Smoothing Variance Estimates for Price Indexes Over Time

Richard Valliant
U.S. Bureau of Labor Statistics
Room 2126, 441 G St. NW
Washington DC 20212

April 1992

ABSTRACT

This paper explores the possibility of developing generalized variances for price indexes by applying nonparametric scatterplot smoothers to time series of point variance estimates. The goal here is to formulate smoothed variances which are approximately unbiased, which provide acceptable confidence interval coverage, and which are more stable than the point variance estimates. Smoothing methods are applied to time series of point variance estimates in a simulation study using data from the U.S. Consumer Price Index program.

Key words: Generalized variance function, Laspeyres price index, linearization variance estimator, loess, super smoother.

ACKNOWLEDGEMENT

Any opinions expressed are those of the author and do not constitute policy of the Bureau of Labor Statistics. I thank the editors and referees for their useful comments.

1. INTRODUCTION

Price index series are some of the most important statistics published by national governments. Having a measurement of inflation, for example, is a fundamental requirement for tracking the health of an economy. When index series are estimated from sample surveys, the indexes are subject to sampling error, and an important statistical question is how to best estimate their variances. The importance of calculating variances of indexes was illustrated by an experience in Sweden related by Andersson, Forsman, and Wretman (1987). In January 1987 major Swedish labor agreements were invalidated when the consumer price index increased 3.26% during 1986. If the increase had been .02 percentage points less, the threshold specified in the labor agreements would not have been exceeded and the agreements would have remained in effect. Although the increase may have been within sampling error of the threshold, at the time, procedures were not in place for variance estimation so that no testing of that possibility could be done.

Index series are characterized by seasonal and irregular fluctuations in addition to underlying trends. The literature is replete with methods for decomposing and smoothing such time series. Point estimators of variance, obtained by linearization, replication, or another method, may be subject to the same types of seasonal and irregular variations as the index series themselves. The variable nature of point variance estimates was illustrated by Leaver (1990) for indexes. This paper explores the possibility of developing generalized variances for price indexes by applying nonparametric scatterplot smoothers to series of point variance estimates. The goal here is to formulate smoothed variances which are approximately unbiased, which provide acceptable confidence interval coverage, and which, most importantly, are more stable than the point variance estimates.

The approach taken here is somewhat different than that which is sometimes used in household surveys for estimating generalized variance functions (*GVF*'s). That method is described in Wolter (1985) with some justifying theory given in Valliant (1987). The general idea is to use models to approximate variances. Given a set of survey variables whose variances all follow the same model, parameters of the model are estimated by least squares. The parameter estimates are then provided to users rather than individual variance estimates in order to condense survey publications. Ideally, the models will also lead to more stable estimates of variance. Applications of *GVF*'s in two particular surveys can be found in Hanson (1978) and Johnson and King (1987). In the case of price indexes, finding multiple indexes whose variances follow the same model may be difficult. However, smoothing the variances of a particular index series over time is a practical alternative. For a given index series this is a two-step process consisting of estimating variances at a number of points in time and of smoothing the series of point variance estimates. As will be illustrated, this approach can produce more stable variance estimates that are approximately unbiased and that provide near nominal confidence interval coverage.

Section 2 defines the population Laspeyres price index, a class of index estimators, and a superpopulation model which is used to study the variance of the index estimators. In section 3, an approximation to the variance of a long-term price change estimator is discussed. The fourth section presents the methods that were tested for estimating generalized variances. A simulation study, described in section 5, was conducted using data from the U.S. Consumer Price Index to determine how well the proposed variance estimators would work in practice. Finally, section 6 gives conclusions.

2. INDEX ESTIMATORS AND A SUPERPOPULATION MODEL

The population is divided into H strata with stratum h containing N_h establishments. Establishment (hi) contains M_{hi} items, and the total number of items in all establishments in stratum h is $M_h = \sum_{i=1}^{N_h} M_{hi}$. At time t the price of item j in establishment (hi) is p_{hij}^{t}, and the price relative between time t and the base period time 0 is $r_{hij}^{t,0} = p_{hij}^{t} \big/ p_{hij}^{0}$. The quantity of item (hij) purchased in the base period is q_{hij}^{0}. The finite population value of the long-term fixed base Laspeyres price index for comparing period t to period 0 is

$$
\begin{aligned}
I^{t,0} &= \sum_{h=1}^{H}\sum_{i=1}^{N_h}\sum_{j=1}^{M_{hi}} p_{hij}^{t} q_{hij}^{0} \Bigg/ \sum_{h=1}^{H}\sum_{i=1}^{N_h}\sum_{j=1}^{M_{hi}} p_{hij}^{0} q_{hij}^{0} \\
&= \sum\nolimits_{h}\sum\nolimits_{i}\sum\nolimits_{j} W_{hij}^{0} r_{hij}^{t,0} \, ,
\end{aligned} \tag{1}
$$

where $W_{hij}^{0} = p_{hij}^{0} q_{hij}^{0} \Big/ \sum_{h,i,j} p_{hij}^{0} q_{hij}^{0}$ is the fraction of total base period cost or value accounted for by item (hij). For later reference, it is also convenient to define the stratum index $I_h^{t,0} = \sum_{i=1}^{N_h}\sum_{j=1}^{M_{hi}} W_{hij}^{0} r_{hij}^{t,0} \Big/ W_h^{0}$ where $W_h^{0} = \sum_{i=1}^{N_h}\sum_{j=1}^{M_{hi}} W_{hij}^{0}$. Using long-term indexes, the population short-term index for comparing periods t_2 and t_1 $(t_1 < t_2)$ is defined as $I^{t_2,t_1} = I^{t_2,0} \big/ I^{t_1,0}$. Monthly, quarterly, semiannual, and annual changes are commonly published by index programs.

In order to analyze the properties of index estimators, we will consider the superpopulation model defined below, which was also used in Valliant (1991).

$$
\begin{aligned}
r_{hij}^{t,0} &= \alpha_{th} + \omega_{thi} + \varepsilon_{thij} \\
\varepsilon_{thij} &= \rho_h \varepsilon_{t-1,hij} + \xi_{thij}
\end{aligned} \tag{2}
$$

where $E(\omega_{thi}) = E\left(\omega_{t_1hi}\omega_{t_2h'i'}\right) = 0$ for all t, h, i, and $(t_1hi) \neq (t_2h'i')$; $E\left(\omega_{thi}^2\right) = \sigma_{\omega h}^2$; $E\left(\xi_{thij}\right) = E\left(\xi_{t_1hij}\xi_{t_2h'i'j'}\right) = 0$ for all t, h, i, j and $(t_1hij) \neq (t_2h'i'j')$; $E\left(\xi_{thij}^2\right) = \sigma_{\xi h}^2$; and $-1 < \rho_h < 1$. By convention, define $\alpha_{0h} \equiv 1$ and $\varepsilon_{0hij} \equiv 0$. Considering times only back to the base period and not beyond, (2) implies that $\varepsilon_{thij} = \sum_{k=0}^{t-1} \rho_h^k \xi_{t-k,hij}$. Using this expression and the properties of ξ_{thij}, the covariance structure implied by model (2) is

$$
\operatorname{cov}\left(r_{hij}^{t_2,0}, r_{h'i'j'}^{t_1,0}\right) = \begin{cases}
\sigma_{\omega h}^2 + \left(1-\rho_h^{2t_2}\right)\Delta_h^2 & t_2 = t_1, h = h', i = i', j = j' \\
\rho_h^{t_2-t_1}\left(1-\rho_h^{2t_1}\right)\Delta_h^2 & t_1 < t_2, h = h', i = i', j = j' \\
\sigma_{\omega h}^2 & t_2 = t_1, h = h', i = i', j \neq j' \\
0 & \text{otherwise}
\end{cases} \tag{3}
$$

where $\Delta_h^2 = \sigma_{xh}^2 \big/ \left(1-\rho_h^2\right)$. Expression (3) implies that price relatives for a particular item are correlated over time. At a given time period, items within a particular establishment are also correlated, while other items are not.

The sample design addressed here is a rotating panel survey in which establishments are sampled as the first-stage units. Establishments are retained in the sample for a specified period of time and then rotated out and replaced by new units. At each time t (t=1,...,T), we have a sample s_{th} of n_h establishments from the N_h establishments in stratum h and a sample s_{thi} of m_{hi} items from the M_{hi} items in sample establishment (thi). A two-stage sampling plan, often approximated in practice, is one in which establishments are selected with probabilities proportional to $W_{hi}^0 = \sum_{j=1}^{M_{hi}} W_{hij}^0$. Items within establishments are then selected with probabilities proportional to W_{hij}^0. Surrogate measures of size, felt to be closely related to W_{hi}^0 and W_{hij}^0, such as current sales values or employment are often used in practice. At each time period, the total establishment sample size is assumed to be constant at $n = \sum_h n_h$ with the total number of sample items in stratum h being $m_h = \sum_{i \in s_{th}} m_{hi}$. At each time period a proportion δ_h of the sample establishments is rotated out in stratum h and an equal number rotated in. The size of the overlap, $s_{tuh} = s_{th} \cap s_{uh}$, between samples from time t and u ($t \neq u$) is $\max\left\{0, n_h\left[1 - \left|t-u\right|\delta_h\right]\right\}$.

The class of estimators considered here was introduced in Valliant and Miller (1989) for one-stage sampling and generalized in Valliant (1991). For the long-term index, define

$$\hat{I}^{t,0} = \sum_h \bar{z}_{th}^t \prod_{u=1}^{t-1} \left(\frac{\bar{z}_{uh}^u}{\bar{z}_{u+1,h}^u} \right)^{\gamma_h^{tu}} \tag{4}$$

where $\bar{z}_{kh}^u = \sum_{i \in s_{kh}} \lambda_{hi} \bar{r}_{khi}^{u,0}$, $\bar{r}_{khi}^{u,0} = \sum_{j \in s_{khi}} r_{hij}^{u,0} \big/ m_{hi}$ for k=u or u+1 (u=1,...,t-1), and γ_h^{tu} is a real number. The term λ_{hi} is a coefficient which does not depend on the model random variables $r_{hij}^{t,0}$. For the two-stage probability-proportional-to-size design mentioned above, for example, $\lambda_{hi} = W_h^0 \big/ n_h$. We restrict consideration to cases where

$$\sum_{i \in s_{uh}} \lambda_{hi} = W_h^0$$

in which case $E\left(\bar{z}_{kh}^{u} - W_h^0 I_h^{u,0}\right) = 0$. When all within-stratum samples of establishments are large, $\hat{I}^{t,0}$ is approximately model-unbiased under (2). Short-term estimators are defined by taking ratios of long-term estimators. The price change from time t_1 to t_2 $(t_1 < t_2)$ is estimated by $\hat{I}^{t_2,t_1} = \hat{I}^{t_2,0} / \hat{I}^{t_1,0}$.

A number of estimators in class (4) are listed in Valliant (1991). Three are of particular interest. If $\gamma_h^{tu} \equiv 1$, then (4) is the product estimator, which can be written as

$$\hat{I}_1^{t,0} = \sum_h \prod_{u=1}^{t} \left(\frac{\bar{z}_{uh}^{u}}{\bar{z}_{uh}^{u-1}}\right)$$

with $\bar{z}_{1h}^{0} \equiv 1$. If $\gamma_h^{tu} \equiv 0$, (4) reduces to the simple index estimator

$$\hat{I}_2^{t,0} = \sum_h \bar{z}_{th}^{t} .$$

A third choice of γ_h^{tu} is the one which minimizes the approximate variance of $\hat{I}^{t,0}$ under model (2). The optimum is complicated in general, but in the special case of a constant number of sample items per establishment, $m_{hi} = \bar{m}_h$, and λ_{hi} a constant for all sample establishments in stratum h, then the optimum reduces to

$$\gamma_h^{*tu} = \frac{1}{2}\frac{\alpha_{uh}}{\alpha_{th}}\rho_h^{t-u}\left[1 + \bar{m}_h \frac{g_{uh}}{1 - g_{uh}}\right]^{-1}$$

for $1 \le u \le t-1$ and $g_{uh} = \sigma_{wh}^2 \Big/ \left[\sigma_{wh}^2 + \left(1 - \rho_h^{2u}\right)\delta_h^2\right]$. The long-term estimator which uses the optimal γ_h^{*tu} will be denoted as $\hat{I}_3^{t,0}$.

3. APPROXIMATE VARIANCES UNDER THE MODEL

When the establishment sample size n_h is large in each stratum, the long-term index estimator can be approximated, as shown in Appendix A of Valliant (1991), by

$$\hat{I}^{t,0} \cong \sum_h \left\{\bar{z}_{th}^{t} + \sum_{u=1}^{t-1} \gamma_h^{tu} \frac{\alpha_{th}}{\alpha_{uh}} \left(\bar{z}_{uh}^{u} - \bar{z}_{u+1,h}^{u}\right)\right\} \qquad (5)$$

Using results in the appendix of that paper, we can write the approximate variance of the long-term estimator as

$$\mathrm{var}\left(\hat{I}^{t,0}\right) \cong \sum_h \left\{\sum_{u=1}^{t-1} a_{tuh}\left(\gamma_h^{t,u}\right)^2 + 2\sum_{u=1}^{t-1} b_{tuh}\gamma_h^{t,u} + c_{th}\right\} \qquad (6)$$

where

$$\iota_h^{t,u} = \alpha_{th}/\alpha_{uh} = E\left(T_h^{t,0}\right)/E\left(T_h^{u,0}\right),$$

$$a_{tuh} = \left(\gamma_h^{tu}\right)^2\left[\sum_{i\in C_{uh}}\frac{\lambda_{hi}^2}{m_{hi}}v_{uhi} + \sum_{i\in D_{uh}}\frac{\lambda_{hi}^2}{m_{hi}}v_{uhi}\right]$$

$$b_{tuh} = \gamma_h^{tu}\rho_h^{t-u}\left(1-\rho_h^{2u}\right)\Delta_h^2\left[\sum_{i\in s_{tuh}}\frac{\lambda_{hi}^2}{m_{hi}} - \sum_{i\in s_{t,u+1,h}}\frac{\lambda_{hi}^2}{m_{hi}}\right],$$

$$c_{th} = \sum_{i\in s_{th}}\frac{\lambda_{hi}^2}{m_{hi}}v_{thi},$$

where $v_{uhi} = v_{uh}\left[1+\left(m_{hi}-1\right)\rho_{uh}\right]$, $v_{uh} = \sigma_{wh}^2 + \left(1-\rho_h^{2u}\right)\Delta_h^2$, $C_{uh} = s_{uh} - s_{u+1,h}$, i.e. the part of s_{uh} that is not contained in $s_{u+1,h}$, and $D_{uh} = s_{u+1,h} - s_{uh}$.

An expression similar to (6) can also be worked out for the approximate variance of the short-term index estimator $\hat{I}^{t,u}$.

4. GENERALIZED VARIANCE FUNCTIONS FOR INDEXES

Judging from (6), the approximate variance is a second order polynomial in the stratum superpopulation short-term indexes $\iota_h^{t,u}$. This is analogous to the relationship between an estimator $\hat{T}$ of the population total T in two-stage sampling and its approximate variance derived in Valliant (1987) for a particular class of models in which the variance of a unit was a quadratic function of the unit's mean:

$$\operatorname{var}\left(\hat{T}\right) \cong aE(T)^2 + bE(T) \tag{7}$$

The terms a and b are coefficients that depend on various quantities, such as intracluster correlations, numbers of population and sample units within clusters, and coefficients in the estimator $\hat{T}$. In fitting the *GVF* model defined by (7), the usual procedure is to select a group of variables which all have the same a and b coefficients, calculate point estimators of variance for each of the variables, and then to estimate a and b by some form of least squares. Application of this course to (6) would be fraught with practical difficulties. In (7) there are only two regression coefficients to be estimated — a and b. In (6) there are $2(t\text{-}1) + 1$. As t increases so does the number of coefficients. The components of the coefficients, a_{tuh} , b_{tuh} , and c_{th}, are also complex, so that identifying different indexes that all follow model (6) would be a problem.

An alternative approach is to work with a particular index series and attempt to model the behavior of its variance over time. If $i_h^{t,u}$ is a smooth function of time, e.g. a polynomial in *t-u*, then the variance (6) will also be a smooth function of time, say, $f(t)$. If an unbiased, or approximately unbiased, variance estimator is used for $\hat{I}^{t,0}$, then its expectation can also be described by $f(t)$. As data are accumulated over time, a time series of point variance estimates is developed and the function $f(t)$ can be fitted by a scatterplot smoother without having to know the explicit form of the function. A number of such smoothers are available, and we will consider two that have proved to be useful in other situations.

The two smoothers used here are the super smoother (Friedman 1984) and loess (Cleveland 1979, Cleveland, Cleveland, McRae, and Terpenning 1990). The two algorithms are fairly complex to describe in detail, so that only rough sketches will be given here. Both methods use local linear fits in neighborhoods around each point *t*. A critical parameter in both algorithms is the span, the size of the neighborhood around *t*, which is used to estimate $f(t)$. In loess the span is fixed while for the super smoother, spans can be variable. Of the two, loess explicitly incorporates features to reduce the effects of outlying values and tends to produce a smoother looking curve of estimates. The variable span used by super smoother allows it to adapt more readily to changing curvature in $f(t)$. Super smoother also has the advantage of being computationally faster than loess.

5. AN EMPIRICAL STUDY

A simulation study, using a population derived from data collected for the U.S. Consumer Price Index program by the BLS, was undertaken to test the usefulness of the proposed method of calculating *GVF*'s. The population was composed of establishments and items and was described in detail in Valliant (1991). Its main features are briefly recounted here. The population was divided into the five strata, which are listed in Table 1 along with various population and sample allocation numbers. The six hundred and fifty-nine establishments contained an average of just under 10 items each. Each item had prices for 42 consecutive months.

Two sets of 500 stratified two-stage samples were selected with the number of sample establishments allocated to each stratum being roughly proportional to W_h^0. The total establishment sample sizes in the two sets of samples were n = 50 and 100. Samples were selected in such a way that 20% of the sample establishments were rotated in each 12-month period. This was done by first selecting a large systematic, random-start sample of establishments in each stratum with probabilities proportional to W_{hi}^0. For samples of size n = 50, the initial, large sample size was 84 and was 168 for the samples of size n = 100. These initial samples were large enough to accommodate all 42 months accounting for the amount of establishment rotation. The initial sample from each stratum was then sorted in a random order. For a particular time period t, the stratum establishment sample consisted of establishments $1+(t-1)\delta_h n_h, \ldots, n_h+(t-1)\delta_h n_h$, where δ_h was the proportion of establishments rotated in a month. For both the cases of n = 50 and n = 100, $\delta_h = 1/60$ which resulted in an annual turnover of $12\left(n_h/60\right) = n_h/5$ establishments or 20%. From each sample establishment, $\bar{m}_h = 2$ sample items were selected systematically with probability proportional to W_{hij}^0.

From each sample, the long-term estimators $\hat{P}_j^{t,0}$ (j=1,2,3; t=1,...,42), and the short-term estimators of 1-month and 12-month change $\hat{P}_j^{t_2,t_1}$ (j=1,2,3; t_1=t_2-1, t_2-12 for $t_1 \neq 1$) were computed. The special case of $\lambda_{hi} = W_h^0 / n_h$ was used, which produces a design-unbiased estimator under the simulation study sampling plan. The parameters needed for the optimal estimator $\hat{P}_3^{t,0}$ were approximated as described in Valliant (1991). Empirical results for the simple estimator were similar to those for the optimal so that only the latter is discussed subsequently.

Point variance estimates were obtained by the linearization method and were described in detail in Valliant (1991). It should be emphasized that the results here do not depend on the use of any particular method of point variance estimation. Estimates obtained by balanced repeated replication, the jackknife, or another approach would work just as well as long as consistent or approximately unbiased variance estimates were used. For each sample the linearization variance estimate was computed for each of the long-term and short-term index estimates and time periods named above. Two *GVF*'s — super smoother and loess — were then computed for each index series. For example, for the product long-term index estimate, a series of 42 point variance estimates was produced for each sample. The super smoother and loess estimates were calculated in each sample by applying those methods to the series of 42 linearization estimates for each index estimate. The simulation calculations were performed in double precision using Borland's *Turbo Pascal*. *GVF*'s were calculated with the software package *S-PLUS for DOS* by Statistical Sciences Inc.

Summary statistics were then calculated across all 500 samples. The square roots of the empirical mean squared errors were computed as $\left[\sum \left(\hat{I}-I\right)^2/500\right]^{1/2}$ with the summation being over the 500 samples, $\hat{I}$ being one of the long- or short-term estimators, and I being the population index defined in section 2. Square roots of the average of the variance estimates were computed at each time period as $\sqrt{\bar{v}}$ where $\bar{v} = \sum_{s=1}^{500} v_s/500$ and v_s is one of the three types of variance estimate (linearization, super smoother, or loess) at a particular time period from sample *s*. This was done separately for the product and optimal estimates for long-term, 1-month, and 12-month price change.

Summary results across all samples and time periods are listed in Tables 2 and 3. The ratios (in percent) of the square root of the average variance estimate to the root of the empirical mean squared error (RMSE) are generally somewhat less than 100 in all cases, i.e. both the point variance estimate ($\hat{v}$) and the *GVF*'s are underestimates, but the problem is minor. The exception is the long-term optimal estimator for *n*=50, where the variance estimates are slight overestimates. In all cases, for both the product and optimal estimators, the *GVF*'s are more stable than $\hat{v}$. For example, in Table 2, the standard deviation of the super smoother *GVF* is 61% of that of $\hat{v}$ for 1-month change when *n*=100. For the same case the loess GVF has a standard deviation which is 57% of that of $\hat{v}$. The biggest gains in stability are for 1-month price change while the smallest gains occur for long-term change. The loess estimates are generally more precise than the super smoother estimates with the improvement compared to the linearization estimate being somewhat less for the larger sample size. Tables 2 and 3 also list empirical coverage of 95% confidence intervals across the 42 time periods. Normal approximation confidence intervals were computed in the usual way as $\hat{P} - 1.96\sqrt{v}$ where $\hat{P}$ is one of the long- or short-term indexes and *v* is one of the variance estimates. Although all variance estimates provide slightly less than the nominal 95% coverage, the smallest percentage in Tables 2 and 3 is 92.0%, and the *GVF*'s are quite competitive with $\hat{v}$.

Figures 1-4 are plots of summary statistics over the 500 samples by time period for *n*=100. Figures are given only for the long-term and 1-month product and optimal estimators. Plots for the 12-month price change estimators are qualitatively similar. The upper left panel in each figure plots empirical RMSE's and the square root of the average of each *GVF* versus time. The *GVF*'s are much smoother than $\hat{v}$, as might be expected. Although both smoothers are not inordinately influenced by outliers among the $\hat{v}$'s, the super smoother does follow the fluctuations of the $\hat{v}$ curves more closely than does loess. This leads to the super smoother's generally having a larger standard deviation than loess, as shown in the upper right-hand panel of each figure. The lower left-hand panel shows the ratio of the *GVF* standard deviation over the 500 samples to the standard deviation of $\hat{v}$. This again simply illustrates that the two *GVF*'s are more precise than the linearization estimate with gains being especially large for 1-month change. The lower right-hand panel of each figure charts the coverage of 95% confidence intervals over time. The *GVF*'s give reasonably good coverage which is almost equal to that of the point variance estimates. The time periods where the *GVF*'s provide noticeably poorer coverage than $\hat{v}$ are ones where the smoothers do not closely follow upward fluctuations in $\hat{v}$.

A further possibility, which we have not pursued, would be to calculate a weighted average of a smoothed variance and the point variance estimate at each time point. This could be advantageous when the point variance estimators are felt to be more nearly unbiased than the smoothed estimates because of failure of the approximate variance (6) to be a smooth function of time.

6. CONCLUSION

In continuing surveys which produce time series of estimates, the methods studied here for smoothing variance estimates appear to be quite useful. For continuing surveys in which the sample design and sample size are the same for long periods of time, users expect variances to be smooth over time, a feature which point variance estimates generally do not have. Such expectations by users may seem, at first, to be statistically unreasonable since actual mean square errors may vary over time. However, for price indexes we have shown, using large-sample theory and simulations, that smoothed, approximately unbiased variance estimates can be obtained which are more stable than point variance estimates for both long-term and short-term price change, and which also provide near nominal confidence interval coverage. Thus, in the situation studied here, smoothed variances have a statistical justification, in addition to having considerable cosmetic appeal to data users.

The methods explored here also should apply to other types of panel surveys, like labor force surveys, which publish time series of employment estimates. The linearized form of the index estimator given by (5) in section 3 is similar to the composite estimators described by Cantwell (1990) for household surveys. Since, in large samples, the index estimator and the composite estimators have similar structure, the possibility of smoothing variances of the latter over time appears to be worth pursuing.

The nonparametric approach may also have use in sample design problems for continuing surveys. If components of variance are estimated at a number of points in time, one alternative is to average the components over time. However, when components of variance are subject to seasonal factors, averaging of variance components across time periods may obscure the seasonality. Alternatively, the smoothers can be used to obtain more stable estimates of components which could subsequently be used for determining sample allocations. Smoothing can be done in such a way that seasonal differences in components are preserved in order to study their effects on computed allocations.

One useful feature of the *GVF*'s, described in Wolter (1985), for household surveys is the ability to publish a limited number of model parameters from which users calculate their own standard errors. This feature is lost for the smoothed variances because an explicit functional form is not estimated. The main use of the variance smoothers, thus, may be to produce more stable estimates.

References

Andersson, C., Forsman, G., and Wretman, J. (1987). On the Measurement of Errors in the Swedish Consumer Price Index. Bulletin of the International Statistical Institute, **52**, Book 3, 155-171.

Cantwell, P. (1990). Variance Formulae for Composite Estimators in Rotation Designs. Survey Methodology, **16**, 153-163.

Cleveland, W.S. (1979). Robust Locally Weighted Regression and Smoothing Scatterplots. Journal of the American Statistical Association, **74**, 829-836.

Cleveland, R.B., Cleveland, W.S., McRae, J.E., and Terpenning, I. (1990). STL: A Seasonal Decomposition Procedure Based on Loess. Journal of Official Statistics, **6**, 3-32.

Friedman, J.H. (1984). A Variable Span Smoother. Technical Report No. 5, Laboratory for Computational Statistics, Stanford University.

Hanson, R.H. (1978). The Current Population Survey: Design and Methodology, Technical Paper 40, Washington DC, U.S. Bureau of the Census.

Johnson, E.G. and King, B.F. (1987). Generalized Variance Functions for a Complex Sample Survey. Journal of Official Statistics, **3**, 235-250.

Leaver, S.G. (1990). Estimating Variances for the U.S. Consumer Price Index for 1978-1986. Proceedings of the Section on Survey Research Methods, American Statistical Association, 290-295.

Valliant, R. (1987). Generalized Variance Functions in Stratified Two-stage Sampling. Journal of the American Statistical Association, **82**, 499-508.

———— (1991). Variance Estimation for Price Indexes from a Two-stage Sample with Rotating Panels. Journal of Business and Economic Statistics, **9**, 409-422.

Valliant, R. and Miller, S.M. (1989). A Class of Multiplicative Models for Laspeyres Price Indexes. Journal of Business and Economic Statistics, **7**, 387-394.

Wolter, K.M. (1985). Introduction to Variance Estimation, Springer-Verlag, New York.

FIGURE TITLES

Figure 1. Summary plots of simulation results for the long-term product estimator from 500 samples of size *n*=100 establishments. Legend for each of the panels is in the upper left-hand panel. v denotes the linearization estimator; supsmu denotes the super smoother.

Figure 2. Summary plots of simulation results for the 1-month product estimator from 500 samples of size *n*=100 establishments. Legend for each of the panels is in the upper

left-hand panel. v denotes the linearization estimator; supsmu denotes the super smoother.

Figure 3. Summary plots of simulation results for the long-term optimal estimator from 500 samples of size *n*=100 establishments. Legend for each of the panels is in the upper left-hand panel. v denotes the linearization estimator; supsmu denotes the super smoother.

Figure 4. Summary plots of simulation results for the 1-month optimal estimator from 500 samples of size *n*=100 establishments. Legend for each of the panels is in the upper left-hand panel. v denotes the linearization estimator; supsmu denotes the super smoother.

Table 1. Universe and sample characteristics for the study population

Stratum	W_h^0	N_h	M_h	n_h/n	f_h n=50	f_h n=100
1 Beef	.32	154	1800	.32	.10	.20
2 Eggs	.13	57	653	.12	.10	.21
3 Milk, other dairy	.33	155	1800	.32	.10	.20
4 Fresh vegetables	.10	193	1013	.12	.03	.06
5 Sugar	.12	100	1175	.12	.06	.12
Total	1.00	659	6441	1.00		

Table 2. Simulation results for the product estimator from 500 two-stage cluster samples averaged over 42 time periods. All figures are in percent. $\hat{v}$ denotes the linearization variance estimate.

	$\sqrt{\bar{v}}/$RMSE			Std. dev. (*GVF*) /Std. dev.($\hat{v}$)		95% CI coverage		
	$\hat{v}$	Supsmu	Loess	Supsmu	Loess	$\hat{v}$	Supsmu	Loess
<u>*n*=50</u>								
LT	99.5	99.0	98.0	90.8	86.8	93.7	93.1	92.8
1-month	99.3	99.4	96.8	56.8	50.8	93.3	93.6	92.9
12-month	95.0	94.8	93.9	74.3	71.0	92.0	92.3	92.1
<u>*n*=100</u>								
LT	99.8	99.2	98.5	96.1	92.1	94.2	93.4	93.3
1-month	99.3	99.8	98.1	61.0	57.0	93.8	93.6	93.3
12-month	96.1	95.8	96.0	79.1	79.6	93.2	93.4	93.3

Table 3. Simulation results for the optimal estimator from 500 two-stage cluster samples averaged over 42 time periods. All figures are in percent. $\hat{v}$ denotes the linearization variance estimate.

	$\sqrt{\bar{v}}/$RMSE			Std. dev. (*GVF*) /Std. dev.($\hat{v}$)		95% CI coverage		
	$\hat{v}$	Supsmu	Loess	Supsmu	Loess	$\hat{v}$	Supsmu	Loess
n=50								
LT	104.2	103.7	102.5	86.6	82.6	94.4	93.9	93.7
1-month	98.8	99.1	96.1	50.4	42.9	93.1	93.6	92.9
12-month	95.3	94.9	94.3	71.8	69.4	92.1	92.9	92.6
n=100								
LT	99.9	99.4	98.7	91.5	86.9	94.3	93.5	93.4
1-month	98.7	99.3	97.7	55.8	51.4	93.8	93.5	93.2
12-month	94.5	94.2	94.3	78.3	77.9	92.7	92.8	92.7