitution was unlikely to provide the full explanation suggests looking beyond lower level substitution.

The Commission was evidently unimpressed with the BLS analysis of formula bias, the bias of the estimator with respect to the population Laspeyres index. How highly does one rank exact computation of a Laspeyres index if the objective is to produce a COLI? I agree with the Commission that the COLI is the way to think about the problem of CPI basic components.

Shopping behavior vs. substitution behavior in a COLI

Commodity substitution behavior is undoubtedly one relevant concern for basic components since many components are made up from samples of substitutable commodities. When consumers substitute in response to relative price changes within a basic component, then the geometric index is a better approximation to the COLI, and that is all that needs to be said.

A theory of basic components, however, must be applicable to all basic components, not just some of them. It must explain differences between arithmetic and geometric means for components such as the CPI banana price index, or the Schultz (1994) study of a single brand of soft drink, cited above. Surely there is no room for commodity substitution within a single size and brand of one product.

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. The standard model of commodity substitution — though clearly relevant to the construction of basic components — must be supplemented by a model of consumer search with costly and imperfect information, combined with inventory and storage behavior, as Pollak (1998), Feenstra and Shapiro (2003), and Triplett (2003) contend.[14] The Committee on National Statistics Panel (Schultze and Mackie, 2002:24) called for more research on CPI basic components because: "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."

Search, storage and so forth are not necessarily modeled by simply switching to a superlative index or a geometric index, since the theory that lies behind those index number formulas is not the theory that explains consumer search and storage behavior. To apply to basic component index numbers, a theory of consumer behavior must model consumers' choices across *sellers* of a homogeneous commodity, not just consumers'

14 See Baye (1985), Anglin and Baye (1987), and Reinsdorf (1993).

choices across different (substitutable) commodities. A geometric mean of matched sellers prices is not in any way related to household search and inventory behavior, nor does it effectively incorporate it into the CPI.

The CPI collection strategy implies that the price frontiers faced by individual consumer units can be calculated from *prices collected from matched retail outlets*. This is demonstrably wrong. Indeed, Triplett (2003) presents a simple numerical example to show that with an imputation for search costs, no standard price index formula applied to prices collected from matched retail outlets will measure the COLI of households who shop. Hendel and Nevo (2002) show that neglect of consumer storage and shopping behavior results in an over-estimate of ordinary demand elasticities, surely a fatal problem if one proposes to model index number substitution bias at the basic component level with a simple Konüs system.

In some cases characterized by shopping behavior, a unit value index may perhaps be justified. If the average price paid drops because information has become easier to obtain, so that consumers are better able to find the lowest prices, the COLI should decline, even though no price has changed. On the other hand, Bradley (2005) shows deficiencies of unit value indexes for almost every purpose, so more analysis of the shopping/search behavior question will be required.

BLS Change to Geometric Means

Implementing the recommendation of the Commission, the BLS changed most (but not all) components of the CPI to the geometric mean formula in 1999. Their unpublished study in support of the change analyzed commodity substitution at a detailed level, thus showing BLS acceptance of the commodity substitution paradigm at the basic component level, as proposed by the Commission. The unpublished study was exemplary. But the difference between geometric and arithmetic mean indexes for commodities such as bananas cannot, as we have noted above, be explained by commodity substitution. The study was incomplete.[15]

A crucial issue is the source of the price data. Laspeyres and Paasche indexes computed from matched sellers prices are not necessarily bounds on the index needed for a COLI in the presence of costly information, because consumers switch between sellers. Studies of commodity search and shopping behavior and also consumer storage behavior (stocking up at sales price time) will need data collected from households — that is, purchasers' prices — not just data obtained from matched sellers.

Attempts to fit the basic component problem into the standard Konüs commodity substitution model lack insight into the nature of the problem and have yielded misleading conclusions. Pollak (1998:73) 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." Both the Commission and the BLS rushed to judgment on the geometric mean. Likely, the geometric mean will prove better than the arithmetic mean of price relatives, for statistical reasons, if no other. However, COLI theory supports neither the Commission's recommendation nor the BLS action: More research is required to determine why the geometric mean is a better measure, if it is one.

15 BLS found no studies that could be used to analyze commodity substitution within the individual basic components, so the analysis proceeded at a higher level, e.g., fruit (a first-level aggregation), not bananas and citrus fruits, which are basic components in this case. BLS accepted the geometric mean for a basic component when substitution was non-neglible at the first-level aggregate, but retained the arithmetic mean when less substitution was found (e.g., medical items). Strictly, the BLS findings supported abandoning the Laspeyres index for the first-level aggregate in the CPI — combining apples and oranges, say, into an index for fruit. It did not support the decision to change the formula for the basic components to the geometric mean.

A Neglected Recommendation: Classification of CPI Commodities.

The Commission's eighth recommendation to the BLS reads: "The BLS should investigate the impact of classification, that is item group definition and structure..." So far as I can determine, this seems to have brought no comment elsewhere and no response of any kind from BLS. The Commission's recommendation is a good one, and BLS ought to take it seriously.

Classifications are the kernels of economic statistics, they group economic data into the units that economists use for their analysis. Classifications — that is to say, groupings — are vital to economic analysis because the way that data are grouped limits the analyses that can be done with them. It is accordingly surprising how little attention economists pay to their classifications — unlike biologists, who understand that classifications need to be done according to a theory.

The theory of economic classifications is the economic theory of aggregation. For the CPI, the relevant research is small, the main items being Pollak (1975), Blackorby and Russell (1978), and Triplett (1990). The theory says that CPI groupings ("fruit," or the higher level aggregate "food") depend on separability of the direct or indirect utility function. Obviously, the theory is hard to implement empirically, but other theoretical abstractions (substitution bias) are now tractable, and so too might this one become, with sufficient work.

As the Commission suggested, better groupings might facilitate analysis of lower level substitution. Regrettably, the BLS has recently implemented a new CPI classification that makes no reference to economic theory. As well, the international agencies seem committed to atheoretical approaches, as demonstrated in the new international price index manuals. More work needs to be done on the theory, and BLS (and others) should try to work out implementations that are more consistent with the theory, which will be a challenging task.

Conclusion (and Politics)

The Commission Report has been very influential and many of its conclusions stand up to research in the intervening period. Sections three and four of this paper suggest that its reviews of quality change bias and of methods for estimating CPI basic components would need rethinking if the Commission were writing its report now. But I suspect that members of the Commission did not believe they were writing a timeless document. Moreover, they are not to blame if their imitators at making guestimates have not always followed best practice (Section two).

Greg Mankiw, in commenting on the paper by Shapiro and Wilcox (1996:154), "expressed the view that the current debate about the CPI was really a political debate about how, and by how much, to cut real entitlements." Quite so.

The Commission wrote a technical document about CPI measurement. All of the members, no doubt, were aware why Congress was so interested in arcane details of the CPI: Senator Moynihan's complaints about deficits that stretched "as far as the eye can see" were not made in a private setting. Being aware of the political setting is not the same thing as being a captive of it, though the Commission Report did repeat much of the rhetoric of the time, and indeed succumbed to the lure of political statements in its choice of language to describe the effect of CPI measurement errors on Social Security expenditures.

In the debate after release of the Commission Report, political charges were made. It should have been possible to have a debate on the merits of the Commission's report on the CPI that was separate from a debate on whether the depen-

dent part of the population was or was not getting a fair share (whatever that means, which is part of the debate) of the social product. Professionals, at any rate, should understand that improving the accuracy of the CPI is not the same thing as improving the basis for the allocation to the dependent population, and not even the same thing as improving the basis for escalation of their Social Security payments. Zvi Griliches made that point in his testimony before the Senate Finance committee, and it is perhaps instructive that his wisdom on this won few adherents (but see the well stated passages in the Berndt and Baily papers in this symposium).

Many criticisms of the Commission's report were motivated by the belief that benefits to Social Security recipients should not be cut, rather than balanced judgments about the CPI or about the Commission Report. To be sure, parts of the report could be criticized by non-professionals, and that contributed to the politics of the debate: One real estate agent was quoted in the press at the time to the effect that (put in economists' language, not the language the agent used) hedonic functions for housing did not go through the origin.[16]

It was inevitable. Congress was not so much interested in the CPI as such, but rather (as one Congressman candidly put it) in finding a way to get BLS to cut the deficit so they did not have to vote on it. I was reminded of an earlier (1981) controversy on CPI escalation of Social Security benefits: In the midst of that one, Senator Goldwater introduced a bill that would have prevented any change in CPI methods if the effect was to lower Social Security benefits. A mix of politics and statistics seldom produces an outcome that is favorable to economic statistics. I have seen many political debates about economic statistics: They are always debates about something, other than statistics, that is under the table.

We still need discussion of CPI methods. But we need even more a debate on principles for allocation of resources to the dependent population, not only Social Security payments but also Medicare. The two debates need to be conducted separately. The debate ten years ago on the CPI effectively served to thwart the debate on the more important issue.

References

Adelman, Irma (1958) "A New Approach to the Construction of Index Numbers," *Review of Economics and Statistics* 40, pp. 240-9.

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

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, pp. 25-33.

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.

Armknecht, Paul and Donald Weyback (1989) "Adjustments for Quality Change in the U.S. Consumer Price Index," *Journal of Official* Statistics 5:2, pp. 107-23.

Baye, Michael R. (1985) "Price Dispersion and Functional Price Indices," *Econometrica* 53:1, January, pp. 213-223.

Berndt, Ernst R., Zvi Griliches, and Joshua G. Rosett (1993) "Auditing the Producer Price Index: Micro Evidence from Prescription Pharmaceutical Preparations," *Journal of Business and Economic Statistics*, 11(3), July, pp. 251-64.

Blackorby, Charles and Robert R. Russell (1978) "Indices and Subindices of the Cost-of-Living and the Standard of Living," *International Economic Review* 19:1, February, pp. 229-240.

Blow, Laura and Ian Crawford (2001) "The Cost of Living with the RPI: Substitution Bias in the UK Retail Prices Index," *Economic Journal*, 111(472), June, pp. F357-82.

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 Senate

16 The agent was criticising the Commission's price per square foot measure in its discussion of housing. "Price per square foot goes down as square footage increases." Bette Gorman, as quoted by John M. Berry, Washington Post, December 19, 1996, page E10.

Finance Committee, Advisory Commission to Study the Consumer Price Index, December 4, http://www.ssa.gov/history/reports/boskinrpt.html.

Boskin, Michael (2005) "Causes and Consequences of Bias in the Consumer Price Index as a Measure of the Cost of Living," *Atlantic Economic Journal* 33, pp. 1-13.

Bradley, Ralph (2005) "Pitfalls of Using Unit Values as a Price Measure or a Price Index," *Journal of Economic and Social Measurement* 30 (1), pp. 39-61.

Braithwait, Steven D (1980) "The Substitution Bias of the Laspeyres Price Index: An Analysis Using Estimated Cost-of-Living Indexes," *American Economic Review* 70(1), March, pp. 64-77.

Camba-Mendez, Gonzalo, Vitor Gaspar, and Mark Wynne (2001) *Measurement Issues in European Consumer Price Indices and the Conceptual Framework of the HICP* (Frankfort and London: European Central Bank and Centre for Economic Policy Research).

Carruthers, A.G., D.J. Sellwood, and P.W. Ward (1980) "Recent Developments in the Retail Prices Index," *The Statistician* Vol 29, No. 1, pp. 1-32.

Christensen, Laurits R. and Marilyn E. Manser (1976) "Cost-of-Living Indexes and Price Indexes for U.S. Meat and Produce, 1947-1971," In Nestor E. Terleckyj, ed. *Household Production and Consumption*, Conference on Research in Income and Wealth, Studies in Income and Wealth 40: 399-446 (New York: Columbia University Press for the National Bureau of Economic Research).

Covas, Francisco, and Joao Santos Silva (1999) "Outlet Substitition Bias," *Bank of Portugal Economic Bulletin*, September, pp. 77-85.

Dalén, Jorgen (2002) "Personal Computers in Different HICPs," paper presented at the Brookings Workshop, "Hedonic Price Indexes: Too Fast? Too Slow? Or Just Right?" Washington, D.C., February 1, available on the Brookings Institution website.

Diewert, W. Erwin (1995) "Axiomatic and Economic Approaches to Elementary Price Indexes," NBER Working Paper No. 5104, May (Cambridge, MA: National Bureau of Economic Research).

Ducharme, Louis Marc, ed. (1997) *Bias in the CPI: Experiences from Five OECD Countries*. Statistics Canada, Analytical Series, Prices Division No. 10, Ottawa: Statistics Canada, Catalogue no. 62F0014MPB, No. 10.

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.

Gordon, Robert J. (2003) "A Century of Downward Bias in the Most Important Component of the CPI: The Case of Rental Shelter, 1914-2003," in Ernst R. Berndt and Charles R. Hulten (eds.), *Hard to Measure Goods and Services: Essays in Honor of Zvi Griliches*," forthcoming CRIW-NBER conference volume.

Gordon, Robert J. (2004) "Apparel Prices and the Hulten-Bruegel Paradox," in Carol Corrado, W. Erwin Diewert, and Charles R. Hulten, (eds.), *Price Index Concepts and Measurement*, forthcoming CRIW-NBER conference volume.

Hendel, Igal and Aviv Nevo (2002) "Sales and Consumer Inventory Behavior," National Bureau of Economic Research Working paper 9048.

Hill, Peter (1997) "The Measurement of Inflation and Changes in the Cost of Living," prepared for the Conference of European Statisticians, Joint ECE/ILO Meeting on Consumer Price Indices, Geneva, November 24-27.

Hoffmann, Johannes (1998) "Problems of Inflation Measurement in Germany," Discussion Paper 1/98, Economic Research Group of the Deutsche Bundesbank (Frankfurt, Germany: Deutsche Bundesbank).

Konüs, A.A. (1925) Translated into English and published as "The Problem of the True Index of the Cost of Living," *Econometrica* 7, January 1939, pp. 10-29.

Krone, Theodore M., Leonard I. Nakamura, and Richard Voith (2002) "The CPI for Rents: A History," Brookings Program on Economic Measurement, paper available on the Brookings Institution Website.

Lebow, D. E., Roberts, J. M., and Stockton, D. J. (1994) "Monetary Policy and 'The Price Level'," Washington D.C., Board of Governors of the Federal Reserve System.

Lebow, David E. and Jeremy B. Rudd (2003) "Measurement Error in the Consumer Price Index: Where Do We Stand?" *Journal of Economic Literature* (43), pp. 157-201.

Liegey, Paul R. Jr (1993) "Adjusting Apparel Indexes in the Consumer Price Index for Quality Differences," in Murray F. Foss, Marilyn E. Manser, and Allen Young (eds.), *Price Measurements and Their Uses*, conference on Research in Income and Wealth, Studies in Income and Wealth 57: 209-26 (Chicago: University of Chicago Press for the National Bureau of Economic Research).

Lequillier, Francois (1997) "Does the French Consumer Price Index Overstate Inflation?" Chapter 3 in Ducharme (ed.).

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. and Richard J. McDonald (1988) "An Analysis of Substitution Bias in Measuring Inflation, 1959-85," *Econometrica*, July, pp. 909-30.

McCarthy, Jonathan and Richard Peach (2002) "Housing Trends in the 1990s: Measurement in the CPI," Brookings Program on Economic Measurement, paper available on the Brookings Institution website.

McCarthy, Phillip J. (1961) "Sampling Considerations in the Construction of Price Indexes with Particular Reference to the United States Consumer Price Index," Staff Paper number 4 in Price Statistics Review Committee.

Moulton, Brent R. and Karin E. Moses (1997) "Addressing the Quality Change Issue in the Consumer Price Index," *Brookings Papers on Economic Activity* (1), pp. 305-49.

Organisation for Economic Co-operation and Development (1997) "Synthesis Paper on Shortcomings of the Consumer Price Index Measure of Inflation for Economic Policy Purposes," document number ECO/CPE/WP1(97)12, (Paris: OECD, Economics Department) September.

Oulton, Nicholas (1995) "Do UK Price Indexes Overstate Inflation?" *National Institute Economic Review* (152), pp. 60-75.

Pollak, Robert A. (1975) "Subindexes of the Cost of living," *International Economic Review*, 16(1) February, pp. 135-50. Reprinted in Robert A. Pollak, *The Theory of the Cost-of-Living Index*, Chapter 2, pp. 53-70, 1989 (Oxford, U.K.: Oxford University Press).

Pollak, Robert A (1998) "The Consumer Price Index: A Research Agenda and Three Proposals," *Journal of Economic Perspectives* 12(1), pp. 69-78, Winter.

Price Statistics Review Committee (1961) *The Price Statistics of the Federal Government*, U.S. Congress, Joint Economic Committee, Government Price Statistics, Hearings, Part 1. 87th Congress, 1st Session, 1961; also published as: National Bureau of Economic Research, General Series, Number 73.

Randolph, William C (1988) "Housing Depreciation and Aging Bias in the Consumer Price Index," *Journal of Business and Economic Statistics*, July, pp. 359-71.

Randolph, William C. (1988) "Housing Depreciation and Aging Bias in the U.S. Consumer Price Index," *Journal of Economic and Business Statistics*. 6-3, pp. 241-246, May.

Reinsdorf, Marshall (1993) "The Effect of Outlet Price Differentials on the U.S. Consumer Price Index," In Murray F. Foss, Marilyn E. Manser, and Allen Young (eds.), *Price Measurements and Their Uses.* Conference on Research in Income and Wealth, Studies in Income and Wealth 57: 227-60 (University of Chicago Press for the National Bureau of Economic Research).

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 and Brent R. Moulton (1997) "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) pp. 397-423.

Reinsdorf, Marshall and Jack E. Triplett (2004) "A Review of Reviews: Eighty Years of Professional Thinking About the CPI," presented at the NBER-CRIW conference on Price Measurement, Vancouver, June.

Ribe, Martin (2002) "Quality Adjustment (QA) for New Cars in Austria and Sweden," paper presented at the Brookings Workshop, "Hedonic Price Indexes: Too Fast? Too Slow? Or Just Right?" Washington, D.C., February 1, available on the Brookings Institution website.

Rossiter, James (2005) "Measurement Bias in the Canadian Consumer Price Index," Ottawa: Bank of Canada, Working Paper 2005-39, December.

Saglio, A. (1994) "Comparative Changes in Average Price and a Price Index: Two Case Studies," in Statistics Canada (1994).

Schultz, Bohdan J. (1994) "Choice of Price Index Formulae at the Micro-Aggregation Level: The Canadian Empirical Evidence," in Statistics Canada (1994).

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).

Shapiro, Matthew D. and David W. Wilcox (1996) "Mismeasurement in the Consumer Price Index: An Evaluation," in Ben S. Bernanke and Julio Rotemberg (eds.), *NBER Macroeconomics Annual* (Cambridge, MA: MIT Press) pp. 93-142.

Shiratsuka, Shigenori (1999) "Measurement Errors in Japanese Consumer Price Index," Federal Reserve Bank of Chicago, Working Paper Series (WP-99-2), February (this paper is a summary of a longer work published in Japanese).

Silver, Mick and Saeed Heravi (2001) "Scanner Data and the Measurement of Inflation," *Economic Journal*, 111(472), June, pp. 383-404.

Statistics Canada (1994) *International Conference on Price Indices: Papers and Final Report*, First Meeting of the International Working Group on Price Indices, Ottawa, Canada, November.

Statistics Canada (1995) *The Consumer Price Index Reference Paper*, Ottawa, Canada: Minister of Industry, catalog No. 62-553, July.

Timmer, Marcel, and Robert Inklaar (2005) "Productivity Differentials in the U.S. and EU Distributive Trade Sector: Statistical Myth or Reality?" Research Memorandum GD-76, Gronigen Growth and Development Centre, University of Gronigen, April.

Triplett, Jack E. (1971) "Quality Bias in Price Indexes and New Methods of Quality Measurement," in Zvi Griliches, ed., *Price Indexes and Quality Change: Studies in New Methods of Measurement* (Cambridge, MA: Harvard University Press) pp. 190-94.

Triplett, Jack E. (1990) "The Theory of Industrial and Occupational Classifications and Related Phenomena," 1990 *Annual Research Conference Proceedings*: 9-25. U.S. Bureau of the Census August.

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, pp. 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. (2004) *Handbook on Hedonic Indexes and Quality Adjustments in Price Indexes*, Directorate for Science, Technology and Industry, Paris: Organization for Economic Co-operation and Development, STI Working Paper 2004/9, Document DST/DOC2004)9.

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).

Wynne, Mark A. and Fiona Sigalla (1993) "A Survey of Measurement Biases in Price Indexes," *Federal Reserve Bank of Dallas* Research Paper 9340, October.

Wynne, Mark A. and Diego Rodrigiez-Palenzuela (2001) "Measurement Bias in the HICP: What Do We Know and What Do We Need to Know?" paper presented at the CEPR/ECB Workshop on Issues in the Measurement of Price Indices, Frankfort, November.