# A RIDGE-SHRINKAGE METHOD FOR RANGE-RESTRICTED WEIGHT CALIBRATION IN SURVEY SAMPLING

J.N.K. Rao and A.C. Singh
Carleton University and Statistics Canada, Ottawa
A.C. Singh, Methodology Research Advisor, Statistics Canada, 16-A R.H.Coats, Ottawa, Ontario K1A 0T6

## ABSTRACT

The generalized regression (GR) method is often used to adjust sampling weights to satisfy benchmark constraints (BC), but some of the calibrated weights may not satisfy range restrictions (RR). BC are needed in the interest of efficiency due to correlated auxiliary information and also to make estimates internally consistent with published population totals, while RR are needed to avoid extreme weights which may render domain estimates inefficient. To address this problem, several iterative methods that attempt to meet both RR and BC have been proposed in the literature. However, for given RR and BC, and a specified number of iterations, these methods may not converge even if RR are mild, *e.g.*, restrictions to only nonnegative weights. This is likely to happen in practice if there is a high discrepancy between a BC and its estimate due to sample being not large enough, or if there are too many BC or multicollinear auxiliary variables implying instability in the estimated regression coefficients. A third possibility of course is if RR are too tight. For given RR, a natural and practical way out is to relax a few BC, *i.e.*, make them nonbinding within specified tolerances, while keeping other BC as binding *(i.e.*, with zero tolerance). An important requirement while relaxing BC is that for given tolerance levels, the calibration method should ensure asymptotic design consistency (ADC) like GR. Note however that since the extreme weight problem is due to sample being not large enough, asymptotically the problem disappears. This implies a possible loss in efficiency by making BC nonbinding. Therfore, in the interest of efficiency, the tolerances should be specified adaptively so that asymptotically they tend to zero implying, in turn, that the calibrated weights tend to GR weights. The RR themselves can be further relaxed if necessary to get lower tolerance levels. In this article, for complex surveys, we consider first Rao's (1992) modification of the ridge-regression method of Bardsley and Chambers (1984) so that the resulting estimator has the ADC property in spite of the presence of the ridge matrix which makes BC nonbinding. We then establish an important relation between the ridge (or inverse cost) matrix and the matrix of specified tolerances, and show that the above method can be adapted to meet BC up to specified tolerances while maintaining ADC. This method like GR is noniterative, and can be easily implemented. However, in spite of relaxing BC, the method may not meet RR. We, therefore, propose an iterative method termed ridge-shrinkage, which generalizes the above ridge-regression method in a manner analogous to the generalization of the usual GR by calibration methods to meet RR. The proposed method is designed to force convergence for a given number of iterations by using a built-in tolerance specification procedure to relax BC while satisfying RR and maintaining design consistency. Numerical results on the relative performance of several related methods are also presented.

**Key Words**: Asymptotic design consistency; Binding and nonbinding benchmark constraints; Range restrictions; Ridge regression; Shrinkage-minimization.

## 1. INTRODUCTION

In survey sampling, perfect auxiliary information in the form of benchmark constraints (BC) is commonly incorporated by means of generalized regression (GR) or raking methods of estimation. Use of BC is not only desirable from the efficiency perspective, but also due to the need to make estimates internally consistent with published population/domain totals. It is also known that regression or raking methods may lead to extreme calibrated weights, which may render domain estimates highly unreliable. To get around this problem, several methods are proposed in the literature to adjust sampling weights to meet BC while satisfying certain range restrictions (RR); see Huang and Fuller (1978), Deville and Särndal (1992) and Singh (1993). For an expository review, see Singh and Mohl (1996).

All these methods are iterative in nature and may not lead to a solution in a fixed number of iterations. Note that at any iteration, we can always ensure that RR are met by shrinking the weights to the boundaries of RR. (Some methods automatically satisfy RR after each iteration and iterations are continued to meet BC, while others automatically satisfy BC after each iteration and iterations are continued to meet RR.) Now if at the maximum number of iterations allowed, BC are not met but RR are, the process can be terminated and the resulting estimator will have the ADC property. However, this approach is seriously deficient in that there is no control on the extent of discrepancy in meeting BC. The nonconvergence problem is likely to happen in practice if RR are too tight, or if there is a high discrepancy between a BC and its estimate due to sample being not large enough, or if there are too many BC or multicollinear auxiliary variables implying instability in the estimated regression coefficients. One way out might be to drop some BC as suggested by Bankier et al. (1992) who encountered the problem of negative weights in weight calibration for census 2B sample because of a large number of BC at the enumeration and weighting area levels. This approach may seem somewhat drastic in that most BC are treated as binding (*i.e.*, with zero tolerance) while some as nonbinding in the extreme sense (*i.e.*, with infinite tolerance). A practically appealing alternative might be to allow most BC to be nonbinding with possibly varying tolerance except for a few binding ones based on subject matter considerations.

A solution to the above problem may be motivated by drawing analogy with the problem of instability in regression estimates due to insufficient sample in the presence of multicollinearity or too many auxiliary variables in regression modelling, $y = \mu_y + \beta'(\boldsymbol{x} - \boldsymbol{\mu}_x) + \varepsilon$. A standard solution in classical statistics is to use ridge regression in which the least squares criterion is modified by a penalty function involving a cost matrix; the inverse cost matrix appears as the ridge matrix in the ridge regression estimate. It is interesting to note that although the resulting estimate of the regression coefficient becomes biased (but stable), the regression estimate of the unconditional mean $\mu_y$ of the dependent variable ($y$) remains approximately unbiased and consistent. However, unlike the usual regression estimate, it does not reproduce perfectly the mean $\boldsymbol{\mu}_x$ of the auxiliary vector ($\boldsymbol{x}$) when $y$ is replaced by $\boldsymbol{x}$. This would tend to reduce the estimator's efficiency, had true $\beta$ been known. This, however, is compensated by using a more stable estimate of $\beta$ when it is unknown.

In survey statistics, the problem of instability in GR (GR is defined by a difference-type estimator where the difference coefficient corresponds to the regression coefficient $\beta$) is often due to the large dimension of $\beta$ in view of many predictors needed to satisfy multipurpose needs of the user For this case, an important and interesting model-based method using ridge-regression was proposed by Bardsley and Chambers (1984) to obtain a set of adjusted weights for the estimator. They showed that in order to satisfy RR, extreme weights can be avoided by choosing the parameter in the ridge matrix appropriately. However, no guidance was provided on choosing the ridge matrix to meet a desired tolerance on benchmark controls corresponding to the auxiliary variables. Moreover, they did not consider the use of survey weights for a design-based approach. Rao (1992) modified the above method to take account of the survey weights to ensure the important property of ADC. However, the problem of a suitable choice of the ridge matrix to meet specified tolerances on BC for given RR was not addressed.

In this paper, we first establish a relation between ridge (or inverse cost) and tolerance matrices so that for a given set of upper bounds on tolerances, the corresponding cost matrix for ridge-regression can be specified. In particular, zero tolerance would correspond to infinite cost (*i.e.*, zero inverse cost), and infinite tolerance would lead to zero cost in the limit. This can be used to modify the Bardsley-Chambers method so that BC are met within tolerances, although in this article we have not pursued model-based methods. The above result on the relation between ridge and tolerance matrices enables us to modify several existing (iterative) calibration methods via the ridge-regression idea to meet RR while relaxing those BC which are not deemed binding. For given RR, the tolerance levels are chosen adaptively in order to relax BC only when necessary in the interest of efficiency and internal consistency. We can also relax RR if it is desired to get lower tolerances. The proposed method, termed ridge-shrinkage, starts with GR weights (corresponding to zero tolerance) which are then shrunk to meet RR. Next, at each cycle of iterations, tolerance levels for discrepant BC (up to the specified tolerance) are raised in increments, and thus are defined adaptively. Assuming that asymptotically GR weights meet RR, tolerance levels so defined tend to zero, and the ridge-shrinkage method becomes asymptotically equivalent to GR. Therefore, the asymptotic variance of the proposed estimator can be estimated using the GR formula with ridge-shrinkage residuals replacing the GR residuals.

Section 2 provides a brief review of existing methods for the cases of binding and nonbinding BC. The proposed method is described in Section 3 in which its asymptotic equivalence to GR is outlined. A numerical example based on Statistics Canada's FAMEX data is presented in Section 4, and finally some remarks in Section 5.

## 2. EXISTING CALIBRATION METHODS

For a sample of size $n$, let $h_k$ denote the initial design weight, and $c_k$ the final calibrated weight for the $k^{th}$ sample unit, $1 \le k \le n$. Let the RR be given by lower and upper bounds $[L, U]$ such that $Lh_k \le c_k \le U h_k$ for all $k$. The BC are given by $\boldsymbol{X}'\boldsymbol{c} = \tau_x$, where $\boldsymbol{X}$ is the $n \times p$ matrix of values of the $p$-auxiliary variables, and $\tau_x$ is the $p$-vector of control totals. Assume without loss of generality that $\tau_{xi} \ge 0$ for $1 \le i \le p$. The tolerance matrix $\Delta$ for BC is $\Delta = \text{diag}(\delta_1, ..., \delta_p)$, where $\delta_i$ is the tolerance for the $i$th BC and is defined as

$$| \boldsymbol{x}_i' \boldsymbol{c} - \tau_{xi} | \le \delta_i \tau_{xi}, \ 1 \le i \le p. \tag{2.1}$$

Note that $\delta_i > 0$ implies a nonbinding constraint and $\delta_i = 0$ corresponds to a binding constraint. The limiting case of $\delta_i = \infty$ implies discarding of constraint. Also, note that if the control total $\tau_{xi} = 0$, then $\delta_i$ can be defined as $|x_i' c| \le \delta_i N$ where $N$ is the population size ascertained from external sources.

## 2.1 Nonbinding Case

We consider a modified version of the ridge regression of Bardsley and Chambers (1984) as proposed by Rao (1992) for survey data. Suppose the inverse cost matrix is $\Lambda = \text{diag}(\lambda_1, ...., \lambda_p)$. Then consider minimization of the objective function in the form of a penalized least squares criterion:

$$\Delta^{rid}(c, h) = (c - h)'\Gamma^{-1}(c - h) + (X'c - \tau_x)'\Lambda^{-1}(X'c - \tau_x) \quad (2.2)$$

where $\Gamma = \text{diag}(h)$. The solution is given by

$$c^{rid} = h + \Gamma X (X'\Gamma X + \Lambda)^{-1} (\tau_x - X'h) \quad (2.3)$$

The above solution minimizes $\Delta^{rid}$ provided the second derivative, $\Gamma^{-1} + X\Lambda^{-1}X'$, is nonnegative definite. Note that technically the cost matrix $\Lambda^{-1}$ need not be nonnegative for the above condition to be satisfied, although then it loses its usual interpretability.

The estimator $\hat{\tau}_y^{rid}$ of the population total $\tau_y$ for the study variable $y$ is

$$\hat{\tau}_y^{rid} := y'c^{rid} = y'h + y'\Gamma X (X'\Gamma X + \Lambda)^{-1} (\tau_x - X'h) \quad (2.4a)$$

$$= (1 - \alpha)\hat{\tau}_y^{\text{HT}} + \alpha\,\hat{\tau}_y^{\text{GR}}, \quad (2.4b)$$

where $\hat{\tau}_y^{\text{HT}}$ is the Horvitz-Thompson (HT) estimator, $\hat{\tau}_y^{\text{GR}}$ is the GR-estimator, and $\alpha$ is the shrinkage coefficient which shrinks GR towards the HT-estimator. The corresponding expressions are

$$\hat{\tau}_y^{\text{HT}} = y'h,\ \hat{\tau}_y^{\text{GR}} = y'h + y'\Gamma X(X'\Gamma X)^{-1}(\tau_x - X'h),$$
$$\alpha = [y'\Gamma X(X'\Gamma X + \Lambda)^{-1}(\tau_x - x'h)] \times [y'\Gamma X(X'\Gamma X)^{-1}(\tau_x - X'h)]^{-1} \quad (2.5)$$

Thus, the ridge-regression estimator $\hat{\tau}_y^{rid}$ is a linear combination of HT and GR estimators. If $\lambda_i \to 0$, which implies that all the BC become binding, then $\hat{\tau}_y^{rid}$ tends to GR as expected. If $\lambda_i \to \infty$, implying that BC can be nonbinding with unlimited tolerance, then $\hat{\tau}_y^{rid}$ tends to HT as expected.

Moreover, we have for the ridge weights $c^{rid}$,

$$X'c^{rid} = X'h + X'\Gamma X(X'\Gamma X + \Lambda)^{-1}(\tau_x - X'h)$$
$$= \tau_x - \Lambda(X'\Gamma X + \Lambda)^{-1}(\tau_x - X'h) \quad (2.6)$$

It follows that if $\lambda_i = 0$, the corresponding BC $\tau_{xi}$ is exactly satisfied by $c^{rid}$. For $\lambda_i > 0$, the $\tau_{xi}$ is not treated as perfect auxiliary information. As $\lambda_i \to \infty$, the corresponding control total $\tau_{xi}$ is automatically discarded. This can be seen as follows. Assuming all $\lambda_j > 0$, we can write $c^{rid}$ as

$$c^{rid} = h + \Gamma[I - \Lambda^{-1}X'(\Gamma^{-1} + X\Lambda^{-1}X')^{-1}] \times X\Lambda^{-1}(\tau_x - X'h) \quad (2.7)$$

Now, denoting $X, \Lambda, \tau_x$ without the $i$th row as $X_{(i)}, \Lambda_{(i)}$ and $\tau_{x(i)}$, and letting $\lambda_i \to \infty$, (2.7) reduces to

$$c^{rid} = h + \Gamma[I - \Lambda_{(i)}^{-1}X'_{(i)}(\Gamma^{-1} + X_{(i)}\Lambda_{(i)}^{-1}X'_{(i)})^{-1}] \times X_{(i)}\Lambda_{(i)}^{-1}(\tau_{x(i)} - X_{(i)}h) \quad (2.8)$$
$$= h + \Gamma X_{(i)}(X'_{(i)}\Gamma X_{(i)} + \Lambda_{(i)})^{-1}(\tau_{x(i)} - X_{(i)}h)$$

which proves the result. If some $\lambda_j = 0$, the above proof still goes though if we initially set $\lambda_j$ equal to a small positive value $\epsilon_j$ and then take the limit in (2.8) as $\epsilon_j \to 0$.

The ADC of $\hat{\tau}_y^{rid}$ follows easily by arguments similar to those used for $\hat{\tau}_y^{GR}$ under the asymptotic framework of Isaki and Fuller (1982). Specifically, let $\hat{\beta}_{rid}$ denote $(X'\Gamma X + \Lambda)^{-1}X'\Gamma y$ and assume $\lambda_i$ random and chosen adaptively such that $\lambda_i / N = O_p(1)$, then $\hat{\beta}_{rid}$ tends in probability to a limit, $\beta_{rid}$ ,say, and

$$\hat{\tau}_y^{rid} \doteq y'h + \beta'_{rid}(\tau_x - X'h) \quad (2.9)$$

and the RHS is ADC for $\tau_y$ under the assumption of ADC of $y'h$ and $X'h$. We remark that the only difference between $\hat{\tau}_y^{GR}$ and $\hat{\tau}_y^{rid}$ is that $\hat{\beta}_{GR}$ is replaced by $\hat{\beta}_{rid}$. The predictors $\tau_x - X'h$ are unaltered even though $\tau_x$ are not perfectly satisfied. This is the reason why ADC of $\hat{\tau}_y^{GR}$ is maintained in ridge-regression. Note that if $\lambda_i / N = o_p(1)$, then the ridge regression estimator will be equivalent to the GR-estimator.

Clearly, the behaviour of $c^{rid}$ depend on $\Lambda$. In particular, for $\lambda_i$ sufficiently large, all $c$-weights should behave well, *i.e.*, should be free from extremes. Bardsley and Chambers (1984) consider $\Lambda = \lambda I$, and use a graphical tool (ridge trace as $\lambda$ varies) to find a suitable value of $\lambda$ so that $c$-weights behave well. Thus, $\lambda$ is chosen adaptively. A more satisfactory solution would be to set tolerances $\delta_i$'s on BC and then find the corresponding $\Lambda$ that meets these tolerances. A method to achieve this is presented in Section 3.

Finally, we note that Bankier et al.'s (1992) method of discarding some BC while the remaining BC are perfectly satisfied, is easily seen as a special case of the ridge-regression method when some $\lambda_i$'s tend to $\infty$ while others are set at 0.

## 2.2 Binding Case

Suppose all the BC are binding so that $\delta_i = 0$ for all $i$. As mentioned in the introduction, there exist several iterative methods whose aim is to meet BC for a given set of RR. We will briefly describe only three methods which consist of GR-like steps in iteration. The proposed ridge method to meet RR and BC can be applied to any of these three methods, although it is the first method which is considered in detail here.

### 2.2.1 Shrinkage-Minimization (SM)

This method was proposed by Singh (1993); see also Singh and Mohl (1996). Each iteration consists exactly of a GR-step for a suitable chi-square distance. Let $c^{(\nu)}$ be the final weights obtained at iteration $\nu$. These weights satisfy BC by construction. If they satisfy RR, we stop. If not, they are shrunk to $c^{(\nu)*}$ to meet RR. Then treating $c^{(\nu)*}$ as initial weights for the next iteration, we minimize the chi-square distance.

$$\Delta^{SM}_{\nu+1}(c, c^{(\nu)*}) = \sum\nolimits_{k \in s} (c_k - c_k^{(\nu)*})^2 / c_k^{(\nu)*}, \qquad (2.10)$$

subject to BC. Since each iteration is like GR (except that the distance function varies from iteration is iteration), it follows easily from section 2.1 how this can be converted into a ridge-regression to allow nonbinding BC. In fact, this is what is done in the proposed ridge-shrinkage method described in Section 3. The above minimization step at iteration $(\nu + 1)$ leads to weights $c^{(\nu+1)}$ given by

$$c^{\nu+1} = c^{(\nu)*} + \Gamma_\nu X (X' \Gamma_\nu X)^{-1} (\tau_x - X' c^{(\nu)*}), \qquad (2.11)$$

where $\Gamma_\nu = \mathrm{diag}(c^{(\nu)*})$. These weights satisfy BC but RR may not be satisfied. If RR are satisfied, we stop. If not, then we perform the shrinkage step for iteration $\nu + 2$ to get the initial weights $c^{(\nu+1)*}$. It is defined as follows. Let $[L, U], L<1<U$ denote the lower and upper bounds specified by RR which the calibrated weights $c$ must satisfy, *i.e.*,

$$L h_k \le c_k \le U h_k,\ 1 \le k \le n. \qquad (2.12)$$

Now, to speed convergence, the weights $c^{(\nu+1)}$ are shrunk more than necessary. For this purpose, two parameters $\alpha$ and $\eta$ are defined, $0 < \alpha < \eta < 1$ (*e.g.*, $\alpha = 2/3$ and $\eta = 9/10$.) Let $L' = \alpha L + (1 - \alpha) 1$, $U' = \alpha U + (1 - \alpha) 1$, and $L'' = \eta L + (1 - \eta) 1$, $U'' = \eta U + (1 - \eta) 1$. Then, we shrink $c^{(\nu+1)}$ weights that are outside the interval $[L h_k, U h_k]$ and also those which are inside but near the boundary, to points further inside the interval. Specially,

$$c_k^{(\nu+1)*} = \begin{cases} L' h_k & \text{if } c_k^{(\nu+1)} \le L'' h_k; \\ U' h_k & \text{if } c_k^{(\nu+1)} \ge U'' h_k; \\ c_k^{(\nu+1)} & \text{otherwise.} \end{cases} \qquad (2.13)$$

The above shrinkage step of iteration $(\nu + 2)$ is followed by the minimization step with distance function $\Delta^{SM}_{\nu+2}(c, c^{(\nu+1)*})$ analogous to (2.10), to get $c^{(\nu+2)}$. If RR are satisfied, we stop; else iterations are continued until the maximum number $\nu_{max}$ of iterations is reached. Clearly, there may not be convergence within $\nu_{max}$ of iterations if the RR are too tight, or if there are too many BC or if there is multicollinearity in the variables defining BC.

Using suitable regularity conditions, the SM-estimator, $\hat{\tau}_y^{SM} = \sum_{k \in s} y_k c_k^{SM}$, can be shown to be asymptotically equivalent to the GR-estimator. This result is analogous to the Deville-Särndal's (1992) result on the asymptotic equivalence of a family of calibration estimators to GR. It follows that $\hat{\tau}_y^{SM}$ is ADC and its asymptotic variance can be estimated from the familiar expression for GR.

### 2.2.2. Modified Huang-Fuller Method (or SMCS)

This is a slightly modified version of the method of Huang and Fuller (1978), and was termed as the Scaled Modified Chi-square (SMCS) method in Singh (1993). In SMCS, at iteration $(\nu + 1)$, a chi-square-type distance function is minimized subject to BC. It is given by

$$\Delta^{\mathrm{SMCS}}_{\nu+1}(c, h) = \sum\nolimits_{k \in s} (c_k - h_k)^2 / q_k^{[\nu]} h_k, \qquad (2.14)$$

where $q_k^{[\nu]}$ is a scaling factor designed such that the $h$-weights for those units that tend to disobey RR are adjusted only a little. This is accomplished by making $q_k^{[\nu]}$ smaller for the next iteration. Note that, unlike SM, at each iteration $\nu$ the SMCS-estimator is not like the usual GR because of the scaling factor. However, it does satisfy BC at each iteration, and iterations are continued until RR are met or $\nu \geq \nu_{max}$. The form of $c$-weights at iteration $(\nu + 1)$ is given by

$$c^{(\nu+1)} = h + \Gamma_\nu X (X' \Gamma_\nu X)^{-1} (\tau_x - X' h), \qquad (2.15)$$

where $\Gamma_\nu = \text{diag}(q_k^{[\nu]} h_k,\ 1 \leq k \leq n)$. Again, a solution may not exist for a specified $\nu_{max}$.

As before, the estimator $\tau_y^{SMCS}$ from the SMCS method is asymptotically equivalent to the GR-estimator.

### 2.2.3. Truncated Linear Method (or MCS-$r$)

This method is due to Deville and Sarndal (1992), and was termed as the restricted Modified Chi-square (MCS-$r$) method in Singh (1993). Unlike the previous two methods, here distance function does not change from iteration to iteration; and at each iteration RR are satisfied, but iterations are continued to meet BC. In MCS-$r$, the distance function to be minimized subject to BC is given by

$$\Delta^{MCS\text{-}r}(c,h) = \sum_{k \in s} (c_k - h_k)^2 / h_k \quad \text{if } Lh_k \leq c_k \leq Uh_k;\ \infty \text{ otherwise.} \qquad (2.16)$$

With the initial weights, $c^{(0)} = h, \Gamma_0 = \text{diag}(h)$ and letting $\Gamma_\nu = \text{diag}(h_k^{(\nu)})$ where $h_k^{(\nu)} = h_k$ if $Lh_k \leq c_k^{(\nu)} \leq Uh_k$; 0 otherwise, the $c$-weights at iteration $(\nu + 1)$ are

$$c^{(\nu+1)} = c^{(\nu)} + \Gamma_0 X (X' \Gamma_\nu X)^{-1} (\tau_x - x' c^{(\nu)}), \qquad (2.17)$$

provided $c_k^{(\nu+1)}$ is inside $[Lh_k, Uh_k]$. If outside, it is truncated at the left or the right boundary as the case may be.

Note that the expression (2.17) for $c^{(\nu+1)}$-weights is somewhat similar to that for GR-weights except for the term $\Gamma_0$ and the use of truncation. However, as was shown by Deville and Särndal, the MCS-$r$ estimator is asymptotically equivalent to the GR-estimator.

## 3. THE PROPOSED METHOD OF RIDGE-SHRINKAGE

As mentioned in Section 2.2.1, the proposed method combines in a fairly straightforward way the idea of ridge-regression with each of the iterations of the SM method because each SM iteration is simply a version of GR. Before we describe the proposed ridge-shrinkage (RS) method, we need to establish a link between the tolerance matrix $\Delta = \text{diag}(\delta_i)$ and the inverse cost matrix $\Lambda = \text{diag}(\lambda_i)$ which will be used at each iteration of the RS method.

### 3.1 Link between tolerance and cost matrices

In the ridge approach, it is probably easier to specify the tolerance matrix $\Delta$ in practice than the inverse cost matrix $\Lambda$. Now, it follows from (2.6) that for the $(\nu+1)$st iteration of RS,

$$\Lambda (X' \Gamma_\nu X + \Lambda)^{-1} (X' c^{(\nu)*} - \tau_x) = X' c^{(\nu+1)} - \tau_x \qquad (3.1)$$

For each $i$, $1 \leq i \leq p$, we want the $i$th element of the RHS of (3.1) to be less than or equal to $\delta_i \tau_{xi}$ in absolute value. To find appropriate $\Lambda_\nu$ for the $(\nu+1)^{st}$ iteration, we solve for $\Lambda_\nu$ from (3.1) by setting the RHS equal to the boundary values $\delta_i \tau_{xi}$ with appropriate signs. In other words, we set the RHS equal to $\nabla_\nu \tau_x$ where

$$\nabla_\nu = \text{diag}\{sgn(x_i' c^{(\nu)*} - \tau_{xi}) \delta_i,\ 1 \leq i \leq p\} \qquad (3.2)$$

In practice, in the interest of convergence, it would be better to modify $\nabla_\nu$ somewhat; see section 4 for details. We now have

$$\Lambda_\nu (X' \Gamma_\nu X + \Lambda_\nu)^{-1} (X' c^{(\nu)*} - \tau_x) = \nabla_\nu \tau_x$$

or

$$(X' c^{(\nu)*} - \tau_x) = (X' \Gamma_\nu X \Lambda_\nu^{-1} + I) \nabla_\nu \tau_x$$

or

$$X' c^{(\nu)*} - (I + \nabla_\nu) \tau_x = X' \Gamma_\nu X \Lambda_\nu^{-1} \nabla_\nu \tau_x$$

which implies that

$$\Lambda_\nu (X' \Gamma_\nu X)^{-1} (X' c^{(\nu)*} - (I + \nabla_\nu) \tau_x) = \nabla_\nu \tau_x . \qquad (3.3)$$

So, $\Lambda_\nu$ being diagonal, can be obtained by element-wise division of the $p$-vector on the RHS by the $p$-vector on the LHS of (3.3). Note that $\Lambda_\nu$ has zero on its diagonal when $\delta_i = 0$. The above method does not ensure nonnegative $\lambda_i$. This is not essential in view of the comment below (2.3). Also note that if the choice of $\delta_i$ depends on the sample (thus rendering it random) then (3.3) implies that the diagonal elements of $\Lambda_\nu / N$ are $O_p(n^{1/2} N^{-1} \delta_i \tau_{xi})$. This, in turn, implies that if $n^{1/2} N^{-1} \delta_i \tau_{xi} = o_p(1)$, the ridge regression estimator at the $\nu th$ iteration will be asymptotically equivalent to the GR-estimator in view of the comment below (2.9).

### 3.2 The RS Method

Similar to SM, each iteration of RS consists of two steps: the ridge step and the shrinkage step. It consists of cycles of iterations, the $q$th cycle corresponds to a given tolerance $\Delta_q$. For each cycle $q$, there is a prescribed maximum number of iterations. For the initial cycle $q = 0$, the usual SM is performed except for the reverse ordering of steps, *i.e.* first the minimization step and then the shrinkage step. (Note that the minimization step can be viewed as the ridge step with all $\lambda_i = 0$). The order of steps is reversed because of the introduction of tolerance on BC. Thus, after each iteration, RR are necessarily met. If the BC are satisfied within the tolerance levels (for checking this, it is better not to shrink the weights more than necessary, i.e., truncate outlying weights to the boundaries only), then we stop the iterations. Else, iterations are continued until $\nu_{max}$. Denote the final SM-weight after shrinkage as $c^{SM*}$. If at this point BC are not met within tolerance, then we start the next cycle $q = 1$ with $c^{SM*}$ as the initial weights. For this cycle, the tolerance matrix $\Delta_q$ with $q = 0$ is used to specify $\Lambda_{q=0}$ using equation (3.3). Each iteration $\nu$ within this cycle consists of two steps:

**Step I** (Ridge): Do ridge-regression on $c^{(\nu)*}$ with the inverse cost matrix $\Lambda_q$ to obtain $c^{\nu+1}$ from formula (2.3). Now, all BC are met within the prescribed tolerance. If RR are met, stop; else perform the next step.

**Step II** (Shrinkage): First truncate outlying weights to the boudaries only so that they just meet RR. Stop if BC are met within tolerance. If not, shrink $c^{(\nu+1)}$ to $c^{(\nu+1)*}$, and then repeat steps I and II until convergence or $\nu \geq \nu_{max}$.

Similarly cycle 2 is performed if there is no convergence after cycle 1. Note that each cycle is started with the same $c^{SM*}$-weights for initialization. However, the tolerance levels are revised adaptively (in increments such as 1%) so that BC showing higher discrepancy are assigned higher tolerances. With $\Delta_{q=1}$ so chosen after cycle 1, the inverse cost matrix $\Lambda_{q=1}$ is obtained and then the iterations for cycle 2 are conducted as in cycle 1. This process is continued until convergence (within revised tolerances) after each cycle. Note that in the absence of convergence, the process can be terminated after reaching $\nu_{max}$ of iterations in the maximum allowed number of cycles $q_{max}$. At this point, RR are of course met, but BC can be deemed as satisfied with tolerances suitably increased. Note that from (2.5), the g-weights ($g$-weights are simply defined as $c_k/h_k$, $1 \leq k \leq n$) for GR can be shown to be $1 + O_p(n^{-1/2})$ uniformly in $k$, and therefore, GR-weights satisfy RR asymptotically. This implies that as $n \to \infty$, with high probability RS will converge after the initial cycle itself. Thus $\delta_i = 0$ with high probability from which it follows that $\Lambda_q / N$ tends to zero in probability using the comment below (3.3).

One can also define ridge-versions of the other two methods, SMCS and MCS-$r$ by introducing the inverse cost matrix $\Lambda_q$ in (2.15) for ridge-SMCS and in (2.17) for ridge-MCS-$r$. The specification of $\Lambda_q$ from $\Delta_q$ is quite similar to (3.3) for ridge-SMCS, but somewhat different for ridge MCS-r; see Section 4.

### 3.3 Asymptotic Properties

The RS-estimator $\hat{\tau}_y^{RS}$ of $\tau_y$ is asymptotically equivalent to $\hat{\tau}_y^{GR}$ if at the final cycle $q_0$ of iterations, $N^{-1} \lambda_i^{(q_0)} \to_p 0$ for all i. A sufficient condition for this to hold is that tolerances $\delta_i$ be initialized at 0, and be revised adaptively as described above. The proof for the asymptotic equivalence is outlined below. First note that for given $\Lambda = \Lambda_{q0}$, the RS-estimator $\hat{\tau}_y^{RS}$ is simply an iterative modification of $\hat{\tau}_y^{rid}$ to meet RR which is analogous to the modification of $\hat{\tau}_y^{GR}$ by $\hat{\tau}_y^{SM}$. Thus it can be shown that $\hat{\tau}_y^{RS} \doteq \hat{\tau}_y^{rid}$ by parallel arguments used for showing $\hat{\tau}_y^{SM} \doteq \hat{\tau}_y^{GR}$. Now it remains to show that $\hat{\tau}_y^{rid} \doteq \hat{\tau}_y^{GR}$ if $N^{-1} \lambda_i^{(q_0)} \to_p 0$, $1 \leq i \leq p$. This follows easily from the expression

$$\hat{\tau}_y^{rid} = y'h + (y'\Gamma X)(X'\Gamma X)^{-1}(\tau_x - X'h) + (y'\Gamma X)[(X'\Gamma X + \Lambda_{q0})^{-1} - (X'\Gamma X)^{-1}] \times (\tau_x - X'h) \tag{3.4}$$

and the fact that the last term in the RHS of (3.4) is of smaller order because $N^{-1} \lambda_i^{(q_0)} \to_p 0$.

In view of the above asymptotic equivalence, the asymptotic variance of $\hat{\tau}_y^{RS}$ can be obtained from that for GR, using the RS $g$-weights rather than the GR $g$-weights. One can then construct confidence intervals for $\tau_y$ using a finite population central limit theorem for $\hat{\tau}_y^{RS}$.

For ridge-versions of the other two methods SMCS and MCS-$r$, above asymptotic properties can be similarly obtained.

## 4. APPLICATION TO THE FAMEX DATA

The numerical results are based on the work of Yannick Janneau completed for a M.Sc directed studies course in 1996 at Carleton University. He extended the numerical comparison of Singh and Mohl (1996), based on Statistics Canada's family expenditure (FAMEX) survey data, to include ridge methods. All the three methods: RS, ridge-SMCS, and ridge MCS-r, were compared although full details for ridge SMCS and MCS-$r$ are not given here. The tolerances were set either at 0 or at a common value of $\delta>0$. This proved to be convenient in practice. The matrix $\nabla_v$ of (3.2) was modified by replacing $\delta_i$ by $\psi a_i \delta_i$ where $a_i = 0$ if the discrepancy $(X_i' c^{(v)*} - \tau_{xi})$ is $\leq \psi \delta_i$, and 1 otherwise. The parameter $\psi$, $0 < \psi < 1$ makes tolerances conservative, and helps to speed up the convergence. We used $\psi = .9$ in the example. The indicator variable $a_i$ treats the $i$th control as binding in the ridge step of the $(v+1)$st iteration if the discrepancy at the $v$th iteration is within tolerance. This modification is again in a conservative sense, and helps to speed up the convergence. Details about description of the FAMEX data, choice of BC and RR, and behaviour of existing calibration methods are given in Singh and Mohl (1996).

Now, along the lines leading to the equation (3.3) for establishing the link between tolerance and the cost matrices, the corresponding equation for ridge SMCS can be obtained. It is very similar and given by

$$\Lambda_v (X'\Gamma_v X)^{-1} [X'h - (I + \nabla_v)\tau_x] = \nabla_v \tau_x, \qquad (4.1)$$

where $\Gamma_v$ is now $diag(q_k^{[v]} h_k, 1 \leq k \leq n)$. Note that $c^{(v)*}$ in the LHS of (3.3) is now replaced by $h$. However, for ridge MCS-$r$, the equation (3.3) changes somewhat, and can be obtained by using the same line of argument as

$$\begin{aligned} &\Lambda_v (X'\Gamma_0 X)^{-1} [X'c^{(v)} - (I + \nabla_v)\tau_x] \\ &= \nabla_v \tau_x + (X'(\Gamma_0 - \Gamma_v)X)(X'\Gamma_0 X)^{-1} \times \qquad (4.2) \\ &\qquad [X'c^{(v)} - (I + \nabla_v)\tau_x], \end{aligned}$$

where $\Gamma_v$ and $c^{(v)*}$ in the LHS of (3.3) are now replaced by $\Gamma_0$ and $c^{(v)}$ respectively, and an additional term is added on the RHS of (3.3). The vector $\lambda_v$ from each of (4.1) and (4.2) can be solved as before by dividing element-wise the $p$-vector on the RHS with the $p$-vector on the LHS.

For the sake of illustrating the ridge methods, the three methods were applied to the 1990 FAMEX data for the city of Regina. Since there were only a few BC, the RR bounds [L,U] were made quite tight so that none of the existing calibration methods converged . For L=.5, U=2, even after 100 iterations, the % discrepancy in respecting the four BC were 21.64, 16.94, 75.17, and19.61 for SM, 24.17, 18.73, 75.17, and 21.01 for SMCS, and 97.28, -21.09, -12.42, and 2.62 for MCS-r.

For L=.5, U=2, Table 1 shows the CV($g$) (coefficient of variation of $g$-weights) and percentage discrepancy in respecting BC. Here $\delta_{min}$ denotes the minimum tolerance required for a given ridge method so that all the BC are met within tolerance. It is seen that all the three methods behave quite similarly and the discrepancy in respecting BC can be considerably reduced in comparison to non-ridge methods. Table 2 shows the relative difference (RD) and relative precision (RP) in point estimates for four study variables. RD is defined as the ridge-calibration estimator minus the regression estimator divided by the regression estimator, while RP is the SE (regression estimator) divided by the SE (ridge-calibration estimator). The variances were computed using jackknifing, see Singh and Mohl (1996) for further explanation. It is seen that all the estimates give higher estimated variance as compared to the usual regression estimator. This is expected as explained below.

Observe that even if some weights are extreme (e.g., negative or very high), there may or may not be instability in the GR-estimator depending on the study variable. Now with respect to the variables studied, GR does not seem to have the instability problem because with loose bounds [.2, 5], it is known from Singh and Mohl (1996) that all the usual calibration methods converge fairly quickly with estimated variance similar to that of GR. For our example, the bounds are further tightened to [.5, 2] so that the calibration methods no longer converge for the sake of illustrating the ridge methods. Now since the ridge methods do not satisfy BC perfectly, ridge-calibration estimates, although asymptotically equivalent to GR, are expected to have higher variances for finite samples whenever GR is not unstable. There is likely to be a further loss in efficiency if we drop a BC (*i.e.* increase the tolerance to $\infty$) as an alternative approach to get rid of the problem of extreme weights. For example, one can perhaps drop the second BC and then attempt to satisfy perfectly the remaining three BC in this alternative approach. However, note that the RS method also satisfies almost perfectly these three BC and within 4% the second BC. With only three BC (i.e., when the second one out of four is dropped), all the three non-ridge methods converge in one iteration which implies that GR also satisfies RR and the three BC. In this case, the discrepancy with respect to the dropped BC is -13.77%, much higher in magnitude than the 4% tolerance required by the ridge method RS.

## 5. CONCLUDING REMARKS

The proposed method of ridge-shrinkage is a simple iterative method of adjusting sampling weights to meet RR and BC within tolerances. Each iteration involves a ridge step which modifies the usual GR-formula by introducing an inverse cost matrix $\Lambda$. A simple relation was established to choose $\Lambda$ corresponding to a specific tolerance matrix $\Delta$. The RS-method, like the result of Deville and Särndal (1992), for other calibration methods, remains asymptotically equivalent to GR if the matrix $N^{-1}\Lambda$ tends to zero in probability. The condition $N^{-1}\Lambda \to_p \mathbf{0}$ is satisfied by the adaptive choice of the tolerance matrix $\Delta$ as proposed in the paper because the GR-weights meet RR with high probability for large samples in view of the ADC property. This shows that RS is ADC, and its asymptotic variance can be conveniently obtained from the variance expression for GR.

The RS method generalizes the existing shrinkage-minimization calibration method by allowing BC to be nonbinding while meeting RR. Some other calibration methods, namely, the Huang-Fuller and the truncated linear can also be generalized in a similar way. The RS method also generalizes the existing ridge methods (which include the method of discarding BC as a special case) by allowing iterations to meet RR while satisfying BC within tolerances. Thus, it is expected to provide a useful practical calibration tool as it combines strengths of various existing methods dealing with both binding and nonbinding BC under RR.

## ACKNOWLEDGEMENTS

Both authors' research were supported in part by individual NSERC grants, the second author's grant being held at Carleton University under an adjunct research professorship.

## REFERENCES

Bankier, M.D., Rathwell, S., and Majkowski, M. (1992). Two step generalized least squares estimation in the 1991 Canadian Census. Methodology Branch Working Paper Series SSMD 92-007E, Statistics Canada, Ottawa.

Bardsley, P and Chambers, R.L. (1984). Multipurpose estimation from unbalanced samples. *Applied Statistics*, 33, 290-299.

Deville, J.-C., and Särndal, C.-E. (1992). Calibration estimation in survey sampling. *Journal of the American Statistical association*, 87, 376-382.

Huang, E.T., and Fuller, W.A. (1978). Nonnegative regression estimation in sample survey data. *Proceedings of the Social Statistics Section, American Statistical association*, 300-305.

Isaki, C.T,, and Fuller, W.A. (1982). Survey design under the regression superpopulation model. *Journal of the American Statistical Association*, 77, 89-96.

Rao, J.N.K. (1992). Internal memorandum to M.D. Bankier. Social Survey Methods Division, Statistics Canada, Ottawa, November 10.

Singh, A.C. (1993). On weight adjustment in survey sampling. Discussion paper for the 18th meeting of the Advisory Committee on Statistical Methods, Statistics Canada, Ottawa, October 25-26.

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

**Table 1: CV(g) and Discrepancy in respecting BC (FAMEX-Regina City)**
$(\alpha = .67, \beta = .8, \eta = .9, \psi = .9, \nu_{max} = 10,\ q_{max} = 10)$
L = 0.5, U = 2.0, #BC = 4

| Method | CV (g) | Dicrepancy in respecting BC in % | | | | |
|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | 4 | $\delta_{min}(\%)$ |
| RS | .520 | 0.00 | -3.88 | 0.00 | -0.07 | 3.88 |
| Ridge-SMCS | .524 | 3.51 | -3.51 | 0.00 | 0.00 | 3.51 |
| Ridge-MCS-*r* | .589 | -9.40 | -5.3 | 0.00 | -3.59 | 9.40 |

**Table 2: Difference in Point Estimates and Precision Relative to Regression Estimator**

| | **Owned Dwelling** | | **Furniture\Equipment** | |
|---|---|---|---|---|
| **Method** | RD | RP | RD | RP |
| RS | -.070 | .881 | -.008 | .888 |
| Ridge-SMCS | -.062 | .869 | -.004 | .893 |
| Ridge-MCS-*r* | -.096 | .893 | -.033 | .894 |
| | **Women's Clothing** | | **Men's Clothing** | |
| RS | -.019 | .869 | -.032 | .899 |
| Ridge-SMCS | -.015 | .874 | -.025 | .902 |
| Ridge-MCS-*r* | -.036 | .870 | -.032 | .894 |

Note: RD and RP denote respectively "relative difference" and "relative precision".