Handling Program Constraints in the Sample Design for the Commodities and Services Component of the U.S. Consumer Price Index

August 2005

Sylvia G. Leaver, Darin T. Solk
United States Bureau of Labor Statistics, Washington, D.C.

Abstract

Sampling for the Commodities and Services (C&S) component of the U.S. Consumer Price Index involves selection of outlets from establishment frames and individual items from a highly stratified item frame. The methodology employed in this process relies primarily on models of survey operation costs and sampling variance. These, in turn, are used to find local solutions to a nonlinear programming problem as a first stage sample resource allocation.

Explicit constraints in this setting involve the pre-selected area design, a total survey cost ceiling and minimum and maximum sample size requirements. However, there are additional important constraints that inform and challenge this process. These include the organization of the sample rotation scheme for C&S, including the grouping of items for simultaneous rotation, the number of sample size variables for which the problem is tractable as a nonlinear programming problem, publication requirements, availability of program cost accounting data, the variability of the character of outlet frames, and the need to retain consistency in allocation from one sample rotation to the next. This paper discusses these additional constraints and the approaches taken in this design setting.

Any opinions expressed in this paper are those of the authors and do not constitute policy of the Bureau of Labor Statistics.

Keywords: Nonlinear optimization, Cost modeling, Components of variance, Simulation

1. Introduction

We first describe the methods used to allocate data collection resources for the sample for the Commodities and Services (C&S) component of the U.S. Consumer Price Index. These methods rely on models relating price change sampling variance and data collection costs to design variables which are the number of items to price and outlets to visit per item group in each sample city. With these models, an optimal allocation of data collection resources to minimize sampling variance of price change, subject to budgetary and operational constraints, is found using nonlinear programming techniques. Models for sampling variance and costs are given, and approaches taken to accommodate program and design constraints are discussed. A closing section characterizes the changes in sample allocation some of these constraints have driven over the last several years, and adaptations we are considering to improve sample efficiency.

1.1 Background

The CPI is calculated monthly for the total U.S. metropolitan and urban non-metropolitan population for all consumer items, and it is also estimated at other levels defined by geographic area and item groups such as cereal, women's suits, and tobacco products (BLS, 2003).

An index area is the most basic geographic area for which a price index is computed. There are two types of index areas: self-representing areas, such as New York, which were selected with certainty; and non-self-representing areas, whose sample comprises two or more primary sampling units (PSUs) selected according to a probability sample. Geographic samples are revised periodically. The U.S. All Cities CPI is a weighted average of 38 index area CPI's; 31 from self-representing and 7 from non-self-representing areas. For purposes of variance estimation and operational manageability, the sample for each self-representing PSU is segmented into two or more subsets, called sample replicates. The C&S sample is refreshed on a rotating basis with approximately one-eighth of the item and outlet sample in each PSU being reselected in each of two semiannual sample rotations.

The CPI is estimated for items grouped into 211 strata for each index area, and for higher item and area aggregates. Each item stratum is composed of one or more narrowly defined classes called entry level items (ELIs.) In CPI item selection, ELIs are selected from each stratum by a systematic probability-proportionalto-size (pps) procedure, where the ELI weights are derived from expenditures reported in the two most recent years of the Consumer Expenditure Survey. ELI selections are independently drawn from each stratum for each sample replicate within each PSU.

Sample frames and weights used in outlet selection are derived from a point-of-purchase survey. Prior to 1998, this survey was the Continuing Point of Purchase Survey (CPOPS), a household personal visit survey conducted by the U.S. Bureau of the Census (USBOC) for BLS. The CPOPS was conducted in each PSU on a five-year rotation cycle. All frames for all applicable items in a PSU were refreshed simultaneously with its rotation. Beginning in 1998, this survey was replaced with the Telephone Point of Purchase Survey (TPOPS), a random-digit telephone survey, also conducted by USBOC for BLS. The structure of the TPOPS survey is remarkably different from CPOPS in that frames are refreshed on a continuing basis in each PSU. The rotation cycle for TPOPS is 4 years, rather than 5 years and approximately one-eighth of the frames and sample for any PSU is refreshed with each semiannual sample rotation. Item categories are staggered across the sample PSUs so that in any one semiannual rotation, the sample for each item category is being rotated in at least one PSU.

TPOPS provides the names and addresses of outlets, and the dollar amounts, of purchases for classes of items known as POPS categories. A POPS category is a group of items normally sold in the same kind of outlet, and each ELI maps to one POPS category. Outlet frames, total daily expenditure estimates, and selection probabilities are derived from TPOPS data for each PSU-POPS category-sample replicate. Outlets are then selected via systematic pps from frames for each PSU-sample replicate for POPS categories corresponding to ELIs selected in item sampling. Selected items are then priced in sample outlets on a monthly, bimonthly, or seasonal basis.

The CPI is constructed in two stages. In the first, or elementary cell stage, the price index for an item-area is updated every one or two months via a function of sample quote-level price changes called a price relative. Let X_{ia}^{t} denote the index at time *t*, in item stratum *i*, area *a*, relative to time period *0*. Then

$X_{ia}^{t} = R_{ia}^{t,t-1} X_{ia}^{t-1}$, where $R_{ia}^{t,t-1}$ denotes the price relative between times *t* and *t*-1 . Since 1999, price relatives for most commodities and services have been computed using a weighted geometric average (BLS, 1997):

$$R_{ia}^{t,t-1} = e^{\sum_{j \in S_{ia}} w'_{iaj} \ln \left(P_{iaj,t} / P_{iaj,t-1} \right)}$$

Here S_{ia} represents the sample for item *i* in area *a*, *P* represents the price, and w' represents the quote-level sampling weight of sample item *j,* normalized to the same sample rotation base for all quotes in the item-index area.

The second stage of aggregation for the CPI is a Laspeyres aggregation of sub-indexes:

$X_{IA}^{t} = \sum_{i,a \in IA} RI_{ia} X_{ia}^{t}$, where RI_{ia} is the expenditure-weighted relative importance of item *i* in area *a*.

In the C&S sample design application, we were concerned with the short term or δ-month percentage price change: $PC(I, A\cdot, t, t-\delta) = 100\left[\frac{X_{IA}^{t}}{X_{IA}^{t-\delta}} - 1\right]$.

1.2 History

Hansen, Hurwitz, and Madow (1953), Kish (1965), and Cochran (1977) present several examples of sample design optimization via cost and error modeling. Groves (1990) discusses sample design for social surveys.

Cost and sampling error models were first formulated for the C&S sample design for the 1978 CPI Revision (Westat, 1974). Item classes comprised two categories—food, and other goods and services—and sample size allocation were made for six PSU classes. Selection of the sample design implemented in that revision was based on evaluation of a number of alternative designs. The 1987 CPI Revision (CPIR) redesign (Leaver, et al., 1987) expanded on this approach, refining models for eight item groups and ten PSU classes. This implementation relied on use of administrative records and modeled estimates for parameters in cost and variance functions. Solution methods used nonlinear programming techniques to identify local minimizers of a modeled relative variance function, under varying assumptions of annual inflation and price change interval. This was further expanded upon for the 1998 CPI sample revision, in which frames from the TPOPS survey were first used. (Leaver, et al, 1999.) The approach maintained with this sample design generally follows that taken for the 1998 sample and weighting revision.

2. The Design Problem

The primary objective of the C&S sample allocation process is to determine values for all sample design variables which will minimize the sampling variance of price change for the C&S component of the CPI. Sample design variables for this problem are the number of ELIs to select in each item stratum and the number of outlets to select per POPS category-replicate panel in each sample PSU. Though only one-eighth of the item and outlet sample is reselected in any semiannual rotation, we resolve the entire allocation each time. In this process, the number of PSUs, the number of replicate panels per PSU, and the item stratification are treated as previously determined and thus design constraints (Williams et al., 1993; Lane, 1996; Williams, 1996.) Item stratification is based on similarity of price movement, price change variability, and historic publication requirements (Lane 1996.) Similarly, the geographic sample designs for the 1998 revision and its planned successor were based on known publication requirements and knowledge of variability of price change among urban population centers. Cost and variance models provided herein assisted in determining the optimal mix between self- and non-self representing cities in the most recent area redesign (Johnson, 2003.)

To render the problem tractable to nonlinear programming solvers available, we made some simplifying assumptions. Item strata were divided into thirteen item groups: four subgroups for food at home, food away from home and alcoholic beverages, household furnishings and operations, fuels and utilities, apparel, transportation less motor fuel, motor fuel, medical care, education and communications, and the combined group of recreation and other commodities and services. PSUs were divided into 15 groups according to size and number of replicate panels. It was assumed that the same outlet sample sizes and item sample selection sizes would apply to all PSUs within the same group. This reduced the first-level allocation problem to one of determining the values of the design variables: (a) the number of ELI selections per item group-replicate panel within each PSU group, denoted by $\{K_{ij}$, i=1,...,13, j =1,...,15$\}$, and (b) the designated number of outlet selections per item group-POPS category-replicate within each PSU group, denoted by $\{M_{ij}$, i=1,...,13, j =1,...,15$\}$, which would minimize a modeled price change sampling variance, subject to additional allocation and cost constraints.

The variance of price change for all C&S items was modeled as a function of the design variables, as were total annual data collection costs. Nonlinear programming methods were then used to determine optimal values for the design values under various cost, variance, and sample share constraints.

3. The Sampling Variance Function

For the purposes of the allocation problem, we write the All U.S. City Average C&S price change estimator as $PC(\cdot,\cdot,t,t-\delta) = \sum_i \sum_k RI_{i,k} PC(i,k,t,t-\delta)$, where $PC(i,k,t,t-\delta)$ is the estimated price change from time t-δ to t for item group i and index area k, and $RI_{i,k}$ is the population-expenditure-weighted relative importance of item group i in index area k. Deriving a component form of the variance of this price change estimator, accounting for the stages of sampling described above, would be extremely difficult. Rather than this direct route, we have taken a more indirect, modeling approach described below. Four sources of variation were modeled: PSU selection, item selection, outlet selection, and other sources, such as sampling within the outlet.

The variance function for the CPI revision was modeled for index areas. The variance model assumes that the total variance of price change for item group i within index area k can be expressed as a sum of four components:

$$\sigma^2_{i,k} = \sigma^2_{psu,i,k} + \sigma^2_{eli,i,k} + \sigma^2_{outlet,i,k} + \sigma^2_{error,i,k}$$

where

$\sigma^2_{i,k}$	is the total variance of price change for item group i in index area k,
$\sigma^2_{psu,i,k}$	is the component of variance due to sampling PSU's in non-self-representing areas, 0 for self-representing areas,
$\sigma^2_{eli,i,k}$	is the component of variance due to sampling of ELIs within item strata,
$\sigma^2_{outlet,i,k}$	is the component of variance due to sampling of outlets, and
$\sigma^2_{error,i,k}$	is a residual component of variance attributable to other aspects of the sampling process, including the final stage of within-outlet item selection, called disaggregation .

We assume that the variance of price change of an individual sampled unit or quote has the same structure:

$$\sigma^2_{unit,i,k} = \sigma^2_{unit,psu,i,k} + \sigma^2_{unit,eli,i,k} + \sigma^2_{unit,outlet,i,k} + \sigma^2_{unit,error,i,k},$$

where

$\sigma^2_{unit,i,k}$	is the total variance of price change of an individual sampled unit or quote for item i in area k,
$\sigma^2_{unit,psu,i,k}$	is the component of unit variance due to sampling PSU's in non-self-representing areas,
$\sigma^2_{unit,eli,i,k}$	is the component of unit variance due to sampling of ELIs within item strata,
$\sigma^2_{unit,outlet,i,k}$	is the component of unit variance due to sampling of outlets, and
$\sigma^2_{unit,error,i,k}$	is the corresponding residual component of unit variance.

It follows that each component of $\sigma^2_{i,k}$ can be written in terms of its corresponding unit variance components:

$$\sigma^2_{i,k} = \sigma^2_{unit,psu,i,k}/N'_k + (\sigma^2_{unit,eli,i,k}/(N_k H_k K_{i,k}))NC_i$$

$$+\sigma^2_{unit,outlet,i,k}/(N_k H_k M'_{i,k}) + \sigma^2_{unit,error,i,k}/(N_k H_k K_{i,k} M_{i,k})$$

where

N_k	is the number of PSU's in index area k,
N'_k	is the number of non-self-representing PSU's in the index area,
H_k	is the number of replicate panels per PSU in the index area,
$M'_{i,k}$	is the number of unique in-scope outlets selected per PSU-replicate
NC_i	is the percent of strata in item group i which are non-certainty strata.

Note that the expected number of quotes per PSU-replicate panel-item group is estimated by the product of the designated outlet sample size and the number of item stratum selections, $M_{ik} \cdot K_{ik}$.

Thus the sampling variance of price change for the All U.S. City Average C&S index is

$$\sigma^2_{TOTAL} = \sum_k \sum_i RI^2_{i,k}\sigma^2_{i,k}$$

4. The Cost Function

The total annual cost of the C&S portion of the CPI includes costs of initiation data collection and processing, personal visit and telephone pricing, and pricing data processing, each of which were developed in terms of outlet and quote related costs. For PSU group j and item group i, outlet related costs for initiation are:

$$CI_O(M_{ij}, K_{ij}) = 0.25 N_j \cdot H_j \cdot C_{O,i} \cdot (a_{ij} M^2_{ij} + b_{ij} M_{ij}),$$

where

$CI_O(M_{ij}, K_{ij})$	is the outlet-related initiation cost for item group i in PSU group j
N_j	is the number of PSUs in group j,
H_j	is the number of replicates per PSU in PSU group j,
$C_{O,i}$	is the initiation cost per outlet for item group i,

and $(a_{ij} M^2_{ij} + b_{ij} M_{ij})$ is an overlap function used to predict the number of unique sample outlets, accounting for the overlap of elements in the outlet sample within and between item groups for a replicate panel. The number 0.25 accounts for the rotation or reselection of one-fourth of the sample each year.

Quote related initiation costs are:

$$CI_Q(M_{ij}, K_{ij}) = 0.25 N_j H_j \cdot SeasI_i \cdot C_{Q,i} \cdot M_{ij} \cdot K_{ij} \cdot NR_i$$

where

$CI_Q(M_{ij}, K_{ij})$	is the quote-related cost of initiation for item group i in PSU group j,
$SeasI_i$	is a seasonal items initiation factor for item group I,
$C_{Q,i}$	is the initiation cost per quote for item group i, and
NR_i	is the outlet initiation response rate for item group i.

The costs of ongoing price data collection and processing were also developed in a similar form. Outlet related costs are:

$$CP_O(M_{ij}, K_{ij}) = MB_{ij} \cdot N_j \cdot H_j \cdot NR_i \cdot (a_{ij} M^2_{ij} + b_{ij} M_{ij}) \cdot [(C_{PV,O,i} + C_{PV,T,i}) \cdot (1 - R_{T,O,i}) + C_{T,O,i} \cdot R_{T,O,i}]$$

where

$CP_O(M_{ij}, K_i)$	is the total outlet-related cost for ongoing pricing,
$C_{PV,O,i}$	is the cost for a personal visit for pricing per outlet for item group i,

$C_{PV,T,i}$	is the travel cost for a personal visit for pricing per outlet for item group *i*,
$R_{T,O,i}$	is the proportion of outlets priced by telephone for item group *i*,
$C_{T,O,i}$	is the per outlet cost for telephone collection,
MB_{ij}	is a factor to adjust for the monthly/ bimonthly mix of outlets and quotes by PSU and major product group.

And quote related costs are:

$$CP_Q(M_{ij}, K_i) = MB_{ij} \cdot N_j \cdot H_j \cdot M_{ij} \cdot K_i \cdot NRQ_i \cdot SeasR_i \cdot [C_{PV,Q,i} \cdot (1 - R_{T,Q,i}) + C_{T,Q,i} \cdot R_{T,Q,i}]$$, where

$CP_Q(M_{ij}, K_{ij})$	is the total quote-related cost for ongoing pricing,
$C_{PV,Q,i}$	is the per-quote cost for a personal visit for pricing,
$R_{T,Q,i}$	is the proportion of telephone collected quotes for item group *i*,
$C_{T,Q,i}$	is the per-quote cost for telephone collection for item group *i*, and
$NRQ_{,i}$	is the quote level pricing response rate for item group *i*.
$SeasR_i$	is a seasonal items ongoing pricing factor for item group *i*.

The total cost function associated with data collection and processing for C&S, summed over all item groups and PSU groups, is then given by:

$$C_{Total} = \sum_{i,j} [CI_O(M_{ij}, K_i) + CI_Q(M_{ij}, K_i) + CP_O(M_{ij}, K_i) + CP_Q(M_{ij}, K_i)]$$

With this we write the sample design problem as the nonlinear programming problem:

Minimize $\sigma^2_{Total}(\{K_i\},\{M_{ij}\})$ subject to:

$C_{Total} \le \$5,300,000$

$K_{ij} \ge$ *Number of item strata in item group I, PSU group j*

$K_{ij} \le$ *Maximum #r of item hits for item group i,*

$M_{ij} \ge 2$, *i=1,..., 13, j=1,...,15*

Average item hits per stratum ≥ 9

5. Model coefficients

Components of price change variance were computed using weighted restricted maximum likelihood components of variance estimation methods and C&S price micro-data collected and updated periodically (Shoemaker and Johnson, 1999 and Shoemaker, 2001). Though estimates were developed for 6- and 12-month price change for the 13 item groups for each index area, we chose to work with the 6-month estimates for these models, as they appeared to be more stable.

Estimates of components of the cost function were developed using a variety of agency administrative records and the C&S database. Fiscal year data were used to obtain a total cost per outlet to initiate, and then hourly staff utilization data provided by the BLS Office of Field Operations (OFO) and hourly compensation rates from the CPI program office were used to produce a per-hour cost of initiation. Outlet unit costs and quote unit costs of initiation, by item group, were derived by taking these per-outlet and per-hour costs and combining them with data obtained from a data collection time-and-travel study conducted in 1987. Travel costs per quote, by item group, were estimated by using an overall travel cost per outlet and again comparing it to data from the time-and-travel study.

Pricing costs were figured in a similar manner. Distinctions between personal visit and telephone collection of data were made based on total personnel hours cost accounting information from OFO, estimates of hourly compensation rates from the CPI program office, and an analysis conducted within the Price Statistical Methods Division. Outlet initiation survival rates and quote and outlet retention rates for each item group were developed from field initiation records and ongoing pricing records, and are periodically updated with data from the C&S production data base.

To date, developing and maintaining cost accounting estimates for CPI initiation and data collection remain among the more challenging elements of the sample allocation process. Agency cost accounting currently occurs at a higher level than the analysis described herein. OFO record-keeping on staff utilization for data collection exists on separate systems for full- and part-time staff. An interdivisional team is currently being formed within the CPI program to describe and specify system requirements to more readily access current staff utilization records and tie them to detailed computer assisted data collected (CADC) initiation and repricing schedule data. It is envisioned that CADC record-keeping will facilitate updating the time-and-travel estimates still in use.

"Overlap" functions are modeled to project the number of unique outlets realized in sample selection as a function of designated sample size. These are obtained by regressing the number of unique outlets on outlet sample sizes obtained in simulations of sampling procedures for each PSU and item group, using CPOPS and TPOPS sampling frames for the most recent rotations for each PSU-item group (Johnson et al., 1999). As the introduction of TPOPS frames has been gradual and staggered among PSUs, overlap functions are re-estimated with each semiannual sample rotation, and incorporate information for the most recently collected frames.

An unforeseen consequence of the introduction of the TPOPS in the CPI resulted from the expansion of the number of POPS categories from 140 in CPOPS to 215 in TPOPS, representing over a 50% increase in categories and frames. This effectively introduced a proliferation of unique outlets in real and simulated sample selections due to the fact that outlet sampling is performed independently for each PSU-replicate panel and POPS category. As outlet-related costs—and in particular, travel-related costs—dominate the allocation cost function, first-level allocation solutions tended over time to more rapidly consume sample resources. This resulted in fewer sample outlets per PSU-POPS being allocated, with an attendant shift toward greater numbers of item selections. Over time, the number of designated outlets per PSU-replicate panel has dropped to the minimal allocation in many item-PSU groups and the number of quotes collected per sample outlet has dropped as well.

We are now considering recombining some TPOPS categories, where doing so seems reasonable, in order to increase sample sizes and sampling efficiency in outlets where several ELIs can be priced. The first group of items we have considered is the group comprising the eight fresh fruits and vegetables (FFV) strata. Under the CPOPS structure, these eight item-strata belonged to one category, but under the new TPOPS structure are assigned to eight separate categories, which, in any given PSU, are rotated at the same time. Analysis of TPOPS frames indicate a reasonable degree of overlap among the frames for these categories, and experiments in allocation simulations with these categories treated as one category indicate a greater level of efficiency in data collection for this set of items. See Table 1.

6. Problem Solution

SAS NLP is used to find a local minimum to the design problem. For each semiannual rotation, a solution is found using components of variance estimates for 6-month price change. For each item group, the number of item selections is bounded below by the number of strata in the item group and above by a ceiling of 133% of the item group's pre-1998 revision item-stratum hits allocation. Additional constraints include a minimum average per-stratum sample size of 9, designated to avoid small sample bias.

Item hits are then distributed among item strata within each item group, initially using a Neymann allocation within major group, which takes into account stratum-level price change variances and expenditure relative importances. Then these allocations are adjusted, with consideration given to differences in response rates among the item strata within each item group, as well as to special problems identified by commodity analysts and field staff.

Outlet hits are initially uniformly allocated to each POPS category in each major group in each PSU-replicate panel. Designated outlet sample sizes are then adjusted among the various TPOPS categories within item groups to manage variation in expected response rates and respondent burden. In the case that item hit allocations are trimmed to accommodate respondent cooperation considerations, outlet sample sizes are adjusted upwards. Final item and outlet allocations are then compared with those from the previous rotation to determine a reasonable degree of consistency across time.

Although weighting revisions in the CPI occur every two years, incremental revisions occur in the sample allocation with each semiannual TPOPS rotation.. Table 1 below characterizes the TPOPS rotation sample design for the August 2004 rotation, contrasting it with the design produced for same rotation, but using overlap functions based on the collapsed fresh fruit and vegetables TPOPS frames.

Fresh fruits and vegetables belong to the design item group 3, an item group that exhibits considerable price change volatility and variability. The fine degree of stratification for this group, coupled with its price change variability and the program constraint that two outlets be selected per TPOPS category per PSU-sample replicate, drove the sample cost share for this item group to be over 6.5 percent, even though the expenditure share for this group is only 1.72 percent. Collapsing the eight categories produced a design solution with a lower predicted variance, in which sample resources "freed" from FFV were redistributed over the other item groups. The modeled variance for item group 3 increased, due largely to the reduction in unique outlets, but modeled variances for other item groups decreased. The relative quote share for item group 3 maintained at about the same level, however, as more quotes were allocated for each selected sample outlet. The average number of quotes collected per outlet jumped from 2.2 to 6.3, but because of the drop in the number of unique sample outlets allocated to FFV, the relative sample resource cost share for this group dropped to below 3 percent.

We proceeded to investigate the potential for further sample efficiency gains by collapsing additional TPOPS categories, in particular among food at home and some household furnishings and operations item strata. These results are presented in summary in Table 1 below. Collapsing the FFV categories contributed the greatest boost to sample efficiency, yielding a nearly 6 percent greater sample size, in terms of available quotes, for the same sampling budget. However, collapsing the additional 30 categories into 8 categories provided an additional boost in efficiency, yielding, with FFV, a nearly 20 percent increase in available quotes.

7. Conclusions

Cost-variance modeling has proven useful in allocating the C&S sample for the CPI, which by nature of its size and program and publication requirements, demands strategies to handle the constraints so imposed. Our most significant program constraints include the high degree of stratification with which we must deal, both in our item structure and outlet category structure and programmatic difficulty in accessing management information data to develop and validate our cost functions. Directions for further inquiry include further revision of TPOPS category definitions to improve sample efficiency, revision of item groupings for design purposes, and adaptation of our current available management information devices to meet our data needs.

8. Acknowledgments

The authors would like to thank Alan Dorfman, Kirk Hagemeier, and William Johnson for their careful reading of earlier drafts of this paper and for their helpful comments. The authors also would like to thank William Johnson and Robert Eddy for their support on this project.

References

Bureau of Labor Statistics (2003), *BLS Handbook of Methods*, Washington. DC, http://www.bls.gov/opub/hom/home.htm

Cochran, W. G. (1977). *Sampling Techniques,* Wiley, New York.

Groves, Robert (1990) *Cost and Error Modeling in Social Science Surveys*, Wiley, New York.

Hansen, Morris G., Hurwitz, William N., and Madow, William G. (1953) *Sample Survey Methods and Theory*, Wiley, New York.

Johnson, William H., Leaver, S.G., and Benson, T.S. (1999) "Modeling the Realized Outlet Sample for the Commodities and Services Component of the U.S. Consumer Price Index," *Proceedings of the Section on Government Statistics, American Statistical Association,* pp 304-308.

Kish, Leslie (1965) *Survey Sampling*, Wiley, New York.

Lane, Walter (1996) "Changing the CPI Item Structure," U.S. Bureau of Labor Statistics http://www.stats.bls.gov/mlr/cpiwl001.htm

Leaver, S., Weber, W., Cohen, M., and Archer, K. (1987) "Item-Outlet Sample Redesign for the 1987 U.S. Consumer Price Index Revision," *Proceedings of the46th Session, International Statistical Institute.* Tokyo, Vol. 3, pp.173-185.

Leaver, S.G., Johnson, W. H., Baskin, R., Scarlett, S., and Morse, R. (1996) "Commodities and Services Sample Redesign for the 1998 Consumer Price Index Revision," *Proceedings of the Section on Survey Research Methods, American Statistical Association*, Vol. I, pp.239-244**.**

Leaver, S. G., Johnson, W. H., Shoemaker, O. J., and Benson, T. S. (1999) "Sample Redesign for the Introduction of the Telephone Point of Purchase Survey Frames in the Commodities and Services Component of the U.S. Consumer Price Index," *Proceedings of the Section on Government Statistics and Section on Social Statistics*, *American Statistical Association*, pp. 292-297**.**

Leaver, S. and Valliant, R. (1995) "Chapter 28: Statistical Problems in Estimating the U.S. Consumer Price Index," *Business Survey Methods*, Brenda G. Cox, et al., editors, Wiley, New York.

Shoemaker, O. J., and Johnson, W. H. (1999) "Estimation of Variance Components for the U.S. Consumer Price Index", *Proceedings of the Section on Government Statistics, American Statistical Association*, pp. 298-303.

Shoemaker, O. J. (2001) "Estimation of Variance Components for the U.S. Consumer Price Index: A Comparative Study", *Proceedings of the Joint Statistical Meetings, American Statistical Association*.

Westat, Inc. (1974). "Proposals for and Evaluation of Alternative Designs for Allocation of CPI Pricing

Efforts to Items, Outlets, and within Outlets," CPIR-WS-4, Rockville, Maryland.

Williams, J. L., Brown, E. F., Zion, G. R. (1993), "The Challenge of Redesigning the Consumer Price Index Area Sample," *Proceedings of the Survey Research Methods Section, American Statistical Association* (Vol. 1), pp. 200-205.

Williams, Janet L. (1996), "The Redesign of the CPI Geographic Sample," U.S. Bureau of Labor Statistics http://www.stats.bls.gov/mlr/cpijw001.htm

Table 1. Comparison of Modeled Variance and the Distribution of Sample Resources between August 2004 TPOPS Allocation of the CPI C&S Design and Revised Version with Fresh Fruits and Vegetables and 30 Additional TPOPS Categories Collapsed into 9 Categories

Item Group	Modeled Variance without FFV Collapsed, Aug 2004	Reduction in Variance due to Collapsing FFV (%)	Reduction in Variance due to Collapsing in All 9 Super Categories (%)	Share,Total Costs, Aug 2004 TPOPS Design (%)	Share,Total Costs, Aug 2004, FFV Collapsed (%)	Share, Total Costs, Aug 2004, 38 Categories Collapsed to 9 (%)	Share, Total Quotes, Aug 2004 TPOPS Design (%)	Share, Total Quotes, Aug 2004 FFV Collapsed (%)	Share, Total Quotes, Aug 2004, 38 Categories Collapsed to 9 (%)	2002-2003 CE Relative Importance (relative to total C&S)
Total, All Items less Rent and Owners' Equivalent Rent	0.0119	3.55	8.93	100.00	100.00	100.00	100.00	100.00	100.00	100.00
Food at home MG1	0.2077	2.78	-24.66	8.06	8.18	3.57	7.32	7.20	7.42	3.06
Food at home MG2	0.2762	2.44	-4.94	4.66	4.73	3.32	8.31	8.25	8.66	3.08
Food at home MG3	0.3221	-27.99	-16.22	6.59	2.97	3.17	6.64	6.49	6.37	1.72
Food at home MG4	0.1453	0.58	-2.18	10.93	11.05	9.89	10.11	9.65	8.75	4.68
				30.24	26.94	19.95	32.39	31.60	31.20	12.54
Food away + Alcoholic Beverages	0.1404	5.99	14.54	3.83	4.34	4.84	3.88	4.18	4.10	9.18
Household Furnishings and Operations	0.0866	3.83	8.53	11.55	11.91	12.09	14.25	14.21	13.09	11.74
Fuels and Utilities	0.2322	3.62	8.05	3.67	3.96	4.56	5.66	5.85	6.04	6.80
Apparel & Upkeep	0.1552	4.20	13.06	11.28	11.42	12.57	6.42	6.40	7.84	5.46
Transportation less Motor Fuels	0.0689	5.84	14.66	11.78	12.90	15.71	10.08	10.64	11.64	19.26
Motor Fuels	0.8276	1.40	3.78	1.68	1.90	2.04	2.36	2.54	2.39	4.90
Medical Care	0.1283	5.71	13.93	3.89	4.27	4.51	4.75	4.97	4.69	8.61
Education and Communication	0.0932	2.76	8.56	7.51	7.66	7.99	7.45	7.29	6.76	8.40
Recreation + Other C&S	0.0581	1.28	8.02	14.55	14.70	15.73	12.76	12.33	12.25	13.11
Total Quotes, Repricing,No Categories Collapsed	827,663	% Increase Quotes for Repricing								
Total Quotes, Repricing, FFV Collapsed	871,081	5.25								
Total Quotes, Repricing, All 9 Categories Collapsed	987,039	19.26								