Proceedings of the Annual Meeting of the American Statistical Association, August 5-9, 2001

An Extension of Fay's Method for Variance Estimation to the Bootstrap

Pedro J. Saavedra
Pedro J. Saavedra, ORC Macro, 11785 Beltsville Dr., Calverton, MD 20705

Key Words: **replication, simulation, BRR, ratios, subpopulation**

Balanced Repeated Replications (BRR) ia a variance estimation method usually implemented for cases where there are two units (be they PSUs or otherwise) per stratum. For each replication one of the units is dropped from the sample and the weights of the other adjusted. BRR can be applied when there are more than two units per stratum, but it is more difficult to implement.

Fay's method for variance estimation (Judkins, 1990) modifies the balanced repeated replications approach by instead of deleting one half-sample, multiplying the weights of one half-sample by k (where $0 < k < 2$) and the other by 2-k. Thus when one uses Fay's method every respondent is used in every replicate, but the weights will be different from replicate to replicate.
This approach permits for ratio and small subpopulation estimates where the denominator may be zero for some half-sample or the subpopulation may be not present in the half-sample. This paper examines an extension of Fay's concept to the bootstrap.

One of the advantages of Fay's method is that a denominator which could otherwise disappear for a given replicate will not disappear, for the units will be there with a smaller weight. An estimate will be obtained for every replicate if one was obtained for the original sample. Furthermore, under certain circumstances for linear estimates where the estimate would not disappear, the BRR estimate will be the same as the estimate using Fay's method.

The heuristic motivation for this paper came as several variance estimation approaches were compared as part of a simulation study. While the design used did not permit easy use of balanced half sample replications, Monte Carlo half-sample simulations were used. This approach essentially selects half the sample and weighs up the estimate to the total sample. It takes more computer time than BRR , but is easy to program and yields results comparable to other methods. But it seems that Fay's approach can be easily extended to the Monte-Carlo half sample, weighting half the sample up and the other half down by the same percent.

This method led to consideration of whether the Fay approach could be applied to other forms of variance estimation. In the case of the bootstrap one can keep all the units in each replicate and merely change the weights. One has essentially replicate weights which do not vanish for any case.

Why would this method be expected to work at all? Consider an estimate X from a sample. Let X' be an estimate of the same parameter derived from a bootstrapped sample. Now let $0 < k < 1$. It follows that one can obtain an estimate of the same parameter by letting X"=kX'+(1-k)X. The extension can be conceived either as a weighted average of each bootstrapped sample and the original sample, or as a transformation of bootstrap weights where the new weights are kw'+(1-k)w (where w' is the untransformed bootstrapped weight, and w is the original weight.

Using this method every subpopulation present in the original sample will be present in each modified bootstrapped sample. It can be shown that for linear estimates the modified bootstrap's standarderror estimate divided by k yields the same standard error estimate as theregular bootstrap's standard error estimate. Following the Fay terminology, 100k will be called the perturbation factor of the modified bootstrap, where a perturbation factor of 100% gives one a standard bootstrap.

Thus the variance of this modified bootstrap will be

$$\frac{\sum (k\bar{x}_j + (1-k)\bar{x} - \bar{x})^2}{n} = \frac{\sum (k\bar{x}_j - k\bar{x})^2}{n} = k^2 \frac{\sum (\bar{x}_j - \bar{x})^2}{n}$$

and hence the result can be divided by k^2 to obtain the same result as one would for the ordinary bootstrap, provided one has a linear estimate where the denominator never vanishes for each bootstrapped sample. This, however, cannot be said for other estimates. Indeed, if there are weight adjustments and other deviations from linearity, there will be differences in the two sets of estimates.

Example 1: The EIA-878

The first example where the procedure was tested is the EIA-878 gasoline survey. This is a survey of gasoline prices at the pump. For each station the weighted volumes add to known values per cell. The survey is post-stratified separately for each gasoline grade, and the total gasoline price (combining grades) is obtained at the basic stratum level. A regular bootstrap is used to obtain the variance of the prices for publication cells which are in general combinations of sampling strata The sampling strata are used to adjust weights to obtain correct volumes for different regions. The term publication cell is used for cells where in principle one might consider publishing prices, though some are too small to do so.

Given this design there are two kinds of estimates where one might expect a different between the simple bootstrap and the modified bootstrap. One such estimate is for cells that are particularly small. In this survey the geographical regions (PADDs) served as sampling strata, but attainment status regions within geographic regions were post-stratification cells. Because the sampling design allows for varying sample sizes within post-stratification cells (and thus publication cells) the bootstrap allowed for the occasional bootstrap sample where the cell vanished. Estimates for these cells are among the ones where the two bootstrap methods can be expected to differ.

The second type of estimate where the two approaches are likely to differ is in estimating prices for total gas, combining the three grades. Because the post-stratification adjusts separately for total volume for each of the three grades, the modified bootstrap will yield different results than the regular bootstrap. Using a single week, bootstrapping 1,000 samples, and using perturbation factors of 100% and 80% one finds indeed that the modified bootstrapped yields identical results as the simple bootstrap for regular, midgrade and premium gasoline, except in very small cells (cells with only four to six outlets per cell). In these cases, the variance estimates are lower using the modified bootstrap (perturbation of 80%).

Likewise, the variance estimates for total gasoline are also lower for the modified bootstrap than for the simple bootstrap. In each case the results are still of a similar magnitude. For example, estimates for a very small cell went down from a standard error of 4.0 cents to 3.7 cents. And one total gasoline estimate in a larger cell went from 1.8 cents to 1.5 cents. Decreasing the perturbation factor continued to decrease the non-linear estimates, while leaving the linear estimates unaffected. The estimate which went from 4 to 3.7 went down to 3.5 with a perturbation factor of 50%.

This led us to try a perturbation factor of 99%. The results were practically identical to the perturbation factor of 100%. Unlike the regular bootstrap, this method guarantees a denominator for every bootstrapped estimate (the regular bootstrap -- or perturbation factor of 100% -- computes variances using only the bootstrapped samples where the denominator does not vanish) but does not alter the results, at least for the kind of estimates that have been examined here.

Example 2: The Zip Code simulation

The second example uses a file of zipcode areas. One thousand samples of 100 zip code areas with 1990 census populations greater than zero and at least one housing unit were drawn. For each sample, estimates were obtained for average population per zip code area, average number of housing units per zip code area and the ratio of population to housing units. The estimates were obtained for the total sample and for states with at least three zip code areas in the sample.

One thousand bootstraps were conducted for each sample, using perturbation factors of 100% and 80%. In addition standard errors were estimated from the frame by calculating them directly using only those samples where at least three units entered the sample. This approach means that there were fewer than 1,000 samples for which estimates were possible for each state (though 1,000 samples yielded national estimates).

Thus there are four standard error estimates for states and the nation:

1) Estimate obtained from the variance of sample estimates. This is one estimate, while the others are estimated separately for each sample.

2) Use of the standard error formula.

3) Regular bootstrap (perturbation factor of 100%).

4) Modified bootstrap (perturbation factor of 80%).

The standard error differences for pairs of estimates are divided by the first of these, and then the results are tested for significance across all the samples for which the estimates were possible. The state estimates were also pooled to examine the results, even though the different states were not independently sampled.

Table 1 summarizes the results. As expected, the total units and the total population national estimates were identical for the two bootstraps. The standard error for the ratio estimate was lower using the modified bootstrap. All sample based estimates were on the average lower than the frame based estimate, and the standard formula yielded slightly higher standard errors than either bootstrap. The sample based estimates were off by 0.9% to 1.5% for population and units, and by about 10% for the ratio.

Absolute relative errors were calculated for the three sample based methods and the differences were tested for significance across the 1,000 simulations. Each bootstrap sample used in the regular bootstrap was then combined with the original sample to form a modified bootstrap sample. Matched pairs t-tests were used to test for significant differences across the simulations. The regular bootstrap outperformed the modified bootstrap as well as the standard error formula at the .01 level.

The state estimates were then pooled. No state had three zip codes in every sample. It was however, possible to compare results across all samples with at least three representatives for that state. In addition, the results for the different states were pooled to obtain a global picture. Division of differences by the frame estimate permitted an examination of the results across states.

All sample based estimates tended to underestimate variances (compared with the frame estimate) except the simple bootstrap for the ratio, but the modified bootstrap outperformed the regular bootstrap and the standard error formula for the two linear estimates.

The simple bootstrap had the largest value for the ratio estimate, but the standard deviation was much larger than for the others so the results were not significant. The results are probably due to some extreme estimates.

What is true of the pooled state estimates is not necessarily true for all individual states and none of three estimates from samples tended to bevery close to the criterion.

Conclusions:

The simulations are for now inconclusive. One issue is the importance of the perturbationfactor. In BRR, Fay started using one factor and subsequent simulations showed that a different factor was appropriate. For the bootstrap, there may be a relationship between the perturbation factor and the accuracy -- the data suggests at least a relationship between the perturbation factor and the magnitude of the estimates. The original motivation of the modified bootstrap -- avoiding the vanishing denominator -- is accomplished by using aperturbation factor close to 100% without altering the simple bootstrap estimates that much. But if one uses an estimate drawn through multiple samples from the frame, the simulations suggest that an 80% perturbation factor improves accuracy of estimates when the sample sizes are such that vanishing denominators are likely. Further simulations and theoretical work may be in order.

References

Judkins, D.R. 1990 Fay's method for variance estimation, *Journal of Official Statistics* 6: 223-239.

ACKNOWLEDGEMENTS

The author wishes to thank Ronaldo Iachan and Tigran Markaryan for acting as peer reviewers.

Table 1: Means of Absolute Relative Errors

National Estimates

METHOD	**Total Population**	**Total Units**	**Ratio**
Formula	.113	.110	.230
Regular Bootstrap	.114	.111	.227*
Modified Bootstrap	.114	.111	.228

* Significantly lower than the other two methods (p<.001)

Table 2: Means of Absolute Relative Errors

Pooled State Estimates

METHOD	**Total Population**	**Total Units**	**Ratio**
Formula	.572	.572	.592
Regular Bootstrap	.563	.564	.948*
Modified Bootstrap	.558	.558	.550

* Only value *not* significantly different from the other two in the column.