A note on partially penalized calibration

Camelia Goga [1] & Muhammad Ahmed Shehzad [2]

[1] *IMB, Université de Bourgogne, Dijon, France. E-mail: camelia.goga@u-bourgogne.fr*
[2] *Department of Statistics, Bahauddin Zakariya University, Multan, Pakistan. E-mail: mdjan708@gmail.com*

Abstract. Partially penalized calibration is a method to obtain weights that allow exact estimation of the total of several auxiliary variables and the approximate estimation of the total of other variables. It was suggested by Bardsley and Chambers (1984), in a model-based setting, and by Guggemos and Tillé (2010), in a calibration setting. They proposed two different optimization problems to obtain such weights. In this paper, we show that the two optimization problems lead to the same solution.

Key words. Biased estimators, Calibration, Penalized calibration, Ridge estimation, Survey sampling.

1 Introduction

The calibration technique as suggested by Deville and Särndal (1992) is a widely used method to derive weights w_k that incorporate auxiliary information given by p variables $\mathcal{X}_1, \ldots, \mathcal{X}_p$. The main goal is to improve the estimation of the finite population total of the study variable $\mathcal{Y}$,

$$t_y = \sum_U y_k$$

while estimating exactly the known population totals of $\mathcal{X}_j, j = 1, \ldots, p$

$$\sum_{k \in s} w_k \mathbf{x}_k = \sum_{k \in U} \mathbf{x}_k \tag{1}$$

where s is a sample selected from the population U according to a sampling design $p(\cdot)$ and $\mathbf{x}_k = (X_{k1}, \ldots, X_{kp})^t$ is the vector of values of $\mathcal{X}_1, \ldots, \mathcal{X}_p$ recorded for the k-th individual. In a model-based approach, a sample s satisfying the relation (1) is called a *balanced sample* (Valliant *et al.*, 2000, p. 54) and Royall and Herson (1973) have shown how balanced sampling can protect from model misspecification.

In practice, the survey statistician usually deals with multipurpose surveys, namely means or totals of a large number of variables are to be estimated. Thus, a large number of auxiliary variables are to be considered and the requirement that the constraints given by (1) are satisfied exactly becomes too severe. Guggemos and Tillé (2010) called it an *over-calibration* situation. In such conditions, Bardsley and Chambers (1984) suggested dropping the exactly balanced requirement by imposing a cost function which penalizes large values of $\sum_{k \in s} w_k \mathbf{x}_k - \sum_{k \in U} \mathbf{x}_k$. This construction leads naturally to a ridge-type class of estimators (Hoerl and Kennard, 1970). The method has a design-based version suggested by Chambers (1996) and afterwards, by Rao and Singh (1997, 2009), Beaumont and Bocci (2008) and Guggemos and Tillé (2010). Very recently, Goga and Shehzad (2013) give a comprehensive review of the ridge principle and of its use in survey sampling. They also give geometrical interpretations.

An intermediate situation is when several auxiliary variables are considered important and their totals must be estimated exactly. Bardsley and Chambers (1984), in a model-based framework, and Guggemos and Tillé (2010), in a calibration framework, suggested independently two optimization problems to obtain weights that satisfy this requirement. The Guggemos and Tillé's optimization problem states clearly the exact calibration on certain totals, but it should be shown for the Bardsley and Chambers's method. However, the weights derived by the Guggemos and Tillé's constrained optimization problem are very complicated and so is, the weighted estimator built on them. On the contrary, the Bardsley and Chambers's estimator is a ridge-type regression estimator with a particular cost matrix.

We show in this paper that the weights determined by these two different optimization problems are in fact equal. Thus, the same weighted-estimators are obtained. A combined strategy may then be used: weights are sought such that they verify the Guggemos and Tillé's constrained optimization problem but their expression is given by the one obtained with the Bardsley and Chambers's optimization problem.

2 Penalized or ridge estimators

Let $\mathbf{y} = (y_1, \ldots, y_N)'$ be a $N \times 1$ vector of population values and let $\mathbf{X} = (\mathbf{X}_1, \ldots, \mathbf{X}_p)$ be the $N \times p$ matrix with $\mathbf{x}_k' = (X_{k1}, \ldots, X_{kp})$ as rows. Let ξ be a superpopulation model explaining the relationship between the auxiliary variables $\mathcal{X}_1, \ldots, \mathcal{X}_p$ and the variable of interest $\mathcal{Y}$,

$$\xi : \quad \mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \tag{2}$$

The errors ε_k, for all $k \in U$ are independent of one another, with mean zero and variance $\mathrm{Var}(\varepsilon_k) = \sigma^2 v_k^2$, where v_k are positive and known quantities. Let $\mathrm{Var}_\xi(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{V}$ with $\mathbf{V} = \mathrm{diag}(v_k^2)_{k \in U}$. Some further notations are needed. Let $\mathbf{X}_s = (\mathbf{x}_k')_{k \in s}$, respectively $\mathbf{y}_s = (y_k)_{k \in s}$, be the restriction of $\mathbf{X}$, respectively of $\mathbf{y}$, on the sample s. Let $\mathrm{Var}_\xi(\boldsymbol{\varepsilon}_s) = \sigma^2 \mathbf{V}_s$ be the variance of $\boldsymbol{\varepsilon}_s$, the restriction of $\boldsymbol{\varepsilon}$ on the sample s, and $\mathrm{Var}_\xi(\boldsymbol{\varepsilon}_{\overline{s}}) = \sigma^2 \mathbf{V}_{\overline{s}}$ be the variance of $\boldsymbol{\varepsilon}_{\overline{s}}$, the restriction of $\boldsymbol{\varepsilon}$ on $\overline{s} = U - s$.

Model-based approach

Bardsley and Chambers (1984) suggested using the weights $\mathbf{w}_s = (w_k)_{k \in s}$ that verify the following optimization problem

$$\textbf{(P1)}: \quad \mathbf{w}_{MB,R} = \mathrm{argmin}_{\mathbf{w}_s} (\mathbf{w}_s - \mathbf{1}_s)' \mathbf{V}_s (\mathbf{w}_s - \mathbf{1}_s) + (\mathbf{w}_s' \mathbf{X}_s - \mathbf{1}_U' \mathbf{X}) \mathbf{C} (\mathbf{w}_s' \mathbf{X}_s - \mathbf{1}_U' \mathbf{X})'$$

where $\mathbf{1}_s, \mathbf{1}_U$ are vectors of ones and $\mathbf{C} = \mathrm{diag}(C_j)_{j=1}^p$ is a diagonal matrix where $C_j \geq 0$ is a user-specified cost associated with not satisfying the j-th calibration equation. The weights $\mathbf{w}_{MB,R}$ are determined by taking the derivative of the loss function

$$\mathcal{L}(\mathbf{w}_s) = (\mathbf{w}_s - \mathbf{1}_s)' \mathbf{V}_s (\mathbf{w}_s - \mathbf{1}_s) + (\mathbf{w}_s' \mathbf{X}_s - \mathbf{1}_U' \mathbf{X}) \mathbf{C} (\mathbf{w}_s' \mathbf{X}_s - \mathbf{1}_U' \mathbf{X})'.$$

with respect to $\mathbf{w}_s$ and solving $\frac{\partial \mathcal{L}(\mathbf{w}_s)}{\partial \mathbf{w}_s} = 0$. We get after some algebra:

$$\mathbf{w}_{MB,R} = \mathbf{1}_s - (\mathbf{V}_s + \mathbf{X}_s\mathbf{C}\mathbf{X}_s')^{-1}\mathbf{X}_s\mathbf{C}(\mathbf{X}_s'\mathbf{1}_s - \mathbf{X}'\mathbf{1}_U).$$

Using Lemma 3 from the Appendix, we obtain

$$(\mathbf{V}_s + \mathbf{X}_s\mathbf{C}\mathbf{X}_s')^{-1}\mathbf{X}_s\mathbf{C} = \mathbf{V}_s^{-1}\mathbf{X}_s(\mathbf{C}^{-1} + \mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1}$$

leading to the following expression of the ridge model-based weights

$$\mathbf{w}_{MB,R} \quad = \quad \mathbf{1}_s - \mathbf{V}_s^{-1}\mathbf{X}_s\left(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s + \mathbf{C}^{-1}\right)^{-1}(\mathbf{X}_s'\mathbf{1}_s - \mathbf{X}'\mathbf{1}_U). \tag{3}$$

Thus, the ridge model-based estimator $\hat{t}_{MB,R}$ of t_y is given by,

$$\hat{t}_{MB,R} = \mathbf{w}'_{MB,R}\mathbf{y}_s = \sum_{k\in s} y_k + \left(\sum_{k\in U}\mathbf{x}_k - \sum_{k\in s}\mathbf{x}_k\right)'\hat{\boldsymbol{\beta}}_{MB,R} \tag{4}$$

where $\hat{\boldsymbol{\beta}}_{MB,R} = \left(\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s + \mathbf{C}^{-1}\right)^{-1}\mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{y}_s$. One may remark that $\hat{\boldsymbol{\beta}}_{MB,R}$ is in fact the ridge-type estimator (Hoerl and Kennard, 1970) of the regression coefficient $\boldsymbol{\beta}$ from the superpopulation model ξ.

Calibration approach

Let $\pi_k = Pr(k \in s)$ for $k \in s$ be the first-order inclusion probability and let $d_k = 1/\pi_k$ be the sampling weights. Chambers (1996) and Rao and Singh (1997, 2009) suggested finding weights that satisfy the following optimization problem

$$(\mathbf{P2}): \quad \mathbf{w}_{CAL,R} = \text{argmin}_{\mathbf{w}_s}(\mathbf{w}_s - \mathbf{d}_s)'\widetilde{\mathbf{\Pi}}_s(\mathbf{w}_s - \mathbf{d}_s) + (\mathbf{w}_s'\mathbf{X}_s - \mathbf{1}_U'\mathbf{X})\mathbf{C}(\mathbf{w}_s'\mathbf{X}_s - \mathbf{1}_U'\mathbf{X}) \tag{5}$$

where $\mathbf{d}_s = (d_k)_{k\in s}$ and $\widetilde{\mathbf{\Pi}}_s = \text{diag}(q_k^{-1}d_k^{-1})_{k\in s}$ with q_k positive constants. Usually, $q_k = 1$ for all $k \in s$ is considered. Following the same lines as in the model-based case, we determine the penalized calibration weights as follows

$$\mathbf{w}_{CAL,R} = \mathbf{d}_s - \widetilde{\mathbf{\Pi}}_s^{-1}\mathbf{X}_s(\mathbf{X}_s'\widetilde{\mathbf{\Pi}}_s^{-1}\mathbf{X}_s + \mathbf{C}^{-1})^{-1}(\mathbf{X}_s'\mathbf{d}_s - \mathbf{X}'\mathbf{1}_U). \tag{6}$$

The penalized calibration estimator of the population total t_y is given by

$$\hat{\mathbf{t}}_{CAL,R} = (\mathbf{w}_{CAL,R})'\mathbf{y}_s \quad = \quad \sum_{k\in s} d_k y_k + \left(\sum_{k\in U}\mathbf{x}_k - \sum_{k\in s} d_k\mathbf{x}_k\right)'\hat{\boldsymbol{\beta}}_{CAL,R} \tag{7}$$

where $\hat{\boldsymbol{\beta}}_{CAL,R} = (\mathbf{X}_s'\widetilde{\mathbf{\Pi}}_s^{-1}\mathbf{X}_s + \mathbf{C}^{-1})^{-1}\mathbf{X}_s'\widetilde{\mathbf{\Pi}}_s^{-1}\mathbf{y}_s$ is the ridge-type estimator of $\boldsymbol{\beta}$ derived in a design-based or calibration framework.

Considering a cost C_j going to zero means discarding from (1) the constraint corresponding to X_j and considering a cost C_j going to infinity, means satisfying this constraint exactly. This is true for the model-based approach as well as for the calibration approach. In practice, one needs only to specify the inverse of the matrix $\mathbf{C}$ in order to compute the penalized weights (see relations 3 and 6). So, if we desire having exact calibration on the total of X_j, then we consider $C_j^{-1} = 0$ unless we take a large C_j^{-1}. This idea has been used by Bardsley and Chambers (1984) to obtain partially penalized estimators and it is developed in the next section.

3 Partially penalized estimators

Consider now that we want to estimate exactly the total of q auxiliary variables contained in $\mathbf{X}$ and that we are less restrictive concerning the estimation of the other $p-q$ totals. We may consider that the matrix $\mathbf{X}$ has the following expression after re-ordering the variables $\mathbf{X}_1, \ldots, \mathbf{X}_p$,

$$\mathbf{X} = \left(\widetilde{\mathbf{X}}_1, \widetilde{\mathbf{X}}_2\right),$$

where $\widetilde{\mathbf{X}}_1 = [\mathbf{X}_1, \ldots, \mathbf{X}_q]$ and $\widetilde{\mathbf{X}}_2 = [\mathbf{X}_{q+1}, \ldots, \mathbf{X}_p]$. The variables contained in $\widetilde{\mathbf{X}}_1$ may be related for example to socio-demographic criteria. Let the cost matrix be of the following form,

$$\mathbf{C} = \left(\begin{array}{ll} \mathbf{C}_1 & \mathbf{0}_{(q,p-q)} \\ \mathbf{0}_{(p-q,p)} & \mathbf{C}_2 \end{array}\right),$$

where $\mathbf{C}_1$, respectively $\mathbf{C}_2$, is the relative diagonal cost matrix of size $q \times q$ associated to $\widetilde{\mathbf{X}}_1$, respectively of size $(p-q)\times(p-q)$ associated to $\widetilde{\mathbf{X}}_2$.
In order to estimate exactly the total of variables from $\tilde{\mathbf{X}}_1$, Bardsley and Chambers (1984) suggested using the optimization problem **(P1)** for a particular cost matrix $\mathbf{C}$. The *partially penalized ridge* weights may be derived as solution of the following optimization problem

$$\textbf{(P1')}: \quad \mathbf{w}_{ppr}^{(1)} = \operatorname{argmin}_{\mathbf{w}_s}(\mathbf{w}_s - \mathbf{1}_s)'\mathbf{V}_s(\mathbf{w}_s - \mathbf{1}_s) + (\mathbf{w}_s'\mathbf{X}_s - \mathbf{1}_U'\mathbf{X})\mathbf{C}(\mathbf{w}_s'\mathbf{X}_s - \mathbf{1}_U'\mathbf{X})'$$

with the inverse cost matrix given by

$$\mathbf{C}^{-1} = \left(\begin{array}{ll} \mathbf{0}_{(q,q)} & \mathbf{0}_{(q,p-q)} \\ \mathbf{0}_{(p-q,p)} & \mathbf{C}_2^{-1} \end{array}\right) \tag{8}$$

The optimization problem **(P1')** is a particular case of **(P1)**. Thus, the weights $\mathbf{w}_{ppr}^{(1)}$ may be determined from relation (3) for the inverse cost matrix given in (8),

$$\begin{aligned} & \mathbf{w}_{ppr}^{(1)} \\ = \; & \mathbf{1}_s - \left(\mathbf{V}_s^{-1}\widetilde{\mathbf{X}}_{1s}, \quad \mathbf{V}_s^{-1}\widetilde{\mathbf{X}}_{2s}\right)\left(\begin{array}{cc} \widetilde{\mathbf{X}}_{1s}'\mathbf{V}_s^{-1}\widetilde{\mathbf{X}}_{1s} & \widetilde{\mathbf{X}}_{1s}'\mathbf{V}_s^{-1}\widetilde{\mathbf{X}}_{2s} \\ \widetilde{\mathbf{X}}_{2s}'\mathbf{V}_s^{-1}\widetilde{\mathbf{X}}_{1s} & \widetilde{\mathbf{X}}_{2s}'\mathbf{V}_s^{-1}\widetilde{\mathbf{X}}_{2s} + \mathbf{C}_2^{-1} \end{array}\right)^{-1}\left(\begin{array}{c} \widetilde{\mathbf{X}}_{1s}'\mathbf{1}_s - \widetilde{\mathbf{X}}_1'\mathbf{1}_U \\ \widetilde{\mathbf{X}}_{2s}'\mathbf{1}_s - \widetilde{\mathbf{X}}_2'\mathbf{1}_U \end{array}\right), \end{aligned} \tag{9}$$

where $\widetilde{\mathbf{X}}_{1s}$, respectively $\widetilde{\mathbf{X}}_{2s}$, is the sample restriction of $\widetilde{\mathbf{X}}_1$, respectively of $\widetilde{\mathbf{X}}_2$. The total t_y is estimated by

$$\hat{t}_{ppr}^{(1)} = (\mathbf{w}_{ppr}^{(1)})'\mathbf{y}_s = \sum_{k\in s} y_k + \left(\sum_{k\in U}\mathbf{x}_k - \sum_{k\in s}\mathbf{x}_k\right)'\hat{\boldsymbol{\beta}}_{MB,R}$$

where $\hat{\boldsymbol{\beta}}_{MB,R} = \left(\mathbf{X}_s' \mathbf{V}_s^{-1} \mathbf{X}_s + \left(\begin{array}{cc} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_2^{-1} \end{array} \right) \right)^{-1} \mathbf{X}_s' \mathbf{V}_s^{-1} \mathbf{y}_s$.

One may verify (see also Beaumont and Bocci, 2008) that using the weights $\mathbf{w}_{ppr}^{(1)}$ allows estimating exactly the total of the variables from $\widetilde{\mathbf{X}}_1$, namely we have

$$(\mathbf{w}_{ppr}^{(1)})' \widetilde{\mathbf{X}}_{1s} = \mathbf{1}_U' \widetilde{\mathbf{X}}_1. \tag{10}$$

In a design-based framework, Guggemos and Tillé (2010) suggested determining weights that are as close as possible to the sampling weights $d_k, k \in s$ while exactly estimating the total of the auxiliary variables contained in $\widetilde{\mathbf{X}}_1$ and estimating approximatively the total of the variables from $\widetilde{\mathbf{X}}_2$. We transpose the Guggemos and Tillé's idea in a model-based framework. This means that we are looking for weights that satisfy

$$\textbf{(P2')}: \quad \mathbf{w}_{ppr}^{(2)} = \text{argmin}_{\mathbf{w}_s} (\mathbf{w}_s - \mathbf{1}_s)' \mathbf{V}_s (\mathbf{w}_s - \mathbf{1}_s) + (\mathbf{w}_s' \widetilde{\mathbf{X}}_{2s} - \mathbf{1}_U' \widetilde{\mathbf{X}}_2) \mathbf{C}_2 (\mathbf{w}_s' \widetilde{\mathbf{X}}_{2s} - \mathbf{1}_U' \widetilde{\mathbf{X}}_2)'$$

subject to

$$\mathbf{w}_s' \widetilde{\mathbf{X}}_{1s} = \mathbf{1}_U' \widetilde{\mathbf{X}}_1. \tag{11}$$

The Lagrangian function is

$$\mathcal{L}(\mathbf{w}_s, \boldsymbol{\lambda}) = (\mathbf{w}_s - \mathbf{1}_s)' \mathbf{V}_s (\mathbf{w}_s - \mathbf{1}_s) + (\mathbf{w}_s' \widetilde{\mathbf{X}}_{2s} - \mathbf{1}_U' \widetilde{\mathbf{X}}_2) \mathbf{C}_2 (\mathbf{w}_s' \widetilde{\mathbf{X}}_{2s} - \mathbf{1}_U' \widetilde{\mathbf{X}}_2)' - (\mathbf{w}_s' \widetilde{\mathbf{X}}_{1s} - \mathbf{1}_U' \widetilde{\mathbf{X}}_1) \boldsymbol{\lambda}.$$

and the solution is obtained by solving $\frac{\partial \mathcal{L}(\mathbf{w}_s, \boldsymbol{\lambda})}{\partial \mathbf{w}_s} = 0$ under the constraint $\mathbf{w}_s' \widetilde{\mathbf{X}}_{1s} = \mathbf{1}_U' \widetilde{\mathbf{X}}_1$ (Guggemos and Tillé, 2010). We obtain

$$\begin{aligned} & \mathbf{w}_{ppr}^{(2)} \\ = \; & \boldsymbol{\Omega}_s^{-1} \left[\tilde{\mathbf{X}}_{1s} (\tilde{\mathbf{X}}_{1s}' \boldsymbol{\Omega}_s^{-1} \tilde{\mathbf{X}}_{1s})^{-1} \left(\tilde{\mathbf{X}}_1' \mathbf{1}_U - \tilde{\mathbf{X}}_{1s}' \boldsymbol{\Omega}_s^{-1} (\mathbf{V}_s \mathbf{1}_s + \tilde{\mathbf{X}}_{2s} \mathbf{C}_2 \tilde{\mathbf{X}}_2' \mathbf{1}_U) \right) + \mathbf{V}_s \mathbf{1}_s + \tilde{\mathbf{X}}_{2s} \mathbf{C}_2 \tilde{\mathbf{X}}_2' \mathbf{1}_U \right] \end{aligned} \tag{12}$$

where $\boldsymbol{\Omega}_s = \mathbf{V}_s + \widetilde{\mathbf{X}}_{2s} \mathbf{C}_2 \widetilde{\mathbf{X}}_{2s}'$. The partial penalized estimator for the total t_y becomes

$$\hat{t}_{ppr}^{(2)} = (\mathbf{w}_{ppr}^{(2)})' \mathbf{y}_s = \mathbf{1}_s' \mathbf{y}_s + (\mathbf{1}_U' \widetilde{\mathbf{X}}_1 - \mathbf{1}_s' \widetilde{\mathbf{X}}_{1s}) \hat{\mathbf{b}} + (\mathbf{1}_U' \widetilde{\mathbf{X}}_2 - \mathbf{1}_s' \widetilde{\mathbf{X}}_{2s}) \hat{\mathbf{u}} \tag{13}$$

where $\hat{\mathbf{b}} = \left(\widetilde{\mathbf{X}}_{1s}' \boldsymbol{\Omega}_s^{-1} \widetilde{\mathbf{X}}_{1s} \right)^{-1} \widetilde{\mathbf{X}}_{1s}' \boldsymbol{\Omega}_s^{-1} \mathbf{y}_s$, and $\hat{\mathbf{u}} = \mathbf{C}_2 \widetilde{\mathbf{X}}_{2s}' \boldsymbol{\Omega}_s^{-1} (\mathbf{y}_s - \widetilde{\mathbf{X}}_{1s} \hat{\mathbf{b}})$.

*Remark 1: By imposing (11), the constrained optimization problem **(P2')** states clearly the fact that the weights allow exact calibration to $\widetilde{\mathbf{X}}_1$ while this property should be shown in the case of **(P1')** (relation 10). However, the estimator $\hat{t}_{ppr}^{(2)}$ has a much more complicated formula compared to $\hat{t}_{ppr}^{(1)}$.*

*Remark 2: Under a calibration approach, the weights may be derived from the optimization problem **(P1')** or **(P2')** with $\mathbf{V}_s$ replaced by $\widetilde{\boldsymbol{\Pi}}_s$ and $\mathbf{1}_s$ by $\mathbf{d}_s$.*

*Remark 3: Park and Yang (2008) suggested the optimization problem **(P2')** in order to derive weights for estimating the mean $\bar{y}_U = \sum_U y_k/N$ of the variable of interest $\mathcal{Y}$. They used the model ξ given in (2) with an intercept and they looked for a weighted estimator with sum of weights equal to the unity and being as close as possible to the Hájek (1971) weights,*

$$\alpha_i = \frac{\pi_i^{-1}}{\sum_s \frac{1}{\pi_i}}.$$

*This means that the optimization problem **(P2)** is used with $\mathbf{1}_s$ replaced by $\boldsymbol{\alpha}_s = (\alpha_i)_{i\in s}$.*
We give now the main result of the paper. We prove in the below result that the weights $\mathbf{w}_{ppr}^{(1)}$ and $\mathbf{w}_{ppr}^{(2)}$ are equal. Thus, $\hat{t}_{ppr}^{(1)} = \hat{t}_{ppr}^{(2)}$ and one may use a combined strategy: the constrained optimization problem (**P2'**) with the weights given by the constrained optimization problem (**P1'**), namely $\mathbf{w}_{ppr}^{(1)}$.
Result *The ridge weights $\mathbf{w}_{ppr}^{(1)}$ verifying the optimization problem **(P1')** are equal to the weights $\mathbf{w}_{ppr}^{(2)}$ that verify the optimization problem **(P2')**.*
Proof Let $\mathbf{w}_{ppr}^{(1)}$ given by (9) and let $\mathbf{R} = \mathbf{C}^{-1} + \mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s$ with $\mathbf{C}^{-1}$ given by (8). In order to prove the result, we need to compute the inverse of $\mathbf{R}$,

$$\mathbf{R}^{-1} = \begin{pmatrix} \tilde{\mathbf{X}}_{1s}'\mathbf{V}_s^{-1}\tilde{\mathbf{X}}_{1s} & \tilde{\mathbf{X}}_{1s}'\mathbf{V}_s^{-1}\tilde{\mathbf{X}}_{2s} \\ \tilde{\mathbf{X}}_{2s}'\mathbf{V}_s^{-1}\tilde{\mathbf{X}}_{1s} & \mathbf{C}_2^{-1} + \tilde{\mathbf{X}}_{2s}'\mathbf{V}_s^{-1}\tilde{\mathbf{X}}_{2s} \end{pmatrix}^{-1} := \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{L} \end{pmatrix}^{-1}$$

with

$$\begin{aligned} \mathbf{A} &= \tilde{\mathbf{X}}_{1s}'\mathbf{V}_s^{-1}\tilde{\mathbf{X}}_{1s} \\ \mathbf{B} &= \tilde{\mathbf{X}}_{1s}'\mathbf{V}_s^{-1}\tilde{\mathbf{X}}_{2s} \\ \mathbf{L} &= \mathbf{C}_2^{-1} + \tilde{\mathbf{X}}_{2s}'\mathbf{V}_s^{-1}\tilde{\mathbf{X}}_{2s} \end{aligned}$$

Using Lemma 1 from the Appendix, we get

$$\mathbf{R}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F}' & -\mathbf{F}\mathbf{E}^{-1} \\ -\mathbf{E}^{-1}\mathbf{F}' & \mathbf{E}^{-1} \end{pmatrix} \tag{14}$$

where $\mathbf{E} = \mathbf{L} - \mathbf{B}'\mathbf{A}^{-1}\mathbf{B}$ and $\mathbf{F} = \mathbf{A}^{-1}\mathbf{B}$. We may write $\mathbf{A}^{-1} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F}'$ in a shorter form. We have $\boldsymbol{\Omega}_s = \mathbf{V}_s + \tilde{\mathbf{X}}_{2s}\mathbf{C}_2\tilde{\mathbf{X}}_{2s}'$ and its inverse may be computed by using Lemma 2,

$$\boldsymbol{\Omega}_s^{-1} = \mathbf{V}_s^{-1} - \mathbf{V}_s^{-1}\tilde{\mathbf{X}}_{2s}(\mathbf{C}_2^{-1} + \tilde{\mathbf{X}}_{2s}'\mathbf{V}_s^{-1}\tilde{\mathbf{X}}_{2s})^{-1}\tilde{\mathbf{X}}_{2s}'\mathbf{V}_s^{-1}. \tag{15}$$

We multiply $\boldsymbol{\Omega}_s^{-1}$ at left by $\tilde{\mathbf{X}}_{1s}'$ and at right by $\tilde{\mathbf{X}}_{1s}$. We get $\tilde{\mathbf{X}}_{1s}'\boldsymbol{\Omega}_s^{-1}\tilde{\mathbf{X}}_{1s} = \mathbf{A} - \mathbf{B}\mathbf{L}^{-1}\mathbf{B}'$ and its inverse may be determined using again Lemma 2,

$$\begin{aligned} (\tilde{\mathbf{X}}_{1s}'\boldsymbol{\Omega}_s^{-1}\tilde{\mathbf{X}}_{1s})^{-1} &= \mathbf{A}^{-1} + \mathbf{A}^{-1}\mathbf{B}(\mathbf{L} - \mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{B}'\mathbf{A} \\ &= \mathbf{A}^{-1} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F}' \end{aligned} \tag{16}$$

For ease of notation, denote $\mathbf{H} = \tilde{\mathbf{X}}_{1s}'\boldsymbol{\Omega}_s^{-1}\tilde{\mathbf{X}}_{1s}$. The inverse $\mathbf{R}^{-1}$ given in relation (14) becomes

$$\mathbf{R}^{-1} = \begin{pmatrix} \mathbf{H}^{-1} & -\mathbf{F}\mathbf{E}^{-1} \\ -\mathbf{E}^{-1}\mathbf{F}' & \mathbf{E}^{-1} \end{pmatrix} \tag{17}$$

For the above value of $\mathbf{R}^{-1}$, the partially penalized calibrated weights given in expression (9) become,

$$\begin{aligned} \mathbf{w}_{ppr}^{(1)} &= \mathbf{1}_s - \mathbf{V}_s^{-1}\mathbf{X}_s(\mathbf{C}^{-1} + \mathbf{X}_s'\mathbf{V}_s^{-1}\mathbf{X}_s)^{-1}(\mathbf{X}_s'\mathbf{1}_s - \mathbf{X}'\mathbf{1}_U) \\ &= \mathbf{1}_s - \mathbf{V}_s^{-1}(\tilde{\mathbf{X}}_{1s}, \tilde{\mathbf{X}}_{2s}) \begin{pmatrix} \mathbf{H}^{-1} & -\mathbf{F}\mathbf{E}^{-1} \\ -\mathbf{E}^{-1}\mathbf{F}' & \mathbf{E}^{-1} \end{pmatrix} \begin{pmatrix} \tilde{\mathbf{X}}_{1s}'\mathbf{1}_s - \tilde{\mathbf{X}}_1'\mathbf{1}_U \\ \tilde{\mathbf{X}}_{2s}'\mathbf{1}_s - \tilde{\mathbf{X}}_2'\mathbf{1}_U \end{pmatrix} \end{aligned}$$

Consider now the optimization problem **P2**. The weights $\mathbf{w}_{ppr}^{(2)}$ are given in (12),

$$\begin{aligned} &\mathbf{w}_{ppr}^{(2)} \\ &= \boldsymbol{\Omega}_s^{-1}\left[\tilde{\mathbf{X}}_{1s}\mathbf{H}^{-1}\left(\tilde{\mathbf{X}}_1'\mathbf{1}_U - \underbrace{\tilde{\mathbf{X}}_{1s}'\boldsymbol{\Omega}_s^{-1}(\mathbf{V}_s\mathbf{1}_s + \tilde{\mathbf{X}}_{2s}\mathbf{C}_2\tilde{\mathbf{X}}_2'\mathbf{1}_U)}_{(*)}\right) + \underbrace{\mathbf{V}_s\mathbf{1}_s + \tilde{\mathbf{X}}_{2s}\mathbf{C}_2\tilde{\mathbf{X}}_2'\mathbf{1}_U}_{(**)}\right] \end{aligned}$$

We have that $\mathbf{V}_s = \boldsymbol{\Omega}_s - \tilde{\mathbf{X}}_{2s}\mathbf{C}_2\tilde{\mathbf{X}}_{2s}'$. The terms $(*)$ and $(**)$ become

$$(*) = \tilde{\mathbf{X}}_{1s}'\mathbf{1}_s + \tilde{\mathbf{X}}_{1s}'\boldsymbol{\Omega}_s^{-1}\tilde{\mathbf{X}}_{2s}\mathbf{C}_2\left(\tilde{\mathbf{X}}_2'\mathbf{1}_U - \tilde{\mathbf{X}}_{2s}'\mathbf{1}_s\right) \tag{18}$$

$$(**) = \boldsymbol{\Omega}_s\mathbf{1}_s + \tilde{\mathbf{X}}_{2s}\mathbf{C}_2(\tilde{\mathbf{X}}_2'\mathbf{1}_U - \tilde{\mathbf{X}}_{2s}'\mathbf{1}_s). \tag{19}$$

From (18) and (19), we get after some algebra,

$$\begin{aligned} \mathbf{w}_{ppr}^{(2)} &= \mathbf{1}_s + \boldsymbol{\Omega}_s^{-1}\tilde{\mathbf{X}}_{1s}\mathbf{H}^{-1}\left(\tilde{\mathbf{X}}_1'\mathbf{1}_U - \tilde{\mathbf{X}}_{1s}'\mathbf{1}_s\right) \\ &+ (\boldsymbol{\Omega}_s^{-1}\tilde{\mathbf{X}}_{2s}\mathbf{C}_2 - \boldsymbol{\Omega}_s^{-1}\tilde{\mathbf{X}}_{1s}\mathbf{H}^{-1}\tilde{\mathbf{X}}_{1s}'\boldsymbol{\Omega}_s^{-1}\tilde{\mathbf{X}}_{2s}\mathbf{C}_2)\left(\tilde{\mathbf{X}}_2'\mathbf{1}_U - \tilde{\mathbf{X}}_{2s}'\mathbf{1}_s\right). \end{aligned} \tag{20}$$

Furthermore, we obtain from (15) that

$$\boldsymbol{\Omega}_s^{-1}\tilde{\mathbf{X}}_{1s}\mathbf{H}^{-1} = \mathbf{V}_s^{-1}(\tilde{\mathbf{X}}_{1s}\mathbf{H}^{-1} - \tilde{\mathbf{X}}_{2s}\mathbf{L}^{-1}\mathbf{B}'\mathbf{H}^{-1}) \tag{21}$$

and by Lemma 3,

$$\begin{aligned} \boldsymbol{\Omega}_s^{-1}\tilde{\mathbf{X}}_{2s}\mathbf{C}_2 &= \mathbf{V}_s^{-1}\tilde{\mathbf{X}}_{2s}(\mathbf{C}_2^{-1} + \tilde{\mathbf{X}}_{2s}'\mathbf{V}_s^{-1}\tilde{\mathbf{X}}_{2s})^{-1} \\ &= \mathbf{V}_s^{-1}\tilde{\mathbf{X}}_{2s}\mathbf{L}^{-1}. \end{aligned} \tag{22}$$

Hence, relations (21) and (22) yield

$$\boldsymbol{\Omega}_s^{-1}\tilde{\mathbf{X}}_{1s}\mathbf{H}^{-1}\tilde{\mathbf{X}}_{1s}'\boldsymbol{\Omega}_s^{-1}\tilde{\mathbf{X}}_{2s}\mathbf{C}_2 = \mathbf{V}_s^{-1}(\tilde{\mathbf{X}}_{1s}\mathbf{H}^{-1}\mathbf{B}\mathbf{L}^{-1} - \tilde{\mathbf{X}}_{2s}\mathbf{L}^{-1}\mathbf{B}'\mathbf{H}^{-1}\mathbf{B}\mathbf{L}^{-1}). \tag{23}$$

Finally, using relations (21), (22) and (23) in equation (20) yields

$$\begin{aligned}
\mathbf{w}^{(2)}_{ppr} &= \mathbf{1}_s - \mathbf{V}_s^{-1}\Big[\left(\tilde{\mathbf{X}}_{1s}\mathbf{H}^{-1} - \tilde{\mathbf{X}}_{2s}\mathbf{L}^{-1}\mathbf{B}'\mathbf{H}^{-1}\right)\left(\tilde{\mathbf{X}}'_{1s}\mathbf{1}_s - \tilde{\mathbf{X}}'_1\mathbf{1}_U\right) \\
&- \left(-\tilde{\mathbf{X}}_{1s}\mathbf{H}^{-1}\mathbf{B}\mathbf{L}^{-1} + \tilde{\mathbf{X}}_{2s}\left(\mathbf{L}^{-1} + \mathbf{L}^{-1}\mathbf{B}'\mathbf{H}^{-1}\mathbf{B}\mathbf{L}^{-1}\right)\right)\left(\tilde{\mathbf{X}}'_{2s}\mathbf{1}_s - \tilde{\mathbf{X}}'_2\mathbf{1}_U\right)\Big] \\
&= \mathbf{1}_s - \mathbf{V}_s^{-1}(\tilde{\mathbf{X}}_{1s}, \tilde{\mathbf{X}}_{2s})\begin{pmatrix} \mathbf{H}^{-1} & -\mathbf{H}^{-1}\mathbf{B}\mathbf{L}^{-1} \\ -\mathbf{L}^{-1}\mathbf{B}'\mathbf{H}^{-1} & \mathbf{L}^{-1} + \mathbf{L}^{-1}\mathbf{B}'\mathbf{H}^{-1}\mathbf{B}\mathbf{L}^{-1} \end{pmatrix}\begin{pmatrix} \tilde{\mathbf{X}}'_{1s}\mathbf{1}_s - \tilde{\mathbf{X}}'_1\mathbf{1}_U \\ \tilde{\mathbf{X}}'_{2s}\mathbf{1}_s - \tilde{\mathbf{X}}'_2\mathbf{1}_U \end{pmatrix} \\
&= \mathbf{1}_s - \mathbf{V}_s^{-1}(\tilde{\mathbf{X}}_{1s}, \tilde{\mathbf{X}}_{2s})\begin{pmatrix} \mathbf{H}^{-1} & -\mathbf{F}\mathbf{E}^{-1} \\ -\mathbf{E}^{-1}\mathbf{F}' & \mathbf{E}^{-1} \end{pmatrix}\begin{pmatrix} \tilde{\mathbf{X}}'_{1s}\mathbf{1}_s - \tilde{\mathbf{X}}'_1\mathbf{1}_U \\ \tilde{\mathbf{X}}'_{2s}\mathbf{1}_s - \tilde{\mathbf{X}}'_2\mathbf{1}_U \end{pmatrix} \\
&= \mathbf{w}^{(1)}_{ppr}
\end{aligned}$$

since $\mathbf{E}^{-1}\mathbf{F}' = \mathbf{L}^{-1}\mathbf{B}'\mathbf{H}^{-1}$ and $\mathbf{E}^{-1} = \mathbf{L}^{-1} + \mathbf{L}^{-1}\mathbf{B}'\mathbf{H}^{-1}\mathbf{B}\mathbf{L}^{-1}$.

□

Appendix

We give below several lemmas useful for the proof of the result.

Lemma 1 *(Rao, 1965; p.29). Let* $\mathbf{A}$ *and* $\mathbf{L}$ *be symmetric matrices such that the inverses which occur in the below expression exist. Then,*

$$\begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{L} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F}' & -\mathbf{F}\mathbf{E}^{-1} \\ -\mathbf{E}^{-1}\mathbf{F}' & \mathbf{E}^{-1} \end{pmatrix}$$

where $\mathbf{E} = \mathbf{L} - \mathbf{B}'\mathbf{A}^{-1}\mathbf{B}$ *and* $\mathbf{F} = \mathbf{A}^{-1}\mathbf{B}$.

Lemma 2 *(Rao, 1965; p.29) Let* $\mathbf{A}$ *and* $\mathbf{L}$ *be nonsingular matrices of orders* m *and* n *and* $\mathbf{B}$ *be* $m \times n$ *matrix. Then,*

$$(\mathbf{A} + \mathbf{B}\mathbf{L}\mathbf{B}')^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{B}'\mathbf{A}^{-1}\mathbf{B} + \mathbf{L}^{-1})^{-1}\mathbf{B}'\mathbf{A}^{-1}.$$

As a consequence, we get

Lemma 3 *(Henderson and Searle, 1981). Let* $\mathbf{A}$ *and* $\mathbf{L}$ *be nonsingular matrices of orders* m *and* n *and* $\mathbf{B}$ *be* $m \times n$ *matrix. Then,*

$$(\mathbf{A} + \mathbf{B}\mathbf{L}\mathbf{B}')^{-1}\mathbf{B}\mathbf{L} = \mathbf{A}^{-1}\mathbf{B}(\mathbf{L}^{-1} + \mathbf{B}'\mathbf{A}^{-1}\mathbf{B})^{-1}.$$

Bibliography

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

Goga, C. and Shehzad, M. A. (2013), Ridge regression estimators in survey sampling: a review, *to appear in Mathematical Population Studies*.

Guggemos, F. and Tillé, Y. (2010), Penalized calibration in survey sampling: Design-based estimation assisted by mixed models, *Journal of Statistical Planning and Inference*, **140**, 3199-3212.