May 23, 2013

Addendum Number 2

"Projected Variance for the Model-Based Classical Ratio Estimator,"
http://interstat.statjournals.net/YEAR/2012/abstracts/1209001.php?Name=209001

Additional Application Consideration:

Equation 4 on page 15 in this article should yield reasonable volume coverage requirements when employing the model-based classical ratio estimator (CRE), a robust estimator for sample surveys, if the input values for the variables shown are sufficiently close to reality. A range of reasonable values for those input variables may be used in a sensitivity analysis. Further considerations regarding quasi-cutoff sampling, cutoff sampling, and balanced sampling, employing such coverage to determine sample size requirements, are scheduled to appear in the *ASA Survey Research Methods Section Proceedings* following the 2013 Joint Statistical Meetings.

In practice, in cases where past samples have been very large, there may have been a great deal of nonsampling error, due to the overburdened nature of such data collection, and quality control problems that often result when there is an attempt to collect data on many small members of a population, on a frequent basis. Data quality for the smallest observations, even just a few, collected so frequently, is often problematic. The key is to estimate $\sigma^*_{e_0}$ (the RMSE in SAS PROC REG, for example, when using one regressor, through the origin, with w=1/x (the CRE)). In such a case as just described, the author suggests using historical data, where the smallest observations might be dropped, depending upon (1) sample sizes, (2) variability of the $\sigma^*_{e_0}$ (the estimated standard error of the estimated residuals), and hopefully (3) sufficient experimentation to study this. This is recommended because residuals for observed y-values (sample values), associated with the smallest x-values, are generally inflated disproportionately by nonsampling error, and may not yield a good $\sigma^*_{e_0}$ (RMSE) estimate. Note that we even see this kind of nonsampling error problem, near the origin, in analytical chemistry. See Lee(2012). Also, note Knaub(2002), pages 17 and 20, where points on a scatterplot are described as reminiscent of a thermometer. However, a horn with a bulb at one end and then flaring outward would be a better analogy to the appearance that these data may have when plotted. This is also why the CRE is suggested when the estimated coefficient of heteroscedasticity would generally be greater than that for the CRE**:** The CRE is robust against this problem.

Note that in some cases it may be assumed that the smaller respondents actually do behave differently. In such cases, (1) stratification, (2) a regression weight step-function, or perhaps even (3) multiple regression could be considered – but that is a different matter.

Knaub, J.R., Jr. (2002), "Practical Methods for Electric Power Survey Data," detailed version, *InterStat*, July 2002, http://interstat.statjournals.net/YEAR/2002/abstracts/0207001.php?Name=207001. (Note shorter version in ASA Section on Government Statistics proceedings, 2002, JSM CD.)

Lee, C.R.(2012), "Use of replicate calibration samples in analytical chemistry: uncertainties due to lack of knowledge of heteroscedasticity," found at http://www.analyt.chrblee.net/calibration/calibscedastpost2.pdf