Federal Reserve Bank of Dallas

presents

RESEARCH PAPER
No. 9340

A Survey of Measurement Biases in Price Indexes

by

Mark A. Wynne, Research Department
Federal Reserve Bank of Dallas

Fiona D. Sigalla, Research Department
Federal Reserve Bank of Dallas

October 1993

A survey of measurement biases in price indexes[1]

Mark A. Wynne and Fiona Sigalla

Research Department
Federal Reserve Bank of Dallas
2200 North Pearl Street
Dallas TX 75201

This draft: October 19, 1993

Preliminary
Comments welcome

[1]This paper was prepared for a joint meeting of the System Committees on Business and Financial Analysis held at the Board of Governors on September 22-24, 1993. The views expressed in this paper are those of the authors and should not be attributed to the Federal Reserve Bank of Dallas or the Federal Reserve System.

"Without the notion of price there would be no economic science. The concept is of absolutely central significance. It is not as easy and trivial a concept as it appears to be at first sight. A satisfactory measurement of price is, as a consequence, a difficult undertaking, and it is not surprising that price statistics, abundant as they are, have to be approached with utmost caution." (Oskar Morgenstern On The Accuracy of Economic Observations second edition 1963 p.194)

1. Introduction

This paper surveys the existing research on the accuracy of the various price indices that the Fed relies on to gauge the extent of inflation in the US economy. Specifically, we examine the usefulness of the Consumer Price Index (CPI), the Producer Price Index (PPI) and the GDP deflators as measures of inflation.

The most comprehensive existing survey is Triplett (1975), which was updated somewhat in Triplett (1988). Less comprehensive, though more recent, reviews of the issues involved in price level measurement are contained in Lebow, Stockton and Roberts (1992) and Gordon (1992).[2] We take the surveys by Triplett as our point of departure, reorganizing the issues in a way that better addresses the concerns of the Fed, and update them to include studies of the accuracy of price indexes that have been published in the past five years, especially the monumental study of durable goods prices by Gordon (1990).

The first major outside review of the quality of the various price measures published by the Federal government was the Stigler Committee Report (1961). The Stigler Committee advocated a number of changes in the way the

---

[2]Similar reviews from the Canadian perspective are Fortin (1990) and Crawford (1993). Crawford concludes that "the rate of change in the CPI probably overstates the rate of increase in the cost of living in Canada, but ...the bias is probably less than 0.5 percentage points a year."

BLS compiled and published it price statistics, some of which have been adopted and others of which have not. Some of the issues raised by the Stigler Committee are still relevant and require further research, such as how to deal with quality change, while others, such as publication of the methods used by the BLS in constructing its price indices have been dealt with (see U.S. Department of labor (1992)). One of the most important recommendations of the Stigler Committee was the establishment of a research division within the BLS to conduct ongoing research on how to improve the price indexes. This suggestion was also adopted, with the result that there have been a number of improvements over the past thirty years in how the CPI and PPI are constructed.

Backman and Gainsbrugh (1966) also reviewed the major price indexes produced by the Federal government. One of the more interesting of their conclusions was that "There is no solidly based evidence to support the conclusion that the net quality bias was large enough to offset fully, or even to a substantial extent, the steady rise in the CPI between 1958 and 1964." (Backman and Gainsbrugh, 1966, p.29). They also argued that the three major price indexes (the CPI, PPI and GNP deflator) provided an incomplete picture of inflation, a point we shall return to in more detail below, and concluded that despite its shortcomings, the CPI provided the best measure of inflation. A comprehensive review of the PPI was carried out by the Council on Wage and Price Stability (U.S. Executive Office of the President (1977)), which is also known as the Ruggles Report. The Ruggles Report drew attention to the failure of the BLS to collect transactions rather than list prices and the failure to properly deal with quality change as among the more serious shortcomings of the PPI.

Surveys of price measurement problems for particular classes of goods are Gordon (1971) on capital goods, Triplett (1989) on computer prices, and Gordon (1990) on durable goods. There are also a number of surveys by Fed economists of the various price indexes, including Wallace and Cullison (1981), Webb and Willemse (1989), Carlson (1989) and Brauer and Wu (1991).

This paper begins with a review of problems that are common to the measurement of all prices and price indexes, such as quality change and the measurement of list versus transactions prices. We then consider the construction of the three major price indices in greater detail. We start with the CPI, explaining how it is constructed (following Bureau of Labor Statistics (1992)). We then review what we know about the accuracy of the CPI, focusing primarily on problems that have been identified and quantified, and then discussing the problems that still remain and about which we know relatively little. We then consider the construction and shortcomings of the PPI, and pay particular attention to the recent study by Gordon (1990) that attempts to arrive at an overall assessment of the accuracy of the PPI. Since the GDP deflators are derived from the CPI and PPI rather than being independent measures of prices, our discussion of them focuses on the recent introduction of alternative chain weighted versions of these deflators, and the potential usefulness of these deflators in tracking inflation trends. The penultimate section of the paper considers issues related to the choice among the various price indexes in monitoring inflation trends.

2. Generic problems in the measurement of the price level

The problems that arise in the construction of measures of the price level can be divided into two broad categories. First, the problems arising from the measurement of the individual prices that make up the overall price index. The second set of problems has to do with how best to combine the prices of the various goods and services into a single measure of the price level. This latter problem is also known as the index number problem. We will review the main aspects of each of these issues before turning to a review of the major price indexes.

2.1 Measuring individual prices

> "[The] reliability of an index number obviously depends upon the judgement and accuracy with which the original price quotations were collected. This field work is not only fundamental, it is also laborious, expensive, and perplexing beyond any other part of the whole investigation. Only those who have tried to gather from the original sources quotations for many commodities over a long series of years appreciate the difficulties besetting the task.... To judge from the literature about index numbers, one would think that the difficult and important problems concern methods of weighing and averaging. But those who are practically concerned with the whole process of making an index number from start to finish rate this office work lightly in comparison with the field work of getting the original data." (Wesley Clair Mitchell, Index Numbers of Wholesale Prices in the U.S. and Foreign Countries, BLS Bulletin 173, 1915, quoted in Foss (1993), p.278.)

The most basic problem in the construction of a measure of the price level or average prices has to do with specifying and measuring the prices of the individual goods and services that will go into the aggregate index. Consider the seemingly simple task of measuring the price of steel. The following quote from Stigler and Kindahl (1970) illustrates the difficulty of measuring the price of what is considered by most people to be a relatively standardized, "low-tech" product:

"Consider hot rolled carbon steel sheets, which are arbitrarily defined

to be 0.23 inches or less in thickness if 12 to 48 inches wide, and 0.18 inches or less in thickness if over 48 inches wide. (Other dimensions are called bar, strip, or plate.) The buyer may choose among

1. Seven gauges of thickness
2. Ten classes of width
3. Four classes of length, if cut
4. Two classes of flatness
5. Two classes of squaring ends
6. Six lot sizes
7. Three classes of oil treatment
8. Ten classes of carbon maximum
9. Seven classes of manganese maximum
10. Five classes of sulfur maximum
11. Three classes of silicon maximum
12. Seven classes of packaging.

And a dozen other dimensions of product variety! Many of the 135 million varieties implied by the twelve attributes have never been produced, one may reasonably conjecture, but the varieties produced in one year must be immense. Each class has its own price. This is one product category in steel.

Tier after tier of further differences may be piled upon the physical varieties. There are terms of credit, transportation charges, guarantees of performance, facilities for replacement, techniques for arbitration of conflicts, promptness of delivery in normal times and in crisis. To be told that the base price of hot rolled carbon steel sheets was $5.30 per 100 pounds on May 15, 1963, is rather an

oversimplification of the price structure."(Stigler and Kindahl, 1970, pp.5)

Thus even for tangible goods subject to relatively few technical innovations or quality changes, it is far from being a straightforward matter to measure the price of the good. The problem is immediately complicated when the goods that are being priced are subject to rapid quality change due to technical progress. Obviously we do not want to count as price increases those increases in price that accompany the introduction of new models of existing products that are superior to the existing goods. How then do we adjust for quality changes? Quality bias is considered by many to be the most serious shortcoming of the official price series, and is reckoned to cause them to overstate inflation by as much as a couple of percentage points a year. As we shall see, the evidence does not support such claims, but the issue of how best to control for quality change is an open question. Hedonic techniques were once considered to be the best way to deal with this problem, but experience has shown that hedonic methods are of limited value in practice.

Then there is the problem of measuring price when we do not know what is being exchanged, as in the case of most services. A doubling of the fee for some service does not count as a price increase if the fee increase is accompanied by an increase in the amount of service rendered. But if we do not observe or cannot measure what is exchanged in return for the fee, we may not be able to determine whether a price change has in fact occurred. The rapid rates of price inflation recorded in the medical services category may be due to our inability to quantify the amount of service being rendered or due to our inability to correct for quality changes in the provision of health

care.

A further problem arises when we wish to price an imputed flow of services from a durable good such as a house or car. In this case we observe neither the price nor the quantity. At present this is only an issue for the measurement of homeowners rental equivalence costs in the CPI. Month to month movements in this component of the CPI are estimated using rents paid by renters of housing units that are similar to the typical owner-occupied unit. However, the absence of rental markets for most types of durable goods prevents the rental equivalence concept from being employed more widely in the CPI.

Given the impossibility of pricing every good and service exchanged in every transaction, it is necessary to rely on a sample of prices to infer trends in all prices. This raises the issue of how representative the sampled prices are of all prices. Is the sampling methodology biased in some ways towards older products or outlets, or towards products whose prices are more easily measured? The PPI has been criticized in the past for having a tendency to price older products and for being slow to include newer products. Computer prices were not introduced into the PPI until 1990. Recently the issue of outlet substitution bias has been raised in connection with the measurement of prices for the CPI. By oversampling prices at higher price outlets, Reinsdorf (1993) estimates that the CPI overstates inflation for some categories of goods by as much as 2 percent a year.

Are the sampled prices measuring the true prices at which transactions take place, or are they simply list prices that very few purchasers actually end up paying? Are all relevant discounts and rebates factored in to the price calculations? What sort of considerations would cause firms to report

list rather than transactions prices? The problem of accurate measurement of transactions prices was raised by Morgenstern (1963), who noted that price quotes obtained from buyers and sellers ought to be the same. Morgenstern (1931) noted how the structure of industry might hinder the collection of accurate information on transactions prices. More recently Foss (1993) has examined ways in which the regulatory environment may inhibit the reporting of transactions prices.

2.2 Combining individual prices into a single measure of the "price level"

The second broad group of problems that arise in constructing measures of the price level have to do with how best to aggregate the prices of individual commodities to obtain a single measure of the price level. There are two aspects of this problem - (i) what do we mean by the "price level"? and (ii) the index number problem. The first of these questions requires that we define exactly what it is that we are trying to measure. For example, the consumer price index is supposed to measure the cost of living, but as we shall argue below, the cost of living is not necessarily synonymous with the price level of monetary theory.[3] Likewise, the appropriate theory that forms the basis of the producer price index is the theory of output prices or input prices depending on the use to which the price series is to be put. The issue of how best to combine the individual prices is generally referred to as the index number problem, and the theory of the cost of living index is probably the best developed part of the economic theory of index numbers (see, for example, Allen (1975), Afriat (1977), Banerjee (1975), Pollak (1989) and also the surveys by Diewert (1987, 1988)).

[3]See also Fisher (1920) chapter X.

The nature of the index number problem is stated succinctly by Diewert (1987): Suppose we have price data $p^t \equiv (p_1^t,\ldots,p_N^t)$ and quantity data $x^t \equiv (x_1^t,\ldots,x_N^t)$ on N commodities that pertain to the same economic unit at time period t for $t = 1,\ldots,T$. The index number problem is to find T numbers P^t and T numbers X^t such that

$$P^tX^t = \sum_{i=1}^{N} p_i^t x_i^t$$

P^t is the price index for period i and X^t is the corresponding quantity index. The price index P^t is supposed to be in some sense representative of all of the prices p_i^t, $i = 1,\ldots,N$, while the quantity index is supposed to be in some sense representative of the quantities x_i^t, $i = 1,\ldots,N$. The different approaches to index number theory originate in trying to make the sense in which the index numbers are representative more precise.

The four most commonly encountered index number formulas in the contemporary literature on price measurement are those due to Laspeyres, Paasche, Fisher and Tornqvist. The Laspeyres index is defined as:

$$\frac{P^2}{P^1} = \frac{\sum_i p_i^2 x_i^1}{\sum_i p_i^1 x_i^1} \equiv P_L(p^1,p^2,x^1,x^2)$$

The Paasche index is defined as:

$$\frac{P^2}{P^1} = \frac{\sum_i p_i^2 x_i^2}{\sum_i p_i^1 x_i^2} \equiv P_P(p^1, p^2, x^1, x^2)$$

The Laspeyres and Paasche index number formulas are the two that are most frequently discussed in textbooks, for the simple reason that they are the most commonly used. As we shall see, both the CPI and the PPI are constructed using the Laspeyres formula, while the various implicit deflators in the national income and product accounts are (modified) Paasche index numbers.[4] The two formulas have competing merits. When used to construct measures of the change in the cost of living, the Laspeyres formula tends to overstate the increase in the cost of living while the Paasche formula tends to understate the increase.

The Laspeyres index formula has an important property that makes it preferable under certain circumstances to the Paasche formula. Consider two Laspeyres indexes where, for simplicity, the base and reference periods coincide:[5]

$$P_L(p^t, p^0, x^t, x^0) = \frac{\sum_i p_i^t x_i^0}{\sum_i p_i^0 x_i^0}$$

$$P_L(p^{t+1}, p^0, x^{t+1}, x^0) = \frac{\sum_i p_i^{t+1} x_i^0}{\sum_i p_i^0 x_i^0}$$

It is straightforward to show that the ratio of these two indexes

---

[4]See Kipnis (1966).

[5]This discussion draws on Triplett (1981)

$$\frac{P_L(p^{t+1},p^0,x^{t+1},x^0)}{P_L(p^t,p^0,x^t,x^0)} = \frac{\dfrac{\sum_i p_i^{t+1}x_i^0}{\sum_i p_i^0 x_i^0}}{\dfrac{\sum_i p_i^t x_i^0}{\sum_i p_i^0 x_i^0}} = \frac{\sum_i p_i^{t+1}x_i^0}{\sum_i p_i^t x_i^0}$$

is a Laspeyres index which shows the change in cost between periods t and t+1 of purchasing a basket of goods typifying period 0 consumption levels. This useful property of the Laspeyres index (that the ratio of two Laspeyres indexes with common weights is itself a Laspeyres index) partly accounts for the popularity of the Laspeyres formula.

By contrast, when we consider two Paasche indexes

$$P_P(p^t,p^0,x^t,x^0) = \frac{\sum_i p_i^t x_i^t}{\sum_i p_i^0 x_i^t}$$

$$P_P(p^{t+1},p^0,x^{t+1},x^0) = \frac{\sum_i p_i^{t+1} x_i^{t+1}}{\sum_i p_i^0 x_i^{t+1}}$$

and take the ratio of these two index numbers

$$\frac{P_P(p^0,p^{t+1},x^0,x^{t+1})}{P_P(p^0,p^t,x^0,x^t)} = \frac{\dfrac{\sum_i p_i^{t+1}x_i^{t+1}}{\sum_i p_i^0 x_i^{t+1}}}{\dfrac{\sum_i p_i^t x_i^t}{\sum_i p_i^0 x_i^t}}$$

we see that it does not reduce to any index number formula because the two index numbers have different weights. Thus the change in any index number calculated using the Paasche formula cannot be give an interpretation from the theory of index numbers.[6]

However, if we consider the expression for a Laspeyres chain-linked[7] index

$$P_L(p^t,p^{t+1},x^t,x^{t+1}) = \frac{\sum_i p_i^{t+1}x_i^t}{\sum_i p_i^t x_i^t}$$

which we can rearrange as

$$\sum_i p_i^t x_i^t = \frac{\sum_i p_i^{t+1}x_i^t}{P_L(p^t,p^{t+1},x^t,x^{t+1})}$$

Substituting this into the right hand side of the expression for the change in the Paasche indexes above we obtain

$$\frac{P_P(p^0,p^{t+1},x^0,x^{t+1})}{P_P(p^0,p^t,x^0,x^t)} = \frac{\dfrac{\sum_i p_i^{t+1}x_i^{t+1}}{\sum_i p_i^{t+1}x_i^t}}{\dfrac{\sum_i p_i^0 x_i^{t+1}}{\sum_i p_i^0 x_i^t}} \times P_L(p^t,p^{t+1},x^t,x^{t+1})$$

that is, the period to period change in the Paasche index is equal to the Laspeyres chain weighted index multiplied by the ratio of two <u>quantity</u> indexes.

---

[6]For further discussion of this point with reference to the various GDP deflators, see Los (1985).

[7]Chain linked indexes are explained in more detail below.

The other two index numbers that are commonly encountered in the index number literature are those due to Fisher and Tornqvist. The Fisher formula is defined:

$$\frac{P^2}{P^1} = \sqrt{P_L P_P} \equiv P_F(p^1, p^2, x^1, x^2)$$

i.e. the Fisher index is simply the square root of the product of the Laspeyres and Paasche indexes.

The Tornqvist index is defined as:

$$\frac{P^2}{P^1} = \prod_i \left(\frac{p_i^2}{p_i^1}\right)^{s_i} \equiv P_T(p^1, p^2, x^1, x^2)$$

where

$$s_i \equiv \left(\frac{1}{2}\right)\left[\frac{p_i^1 x_i^1}{\sum_i p_i^1 x_i^1} + \frac{p_i^2 x_i^2}{\sum_i p_i^2 x_i^2}\right]$$

The popularity of the Fisher and Tornqvist index number formulas comes from the fact that both belong to the class of "superlative" index numbers introduced by Diewert (1976). Superlative index numbers have the property that they provide a good approximation to the exact cost of living index. Diewert showed that superlative indexes such as the Fisher and Tornqvist have the desirable property of accommodating substitution in consumer spending while holding living standards or welfare constant.

As Triplett (1992) points out, the Fisher and Tornqvist formulas are equally good on theoretical grounds,[8] so that the choice between them comes down to practical considerations. For the purposes of introducing flexible

[8]See also Diewert (1987).

weighted price and quantity indexes into NIPA (discussed in more detail below), the dual property of the Fisher index, namely that a Fisher price index implies a Fisher quantity index and vice versa, caused it to be chosen over the Tornqvist index. The Tornqvist index does not possess this property: a Tornqvist price index multiplied by a Tornqvist quantity index does not equal the change in value between two periods.[9]

Superlative index numbers are not without their problems, however. The most important shortcoming of superlative index numbers is that they do not satisfy circularity, that is, $P(p^1,p^2,x^1,x^2) \times P(p^2,p^3,x^2,x^3) \neq P(p^1,p^3,x^1,x^3)$.

When it comes to constructing index numbers to measure price movements over more than two time periods, we have a choice between using either the fixed base year approach or the chain principle. Under the fixed base year approach, which is used to construct the CPI and the PPI, the index tracks the average price of the quantities consumed at some fixed point in time (the base period). The weight of each good in the index is held constant for long periods of time. As we get further away from the base period, the basket of good being priced, and the weights attached to them, may become increasingly unrepresentative of actual expenditure patterns, depending on how (relative) prices change over time. For fixed base year indexes, the choice of base year can also distort the index if the base year is unusual in some way. Thus, prior to the most recent revision of the PPI, the base year was 1982, which was a recession year, with the result that cyclically sensitive products were underweighted in the index, while cyclically insensitive products were overweighted.

---

[9]See also Diewert (1987).

The chain principle (originally due to Alfred Marshall) provides an alternative to the fixed base year approach to constructing index numbers over time. This principle makes use of the natural order provided by the march time to construct a time series of index numbers. We start by choosing some index formula P, setting the period 1 (initial period) price level equal to unity. The period 2 price level is set equal to $P(p^1,p^2,x^1,x^2)$, while the period 3 price level is set equal to $P(p^1,p^2,x^1,x^2) \times P(p^2,p^3,x^2,x^3)$. The period 4 price level is set equal to the period 3 price level times $P(p^3,p^4,x^3,x^4)$, and so on. Thus the period t price level is not obtained from direct comparison of period t prices with period 1 prices, $P(p^1,p^t,x^1,x^t)$, as in the fixed base period approach, but rather as the product of the period by period relative price levels. The major drawback of the chain principle is that if $p^t = p^s$ and $x^t = x^s$ and periods t and s are not adjacent, then it is not necessarily the case that the price level in period t will coincide with the price level in period s if the bilateral formula $P(p^1,p^2,x^1,x^2)$ does not have the circularity property.[10] However, experience has shown that deviations from circularity for superlative index numbers such as the Fisher and the Tornqvist are small in the time series context.

Before concluding this introductory review of the issues involved in price measurement, we might note that a number of recent papers (Clements and Izan (1987), Bryan and Cecchetti (1993a,b) and Dow (1993)) have advocated alternative approaches to extracting information about the price level or

[10]Diewert (1987) notes that the Laspeyres, Paasche, Fisher and Tornqvist index number formulas do not have the circularity property.

inflation rate from observations on individual prices.[11] Essentially, all four papers advocate a return to the statistical approach to index number construction as opposed to the currently more popular behavioral approach (see Diewert (1987)). Clements and Izan explicitly relate their approach to the older work on index numbers by Jevons and Edgeworth. Their approach is to measure inflation as an average of the price changes of individual commodities within a statistical model that also allows for changes in relative prices. Bryan and Cecchetti (1993a) use the weighted median of the rates of increase of individual components of the CPI to arrive at a measure of "core" inflation.[12] Dow (1993) constructs a measure of core inflation as a weighted average of individual prices, where the weights are inversely related to variance of the individual price series. This seems a more desirable approach than the current practice of simply excluding the two most volatile categories (i.e. food and energy) to arrive at a measure of core inflation. Dow also estimates a common factor model of the inflation rate, as do Bryan and Cecchetti (1993b). We will argue below that these approaches to inflation measurement deserve further exploration, although Diewert (1987) expresses skepticism about the usefulness of neostatistical approaches to index number construction, noting that such methods tend to be arbitrary and lack economic interpretation.

---

[11]The papers by Lichtenberg and Griliches (1989) and Siegel (1991) that are considered in more detail below could also be included in this group.

[12]We might note that the weighted median was also suggested by Fisher (1920, 1922) as one of the best available measures of the price level, who cites with approval Wesley Mitchell's use of the median as a measure of price change.

3. The Consumer Price Index

The purpose of the Consumer Price Index is to measure the average change in the prices paid by urban consumers for a fixed market basket of goods and services of constant quality. The CPI is calculated for two population groups, namely wage earners and clerical workers (CPI-W) and all urban consumers (CPI-U). The former is representative of the buying habits of about 32 percent of the noninstitutional population of the US, while the latter (which was introduced in 1978) covers about 80 percent of the population (BLS, 1992, p.176). Since it is obviously impossible to obtain price data for all consumer transactions in the US, the CPI is estimated from a series of samples. These samples are designed with the intention of making the CPI representative of the prices paid for all goods and services purchased by consumers in all US urban areas. The use of a sample, of course, introduces a source of error into the index, but this may be more than offset by the errors that sampling eliminates. New item and outlet samples are selected each year for 20 percent of the primary sampling units on a rotating basis, the intention being that such rotation should enable the CPI to reflect new developments in the market for consumer goods.

Individual commodities in the CPI are weighted by the share of expenditure on the item as estimated in the Consumer Expenditure Survey. Currently, the CPI is constructed using expenditure shares obtained from the 1982-84 Consumer Expenditure Surveys, and is formally defined as:

$$CPI^t = 100 \times \frac{\sum_i p_i^t x_i^b}{\sum_i p_i^r x_i^b} = 100 \times \sum_i \left(\frac{p_i^t}{p_i^r}\right) \frac{p_i^r x_i^b}{\sum_i p_i^r x_i^b} = 100 \times \sum_i \left(\frac{p_i^t}{p_i^r}\right) \omega_i^b$$

where p_i^t is the price of the i'th good in the comparison period (t), p_i^r is

the price of the same good in the reference period (r), and x_i^b is the quantity of the good consumed in the expenditure base period, (b). When the expenditure base and the reference periods coincide, this of course becomes the standard Laspeyres price index formula. Note that, in general, there is a difference between the base period for the expenditure weights, and the numeric reference base period, although at present both are 1982-84=100.

A concept that is related to the weight of an item in the CPI is the relative importance of a commodity, which shows the share of total expenditure that would be accounted for by a particular commodity if quantities consumed were unaffected by changes in relative prices. The relative importance of a commodity as of a particular date is given by

$$\omega_i^0 \frac{p_i^{t-1}}{CPI^{t-1}}$$

The relative importance of an item in the CPI grows over time if it experiences greater than average price increase; conversely the relative importance of an item declines if it experiences slower than average price increase.

The importance of the distinction between the relative importance of an item in the CPI and its weight is often overlooked, and a number of critics have suggested that the BLS publish the weights used in the construction of the CPI as well as the relative importances. Triplett (1988) suggests that the BLS should publish the weights as well as the relative importances, as in many cases it is the weights that users of the CPI need and the fact that they

are not available leads some researchers astray.[13]

The prices that are used to calculate the CPI are collected from about 21,000 retail and service establishments in 85 urban areas across the US. Data on rents are collected from about 40,000 landlords or tenants, and 20,000 owner-occupants are asked about their housing units. (BLS, 1992, pp. 178) All price information is collected by BLS field agents through visits or telephone calls.[14]

The problems that are typically thought to bias the CPI as a measure of inflation or the cost of living are substitution bias, which has recently been extended to include the concept of outlet substitution bias, quality adjustment bias, and new goods bias, although the distinction between quality bias and new goods bias is far from clear. The problems that arise from the use of list rather than transactions prices in constructing an index are less frequently discussed in relation to the CPI than they are in relation to the PPI. We will argue below that the transactions-list problem is probably an issue in consumer price measurement, although for different reasons than in producer price measurement.

We now consider the various biases that are known or conjectured to exist in the CPI and review the existing literature on these problems.

---

[13]See for example the recent paper by Bryan and Cecchetti (1993). Bryan and Cecchetti construct a measure of core inflation as a weighted median of 36 components of the CPI using the relative importance of the components as of December 1991 as weights rather than 1985 CPI weights that they claim to be using. See their Table 3.

[14]For a description of the work of a BLS field agent, see Madigan (1991).

3.1 Substitution bias

Substitution bias is probably the best known and most frequently studied problem in the CPI. Triplett concluded his 1975 survey of measurement bias by noting that "Estimates of substitution bias that have so far been made indicate that it is extremely small, so small that substitution bias cannot be viewed as an important empirical defect of fixed-weight consumption price indexes."(Triplett, 1975, p.66) Substitution bias is inherent in any fixed weight estimate of the cost of living, and arises from the fact that households typically alter the bundle of goods they purchase in response to relative price changes. Essentially substitution bias is the difference between the CPI and the true cost of living index. The true cost of living index is defined using household cost or expenditure functions. The expenditure function for a household is defined as the minimum expenditure needed to attain a given level of welfare (or standard of living) and is defined by

$$e(p,u) = \min_{x_i}\{ \sum_i p_i x_i \ : \ u(x) \leq u\}$$

where $p = (p_1,\ldots,p_N)$. This in turn forms the basis for a natural definition of the change in the cost of living as the minimum expenditure (or cost) needed to attain some given base level of utility u^b in year t relative to the minimum expenditure needed to attain the same level of utility in some reference period r:

$$COL(p^t,p^r,u^b) = \frac{e(p^t,u^b)}{e(p^r,u^b)}$$

where $p^t = (p_1^t,\ldots,p_N^t)$ and $p^r = (p_1^r,\ldots,p_N^r)$.

The size of the substitution bias depends on the extent to which households substitute between goods in response to relative price changes and the extent to which relative prices change over time. Absent either of these factors, a fixed-weight Laspeyres index will give an unbiased estimate of the true cost of living index. Thus if household preferences are of the fixed-coefficient or Leontief type, no substitution occurs in response to relative price changes and the fixed-weight Laspeyres index is equal to the true cost of living index. Likewise, if all prices increase or decrease together, relative prices never change and again the fixed-weight Laspeyres index coincides with the true cost of living index. It is unlikely, however that either of these conditions hold in practice.

At the time of Triplett's 1975 survey, the principal studies of substitution bias in the CPI were Noe and von Furstenburg (1972), Christensen and Manser (1976) and Braithwait (1980) (which circulated as a BLS working paper in 1975). The major conclusions of these studies were that the size of the substitution bias was small, probably no more than 0.1 percent a year, that the estimated magnitude of the bias was relatively insensitive to the choice of functional form for household preferences, that the estimated magnitude of the bias was sensitive to the level of commodity disaggregation, and that the size of the bias was greater during periods of high inflation when relative price fluctuations were greater.

Since Triplett's survey, the most thorough (published) study of the substitution bias problem in Laspeyres type indexes is that of Manser and McDonald (1988) which used NIPA consumption data for 101 commodities over the period 1959-85. (Analysis was performed at the most disaggregated level for which data is available, which is 149 commodities. However, for some of the

commodities price data were not available, and the categories were dropped) Manser and McDonald present two sets of estimates of the substitution bias. First, using nonparametric methods and maintaining the assumption of homothetic preferences (which they test and are unable to reject) they calculated bounds on the size of the substitution bias in the Laspeyres index for the period 1959-85. They calculate a maximum possible bias of 0.22 percent per year, and a minimum possible bias of 0.14 percent per year. Second, using superlative index numbers to calculate changes in the cost of living, they estimate substitution bias of about 0.18 percent per year over the same period.

The findings of these studies of substitution bias in the CPI are summarized in Table 1.[15] Table 1 also reports the estimates of the "outlet substitution bias" obtained by Reinsdorf (1993), since bias in the CPI as a result of consumer substitution towards lower price retail outlets is in some sense analogous to the standard substitution bias in a fixed weight index as a result of commodity substitution. The idea that the CPI might overstate inflation because of its failure to allow for changes in consumers shopping patterns was first raised by Denison (1962).[16] Denison's concern was that the shift towards low-cost high-volume retail outlets yielded substantial benefits to consumers that were not captured by the CPI. The source of the problem had to do with the manner in which the BLS incorporates new stores

---

[15] We do not review the results of two recent studies by Kokoski (1987) and Blisard and Blaylock (1991) that examine the relationship between demographic factors and substitution bias.

[16]The issue of outlet substitution was also discussed by the Stigler Committee (NBER, 1961, p.58), and in more detail, by Backman and Gainsbrugh (1966, pp. 29-31). Backman and Gainsbrugh cite studies by the BLS and Willard Arant that put the magnitude of this bias at no more than a couple of tenths of a percentage point a year.

into CPI outlet samples, with the result that the CPI probably overstated the increase in the cost of living. Reinsdorf tries to quantify the extent to which shifting patterns of retailer patronage by consumers lowers the average prices they pay and their cost of living. His empirical analysis focuses on the food-at-home and gasoline components of the CPI. Persistent price dispersion in the retail market is essential for consumer gains from switching outlets not to be reflected in the CPI. Such persistence may arise due to market disequilibria or information costs. An index that tracks prices at incumbents is biased if quality-adjusted prices at incumbent retailers fail to decline to match those of the new retailer.

Reinsdorf offers two types of evidence on the size of the outlet substitution bias in the CPI. First, he compares the prices at incoming and outgoing CPI outlet samples, and second, he compares the evolution over time of unlinked sample average prices and their (linked) CPI component counterparts. Comparing prices at new and old outlets, Reinsdorf obtains an estimate of upward bias in the food at home component of the CPI of 0.25 percent a year, and a comparable figure for the motor fuel component. Reinsdorf notes that this figure may overstate the true size of the bias if average quality declines along with average prices. Reinsdorf's second test compares the growth of the Average Price series published by the BLS with that of the corresponding components of the CPI. The Average Price series for an item tracks the price paid on average for a representative variety of the good, and is defined as follows (see BLS, 1992, p.199):

$$\bar{p}^t = \frac{\sum_i \frac{p_i^t}{p_i^0}\omega_i^t}{\sum_i \frac{1}{p_i^0}\omega_i^t}$$

where the ω_i^t denote expenditure weights. That is, the average price is essentially a weighted average price, where the weights are amounts purchased. For food during the 1980's the CPI rose more rapidly than the Average Price series, implying an outlet substitution bias of 2.0 percent a year. While quality adjusting the Average Price food indexes might reduce some of the difference with the CPI, it would not eliminate it completely since the willingness of consumers to shift to low cost retailers presumably indicates that the quality difference must be more than compensated for by the better prices. For gasoline the estimated bias is 0.3 percent per year. Reinsdorf notes that the estimates obtained from a comparison of Average Prices with the CPI should be taken as upper bounds on the amount of substitution bias, since no attempt is made to control for the possibility that the average quality of outlets may have declined.

In view of the importance that some analysts have attached to Reinsdorf's findings (see, for example, Gordon (1992)) it is important to be explicit about the caveats that accompany his results.[17] One is that outlet substitution and commodity substitution typically occur at the same time. To the extent that this allows the substitution in the CPI sample of say, a (low-cost) store-brand item for a name brand item, some of difference found by Reinsdorf may be simply due to switching brands. Also, it is important to try to quantify the quality difference between different retail outlets to get a

[17]See, for example, Popkin (1993).

proper handle on the size of the bias. Popkin suggests the use of hedonic regressions of the sort used to make adjustments in the apparel indexes to determine the appropriate quality adjustments for outlets.

Table 1 also reports estimates of substitution bias from Bryan and Cecchetti (1993b). Bryan and Cecchetti adopt a novel approach to estimating the bias that arises from the use of fixed weights in the CPI. The essence of their approach is to estimate the common (inflation) factor in a number of the sub-components of the CPI, and interpret the difference between the estimated common factor and the CPI as an estimate of the substitution bias in the CPI. The estimates they obtain averages 0.7 percent over their full sample, which ranges from 1967:1 to 1992:12. They also estimate biases for two sub samples, finding that the CPI overstates inflation by about 1.2 percentage points annually during 1967-82, whereas it understates inflation by about 0.1 percentage point during the 1983-92 period.[18]

Note that in concluding his survey of findings on substitution bias in the CPI, Triplett (1988) argued that the fact that there is very little difference between Laspeyres and Paasche CPI's is further reason to believe that the substitution bias in the CPI is probably quite small. Diewert (1988) discounts this argument, however, on the grounds that the indexes that Triplett has in mind are not "true" Laspeyres and Paasche indexes. That is, these indexes do not include every commodity, and invariably use aggregates or composites of goods which may disguise genuine substitution.

The bottom line on substitution bias is that this particular form of

---

[18]Bryan and Cecchetti rationalize the possibility of a negative or downward bias in the CPI by pointing out that for some time periods and some goods expenditure shares and price changes might be positively correlated if there is a taste shock that raises demand for the good.

bias in the CPI is probably relatively unimportant quantitatively. This conclusion is shared by Triplett (1975, 1988) and Gordon (1992), among others. However, it remains to be seen how serious the problem of outlet substitution bias identified by Reinsdorf is in the long run. The evolution of the retail industry means that the BLS needs to pay attention to the emergence of new, lower cost outlets for consumer goods. It also remains to be seen how important new sales outlets such as direct shopping from home via cable TV or computer is for the average consumer.

Insofar as substitution bias is considered a problem in the CPI, it could be handled in one of two ways. One is to update the expenditure weights more frequently: currently the expenditure weights are updated about once a decade, the most recent revision occurring in 1987, when expenditure patterns from the 1982-84 Consumer Expenditure Survey replaced those from the 1972-73 survey. An alternative is to construct an index with multiple expenditure weights, or multiweighted index (such as fixed base and chain linked Tornqvist and Fisher indexes). Both of these possibilities are now feasible since the Consumer Expenditure Survey has been conducted on an ongoing basis since 1980. Kokoski (1989) describes research currently underway at the BLS in using multiweighted indexes to keep track of price movements without problems of substitution bias.

3.2 Quality bias

> "If a poll were taken of professional economists and statisticians, in all probability they would designate (and by a wide majority) the failure of the price indexes to take full account of quality changes as the most important defect in these indexes. And by almost as large a majority, they would believe that this failure introduces a systematic upward bias in the price indexes - that quality changes have on average been quality improvements." (NBER, 1961, p.35 [emphasis added])

As we have already noted there is a general perception that quality bias is probably the most serious shortcoming of the CPI. This probably reflects two other beliefs, first, that the average quality of all goods is improving over time and, second, that the BLS does little or nothing to take this quality improvement into account when calculating its price indexes. As we will see, neither of these statements is completely true. While it is the case that the quality of many products does seem to improve over time, we should not overlook the obvious examples where quality seems to be deteriorating - for example the increased use of graduate students for undergraduate instruction at major universities, the disappearance of full service gas stations, and the decline in the quality of in-flight service on some airlines.[19] Second, the BLS has a number of methods for dealing with quality change (which we will review shortly). The question then becomes how adequately do these methods capture changes in the quality of the products sampled for the price indexes - we will see that in some cases the BLS may in fact over-adjust for quality change in calculating the price indexes, thereby imparting a downward bias to the reported figures.

3.2.1 How the BLS deals with quality change

The quality adjustment problem in constructing a price index may be

---

[19]On the latter, see for example, James S. Hirsch "With Fewer Attendants Aboard Jets, Mood of Passengers Turns Turbulent" Wall Street Journal, Friday July 23, 1993, pages B1-B3, who notes "Seeking to stanch their financial hemorrhaging, most airlines are putting fewer attendants on board their aircraft. The result: Passengers wait longer for meals and beverages; meal carts clog the aisles longer; dirty trays stack up; and obtaining the little extras of life aloft - a pillow, a magazine, a drink of water - is often a do-it-yourself experience."

stated as follows.[20] Suppose some particular variety ν_0 of a good $x_i(\nu_0)$ is selected for inclusion in a price index. Suppose further that at some date t the chosen variety disappears and is replaced by a newer model $x_i(\nu_1)$. How do we compare the price of the old variety of the good in period t-1, $p_i^{t-1}(\nu_0)$ with the price of the new variety in period t, $p_i^t(\nu_1)$? The BLS categorizes the methods it uses to deal with new product varieties or quality change as follows: (1) direct comparison (2) direct quality adjustment and (3) imputation.

In the first case, direct comparison, if the varieties ν_0 and ν_1 are judged to be sufficiently close in some well defined sense, all of the observed difference in price between the new and old varieties of a product is counted as a price change, and nothing more is done. The risk here of course is that some unnoticed quality change is inadvertently being treated as a price change.

If however the varieties are judged to be different in some meaningful sense, the BLS makes some form of direct quality adjustment using one of a number of different methods. The simplest case is when the two varieties are observed in some common period, in which case the ratio of the two prices in the period of overlap is taken as the quality adjustment. That is, $A_i \equiv p_i^\tau(\nu_1)/p_i^\tau(\nu_0)$ where τ is the period of overlap. The problem with this adjustment procedure is that we rarely observe overlapping prices for new and old varieties. Furthermore, the theory of the cost of living index implies that quality adjustments of this type are biased, although by only a small

[20]This discussion follows Triplett (1988). See also Armknecht and Weyback (1989), Armknecht and Ginsburg (1993), BLS (1992) and Gordon (1992).

amount.

An alternative is to make an adjustment based on the manufacturers production cost differences for the two varieties. In this case the manufacturer is asked to estimate the cost difference for $x_i^{t-1}(\nu_0)$ and $x_i^t(\nu_1)$, which is then scaled up to the retail level and added to $p_i^{t-1}(\nu_0)$ to obtain an estimated price for $x_i(\nu_1)$ in the previous period, $\hat{p}_i^{t-1}(\nu_1)$. The quality adjustment is then $A_i \equiv \hat{p}_i^{t-1}(\nu_1)/p_i^{t-1}(\nu_0)$, and the price relative that goes into the calculation of the overall index is $p_i^t(\nu_1)/\hat{p}_i^{t-1}(\nu_1)$. This form of quality adjustment is used most frequently in valuing quality changes in automobiles, and has been in use since around 1960.[21] The problem with using cost as a measure of quality is that if productivity is growing cost based quality adjustments will lead to overestimation of price change.

Another form of quality adjustment, and the one that has formed the basis of most attempt to estimate the size of the quality bias in the CPI, entails estimating a hedonic regression that relates the price of different varieties of a good to the characteristics of the different varieties:[22]

---

[21]Gordon (1981) notes that a problem with this procedure is that a manufacturer may overstate the cost of a quality improvement in order to disguise some part of an actual price increase, especially in periods where price controls or guidelines are in force. Triplett (1990) notes that at least since the early 1970's BLS staff have frequently not fully allowed manufacturer's cost claims.

[22]Hedonic methods are reviewed in Triplett (1986, 1987) and Griliches (1971). Griliches (1961) is the seminal application of hedonic methods to evaluating the quality bias in the CPI.

$$p_i(\nu_j) = \alpha_0 + \sum_{k=1}^{K} \alpha_k z_k(\nu_j)$$

The estimated parameters from such a regression $(\alpha_1,\ldots,\alpha_K)$ provide implicit prices for each of the price determining characteristics of the good. Thus, when a new variety of product i becomes available, ν_1, which differs from the existing variety, ν_0, in terms of some or all of the relevant characteristics, it is straightforward to make a quality adjustment on the basis of the hedonic regression which is $\sum_{k=1}^{K} \alpha_k(z_k(\nu_1) - z_k(\nu_0))$.

Recently the BLS has begun using hedonic indexes to adjust for quality change in apparel. The BLS is currently exploring ways to use hedonic methods to improve quality adjustments in a number of service categories (see Armknecht and Ginsburg (1992)). Despite the initial promise of hedonic techniques, however, a number of commentators have recently suggested that they are of limited usefulness in dealing with the problem of quality change (see, for example, Triplett (1988) and Gordon (1990)). Specifically, hedonic techniques are not able to deal with quality changes that are not easily quantified (such as the handling characteristics of a car, or the multitasking ability or portability of a personal computer, or the quality of care during a hospital stay, or whether an item of clothing is in or out of fashion.)

However, in some situations the BLS has not yet determined how best to make quality adjustments. New product varieties that can be neither directly compared nor quality adjusted are called noncomparable, and in these situations the BLS estimates the constant quality price change by imputation. Two type of imputation are used to make adjustments. The first type of

imputation is commonly used for noncomparable substitutions in the food and services categories and consists of setting the rate of price change for the new and old varieties equal to the average price change for similar goods. This yields an implicit quality adjustment

$$A_1 \equiv \frac{p_1^t(\nu_1)}{p_1^{t-1}(\nu_0)} - \sum_{j \neq 1} \omega_j \frac{p_j^t(\nu)}{p_j^{t-1}(\nu)}$$

The implicit assumption that treating noncomparable substitutions in this manner is benign seems to be incorrect as we shall see below. For example, it is quite likely that the new product is in the early stage of its product cycles and experiencing substantial price declines, while the products used to impute the price change are probably mature products that are experiencing price increases.

The second type of imputation is designed to handle quality change in situations where price change is primarily associated with the introduction of new lines or models, as in the case of cars. For cars, price changes of models within the same model year are not the best estimate of the price change for noncomparable substitutions across model years. A better estimate is obtained by looking at the average price change between model year changeovers. Since 1989 the BLS has relied on constant quality price changes for comparable model changeovers to impute the price change for noncomparable new model vehicles, and intends to use this type of imputation more extensively for other noncomparable substitutions in the future (BLS, 1992, p.193).

The inherent difficulty of deciding whether new products, or new varieties of existing products, are comparable to old is the essential source

of quality error in the CPI. When a new product is deemed to be comparable with an old product, some quality change may be incorrectly treated as price change, leading to an upward bias in the CPI. When the new product is deemed noncomparable, some price change may be incorrectly treated as quality change. leading to a downward bias in the CPI.[23]

3.2.2 How big is the quality bias in the CPI?

Since the Stigler Committee Report in 1961 a number of studies have been carried out that have attempted to estimate the extent of quality bias in the CPI. It is not a simple matter to use the results of these studies to infer the extent of quality bias in the CPI as it is currently constructed because the methods used by the BLS have evolved over time, partly in response to these studies. For example, Griliches (1961) seminal study of hedonic price indexes was motivated by the failure of the BLS at that time to properly account for quality change in new cars, a problem that has since, to some extent, been corrected.

We need to be careful when interpreting estimates of the quality bias in an official price index. How comparable are the price data used in the study to those used in the official index?[24] Do they come from the same sources as the data used in the official index? Are they list or transactions prices?[25] Is the index formula the same as that used to construct the official index to

---

[23]See Armknecht (1984) p.58.

[24]Gordon (1990) makes a virtue of using non-standard data sources.

[25]For example Triplett (1975) points out that the different conclusions reached by Griliches (1961) and Cagan (1966) concerning quality bias in the auto component of the CPI were mainly the result of Cagan's use of list prices and Griliches use of published CPI prices.

which the new index is being compared? Triplett has consistently argued that we need to be careful when interpreting differences between official and alternative price indexes, and concluded his (1971) review of studies of quality bias with the observation that "...just because an economist produces an index that differs from an official index, this does not necessarily imply that it is the official index that must be the incorrect one."(Triplett, 1971, p.212)

Triplett (1975) concluded his survey of quality bias in the CPI by noting that "...the research results imply that no simple estimate of the overall quality error can be made, and, moreover, even the sign of the error is in doubt."(Triplett (1975), p.48) In concluding his 1988 survey, Triplett was even more emphatic about the uncertainty surrounding the sign and magnitude of the quality bias in the CPI, noting that "...because a number of large CPI components appear quite clearly downward biased, I suspect that the CPI has, if anything, understated inflation in the last several years." (Triplett, 1988, p.67). Among those categories of the CPI that Triplett singled out as being clearly downward biased were housing, clothing, new cars, and restaurant meals, with downward bias in these categories possibly being offset by upward bias in services, used cars, and certain other components such as airline fares. Let us review the evidence cited by Triplett to support these positions and see if BLS practices have changed to render them less valid. Table 2 summarizes the main studies of quality bias in the CPI that have appeared in the past 5 years.

On the issue of downward bias in housing Triplett cited the study by Randolph (1988) which showed that failure to allow for depreciation in the housing components of the CPI caused the shelter cost indexes to be downward

biased by as much as 0.3 to 0.4 percent annually. However, since 1988 the BLS has made adjustment to the rental components of the CPI to allow for aging - these adjustments are described in more detail in Lane, Randolph and Berenson (1988).

On clothing, Triplett cites the almost impossible task of separating taste or fashion changes from quality changes and the strong seasonal pattern in clothing styles as sources of error in the CPI treatment of clothing. One piece of evidence he offers comes from a comparison of the rates of inflation for infants and toddlers apparel, which presumably is less subject to fashion cycles, with those for men's and boys apparel and women's and girls apparel. In the 1967-87 period the index for infants and toddlers apparel grew at a 6.0 percent average annual rate, while those for men's and boys apparel and women's and girls apparel grew at 3.4 percent and 2.9 percent respectively, suggesting that the BLS may have over-adjusted for quality change in these categories.

Further evidence on the likely downward bias in the apparel components of the CPI was provided by Armknecht and Weyback (1989).[26] Armknecht and Weyback note the high rate of product substitution in the apparel category of the CPI: over 17 percent of the prices collected for this category in 1983 and 1984 reflected product substitutions. For the CPI as a whole, the number was under 4 percent. The breakdown of the substitutions as between comparable and noncomparable for apparel was about 7.0 percent comparable and about 10.0 percent noncomparable. Note that the former type of substitution raises the possibility of some quality change being inadvertently counted as price change, producing an upward bias in the CPI, whereas for the latter the risk

[26]See also Armknecht (1984).

is that some price change is inadvertently attributed to quality change, imparting a downward bias to the CPI. For the CPI as a whole, slightly more than 1.5 percent of the product substitutions are comparable, while just under 2.0 percent are noncomparable.

Armknecht and Weyback report results from hedonic regressions for women's coats and jackets, and women's suits and compare the implied quality adjusted price series with the official CPI-U series for each of these categories. For women's coats and jackets they find that the CPI rises 3.5 percent more over the July 1987 to June 1988 period than the adjusted series, indicating a potential upward bias in this series. For women's suits, they reached the opposite conclusion: the official CPI series for this category rose 3.4 percent less than the quality adjusted series, suggesting a potential downward bias in this series. However, they conclude by emphasizing the inconclusive nature of their results.

Leigey (1993) also considered the problem of quality adjustment in the apparel indexes. He points out that recent research using hedonic regressions has enhanced the quality of the raw data being collected for the apparel components of the CPI by facilitating the identification of comparable quality substitutes for discontinued items and improving the decision rules used to determine whether an item is a comparable or noncomparable substitute for a discontinued commodity. Leigey reports results from a study of women's coats and jackets, and women's suits (the same categories studied by Armknecht and Weyback). The choice of these categories was motivated by the fact that the price indexes for these categories indicated minimal and even declining price changes over long periods of time, and by the presence of more noncomparable substitutions in these categories than in other apparel categories.

(Armknecht and Weyback point out that the women's suits collection document is the most complex in terms of the number and correlation of specification elements (Armknecht and Weyback, 1989, p.115)).Liegey finds both upward and downward bias in the apparel indexes he studies. Liegey concludes his article noting the increased use of hedonic methods at the BLS to get around the biases introduced by inappropriate quality adjustment.[27]

Returning to the other categories of goods that Triplett identified as being possibly downward biased, the downward bias in auto prices comes from the BLS treatment of mandatory safety and pollution devices as quality improvements rather than as price increases, as the theory of the cost of living index indicates they should be treated. Triplett (1993a) argues that by following BLS convention on this issue, Gordon (1990) overestimates the size of the quality bias in durable goods prices. As for restaurant meals, the only evidence Triplett offers is anecdotal, and further Triplett notes that the growth in fast food, which is highly standardized, means that the bulk of the food-away-from-home component of the CPI is probably reasonably accurate.

The candidates for upward bias that Triplett identifies are services, used cars and miscellaneous other. The problems with measuring the prices of services accurately, and especially the price of medical care, are well known, but Triplett concludes that "existing research on the subject is insufficient to indicate whether the medical care components are upward biased." (Triplett, 1988, p.70) We will return to the measurement of service prices in more detail below.

---

[27]See Triplett (1988), part of which was published as Triplett (1992), for a discussion of why it has taken so long for the BLS and other statistical agencies to adopt hedonic methods.

For used cars, the problems stem from the inadequate quality adjustments made to price observations for used cars, although since xxxx the BLS has been making adjustments based on the quality adjustments used for new cars. As for the final miscellaneous other category, Triplett cites the difficulty of dealing with the subtle substitutions between restricted and unrestricted airline fares as a source of possible upward bias in the CPI.

However, probably the single most important piece of research on the problem of price measurement recently is Gordon's (1990) study of durable goods prices. While the primary objective of this study was to obtain improved estimates of the prices of producer durables, Gordon also looked at selected categories of consumer durables.[28] Gordon concluded that the CPI overstated the rate of increase in durables prices by an average of 1.54 percent a year over the full 1947-83 sample, with the largest errors occurring prior to 1960. Table 3 summarizes the details of Gordon's findings on the biases in the official estimates of consumer durables prices. Gordon notes that while the rate of drift or error for the appliance and radio-TV categories of durables may appear surprisingly high, the data used to construct these indexes are among the most accurate and comprehensive parts of his study.

It is important to note that Gordon explicitly cautions against inferring that all of the official price indexes are biased upward on the basis of the findings of his study:

> "No claim is made here that there is a consistent upward bias in the full range of government price indexes. In particular, there is no

[28]Indeed, it is arguable that, given the data sources Gordon relied upon (the Sears Catalog, Consumer Reports), his results are more relevant for assessing potential problems in the CPI that the PPI.

necessary conflict between the new results and the longstanding claim by Triplett (1975, 1988) that the overall bias in official price indexes is just as likely to downward as upward. Triplett's claim has always been explicitly with reference to the CPI as a whole and does not conflict with the new results showing a major upward bias for durable goods in the PPI, and to a lesser extent for durable goods in the CPI. We cannot assess the direction of bias in the full CPI, and hence we cannot assess Triplett's claim, since we have no new data for the prices of consumer nondurables or services." (Gordon, 1990, p.6)

3.3 New goods bias

In some respects the new goods problem is simply another version of the quality adjustment problem: the distinction between a new variety of an existing product and an entirely new product is not always obvious. For example, a personal computer can be considered to be a new product, or it can be viewed as an extraordinarily efficient combination calculator and typewriter. The essence of the problem posed by the arrival of new goods is as follows.[29] Suppose we are in a time series context and we have price and quantity data for N-1 commodities in periods 1 and 2, p_n^t and x_n^t for $t = 1,2$ and $n = 1, \ldots, N-1$. Suppose in addition that x_N^2 units of a new good are sold at price p_N^2 in period 2. How are we to calculate the bilateral price index $P(p^1,p^2,x^1,x^2)$ when we do not know p_N^1 the price of the new good in period 1? Stated this way, the similarity between the problem posed by new goods and

[29]This discussion follows Diewert (1987).

quality change in existing goods becomes clear. A formal solution to this problem was proposed by Hicks (1940) - if we are in the consumer context, simply calculate the shadow price that would just make the consumer's demand for the good in period 1 equal to zero. The calculation of this shadow price requires knowledge of the consumer's preferences, which might be obtained by econometric techniques. In practice, however, it is too costly to resort to such techniques and official indexes frequently ignore new goods. Cars were not introduced into the CPI until 1940, while the PPI did not include computer prices until 1990.

While there do not appear to be empirical studies of the new goods problem as such, the following example from Diewert (1987) is illuminating, and gives us some sense of the potential magnitude of the problem. Consider a three period situation - there is one "old" good, x_1, with constant price and quantity $p_1^t = x_1^t = 1$ for $t = 1,2,3$. Thus, a price index that ignores the existence of the new good will also be equal to 1 in every period. A new good, x_2, appears in period 2, so that $x_2^1 = 0$. Assume that the new good follows a standard product cycle, being introduced at a relatively high price which subsequently declines over time. Thus, let us assume that $p_2^2 = 2$ and $p_2^3 = 1$. Assume that the quantity purchased of the new good in period 2 is $f > 0$, where f is a fraction representing the period 2 proportion of new goods to old goods. Assume that 2f of the new good is purchased in period 3, and that the shadow price in period 1 that would make demand for the new good in that period equal to zero is 4.[30]

[30]The same indexes are obtained no matter what shadow price is chosen.

The true chain Laspeyres price indexes which do not ignore the new good are $P^1 = 1$, $P^2 = 1$ and $P^3 = (1+f)/(1+2f)$. Table 4 summarizes the true values for P^3 obtained using different index number formulae, along with the bias (difference between actual value computed ignoring the new good and the true value of the index) for different values of f. The bottom line is that ignoring new goods can lead to substantial overstatement of price inflation.

Lebow, Stockton and Roberts (1992) attempt to put more concrete numbers on the size of the new goods bias by making some simple assumptions. They isolate the categories in the CPI where they think rapid product innovation is most likely to be important, finding that the relevant categories account for about 2.4 percent of the CPI (based on December 1991 relative importance figures). Assuming that new products experience price declines at a comparable rate as computers, that is, about 20 percent a year on average, they arrive at an estimate of new goods bias in the overall CPI of 0.5 percent a year. However, their assumption that new goods are only important in the Appliances, Lawn Equipment and Power Tools, and Medical Care Commodities categories of the CPI, along with the assumption that all new goods experience price declines comparable to those experienced by computers, suggest that we should interpret their estimate of new goods bias with caution. Computer prices seem to have been examined so frequently mainly because they have declined at such extraordinarily rapid rates. New goods bias could well be a lot higher or a lot lower than their calculation suggests: we simply do not know.

3.4 List versus transactions prices

As we have already noted, the price information that goes into the CPI

is collected by BLS field representatives through visits or telephone calls. The BLS puts great emphasis on obtaining price quotes that reflect the actual prices paid by consumers, and to this end it makes a number of adjustments to some of the raw price data to obtain better estimates of transactions prices. For example, in pricing new cars, the BLS agents obtain estimates of the base price for the vehicle, along with estimates of the prices of various options and dealer preparation and delivery. The BLS agents also obtain estimates of the average concession or markup during the previous thirty days to arrive at an estimate of the transaction price of the vehicle. The BLS also tries to take account of manufacturers rebates, bonus merchandise, quantity discounts and utility refunds.[31] However, no adjustment is made for the use of cents-off coupons by consumers, except when the coupons are attached to the product for immediate redemption at the time of purchase.

There seems to be relatively little research (in fact, none) on how accurately the prices that go into the CPI reflect the actual prices paid by consumers. (In contrast there are a number of studies addressing this problem for the PPI, as we shall see.) It would appear that the BLS does make a reasonable attempt to ensure that the prices are accurate, but the failure to account for the use of cents-off coupons does raise some questions. Data on coupon use is difficult to come by, although we can get some sense of its potential importance from the Nielsen Report. According to the Nielsen Report, consumers redeemed about 7.7 billion manufacturer-issued coupons in 1992 with an average face value of $0.58, for total "savings" of $4.5 billion, which is slightly less that 1.0 percent of consumer spending in the relevant categories. Somewhat more than 14.0 percent of total grocery volume was

[31]BLS, 1992, pp. 193-96.

purchased with a coupon in 1992. Growth in the average face value of coupons redeemed has consistently exceeded growth in the CPI since 1980. It is an open question whether failure to allow for the use of cents-off coupons by consumers leads to a substantial upward bias in the CPI.

3.5 Treatment of durable goods

The theory of the cost of living index that forms the theoretical basis for the CPI is essentially a static theory. The appropriate treatment of durable goods in such an index requires the measurement and pricing of the flow of services obtained by the consumer from the durable good over its useful lifetime. That is, since a durable good yields a flow of consumption services valued by the consumer over several time periods, we do not want to price the purchase of the good but rather the flow of services that it yields each time period.

In 1983 the BLS switched to the rental equivalence concept to measure housing costs in the CPI (1985 for the CPI-W). The motivation for this change was the large discrepancy that emerged between the CPI and the PCE deflator in the late 1970's due to their different treatment of housing costs and the general perception that the rental equivalence approach employed in the construction of the PCE deflator was superior.[32] Prior to the change the BLS was accused of mixing the consumption and investment components of the cost of housing. The appropriate concept for a cost of living based index was felt to be the cost of the flow of housing services consumed over the measurement interval. However the rental equivalence approach is not without its

---

[32]See, for example, Blinder (1980), Gordon (1981), and Dougherty and van Order (1982).

problems. There are important differences between the markets for owner-occupied homes and the markets for rental units, not least of which is the difference in quality. While initially the BLS simply took figures from the CPI's rental component to estimate the value of owner-occupied housing services, currently they make an attempt to measure the rents of homes that better approximate the typical owner-occupied home.[33]

Francois (1989) examined the issue of whether the manner in which the weight or relative importance of the implicit rental payments of owner-occupiers is calculated leads to biased estimates. The weight for owner-occupied housing in the CPI is determined by asking a sample of owners to estimate how much they believe their house would rent for on a monthly basis.[34] It might legitimately be wondered whether such a question elicits the right answer, or whether owner occupiers are prone to overestimate what their properties would rent for. Preliminary hedonic research had suggested that implicit rent estimates could be biased upward by as much as 10 to 30 percent, but Francois shows that when proper account is taken of certain features of the rental market, the hedonic studies and the owners estimates are in close agreement.

Rogers, Henderson and Ginsburg (1993) evaluate the performance of the CPI's rental indexes during the 1990-91 recession and recovery. Their

---

[33] Pollin and Stone (1991) also argue that there is an important consumption aspect that derives from ownership in and of itself - "freedom from the landlord" is the catch-all term they suggest to capture the notion of utility of security and privacy stemming from homeownership. These benefits are also accompanied by costs that the CPI currently ignores - they suggest the capital gains and losses that accompany fluctuations in property vales.

[34] By contrast, observed changes in the rents paid on rental units that are similar to owner-occupied units are used to track monthly changes in the implicit rents paid by owners.

analysis is motivated by the observation that the rental components of the CPI were a lot stronger than might have been expected during this time period, given that housing prices were depressed, apartment vacancy rates were high and hotel occupancy rates were falling. They model the relationship between economic fundamentals and the CPI rental components and conclude that the behavior of the rental price series in recent years is consistent with what would be expected, once appropriate account is taken of the dynamics of the market for shelter.

Armknecht and Ginsburg (1992) describe research currently being undertaken at the BLS to shift the treatment of autos in the CPI to the theoretically more appropriate flow of services approach. Under ideal circumstances all durable goods would be priced on a flow-of-services basis. In reality this is not possible, primarily because of the absence of rental markets for many types of durable goods.[35] For autos, however, there are two very active and distinct rental markets that may enable the rental equivalence approach to auto prices to be implemented. Armknecht and Ginsburg point out that car use in the traditional (short-term) rental market is very different to normal use, making it unrepresentative of the general population. However, the long-term rental market for auto leases may provide more appropriate measures of the rental equivalence of auto services. The feasibility of pricing auto transportation services on this basis is currently being examined by the BLS.

3.6 Measuring the prices of services

Triplett's 1975 survey reviewed a small number of studies that attempted

[35]As Armknecht and Ginsburg note, there is no rental market for shoes.

to assess the quality of the CPI service price indexes. Four of the six studies he reviewed examined the medical services component of the CPI, and three of these four found upward bias in the CPI components. However, the main conclusion that Triplett draws from these studies is that the appropriate pricing concept in the medical services area is not very well defined. One simple way of stating this is to note that it is unclear whether we should be pricing treatments or cures. Triplett concludes his 1988 review by noting that "...existing research on the subject is insufficient to indicate whether the medical care components are upward biased."(Triplett, 1988, p.70). However, the BLS handbook quite explicitly states that in many instances quality changes are treated as price changes, either because the BLS is unaware of the quality change or because it has no method for dealing with the quality change.(BLS, 1992, p.193).[36] In instances where quality adjustment is feasible, it is carried out. The other studies of measurement problems in services that Triplett reviewed had to do with pricing visits to the movies, and the problem of aging bias in housing.

Recently Armknecht and Ginsburg (1992) have reviewed the manner in which the BLS has tried to deal with the increasing importance of services in consumers budgets, and research currently underway to improve measurement and quality adjustment of services prices. They note attempts to apply hedonic techniques to the measurement of the price of hospital room stays, and cite the problem of properly accounting for and measuring differences in the level of nursing care among hospitals as a major obstacle to the implementation of

---

[36]Thus Tregarthen (1993) points out that recent concerns over the rising cost of health care may be misplaced, in part because of the failure of the BLS to properly account for quality change in health care, and in part because of the reliance of the BLS on list rather than transactions prices in pricing the health components of the CPI.

hedonic methods.

Kroch (1991) reviews the problems of price measurement in the service sector, and argues that the true rate of service sector inflation is lower than the measured rate, and closer to the rate in the goods sector. Kroch's critique of service price measurement extends beyond the services component of the CPI to include services in the national income and product accounts. He correctly notes the problems that arise when price movements are inferred from earnings data, especially when productivity is increasing. This is an important caveat accompanying services deflators in the NIPA, but is of less relevance to the treatment of services in the CPI. Kroch argues that the slower rate of inflation for medical equipment is indicative of upward bias in the CPI medical services category, and suggests that a comparison of the inflation rates of the two series indicates that medical services inflation was overstated by as much as 1.0 percent a year during the 1980's. Kroch also argues that the education price category overstates inflation when compared to an index of tuition for higher education. The last category of services that Kroch considers is the rental equivalence measure of owner-occupied housing, and while he suggests that the failure of the rental equivalence index to track house prices in recent years may suggest that the CPI is overstating inflation in the housing services category, he refrains from drawing a firm conclusion.[37]

It is of course important to keep in mind that despite the problems that may accompany price measurement for many services, the influence of these problems on the overall CPI is determined by the importance of the problem

[37]However, the study by Rogers, Henderson, and Ginsburg (1993) suggests that this is not a problem.

categories in the consumers budget. Thus, even if it were true that inflation in the medical care component of the CPI was overstated, the fact that medical care only accounts for 4 percent of consumers expenditure would greatly limit the influence of mismeasurement in this component on the overall CPI.[38] The categories of the CPI that Kroch argues may be overmeasuring inflation together account for a mere 5.6 percent of budget outlays in the base period (see Kroch, 1991, p.32)

3.7 Conclusions about measurement bias in the CPI

We began this review of the CPI quoting form the earlier survey by Triplett (1975), who concluded that as of the mid 1970's not enough was known to say whether there was a clear overall bias in the CPI, nor its sign. Triplett repeated this opinion in his unpublished 1988 survey. Have we learned anything in the intervening period that would lead us to draw different conclusions?

It seems clear that the issue of substitution bias has been settled - it is unlikely that the Manser and McDonald study will be improved upon (in the sense of achieving an even finer level of commodity disaggregation) any time soon, and we can safely conclude that the substitution bias arising from the use of fixed weight index is at most 0.2 percent a year, and probably less. Many authors have made a lot of the outlet substitution bias recently quantified by Reinsdorf (1993). In view of this we gave it more detailed coverage in our review of the recent literature, and argued that it needs to backed up by further work before the figure of an upward bias of up to 2

[38]Medical care accounts for a larger percentage of personal consumption expenditures in the National Income and Product Accounts (NIPA) because of the inclusion of employer provided benefits.

percent a year can be accepted as true.

We noted that some of the categories of the CPI that Triplett pointed to as having potential downward bias, such as housing, are now treated differently, and aging bias is less likely to cause inflation of housing costs to be understated. Apparel remains a problem, although it is interesting that the studies by Armknecht and Weyback (1989) and Liegey (1993) find both upward and downward bias in this category. Auto prices are still biased downward for the reasons stated by Triplett, and will remain so until the BLS changes its methodology. We have no firm evidence on the size and nature of the biases in pricing medical care and so cannot draw any firm conclusions about this category.

In conclusion, while we have made further inroads into understanding the problems with the CPI over the past 5 to 10 years, the recent research does not allow us to draw any firmer conclusions about the overall bias in the CPI than Triplett was able to draw almost 20 years ago. There has yet to be a comprehensive independent study of the problems in the CPI similar to that carried out by Gordon (1990) for the PPI, which we will discuss in more detail below.

4. The Producer Price Index

The purpose of the Producer Price Index (PPI) is to measure the average change in selling prices received by domestic producers for their output. The PPI or industrial price program is based on prices received by domestic producers from whoever makes the first purchase. The PPI is one of the oldest price series published by the Federal Government, extending back to 1890, and was known as the Wholesale Price Index (WPI) until 1978. The methodology used to construct the various price indexes used to construct the PPI has been revised many times, with the most recent extensive revision occurring in the late 1970's and early 1980's. The methodology currently in use has been in place since 1986. The universe that forms the basis for PPI surveys consists of all industries in the goods making sectors of the US economy, along with gas and electricity and other goods that are competitive with those made in the goods sectors. There is also partial coverage of the service sectors.

As with the CPI, definition of "price" is an important first step in the construction of the PPI. According to the BLS Handbook of Methods

> "For the purposes of the industrial price program, a price is defined as the net revenue accruing to a specified producing establishment from a specified kind of buyer for a specified kind of product shipped under specified transaction terms on a specified day of the month."(BLS, 1992, p.141)

The BLS stresses to firms that cooperate in the survey that they report transactions prices that include all discounts, rebates etc. rather than list or book prices, and claims that the use of list as opposed to transactions prices in the PPI program has been the exception rather than the rule (BLS,

1992, p.141).[39] Most prices that go into the PPI refer to the Tuesday of the week containing the 13th of the month.

The PPI is calculated using a modified Laspeyres formula:

$$PPI_t = 100 \times \frac{\sum_i p_i^t q_i^b}{\sum_i p_i^0 q_i^b}$$

where p_i^0 is the price of a commodity in the comparison period, p_i^t is the price of a commodity in the current period, and q_i^b is the quantity of the commodity shipped during the weight-base period. When the weight-base period and the comparison or reference period coincide, the above collapses to the traditional Laspeyres index. The actual computation procedure is more closely approximated by a slightly different formula:

$$PPI_t = 100 \times \sum_i \left[\frac{p_i^t}{p_i^0}\right] \frac{p_i^0 q_i^b}{\sum_i p_i^0 q_i^b}$$

This formula simply restates the PPI as a weighted index of "price relatives" or price ratios for each item (p_i^t/p_i^0). The advantage of rewriting the formula this way is that the index can be calculated using only data on prices and actual expenditures, both of which are typically more readily available than quantities purchased.

As with the CPI, the BLS does not publish the weights used to construct

---

[39]The BLS claims that prior to the adoption of the current methodology for collecting prices, less than 20 percent of the commodity price indexes were based on list prices, and expresses confidence that the use of list prices is even less of a problem now. A number of authors have expressed skepticism about the 20 percent figure, including Foss (1993).

the aggregate indexes but rather the relative importance of various commodities and commodity groups. The relative importance data are calculated for December of each year, and change from one year to the next as relative prices change.

4.1 Quality adjustment in the PPI

The procedures used by the BLS to make quality adjustments in the PPI are described in U.S. Department of Labor (1992). The procedures are essentially the same as those used in the CPI, with some subtle differences. Part of the reason for the hybrid methodology that evolved for handling quality change in the PPI may have to do with the fact that the PPI does not have as clear a conceptual basis as does the CPI. While the theory of the cost of living index is the guiding principle behind the construction of the CPI, the appropriate theory for use in construction of the PPI is the theory of output price indexes or input price indexes, depending on the use to which the price series is to be put. However, the revised (post 1986) PPI takes the theory of output price indexes as its conceptual basis.[40]

Changes in the quality of the items being priced are dealt with in one of four ways.[41] First, if the change in the specification of the product is so minor that there are no differences in the cost of the product, the BLS ignores the change and the new price is directly compared to the old price. Second, if the cost of the product does change, the cost difference is used to

[40]See Gousen, Monk, and Gerduk (1986).

[41]See the discussion above on quality adjustment in the CPI. The methods used for quality adjustment in the PPI are essentially the same as those used for quality adjustments in the CPI, with one interesting exception.

estimate the value of the change in product quality and a quality adjusted price is put into the index. Third, in situations where it is not possible to estimate the production cost difference between a new and an old model, or a comparison between an entirely new and an old product is not feasible, the BLS may simply assume that all the difference in price between the old and the new products is due to quality differences. This contrasts with the method used in the CPI, where price change is imputed from the price changes of similar goods.[42] This is known as the overlap method, or linking to show no change in price. Finally hedonic methods may be used in situations where none of the above are appropriate. Since 1991 the BLS has used hedonic methods to estimate quality adjustments for computer prices.

4.2 How accurate is the PPI?

As with the CPI we take as our point of departure the state of knowledge as surveyed by Triplett (1975) and updated in Triplett (1988). Triplett concluded his 1975 survey of the PPI "As with the studies surveyed in the CPI section, no clear pattern of the direction of index bias emerges from the studies on quality error in the WPI."(Triplett, 1975, p.58) Triplett essentially re-iterated this point in his 1988 survey, noting that so few of the component series of the PPI had been examined in any detail that it was not possible to come to an overall judgement as to the quality of the PPI. Triplett did note the results of the Lichtenberg and Griliches (1989) study that we will consider in more detail below, expressing surprise at the size of the overall quality bias they claim to find in the PPI. However he neglects

[42]The reasons for this difference between the CPI and the PPI are unknown. See Triplett (1988), p.38.

to mention the results of Gordon's study of durables prices, since published as Gordon (1990), which apparently circulated in draft form for quite some time before their official publication.[43]

Two recent studies of how well the BLS handles quality adjustment in the PPI are Lichtenberg and Griliches (1989), which was extended by Siegel (1991). The approach used in both of these papers is to estimate simple multiple-indicator multiple-cause (MIMIC) or factor analytic models of price change. The first equation in such a model is a measurement model, relating a vector of observable variables, Z_t, to a vector of unobservables, X_t^*:

$$Z_t = AX_t^* + \epsilon_t$$

The observables in Lichtenberg and Griliches' study are the measured PPI, the unit value relative (UVR) collected by the Census Bureau,[44] and a measure of the rate of introduction of new products. Siegel also includes changes in capacity utilization as an indicator of industry supply shocks, and adds two other indicators of quality change based on the price changes of new products as reflected in the PPI and the UVR. Lichtenberg and Griliches consider X_t^* as consisting of two variables: the unobserved true PPI, P_t^*, and quality change, Q_t^*. Siegel adds a third unobservable, S_t^*, an industry-specific supply shock.

The second equation is a structural model, relating the vector of

---

[43]See Gordon (1990) p.35, footnote 22.

[44]The unit value collected by the Census Bureau is defined as the ratio of the value of shipments of a product to the number of units shipped. The unit value relative (UVR) is simply the ratio of two unit values - in the Lichtenberg and Griliches study it is the ratio of the 1977 to the 1972 unit value.

unobservables, X_t^*, to a vector of causal variables, Y_t:

$$X_t^* = BY_t + \eta_t$$

Lichtenberg and Griliches include measures of R&D in the vector of causal variables, while Siegel extends the vector of causal variables to include the prices of energy and materials, and wages.

Lichtenberg and Griliches (1989) analyze data based on the 1977 Indexes of Production at the four-digit SIC level - Siegel updates their study to the 1982 Indexes. While Lichtenberg and Griliches conclude that the PPI adjusts for about two-thirds of quality change (which they estimate occurs at an average annual rate of 0.9 percent), Siegel concludes that the PPI adjusts for 57 percent of quality change.[45] We can obtain an implicit estimate of the quality bias in the PPI based on Lichtenberg and Griliches results as being around 0.3 percent a year (i.e one-third of the estimated average annual rate of quality improvement). Siegel does not provide an estimate of the average annual rate of quality improvement over the sample he studies, so we cannot perform a similar calculation. However, given the similarity of his estimate of the extent to which the PPI adjusts for quality change to the finding of Lichtenberg and Griliches, and assuming that the rate of quality change was roughly similar in the 1972-77 and 1977-82 time periods, we would obtain a

[45]The results of the Lichtenberg and Griliches study are stated in contradictory ways in the Abstract and Conclusions of their paper. In their Abstract, they state that "Apparently, less than half of quality change is adjusted for in the PPI" while in their Conclusion they state that "Almost two-thirds of quality change, which we estimated to occur in our sample at an average annual rate of.9%, appears to be adjusted for in the PPI." The results reported in their Table 6 would suggest that the latter is the correct figure.

figure close to 0.3 percent a year.

One caveat that must accompany these results is that they are not based on a random sample of manufacturing firms. The sample of firms or products included in the studies was determined by the availability of both price measures, i.e. PPI's and UVR's. Lichtenberg and Griliches note that to the extent that the problem of measurement error is less serious for the products included in their sample than for those not in their sample, their estimates of the average rate of quality change (and hence of quality bias) should be interpreted as lower bounds. Of course, to the extent that the problems of measurement error are more serious for the products in their sample, the opposite conclusion holds. Lichtenberg and Griliches do not provide any additional evidence as to which of the two possibilities is more likely. Siegel, however, presents evidence that suggests that sample selection bias is not a problem.

It is also important to note that the methodology of quality adjustment now used in the PPI is different to what was in place during the time periods covered by these studies. Siegel notes Triplett's (1988) comment that the shift in PPI methodology towards more linking (and hence possibly overadjustment for quality change) may well offset Lichtenberg and Griliches' finding of upward bias in the PPI.

4.3 Computers and related products

Computers posed unique problems for quality adjustment in the PPI.[46] The standard quality adjustment procedure used in the PPI (particularly with

[46]The state of research on computer prices as of the mid-1980's is reviewed in Triplett (1989).

automobiles) is to make quality adjustment on the basis of manufacturers cost differential between new and old product. The implicit assumption underlying this approach is that increased costs of production are typically associated with improvements in quality of the final product, and declining costs are associated with deterioration in product quality. The computer industry posed special problems because rapid improvements in quality were associated with rapid productivity growth in computer manufacturing.

The BLS introduced its computer price series in January 1991, nearly forty years after the appearance of the first commercial computers, and discontinued publication of the experimental computer price indexes that had been introduced in 1988. The experimental PPI computer price indexes are described in Sinclair and Catron (1990). The BLS Handbook of Methods (1992) gives no other reference for the computer price series in the PPI other than the Sinclair and Catron article on the experimental indexes, despite some obvious differences between the two (e.g. the experimental indexes were calculated on a quarterly rather than on a monthly basis, and the experimental sub-indexes do not correspond with the sub-indexes currently being published).

Evidence that the PPI for semiconductors is subject to error is presented in the papers by Dulberger (1993), Norsworthy and Jang (1993), and Flamm (1993). Dulberger (1993) shows that a chain price index of semiconductors falls a lot faster than the fixed-weight PPI for semiconductors, probably because of delays in introducing new products into the PPI - other evidence is offered by Norsworthy and Jang (1993). Triplett (1993) commenting on all three papers (Dulberger, Norsworthy and Jang, and Flamm) notes that all three imply that the PPI for semiconductors overstates the rate of increase in their price or understates the price decline, although

he notes problems of comparability between the results of the studies and the PPI. Berndt and Griliches (1993) (using a hedonic regression approach) calculate that the quality adjusted price of PC's fell by about 25 percent per year from 1982-1988, which is notably faster than the 16 percent annual decline in the matched model index used by the BEA. However they characterize their findings as preliminary, on the basis of the incomplete list of characteristics included in their hedonic regressions. In her comments on the Berndt and Griliches paper, Cole (1993) notes the difficulty in applying hedonic methods to computers because of the problem of obtaining accurate measures of important characteristics - the example she gives is the difficulty of measuring speed, and especially in constructing measures of computer speed that are comparable across products. Significant differences can exist between clock rate speed and speed measures based on specific applications (e.g. spreadsheet or word processing).

Triplett (1993b) refers to the delay in introducing new chips into the semiconductor index as "new introductions" bias and argues that it is distinct from the quality adjustment problem. The new introductions problem does not just exist in the semiconductor industry but also in other technologically dynamic sectors such as pharmaceuticals. New introductions bias is a quality problem, but not a quality adjustment problem, although inadequate quality adjustment methods may well exacerbate it. Rather new introductions bias is a sampling problem, where rapid technological change creates a sample that is not representative of current price change in the industry. While more frequent initiations may alleviate the problem they cannot eliminate it.

4.4 Pharmaceuticals

The problems of price measurement in the pharmaceuticals industry have been studies in some detail in Berndt, Griliches, and Rosett (1993) and Griliches and Cockburn (1992). Berndt, Griliches and Rosett try to explain why the official PPI for pharmaceuticals grows at an average annual rate of 9.09 percent between 1984 and 1989 when a Divisia index calculated using data on more than 2,000 products sold by four of the ten largest pharmaceutical companies grows at an average annual rate of only 6.03 percent over the same period.[47] They consider and rule out a variety of possible explanations, such as that the four firms are unrepresentative of the industry as a whole, or that the difference is due to the use of a Divisia rather than the (official) Laspeyres index. Berndt, Griliches and Rosett find that about half of the discrepancy "can be attributed to the fact that the BLS tends to undersample younger products that experience less than average price increases and oversample medium-age products that undergo above average price increases."(Berndt, Griliches and Rosett, 1993, p.262).

Griliches and Cockburn (1993) have studied the quality change problem for the case of generic pharmaceuticals. Griliches and Cockburn argue that in their attempts to deal with the general quality change problem, the BLS (and other statistical agencies) may inadvertently aggravate it by defining goods too narrowly and linking-in new goods that are in fact comparable to older, existing goods. As a result many price declines are missed by the official price indexes. Studying generic pharmaceuticals gives useful insights into the extent of this problem, as generic drugs are equivalent in almost all

---

[47]A Divisia index is defined as $d\log P_D = \left(\sum_i p_i^t x_i^t d\log p_i^t\right) / \left(\sum_i p_i^t x_i^t\right)$.

relevant respects to the previously patented brand name drugs. Griliches and Cockburn show that using BLS methods the "official" price index records an increase of 14 percent in the price of cephalexin (one of the drugs they study) over their 45 month sample period, while their preferred "adjusted Paasche" index falls by 48 percent. They conclude by advocating more current sampling of new products, faster introduction of new products into the official indexes, and more current weighting of products in the official indexes.

4.5 Computed Tomography (CT) Scanners

Trajtenberg (1990) proposes a new way to construct quality adjusted prices indexes that is based on discrete choice models and the "characteristics approach" to demand theory. To implement his proposed methodology, Trajtenberg studies the case of Computed Tomography (CT) Scanners during the first decade of their existence. He estimates that the quality adjusted price of CT Scanners fell at an average annual rate of 55 percent a year between 1973, when they were first introduced, and 1982. A hedonic index captures only part of this decline in the quality adjusted price, while the raw (nominal) price posts an increase over the same period. Trajtenberg's findings are summarized in Table 6. Note that if we start the comparison in 1977, the extent of the discrepancy between Trajtenberg's preferred (ΔW - based) quality adjusted price series and the hedonic series is reduced, which Trajtenberg interprets as suggesting that hedonic methods may work reasonably well for technologically mature industries. The difference between the two indexes is attributed to the fact the ΔW - based index allows for discrete product innovations, while the hedonic index assumes that changes occur

continuously. Trajtenberg notes that his alternative method of quality adjustment are probably not applicable on a wide scale because of its "voracious data requirements."

4.6 List versus transactions prices

One of the oldest concerns of economists with regard to the PPI is the possibility that the BLS uses list rather than transactions prices in constructing the index. The concern here is that this would impart a bias to the PPI insofar as transactions prices may be more flexible than list prices over the course of the business cycle, declining relative to list prices as demand contracts and rising relative to list prices as demand expands. The Stigler Report in 1961 highlighted the problem of list-transactions prices as one of the major shortcoming of the WPI, recommending that "The individual product prices should, where feasible, be collected from buyers (not from sellers, as at present) to get more accurate information on actual transactions prices."(NBER, 1961, p.21) A similar conclusion was reached by the Ruggles Report in 1977, which noted that "In many instances, the wholesale price data fail to capture changes in actual transactions prices; rather they reflect changes in manufacturers' list prices." (U.S. Executive Office of the President: Council on Wage and Price Stability, 1977, p.ii) In his 1975 survey, Triplett refers to Stigler and Kindahl (1970) as the most important piece of empirical research on the transactions-list price problem.[48] Stigler and Kindahl gathered prices from buyers of a selected group of products rather than sellers (as the BLS does) and constructed indexes that

[48]Triplett also references earlier studies by Flueck (1961) and McAllister (1961), both of which were sponsored by the Stigler Committee.

they then compared to the comparable PPI's (then called the WPI). They found differences between the movements of their price indexes and the WPI, which have been interpreted by some as indicating that the WPI relies on list rather than transactions prices.

Two recent studies addressing the transactions-list price problem are Betsock and Gerduk (1993) and Foss (1993). Betsock and Gerduk consider the problem of obtaining accurate measures of transactions prices in the steel industry (which was also studied by Stigler and Kindahl (1970), Flueck (1961) and McAllister (1961)). They document the use of list prices through the 1980's and show how this imparted a substantial upward bias to the index for steel. Specifically, in 1982 and 1986 list prices and transactions prices moved in opposite directions. These episodes motivated the BLS to review its pricing of steel products. Since January 1990, the steel index has been based only on net transactions prices. The BLS has refused to accept list prices, even at the risk of not being able to publish certain indexes. Prior to this date the BLS accepted list prices, as the steel producers simply refused to supply transactions prices. Since 1990, steel producers have generally been more cooperative in supplying net transactions prices.

Foss (1993) attempts to explain what factors might motivate firms to report list rather than transactions prices to the BLS. Foss's paper was motivated by the transactions-list problem in the steel industry in the 1980's, and he expresses the opinion that while recent changes in the way steel prices are reported may have alleviated the problem, it is still a very real problem. He points to the existence of the Robinson-Patman Act, which outlaws price discrimination, as a major impediment to the accurate reporting

of transactions prices by firms.[49] As evidence Foss compares response rates from businesses to a number of large scale voluntary surveys conducted by the BLS and the Bureau of the Census. These are summarized in Table 7. Foss interprets these figures as indicating that the response rate for the PPI is low relative to comparable voluntary surveys carried out by the Federal government, and argues that the fear of lawsuits under the Robinson-Patman Act is partly responsible.

Lichtenberg and Griliches (1989) also address the issue of whether the PPI or unit-value relatives is a better price signal. Based on a sample of products from 238 industries they conclude that the PPI is a far more reliable indicator of long-term price change, on the basis of an estimated signal-to-noise ratio 2.72, as opposed to a signal to noise ratio of 0.53 for UVR's. Siegel (1991) confirms the result that the UVR is a noisier indicator of price change than the PPI. Gordon (1990) also reviews the usefulness of unit value indexes as measures of price change, and concludes that they are useful indicators of the existence of deviations between transactions and list prices as long as the problem of shifting product mix is properly accounted for.

4.7 "The Measurement of Durable Goods Prices" by Robert J. Gordon

It is worth reviewing in some detail Gordon's work on the measurement of durable goods prices, as it constitutes the only recent attempt to conduct a comprehensive review of the PPI and draw conclusions about its overall reliability. Gordon constructs a new series of price indexes of producers and consumers durable goods using data collected from an assortment on non-

[49]Stigler and Kindahl also mention the existence of the Robinson Patman Act as an impediment to the accurate reporting of transactions prices. See Stigler and Kindahl (1970) p.26.

standard data sources, such as the Sears Catalog, Consumers Reports, PC magazine, and NADA Official Used Car Guide, among others. The price indexes that he calculates as alternative deflators of the various categories of producers durable equipment (PDE) "are weighted averages of 25,650 separate price quotations selected to satisfy the twin criteria of careful adjustment for quality change and measurement of transactions rather than list prices."(Gordon, 1990, p.7). Adjustment for quality change was done using either hedonic methods or the conventional BLS specification method. Gordon checked the accuracy of his alternative price indexes by developing price comparisons for "closely similar" products over long periods of time. The purpose of this is to provide concrete examples of goods with roughly the same quality characteristics whose price trends matched those of the new indexes.

Table 8 presents the overall results from Gordon's study. The table presents the difference in the growth rates of the new and official price indexes for the twenty-two categories of goods that make up producers durable equipment (PDE). Within each of the twenty-two product categories NIPA weights were used to aggregate individual price series using the Tornqvist index number formula. The figures in the table show the average annual growth rate of the ratio of the alternative to the official price indexes for the entire sample period and various sub-samples. The prevalence of negative numbers indicates that in almost every case, the official price series for the indicated categories grew faster than the alternative quality adjusted price series. Gordon claims that his results decisively refute the possibility of downward quality bias for durable goods advanced by Triplett.

Table 9 presents the weighted average drift of this ratio, which is obtained by applying the Tornqvist formula to each of the twenty-two

categories of goods and using NIPA nominal annual investment in each of the twenty two categories as weights. The line in Table 9 labelled "Tornqvist" is, in Gordon's words, "the bottom line of the whole study."(Gordon, 1990, p.535) This indicates an overall drift in the ratio of the alternative and official indexes of 2.96 percent a year, that is, the new alternative index increases at an average annual rate that is 2.96 percent slower than the rate of increase of the official deflators when aggregated in a comparable manner. Since the same weighting and aggregation methods are used for the alternative and official indexes, all of the difference is interpreted by Gordon as being due to a failure of the official indexes to properly account for quality change. Note that the aggregate quality bias is greater in the period before 1960 when the BLS made comparatively few quality adjustments to the price data used in the construction of the PPI. Gordon notes that there is a significant correlation between the rate of growth of investment in a sector and the rate of drift of the ratio of the price and interprets this as supporting the hypothesis that the drift is due to unmeasured quality change in the durable goods that are most in demand.

Gordon claims that his new price indexes may still be biased upwards because of unmeasured quality change. The other caveats that accompany his conclusions are that only about three-quarters of producers durable equipment (by 1967 value) is covered by his new indexes. For the uncovered products, he assumes that within any of the 22 categories of producers durable equipment the observed drift for the covered goods can be applied to the uncovered goods. Triplett (1993a) argues that by following BLS practice of treating mandated pollution devices, safety equipment and so on as quality improvements, Gordon produces figures for the quality bias in the PPI that are

too high. Furthermore as Triplett noted in his 1975 survey and again in his review of Gordon (1990) (see Triplett (1993a)), while studies based on alternative data sources may yield interesting insights into the behavior of the official indexes, the difference between the behavior of an alternative index based on an alternative data source and the PPI cannot always be unambiguously attributed to some problem in the official index - inappropriate quality adjustment in the case of Gordon's study.) This leaves outside reviewers and critics of the PPI in an awkward position, as BLS confidentiality rules generally prohibit access to the raw price series that are used in the PPI.

4.8 Conclusions about measurement bias in the PPI

Triplett concluded his 1975 and 1988 surveys by noting that so few components of the PPI had been studied in any detail that it was impossible to reach any conclusions about the likely overall direction of bias. Since then, the monumental study of durable goods prices by Gordon (1990) has appeared, seeming to suggest the existence of substantial upward bias in the PPI (of the order of magnitude of around 3.0 percent a year). Berndt, Griliches and Rosett (1993) find an upward bias of a similar order of magnitude for the pharmaceuticals industry, while Lichtenberg and Griliches (1987) and Siegel (1992) obtain smaller estimates of bias in the PPI.

Unlike the CPI, there do not appear to be many documented cases of downward bias in the PPI,[50] although Triplett (1988) does suggest that the change in PPI methodology in 1986 may make this more likely.

---

[50]Although note the instances of positive drift in the ratios of Gordon's alternative and official price indexes in Table 8.

We conclude that while those studies of the PPI that have been carried out since Triplett's surveys have tended to find more examples of upward than of downward bias in the PPI, the limited nature of these studies, the evolution of the methodology of correcting for quality change in the PPI and the almost complete absence of replication make it difficult to draw any firm conclusions about the direction or magnitude of biases in the PPI.

5. The GDP deflators

The most comprehensive set of price series available for tracking inflation trends in the US economy are the various GDP deflators, in that these deflators cover all sectors of the economy. By contrast, the CPI only tracks the average price paid by the urban population for final consumption purchases, whereas the PPI only tracks prices received by domestic producers of tangible goods. However, as Triplett (1975) noted, the comprehensiveness of the GDP deflators is partly illusory. The BEA relies heavily on the CPI and PPI series published by the BLS to deflate many components of GDP. In some cases the BEA constructs its own deflators when the BLS series are inadequate. The best example of this is of course the BEA computer price deflator, which was introduced many years before the PPI included a price series for computers.[51] When the BEA introduced its new price index for computers in 1985, the growth rate of real GNP for the period 1982-88 was revised up from 3.8 percent a year to 4.1 percent a year. Since 1990, the BEA has used the new PPI's for computer prices to interpolate quarterly values of its annual price series for mainframe computers. The BEA also constructs its own deflators for housing (see, for example, de Leeuw (1993)) and defense purchases (see, for example, U.S. Department of Commerce (1988) and Ziemer and Alexander (1993)). The sources of the price data used to deflate the various components of GDP are illustrated in Table 10. The CPI and the PPI, or rather, components of the CPI and PPI, are used to deflate more than three-quarters of GDP.

Because the GDP deflators are not primary source data, they are only as

[51]The BEA computer price deflator is described in Cartwright (1986), Cole, Chen, Barquin-Stolleman, Dulberger, Helvacian, and Hodge (1986), and Cartwright and Smith (1988).

good as the more basic indexes used in their construction. Thus all of the shortcomings of the CPI and the PPI discussed in the previous sections apply to the GDP deflators as well. At times, the BEA may do slightly better than the BLS in handling problems of quality change: the earliest use of a hedonic index in official price series was in the deflation of residential construction in the NIPA, and as we have already noted, the BEA was ahead of the BLS in introducing a quality-adjusted price series for computers.[52] At the aggregate level, the implicit deflators are constructed using the Paasche formula. However, within categories of expenditure the deflators are combined using Laspeyres weights, with the result that the overall indexes are of a mixed type.

For some components of GDP, the BEA relies on cost indexes to deflate nominal product. The use of cost indexes for deflation is predicated on the questionable assumption of constant productivity. If productivity is growing, cost indexes will result in over deflation, that is they will be upward biased. Pieper (1989, 1993) reviews the problems that arise in deflating construction expenditures, and shows that the BEA deflators for construction were upward biased for the period 1963-82.[53]

Recently the BEA introduced alternative measures of real GDP and the GDP deflator (see Young (1989,1992,1993) and Triplett (1992b)) to supplement the

---

[52]The BLS computer price index has been subject to criticism in some quarters. See Denison (1989), who argues that the BEA index overstates the price decline in computers by focusing solely on the characteristics of computers and ignoring the additional costs that the user incurs.

[53]Suspicions about the quality of the construction deflators are heightened by the fact that measured productivity in construction has been falling since 1963. See Pieper (1989) pp. 293-94.

existing fixed weight deflator.[54] The motivation for introducing chain weighted indexes was twofold. First, the experience with food and energy price shocks in the 1970's, and second, the discovery that choice between 1982 and 1987 price weights made a difference of 0.3 percent in the estimated average annual growth rate of real GNP over the 1982-88 period because of the rapid decline in the price of computers. The new or alternative measures of GDP prices are based on the Fisher Ideal index number formula, and are designated the "chain-type annual-weighted price index" and the "benchmark-years-weighted price index".

The chain-type annual-weighted price index is defined as follows:

$$\left[\frac{\sum_i p_i^t x_i^{t-1}}{\sum_i p_i^{t-1} x_i^{t-1}} \times \frac{\sum_i p_i^t x_i^t}{\sum_i p_i^{t-1} x_i^t}\right]^{0.5}$$

For the quarterly indexes, annual quantities for adjacent years are used as weights.

The benchmark-years-weighted GDP price index is defined thus:

$$\left[\frac{\sum_i p_i^t x_i^a}{\sum_i p_i^{t-1} x_i^a} \times \frac{\sum_i p_i^t x_i^b}{\sum_i p_i^{t-1} x_i^b}\right]^{0.5}$$

where a and b are benchmark years and $t = a+1, a+2, \ldots, b$. The benchmark years are used as weighting periods because the benchmark year price and quantity data obtained from quinquennial economic censuses are considered superior to those for other years. The cumulation of the benchmark-years-weighted index

[54]See also Motley (1992) for a description of these changes.

values for the years between the benchmark years a and b is equal to the Fisher Ideal value calculated directly from year a to year b:

$$\left(\frac{\sum_i p_i^{a+1}x_i^a}{\sum_i p_i^a x_i^a}\times\frac{\sum_i p_i^{a+1}x_i^b}{\sum_i p_i^a x_i^b}\right)^{0.5}\times\left(\frac{\sum_i p_i^{a+2}x_i^a}{\sum_i p_i^{a+1}x_i^a}\times\frac{\sum_i p_i^{a+2}x_i^b}{\sum_i p_i^{a+1}x_i^b}\right)^{0.5}\times\ \ldots\ \times\left(\frac{\sum_i p_i^b x_i^a}{\sum_i p_i^{b-1}x_i^a}\times\frac{\sum_i p_i^b x_i^b}{\sum_i p_i^{b-1}x_i^b}\right)^{0.5}$$

$$=\left(\frac{\sum_i p_i^b x_i^a}{\sum_i p_i^a x_i^a}\times\frac{\sum_i p_i^b x_i^b}{\sum_i p_i^a x_i^b}\right)^{0.5}$$

For years beyond the most recent benchmark year, the benchmark years price index is calculated using weights from the most recent benchmark year and weights from the most recent year. Quarterly values of the benchmark years weighted index are calculated using weights from adjacent benchmark years. Quarterly figures from both series are adjusted so that the annual average of the four quarterly figures equals the annual figure.

The choice between the two new indexes is dictated by the question one wants the index to answer. For quarter-to-quarter comparisons of inflation rates, the chain-type annual-weighted index is preferable, while for longer term analysis of inflation trends the benchmark-years weighted index is probably more desirable.[55]

With the introduction of these new chain-linked indexes the BEA would appear to be slightly ahead of the BLS in terms of the quality of the price

[55]For further discussion of the choice between the two, see Triplett (1992b).

series they produce.[56] The superior properties of the new alternative deflators make them candidates for consideration as alternative measure of inflation.

Conclusions about the GDP deflator

Since three quarters of GDP is deflated using either the CPI or PPI, any conclusions that we draw about biases in those price series apply with equal force to the GDP deflators as well. The remainder of GDP is deflated using price series of unknown quality. Construction and defense pose special problems for price measurement that may result in price changes in those sectors being overstated.

[56]Although, as we have already noted, the BLS is also experimenting with chain-linked versions of the CPI. See Kokoski (1989). We might note that the U.K. retail price index is a chain-linked index - see Layard and Walters (1978), and Allen (1975).

6. What price level to stabilize?

The original purpose of the PPI was to measure changes in the purchasing power of money. However, more recently attention has switched to the CPI and the GDP deflator as the most closely watched indicators of inflation. As we have already noted, the BLS explicitly states that the CPI is intended to approximate the behavior of a cost of living index. Many of the improvements in the CPI over the years have come about as a result of its being more firmly based in the economic theory of the cost-of-living index. However, it is worth considering whether the appropriate theory of the price level that is relevant for monetary policy is the theory of the cost of living. Rather, it can be argued that the appropriate conceptual basis for the price level relevant for monetary policy is the theory of the demand for money, or quantity theory of money as outlined Fisher (1922) and Friedman (1987).[57] It is interesting to note that Fisher was quite explicit in stating that it is not just the price of consumption goods that should be considered when trying to determine whether money is losing its value, but rather the average price of all transactions that are conducted through the medium of money:

> "...perhaps the best and most practical scheme [for the construction of an index number] is that which has been used in the explanation of P in our equation of exchange, an index number in which every article and service is weighted according to the value of it exchanged at base prices in the year whose level of prices it is desired to find. By this

[57]We note that recent papers by Dow (1993) and Bryan and Cecchetti (1993b) also distinguish between the cost of living and the price level that is relevant for monetary theory and policy.

means, goods bought for immediate consumption are included in the weighting, as are also all durable capital goods exchanged during the period covered by the index number. What is repaid in contracts so measured is the same general purchasing power. This includes purchasing power over everything purchased and purchasable, including real estate, securities, labor, other services, such as the services rendered by corporations, and commodities."(Fisher, 1922, pp.217-218)

However, Fisher noted the practical difficulty of collecting data on all of the relevant prices, and concluded that

"It is, of course, utterly impossible to secure data for all exchanges, nor would this be advisable. Only articles which are standardized, and only those the use of which remains through many years, are available and important enough to include. These specifications exclude real estate, and to some extent wages, retail prices, and securities, thus leaving practically nothing but wholesale prices of commodities to be included in the list of goods, the prices of which are to be compounded into an index number."(Fisher, 1922, pp.225-226)

Fisher's conclusion that the wholesale price index might be an adequate indicator of movements in the general price level seems to be motivated at least in part by the problem of coming up with a constant quality basket of goods to include in an index. As already noted, since the time when Fisher wrote, the CPI and the GDP deflator have replaced the WPI as the preferred measure of inflation.

But it is worth bearing in mind that while inflation will generally be

reflected in these indexes, it is not necessarily the case that increases in these indexes imply that inflation is occurring. It is possible to conceive of situations where an increase in the price of consumption goods relative to investment goods is consistent with a stable "price level" appropriately defined, even though the CPI is increasing. In a situation such as this the GDP deflator might give a better indication of whether the price level is rising or falling. But now consider a situation where the price of final goods transaction rises relative to the price of intermediate goods transactions, but again with a constant money stock. Here the GDP deflator (and the CPI) will fail to give an accurate picture of the underlying monetary reality.[58]

The implicit assumptions underlying the use of the average price of consumption transactions only or the average price of final goods transactions only as measures of the price level need to be spelled out and examined. It is interesting to note that early attempts to measure the price level tried to include the prices of more than just consumption transactions or final goods transactions - see, for example, Snyder (1924).

Friedman and Schwartz (1982) note the important difference that the choice of scale variable makes in the quantity theory approach:

> "[The income form of the quantity equation] is conceptually and empirically more satisfying than [the transactions form]. However [it] has the disadvantage that [it] completely neglects the ratio of intermediate to final transactions and transactions in existing capital assets.

---

[58]Alchian and Klein (1973) also criticize the use of the CPI and the GDP deflators as measures of the price level on the grounds that they do not include the prices of existing assets.

In the transactions version of the quantity equation, each intermediate transaction - that is, purchase by one enterprise from another - is included at the total value of the transaction, so that the value of wheat, for example, is included once when it is sold by the farmer to the mill; a second time when the mill sells flour to the baker, a third time when the baker sells bread to the grocer, a fourth time when the grocer sells bread to the consumer. In the income version, only the net value added by each of these transactions is included. To put it differently, in the transactions version the elementary event is an isolated exchange of a physical item for money - an actual, clearly observable event. In the income version, the elementary event is a hypothetical event that can be inferred from observation but is not directly observable. It is a complete series of transactions involving the exchange of productive services for final goods, via a sequence of money payments, with all the intermediate transactions in the income circuit netted out. The total value of all transactions is therefore a multiple of the value of income transactions only.

For a given flow of productive services, or alternatively, of final products (two of the multiple faces of income), the volume of transactions will be affected by vertical integration or disintegration of enterprises, which reduces or increases the number of transactions involved in a single income circuit, or by technological changes that lengthen or shorten the process of transforming productive services into final products. The volume of income will not be thus affected.

Similarly, the transactions version includes the purchase of an

existing asset - a house or a piece of land or a share of equity stock - precisely on a par with an intermediate or final transaction. The income version excludes such transactions completely.

Are these differences an advantage or disadvantage of the income version? That clearly depends on what it is that determines the amount of money that people want to hold. Do changes of the kind considered in the preceding paragraphs, changes that alter the ratio of intermediate and capital transactions to income, also alter in the same direction and by the same proportion the amount of money that people want to hold? Or do they tend to leave this amount unaltered? Or do they have a more complex effect?

Clearly, the transactions and income versions of the quantity theory involve very different conceptions of the role of money. For the transactions version, the most important thing about money is that it is transferred. For the income version, the most important thing is that it is held." (Friedman and Schwartz, 1982, pp. 23-24.)

Friedman (1987) seems to suggest that the shift away from the transactions approach to the income approach was due in part to the availability of income measures following the invention of national income accounting. This is somewhat ironic, in that while income measures are of value in and of themselves, they are a lot more difficult to construct than measures of gross transactions.

Emphasis on the transactions approach to price level measurement also suggests important distinctions between the price level of monetary theory and the cost of living in terms of how we ought to treat consumer durables. In

the theory of the cost of living index, the focus is on the flow of consumption services derived from a durable good, and the challenge is to somehow measure the value of the flow of services consumed each period. However, if we are only interested in the average prices paid in monetary market transactions, we would look only at the price of the durable good when it is purchased, and ignore the imputation of the service flows over the life of the good as these do not involve market transactions or more importantly money. In this sense the older treatment of housing in the CPI may have been more appropriate than the current treatment.[59]

As a practical matter we do not now, nor are we likely to have at any time in the near future, a measure of the average price of all transactions conducted in the economy. We are left then with the problem of deciding which of the existing price indexes best serves the purposes of monetary policy. The CPI has the advantage that it is well grounded in economic theory and is subjected to continuous scrutiny and refinement, in part because of its widespread use in contracts. It is also considered by many economists to be the best statistical series produced by the U.S. government.[60] The primary drawback of the CPI is of course that it only covers part of the transactions conducted in the economy by part of the population. In contrast the GDP deflator is more comprehensive, covering the full range of (final goods) transactions conducted by the entire population. Its prime drawback is that it is not a completely independent measure of the price level, being based on the CPI, the PPI and a variety of other price series of unknown quality. Both the CPI and the GDP deflator suffer from the inclusion of imputed transactions

---

[59]These ideas are explored further in Wynne (1993).

[60]See for example, Griliches (1992).

(specifically for the services of owner-occupied houses) that do not really belong in any index of the price level based on market exchange. The advantages and disadvantages of the different price indexes are summarized in Table 11.

Fortin (1990) advocates focusing on the CPI as the appropriate measure of inflation, on the grounds that "...the cardinal economic objective of society is to improve the standard of living of its members, which, in turn, come from the activity of real consumption, whose average price the CPI attempts to measure."(Fortin, 1990, p.113).

Poole (1992) argues that price stability ought to be defined in terms of goods prices, on the grounds that the problem of properly adjusting services prices for quality change is virtually impossible.[61] He cites the example of medical care prices to illustrate the nature of the problem. Obviously there have been enormous advances in medical technology that have led to improvements in the quality of medical care, and it is not clear that the prices that go into the CPI make adequate adjustments for these advances. A major reason for this is that it is very difficult to define the service that is being performed.[62] An additional complication is that health care is increasingly subject to government involvement, and insofar as this has been accompanied by attempts to force cost containment of private suppliers, some quality deterioration is likely, creating problems in the opposite direction. Poole argues that zero increase in the CPI or the PCE deflator would probably

---

[61]More specifically, the goods component of either the CPI or the PCE deflator.

[62]"If we cannot define the service in the first place, we cannot make sense of the question of whether quality is improving or deteriorating and we do not really know what the price indexes are measuring."(Poole, 1982, p. 55)

mean falling goods prices, which could cause problems if the rate of decline exceeded the real rate of interest consistent with full employment equilibrium (since nominal interest rates cannot go below zero). And experience suggests that the riskless real rate in the U.S. is about 2-3 percent. Poole concludes that economic activity is likely to be more stable if monetary policy yields stable goods prices rather than falling goods prices, which would in turn mean that price indexes that include services prices would be rising.

Our review of the evidence on the extent of quality bias in the CPI does not support the strong position taken by Poole. While the existence of upward bias in the prices of services can plausibly be conjectured, we simply do not have any strong evidence concerning its size or significance.

One issue that we did not address here has to do with the bias in the current CPI relative to the wealth-based definition of inflation or cost of living proposed by Alchian and Klein (1973). Alchian and Klein argue that the appropriate definition of the cost of living is in terms of the minimum cost of attaining a given lifetime flow of consumption services.[63] This approach to the measurement of inflation has received relatively little attention, primarily because it is almost impossible to implement empirically. One interesting implication of Alchian and Klein's argument, however, is that the rental equivalence approach to the treatment of consumer durables in cost of living indexes is inappropriate. A number of authors have explored other aspects of Alchian and Klein's argument, including Brown and Santoni (1981), Cukierman and Santoni (1987), Santoni and Moehring (1993).

---

[63]Pollak (1989) also contains a discussion of the theory of the intertemporal cost of living index.

7. Conclusions

What can we conclude from this review of the literature on measurement problems in the major price indexes? Is it the case, as some have claimed, that because of measurement problems, "price stability" is consistent with CPI inflation in the 0-2.0 percent range? Or is it the case that the various measurement biases probably cancel each other out, so that price stability really does mean no increase in the CPI or whatever price index is chosen as the preferred measure of inflation?

Lebow, Roberts and Stockton (1992) conclude their survey of measurement bias in the CPI by noting that under extreme assumptions, the upper bound on the measurement bias is about 1.8 percent a year:

> "We can combine all of the extreme assumptions in one example to get a rough upper bound on the measurement problem. Existing research suggests that commodity substitution bias is adding at most 0.2 percent annually to the CPI. The other calculations are more problematic. Our extreme estimate for retail substitution bias is 0.8 percent per year. For year-to-year quality adjustment difficulties we estimate an upper bound of 0.3 percent and for the lag in the introduction of new goods we estimate an upper limit of 0.5 percent per year. The total of these high-end estimates is 1.8 percent. Less extreme assumptions would lead to an estimate of around 1.0 percent, about midway between zero and 1.0 percent. Notions that this figure could be as high as the 3 to 4 percent inflation experienced over the past decade do not appear to be supportable by existing research." (Lebow, Stockton and Roberts, 1992, p.21)

Similar calculation by Crawford (1993) for the Canadian CPI lead him to

conclude that the upward bias is at most 0.5 percent a year.

While calculations of this sort are suggestive, it is important to consider what limitations and caveats surround them. Specifically, we need to ask whether calculations of this sort may give us figures on the overall bias in, say, the CPI that is too high because of double counting of some of the biases. Can we simply add together estimates of the quality adjustment bias and the new goods bias, given that the distinction between the two is elusive? Is it possible that traditional substitution bias and quality adjustment bias are also aspects of the same phenomenon? The same question can be raised for outlet-substitution bias recently quantified by Reinsdorf (1993): how do we disentangle this from other more traditional forms of quality and substitution bias? As Gordon (1990) notes in his study of producer durable prices, and Triplett emphasized in his 1975 and 1988 surveys, the problem is that many of the potential problems with the different price indexes are considered in isolation from one another, without any regard to the possible interaction between them.[64] This is not to criticize the calculations carried out by Lebow, Roberts and Stockton: such calculations are essential if any sort of conclusions are to be drawn about the potential biases in the most closely watched price index. However, it is important to be aware of, and hopefully at some point do something about, the limitations that surround calculations of this type.

A major problem that confronts us in trying to draw firm conclusions about the overall direction and magnitude of the biases that may exist in the official price indexes stems from the fact that research on the problems has

---

[64]Kokoski (1987) is one attempt to consider more than one of the traditional problems of price level measurement in a single study.

tended to be concentrated in the durable goods area. While there is a lot of evidence suggestive of significant upward bias in the prices of some durable goods, the irony is that there are serious doubts as to the adequacy of the current treatment of durables prices in price indexes such as the CPI. While the treatment of housing costs in the CPI now recognizes that the appropriate concept of consumption for a (static) cost of living index is the flow of services derived from the durable good, this approach has yet to be extended to other durables.

In the (almost) twenty years since Triplett's comprehensive 1975 survey there has been very little new research on the problems of price measurement, and certainly not enough, in our view, to warrant settling for 0 to 2.0 percent CPI inflation as being acceptable because of supposed measurement errors in the CPI. The most important single contribution to the field of price measurement in recent years is Gordon's (1990) study of producer durables prices, which also included some analysis of consumer durables prices. Other components of the PPI have also been studied in some detail by Griliches and various co-authors, especially pharmaceuticals and computers. For the CPI, the most significant recent studies have been those of Manser and McDonald (1988) on substitution bias and Reinsdorf (1993) on outlet substitution bias. Remarkably, there are no recent studies of quality bias for the nondurables and services components of the CPI. For example, we still have no sense of how large the potential bias is in the measurement of health care costs.

In view of the increased importance that accurate measurement of inflation takes on as we get closer to the goal of zero inflation, it might be worthwhile for the Federal Reserve System to sponsor more research on these

issues. The last such attempt (Griliches, 1971) was over twenty years ago, since when there have been substantial changes in how the CPI and PPI are constructed. It is also worth noting that the focus of the papers in the Griliches volume was relatively narrow, being primarily concerned with the issue of quality change and the application of hedonic methods to deal with it. The focus was further confined to durable goods, with little or no mention of nondurables, or more significantly, services.

To close, we repeat the conclusion of Foss, Manser, and Young (1993) that "...price research is an area where much remains to be done", especially in the area of price level measurement as it pertains to monetary policy.

Table 1
Studies of substitution bias in CPI

Study	Sample period	Level of disaggregation	Size of bias
Noe and von Furstenberg (1972)	1964-70	All	0.03 to 0.11
Christensen and Manser (1976)	1947-71	Meat and produce (4 categories of each)	0.1 to 0.2
Braithwait (1980)	1958-73	All	0.1
Manser and McDonald (1988)	1959-85	All	0.18
Bryan and Cecchetti (1993b)	1967-92	All	-0.1 to 1.2
Reinsdorf (1993)	1980-89	Food and gasoline	0.25 to 2.0

Notes to Table 1. Bias expressed in terms of the average value over the sample period of the difference between the Laspeyres index and the true cost of living, expressed as a fraction of the true cost of living.

Table 2 Recent Studies of quality bias in the CPI			
Study	Categories studied	Sample period	Estimated bias
Armknecht and Weyback (1988)	Women's coats & jackets, women's suits	1987–1988	-3.4 to 3.5
Randolph (1988)	Housing	1983	-0.3 to -0.4
Gordon (1990)	Durables	1947–83	1.54
Liegey (1993)	Women's coats and jackets, women's suits	1989	-1.3 to 6.0

Notes to Table 2. A negative bias means that the CPI understates the rate of inflation of the item.

Table 3
Drift of ratio of Tornqvist indexes
Gordons study and corresponding NIPA implicit deflators for selected consumer durables, 1982 base

	Annual growth rates			
	Full sample	1947-60	1960-73	1973-83
Motor vehicles and parts	-1.71	-2.39	-1.69	-0.85
Furniture and Household Equipment	-1.79	-2.52	-1.26	-1.55
-Kitchen and other household appliances	-3.22	-4.39	-2.37	-2.83
-Radios and TV's	-5.94	-9.07	-3.77	-4.69
Total Consumer Durables	-1.54	-2.21	-1.24	-1.05

Notes to Table 3. Source Gordon (1990) Table 1.2. The table shows the drift in the ratio of Gordon's alternative price indexes to the official price indexes for different periods. Thus negative numbers are interpreted as showing that the official series are upward biased, that is, they overstate inflation.

Table 4 Hypothetical example of new goods bias						
	Laspeyres		Paasche		Fisher	
f	p^3	Bias	p^3	Bias	p^3	Bias
0.01	0.9902	0.98%	0.9619	3.81%	0.9759	2.41%
0.02	0.9808	1.92%	0.9273	7.27%	0.9537	4.63%
0.05	0.9545	4.55%	0.8403	15.97%	0.8956	10.44%

Notes to **Table 4**. Source: Diewert (1987)

Table 5 Recent studies of bias in the PPI				
Problem	Study	Categories studied	Sample period	Estimated bias
Quality bias	Lichtenberg and Griliches (1989)	All	1972-77	0.3
	Gordon (1990)	All	1947-83	-2.96
	Trajtenberg (1990)	Computed Tomography Scanners	1973-82	na
	Siegel (1991)	All	1977-82	0.3*
	Berndt, Griliches and Rosett (1993)	Pharmaceuticals	1984-89	3.0
	Griliches and Cockburn (1993)	Pharmaceuticals	1986-90	na
List prices vs transactions prices	Betstock and Gerduk (1993)			

Notes to **Table 5**. Bias is stated in terms of percentage points per year. * See text for derivation of this figure.

Table 6 Comparing different price indexes for CT Scanners						
Year	Nominal Index		Hedonic		ΔW - based	
1973	10,000		10,000		10,000	
1974	11,940		10,770		800	
1975	12,000		6,130		240	
1976	14,450		4,600		31	
1977	17,450	100	3,850	100	15	100
1978	15,940	91	3,050	79	13	87
1979	16,610	95	2,780	72	11	73
1980	20,190	116	2,840	74	10	67
1981	24,840	142	3,020	78	8	53
1982	25,940	149	2,730	71	7	47

Notes to Table 6. Source: Trajtenberg (1990) Table 5.

Table 7 Response rates for different surveys			
	Frequency	Date	Response Rate
BLS: -PPI	Monthly	Late 1989	64
Census: -Retail sales	Monthly	August 1989	84
-Wholesale sales	Monthly	August 1989	83
-Private nonresidential construction	Monthly	1988 average	73

Notes to Table 7. Source: Foss (1993) Table 9.8.

Table 8
Drift of the ratio of Tornqvist indexes for NIPA categories, Gordon study and corresponding PPI's

NIPA Category	Annual growth rates			
	Full sample	1947-60	1960-73	1973-83
Office, computing, and accounting machinery	-9.32	-3.94	-16.61	-6.83
Communications Equipment	-5.84	-8.44	-2.89	-6.28
Instruments, photocopy and related equipment	-3.49	-3.18	-4.21	-2.97
Fabricated metal products	-1.80	-4.08	-1.28	0.49
Engines and Turbines	-3.53	-7.16	-2.27	-0.44
Metalworking machinery	-1.15	-3.01	0.58	-0.96
Special industry machinery, n.e.c.	-2.48	-3.70	-1.01	-2.79
General industrial equipment	-1.81	-2.87	-1.15	-1.29
Electrical transmission, distribution and industrial apparatus	-2.11	-3.62	-1.89	-0.43
Trucks, buses, and truck trailers	-2.97	-5.74	-2.04	-0.59
Autos	-1.33	-5.02	-0.27	2.09
Aircraft	-8.29	-12.69	-7.48	-3.63
Ships and boats	-1.93	-3.17	-1.11	-1.39
Railroad equipment	-1.45	-1.24	-1.43	-1.76
Furniture and fixtures	-1.41	-2.72	-0.84	-0.46
Tractors	-1.35	-0.05	-1.28	-3.14
Agricultural machinery	-0.70	-2.80	0.69	0.21
Construction machinery	-1.63	-2.35	-1.63	-0.68
Mining and oilfield machinery	-1.63	-2.35	-1.63	-0.68
Service industry machinery	-3.15	-4.06	-1.91	-3.59
Electrical equipment, n.e.c.	-1.01	-2.56	-0.09	-0.20
Other	-1.99	-3.90	-0.30	-1.69

Note to Table 8. Source Gordon (1990) Table 1.1

Table 9 Drift of the Ratio of Tornqvist Indexes Aggregates				
	Full sample	1947-60	1960-73	1973-83
Tornqvist	-2.96	-4.13	-2.44	-2.07
Implicit deflator, 1972 base	-2.90	-3.17	-1.88	-3.87
Implicit deflator, 1982 base	-1.97	-3.12	-1.20	-1.48

Notes to Table 9. Source: Gordon (1990), Table 1.1.

Table 10 Sources prices used to deflate components of GDP	
Component	Sources of deflators
Personal Consumption Expenditures	CPI, BEA computer price index, BEA index of prices received by farmers, Composite index of CPI's for foreign countries, BEA toll price index, BEA airline price index, BEA index of hospital input prices, Implicit deflator for stock brokerage charges, PPI, plus a number of interpolated series.
Nonresidential structures	PPI, Census Bureau price index of new houses, cost indexes from various trade and government agencies
Nonresidential producers' durable equipment	PPI, Department of Transportation index for aircraft, Maritime Administration index for ship prices, BEA computer price index, CPI,
Residential investment	CPI, PPI, price of new houses from Census Bureau.
Change in business inventories	PPI, CPI, Energy Information Administration crude petroleum index, BEA unit labor cost index
Net exports of goods and services	CPI, PPI, BLS import and export price indexes,
Government purchases of goods and services	CPI, PPI, BEA deflators for defense purchases, BEA computer price index,

Notes to Table x. Source U.S. Department of Commerce (1987, 1988, 1990).

Table 11
Comparison of the major price series

Price Series	Advantages	Disadvantages
Consumer Price Index	1. Strong conceptual basis. 2. Subject to intensive scrutiny due to its use in legal contracts. 3. Available on a timely basis.	1. Coverage limited to final consumption prices. 2. Represents prices paid by only 80 percent of the population. 3. Never revised. 4. Includes prices of imputed service flows. 5. Only available as a fixed weight index. 6. Includes changes in sales and indirect taxes.
Producer Price Index	1. Excludes difficult-to-measure service prices. 2. Excludes prices of imputed service flows. 3. Available on a timely basis. 4. Only one of the price series that has been subject to a comprehensive outside review (i.e. Gordon (1990)). 5. Excludes changes in excise taxes.	1. Lacks strong conceptual basis. 2. Coverage limited mainly to goods sector. 3. Never revised. 4. Only available as fixed weight index.
GDP Deflator	1. Most comprehensive coverage of all. 2. Available as both a fixed-weight and chain linked index. 3. Regularly revised on the basis of new information.	1. Only as good as the CPI and PPI data that deflate most of GDP. 2. Some components of deflator of unknown or dubious quality e.g. construction, defense. 3. Includes prices of imputed service flows. 4. Not available on a timely basis.

Bibliography

Afriat, Sidney N., 1977, The Price Index, Cambridge: Cambridge University Press.

Alchian, Armen A., and Benjamin Klein, 1973, On a correct measure of inflation, Journal of Money, Credit, and Banking, 5, 173-191.

Allen, R.G.D., 1975, Index Numbers in Theory and Practice, Chicago: Aldine Publishing Company.

Armknecht, Paul A., 1984, Quality adjustment in the CPI and methods to improve it, Journal of the American Statistical Association: Proceedings of the Business and Economic Statistics Section, ...

-----, and Donald Weyback, 1989, Adjustments for quality change in the U.S. consumer price index, Journal of Official Statistics, 5(2), 107-123.

-----, and Daniel H. Ginsburg, 1992, Improvements in measuring price changes in consumer services: past, present, and future, in Output Measurement in the Service Sectors, NBER Studies in Income and Wealth Volume 56, Zvi Griliches (ed), Chicago: University of Chicago Press.

Backman, Jules, and Martin R. Gainsbrugh, 1966, Inflation and the Price Indexes, U.S. Congress, Joint Economic Committee, Subcommittee on Economic Statistics, 89th Congress, 2nd Session, July. Washington DC: Government Printing Office.

Banerjee, Kali S., 1975, Cost of Living Index Numbers: Practice, Precision and Theory, New York: Marcel Dekker.

Berndt, Ernst R. and Jack Triplett, 1990, Fifty Year of Economic Measurement: The Jubilee of the Conference on Research in Income and Wealth, NBER Studies in Income and Wealth Volume 54, Chicago: University of Chicago Press.

-----, Zvi Griliches, and Joshua G. Rosett, 1993, Auditing the producer price index: Micro evidence from prescription pharmaceutical preparations, Journal of Business and Economic Statistics,

Betstock, Thomas, and Irwin B. Gerduk, 1993, The problem of list prices in the producer price index: the steel mill products case, in Price Measurements and Their Uses, NBER Studies in Income and Wealth Volume 57, Murray F. Foss, Marilyn E. Manser and Allan H. Young (eds), Chicago: University of Chicago Press.

Blinder, Alan S., 1980, The consumer price index and the measurement of recent inflation, Brookings Papers on Economic Activity, 539-565

Blisard, William Noel, and James R. Blaylock, 1991, Construction of true cost of food indexes from estimated Engel curves, American Journal of Agricultural

Economics, ,August, 775-783.

Braithwait, Steven D., 1980, The substitution bias of the Laspeyres Price Index, American Economic Review, 70, 64-77.

Brauer, David and Lucille Wu, 1991, An overview of inflation measurement, Federal Reserve Bank of New York Quarterly Review, Summer, 25-29.

Brown, W.W. and G.J. Santoni, 1983, Unreal estimates of the real rate of interest, Federal Reserve Bank of St. Louis Review, 16-25.

Bryan, Michael F., and Stephen G. Cecchetti, 1993a, Measuring core inflation, mimeo, Federal Reserve Bank of Cleveland, March.

-----, and -----, 1993b, The consumer price index as a measure of inflation, mimeo, Federal Reserve Bank of Cleveland, August 5.

Burns, Roger, Harry Briggs and William Thomas, 1993, Effect of updated weights on Producer Price Indexes, Monthly Labor Review, 116(3), March, 36-48.

Carlson, Keith M., 1989, Do price indexes tell us about inflation? A review of the issues, Federal Reserve Bank of St. Louis Economic Review, November/December, 12-30.

Cartwright, David W., 1986, Improved deflation of purchases of computers, Survey of Current Business, 66, March, 7-10.

-----, and Scott Smith, 1988, Deflators for purchases of computers in GNP: revised and extended estimates, Survey of Current Business, 68, November, 22-23.

Christensen, Laurits R. and Marilyn E. Manser, 1976, Cost-of-living indexes and price indexes for U.S. meat and produce, 1947-1971, in Nestor Terleckyj (ed) Household Production and Consumption, NBER Studies in Income and Wealth Volume 40, New York: Columbia University Press.

Cole, Rosanne, Y.C. Chen, Joan A. Barquin-Stolleman, Ellen Dulberger, Nurhan Helvacian, and James H. Hodge, 1986, Quality-adjusted price indexes for computer processors and selected peripheral equipment, Survey of Current Business, 66, January, 41-50.

Cox, Reavis, 1946, Non-price competition and the measurement of prices, The Journal of Marketing, X, April, 370-383.

Crawford, Allan, 1993, Measurement biases in the Canadian CPI: A technical note, Bank of Canada Review, Summer, 21-36.

Cukierman, Alex and G.J. Santoni, 1987, Some problems with current price indices, Working paper: Federal Reserve Bank of St. Louis.

Davis, Richard G., 1991, Inflation: measurement and policy issues, Federal Reserve Bank of New York Quarterly Review, Summer, 13-24.

Diewert, W. Erwin, 1976, Exact and superlative index numbers, Journal of Econometrics, 4, 114-45.

-----, 1987, Index numbers, in The New Palgrave: A Dictionary of Economics vol. 2 John Eatwell, Murray Milgate and Peter Newman (eds), London: Macmillan.

-----, 1988, The early history of price index research, NBER Working Paper No 2713.

-----, and Claude Montmarquette, 1983, Price Level Measurement: Proceedings From A Conference Sponsored By Statistics Canada, Ottawa: Statistics Canada.

Domberger, Simon, 1987, Relative price variability and inflation: a disaggregated analysis, Journal of Political Economy, 93(3), June, 547-566.

Dow, James P., Jr, 1993, Measuring inflation using multiple price indexes, mimeo, March, paper presented at the 1993 Meetings of the Western Economic Association.

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

Fisher, Irving, 1920, The Purchasing Power of Money: Its Determination and Relation to Credit, Interest and Crises, (second edition) New York: Macmillan.

Flueck, John A., 1961, A study in validity: BLS wholesale price quotations, Staff Paper No. 9 in The Price Statistics of the Federal Government: Review, Appraisal, Recommendations, NBER, New York: NBER.

Fortin, Pierre, 1990, Do we measure inflation correctly?, in Zero Inflation: The Goal of Price Stability, Richard G. Lipsey (ed), Ottawa: C.D. Howe Institute.

Foss, Murray, 1983, The U.S. National Income and Product Accounts: Selected Topics, NBER Studies in Income and Wealth Volume 47, Chicago: University of Chicago Press.

-----, 1993, Does government regulation inhibit the reporting of transactions prices by business?, in Price Measurements and Their Uses, Murray F. Foss, Marilyn Manser and Allan H. Young (eds) NBER Studies in Income and Wealth Volume 57, Chicago: University of Chicago Press.

-----, Marilyn Manser, and Allan H. Young, 1993, Price Measurements and Their Uses, NBER Studies in Income and Wealth Volume 57, Chicago: University of Chicago Press.

Francois, Joseph F., 1989, Estimating homeownership costs: owner's estimates of implicit rents and the relative importance of rental equivalence in the consumer price index, AREUEA Journal, 17(1), 87-99.

Friedman, Milton, 1987, The quantity theory of money, in The New Palgrave: A Dictionary of Economics, John Eatwell, Murray Milgate, and Peter Newman (eds),

London: Macmillan.

-----, and Anna J. Schwartz, 1982, Monetary Trends in the United States and the United Kingdom: Their Relation to Income, Prices, and Interest Rates, 1867-1975, Chicago: University of Chicago Press.

Gordon, Robert J., 1971, Measurement bias in price indexes for capital goods, Review of Income and Wealth, 17(2), June, 121-74.

-----, 1981, The Consumer Price Index: measuring inflation and causing it, The Public Interest, , Spring, 112-134.

-----, 1990, The Measurement of Durable Goods Prices, Chicago: University of Chicago Press.

-----, 1992, Measuring the aggregate price level: implications for economic performance and policy, NBER Working Paper No. 3969.

Gousen, Sarah, Kathy Monk, and Irwin Gerduk, 1986, Producer Price Measurement: Concepts and Methods, U.S. Department of Labor: Bureau of Labor Statistics.

Griliches, Zvi, 1961, Hedonic price indexes for automobiles: an econometric analysis of quality change, in The Price Statistics of the Federal Government, NBER, New York: National Bureau of Economic Research, reprinted in Griliches 1971.

-----, (ed), 1971, Price Indexes and Quality Change: Studies in New Methods of Measurement, Cambridge MA: Harvard University Press.

-----, 1992a, Output Measurement in the Service Sectors, NBER Studies in Income and Wealth Volume 56, Chicago: University of Chicago Press.

-----, 1992b, Introduction, in Output Measurement in the Service Sectors, NBER Studies in Income and Wealth Volume 56, Chicago: University of Chicago Press.

-----, and Iain Cockburn, 1993, Generics and new drugs in pharmaceutical price indexes, NBER Working Paper No. 4272, February.

Hicks, John R., 1940, The valuation of social income, Economica, 7, 105-24.

Kipnis, Gregory, 1966, Implicit Price Index (IPI), Appendix C of Backman and Gainsbrugh (1966).

Kokoski, Mary F., 1987, Problems in the measurement of consumer cost-of-living indexes, Journal of Business and Economic Statistics, 5(1), 39-46.

-----, 1989, Experimental cost-of-living indexes: a summary of current research, Monthly Labor Review, , July, 34-39.

Lane, Walter F., William C. Randolph, and Stephen A. Berenson, 1988, Adjusting the CPI shelter index to compensate for effect of depreciation, Monthly Labor Review, 111(10), 34-37.

Layard, P.R.G., and A.A. Walters, 1978, Microeconomic Theory, London: McGraw Hill.

Lebow, David E., John M. Roberts, and David J. Stockton, 1992, Understanding the goal of price stability, mimeo, Board of Governors of the Federal Reserve System, October.

Lichtenberg, Frank R., and Zvi Griliches, 1989, Errors of measurement in output deflators, Journal of Business and Economic Statistics, 7(1), 1-9.

Liegey, Paul F., Jr., 1993, Adjusting apparel indexes in the consumer price index for quality differences, in Price Measurements and Their Uses, NBER Studies in Income and Wealth Volume 57, Murray F. Foss, Marilyn E. Manser and Allan H. Young (eds), Chicago: University of Chicago Press.

Madigan, Kathleen, 1991, How reliable is the consumer price index?, Business Week, April 29, 70-71.

Manser, Marilyn E., and Richard McDonald, 1988, An analysis of substitution bias in measuring inflation, 1959-1985, Econometrica, 56, 909-930.

McAllister, Harry E., 1961, Statistical factors affecting the stability of the wholesale and consumer price indexes, Staff Paper 8 in Price Statistics of the Federal Government: Review, Appraisal, and Recommendations, NBER, New York: NBER.

Morgenstern, Oskar, 1931, Free and fixed prices in the Depression, Harvard Business Review, X, 62-68 (October).

-----, 1963, On the Accuracy of Economic Observations, (second edition), Princeton: Princeton University Press.

Motley, Brian, 1992, Index numbers and the measurement of real GDP, Federal Reserve Bank of San Francisco Economic Review, ,3-13.

NBER, 1961, The Price Statistics of the Federal Government: Review, Appraisal, and Recommendations, General Research Series 73, New York: National Bureau of Economic Research.

Noe, Nicholas N., and George M. von Furstenberg, 1972, The upward bias in the consumer price index due to substitution, Journal of Political Economy, 80, 1280-1286.

Pieper. Paul E., 1989, Construction price statistics revisited, in Technology and Capital Formation, Dale W. Jorgenson and Ralph Landau (eds), Cambridge MA: MIT Press.

-----, 1990, The measurement of construction prices: retrospect and prospect, in Fifty Years of Economic Measurement: The Jubilee of the Conference on Research in Income and Wealth, NBER Studies in Income and Wealth Volume 54, Ernst R. Berndt and Jack E. Triplett (eds), Chicago: University of Chicago Press.

Pollak, Robert A., 1989, The Theory of the Cost of Living Index, New York: Oxford University Press.

Pollin, Robert, and Michael Stone, 1991, The illusion of an improved CPI, Challenge, 30(1), January/February, 53-57.

Poole, William, 1992, Where do we stand in the battle against inflation?, Shadow Open Market Committee Policy Statement and Position Papers, March.

Popkin, Joel, 1993, Discussion of The effect of outlet price differentials on the U.S. Consumer Price Index by Marshall Reinsdorf, in Price Measurements and Their Uses NBER Studies in Income and Wealth Volume 57, Murray F. Foss, Marilyn E. Manser and Allan H. Young (eds), Chicago: University of Chicago Press.

Randolph, William C., 1988, Housing depreciation and aging bias in the consumer price index, Journal of Business and Economic Statistics, volume 6, number 3, July, 359-371.

Reinsdorf, Marshall, 1994, The effect of price dispersion on cost of living indexes, International Economic Review, forthcoming.

Reinsdorf, M., 1993, The effect of outlet price differentials on the U.S. consumer price index, in Price Measurements and Their Uses, NBER Studies in Income and Wealth Volume 57, Murray F. Foss, Marilyn E. Manser and Allan H. Young (eds), Chicago: University of Chicago Press.

Rogers, R. Mark, Steven W. Henderson, and Daniel H. Ginsburg, 1993, Consumer prices: examining housing rental components, Federal Reserve Bank of Atlanta Economic Review, 78(3), May/June, 32-46.

Santoni, Gary J., and H. Brian Moehring, 1993, Asset returns and measured inflation, mimeo, Ball State University.

Selody, Jack, 1990, The goal of price stability: a review of the issues, Bank of Canada Technical Report No. 54, May.

Siegel, Donald, 1991, Errors in output deflators revisited: unit values and the PPI, NBER Working Paper No. 3935.

Sinclair, James, and Brian Catron, 1990, An experimental price index for the computer industry, Monthly Labor Review, , October, 16-24.

Snyder, Carl, 1924, A new index of the general price level from 1875, Journal of the American Statistical Association, XIX, June, 189-195.

Stigler, George J., and James Kindahl, 1970, The Behavior of Industrial Prices, New York: Columbia University Press (for the National Bureau of Economic Research).

Tregarthen, Suzanne, 1993, Statistics overstate health care costs, Wall Street

Journal, Wednesday, August 18, A10.

Triplett, Jack E., 1971, Quality bias in price indexes and new methods of quality measurement, in Price Indexes and Quality Change: Studies in New Methods of Measurement, Zvi Griliches (ed), Cambridge MA: Harvard University Press.

-----, 1975, The measurement of inflation: a survey of research on the accuracy of price indexes, in Analysis of Inflation, Paul H. Earl (ed), Lexington: Lexington Books.

-----, 1981, Reconciling the CPI and the PCE deflator, Monthly Labor Review, , September, 3-15.

-----, 1986, The economic interpretation of hedonic methods, Survey of Current Business, 64(1), January, 36-40.

-----, 1987, Hedonic functions and hedonic indexes, in The New Palgrave: A Dictionary of Economics, John Eatwell, Peter Newman and Murray Milgate (eds), London: Macmillan.

-----, 1988, Price index research and its influence on data: a historical review, paper presented to the 50th anniversary conference of the Conference on Research on Income and Wealth.

-----, 1989, Price and technological change in a capital good: a survey of research on computers, in Technology and Capital Formation, Dale W. Jorgenson and Ralph Landau (eds), Cambridge: MIT Press.

-----, 1990, Hedonic methods in statistical agency environments: an intellectual biopsy, in Fifty Years of Economic Measurement: The Jubilee of the Conference on Research in Income and Wealth, NBER Studies in Income and wealth Volume 54, Ernst R. Berndt and Jack E. Triplett (eds), Chicago: University of Chicago Press.

-----, 1992, Economic theory and the BEA's alternative quantity and price measures, Survey of Current Business, , April, 49-52.

-----, 1993a, Review of Robert J. Gordon "The Measurement of Durable Goods Prices", Journal of Economic Literature, XXXI, March, 244-246.

-----, 1993b, Comment on papers by Dulberger, Norsworthy and Jang, and Flamm, in Price Measurements and Their Uses, NBER Studies in Income and Wealth Volume 57, Murray F. Foss, Marilyn E. Manser, and Allan H. Young (eds), Chicago: University of Chicago Press.

-----, and Richard J. McDonald, 1977, Assessing the quality error in output measures: the case of refrigerators, Review of Income and Wealth, 23, 137-56.

Tullock, Gordon, 1979, When is inflation not inflation?, Journal of Money, Credit, and Banking, 11(2), May, 219-221.

Turvey, Ralph, et. al., 1989, Consumer Price Indices: An ILO Manual, International Labor Office: Geneva.

U.S. Department of Commerce Bureau of Economic Analysis, 1987, GNP: An Overview of Source Data and Estimating Methods, Methodology Paper Series MP-4, Washington, DC: U.S. Government Printing Office.

-----. 1988, Government Transactions, Methodology Paper Series MP-5, Washington, DC: Government Printing Office.

-----, 1990, Personal Consumption Expenditures, Methodology Paper Series MP-6, Washington, DC: Government Printing Office.

U.S. Department of Labor, 1987, The Consumer Price Index: 1987 Revision, (Bureau of labor Statistics report 736), Washington DC: Government Printing Office.

U.S. Department of Labor, 1992, BLS Handbook of Methods, (Bureau of Labor Statistics Bulletin 2414) Washington DC: Government Printing Office (September).

U.S. Executive Office of the President: Council on Wage and Price Stability, 1977, The Wholesale Price Index: Review and Evaluation, Washington DC: June.

Wahl, Richard W., 1982, Is the CPI a fair measure of inflation?, Journal of Policy Analysis and Management, 1(4), 496-511.

Webb, Roy H., and Rob Willemse, 1989, Macroeconomic price indexes, Federal Reserve Bank of Richmond Economic Review, July/August, 22-32.

Wallace, William H., and William E. Cullison, 1981, Measuring Price Changes: A Study of the Price Indexes, (fourth edition), Richmond: Federal Reserve Bank of Richmond.

Wynne, Mark A., 1993, Measuring inflation, manuscript, Federal Reserve Bank of Dallas.

Young, Allan H., 1989, Alternative measures of real GNP, Survey of Current Business, 69(4), April, 27-34.

-----, 1992, Alternative measures of change in real output and prices, Survey of Current Business, 72(4), April, 32-48.

-----, 1993, Alternative measures of change in real output and prices, quarterly estimates for 1959-92, Survey of Current Business, 73(3), March, 31-41.

Ziemer, Richard C., and Pamela A. Kelly, 1993, The deflation of military aircraft, in Price Measurements and Their Uses, NBER Studies in Income and Wealth Volume 57, Murray F. Foss, Marilyn E. Manser, and Allan H. Young (eds), Chicago: University of Chicago Press.

RESEARCH PAPERS OF THE RESEARCH DEPARTMENT
FEDERAL RESERVE BANK OF DALLAS

Available, at no charge, from the Research Department
Federal Reserve Bank of Dallas, P.O. Box 655906
Dallas, Texas 75265-5906

9201 Are Deep Recessions Followed by Strong Recoveries? (Mark A. Wynne and Nathan S. Balke)

9202 The Case of the "Missing M2" (John V. Duca)

9203 Immigrant Links to the Home Country: Implications for Trade, Welfare and Factor Rewards (David M. Gould)

9204 Does Aggregate Output Have a Unit Root? (Mark A. Wynne)

9205 Inflation and Its Variability: A Note (Kenneth M. Emery)

9206 Budget Constrained Frontier Measures of Fiscal Equality and Efficiency in Schooling (Shawna Grosskopf, Kathy Hayes, Lori Taylor, William Weber)

9207 The Effects of Credit Availability, Nonbank Competition, and Tax Reform on Bank Consumer Lending (John V. Duca and Bonnie Garrett)

9208 On the Future Erosion of the North American Free Trade Agreement (William C. Gruben)

9209 Threshold Cointegration (Nathan S. Balke and Thomas B. Fomby)

9210 Cointegration and Tests of a Classical Model of Inflation in Argentina, Bolivia, Brazil, Mexico, and Peru (Raúl Anibal Feliz and John H. Welch)

9211 Nominal Feedback Rules for Monetary Policy: Some Comments (Evan F. Koenig)

9212 The Analysis of Fiscal Policy in Neoclassical Models' (Mark Wynne)

9213 Measuring the Value of School Quality (Lori Taylor)

9214 Forecasting Turning Points: Is a Two-State Characterization of the Business Cycle Appropriate? (Kenneth M. Emery & Evan F. Koenig)

9215 Energy Security: A Comparison of Protectionist Policies (Mine K. Yücel and Carol Dahl)

9216 An Analysis of the Impact of Two Fiscal Policies on the Behavior of a Dynamic Asset Market (Gregory W. Huffman)

9301 Human Capital Externalities, Trade, and Economic Growth (David Gould and Roy J. Ruffin)

9302 The New Face of Latin America: Financial Flows, Markets, and Institutions in the 1990s (John Welch)

9303 A General Two Sector Model of Endogenous Growth with Human and Physical Capital (Eric Bond, Ping Wang, and Chong K. Yip)

9304 The Political Economy of School Reform (S. Grosskopf, K. Hayes, L. Taylor, and W. Weber)

9305 Money, Output, and Income Velocity (Theodore Palivos and Ping Wang)

9306 Constructing an Alternative Measure of Changes in Reserve Requirement Ratios (Joseph H. Haslag and Scott E. Hein)

9307 Money Demand and Relative Prices During Episodes of Hyperinflation (Ellis W. Tallman and Ping Wang)

9308 On Quantity Theory Restrictions and the Signalling Value of the Money Multiplier (Joseph Haslag)

9309 The Algebra of Price Stability (Nathan S. Balke and Kenneth M. Emery)

9310 Does It Matter How Monetary Policy is Implemented? (Joseph H. Haslag and Scott E. Hein)

9311 Real Effects of Money and Welfare Costs of Inflation in an Endogenously Growing Economy with Transactions Costs (Ping Wang and Chong K. Yip)

9312 Borrowing Constraints, Household Debt, and Racial Discrimination in Loan Markets (John V. Duca and Stuart Rosenthal)

9313 Default Risk, Dollarization, and Currency Substitution in Mexico (William Gruben and John Welch)

9314 Technological Unemployment (W. Michael Cox)

9315 Output, Inflation, and Stabilization in a Small Open Economy: Evidence From Mexico (John H. Rogers and Ping Wang)

9316 Price Stabilization, Output Stabilization and Coordinated Monetary Policy Actions (Joseph H. Haslag)

9317 An Alternative Neo-Classical Growth Model with Closed-Form Decision Rules (Gregory W. Huffman)

9318 Why the Composite Index of Leading Indicators Doesn't Lead (Evan F. Koenig and Kenneth M. Emery)

9319 Allocative Inefficiency and Local Government: Evidence Rejecting the Tiebout Hypothesis (Lori L. Taylor)

9320 The Output Effects of Government Consumption: A Note (Mark A. Wynne)

9321 Should Bond Funds be Included in M2? (John V. Duca)

9322 Recessions and Recoveries in Real Business Cycle Models: Do Real Business Cycle Models Generate Cyclical Behavior? (Mark A. Wynne)

9323* Retaliation, Liberalization, and Trade Wars: The Political Economy of Nonstrategic Trade Policy (David M. Gould and Graeme L. Woodbridge)

9324 A General Two-Sector Model of Endogenous Growth with Human and Physical Capital: Balanced Growth and Transitional Dynamics (Eric W. Bond, Ping Wang, and Chong K. Yip)

9325 Growth and Equity with Endogenous Human Capital: Taiwan's Economic Miracle Revisited (Maw-Lin Lee, Ben-Chieh Liu, and Ping Wang)

9326 Clearinghouse Banks and Banknote Over-issue (Scott Freeman)

9327 Coal, Natural Gas and Oil Markets after World War II: What's Old, What's New? (Mine K. Yücel and Shengyi Guo)

9328 On the Optimality of Interest-Bearing Reserves in Economies of Overlapping Generations (Scott Freeman and Joseph Haslag)

9329* Retaliation, Liberalization, and Trade Wars: The Political Economy of Nonstrategic Trade Policy (David M. Gould and Graeme L. Woodbridge) Reprint of 9323

9330 On the Existence of Nonoptimal Equilibria in Dynamic Stochastic Economies (Jeremy Greenwood and Gregory W. Huffman)

9331 The Credibility and Performance of Unilateral Target Zones: A Comparison of the Mexican and Chilean Cases (Raul A. Feliz and John H. Welch)

9332 Endogenous Growth and International Trade (Roy J. Ruffin)

9333 Wealth Effects, Heterogeneity and Dynamic Fiscal Policy (Zsolt Becsi)

9334 The Inefficiency of Seigniorage from Required Reserves (Scott Freeman)

9335 Problems of Testing Fiscal Solvency in High Inflation Economies: Evidence from Argentina, Brazil, and Mexico (John H. Welch)

9336 Income Taxes as Reciprocal Tariffs (W. Michael Cox, David M. Gould, and Roy J. Ruffin)

9337 Assessing the Economic Cost of Unilateral Oil Conservation (Stephen P.A. Brown and Hillard G. Huntington)

9338 Exchange Rate Uncertainty and Economic Growth in Latin America (Darryl McLeod and John H. Welch)

9339 Searching for a Stable M2-Demand Equation (Evan F. Koenig)

9340 A Survey of Measurement Biases in Price Indexes (Mark A. Wynne and Fiona Sigalla)