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

EVALUATING ALTERNATIVE RAKING APPROACHES

Douglas Willson, Anka Wagner
National Analysts Inc.

Douglas Willson, National Analysts Inc.
1700 Market Street, Philadelphia, PA, 191031

Key Words: Calibration; Raking.

1 Introduction

Survey researchers often adjust preliminary survey analysis weights so that sample estimates match known control totals for auxilliary variables. These adjustments are attractive in that the resulting statistical estimates have desirable properties, including reduced bias and increased efficiency in some circumstances. Over the years, many different approaches for raking or calibration have been proposed. Singh and Mohl (1996) provide a detailed description for many of these methods.

There are two types of decisions that must be made during any practical raking application. First, a specific raking algorithm must be chosen from the rather large set of available methods. Deville and Sarndal (1992) showed that a wide class of calibration estimators are (first-order) asymptotically equivalent to the generalized regression estimator. While this may appear to suggest that choosing a method is not important, in practice one often finds that the different raking/calibration algorithms produce very different results in finite samples.

Historically, researchers have addressed this issue using summary measures of fit that compare the raked/calibrated weights relative to the base analytic weights. Examples of comparative studies on calibration methods are given in Deville, Sarndal and Sautory (1993) and Stukel and Boyer (1993). Intuitively, a set of calibrated weights that match control totals and that is "close" to the base weights would be preferred to a set that was "farther away". While these measures have some appeal, these decisions rules are usually *ad hoc* and do not have a formal statistical basis.

A second decision that must be made concerns the selection of raking targets. In many situations there are a large number of potential targets, and a researcher must ones to include. If one set of raked weights is constructed that simultaneously matches all control totals, including a large number of targets may substantially increase the variability of the raked weights and the associated sampling variability of the survey estimates. From this perspective, increased variability is the "price" associated with including a large number of targets. Although research concerning approaches for selecting targets has been limited, Chambers, Skinner, and Wang (2000) (henceforth CSW) have recently described a number of statistical approaches for selecting raking targets.

In this paper we examine the behavior of a variety of different raking algorithms within the context of two examples from market research. The paper illustrates the range of differences in raked weights that can result in practice because of differences in (and assumptions underlying) raking algorithms. In addition, we develop several statistical decision-rules for selecting raking targets, and evaluate their performance within the context of the two examples.

The paper will proceed as follows: Section 1 provides some background concerning raking methods and (1992) results concerning the asymptotic equivalence of the various raking/calibration algorithms. Section 2 develops two methods for selecting raking targets, following the suggestions of Chambers , Skinner, and Wang (2000). Section 3 presents some empirical results using two examples from market research surveys. Section 4 concludes and suggests avenues for future research.

2 Background

Deville and Sarndal (1992) and Deville, Sarndal and Sautory (1993) (henceforth DSS) consider the following notation. Let n, N denote the sample size and population size respectively. Let d_k represent the usual design-based survey weight (the base weight) for respondent k. Let y_k be the value of a variable of interest for the k^{th} population element, and let

x_k=$\{x_{k1}, ...x_{kJ}\}$ be a vector of J auxiliary variables. For the auxilliary variables, we assume that the population totals or benchmark constraints are known, *i.e.* τ_j=$\sum_{i=1}^{N} x_{ij}$.

The basic idea behind calibration is to develop new weights $\{w_k$, k=1...n$\}$ for each respondent such that the survey sample produces estimates that match the population or benchmark totals. Following D-S, this can operationalized as a minimum-distance problem, with different calibration estimators employing different distance measures.

To illustrate, DSS consider distance measures $G_k(w, d)$ satisfying certain regularty conditions with $g_k(w, d)$= $\partial G_k/\partial w$. Calibration estimators are chosen to minimize distance measured as $\sum_{k=1}^{n} G_k(w_k, d_k)$ subject to the J calibration constraints. Let λ be a vector of lagrange multipliers. It follows that

$$g_k(w_k, d_k - x_k'\lambda) = 0. \tag{1}$$

In what follows, it is useful to write this as

$$w_k = d_k F_k(x_k'\lambda). \tag{2}$$

It is informative to examine the minimization using G=$(w_k$-$d_k)^2/d_k$ which is the linear regression or unrestricted modified minimum chi-square method. In this situation, Equation (2) implies that

$$w_k = d_k(1 + x_k'\lambda) \tag{3}$$

where

$$\lambda = (\sum_{k=1}^{n} d_k q_k x_k x_k'^{-1})(t_x - \hat{t_{x\pi}} \tag{4}$$

The regression estmator can be written as:

$$\hat{t_{yreg}} = \hat{t_{y\pi}} + (t_x - \hat{t_{x\pi}})'\hat{B} \tag{5}$$

where

$$\hat{B} = (\sum_{k=1}^{n} d_k x_k x_k'^{-1}) \sum_{i=1}^{n} d_k x_k y_k \tag{6}$$

The variance of the generalized regression estimator is

$$\sum_k \sum_l (\pi_{kl} - \pi_k \pi_l)\pi_{kl}^{-1}(e_k d_k)(e_l d_l) \tag{7}$$

where $e_k = y_k - x_k'B$. This quantity can be estimated using $\hat{e_k} = y_k - x_k'\hat{B}$ with w_k replacing d_k. D-S show that all of the calibration estimators corrsponding to different distance measures have the same first-order asymptotic distribution, in the sense that $N^{-1}(\hat{t_{yw}} - \hat{t_{yreg}})$ is $O_p(n^{-1})$.

The formula in (5) provides some intuition concerning the source of increased variability associated with raking. $\hat{B}$ is the vector of design-based regression coefficients, and cacluating $\hat{B}$ requirs inverting the X-matrix. As targets are added, X eventually becomes more and more ill-conditioned, inflating the varaicne $(\hat{b})$ and the regresison estimator of the total.

3 Selecting Targets

Partition the X matrix associated with a set of targets into two subsets (X_1, X_2). We consider the situation where we would like to evaluate the raked solution using X_1 only, with the solution using both X_1 and X_2. Perhaps the simplest approach is to consider the design-based regression of Y on X and to test the null hypothesis B_2=0.

consider two procedures for selecting raking targets

4 Examples

We consider data from two market research surveys. Both surveys investigated business purchasing behavior. Samples were drawn from a national list of business locations. Both samples followed simple stratified designs, with primary stratifictaion by industry group and number of employees. Locations were sampled with equal probabilities within primary strata. Base weights to be used as starting values in the raking algorithms include a simple within-stratum nonresponse adjustment. In the first survey, weights were calibrated to obtain benchmark spending targets for the industry under consideration. In the second survey, weights were calibrated to obtain the spending targets for three product categories, and to match benchmark totals for geographic representation in 4 census regions.

Table 1: Survey 1: Distribution of Weights, Alternative Methods

Method	Mean	Max	Std. Dev.
1	12,950	137,948	25,211
2	12,950	110,841	23,279
3	12,950	121,445	24,123

Tables 1 and 2 present summary statistics for the distribution of the weights for Methods 1-3. Note that the mean weight is identical across the methods in each table, but that the extreme values and standard deviation vary substantially across the mthods. Tables 1-2 clearly demonstrate that the different algorithms produce very different results. Overall vari-

Table 2: Survey 2: Distribution of Weights, Alternative Methods

Method	Mean	Max	Std. Dev.
1	123	1923	261
2	123	1108	216
3	123	987	205

ability of the raked weights tends to lower for Methods 2 and 3, but there appears to be no clear clear ranking between Methods 2 and 3 across studies.

We also compare the second order asymptotic variance of the estimates, expressed relative to the variance of the generalized regression estimator, for several survey items. For the first study, we examined a variable concerning future purchase intentions. For this item, the variances for Methods 2 (96%) and 3 (94%) were lower in each case. For the second survey, we examined a general purchase propensity measured on a 100 point scale. For this survey, estmated variances were lower for Method 2(92%) and Method 3 (96%).

5 Conclusions

This paper compared the performance of various raking/calibration algorithms by exmainig the second order asymptotic distribution of the differetn estimates. Second order approximations were found to provide some assistance in choosing between the various methods. Future research will broaden the set of raking algorithms examined in this study, and will examine (through simulation) the actual finite sample performacne of

6 References

Chambers, R. L., Skinner, C. J., and Wang, S. (1999) Intelligent calibration, *Bulletin of the International Statistical Institute*, 58(2), pp 321-324.

Deville, J. C., and Sarndal, C. E. (1992) Calibration estimators in survey sampling, *Journal of the American Statistcial Association*, 87, pp. 376-382.

Deville, J. C., Sarndal, C. E., and Sautory, O. (1993) Generalized raking procedures in survey sampling, *Journal of the American Statistical Association 88, pp. 1013-1020.*

Rothenberg, T., (1984) Approximating the distributions of econometric estimators and test statistics, in Handbook of Econometrics, II*, Grilliches, Z., and intrilligator, M., (eds.), pp. 882-911.*

Singh, A.C., and C.A. Mohl (1996) Undertsanding calibration estimators in survey sampling, Survey Methodology 22*, pp. 107-115.*

Stukel, D. M., and Boyer, R. (1992) Calibration estimation: an application to the CAnadian Labour Force Survey, Methodology Branch Working Paper, SSMD 92-009E, Statistics Canada