Journal of Official Statistics
Vol. 6. No. 2, 1990, pp. 179–204

Miscellanea

Under the heading Miscellanea, essays will be published dealing with topics considered to be of general interest to the readers. All contributions will be refereed for their compatibility with this criterion.

Improving Consumer Price Indexes

John F. Early[1]

Abstract: Design and procedural improvements of Consumer Price Indexes are suggested. Specific consequences of economic theory are developed. The results are applied to homeowner shelter costs and other measurement issues. Advances in sample design and estimation techniques are reviewed. The contributions of quality management, technological applications, and error profiles are also addressed.

Key words: Consumer price index; consumer theory; sample design; non-sampling error; quality; cost of living index.

1. Introduction

A consumer price index (CPI) is among the two or three most important economic statistics which a nation's statistical system produces. It may also be the most complex. This paper is intended as a comprehensive identification and analysis of the design and procedural factors which determine the quality of a CPI. Drawing on examples of recent developments and research currently underway, it provides a framework in which systematic improvements in a CPI may be developed. While it is based heavily on recent experience in the United States, it also draws on advances in other countries. This paper tries to express a number of very complex technical matters without recourse to excessive jargon and mathematical notation. It is hoped that this article will be accessible to the diverse audience interested in consumer price indexes.

Section 2 discusses a conceptual framework which is based on economic theory. Using the conceptual underpinning, Section 3 derives operational measurement objectives – especially for difficult items like homeowner shelter costs and insurance. Section 4 establishes estimation and pricing techniques that minimize the mean squared error of the statistics resulting from the measurements. Section 5 discusses design issues pertinent to probability sampling. Section 6 looks at the effects that substitutes for disappearing products have on the index and develops techniques to minimize the errors caused by substitution. Section 7 examines alternative estimators that measure

[1] Vice President, Juran Institute, Inc., 11 River Road, Wilton, CT 06897, U.S.A.

the effects of different weights, exclusion of certain taxes, and different population definitions, all of which reflect practical choices inherent in defining a CPI. Section 8 argues for the development and publishing of methodological statements and error profiles of both sampling and nonsampling errors. Section 9 provides guidelines for establishing a comprehensive quality control where empirical techniques are used to evaluate the survey process. Error prevention is preferred to error correction. Section 10 discusses some recent advances that led to improved quality and productivity in operations.

2. A Conceptual Basis

No other factor is more important for the quality of any statistic than a full and accurate definition of the concept to be measured.

2.1. Economic foundation: Cost of living index

While one can define a price index as a configuration of prices and weights, that type of nominalist approach provides only minimal information about the meaning of the final result. Meaningful economic statistics must be related to underlying economic theory and measurement. Over the last 15 years, a rigorous and widely accepted framework has been developed for the CPI. This framework provides a basis for the answers to fundamental questions about index construction and provides the context in which results can be interpreted meaningfully.

Meaningful price indexes can be constructed for many economic phenomena – industry input, industry output, imports, exports, etc. For the CPI, the relevant economic process is consumption. A consumer derives satisfaction (or welfare or utility) from economic activity in one of three ways: (1) direct consumption of a service, such as riding on a bus, (2) consumption of a service incidental to the consumption of a nondurable good, such as eating a banana, and (3) consumption over time of the flow of services provided by a durable good such as a house (Gillingham 1974). Consumption for a time period is the quantity of all services used in that period. Consumption expenditure is the sum of the prices for the service units consumed. Consumption expenditures change over time because (1) the prices of the items consumed change and (2) the quantities consumed change. The objective of a consumer price index is normally the measurement of that first factor alone.

An expenditure-based cost-of-living index (in economic terms) has been defined as the ratio of the cost of achieving the same level of consumer satisfaction (welfare or utility) within a fixed set of tastes at two different points in time (Pollak 1971, 1983). The two points in time are the reference period and the comparison period, with the reference period coming first in time and the comparison coming later. Normally, the numerator of the index ratio is the cost for the given satisfaction level at comparison period prices, and the denominator, the cost at reference prices. The index can then be interpreted in terms of the percent change in prices from the one period to the other.

But the prices themselves are only a part of the index. What pattern of consumption is being priced? For time series cost-of-living indexes, there are usually one of two choices – either the consumption at the reference time or the consumption at the comparison time – and we can call these respectively the reference-expenditure-based cost-of-living index and the comparison-expenditure-based cost-of-living index. The importance of the difference between these two types of indexes can be seen intuitively by considering the difference between (1) a price index measuring price change from 1967 to 1987 based on consumption in 1967, and (2) a price index for the same period based on consumption in 1987. In many countries, consumer preferences for such things as day care, home electronics, motor fuels and restaurant meals have changed radically in those two decades. As a result, the choice of consumption pattern could have a significant effect on the indexes.

In either case, the comparison for a cost-of-living index is between the costs for achieving the same consumer satisfaction in two time periods. While such a theoretical construct may be difficult to create in practice, it provides the necessary conceptual anchor for other more tractable calculations.

2.2. *Price indexes*

The most widely used formula for consumer price indexes is the Laspeyres index, which is constructed by (1) identifying actual consumption amount in the reference period, (2) calculating the cost of that amount of consumption in both the reference and comparison periods, and (3) dividing the comparison cost by the reference cost. Symbolically, this can be expressed as follows:

$$I_L(r, c) = \sum P_c Q_r / \sum P_r Q_r$$

where

$I_L(r, c)$ = the Laspeyres index for period c with a reference period r

P = Price

Q = Quantity

r = reference period

c = comparison period

The difference between a reference-expenditure-based cost-of-living index and a Laspeyres formula is in the numerator. The denominator is the same for both. The numerator for the Laspeyres index is the comparison period cost of consuming the same quantities of items that were consumed in the reference period. The numerator of a reference-expenditure-based cost-of-living index is the cost of achieving the same level of satisfaction within a fixed preference ordering (or fixed tastes). While one can always be certain that purchasing the same set of items will provide at least the same level of satisfaction as during the reference period, if relative prices have changed, there may be a less costly combination of purchases that will produce the same satisfaction. For example, one could construct a hypothetical index for chicken and ground beef using the following data:

item	reference quantity	reference price	comparison price
chicken	20 kg.	$1.00	$5.00
ground beef	10 kg.	$1.00	$1.00

A Laspeyres index would be constructed as follows:

$$I_L(r, c) = ((20 \times 5) + (10 \times 1)) / ((20 \times 1) + (10 \times 1)) = 3.667.$$

But suppose the consumers are indifferent between the combination of 20 kg. of chicken and 10 kg. of ground beef on the one hand and 10 kg. of chicken and 30 kg. of ground beef of the other. (The second combination is, after all, 10 kg. more total weight.) A cost of living index would then be calculated as follows:

$$I_c(r, c) = ((10 \times 5) + (30 \times 1)) / ((20 \times 1) + (10 \times 1)) = 2.667.$$

While this ia a didactic example, it illustrates an important result: that a Laspeyres-type CPI is the upper bound to a true expenditure-based cost-of-living index.

Another widely known index formula is the Paasche index, which can be expressed as follows:

$$I_p(r, c) = \sum P_c Q_c / \sum P_r Q_c .$$

The Paasche index differs from the Laspeyres in that the quantity weights for the Paasche belong to the comparison rather than the reference period. Paasche indexes form the lower bound for comparsion-expenditure-based cost-of-living indexes.

Because current period expenditure weights are difficult to develop in a timely fashion, Paasche indexes are rarely calculated except as by-products of the national income and product accounts. Paasche indexes are also inconvenient for analytical purposes. A time series of Laspeyres indexes on a given reference base can be used to make meaningful price change comparisons between any two time periods. The same is not true for Paasche indexes. Except for the reference-comparison change, all other changes in Paasche indexes of a common reference base are composites of price and quantity change and cannot be used to calculate price-only changes (Triplett 1981).

2.3. *Derivative data series*

There is often strong, but subsidiary, interest in price indexes for particular items or for specific local areas. Because of the preeminent importance of the national all items CPI for national policy and adjustments of incomes and payments, the U.S. has adopted a sample design that minimizes the error on the national all items index. It provides only those detailed indexes which can be produced as by-products. The additional cost for more detailed local data, in particular, has been judged by the governmental budget policy process to be too high for the potential value (Marcoot 1985).

The data requirements for a good CPI are not necessarily the same as those for average prices or place-to-place price comparisons. Some food and energy average prices can be constructed from a CPI-only set of prices because of the great homogeneity of these products. Data on place-to-place comparisons were discontinued after 1980 in the U.S. because adequate measures could not be constructed from CPI data alone. Emerging advanced econometric techniques, however, may make it possible to construct some place-to-place and average price estimates from the CPI data (Blanciforti 1986).

3. Operational Measurement Objectives

Many of the most perplexing problems of price measurement can yield satisfactory results with appropriate, systematic application of a consistent economic framework like that cited above. One cannot hope to review many of these in the span of a single article. But one case – that of shelter costs for homeowners – is an excellent example that illustrates how proper conceptual solutions are perhaps the most significant contributors to the quality of a CPI.

Systematic progression from conceptual basis to operational measurement objectives improves the final index by insuring that the estimates produced will meet the overall user needs. It also makes the resolution of issues easier because the rules and objectives are clearly stated; there is little need to seek new criteria in making design or operation decisions.

3.1. *Measurement objective: The case of homeownership*

For many years, the shelter cost index for homeowners in the U.S. CPI was measured as the combined effects of changes in current house prices, mortgage interest costs (over half the life of the mortgage), property tax, property insurance, and maintenance and repair costs. These were the costs of purchasing and maintaining a housing unit as an asset. Application of consumer economic theory to this problem, however, tells us that the proper price for homeownership is the price of the flow-of-services which the homeowner receives from the asset (the house) (Gillingham 1980). A housing asset is very different from a service, such as a bus ride, which the purchaser consumes directly or a nondurable good, such as a banana, which is totally consumed in the process of providing service to the purchaser. Not only is the housing unit not consumed in the process of providing the shelter service, it may actually increase in value when property values rise. Distortions that result from using house prices in the CPI can be illustrated by the following example. If house prices jump 15% in a year and are included in the CPI, it would appear that homeowner costs have gone up sharply. But that increase is experienced only by the small minority who buy a home in that year. Many other homeowners, however, who are also consuming shelter services, may actually experience a decline in shelter costs because they have received a capital gain equal to 15% of the value of the home that they own. No matter what weights are applied to house prices, and a number of proposals have been offered, the central points are (1) a house price is an asset price and (2) changes in asset prices represent not only changes in the cost of acquiring the asset but also changes in the incomes of asset holders. Asset prices are not consumption prices.

What is needed, then, is a cost of current shelter services for all homeowners and that cost can be estimated as follows:

$$C = D - K + i_m M + i_e E + Z, \quad (3.1)$$

where

C = The cost of current shelter services
D = The depreciation in the house by virtue of its use and aging in the current period
K = Capital gain (or loss) in the market value of the house (adjusted for depreciation) during the current period
M = Mortgage balance owed
E = Owner equity in the home
i_m = interest rate paid
i_e = interest rate foregone on owner equity
Z = Operating costs such as taxes, maintenance and repair, and insurance on the structure.

The depreciation, mortgage interest, and operating costs are straightforward, easily interpreted costs. In addition, if the prices for houses are rising (or falling) their owners are receiving capital gains (or losses). Typically, in recent U.S. history, there have been strong gains in the capital value of houses so that appreciating values have off-set other costs. The noneconomist may argue that this appreciation is only "income on paper" and not a real cost offset. But it must be realized that this is real income from an asset (just like rising values of a stock portfolio).

Another way to understand this capital gain is to consider an individual who buys a home for $70,000 and sells it five years later for $140,000 in order to buy a new identical home in another city. If one is using an asset measurement approach like the pre-1983 U.S. CPI, this would be measured as a price increase of 100%. But for this individual there was no cost increase (from the house price) because of the appreciated value of the asset that he sold. Capital gains are a negative component of homeowner costs.

Of course, capital invested in the purchase of a home, cannot earn other investment income. If home owners sold their homes, became renters, and invested the sale proceeds, they would obtain some rate of return (i_e). The alternative investment income (interest) that the owners foregoe is a cost of owning their housing assets.

While some of the terms in the user cost function (3.1) present a number of practical measurement difficulties, they can all be estimated in principle – all but one. The estimation of the interest rate for foregone returns on equity (i_e) has no immediately obvious data source. Interest rates are far from homogeneous, varying radically depending on the risk and liquidity characteristics of the asset. What asset has the same liquidity and risk of a housing unit? The answer is – a housing unit. Nearly half the housing in the U.S. is investor-owned rental property. For those units, the rent (R) will equal the cost (C) of owning the asset as defined in (3.1) above.[2] If rent and cost are equal, then (3.1) can be restated and rearranged such that

$$i_e = (R - i_m M - D - Z + K)/E. \tag{3.2}$$

In other words, because all the other terms are known, it becomes possible to solve for the internal rate of return for rental housing as the lost equity-return rate for owners. But since, in order to do that, one must first determine the rental value of the owner property and since that rental value is equal to the cost of the service from the property, the easiest, most straight-forward measurement of the cost of service from owner occupied housing is the rental equivalence of that property (Gillingham and Lane 1982).

While this is but a sketch of the entire issue, I am convinced that a flow of services approach provides unambiguously the best concept of shelter costs. Rental equivalence is, for the U.S., the most viable available technique for measurement. There can never again be any question of returning to a house price-based asset approach.[3]

3.2. *The effects of selecting objective*

The importance of the proper choice of measurement objectives in this case cannot be overstated. The differences in measured inflation between the official U.S. CPI using the asset cost approach and an experimental flow-of-service index using rental equivalence are shown in Table 1.

4. Estimation and Pricing Technique with Minimum Mean Squared Error

By continuing the orderly progression from concept to measurement objective to estimator, the quality of the final product is clearly improved. In designing estimators and techniques, statisticians need to be aware of the total error structure of the

[2] For this equality to be strictly true, one must make certain assumptions about property management costs and market efficiencies which are beyond the scope of this presentation but do not affect the final conclusions. See Gillingham (1980) and Gillingham and Lane (1982) for more detailed treatment.

[3] Other durable goods – especially vehicles – offer similar opportunities for improvement and appropriate research is underway. The implication for most other consumer durables are less severe, however, because the depreciation term usually overwhelms the capital gain and the current purchases can be viewed as a first order approximation of the replacement of depreciated stock over the aggregation of all households.

Table 1. Annual percent changes, U.S. consumer prices 1979–1982

Year	Official CPI	Experimental CPI using rental equivalence	Difference between two measures
1979	13.3	10.8	2.5
1980	12.4	10.8	1.6
1981	8.9	8.5	0.4
1982	3.9	5.0	–1.1

process – both sampling and nonsampling. Neither one of these general sources of error is necessarily dominant, and it is their total mean squared error that needs to be considered.

In this section, I shall highlight a few which I consider to be particularly important as well as illustrative of that broad range. I almost hesitate to treat these topics as though they were independent of the sample design or preceeded it in some logical sense. It is, in fact, necessary that the estimators and samples be designed jointly. They are separated here only for expository convenience.

4.1. Owners' equivalent rent

If there were not an active rental housing market, it would not be possible to construct a direct reasonable measure of owners' equivalent rents. As a first approximation, the CPI rent index itself can be taken as the measure of change in owners' equivalent rents and, in fact, the experimental index cited above used such an approach. But, in most housing markets, there are some distinctions between the set of homes that are owner occupied and the set of those rented. Often, owners are more likely to live in single-family detached units, and renters in apartments. The two types of housing may even be concentrated in different parts of the city. It would, thus, not be best to build an estimator which assumes homogeneity across both markets.

In the U.S., owners' equivalent rent changes are estimated using both a sample of owners and a sample of renters. A portion of the renter sample was selected explicitly to measure primarily the changes in owners' equivalent rents. Each owner housing unit in the sample is "matched" with a set of about three renter units. The two most important elements of this matching are the geographic location of the units and the type of structure – single family detached, apartment, etc. These two elements have been identified empirically as determinants of rent change. The changes in the rental value of an owner unit over time are estimated from the average percent changes in the rent of the rental units which have been matched with the owner unit (Lane and Sommers 1984).

4.2. Composite estimation

In the U.S., the CPI program makes use of composite estimation in a number of applications. Generically, these composites are the weighted average of two estimates for a variable – one essentially unbiased with relatively high variance and the other with some bias and substantially less variance. The weighting of the two estimates is selected so as to minimize the mean squared error of the result.

One class of composites is engaged to provide lower mean squared error estimates for local areas than would otherwise be possible with simple estimators. Simple estimates for the individual geographic areas are often based on very small samples and, thus, tend to have very high sampling variability. For example, the expenditure weight for each item in each geographic area is a composite of the mean expenditures developed for that area (unbiased with large variance) and the mean for a larger regional grouping (smaller variance but biased as an estimate for the individual local area) (Cohen and Sommers 1984). A regression-based composite between local and regional estimates has also been developed for the local parameters used in the housing aging adjustments discussed below (Randolph and Zieschang 1987).

Another type of composite is used in the CPI rent and owners' equivalent rent indexes. The rent sample consists of six panels, each priced at a six-month interval, at which time rents are collected for both the current and previous months. The index is estimated as follows:

$$I_t = .65 I_{t-1} R_{1,t} + .35 I_{t-6} R_{6,t},$$

where I_t is the rent index in month t.

$R_{1,t}$ is the average relative of rent change over a one-month period and $R_{6,t}$ is the average relative of rent change over a six-month period. The one-month relative of change has a higher variance than the six-month because rent changes tend to occur as infrequent discrete jumps. Over a six-month period these changes will show less variability than in a single month when only a very few units will change. Exclusive use of the six-month relative would, however, create certain time-lag problems. For example, if rents should be frozen by government action, the six-month comparison would continue to show rent changes, even if the rent freeze had full compliance. The 65/35 relative weighting seems to give a reasonable reduction in mean squared error compared with other options (Kosary, Sommers, and Branscome 1984).

4.3. *Vacancy nonresponse*

In a survey of rent, there is a particular form of nonresponse that raises unusual nonresponse adjustment problems – viz., that of the unit which is for rent but vacant. If the same panel of renters were priced month after month, any price change that occurred between the last month of one renter and the first month of the next renter would eventually show up in the index. The same is true if one used exclusively six-month comparisons, the only problem being one of timing. However, the particular composite form used in the U.S. produces a more difficult problem. While the six-month relatives cover the full elapsed time completely and will eventually capture any rent change of vacant units that became occupied, the six-month relative has only 0.35 of the weight. The one-month relatives, however, do not chain together to cover all time because each unit has only one such relative every six months. As a result, there will be no observed one-month rent changes for any units which are vacant in the current or previous month. The rent changes for these vacancy cases will be the result of whatever nonresponse adjustment is used.

There are three possibilities for this adjustment. The first is the "naive" adjustment applied to other nonresponse cases – i.e., the average rent change in the stratum. This method was used for many years in the U.S. CPI until Sommers and Rivers (1983) showed that this procedure biased rent change toward zero because vacant units are generally undergoing a change of tenant, and rent changes are more likely to accompany tenant changes.

This determination led to the second and current method for imputing rents of vacant units. A model of rent change is estimated for the set of rental units that have changed occupants in the last six-months. From this model, the average rent change at the time of the new tenants' occupancy is estimated. With some adjustments for the expected length of vacancies, this average jump in rents for new tenants is used to estimate the one-month rent change for vacant units.

4.4. Adjustment for aging of housing

A Laspeyeres price index relies on being able to price exactly the same item from one time period to the next. On the surface, the comparison of rents would seem to be the easiest case of all; except for a very few cases, the unit is exactly the same month after month. It is the same in all but one way – it becomes older. This aging or physical depreciation of a housing unit may make the unit worth less rent, all other things being equal. A simple comparison for the CPI will understate the price change because in the later period, the rent for the unit will be the rent for a unit of greater age, and, on average, lower quality.

This problem is not a new one. (See, for example, U.S. Department of Labor (1965); Gillingham (1983).) But only recently has the BLS made significant progress toward a sound practical solution. There are two general approaches to measuring depreciation or aging effects. One would be to observe units longitudinally. But that approach would not be very fruitful since the measure of depreciation would require adjustments for inflation, which of course is the original variable the CPI is seeking to measure.

A cross-sectional approach is plagued by the vintage effect – namely that the quality of the housing built in a given year may differ from the quality of that built in previous years. Measurement of aging effects requires satisfactory removal of these vintage factors. Fortunately, in the case of the United States, Randolph (1988) has shown that these vintage effects are quite small and can be largely controlled by using such characteristics as number of rooms and location. As a result, regression equations which explain the log of rent by a large number of characteristic and location variables plus variables for age and age crossed with other variables can yield good estimates of the size of the aging effect. Randolph (1987) has estimated that the U.S. rent index is understated by about 0.3 percentage points per year because of lack of an adjustment for aging of the housing units. Such an adjustment, based on further refinements, was introduced with the January 1988 CPI.

4.5. New cars and transaction prices

While the prices for most items are reasonably straightforward to define and obtain, others present interesting challenges. At least in the United States, new cars do not have posted transaction prices that one can simply copy down from a price tag. Manufacturers place "sticker prices" on the car which include not only the cost of the basic vehicle but also the cost of various "options" from power steering to air conditioning to 100 watt stereo. Dealers may add other items such as protective finishes. Depending on current market conditions, the dealer may offer certain "concessions" on the sticker price. If a car is particularly scarce (for example, resulting from import restrictions), the dealer may actually levy a surcharge on top of the sticker price. In general, the exact sizes of these concessions and surcharges cannot be easily observed by the individual consumer (or CPI price collection agent). The final price for a particular sale is typically the result of a "negotiating" process, which may bear more resemblance to street theatre than to a rational economic phenomenon. Whatever the marketing virtues of this process, the practical problems for the CPI are considerable. The CPI requires a transaction price.

In the U.S., BLS prices the same fully specified automobiles on each visit to a specific auto dealer. Because of the wide variety of available options as well as special "packages" in which these options may be bundled, the dealer may not have in stock or have recently sold exactly the specified automobile. But, the sticker price of the specified car can be readily computed. With the sticker price computed, the next step is to compute the average concession/surcharge for that dealer, during the last month, on the cars sold bearing the specified name plate. (A name plate is a variety within a brand – for example a Chevrolet Caprice.) This method represents a lot of work for the dealer, but it is the most reliable process we have been able to find thus far to ensure that the CPI reflects actual consumer experience. There are other cases of similar complications. I cannot catalog them all here, but this example illustrates a method of obtaining critical transaction prices.

5. Design Efficient, Full-Probability Samples

A proper sample design for a CPI is an extremely complex matter, one which has reached a mature, satisfactory solution only in the past decade. The universe of inquiry is the set of all prices paid by consumers for all varieties of consumer commodities and services. A proper probability sample will give each price in that universe a known, finite probability of selection. This kind of process will produce an index that is unbiased and efficient. It also makes it possible to construct estimates of sampling error. Designs that do not rely on probability selection can produce indexes that are not representative of that universe and use data and resources very inefficiently. Nonprobability designs cannot support estimates of sampling error because there is no way of knowing whether the dispersion in the data represent the dispersion in the universe.

Historically, the most difficult problems in selecting full probability samples for a CPI have been the lack of appropriate sampling frames. A discussion follows of the approach used by the United States since 1978 to overcome those problems (U.S. Department of Labor 1984). This general approach was optimized in 1987 to produce the most precise index possible, subject to the constraints of budgeted resources. The U.S. approach is not the only possible design, but it illustrates systematic application of probability techniques throughout the selection.

5.1. Stage I: Geographic sample

The first stage of sampling is geographic – selecting urban areas throughout the country in which pricing will be done. The nation is divided into exclusive and exhaustive primary sampling units (PSUs), each of which is relatively homogeneous. These PSUs are assigned to strata according to region, size, and a variety of demographic characteristics (Dippo and Jacobs 1983). Given the work associated with the minimum per-PSU collection effort and the size of the budget available for the CPI, BLS determined that the budget would support pricing in 127 PSU-equivalents. Each PSU-equivalent would, then, represent 1/127th of the U.S. urban population. Each PSU in the universe had a probability of selection proportional to its population size (PPS). Areas with a population greater than 1/127th of the total were selected with certainty and the very largest were assigned (proportionate to their population) additional PSU-equivalents of pricing – for example, the New York Consolidated Statistical Area has 11 PSU-equivalents. By allocating the pricing sample in a strictly proportional fashion among areas, maximum efficiency can be gained for the national estimates.[4]

5.2. Stage II: Items within item strata

All consumption expenditures from the Consumer Expenditure Survey (CES) are classified into 207 exclusive and exhaustive categories – for example ground beef, hospital rooms, and new cars. Each of these item strata may be further divided into entry level items[5] (ELIs) – for example laundry equipment is divided into washers and dryers. Prices are collected for each item stratum within each sampled PSU-equivalent. However, for item strata with more than one ELI, the ELIs are sampled within the item stratum for each PSU-equivalent. The ELI is selected with a probability proportional to its relative expenditure within the item stratum in the geographic region.

The number of prices to be collected for each item stratum were allocated so as to produce the smallest possible sampling error for the U.S. all items CPI within the overall available budget. This allocation required a nonlinear programming approach that considered the relative expenditure importance of each item stratum, the unit variance, the components of the variance, the unit cost of data collection and processing, and models of outlet selection overlap and response rates (Leaver, Weber, Cohen, and Archer 1987). Strata with greater expenditures have a greater effect on the CPI and, thus, require larger sample sizes than strata with smaller expenditures. Similarly, item strata with large unit variances also require larger sample sizes to reduce their contribution to the overall variance. Finally, the optimization problem is one of achieving roughly equal reduction of the all items CPI variance for each dollar expended, so that prices which are relatively inexpensive to collect (e.g., additional food items in super markets) are collected more in greater numbers than prices which are costly to collect (e.g., new cars).

5.3. Stage III: Sampling outlets

In the next stage of sampling, it is necessary to select the outlets (stores, restaurants, hospitals, physicians, utilities, etc.) from which prices are to be collected for each of the item strata. Unlike many other statistics that are strictly household or industry based, there are no readily available sampling frames, so a special frame construction process is required. In the U.S., a Point-of-Purchase Survey (POPS) is conducted in each urban area included in the CPI sample. This is a household survey. It is similar to the Consumer Expenditure Survey in that it asks consumers how much they spend in certain categories, but the categories are less detailed. Within each spending category, the survey obtains the outlets at which the items were purchased and the amount spent at each outlet for the specific set of items.

[4] BLS deviated from this strict proportionality only in designating Honolulu, Hawaii and Anchorage, Alaska as certainty sample selections because of their unique geographic character.

[5] The import of this name will be clarified shortly.

The product of the POPS is a list of outlets for each item category with a dollar value of purchases for the outlet. Each list then serves as the sampling frame for outlets for the item category. Outlets are selected with probability proportional to size (PPS) of expenditures from this list and assigned to the field for initiation into the survey.

5.4. Stage IV: Specification selection

Once an outlet has been assigned for pricing, the next task is to select for pricing one specific variety among all those available in the outlet. This selection process also needs to be conducted within a probability framework to insure an unbiased and representative sample.

To guide in this process, a checklist is developed for each ELI in the index. These checklists identify the characteristics which are necessary to describe the selected variety fully and completely. The factors which are most important in determining the price of the item are listed first. For example, for the item stratum "white bread" the first factor of the checklist is "type of bread," which includes regular, Italian, Vienna, and raisin. The CPI pricing agent obtains for each bread type the proportion which it is of all white bread sales. Suppose the sales proportions are those shown in Figure 1. A cumulative proportion is computed for each type and a random number between 1 and 100 is supplied. This random number will be within a range of the cumulative distribution that is associated with one of the types of bread – in our example, Italian bread. We have now selected a type of bread for pricing in this specific outlet, and that selection in all its stages has been probability proportional to sales. But the probability selection continues through additional stages, at each stage making a selection among the available values for each of the other bread characteristics. This could mean making such selections as between dietetic and non-dietetic; between salt and salt free; between ready-to-serve and brown-and-serve; among the various loaf sizes; and among the available brands.

Fig. 1. Disaggregation

ELI = White bread

Type	Percent sales	Cumulative percent	Random number
Regular	65	65	
Vienna	5	70	
Italian*	10	80	73
Raisin	20	100	

*Selected

With all stages of sampling complete, we now have an efficient full-probability sample of all consumer prices. Of course, this sampling process within ELI is an ideal. There are some practical limitations on the final product. Experience has suggested that in most cases, sampling to five stages below the ELI will produce a unique item. In those few cases where it does not, a unique item is arbitrarily selected from those remaining within the class.

Especially at the lower sampling stages within the ELI, the outlet cannot or will not provide data on proportions of sales within the selected class. In such cases, rankings are used to provide for improved efficiency over simple random sampling. In the absence of rankings, simple random selection is used, thereby preventing bias in selection. In almost all cases, at least one stage of sampling within ELI is achieved with PPS, and probability selection is preserved to the unique item. While there is some loss of efficiency with these fall-back procedures, the biases that might result from the alternative nonresponse or nonprobability selection would be far more serious.

5.5. *The shelter sample*

Sampling for shelter is a little different. The first stage PSU selection is the same. After that the objective is to obtain a sample of housing units to support estimation of indexes both for residential rent and for homeowners' equivalent rents. For the residential rent index alone, this would be a fairly straightforward area segment sampling process, using census data to select blocks or clusters of housing units within each PSU with a probability proportional to the number of renters. Field staff would then visit the location, list all the housing units, select a sample, visit the sample, and initiate into the survey those sampled units that are renters. However, the sample is needed to support both indexes, and, for that purpose, requires a more complex joint optimization (Lane and Sommers 1984).

The estimator for owners' equivalent rents relies on a sample of owner-occupied dwellings. An initial market rental value is estimated for each unit. From then on the rate of change in that rent is estimated by a set of rental units matched to each owner unit by geographic location, structure type, and unit characteristics. Consequently, the sample of housing units must yield us three types of sample units.

1. owner-occupied units,
2. rental units primarily selected to match the owner units for the owner's equivalent rent index, and
3. rental units primarily selected to calculate the residential rent index.

The two set of renters are not disjoint. Any renter in the sample will contribute to the residential rent index and many will have at least some influence on the owners' equivalent rent index. But, because some neighborhoods tend to be more heavily owner and others more heavily renter and because the proportion of owners living in single family homes is larger than for renters, some disproportionate sampling of renters is required in areas that are more heavily owner-occupied.

In outline, the optimum sample involves selecting census blocks and block groups with probabilities proportional to the total number of housing units in them. Once the roster of housing units is completed for a block, two different sampling rates are applied to the list – one rate for owners and another for renters. Not only are there two different rates, but the rates for renters vary by the proportion of renters in the block. In areas which are predominantly owner-occupied, the sampling rate for renters is high – often as high as one in one – because renters are rare in those areas and are needed for estimating equivalent rent changes for the owners in the block group.

5.6. *Sample updating*

Although expenditure weights in a CPI are usually fixed for extended periods, it is both possible and desirable to keep the sample which supports each detailed item index up to date. In the U.S., this is achieved by reselecting one-fifth of the sample of outlets and unique items every year. Currently, the process is to update most of the sample in a given geographic area at one time, but there may be operational and other advantages to alternative arrangements which have some sample updating in each local area every year.

The updating process is highly desirable. First, it updates the outlets to reflect current buying patterns. For example, as traditional department stores became less important and "discount outlets" became more significant in the U.S., the sample was updated to reflect that fact. Second, it updates the varieties of items priced. For example, as consumers purchase home appliances with more advanced features, the sample changes to reflect the distribution of varieties currently being purchased. The base expenditure weights remain the same between major market basket revisions, but the price relatives are based on continuously updated outlet and variety samples.

6. The Effects of Substitutes

The one factor which sets price index estimation apart from all other economic statistics is the need to ensure that only pure price change is measured. When the same item is available in adjacent time periods, price change measurement is relatively straightforward. But when the described item disappears, one must then select the closest substitute, determine how comparable the new item is, and make such quality adjustments as necessary to prevent quality changes from affecting the index.

6.1. *Comparable substitutions*

Sometimes the original and substitute items are comparable – that is, the physical differences between the two are of no economic consequence. Frequent examples of this type are new model years for appliances, when the model number and exterior appearance may change but the functional performance is unchanged.

6.2. *Overlap prices*

Sometimes it is possible to collect a price for both the original and substitute variety in the same time period. So long as both are regular market prices (e.g., one is not a clearance sale price), the difference in the prices of the two varieties is the market valuation of the quality difference between the two. The index can be estimated through the overlap month using the original variety and for all subsequent months using the new variety.

6.3. *Direct quality adjustment*

In some cases it is possible to develop direct measures of the economic value of the quality differences between two varieties. Some of the easiest cases are, of course, when varieties differ from each other in weight or volume. As long as these differences are not too extreme, these cases are handled in the index by using prices per unit of volume or unit weight.

Automobiles are perhaps the most celebrated cases. In the U.S., when a new model year is introduced, the BLS works closely with automobile producers to identify all functionally meaningful physical changes and the associated production costs, marked up to retail. Under appropriate assumptions of efficient markets, these production costs will be reasonable estimates of the market value of the quality change.

One type of automobile quality change is of special interest – the addition of governmentally required anti-pollution devices. From a production point of view, these devices are additional production and not changes in the output price. From a consumer perspective, however, these devices represent extra cost which the consumer would usually not pay if given the option. It is more of a tax for the individual consumer. Nevertheless, in the U.S., a government-wide decision was made that all legislatively required anti-pollution devices would be treated as quality improvements in all statistical series. The reasoning was that through

Table 2. Percent of possible price comparisons which are substitutes, selected items 1984

CPI items	Substitutes percent of possible price comparison
All items	3.9
Food and beverages	1.9
Housing	4.8
Apparel	17.6
Transportation	5.8
Medical care	2.2
Entertainment	6.1
Other	4.0

the democratic process all consumers had decided that anti-pollution devices were worth their cost. This is an unusually important procedure to document. It is illustrative of the need to document fully the quality adjustment procedure so users can assess the appropriateness for their particular applications.

6.4. Other noncomparable substitutes

Frequently, noncomparable substitutes have no overlap prices and the data are not available for direct quality adjustment. In these cases, the change in price between the two varieties cannot be used in the index. The price change in the index for this nonresponse is estimated as the average rate of price change for those quotations with comparable or quality adjusted comparisons in the same item stratum and geographic area.

This long-standing and wide-spread procedure, however, may understate pure price change associated with introducing new varieties. Research is needed to investigate in detail the effect of this procedure and develop, if necessary, alternative nonresponse adjustments.

Table 3. Percent of substitutes by type of treatment in CPI, 1984

Type of substitute	Percent of all substitutes
Comparable	43.2
Overlap price	5.8
Direct quality adjustment	7.6
Noncomparable, not used	43.4
Total substitutes	100

6.5. Effects of substitutes and quality adjustments

Substitutes usually occur in only about 3% or 4% of the possible price comparisons, but the effects are uneven across items (See Table 2), with substitutions being greatest among apparel items. Unfortunately, data are available only infrequently for overlap pricing and direct quality adjustment. (See Table 3.) Most direct quality adjustments are for new vehicles. Comparable substitutes and noncomparables that cannot be used occur with approximately equal frequency.

When BLS introduced full probability sampling for varieties within outlets in 1978, concern was expressed by some that this departure from the "typical volume seller" approach would result in more frequent substitutions. In fact, because the sampling was done PPS within each outlet, greater stability in availability was actually achieved (Armknecht 1984).

Measurement error arising from substitutions is likely the most significant source of error remaining in the U.S. CPI. Sampling error and operational errors are easier to evaluate and control. Great progress has been made in the design of the substitution process but much remains to be done.

The first key to a good substitution process is the empirical identification of factors that determine variety comparability for each item. The second key, which follows the first both in importance and logical dependence, is the measurement of the value of characteristic differences between varieties. Historically, BLS has relied on a corps of commodity economists to estimate these values. They developed individual expertise in the characteristics and marketing of the commodities and services in the CPI. Their judgement determined what characteristics were most significant, and they worked with producers of the items to identify possible data sources for valuing differences in these characteristics.

BLS realized the need to build on this item expertise with the application of systematic empirical methods to identify and evaluate the characteristics of quality change. Generally speaking, we have attempted to use regression analysis for this purpose, building equations that explain differences in price levels among varieties of an item by the different characteristics of the varieties – a technique sometimes called hedonic regressions. For example, one might build a model that explains automobile prices in terms of characteristics like acceleration, turning radius, trunk space, leg room, and the presence of certain luxury features. To date most applications of this technique have been used to evaluate the results already achieved by more traditional analysis. (For example, Blanciforti and Galvin (1984); Triplett and McDonald (1977); Early and Sinclair (1983)). We have recently begun to develop and use these hedonic results to prepare the "checklists" used by data collection field staff when they select a substitute for a discontinued item. These checklists indicate the relative importance attached to each characteristic, and the field staff is instructed to select a substitute in such a way that it shares the original item's most important characteristics. Use of hedonic regression provides a structured and empirical foundation for the judgement of the commodity economists in constructing this ordering (Armknecht and Weyback 1989). Actual use of coefficients developed from the regressions to adjust for quality change is, for now, confined to the shelter component of the index, where estimated equations produce adjustments for rent when there is a change in the inclusion of utilities or furnishings in the rent and parameters are being estimated that will permit adjustments for aging of the sample rental units (Randolph 1987).

6.6. *Criteria for quality adjustment*

The substitution and quality adjustment processes are important for a high quality CPI. The existing procedures need to be evaluated carefully and improved methods need to be developed. Adequate procedures require that two conditions be met. First, only price change should be reflected in the CPI; all quality and quantity effects must be avoided. Second, all true price change must be captured for the CPI. The two criteria are equally important, but at times CPI programs may focus on the first to the detriment of the second. In earlier years, the apparel component of the U.S. CPI may have suffered from this problem, drawing comparability criteria so tightly that price increases which accompanied introduction of new fashion seasons may have been partially lost – imparting some downward bias to the index. (Armknecht and Weyback 1989, pp. 114–115).

7. Alternative Estimates

Because of the importance and variety of applications for the CPI, the conceptual characteristics of the official CPI should be fully understood, and the empirical consequences of that conceptual design should be documented as completely as possible. Frequently, it is desirable to achieve this end by producing experimental or supplementary indexes that measure alternative formulations.

Although the CPI should be constructed in a cost-of-living framework, there are important differences between a CPI and a cost of living index that need to be documented. As noted above, a Laspeyres-type CPI is the upper bound of a fixed-expenditure cost-of-living index. The difference between the official CPI and a fixed-expenditure cost of living index has been labeled the "substitution bias" of the CPI. While there is no simple, perfect method for estimating the substitution bias, a number of approaches have been developed using "superlative index numbers" which estimate expenditure-based cost of living indexes under a variety of assumptions that are not terribly restrictive (Diewert 1976). Research on U.S. data indicates that substitution bias causes the official CPI to have annual rates of price change that are in the neighborhood of 0.1 to 0.2 percentage points larger than an expenditure-based cost of living index would have (Kokoski 1987a). It is useful to know that the substitution bias is, at least in the U.S. and for the short run, quite small. Other sources of measurement uncertainty are undoubtedly more important.

Substitution bias should not be confused with the effects of updating the expenditure weights in a Laspeyres CPI. In fact, if one were to update the CPI weights frequently, the potential substitution bias could actually increase. Because the Laspeyres index for each weight period is an upper bound for the cost-of-living index for that period, chaining together a series of Laspeyres indexes on different reference periods over a longer period of time could produce results that are higher than Laspeyres indexes based on a single reference expenditure pattern for the entire period. Nevertheless, there is great interest in the price changes for more up-to-date market baskets, and Kokoski (1987a) has found that, in practice, for the U.S. chained indexes reduce substitution bias. As part of its research program, the BLS is preparing to issue experimental indexes with updated expenditure weights between CPI revisions.

The comparison of a CPI with expenditure-based cost of living indexes to identify substitution biases is only a partial consideration of the larger question of cost-of-living indexes. So far in this article, we have considered only the change in expenditure needed to achieve a fixed degree of satisfaction from consumption. A more general consideration is to measure the change in income needed to achieve a fixed degree of satisfaction from consumption. This more general formulation requires a systematic treatment of taxes.

Most CPI constructs include the effects of indirect taxes (those levied on the producers and distributors of consumption items) and "sales taxes" (those collected from the consumer at the time of the sale of an item on an ad valorem or per unit basis). Taxes on the income or assets of the consumer do not generally enter the calculation of the CPI. But in order to consider income-based cost-of-living measures, these direct taxes must be included. The tax effects on the income needed to maintain a given level of consumer satisfaction is of considerable economic interest and policy importance. If some of the government revenue is shifted from personal income taxes to excise or value-added taxes, the construction of such a tax and price index can be of special importance. (For example, see U.K. Central Statistical Office 1979. See also Gillingham and Greenless 1987; and Kokoski 1987a).

The cost of living index framework outlined in Section 2 has sometimes been called "plutocratic" because by using aggregate consumption of each item for the weighting pattern, it implicitly gives higher weight to consumer units that have greater expenditures. Prais (1959) has proposed an alternative "democratic" index in which the all items index is the simple arithmetic average of the all items indexes from each consumer unit. Each consumer unit, therefore, has equal weight in the final summary index.

Diewert (1983) has expanded theories relating the "plutocratic" and "democratic" indexes. The most highlighted argument for using the "democratic" formulation seems to spring from a philosophical desire not to obscure the differential effects of inflation on low income groups. There are, of course, substantial practical difficulties in creating a "democratic" index. Such an index is also less useful as a macroeconomic indicator because it does not relate in any clear fashion to other economic aggregates.

As is the case for any average, a "democratic" index for the entire population does not really give us any better indication of the fate of a particular expenditure group. The best way to deal with the differential effects of inflation on different demographic groups is to construct specific indexes based on their expenditure patterns. In addition to the policy concerns with such differences, the wide-spread practice of escalating transfer payments by a CPI has also created an interest in CPIs for specific demographic groups. Strictly speaking, these special population indexes should be based not only on the subpopulations' spending and geographic patterns but also on the specific outlets and varieties of items bought by the subpopulation. The substantial cost of such fully independent subpopulation CPIs has generally precluded their development. Nevertheless, useful and interesting experimental indexes for demographic groups have highlighted some of the issues associated with CPI escalation applications (Hageman 1982; and Kokoski 1987b).

8. Publish Methodology and Error Profiles

All statistical series need to be accompanied by comprehensive statements of the methods used. Such documentation improves the usefulness of the data for the user as well as enhances his or her confidence in the final product. Beyond these normative statements of method, there is also a need to document the effects of the application of the methods in error profiles.

Error profiles are comprehensive descriptions of the sources and magnitude of errors in statistical series. All too often one thinks only of the sampling error (variances), but in a well-designed CPI, sampling error may be smaller than other error sources. A good error profile serves two functions. First, it provides insight to the survey managers on the aspects of the survey that could benefit most from improvement. Second, it provides users with valuable data on the limitations that might apply to their applications.

In a complex design such as the CPI, variance estimates cannot be derived easily or analytically. Involved procedures, such as half-sample replication, are generally required (U.S. Department of Labor 1984, pp. 22–24). Unfortunately, the cost for such procedures is high.

The most general categories of nonsampling error are coverage error, nonresponse error, response error, processing error, and estimation error. Coverage error is the error in an estimate that results from the omission of part of the target population (undercoverage) or the inclusion of units from outside of the target population (overcoverage). Coverage errors in the CPI would result from the omission of cities, households, outlets, and items that are part of the target populations from the relevant sampling frames or the double counting or inclusion in the frames of such sample units when they should not be. A potential source of coverage error in the CPI is the time-lag between the Point of Purchase Survey and the initiation of commodities and services price collection. Because of the time lag, the products offered by the outlet at the time it is initiated may not exactly coincide with the set from which the POPS respondents were purchasing.

Nonresponse error results when data is not collected for some sampled units. In the CPI, nonresponse errors relate to the failure to interview householders or to price outlets. This can occur when selected households and outlets cannot be contacted or refuse to participate. This nonresponse could bias the CPI if the rate of price change at the lost survey units differs from the rate of price change at the survey units successfully initiated. Nonresponse rates for the U.S. CPI are published periodically (U.S. Department of Labor 1984, p. 13). These data are most useful if they identify the point in the survey at which the nonresponse occurred and the type of nonresponse involved.

Response error results from the collection and use in estimation of incorrect, inconsistent, or incomplete data. In the CPI, response error may arise because of the collection of data from inappropriate respondents, respondent memory or recall errors, deliberate distortion of responses, interviewer effects, misrecording of responses, the pricing of wrong items, the misunderstanding or misapplication of data collection procedures, or the misunderstanding of the survey needs or lack of cooperation from respondents. BLS is beginning use of a reinterview methodology for investigation of various kinds of response variance in the CPI. Response variance can be measured through a reinterview conducted under identical conditions to the original interview. Work in the area of response biases has not yet been fully developed for the CPI. The pricing methodology in the Commodities and Services component of the CPI provides the previous periods's price to the pricing agent at the time of collection. This dependent pricing methodology is believed to reduce response variance for measuring change, but it may cause response bias and lag in the measurement of price change.

Processing error arises from incorrect editing, coding, and data transfer. Such errors are introduced in converting the survey responses into machine-readable form for analysis. Errors can also result from software problems in the computer processing which cause correctly keyed data to be lost. Computer screening and professional review of the data provide additional checks on processing accuracy. Special studies of these processing errors in the U.S. CPI have shown them to be extremely small.

Estimation error results when the survey process does not measure accurately what is intended to be measured. Such errors may be conceptual or procedural in nature, arising from a misunderstanding of the underlying survey measurement concepts or a misapplication of rules and procedures. In the CPI, a source of estimation error due to conceptual problems was the pre-1983 treatment of housing, discussed earlier, which failed to distinguish between the consumption and investment aspects of homeownership.

As already discussed, substitutions and

adjustments for quality change in the items priced for the CPI are possibly the largest sources of estimation error in the CPI. This is a unique problem with price indexes and deserves careful attention. Estimation error will result if the adjustment process, which necessarily has a significant judgement component and may have key data unavailable, is misapplied, or if it consistently over or underestimates quality change for particular kinds of items. While individual substitution estimation problems have been identified, the evidence to date for the U.S. is that, on average, there is no systematic bias from this process. Cases where price change is overestimated are about as frequent as those where it is underestimated.

9. Comprehensive Quality Management

The CPI can also benefit from what managers throughout the world have been learning about making all types of better products – automobiles, electronics, or CPI's – through comprehensive quality management (for example, Juran 1989; Juran and Gryna 1980; Imai 1986). Traditional quality control and quality assurance tools are basic to this process, but quality management is more inclusive than that. First, quality and the cost of quality become the chief driving forces for the entire effort. Second, all elements of the program are subject to quality audits and quality improvement activities – concept design, sample design, frame construction, data collection, data capture (conversion to machine-readable form), data processing, data presentation, and data analysis (Early and Dmytrow 1986). The quality of the final product is actually more dependent on the way we design it (quality of design) than on the precision with which the operating forces follow the design (quality of conformance). Historically, most survey management praxis has focused on whipping the operating forces into shape when usually the greatest gains are to be had from improving our designs in the first place – more appropriate concepts, better collection instruments, and more rigorous system designs.

Careful study of each element of the survey process can identify the points where the greatest quality improvements can be achieved. (Sometimes this involves lowering the cost for achieving a particular level of quality.) These process studies provide the basis for redesign of the process to provide better quality. This improvement process works, however, only with the full openness and support of top management of the statistical agency.

Most statistical agencies have some form of survey reinterview process. This process is usually viewed as a way to make sure that field staff are conducting their work properly and as a way to measure response variance. In an effective quality management environment, however, such reinterview information first and foremost is used to identify design problems – question wording, instrument structure, concept definition, operational instructions, or training. Then, after the designs are under adequate control, and only then, does reinterview data serve as a good basis for measuring individual field representative performance and response variance.

While it is fairly commonly accepted that "quality control" procedures are required for "nonprofessional" operations like data entry or data collection, statisticians and economists tend to believe that their contributions to the survey process cannot be subjected to quality studies. Nothing could be farther from the truth. Not only are the quality of their designs the primal issue in the quality of a CPI, but many of their efforts can be measured and evaluated more directly. In the U.S. CPI, direct studies of the data editing, adjustment, and correction actions by staff economists revealed meaningful patterns that permit managers to monitor performance and identify redesign needs (Early and Galvin 1987).

10. Research in Operational Enhancements

An appropriate quality management process results in literally hundreds of operational improvements. Experience with that type of process suggests that these incremental improvements are actually more important than the less frequent but more dramatic innovative breakthroughs (Imai 1986, pp. 33–38 and Juran 1989, pp. 28–38). Nevertheless, there are three examples of potentially dramatic innovations that should be mentioned.

10.1. Interactive data capture

In the United States, the CPI program has recently installed an interactive data capture process. To understand how this process could provide better data capture at a lower cost, one must start with the previous data capture process, which is summarized in the first column of Figure 2. The new system uses a network of microcomputers and data bases to capture the data and prepare it for index calculation (Helliwell 1987). It is summarized in the second column.

Even with this brief outline, one can imagine how the keying can have fewer errors, be completed in a more timely fashion, and still cost less to operate. The most fundamental characteristics here are (1) the elimination of opportunities to make transcription errors, (2) placing the editing power of the computer so that it is available immediately and interactively, and (3) replacing endless checking and verification with independent keying and discrepancy resolution. After cumulative analysis of keying error rates, a more sophisticated process may be built that can increase the editing power of the original data entry and make use of acceptance sampling of batches to reduce further the amount of keying and resolution.

A fundamental fact in survey processing is that most editing and checking serves primarily to prevent deterioration of the originally collected data. The closer the data preparation is to the original collection and the fewer the hands through which it must pass, the more accurate and less costly it is. Two other techniques being explored by BLS for the CPI start with the lessons learned from microcomputer networks for data capture and move their power even closer to the original source of the data.

10.2. Computer assisted telephone interviewing (CATI)

CATI is not a new technique, nor is it appropriate for all kinds of survey work. But the CPI shelter survey is a good candidate for a CATI application. Since the shelter survey is longitudinal over several years, it can benefit from both personal contact at initiation and the efficiency and control of CATI for subsequent interviews. A small pilot test to date has provided encouraging results (Kosary and Sommers 1987).

The survey instrument is presented to the telephone interviewer on a screen, one question at a time, with skip patterns and probes controlled by the computer. The responses are automatically edited for legal values and internal consistency. The results of the entry may generate clarifying probes and control the sequence of the interview. The computer can also perform extensive workload scheduling and management functions.

Such a system has obvious advantages. It saves the time and cost associated with travel for personal interviews. It provides better control of the sequence and wording of the interview process. It performs edit checks immediately, at the point in time

Fig. 2. Comparison of two data keying technologies

Old Technology	*New Technology*
1. Data collection schedules received from field	1. Data collection schedules received from field
2. Schedules entered in a manual log	2. Schedules logged-in using automated optical character recognition
3. Schedules sent to keying operation	3. Clerk keys data directly on a micro-computer
4. Schedules keyed	4. The data are edited while being keyed and corrections made
5. Verification keying performed	5. A second clerk keys the data again independently subject to the same editing
6. Listing made of keyed data	6. The two keyings are compared by computer and differences resolved by a supervisor directly through a net-work micro-computer
7. Listing reviewed line by line for accuracy of keying	7. Professional review listing is created
8. Corrections based on listing entered on paper	8. Economists use their own micro-computers to access the data on the network and make their changes and adjustments, which are edited while they are being entered
9. Corrections keyed	
10. Computer edits performed	
11. Listing of edits reviewed	
12. Corrections based on edits entered on paper	
13. Corrections keyed	
14. Professional review listing created	
15. Economists write up special transmittal forms to specify changes to data and adjustments	
16. Adjustments and changes keyed	
17. Keying of adjustments and changes reviewed	

when it is easiest to correct the collected data. Once the interview is complete, the data became available almost immediately for professional review and survey-status summaries.

10.3. Computer assisted personal interviewing (CAPI)

Most price collection relies on the personal observation of the individual field agent, not only for determining the price, but also for validating or changing the specification. Emerging laptop microcomputer and telecommunications techniques offer substantial opportunities to acquire many of the other virtues of a CATI operation in a personal interview environment. With a laptop computer, the collection sequence and operation can be more rigorously defined and controlled. The computer logic can be substituted for voluminous data collection manuals. Data can be edited and validated on the spot. With nightly hook-ups to the central computing facility, professional review of the data and response to field queries can be just hours away, rather than separated by days or even weeks from the original collection. Early dutch experiments (Bemelmans-Spork and Sikkel 1985) found some technical limits to the equipment which would be even more problematic in the U.S. with its store-specific specifications. But both the storage capacity and speed of available hardware have grown while size and weight have declined. Prospects now seem sufficiently encouraging so that the U.S. has begun development of a pilot effort to test the application of CAPI to price collection.

11. Success and Challenge

The United States CPI program is a relatively large one, employing the equivalent of about 500 full-time persons and spending about 26 million U.S. dollars annually. The sample sizes of about 90,000 prices each month for goods and services and 20,000 housing units for shelter costs, however, are certainly not the largest among industrialized nations. While sample size is one important determinant of the precision, we have seen above that proper methods and processes are far more important. Although it is dangerous to make sweeping generalizations, within very broad limits, it is this author's conviction that when faced with budget reductions or with opportunities to restructure a CPI program, smaller samples selected properly and processed appropriately are to be preferred over larger samples lacking one or more of the quality improvements discussed in this paper.

As a rough indication of where the efforts in the United States CPI are focused, the following are the approximate percent distributions of the annual CPI program resources with supervision, administration and planning apportioned across activities:

Collection and processing of monthly data	66%
Annual selection and updating of samples	17%
Professional review and adjustment of data	5%
Quality improvement and assurance	8%
Research in theory and methods	2%
Provision of current data and analysis	2%

Over the last two decades, there have been mammoth strides in improving the quality of CPIs. A proper conceptual framework has been clearly enunciated. The consequences of alternative conceptual designs have been widely explored. More appropriate and effective estimation techniques have been established. An effective, full probability sample design has been demonstrated to be effective.

There remain many important challenges for managers of CPI programs – challenges that can be effectively addressed in an open quality management environment which focuses on the needs of the diverse multitude of CPI users. The most important challenges lie in the need to measure and reduce the error associated with substitutions. Fuller quality measurement and assessments will undoubtedly open up many opportunities for improvement, both through the application of new technology and the operational improvement of existing techniques.

12. References

Armknecht, P. (1984). Quality Adjustment in the CPI and Methods to Improve It. Proceedings of the Business and Economic Statistics Section, American Statistical Association, 57–63.

Armknecht, P.A. and Weyback, D. (1989). Adjustments for Quality Change in the U.S. Consumer Price Index, Journal of Official Statistics, 5, 107–123.

Bemelmans-Spork, M.E.J. and Sikkel, D. (1985). Observation of Prices with Handheld Computers. Statistical Journal of the United Nations Economic Commission for Europe, 3, 153–160.

Blanciforti, L.A. and Galvin, J.M. (1984). New Approaches for Automobiles in the CPI. Proceedings of the Business and Economic Statistics Section, American Statistical Association, 64–73.

Blanciforti, L. A. (1986). Price Measurement for Interarea Comparisons. Division of Price Index Number Research, Bureau of Labor Statistics, Presented at the December 1986 American Statistical Association Meeting, New Orleans.

Cohen, M.P. and Sommers, J.P. (1984). Evaluation of the Methods of Composite Estimation of Cost Weights for the CPI. Proceedings of the Business and Economic Statistics Section, American Statistical Association, 466–471.

Diewert, W.E. (1976). Exact and Superlative Index Numbers. Journal of Econometrics, 4, 115–145.

Diewert, W.E. (1983). The Theory of the Cost-of-Living Index and the Measurement of Welfare Changes. In Price Level Measurement, eds. W.E. Diewert and C. Montmaquette, Ottawa: Statistics Canada, 163–223.

Dippo, C.S. and Jacobs, C.A. (1983). Area Sample Design for the Consumer Price Index. Proceedings of the Survey Research Methods Section, American Statistical Association, 118–123.

Early, J.F. and Dmytrow, E.D. (1986). Managing Information Quality in the CPI. Fortieth Anniversary Quality Congress Transactions, American Society for Quality Control, 2–9.

Early, J.F. and Galvin, J.M. (1987). Comprehensive Quality Measurement in the Consumer Price Index. IMPR087 Conference Proceedings, Juran Institute, Inc., Wilton, Connecticut.

Early, J.F. and Sinclair, J.H. (1983). Quality Adjustment in the Producer Price Indexes, In The U.S. National Income and Product Account Selected Topics, eds. M.F. Foss, Chicago: University of Chicago Press.

Gillingham, R.F. (1974). A Conceptual Framework for the Consumer Price Index. Proceedings of the Business and Economics Statistics Section, American Statistical Association, 246–252.

Gillingham, R.F. (1980). Estimating the User Cost of Owner-Occupied Housing. Monthly Labor Review, 103 : 2, 31–35.

Gillingham, R.F. and Lane, W. (1982).

Changing the Treatment of Shelter Costs for Homeowners in the CPI. Monthly Labor Review, 105:6, 9–14.

Gillingham, R.F. (1983). Measuring the Cost of Shelter for Homeowners: Theoretical and Empirical Considerations. The Review of Economics and Statistics, 254–265.

Gillingham, R.F. and Greenless, J.S. (1987). The Impact of Direct Taxes on the Cost of Living. Journal of Political Economy, 95, 775–796.

Hageman, R.P. (1982). The Variability of Inflation Rates Across Household Types. Journal of Money, Credit and Banking, 14, 494–510.

Helliwell, J. (1987). New Network Behind Improved CPI. PC Week, February 3, pp. C/4–C/10.

Imai, M. (1986). Kaizen: The Key to Japan's Competitive Success. New York: Random House.

Juran, J. M. (1989). Juran on Leadership for Quality, an Executive Handbook. New York: The Free Press.

Juran, J.M. and Gryna, F.M. (1980). Quality Planning and Analysis. New York: McGraw-Hill.

Kokoski, M.F. (1987a). Experimental Indexes of the Cost of Living. Report, Bureau of Labor Statistics.

Kokoski, M.F. (1987b). Consumer Price Indexes by Demographic Group. BLS Working Paper 167, U.S. Department of Labor, Bureau of Labor Statistics.

Kosary, C.L., Sommers, J.P., and Branscome, J.M. (1984). Evaluating Alternatives to the Rent Estimator. Proceedings of the Business and Economic Statistics Section, American Statistical Association, 410–412.

Kosary, C.L. and Sommers, J.P. (1987). The Effect of Computer Assisted Telephone Interviewing on Response Inconsistency. Proceedings of the Survey Research Methods Section, American Statistical Association, 627–632.

Lane, W.F. and Sommers, J.P. (1984). Improved Measures of Shelter Costs. Proceedings of the Business and Economic Statistics Section, American Statistical Association, 49–56.

Leaver, S.G., Weber, W.L., Cohen, M.P., and Archer, K.P. (1987). Determining an Optimal Item-Outlet Sample Design for the 1987 U.S. Consumer Price Index Revision. Proceedings, International Statistical Institute.

Marcoot, J.L. (1985). Revision of the Consumer Price Index Now Underway. Monthly Labor Review, 108:4, 27–38.

Pollak, R.A. (1971). The Theory of the Cost of Living Index. BLS Working Paper 11, U.S. Department of Labor, Bureau of Labor Statistics.

Pollak, R.A. (1983). The Theory of the Cost of Living Index. Conference Proceedings, Statistics Canada Conference on Price Measurement, Ottawa: Statistics Canada.

Prais, S. (1959). Whose Cost-of-Living? Review of Economic Studies, 26, 126–134.

Randolph, W.C. and Zieschang, K.D. (1987). Aggregation Consistent Restriction-Based Improvement of Local Area Estimation. Proceedings of Business and Economic Statistics Section, American Statistical Association.

Randolph, W.C. (1987). Housing Depreciation and Aging Bias in the Consumer Price Index. BLS Working Paper 166, U.S. Department of Labor, Bureau of Labor Statistics.

Randolph, W.C. (1988). Short Term Quality Change and Long Term Vintage Effects. Journal of Urban Economics, 23:1, 162–178.

Sommers, J.P. and Rivers, J.D. (1983). Vacancy Imputation Methodology for

Rents in the CPI. Proceedings of the Business and Economic Statistics Section, American Statistical Association, 201–205.

Tripplet, J.E. and McDonald, R.J. (1977). Assessing the Quality Error in Output Measures: The Case of Refrigerators. Review of Income and Wealth, 23, 136–157.

Triplett, J.E. (1981). Reconciling the CPI and PCE Deflator. Monthly Labor Review, Vol. 104:9, 3–15.

U.K. Central Statistical Office (1979). The Tax and Price Index – Sources and Methods. Economic Trends, No. 310, August, 81–89.

U.S. Department of Labor (1965). The Consumer Price Index: History and Techniques. Bulletin 1517, Bureau of Labor Statistics, Washington, D.C., p. 70.

U.S. Department of Labor (1984). BLS Handbood of Methods: Volume II, The Consumer Price Index. Washington D.C.: Bureau of Labor Statistics.

Received April 1988
Revised May 1989