corrected: dec. 13, 1999 Remarks Presented at the Panel Discussion on Research in Price, Quality, and Quantity Measurement: What Agenda for the Next Twenty Years?

NBER Productivity Program Ernst Berndt & Zvi Griliches, Organizers July 21, 1999 Cambridge, Massachusetts

> Jack E. Triplett Brookings Institution

Members of the panel were invited to discuss improvements in the U.S. Consumer Price Index (CPI) and to suggest an agenda of research issues in price measurement for the future.

I. Improvements

A. COL index framework. Controversy has broken out anew on an old issue: Does the concept of the cost of living (COL) index provide the underlying conceptual framework for the CPI?

The United States is one of a small number of countries (which also include the Netherlands and Sweden) that accept the COL index framework. Most countries' statistical agencies do not. The Boskin Commission report stated unequivocally that the U.S. CPI should be interpreted as an approximation to a COL index. I agree. But a no doubt unintended side effect of the report was to heighten an international dispute on this matter, a dispute in which the U.S. position—shared by the BLS and the CPI's critics—is decidedly in the minority.

Incidentally, the Bureau of Labor Statistics (BLS) endorsement of the COL index is not a recent matter. It dates to the 1970's when Joel Popkin, Robert Gillingham, and I were all in the BLS Office of Prices and advocated the COL index framework, and to the 1978 CPI revision that was managed by John Layng, who was also an advocate of the COL framework for the CPI. In the intervening 20 years, I have never heard of anyone in a responsible position among the BLS *staff* who opposed the COL position, some rather sloppy and ambivalent writing for BLS publications notwithstanding. The Boskin Commission's endorsement of a COL concept was not something

that BLS *staff*, present or in the past 20 years, opposed. It is important to emphasize that point because there has been substantial confusion on the matter in the public discussion of the Boskin Commission report and its recommendations.

The international dispute about the COL framework has become important because there is a tremendous push for international comparability of economic statistics. Alan Greenspan, in a recent speech in Europe, called for more international comparability of CPIs, essentially in the pursuit of international harmonization of monetary policies. For very different reasons, the preparation of a new international manual for CPIs, which would be comparable to the huge international manual for construction of national accounts (*SNA*, 1993), is just getting underway.

Significantly, a set of internationally comparable CPIs exists now, in the European *Harmonized Indexes of Consumer Prices* (HICP), which are mandatory for all the countries in the European Union, and will probably in the future be followed as well by a set of associated countries and candidate countries, some 29 countries in all.

Additionally, Eurostat (the European statistical agency) has proselytized among countries of Central and South America, proposing that they adopt the Eurostat HICP framework for their indexes, in particular, for the harmonized CPIs that are now being constructed for the Mercosur countries (Argentina, Brazil, Paraguay, Uruguay), with Chile and Bolivia as associated participants. It does not take much foresight to see that the U.S. may soon be surrounded by world CPIs that take quite different approaches from the COL oriented approach of the BLS.

These HICP indexes are emphatically *not* COL index oriented. They follow the intellectual parentage of Hill (1997), who contends that a COL index is not appropriate as a price index for measuring inflation. The HICP indexes do not subscribe to a flow of services approach to measuring owner-occupied housing, and under a "no imputations" rule, they will not use the rental equivalence approach for measuring owner–occupied housing that was adopted in the U.S. CPI in 1983. Indeed, one possible HICP measure of owner-occupied housing will resemble the old BLS approach, abandoned (for good reason) in 1983.

The idea that the CPI should approximate a COL index is not without controversy in the United States. For example, Angus Deaton has written (in *The Journal of Economic Perspectives*, Winter 1998): "The Boskin Commission's . . . recommendation that the Bureau of Labor Statistics should establish a cost of living index as its objective in measuring consumer prices, taken by them as essentially obvious, is a contentious proposition that requires serious argument. In fact it is unclear that a quality-corrected cost of living index in a world with many heterogeneous agents is an operational concept."

If there is not agreement that the COL index provides the underlying conceptual framework for the CPI, then we have no way to determine whether changes to the CPI are improvements, and no way to determine what improvements to the CPI are needed. So my first point is that—although it is neither research nor CPI improvement, strictly speaking—we need more international consensus on what a COL index means in practical application, and more enlightenment about what alternative conceptual frameworks imply. I am at work on a contribution to that topic myself [subsequently circulated as "Should the Cost of Living Index Provide the Conceptual Framework for a Consumer Price Index?" presented at the Measurement of Inflation International Conference, Cardiff University, September 1999].

B. Escalation Uses. The issues surrounding the COL index have become more muddled by the fact that the Boskin report was commissioned, not out of a pure concern for inflation measurement, as was the NBER's Stigler Committee report of 1961, but out of concern over escalation of social security payments. There is a difference between inflation measurement, on the one hand, and on the other, the principles that determine how we want social security *incomes* to behave. In this matter, I heartily endorse Zvi Griliches' Congressional testimony and his remarks in other places, which I might paraphrase as follows.

Suppose granny now has access to a VCR with a remote that lets her channel surf without causing misery to her aching hip, or that she now has access to hip replacement surgery that (in this context) substitutes for the VCR. *Whether one wants to reduce granny s social security*

payments because of such improvements is a wholly different question from whether quality adjustments for them should be made in her COL index.

Quality adjustments for improved electronics and improved medical procedures are necessary because the COL index holds granny's standard of living constant. Escalation issues, on the other hand, concern partly whether we want granny to be able to improve her living standard, to consume at the higher level that improvements in technology permit.

Part of the issue here is the definition of income or wealth that is being deflated with the COL index, a point that I made many years ago (Triplett, 1983). But a more important part concerns equity in sharing the fruits of technological change that increases average living standards. It is not so surprising that noneconomists get escalation issues tangled up with price measurement issues, but it is a bit more distressing when economists do likewise.

Improving the CPI, therefore, requires more attention to the principles for equity in escalation policy and to income measurement issues for escalation. Otherwise, mistaken proposals for "improving" the CPI will undoubtedly emerge from concerns that are, at their root, questions about fairness in the distribution of incomes between different social groups. Both inflation measurement and escalation issues are important—it is just that we should avoid tangling one issue with the other.

C. Statistical and sampling techniques. Statistical issues in the CPI need far more attention. In contrast to the Stigler Committee report, the Boskin Commission report said nothing substantive about statistical matters. No other country does probability sampling of outlets and items or uses a household point of purchase survey in conjunction with a household consumption survey as the U.S. does. If you press them on it, most will say, first, that they cannot afford it and, second, that they do not think that probability sampling would make much difference anyway.

The BLS approach has been in place for twenty years. To my knowledge, it has never had a comprehensive statistical review, which is long overdue. I am not fully convinced that the BLS

statistical apparatus is necessarily the best way to go, or the only way to go, but I lack the statistical competence to say very much about it.

For example, the Australians, who do not use probability sampling, subdivide their CPI into some 1,500 different components, which they justify largely on the grounds that they want CPI components to be as homogeneous as possible with respect to price behavior. Sample sizes get very small under the Australian procedure, so one gets apprehensive about the representativeness and the statistical properties of the samples. The BLS probability sampling route requires, by its nature, much broader CPI aggregates. The BLS avoids the representativeness problems that may arise in the Australian index, but the BLS approach invites the kinds of statistical problems that arise out of the extreme heterogeneity in a large number of the CPI's 207 basic components.

Intuitively, I have reservations about both the Australian judgmental and the BLS probability approach, but I have never seen an analysis presenting the statistical tradeoffs between them, and reviewing whether some midway position exists that might be as good at lower cost or perhaps better at the same cost.

D. Consumer expenditure survey. It is time to improve the BLS consumer expenditure survey (CEX). The CEX provides the weights for the CPI. It should be obvious that one cannot have an accurate CPI without having accurate weights.

One cannot, as well, estimate the substitution bias in a fixed-weight index without having accurate weights. For example, most recent estimates of the bias in a Laspeyres index number do it by comparing the Laspeyres index to a superlative index number, such as the Fisher index (which is the geometric average of Laspeyres and Paasche indexes). How much of the difference between these two indexes is statistical noise arising out of inaccurately estimated weights?

The CEX is far too small. I applaud the BLS decision to increase it by 50 percent, but that still is only 7,500 households, and that is a small survey. Many countries have larger surveys, although they often support them by collecting less detailed information. We need a CEX that is substantially larger than the survey's present size.

Additionally, other CEX problems need attention. For example, the within-sample attrition rate is a cause for concern, and there is also a need to examine the obtrusiveness of the survey instrument to determine if it has something to do with the rapid falloff in responses as the survey's panel extends. This, again, is a set of issues that affects the accuracy of the CPI, and it is a set of issues on which CEX users *and* CEX staff have largely been in agreement over many years. Some coordinated, cross-government action needs to be initiated. The BLS is the natural place, in the absence of any centralized statistical decision-making in the U.S., to look for leadership on CEX issues.

II. The Research Agenda

A. Substitution bias. The Boskin Commission estimated the total bias in the CPI at 1.1 index points; only 0.1 of that was what it called "upper-level" substitution bias. Thus, bias within the components outweighs between-component substitution bias by a ratio of 10 to 1. Interestingly, substitution bias estimates for other countries that have a greater number of basic components than the 207 in the U.S. CPI show roughly the same amount of upper-level substitution bias.

I conclude from the 10 to 1 relative sizes of substitution bias and non-substitution bias estimates that substitution bias between CPI basic components—for the past 65 years the dominant issue in the price index number literature—is dead as a research issue. So long as the CPI weights are kept up to date and are changed frequently, substitution bias among the components *for which the weights are fixed* is small enough to be neglected. Although I raised some controversy a dozen years ago with the phrase "all the fruit has been picked from that tree," professional opinion now concurs.

Future price index research should therefore focus on measuring CPI components. It is at the component level, measuring the prices of cars and computers and—yes—bananas, where the problems lie. The Boskin Commission's focus on component price indexes, on new products and quality change, was right on the mark.

B. Lower level price aggregation. How one aggregates the CPI's 207 basic components into the overall CPI may be a dead issue, but the question of how one aggregates the individual price quotations into a price index for cars or for bananas is certainly not resolved. I want to raise some serious reservations about what the Boskin Commission called "lower-level" substitution bias within CPI components. Faced with evidence that geometric means of price relatives showed lower rates of inflation than arithmetic means of price relatives, the Boskin Commission decided that this looked like the difference between a Laspeyres index (arithmetic mean) and a Cobb-Douglas index (geometric mean), and that therefore the geometric mean-arithmetic mean issue was just another form of substitution bias (between red delicious and yellow delicious apples, as it put it).

This interpretation is questionable. Perhaps the consumer behavior that is relevant to the price index for apples is nothing more than the same long-term commodity substitution that has so long dominated the price index literature—substitution between, say, apples and bananas, as the price of the latter has fallen relative to the former. But I doubt it.

In the long run, I may substitute more bananas for apples, as the relative price of bananas falls. This is classic commodity substitution, which leads to substitution bias in the aggregate fixedweight Laspeyres CPI. But in the short run, I may exercise my taste for variety in diet by buying bananas when they are on sale, rather than when they are not, or I might (but don't) go across the street where they are on sale in preference to buying from the store where they are not on sale. This is consumer shopping and search behavior, not commodity substitution behavior. For food commodities that are more durable than bananas, I might choose to stock up and to rebuild my inventory of consumption goods when they are on sale, and buy at nonsale prices only when I have made an inventory error or confront unforeseen circumstances. Consumer inventory behavior probably does not resemble the behavior that is modeled as commodity substitution. The correct formula for lower-level indexes—the price index for bananas—cannot be determined solely by the principles of commodity substitution that guide the choice of upper-level CPI formulas.

Thus, as Pollak (1998) contends, consumer behavior that matters *within the price index for bananas* (bananas are, after all, in the United States, as close to a homogeneous commodity as one can find) is not consumer commodity-substitution behavior, but consumer search, shopping, inventory, and related behaviors. We need much more research into the nature of the index number problem that the Commission discussed under the "lower-level substitution bias" name, research that will explain how consumer search and shopping behaviors should be built into a COL index, and therefore how the BLS should compute the price index for "bananas." The Boskin Commission (and the BLS, as well) did us a disservice by making this lower-level aggregation problem appear to be a simple and conventional problem, when it is in fact a complex one on which little is known.

Having said this, I should make clear that I am not opposed to the BLS' recent move to a geometric mean for the CPI as an interim step, nor am I asserting that there is no commodity substitution in any of the 207 item indexes for the CPI (some of which are quite heterogeneous). It is, rather, that the problem to be solved is a more complex problem than the one that the term "lower level substitution" suggests.

C. Quality Change. Quality change was a major concern of the Stigler Committee and of the Boskin Commission 25 years later. It undoubtedly will still be on the agenda 20 years from now.

A recently-developed distinction (that is not yet very clear) is between quality change in the universe and quality change in the CPI sample. Almost the entire history of research on quality change has concerned the questions: What happens when there is quality change among the goods that are included in the CPI sample? And, do the methods used by the BLS create bias in the index?

Despite some recent assertions that this within-sample quality change issue is "irrelevant" (that seems to be the interpretation of Gordon, 1998), quality change within the sample remains an important question. Moreover, it still remains true that the implications of methods actually used to

control for quality change within the sample are widely misunderstood. When quality is improving, the bias from the most widely employed quality adjustment method (in Triplett, 1990, I called it "deletion," but this might not be good terminology) is downward when prices are rising and upward when prices are falling. It is not in general true that the direction of the bias in the index depends on the direction of quality change, it is more nearly the case that it depends on the direction of the true price change.

On useful research on quality change within the CPI sample, I would like to repeat a proposal that I made at the NBER meeting on the CPI that Zvi Griliches organized in April, 1995. I would like to see BLS conduct statistical "audits" on the treatment of quality changes in the CPI. These should be carried out on a sample of commodities and services, taking care to select a sample that contains not only commodities for which quality change is thought to be rapid, but also some others where it is not so evident. William Nordhaus (in BPEA, 1998) has now proposed the same thing, so I join forces with him.

Price index research outside government tends to be biased toward those goods and services where researchers think rapid quality change has occurred. A doctoral student who wanted to study price indexes for computers or for banking or for medical care will, understandably, get a far different reception from his advisor than the student who proposes research on price indexes for hair brushes or for hair cuts. Research ought to be directed, to be sure, where the payoff is likely. But an estimate of the bias in the aggregate CPI will itself be biased if based on existing research, because the research topics are not randomly selected with respect to the components of the CPI. Thus, the BLS should conduct a representative series of audits, taking care to cover components in which there has been rapid technological change and components in which there has been little.

The Boskin Commission stated that its extrapolation of research studies to areas of the index in which no research had been done was justifiable because "An estimate of zero bias is itself biased." What the commission apparently had in mind was a world in which the sign of the

quality change bias is known, and that it is always upward. That is a serious misreading of research. Studies that have shown downward bias in CPI components at various times include rent, automobiles, and clothing, to name some items with major weight. If the unknown bias also has an unknown sign, an estimate of zero for items that have not been studied is not such a bad one.

We need better estimates of the effect of CPI quality adjustment procedures. A CPI audit program would be very helpful. Only BLS can carry out such a research project. The project should of course bring in external data as well as carrying out a careful examination of the quality changes that are encountered within the CPI sample. Moulton and Moses (1998) is an excellent place to start on developing data on quality changes that are conducted within the CPI sample. My old paper with McDonald (Triplett and McDonald, 1977) was a kind of audit, on an actual BLS price index (it was from the PPI, not the CPI). Another example is Berndt, Griliches and Rosett (1993) on drug prices. This is a very small literature that needs to be expanded.

Recently, another problem has come to the fore. Rapid quality change may mean that price change is more likely to occur on newly introduced, improved products that might be outside the CPI sample. For example, a new computer enters the market at a lower quality adjusted price than the old ones. Even if the new computer is put into the CPI market basket immediately (which is in fact seldom the case because of sampling and other considerations), the price change that occurs when the product enters will be missed, unless the prices of the computers that are included in the CPI sample fall to meet the quality/price ratio of the newly introduced computer. This scenario might have been implicit in what was discussed in years past, but making it explicit clarifies some of the issues. This implies some research on the ways that quality change and new product varieties enter the market, their effects on the price behavior of old varieties, and the extent that new varieties push old ones out of the market, without necessarily producing price/quality equilibrium among new and old product prices.

D. New products. Another major research agenda item is the new product, consumer surplus problem, as discussed by Hausman (1996) and the Boskin Commission. This is clearly a long-run research agenda item. Although one might hope that hedonic methods offer hope for ameliorating the quality change problem in consumer price indexes, estimation of consumer surplus and the effects of new products lacks at this writing an effective, practical method for implementation in the CPI.

There is a useful parallel between the Stigler Committee's discussion of substitution bias and the Boskin Commission's discussion of consumer surplus. Looking back at the Stigler Committee's 1961 proposal to estimate substitution bias, it is remarkable how deficient was 1961 methodology for carrying out that suggestion. Only in the 1970's did computer capacity finally exist to permit estimation of substitution effects econometrically. Even then (as we soon found out when in the early 1970's we began the BLS project to estimate the substitution bias in the CPI), there were huge problems that were not tractable with the consumer demand systems, the econometric methodologies, and the computer capacities that existed at that time.

The solution to the commodity substitution problem really came with Erwin Diewert's (1976) remarkable demonstration that one could approximate the substitution bias very closely with an extremely simple and easy to carry out calculation of a superlative index number. As a result of both BLS econometric estimates (such as Braithwait, 1980, and Manser and McDonald, 1986) and the superlative index number approaches, we now have considerable confidence in that 0.1-0.2 substitution bias estimate cited by the Boskin Commission. In contrast, before any of this work was done, it was quite common for economists to "guesstimate" far higher substitution biases in the CPI (perhaps three or more percentage points annually).

The Boskin Commission "guesstimated" the effect of new products (and quality change), in the absence of firm empirical estimates. Although research of the type pioneered by Hausman (1997) will be valuable, it is hardly conceivable that we will accumulate very rapidly studies on a very high proportion of new products that are introduced. And the substantial difficulties exposed

by Bresnahan (1997) suggest that prospects are remote for an aggregate empirical estimate along the lines that Hausman pioneered.

Thus, I suspect that we will not make that much headway on the new product issue until someone produces a practical innovation for estimating reservation prices and consumer surplus comparable to the one that Diewert introduced for estimating the substitution bias. That is another way of saying that measurement improvements in the CPI may require the search for simpler procedures rather than more complex ones.

E. Services. Another topic is measuring the prices of services. Surely, this is a question that has waited for a long time. The book by Griliches (1992) reminds us not only of the long period that passed since the earlier flurry of interest provoked by Fuchs (1969), but also the lack of progress over that interval. Although Barry Bosworth and I have a substantial project at the Brookings Institution on improving the measurement of output and productivity in the service sector, as well as money to fund projects, I must report that it has been extremely difficult to find researchers who want to commit to the task. Throwing money at the problem doesn't help if the problem does not attract good minds. I hope that twenty years from today we are not still saying that meaningful research needs to be done on the concepts of prices and outputs for services, but there are grounds for pessimism on real progress.

F. Consumer heterogeneity. Existing research that computes price indexes for individual households shows clearly that individuals have quite different preference functions. The variation in individual price indexes among households within any group that has been examined (rich versus poor, elderly versus young, single versus married, and so forth) completely swamps the between group effects. Despite this, we ignore the long-standing distinction in the existing price index literature between "democratic" (households are weighted equally) and "plutocratic" (households are weighted according to their expenditure shares) aggregate price indexes, on the presumption that the alternative weighting patterns do not matter very much.

However, even if the aggregate weighting effects do not matter much, heterogeneity matters a great deal for quality change, new products, and even for selecting the items that should be priced within any CPI basic component. Existing work on quality change, such as hedonic price indexes, as well as existing estimates of consumer surplus for new products, explicitly rely on the presence of heterogeneity among consumers. One cannot work on these topics using the model of the "representative consumer" that underlies much of the current CPI. The representative consumer buys a car with 0.8 of an air conditioner and equipped with 0.1 manual and 0.9 automatic transmission. That representative consumer fiction simplifies the marginal analysis necessary to handle quality change in the CPI, but does such great damage to reality that the fiction is unappealing. Alternative approaches need to be explored.

III Conclusions

The future research agenda is rich. No doubt other topics will emerge as research proceeds, so this forecast is likely to be about as valid as other economic forecasts—not without some usefulness, I would hope, but not all that prescient, either.

###

References

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): 251-264.

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): 64-77, March.

Bresnahan, Timothy F. 1997. "Comment." In Timothy F. Bresnahan and Robert J. Gordon, eds. 1997. The Economics of New Goods. National Bureau of Economic Research (NBER) Studies in Income and Wealth, vol.58. Chicago and London: University of Chicago Press.

Commission of the European Communities, International Monetary Fund, Organisation for Economic Co-operation and Development, United Nations, and World Bank. 1993. *System of National Accounts 1993.* Office for Official Publications of the European Communities Catalogue number CA-81-93-002-EN-C, International Monetary Fund Publication Stock No. SNA-EA, Organisation for Economic Co-operation and Development OECD Code 30 94 01 1, United Nations publication Sales No. E.94.XVII.4, World Bank Stock Number 31512.

Deaton, Angus. 1998. "Getting Prices Right: What Should Be Done?" *Journal of Economic Perspectives* 12(1): 37-46.

Diewert, W. Erwin. 1976. "Exact and Superlative Index Numbers." *Journal of Econometrics* 4(2): 115-45, May.

Fuchs, Victor R. 1969. Production and Productivity in the Service Industries. NBER Studies in Income and Wealth, vol. 34. New York: Columbia University Press.

Gordon, Robert J. 1997. "Addressing the Quality Change Issue in the Consumer Price Index: Comment." *Brookings Papers on Economic Activity* 0(1): 350-57.

Griliches, Zvi, ed. 1992. Output Measurement in the Service Sectors. NBER Studies in Income and Wealth, vol. 56. Chicago, IL: University of Chicago Press.

Hausman, Jerry A. 1997. "Valuation of New Goods Under Perfect and Imperfect Competition." In Timothy Bresnahan and Robert Gordon, eds. <u>The Economics of New</u> <u>Goods</u>. Conference on Research in Income and Wealth, Studies in Income and Wealth 58: 209-237. University of Chicago Press for the National Bureau of Economic Research.

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.

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

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

Nordhaus, William D. 1997. "General Discussion of Addressing the Quality Change Issues in the Consumer Price Index." *Brookings Papers on Economic Activity* 0(1): 361-63.

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

Triplett, Jack E. and Richard J. McDonald. 1977. "Assessing the Quality Error in Output Measures: The Case of Refrigerators." *Review of Income and Wealth* 23(2): 137-56.

Triplett, Jack E. 1983. "Escalation Measures: What is the Answer? What is the Question?" In <u>Price Level Measurement</u>, edited by W. Erwin Diewert and C. Monmarquette. Ottawa, Canada: Minister of Supply and Services Canada, 457-82.

Triplett, Jack E. 1990. "Hedonic Methods in Statistical Agency Environments: An Intellectual Biopsy." In *Fifty Years of Economic Measurement: The Jubilee of the Conference on Research in Income and Wealth,* Ernst R. Berndt and Jack E. Triplett, eds. Chicago: University of Chicago Press for the National Bureau of Economic Research.

Triplett, Jack E. Forthcoming. "Should the Cost-of-Living Index Provide the Conceptual Framework for a Consumer Price Index?" In Measurement of Inflation International Conference Volume. Wales: Cardiff University.