

Editor's Introduction

Measurement is an essential ingredient for the design and implementation of good economic policies. Recent controversy surrounding the Consumer Price Index (CPI) shows the importance of accurate measurement for both the making of monetary policy and the indexation of government programs. This controversy, although seemingly about measurement, has also been heated by ideological disagreements about government programs and the public's desire to discipline the budget process. It is important, however, to distinguish political decisions involving redistribution from economic decisions involving accurate measurement of theoretical concepts.

The twenty-first annual policy conference at the Federal Reserve Bank of St. Louis examined research aimed at improving measures of inflation and output growth. The conference generated both theoretical and practical insights for improving measurement. While it is important to improve measurement, it is also important to recognize that there is, and always will be, a large and uncertain gap between what statistical agencies measure and what actually goes on in the U.S. economy. Much of what is important for social welfare is inherently unmeasurable. Whenever possible, economists and policymakers have tried to finesse the information problems by incorporating market mechanisms into the design and implementation of policies. Yet it remains imperative that economists continue to develop better measures of inflation and output. High-quality data on prices and quantities are needed both to test economic theories and to evaluate the effects of policies.

MEASURING OUTPUT

The first session of the conference dealt with the measurement of the economy's real output of goods and services. Because most economic activity is

measured in current dollar units, improving our measures of real output usually means improving our measures of prices. In the first conference paper, "Measuring Consumption: The Post-1973 Slowdown and the Research Issues," Jack E. Triplett examines whether the consumption slowdown that began around 1973 can be explained by poor measurement. He begins by documenting the consumption slowdown, noting that per capita real consumption growth averaged only 1.7 percent per year from 1973 to 1995, after having grown at a 3 percent annual rate from 1959 to 1973, and that the slowdown was evident in all the sub-components: durables, non-durables, and services.

Triplett asks if measurement error could account for some of the observed slowdown. He notes that the Consumer Expenditure (CE) survey data used by the Bureau of Labor Statistics (BLS) to weight component prices in the CPI provides an independent source of consumption data. Comparing measures of consumption for various items from that survey with the Personal Consumption Expenditure (PCE) data used in the National Income and Product Accounts (NIPA), he finds no evidence in the CE data to support the notion that PCE data understate consumption growth. Unfortunately, the two series differ widely in coverage; each is better in some respects, and in many cases there is not enough information to make an informed judgment about which series is more accurate. Triplett recommends reconciling the NIPA and CPI data collection methods to save resources and to ensure that the common price and expenditure components are used in the construction of the relevant indexes.

Triplett then turns to a discussion of whether the deflators used to derive real consumption data from nominal expenditures might be biased. If the inflation rate is overstated, real growth might be understated. He examines four sources of bias:

upper-level substitution, lower-level substitution, new goods, and unmeasured quality improvements. He notes that although many components of the PCE deflator come from the CPI, the biases in the CPI recently emphasized by the Advisory Commission to Study the Consumer Price Index (Boskin Commission) may not be significant in the PCE deflator. Substitution bias in the PCE deflator will be smaller than in the CPI because the PCE deflator is a chain-weighted Fisher Ideal index, rather than a fixed-weight Laspeyres index. After discussing BLS adjustments for quality change, Triplett concludes that none of the Boskin Commission's biases support a plausible explanation for the 1973 consumption growth slowdown.

In his comments, Peter J. Klenow points out that services have been a growing share of the economy, and the areas growing fastest seem to be the ones most badly measured. He develops a simple method for estimating a "quality residual" that shows quality in services improving 1.6 percent faster after 1975 than before, implying that mismeasurement of services alone may be large enough to explain a large part of the consumption slowdown.

In a short article, "On Defining Real Consumption," Edward C. Prescott argues "that theory can and should be used to better define real consumption." He implies that the consumption slowdown is a result of poor measurement, resulting from an inadequate use of theory in defining goods and services. He shows that the price deflators for the "badly-defined" sectors—including owner-occupied housing, personal business services, medical care, and private education and research—grew more quickly than the price deflators in the "well-defined" categories.

In the final paper of the first session, "Measuring and Analyzing Aggregate Fluctuations: The Importance of Building from Microeconomic Evidence," John C. Haltiwanger argues that understanding establishment-level data is necessary to

understanding aggregate fluctuations of investment, employment, and productivity growth. He also suggests ways to improve the data collection process.

By looking at disaggregated data, Haltiwanger shows that gross job changes dwarf the net changes. At cyclical frequencies, gross job creation is more stable over the business cycle than gross job destruction. Haltiwanger argues that theories based on the representative-firm model cannot explain the aggregate movements in investment, employment, and productivity growth because the response of an industry aggregate to a common shock varies over time with the distribution of labor and capital needs among firms in a heterogeneous sample. He emphasizes that productivity gains are not distributed evenly across firms within an industry. Aggregate industry productivity generally grows because new establishments, with productivity levels matching existing plants, enter the industry while the least efficient establishments exit.

Haltiwanger advocates using this microeconomic detail to guide the search for useful models. He argues that knowledge gleaned from disaggregate data suggests fruitful ways to improve the data-collection process. He urges federal statistical agencies to develop a common master business establishment list as the sample frame for all surveys, to include identifiers that permit aggregation of establishment data to the firm and industry level, and to introduce sample rotation procedures that permit the construction of longitudinal statistics. He supports proposed legislation (H.R. 3924) that would standardize confidentiality restrictions across statistical agencies so that researchers and statisticians could share the establishment-level data.

Jeffrey R. Campbell focuses his comments on the importance of using modern macroeconomic theory to organize data collection. He supports Haltiwanger's call for more longitudinal data so that researchers can study the behavior of individual establishments, and he notes that it is important to identify which establish-

ments belong to which firms if we hope to match longitudinal data on financial variables with longitudinal data on prices and output.

Alan Heston agrees with Haltiwanger that looking at the disaggregated data can lead to a very different view of the aggregate facts. He urges caution when interpreting the micro evidence, however, noting that in a regional dissection of the data one finds more variation in job creation than in job destruction. If there are regional concentrations of creation and destruction, or if there is a large variance in a regional cross-section of job creations, then the theory required to explain the aggregate fluctuations becomes even more complex. Heston notes that evidence from European countries (where gross job allocations are even more volatile than in the United States) seems to belie the simple notion that high unemployment results from stagnation of European economies. He applauds Haltiwanger's suggestions for improved data collection but notes that, because of budget realities, many of these proposals may never be implemented. He suggests that special-purpose surveys may provide an inexpensive alternative to a more elaborate data-collection process. Such surveys would be especially useful if they were designed to help understand the nature of the economy as it goes through periods of rapid change.

QUALITY AND PRICES

The second session of the conference focused on problems in using changes in the CPI as a measure of changes in the cost of living. In the first paper of this session, Charles R. Hulten describes errors that arise in measuring quality change. He discusses the methods that the BLS uses to price new items that are brought into the CPI to replace items that are no longer available. He develops an economic/accounting framework to illustrate the potential size of the bias when a higher-quality product is introduced but the market price does not fully reflect the quality improvement. If this mispricing is serious, Hulten argues that the CPI

inflation may overstate the cost of living by even more than the Boskin Commission has estimated. By his calculations, this mispricing of improved quality would double the upper limit of the Boskin Commission's estimated range of bias.

On the other hand, Hulten also notes that there are reasons to think that the BLS overstates quality change in some cases. He argues that one method in particular, the "link" method, may introduce a negative bias that would lower the bottom of the Boskin Commission's estimated range of bias to zero. Hulten emphasizes the enormous, and perhaps unresolvable, uncertainty surrounding estimates of quality bias.

In his comments, Robert J. Gordon disputes both major premises in the Hulten paper. He challenges the economic framework used by Hulten, arguing that it ignores the distinction between movements along a supply curve for a product's characteristics (a measure of quality) and shifts in the supply curve brought about by technological change. When this distinction is made, Gordon argues that the analysis supports the hedonic methods used by the BLS. Gordon also argues that Hulten misinterprets the "quality" adjustments incorporated in the "link" method.

Per Krusell agrees with Hulten that market prices may not reflect quality improvements adequately; however, he argues, there is little if anything to be done about it. There is no theory to determine when and by how much prices may fail to capture quality differentials. Therefore, any adjustment is arbitrary. Krusell notes the correspondence between quality change and technological growth and discusses the importance of measuring technological change to understand the source of ongoing changes in worker productivity and the observed increases in wage inequality. He argues that accurate measurement of quality adjustments is crucial for measuring technological change and therefore crucial for understanding the determinants of real growth and the distribution of income among different types of labor.

In the second paper of this session, “Alternative Strategies for Aggregating Prices in the CPI,” Matthew D. Shapiro and David W. Wilcox use a BLS data set to calculate various price indexes, allowing a detailed analysis of the substitution bias in the “upper-level” stage of aggregation. The BLS aggregates prices in two separate stages. In the lower-level stage, a large number of individual prices are aggregated into strata (item-area) price indexes—207 item categories for more than 40 different geographic areas. In the upper-level stage, the strata indexes are combined into the all-item CPI by means of a Laspeyres index number formula. At this stage, substitution bias results from using the fixed-weight Laspeyres formula instead of a superlative index number. Shapiro and Wilcox find that, on average, superlative indexes such as the Fisher Ideal and Törnqvist-Theil grow 0.3 percentage points slower per year than the CPI with 1982-84 weights.

The main reason the BLS does not use a superlative index number is that superlative indexes require more frequent data on consumer expenditures to match the monthly data on prices. Consequently, Shapiro and Wilcox ask whether the BLS could use currently available data on household spending to approximate a superlative price index on a timely basis. They propose a method that could be used to implement the Boskin Commission’s recommendation that the BLS report two price indexes: one based on the current Laspeyres formula—and not subject to revision—and another that would use more frequent data on expenditures and would require revisions of the sort found in the Bureau of Economic Analysis measure of GDP.

In his comments, W. Erwin Diewert praises Shapiro and Wilcox for their useful research but suggests that there are many more measurement problems related to the CPI. He reviews arguments surrounding the choice of a superlative index-number formula, citing his own research showing that the Fisher Ideal index has many desirable properties.

Diewert also discusses problems that arise when statistical agencies are not able to make price comparisons for products that are seasonal or new. He suggests excluding such goods from the domain of goods used to measure short-run inflation. He also suggests that one might want to exclude durable goods where the price refers to a stream of service flows rather than the price for consumption in the reference period. In general, he encourages a more systematic approach to the problem of deciding on a specific domain for the measurement of a price index.

Peter Howitt takes a macroeconomic perspective, arguing that the size of the bias estimated by Shapiro and Wilcox is small compared to the uncertainty about the optimal inflation rate. He argues that it is important to understand the sources of costs of inflation before choosing an index. He notes that the CPI has been chosen as a policy target by some central banks, not because these central banks have worked out theories of the cost of inflation, but because the CPI is familiar, published with a short lag and never revised, contributing to the transparency and accountability of monetary policy.

POLICY IMPLICATIONS

The final session of the conference focused on the lessons of recent measurement research for policymakers. In “Measuring Short-Run Inflation for Central Bankers” Stephen G. Cecchetti examines various measures of inflation for their information about the underlying inflation trend. He notes two problems with inflation data: noise and bias. Bias is a mismeasurement that persists, while noise is temporary. Policymakers need to reduce the noise to get timely information about changes in trends. Noise can be eliminated by averaging over longer time intervals, but doing so reduces the timeliness of the information. To avoid this problem, Cecchetti recommends limited information estimators. Examples of such measures include the CPI less food and energy, the weighted median, and a trimmed mean—one that trims a certain

percentage off the tails of the monthly distribution of price changes. He examines means trimmed by 10 percent and by 25 percent.

Cecchetti also examines the statistical efficiency of inflation estimators, the distribution of deviations of the means from the trend, and the ability of seasonal adjustment to reduce noise in the estimator. Using a bootstrap method in Monte Carlo experiments, he finds that the mean change in the CPI less food and energy is actually a less efficient estimator of the trend inflation rate than is the mean change in the all-items CPI. Using CPI component data, Cecchetti finds that the 10 percent trimmed mean is the most efficient estimator among the candidates examined. He then examines the distribution of the deviation of the alternative measures from a 36-month centered moving average. Again, he finds that the 10 percent trimmed mean tends to provide the best estimate of the inflation trend.

In his comments, Alan S. Blinder argues that Cecchetti's research, although not glamorous, is a good example of the type of research that economists should be doing for policymakers. Blinder agrees with Cecchetti that the bias seems to be time-varying and asserts that a large bias in the measure of inflation might lead the Fed to make a costly mistake of pushing inflation down too much. He suggests that the policymakers should fix the bias problem before applying Cecchetti's noise-reduction methods.

Blinder suggests that the ability to forecast the future, not predict a centered moving average, would be a more useful criterion for evaluating alternative indicators of the inflation trend. He recommends replacing expenditure-share weights with stochastic weights based on a component's ability to predict the future CPI. In conclusion, he reasons that excluding food and energy from the core inflation rate is a good idea, not because these components are more volatile, but because, unlike "the rest of the price index—the part that comes out of the

industrial core of the economy, so to speak"—the Fed cannot control them.

In his comments, Mark A. Wynne notes that Cecchetti's research differs from much research on price index theory because it is aimed at understanding the dynamic macroeconomic implications of alternative price indexes rather than static measures of a representative agent's cost of living. The fundamental goal of the program is to identify the common element, due to excessive monetary growth, in all individual price changes. The research assumes that the price index the central bank should care about may not be the cost-of-living index sought by the BLS. The research program also differs from price index research because it includes a stochastic approach to finding weights as well as the theoretical approach, which uses expenditure weights. Wynne notes that historical studies of the stochastic approach go back over 100 years to Jevons and Edgeworth.

The conference's final session included a panel discussion of the policy issues related to the debate about measurement. Zvi Griliches notes that the reasons for the appointment of the Boskin Commission and the current public debate over bias in the CPI are the same: Adjusting the CPI downward would raise taxes and lower spending, thus making it easier for Congress to balance the federal budget. He reminds us that different price indexes are constructed for different reasons, and that one should be clear about the questions that might be answered by an index before deciding how to construct it. He notes that many measurement problems have been with us for a long time, and there are not good reasons to think that the problems have become substantially worse.

Griliches also discusses the problems inherent in the CPI that arise from its conceptual basis—the representative consumer paradigm. In particular, the CPI may be a good measure of the average consumer preferences, but it may not be a good measure for any particular person. Griliches encourages policymakers to make changes in real transfers and taxes

openly, rather than disguising them as concern about measurement problems in the CPI.

John S. Greenlees presents the BLS perspective on the controversy surrounding bias in the CPI. He begins by summarizing some of the actions that the BLS has taken to address measurement problems and notes that once the BLS identified the “formula” bias in the lower stage of the CPI process, it changed its aggregation methods. As a result, he estimates that the CPI growth rate has been reduced by about 0.24 percent per year. Greenlees notes that the substitution bias is relatively easy for the BLS to measure and correct. BLS research suggests that approximately one-half of the upper-level substitution bias in the CPI could be eliminated by updating the expenditure shares annually—calculating a Laspeyres index with recent expenditure weights rather than the 1982-84 weights currently used. The other half might be eliminated by switching to a superlative index.

The bias induced by quality changes and new goods is more problematic. Greenlees argues that the BLS makes adjustments for quality change whenever and wherever the changes can be measured, and he welcomes research from outside the BLS.

William A. Allen concluded the conference with a discussion of the UK experience with the measurement and targeting of inflation. In his article, he describes the indexes used at the Bank of England, noting adjustments made to make the index more appropriate for monetary policy. For example, alternative computations of housing expenditure can make a big difference—up to one-half of a percentage point per year—for the reported inflation rates. Allen also questions whether asset prices contain special information about the stance of policy that might be missing from the consumer price indexes.

SUMMARY

The conference brought together researchers interested in developing better measurement and those who are developing new theories that challenge the methods and output of the government’s statistical

agencies. It is clear that a consensus is forming on several fronts: (1) The economy’s output of services is poorly measured. (2) Longitudinal data on individual firms and households is needed to explain aggregate dynamics. (3) It is very difficult to measure quality change accurately. (4) The adoption of a superlative index number formula could reduce the upward bias in the CPI by at least one quarter of a percentage point per year. (5) The choice of the appropriate price index depends on its use. (6) And, finally, more research is needed to develop better measures of inflation and output growth. We thank the participants in the conference for their contributions to this effort. We also thank the analysts in the Research Department at the Federal Reserve Bank of St. Louis who helped to review the text and data for each of the articles: Heidi L. Beyer, Cindy A. Gleit, Joshua D. Feldman, Eran Segev, Daniel R. Steiner, and Robert Webb.

William T. Gavin
St. Louis, Missouri
June 4, 1997