Improvements to the Food at Home, Shelter and Prescription Drug Indexes in the U.S. Consumer Price Index

Paul A. Armknecht, Brent R. Moulton and Kenneth J. Stewart

Introduction

A cost-of-living index (CLI) is used as a conceptual framework for dealing with practical questions that arise in the construction of the Consumer Price Index (CPI) of the U.S. Bureau of Labor Statistics (BLS) (*BLS Handbook of Methods, 1992*). While the Laspeyres approach used in the U.S. CPI can provide an approximation to a CLI, measuring the average change in the prices paid by urban consumers for a fixed market basket of goods and services has many limitations when interpreted relative to a true cost-of-living index. For example, consumers shift spending patterns in response to changes in relative prices, items and outlets available in the original or base period disappear, and new items and outlets enter the marketplace. To alleviate some of these problems, the U.S. CPI uses a modified Laspeyres approach, which allows for product substitution and the introduction of new samples of outlets and items.[1]

The universe of consumer goods and services in the U.S. CPI is divided into 207 component item strata and 44 geographic areas,[2] stratifying the CPI into 9108 basic item-area components (207 item components x 44 areas). A two-tiered weighting system is used to calculate the CPI. At the first or lowest level, within each basic item-area component, price relatives are calculated using a weighted average of price change for sampled items. At the second or higher level of aggregation, those price relatives are used to update expenditures across items and geographic areas. At each level, we periodically update the samples and/or expenditure weights to keep them representative of consumer purchases.

Within basic item-area components, the CPI has used a Point-of-Purchase Survey (POPS) since 1978 to provide the sampling frame of outlets for most goods and services. Item and outlet samples are scientifically selected using probability proportionate to sales. To keep CPI samples representative of consumer purchases, the POPS survey is conducted annually in approximately twenty percent of the Primary Sampling Units (or urban areas) in the CPI, where new outlets are selected to replace the old samples. (*BLS Handbook of Methods, 1992*).

At the second or higher level of weighting, base period expenditure weights for each basic item-area component of the CPI are derived from the Consumer Expenditure (CE) Surveys. The index (and thus the expenditure weight) of each component item-area is effectively updated each month by its basic component price relative, and the updated expenditure weights are aggregated across items and geographic areas. To keep these expenditure weights representative of consumer purchases, CE weights are updated every ten years or so. The present CPI is based on 1982-84 CE weights; in January 1998, we are planning to introduce 1993-95 CE weights into the CPI.

The effect of consumers shifting spending patterns in response to price change at the aggregate level in the U.S. CPI has been investigated at length. The "substitution bias" caused by shifting purchasing patterns has been measured at between 0.1% and 0.22% per year (Braithwait, 1980; Manser and McDonald, 1988; Aizcorbe and Jackman, 1993). A portion of these substitution effects could be eliminated with more frequent updating of the CPI expenditure weights (Schmidt, 1993).

The potential effects of substitutions by consumers at the lowest level of estimation within item and area has been explored recently by Armknecht (1993), Moulton (1993), Reinsdorf (1993 and 1994), and Reinsdorf and Moulton (1994). Much of this research examines the assumptions of Leontief preferences in which consumers choose to purchase fixed quantities of goods and services over time, versus Cobb-Douglas preferences in which consumers will substitute goods and services across outlets but keep expenditure shares constant. Leontief preferences imply no substitutions and that consumers have demand elasticities equal to zero. With Cobb-Douglas preferences consumers' demand elasticities are assumed to be unitary. The Leontief or fixed quantity preferences are consistent with the Laspeyres price index for the CLI, while the Cobb-Douglas or constant expenditure shares preferences are consistent with the geometric mean form of the Laspeyres concept for the CLI.

While researching in this area of the appropriate functional form of the estimator for the basic indexes or elementary aggregate, Reinsdorf (1994) has uncovered a systematic bias with the estimator used in the U.S. CPI when new samples are introduced. Moulton (1993) summarizes the type of estimation bias inherent in the U.S. CPI at the first or micro-level of estimation -- the calculation of basic item-area price relatives -- caused by the introduction of new outlet and item samples into the index. Further investigation of the functional form bias has uncovered an analogous problem with the estimator used for imputation of implicit rent changes for owner occupied housing.

Another potential source of bias in the U.S. CPI is the treatment of brand name prescription drugs after their patent expires. Griliches and Cockburn (1993) present evidence that there is substantial substitution of generic drugs for brand drugs when patents expire. The U.S. CPI does not pick up such substitution in a timely manner. To address this situation, BLS will institute new substitution procedures for the prescription drug component of the CPI.

This paper summarizes BLS research into the causes of these biases, presents estimates of their magnitude, and discusses actions to reduce these effects in the U.S. CPI. In section I we discuss the sample rotation problem and alternative solutions, including the one we are implementing. In section II we demonstrate the shelter estimation biases and present the new estimators we plan to introduce. In section III we discuss the new procedures that are being introduced to improve our ability to capture consumers switching to generic drugs as patents on brand drugs expire. In section IV we present a summary and suggestions for future research.

I. Sample rotation in the U.S. Consumer Price Index

Within most basic item-area components, expenditure estimates for each sampled item are derived from the POPS. Retail outlets are also drawn from the POPS, and are introduced into the index usually a year or two after the POPS survey is conducted.

A typical timeline:

Month	1	4	8	12	16	20	24
	Month 1: POPS survey taken to estimate EX_b					Month 20: Initiation prices for new sample collected	Month 24 : Link month prices collected

Following the methodology presented by Moulton (1993) and others, the price index for each item-area component is updated each month by a price relative, as in

$$Index_t = Index_{t-1} \times R_{t,t-1}$$

Ideally, $R_{t,t-1}$ is a Laspeyres-based estimate using the ratio of the current cost of purchasing the base period quantities in t to the cost in the previous period t-1,

$$R_{t,t-1} = \frac{\sum P_t Q_b}{\sum P_{t-1} Q_b}$$

where P is the price of an item and Q_b its base period quantity.

In the U.S. CPI, we do not have base period quantities (Q_b); we only have estimates of base period expenditures (EX_b). The Laspeyres-based formula can be rewritten in terms of expenditures using the relation:

$$Q_b = (EX_b / P_b)$$

Expenditure information is then used to draw the sample of specific items. Because the sample is drawn after the base period, the base-period prices are not known. A proxy base price ($\hat{P}_b$) for each selected item is estimated by dividing the link month price of the item by the price change of the component item-area relative from the POPS reference period to the link month, as in

$$\hat{P}_b = \frac{I_b}{I_{link}} P_{link},$$

where I_b is the component index value in the POPS base period and I_{link} is the index value at link month.

The Laspeyres formula from above

$$R_{t,t-1} = \frac{\sum P_t Q_b}{\sum P_{t-1} Q_b}$$

is then rewritten as:

$$R_{t,t-1} = \frac{\sum P_t(EX_b / \hat{P}_b)}{\sum P_{t-1}(EX_b / \hat{P}_b)}$$

Again, after the new sample has been linked into the index, price relatives are chained together to effectively update fixed-quantity base-period expenditures to the current period. In the first month after the new sample is introduced then, the price relative is

$$R_{link+1,link} = \frac{\sum P_{link+1}(EX_b / \hat{P}_b)}{\sum P_{link}(EX_b / \hat{P}_b)}$$

Therefore,

$$R_{link+1,link} = \frac{\sum P_{link+1}[EX_b / (X * P_{link})]}{\sum P_{link}[EX_b / (X * P_{link})]}$$

where $X = I_b / I_{link}$, the index change for that item-area from the link month back to the POPS period. This can be further simplified to :

$$R_{link+1,link} = \frac{\sum [P_{link+1} / P_{link}][EX_b / X]}{\sum EX_b / X}$$

Reinsdorf (1993) noted that price indexes were rising at significantly faster rates than average prices in the U.S. CPI, particularly for food items. Most of the divergence between indexes and average prices can probably be attributed to a bias in the functional form for estimating price change at the basic component index level, as noted in a number of recent articles and papers (Armknecht, 1993; Moulton, 1993; Reinsdorf, 1994). Related discussions of functional form bias in basic component indexes also appear in Dalén (1992), Sculz (1983), and Reinsdorf and Moulton (1994).

This occurs because the current method of setting base period prices in the U.S. CPI often creates a significant positive correlation between the weight of an item priced and its price change immediately after introducing the new sample. Forsyth and Fowler (1981), Szulc (1989) and Dalén (1992) also note that the "bouncing" of prices from the base expenditure period to their introduction into the sample often results in a correlation between the weight of an item introduced into the sample and its subsequent price change. This can occur with seasonal prices (e.g., strawberries), sale prices, as well as when product substitutions occur (e.g., apparel). Szulc (1983) and Reinsdorf (1994) also note that the effect is particularly severe when indexes are chained, as they are in the U.S. CPI.

As seen from the equation above, in the U.S. CPI, newly sampled items that are on sale or have unusually low prices at sample rotation will tend to have unusually low base period prices, and thus unusually high implicit quantity weights. These items are likely to go off sale the next period, so that a high weight is given to items that are increasing in price. Since the weights are inversely proportional to the starting price P_{link}, Reinsdorf (1994) notes that this estimator has properties similar to that of a Sauerbeck index immediately after the new sample is linked into the index. A Sauerbeck index is derived from the unweighted average of price relatives.

For example, presume a set of banana prices where the average price in an area is always $1.00, but prices within a specific outlet can vary between $0.50 to $1.50.
If a new sample of two equally weighted banana quotes in that area is introduced into the CPI where

	link month	link month + 1
Banana #1	$1.50	$1.00
Banana #2	$0.50	$1.00

then the price relative in link month +1 would be

$$\frac{(1.00/1.50+1.00/0.50)}{(1.50/1.50+0.50/0.50)}=\frac{0.67+2.0}{2}=1.33$$

Even without the correlation between new weights and subsequent price change, this estimator fails the permutation test suggested by Dalén (1992).

Alternatives for alleviating sample rotation bias in the U.S. CPI

Many proposals have been forwarded to alleviate the bias caused by the rotation of new item and outlet samples in the U.S. CPI. If we had unlimited resources, we could expand our surveys to get more timely information on expenditures or even actual quantities sold. These are options that will be explored in the future for specific commodities as we examine data available from private market research groups. In the interim, there are three ways that have been systematically investigated in which the current bias in the CPI sample rotation process may be alleviated:

1) using geometric means to calculate basic item-area price relatives;
2) setting base period prices using pre-link month "initiation" prices; and
3) pricing both the old and new samples for a period of time before introducing the new samples into the CPI.

Using geometric means to calculate basic item-area price relatives.

Price relatives constructed using geometric means have many desirable statistical properties (Dalén, 1992; Turvey, 1989; Szulc, 1989) and, under certain conditions, can be demonstrated to be superior to Laspeyres when approximating cost-of-living indexes (Moulton, 1993).

In Moulton (1993), most item categories within the CPI were recalculated from June 1992 through June 1993 using geometric means to calculate basic item-area price relatives. Mathematically, the geometric mean price relative is:

$$R_{t,t-1} = \prod (P_t/P_{t-1})^{S_b}$$

where S_b is the base-period expenditure share for that item. Since items are effectively weighted by expenditure shares, it is not necessary to estimate (or misestimate) base period quantities. Unlike the true form of the Laspeyres concept, where base period quantities are fixed, here relative base period expenditure shares are fixed.

For items included in the simulation by Moulton, the simulated Laspeyres arithmetic mean CPI went up 2.95 percent; the index using geometric means rose 2.48 percent. The differences between the two indexes were most striking for food and apparel items.[3]

It is the intention of the Bureau of Labor Statistics to create experimental indexes for the U.S. CPI using geometric means in calculating basic item-area price relatives starting sometime in 1996. These test indexes will be used to generate further research. We are hopeful that a parallel production-quality index using geometric means could be published on a monthly basis late in the year 2000.

In addition, the U.S. CPI is seeking additional research into the conceptual underpinnings associated with using geometric means at the basic component level. At a recent conference sponsored by the National Bureau of Economic Research on the economics of new products, this approach to estimation was raised and received considerable encouragement for future research efforts.

Using initiation prices to set base period prices.

Changes have been suggested in our current estimator to create an independence between the price used to set a base price and the price change following the introduction of a new sample. One potential way to reduce the correlation between the link month price and the base-period price would be to set the base price by taking the *initiation price,* usually observed several months before the sample is linked, and rebasing it to the POPS year. The initiation price for a quote is typically collected about three to five months before link month, but approximately one-sixth of quotes are initiated during link month (see table 1).

Table 1. Number of months between initiation price and link month price, drawn from quotes in the April 1993 CPI database.

Number of Months	Percent
0	17.2
1	0.9
2	2.3
3	26.3
4	26.5
5	18.1
6	5.9
7 or more	2.9

In this approach, the base period price would be defined as

$$\hat{P}_b = X_z P_{init},$$

where $X_z = I_b / I_{init}$, and the price relative would become

$$R_{link+1,\,link} = \frac{\sum P_{link+1}[EX_b / (X_z * P_{init})]}{\sum P_{link}[EX_b / (X_z * P_{init})]}$$

To test the impact on the CPI of setting base period prices on initiation prices, indexes were recalculated for most item categories using the new base period prices (see Table 2).

Table 2. Percent change from May 1993-May 1994, selected item categories, current CPI versus alternative.

Expenditure group (defined by the set of strata included in test)	Current CPI	Alternative CPI (with base prices recalculated)
All items included in test	2.13%	2.02%
Food and Beverages	1.59	1.38
Food at home	1.59	1.23
Food away from home	1.66	1.65
Beverages	1.32	1.33
Housing (excl. shelter)	1.82	1.69
Transportation items	1.76	1.72
Medical care items	4.34	4.30
Entertainment items	3.50	3.42
Other goods and services	1.94	1.90

If this change had been implemented in the CPI for the items included in the test, the all-items CPI would have been reduced by around 0.07% per year [(2.13%-2.02%)x64%]=0.07, where 64% is the approximate weight of the items included in the test. About half of the effect would occur in the "Food at home" category, which has approximately 10% of the weight in the CPI-U population [(1.59%-1.23%)x10%]=0.036.

Redesigning the computer system to set base prices on initiation prices, however, was not feasible because the required resource commitment was too large. A similar approach, discussed in the next section, was chosen because it was more cost effective.

Other proposals aimed at creating an independence between the price used to set a base-period price and the price collected at sample introduction have not been fully investigated at this point. These proposals include using an average price calculated for a similar cluster of items, using an average price over the past year supplied by the respondent, and obtaining an average price from private market research data for a particular item. Preliminary research into these ideas have suggested they would either be difficult operationally and, therefore, costly to implement, or may introduce other errors into the CPI.

<u>Pricing both the old and new samples for a period of time before introducing the new samples into the CPI.</u>

An alternative solution is to price overlap samples for food-at-home categories for three months. This means that the CPI would continue to price and use in the index calculation the old samples for food-at-home observations for three months after it had discontinued the rest of the old sample for other items; at the same time it would price but not use in the calculation the new sample for food-at-home observations. By doing this, base-period prices for items in the new sample can be calculated using these earlier prices, and are much less likely to be correlated with price change after the new sample is linked into the index. Pricing double samples for food at home for three months is expected to have the same effect as setting the base price on the initiation price. Although food at home represents only about 10% of the weight of all items, it would benefit the most by the institution of overlap samples

Mathematically, the calculation of the price relative would be similar to that for setting base prices from initiation prices:

$$R_{link+1,\,link} = \frac{\sum P_{link+1}[EX_b / (X_z * P_{link-3})]}{\sum P_{link}[EX_b / (X_z * P_{link-3})]}$$

The only difference is that the initiation price can typically be collected as early as P_{link-6} and as late as P_{link}, as seen in Table 1 above.

While we cannot make a prospective estimate, our research into setting base prices on initiation prices indicates that the bias in the entire CPI by delaying the linking of new samples for food at home could be reduced around 0.04% per year. We should have a retrospective 12-month impact of this change in the middle of 1996.

We will implement this method for food at home items in those cities whose samples will be linked in January 1995 (St. Louis, Missouri and Nashville, Tennessee). A new sample for Chicago will be linked in February 1995.

II. Rent and owners' equivalent rent

The shelter components of the U.S. CPI also present some unusual estimation problems. The most prominent of these is the estimation of changes in homeowners' equivalent rent, a concept that must be estimated indirectly because the owners' implicit rents are not directly observable. Another issue is that rent data are collected less frequently than other prices, at 6-month intervals. This led to adoption in 1978 of a composite estimator that combines information from the 6-month collection intervals with respondent-recall information about rent changes during the preceding month.

Owners' implicit rent.

One outcome of research by BLS staff into the sample rotation and chaining properties of CPI estimators has been an examination of the chaining properties of the imputation formula used to calculate owners' implicit rents. BLS adopted the concept of owners' equivalent rent in 1983 as the economic concept to be measured by the price index for owner-occupied housing (Gillingham and Lane, 1982). From 1983-1986 the index was calculated by simply reweighting the rent sample to represent owner-occupied units. Beginning in January 1987, BLS began estimating owners' equivalent rent using a sample of owner-occupied units. Each owner unit was assigned an initial implicit rent based on a rent level estimated by the field representative. Subsequently, every six months each owner unit was matched to one or more rental units, and the implicit rent was imputed forward by the average rent change of the matched rental units. The matching procedure attempts to match the owner to rental units in the same neighborhood with similar structural attributes. See *BLS Handbook of Methods* (1992) for details on the matching and imputation procedures and Lane and Sommers (1984) for a discussion of the model underlying the matching and imputation methodology.

The formula currently used to measure average rent change of matched rental units is an equally weighted average of renter relatives, or Sauerbeck formula. These relatives are chained together each time an owner unit is matched, every six months. The tendency of the Sauerbeck formula to exhibit chain drift could therefore lead to substantial overstatement of the change in owners' implicit rent. Our research, presented below, suggests that the overstatement relative to other formulas with better index-number properties has been about 0.5% per year.

To understand how the chaining of imputed owner rent changes causes upward drift, consider the following example. Suppose that an owner with an initial implicit rent of $500 is matched to two renters, each of whom also has an initial rent of $500. The first period, one of the renters experiences a rent increase to $600, and the other rent remains unchanged. The next period, the first rent remains equal to $600, but the second rent increases from $500 to $600. An appropriate implicit rent formula should have the owner's implicit rent increase to $600, or in other words satisfy the "proportionality" property. The current formula does not satisfy this property, however, and the owner's implicit rent would increase to $605.

Date	Matched rent # 1	Matched rent # 2	Owner implicit rent: Sauerbeck formula	Owner implicit rent: ratio-of-sums formula
Jan-93	$500	$500	$500	$500
Jul-93	600	500	550	550
Jan-94	600	600	605	600

This example illustrates that the current formula causes the implicit rent (and therefore the index) to overstate the rent change of matched rental units. The current implicit rent formula was introduced in January 1987—since then the growth rate of the homeowners' index has exceeded the growth rate of the renters' index by 0.9% per year. Approximately half of the difference can be attributed to using an inappropriate imputation formula.

The current Sauerbeck imputation formula for the 6-month owner relative is (except for an age factor):

$$S_j = \frac{1}{n_j} \sum_{i \in Q_j} \frac{P_{i,t}}{P_{i,t-6}},$$

where $P_{i,t}$ is the current period pure rent of matched renter i, $P_{i,t-6}$ is the pure rent from 6 months previous, Q_j is the set of renters matched to owner j, and n_j is the number of renters in that set.

An alternative formula that satisfies the proportionality criterion is the ratio of sums (A):

$$A_j = \frac{\sum P_{i,t}}{\sum P_{i,t-6}},$$

The fact that this formula satisfies the proportionality principle is illustrated using the earlier example, in which the ratio of sums moves the owner's implicit rent to the correct long-run value of $600.

In the simulations reported below, the age (depreciation) factor was applied to the renter relatives, as is done in the current index, and then formulas S and A were used to calculate implicit rents from the matched renters. Similar expressions were applied to obtain the 1-month changes in implicit rent.

Composite Estimation.

The current composite estimator for shelter was introduced with two objectives: to improve the timeliness of the index and to reduce the variance. The form of the composite estimator currently used for rent and homeowners' equivalent rent is

$$I_t = 0.65 \times R_{t,t-1} \times I_{t-1} \quad + \quad 0.35 \times R_{t,t-6} \times I_{t-6},$$

where I_t, I_{t-1}, and I_{t-6} are the current period, previous period, and period $t-6$ indexes for the area, $R_{t,t-1}$ is the 1-month relative, which is calculated using respondent recall of the previous month's rent, and $R_{t,t-6}$ is the 6-month relative, which is calculated from comparing current rent to the rent reported during the last interview, six months previously.

The composite estimator was intended to produce an unbiased estimate of the current index, under the assumption that the 1-month and 6-month relatives are each unbiased estimates of relative rent change during the corresponding interval. Research by staff of the Price Statistical Methods Division of BLS, however, has found that the 1-month relatives tend to underestimate rent change. One of the reasons for this is that rent changes often occur when the tenant changes, and the new tenant may not be aware that a rent change has taken place. The CPI uses a vacancy imputation procedure to try to adjust for units that were vacant during the previous month, but these imputations are necessarily imperfect (Baskin, 1994). Even among long-term tenants, however, the evidence indicates that there is significant under reporting of 1-month rent changes (Jacobson, 1994a).

Because the 6-month relative is considered much more reliable, in general, than the 1-month relative, the following "6-month chained" formula has been studied as an alternative to the current composite estimator:

$$I_t = (R_{t,t-6})^{1/6} \times I_{t-1}$$

This index formula also eliminates the oscillating patterns that occur in the composite formula because the 6-month relative is multiplied by the 6-months previous index. For analysis of the properties of the composite index and various alternatives see Kosary, Branscome, and Sommers (1984) and Jacobson (1994b). Research has shown that the variance of the 6-month chained formula is much lower than for the composite formula, and furthermore the validity of the measure is improved to the extent that under reporting of 1-month rent changes is eliminated (Leaver, 1994). The only potential disadvantage is a possible reduction in the timeliness of the indexes, since it would take an average of 3-4 months for rent changes to appear in the index. It should be noted, however, that some of the improved timeliness of the current composite estimator is illusory, since much of the 1-month rent change is derived from vacancy imputations which in turn are derived from 6-month rent changes. Since rents are largely determined by long-term contracts and tend to move gradually, the disadvantage of reduced timeliness is not so great for shelter as it would be for other components of the CPI.

Simulations

To test these alternative formulas using actual production data, we simulated the index calculations for owners' equivalent rent under alternative implicit rent imputation formulas and under the composite and 6-month chained index estimation formulas over the period from March 1992 to June 1994. These simulations use the same owner-renter matches that were assigned when the published indexes were produced, and thus these simulations give us accurate estimates of what the indexes would have been had the alternatives been adopted in April 1992. The alternative implicit rents and indexes were carried forward for each subsequent month's calculations.

Chart 1 shows the 1-month index changes for rent using the current composite estimator and the alternative 6-month chained formula. The lower variability of the 6-month chained formula is easily seen in the chart. Table 3 shows that the composite estimator produces an annual growth rate for the index of about 2.1%, compared to 2.3% for the 6-month chained estimator. This difference is most likely attributable to under reporting of 1-month rent changes. Chart 2 shows that the 6-month chained formula results in a similar reduction in variability for homeowners' equivalent rent. Table 3 shows that changing the imputation formula (while continuing to use the composite estimator) reduces the annual growth rate of the owners' equivalent rent index from 3.0% to 2.5%, which is attributable to the chain drift problem of the Sauerbeck formula. Changing from the composite to the 6-month chained estimator increases the index growth rate from 2.5% to 2.6%, which can be attributed to the under reporting of 1-month rent changes. Overall, the alternative estimator is expected to reduce the growth rate of owners' equivalent rent by 0.4% per year.

Table 3. Percent change from March 1992-June 1994, rent and homeowners' equivalent rent, current CPI versus alternatives.

Index / Estimator	Cumulative change (27 months)	Annual rate of change
Rent: current, composite estimator	4.82%	2.11%
Rent: 6-month chained estimator	5.32	2.33
Owners' equivalent rent: composite estimator with Sauerbeck imputation	6.98	3.04
Owners' equivalent rent: composite estimator with ratio-of-sums imputation	5.75	2.52
Owners' equivalent rent: 6-month chained estimator with ratio-of-sums imputation	6.02	2.63

Our research has also examined two other imputation formulas for owners' implicit rents that also satisfy the proportionality and reciprocity properties (the geometric mean and a generalized ratio-of-means formula). The indexes derived from these two formulas moved very similarly to the index based on the ratio of sums. On the basis of these findings, the simpler ratio-of-sums imputation formula, along with the 6-month chained estimator is scheduled to be implemented by the CPI shelter indexes effective with the January 1995 CPI. Research is continuing for further improving the housing components during the next major revision of the U.S. CPI.

Chart 1. Comparison of Monthly Changes (%) for Rent:
Current, Composite Estimator vs. 6-Month Chained Estimator

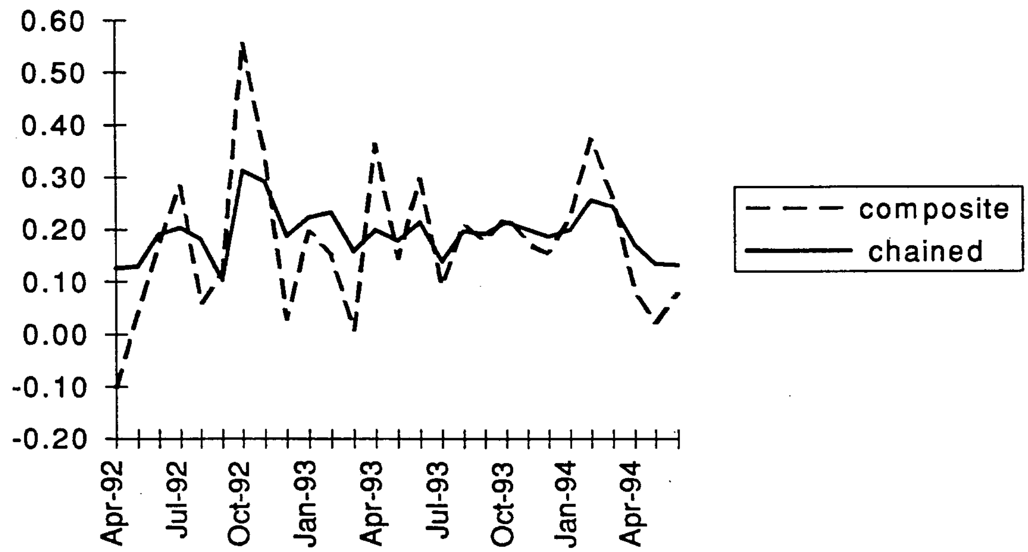
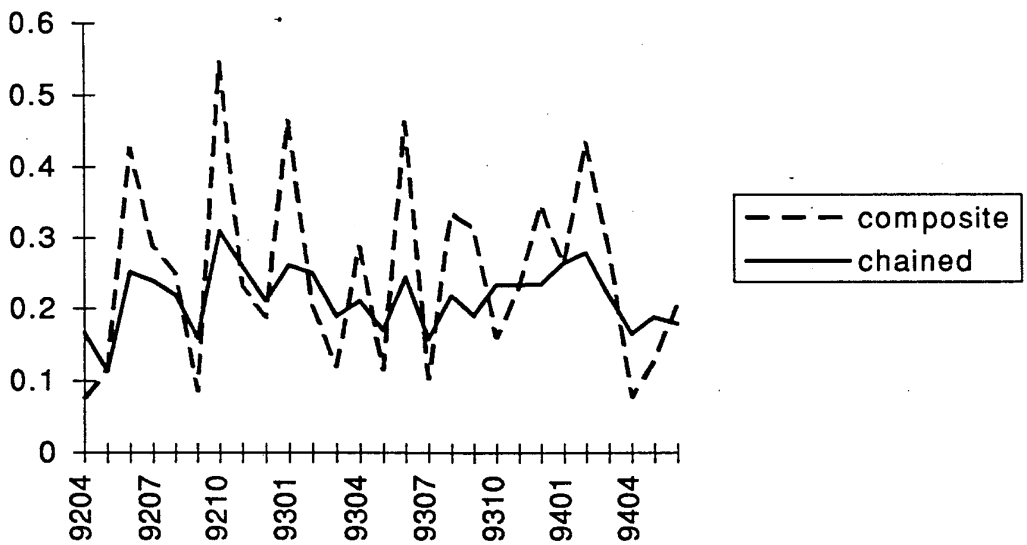


Chart 2. Comparison of Monthly Changes (%) for Owners' Equivalent Rent:
Current Composite Estimator with Sauerbeck Imputation vs. 6-Month Chained Estimator with Ratio-of-Sums Imputation Formula



III. Prescription Drugs

Griliches and Cockburn (1993) note that traditional price indexes do not capture price declines enjoyed by consumers who switch from a branded drug to its generic "equivalent" when the branded drug loses its patent. Indeed, in the U.S. CPI, when a branded drug under patent is sampled, we have historically continued to price that brand even when it loses its patent, as long as it remains for sale in the outlet. Generic drugs typically enter the CPI only through sample rotation, where any price change from the previous sample is linked. On the rare occasions when a pharmacy has discontinued selling the branded drug but offered the generic equivalent, we have substituted to the generic form and shown a price decline.

Effective with the calculation of the index for January 1995, however, the U.S. CPI will change its treatment of prescription drugs that lose patent protection. The CPI prescription drug analyst will keep track of when all branded drugs lose patent protection. Six months after a drug loses its patent, that particular drug will be resampled using probability proportional to expenditures for that drug in that specific outlet.

For example, the patented brand Capoten (generically, Captopril), used in the treatment of hypertension and congestive heart failure, will go off patent August 8, 1995. In March 1996, we will select between Capoten and other, generic versions of Captopril, based on revenues between these. When a generic version is selected in place of Capoten, we will reflect in the U.S. CPI the entire price difference (i.e., price decline) between the brand drug and the "equivalent" generic. This change in CPI procedures will have the effect of slightly slowing the rate of growth in the CPI prescription drugs component.

This method is in one respect less conservative than that laid out by Griliches and Cockburn. While some consumers, even those who switch to generics, perceive some quality difference between the brand and the generic, we will treat use the entire price difference between the brand and the generic as pure price change. On the other hand, by basing our substitution procedures for prescription drugs on revenues, there will be somewhat less substitutions taking place in the CPI sample for prescription drugs than if quantities were used to substitute.

IV. Summary and Future Research

The U.S. CPI uses a Laspeyres-type estimator to measure price change in consumer markets. There are many concerns about fixed market basket indexes being used to measure changes in the cost of living. Some of these concerns revolve around ways in which such indexes can incorporate new products as well as reflect changes in consumer tastes and preferences that occur over time. The U.S. CPI approaches these problems by holding expenditure patterns fixed at the basic index level, but allows the sample of items and weights to vary within the elementary index. Thus, new samples of products and services are introduced through a rotating sample scheme that updates 20 percent of the sampled items in geographic areas on an annual basis. The base period expenditures are escalated every month by chaining together weighted averages of monthly price changes using the products and services that are actually being purchased by consumers.

While this approach helps resolve problems of declining samples with less representative products and services in the market place, it has been discovered that the procedures used to introduce the new samples are not fully congruent with the type of Laspeyres estimator employed at the basic index level. The combination of introducing new weights and the price oscillations in several components of the CPI contributes to a problem of chain index drift. Previously, a number of researchers have observed this phenomenon for indexes computed at higher levels of aggregation. This fact demonstrates the importance of knowing the relationship between the estimator used at the micro-index level and the weighting structure used for combining price observations.

A similar problem also occurs in the shelter component of the U.S. CPI. The use of a Sauerbeck estimation formula (average of relative price change) for imputing the implicit rents of owner-occupied housing causes these values to be overstated in comparison to a more traditional Laspeyres estimator (ratio of average prices). This fact highlights the importance of being cognizant of the statistical properties of index imputation formula as well.

Another issue that has become of increasing interest in calculating consumer price indexes is the relationship of the a CPI to a true cost-of-living index (CLI). The Laspeyres-type indexes that are predominantly used in most countries attempt to keep quantities fixed and, thus, only allow prices to change in response to changes in relative prices. The CLI should reflect the substitution effects of consumer responses in such situations. Where the Laspeyres index assumes no substitution (price elasticity of demand is zero), the geometric mean index assumes the consumer will substitute with unitary price elasticity. This has been an area of research to which we have devoted considerable resources. We will continue this research to determine if the geometric mean form of the Laspeyres index is a better estimator of the components of a CLI than our current Laspeyres formulation.

In the future we will publish a number of alternative research indexes on an annual basis so that users can see the effects of different index estimators and understand the different assumptions about consumer behavior that underlie each. We have published both Fisher Ideal and Tornqvist indexes (Aizcorbe and Jackman, 1993) and will update these annually. In 1996, we plan to publish a complete set of test indexes that use the geometric mean estimator. To date, we have only published indexes for selected components of the CPI with the shelter component being the most prominent area missing (Moulton, 1993 and Reinsdorf and Moulton, 1994). We will be very interested in research conducted by other statistical agencies in this area.

The authors wish to thank Karin Smedley, who developed most of the non-shelter index simulations presented in this paper. We also wish to thank Steve Henderson, Walter Lane, Sylvia Leaver, and Marshall Reinsdorf for their input, but any errors or omissions are our own.

Footnotes

[1]A true cost-of-living index would also take into account the effect on consumers of such things as income taxes, changes in real income, and the consumption of non-market items.

[2]The 44 geographic areas of the country are represented by 88 primary sampling units (PSUs). Of these 88 PSUs, 32 are self-representing cities such as Washington, DC and St. Louis, MO. The remaining 56 PSUs are probability selected to represent the four U.S. Census regions and 3 city-size classes within each region, for a total of 12 additional geographic areas.

[3]Only about 70 percent of CPI items were included in Moulton's simulation. The most notable missing component was shelter, which accounts for over 25 percent of the market basket.

References

Aizcorbe, Ana M., and Jackman, Patrick: "The commodity substitution effect in CPI data, 1982-1991," *Monthly Labor Review,* December 1993, pp. 25-33.

Armknecht, Paul A.: "Effects of changing market baskets and product substitution in the U.S. Consumer Price Index," Economic Commission for Europe/International Labour Organization Meeting of Consumer Price Indexes, Geneva, October 1993.

Armknecht, Paul A. and Ginsburg, Daniel H.: "Improvements in measuring price changes in consumer services: past, present, and future," *Output Measurement in Services Sectors*, National Bureau of Economic Research, Studies in Income and Wealth Volume 56, University of Chicago Press, 1992, pp. 73-156.

Baskin, Robert: "Imputation of Vacant Units in the Residential Rent Index of the U.S. Consumer Price Index," *1994 American Statistical Association Proceedings of the Survey Research Methods Section* (to appear).

Braithwait, Steven D.: "The substitution bias of the Laspeyres price index: An analysis using estimated cost-of-living indexes." *American Economic Review*, March 1980, pp. 64-77.

Dalén, Jörgen: "Computing elementary aggregates in the Swedish Consumer Price Index," *Journal of Official Statistics* 8 (1), 1992, pp. 129-147.

Diewert, W. Erwin: "Index numbers," *The New Palgrave Dictionary of Economics*, Vol. 2, The Stockton Press, New York, 1987, pp. 767-780.

Forsyth, F.G. and Fowler, R.F.: "The Theory and Practice of Chain Price Index Numbers," Journal of the Royal Statistical Society (A Series) 144(2): 224-46.

Gillingham, Robert: "A Conceptual Framework for the Revised Consumer Price Index," Proceedings of the Business and Economic Statistics Section, American Statistical Association, 1974, pp. 46-52.

Gillingham, Robert, and Lane, Walter: "Changing the treatment of shelter costs for homeowners in the CPI," *Monthly Labor Review,* Vol. 105, June 1982, pp. 9-14.

Griliches, Zvi and Cockburn, Iain.: "Generics and New Goods in Pharmaceutical Price Indexes," National Bureau of Economic Research, Working Paper No. 4272, February 1993.

Jacobson, Shawn: Memorandum to CPI Housing Steering Committee, January 11, 1994 (a).

Jacobson, Shawn: "Evaluation of alternative rent estimators for the Consumer Price Index," *1994 American Statistical Association Proceedings of the Government* Statistics Section (to appear), referenced as 1994 (b).

Kosary, Carol, Branscome, James M., and Sommers, John P.: "Evaluating alternatives to the rent estimator," *1984 American Statistical Association Proceedings of the Business and Economic Statistics Section.*

Lane, Walter F., and Sommers, John P.: "Improved measures of shelter costs," *1984 American Statistical Association Proceedings of the Business and Economic Statistics Section.*

Leaver, S.G.: Internal communication to Housing Sample Design and Estimation Teams, July 6, 1994.

Manser, Marilyn E. and McDonald, Richard J: "An analysis of the substitution bias in measuring inflation, 1959-85," *Econometrica*, Vol. 56, July 1988, pp. 909-930.

Moulton, Brent: "Basic components of the CPI: Estimation of price changes," *Monthly Labor Review*, December 1993, pp. 13-24.

Pollak, Robert A.: "Subindexes in the cost-of-living index," *International Economic Review*, February 1975, pp. 135-150.

Reinsdorf, Marshall: "The effect of outlet price differentials on the U.S. Consumer Price Index," *Price Measurements and their Uses*, Vol. 57, National Bureau of Economic Research, University of Chicago Press, 1993.

Reinsdorf, Marshall: "Price dispersion, seller substitution, and the U.S. CPI," U.S. Bureau of Labor Statistics Working Paper 252, March 1994 (revised version of paper presented at Statistics Canada's Price Measurement Advisory Committee, Ottawa, Ontario, Canada, May 1993).

Reinsdorf, Marshall and Moulton, Brent: "The construction of basic components of cost of living indexes," National Bureau of Economic Research, Conference on Research in Income and Wealth - The Economics of New Products, April 1994.

Schmidt, Mary Lynn: "Effects of updating the CPI market basket," *Monthly Labor Review*, December 1993, pp. 59-62.

Szulc, Bohdan J.: "Linking price index numbers," *Price Level Measurement: Proceedings from a Conference Sponsored by Statistics Canada*, Ottawa, Ontario, 1983.

Szulc, Bohdan J.: "Price indices below the basic aggregation level," in *Consumer Price Indexes: An ILO Manual,* International Labour Office, Geneva, 1989.

Turvey, Ralph: *Consumer Price Indices: An ILO Manual*, International Labour Office, Geneva, 1989.

U.S. Department of Labor, Bureau of Labor Statistics, *Consumer Price Index Detailed Report*, September 1994.

U.S. Department of Labor, Bureau of Labor Statistics, *Handbook of Methods*, Bulletin 2414, 1992.