Pak. J. Statist
2009 Vol. 25(4), 371-384

RANGE RESTRICTED WEIGHT CALIBRATION FOR SURVEY DATA USING RIDGE REGRESSION

J. N. K. Rao[1] and **A. C. Singh**[2]

[1] Distinguished Research Professor, School of Mathematics and Statistics, Carleton University, Ottawa, Ontario K1S 5B6, Canada.
[2] Senior Research Advisor, Statistical Research and Innovation Division, Statistics Canada, Ottawa, ON, K1A 0T6.

ABSTRACT

Generalized regression is commonly used to adjust sampling weights to satisfy benchmark constraints. The resulting estimator of a total is design-consistent but some of the calibrated weights may not satisfy specified range restrictions. Several iterative modifications of generalized regression that attempt to meet both benchmark and range constraints have been proposed, but these methods may not yield a solution even if the range restrictions are mild, for example, restriction to nonnegative weights. A natural and practical way out of this difficulty is to make some of the benchmark constraints nonbinding up to specified tolerances while maintaining the other benchmark constraints as binding with zero tolerance. The range restrictions themselves can be relaxed if necessary to get lower tolerance levels on benchmark constraints. For complex surveys, Chambers (1996) proposed a modification of the ridge regression method of Bardsley and Chambers (1984) that leads to a design-consistent estimator of a total in spite of the presence of a ridge (or inverse cost) matrix which makes the weights stable, although the benchmark constraints become nonbinding. We establish a relationship between the ridge matrix and the matrix of specified tolerances and then show that Chamber's method can be adapted to meet benchmark constraints up to specified tolerances while maintaining design-consistency. This procedure is noniterative and easy to implement like the generalized regression, but it may not meet range restrictions despite relaxing of benchmark constraints. To address this problem, we propose an iterative method, termed ridge shrinkage, which generalizes the ridge regressions method in a manner similar to the iterative modifications of generalized regression, to meet range restrictions. This method is designed to force convergence for a specified number of iterations by using a built-in tolerance specification procedure to relax benchmark constraints while satisfying range restrictions and maintaining design consistency. Empirical results on the performance of the proposed method are also presented.

KEYWORDS

Benchmark constraints; Generalized regression; Jackknife; Ridge-shrinkage.

1 INTRODUCTION

In survey sampling, benchmark totals $T(x_j)$ of auxiliary variables x_j, $1 \leq j \leq p$, produced through external administrative sources, are commonly incorporated into the estimation of totals of survey variables by means of generalized regression (GR) (see e.g., Deville and Särndal, 1992). Use of such benchmark constraints (BC) is desirable not only from the efficiency perspective, but also due to the need to make survey estimates $\hat{T}(y_a)$ of totals $T(y_a)$ consistent with the external benchmark constraints, where y_a is a survey variable; that is, $\hat{T}(x_j) = T(x_j)$ for $j = 1,\ldots,p$. But generalized regression may lead to weights not satisfying desired range restrictions (RR). This could happen especially when the number of BC is large. To get around this problem, several modifications to GR have been proposed in the literature that attempt to meet both BC and RR by adjusting the GR weights (Huang and Fuller, 1978; Deville and Särndal, 1992; Jayasuriya and Valliant, 1996). Singh and Mohl (1996) provide an empirical evaluation of these methods.

The modified methods are iterative and may not yield a convergent solution that satisfies both BC and RR even if the RR are mild. Some of these methods automatically satisfy RR after each iteration and the iterations are continued to meet BC, while the others automatically satisfy BC after each iteration and the iterations are continued to meet RR. If the BC are not met in the former case after a specified number of iterations, then there is no control on the extent of discrepancy in meeting BC. Similarly, in the latter case there is no control on the extent of discrepancy in meeting RR.

Failure to achieve a solution may be due to RR being too tight, large discrepancies between BC and their estimates due to the sample size being not large enough, too many BC or near multi-collinearity of auxiliary variables implying instability of the estimated regression coefficients. One way out of this difficulty is to drop some of the BC, as proposed by Bankier et al. (1992). They encountered the problem of negative weights in the context of weight calibration for the Canadian Census 2B sample because of a large number of BC at the enumeration and weighting area levels. But the approach of Bankier et al. seems to be somewhat drastic in that some BC are treated as binding (i.e., zero tolerance) while the remaining as non-binding in the extreme sense (i.e., with infinite tolerance). A practical alternative that is appealing might be to allow most BC to be non-binding with possibly varying tolerances; the remaining BC would be binding based on subject matter requirements.

The instability of estimated regression coefficients mentioned above can be reduced by using ridge regression. Bardsley and Chambers (1984) proposed a model-based method using ridge regression to handle over-specified models. Such models are employed for multipurpose surveys with many survey variables because the same set of auxiliary variables, $\mathbf{x}$, can then be used in formulating linear regression models for all the survey variables $y_a (a = 1, \ldots, A)$. That is, $E(y_a|\mathbf{x}) = \mathbf{x}'\boldsymbol{\beta}_a, 1 \le a \le A$ which in turn leads to a unique case weight, m_k, for each unit or case k in the sample s. The resulting case-weighted estimators $\hat{T}_m(y_a) = \Sigma_s m_k y_{ak}$ of the totals $T(y_a)$ are internally consistent in the sense that $\hat{T}_m(y_a) + \hat{T}_m(y_b) = \hat{T}_m(y_a + y_b)$. Note that generalized regression and its modifications also lead to a unique set of calibration weights because the initial design weights, d_k, are calibrated to meet specified BC (see Section 2). Chambers (1996) modified Bardsley-Chambers' method to take account of the design weights d_k and obtained ridge case weights w_k. The resulting ridge case-weighted estimator, say $\hat{T}_w(y_a) = \Sigma_s w_k y_{ak}$, is asymptotically design-consistent even though the ridge estimator of regression coefficient vector is inconsistent.

The ridge case weights m_k or w_k involve the inverse of a diagonal cost matrix $\mathbf{C}$ with diagonal elements $c_j (\ge 0)$, where c_j represents the cost of the case-weighted estimator not satisfying the j-th specified calibration constraint, and a user specified scale factor λ. The choice $c_j = \infty$ ensures that the j-th BC is binding; see Section 3. The choice of λ is decided by examining the ridge trace plot of all the case weights $m_k = m_k(\lambda)$ or $w_k = w_k(\lambda), k \in s$ as log λ changes, as well as a plot of change in the relative calibration errors $\{\hat{T}_m(x_j) - T(x_j)\}/T(x_j)$ or $\{\hat{T}_w(x_j) - T(x_j)\}/T(x_j)$ as log λ changes (Chambers, 1996). This diagnostic approach should lead to a compromise solution in terms of both BC and RR. But we have no control on the extent of discrepancy in meeting BC, although the method can handle multicollinearity unlike GR-based methods. Moreover, for large-scale surveys the ridge trace plot of all the case weights may not be operationally feasible because of very large sample sizes.

In this paper we address the problem of a suitable choice of the cost elements "c_j" to meet tolerances on BC for specified RR. We first establish a relationship between the ridge matrix and the diagonal tolerance matrix $\boldsymbol{\Delta} = \text{diag}(\delta_1, \ldots, \delta_p)$, where δ_j is the tolerance for the j-th BC. This enables us to determine the cost elements for a given set of δ_j's; in particular, zero tolerance $(\delta_j = 0)$ corresponds to infinite cost $(c_j = \infty)$ and infinite tolerance $(\delta_j = \infty)$ to zero cost $(c_j = 0)$. Using the above relationship, we modify existing iterative calibration methods to meet RR while relaxing those BC that are deemed not binding. For given RR, the tolerance levels are chosen adaptively to relax BC only when necessary. We can also relax RR if lower tolerance levels are needed. Assuming that the GR weights meet RR asymptotically as sample size increases, the proposed method becomes asymptotically equivalent to GR. As a result, the asymptotic variance of the proposed case-weighted estimator can be estimated using the GR formula for the variance estimator with GR "residuals" replaced by the residuals of the proposed method.

Sections 2 and 3 provide a brief account of existing methods for binding and nonbinding BC. The proposed method, named ridge-shrinkage, and associated variance estimation by jackknife are described in Section 4. A numerical example based on Statistics Canada's Family Expenditure (FAMEX) survey data is presented in Section 5. Our method is design-based but it can be modified to handle model-based estimation.

2 GR-BASED METHODS

Let d_k be the (initial) design weight and w_k be the (final) calibrated weight for the k-th unit in a sample, s, of size $n(k = 1, \ldots, n)$. The benchmark constraints are given by $\hat{\mathbf{T}}_w = \mathbf{T}$, where $\mathbf{T}$ is the vector of x-totals $T(x_j)$ and $\hat{\mathbf{T}}_w$ is the corresponding vector of estimated totals $\hat{T}_w(x_j) = \mathbf{x}_j'\mathbf{w}$ with $\mathbf{w} = (w_1, \ldots, w_n)'$ and $\mathbf{x}_j = (x_{j1}, \ldots, x_{jn})'$. The range restrictions on the calibrated weights w_k are given by lower and upper bounds L and $U(L < 1 < U)$ such that $Ld_k \leq w_k \leq Ud_k$ for all $k \in s$. The tolerance matrix for benchmark constraints is given by $\mathbf{\Delta} = diag(\delta_1, \ldots, \delta_p)$, where δ_j is the tolerance for the j-th BC: $|\mathbf{x}_j'\mathbf{w} - T(x_j)| \leq \delta_j T(x_j), 1 \leq j \leq p$, assuming $T(x_j) > 0$ for all j. Note that $\delta_j = 0$ means a binding constraint while $\delta_j > 0$ implies a nonbinding constraint. The limiting case $\delta_j = \infty$ means that the constraint is discarded.

2.1 Generalized Regression

The GR weights, $\mathbf{w}^{gr}$, are obtained by minimizing a modified chisquare distance function, $\Sigma_s(w_k - d_k)^2/d_k$, subject to the benchmark constraints $\mathbf{x}_j'\mathbf{w} = T(x_j), j = 1, \ldots, p$. This leads to

$$\mathbf{w}^{gr} = \mathbf{d} + \mathbf{D}_s\mathbf{X}_s(\mathbf{X}_s'\mathbf{D}_s\mathbf{X}_s)^{-1}(\mathbf{T} - \hat{\mathbf{T}}_d), \quad (1)$$

where $\mathbf{d} = (d_1, \ldots, d_n)', \mathbf{D}_s = diag(d_1, \ldots, d_n), \hat{\mathbf{T}}_d$ is the vector of estimated x-totals $\hat{T}_d(x_j) = \Sigma_s d_k x_{jk}$ based on the initial design weights d_k and $\mathbf{X}_s$ is the $n \times p$ full rank matrix with k-th row given by $\mathbf{x}'_{(k)} = (x_{1k}, \ldots, x_{pk})$, see Deville and Särndal (1992). The generalized regression estimator, $\hat{T}_w^{gr}(y) = \Sigma_s w_k^{gr} y_k$, is design-consistent for the total $T(y)$, provided the basic estimator $\hat{T}_d(y) = \Sigma_s d_k y_k$ is design-consistent.

More general distance measures of the form $\Sigma_s (w_k - d_k)^2 \gamma_k / d_k$ have also been studied, where the weight γ_k is motivated through a "working" model $E(y_{ak}|\mathbf{x}_{(k)}) = \mathbf{x}'_{(k)}\boldsymbol{\beta}_a$ with $var(y_{ak}|\mathbf{x}_{(k)}) = \sigma_a^2 \gamma_k$ and known γ_k. In this case, the diagonal elements of $\mathbf{D}_s$ in (1) are changed to d_k/γ_k. We confine ourselves here to the case of uniform weighting, $\gamma_k = 1$, which is used in many applications.

The GR weights, $\mathbf{w}^{gr}$, satisfy BC, that is, $\mathbf{x}'_j \mathbf{w}^{gr} = T(x_j)$ for $j = 1, \ldots, p$, but may not satisfy range restrictions. In particular, some weights w_k^{gr} can be negative which may be unacceptable to some users. An alternative distance function, $\Sigma_s \{w_k \log(w_k/d_k) - w_k + d_k\}$, guarantees positive weights but it can lead to unrealistic or extreme weights (Deville and Sarndal, 1992). These weights are known as generalized raking weights.

2.2 Scaled Modified Chisquare

A slightly modified version of the method of Huang and Fuller (1978), termed scaled modified chisquare, was proposed by Singh and Mohl (1996). This method is iterative and satisfies BC after each iteration. At iteration $\nu + 1$, calibration weights $w_k^{(\nu+1)}$ are obtained by minimizing $\Sigma_s (w_k - d_k)^2 / (q_k^{(\nu)} d_k)$ subject to the benchmark constraints, where $q_k^{(\nu)}$ is a scale factor designed to control the weights $w_k^{(\nu)}$ that do not meet RR.

The weights $w_k^{(\nu+1)}$ are given by

$$\mathbf{w}^{(\nu+1)} = \mathbf{d} + \mathbf{D}_s^{(\nu)} \mathbf{X}_s (\mathbf{X}'_s \mathbf{D}_s^{(\nu)} \mathbf{X}_s)^{-1} (\mathbf{T} - \hat{\mathbf{T}}_d), \tag{2}$$

where $\mathbf{D}_s^{(\nu)} = diag(q_1^{(\nu)} d_1, \ldots, q_n^{(\nu)} d_n)$. Iterations start with $q_k^{(0)} = 1, 1 \leq k \leq n$ so that $\mathbf{w}^{(1)} = \mathbf{w}^{gr}$, the vector of GR-weights. The weights $\mathbf{w}^{(\nu+1)}$ satisfy BC at each iteration, and iterations are continued until RR are met or $\nu \geq \nu_{\max}$, a specified maximum number of iterations. A solution that satisfies both BC and RR may not exist, and in that case there is no control on the extent of discrepancy in meeting RR.

2.3 Restricted Modified Chisquare

An iterative method that satisfies RR after each iteration, termed restricted modified chisquare method by Singh and Mohl (1996), was proposed by Deville and Sarndal (1992). Here the distance function to be minimized remains unchanged over iterations, and iterations are continued to meet BC. The distance function is given by $\Sigma_s (w_k - d_k)^2 / d_k$ if $L d_k \leq$

$w_k \leq Ud_k; = \infty$ otherwise. At the $(\nu+1)$-th iteration the weights $\mathbf{w}^{(\nu+1)}$ are given by $w_k^{(\nu+1)} = \bar{w}_k^{(\nu+1)}$ if $Ld_k \leq \bar{w}_k^{(\nu+1)} \leq Ud_k; Ld_k$ if $\bar{w}_k^{(\nu+1)} < Ld_k; Ud_k$ if $\bar{w}_k^{(\nu+1)} > Ud_k$, where

$$\bar{\mathbf{w}}^{(\nu+1)} = \mathbf{w}^{(\nu)} + \mathbf{D}_s^{(\nu)} \mathbf{X}_s (\mathbf{X}_s' \mathbf{D}_s^{(\nu)} \mathbf{X}_s)^{-1} (\mathbf{T} - \bar{\mathbf{T}}_w^{(\nu)}) \tag{3}$$

with $\bar{\mathbf{T}}_w^{(\nu)} = \Sigma_s \bar{w}_k^{(\nu)} \mathbf{x}_{(k)}$ and $\mathbf{D}_s^{(\nu)} = diag(d_1^{(\nu)}, \ldots, d_n^{(\nu)})$ where $d_k^{(\nu)} = d_k$ if $Ld_k \leq \bar{w}_k^{(\nu)} \leq Ud_k; 0$ otherwise. Iterations start with $\bar{w}_k^{(0)} = d_k$. Again, a solution that satisfies both BC and RR may not exist and in this case there is no control on the extent of discrepancy in meeting BC.

2.4 Shrinkage Minimization

Singh and Mohl (1996) proposed another method, termed shrinkage minimization (SM), in which each iteration consists of a GR-step for a suitable chisquare distance. At the ν-th iteration, we minimize $\Sigma_s \{w_k - w_k^{(\nu)}\}^2 / w_k^{(\nu)}$ subject to BC. This leads to weights $\tilde{w}_k^{(\nu+1)}$ given by

$$\tilde{\mathbf{w}}^{(\nu+1)} = \mathbf{w}^{(\nu)} + \mathbf{D}_s^{(\nu)} \mathbf{X}_s (\mathbf{X}_s' \mathbf{D}_s^{(\nu)} \mathbf{X}_s)^{-1} (\mathbf{T} - \hat{\mathbf{T}}_w^{(\nu)}) \tag{4}$$

with $\mathbf{D}_s^{(\nu)} = diag(w_1^{(\nu)}, \ldots, w_n^{(\nu)})$ and $\hat{\mathbf{T}}_w^{(\nu)} = \Sigma_s w_k^{(\nu)} \mathbf{x}_{(k)}$, where $w_k^{(\nu)} = \tilde{w}_k^{(\nu)}$ if $Ld_k \leq \tilde{w}_k^{(\nu)} \leq Ud_k$; Ld_k if $\tilde{w}_k^{(\nu)} < Ld_k$; Ud_k if $\tilde{w}_k^{(\nu)} > Ud_k$. Iterations start with $w_k^{(0)} = d_k$ so that $\tilde{\mathbf{w}}^{(1)} = \mathbf{w}^{gr}$. Again, a solution that satisfies both BC and RR may not exist, and in that case there is no control on the extent of discrepancy in meeting BC. Note that $\tilde{\mathbf{w}}^{(\nu+1)}$ satisfies BC at each iteration.

3 RIDGE CASE WEIGHTS

We now turn to ridge case weights that take account of the design weights d_k. Let $c_j (\geq 0)$ be the cost of the case-weighted estimator not satisfying the j-th BC and let $\lambda (\geq 0)$ be a user-specified scale factor. The ridge case weights are obtained by minimizing the "λ-scaled and cost-ridged" loss function

$$L_\lambda(\mathbf{w}, \mathbf{d}) = \Sigma_{k=1}^n (w_k - d_k)^2 / d_k + \lambda^{-1} \Sigma_{j=1}^p c_j \{\hat{T}_w(x_j) - T(x_j)\}^2.$$

The resulting case weights are given by

$$\mathbf{w}(\lambda) = \mathbf{d} + \mathbf{D}_s \mathbf{X}_s (\mathbf{X}_s' \mathbf{D}_s \mathbf{X}_s + \lambda \mathbf{C}^{-1})^{-1} (\mathbf{T} - \hat{\mathbf{T}}_d), \tag{5}$$

where $\mathbf{C} = diag(c_1, \ldots, c_p)$. Note that for $\lambda = 0, \mathbf{w}(\lambda) = \mathbf{w}^{gr}$.

The case-weighted estimator of the total $T(y)$ is given by

$$\hat{T}_{w(\lambda)}(y) = \Sigma_s w_k(\lambda) y_k = \hat{T}_d(y) + \hat{\mathbf{B}}'_{s\lambda}(\mathbf{T} - \hat{\mathbf{T}}_d), \tag{6}$$

where

$$\hat{\mathbf{B}}_{s\lambda} = (\mathbf{X}'_s \mathbf{D}_s \mathbf{X}_s + \lambda \mathbf{C}^{-1})^{-1} \mathbf{X}' \mathbf{D}_s \mathbf{y} \tag{7}$$

is the ridge regression coefficient vector and $\mathbf{y} = (y_1, \ldots, y_n)'$. Note that as $\lambda \to 0$ we get the GR estimator

$$\hat{T}_w^{gr}(y) = \hat{T}_d(y) + \hat{\mathbf{B}}'_s(\mathbf{T} - \hat{\mathbf{T}}_d),$$

where $\hat{\mathbf{B}}_s$ is the vector of case-weighted regression coefficients obtained by setting $\lambda = 0$ in (7). It follows from the form (6) that $\hat{T}_{w(\lambda)}(y)$ is design-consistent for $T(y)$, even though $\hat{\mathbf{B}}_{s\lambda}$ is not design-consistent for the "census" regression coefficient vector unlike $\hat{\mathbf{B}}_s$.

From (5) we have

$$\mathbf{X}'_s \mathbf{w}(\lambda) = \mathbf{T} - \mathbf{\Lambda}(\mathbf{X}'_s \mathbf{D}_s \mathbf{X}_s + \mathbf{\Lambda})^{-1}(\mathbf{T} - \hat{\mathbf{T}}_d), \tag{8}$$

where $\mathbf{\Lambda} = diag(\lambda_1, \ldots, \lambda_p)$ with $\lambda_j = \lambda c_j^{-1}$. It follows from (8) that if $c_j = \infty$ (or $\lambda_j = 0$), then the j-th BC is exactly satisfied by $\mathbf{w}(\lambda)$. For a fixed $c_j > 0, \mathbf{w}(\lambda)$ does not satisfy the j-th BC. As $c_j \to 0$ (or $\lambda_j \to \infty$), the j-th BC is automatically discarded. To show this, assume all $\lambda_j > 0$ and rewrite (5) as

$$\mathbf{w}(\lambda) = \mathbf{d} + \mathbf{D}_s[I - \mathbf{X}_s \mathbf{\Lambda}^{-1} \mathbf{X}'_s(\mathbf{D}_s^{-1} + \mathbf{X}_s \mathbf{\Lambda}^{-1} \mathbf{X}'_s)^{-1}] \mathbf{X}_s \mathbf{\Lambda}^{-1}(\mathbf{T} - \hat{\mathbf{T}}_d), \tag{9}$$

using the matrix result $(\mathbf{I} + \mathbf{PQ})^{-1} = \mathbf{I} - \mathbf{P}(\mathbf{I} + \mathbf{QP})^{-1}\mathbf{Q}$. Now denoting $\mathbf{X}_s, \mathbf{\Lambda}$ and $\mathbf{T}$ and $\hat{\mathbf{T}}_d$ without the j-th row as $\mathbf{X}_{s(j)}, \mathbf{\Lambda}_{(j)}, T_{(j)}$ and $\hat{\mathbf{T}}_{d(j)}$ and letting $\lambda_j \to \infty$, (9) reduces to

$$\begin{aligned}\mathbf{w}(\lambda) &= \mathbf{d} + \mathbf{D}_s[\mathbf{I} - \mathbf{X}_{s(j)} \mathbf{\Lambda}_{(j)}^{-1} \mathbf{X}'_{s(j)}(\mathbf{D}_s^{-1} + \mathbf{X}_{s(j)} \mathbf{\Lambda}_{(j)}^{-1} \mathbf{X}_{s(j)})^{-1}] \mathbf{X}_{s(j)} \mathbf{\Lambda}_{(j)}^{-1}(\mathbf{T}_{(j)} - \hat{\mathbf{T}}_{d(j)}) \\ &= \mathbf{d} + \mathbf{D}_s \mathbf{X}_{(j)}(\mathbf{X}'_{s(j)} \mathbf{D}_s \mathbf{X}_{s(j)}) + \mathbf{\Lambda}_{(j)})^{-1}(\mathbf{T}_{(j)} - \hat{\mathbf{T}}_{d(j)})\end{aligned}$$

which shows that the j-th BC is discarded. It follows that Bankier et al's (1992) method of discarding some BC and retaining the remaining as binding is a special case of the ridge regression method with λ_j either 0 or ∞ for $j = 1, \ldots, p$.

The choice of λ, for specified c_j's, is decided by examining the ridge trace plot of case weights $w_k(\lambda), k = 1, \ldots, n$ as $\log \lambda$ changes, as well as a plot of change in the relative calibration errors $\{\hat{T}_{w(\lambda)}(x_j) - T(x_j)\}/T(x_j), j = 1, \ldots, p$ as $\log \lambda$ changes (Bardsley and Chambers, 1994). This diagnostic approach should lead to a compromise solution in terms of BC and RR. But there is no control on the extent of discrepancy in meeting BC, as in the case of modified GR methods, except for those with $c_j = \infty$. Moreover, for large-scale surveys the ridge trace plot of case weights may not be operationally feasible because of very large sample sizes.

We propose an alternative ridge calibration method in Section 4. This method sets tolerances γ_j on BC adaptively and then finds the corresponding λ_j that meets these tolerances. We regard $\lambda_1, \ldots, \lambda_p$ as parameters to be determined from the tolerances so that no apriori costs c_j are involved, except for the binding constraints with $c_j = \infty$ or $\delta_j = 0$ fixed.

4 RIDGE-SHRINKAGE

The ridge-shrinkage method combines in a fairly straightforward way the ridge regression method of Section 3 with the shrinkage-minimization method of subsection 2.4. Before describing the method, it is necessary to establish a link between the tolerance matrix $\mathbf{\Delta} = diag(\delta_1, \ldots, \delta_p)$ and the inverse cost matrix $\mathbf{\Lambda} = diag(\lambda_1, \ldots, \lambda_p)$ which will be used at each iteration of the method.

4.1 Link Between $\mathbf{\Delta}$ and $\mathbf{\Lambda}$

In the ridge approach, it may be easier in practice to specify $\mathbf{\Delta}$ than $\mathbf{\Lambda}$. Consider the ridge weights (5) with $\mathbf{d}$ changed to some weights $\mathbf{w}_{(0)}$ that satisfy RR, $\lambda\mathbf{C}^{-1}$ to arbitrary $\mathbf{\Lambda} = diag(\lambda_1, \ldots, \lambda_p)$ and $\mathbf{w}(\lambda)$ to $\mathbf{w}^r$. Then (8) may be written as

$$\mathbf{\Lambda}(\mathbf{X}_s'\mathbf{D}_{s(0)}\mathbf{X}_s + \mathbf{\Lambda})^{-1}(\mathbf{X}_s'\mathbf{w}_{(0)} - \mathbf{T}) = \mathbf{X}_s'\mathbf{w}^r - \mathbf{T}, \tag{10}$$

where $\mathbf{D}_{s(0)} = diag\{w_{1(0)}, \ldots, w_{n(0)}\}$. For each $j, 1 \leq j \leq p$, we want the j-th element of $\mathbf{X}_s'\mathbf{w}^r - \mathbf{T}$ to be less than or equal to $\delta_j T(x_j)$ in absolute value. We solve for $\mathbf{\Lambda}$ from (10) by setting $\mathbf{X}_s'\mathbf{w}^r - \mathbf{T}$ equal to the boundary values $\delta_j T(\mathbf{x}_j)$ with appropriate signs, i.e., equal to $\nabla\mathbf{T}$, where $\nabla = diag[sgn\{\mathbf{x}_j'\mathbf{w}_{(0)} - T(x_j)\}\delta_j; 1 \leq j \leq p]$. This gives

$$\mathbf{\Lambda}(\mathbf{X}_s'\mathbf{D}_{s(0)}\mathbf{X}_s)^{-1}[\mathbf{X}_s'\mathbf{w}_{(0)} - (\mathbf{I} + \nabla)\mathbf{T}] = \nabla\mathbf{T}. \tag{11}$$

Noting that $\mathbf{\Lambda}$ is diagonal, the elements λ_j can be obtained by element-wise division of the p-vector on the right hand side of (11) by the p-vector on the left hand side of (11). Note that $\lambda_j \to 0$ if $\delta_j \to 0$, but some λ_j's may be negative for other values of δ_j. This does not affect the solution although the interpretability of λ_j's as costs is lost.

4.2 Proposed Method

Each iteration of the proposed ridge shrinkage involves a ridge step followed by a shrinkage step. The method consists of cycles of iteration with tolerance matrix $\boldsymbol{\Delta}_q$ for the q-th cycle chosen adaptively, where q denotes the iteration index. For the initial cycle $q = 0$, we take $\delta_j = 0$ for all j (i.e., all binding BC) and perform shrinkage minimization (see subsection 2.4); the minimization step can be regarded as the ridge step with all $\lambda_j = 0$. At each iteration the calibration weights $\tilde{w}_k^{(\nu+1)}$, given by (4), meet BC. If $\tilde{\mathbf{w}}^{(\nu^*)}$ also meets RR for some $\nu^* \leq \nu_{\max}$, the procedure is terminated and $\tilde{\mathbf{w}}^{(\nu^*)}$ is taken as the desired solution. Otherwise, we shrink $\tilde{\mathbf{w}}^{(\nu^*)}$ to $\mathbf{w}^{(\nu^*)}$ to meet RR and specify tolerances $\delta_j^*(1)$ obtained by taking small increments on $\delta_j = 0$, say $\delta_j^*(1) = 0.001$. If $\mathbf{w}^{(\nu^*)}$ satisfy BC within tolerances $\delta_j^*(1)$, then again it is taken as the desired solution. If not, we go to cycle $q = 1$. For all cycles $q \geq 1$, the tolerances δ_j are always set to zero for controls that are deemed as binding or that are satisfied by $\mathbf{w}^{(\nu^*)}$ within a fraction of the tolerances specified at the previous cycle; this is done in the interest of faster convergence. Also, for each new cycle, we always start with the same set of weights $\mathbf{w}^{(\nu^*)}$ which is not from the previous cycle.

Now perform the ridge step by determining the elements λ_j of the inverse cost matrix Λ from (11) using $\mathbf{w}_{(0)} = \mathbf{w}^{(\nu^*)}$ and $\boldsymbol{\nabla}$ obtained from the above δ_j's. The ridge case weights $\mathbf{w}^r$ are then obtained from (5) and the ridge estimator $\mathbf{X}_s'\mathbf{w}^r$ meets BC within specified tolerance δ_j. If the weights $\mathbf{w}^r$ also satisfy RR then the procedure is terminated, otherwise the weights $\mathbf{w}^r$ are shrunk to meet RR to get the next set of weights. If these weights satisfy BC within the tolerances $\delta_j^*(1)$, the procedure is terminated. Otherwise, iterations are continued as above. If a solution is obtained in $\nu^* \leq \nu_{\max}$ iterations the procedure is terminated. Otherwise we go to the next cycle $q = 2$ and specify tolerances $\delta_j^*(2)$ obtained by taking small increments on $\delta_j^*(1)$. By choosing progressively higher tolerances $\delta_j^*(q)$ over the cycles q, we can achieve a solution for some q_0 that meets RR and BC within $\delta_j^*(q_0)$.

Denote the case weights obtained by the proposed method as $\mathbf{w}[\boldsymbol{\lambda}(q_0)]$. The resulting estimator of $T(y)$, say $\tilde{T}(y)$, is of the form (3) with $\lambda\mathbf{C}^{-1}$ changed to $\boldsymbol{\Lambda}(q_0) = \text{diag}[\lambda_1(q_0), \ldots, \lambda_p(q_0)]$. It now follows that the proposed estimator of $T(y)$ is design-consistent, noting that $\hat{T}_d(x_j)$ is design-consistent for $T_d(x_j)$.

One can also define ridge-versions of the other two methods, scaled modified chisquare and restricted modified chisquare, by introducing the inverse cost matrix $\boldsymbol{\Lambda}$ into (2) and (3). The specification of $\boldsymbol{\Lambda}$ from $\boldsymbol{\Delta}$ is quite similar to (11) for the ridge version of scaled modified chisquare. It is given by

$$\mathbf{\Lambda}(\mathbf{X}_s'\mathbf{D}_{s(0)}\mathbf{X}_s)^{-1}[\mathbf{X}_s'\mathbf{d}-(\mathbf{I}+\mathbf{\nabla})\mathbf{T}]=\mathbf{\nabla}\,\mathbf{T} \tag{12}$$

Note that $\mathbf{X}_s'\mathbf{w}_{(0)}$ in (11) is changed to $\mathbf{X}_s'\mathbf{d}$ in (12). For the ridge version of restricted modified chisquare, the specification of $\mathbf{\Lambda}$ from $\mathbf{\Delta}$ is somewhat different. It is given by

$$\begin{aligned}&\mathbf{\Lambda}(\mathbf{X}_s'\mathbf{D}_s\mathbf{X}_s)^{-1}[\mathbf{X}_s'\bar{\mathbf{D}}_{s(0)}\mathbf{X}_s-(\mathbf{I}+\mathbf{\nabla})\mathbf{T}] \\ =&\mathbf{\nabla}\,\mathbf{T}+[\mathbf{X}_s'(\mathbf{D}_s-\bar{\mathbf{D}}_{s(0)})\mathbf{X}_s](\mathbf{X}_s'\mathbf{D}_s\mathbf{X}_s)^{-1}[\mathbf{X}_s'\bar{\mathbf{w}}_{(0)}-(\mathbf{I}+\mathbf{\nabla})\mathbf{T}],\end{aligned} \tag{13}$$

where $\bar{\mathbf{D}}_{s(0)}=diag\{\bar{w}_{1(0)},\ldots,\bar{w}_{n(0)}\}$ with $\bar{\mathbf{w}}_{(0)}$ obtained from (3), i.e., the weights before truncation. As before, the elements λ_j can be obtained by element-wise division of the p-vector on the right hand side of (13) by the p-vector on the left hand side of (13). Note that an extra term is added on the right hand side of (13).

4.3 Jackknife Variance Estimation

The jackknife method can be employed to estimate the variance of the proposed estimator for stratified simple random sampling and stratified multistage sampling. We refer the reader to Shao and Tu (1995, Chapter 6) for an excellent account of resampling methods for survey data.

In the application given in Section 5, stratified multistage sampling was used. We now give a brief account of the jackknife for stratified multistage sampling. Suppose we have L strata with m_h primary units (clusters) i sampled from stratum h $(i=1,\ldots,m_h; h=1,\ldots,L)$. Denote the jackknife basic weights as $d_k^{(hi)}$ when the (hi)-th sample cluster is deleted: $d_k^{(hi)}=0$ if k belongs to (hi); $=m_h/(m_h-1)d_k$ if k belongs to (hj), $j\neq i$; $=w_k$ if k belongs to stratum g $(\neq h)$. We apply the proposed method to the jackknife weights $d_k^{(hi)}$ to obtian the estimators $\tilde{T}^{(hi)}(y)$, $i=1,\ldots,m_h$; $h=1,\ldots,L$. A jackknife estimator of the proposed estimator $\tilde{T}(y)$ is given by

$$v_J[\tilde{T}(y)]=\sum_h\frac{m_h-1}{m_h}\sum_i\left\{\tilde{T}^{(hi)}(y)-\tilde{T}(y)\right\}^2. \tag{14}$$

Jackknife variance estimators for the existing methods are obtained in a similar manner (Singh and Mohl, 1996). Asymptotic consistency of the proposed jackknife variance estimator (14) remains to be investigated.

5 APPLICATION

In this section we apply the proposed method to Statistics Canada's Family Expenditure (FAMEX) survey data (Singh and Mohl, 1996). We compared all the ridge versions but details for ridge versions of scaled modified chisquare (SMC) and restricted modified chisquare (RMC) are not given here. A common value 0.001 for the tolerances $\delta_j^*(1)$ with successive increments of 0.001 was chosen. For illustration, the three ridge versions were applied to the 1990 FAMEX data for the city of Regina with $n = 797$ persons. The benchmark constraints correspond to population counts for the following four groups: age < 15, age ≥ 15, one person household and household with two or more persons. We refer the reader to Singh and Mohl (1996) for details about FAMEX data.

Since the number of BC, p, is only 4, RR were made quite tight: $L = 0.5, U = 2.0$. For this choice of RR, none of the existing iterative methods satisfied BC even after 100 iterations. The percent discrepancy in respecting the four BC were 21.6, 16.9, 75.2 and 19.6 for shrinkage minimization (SM), 211, 18.7, 75.2 and 21.0 for SMC, and 97.2, -21.1, -14 and 2.6 for RMC.

Table 1. Values of $CV(g), \delta_{\min}$ and D for SM-r, RMC-r

Method	% CV(g)	$\delta_{\min}(\%)$	D
SM-r	52.0	3.9	210
SMC-r	52.4	3.5	221
RMC-r	58.9	9.4	813

Note: SM-r, SMC-r and RMC-r denote ridge versions of shrinkage minimization, scaled modified chisquare and restricted modified chisquare respectively

To study the performance of the three ridge versions, denoted by SM-r, SMC-r and RMC-r, we used the following measures: (a) CV(g): coefficient of variation of the ratios $g_k = w_k/d_k$, where w_k is the final weight and d_k is the initial design weight; (b) $\delta_{\min}$: Minimum tolerance required so that all BC are met within tolerance; (c) $D = \Sigma_j[\mathbf{x}_j'\mathbf{w} -$

Table 2. % Relative difference (RD) and % relative precision (RP) of ridge methods

Method	RD	RP	RD	RD
	Owned Dwelling		Furniture and Equipment	
SM-r	-7.0	88.1	-0.8	88.8
SMC-r	-6.2	86.9	-0.4	89.3
RMC-r	-9.6	89.3	-3.3	89.4
	Women's Clothing		Men's Clothing	
SM-r	-1.9	86.9	-3.2	89.9
SMC-r	-1.5	87.4	-2.5	90.2
RMC-r	-3.6	87.0	-3.2	89.4

$T(x_j)\}^2/T(x_j)$, a chisquare measure of distance between the estimates $\mathbf{x}_j'\mathbf{w}$ and the corresponding x-totals $T(x_j)$. Table 1 reports these values. It is seen from Table 1 that the ridge versions of shrinkage minimization, SM-r, and scaled modified chisquare, SMC-r, behave quite similarly in terms of CV(g), $\delta_{\min}$ and D. Note that $\delta_{\min}$ is quite small here for SM-r and SMC-r even under tight range restrictions. The ridge version of restricted modified chisquare, RMC-r, leads to $\delta_{\min}$ and D considerably bigger than those for SM-r and SMC-r.

Table 2 shows the relative difference (RD) and relative precision (RP) of point estimators for four survey variables, where RD = (ridge estimator-GR estimator)/GR estimator and RP = s.e. (GR estimator)/s.e. (ridge estimator). Standard errors (s.e.) were computed using the jackknife method but using only $q_{\max} = 10$ cycles for each jackknife replicate to reduce the computations. It is seen that the ridge-calibration estimators are less efficient than the GR estimator: the RP is about 85%. Some efficiency loss is to be expected because BC are not exactly satisfied by the ridge-calibrations methods, even though the ridge-based weights are more stable. Also, the jackknife final weights may lead to larger discrepancies in meeting BC because iterations are terminated at $q = 10$ cycles. As a result, the jackknife variance estimate may be overestimating the variance of the estimator.

To study the performance of drop-some-BC method of Bankier et al. (1992), we dropped the second BC. In this case, all the three non-ridge methods converged in one iteration which implies that GR-based weights satisfy RR and the three BC's. But the discrepancy with respect to the dropped BC is -13.8% which is much higher than the 3.9% discrepancy for the SM-r weights which also satisfy the other three BC's almost perfectly.

CONCLUDING REMARKS

An earlier version of this paper has appeared in the 1997 Proceedings of the Section on Survey Research Methods, American Statistical Association, Washington DC, pages 57-65. Chen, Sitter and Wu (2002) used an alternative method that meets pre-specified range restrictions (RR). Beaumont and Bocci (2008) established a connection between the two methods and justified the use of ridge calibration.

ACKNOWLEDGEMENT

We wish to thank a referee for useful comments. This research was supported by grants from the Natural Sciences and Engineering Research Council of Canada.

REFERENCES

1. 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.

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

3. Beaumont, J.F. and Bocci, C. (2008). Another look at ridge calibration. *Metron*, **66**, 5-20.

4. Chambers, R.L. (1996). Robust case-weighting for multipurpose establishment surveys. *J. Off. Statist.*, **12**, 3-32.

5. Chen, J., Sitter, R. and Wu, C. (2002). Using empirical likelihood methods to obtain range restricted weights in regression estimators for surveys. *Biometrika*, **89**, 230-237.

6. Deville, J.-C. and Särndal, C-E. (1992). Calibration estimation in survey sampling. *J. Amer. Statist. Assoc.*, **87**, 376-382.

7. Huang, E.T. and Fuller, W.A. (1978). Nonnegative regression estimation for sample survey data. *Proceedings of the Social Statistics Section, American Statistical Association*, Washington, D.C., pp. 300-305.

8. Jayasuriya, B. and Valliant, R. (1996). An application of restricted regression estimation in a household survey. *Survey Meth.,* **22**, 127-137.

9. Rao, J.N.K. and Singh, A.C. (1997). A ridge-shrinkage method for range-restricted weight calibration in survey sampling. *Proceedings of the Section on Survey Research Methods, American Statistical Association*, Washington, D.C., pp. 57-65.

10. Shao, J. and Tu, D. (1995). *The jackknife and Bootstrap*. New York: Springer-Verlag.

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