Choosing a Variance Computation Method for the Revised Consumer Price Index

Shawn Jacobson, Sylvia G Leaver, & David Swanson
Shawn Jacobson, Bureau of Labor Statistics, Rm. 3655, 2 Mass. Ave. NE, Washington, DC 20212
Jacobson_S@BLS.GOV

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

KEY WORDS: Jackknife, Balanced Repeated Replication, Random Group, Linearization

1. Introduction

A team of CPI staff was chartered to determine the best combination of estimation methodology, and software package for computing variances for the Consumer Price Index (CPI). This paper gives results of testing six combinations of methodology and package on 16 index series. Three of these combinations used linearization (SUDAAN and two versions of an in house package called VCS) and the other three used replication (VPLX and two versions of WesVarPC).

One criterion the team used to evaluate the variance software packages was stability of the variance estimates. This is because variance estimates are used as a publication criterion for CPI index series; they are also used to allocate sample among priced items and outlets. These uses put a premium on smooth estimates of standard error for index series. Stability is measured by the square root of the average squared difference between the estimated standard error in month t, SE_t, and the smoothed standard error in month t, SE_{st}, calculated as:

$$SE_{st} = median\left(SE_{t-2}, SE_{t-1}, SE_t, SE_{t+1}, SE_{t+2}\right) \quad (1)$$

Plots 1A & 1B show SE_t and SE_{st} for the most and least stable series of standard errors for an index series; the best standard error series is three times as stable as is the worst. The standard error series in Plot 1A can be smoothed with a minimum of effort; however, the standard error series in Plot 1B would be extremely hard to smooth.

In addition to the stability of variance estimates, the team addressed three practical concerns:

1. Can software packages produce variances for the volume of indexes produced by the CPI?
2. Can the results from a package be placed in variance tables with a minimum of processing?
3. How expensive is it to run the packages?

Section 2 lists the index series tested by the variance team. Section 3 gives the formulas used to compute variances for each candidate method. Section 4 gives the results of variance computations for 12 published index series. Section 5 discusses other non-numerical criteria. Conclusions are given in section 6.

2. Test index series

The team chose to study estimates of standard error for a variety of CPI index series. These series were of two types. The first type included combinations of four area levels (All US, all A-size PSUs, the Midwest Region, and Chicago) and three item levels (All Items, Transportation, and Motor Fuels). The resulting 12 index series represent index estimates diceminated to the public. The second type included 4 index series that represent a worst case scenario, a combination of a problem item and four relatively small A-size PSUs.

To simplify the choice of a variance computing methodology, the team decided to study only 12-month price relative (considered more important than other lengths of price relative). The time period under study was December 1986 through December 1995. Thus, the first price relative in a series would run from December 1986 through December 1987 and the last relative in the series would run from December 1994 through December 1995.

.

3. Description of candidate variance methods

Of the six studied variance calculation methods three used linearization and the other three methods used replication.

Replication methods use estimates of the function of interest based on subsets of the full sample. These subsamples, variance replicates, consist of the full sample or a stratum replicate from one or more strata. In replication techniques the variance estimate is a measure of the squared distance between the function of interest (a weighted ratio of price indexes) computed for a variance replicate and the same function computed for the full sample. WesVarPC and VPLX use replication methods.

Linearization methods use a first order Taylor series expansion about a central point to approximate the function of interest. The variance of the approximating linear function is then found. The VCS methods use a hybrid of replication and linearization where modified stratum replicates take the place of observations; this approach is briefly discussed in Wolter (1985). Since CPI index data is stored at the stratum replicate or full sample level, SUDAAN, as applied here, must also use a hybrid of replication and linearization.

The first VCS method studied, method 1, produces standard error estimates currently published in the *CPI Detailed Report*. For the 12-month price relative from months t-12 to t, the method 1 variance ($V'_{t,t-12}$) is:

$$V'_{t,t-12} = \frac{100^2}{2}\left\{\frac{\sum_{r=1}^{2}\left(Y_r - R_{ts}X_r\right)^2}{\left(\sum_{i\in A} C_{0is}\right)^2}\right\} \qquad (2)$$

Where:

$s = t - 12$,

$Y_r = \sum_{i\in A}\left(C_{rit} - C_{0it}\right)$,

$X_r = \sum_{i\in A}\left(C_{ris} - C_{0is}\right)$,

$R_{ts} = \sum_{i\in A} C_{0it} \Big/ \sum_{i\in A} C_{0is}$,

C_{rim} is the month m (either t or s) cost weight in stratum replicate r in stratum i,

C_{0im} is the month m full sample cost weight in stratum i, and

A is the set of strata included in the index series.

The "100" in the numerator scales variance estimates in units of percent of price change. The "cost weight" of an item in month t is an estimate of the amount of money spent on an item computed by multiplying the amount spent on that item in a base time period by the price change for that item from the base period to month t.

The equation for VCS method 2, $V''_{t,t-12}$, is the same as that for method 1, except that the full sample estimate, C_{0im}, is replaced by $C_{.im}$ where:

$$C_{.im} = \sum_{r=1}^{2}\frac{C_{rim}}{2},$$

and R_{ts} is replaced by R^*_{ts} where:

$$R^*_{ts} = \sum_{i\in A} C_{.it} \Big/ \sum_{i\in A} C_{.is} .$$

For VCS, strata were defined as areas for which indexes are published, index areas. For details, and a matrix version of these equations, see Leaver and Valliant (1995).

Like VCS, SUDAAN uses a linearization technique to compute variances. Sampling with replacement is assumed because without replacement estimates require a "TOTCNT" variable that is not relevant to the replicate observations used here. The SUDAAN, variance $S_{t,t-12}$, is;

$$S_{t,t-12} = \sum_{i=1}^{G} W_i \sum_{r=1}^{n_i}\frac{n_i}{n_i - 1}\left(\frac{d_{rit} - R_{its}d_{ris}}{n_i C_{.is}}\right)^2 \qquad (3)$$

Where:

$W_i = C_{.is} \Big/ \sum_{i=1}^{G} C_{.is}$ is the stratum i weight,

$d_{rim} = C_{rim} - C_{.im}$,

$R_{its} = C_{.it} / C_{.is}$ is the stratum i price relative, and

n_i is the number of stratum replicates in stratum i.

For SUDAAN, strata are defined as combinations of index areas and major groups; see The BLS Handbook of Methods (1996) for details about area and item classification systems. Apart from the way strata are defined, SUDAAN differs from VCS method 1 in three ways:

- VCS uses a combined estimate of the weighted index ratio; SUDAAN calculates a weighted index ratio for each stratum.
- VCS allows only two stratum replicates per stratum; SUDAAN does not restrict the number of observations per stratum; and
- VCS method 1 uses the full sample estimate as the point where the linear approximation is computed; SUDAAN approximates the function at the mean of observations.

Unlike VCS and SUDAAN, VPLX uses replication techniques to estimate variances. VPLX standard errors were computed using a stratified random group methodology. The VPLX variance $X_{t,t-12}$ is:

$$X_{t,t-12} = \sum_{i\in A}\frac{1}{G_i\left(G_i - 1\right)}\sum_{r=1}^{G_i}\left(E^*_{ri} - E\right)^2 \qquad (4)$$

Where:

A is the set of index areas included in the index series,

E is the full sample price relative estimate, for A,

E^*_{ri} is the price relative estimate for variance replicate r in stratum i, and

G_i is the number of replicates in stratum i.

The estimate of E for variance replicate r in stratum i is computed using the full sample in all strata except the ith stratum, and using the sample for stratum replicate r, appropriately weighted, in place of the stratum i full sample. For this analysis, strata are defined as combinations of A-size PSUs and major groups, or index areas outside of A-size PSUs.

Two variance calculation methods offered by WesVarPC were tested, the unstratified Jackknife method and the balanced repeated replication (BRR) method.

The unstratified Jackknife variance, $W_{t,t-12}^{JK}$ is:

$$W_{t,t-12}^{JK} = \frac{(G-1)}{G} \sum_{r=1}^{G} \left(E_r^* - E\right)^2 \qquad (5)$$

The BRR variance $W_{t,t-12}^{BRR}$ is:

$$W_{t,t-12}^{BRR} = \frac{1}{G} \sum_{r=1}^{G} \left(E_r^* - E\right)^2 \qquad (6)$$

Where:
E is the price relative estimate for the full sample,
E_r^* is the price relative estimate for replicate r, and
G is the number of replicates.

For the jackknife method of WesVarPC, variance replicate r is formed by removing the rth combination of index area and stratum replicate from the full sample; however, for BRR, the rth variance replicate is formed by using combinations of index area and stratum replicate in the rth half sample. For the BRR option of WesVarPC, index areas were used as strata.

The three replication methods are distinguished from each other in four ways.

- They assume differing stratum definitions.
- They create variance replicates differently.
- They used different central values; and
- They multiply squared differences about the

4. Findings from the analysis

Plot 2A shows the smoothed estimates of standard error from the three linearization methods for the All items All U.S. CPI. The standard error estimates from VCS method 1 tend to be high and are prone to spiking; by contrast, VCS method 2 produces low standard error estimates. This is because VCS version 1 centers standard error estimates about the full sample, while VCS version 2 centers them about the mean of the observations. Hence, method 1 is more conservative than method 2. Since method 1 is more conservative than method 2, this result is not surprising. Standard error estimates from SUDAAN are smaller than estimates from VCS method 1.

Plot 2B shows smoothed All Items All US standard errors for the three replication methods. The jackknife version of WesVarPC tends to give higher standard error estimates than does the BRR version; this is probably because the BRR method takes advantage of the CPIś sample design while the jackknife method does not. Standard error estimates from VPLX tend towards the middle of the plot.

Table 1 shows the mean of the stability measure for each publishable index series. For the All Items category, SUDAAN has the most stable estimates. For Transportation, and for Motor Fuels, VPLX has the most stable estimates. The one exception is the Transportation index for the Midwest where VCS version 2 has the most stable estimates.

Overall, there was little variation between the three replication methods; in fact, for standard error estimates for index series computed for Chicago, the two WesVarPC methods produced the same results. When comparing Linearization methods, SUDAAN won the stability criterion for every studied index except for Transportation in Chicago where VCS method 2 was slightly better.

The linearization methods had problems for the worst case scenario where the linearization assumptions break down. Linearization assumes that the function being approximated is smooth over the sphere centered at the point where the function of interest is approximated and containing all observations. For the worst case index series, this assumption does not always hold. When the assumption is violated, SUDAAN produces standard error estimates that are dangerously small while VCS produces extremely large standard error estimates. In one instance the VCS method 1 standard error exceeded the SUDAAN standard error by a factor of 500. This problem leads the authors to conclude that linearization may not suit the needs of our survey.

VPLX and SUDAAN were the two best packages from the standpoint of stability. Overall, VPLX does better than SUDAAN, though not by much, at producing stable standard error series; furthermore it requires fewer assumptions than does linearization. Thus, VPLX wins the stability criterion.

5. Non-numeric criteria

As mentioned earlier, the stability criteria had to be balanced against other criteria for choosing a variance package.

In the volume test, SUDAAN fared best; all desired variances could be run in 10 minutes. VPLX and WesVarPC were second best. Not all desired variances could be computed in one job using either package, but both packages allowed the combining of jobs into a single batch stream. WesVarPC took about 30 minutes to run all variances and VPLX required about 3 hours. VCS had to be run over night; thus, it was the worst package by this criterion.

Any required data processing for VCS is in place; thus VCS wins this criterion. WesVarPC is second best because it requires relatively little data processing before running variances and only requires minor formatting of the output. SUDAAN requires the most data processing before running variances because alphanumeric variables must be changed to ordered numbers; thus, a concordance is required between these numbers and index series identifiers. VPLX requires alphanumeric values to be changed to numbers, but it does not require that the numbers be in order. Unfortunately, the data set output by VPLX includes an estimate number instead of index series identifiers. To get output from VPLX into publications, requires a concordance between estimation numbers and index series identifiers for each VPLX job.

A related issue concerns required change if the basic index formulae for all or part of the CPI is changed from a Laspeyres formula to a Geometric means formula. Any linearization method would need to have its ratio approximating formula replaced by the approximating formula for the geometric mean. Also, if a Laspeyres Geometric means hybrid is used, the issue of how to compute variances for index series that aggregate sub-indexes computed in different ways must be settled.

WesVarPC, SUDAAN, and VPLX can be run on a personal computer; the only cost is the cost of the package and the cost of downloading data from the mainframe. VPLX is free software; WesVarPC is also free, but the free version is being phased out. VCS, however, runs on a mainframe at a considerable cost. Thus, VCS loses this criterion.

6. Conclusion

No package wins for all criteria; different packages have their respective strengths and weaknesses. Table 2 gives a summary of how the different packages fare when measured by the criteria set by the variance team.

VPLX gives stable variance estimates, is flexible, is inexpensive, and does sufficiently well in the volume test. The CPI program has experience with VPLX and feels that it works well. However, it also requires the most work to get variance estimates into a usable form.

WesVarPC gives acceptable variance estimates, requires relatively little data processing, is flexible, and is inexpensive. However, VPLX usually gives more stable estimates for indexes aggregated over multiple items (All areas and Transportation). Thus, VPLX is preferred over WesVarPC.

SUDAAN did best in the volume test, is inexpensive, and gives the most stable variance estimates when indexes are aggregated. However, it requires substantial data processing to get data to it from the index system and any change in index formula would require a change in how SUDAAN is run. Also, any required upgrades of SUDAAN, would need to be purchased at an unknown cost. Of greatest importance, however, is that SUDAAN can give invalid estimates of variance when there are extreme differences between replicate and full sample estimates; hence, SUDAAN is not appropriate for CPI variances estimation.

VCS is expensive, inflexible, and gives poor variance estimates subject to the validity and implementation problems faced by SUDAAN, hence, it is not recommended.

VPLX is recommended by the CPI team chartered to determine the best combination of variance estimation methodology and software package.

Acknowledgments

The authors would like to thank William Smyth and Claire Gallagher of BLS for their participation on the Variance Team. In addition, the authors would like to thank Janet Williams and Janice Lent of BLS for their review of this paper and for their helpful comments.

References

Bureau of Labor Statistics, *BLS Handbook of Methods* (1996), Washington. DC: U.S Government Printing Office, 185 and 217-225.

Fay, Robert, “VPLX: Variance Estimation for Complex Samples”, US Bureau of the Census, 1990.

Jacobson, Shawn, “Recommendation of a Variance Package for the CPI”

Leaver, S. G. and Valliant, R. L. (1995) "Chapter 28: Statistical Problems in Estimating the U.S. Consumer Price Index," *Business Survey Methods.* Wiley & Sons, Inc., 1995.

Shah, Babubhai, Barnwell, Beth, Hunt, P. Nileen, & LaVange, Lisa, *SUDAAN User’s Manual*, Release 6.0, Research Triangle Institute, 1992.

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

Table 1
Square Root of the Average Squared Difference Between Smoothed and Unsmoothed Standard Errors
Of 12 Month Price Relatives Produced by Candidate Variance Packages

		Stability Criterion					
		Linearization Methods			Replication Methods		
Item	Area	VCS,1	VCS,2	SUD.	VPLX	Wes,BRR	Wes,JK
All Items	All US	*.015*	*.015*	.009	.008	*.015*	**.008**
	A-size Cities	*.025*	*.026*	**.011**	.015	*.017*	.013
	Midwest	*.043*	*.042*	**.024**	.028	.027	.026
	Chicago	*.177*	*.162*	**.085**	.121	*.146*	*.146*
Transportation	All US	*.015*	*.016*	.010	**.007**	*.024*	.010
	A-size Cities	*.026*	*.031*	*.018*	**.010**	*.022*	*.016*
	Midwest	.045	*.051*	.032	**.031**	.053	.034
	Chicago	.183	**.169**	.172	.189	.173	.173
Motor Fuels	All US	*.039*	*.039*	.028	**.024**	*.049*	*.044*
	A-size Cities	*.045*	*.045*	.026	**.021**	.029	*.083*
	Midwest	*.149*	*.150*	.100	**.085**	*.189*	*.115*
	Chicago	.258	.259	.252	**.245**	.253	.253

- **Bold indicates most stable variance estimate.**
- *Italics indicates stability criterion > 1.5*l*

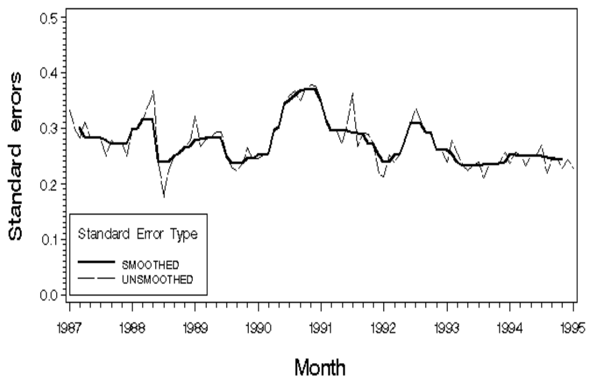
Table 2
Comparison of Variance Packages by Selection Criteria

Package	Stability	Volume	Minimize Processing	Cost
VCS	*Very unstable and assumptions can break down.*	*Must be run overnight.*	**Processing system is in place, but may require changes.**	*Very expensive, because it is a mainframe system.*
SUDAAN	Is most stable for some series, but assumptions can break down.	**Best, can run all indexes in 10 minutes.**	*Extensive processing is required.*	Costs some money to buy package. Once package is bought, it can run on a PC at no additional charge accept for downloading data.
WesVarPC	Is never the most stable index, but is seldom terrible.	Can run all indexes in a short time.	Some processing required.	Free except for cost of downloading data but free version will no longer be updated.
VPLX	**Is best more than half the time.**	Can run all indexes in a short time.	*Extensive processing is required.*	**Free except for cost of downloading data.**

- **Bold type indicates that package is best at satisfying the criterion**.
- *Italic type indicates that package is worst at satisfying the criterion*

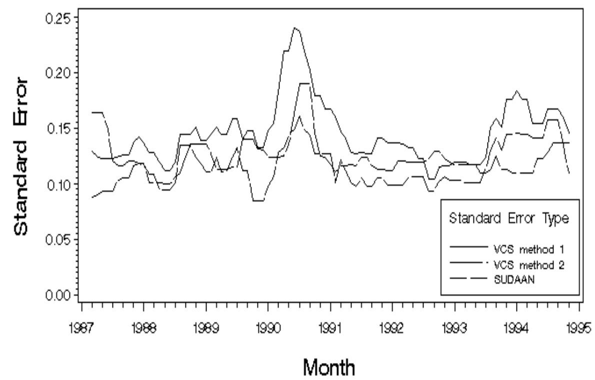
Plot 1A

SMOOTHED AND UNSMOOTHED STANDARD ERRORS
FOR TRANSPORTATION IN LARGE PSUS
USING VPLX METHOD–STABILITY CRITERION=.010



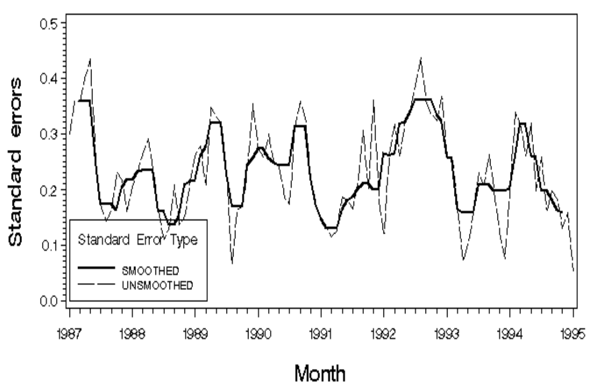
Plot 2A

SMOOTHED STANDARD ERRORS, DEC. 1987–DEC. 1995
USING LINEARIZATION METHODS
INDEX AREA=ALL US, PRICED ITEM=ALL ITEMS



Plot 1B

SMOOTHED AND UNSMOOTHED STANDARD ERRORS
FOR TRANSPORTATION IN LARGE PSUS
USING VCS VERSION 2–STABILITY CRITERION=.031



Plot 2B

SMOOTHED STANDARD ERRORS, DEC. 1987–DEC. 1995
USING REPLICATION METHODS
INDEX AREA=ALL US, PRICED ITEM=ALL ITEMS