ESTIMATION OF VARIANCE COMPONENTS FOR THE U.S. CONSUMER PRICE INDEX: A COMPARATIVE STUDY

Owen J. Shoemaker, U.S. Bureau of Labor Statistics, 2 Mass Ave., NE, Room 3655, Washington, DC 20212

Key Words: Weighted REMLs; Sample Design; PSU

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

For every new sample for the commodities and services (C&S) segment of the U.S. Consumer Price Index (CPI), the Bureau of Labor Statistics produces a C&S sample design in which the outlets and items are allocated in an optimal fashion. This item-outlet optimization requires the estimation of components of variance for the three factors in the design: non-certainty primary sampling units (PSUs), item-strata and outlets. A fourth component of variance is the error term. The total variance of these unit components of variance, divided by their respective number of PSUs, item-strata, outlets and quotes, is then minimized by the optimal number of respective outlets and item hits, as constrained by a cost function.

1. The Design

Commodities and Services (C&S) accounts for 72.5% of the CPI (as measured in expenditure shares), with Housing accounting for the remaining 27.5%.[1] The first stage of the overall design is the PSU sample selection. This stage is common to both Housing and C&S. The CPI survey is conducted in 87 PSUs. The 31 largest A-level PSUs are selected with certainty. The 56 smaller (B- and C-level) PSUs are then selected with probability proportional to size[2] (pps) within their respective regions: Northeast, Midwest, South or West. The components of variances themselves are calculated at the Area by Major Group level, where each A-level PSU is its own Area, and where non-certainty PSUs are grouped together by size and region to form seven other Areas. Thus, PSU is a random variable only in the Areas formed from non-certainty PSUs.

Two further independent sampling stages then occur in C&S: outlet allocation and item allocation, within each Area-Replicate combination. (Each Area sample is made up of at least two independently sampled replicates, with each replicate containing all the items in the CPI.) The outlet sample is based on the Telephone (or Consumer) Point of Purchase Survey (TPOPS or CPOPS), both of which are conducted by the Bureau of the Census for the BLS. (For the 1998 Revision, TPOPS began replacing CPOPS as the BLS survey methodology of choice.) The item sample is based on the Consumer Expenditure Survey (CE), also conducted by Census for the BLS. Outlets are selected in 217 TPOPS categories using a systematic pps[3] sampling scheme. Items are selected in 13 Major Groups (see **Table 2**) using a stratified systematic pps sampling scheme. The variance components capture the variability resulting from each of these sampling stages. For example, a set of three components of variance (Item, Outlet & Error) is calculated for Apparel (Major Group 7) in Atlanta (an A-level Area), and a set of four components of variance (PSU, Item, Outlet & Error) is calculated for, say, Medical in the B-level cities in the Northeast. The estimator is the average price change for each Area by Major Group category. The total variance of each estimator is modeled as the sum of the four components (or three, in the certainty A-level PSUs):

$$\sigma^2_{j,k} = \sigma^2_{psu,j,k} + \sigma^2_{item,j,k} + \sigma^2_{outlet,j,k} + \sigma^2_{error,j,k}$$

for each Major Group j and Area k. We then assume that the variance of price change of an individual sampled unit or quote has the same structure:

[1] Bureau of Labor Statistics, *CPI Detailed Report (Dec 2000),* p. 16.
[2] Size here equals population.

[3] For outlet and item-within-TPOPS-group sample selection, size equals expenditures.

$$\sigma^2_{unit,j,k} = \sigma^2_{unit,psu,j,k} + \sigma^2_{unit,item,j,k} + \sigma^2_{unit,outlet,j,k} + \sigma^2_{unit,error,j,k}, \text{ and finally}$$

$$\sigma^2_{j,k} = \sigma^2_{unit,psu,j,k} / N_p + \sigma^2_{unit,item,j,k} / N_h + \sigma^2_{unit,outlet,j,k} / N_o + \sigma^2_{unit,error,j,k} / N_e$$

where each N_i stands for an appropriate number of PSUs, items, outlets or quotes.

The sampling variance of the price change for the All U.S. C&S Index is computed as

$$\sigma^2_{TOTAL} = \sum_j \sum_k RI^2_{j,k} \cdot \sigma^2_{j,k}$$

where the $RI_{j,k}$'s are the relative importances for each Area by Major Group combination, factoring in both the relative expenditures and the relative populations. ($\sum_j \sum_k RI_{j,k} = 1.$) It is this σ^2_{TOTAL} that is minimized in the optimization procedure. It is, however, the unit-level components of variance (the $\sigma^2_{unit,\bullet,j,k}$'s above) that are calculated and analyzed. We use weighted REML (Restricted Maximum Likelihood) estimation as our chosen methodology.[4]

2. The Weights

In the CPI, price changes are not equally weighted. A full system of weights, essentially based on expenditure shares, is used at both the basic and aggregate levels of index calculation. At the basic item-stratum level, within each Area, a group of prices-with-weights are combined into a sub-index (i.e., a price relative), using either a Laspeyres or a Geometric mean formula. These price relatives update the lower level indexes. These updated indexes are then aggregated, using weighted sums (called cost weights), into various higher level indexes, which then become the published indexes of the CPI, including the All US–All Items index itself.

The random variable of interest for calculating our components of variance is not, however, a composite price relative, but an individual price relative (PR) expressed as an individual percent price change (PC). Thus, PC = (PR–1)*100. A price relative is generally a weighted composite of several individual price changes. An outlet variance component could not, however, be produced without having price change observations at the quote level. Moreover, a unit component of variance requires exactly that: unit-level observations. Fortunately, the individual price quotes have basic "final" weights (FNLWs) assigned to them in the database. In addition, we have cost weights for every item-stratum in each Area by Major Group category, from which we can calculate a relative importance ($RI_{j,k}$) for each Area by Major Group category. The resulting individual quote weight is simply: $QWT_{i,j,k} = FNLW_{i,j,k} * RI_{j,k}$, with *i* referring to the individual quote. These QWTs constitute the weight structure for the REML estimations of the variance components.

One last crucial decision, however, has to be made: whether to utilize these QWTs as computed, or to re-scale them in some way. In their paper, "Weighting For Unequal Selection Probabilities In Multilevel Models", Pfeffermann, et al (1998), offer, as a candidate re-scaling factor, the mean of the weights: $W_{i,j,k} = QWT_{i,j,k} / \lambda$, where $\lambda = \sum_i QWT_{i,j,k} / n_{j,k}$, where $n_{j,k}$ is the sample size. Besides retaining consistency and probably reducing small sample bias[5], this particular mean-scaled weight structure leaves the linear model results themselves at the same unit level at which we want our components of variance. I.e., we get $\sum_i W_{i,j,k} = n_{j,k}$, just as the trace(**I**) = $n_{j,k}$ in any *un*weighted model (with **I** being the identity matrix). This re-scaling clearly retains the full information content of the original weights. Moreover, the use of these weights seems to have produced two additional side-benefits: (1) more stable components of variance across time periods, and (2) a final result, the σ^2_{TOTAL} , that compares favorably in magnitude with an independent variance calculation for the All US– All Items Index.[6]

3. The Model

---

[4] O. J. Shoemaker & W. H. Johnson, *Estimation of Variance Components for the U.S. Consumer Price Index (1999)*, pp. 3-5. In Section 4, the reasons for choosing REML estimation are outlined; in Section 5, the derivation of the components of variances using REML estimation is fully analyzed.

[5] The reduction of small sample bias is the main object of the inquiries and recommendations in the Pfeffermann, et al paper.

[6] Using a stratified random group methodology from a replicate structured index data base, the All-US–All-Items Index standard error for a 6-month price change (using Year 2000 data) was ≈ 0.12, as compared to the design's final optimal SE of ≈ 0.10.

The model we use treats all three effects as random. The design is unbalanced. We let y_{ijkl} be the observed unit percent price change, between time t and time t–6, for quote *i* within PSU *l*, item *j*, and outlet *k*. (Note that the *j* and *k* subscripts here are different from the *j* and *k* subscripts in Section 3.)

$$y_{ijkl} = \mu + p_l + h_j + o_k + e_{ijkl} ,$$

where μ is a fixed effect
$p_l \sim N(0, \sigma^2_{unit, psu})$
$h_j \sim N(0, \sigma^2_{unit, item})$
$o_k \sim N(0, \sigma^2_{unit, outlet})$
$e_{ijkl} \sim N(0, \sigma^2_{unit, error})$,

with p, h, o and e all independent of each other.

A special case of this model, when there is only one PSU in play, can be written:

$$y_{ijk} = \mu + h_j + o_k + e_{ijk} \qquad \text{(since } p_l = 0\text{)}$$

This special case model applies to all the A-sized Areas. The main model encompasses all the rest: i.e., the B- and C-sized Areas.

4. Dataset Comparisons across the 1998 Revision — Area Level Comparisons

Price data from the C&S Archive Database, from mid-1993 through mid-1997, were compiled, month by month, into one pre-Revision dataset. A second dataset was similarly compiled from mid-1997 through 2000 quote-level data, with the data in the second dataset coming from not only a new 3½ year time frame but from an entirely new C&S sample (the 1998 Revision), which includes a newly revamped Item-strata structure as well. In each dataset, like price quotes (collapsed across different versions of a unique quote where permitted) were tracked to each other over time. From these two databases, we then produced two separate sets of 2-, 6- and 12-month variance components The 6-month results, for a variety of reasons, were selected as the variance components of record. The 2-month price changes were considered too variable, and maintaining a high enough level of sample size over the 12-month periods was considered too problematic. No month that held fewer than 20 quotes at the Area–Major Group cell level was used in the final results.

For the 6-month variance components the average sample size for each set of REML estimates was around 80 quotes for the A-level Areas in the earlier dataset and around 70 quotes in the later dataset. The B-level Areas averaged around 500 quotes per cell in the first dateset, around 360 per cell in the later dataset. The C-level Areas averaged around 135 quotes per cell in the first, 100 in the second. (See **Table 1** below.) Every attempt was made to use similar methods in both time frames in the extraction of the data and construction of the datasets. The bulk of both datasets, for example, came from comparable three-year periods: July 1993 through Dec 1996 for the pre-Revision dataset, July 1997 through Dec 2000 for the post-Revision dataset. There was a major revamping of the Item structure of the C&S sample in the 1998 Revision, but we scrupulously worked at conforming the Item structure of the pre-Revision with the new Revision Item structure, especially in regard to fitting Item-strata similarly into the same set of Major Groups for both datasets.

The noticeable difference in sample size at the (model) cell level between the two datasets becomes our first important result. The lowest cell level for CPI index calculations is the Area–Item-Stratum level. The cell level for our REML estimates, and thus for our variance components, is the Area–Major Group level. At this level the sample sizes in the first dataset range anywhere from 1 to 50 percent higher than the sample sizes in the second and more recent dataset. Since variances in general are proportional to *n*, it might be assumed that the new set of variance components would reflect this difference in sample sizes. And they do. In the **PCT INC** column, in **Table 1** below, the percentage increase in the mean total variances at the various aggregate Area levels are listed. The +35% increase overall does seem to match up with the overall percentage increase of sample size in the first dataset over the second. The overall sample size increase is approximately +23%, with the three Area sizes (A-, B-, and C-) being +12%, +38%, and +32%, respectively.

Note that the number of monthly collection periods (**CP**s), on average, range higher for the second dataset. The first dataset is based on an exclusively CPOPS rotation schedule which (unlike TPOPS) allows for complete rotation, in and out, of entire PSUs. These continuity breaks in CPOPS produce more **CP**s which fail to meet the $n \geq 20$ criterion in the first dataset than in the second.

Since the original sets of variance components (as used in the actual C&S Sample Design) are averaged across these monthly time periods, this difference may have had some mitigating effect on the variance differentials in the opposite direction. An analysis of the variation across these collection periods, however, shows a higher variability (nearly 40% higher standard error) across the second set of variance components -- where the number of collection periods is on average larger -- than across the first set. The straightforward statistical relationship between a larger sample size and a smaller variance seems to predominate here.

The Mean Total Variance columns in **Table 1** list the Area-level averaged variance components (after the components themselves have been summed at the lowest Area-Major Group cell level). Note the remarkable similarity of the results within the two respective time periods for any of these Area-level results. Aside from a uniformly higher (35% on average) set of variance components, there are no significant differences across any of these aggregated Area levels so far as the variance components are concerned.

Table 1. Average Variance Components —— Aggregate AREA Levels

PSU-GROUP	Mean Total Variance ‘94-‘96	Mean Total Variance ‘98–‘00	PCT INC	Mean Sample Size ‘94-‘96	Mean Sample Size ‘98-‘00	Mean # of CP’s 94-96	Mean # of CP’s 98-00
All A-Size Cities	**0.02620**	**0.03541**	**+35**	**79.6**	**71.1**	**19.4**	**28.0**
All B-Size Cities	**0.02509**	**0.03209**	**+28**	**499.5**	**361.4**	**21.0**	**29.9**
All C-Size Cities	**0.02453**	**0.03251**	**+33**	**135.6**	**102.6**	**18.3**	**29.7**
ALL U.S.	**0.02595**	**0.03483**	**+35**	**128.2**	**104.2**	**19.5**	**28.4**

Note **CP** = Collection Period

Table 2. Average Variance Components —— by MAJOR GROUP

MAJOR GROUP	Mean Total Var ‘94-‘96	Mean Total Var ‘98-‘00	PCT INC		Mean Sample Size ‘94-‘96	Mean Sample Size ‘98-‘00	PCT DEC
Food – Staples	0.0305	0.0395	**+30**		125.3	105.7	**–16**
Food – Meats	0.0482	0.0556	**+15**		192.6	151.8	**–21**
Food – Fruits / Vegetables	0.0691	0.0654	**– 5**		153.5	118.6	**–23**
Food – Other	0.0276	0.0341	**+23**		168.6	157.6	**–7**
Food Away from Home	0.0160	0.0188	**+17**		176.4	64.9	**–63**
Household Furnishings	0.0191	0.0433	**+127**		128.8	105.6	**–18**
Utilities / Fuels	0.0131	0.0331	**+152**		86.9	89.5	**+3**
Apparel	0.0396	0.0445	**+12**		94.1	93.6	**–1**
Transportation	0.0199	0.0260	**+31**		128.8	103.3	**–20**
Gasoline	0.0130	0.0147	**+12**		55.9	54.3	**–3**
Medical	0.0144	0.0272	**+89**		147.9	98.2	**–34**
Educ / Communications	0.0093	0.0221	**+137**		75.7	83.6	**+10**
Entertainment / Miscellan	0.0173	0.0285	**+64**		132.1	127.4	**–4**

5. Major Group Level Comparisons

In **Table 2** we break out the analysis of our two sets of variance components by Major Group. These Mean Total Variances have been summed (within Major Group) and averaged (over all Areas) in a similar fashion as was done with the Area breakouts shown in **Table 1**. But now we are uncovering significant differences in the variance components, both across the two sets of components and within each set. Within the first set, the range of variances are greatly expanded, from a low of 0.013 in Gasoline to a high of 0.069 in Fruits & Vegetables. Similarly in the second set (**'98-'00**), the range runs from Gasoline's 0.0147 to Fruits and Vegetables' 0.0654. In the **'94-'96** variances, the Food groups and Apparel are significantly larger, by a factor of two to three, than any of the other Major Groups. But then these differences tighten up when we move across the 1998 Revision to the second set of variance component results. The Food groups and Apparel do not appreciably change from the first time frame to the second. The rest of the Major Groups' variances, however, do increase appreciably. That 35% increase of variance noted above seems to have come from all the Major Groups *other* than Food and Apparel (with the exception of Gasoline and Food Away from Home, both of which stayed pretty much the same across the Revision). Variances in Household Furnishings doubled; in Utilities they tripled; in the Education & Communications group they nearly tripled; the variances in Medical and in Entertainment & Miscellaneous nearly doubled.

One might hope to see these variance shifts as a function of sample size reduction after the 1998 Revision, but a closer analysis of the corresponding mean sample sizes (by Major Group) does not warrant such a conclusion (see **Mean Sample Size** columns in **Table 2**). The starkest drop in sample size occurred in the Food Away from Home group, yet its corresponding mean total variances shifted up only from 0.0160 to 0.0188 (a 17% increase in **'98-'00** compared with a 170% greater sample size in **'94-'96**). Similar anomalies occur. The Utilities group's sample size stayed the same yet its variance (albeit one of the smallest variances to begin with) nearly tripled; the Education/Communications group's variance nearly tripled yet its sample size actually increased by 10%; Fruits & Vegetables' variance actually decreased a little, even while its sample size fell off by 23%. The rest of the groups, however, tended to fall into line, with an increase in variance corresponding to a somewhat similar decrease in sample size. The hard conclusion must be reached, however, that in several of these Major Group categories variability has increased with the introduction of the new 1998 Revision sample. In Household Furnishings, in Utilities, in Education & Communications, and in Entertainment & Miscellaneous, the variances have increased substantially without any concomitant decrease in sample size to explain this behavior.

The components of variance at the Area–Major Group level are the constituent variances that help determine how the Outlets and Items are (optimally) allocated in the C&S Sample Design. In **Table 3** below, we compare the results of the optimization procedure when, first, the old set of variance components are used and, second, when the new set is used. (The **LO** and **UP** columns give the pre-determined lower and upper constraining bounds.) Largely due to the upper bound constraints, all of the Food categories are the same using either set of variance components. The only notable differences occur in Household Furnishings, in Transportation, in Medical, in Education/Communications, and in Entertainment/Miscellaneous. In each of these categories there is a substantial *increase* in the number of Item hits to be allocated. Combining the results from **Table 2** and **Table 4** in these categories, it should be clear why these particular Item hits would have been increased. These four categories show marked percentage increases from the first time period to the second. Then, in **Table 4**, there is clear evidence that the general *increase* in the variances across the two time frames is due to a doubling of the Item-Stratum variance component. The PSU component is really too small (and only occurs in the B- and C-Size cities anyway) to worry about its relative increase, and effectively both the Outlet and the Error (Residual) components have remained about the same. And so, in the latest Sample Design, the Item hits have been increased and, concomitantly, the Outlet hits have decreased. (The longer list of Outlet allocation changes have not been presented here, due to the length of the series.) More variability has crept into the CPI due to the new Item structure and the new sample that was introduced with the 1998 Revision. The optimization procedures only reflect this new reality.

Table 3. C&S Sample Design Comparisons

MAJOR GROUP	Opt. Item Hits 94–96	Opt. Item Hits 98–00	LO	UP

Food–Staples	28	28	12	28
Food–Meats	28	28	13	28
Food–Fr/Vegs	27.5	27.5	11	28
Food–Other	42	42	20	42
Food Away	16.6	16.8	6	16.8
Household	70.0	79.6	27	134
Utilities/Fuels	23.2	25.2	9	25.2
Apparel	39.5	40.6	16	101
Transportn	39.9	54.3	16	67.2
Gasoline	6	6	2	6
Medical	22.6	33.1	9	33.6
Educ/Comms	20.0	32.7	13	44.8
Entr/Misc	76.6	87.5	29	134

Table 4. All-US–All-Items Averaged Variance Components

Variance Components	meanVC '94–'96	meanVC '98–'00	PCT INCR
PSU	0.00099	0.00218	**+121**
ITEM-STR	0.00559	0.01105	**+98**
OUTLET	0.00873	0.01022	**+17**
Error	0.01145	0.01316	**+15**
All-US VARS			
VC Total	0.02595	0.03483	**+34**
Optimal Var	0.00684	0.01013	**+48**
CPI Variance	0.01348	0.01472	**+10**

6. All US–All Items Variance Results

Finally, we would like to compare some All-US–All-Items variance results. The first set of All-US–All-Items results we already have. From **Table 1,** in the last row, we can find two All-US–All-Items variance results to compare across our two time periods: 0.02595 for **'94-'96** and 0.03483 for **'98-'00**. Taking the square root of these two numbers we get two standard error results: 0.161 and 0.187, respectively. The C&S Sample Design's optimization procedure also produces an All-US–All-Items variance result. The final objective function value is a measure of an optimal variance result (should the optimal allocations be implemented precisely). If the **'94-'96** variance components are used, the final optimal variance is 0.0068 (standard error 0.083); if the **'98-'00** variance components are used, the final optimal variance is 0.0101343 (standard error 0.101). Finally, we thought it would be useful to compare these summary variance results that are based on components of variances with some actual CPI variances. To that end, we used the jackknife procedure to calculate 6-month variances (and standard errors) at the All-US–All-Items level[7] across two sets of 24 months, the first set from 1995 and 1996 and the second set from 1999 and 2000. Averaging the variances across the first ('95-'96) set gives a variance of 0.01348 (standard error 0.116); the second ('98-'00) set gives a 6-month variance value of 0.1472 (standard error 0.121). Thus, we have three sets of comparative variances, each of which reveals that an *increase* in overall variance occurred with the introduction and implementation of the 1998 Revision.

7. Acknowledgments

The author would like to thank Janet Williams for her careful reading of this paper and her helpful comments. I would also like to thank my BLS colleagues, Bill Johnson and Sylvia Leaver.

8. References

Baskin, R.M. (1995), "Estimation of Variance Components for the U.S. Consumer Price Index", *Proceedings of the Survey Research Methods Section*, American Statistical Association (Vol. 1), 126-131.

Bureau of Labor Statistics, *CPI Detailed Report (Feb 1999)*, Washington, DC, U.S. Government Printing Office.

Pfeffermann, D., Skinner, C.J., Goldstein, H., Holmes, D.J., Rasbash, J. (1998), "Weighting for Unequal Selection Probabilities in Multilevel Models (with discussion)", Journal of the Royal Statistical Society, Series 60, pp. 23-40.

[7] The variance components are from C&S data exclusively. The data used to compute the jackknife variance results were similarly drawn from data which excluded Rent and Rental Equivalency data.