MEASURING THE AGGREGATE PRICE LEVEL: IMPLICATIONS
FOR ECONOMIC PERFORMANCE AND POLICY

Robert J. Gordon

MEASURING THE AGGREGATE PRICE LEVEL:
IMPLICATIONS FOR ECONOMIC PERFORMANCE AND POLICY

ABSTRACT

Inaccurate measures of the aggregate price level may distort short-run policy decisions and may produce misleading comparisons of productivity growth across decades and among nations. This paper serves a dual purpose of reviewing compactly the vast American literature on price and output measurement, and of identifying special aspects of American methods which affect international comparisons of inflation and output growth.

The traditional problem of substitution bias in the Consumer Price Index (CPI) is of minor importance compared with the bias introduced by new products, changes in the quality of existing products, and outlet substitution bias. The quality bias for U.S. consumer durables has recently been estimated to be roughly 1.5 percent per year for the postwar period, and roughly 3 percent per year for consumer durables. The only available study of outlet substitution bias estimates a 2 percent annual rate for food in the 1980s.

Cross-country differences in measurement methods tend to overstate the recent productivity performance of U.S. relative to European manufacturing, with an understatement for U.S. nonmanufacturing. Both European and U.S. manufacturing performance are probably understated relative to Japan, which seems to do the best job of incorporating new products and correcting for quality change of high tech goods.

Robert J. Gordon
Department of Economics
Northwestern University
Evanston, IL 60208-2600
and NBER

I. INTRODUCTION

A comparison of the economic performance of nations usually involves a minimum of three measures, the unemployment rate, inflation rate, and growth rate of real output per person (or per employee-hour).[1] Of these three, the measurement of two (inflation and real output per person) requires an accurate estimate of the aggregate price level. The inflation rate, of course, is simply the rate of change of the aggregate price level, while real output is equal to nominal expenditure or income divided by the aggregate price level. In this sense accurate measurement of the aggregate price level is one of the most important tasks of national accounting.

Implications of Measurement Error

Inaccurate measures of the price level can lead policymakers astray. In the short run, for instance, an overstatement of the true rate of inflation can lead policymakers to adopt restrictive policies when none may be required. In the long run, errors in measuring the change in the price level lead to incorrect measures of the growth of output and productivity from one decade to the next. Measurement errors can also distort comparisons of the economic performance of different countries.

This paper is an introduction to the main principles and pitfalls in the measurement of the aggregate price level. While most of the examples are drawn from the United States, the principles apply universally and include the distinction between cost-of-living and fixed market-basket indexes for consumer prices, the difference between consumption and output

---

1. The best graphical measure of economic performance is the "diamond" invented by OECD economists in the early 1980s; this adds the current-account balance as a percent of GDP to the three measures listed in the text. See "An Economy's Best Friend," *The Economist*, November 16, 1991, p. 87.

price indexes, and the contrast between welfare-oriented and production-oriented criteria, e.g., in the treatment of environmental and anti-pollution expenditures. Pitfalls can plague measurement in every country and include bias introduced by quality change, substitution among products, and substitution among sales outlets, as well as spurious short-run volatility of inflation measures that can be caused by misleading measures of housing costs.

For non-U. S. readers, the paper serves a dual purpose of reviewing compactly the vast American literature on price and output measurement, and of identifying special aspects of American methods (both admirable and dubious) which affect international comparisons of inflation and output growth. A central focus of my present research is to uncover differences in measurement methods between the U. S. and the other major industrial countries, particularly Japan, Canada, and the large European nations, and I hope that wide distribution of this paper will stimulate an interchange of both methodological descriptions and suggestions for improvement that allow the statistical experts of each country to learn from each others' successes and failures.

A universal feature of price measurement is that errors become more important the shorter the time period examined and the narrower the industry definition that is of concern. For instance, alternative weighting schemes can give quite different signals to policymakers about the extent of an acceleration of inflation over a month or a quarter, yet the ranking of inflation over successive decades in a single country is unlikely to be affected much by measurement error. Nor are measurement errors likely to overturn the widespread impression that Argentina has had much more inflation over the past decade than any

member country of the OECD. Similarly, the treatment of computer prices makes a great difference in the output and productivity performance of the part of manufacturing that makes computers, less for manufacturing as a whole, and even less for the economy as a whole. Nevertheless, errors in measuring the price level for the aggregate economy can be significant enough to distort comparisons of productivity growth over time and across nations.

Plan of the Paper

The paper begins in Part II with general principles that apply to the computation of consumption price indexes, treating such topics as cost-of-living vs. market basket, substitution bias, weighting, and other general issues. Part III summarizes the general principles involved in developing output price indexes and places special emphasis on the relatively recent literature on "nonproportional" quality change. Then Part IV assesses the advantages and disadvantages of the major methods of measuring individual prices, including the specification and hedonic regression methods, and covers the use of nonstandard data sources. Part V focusses on measurement issues that are of particular interest in evaluating differences among countries in the measured growth rate of output and productivity. Part VI examines special issues that have been of major importance in the interaction between price indexes and monetary policy, particularly the measurement of homeownership costs. Finally, Part VII concludes with an agenda for those improvements in methods that deserve the highest priority for additional resources.

The Decentralized U. S. Statistical System

Nations differ in the degree of centralization of their national economic statistical agencies, and the U. S. is on the decentralized end of this spectrum.[2] Price data collected from original sources are the responsibility of the Bureau of Labor Statistics (BLS), which compiles and publishes the Consumer Price Index (CPI, with price data based on thousands of reports from field agents) and the Producer Price Index (PPI, based mainly on mailed-in questionnaires from manufacturers). Data on nominal and real output and the associated price deflators are the responsibility of the Bureau of Economic Analysis (BEA) in the Department of Commerce, and this agency traditionally has collected no primary price data of its own but rather has taken indexes at the level of individual commodities and services from the two BLS indices and has then combined them with a different set of weights. The deflators for components of consumption expenditures in the BEA's National Income and Product Accounts (NIPA) are based entirely on the CPI, and for equipment investment on parts of the PPI. For investment in structures the BEA combines a variety of indexes from the private sector and from still other government agences, including a hedonic regression index for single-family housing developed by the Census Bureau and an index for public roads developed by the Federal Highway Administration. Most of the deflators for exports and imports are based on the BLS international price program, but the BEA in recent years has been careful to integrate carefully its own hedonic price indices for computers and

---

2. Canada, with its centralized system, was recently awarded top international rank for the quality of its national statistics; see "Official Numbers: the Good Statistics Guide," *Economist*, September 7, 1991, p. 88. The U. S. was ranked seventh among ten countries listed.

peripherals into the appropriate "end-use" categories of exports and imports. In addition to developing the computer price index, the BEA has been active in creating primary price indexes for government purchases of military equipment and in developing methods of adjusting for the pervasive quality changes that occur in such equipment.

One side-effect of this decentralization is that price research within the government has been rather unbalanced across components of expenditure. The BLS has devoted substantial research to the CPI, and the Census Bureau has studied carefully the properties of its housing price index. The BEA's computer price index was the subject of extensive research and discussion before its introduction into the NIPA. But there has been much less research on price indexes for nonresidential construction, government purchases, and the foreign sector. One may question, for instance, whether the intensive effort by the BEA to develop price indexes for computers (used to deflate the relevant components of producers' durable equipment, exports, and imports) is consistent with the BEA's reliance on BLS producer price indexes for numerous other types of high-tech equipment that appear to incorporate only a fraction of the needed corrections for quality change.[3] Another side-effect of decentralization is that the ultimate decision power on the investment of resources in price research and primary price collection lies with yet another agency, the Office of Management and Budget (OMB).[4] Since improvement of the price deflators in the NIPA

---

3. For instance, in my book (1990, pp. 395-408) on durable goods prices, I developed evidence, based largely on the work of Kenneth Flamm, suggesting that the BLS price indexes used by the BEA to deflate telephone switching equipment are subject to a bias of at least five percent per year, and probably more.

4. For instance, Robert Parker states that the main obstacle to collection of improved price statistics for construction is the low priority that OMB places on this activity. See p. 270 of his comment on Pieper (1990).

inevitably requires inter-agency cooperation between BEA and the primary price collecting

agencies, in the end the OMB determines the degree of progress that can be made in

making improvements that everyone now agrees need to be introduced.[5]

II. CONSUMER PRICES: BASIC ISSUES

In Part II we review those issues that are of primary importance to the development

of a consumer price index, and Part III those of more importance to an output price index.

We defer to Part IV issues that are common to all price indexes, particularly specific

methods of adjusting for changes in quality.

The Cost of Living vs. the Consumer Price Index

The concept of the "cost of living" measures the cost of maintaining a constant level

of utility, or in two dimensions of remaining on the same "indifference curve." Technically,

consider two hypothetical situations, A and B, in which prices differ. The criterion based

on the indifference surface is to ask, given some stated income and set of prices in situation

A, how much income the consumer in situation B would require to make him or her just

indifferent between situations A and B. The resulting income level for B, divided by the

stated income level in A, is a "true cost-of-living index" which is the appropriate deflator of

situation B income (Fisher-Shell, 1972, p. ix). In most countries, including the U. S., the CPI

measures something quite different, the price of purchasing a fixed market basket of goods

---

5. The word "now" refers explicitly to the longtime opposition of George Jaszi, until the early 1980s the head of the BEA, to hedonic price indexes for computers and other products. Triplett (1990) presents a deep and probing discussion of the reluctance of statistical agencies to embrace hedonic price techniques.

and services. The CPI would be equivalent to a true cost-of-living index only under the extreme and unrealistic assumptions that there are no new products, no changes in the quality of existing products, no changes in consumer tastes, no changes in the relative prices of goods and services, and no changes in the distribution of retail outlets across conventional and discount stores.

However once these assumptions are dropped, we recognize that the real-world consumer can make choices which cause the CPI to diverge from a cost-of-living index. Starting with a market basket that has fixed shares of spending on chicken and beef, a decline in the relative price of chicken will lead consumers to increase the quantity of chicken consumed relative to beef; any Laspeyres index like the CPI will fail to reflect these shifts in shares and will overstate the overall increase in prices. In the case of new products that offer a new way of achieving a given level of utility, e.g., the introduction of the VCR, it is possible that the cost of maintaining a given level of utility can go down while the price of a fixed market basket can go up.

This basic distinction between a cost-of-living index and the CPI leads to the general conclusion that the CPI overstates the increase in the cost of living. The main exception is during wartime, when the fixed market basket of the CPI may include goods that are subject to price controls and rationing or are simply unavailable. In this case the CPI may substantially understate the increase in the cost of living.

Agencies responsible for the CPI in various nations are usually explicit that it is not their intention to measure the cost of living, but there is no escaping the underlying virtue

of mimicking a cost-of-living index as closely as possible. For instance, a longstanding use for the CPI is the escalation of wages under indexed contracts, and here the explicit desire is to compensate workers for inflation so that they might maintain a constant "standard of living," i.e., level of utility. The same idea motivates the use of the CPI to escalate social security benefits and other transfer payments. If the CPI overstates the rate of increase of a pure cost-of-living index, then workers and retirees are being overcompensated for inflation.

Agencies are hard pressed by critics who urge that weights be revised more frequently and that new products be introduced sooner, and in fact such agencies usually defend their practices not on the grounds of any principle but rather by appealing to insufficient government funding in their agency as the source of most problems.

Sources of Bias in the CPI Relative to a Cost-of-Living Index

The accusation that the CPI systematically overstates the increase in the cost of living, and thus leads to an understatement of increases in the standard of living, is serious but longstanding.[6] Although they all lead to a bias in the same direction, the problems can be subdivided into separate categories:

• *New Products.* If a new product is developed that is eagerly embraced by the consumer, e.g., room air conditioners, TV sets, or VCRs, a CPI that includes only a fixed market basket of older products will understate the gain in consumer utility, even if the

---

6. A similar statement is contained in the 1961 "Stigler Report" (NBER, 1961), pp. 51-4.

prices of these older products are measured with perfect accuracy. In the past, the CPI was notorious for the late introduction of products, e.g., the CPI did not introduce autos until 1940, more than two decades after Ford's "Model T" brought the automobile to the average family. Penicillin entered the CPI in 1951, after it had already experienced a 99 percent decline from its initial price. Air conditioning entered the CPI in 1964, more than a decade after the widespread sale of such products.[7] The pocket calculator entered the CPI in 1978, after it had declined in price about 90 percent from early models introduced in 1970.

Most newly introduced commodities follow the "product cycle," falling in relative price at first as economies of scale and learning-by-doing reduce their cost of production, and then rising in relative price late in the product cycle as they are supplanted by other new products. While the CPI cannot be expected to quantify the gain in consumer welfare at the moment such a product is introduced, it must follow the relative price decline at the early stage of the product cycle with appropriate expenditure weights if the entire benefit of the innovation is not to be missed entirely. Further, statistical agencies have tended to be extremely conservative in assessing the function of new products and have been reluctant to incorporate directly the price decline involved in supplanting an old product by a new product with a comparable function. For instance, the VCR allowed rental movies to be enjoyed in the home on demand and substantially supplanted the old activity of hiring baby sitters for a trip to the local movie theater. An even more dramatic case largely involving a producer durable involved the supplanting of the old rotary electric calculator by the

---

7. My book (1990) develops price indexes for room air conditioners from two sources, the Sears catalogue and *Consumer Reports*, that indicate that this product was sold in the millions as early as 1952-53.

electronic calculator; all of us can purchase for $10 or so a calculator that can perform all the functions (in a fraction of the elapsed time) of an old 1970-vintage $1000 rotary electric calculator. Yet in the U. S. the electronic calculator was treated as a new product, and the decline in price from the obsolete rotary electric model to the early models of the electronic calculator was "linked out" in the official price indexes.[8]

    • *Quality Changes in Existing Products.* There is no sharp distinction between the problem posed by new products and quality change in existing products, because some new products can be viewed as alternative methods of obtaining a given level of consumer utility. In the previous example, the electronic calculator is less a new product than a price reduction in the old generic product called a "calculator." As another example, piped natural gas central heat may be treated as a new product, yet may be viewed as representing a quality improvement in the old product "heat" that was previously represented by dirty and inconvenient coal heating.

    Most research by economists on quality change, dating back to Griliches' classic (1961) paper, has involved the application of the hedonic regression technique to particular varieties of durable goods, and the ambiguity involved in the definition of a "new product" has not been directly addressed. While a decade ago there was a tendency for analysts to doubt that there had been any systematic quality bias, there is now general recognition that there is a pervasive bias not only resulting from improvements in product quality but from

___

8. The only ambiguity in the calculator example is the extent to which the electronic calculator substituted for the slide rule, as opposed to substituting for the rotary electric calculator. The fact that both of the old products have been made obsolete indicates that the new product was superior to both, and the only practical question is one of obtaining appropriate weights for the relative importance of the old models.

innovations in distribution that allow retail markups to decline.[9] We return to these issues

below.

 • *Changes in Relative Prices.* The classic bias in a fixed-market-basket index

is caused by shifts in consumer purchases in response to changes in relative prices. This

involves not just long-run shifts in the prices of, say, chicken or beef, but also the tendency

of consumers to "bunch" purchases at the time of sales. Even if the CPI meticulously tracks

the price decline of every product during every sale, its fixed-market-basket weights

(applying to average purchase patterns over a time period of a year or more) will miss the

shift of weights implicit in the bunching of consumer purchases. This problem does not

cause a bias in the measurement of the rate of inflation if sales occur at the same frequency

and with the same relative price reduction year after year. But an upward secular bias in

the measured rate of inflation could occur if sales become more frequent, and a cyclical bias

could occur if sales were more frequent in recessions than during booms.

 • *Changes in the Availability of Discount Outlets.* Attention has recently been

drawn in U. S. research to a potentially important form of substitution bias, the shifting of

consumers from conventional retail outlets to discount stores. At present the CPI treats a

container of toothpaste purchased at a "ma and pa" drugstore at a high price as a different

commodity from the same container of toothpaste purchased at a discount store. The

decline in the average price paid when the consumer shifts to the discount store is "linked

---

9. Triplett (1975) argued that there were as many examples of downward as upward bias in the CPI. However at that time he did not have access to the evidence of pervasive quality bias developed in Gordon (1990) and of substantial outlet bias in the CPI developed by Reinsdorf (1991a).

out," that is, not treated as a price decline.

A difficult theoretical issue is involved here, since the advent of self-service several decades ago did create the presumption, argued most forcefully by Schwartzman (1969), that the new outlets were providing a different and inferior type of retail service than the older outlets, thus justifying the linking out of the price decline. However, more recently the Schwartzman position has come under fire, since most of the outlet shifting in the past decade has been from established self-service outlets to newer and more efficient self-service outlets (e.g., Wal-Mart) that can sell the same goods with the same service level at lower prices. Unless consumers perceive the new outlets as inferior in service quality to the established outlet by the same amount as the price differential, the CPI procedure of linking out the shift in purchase patterns introduces an "outlet substitution bias."

Estimates of Traditional Substitution and Outlet Substitution Bias

There are no comprehensive estimates of the four main sources of bias in the CPI relative to a cost-of-living index. This results from the difficulty of obtaining estimates of the true price change implicit in the introduction of new products and quality changes in existing products. In a suggestive but partial investigation of the quality change bias for automobiles, major appliances, and a few types of minor appliances, I concluded (1990) that over the period 1947-83 the CPI overstated the rate of price increase in consumer durables by an average of 1.5 percent a year. This figure is an average of a conservatively assumed zero bias in those product classes for which I had no information, and higher rates of bias in those categories that I did cover (e.g., annual rates of 1.7 percent for automobiles, 3.2

percent for major appliances, and 5.9 percent for radio and TV sets).[10]   Only quality
changes on existing products are included, and no attempt was made to estimate the value
of totally new products.   In Part III and IV below we review some of the theoretical and
practical issues involved in making such estimates of the quality change bias.

Substitution bias from changes in relative prices, the third source of deviation
between the CPI and a cost-of-living index, has been frequently studied. One of the most
careful recent studies indicates that the problem is relatively minor, with an upward bias of
0.18 percent per year (Manser and McDonald, 1988).[11] An alternative to relying on such
internal BLS studies for the measurement of substitution bias is to examine indirect
evidence provided by the BEA, which computes three alternative price indexes for total
output and for consumption. These are the implicit deflator, the fixed-weight price index
(comparable to the CPI for years after the base year for which the fixed weights are
calculated), and a "chain-weight" index that approximates an ideal Divisia or Törnqvist-type
index. Table 1 illustrates the calculation of the three indices for a hypothetical example
applying to the first and second quarters of 1980, when the base year was 1972. As shown,

---

10. Gordon (1990, Table 12.10, p. 552). My automobile index is based half on a hedonic index adjusted for
changing retail discounts and half on the CPI prior to its adjustments for safety and pollution devices; the
combined index is then adjusted for the effect of these devices and of changing fuel economy. Marshall
Reinsdorf in correspondence has suggested that the CPI, the source of half of my index, may incorporate a
downward secular bias caused by an attempt by auto manufacturers to mask true price increases by timing them
to coincide with important changes in model characteristics. Since the CPI deletes some price changes involving
major changes in model characteristics, it relies disproportionately on a sample of models undergoing little or
no design change and hence having an unrepresentative low rate of price increase.

11. The Manser-McDonald study presents alternative measures of the bias ranging in size between 0.14 and
0.22 percent per year; the estimate of 0.18 is based on a superlative index as the measure of the cost-of-living.
For other research in this area see Christensen and Manser (1976) and Braithwait (1980).

large differences can result, particularly when there are sharp quarter-to-quarter changes in relative prices or weights.

Since the currently published BEA fixed-weight indexes use a base year of 1982, we would expect the superior chain-weight index to increase less after 1982 and more before 1982 than the fixed-weight index. Indeed, this is what occurs. The average bias indicated in the fixed-weight consumption deflator as compared with the chain-weight consumption deflator for 1982-90 is only 0.08 percent per year, although this may be an understatement because the alternative deflators are not based on as fine a level of disaggregation as the BLS studies cited above. However, in the decade before 1982, the bias is 0.3 percent per year (the excess in the increase of the chain index over that in the fixed-weight index), indicating a bias in the same range as the BLS studies.[12]

Reinsdorf's (1991a) study of outlet substitution bias in the CPI for food and fuel finds a much more significant problem, at least for food, than have studies of traditional substitution bias. His most important finding comes from comparing the average price of an item and the most closely corresponding CPI expenditure class index for individual narrowly defined food items (e.g., "apples, red delicious"; "cola, non-diet, can, 72 oz., 6 pk."). He finds that:

> " . . . means of the relative CPI food indexes . . . show an average annual increase of 4.2 percent while the weighted mean of the average prices grows at only a 2.1 percent per year rate.

---

12. The BEA shifted to a 1987 base year in December, 1991. The new indexes required to update the bias estimates in this paragraph were not available at the time this paper was completed.

This implies an outlet substitution bias for food in the CPI of

about 2 percent per year during the Eighties."

Reinsdorf rejects numerous possible explanations of this result which would not involve a

bias in the CPI, including a geographical shift to lower priced regions or a shift away from

branded goods. The marked increase in the market share of discount food outlets is

interpreted as evidence that consumers consider the lower prices as more than sufficient to

offset any decline in the quality of service at such outlets. He leaves open whether the

entire 2 percent per annum drift should be treated as an error in the CPI.[13]

While the Reinsdorf study covers only food and fuel, it is plausible that other future

studies may find a substantial outlet substitution bias in apparel, housewares, appliances, and

other products for which discount outlets like Wal-Mart have been gaining market share.

Recent reports suggest a quantum leap in the fraction of brand-name apparel and other

goods being sold by "outlet malls" at discounts of 33 percent to 70 percent relative to

traditional merchants. These lower prices are being obtained with equal or better service

to that in traditional stores.[14]

An important qualification to our discussion of outlet substitution bias is that a bias

is introduced only if there is a permanent shift to new lower-priced outlets offering similar

---

13. While this would imply that consumers may receive a lower level of service in the new types of discount stores, they have also shifted purchases toward convenience stores that provide more service and convenience, so the two types of shift in service quality may wash out.

14. See Helliker (1991). Regarding service, this source cites a survey of 818 outlet shoppers that found a "majority" deeming service at outlet stores superior to that at department stores. Other examples of shifts to higher quality and/or lower priced marketing are provided in "Value Marketing," *Business Week*, November 11, 1991, pp. 132-40.

or superior service. There mere fact that a given type of outlet sells some or most of its goods at "sale" prices implies no secular bias in the CPI, even if the frequency of such "sales" increases, since CPI price reporters will encounter these lower sales prices as part of their regular process of checking a given set of outlets.[15]

Overall, we conclude that the traditional problem of substitution bias in the CPI is probably of minor importance compared with the bias introduced by the new products, changes in the quality of existing products, and outlet substitution bias. A broader perspective across countries that searches for sources of systematic bias in measured rates of inflation or rates of output growth supports this conclusion. Substitution bias is likely to be of a similar size across countries. But new product, quality change, and outlet substitution bias may differ substantially. However, the extent to which measured rates of inflation differ across countries because of different measurement methods is largely a mystery. And we need to learn more about how the CPIs of other countries handle outlet substitution.

III. OUTPUT PRICES: THE BASIC ISSUES

Fisher and Shell (1972) have argued for a conceptual basis for an output price index that is analogous to the cost-of-living index concept underlying the CPI. In their interpretation, the indifference surface that serves as the reference point for a cost-of-living

---

15. A formal exposition of the point that an increase in sales, interpreted as an increased dispersion of prices, has an ambiguous efect on the bias in a cost of living index that ignores shopping behavior, is contained in Reinsdorf (1991b).

index is replaced by a production possibility surface as the reference point for an output price index. That is, the task is to determine what revenue the multi-product firm must receive in situation B after optimal movement (in response to a new set of output and input prices) along the initial situation-A production possibility frontier, while holding all input *quantities* fixed. The output price deflator is the ratio of the required revenue in situation B to the stated revenue in a situation A. This deflator can then be used to convert measures of nominal output into measures of real output.[16]

Accurate measures of output prices involve the same general set of issues as do consumer price indexes, but there is a different emphasis. The primary focus is on components of output other than consumer goods, particularly producers' durable goods and structures. Here the price indices of concern are the investment deflators in the national accounts. Since these are not based on the fixed market-basket approach of the CPI, there is no need to discuss the traditional form of substitution bias, and since price quotes are obtained directly from manufacturers, there is much less room for outlet substitution bias. Thus we are left with new products and quality change as the main issues, and these have been the subject of intensive debate recently as a result of dramatic changes in the performance-to-price ratio of electronic computers, an increasingly important type of capital good.

---

16. Suggestions from Marshall Reinsdorf helped to improve the exposition of this paragraph.

The Output Price Index in Characteristics Space

Triplett (1983) has restated the Fisher-Shell theory in terms of output characteristics (e.g., computer calculations), in order to explore situations in which quality change can be quantified as an increase in the quantity of characteristics embodied in the good. Obviously, this approach is easier to apply to existing goods like computers in which there are a few easily measurable characteristics than to quantifying the output represented by totally new products with new types of characteristics. Like that of Fisher and Shell, Triplett's definition of the output price index holds constant the production function and the quantity of inputs, and allows the price regime to vary, as well as the optimally chosen quantities of output characteristics.

Triplett's analysis clarifies the treatment of quality change, which had previously incorporated a confusing distinction between the "resource cost" and "user value" criteria. The old debate centered around whether a durable good had the same quality if it (a) had the same resource cost in the base period or (b) had the same value to the user in the base period. Triplett showed that the proper quality adjustment in an output price index is:

> ". . . equal to the value of the resources required to move the
> set of output characteristics included in the index back to the
> same production possibility curve. This is precisely the resource
> cost quality measurement rule that has been argued in the
> literature" (1983, p. 299).

Note that technology must be held constant at the base year level, but now the criterion of

comparison has shifted to the resource cost *of the characteristic*, e.g., the computer calculation, and is no longer the resource cost of the good itself, e.g., the computer "box." In effect, under Triplett's criterion two computers are considered equivalent if their computation power would have cost the same to produce in a base year. Yet two such computers with the same computation power also have the same "user value," so that the resource cost and user value criteria amount to the same thing.

While the Triplett clarification may seem from today's perspective a restatement of the obvious, in fact the BEA had long maintained a nominal-fixed computer price index throughout the 1960s and 1970s ("always and forever fixed at 1.0"), reflecting the belief of former BEA head George Jaszi that two computer "boxes" should be treated as equivalent if they cost the same to produce in a base year, even if their computational ability was vastly different.

Adjusting for Nonproportional Quality Change

Most of the practical problems in developing output price indexes are posed by quality changes in durable goods, and we use the example of computers and aircraft, important types of durables, in the illustration of methodology in this section. Subsequently we discuss the treatment of quality change in nondurable goods and services.

A rigorous statement of criteria that should be used for quality adjustment of durable goods begins by asking what function producers' durables perform and recognizes that producers' durables are normally an input, along with labor, structures, energy, and materials, into the production of goods and services consumed by final users. Consumer

durables may also be considered an input, producing the services of durable goods. The crucial step in developing adequate deflators for durable goods is to recognize that final market product (Q) is produced by a vector of market-purchased input characteristics (X):

(1) $Q = Q(X)$, $Q_X > 0$, $Q_{XX} < 0$.

An input characteristic is defined as any attribute of a market-purchased input that has a positive marginal product, including in the case of durable goods the horsepower and physical dimensions for a truck, or memory size and calculations per unit of time for a computer. In Triplett's more precise definition, a quantity is an input characteristic if it reduces the unexplained variation in output, given the explanation contributed by all the other arguments in the production function. In simpler language, two durable goods are considered equivalent if they have the same marginal product, i.e., the criterion of "user value" is the marginal product.

In determining the proper treatment of innovations in durable goods, we start with the types of shifts in the performance-to-price ratio that have been typical throughout the evolution of the computer industry, and then apply the same ideas to changes in energy efficiency or other aspects of user cost. We can call technical innovations "nonproportional" when they raise the performance of a good by increasing its built-in quantity of characteristics (X) relative to the resources used by the supplying industry.[17] While process

---

17. This discussion is adapted from Baily-Gordon (1988, pp. 377-81) and summarizes part of chapter 2 of Gordon (1990).

innovations occur continuously which improve the productivity of inputs in the manufacture of *given models* of durable goods, the shifts which concern us here are those innovations involving both changes in processes and in product specifications which occur when a *new model* is introduced. Such an innovation takes the form of a downward shift in the real cost of producing a given quantity of characteristics, say computer calculations. There is no shift in the user firm's production function (equation 1), since a single calculation still produces the same amount of final output (Q). The quality change, though "nonproportional," is not "costless," since the reduction in cost must consume managerial and R&D resources, or else it would have occurred long ago.

We depict a nonproportional new-model introduction in Figure 1 by plotting two upward sloping lines plotting the cost function $C(Q,\lambda)$ of "old" and "new" models of a particular type of durable good corresponding to two different values of the technical shift parameter λ. Initially, output level Q^* is produced at an input cost of C_0 at point A. The technological change represented by the shift from λ_0 to λ_1 improves quality by raising the quantity of input characteristics relative to their cost. This raises the demand for characteristics and the level of output, depicted by Q_1 in the diagram. The unit cost of the durable good (C_1) could be either higher or lower than in the initial situation (C_0).

The change in the price index of the durable good is simply -AF, the vertical downward shift in the supply schedule itself. This would be measured in practice by taking the observed reduction in the price of the machine (-AE) and adding an adjustment factor equal to the change in output produced by the extra characteristics (EB) times the marginal

cost (EF/EB) of building the extra input characteristics. This extra adjustment is what is accomplished by estimating regression coefficients for the value of greater speed and memory in the case of the hedonic price index for computers used now in the NIPA. For instance, if a new model computer costs 10 percent less than an old model, but the regression coefficients imply that its greater speed and memory represent 40 percent higher quality, the measured price decrease is not 10 percent but rather 50 percent. In the example of the hedonic price index for computers now used in the NIPA, such calculations lead to an annual rate of price decline of about 20 percent. Shifts in the supply curve like that depicted in Figure 1 have greatly increased the performance of mainframe computers without reducing their average price. For instance, the IBM model 8040, introduced in 1983, cost the same $1.3 million as the IBM model 7080 introduced in 1961, but had a calculating power 1452 times as great.[18]

The idea of nonproportional quality change can be extended to changes in energy efficiency and other characteristics that affect user cost. Now a nonproportional quality change is one which raises the "net revenue" earned by a machine (gross revenue minus operating cost, prior to the deduction of depreciation and interest) relative to the machine's cost, holding constant the price of output, energy, and labor when comparing the net revenue of two alternative models incorporating different technologies. For instance, simply building a more expensive refrigerator to save energy with existing technology, with the extra cost of the refrigerator equal to the present value of energy saving, does not represent a

---

18.   This example is taken from Gordon (1990, Table 6.8, pp. 217-8). Over the same interval, 1972-83, the average unit price of a mainframe computer "box" increased from $0.3 to $1.0 million.

nonproportional improvement. But the invention of a new form of insulation that allows all refrigerators to be more efficient at the same refrigerator price would call for a quality adjustment and a decline in a properly measured price index. The most dramatic example of such a change was the invention of the jet plane, which raised the net-revenue-generating ability of a DC-8 compared to a DC-7 by a factor of 10 at a capital cost only three times higher.

The ultimate test of this approach to the measurement of durable goods prices is to compare its predictions with the verdict of the used asset market, as has been done in the case of used aircraft and personal computer prices.[19] How does the approach differ from the deflation procedures now used in the NIPA? The NIPA computer deflator treats nonproportional quality change by the method that we recommend, so that there is no dispute in principle. But as a practical matter many improvements in the ratio of performance to price for durable goods are missed, either because official price indexes (mainly components of the Producer Price Index) make inadequate allowances for quality change, or because price indexes do not exist for important types of capital equipment where technological progress has been rapid, e.g., electronic telephone switching equipment. As for improvements in energy efficiency, there are no explicit adjustments in the NIPA.[20]

---

19.   On used aircraft, see Gordon (1990, Chapter 4). On used PC prices see Berndt-Griliches (1990).

20.   Improving fuel economy for automobiles has been implicitly taken into account, at least in part, by the dedcision of the BLS not to treat the shrinking dimensions of "downsized" automobiles as a decline in quality. If automobiles of a *given* size and performance have better fuel economy now than 20 years ago, however, this aspect of quality improvement has not been taken into account.

The Distinction between Input and Output Indexes

The work of Fisher-Shell (1972) and Triplett (1983) places major emphasis on the distinction between input and output price indexes. In the case of a machine that is both an input to a machine-using industry and the output of a machine-producing industry, the input price index should be used in calculating measures of real capital input in the using industry, while the output price index should be used in calculating output and productivity indexes for the machine-producing industry.

The distinction between input and output indexes for producers' durable goods creates the potential for inconsistency in measures of investment and capital. The real net investment component of national product is defined as the change in the real capital stock. If the price change of a machine is measured differently by the input and output price indexes, then the resulting change in real capital input will not be compatible with the computed change in net investment. Fortunately, both cost-increasing and nonproportional quality changes are treated identically by input and output price index measures.[21]

Why, then should we bother to distinguish between the input and output price index concepts? The only justification for a distinction arises when output and input change are not identical. The primary example concerns the addition to a machine of a pollution-control device (e.g., a catalytic converter on a truck or environmental "scrubber" on an electric generating station). Such a device does not actually produce output and thus consumes resources in the machine-producing industry without raising input in the machine-

---

21. This equivalence is demonstrated formally in Gordon (1990, pp. 48-52).

using industry. In such cases, input and output price indexes can diverge, and explicit accounting is required. As an example, I have published price indexes for trucks that treat the extra cost of anti-pollution equipment as a quality improvement rather than as a price increase. This is the appropriate index if we are interested in measuring the output and productivity of the truck-producing industry. Yet in a study (1991) of productivity in the trucking part of the service sector, that is, the industry that *uses* trucks, I have recalculated the price indexes to treat the value of the anti-pollution equipment as a price increase rather than a quality improvement, since it does not directly create output in the truck-using industry.

Consumer Durables, Nondurables, and Services

The treatment of quality change in consumer durables is similar to that for producers' durables. Indeed, in my own work in the measurement of quality change for durable goods, I have applied the same techniques interchangeably to both producers and consumer durables. The main problem arise in the treatment of anti-pollution devices, and in the interpretation of corrections for improved energy efficiency.

Cost-of-living price indexes are usually interpreted as input price indexes, measuring the "input" of goods and services into consumer welfare. This might appear to imply that anti-pollution devices should not be treated as quality improvements, since in the case of such devices on automobiles, they do not directly change the service of motoring that a consumer receives from the automobile. However, the U. S. statistical agencies have long treated such devices as improvements in quality rather than in price, generally on the

argument that a rational legislative process would require the adoption of such devices only up to the point where their marginal costs equal their marginal benefits. In practice there is little evidence on the benefits of antipollution devices that would allow such a cost-benefit calculation to be made.

A question may be raised about my earlier suggestion that quality adjustments in price indexes be made for nonproportional improvements in energy efficiency. For a consumer price index, the alternative would be to leave the price index for automobiles unaffected by such improvements, and instead in the CPI measure changes in the input service price per unit of time of the combined costs of capital, fuel, and maintenance. A fuel saving would thus emerge as a reduction in the fuel component of the service price. This approach would be sensible if the only function of official price measurement were to deflate consumer expenditures on a flow of services. But it would leave us without sensible deflators for the output of durable goods. If the automobile industry shifts research resources from achieving improvements in speed or power to improvements in fuel efficiency that are of the same value to the automobile user, we do not want a price index that treats faster speed as a quality improvement but fails to treat improved fuel efficiency as an equivalent improvement.

The vast majority of research on quality change has been devoted to durable goods, and little is known about the nature of possible mismeasurement for nondurable goods and services. For food products, it has been generally assumed that "an apple is an apple," and that the "new model problem" is relatively unimportant. The fact that consumers have

increased the share of food expenditures on frozen and "microwavable" products, and on in-store deli and bakery items, suggests that they perceive the extra benefits of these new products to be worth more than their extra cost. However, there is no readily available method to adjust for the quality change, and I would judge the price index bias, if any, to be relatively small, particularly in relation to the large outlet substitution bias in the CPI for food, as discussed above. There is probably more potential for a quality bias in such nondurable goods as apparel, shoes, and tires, particularly improvements or deteriorations in durability, but thus far there is only scattered research on this issue.[22]

IV. TECHNIQUES OF QUALITY ADJUSTMENT

The two main methods by which statistical agencies adjust for quality changes in existing products are the conventional "specification" method and the hedonic regression method. Econometricians embraced the hedonic technique soon after Griliches' seminal (1961) application to U. S. automobile prices, and, for a while, it seemed only a matter of time before the hedonic regression technique would become pervasive. Indeed, by far the most dramatic change in the price index methodology of the U. S. NIPA was the adoption in 1986 of a hedonic price index for computers. Yet there is more at issue than an apparent debate between an obsolete conventional method and the sophisticated and modern hedonic method. There is a growing recognition that hedonic indexes are not suitable for measuring

---

22. Moore (1978) has calculated that between 1935 and 1978 the price of tires per mile of tire life declined by 9 percent, in contrast to an increase in the CPI tire index of 140 percent. Over the same period, the price of motor oil per mile declined by 52 percent as compared to an increase in the CPI of 234 percent.

the prices of many types of complex goods.

The "Specification" Method

The primary source of U. S. price data for capital equipment is the PPI compiled by the BLS. Each month thousands of price changes are reported to the BLS in questionnaires returned from manufacturers, and these are taken at face value unless the respondent informs the BLS that there has been a model change. If so, the BLS has four different methods to cope with the model change.

• *Ignore it.* This is the "direct comparison method," which attributes none of the price change to a quality difference between models. There is no necessary bias, because quality can decrease as well as increase. But in practice durable goods tend to undergo steady improvement, and often a better model is introduced with no change in price, causing the quality change to be missed entirely.[23]

• *Delete it.* The price change between the old and new models is simply deleted, and the weight attributable to this product is applied to the average price change of other products in the same commodity classification. Again, no necessary bias is implied, although there could be a tendency for the "other products" that proxy for the missing price quotation to be older and further along in the product cycle, and thus prone to above-average price increases. Stated another way, the PPI consists disproportionately of commodities of constant quality.

---

23. A possible counterexample is in the field of apparel, where the BLS found that the "ignore it" approach avoided a serious downward bias that occurred when the "delete it" appropach was used.

• *Link it.* If the new and old model are sold simultaneously, the difference in quality can be taken to be the price differential at the time of introduction of the new model. Unfortunately, new models usually replace old models entirely, and the link prices are not observed. Also, in the context of Figure 1, a quality improvement in model B can occur even if model B costs less or the same as model A.

• *Cost it.* Corresponding most closely to the desired conceptual framework is to estimate the cost of the extra characteristics of the new model, which requires an estimate of the slope (EF/EB) of the marginal cost curve in Figure 1. Problems in practice with the costing method have been its infrequency of use, particularly before 1960, and the fact that it has been applied disproportionately in the case of automobiles relative to other products. This raises the possibility that there is a spurious downward "drift" in the relative price of automobiles due to uneven application of the costing method. As shown in Table 2, the costing method is much more important in transportation equipment than for other types of capital goods. And the drift may have become more important as the BLS has become more diligent about applying the costing method to autos but not to other comparable products. For instance, I showed (1990, p. 89) that there was a decline in the relative price of motor vehicles relative to agricultural machinery of 0.8 percent for 1950-60, 22.9 percent for 1960-70, and 44.8 percent for 1970-80.

Hedonic Regression Technique

The literature on the hedonic technique has become quite large and cannot be reviewed in any detail here; interested readers are referred to important surveys and

interpretations by Griliches (1971, 1990) and Triplett (1987, 1990). In the context of this discussion, the hedonic approach can be viewed as an alternative method to manufacturers' cost estimates in making quality change adjustments. It assumes that the price of a product observed at a given time is a function of its quality characteristics, and it estimates the imputed prices of such characteristics by regressing the prices of different models of the product on their differing embodied quantities of charcteristics. Thus the hedonic approach is less a new method than an alternative to cost estimates to be used when practical factors make it more suitable than the conventional method.

What are these practical factors? The hedonic index is most suitable for products where the main determinant of price differences among models available on the market at the same time is a limited set of easily quantifiable characteristics. Mainframe computers ideally suited for hedonic analysis, as most price differences among models are attributable to differences in speed and memory, and peripherals (disk drives, etc.) are priced separately. Studies of personal computers, e.g., Berndt and Griliches (1990), need a longer list of explanatory variables, as peripherals like disk drives are usually packaged and sold with the basic machine. At the other end of the spectrum are complex goods like autos, where there are literally hundreds of characteristics that determine price. Often these are highly multicollinear, e.g., larger and heavier cars also have more of other characteristics, and the first generation of hedonic studies (e.g., Griliches, 1961) used weight as a proxy for all these other characteristics. Unfortunately, this led to nonsensical results when cars were "downsized" in order to improve fuel economy, and when lighter materials were introduced.

Recently Triplett (1990), a pioneer in the field of automobile hedonics, has concluded that "the automobile may be too complicated for hedonic studies."

Even in the case of mainframe computers, where the list of important characteristics is relatively short, a substantial bias can creep into hedonic indexes. First, the existing hedonic indexes for mainframe and personal computers are separate, treating each as different products. Yet there is an enormous scope for substitution, which has indeed occurred on a massive scale as users have responded to the much lower price per calculation (MIPS) of PCs. One suggestive example for the 1972-87 period indicates that the rate of price decline for users able to switch from a typical 1972 mainframe to a 1987 PC using a 386 chip would have enjoyed an annual rate of price decline of more than 35 percent, not the 20 percent that typically emerges from hedonic studies of mainframe and PC prices (Gordon, 1990, p. 239). Second, our conceptual discussion above suggests that the "net revenue" of the computer center rather than the price of the computer is the correct dependent variable for a hedonic regression, and existing hedonic indexes greatly understate the percentage improvement in net revenue that is made possible by higher calculating power, since there is no corresponding increase (and may in fact be a decrease) in other factor inputs of the computer center, including labor, electricity, and space.

Finally, by their very nature hedonic indexes require large amounts of data. Given the thousands of separate products that are produced in modern industrial societies, the need to collect a full cross-section of data on each product presents an insurmountable obstacle to the full-blown adoption of the hedonic technique. As a result, the hedonic

technique is likely to be used in official price indexes only for a few very important products, e.g., houses and computers.

The Payoff from Using Unconventional Data Sources

Recently (1990) I completed a major study of durable goods prices for the period 1947-83 that combined more than 25,000 annual price observations from sources independent of those used in the two main U. S. price indexes, the CPI and PPI. The main methodological innovation is to equate two investment goods when they make the same contribution to net revenue, i.e., to adjust for quality differences not only in performance but also in fuel efficiency, durability, and maintenance requirements. Such quality adjustments for operating characteristics were made for aircraft, electric generating equipment, railroad locomotives, consumer appliances, and TV sets. For these and many other products, a wide variety of data sources yielded radical conclusions, even though most price comparisons were made by the conventional specification method.[24]

The results of the study yield radical conclusions, as summarized in Table 3. The new producers' durable equipment (PDE) deflator rises 3.0 percent a year more slowly than the NIPA PDE deflator for the full 1947-83 period, with a somewhat larger negative drift in the first half. The most startling impact is on the ratio of real PDE spending to real GNP, as shown in the second section of Table 3. The ratio of total durables spending to

---

24. For the most important single data source, the Sears catalogue, the method used was what the BLS calls "deletion." That is, price comparisons between two years were based only on models which were absolutely identical. There was sufficient model overlap for given products that the product cycle bias discussed above was not important in practice. For other data sources a variety of methods was used, including the hedonic method, and calculations of net revenue (rather than direct price comparisons) for alternative models.

GNP (section 4) rises between 1947 and 1983 by 29 percent in the NIPA version and 170 percent in the new version.

The finding on line 2 of Table 3 that the ratio of equipment investment to output has increased rapidly during the postwar period, instead of remaining roughly constant, has important implications for longstanding debates regarding the interpretation of growth theory and the sources of growth. In contrast to the standard neoclassical growth model with its constant investment-output and capital-output ratios, it appears that there has been steady growth in these ratios, fueled by a steady decline in the relative price of investment goods. However, despite the radical increase in the growth rate of investment implied in these numbers, they have a surprisingly small impact on the contribution of capital as a source of growth. This occurs because the new deflators make not just capital input but also output grow faster, so that the impact on multi-factor productivity growth is much less than on output or capital separately.

V. DIFFERENCES IN METHODS AMONG COUNTRIES

Differences in measurement methods across countries are relevant for at least three issues. First, is the growth of a particular country's output or productivity over- or understated relative to the average for all industrialized countries? Second have methods improved more in one country than another in the last 10 or 20 years, giving the appearance of an increase in the output growth rate that is due to better measurement methods rather than superior economic performance? Third, do the price indexes of some countries give

flawed short-run signals to policymakers, impeding the operation of short-run stabilization policy?

Given the complexity of price measurement in a single country, any discussion of differences across countries quickly becomes mired in details. Accordingly, we can only scratch the surface here. As a general disclaimer, my research thus far has been limited to a detailed study of French and German methods and a sketchy study of British methods.[25] My primary interest in cross-country differences is the extent to which price measurement methods can explain the puzzling difference between the U. S. and other countries in the relative growth of productivity in the manufacturing and nonmanufacturing sectors; the ratio of productivity growth in nonmanufacturing to manufacturing since 1973 has only been about 0.1 in the U. S., and in other major countries ranges from 0.5 to as high as 0.9 in France.

General Approach to Price Measurement and Deflation

At the general level of output and price measurement, the U. S. is ahead in its use of a hedonic price index for computers in price deflators for consumer durables, PDE, exports, and imports. This immediately explains some of the puzzle, since productivity growth in manufacturing in other countries is understated due to the absence of a comparable price index for computers, except perhaps in Japan where the measurement of computer prices by the average "price per MIPS" may be equivalent for practical purposes to the more elaborate U. S. hedonic approach.

---

25. Comments below on Japan are based on the conference discussion of an earlier draft of this paper.

But there are other areas in which U.S. methods seem inferior to those used in the other countries. Japan appears to take a much more aggressive approach in searching out new products and new models in its Producer Price Index, and the turnover of models included in the PPI is much higher than in the U. S.[26] The most pervasive weakness of the U. S. approach involves weighting methods, and this is unfortunate, since weighting issues are more important for the U.S. than for other countries due to the rapid change in relative prices introduced by the hedonic price deflator for computers. Only the U.S. appears to use a single base year for the entire history of the national accounts. This practice causes output growth to be biassed down prior to the base year and up after the base year for goods with rapidly falling relative prices. But also it implies that the relative importance of different goods and services changes whenever the fixed base year is changed. For instance, when the U.S. changed in early 1986 from a fixed 1972 to a fixed 1982 base year, the relative share of energy production and energy imports increased sharply. A more bizarre example reflects the fact that the U.S. maintains its accounts with a single base year back to 1909. Since the relative price of government services has increased steadily through the decades, every time a new base year is chosen, World War I becomes bigger (i.e., the share of government wartime expenditures in World War I increases as a share of GNP)!

The U.S. is apparently the only country that forces the relative price of a single base year to apply throughout the history of the national accounts. The use of shifting base years in France, Germany, and the U.K. increases the accuracy of the relative price structure used

---

26. I am told that there is a 10 percent annual turnover of individual models in the Japanese PPI each year.

to aggregate the output and price indexes. This major advantage comes at a cost for users of the accounts: different sets of tables are provided for each base year, and the user must go to the extra work of linking when a long time series is desired. For instance, the French accounts are published for 1949-59 on a 1956 base, 1959-79 on a 1970 base, and 1970-present on a 1980 base. The need for linking extends to nominal series, not just real series, since new measurement methods are generally introduced as part of a new weighting system. Thus the nominal value of construction output in France for 1970 is different in the 1980-base accounts than in the 1970-base accounts. As the base year is updated, numerous other aspects of the accounts change, again inhibiting links. For instance, the French accounts are available at a progressively greater level of industry detail as the base year shifts from 1956 to 1970 to 1980.

The German system is similar to the French in most respects, although I have obtained only two sets of accounts, with a 1970 base year back to 1960, and 1980 base back to 1970. The U.K. accounts rebase every five years. Historical U.K. data for 1978-83 use 1980 weights, 1983 to the present 1985 weights, and so on. Frequent U.K. rebasing carries with it a cost in terms of ease of data availability; for instance, the U.K. producer price statistics are simply not available after 1980 in a product breakdown comparable to the pre-1980 period.

Until recently, the other major difference between the U.S. and both France and Germany, as well as Japan, was the much greater use of double deflation outside of manufacturing in these three nations. However in January, 1991, the BEA published revised

industry output measures for 1977-88 that extend double deflation to most of the service sector. The new U. S. output measures result in improved productivity performance in some sectors, particularly transportation, but this is not enough to boost the poor overall performance of the nonmanufacturing sector.

Differences in Specific Industries

While the U. S. has made improvements that make its transportation sector output measures comparable to those in France and Germany, there are still numerous other defects in U. S. price deflators outside of manufacturing. Both Germany and France make a much more systematic effort to correct rental price indexes for quality change. U. S. industry deflators are severely upward biased for banking, finance, and insurance relative to France, where physical output measures (shares traded, insurance policies written, etc.) are more widely used. France, Germany, and Canada do a much more careful job of measuring the prices of quality-controlled components of buildings in order to develop sensible measures of construction prices, whereas U. S. structures price deflators lead to the nonsensical conclusion that the absolute level of productivity in construction has fallen steadily since 1965 and is now lower than in 1948.

While it is too early to draw any firm conclusions, it is likely that measurement issues bias downward the growth rate of manufacturing output in every country, due to inadequate adjustment of price indexes for quality change, but more outside of the U. S. due to the absence of a computer price deflator. Service and construction productivity growth is biased downward in the U. S., probably more than in other countries, due to measurement

problems.

VI. THE HOMEOWNERSHIP DEBACLE AND LESSONS FOR POLICY

To this point, the paper has been devoted to price measurement issues that affect the assessment of an economy's long run performance, particularly its growth of output per capita and of productivity in the aggregate and at the sectoral level. However, price measurement problems can also distort the short-run behavior of inflation indices, and send erroneous signals to policymakers.

Interest Rates, Indirect Taxes, and Perverse Policy Signals

A primary goal of monetary policy is to control the rate of inflation, and an upsurge in the measured inflation rate has often convinced policymakers to tighten monetary policy by raising short-term interest rates. Most central banks focus on the consumer or retail price index (CPI) in their policy deliberations, since this is available monthly with a relatively short lag (rather than available quarterly as is the case with the GNP deflator) and is widely publicized. Since the CPI in most countries is a Laspeyres fixed-market-basket index, there is no distortion from short-run changes in spending patterns (albeit there is a longer-run bias from the traditional substitution and outlet substitution problems discussed earlier).

Yet the CPI may contain two elements, mortgage interest rates and indirect taxes, that cause short-term false signals. Tight policy in the form either of higher interest rates or higher indirect taxes can cause a sharp uptick in the CPI rate of inflation, and lead to

further unnecessary tightening of policy or delay subsequent easing of policy. In the U. S. almost all indirect taxation is carried out at the state and local level, and so changes in indirect taxes are not a tool of stabilization policy. Uncoordinated decisions at the local level lead to smoothly changing effective indirect tax rates and little distortion in price indexes. But in other countries indirect taxes, particularly the VAT, are levied by the central government and can be an effective means of shifting the time pattern of consumer spending for policy purposes, so the policy distortion is a potentially important problem almost everywhere outside the U. S. For instance, in retrospect a sharp jump in the British VAT in 1979 may have caused the Thatcher government to prolong tight policy too long and make the British recession of 1980-81 unnecessarily severe.

The perverse policy signal given by the inclusion of mortgage interest rates in the CPI can be illustrated by the American debacle of 1977-81. Because of an error in the computation of homeownership costs, the CPI was severely biased upward during this period. The climax occurred in the first quarter of 1980, when the CPI registered an annual rate of increase close to 18 percent, whereas in contrast the three alternative national accounts deflators for personal consumption expenditures (PCE) registered an average inflation rate for the same quarter of 12.3 percent, or only 10.8 percent for the four quarters ending in that quarter. The bad CPI news led a panicked Carter Administration to impose temporary credit controls, which in turn caused the sharp V-shaped recession of 1980. In this policy debacle it is important to make the distinction between the Administration, which panicked over the early-1980 CPI news, and the Volcker-led Federal Reserve Board, which

maintained a steadfast anti-inflationary policy during the 1979-82 period that ignored temporary shifts in CPI inflation and instead aimed to hold the growth of monetary aggregates well below the 10 percent inflation rate registered by the PCE and GNP deflators.

As we shall see subsequently, this CPI error has been fixed but in a way that permanently imbeds in the CPI an upward historical bias that distorts long-run measures of growth in the U. S. standard of living. Between 1976 and 1981 the CPI increased cumulatively by 10.8 percent more than the fixed-weight PCE deflator, and almost all of this difference is accounted for by the difference in the treatment of homeownership costs. In contrast, the PCE deflator used the CPI rent index as a proxy for all homeownership costs, and indeed in 1983 the CPI shifted to its own rent index as a similar proxy.

But before 1983 the CPI used an approach that made the fatal error of treating the whole homeowning population as if it were paying an adjustable-rate mortgage, whereas in fact at that time almost all mortgages were fixed-rate contracts with fixed monthly payments that protected homeowners from any impact of rising interest rates. Thus the CPI was registering a spurious increase in the cost of the *service price* of homeownership at the very time (1976-81) when, ironically, homeowners were enjoying massive increases in wealth as a result of the fastest rate of real capital gains that occurred in any five-year period of the postwar period. The CPI compounded this error by *double-counting* home purchase and mortgage payments as two separate items of expenditure, both weighted in proportion to the *total* purchase price of the house, instead of treating the home purchase component as a

capital transaction. In effect, the CPI treated homeowners as buying a house outright, and then separately paying variable-rate mortgage payments weighted as heavily as the house purchase despite the fact that if they had purchased the house outright they would have had no mortgage debt at all! Finally, the CPI erred by computing the actual interest rate paid, rather than the real after-tax interest rate, net both of income tax savings as well as capital gains.

During the period when the CPI was being pushed up by its ludicrous treatment of homeownership, the BLS was well aware of the problem but undecided between two alternative replacement approaches. The first was the rental price proxy approach that was ultimately adopted. The second was the user-cost appraoch that measures the current annual capital and operating cost of homeownership. Ultimately the user-cost approach was rejected because of its ambiguity and volatility. Ambiguity arises because there are numerous ways to measure the ingredients in the user cost formula (interest rates, tax rates, and capital gains). Volatility occurs if unrealized capital gains are included, since the user cost then plummets when house prices are rising fast and jumps when house prices remain stable. The user cost approach, by correcting for capital gains, has the additional defect that is inconsistent with the national accounts, which exclude capital gains from income. In contrast, proxying homeownership cost with a rent index provides an adequate proxy, since competition should keep the rate of increase in rents roughly in line with true homeowner-ship costs, including utilities, net taxes, and the long-run trend in real capital gains.

Since 1983, when the CPI switched to the rental proxy, mortgage rates have generally

declined and the rate of increase in housing prices has slowed. Thus the upward bias in the CPI has never been corrected by an offsetting downward bias. As a result, all measures of the increase in U. S. real personal income and consumption over periods that span 1976-81, e.g., a comparison for the last two decades, understate the improvement by at least 10 percent if they use the CPI for deflation.

To avoid perverse policy signals, those other nations that include interest rates and indirect taxes in their CPIs should, as a minimum, calculate an alternative policy-oriented CPI that excludes those items. A precedent is provided by the U. S., which publishes several alternative CPIs, including one version for consumption net of food and energy, and this so-called "core" inflation measure is generally preferred by the makers of monetary policy in order to avoid an undesirable tightening of policy in response to a temporary oil or food price shock.

VII. CONCLUSION

Price measurement matters. This is particularly true for errors in the rate of inflation over periods of three months to two years. In the American case, the late 1970s and early 1980s provided a textbook case of the follies that occur when policymakers rely on flawed price indexes. Even though the annual increase in the fixed-weight GNP deflator never exceeded 9.3 percent during the Volcker era, a string of months in which the CPI rose by 18 percent at an annual rate in early 1980 panicked the Carter Administration into ill-advised credit controls. And yet much of the early-1980 inflation was due to the perverse

feedback from the tightening of monetary policy in the fall of 1979, as a result of the channel from monetary policy to mortgage interest rates to prices. Not only were millions thrown out of work in 1980 and again in 1982, but the distribution of income was permanently distorted as millions of others receiving social security benefits escalated by the CPI enjoyed a windfall gain, since their indexed payments went up by more than the true inflation rate. The unnecessary extra benefit payments caused the government deficit to increase, and higher indexed union wages put more pressure on firms to raise prices and continue the inflation.

Macro textbooks teach now that the surest recipe for transition to Brazilian-style inflation is the unholy trinity of monetary accommodation, supply shocks, and full indexation of wages and government transfer payments. I would make this a quartet, by adding incorrect inflation measures that allow higher nominal interest rates and indirect taxes to feed into measured inflation, fueling the spiral of higher wages, benefits, and prices. In any country where inflation measures are flawed in this sense, policymakers should insist that statistical agencies create alternative CPIs, or policymakers should carry out this task themselves from data on components of the CPI.

However, one cannot recommend without qualification that policymakers go further by purging the influence of energy and food prices from the inflation measure used for monetary policy. As these same textbooks teach us, policymakers must be sure *ex ante* that an upsurge in the real price of oil is temporary (as might have been rightly anticipated in the case of the Kuwait blip in 1990-91). But a permanent increase in the real price of oil

can cause a permanent increase in the inflation rate if any component of wages is indexed to a price index including oil, and policymakers will lock in that permanently higher rate of inflation if they accommodate the oil price increase.

While the short-run twists and turns of inflation are of immediate concern to the makers of monetary policy, accurate measures of prices over a longer period are of importance to the general policy community. Politicians and journalists in the U. S. often despair that there has been no increase in the U. S. real wage for 20 years, but any such conclusion rests on faulty price measures. There is a systematic upward bias in measures of consumption prices that has persisted, and an upward bias in measures of investment goods prices (except perhaps since 1982 due to the overweighting of computer prices). Thus the behavior of real income and productivity growth in the U. S. is not nearly as bad as it looks.

But there is still considerable reason for Americans to worry, both as they look back in history and look across the oceans. Almost all the factors that cause a secular upward bias in price indices were even more serious before 1970 than after, because techniques of adjustment for quality change have improved. Thus our official measures of the growth of productivity before 1970 are biased downward by even more than after 1970. Some other countries may do a better job of measuring prices in the service sector than the U. S., but few (except perhaps Japan) have acceptable price indexes for high-tech goods in the manufacturing sector. As a result, the growth of manufacturing productivity is likely to be understated by even more in Europe than in the U. S., while U. S. productivity growth in

manufacturing may be closer to published Japanese figures than the official U. S. data suggest.

This paper is obliged to end with a set of suggested improvements in the indices. At little cost every statistical agency should prepare an alternative CPI for policymakers that excludes interest rates, indirect taxes and, optionally, food and energy costs. This first step can be taken at a trivial cost. But it appears from the above discussion that most improvements in measures of inflation needed to create better indices of long-run growth and productivity will require substantial additional resources. In the U. S. context it appears that top priority should be an introduction of new techniques to correct for output substitution bias and an extension of the work already done on computer price indexes to the other major components of high-tech capital and consumer goods. The extensive field work currently carried out by the Bank of Japan to locate new products and new models for the Japanese PPI and national accounts deflators may serve as a possible model to guide the U. S. statistical agencies toward improved methods of quality adjustment that have a high ratio of benefits to cost.

REFERENCES

Berndt, Ernst R. and Griliches, Zvi, 1990. Price Indexes for Microcomputers: An Exploratory Study. NBER working paper 3378, June.

_____ and Triplett, Jack E., eds., 1990. *Fifty Years of Economic Measurement.* Chicago: University of Chicago Press for National Bureau of Economic Research.

Braithwait, Steve D., 1980. The Substitution Bias of the Laspeyres Price Index. *American Economic Review* (March) 70: 64-77.

Christensen, Laurits R. and Manser, Marilyn E., 1976. Cost-of-Living Indexes and Price Indexes for U. S. Meat and Produce, 1947-1971. In Terleckyj, Nestor, ed., *Household Production and Consumption.* Studies in Income and Wealth, vol. 40.

Fisher, Franklin M., and Shell, Karl, 1972. *The Economic Theory of Price Indices.* New York: Academic Press.

Gordon, Robert J., 1990. *The Measurement of Durable Goods Prices.* Chicago: University of Chicago Press for National Bureau of Economic Research.

_____, 1991. Productivity in the Transportation Sector. In Griliches (1991).

Griliches, Zvi, 1961. Hedonic Price Indexes for Automobiles: An Econometric Analysis of Quality Change. In NBER (1961): 173-96.

_____, 1971, ed. *Price Indexes and Quality Change.* Cambridge, MA.: Harvard University Press.

_____, 1990. Hedonic Price Indexes and the Measurement of Capital and Productivi-
ty: Some Historical Reflections. In Berndt and Triplett, ed. (1990): 185-202.

_____, ed., 1991. *Output Measurement in the Service Sectors*. Studies in Income and
Wealth. Chicago: University of Chicago Press for National Bureau of Economic
Research, forthcoming.

Helliker, Kevin, 1991. Thriving Factory Outlets Anger Retailers as Stores Suppliers Turn
Into Competitors. *Wall Street Journal*, October 8: B1.

Manser, Marilyn E and McDonald, Richard, 1988. An Analysis of Substitution Bias in
Measuring Inflation, 1959-1985. *Econometrica* (July) 56: 909-30.

Moore, F. Lee, 1978. Index Mischief: Price versus Cost. *Electric Perspectives*, 5: 8-27.

National Bureau of Economic Research, 1961. *The Price Statistics of the Federal Govern-
ment*, General Series, no. 73.

Pieper, Paul E., 1990. The Measurement of Construction Prices: Retrospect and Prospect.
In Berndt and Triplett, ed. (1990): 239-68.

Reinsdorf, Marshall, 1991a. The Effect of Outlet Price Differentials on the U. S. Consumer
Price Index. Unnumbered working paper, Office of Economic Research, U. S.
Bureau of Labor Statistics.

_____, 1991b. The Effect of Price Dispersion on Cost of Living Indexes. Working
paper 213, Office of Research and Evaluation, U. S. Bureau of Labor Statistics,
April.

Schwartzman, David, 1969. The Growth of Sales Per Man-Hour in Retail Trade, 1929-1963.

In Fuchs, Victor, ed., *Production and Productivity in the Service Industries.* New York: Columbia University Press for National Bureau of Economic Research: 201-29.

Triplett, Jack E., 1975. The Measurement of Inflation: A Survey of Research on the Accuracy of Price Indexes. In Earl, Paul H., ed., *Analysis of Inflation.* Lexington, MA: Lexington: 19-82.

_____, 1983. Concepts of Quality in Input and Output Price Measures: A resolution of the User-value Resource-cost Debate. In Foss, Murray F., ed., *The U. S. National Income and Product Accounts: Selected Topics.* Studies in Income and Wealth, vol. 47. Chicago: University of Chicago Press for National Bureau of Economic Research: 269-311.

_____, 1987. Hedonic Functions and Hedonic Indexes. In Eatwell, J., Milgate, M., and Newman, P., eds., *The New Palgrave: A Dictionary of Economics.* London: Macmillan: 630-4.

_____, 1990. Hedonic Methods in Statistical Agency Environments: An Intellectual Biopsy. In Berndt and Triplett, ed. (1990): 207-33.
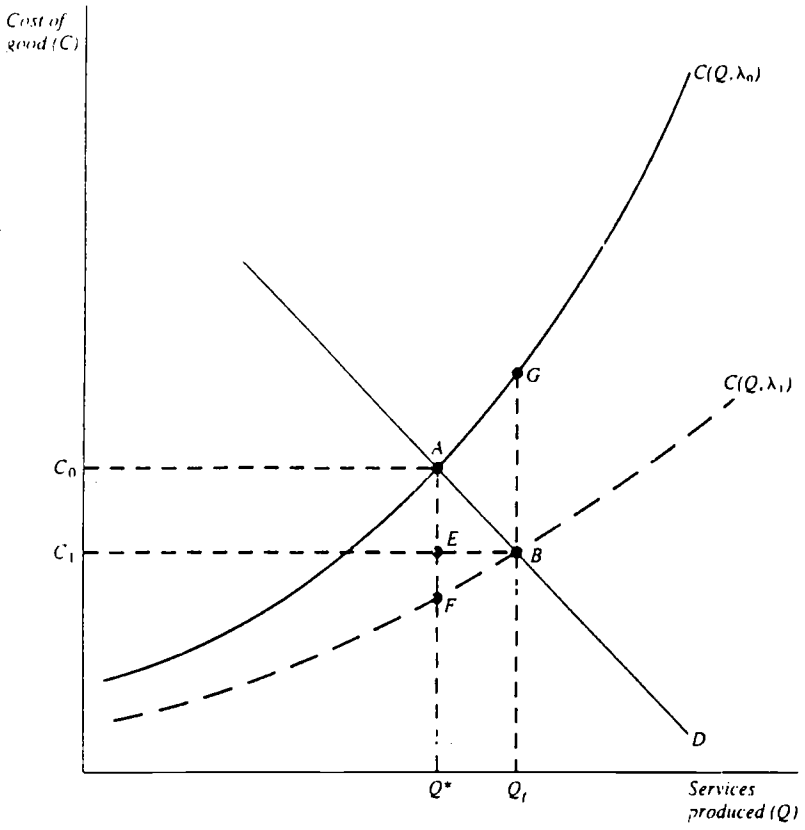
Figure 1

Table 1

Methods of Calculating Price Indexes (for a Hypothetical Economy)

		1980	
	1972	FIRST QUARTER	SECOND QUARTER
THE HYPOTHETICAL ECONOMY			
1. Prices			
Coffee per pound	$ 1	$ 4	$ 5
Whiskey per bottle	$ 5	$ 5	$ 5
2. Units Sold			
Pounds of Coffee	5	3	2
Bottles of Whiskey	1	2	3
3. Actual Expenditures			
Coffee	$ 5	$12	$10
Whiskey	$ 5	$10	$15
Total	$10	$22	$25
4. Real Expenditures in 1972 prices			
Coffee	$ 5	$ 3	$ 2
Whiskey	$ 5	$10	$15
Total	$10	$13	$17
THE EFFECTS ACCORDING TO THREE INDEXES			
5. Implicit PCE Deflator	100	169	147
(percentage change)		69.0%	-13.0%
6. Chain Index of 1980 change		—	11.7%
7. 1972 Fixed-weight Index	100	250	300
(percentage change)		150.0%	20.0%

Notes: The implicit PCE deflator in section 5 is 100 times the ratio of total actual expen-
ditures (section 3) to real expenditures (section 4).
 The Chain Index in section 6 multiplies the price change for the second quarter of
1980 for each item (25 percent for coffee, zero for whiskey) by the average expenditure
share of each product in both quarters of 1980 (22/47 and 25/47, respectively).
 The Fixed-weight Index in line 7 multiplies the level of the item index for each period
(100, 400, and 500 for coffee; 100 each period for whiskey) by that item's share in 1972
expenditures (50 percent for each product in this case).

Table 2
Classification of Quality Adjustments in
Selected Major PPI Commodity Groups

	Major Group and Two-Digit PPI Code			
	Metals and Metal Products 10 (1)	Machinery and Equipment 11 (2)	Furniture and Household Durables 12 (3)	Transportation Equipment 14 (4)
1. Tally as of July 1975:				
Number of items	373	799	94	104
Number of price quotations	1.225	2.893	431	242
2. 1976 quality adjustments as percent of all price quotations	0.4	0.5	1.3	2.4
3. 1976 percent share by quality adjustment method:				
Direct comparison	32.2	32.4	29.0	11.9
Link. no change	49.1	38.8	56.4	20.9
Cost estimate	18.7	28.8	14.5	67.1
	100.0	100.0	99.9	99.9

Table 3

Alternative and NIPA Deflators and Investment-to-GNP Ratios
for Durable Goods, 1947-83, Selected Years

Item	1947	1960	1973	1983
1. PDE deflator (1982 = 100)				
Alternative[a]	59.60	60.10	58.40	99.10
NIPA	20.60	35.70	47.30	99.50
Alternative-NIPA ratio	2.89	1.68	1.23	1.00
2. PDE/GNP (1982 dollars, percent)				
Alternative[a]	2.58	3.07	6.02	7.16
NIPA	6.96	5.00	7.27	7.14
Alternative-NIPA ratio	0.37	0.61	0.83	1.00
3. Consumer durable expenditures/GNP (1982 dollars, percent)				
Alternative[a]	3.27	4.66	7.38	8.64
NIPA	5.30	5.89	8.03	8.65
Alternative-NIPA ratio	0.62	0.79	0.92	1.00
4. Total durables/GNP (1982 dollars, percent)				
Alternative[a]	5.85	7.73	13.40	15.80
NIPA	12.26	10.89	15.30	15.79
Alternative-NIPA ratio	0.47	0.71	0.88	1.00

a. Alternative based on new detailed price data assembled by Gordon, using Törnqvist indexes that weignt the annual percentage change in components of real output in each subcategory between years t and $t + 1$ by the average of the nominal value weights in the two adjacent years.