MODIFIED BALANCED REPEATED REPLICATIONS

David R. Judkins, Westat, Inc.
1650 Research Blvd., Rockville, MD 20850

1. Background and Outline

The variance estimation technique of balanced half-samples or balanced repeated replications (BRR) is well known. Formalized by McCarthy [5] in 1966 with roots in work at the Census Bureau in the 1950's it has long been used as one of a number of techniques to estimate the variance of nonlinear statistics from complex designs. (For a detailed and accessible discussion, see Wolter [6].) In fact, it is often used for linear statistics from simple designs because of the simplicity of the system once it has been set up for more complicated situations.

The relative advantages of this system relative to Taylor linearization and the jackknife have been long debated. Recently Kovar [4] and Hansen and Tepping [3], reported on simulations which indicated that the jackknife appeared to be better than BRR both in terms of bias and stability and that the Taylor linearization may be better yet.[1]

However, a few years ago Robert Fay from the Census Bureau (Dippo, Fay and Morganstein [1]) suggested a modification to BRR that held the promise of improving the method's stability. The technique was motivated by the observation that the standard half-sample variance estimator occasionally runs into problems estimating the variance of ratios because the denominators are zero for some replicates. Less drastic but also a problem, some ratio replicates can be extremely large because of near-zero denominators. This is caused by the fact that when half the sample is zero weighted and half is double weighted, less common groups disappear more frequently than in the full sample. (The jackknife avoids these problems by only dropping one observation at a time.)

Dr. Fay's idea was to use weights of .5 and 1.5 instead of 0 and 2 for the half samples within each stratum, or more generally, weights of k and 2-k for $0 \leq k < 1$.[2] Although the mean square error of the ratio replicates from the full sample ratio becomes much smaller as k approaches unity, it is transformed to a reasonable estimator of the variance by multiplication by $1/(1-k)^2$. For example, if k=.9, then the mean square error of the ratio replicates from the full-sample ratio must be multiplied by 100 to obtain a reasonable estimator of variance. The advantage of this approach is that any population represented in the full sample will also be represented in each replicated re-sample. Thus, any ratio defined in the full sample will also be defined with any of the generated sets of replicate weights.

Another way of thinking about k is that 100(1-k) can be recognized as a perturbation factor. In the standard BRR, weights are being perturbed 100 percent. If k=0.9, then the weights are only being perturbed 10 percent. Heuristically, it makes sense that gentler perturbation should lead to fewer extreme replicate estimates, thereby yielding lower kurtosis of the replicates and hence better stability of the variance estimator.

I first discuss some simple properties of the estimator which I refer to as BRR/Fay. Then I discuss a Monte Carlo type comparison of BRR/Fay with the standard BRR and with the jackknife. Finally, I mention some extensions of the idea also due to Dr. Fay that are useful in some special situations.

2. Summary

There are two sets of considerations: mathematical and practical. Mathematically, the modification to BRR preserves several desirable properties of the BRR estimator. Furthermore, when the number of strata is small and the reliability of the denominator is low, the modification can dramatically reduce the bias and especially the variance of the variance estimates of ratio statistics.

Practically, BRR/Fay is just as easy to implement as the plain BRR. After the replicate weights have been created, the only change from current procedure is in the application of the multiplier. For example, if k=.99, then all variance estimates produced with old software need to be multiplied by 10,000 (standard error estimates by 100).

It is more difficult to make a choice between BRR/Fay with k close to 1 and the jackknife. The simulation results are very similar, and there is hardly any difference in computer time or software development. Either appears to be a good choice for a simple estimator where large numbers of variances are required for non-linear statistics from complex designs. However, Kovar has hinted that there may be some important differences for the median.

3. Equivalency for Linear Statistics

BRR/Fay is identical to BRR for linear statistics. This implies unbiasedness and the equivalency of a balanced set of replicates to the full set of replicates for linear statistics. I show this first for the case of a single stratum and then for the case of multiple strata.

3.1 One Stratum

The BRR/Fay variance estimator in this case is

$$\text{estvar}=(kx_1+(2-k)x_2-x)^2/(1-k)^2,$$

where $0 \leq k < 1$, x_1 and x_2 are independent half-sample estimates for a single stratum, and $x=x_1+x_2$. It is trivial to show that for any k, $\text{estvar}=(x_1-x_2)^2$, which is the standard BRR estimate of variance for the case of a single stratum.

3.2 Multiple Strata

Also in the case of multiple strata, the BRR/Fay is identical to BRR for linear functional statistics. Let d_{th} be the entry for row t and column h in a Hadamard matrix of dimension equal to the number of replicates. (By definition, d_{th} = 1 or -1 and $\Sigma_t d_{th} d_{tm}=0$ for $h \neq m$.) The replicate weights for the two half-samples in stratum h for replicate t may be written as $\delta_{th1}=1+d_{th}(1-k)$ and $\delta_{th2}=1-d_{th}(1-k)$, where k is defined as above. Note that for k=0, this reduces to the familiar 2 and 0 or 0 and 2. Also note that

$$\Sigma_t(\delta_{th1}-1)(\delta_{tm1}-1)=(1-k)2\Sigma_t d_{th}d_{tm}.$$

Thus, for balanced replicates and $h \neq m$, the left-hand expression is zero. Similarly, for $h \neq m$,

$$0=\Sigma_t(\delta_{th}1-1)(\delta_{tm}2-1)$$

$$=\Sigma_t(\delta_{th}2-1)(\delta_{tm}2-1)$$

Now let x_{hi} be the estimate of interest for the i-th half-sample in the h-th stratum and $x=\Sigma\Sigma_{hi}x_{hi}$. The BRR/Fay variance estimator is

$$\Sigma_t\{\Sigma\Sigma_{hi}\delta_{thi}x_{hi}-x\}^2/[T(1-k)^2],$$

where T is the number of replicates. Using the relationships derived above and the linearity of x it is now easy to show that BRR/Fay is identical to BRR.

This equivalency is reassuring since for the traditional BRR, the use of a balanced set of replicates gives the same variance estimate for a linear statistic as the full set of replicates. Of course, use of a balanced set of replicates does not guarantee the same thing for non-linear statistics. This is true both for the traditional BRR and for BRR/Fay.

4. Simulation Studies

To gauge the accuracy of BRR/Fay relative to BRR and the jackknife, I did a simulation study along the same lines as the studies by Hansen and Tepping and by Kovar.[3] The basic situation is 32 strata, each with a bivariate normal pair of variables.

4.1 Background on the Monte Carlo Technique

Each Monte Carlo replicate provides an estimate of variance. By averaging these across many replicates, an estimate of the expected value of the variance estimator is obtained. By taking mean square variation from that average across replicates, an estimate of the variance of the variance estimator is obtained. The true variance of the statistic of interest (such as a ratio or regression coefficient) may be estimated by taking the mean square deviation of the statistic across replicates from its average.

The simulated variance (of the statistic of interest) and the estimated mean of the variance estimator are compared to get an estimate of the bias of the variance estimator. One natural measure is the ratio of the estimated mean of the variance estimator to the simulated variance. Alternatively, one can view the variance estimator as being really aimed at the mean square error. In which case, it is more appropriate to look at the ratio of the estimated mean of the variance estimator to the simulated mean square error. Similarly, there are several reasonable measures of the stability of the variance estimator. The measure used in this report is the ratio of the square root of the observed mean squared deviation of the variance estimator from the observed mean square error (of the underlying statistic) to that same mean squared error.

4.2 Mechanics

Hansen and Tepping created three main artificial populations with a common structure. They then varied the parameters to create a larger set of populations. In this study, I looked at only one of their main populations. The common structure was 32 strata with two independent and identically distributed observations, (x_{h1},y_{h1}) and (x_{h2},y_{h2}), from a bivariate normal population within each stratum. The correlation between x and y was assumed to be constant over all strata. The coefficient of variation (cv) per stratum for x was constant at 10 percent over all strata. The cv per stratum for y was roughly uniform at 24 percent. The means of x and y were allowed to change across strata. Table 1 gives the stratum weight, means of x and y and their standard errors by stratum for population number 1. To create variations within this main population, the correlation (ρ) is varied and the standard error of x is varied by a uniform factor, as is the standard error of y.

When there are multiple strata, the question arises whether to use complimentary balanced half-samples and the "full" jackknife or just balanced half-samples and the half-jackknife. The difference is whether any action taken on half the sample in a stratum is also repeated on the other half. Since Kovar indicated that the "full" versions of BRR and the jackknife gave practically the same results as the "half" versions and since the full versions double the number of replicates and hence the number of calculations, I only simulated the half versions.

Since the earlier authors had not censored stratum half-sample means to be bounded away from zero, I didn't either. However, it may be better in future studies to do so. Variables that can assume negative values are after all seldom used in ratio estimation.

To generate the pseudo-random numbers, I used the standard congruential operator available in a package called GAUSS. When different methods were being tested on the same population parameters, a single seed and multiplier were used for the algorithm so that the methods were compared on the same population. When the population parameters changed, new seeds and occasionally new multipliers were used.

One caution noted by the earlier authors is that this sort of Monte Carlo design yields much smaller variances on the mean of the variance estimator than on the estimated true variance (and mean square error) of the underlying statistic. To correct for this disparity in accuracy, more replicates are generally used to estimate the true variance than to estimate the mean of the variance estimator.

When the underlying statistic was the ratio of y to x, I used a single set of 1000 replicates to estimate both the variance and MSE of the ratio across replicates and the mean and variance of the variance estimator. When the underlying statistic was the regression coefficient of y on x, I used one set of 200 replicates for the mean and variance of the variance estimator and an independent set of 5000 to estimate the variance and MSE of the regression coefficient across replicates.

4.3 Results

Table 2 summarizes the results for the ratio; Table 3 for the regression coefficient. Bias and stability are evaluated relative to the observed mean square error of the underlying statistic rather than relative to the variance of the same. I did this to keep the presentation similar to that used by Hansen and Tepping and by Kovar. Note that there is some abbreviated scientific notation in the tables; for example, $5.3x10^{-4}$ is abbreviated as 5.3-4.

Based on the simulations that I ran, it appears that all the estimators are essentially equivalent for estimating variance on ratios of normal variables where the standard errors per stratum are small. Regardless of the standard errors per stratum, BRR/Fay appears to converge to the jackknife as k approaches unity. As k approaches 0 (the traditional BRR), quality deteriorates markedly if the standard errors per stratum are large. (The value of stability greater than 900 or 90,000% is not a typographical error.)

For estimating the variance on a regression coefficient, there appears to be a monotone relationship between k and the bias of the variance estimator. Kovar already showed that the standard BRR (k=0) tends strongly toward overestimates of the variance while the jackknife may have a slight tendency toward underestimation. My results indicate that increasing k leads to smaller estimates of variance. In fact, k=.99 leads to smaller estimates than the jackknife. There may be some intermediate value of k that gives comparability to the jackknife. The stability figures for the standard BRR are worse than those for the other estimators, among which I see no clear pattern.

5. Handy Notes for Special Situations

"P-weights"

Dr. Fay also figured out how to adjust replicate weights for NSR strata to reflect an adjustment for collapsing strata (only relevant when one PSU has been selected per stratum). The adjustment is the traditional one found in Hansen, Hurwitz and Madow [2]. If there is sufficient correlation between the PSU measure of size and characteristics of interest, this adjustment will ameliorate the bias in the variance estimates caused by unequal sizes of the collapsed strata. (The bias caused by unequal stratum means remains.) Under this procedure, the replicate weights for the two strata in a collapsed NSR stratum are

$1+2d_{th}(1-k)$(MOS for other stratum)/(collapsed MOS) and

$1-2d_{th}(1-k)$(MOS for other stratum)/(collapsed MOS),

where d_{th} is defined as before, $0 \leq k < 1$, and this k must equal that used for SR strata. Note that if k=0 and the two strata are equal in size, then the weights are the familiar 2 and 0 or 0 and 2.

Three Collapsed Strata

It is not uncommon to encounter a stratified multi-stage design with one sample unit per stratum and an odd number of strata. One strategy for using BRR in this case is to collapse three strata into a single collapsed stratum and then assign two first-stage units to one half-sample within the collapsed stratum and one to the other. Dr. Fay suggested using the replicate weights

$1+d_{th}(1-k)\sqrt{2}$ for the half-sample consisting of a single first stage unit and

$1-d_{th}(1-k)/\sqrt{2}$ for the half-sample consisting of two first stage units.

These replicate weights may be used with the same multiplier of $1/(1-k)^2$.

More than 2 sample units per stratum

There are several techniques for applying BRR to designs with more than two sample units per stratum. Some of these are discussed in Wolter[6]. These may also be used with BRR/Fay.

ACKNOWLEDGEMENTS

This work was supported by the National Institute of Dental Research as part of work by WESTAT on the Epidemiologic Survey of Oral Health in Adults.

REFERENCES

[1] C.S. Dippo, R.E. Fay, D.H. Morganstein, "Computing Variances from Complex Samples with Replicate Weights," ASA Srvy Rsr. (1984), pp489-494. (Section 4 of the article is relevant.)

[2] M.H. Hansen, W.N. Hurwitz, and W.G. Madow, Sample Survey Methods and Theory, Vol. 2, (New York: Wiley, 1953), pp218-222.

[3] M.H. Hansen and B.J. Tepping, "Estimation of variance in NAEP," Unpublished manuscript (1985).

[4] J. Kovar, "Variance estimation of nonlinear statistics in stratified samples," Methodology Branch Working Paper No. 85-052E, (Statistics Canada, 1985).

[5] P.J. McCarthy, "Replication: An Approach to the Analysis of Data from Complex Surveys," Vital and Health Statistics, Series 2, No. 14 (Washington D.C.: National Center for Health Statistics, 1966).

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

FOOTNOTES

1 Given the greater software development effort required by Taylor linearization, Hansen and Tepping would probably disagree with Kovar on the later point.

2 Such weights were in fact used for the 1984 panel of the Bureau's Survey of Income and Program Participation.

3 It may be possible to approximate directly the variance of the variance estimator by taking all fourth partial derivatives with respect to all stratum half-sample totals, but it appears much simpler to use Monte Carlo techniques.

Table 1. Parameters for population 1

Stratum (h)	W_h	μ_{xh}	μ_{yh}	σ_{xh}	σ_{yh}
1	0.042	100.000	90.000	10.000	25.000
2	0.042	95.000	75.000	9.500	24.000
3	0.042	90.000	70.000	9.000	22.000
4	0.039	98.000	75.000	9.800	22.000
5	0.039	93.000	70.000	9.300	20.000
6	0.037	98.000	75.000	9.800	24.000
7	0.037	96.000	75.000	9.600	23.000
8	0.037	94.000	75.000	9.400	22.000
9	0.037	92.000	70.000	9.200	24.000
10	0.034	96.000	75.000	9.600	23.000
11	0.034	94.000	70.000	9.400	20.000
12	0.034	92.000	70.000	9.200	22.000
13	0.034	90.000	70.000	9.000	22.000
14	0.031	96.000	75.000	9.600	25.000
15	0.031	94.000	70.000	9.400	20.000
16	0.031	92.000	70.000	9.200	18.000
17	0.031	90.000	70.000	9.000	19.000
18	0.031	88.000	70.000	8.800	20.000
19	0.031	86.000	65.000	8.600	20.000
20	0.031	84.000	60.000	8.400	18.000
21	0.031	82.000	60.000	8.200	16.000
22	0.031	80.000	60.000	8.000	20.000
23	0.028	90.000	70.000	9.000	22.000
24	0.028	85.000	65.000	8.500	18.000
25	0.028	80.000	60.000	8.000	20.000
26	0.025	90.000	70.000	9.000	20.000
27	0.025	85.000	60.000	8.500	18.000
28	0.025	80.000	50.000	8.000	15.000
29	0.025	75.000	50.000	7.500	14.000
30	0.020	75.000	50.000	7.500	16.000
31	0.016	75.000	45.000	7.500	14.000
32	0.013	75.000	45.000	7.500	12.000

Table 2. Estimating the variance of the ratio of two normal variables from a stratified sample -- comparison of bias and stability for BRR/Fay and jackknife

Coefficient of variation per stratum: numerator	Coefficient of variation per stratum: denominator	ρ	Obs mse	k	Bias relative to obs mse	Stability relative to obs mse
24%	10%	0.8	5.3-4	0	1.01	0.31
				0.5	1.01	0.31
				0.99	1.01	0.31
				J	1.01	0.31
		0.5	7.1-4	0	1.01	0.29
				0.5	1.01	0.29
				0.99	1.01	0.29
				J	1.01	0.29
		0.2	9.2-4	0	0.96	0.29
				0.5	0.96	0.29
				0.99	0.96	0.29
				J	0.96	0.29
120%	50%	0.8	1.3-2	0	1.11	0.45
				0.5	1.09	0.42
				0.99	1.09	0.40
				J	1.09	0.40
		0.5	1.7-2	0	1.08	0.38
				0.5	1.07	0.36
				0.99	1.07	0.36
				J	1.07	0.36
		0.2	2.1-2	0	1.11	0.39
				0.5	1.10	0.38
				0.99	1.10	0.38
				J	1.10	0.38
24%	100%	0.8	6.8-3	0	1.26	1.81
				0.5	1.07	0.94
				0.99	1.02	0.81
				J	1.02	0.79
		0.5	9.3-3	0	1.27	1.30
				0.5	1.08	0.86
				0.99	1.03	0.75
				J	1.04	0.77
		0.2	1.0-2	0	1.28	1.16
				0.5	1.11	0.86
				0.99	1.08	0.84
				J	1.08	0.84
240%	150%	0.2	1.1-1	0	35.18	909.08
				0.5	1.17	1.40
				0.99	1.08	0.96
				J	1.09	0.96

The true parameter Θ is known. The estimator of Θ is Θ'. The variance and bias of Θ' across all simulation replicates are S^2 and B. The estimator of $Var(\Theta')$ is s^2. The variance and mean of s^2 across all simulation replicates are $S^2(s^2)$ and $M(s^2)$. Then

"bias" = $M(s^2)/(S^2+B^2)$ and

"stability" = $\sqrt{\frac{[S^2(s^2)+(M(s^2)-S^2-B^2)^2]}{S^2+B^2}}$

Table 3. Estimating the variance of the regression of one normal variable on another from a stratified sample -- comparison of bias and stability for BRR/Fay and jackknife

Coefficient of variation per stratum: numerator	Coefficient of variation per stratum: denominator	ρ	Obs mse	k	Bias relative to obs mse	Stability relative to obs mse
24%	100%	0.8	4.8-4	0	1.10	0.53
				0.5	0.98	0.46
				0.99	0.94	0.44
				J	0.96	0.46
		0.5	8.8-4	0	1.07	0.49
				0.5	0.96	0.43
				0.99	0.93	0.43
				J	0.94	0.45
		0.2	1.1-3	0	1.12	0.63
				0.5	0.98	0.52
				0.99	0.93	0.49
				J	0.94	0.48
		-0.7	7.0-4	0	1.11	0.59
				0.5	0.99	0.51
				0.99	0.95	0.49
				J	0.98	0.53

The true parameter Θ is known. The estimator of Θ is Θ'. The variance and bias of Θ' across all simulation replicates are S^2 and B. The estimator of $Var(\Theta')$ is s^2. The variance and mean of s^2 across all simulation replicates are $S^2(s^2)$ and $M(s^2)$. Then

"bias" = $M(s^2)/(S^2+B^2)$ and

"stability" = $\sqrt{\frac{[S^2(s^2)+(M(s^2)-S^2-B^2)^2]}{S^2+B^2}}$