Investigation of Balanced Repeated Replication (BRR) Variance Estimation for The Survey of Residential Alterations and Repairs (SORAR)

Laura Ozcoskun[1], Katherine Jenny Thompson, and Quatracia Williams

U.S. Census Bureau
Room 2136-4, Washington, DC 20233 / laura.t.ozcoskun@census.gov, katherine.j.thompson@census.gov, and qqwill@yahoo.com

Background

The U.S. Census Bureau publishes quarterly estimates of expenditures for the improvement and repairs of residential property. These estimates are obtained from the Consumer Expenditures Survey (CE) and The Survey of Residential Alterations and Repairs (SORAR). The SORAR sample is subsampled from the CE, which is a probability sample of households designed to represent the total U.S. noninstitutional civilian population. SORAR requests detailed expenditures about the **entire** sampled property and produces estimates of totals and quarter-to-quarter change estimates.

SORAR uses partial (grouped) Balanced Repeated Replication (BRR) to estimate the variances. SORAR totals are expansion estimates computed from adjusted sample weights. The SORAR weight adjustment procedure consists of four separate adjustments to the initial sample weights. The presently-used BRR implementation does not replicate this weight-adjustment procedure: it uses a "shortcut" approach, multiplying the full sample adjusted weights by the appropriate replicate factor. This approach was chosen to save processing time, based on analyst-inspection of the survey's shortcut and fully-replicated variance estimates. This is not, however, an unbiased variance estimation procedure. For example, Canty and Davison (1999) found that this shortcut approach **underestimates** the variance, often by a large amount, especially when the magnitude of the adjustments to the survey weights are fairly large. On the other hand, it is not guaranteed that the shortcut variance estimates are necessarily **underestimates.** The SORAR totals are post-stratified estimates. Such estimates are combined ratio estimates, and many studies have shown that a strict application of the BRR method to smooth non-linear statistics often yields <u>overly-large</u> variance estimates (e.g., Judkins (1990), Rao and Shao (1996), Rao and Shao (1999)). If weighting or estimation cell sizes are small, this positive-bias effect is magnified.

Prior to 2004, the SORAR collected data on a quarterly basis. In 2004, it began a monthly data collection, but continues to publish only quarterly estimates while evaluating the reliability of the monthly data. Obviously, a major component of this evaluation would be the level of the standard errors. The concerns expressed above regarding the precision of the BRR shortcut method variances – namely, the inability to determine whether the shortcut procedure BRR variance estimates are consistently over – or underestimates – motivated this research.

When the weight adjustment procedure is fully replicated, the grouped balanced half-sample replicate variance estimator produces unbiased (but inconsistent) estimates of the true variance (Rao and Shao (1996) and Valliant (1996)). By extension, shortcut BRR variances are biased, although the direction of the cumulative biases of estimator with multiple reweighting steps is difficult – if not impossible – to determine. Consequently, we investigate whether all – or some – components of the SORAR weight adjustment procedure should be replicated for variance estimation using historic SORAR data. From our perspective, differences between corresponding fully replicated and shortcut variance estimates provide empirical evidence for using a fully replicated method. We also explore the use of modified half sample (MHS) replication on SORAR variance estimates; this method is designed to alleviate the small sample sizes in the replicate weighting cells obtained under BRR while maintaining the optimal replicate estimation properties of BRR. We present and discuss two sets of empirical comparison results: one from a quarterly data collection (sixteen consecutive quarters; and one from the monthly collection (fifteen consecutive months). Based on these results, we recommend that SORAR use MHS replication instead of BRR to

[1] This paper reports the results of research and analysis undertaken by the U.S. Census Bureau staff. It has undergone a Census Bureau review more limited in scope than that given to official Census Bureau publications. This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress.

produce variance estimates in future applications.

Survey Design and Estimation

The SORAR is a mail survey of approximately 4,000 properties, subsampled from the Consumer Expenditures Survey (CE). The CE selects a two-stage sample; the first stage is a probability proportional to size sample of counties or Minor Civil Divisions (MCD); the second stage is a systematic sample of housing units. PSUs are stratified by region (Northeast, Midwest, South, and West) and metropolitan status. The SORAR sample consists of properties identified in the CE as out-of-scope because they are vacant (vacant owner units) and properties identified in the CE as being one to four unit properties with no resident owner and all properties (excluding owner occupied condominiums) with five or more housing units (occupied rental units).

SORAR's expansion estimates are non-linear. The SORAR has a four-step weight adjustment procedure. The separate stages of this procedure have remained unchanged for several years. However, the weight adjustment cells described below will be introduced in the 2007 survey year.

The first two steps of the SORAR weight adjustment procedure adjust the CE sample weights and yield unbiased estimates. The first step multiplies the CE survey sampling weight by the unit duplication control factors (a factor that accounts for subsampling in the field, bounded between 1 and 4). These unbiased weights (W_{hi}^{*}) are next adjusted for unit nonresponse. The weighting classes for the unit nonreponse adjustment procedure are defined by region and metropolitan status (8 separate cells). Since the PSUs are defined within region and metropolitan area, the SORAR PSUs are entirely contained within weighting class cell. Let *s* index the SORAR sampled units, and *r* index the respondent sampled units. The nonresponse adjustment factor d_p for the p^{th} weighting class is given by

$$d_p = \frac{\sum_{(hi)\in s} W_{hi}^{*}\eta_{hi}^{p}}{\sum_{(hi)\in s} W_{hi}^{*} a_{hi}\eta_{hi}^{p}} = \frac{\hat{N}_{sp}}{\hat{N}_{rp}} \qquad (2.1)$$

where η_{hi}^{p} is a weighting class indicator variable equal to 1 if unit *i* in PSU *h* is in weighting cell *p* and is 0 otherwise, and a_{hi} is a response status indicator variable. This is an adjustment-to-sample weighting procedure as defined by Kalton and Flores-Cervantes (2003). If a cell contains insufficient respondents (currently less than 5), then the non-response weight adjustment is performed within the region, dropping the metropolitan status classification.

SORAR estimates are further post-stratified to independently obtained estimates of population totals for occupied rental units and for vacant units. This 2nd stage adjustment procedure is an adjustment-to-population weighting procedure as defined by Kalton and Flores-Cervantes (2003). The weighting class adjustment cells are again defined by region and metropolitan status. The post-stratification adjustment factor for a weighting class *p* is given by

$$b_{pt} = \frac{N_{pt}}{\sum_{(hi)\in s} d_p W_{hi}^{*} a_{hi}\eta_{hi}^{p}\varsigma_{hi}^{pt}} = \frac{N_{pt}}{\left(\frac{\hat{N}_{sp}}{\hat{N}_{rp}}\right)\hat{N}_{rp}^{t}} \qquad (2.2)$$

where $t = o$ (occupied) or v (vacant), N_{pt} is the control total value for unit type *t* in weighting cell *p*, and ς_{hi}^{pt} is an indicator variable for weighting class *p* and unit type ***t***. Again, if a cell contains insufficient respondents (currently less than 5), then the metropolitan status classification is dropped from the calculation. The independently obtained totals used in the 2nd stage adjustment have a variance associated with them that we do not account for in this study. Excluding the variance of the control totals from our study does not affect our comparison of the different variance estimation methods because including it would cause a constant increase across all of the variance estimates being considered.

Because the non-response adjustment procedure uses the post-strata as weighting cell classes, when all weighting cells contain more than 5 respondents, the survey estimates reduce to the usual post-stratified estimator,

$$\hat{Y}_i = \sum_p N_{pt}\left(\frac{\hat{Y}_{rp}^{t}}{\hat{N}_{rp}^{t}}\right) \qquad (2.3)$$

Notice that the non-response adjustment factors from (2.2) cancel in (2.3).

SORAR collects expenditures for the **entire** property on each form. To allocate the appropriated expenditure data to the individual housing unit[2] level, the post-stratified weights are divided by the total number of units on the property (u_k). Thus, the SORAR final weights (W^F_{hi}) are the product of the sample weight, the duplication control factor, the non-response adjustment factor, and a post-stratification factor divided by the total number of units on the property.

SORAR monthly estimates are obtained by summing the final-weighted expenditures data. Monthly estimates are then summed to obtain the quarterly estimates.

Due to the difficulty in obtaining information from non-resident owners of vacant housing units, the SORAR unit response rates are quite low in all weighing cells. In addition, the post-stratification adjustments for rental units in metropolitan areas and vacant owned units are generally larger than 10, and the adjustment factors for rental units in non-metropolitan areas are often less than 1 (especially in the Northeast and South), showing the effect of the essentially self-selected SORAR universe.

Variance Estimation

The SORAR uses a *partial* or *grouped* balanced repeated replication (BRR) method to produce variance estimates. Balanced repeated replication (BRR) or balanced half-sample replication is a variance estimation method designed for a two PSU/stratum design. With BRR, a half-sample replicate is formed by selecting one unit from each pair of PSUs and weighting the selected unit by 2 (so that it represents both units). Consequently, estimates from every PSU are in each replicate although half are weighted by zero.

Though the number of half-samples can be quite large (2^L, where L = the number of strata), the BRR method requires that only some of the possible half-sample replicates be created to obtain a variance estimator that reduces to the textbook (approximate sampling formula) variance estimator, which is an unbiased estimator of the true population variance. We use a Hadamard matrix (a $k \times k$ orthogonal matrix consisting of 1's and -1's, where k is a multiple of 4) to specify the replicates, choosing a value of k that is greater than L. To minimize the number of replicates, we usually choose the smallest value of k possible. With a fully balanced design, each pair of sample units is assigned to a unique row in the Hadamard matrix. Within pairs, each PSU is assigned to one panel, and this panel assignment is used in conjunction with each column value to assign replicate factors (Wolter, 1985, pp. 111-115).

To reduce the number of replicates, surveys may assign more than one strata to the same row in the Hadamard matrix. This is known as *partial* balancing (Wolter, 1985, pp. 125-131) or as the *grouped* balanced half-sample method (Rao and Shao (1996) and Valliant (1996). With grouped half-sample replication, units in the L strata are divided into G groups with approximately L/G strata in each group. Units within groups are split into two panels, and half-samples are formed for each group. With grouped balanced half-samples, the replicate variance estimator is not equivalent to the textbook estimator. Although it is still an unbiased estimator for the true variance, its variance is larger than the corresponding fully balanced method because cross-product terms between replicates no longer cancel (Wolter, 1985, pp. 127). Rao and Shao (1996) and Valliant (1996) prove that the grouped half-sample replicate estimator is inconsistent. The SORAR assigns 40 **groups** of PSUs (many consisting of more than two PSUs) to rows in a 44×44 Hadamard matrix.

The SORAR drops the metropolitan status classification within a weighting cell if the weighting cell contains less than 5 respondents. Because initial sample sizes are so small, the BRR replicate samples in these cells (essentially halved from the full sample) are often less than 5. Consequently, replicate weighting cells are often collapsed (more than in the full survey data procedure), which can induce a positive bias in the variance estimates (Rao and Shao, (1999)).

The Modified Half Sample (MHS) replication method developed by Robert Fay (1989) addresses this problem of overly perturbed replicate weights and drastically reduced replicate sample sizes. MHS uses replicate weights of m and 2-m (0 # m # 1) in each half sample. All sample units are explicitly included in each replicate, and all replicates retain orthogonal balance.

[2] The survey's ultimate sampling unit

To construct replicate weights for BRR or MHS replication, let f_{jr} be the value of row *j*, column *r* of the Hadamard matrix for collapsed group *j*. Replicate weights *j* = 1 through *k* for sample unit *i* are computed as:

$$rwt_{ji} = \begin{cases} \left(1 + f_{jr}(1-m)\right) \times W_{hi} \text{ if unit } i \text{ is in panel 1} \\ \left(1 - f_{jr}(1-m)\right) \times W_{\text{hi}} \text{ if unit } i \text{ is in panel 2} \end{cases} \quad (3.1)$$

When $m = 0.5$ and $f_{jr} = 1$, panel 1 units receive replicate factors of 1.5 and panel 2 units receive replicate factors of 0.5; otherwise, panel 2 units receives replicate factors of 1.5 and panel 1 units receive replicate factors of 0.5. Selecting $m = 0$ yields the original BRR replicate factors.

The MHS (BRR) variance estimator is given by

$$\hat{v}(\hat{Y}_0) = \frac{1}{(1-m)^2} \frac{1}{k} \sum_r \left(\hat{Y}_r - \hat{Y}_0\right)^2 \quad (3.2)$$

where $\hat{Y}_r$ is the replicate estimate (with replicate factors specified by column *r*) and $\hat{Y}_0$ is the full-sample estimate.

Currently, the SORAR replicates are the products of the replicate factors ($1 \pm f_{jr}(1\text{-}m)$) and the final weights (W^F_{hi}). The shortcut procedure variances are conditional estimates of the true variance (conditioned on survey response rates and post-stratification factors within weighting cells). In contrast, fully replicating the weight adjustment procedures – using either BRR or MHS – yields unconditional variance estimates. This, however, takes time, especially with 44 replicates. This time is justified when both sets of replicates are sufficiently different, given the above-listed theoretical concerns.

Empirical Data Evaluations

Our first set of available research data were from SORAR survey years 2000 through 2003 (16 quarters total). SORAR data were collected quarterly during this time. In 2004, the survey began a monthly collection, although only quarterly estimates are published. Our second set of research data were from survey years 2004 and 2005. This second set of research data are used to evaluate monthly and quarterly variance estimates. We discuss the results for the two different types of data collection separately.

Quarterly Data Collection (Survey Years 2000 Through 2003)

Our first analysis used fully edited and imputed data from the survey years 2000 through 2003. Our estimates are different from the corresponding published estimates because we implement a slightly modified version of the weighting procedure described above: unfortunately, control counts for vacant units were only available at the census **region** level, affecting the implementation of the 2nd stage procedure. The published estimates use different weighting cells for both the non-response adjustment and 2nd stage adjustments. Using the old weighting procedure to assess reweighting effects on replicate variances would have provided a "pure" comparison in terms of variance estimator effects. However, in doing this, we would be unable to directly compare the quarterly data collection results to the monthly data collection results described in the following (monthly data collection) evaluation.

To assess the effects of replicating the weighting procedure, we examine a series of quarterly variance estimates for six expenditures characteristics using:

- replicate estimates based on full-sample final weights (shortcut procedure)
- replicate estimates that replicate both the 2nd stage ratio adjustment and the units on property adjustment, but does not replicate the nonresponse adjustment (partial replicate reweighting, or the partial procedure)
- replicate estimates that replicate the nonresponse adjustment procedure, the 2nd stage ratio adjustment, and the units on property adjustment (full replicate reweighting or the full procedure)

We consider both BRR and MHS replication.

Table 1 presents the median, maximum, and minimum standard error estimates of the six survey expenditures characteristics computed using BRR and MHS, with full, partial, and shortcut procedures for the sixteen quarters. Note that the shortcut variances are equivalent for BRR and MHS. Additions, Improvements to Structure, and Improvements Outside of Structure are rarely reported expenditures (on the average, 7-percent, 40-percent, and 14-

percent, respectively), as is Total Improvements (the sum of these three values); Maintenance and Repairs has a very high item response rate (about 89-percent) as does Total Expenditures.

Table 1: Standard Error Statistics for Full, Partial, and Shortcut Procedures (In Millions)

Item	Reweighting	Method	Median	Minimum	Maximum
Additions	Shortcut	BOTH	2950	897	6030
	Partial (2nd Stage Only)	BRR	3160	1020	6110
	Partial (2nd Stage Only)	MHS	2300	947	5680
	Full	BRR	3220	1020	6200
	Full	MHS	2300	941	5680
Total Improvements	Shortcut	BOTH	12000	6630	159000
	Partial (2nd Stage Only)	BRR	11900	6100	154000
	Partial (2nd Stage Only)	MHS	11300	5120	105000
	Full	BRR	11600	5920	146000
	Full	MHS	11300	4550	105000
Improvements to Structure	Shortcut	BOTH	9730	3250	45000
	Partial (2nd Stage Only)	BRR	9110	4770	40700
	Partial (2nd Stage Only)	MHS	8870	2590	45300
	Full	BRR	9190	4700	40900
	Full	MHS	8790	2320	45300
Improvements Outside of Structure	Shortcut	BOTH	3430	235	157000
	Partial (2nd Stage Only)	BRR	3010	883	154000
	Partial (2nd Stage Only)	MHS	2650	195	102000
	Full	BRR	2960	899	146000
	Full	MHS	2460	198	102000
Maintenance and Repairs	Shortcut	BOTH	6690	3790	17600
	Partial (2nd Stage Only)	BRR	6370	4590	10500
	Partial (2nd Stage Only)	MHS	5570	3540	9630
	Full	BRR	6600	4340	10700
	Full	MHS	5440	3550	9750
Total Expenditures	Shortcut	BOTH	16900	10300	162000
	Partial (2nd Stage Only)	BRR	14000	9340	155000
	Partial (2nd Stage Only)	MHS	12700	8700	106000
	Full	BRR	14000	9360	147000
	Full	MHS	13300	8220	106000

Except for additions, on average the shortcut method standard errors are larger than the corresponding fully or partially replicated standard errors. This pattern is similar to the findings in Yung and Rao (1996) for post-stratified estimators with a stratified jackknife estimator and makes sense. Why? The shortcut procedure could overestimate the variability induced by post-stratification since it treats the 2nd stage-adjusted estimates as expansion estimates and does not include covariance between the numerator and denominator in the calculations (unlike the fully and partially replicated procedures). For five of the six items, the corresponding fully and partially replicated BRR standard errors are larger than the MHS counterparts. Again, these results are consistent with Rao and Shao (1996), indicating the presence of a positive bias in the BRR replicate standard errors caused by overly small sample sizes in adjustment cells.

Note that the partial and the fully replicated standard errors are very close within replication method. Figures 1 and 2 present time-series plots of the standard errors for Additions estimated using BRR and MHS replication respectively. Figures 3 and 4 present similar time-series plots of the standard errors for Total Expenditures.

Figure 1: Time Series Plots of Balanced Repeated Replication (BRR) Standard Errors for Additions (In Millions)

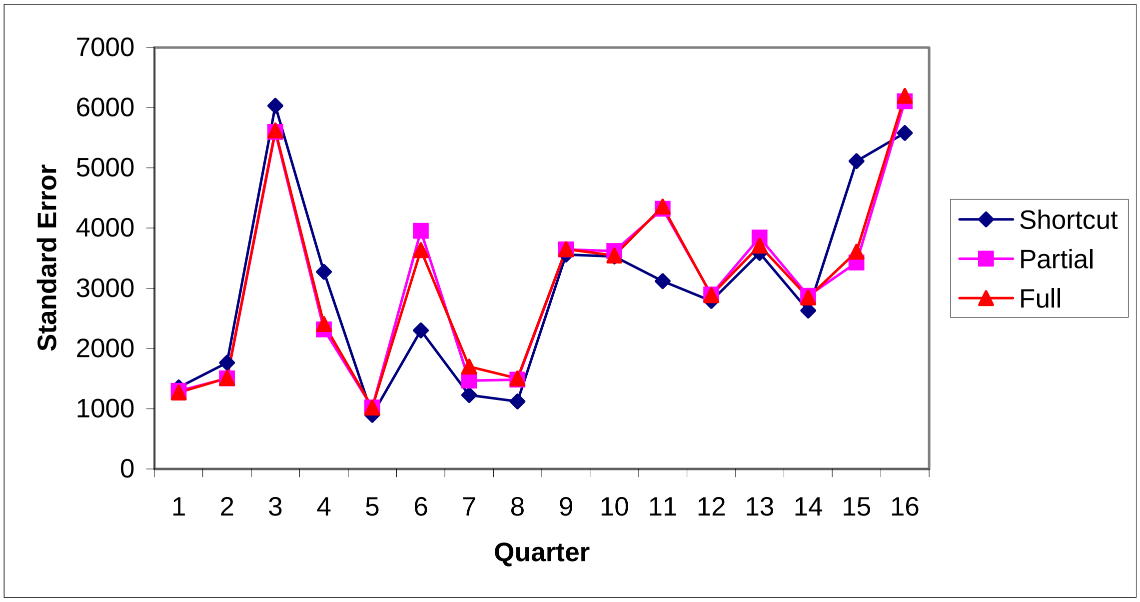


Figure 2: Time Series Plots of Modified Half-Sample Replication (MHS) Standard Errors for Additions (In Millions)

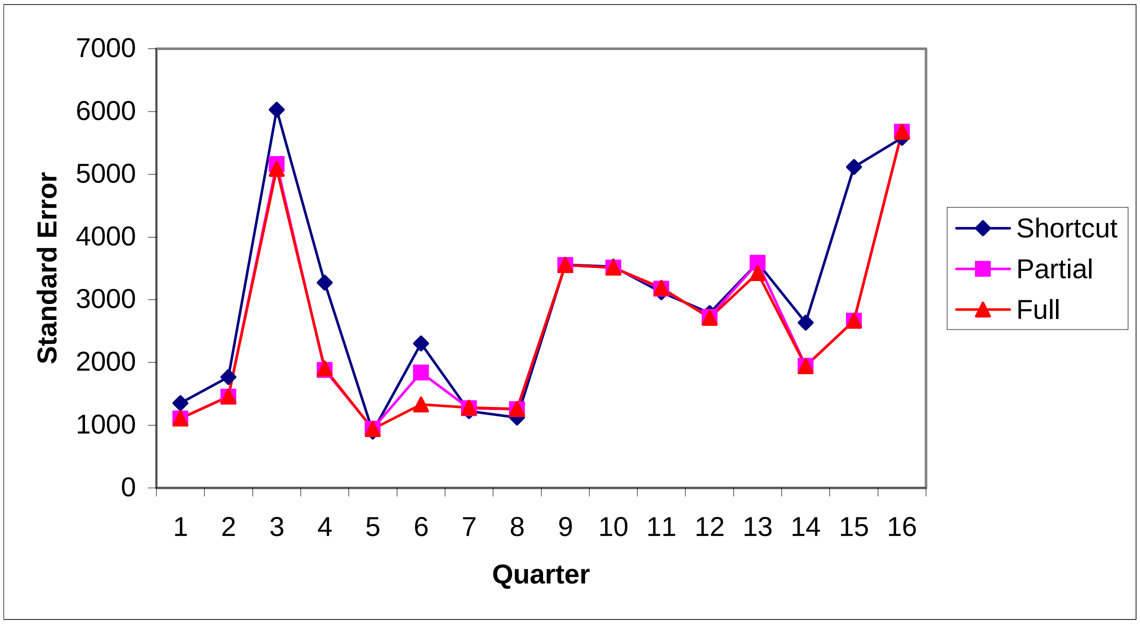


Figure 3: Time Series Plots of Balanced Repeated Replication (BRR) Standard Errors for Total Improvements

(In Millions)

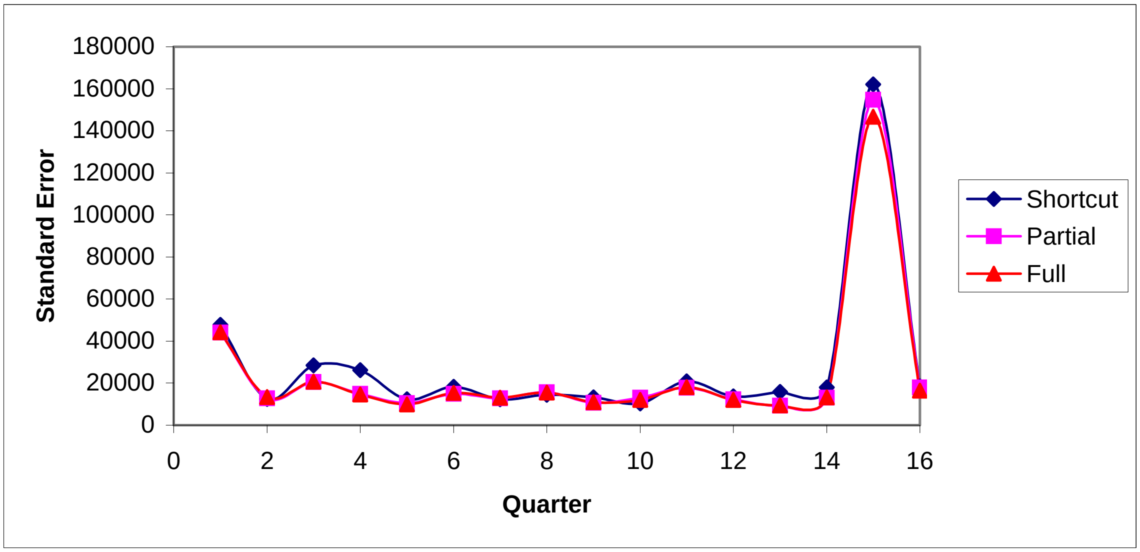
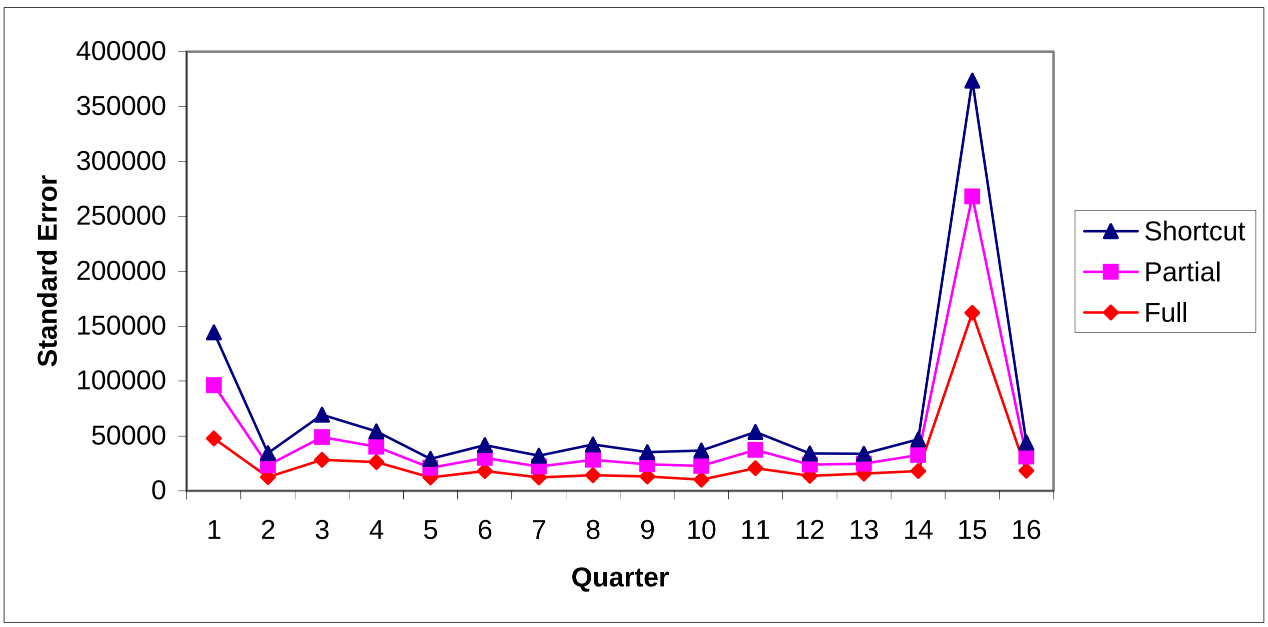


Figure 4: Time Series Plots of Modified Half Sample (MHS) Replication Standard Errors for Total Improvements

(In Millions)



For our studied items, the two corresponding sets of BRR standard error estimates are very close. In contrast, the fully replicated MHS standard errors are generally smaller than the corresponding partially replicated standard errors. The latter results is more intuitively appealing: the fully replicated standard errors directly compensate for **each** stage of weight adjustment, whereas the partially replicated standard errors only indirectly account for the non-response variance via the 2nd stage adjustment and the assignment of non-respondent cases to replicates.

Table 2 presents the median, maximum, and minimum coefficient of variation estimates of the six survey expenditures characteristics computed using BRR and MHS, with full, partial, and shortcut procedures for the sixteen quarters.

Table 2: Coefficient of Variation Statistics for Full, Partial, and Shortcut Procedures

Item	Reweighting	Method	Median	Minimum	Maximum
Additions	Shortcut	BOTH	0.55	0.37	0.88
	Partial (2nd Stage Only)	BRR	0.56	0.41	1.14
	Partial (2nd Stage Only)	MHS	0.49	0.31	0.81
	Full	BRR	0.56	0.41	1.05
	Full	MHS	0.45	0.31	0.81
Total Improvements	Shortcut	BOTH	0.29	0.22	0.7
	Partial (2nd Stage Only)	BRR	0.28	0.16	0.68
	Partial (2nd Stage Only)	MHS	0.27	0.16	0.46
	Full	BRR	0.28	0.16	0.64
	Full	MHS	0.27	0.16	0.46
Improvements to Structure	Shortcut	BOTH	0.33	0.19	0.51
	Partial (2nd Stage Only)	BRR	0.32	0.21	0.49
	Partial (2nd Stage Only)	MHS	0.3	0.17	0.51
	Full	BRR	0.31	0.2	0.49
	Full	MHS	0.29	0.17	0.51
Improvements Outside of Structure	Shortcut	BOTH	0.44	0.29	0.97
	Partial (2nd Stage Only)	BRR	0.44	0.27	1.1
	Partial (2nd Stage Only)	MHS	0.44	0.24	0.78
	Full	BRR	0.45	0.27	1.12
	Full	MHS	0.43	0.24	0.79
Maintenance and Repairs	Shortcut	BOTH	0.22	0.13	0.41
	Partial (2nd Stage Only)	BRR	0.18	0.11	0.33
	Partial (2nd Stage Only)	MHS	0.18	0.12	0.21
	Full	BRR	0.18	0.11	0.32
	Full	MHS	0.18	0.11	0.21
Total Expenditures	Shortcut	BOTH	0.22	0.14	0.6
	Partial (2nd Stage Only)	BRR	0.19	0.13	0.57
	Partial (2nd Stage Only)	MHS	0.17	0.12	0.39
	Full	BRR	0.19	0.13	0.54
	Full	MHS	0.17	0.12	0.39

Table 2 demonstrates that the method of variance estimation **does not** affect ability to detect significant differences from zero (c.v. > 1/1.645 ≈ 0.6 indicates no statistical difference from zero at the 90% confidence level).

Table 3 presents the median, minimum, and maximum ratios of BRR to MHS **fully replicated** procedure standard errors. Except for total expenditures, all of the BRR standard errors are larger than the corresponding MHS standard errors, indicative of the presence of positive bias caused by small replicate sample sizes in several weighting cells with BRR.

Table 3: Ratios of BRR/MHS Fully Replicated Procedure Standard Errors

Item	Median	Minimum	Maximum
Additions	1.04	0.68	1.58
Total Improvements	1.07	0.68	2.01
Improvements to Structure	1.11	0.52	4.58
Improvements Outside of Structure	1.06	0.80	1.85
Maintenance and Repairs	1.11	0.84	1.39
Total Expenditures	0.94	0.37	1.80

SORAR unit response rates can be quite low (as low as 15% in some weighting cells). Consequently, the similarity of the partial and fully replicated BRR standard errors was unexpected. To examine the effect of the non-response adjustment procedure alone on SORAR standard errors, we computed SORAR estimates and standard errors without the post-stratification (2nd stage) weight adjustment procedure. Table 4 presents these median, maximum, and minimum standard error estimates, again using both BRR and MHS replication.

Table 4: Standard Errors Statistics for Full and Shortcut Procedures – No 2nd Stage Adjustment (In Millions)

Item	Reweighting	Method	Median	Minimum	Maximum
Additions	Shortcut	BOTH	661	135	2143
	Full	BRR	626	128	2094
	Full	MHS	532	127	2080
Total Improvements	Shortcut	BOTH	2055	1568	31374
	Full	BRR	2066	1483	29394
	Full	MHS	2051	1425	28889
Improvements to Structure	Shortcut	BOTH	1634	841	4469
	Full	BRR	1595	878	4537
	Full	MHS	1575	841	4359
Improvements Outside of Structure	Shortcut	BOTH	641	152	31106
	Full	BRR	594	137	29008
	Full	MHS	595	135	28693
Maintenance and Repairs	Shortcut	BOTH	1272	330	2811
	Full	BRR	1277	320	2774
	Full	MHS	1251	313	2537
Total Expenditures	Shortcut	BOTH	2544	1846	31897
	Full	BRR	2588	1514	29589
	Full	MHS	2590	1452	29129

For a given item, all three standard error estimates are very close (within 4 to 10 percent of each other). It appears that the contribution to the variance due to non-response is negligible for this survey. Clearly, the majority of the unaccounted-for variance in the shortcut procedure is due to the 2nd stage adjustment. This is reasonable, because the non-response adjustment factors cancel in the combined ratio estimates as shown in (2.3).

For quarterly data, the partial replication procedure does not provide improvements in quality of variance estimates or computation time over the fully replicated procedure. Consequently, the evaluations described in the following sections compare only the fully replicated BRR and MHS variance estimates to corresponding shortcut procedure estimates.

Monthly Data Collection (Survey Years 2004 and 2005)

Quarterly Estimates. While there is some question about the suitability of the monthly SORAR estimates for publication, there is no question that SORAR will continue to publish quarterly estimates. As of the 2004 survey year, quarterly estimates are computed as the sum of the three contributing months' estimates. The weight adjustment procedure is performed each month. This contrasts with the weighting procedure used in the previous survey years, which performed weight adjustment for the quarter. Furthermore, the monthly SORAR estimates are not independent, and the summed-from-monthly-data quarterly estimates implicitly include some covariance.

Table 5 presents the standard error estimates and c.v.'s of five key expenditure items computed using BRR and MHS, with full and shortcut procedures using data from the four available quarters.

Table 5: Standard Errors and C.V.' s of Quarterly Estimates (2004-2005 Survey Years)

(Standard Errors are in Millions)

Item	Quarter	Year	Standard Error			Coefficient of Variation (CV)		
			Shortcut	BRR	MHS	Shortcut	BRR	MHS
Additions	2	2004	431	431	213	0.53	0.53	0.26
	3	2004	467	699	289	0.47	0.70	0.29
	4	2004	589	688	328	0.65	0.76	0.36
	1	2005	142	157	76	0.49	0.54	0.26
Total Improvements	2	2004	945	1000	493	0.29	0.30	0.15
	3	2004	672	993	414	0.28	0.42	0.17
	4	2004	593	685	321	0.43	0.50	0.23
	1	2005	2470	3550	1800	0.60	0.85	0.43
Improvements To Structure	2	2004	139	156	72	0.37	0.42	0.19
	3	2004	339	385	178	0.41	0.47	0.21
	4	2004	104	111	52	0.32	0.34	0.16
	1	2005	2380	3500	1840	0.65	0.96	0.50
Improvements Outside Of Structure	2	2004	1190	1160	623	0.56	0.54	0.29
	3	2004	305	332	157	0.54	0.59	0.28
	4	2004	73	69	53	0.49	0.46	0.35
	1	2005	185	180	180	0.93	0.90	0.90
Maintenance And Repairs	2	2004	1170	1430	522	0.32	0.38	0.14
	3	2004	731	756	370	0.25	0.26	0.13
	4	2004	333	342	163	0.17	0.17	0.08
	1	2005	859	561	538	0.58	0.38	0.36
Total Expenditures	2	2004	1720	2180	845	0.25	0.31	0.12
	3	2004	1130	1380	611	0.22	0.26	0.12
	4	2004	709	821	382	0.21	0.24	0.11
	1	2005	2780	3470	1660	0.49	0.62	0.29

As in the quarterly (2000-2003) data collection, the corresponding fully replicated BRR standard errors are **consistently** larger than their MHS counterparts. In contrast to the quarterly data collection results, however, the BRR standard errors are generally slightly **larger** than the shortcut standard errors. This difference is probably caused by the computation of **monthly** weight adjustment factors (three per quarter compared to one per quarter in the 2000-2003 data sets). Finally, the fully replicated MHS standard errors are always the smallest of the three competing variance estimates.

The change to monthly data collection from quarterly has led to a level shift in expenditure **estimates**, complicating a direct comparison of standard errors from the 2000-2003 survey years to those displayed in Table 5. The c.v.'s are more comparable. As with the 2000-2003 quarterly estimates, the c.v.'s computed with fully-replicated MHS variance estimates are smaller than those computed from the two other variance estimates, often by a factor of as much as two.

Monthly Estimates. The standard error patterns displayed with either set of quarterly estimates are quite consistent. The monthly estimates' results are considerably less so. Figures 5 through 7 present time-series plots of each standard error estimate for total expenditures, maintenance and repairs, and for improvements to structure.

Figure 5: Time Series Plot of Standard Error Estimates for Total Expenditures (In Millions)

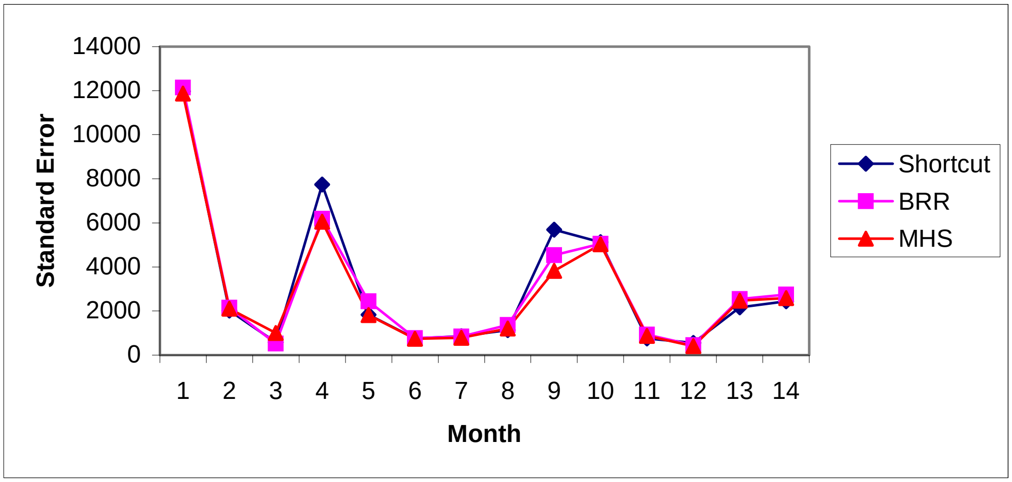


Figure 6: Time Series Plot of Standard Error Estimates of Maintenance and Repairs (In Millions)

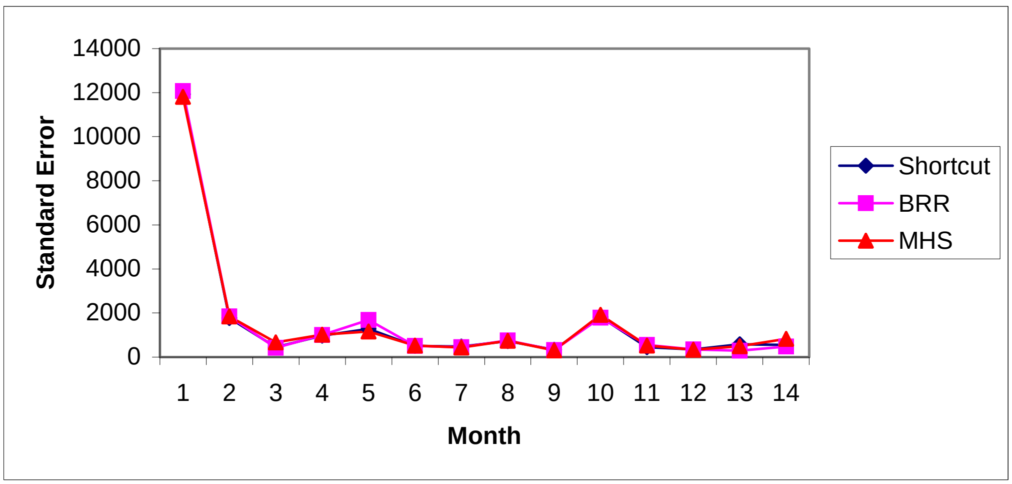
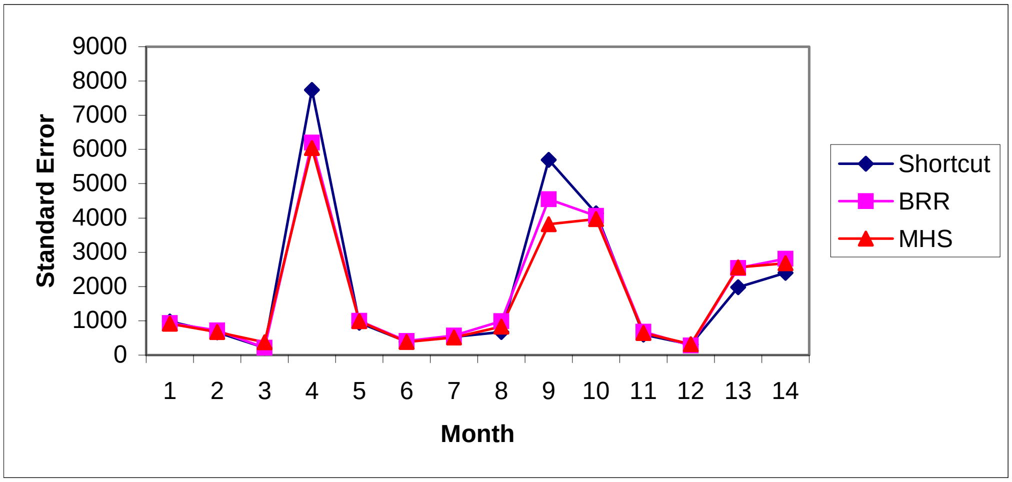


Figure 7: Time Series Plot of Total Improvements

(In Millions)



In general, the three corresponding standard error estimates are quite close, regardless of characteristic. Again, the fully replicated BRR standard error estimate is generally the largest of the three. However, the shortcut and fully replicated MHS variance estimates are quite similar, and it fluctuates from month to month as to which one has the higher value. Table 6 reinforces these observations, presenting median, minimum, and maximum ratios of BRR to MHS, BRR to shortcut, and MHS to shortcut standard error ratios.

Table 6: Standard Error Ratio Statistics for Monthly Standard Error Estimates

	BRR/MHS			BRR/Shortcut			MHS/Shortcut		
Item	Median	Minimum	Maximum	Median	Minimum	Maximum	Median	Minimum	Maximum
Additions	1.04	0.90	1.21	1.03	0.79	1.50	1.01	0.41	1.24
Total Improvements	1.04	0.56	1.20	1.06	0.80	1.48	1.01	0.38	1.75
Improvements To Structure	1.05	0.97	1.26	1.09	0.80	1.30	1.03	0.38	1.55
Improvements Outside Of Structure	1.02	0.65	1.16	1.03	0.87	1.12	1.03	0.67	1.95
Maintenance And Repairs	1.01	0.57	1.46	0.99	0.49	1.31	1.03	0.04	1.82
Total Expenditures	1.05	0.52	1.35	1.02	0.80	1.33	0.98	0.05	1.79

Conclusion

Ask the statistician on the street, and she will tell you that in these days of cheap and intensive computing power, there is no excuse for <u>not</u> completely replicating a weight adjustment procedure. Doing otherwise yields biased variance estimates, and the time-savings are probably not sufficient to justify the departure from the theory.

Having said that, we found very little difference between the shortcut procedure and fully replicated standard errors using BRR with the SORAR data. This justifies the survey analysts' original position – why not save time when there's no practical difference? From a methodologist's perspective, the similarity between the shortcut and fully replicated BRR standard errors are a red flag, especially given the low survey response rate and the magnitude of the 2nd stage (post-stratification) adjustments. In fact, this made us question the choice of variance estimator and led us to consider and ultimately recommend the modified half-sample (MHS) method.

Our empirical results can be explained theoretically. Recall that SORAR publishes quarterly estimates and consequently we scrutinize differences in quarterly estimates to assess the variance estimators. With BRR, the shortcut and fully replicated standard errors are similar. The small replicate weighting cell sample sizes contribute to a bias in the fully replicated estimates. This bias is obviously not present in the shortcut method standard errors; instead, this procedure ignores variance-reducing post-stratification adjustments. With MHS, the fully replicated standard errors are consistently smaller than their shortcut counterparts, reflecting post-stratification improvements.

More important, the MHS replicate weighting cell sample sizes are the same as in the full sample, avoiding the artificially induced variance increase due to cell collapsing.

There's a twist to our story. Prior to the evaluation, we were concerned with **underestimation** in precision. Instead, we found the reverse. This makes our analysts very happy with our recommendation, even if it does lead to an increase in computing time.

No study is complete without recommendations for future research. Now that we have determined an appropriate variance estimator, there are other variance estimation concerns that should be investigated. First, SORAR modifies outlying expenditures values via winsorization. This study does not perform this outlier-correction technique, and we do not know what effect it has on variance estimates. Second, our evaluation data contains imputed values, and consequently all of our variance estimates are underestimates since we do not explicitly account for imputation. This unaccounted-for variance component could be sizeable for several expenditures items. Finally, non-response adjustment factors cancel with 2nd stage adjustment with the current procedure when there's no cell collapsing. This is convenient, but probably does not reduce the bias of the estimates (Little and Rubin, 2002, p.48). However, we have very little evidence of similar response propensity within the region by metropolitan status non-response adjustment cells; in fact, we suspect that the response propensity is much more correlated with the number of units on a property within region. In the near future, we plan a more thorough investigation of this. It is possible that a change in non-response weighting cells would have negligible effects on variance estimates. However, it is always good to develop techniques that reduce the bias in the estimates.

References

Canty, A.J. and Davison, A.C. (1999). Resampling-based Variance Estimation for Labour Force Surveys. *The Statistician*, **48**, pp. 379-391.

Fay, R.E. (1989). Theory and Application of Replicate Weighting for Variance Calculations. *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 212-217.

Judkins, D.R. (1990). Fay's Method for Variance Estimation. *JOS*, **6**, pp. 223-229.

Kalton, G. and Flores-Cervantes, I. (2003). Weighting Methods. *JOS*, **19**, pp. 81-97.

Little, R.J.A. and Rubin, D.B. (2002). Statistical Analysis With Missing Data. New York: Wiley.

Rao, J.N.K. and Shao, J. (1996) On Balanced Half-Sample Variance Estimation in Stratified Random Sampling. *Journal of the American Statistical Association*, **91**, pp. 343 –348.

Rao, J.N.K. and Shao, J. (1999). Modified Balanced Repeated Replication for Complex Survey Data. *Biometrika*, **86**, pp. 403-415.

Valliant, R. (1996). Limitations of Balanced Half-Sampling. *Journal of Official Statistics*, **12**, pp. 225-240.

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

Yung, W. and Rao, J.N.K. (1996). Jackknife Linearization Variance Estimators Under Stratified Multi-Stage Sampling. *Survey Methodology*, **22**, pp. 23-31.