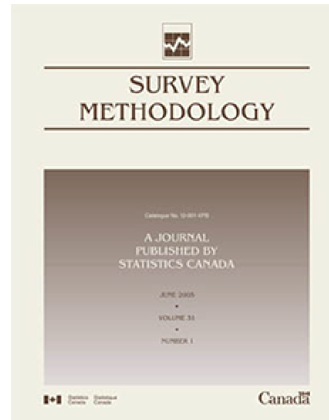
Catalogue no. 12-001-X
ISSN 1492-0921

Survey Methodology

Optimal linear estimation in two-phase sampling



by Takis Merkouris

Release date: December 15, 2022

Statistics Canada Statistique Canada

Canada

How to obtain more information

For information about this product or the wide range of services and data available from Statistics Canada, visit our website, www.statcan.gc.ca.

You can also contact us by

Email at infostats@statcan.gc.ca

Telephone, from Monday to Friday, 8:30 a.m. to 4:30 p.m., at the following numbers:

• Statistical Information Service	1-800-263-1136
• National telecommunications device for the hearing impaired	1-800-363-7629
• Fax line	1-514-283-9350

Standards of service to the public

Statistics Canada is committed to serving its clients in a prompt, reliable and courteous manner. To this end, Statistics Canada has developed standards of service that its employees observe. To obtain a copy of these service standards, please contact Statistics Canada toll-free at 1-800-263-1136. The service standards are also published on www.statcan.gc.ca under "Contact us" > "Standards of service to the public."

Note of appreciation

Canada owes the success of its statistical system to a long-standing partnership between Statistics Canada, the citizens of Canada, its businesses, governments and other institutions. Accurate and timely statistical information could not be produced without their continued co-operation and goodwill.

Published by authority of the Minister responsible for Statistics Canada

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, December 2022
Vol. 48, No. 2, pp. 439-461
Statistics Canada, Catalogue No. 12-001-X

Optimal linear estimation in two-phase sampling

Takis Merkouris[1]

Abstract

Two-phase sampling is a cost effective sampling design employed extensively in surveys. In this paper a method of most efficient linear estimation of totals in two-phase sampling is proposed, which exploits optimally auxiliary survey information. First, a best linear unbiased estimator (BLUE) of any total is formally derived in analytic form, and shown to be also a calibration estimator. Then, a proper reformulation of such a BLUE and estimation of its unknown coefficients leads to the construction of an "optimal" regression estimator, which can also be obtained through a suitable calibration procedure. A distinctive feature of such calibration is the alignment of estimates from the two phases in an one-step procedure involving the combined first-and-second phase samples. Optimal estimation is feasible for certain two-phase designs that are used often in large scale surveys. For general two-phase designs, an alternative calibration procedure gives a generalized regression estimator as an approximate optimal estimator. The proposed general approach to optimal estimation leads to the most effective use of the available auxiliary information in any two-phase survey. The advantages of this approach over existing methods of estimation in two-phase sampling are shown both theoretically and through a simulation study.

Key Words: Auxiliary information; Best linear unbiased estimation; Calibration; Generalized regression estimation; Double sampling.

1. Introduction

The two-phase sampling design, also called double sampling, has traditionally been used in sample surveys as a cost-effective survey method. In the first phase, a relatively large sample is drawn from the target population to provide auxiliary information that is inexpensive to collect. This sample forms a highly informative frame from which a subsample is drawn in the second phase to collect information on the items of interest. Also, two-phase sampling has been increasingly used as a mechanism for handling nonresponse. Särndal, Swensson and Wretman (1992) provide an extensive account of such uses of two-phase sampling. Groves and Heeringa (2006), and Brick and Tourangeau (2017) discuss the important role of two-phase sampling in responsive designs when costly actions are taken for reduction of non-response bias. Other applications of two-phase sampling, which have emerged in recent survey practice, involve various forms of integration of separate surveys. In one such case, a first-phase sample serves as a frame for the second-phase sample for a multitude of similar surveys (Turmelle and Beaucage, 2013). In another case, a primary large survey is used as a frame for another smaller survey with a larger set of survey items (Australian Bureau of Statistics, 2004).

Auxiliary information in two-phase sampling may be available at different levels. Some information is at the level of the whole population, and other information is at the level of the first-phase sample or the second-phase sample. Much research has been devoted to the use of such information for improved estimation of population totals or means; see Särndal et al. (1992), Hidiroglou and Särndal (1998), Hidiroglou (2001), Estevao and Särndal (2002, 2009), Wu and Luan (2003), Chen and Kim (2014), and

1. Takis Merkouris, Department of Statistics, Athens University of Economics and Business, 2 Trias Street, Athens 11362, Greece. E-mail: merkouris@aueb.gr.

references therein. In general, two approaches are identified in this literature for incorporating auxiliary information into the estimation process. The generalized regression approach and the calibration approach; the two phases of sampling imply two regression fits or two successive calibrations. Under certain conditions the two approaches lead to identical estimators, but this is not so in general. Variance estimation of these two-phase estimators has been studied extensively; see, for example, Sitter (1997), Fuller (1998), Kim and Sitter (2003), Kim, Navarro and Fuller (2006), Hidiroglou, Rao and Haziza (2008), Kim and Yu (2011), Beaumont, Beliveau and Haziza (2015).

Irrespective of the regression or calibration formulation of the existing estimation procedures, the resulting estimators for a target variable are in effect linear combinations of Horvitz-Thompson estimators of various totals (or means), including the estimator for the target variable derived from the second-phase sample and estimators for auxiliary variables derived from both first-phase and second-phase sample. Taking a formal approach to optimal estimation, in this paper we consider the most efficient linear combination of available estimators from both phases, based on the principle of best linear unbiased estimation. We show that the derived, in analytic form, best linear unbiased estimator (BLUE) possesses a useful orthogonality property and that it can be alternatively constructed as calibration estimator, linear in the values of the associated variable and incorporating the auxiliary information into the calibrated design weigs. Estimation of the unknown coefficients of this BLUE, using all available auxiliary information from both phases of sampling, gives an “optimal” estimator, analogous to the single-phase optimal regression estimator of Montanari (1987) and Rao (1994). This estimator is a large-sample approximation of the BLUE, with the estimated coefficients minimizing its estimated approximate (large sample) variance, and preserving the orthogonality property of the BLUE. With a proper reformulation of the BLUE, the optimal estimator can also be obtained through a suitable calibration procedure. The distinctive feature of such calibration is the convenient one-step procedure of aligning estimates from the two phases using the combined first-and-second phase samples. Optimal estimation is feasible for certain two-phase designs that are used often in large scale surveys. For general designs, an alternative one-step calibration procedure gives a novel generalized regression estimator as a convenient approximation to the optimal estimator.

The proposed general method of estimation guides the construction of calibration estimators in any particular case of two-phase survey, making the most effective use of the available auxiliary information. It also provides an insig into existing less efficient estimation methods when these are placed into the framework of optimal estimation. The advantages of the proposed method over existing methods are shown both theoretically and through a simulation study.

The paper is organized as follows. The structure of the two-phase sampling design, and notation, are introduced in Section 2. The derivation of the BLUE for the standard type of auxiliary information in two-phase sampling, and its alternative construction as a calibration estimator, are described in Section 3. The two-phase optimal estimator and its calibration equivalent are presented in Section 4. The approximation of the optimal estimator by a generalized regression estimator is discussed in Section 5. Comparisons with existing methods are presented in Section 6. A simulation study is presented in Section 7. The paper concludes with a discussion in Section 8.

2. Two-phase sampling design: Structure and notation

Let $U = \{1, \ldots, k, \ldots, N\}$ denote a finite population of N units. A first-phase sample s_1 of size n_1 is drawn from the population U, using a sampling design that defines inclusion probability $\pi_{1k} = P(k \in s_1)$ for unit $k \in U$, and joint inclusion probability $\pi_{1kl} = P(k, l \in s_1)$ for units $k, l \in U$. Then, a second-phase sample s_2 of size n_2 is drawn from s_1 using a sampling design that defines conditional inclusion probability $\pi_{2k} = P(k \in s_2 | s_1)$ for $k \in s_1$, and joint conditional inclusion probability $\pi_{2kl} = P(k, l \in s_2 | s_1)$ for units $k, l \in s_1$. Assuming that $\pi_{1k} > 0$ for all $k \in U$ and $\pi_{2k} > 0$ for all $k \in s_1$, the first-phase design weig for $k \in s_1$ is $w_{1k} = 1/\pi_{1k}$, the conditional second-phase design weig for $k \in s_2$ is $w_{2k} = 1/\pi_{2k}$, and the overall design weig for $k \in s_2$ is $w_k = w_{1k} w_{2k}$.

The standard type of auxiliary variables in two-phase sampling (see, for example, Särndal et al. (1992)) involves a vector of auxiliary variables $\mathbf{x}$, partitioned as $\mathbf{x} = (\mathbf{x}_1', \mathbf{x}_2')'$ by p and q components of it, with population total $\mathbf{t}_{\mathbf{x}} = \sum_U \mathbf{x}_k$ and known total $\mathbf{t}_{\mathbf{x}_1} = \sum_U \mathbf{x}_{1k}$ of $\mathbf{x}_1$. The value $\mathbf{x}_k$ is observed for every unit $k \in s_1$, whereas for a d -dimensional vector of target variables $\mathbf{y}$, with total $\mathbf{t}_{\mathbf{y}} = \sum_U \mathbf{y}_k$, the value $\mathbf{y}_k$ is observed only for the units $k \in s_2$. In some surveys, components of the vector $\mathbf{x}_2$ are also target variables. An unbiased estimator of the total $\mathbf{t}_{\mathbf{y}}$, the common Horvitz-Thompson (HT) estimator, given by $\tilde{\mathbf{t}}_{\mathbf{y}} = \sum_{s_2} w_k \mathbf{y}_k$, is obtained using the second-phase sample s_2, while two HT estimators of the total $\mathbf{t}_{\mathbf{x}}$, given by $\hat{\mathbf{t}}_{\mathbf{x}} = \sum_{s_1} w_{1k} \mathbf{x}_k$ and $\tilde{\mathbf{t}}_{\mathbf{x}} = \sum_{s_2} w_k \mathbf{x}_k$, are obtained using the samples s_1 and s_2, respectively. In the derivation of results involving these estimators we will use the vector notation $\tilde{\mathbf{t}}_{\mathbf{y}} = \mathbf{Y}_2' \mathbf{w}$, $\hat{\mathbf{t}}_{\mathbf{x}} = \mathbf{X}_1' \mathbf{w}_1$, $\tilde{\mathbf{t}}_{\mathbf{x}} = \mathbf{X}_2' \mathbf{w}$, $\hat{\mathbf{t}}_{\mathbf{x}_1} = \mathbf{X}_{11}' \mathbf{w}_1$, where $\mathbf{w}_1$ and $\mathbf{w}$ denote the vectors of design weigs for samples s_1 and s_2, respectively, $\mathbf{X}_1$ and $\mathbf{X}_{11}$ denote the sample s_1 matrices of $\mathbf{x}$ and $\mathbf{x}_1$ of dimensions $n_1 \times (p + q)$ and $n_1 \times p$, respectively, and $\mathbf{Y}_2$, $\mathbf{X}_2$ denote the sample s_2 matrices of $\mathbf{y}$ and $\mathbf{x}$ of dimensions $n_2 \times d$ and $n_2 \times (p + q)$.

The primary target of estimation is the total $\mathbf{t}_{\mathbf{y}}$. However, for better understanding of the construction of the proposed estimators, and because components of the vector $\mathbf{x}_2$ may also be target variables, a unified approach to the estimation of both $\mathbf{t}_{\mathbf{y}}$ and $\mathbf{t}_{\mathbf{x}}$ will be taken.

3. Best linear unbiased estimation in two-phase sampling

3.1 An analytic form of the best linear unbiased estimator

For more efficient estimation of the totals $\mathbf{t}_{\mathbf{y}}$ and $\mathbf{t}_{\mathbf{x}}$, incorporating all the available information from both phases through the correlation of $\mathbf{y}$ and $\mathbf{x}$, we consider the best linear unbiased estimators (BLUE), denoted by $\hat{\mathbf{t}}_{\mathbf{y}}^B$ and $\hat{\mathbf{t}}_{\mathbf{x}}^B$, which are minimum-variance linear unbiased combinations of the four estimators $\tilde{\mathbf{t}}_{\mathbf{y}}$, $\hat{\mathbf{t}}_{\mathbf{x}}$, $\tilde{\mathbf{t}}_{\mathbf{x}}$, $\mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1}$ and given in matrix form by

$$\left(\hat{\mathbf{t}}_{\mathbf{y}}^{B'}, \hat{\mathbf{t}}_{\mathbf{x}}^{B'}\right)' = \boldsymbol{\mathcal{P}}\left(\tilde{\mathbf{t}}'_{\mathbf{y}}, \hat{\mathbf{t}}'_{\mathbf{x}}, \tilde{\mathbf{t}}'_{\mathbf{x}}, \mathbf{t}'_{\mathbf{x}_1} - \hat{\mathbf{t}}'_{\mathbf{x}_1}\right)', \tag{3.1}$$

where $\boldsymbol{\mathcal{P}} = (\mathbf{W}'\mathbf{V}^{-1}\mathbf{W})^{-1}\mathbf{W}'\mathbf{V}^{-1}$, the matrix $\mathbf{W}$ has entries 1's and 0's and satisfies $E\left[(\tilde{\mathbf{t}}'_{\mathbf{y}}, \hat{\mathbf{t}}'_{\mathbf{x}}, \tilde{\mathbf{t}}'_{\mathbf{x}}, \mathbf{t}'_{\mathbf{x}_1} - \hat{\mathbf{t}}'_{\mathbf{x}_1})'\right] = \mathbf{W}(\mathbf{t}'_{\mathbf{y}}, \mathbf{t}'_{\mathbf{x}})'$, and $\mathbf{V}$ is the covariance matrix of $(\tilde{\mathbf{t}}'_{\mathbf{y}}, \hat{\mathbf{t}}'_{\mathbf{x}}, \tilde{\mathbf{t}}'_{\mathbf{x}}, \mathbf{t}'_{\mathbf{x}_1} - \hat{\mathbf{t}}'_{\mathbf{x}_1})'$. It follows that $\mathrm{Var}\left[(\hat{\mathbf{t}}_{\mathbf{y}}^{B'}, \hat{\mathbf{t}}_{\mathbf{x}}^{B'})'\right] = (\mathbf{W}'\mathbf{V}^{-1}\mathbf{W})^{-1}$. This typical formulation of best linear unbiased estimation has been explored in two other areas of survey sampling; see Wolter (1979), Jones (1980), Fuller (1990), and Chipperfield and Steel (2009). In the present context, a more practical formulation, which leads also to the representation of the BLUE as a calibration estimator, is as follows.

Writing the two linear combinations in (3.1) in expanded form and using the condition of unbiasedness $E(\hat{\mathbf{t}}_{\mathbf{y}}^{B}) = \mathbf{t}_{\mathbf{y}}$ and $E(\hat{\mathbf{t}}_{\mathbf{x}}^{B}) = \mathbf{t}_{\mathbf{x}}$, it is easy to show that the matrix $\boldsymbol{\mathcal{P}}$ of the coefficients in these linear combinations satisfies

$$\boldsymbol{\mathcal{P}} = \begin{pmatrix} \mathbf{B}_{1\mathbf{y}} & \mathbf{B}_{2\mathbf{y}} & \mathbf{B}_{3\mathbf{y}} & \mathbf{B}_{4\mathbf{y}} \\ \mathbf{B}_{1\mathbf{x}} & \mathbf{B}_{2\mathbf{x}} & \mathbf{B}_{3\mathbf{x}} & \mathbf{B}_{4\mathbf{x}} \end{pmatrix} = \begin{pmatrix} \mathbf{I} & \mathbf{B}_{2\mathbf{y}} & -\mathbf{B}_{2\mathbf{y}} & \mathbf{B}_{4\mathbf{y}} \\ \mathbf{0} & \mathbf{B}_{2\mathbf{x}} & \mathbf{I} - \mathbf{B}_{2\mathbf{x}} & \mathbf{B}_{4\mathbf{x}} \end{pmatrix},$$

and then the two components of the BLUE in (3.1) are written in the regression form

$$\begin{aligned} \hat{\mathbf{t}}_{\mathbf{y}}^{B} &= \tilde{\mathbf{t}}_{\mathbf{y}} + \mathbf{B}_{2\mathbf{y}}(\hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}) + \mathbf{B}_{4\mathbf{y}}(\mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1}) \\ \hat{\mathbf{t}}_{\mathbf{x}}^{B} &= \tilde{\mathbf{t}}_{\mathbf{x}} + \mathbf{B}_{2\mathbf{x}}(\hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}) + \mathbf{B}_{4\mathbf{x}}(\mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1}). \end{aligned} \tag{3.2}$$

Now we can write (3.1) as

$$\begin{pmatrix} \hat{\mathbf{t}}_{\mathbf{y}}^{B} \\ \hat{\mathbf{t}}_{\mathbf{x}}^{B} \end{pmatrix} = \begin{pmatrix} \tilde{\mathbf{t}}_{\mathbf{y}} \\ \tilde{\mathbf{t}}_{\mathbf{x}} \end{pmatrix} + \boldsymbol{\mathcal{B}} \begin{pmatrix} \hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}} \\ \mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1} \end{pmatrix}, \tag{3.3}$$

where the matrix $\boldsymbol{\mathcal{B}}$ consists of the second and fourth columns of $\boldsymbol{\mathcal{P}}$, and has the easily derived variance-minimizing value

$$\boldsymbol{\mathcal{B}} = -\mathrm{Cov}\left[\begin{pmatrix} \tilde{\mathbf{t}}_{\mathbf{y}} \\ \tilde{\mathbf{t}}_{\mathbf{x}} \end{pmatrix}, \begin{pmatrix} \hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}} \\ \mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1} \end{pmatrix}\right] \left[\mathrm{Var}\begin{pmatrix} \hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}} \\ \mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1} \end{pmatrix}\right]^{-1}. \tag{3.4}$$

Next write

$$\mathbf{w}^* = \begin{pmatrix} \mathbf{w}_1 \\ \mathbf{w} \end{pmatrix}, \quad \boldsymbol{\mathcal{X}} = \begin{pmatrix} -\mathbf{X}_1 & \mathbf{X}_{11} \\ \mathbf{X}_2 & \mathbf{0} \end{pmatrix}, \quad \boldsymbol{\Psi} = \begin{pmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{Y}_2 & \mathbf{X}_2 \end{pmatrix}, \tag{3.5}$$

so that

$$\boldsymbol{\mathcal{X}}'\mathbf{w}^* = \begin{pmatrix} \tilde{\mathbf{t}}_{\mathbf{x}} - \hat{\mathbf{t}}_{\mathbf{x}} \\ \hat{\mathbf{t}}_{\mathbf{x}_1} \end{pmatrix}, \qquad \boldsymbol{\Psi}'\mathbf{w}^* = \begin{pmatrix} \tilde{\mathbf{t}}_{\mathbf{y}} \\ \tilde{\mathbf{t}}_{\mathbf{x}} \end{pmatrix}, \tag{3.6}$$

and $\mathcal{B}$ may then be expressed as $\boldsymbol{\mathcal{B}} = \mathrm{Cov}(\boldsymbol{\Psi}'\mathbf{w}^*, \boldsymbol{\mathcal{X}}'\mathbf{w}^*)\left[\mathrm{Var}(\boldsymbol{\mathcal{X}}'\mathbf{w}^*)\right]^{-1}$. For the calculation of variances and covariances we define $\mathbf{w}^*$ at the population level as $\mathbf{w}_U^* = (\mathbf{w}'_{1U}, \mathbf{w}'_U)'$, where the k^{th} element of $\mathbf{w}_{1U}$ is $w_{1U_k} = (1/\pi_{1k})\, I_{1k}$, the indicator variable I_1 denoting inclusion of a population unit in

s_1, and the k^{th} element of $\mathbf{w}_U$ is $w_{U_k} = [1/(\pi_{1k}\pi_{2k})] I_{1k} I_{2k}$, the indicator variable I_2 denoting inclusion of a population unit in s_2 conditional on the selection of sample s_1. We may now write $\boldsymbol{\mathcal{X}}'\mathbf{w}^* = \boldsymbol{\mathcal{X}}_U' \mathbf{w}_U^*$ and $\boldsymbol{\Psi}'\mathbf{w}^* = \boldsymbol{\Psi}_U' \mathbf{w}_U^*$, where $\boldsymbol{\mathcal{X}}_U$ and $\boldsymbol{\Psi}_U$ are the population counterparts of $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\Psi}$, respectively; all submatrices in $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\Psi}$ are expanded to population level, having N rows. Then, denoting $\hat{\mathbf{t}}_{\boldsymbol{\Psi}} = \boldsymbol{\Psi}'\mathbf{w}^*$ and $\hat{\mathbf{t}}_{\boldsymbol{\mathcal{X}}} = \boldsymbol{\mathcal{X}}'\mathbf{w}^*$, we get

$$\boldsymbol{\mathcal{B}} = \operatorname{Cov}(\hat{\mathbf{t}}_{\boldsymbol{\Psi}}, \hat{\mathbf{t}}_{\boldsymbol{\mathcal{X}}}) \left[\operatorname{Var}(\hat{\mathbf{t}}_{\boldsymbol{\mathcal{X}}})\right]^{-1} = \boldsymbol{\Psi}_U' \operatorname{Var}(\mathbf{w}_U^*) \boldsymbol{\mathcal{X}}_U \left[\boldsymbol{\mathcal{X}}_U' \operatorname{Var}(\mathbf{w}_U^*) \boldsymbol{\mathcal{X}}_U\right]^{-1}. \tag{3.7}$$

A useful more analytic expression of $\boldsymbol{\mathcal{B}}$ is then obtained using the following Lemma; the proof is in the Appendix.

Lemma 1

$$\operatorname{Var}(\mathbf{w}_U^*) = \begin{pmatrix} \operatorname{Var}(\mathbf{w}_{1U}) & \operatorname{Var}(\mathbf{w}_{1U}) \\ \operatorname{Var}(\mathbf{w}_{1U}) & \operatorname{Var}(\mathbf{w}_U) \end{pmatrix}, \tag{3.8}$$

where $\operatorname{Var}(\mathbf{w}_{1U}) = \{(\pi_{1kl} - \pi_{1k}\pi_{1l}) / \pi_{1k}\pi_{1l}\}$, $\operatorname{Var}(\mathbf{w}_U) = \{(\pi_{1kl}\pi_{2kl} - \pi_{1k}\pi_{2k}\pi_{1l}\pi_{2l}) / \pi_{1k}\pi_{2k}\pi_{1l}\pi_{2l}\}$.

Using (3.7) and (3.8), it is easy to show that (3.4) is expressed as

$$\boldsymbol{\mathcal{B}} = \begin{bmatrix} -\operatorname{Cov}(\tilde{\mathbf{t}}_{\mathbf{y}}, \hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}) \left[\operatorname{Var}(\hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}})\right]^{-1} & \operatorname{Cov}(\tilde{\mathbf{t}}_{\mathbf{y}}, \hat{\mathbf{t}}_{\mathbf{x}_1}) \left[\operatorname{Var}(\hat{\mathbf{t}}_{\mathbf{x}_1})\right]^{-1} \\ \mathbf{I} & \operatorname{Cov}(\tilde{\mathbf{t}}_{\mathbf{x}}, \hat{\mathbf{t}}_{\mathbf{x}_1}) \left[\operatorname{Var}(\hat{\mathbf{t}}_{\mathbf{x}_1})\right]^{-1} \end{bmatrix}. \tag{3.9}$$

Implicit in this representation of $\boldsymbol{\mathcal{B}}$ is the property $\operatorname{Cov}(\tilde{\mathbf{t}}_{\mathbf{x}}, \hat{\mathbf{t}}_{\mathbf{x}}) = \operatorname{Var}(\hat{\mathbf{t}}_{\mathbf{x}})$, following from (3.8), implying that $\operatorname{Var}(\hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}) = \operatorname{Var}(\tilde{\mathbf{t}}_{\mathbf{x}}) - \operatorname{Var}(\hat{\mathbf{t}}_{\mathbf{x}})$, and the property $\operatorname{Cov}(\tilde{\mathbf{t}}_{\mathbf{x}}, \hat{\mathbf{t}}_{\mathbf{x}_1}) = \operatorname{Cov}(\hat{\mathbf{t}}_{\mathbf{x}}, \hat{\mathbf{t}}_{\mathbf{x}_1})$, implying $\operatorname{Cov}(\hat{\mathbf{t}}_{\mathbf{x}_1}, \hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}) = \mathbf{0}$ (this covariance being the off-diagonal block of $\boldsymbol{\mathcal{X}}_U' \operatorname{Var}(\mathbf{w}_U^*) \boldsymbol{\mathcal{X}}_U$). Then (3.2) can be written explicitly as

$$\begin{aligned} \hat{\mathbf{t}}_{\mathbf{y}}^B &= \tilde{\mathbf{t}}_{\mathbf{y}} - \operatorname{Cov}(\tilde{\mathbf{t}}_{\mathbf{y}}, \hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}) \left[\operatorname{Var}(\hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}})\right]^{-1} (\hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}) \\ &\quad + \operatorname{Cov}(\tilde{\mathbf{t}}_{\mathbf{y}}, \hat{\mathbf{t}}_{\mathbf{x}_1}) \left[\operatorname{Var}(\hat{\mathbf{t}}_{\mathbf{x}_1})\right]^{-1} (\mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1}) \\ \hat{\mathbf{t}}_{\mathbf{x}}^B &= \hat{\mathbf{t}}_{\mathbf{x}} + \operatorname{Cov}(\hat{\mathbf{t}}_{\mathbf{x}}, \hat{\mathbf{t}}_{\mathbf{x}_1}) \left[\operatorname{Var}(\hat{\mathbf{t}}_{\mathbf{x}_1})\right]^{-1} (\mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1}). \end{aligned} \tag{3.10}$$

In view of the property $\operatorname{Cov}(\hat{\mathbf{t}}_{\mathbf{x}_1}, \hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}) = \mathbf{0}$, it follows immediately that

$$\begin{aligned} \operatorname{Var}(\hat{\mathbf{t}}_{\mathbf{y}}^B) &= \operatorname{Var}(\tilde{\mathbf{t}}_{\mathbf{y}}) - \operatorname{Cov}(\tilde{\mathbf{t}}_{\mathbf{y}}, \hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}) \left[\operatorname{Var}(\hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}})\right]^{-1} \operatorname{Cov}'(\tilde{\mathbf{t}}_{\mathbf{y}}, \hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}) \\ &\quad - \operatorname{Cov}(\tilde{\mathbf{t}}_{\mathbf{y}}, \hat{\mathbf{t}}_{\mathbf{x}_1}) \left[\operatorname{Var}(\hat{\mathbf{t}}_{\mathbf{x}_1})\right]^{-1} \operatorname{Cov}'(\tilde{\mathbf{t}}_{\mathbf{y}}, \hat{\mathbf{t}}_{\mathbf{x}_1}) \\ \operatorname{Var}(\hat{\mathbf{t}}_{\mathbf{x}}^B) &= \operatorname{Var}(\hat{\mathbf{t}}_{\mathbf{x}}) - \operatorname{Cov}(\hat{\mathbf{t}}_{\mathbf{x}}, \hat{\mathbf{t}}_{\mathbf{x}_1}) \left[\operatorname{Var}(\hat{\mathbf{t}}_{\mathbf{x}_1})\right]^{-1} \operatorname{Cov}'(\hat{\mathbf{t}}_{\mathbf{x}}, \hat{\mathbf{t}}_{\mathbf{x}_1}). \end{aligned} \tag{3.11}$$

Remark 3.1. Every component or linear combination of components of $\hat{\mathbf{t}}_{\mathbf{y}}^{B}$ is BLUE for the corresponding total. Also, as evident from (3.11), the efficiency of $\hat{\mathbf{t}}_{\mathbf{y}}^{B}$, relative to $\tilde{\mathbf{t}}_{\mathbf{y}}$, depends on the strength of correlation of $\mathbf{y}$ with $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2)$, as well as on the difference in sample size (and possibly in sampling design) for the samples s_1 and s_2.

Remark 3.2. Because of the orthogonality property $\mathrm{Cov}(\hat{\mathbf{t}}_{\mathbf{x}_1}, \hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}) = \mathbf{0}$, the coefficient of any of the terms $\hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}$ and $\mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1}$ in (3.10) would not change if the other one would be set equal to $\mathbf{0}$ in (3.2). For instance, the BLUE for $\mathbf{t}_{\mathbf{y}}$ based on $(\tilde{\mathbf{t}}_{\mathbf{y}}, \hat{\mathbf{t}}_{\mathbf{x}}, \tilde{\mathbf{t}}_{\mathbf{x}})$ would be $\hat{\mathbf{t}}_{\mathbf{y}}^{B}$ as in (3.10) but without the last term. This is easily worked out as a special case of the full setup $(\tilde{\mathbf{t}}_{\mathbf{y}}, \hat{\mathbf{t}}_{\mathbf{x}}, \tilde{\mathbf{t}}_{\mathbf{x}}, \mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1})$. This orthogonality property explains the additive reduction of variance noticed in the first equation of (3.11).

Remark 3.3. The BLUE $\hat{\mathbf{t}}_{\mathbf{x}}^{B}$ in (3.10) can also be produced using the reduced setup $(\hat{\mathbf{t}}_{\mathbf{x}}, \mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1})$ in (3.1). The same best linear estimator, for a single target variable, has been derived differently in the context of general single-phase sampling by Fuller and Isaki (1981) and Montanari (1987). In particular, for the auxiliary variable $\mathbf{x}_1$ we have $\hat{\mathbf{t}}_{\mathbf{x}_1}^{B} = \mathbf{t}_{\mathbf{x}_1}$. Next, it can be easily verified that the BLUE in (3.1) can be alternatively derived in two steps of best linear unbiased estimation using the setup $(\tilde{\mathbf{t}}_{\mathbf{y}}^{B}, \hat{\mathbf{t}}_{\mathbf{x}}^{B}, \tilde{\mathbf{t}}_{\mathbf{x}}^{B})$, where $\tilde{\mathbf{t}}_{\mathbf{y}}^{B} = \tilde{\mathbf{t}}_{\mathbf{y}} + \mathrm{Cov}(\tilde{\mathbf{t}}_{\mathbf{y}}, \hat{\mathbf{t}}_{\mathbf{x}_1}) \left[\mathrm{Var}(\hat{\mathbf{t}}_{\mathbf{x}_1})\right]^{-1} (\mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1})$ and $\tilde{\mathbf{t}}_{\mathbf{x}}^{B} = \tilde{\mathbf{t}}_{\mathbf{x}} + \mathrm{Cov}(\tilde{\mathbf{t}}_{\mathbf{x}}, \hat{\mathbf{t}}_{\mathbf{x}_1}) \left[\mathrm{Var}(\hat{\mathbf{t}}_{\mathbf{x}_1})\right]^{-1} (\mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1})$ are the BLUEs generated by the one-phase setups $(\tilde{\mathbf{t}}_{\mathbf{y}}, \mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1})$ and $(\tilde{\mathbf{t}}_{\mathbf{x}}, \mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1})$, respectively. It can be shown through tedious algebra, that another, more explicit, BLUE setup that is equivalent to that in (3.1) is $(\tilde{\mathbf{t}}_{\mathbf{y}}, \hat{\mathbf{t}}_{\mathbf{x}_2}, \tilde{\mathbf{t}}_{\mathbf{x}_2}, \mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1}, \mathbf{t}_{\mathbf{x}_1} - \tilde{\mathbf{t}}_{\mathbf{x}_1})$. This attests that the compact setup in (3.1) provides the most efficient linear estimation of $\mathbf{t}_{\mathbf{y}}$ and $\mathbf{t}_{\mathbf{x}}$ using all available relevant estimates.

3.2 The two-phase BLUE as calibration estimator

Using the notation leading to (3.7), and setting $\hat{\mathbf{t}}_{\boldsymbol{\Psi}}^{B} = (\hat{\mathbf{t}}_{\mathbf{y}}^{B'}, \hat{\mathbf{t}}_{\mathbf{x}}^{B'})'$ and $\boldsymbol{\Delta} = \mathrm{Var}(\mathbf{w}_U^{*})$, we may express the BLUE in (3.3) as $\hat{\mathbf{t}}_{\boldsymbol{\Psi}}^{B} = \hat{\mathbf{t}}_{\boldsymbol{\Psi}} + \boldsymbol{\mathcal{B}}(\mathbf{t}_{\mathcal{X}} - \hat{\mathbf{t}}_{\mathcal{X}})$, where $\boldsymbol{\mathcal{B}} = \boldsymbol{\Psi}_U' \, \boldsymbol{\Delta} \boldsymbol{\mathcal{X}}_U \, (\boldsymbol{\mathcal{X}}_U' \, \boldsymbol{\Delta} \boldsymbol{\mathcal{X}}_U)^{-1}$ and $\mathbf{t}_{\mathcal{X}} = (\mathbf{0}', \mathbf{t}_{\mathbf{x}_1}')'$, or in the more suggestive form

$$\hat{\mathbf{t}}_{\boldsymbol{\Psi}}^{B} = \boldsymbol{\Psi}_U' \left[\mathbf{w}_U^{*} + \boldsymbol{\Delta} \boldsymbol{\mathcal{X}}_U \, (\boldsymbol{\mathcal{X}}_U' \, \boldsymbol{\Delta} \boldsymbol{\mathcal{X}}_U)^{-1} \left(\mathbf{t}_{\mathcal{X}} - \boldsymbol{\mathcal{X}}_U' \, \mathbf{w}_U^{*} \right) \right]. \tag{3.12}$$

It appears from (3.12) that $\hat{\mathbf{t}}_{\boldsymbol{\Psi}}^{B}$ has the form of a calibration estimator, with population vector of calibrated weigs $\mathbf{c}_U^{*} = \mathbf{w}_U^{*} + \boldsymbol{\Delta} \boldsymbol{\mathcal{X}}_U \, (\boldsymbol{\mathcal{X}}_U' \, \boldsymbol{\Delta} \boldsymbol{\mathcal{X}}_U)^{-1} \left(\mathbf{t}_{\mathcal{X}} - \boldsymbol{\mathcal{X}}_U' \, \mathbf{w}_U^{*} \right)$ and vector of calibration totals $\mathbf{t}_{\mathcal{X}}$. This is formalized in the following theorem; the proof is in the Appendix.

Theorem 1. *The vector* $\mathbf{c}_U^{*} = \mathbf{w}_U^{*} + \boldsymbol{\Delta} \boldsymbol{\mathcal{X}}_U \, (\boldsymbol{\mathcal{X}}_U' \, \boldsymbol{\Delta} \boldsymbol{\mathcal{X}}_U)^{-1} \left(\mathbf{t}_{\mathcal{X}} - \boldsymbol{\mathcal{X}}_U' \, \mathbf{w}_U^{*} \right)$ *minimizes the generalized least-squares distance* $(\mathbf{c}_U^{*} - \mathbf{w}_U^{*})' \boldsymbol{\Delta}^{-1} (\mathbf{c}_U^{*} - \mathbf{w}_U^{*})$ *subject to the constraints* $\boldsymbol{\mathcal{X}}_U' \, \mathbf{c}_U^{*} = \mathbf{t}_{\mathcal{X}}$, *i.e.,* $\mathbf{X}_U' \, \mathbf{c}_{1U} = \mathbf{X}_U' \, \mathbf{c}_U$ *and* $\mathbf{X}_{1U}' \, \mathbf{c}_{1U} = \mathbf{t}_{\mathbf{x}_1}$, *where* $(\mathbf{c}_{1U}, \mathbf{c}_U)$ *corresponds to* $(\mathbf{w}_{1U}, \mathbf{w}_U)$.

Theorem 1 states that best linear unbiased estimation using the setup $(\tilde{\mathbf{t}}_{\mathbf{y}}, \hat{\mathbf{t}}_{\mathbf{x}}, \tilde{\mathbf{t}}_{\mathbf{x}}, \mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1})$ is essentially a calibration procedure whereby the two estimates $\hat{\mathbf{t}}_{\mathbf{x}}$ and $\tilde{\mathbf{t}}_{\mathbf{x}}$ of $\mathbf{t}_{\mathbf{x}}$ are calibrated to each other, i.e., they are aligned, and the estimate $\hat{\mathbf{t}}_{\mathbf{x}_1}$ is calibrated to the total $\mathbf{t}_{\mathbf{x}_1}$. We may now write formally the BLUE $\hat{\mathbf{t}}_{\boldsymbol{\Psi}}^{B}$

as a calibration estimator $\hat{\mathbf{t}}_{\boldsymbol{\Psi}}^{B} = \boldsymbol{\Psi}_U' \, \mathbf{c}_U^*$, with its two components given in the simple linear forms $\hat{\mathbf{t}}_{\mathbf{y}}^{B} = \mathbf{Y}_U' \, \mathbf{c}_U$ and $\hat{\mathbf{t}}_{\mathbf{x}}^{B} = \mathbf{X}_U' \, \mathbf{c}_U$.

The alternative two-step construction of the BLUE noted in Remark 3.3 above can also be carried out through a two-step calibration procedure involving $\mathbf{w}_U^*$ in both steps. Indeed, partitioning $\boldsymbol{\mathcal{X}}_U$ by its two column submatrices as $\boldsymbol{\mathcal{X}}_U = (\boldsymbol{\mathcal{X}}_{12U}, \boldsymbol{\mathcal{X}}_{1U})$, and noting that $\boldsymbol{\mathcal{X}}_{12U}' \, \boldsymbol{\Delta} \boldsymbol{\mathcal{X}}_{1U} = \mathrm{Cov}(\hat{\mathbf{t}}_{\mathbf{x}_1}, \hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}) = \mathbf{0}$, it is easy to decompose the vector $\mathbf{c}_U^*$ as

$$\begin{aligned} \mathbf{c}_U^* &= \mathbf{w}_U^* + \boldsymbol{\Delta} \boldsymbol{\mathcal{X}}_{12U} \, (\boldsymbol{\mathcal{X}}_{12U}' \, \boldsymbol{\Delta} \boldsymbol{\mathcal{X}}_{12U})^{-1} \left(\mathbf{0} - \boldsymbol{\mathcal{X}}_{12U}' \mathbf{w}_U^* \right) \\ &\quad + \boldsymbol{\Delta} \boldsymbol{\mathcal{X}}_{1U} \, (\boldsymbol{\mathcal{X}}_{1U}' \, \boldsymbol{\Delta} \boldsymbol{\mathcal{X}}_{1U})^{-1} \left(\mathbf{t}_{\mathbf{x}_1} - \boldsymbol{\mathcal{X}}_{1U}' \mathbf{w}_U^* \right). \end{aligned} \tag{3.13}$$

In the rig hand side of (3.13), the sum of the first and second terms results from calibration with constraint $\boldsymbol{\mathcal{X}}_{12U}' \, \mathbf{c}_U^* = \mathbf{X}_U' \, \mathbf{c}_{1U} - \mathbf{X}_U' \, \mathbf{c}_U = \mathbf{0}$ only, while the sum of the first and third terms results from calibration with constraint $\boldsymbol{\mathcal{X}}_{1U}' \, \mathbf{c}_U^* = \mathbf{t}_{\mathbf{x}_1}$ only.

Now setting $\boldsymbol{\Delta}_1 = \mathrm{Var}(\mathbf{w}_{1U})$ and $\boldsymbol{\Delta}_2 = \mathrm{Var}(\mathbf{w}_U)$, these variances being specified by (3.8), it follows easily from (3.13) that the optimal calibration estimators $\hat{\mathbf{t}}_{\mathbf{y}}^{B}$ and $\hat{\mathbf{t}}_{\mathbf{x}}^{B}$ in (3.10) can be written in the explicit form, which will be recalled later,

$$\begin{aligned} \hat{\mathbf{t}}_{\mathbf{y}}^{B} &= \tilde{\mathbf{t}}_{\mathbf{y}} + \left[\mathbf{Y}_U' \, \boldsymbol{\Delta}_2 \, \mathbf{X}_U - \mathbf{Y}_U' \, \boldsymbol{\Delta}_1 \, \mathbf{X}_U \right] \left[\mathbf{X}_U' \, \boldsymbol{\Delta}_2 \, \mathbf{X}_U - \mathbf{X}_U' \, \boldsymbol{\Delta}_1 \, \mathbf{X}_U \right]^{-1} (\hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}) \\ &\quad + \mathbf{Y}_U' \, \boldsymbol{\Delta}_1 \, \mathbf{X}_{1U} \, (\mathbf{X}_{1U}' \, \boldsymbol{\Delta}_1 \, \mathbf{X}_{1U})^{-1} (\mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1}) \\ \hat{\mathbf{t}}_{\mathbf{x}}^{B} &= \hat{\mathbf{t}}_{\mathbf{x}} + \mathbf{X}_U' \, \boldsymbol{\Delta}_1 \, \mathbf{X}_{1U} \, (\mathbf{X}_{1U}' \, \boldsymbol{\Delta}_1 \, \mathbf{X}_{1U})^{-1} (\mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1}). \end{aligned} \tag{3.14}$$

4. Optimal linear estimation in two-phase sampling

4.1 The two-phase optimal estimator

The matrix $\boldsymbol{\mathcal{B}}$ in (3.7) comprises variances and covariances which need to be estimated. In view of $\mathrm{Var}(\hat{\mathbf{t}}_{\boldsymbol{\mathcal{X}}}) = \boldsymbol{\mathcal{X}}_U' \, \boldsymbol{\Delta} \boldsymbol{\mathcal{X}}_U$ and $\mathrm{Cov}(\hat{\mathbf{t}}_{\boldsymbol{\Psi}}, \hat{\mathbf{t}}_{\boldsymbol{\mathcal{X}}}) = \boldsymbol{\Psi}_U' \, \boldsymbol{\Delta} \boldsymbol{\mathcal{X}}_U$, and recalling (3.8), the obvious unbiased estimates are $\widehat{\mathrm{Var}}(\hat{\mathbf{t}}_{\boldsymbol{\mathcal{X}}}) = \boldsymbol{\mathcal{X}}' \, \hat{\boldsymbol{\Delta}} \boldsymbol{\mathcal{X}}$ and $\widehat{\mathrm{Cov}}(\hat{\mathbf{t}}_{\boldsymbol{\Psi}}, \hat{\mathbf{t}}_{\boldsymbol{\mathcal{X}}}) = \boldsymbol{\Psi}' \, \hat{\boldsymbol{\Delta}} \boldsymbol{\mathcal{X}}$, where the $(n_1 + n_2) \times (n_1 + n_2)$ matrix $\hat{\boldsymbol{\Delta}} = \widehat{\mathrm{Var}}(\mathbf{w}_U^*)$ has diagonal blocks $\hat{\boldsymbol{\Delta}}_1 = \left\{ (\pi_{1kl} - \pi_{1k}\pi_{1l}) \,/\, \pi_{1k}\pi_{1l}\pi_{1kl} \right\}$, $\hat{\boldsymbol{\Delta}}_2 = \left\{ (\pi_{1kl}\pi_{2kl} - \pi_{1k}\pi_{2k}\pi_{1l}\pi_{2l}) \,/\, \pi_{1k}\pi_{2k}\pi_{1l}\pi_{2l}\pi_{1kl}\pi_{2kl} \right\}$, and off-diagonal blocks $\hat{\boldsymbol{\Delta}}_{12}$, $\hat{\boldsymbol{\Delta}}_{21} = \hat{\boldsymbol{\Delta}}_{12}'$ with $\hat{\boldsymbol{\Delta}}_{12} = \left\{ (\pi_{1kl} - \pi_{1k}\pi_{1l}) \,/\, \pi_{1k}\pi_{1l}\pi_{1kl}\pi_{2l} \right\}$, and $\boldsymbol{\mathcal{X}}$, $\boldsymbol{\Psi}$ are the sample matrices in (3.5).

We now obtain, as elements of the matrices $\widehat{\mathrm{Var}}(\hat{\mathbf{t}}_{\boldsymbol{\mathcal{X}}})$ and $\widehat{\mathrm{Cov}}(\hat{\mathbf{t}}_{\boldsymbol{\Psi}}, \hat{\mathbf{t}}_{\boldsymbol{\mathcal{X}}})$, the unbiased estimates of all variances and covariances in (3.9), i.e, $\widehat{\mathrm{Var}}(\hat{\mathbf{t}}_{\mathbf{x}}) = \mathbf{X}_1' \, \hat{\boldsymbol{\Delta}}_1 \, \mathbf{X}_1$, $\widehat{\mathrm{Var}}(\tilde{\mathbf{t}}_{\mathbf{x}}) = \mathbf{X}_2' \, \hat{\boldsymbol{\Delta}}_2 \, \mathbf{X}_2$, $\widehat{\mathrm{Var}}(\hat{\mathbf{t}}_{\mathbf{x}_1}) = \mathbf{X}_{11}' \, \hat{\boldsymbol{\Delta}}_1 \, \mathbf{X}_{11}$, $\widehat{\mathrm{Cov}}(\tilde{\mathbf{t}}_{\mathbf{x}}, \hat{\mathbf{t}}_{\mathbf{x}_1}) = \mathbf{X}_2' \, \hat{\boldsymbol{\Delta}}_{21} \, \mathbf{X}_{11}$, $\widehat{\mathrm{Cov}}(\tilde{\mathbf{t}}_{\mathbf{y}}, \hat{\mathbf{t}}_{\mathbf{x}}) = \mathbf{Y}_2' \, \hat{\boldsymbol{\Delta}}_{21} \, \mathbf{X}_1$, $\widehat{\mathrm{Cov}}(\tilde{\mathbf{t}}_{\mathbf{y}}, \hat{\mathbf{t}}_{\mathbf{x}_1}) = \mathbf{Y}_2' \, \hat{\boldsymbol{\Delta}}_{21} \, \mathbf{X}_{11}$, $\widehat{\mathrm{Cov}}(\tilde{\mathbf{t}}_{\mathbf{y}}, \tilde{\mathbf{t}}_{\mathbf{x}}) = \mathbf{Y}_2' \, \hat{\boldsymbol{\Delta}}_2 \, \mathbf{X}_2$. However, the matrix $\widehat{\mathrm{Var}}(\hat{\mathbf{t}}_{\boldsymbol{\mathcal{X}}})$ includes also the elements $\widehat{\mathrm{Cov}}(\hat{\mathbf{t}}_{\mathbf{x}}, \tilde{\mathbf{t}}_{\mathbf{x}}) = \mathbf{X}_1' \, \hat{\boldsymbol{\Delta}}_{12} \, \mathbf{X}_2$, and $\widehat{\mathrm{Cov}}(\hat{\mathbf{t}}_{\mathbf{x}_1}, \hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}) = \mathbf{X}_{11}' \, \hat{\boldsymbol{\Delta}}_1 \, \mathbf{X}_1 - \mathbf{X}_{11}' \, \hat{\boldsymbol{\Delta}}_{12} \, \mathbf{X}_2$, which clearly do not retain the properties $\mathrm{Cov}(\hat{\mathbf{t}}_{\mathbf{x}}, \tilde{\mathbf{t}}_{\mathbf{x}}) = \mathrm{Var}(\hat{\mathbf{t}}_{\mathbf{x}})$ and $\mathrm{Cov}(\hat{\mathbf{t}}_{\mathbf{x}_1}, \hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}) = \mathbf{0}$, respectively. Unbiased estimates for the variances and covariances in (3.9) could be directly used, but then the estimate of the simple form $\mathcal{B}$ in (3.9) could not be expressed as $\boldsymbol{\Psi}'\hat{\boldsymbol{\Delta}}\boldsymbol{\mathcal{X}}(\boldsymbol{\mathcal{X}}'\hat{\boldsymbol{\Delta}}\boldsymbol{\mathcal{X}})^{-1}$, and thus the resulting estimator would not retain the calibration form of the BLUE in (3.12). This complication is circumvented using the following reformulation. Reset $\mathbf{w}^*$, $\boldsymbol{\mathcal{X}}$ and $\boldsymbol{\Psi}$ as

$$\mathbf{w}^* = \begin{pmatrix} \mathbf{w}_1 \\ \mathbf{w} \\ \mathbf{w}_1 \end{pmatrix}, \qquad \boldsymbol{\mathcal{X}} = \begin{pmatrix} -\mathbf{X}_1 & \mathbf{0} \\ \mathbf{X}_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{X}_{11} \end{pmatrix}, \qquad \boldsymbol{\Psi} = \begin{pmatrix} -\mathbf{Y}_1 & -\mathbf{X}_1 \\ \mathbf{Y}_2 & \mathbf{X}_2 \\ \mathbf{Y}_1 & \mathbf{X}_1 \end{pmatrix}, \tag{4.1}$$

where the sample matrices $\mathbf{X}_1$, $\mathbf{X}_2$, $\mathbf{X}_{11}$ and $\mathbf{Y}_2$ are as before, and $\mathbf{Y}_1$ is the matrix of $\mathbf{y}$ for sample s_1 with dummy values $\mathbf{y}_k$ for $k \notin s_2$. Clearly, $\boldsymbol{\mathcal{X}}'\mathbf{w}^*$ and $\boldsymbol{\Psi}'\mathbf{w}^*$ are exactly as in (3.6). Then, having as before $\hat{\mathbf{t}}_{\boldsymbol{\mathcal{X}}} = \boldsymbol{\mathcal{X}}'\mathbf{w}^*$ and $\hat{\mathbf{t}}_{\boldsymbol{\Psi}} = \boldsymbol{\Psi}'\mathbf{w}^*$, we obtain again $\mathcal{B} = \mathrm{Cov}(\hat{\mathbf{t}}_{\boldsymbol{\Psi}}, \hat{\mathbf{t}}_{\boldsymbol{\mathcal{X}}})\left[\mathrm{Var}(\hat{\mathbf{t}}_{\boldsymbol{\mathcal{X}}})\right]^{-1}$, where $\mathrm{Var}(\hat{\mathbf{t}}_{\boldsymbol{\mathcal{X}}}) = \boldsymbol{\mathcal{X}}_U'\,\mathrm{Var}(\mathbf{w}_U^*)\boldsymbol{\mathcal{X}}_U$ and $\mathrm{Cov}(\hat{\mathbf{t}}_{\boldsymbol{\Psi}}, \hat{\mathbf{t}}_{\boldsymbol{\mathcal{X}}}) = \boldsymbol{\Psi}_U'\,\mathrm{Var}(\mathbf{w}_U^*)\boldsymbol{\mathcal{X}}$, as in (3.7) but with $\mathbf{w}_U^*$, $\boldsymbol{\mathcal{X}}_U$ and $\boldsymbol{\Psi}_U$ being the population counterparts of the redefined $\mathbf{w}^*$, $\boldsymbol{\mathcal{X}}$, $\boldsymbol{\Psi}$. An extension of Lemma 1 to the redefined $\mathbf{w}^*$ gives

$$\mathrm{Var}(\mathbf{w}_U^*) = \begin{pmatrix} \mathrm{Var}(\mathbf{w}_{1U}) & \mathrm{Var}(\mathbf{w}_{1U}) & \mathrm{Var}(\mathbf{w}_{1U}) \\ \mathrm{Var}(\mathbf{w}_{1U}) & \mathrm{Var}(\mathbf{w}_U) & \mathrm{Var}(\mathbf{w}_{1U}) \\ \mathrm{Var}(\mathbf{w}_{1U}) & \mathrm{Var}(\mathbf{w}_{1U}) & \mathrm{Var}(\mathbf{w}_{1U}) \end{pmatrix},$$

where $\mathrm{Var}(\mathbf{w}_{1U})$ and $\mathrm{Var}(\mathbf{w}_U)$ are the same as in Lemma 1. It is easy now to verify that again $\mathcal{B}$ may be expressed analytically as in (3.9), and the two components of the BLUE are identical to those given by (3.10). More importantly, it follows from this special form of $\mathrm{Var}(\mathbf{w}_U^*)$ that we have again $\mathrm{Var}(\hat{\mathbf{t}}_{\boldsymbol{\mathcal{X}}}) = \boldsymbol{\mathcal{X}}_U'\,\boldsymbol{\Delta}\boldsymbol{\mathcal{X}}_U$ and $\mathrm{Cov}(\hat{\mathbf{t}}_{\boldsymbol{\Psi}}, \hat{\mathbf{t}}_{\boldsymbol{\mathcal{X}}}) = \boldsymbol{\Psi}_U'\,\boldsymbol{\Delta}\boldsymbol{\mathcal{X}}_U$, where now $\boldsymbol{\Delta} = \mathrm{diag}(-\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2, \boldsymbol{\Delta}_1)$ and $\boldsymbol{\Delta}_1$, $\boldsymbol{\Delta}_2$ as already defined. Thus we obtain again the BLUE in the calibration form of (3.12), and the retained orthogonal decomposition of the vector of calibrated weigs in (3.13) leads readily to the expression (3.14). Now the orthogonality property $\boldsymbol{\mathcal{X}}_{12U}'\,\boldsymbol{\Delta}\boldsymbol{\mathcal{X}}_{1U} = \mathbf{0}$ is induced by the block-diagonal structure of the redefined $\boldsymbol{\mathcal{X}}_U$, rather than by the special structure of the initial matrix $\boldsymbol{\Delta}$ used in (3.12).

For the reconstructed BLUE we now have the unbiased estimates $\widehat{\mathrm{Var}}(\hat{\mathbf{t}}_{\boldsymbol{\mathcal{X}}}) = \boldsymbol{\mathcal{X}}'\hat{\boldsymbol{\Delta}}\boldsymbol{\mathcal{X}}$ and $\widehat{\mathrm{Cov}}(\hat{\mathbf{t}}_{\boldsymbol{\Psi}}, \hat{\mathbf{t}}_{\boldsymbol{\mathcal{X}}}) = \boldsymbol{\Psi}'\hat{\boldsymbol{\Delta}}\boldsymbol{\mathcal{X}}$, where $\boldsymbol{\mathcal{X}}$, $\boldsymbol{\Psi}$ are the sample matrices in (4.1), and $\hat{\boldsymbol{\Delta}} = \mathrm{diag}(-\hat{\boldsymbol{\Delta}}_1, \hat{\boldsymbol{\Delta}}_2, \hat{\boldsymbol{\Delta}}_1)$ with $\hat{\boldsymbol{\Delta}}_1$, $\hat{\boldsymbol{\Delta}}_2$ as defined at the beginning of the section. From these we rederive easily the unbiased estimates of the variances and covariances in (3.9), but two of the elements of the sample matrix $\boldsymbol{\Psi}'\hat{\boldsymbol{\Delta}}\boldsymbol{\mathcal{X}}$ which involve $\mathbf{Y}_1$, namely $\mathbf{Y}_1'\hat{\boldsymbol{\Delta}}_1\mathbf{X}_1$ and $\mathbf{Y}_1'\hat{\boldsymbol{\Delta}}_1\mathbf{X}_{11}$, require special consideration. The dummy (unobserved) values $\mathbf{y}_k$ for $k \notin s_2$, necessary for expanding $\mathbf{Y}_1$ to the population matrix $\mathbf{Y}_U$ in the reconstructed BLUE, are set equal to zero, and the values $\mathbf{y}_k$ for $k \in s_2$ are then necessarily weiged by $1/\pi_{2k}$. Then $\mathbf{Y}_1'\hat{\boldsymbol{\Delta}}_1\mathbf{X}_1$ and $\mathbf{Y}_1'\hat{\boldsymbol{\Delta}}_1\mathbf{X}_{11}$ reduce to $\mathbf{Y}_2'\hat{\boldsymbol{\Delta}}_{21}\mathbf{X}_1$ and $\mathbf{Y}_2'\hat{\boldsymbol{\Delta}}_{21}\mathbf{X}_{11}$, which are the unbiased estimates $\widehat{\mathrm{Cov}}(\tilde{\mathbf{t}}_{\mathbf{y}}, \hat{\mathbf{t}}_{\mathbf{x}})$ and $\widehat{\mathrm{Cov}}(\tilde{\mathbf{t}}_{\mathbf{y}}, \hat{\mathbf{t}}_{\mathbf{x}_1})$, respectively. The estimated $\mathcal{B}$ in (3.9) is now given by

$$\hat{\mathcal{B}} = \begin{bmatrix} \left[\mathbf{Y}_2'\hat{\boldsymbol{\Delta}}_2\mathbf{X}_2 - \mathbf{Y}_2'\hat{\boldsymbol{\Delta}}_{21}\mathbf{X}_1\right]\left[\mathbf{X}_2'\hat{\boldsymbol{\Delta}}_2\mathbf{X}_2 - \mathbf{X}_1'\hat{\boldsymbol{\Delta}}_1\mathbf{X}_1\right]^{-1} & \mathbf{Y}_2'\hat{\boldsymbol{\Delta}}_{21}\mathbf{X}_{11}\left[\mathbf{X}_{11}'\hat{\boldsymbol{\Delta}}_1\mathbf{X}_{11}\right]^{-1} \\ \mathbf{I} & \mathbf{X}_1'\hat{\boldsymbol{\Delta}}_1\mathbf{X}_{11}\left[\mathbf{X}_{11}'\hat{\boldsymbol{\Delta}}_1\mathbf{X}_{11}\right]^{-1} \end{bmatrix}.$$

The BLUE $\hat{\mathbf{t}}_{\boldsymbol{\Psi}}^{B} = \hat{\mathbf{t}}_{\boldsymbol{\Psi}} + \boldsymbol{\mathcal{B}}(\mathbf{t}_{\mathcal{X}} - \hat{\mathbf{t}}_{\mathcal{X}})$ with estimated $\boldsymbol{\mathcal{B}}$ will be called optimal linear unbiased estimator, optimal estimator in short, denoted by $\hat{\mathbf{t}}_{\boldsymbol{\Psi}}^{O} = \hat{\mathbf{t}}_{\boldsymbol{\Psi}} + \hat{\boldsymbol{\mathcal{B}}}(\mathbf{t}_{\mathcal{X}} - \hat{\mathbf{t}}_{\mathcal{X}}),$ with its two components given by

$$\begin{aligned} \hat{\mathbf{t}}_{\mathbf{y}}^{O} &= \tilde{\mathbf{t}}_{\mathbf{y}} + \left[\mathbf{Y}_2' \hat{\boldsymbol{\Delta}}_2 \mathbf{X}_2 - \mathbf{Y}_2' \hat{\boldsymbol{\Delta}}_{21} \mathbf{X}_1\right] \left[\mathbf{X}_2' \hat{\boldsymbol{\Delta}}_2 \mathbf{X}_2 - \mathbf{X}_1' \hat{\boldsymbol{\Delta}}_1 \mathbf{X}_1\right]^{-1} (\hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}) \\ &\quad + \mathbf{Y}_2' \hat{\boldsymbol{\Delta}}_{21} \mathbf{X}_{11} \left[\mathbf{X}_{11}' \hat{\boldsymbol{\Delta}}_1 \mathbf{X}_{11}\right]^{-1} (\mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1}) \\ \hat{\mathbf{t}}_{\mathbf{x}}^{O} &= \hat{\mathbf{t}}_{\mathbf{x}} + \mathbf{X}_1' \hat{\boldsymbol{\Delta}}_1 \mathbf{X}_{11} \left[\mathbf{X}_{11}' \hat{\boldsymbol{\Delta}}_1 \mathbf{X}_{11}\right]^{-1} (\mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1}). \end{aligned} \tag{4.2}$$

This is the sample version of the BLUEs in (3.14), with estimated coefficients. In particular, $\hat{\mathbf{t}}_{\mathbf{x}}^{O}$ is the customary single-phase optimal estimator of $\mathbf{t}_{\mathbf{x}}$ using $\mathbf{x}_1$ as auxiliary variable, and data from the full first-phase sample s_1; see Montanari (1987) and Rao (1994).

Remark 4.1. When n_2 is very close to n_1, the optimal estimator $\hat{\mathbf{t}}_{\mathbf{y}}^{O}$ can be quite unstable because of the near singularity of the inverted matrix in the coefficient of $\hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}$, and thus can become very inefficient; see, though, later Remark 6.1 on two-phase designs in which this is not an issue. Generally this is not a realistic setting in two-phase sampling, where n_2 is typically much smaller than n_1.

Following the construction of $\mathbf{Y}_2' \hat{\boldsymbol{\Delta}}_{21} \mathbf{X}_1$ and $\mathbf{Y}_2' \hat{\boldsymbol{\Delta}}_{21} \mathbf{X}_{11}$ as two of the estimates in $\hat{\boldsymbol{\mathcal{B}}}$, it transpires that these two bilinear forms can be written alternatively as $\breve{\mathbf{Y}}_1' \hat{\boldsymbol{\Delta}}_1 \mathbf{X}_1$ and $\breve{\mathbf{Y}}_1' \hat{\boldsymbol{\Delta}}_1 \mathbf{X}_{11}$, respectively, where $\breve{\mathbf{Y}}_1$ is a weiged version of $\mathbf{Y}_1$ in which $\breve{\mathbf{y}}_k = \mathbf{y}_k / \pi_{2k}$ if $k \in s_2$ and $\breve{\mathbf{y}}_k = 0$ if $k \notin s_2$. Then $\hat{\mathbf{t}}_{\boldsymbol{\Psi}} = \boldsymbol{\Psi}' \mathbf{w}^* = \breve{\boldsymbol{\Psi}}' \mathbf{w}^*$, where $\breve{\boldsymbol{\Psi}}$ is $\boldsymbol{\Psi}$ in (4.1) with $\breve{\mathbf{Y}}_1$ in place of $\mathbf{Y}_1$, and $\hat{\boldsymbol{\mathcal{B}}}$ can be written compactly as $\hat{\boldsymbol{\mathcal{B}}} = \breve{\boldsymbol{\Psi}}' \hat{\boldsymbol{\Delta}} \boldsymbol{\mathcal{X}} (\boldsymbol{\mathcal{X}}' \hat{\boldsymbol{\Delta}} \boldsymbol{\mathcal{X}})^{-1}$, where $\hat{\boldsymbol{\Delta}} = \operatorname{diag}(-\hat{\boldsymbol{\Delta}}_1, \hat{\boldsymbol{\Delta}}_2, \hat{\boldsymbol{\Delta}}_1)$. Henceforth, $\hat{\boldsymbol{\Delta}}$ will be meant to be the matrix $\operatorname{diag}(-\hat{\boldsymbol{\Delta}}_1, \hat{\boldsymbol{\Delta}}_2, \hat{\boldsymbol{\Delta}}_1)$.

As in Montanari (1987) and Rao (1994) for the single-phase optimal estimator, for large samples s_1 and s_2 the optimal estimator $\hat{\mathbf{t}}_{\boldsymbol{\Psi}}^{O} = \hat{\mathbf{t}}_{\boldsymbol{\Psi}} + \hat{\boldsymbol{\mathcal{B}}}(\mathbf{t}_{\mathcal{X}} - \hat{\mathbf{t}}_{\mathcal{X}})$ approximates the BLUE $\hat{\mathbf{t}}_{\boldsymbol{\Psi}}^{B}$, and thus it is approximately unbiased. Furthermore, the variance of $\hat{\mathbf{t}}_{\boldsymbol{\Psi}}^{O}$ approximates that of $\hat{\mathbf{t}}_{\boldsymbol{\Psi}}^{B}$, which works out easily to be $\operatorname{Var}(\hat{\mathbf{t}}_{\boldsymbol{\Psi}}^{B}) = \operatorname{Var}(\hat{\mathbf{t}}_{\boldsymbol{\Psi}}) - \operatorname{Cov}(\hat{\mathbf{t}}_{\boldsymbol{\Psi}}, \hat{\mathbf{t}}_{\mathcal{X}}) \left[\operatorname{Var}(\hat{\mathbf{t}}_{\mathcal{X}})\right]^{-1} \operatorname{Cov}'(\hat{\mathbf{t}}_{\boldsymbol{\Psi}}, \hat{\mathbf{t}}_{\mathcal{X}})$, i.e., the compact form of (3.11). Then, using the estimates $\widehat{\operatorname{Var}}(\hat{\mathbf{t}}_{\boldsymbol{\Psi}})$ and $\widehat{\operatorname{Cov}}(\hat{\mathbf{t}}_{\boldsymbol{\Psi}}, \hat{\mathbf{t}}_{\mathcal{X}})$, derived earlier, we obtain the estimated approximate variance of $\hat{\mathbf{t}}_{\boldsymbol{\Psi}}^{O}$ as $\widehat{\operatorname{AV}}(\hat{\mathbf{t}}_{\boldsymbol{\Psi}}^{O}) = \widehat{\operatorname{Var}}(\hat{\mathbf{t}}_{\boldsymbol{\Psi}}) - \widehat{\operatorname{Cov}}(\hat{\mathbf{t}}_{\boldsymbol{\Psi}}, \hat{\mathbf{t}}_{\mathcal{X}}) \left[\widehat{\operatorname{Var}}(\hat{\mathbf{t}}_{\mathcal{X}})\right]^{-1} \widehat{\operatorname{Cov}}'(\hat{\mathbf{t}}_{\boldsymbol{\Psi}}, \hat{\mathbf{t}}_{\mathcal{X}})$. From this we derive the computationally convenient expressions $\widehat{\operatorname{AV}}(\hat{\mathbf{t}}_{\mathbf{y}}^{O}) = \mathbf{Y}_2' \hat{\boldsymbol{\Delta}}_2 \mathbf{Y}_2 - \breve{\boldsymbol{\Psi}}_1' \hat{\boldsymbol{\Delta}} \boldsymbol{\mathcal{X}} (\boldsymbol{\mathcal{X}}' \hat{\boldsymbol{\Delta}} \boldsymbol{\mathcal{X}})^{-1} \boldsymbol{\mathcal{X}}' \hat{\boldsymbol{\Delta}} \breve{\boldsymbol{\Psi}}_1$, where $\breve{\boldsymbol{\Psi}}_1$ is the first column submatrix of $\breve{\boldsymbol{\Psi}}$, and $\widehat{\operatorname{AV}}(\hat{\mathbf{t}}_{\mathbf{x}}^{O}) = \mathbf{X}_1' \hat{\boldsymbol{\Delta}}_1 \mathbf{X}_1 - \mathbf{X}_1' \hat{\boldsymbol{\Delta}}_1 \mathbf{X}_{11} \left[\mathbf{X}_{11}' \hat{\boldsymbol{\Delta}}_1 \mathbf{X}_{11}\right]^{-1} \mathbf{X}_{11}' \hat{\boldsymbol{\Delta}}_1 \mathbf{X}_1$.

4.2 The two-phase optimal estimator as calibration estimator

The optimal estimator $\hat{\mathbf{t}}_{\boldsymbol{\Psi}}^{O} = \hat{\mathbf{t}}_{\boldsymbol{\Psi}} + \hat{\boldsymbol{\mathcal{B}}}(\mathbf{t}_{\mathcal{X}} - \hat{\mathbf{t}}_{\mathcal{X}})$, with $\hat{\boldsymbol{\mathcal{B}}} = \breve{\boldsymbol{\Psi}}' \hat{\boldsymbol{\Delta}} \boldsymbol{\mathcal{X}} (\boldsymbol{\mathcal{X}}' \hat{\boldsymbol{\Delta}} \boldsymbol{\mathcal{X}})^{-1}$, takes the form

$$\hat{\mathbf{t}}_{\boldsymbol{\Psi}}^{O} = \breve{\boldsymbol{\Psi}}' \left[\mathbf{w}^* + \hat{\boldsymbol{\Delta}} \boldsymbol{\mathcal{X}} (\boldsymbol{\mathcal{X}}' \hat{\boldsymbol{\Delta}} \boldsymbol{\mathcal{X}})^{-1} (\mathbf{t}_{\mathcal{X}} - \boldsymbol{\mathcal{X}}' \mathbf{w}^*)\right],$$

of a calibration estimator, with vector of calibration totals $\mathbf{t}_{\mathcal{X}}$ and sample vector of calibrated weigs $\mathbf{c}^* = \mathbf{w}^* + \hat{\boldsymbol{\Delta}} \boldsymbol{\mathcal{X}} (\boldsymbol{\mathcal{X}}' \hat{\boldsymbol{\Delta}} \boldsymbol{\mathcal{X}})^{-1} \left(\mathbf{t}_{\mathcal{X}} - \boldsymbol{\mathcal{X}}' \mathbf{w}^*\right)$ satisfying $\boldsymbol{\mathcal{X}}' \mathbf{c}^* = \mathbf{t}_{\mathcal{X}}$. This is established formally by the following theorem; the proof is similar to that of Theorem 1, and is omitted.

Theorem 2. *The vector* $\mathbf{c}^* = \mathbf{w}^* + \hat{\boldsymbol{\Delta}}\boldsymbol{\mathcal{X}}(\boldsymbol{\mathcal{X}}'\hat{\boldsymbol{\Delta}}\boldsymbol{\mathcal{X}})^{-1}(\mathbf{t}_{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}'\mathbf{w}^*)$ *minimizes the generalized least-squares distance* $(\mathbf{c}^* - \mathbf{w}^*)'\hat{\boldsymbol{\Delta}}^{-1}(\mathbf{c}^* - \mathbf{w}^*)$ *subject to the constraints* $\boldsymbol{\mathcal{X}}'\mathbf{c}^* = \mathbf{t}_{\boldsymbol{\mathcal{X}}}$, *i.e.,* $\mathbf{X}_1'\mathbf{c}_1 = \mathbf{X}_2'\mathbf{c}$ *and* $\mathbf{X}_{11}'\mathbf{c}_1 = \mathbf{t}_{\mathbf{x}_1}$, *where* $(\mathbf{c}_1, \mathbf{c})$ *corresponds to* $(\mathbf{w}_1, \mathbf{w})$.

The sample vector $\mathbf{c}^*$ admits the same orthogonal decomposition as its population counterpart $\mathbf{c}_U^*$ in (3.13). We may now write formally the optimal estimator $\hat{\mathbf{t}}_{\boldsymbol{\Psi}}^O$ as a calibration estimator $\hat{\mathbf{t}}_{\boldsymbol{\Psi}}^O = \breve{\boldsymbol{\Psi}}'\mathbf{c}^*$, which in view of $\boldsymbol{\mathcal{X}}'\mathbf{c}^* = \mathbf{t}_{\boldsymbol{\mathcal{X}}}$ is generated by the simultaneous calibration of the two estimates $\hat{\mathbf{t}}_{\mathbf{x}}$ and $\tilde{\mathbf{t}}_{\mathbf{x}}$ of $\mathbf{t}_{\mathbf{x}}$ to each other, and of the estimate $\hat{\mathbf{t}}_{\mathbf{x}_1}$ to the total $\mathbf{t}_{\mathbf{x}_1}$.

Now, in expanded form the vector $\mathbf{c}^*$ is

$$\mathbf{c}^* = \begin{pmatrix} \mathbf{c}_1 \\ \mathbf{c}_2 \\ \mathbf{c}_3 \end{pmatrix} = \begin{pmatrix} \mathbf{w}_1 + \hat{\boldsymbol{\Delta}}_1\mathbf{X}_1[\mathbf{X}_2'\hat{\boldsymbol{\Delta}}_2\mathbf{X}_2 - \mathbf{X}_1'\hat{\boldsymbol{\Delta}}_1\mathbf{X}_1]^{-1}(\hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}) \\ \mathbf{w} + \hat{\boldsymbol{\Delta}}_2\mathbf{X}_2[\mathbf{X}_2'\hat{\boldsymbol{\Delta}}_2\mathbf{X}_2 - \mathbf{X}_1'\hat{\boldsymbol{\Delta}}_1\mathbf{X}_1]^{-1}(\hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}) \\ \mathbf{w}_1 + \hat{\boldsymbol{\Delta}}_1\mathbf{X}_{11}(\mathbf{X}_{11}'\hat{\boldsymbol{\Delta}}_1\mathbf{X}_{11})^{-1}(\mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1}) \end{pmatrix}. \tag{4.3}$$

Then, using the partition $\boldsymbol{\mathcal{X}} = (\boldsymbol{\mathcal{X}}_{12}, \boldsymbol{\mathcal{X}}_1)$, where $\boldsymbol{\mathcal{X}}_{12}$ and $\boldsymbol{\mathcal{X}}_1$ are the two orthogonal column submatrices of $\boldsymbol{\mathcal{X}}$ shown in (4.1), the two constraints are written as $\boldsymbol{\mathcal{X}}_{12}'\mathbf{c}^* = \mathbf{X}_2'\mathbf{c}_2 - \mathbf{X}_1'\mathbf{c}_1 = \mathbf{0}$ and $\boldsymbol{\mathcal{X}}_1'\mathbf{c}^* = \mathbf{X}_{11}'\mathbf{c}_3 = \mathbf{t}_{\mathbf{x}_1}$. It also follows from (4.3) that $\hat{\mathbf{t}}_{\boldsymbol{\Psi}}^O = \breve{\boldsymbol{\Psi}}'\mathbf{c}^*$ implies (4.2). Regarding the two components of $\hat{\mathbf{t}}_{\boldsymbol{\Psi}}^O$ we observe that $\hat{\mathbf{t}}_{\mathbf{x}}^O = -\mathbf{X}_1'\mathbf{c}_1 + \mathbf{X}_2'\mathbf{c}_2 + \mathbf{X}_1'\mathbf{c}_3 = \mathbf{X}_1'\mathbf{c}_3$, and that

$$\hat{\mathbf{t}}_{\mathbf{y}}^O = \breve{\mathbf{Y}}_1'(\mathbf{c}_3 - \mathbf{c}_1) + \mathbf{Y}_2'\mathbf{c}_2 = \sum\nolimits_{s_2}\left[(c_{3k} - c_{1k})/\pi_{2k} + c_{2k}\right]\mathbf{y}_k.$$

The explicit expression of $\hat{\mathbf{t}}_{\mathbf{y}}^O$, in terms of sample units, is

$$\begin{aligned}\hat{\mathbf{t}}_{\mathbf{y}}^O = {} & \tilde{\mathbf{t}}_{\mathbf{y}} + \left[\sum\nolimits_{s_2}\sum\nolimits_{s_2}\hat{\Delta}_{2kl}\,\mathbf{y}_k\mathbf{x}_l' - \sum\nolimits_{s_2}\sum\nolimits_{s_1}\hat{\Delta}_{1kl}\,\breve{\mathbf{y}}_k\mathbf{x}_l'\right] \times \\ & \left[\sum\nolimits_{s_2}\sum\nolimits_{s_2}\hat{\Delta}_{2kl}\,\mathbf{x}_k\mathbf{x}_l' - \sum\nolimits_{s_1}\sum\nolimits_{s_1}\hat{\Delta}_{1kl}\,\mathbf{x}_k\mathbf{x}_l'\right]^{-1}(\hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}) \\ & + \left(\sum\nolimits_{s_2}\sum\nolimits_{s_1}\hat{\Delta}_{1kl}\,\breve{\mathbf{y}}_k\mathbf{x}_{1l}'\right)\left(\sum\nolimits_{s_1}\sum\nolimits_{s_1}\hat{\Delta}_{1kl}\,\mathbf{x}_{1k}\mathbf{x}_{1l}'\right)^{-1}(\mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1}),\end{aligned} \tag{4.4}$$

where $\hat{\Delta}_{1kl}$ and $\hat{\Delta}_{2kl}$ are the kl^{th} elements of $\hat{\boldsymbol{\Delta}}_1$ and $\hat{\boldsymbol{\Delta}}_2$, respectively. Formula (4.4) is simplified in certain two-phase designs employed in important large scale surveys; examples of such surveys are presented in Hidiroglou and Särndal (1998) and Turmelle and Beaucage (2013). Specifically, this is the case when independent sampling (Poisson, or stratified Poisson) is used in one of the two phases, that is, when $\pi_{1kl} = \pi_{1k}\pi_{1l}$ or $\pi_{2kl} = \pi_{2k}\pi_{2l}$. The simplification is considerable in the case of independent sampling in both phases. Then, both $\hat{\boldsymbol{\Delta}}_1$ and $\hat{\boldsymbol{\Delta}}_2$ are diagonal, with diagonal elements $\hat{\Delta}_{1kk} = (1/\pi_{1k})((1/\pi_{1k}) - 1)$ and $\hat{\Delta}_{2kk} = (1/\pi_{1k}\pi_{2k})((1/\pi_{1k}\pi_{2k}) - 1)$, respectively, and (4.4) involves only single summations. Other two-phase designs in which (4.4) involves single summations only, although $\hat{\boldsymbol{\Delta}}_1$ and $\hat{\boldsymbol{\Delta}}_2$ are not diagonal, involve simple random sampling or stratified simple random sampling in either phase; for an example of a survey with such two-phase design see Hidiroglou (2001). In general, however, the optimal estimator may not be practical because it requires the use of first-phase and second-phase joint inclusion probabilities π_{1kl} and π_{2kl}, which are not known for some complex sampling designs. Even when these joint probabilities are known, but the matrices $\hat{\mathbf{\Delta}}_1$ and $\hat{\mathbf{\Delta}}_2$ are not diagonal, the estimated coefficient $\hat{\boldsymbol{\mathcal{B}}}$ and, hence, the optimal estimator may be unstable in very small samples – especially if the dimension of the auxiliary vector $\mathbf{x}$ is large. These difficulties may be overcome, at some loss of optimality, by employing simple approximations of the variances and covariances in $\hat{\boldsymbol{\mathcal{B}}}$; for approximate variance estimates based only on first order inclusion probabilities see, for example, Haziza, Mecatti and Rao (2008) and references therein. A computationally very convenient approximation of $\hat{\boldsymbol{\mathcal{B}}}$ leading to a two-phase estimator that belongs to the class of generalized regression estimators is described in the next section.

5. A two-phase generalized regression estimator

A variant of $\hat{\boldsymbol{\mathcal{B}}} = \breve{\mathbf{\Psi}}' \hat{\Delta} \boldsymbol{\mathcal{X}} (\boldsymbol{\mathcal{X}}' \hat{\Delta} \boldsymbol{\mathcal{X}})^{-1}$ that is computationally efficient, but generally suboptimal, is the generalized regression (GREG) coefficient $\hat{\boldsymbol{\mathcal{B}}}^{\text{GR}} = \mathbf{\Psi}' \mathbf{\Lambda} \boldsymbol{\mathcal{X}} (\boldsymbol{\mathcal{X}}' \mathbf{\Lambda} \boldsymbol{\mathcal{X}})^{-1}$, where $\mathbf{\Psi}$ is as in (4.1) and with $\mathbf{y}_k = 0$ if $k \notin s_2$, and $\mathbf{\Lambda}$ is the "weiging" matrix $\operatorname{diag}(\mathbf{\Lambda}_1, \mathbf{\Lambda}_2, \mathbf{\Lambda}_1)$, with $\mathbf{\Lambda}_1 = \operatorname{diag}\{w_{1k} / q_{1k}\}$ and $\mathbf{\Lambda}_2 = \operatorname{diag}\{w_k / q_{2k}\}$, and with q_{1k}, q_{2k} being positive constants. This gives the GREG estimator

$$\hat{\mathbf{t}}_{\mathbf{\Psi}}^{\text{GR}} = \hat{\mathbf{t}}_{\mathbf{\Psi}} + \hat{\boldsymbol{\mathcal{B}}}^{\text{GR}} (\mathbf{t}_{\boldsymbol{\mathcal{X}}} - \hat{\mathbf{t}}_{\boldsymbol{\mathcal{X}}}) = \hat{\boldsymbol{\mathcal{B}}}^{\text{GR}} \mathbf{t}_{\boldsymbol{\mathcal{X}}} + (\mathbf{\Psi} - \boldsymbol{\mathcal{X}} \hat{\boldsymbol{\mathcal{B}}}^{\text{GR}\prime})' \mathbf{w}^*. \tag{5.1}$$

Note that $\hat{\boldsymbol{\mathcal{B}}}^{\text{GR}}$ is optimal in the sense of least squares, i.e., it minimizes the quadratic distance $(\mathbf{\Psi} - \boldsymbol{\mathcal{X}} \hat{\boldsymbol{\mathcal{B}}}^{\text{GR}\prime})' \mathbf{\Lambda} (\mathbf{\Psi} - \boldsymbol{\mathcal{X}} \hat{\boldsymbol{\mathcal{B}}}^{\text{GR}\prime})$, involving the residuals $\mathbf{\Psi} - \boldsymbol{\mathcal{X}} \hat{\boldsymbol{\mathcal{B}}}^{\text{GR}\prime}$ in $\hat{\mathbf{t}}_{\mathbf{\Psi}}^{\text{GR}}$, whereas the coefficient $\hat{\boldsymbol{\mathcal{B}}}$ minimizes $(\mathbf{\Psi} - \boldsymbol{\mathcal{X}} \hat{\boldsymbol{\mathcal{B}}})' \hat{\Delta} (\mathbf{\Psi} - \boldsymbol{\mathcal{X}} \hat{\boldsymbol{\mathcal{B}}})'$, the estimated approximate variance of the optimal estimator $\hat{\mathbf{t}}_{\mathbf{\Psi}}^{O}$. In this sense $\hat{\mathbf{t}}_{\mathbf{\Psi}}^{\text{GR}}$ is an approximation to $\hat{\mathbf{t}}_{\mathbf{\Psi}}^{O}$. The two components of $\hat{\mathbf{t}}_{\mathbf{\Psi}}^{\text{GR}}$, similar in structure to the components of $\hat{\mathbf{t}}_{\mathbf{\Psi}}^{O}$ in (4.2), are

$$\begin{aligned}
\hat{\mathbf{t}}_{\mathbf{y}}^{\text{GR}} &= \tilde{\mathbf{t}}_{\mathbf{y}} + \left[\mathbf{Y}_2' \mathbf{\Lambda}_2 \mathbf{X}_2 + \mathbf{Y}_1' \mathbf{\Lambda}_1 \mathbf{X}_1\right] \left[\mathbf{X}_2' \mathbf{\Lambda}_2 \mathbf{X}_2 + \mathbf{X}_1' \mathbf{\Lambda}_1 \mathbf{X}_1\right]^{-1} (\hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}) \\
&\quad + \mathbf{Y}_1' \mathbf{\Lambda}_1 \mathbf{X}_{11} \left[\mathbf{X}_{11}' \mathbf{\Lambda}_1 \mathbf{X}_{11}\right]^{-1} (\mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1}) \\
\hat{\mathbf{t}}_{\mathbf{x}}^{\text{GR}} &= \hat{\mathbf{t}}_{\mathbf{x}} + \mathbf{X}_1' \mathbf{\Lambda}_1 \mathbf{X}_{11} \left[\mathbf{X}_{11}' \mathbf{\Lambda}_1 \mathbf{X}_{11}\right]^{-1} (\mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1}).
\end{aligned} \tag{5.2}$$

The GREG estimator $\hat{\mathbf{t}}_{\mathbf{x}}^{\text{GR}}$ is the standard single-phase GREG estimator based on s_1 and the auxiliary variable $\mathbf{x}_1$. The GREG estimator $\hat{\mathbf{t}}_{\mathbf{y}}^{\text{GR}}$, with the two orthogonal regression terms shown in (5.2), is expressed explicitly in terms of sample units as

$$\begin{aligned}
\hat{\mathbf{t}}_{\mathbf{y}}^{\text{GR}} &= \tilde{\mathbf{t}}_{\mathbf{y}} + \left[\sum\nolimits_{s_2} (\Lambda_{1k} + \Lambda_{2k}) \mathbf{y}_k \mathbf{x}_k'\right] \left[\sum\nolimits_{s_2} \Lambda_{2k} \mathbf{x}_k \mathbf{x}_k' + \sum\nolimits_{s_1} \Lambda_{1k} \mathbf{x}_k \mathbf{x}_k'\right]^{-1} (\hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}) \\
&\quad + \left(\sum\nolimits_{s_2} \Lambda_{1k} \mathbf{y}_k \mathbf{x}_{1k}'\right) \left(\sum\nolimits_{s_1} \Lambda_{1k} \mathbf{x}_{1k} \mathbf{x}_{1k}'\right)^{-1} (\mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1}),
\end{aligned}$$

where $\Lambda_{1k} = w_{1k} / q_{1k}$ and $\Lambda_{2k} = w_k / q_{2k}$ are the k^{th} element of $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$, respectively. The constants q_{ik} should be specified as $q_{ik} = n_i$, to account for the differential in the sample size of s_i; see Merkouris (2004) for a justification in the context of calibrating combined samples. An equivalent adjustment of the weigs in Λ_{1k} and Λ_{2k} can be made through the multiplication of w_{1k} in Λ_{1k} by $\phi = n_2 / n_1$. Values of q_{ik}

that convert the GREG estimator $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{GR}}$ to the optimal estimator $\hat{\mathbf{t}}_{\mathbf{y}}^{O}$ can be specified for two-phase sampling designs for which optimal estimation is possible, as in the similar context of matrix sampling (Merkouris, 2015). For the simple example involving Poisson sampling in both phases, this specification is $q_{1k} = \pi_{1k} / (1 - \pi_{1k})$ and $q_{2k} = \pi_{1k}\, \pi_{2k} / (1 - \pi_{1k}\, \pi_{2k})$, rendering $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ identical to $\hat{\mathbf{\Delta}}_1$ and $\hat{\mathbf{\Delta}}_2$.

The vector of calibrated weigs associated with the GREG estimator $\hat{\mathbf{t}}_{\mathbf{\Psi}}^{\mathrm{GR}}$ is $\mathbf{c}^{\mathrm{GR}} = \mathbf{w}^* + \mathbf{\Lambda}\mathcal{X}(\mathcal{X}'\mathbf{\Lambda}\mathcal{X})^{-1}(\mathbf{t}_{\mathcal{X}} - \mathcal{X}'\mathbf{w}^*)$. It has the same form as $\mathbf{c}^*$ in (4.3), but with $\mathbf{\Lambda}_1$ and $\mathbf{\Lambda}_2$ in place of $-\mathbf{\Delta}_1$ and $\mathbf{\Delta}_2$, and minimizes the generalized least-squares distance $(\mathbf{c}^{\mathrm{GR}} - \mathbf{w}^*)'\mathbf{\Lambda}^{-1}(\mathbf{c}^{\mathrm{GR}} - \mathbf{w}^*)$ subject to the constraints $\mathcal{X}'\mathbf{c}^{\mathrm{GR}} = \mathbf{t}_{\mathcal{X}}$. The partition $\mathcal{X} = (\mathcal{X}_{12}, \mathcal{X}_1)$, defined after (4.3), allows the orthogonal decomposition of the vector $\mathbf{c}^*$

$$\begin{aligned} \mathbf{c}^{\mathrm{GR}} \;=\; & \mathbf{w}^* + \mathbf{\Lambda}\mathcal{X}_{12}(\mathcal{X}_{12}'\mathbf{\Lambda}\mathcal{X}_{12})^{-1}(\mathbf{0} - \mathcal{X}_{12}'\mathbf{w}^*) \\ & + \mathbf{\Lambda}\mathcal{X}_1(\mathcal{X}_1'\mathbf{\Lambda}\mathcal{X}_1)^{-1}(\mathbf{t}_{\mathbf{x}_1} - \mathcal{X}_1'\mathbf{w}^*). \end{aligned} \tag{5.3}$$

In the rig hand side of (5.3), the sum of the first and second terms would result from calibration with constraint $\mathcal{X}_{12}'\mathbf{c}^{\mathrm{GR}} = \mathbf{0}$ only, while the sum of the first and third terms would result from calibration with constraint $\mathcal{X}_1'\mathbf{c}^{\mathrm{GR}} = \mathbf{t}_{\mathbf{x}_1}$ only. The practical implication of this is that the vector $\mathbf{c}^*$ could be formed by concatenating the weig vectors generated by two separate calibrations, i.e., calibration of $(\mathbf{w}_1', \mathbf{w}')'$ using $(-\mathbf{X}_1', \mathbf{X}_2')'$ followed by calibration of $\mathbf{w}_1$ using $\mathbf{X}_{11}$. However, the one-step calibration procedure generating $\mathbf{c}^{\mathrm{GR}}$ is more convenient.

On the basis of its Taylor linearization, the GREG estimator $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{GR}}$ in (5.1) is approximately (for large samples) unbiased. Furthermore, denoting by $\mathbf{e}$ the matrix of sample residuals $\mathbf{\Psi} - \mathcal{X}\hat{\mathcal{B}}^{\mathrm{GR}\prime}$, the estimated approximate variance of $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{GR}} = \hat{\mathcal{B}}^{\mathrm{GR}}\mathbf{t}_{\mathcal{X}} + \mathbf{e}'\mathbf{w}^*$ is the estimated variance of $\mathbf{e}'\mathbf{w}^*$, i.e., $\widehat{\mathrm{AV}}(\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{GR}}) = \widehat{\mathrm{Var}}(\mathbf{e}'\mathbf{w}^*) = \mathbf{e}'\hat{\mathbf{\Delta}}\mathbf{e}$, whereas the estimated variance of the HT estimator $\tilde{\mathbf{t}}_{\mathbf{y}}$ is $\widehat{\mathrm{Var}}(\tilde{\mathbf{t}}_{\mathbf{y}}) = \widehat{\mathrm{Var}}(\mathbf{\Psi}_1'\mathbf{w}^*) = \mathbf{Y}_2'\hat{\mathbf{\Delta}}_2\mathbf{Y}_2$, with $\mathbf{\Psi}_1$ being the first column submatrix of $\mathbf{\Psi}$.

Now using the calibration form $\mathbf{\Psi}_1'\mathbf{c}^{\mathrm{GR}}$ of $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{GR}}$ and the orthogonal decomposition (5.3) of $\mathbf{c}^{\mathrm{GR}}$, we easily obtain the decomposition $\mathbf{e} = \mathbf{\Psi}_1 - \mathcal{X}_{12}\hat{\boldsymbol{\beta}}_{\mathbf{x}}' - \mathcal{X}_1\hat{\boldsymbol{\beta}}_{\mathbf{x}_1}'$, where $\hat{\boldsymbol{\beta}}_{\mathbf{x}} = \mathbf{\Psi}_1'\mathbf{\Lambda}\mathcal{X}_{12}(\mathcal{X}_{12}'\mathbf{\Lambda}\mathcal{X}_{12})^{-1}$ and $\hat{\boldsymbol{\beta}}_{\mathbf{x}_1} = \mathbf{\Psi}_1'\mathbf{\Lambda}\mathcal{X}_1(\mathcal{X}_1'\mathbf{\Lambda}\mathcal{X}_1)^{-1}$ are the coefficients of $\hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}$ and $\mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1}$, respectively. Note that $\mathbf{\Psi}_1 - \mathcal{X}_{12}\hat{\boldsymbol{\beta}}_{\mathbf{x}}'$ is the matrix of residuals in the GREG estimator $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{GR}|\mathbf{x}} = \tilde{\mathbf{t}}_{\mathbf{y}} + \hat{\boldsymbol{\beta}}_{\mathbf{x}}(\hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}})$ resulting from calibration involving only $\mathcal{X}_{12}$, and $\mathbf{\Psi}_1 - \mathcal{X}_1\hat{\boldsymbol{\beta}}_{\mathbf{x}_1}'$ is the matrix of residuals in the GREG estimator $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{GR}|\mathbf{x}_1} = \tilde{\mathbf{t}}_{\mathbf{y}} + \hat{\boldsymbol{\beta}}_{\mathbf{x}_1}(\mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1})$ resulting from calibration involving only $\mathcal{X}_1$. Then, using the orthogonality of $\mathcal{X}_1$ and $\mathcal{X}_{12}$, it is shown without difficulty that

$$\widehat{\mathrm{AV}}(\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{GR}}) - \widehat{\mathrm{Var}}(\tilde{\mathbf{t}}_{\mathbf{y}}) = \widehat{\mathrm{AV}}(\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{GR}|\mathbf{x}}) - \widehat{\mathrm{Var}}(\tilde{\mathbf{t}}_{\mathbf{y}}) + \widehat{\mathrm{AV}}(\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{GR}|\mathbf{x}_1}) - \widehat{\mathrm{Var}}(\tilde{\mathbf{t}}_{\mathbf{y}}),$$

which implies that the reduction of variance due to using the two auxiliary variables $\mathbf{x}_1$ and $\mathbf{x}$ in the regression (also calibration) procedure is additive. Thus, recalling Remark 3.2, the generalized regression estimator retains this additivity property of the BLUE of $\mathbf{t}_{\mathbf{y}}$.

6. Comparisons with existing methods

An earlier approach to optimal linear estimation in two-phase sampling designs, involving the standard type of auxiliary information considered in Sections 2 to 5, is described in Hidiroglou (2001). The formulation starts with postulating a regression form for the estimator of $\mathbf{t}_{\mathbf{y}}$, for univariate y, which is identical to the form of the estimator $\hat{\mathbf{t}}_{\mathbf{y}}^{B}$ in the first line of (3.2), and then the two unknown coefficients are determined so as to minimize the variance of this estimator. In estimating the two coefficients, the identities $\operatorname{Var}(\hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}) = \operatorname{Var}(\tilde{\mathbf{t}}_{\mathbf{x}}) - \operatorname{Var}(\hat{\mathbf{t}}_{\mathbf{x}})$ and $\operatorname{Cov}(\hat{\mathbf{t}}_{\mathbf{x}_1}, \hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}) = \mathbf{0}$ were ignored in the first and the second coefficient, respectively, and variances and covariances in both coefficients involving first-phase estimators were estimated using the second-phase sample only, thereby ignoring relevant information from the larger part of the first-phase sample. The resulting estimator was not shown to be a calibration estimator. In fact, this version of optimal estimator cannot be constructed as a calibration estimator. As a practicable variant of this, Hidiroglou (2001) considered a GREG estimator whose two coefficients (of $\mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1}$ and $\hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}$) can be justified either by assuming different regression models for each phase or by two successive calibrations. The same GREG estimator had been proposed earlier by Hidiroglou and Särndal (1998), but with no reference to optimal estimation. In the calibration approach of Hidiroglou and Särndal (1998), the estimator $\hat{\mathbf{t}}_{\mathbf{x}_1}$ is first calibrated to its total $\mathbf{t}_{\mathbf{x}_1}$, using s_1, and the GREG estimator $\hat{\mathbf{t}}_{\mathbf{x}|\mathbf{x}_1}^{\mathrm{GR}}$ of $\mathbf{t}_{\mathbf{x}}$ is then generated using the calibrated weigs, denoted by $\tilde{w}_{1k}$. Then the overall weig for $k \in s_2$ is formed as $\tilde{w}_k = \tilde{w}_{1k} w_{2k}$. In a second calibration involving s_2 and $\tilde{w}_k$, the estimator $\tilde{\mathbf{t}}_{\mathbf{x}}$ is calibrated to $\hat{\mathbf{t}}_{\mathbf{x}|\mathbf{x}_1}^{\mathrm{GR}}$. The resulting calibrated weigs of s_2 are then used to generate the GREG estimator of $\mathbf{t}_{\mathbf{y}}$, denoted here by $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{HS}}$.

Estevao and Särndal (2002, 2009) proposed a simpler version of the estimator $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{HS}}$, in which the overall design weigs $w_k = w_{1k} w_{2k}$ for $k \in s_2$ are used in the second calibration. Using current notation, this estimator, denoted here by $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{ES}}$, can be expressed in regression form as

$$
\begin{aligned}
\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{ES}} &= \tilde{\mathbf{t}}_{\mathbf{y}} + \mathbf{Y}_2' \boldsymbol{\Lambda}_2 \mathbf{X}_2 \left(\mathbf{X}_2' \boldsymbol{\Lambda}_2 \mathbf{X}_2\right)^{-1} (\hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}) \\
&\quad + \mathbf{Y}_2' \boldsymbol{\Lambda}_2 \mathbf{X}_2 \left(\mathbf{X}_2' \boldsymbol{\Lambda}_2 \mathbf{X}_2\right)^{-1} \mathbf{X}_1' \boldsymbol{\Lambda}_1 \mathbf{X}_{11} \left[\mathbf{X}_{11}' \boldsymbol{\Lambda}_1 \mathbf{X}_{11}\right]^{-1} (\mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1}).
\end{aligned}
\tag{6.1}
$$

Here the standard weiging matrices $\boldsymbol{\Lambda}_1 = \operatorname{diag}\{w_{1k}\}$ and $\boldsymbol{\Lambda}_2 = \operatorname{diag}\{w_k\}$ are used. Estevao and Särndal (2009) showed that this estimator is asymptotically equivalent to the estimator $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{HS}}$. For the estimator $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{ES}}$ in (6.1), Estevao and Särndal (2002) provide two linear regression representations corresponding to the two calibration steps. Replacing $\mathbf{y}$ by $\mathbf{x}$ in (6.1) gives $\hat{\mathbf{t}}_{\mathbf{x}}^{\mathrm{ES}}$, which is identical to $\hat{\mathbf{t}}_{\mathbf{x}}^{\mathrm{GR}}$ in (5.2).

In comparison, the regression estimator proposed in Section 5 is motivated by the single-step calibration structure of the optimal two-phase estimator, of which it serves as practical approximation. It derives its statistical and computational efficiency, relative to competing regression estimators assessed in this section, from a single-step calibration procedure involving the combined first-phase and second-phase samples, and in which first-phase and second-phase estimated totals are calibrated to each other. As a consequence, the regression coefficients of the terms $\mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1}$ and $\hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}$ incorporate information from the full sample s_1, as in the optimal estimator, and because of that they are more stable estimates of their population counterparts. An empirical comparison of the proposed regression estimator with the competing regression estimators is included in the simulation study in Section 7.

Replacing $\boldsymbol{\Lambda}_1$ and $\boldsymbol{\Lambda}_2$ by $\hat{\boldsymbol{\Delta}}_1$ and $\hat{\boldsymbol{\Delta}}_2$ in (6.1) converts the coefficient of the GREG estimator $\hat{\mathbf{t}}_{\mathbf{x}|\mathbf{x}_1}^{\mathrm{GR}}$ generated by the first step calibration into the coefficient $\widehat{\mathrm{Cov}}(\hat{\mathbf{t}}_{\mathbf{x}}, \hat{\mathbf{t}}_{\mathbf{x}_1})[\widehat{\mathrm{Var}}(\hat{\mathbf{t}}_{\mathbf{x}_1})]^{-1}$ of the single-phase optimal regression estimator $\hat{\mathbf{t}}_{\mathbf{x}|\mathbf{x}_1}^{O}$, and the coefficient of the GREG estimator $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{ES}}$ generated by the second step calibration into the coefficient $\widehat{\mathrm{Cov}}(\tilde{\mathbf{t}}_{\mathbf{y}}, \tilde{\mathbf{t}}_{\mathbf{x}})[\widehat{\mathrm{Var}}(\tilde{\mathbf{t}}_{\mathbf{x}})]^{-1}$. This latter coefficient may be viewed as pseudo-optimal since $\hat{\mathbf{t}}_{\mathbf{x}|\mathbf{x}_1}^{O}$ is treated as constant in the second step calibration, generating a pseudo-optimal estimator $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{PSO}}$. In turn, if in place of the sample matrices $\hat{\boldsymbol{\Delta}}_1$ and $\hat{\boldsymbol{\Delta}}_2$ in (6.1) we use the population matrices $\boldsymbol{\Delta}_1$ and $\boldsymbol{\Delta}_2$ we construct the pseudo-BLUE

$$\begin{aligned} \hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{PSB}} &= \tilde{\mathbf{t}}_{\mathbf{y}} + \mathbf{Y}_U' \boldsymbol{\Delta}_2 \mathbf{X}_U \left(\mathbf{X}_U' \boldsymbol{\Delta}_2 \mathbf{X}_U\right)^{-1} (\hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}) \\ &\quad + \mathbf{Y}_U' \boldsymbol{\Delta}_2 \mathbf{X}_U \left(\mathbf{X}_U' \boldsymbol{\Delta}_2 \mathbf{X}_U\right)^{-1} \mathbf{X}_U' \boldsymbol{\Delta}_1 \mathbf{X}_{1U} \left(\mathbf{X}_{1U}' \boldsymbol{\Delta}_1 \mathbf{X}_{1U}\right)^{-1} (\mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1}), \end{aligned} \tag{6.2}$$

where the coefficients of $\hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}$ and $\mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1}$ are, respectively, $\mathrm{Cov}(\tilde{\mathbf{t}}_{\mathbf{y}}, \tilde{\mathbf{t}}_{\mathbf{x}})\left[\mathrm{Var}(\tilde{\mathbf{t}}_{\mathbf{x}})\right]^{-1}$ and $\mathrm{Cov}(\tilde{\mathbf{t}}_{\mathbf{y}}, \tilde{\mathbf{t}}_{\mathbf{x}})\left[\mathrm{Var}(\tilde{\mathbf{t}}_{\mathbf{x}})\right]^{-1} \mathrm{Cov}(\hat{\mathbf{t}}_{\mathbf{x}}, \hat{\mathbf{t}}_{\mathbf{x}_1})\left[\mathrm{Var}(\hat{\mathbf{t}}_{\mathbf{x}_1})\right]^{-1}$. Thus, the GREG estimator (6.1) may be viewed as an approximation of $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{PSO}}$, which is the estimator $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{PSB}}$ with estimated coefficients (in analogy with the relationship of the optimal estimator $\hat{\mathbf{t}}_{\mathbf{y}}^{O}$ and the BLUE $\hat{\mathbf{t}}_{\mathbf{y}}^{B}$, in (4.2) and (3.14)). The pseudo-BLUE estimator $\hat{\mathbf{t}}_{\mathbf{x}}^{\mathrm{PSB}}$, obtained from (6.2) by replacing $\mathbf{y}$ with $\mathbf{x}$, is identical to the BLUE $\hat{\mathbf{t}}_{\mathbf{x}}^{B}$, in (3.14). On the other hand, the estimators $\hat{\mathbf{t}}_{\mathbf{y}}^{B}$ and $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{PSB}}$ are identical only under the condition of the following proposition; see proof in the Appendix.

Proposition 1. The estimators $\hat{\mathbf{t}}_{\mathbf{y}}^{B}$ and $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{PSB}}$ are identical only if $\boldsymbol{\Delta}_1 = \delta \boldsymbol{\Delta}_2$, for a constant δ.

Remark 6.1. The condition of Proposition 1 holds if the same equal-inclusion probability design is used in both phases; the constant δ is then a function of the sample inclusion probabilities. Two-phase designs that satisfy this condition are SRS and Bernoulli in both phases, as well as their stratification versions with identical stratification and proportional sample allocation in both phases. The practical importance of this is that for these designs the sample counterparts of $\hat{\mathbf{t}}_{\mathbf{y}}^{B}$ and $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{PSB}}$, i.e., $\hat{\mathbf{t}}_{\mathbf{y}}^{O}$ and $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{PSO}}$, will be for large samples almost identical. Furthermore, $\boldsymbol{\Delta}_1 = \delta \boldsymbol{\Delta}_2$ implies that the minus sign in the coefficient of $\hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}$ in (3.14) and (4.2) could change to plus sign, with $1 - \delta$ factoring out, and thus the singularity problem identified in Remark 4.1 will not exist.

Remark 6.2. It is simple to verify that $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{PSO}}$ is a calibration estimator, constructed by a two-step calibration procedure (as with the $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{ES}}$ estimator). Also, like $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{ES}}$, the estimator $\hat{\mathbf{t}}_{\mathbf{y}}^{\mathrm{PSO}}$ is formed using the calibrated weigs of the second-phase sample only.

Finally, it should be mentioned that a “design-optimal regression estimator”, not having the calibration property, has been proposed by Chen and Kim (2014) for a specific application and auxiliary variable setup. Also, a calibration estimator that is optimal under a model-assisted framework, and with $\mathbf{x}_2$ as the only auxiliary vector, has been proposed by Wu and Luan (2003).

7. Simulation study

We have conducted a simulation study to assess the performance of the proposed two-phase estimators of the total t_y, for scalar variables y, x_1 and x_2, and compare them with the competing regression estimators considered above. Distributions of these variables were specified as follows. The distribution of x_1 is the lognormal with mean and variance parameters $(\mu_{x_1} = 4, \sigma^2_{x_1} = 4)$. The distribution of x_2 is specified by the linear model $x_2 = 5 + x_1 + \epsilon$, where $\epsilon \sim N(0, \sigma^2_\epsilon)$, and the distribution of y is specified by the linear model $y = 10 + 2x_1 + 3x_2 + \eta$, where $\eta \sim N(0, \sigma^2_\eta)$. The value of σ^2_ϵ determines the linear relationship between x_2 and x_1, as defined by the population square correlation coefficient $r^2_{x_1, x_2}$, and the value of σ^2_η determines the linear relationship of y with x_1 and x_2, as defined by the coefficient of determination $r^2 = [r^2_{y,x_1} + r^2_{y,x_2} - 2r_{y,x_1} r_{y,x_2} r_{x_1,x_2}] \big/ (1 - r^2_{x_1,x_2})$.

Three values, 0, 0.25, 0.75, were specified for $r^2_{x_1,x_2}$, and two values, 0.25 and 0.75, for r^2, giving six combinations of values $(r^2_{x_1,x_2}, r^2)$. For the value $r^2_{x_1,x_2} = 0$, in particular, the bivariate lognormal distribution for (x_1, x_2) with parameters $(\mu_{x_1} = 4, \sigma^2_{x_1} = 4)$, $(\mu_{x_2} = 9, \sigma^2_{x_2} = 9)$ and zero correlation was used. The required values of σ^2_ϵ and σ^2_η are readily determined, while values for r^2_{y,x_1} and r^2_{y,x_2} are implicitly specified. For each of these six combinations, a population of size N = 50,000 was simulated by generating values from the distributions of the components of the vector (y, x_1, x_2). Four combinations of first-phase and second-phase sample sizes (n_1, n_2) with fixed n_1 and varying n_2 were specified, i.e., (3,000; 2,000), (3,000; 1,500), (3,000; 1,000), (3,000; 500), thus creating a total of 24 simulation settings.

Three different two-phase sampling designs were considered. Simple random sampling (SRS) without replacement was first used in both phases. For this sampling design, denoted by (SRS, SRS), the BLUE $\hat{t}^B_y$ in (3.10) and its exact variance in (3.11) can be calculated. Using the fact that under SRS the correlation of the HT estimators for two totals is identical to the correlation coefficient of the associated variables, tedious but straigforward algebra gives the relative difference (RDV) of variances of the estimators $\hat{t}^B_y$ and $\tilde{t}_y$ as

$$\frac{\operatorname{Var}(\tilde{t}_y) - \operatorname{Var}(\hat{t}^B_y)}{\operatorname{Var}(\tilde{t}_y)} = \frac{N(n_1 - n_2)}{n_1(N - n_2)} r^2 + \frac{n_2(N - n_1)}{n_1(N - n_2)} r^2_{y,x_1}.$$

The percent RDV is the measure of the efficiency of the BLUE $\hat{t}^B_y$ relative to the HT estimator $\tilde{t}_y$. This exact maximum efficiency will serve to measure the closeness of the approximation of the BLUE by the optimal estimator, for the different sample sizes, as well as the efficiency of the other competing estimators relative to the HT estimator. Notice that as n_2 tends to zero, the RDV tends to r^2, and as n_2 tends to n_1, the RDV tends to r^2_{y,x_1} (the efficiency of the BLUE based on s_1 and x_1). The second two-phase design, denoted by (STRSRS, SRS), was stratified simple random sampling (STRSRS) and SRS in the first and second phase, respectively. The simulated populations were stratified by the size of the variable y, with three strata of sizes N_1 = 30,000, N_2 = 15,000, N_3 = 5,000 and proportional allocation of the sample s_1 to the three strata – giving equal inclusion probabilities in each of the two phases. For this design too, the BLUE $\hat{t}^B_y$ and its exact variance can be calculated. The third two-phase design, denoted by

(SRS, PPSS), involved SRS in the first phase and probability proportional to size systematic sampling (PPSS) in the second phase, using as size measure the simple transformation $z_2 = 15 + 0.5 x_2$ of the variable x_2; using x_2 as size would result in $\hat{t}_{x_2} = \tilde{t}_{x_2}$. In this case the BLUE $\hat{t}_y^B$ (and the optimal estimator $\hat{t}_y^O$) cannot be calculated, because of the unknown probabilities π_{2kl}. However, GREG estimators can be calculated.

For each of these three two-phase designs, and all the 24 simulation settings, sampling was repeated 30,000 times, and each time we computed the estimators $\tilde{t}_y$, $\hat{t}_y^O$, $\hat{t}_y^{\text{GR}}$, $\hat{t}_y^{\text{ES}}$ and $\hat{t}_y^{\text{HS}}$, to obtain their empirical bias and variance. In all these cases, the simulation showed that the bias of all estimators was negligible, even for the smaller subsample sizes n_2. Thus their comparison is based on their variances relative to the benchmark variance of the HT estimator $\tilde{t}_y$. Specifically, the efficiency of each of the competing estimators $\hat{t}_y^O$, $\hat{t}_y^{\text{GR}}$, $\hat{t}_y^{\text{ES}}$ and $\hat{t}_y^{\text{HS}}$ is assessed through the percent relative difference between its empirical variance and the empirical variance of the estimator $\tilde{t}_y$; for example, for $\hat{t}_y^{\text{GR}}$ the relative difference is $\left[\text{Var}(\tilde{t}_y) - \text{Var}(\hat{t}_y^{\text{GR}})\right] \big/ \text{Var}(\tilde{t}_y)$. The relative difference shows the reduction of the variance of the particular estimator relative to the variance of the basic estimator $\tilde{t}_y$.

In the (SRS, SRS) design, the exact efficiency of the BLUE $\hat{t}_y^B$, relative to the HT estimator $\tilde{t}_y$, increases as n_2 decreases and as we move to higher values of (r_{y,x_2}^2, r^2), confirming Remark 3.1; see column 2 of Table 7.1. It is also confirmed that the efficiency of $\hat{t}_y^B$ tends to r^2 as n_2 decreases, faster for higher r_{x_1,x_2}^2. This maximum efficiency is closely approximated by the empirical efficiency of $\hat{t}_y^O$, even for the smaller subsample sizes n_2; see column 3 of Table 7.1. For the (STRSRS, SRS) design the exact efficiency of $\hat{t}_y^B$ is shown in column 6 of Table 7.1, exhibiting a pattern similar to that in the (SRS, SRS). This efficiency is closely approximated by the empirical efficiency of $\hat{t}_y^O$; see column 7 of Table 7.1. In both (SRS, SRS) and (STRSRS, SRS) the approximation of $\hat{t}_y^B$ by $\hat{t}_y^O$ is a little weaker in some settings involving the largest value of n_2, for the reason given in Remark 4.1.

Although the estimator $\hat{t}_y^O$ can be calculated in the (SRS, SRS) and (STRSRS, SRS) designs, the performance of the more practical, and of general applicability, calibration (GREG) estimators $\hat{t}_y^{\text{GR}}$ and $\hat{t}_y^{\text{ES}}$ is of great interest. For (SRS, SRS), the empirical efficiencies of these estimators are shown in columns 4 and 5 of Table 7.1. The negative sign indicates loss of efficiency with respect to the HT estimator. The efficiency of $\hat{t}_y^{\text{GR}}$ approximates closely the efficiency of $\hat{t}_y^O$, except for the four settings specified by $r_{x_1,x_2}^2 = 0$, $r^2 = 0.25, 0.75$ and $n_2 = 2{,}000; 1{,}500$; in particular, when $n_2 = 2{,}000$ the estimator $\hat{t}_y^{\text{GR}}$ is a little less efficient than the estimator $\tilde{t}_y$. In contrast, the estimator $\hat{t}_y^{\text{ES}}$ is less efficient than the estimator $\tilde{t}_y$ in six settings, when $r_{x_1,x_2}^2 = 0$, $r^2 = 0.25, 0.75$ and $n_2 = 2{,}000; 1{,}500; 1{,}000$; substantially less efficient when $n_2 = 2{,}000; 1{,}500$. The highlig in columns 4 and 5 is that the estimator $\hat{t}_y^{\text{GR}}$ is much more efficient than the estimator $\hat{t}_y^{\text{ES}}$ in all settings, more so for higher values of n_2 and for the higher values of (r_{y,x_2}^2, r^2); this indicates that $\hat{t}_y^{\text{GR}}$ is more effective in using information from the complement of s_2 and in exploiting higher correlations of y with x_1 and x_2. The efficiency of the estimator $\hat{t}_y^{\text{HS}}$ was virtually identical with that of $\hat{t}_y^{\text{ES}}$, in all three designs, and hence is not reported in Table 7.1. For (STRSRS, SRS), the empirical efficiencies of the calibration estimators $\hat{t}_y^{\text{GR}}$ and $\hat{t}_y^{\text{ES}}$ are shown in columns 8 and 9 of Table 7.1. It should be noted that the correlations within the strata are much weaker than the correlations for the whole population (shown in Table 7.1). Also, the HT estimator $\tilde{t}_y$ is highly efficient because of the stratification, especially for the larger values of n_2. The estimator $\hat{t}_y^{\mathrm{GR}}$ is less efficient than the estimator $\tilde{t}_y$ in 3 of the 24 settings, involving $n_2 = 2{,}000$, while for the rest its efficiency increases greatly as n_2 decreases, approaching the efficiency of $\hat{t}_y^{O}$. The estimator $\hat{t}_y^{\mathrm{ES}}$ is less efficient than the estimator $\tilde{t}_y$ in 12 settings. The estimator $\hat{t}_y^{\mathrm{GR}}$ is much more efficient than the estimator $\hat{t}_y^{\mathrm{ES}}$ in all settings, more so for higher values of n_2 and as we move from $r^2 = 0.25$ to $r^2 = 0.75$, and considerably more than in the (SRS, SRS) design.

Table 7.1
Percent efficiency of $\hat{t}_y^{B}$, $\hat{t}_y^{O}$, $\hat{t}_y^{\mathrm{GR}}$, $\hat{t}_y^{\mathrm{ES}}$ relative to $\tilde{t}_y$

	(SRS, SRS)				(STRSRS, SRS)				(SRS, PPSS)	
n_2	$\hat{t}_y^{B}$	$\hat{t}_y^{O}$	$\hat{t}_y^{\mathrm{GR}}$	$\hat{t}_y^{\mathrm{ES}}$	$\hat{t}_y^{B}$	$\hat{t}_y^{O}$	$\hat{t}_y^{\mathrm{GR}}$	$\hat{t}_y^{\mathrm{ES}}$	$\hat{t}_y^{\mathrm{GR}}$	$\hat{t}_y^{\mathrm{ES}}$
$\sigma_\eta^2 = 292.41,\ r^2_{x_1,x_2} = 0.00,\ r^2_{y,x_1} = 0.04,\ r^2_{y,x_2} = 0.21,\ r^2 = 0.25$										
2,000	11.54	10.09	-1.66	-26.88	16.74	13.69	-23.49	-79.94	-5.51	-30.38
1,500	15.00	13.84	10.91	-13.61	20.01	17.80	7.22	-38.46	4.57	-20.69
1,000	18.41	17.68	17.79	-0.71	22.22	21.20	19.34	-11.76	9.69	-10.22
500	21.74	20.62	20.77	11.29	23.81	22.34	22.75	7.84	11.24	0.67
$\sigma_\eta^2 = 32.15,\ r^2_{x_1,x_2} = 0.00,\ r^2_{y,x_1} = 0.13,\ r^2_{y,x_2} = 0.62,\ r^2 = 0.75$										
2,000	34.35	31.21	-4.23	-80.22	52.31	48.02	-66.06	-232.15	-21.68	-107.41
1,500	44.84	42.34	33.09	-41.49	61.53	59.02	24.66	-108.30	15.95	-74.74
1,000	55.10	53.80	53.83	-2.25	67.58	65.48	59.17	-30.84	36.49	-39.03
500	65.16	63.87	63.90	34.31	71.85	70.53	70.41	27.83	45.55	2.00
$\sigma_\epsilon^2 = 12.11,\ \sigma_\eta^2 = 632.52,\ r^2_{x_1,x_2} = 0.25,\ r^2_{y,x_1} = 0.12,\ r^2_{y,x_2} = 0.24,\ r^2 = 0.25$										
2,000	16.70	16.89	16.69	9.88	17.14	15.56	2.08	-23.18	10.24	3.03
1,500	18.85	19.02	19.52	13.57	20.26	19.45	16.46	-3.21	13.83	7.08
1,000	20.97	20.79	20.52	16.50	22.38	21.07	21.04	6.84	13.03	8.04
500	23.04	22.28	21.67	19.64	23.91	22.84	23.41	16.40	11.57	9.38
$\sigma_\epsilon^2 = 12.11,\ \sigma_\eta^2 = 70.68,\ r^2_{x_1,x_2} = 0.25,\ r^2_{y,x_1} = 0.36,\ r^2_{y,x_2} = 0.71,\ r^2 = 0.75$										
2,000	49.70	48.33	46.89	25.33	53.11	50.78	20.11	-38.15	35.49	11.64
1,500	56.23	55.48	56.46	38.08	61.86	60.63	53.81	8.36	46.33	22.98
1,000	62.63	62.20	61.35	48.69	67.71	66.38	66.10	34.90	47.91	30.19
500	68.90	68.09	65.58	59.65	71.90	70.99	71.00	55.27	47.68	38.93
$\sigma_\epsilon^2 = 1.33,\ \sigma_\eta^2 = 340.40,\ r^2_{x_1,x_2} = 0.75,\ r^2_{y,x_1} = 0.22,\ r^2_{y,x_2} = 0.24,\ r^2 = 0.25$										
2,000	23.36	23.67	23.01	10.81	18.09	15.47	-6.07	-46.54	16.62	3.38
1,500	23.78	23.63	24.19	13.85	20.83	19.60	14.52	-17.17	19.77	7.95
1,000	24.20	23.83	23.15	16.97	22.68	21.67	21.58	0.86	17.04	10.02
500	24.61	23.54	22.24	19.92	24.01	22.39	22.98	13.24	14.52	11.77
$\sigma_\epsilon^2 = 1.33,\ \sigma_\eta^2 = 37.82,\ r^2_{x_1,x_2} = 0.75,\ r^2_{y,x_1} = 0.67,\ r^2_{y,x_2} = 0.72,\ r^2 = 0.75$										
2,000	69.84	67.98	65.10	26.96	60.26	56.75	32.34	-27.50	56.65	13.24
1,500	71.17	69.57	70.70	38.91	66.25	64.49	59.73	14.39	65.49	26.11
1,000	72.47	71.26	69.17	49.62	70.17	68.80	68.69	40.44	61.00	35.28
500	73.74	72.19	67.58	60.66	72.94	71.12	71.10	56.90	54.67	44.54

SRS = Simple random sampling; STRSRS = stratified simple random sampling; PPSS = probability proportional to size systematic.

For (SRS, PPSS), the empirical efficiencies of the calibration estimators $\hat{t}_y^{\mathrm{GR}}$ and $\hat{t}_y^{\mathrm{ES}}$ are shown in columns 10 and 11 of Table 7.1. The pattern of these efficiencies is very similar to that in the (SRS, SRS) design. This is particularly so for the efficiency of $\hat{t}_y^{\mathrm{GR}}$ relative to $\hat{t}_y^{\mathrm{ES}}$, which is not included in Table 7.1

but can be easily derived using the displayed efficiencies of $\hat{t}_y^{\mathrm{GR}}$ and $\hat{t}_y^{\mathrm{ES}}$ relative to $\tilde{t}_y$. The HT estimator $\tilde{t}_y$ itself is more efficient with this two-phase design, which explains why the efficiency of the two calibration estimators $\hat{t}_y^{\mathrm{GR}}$ and $\hat{t}_y^{\mathrm{ES}}$ relative to $\tilde{t}_y$ is somewhat lower than in the (SRS, SRS) and (STRSRS, SRS) designs.

The whole simulation study was repeated with the simulated population for the vector (y, x_1, x_2) generated from a trivariate lognormal distribution with the specified correlation structures. For all three designs (SRS, SRS), (SRS, PPSS) and (STRSRS, SRS), the results (not shown here) were very similar to those based on the linear model for y used above.

It is of interest to consider the setup of auxiliary variables in which the scalar variable x_1 is augmented to $(1, x_1)$, with known totals (N, t_{x_1}). Then in the (SRS, SRS) design, in which construction of the BLUE $\hat{t}_y^B$ and the optimal estimator $\hat{t}_y^O$ is feasible, using the complete setup $(1, x_1, x_2)$ in calibration gives the same $\hat{t}_y^B$ and practically the same $\hat{t}_y^B$ as when using (x_1, x_2). It would also convert the regression estimator $\hat{t}_y^{\mathrm{GR}}$ to $\hat{t}_y^O$ (using the same adjustment $1/\pi_{2k}$ of $\mathbf{Y}_1$ as in $\hat{t}_y^O$), and the regression estimator $\hat{t}_y^{\mathrm{ES}}$ estimator to the pseudo-optimal estimator $\hat{t}_y^{\mathrm{PSO}}$ (defined in Section 6). These properties are derived from known theory, see for example Merkouris (2004, 2015), more directly for $\hat{t}_y^{\mathrm{ES}}$ and the second regression term of $\hat{t}_y^{\mathrm{GR}}$ and the optimal $\hat{t}_y^O$, irrespective of any specific functional relationship of y with $(1, x_1, x_2)$. Then, the three sample-based estimators would show virtually identical empirical behavior. This follows from Proposition 1, which gives the condition (satisfied by specific designs, including (SRS, SRS)) under which the pseudo-optimal regression estimator $\hat{t}_y^{\mathrm{PSO}}$ is asymptotically equivalent to the proposed optimal estimator $\hat{t}_y^O$. Experimental calculations have confirmed this equivalence. In the (STRSRS, SRS) design too, using $(1, x_1, x_2)$ gives the same $\hat{t}_y^B$ and $\hat{t}_y^O$ as when using (x_1, x_2), and converts the $\hat{t}_y^{\mathrm{GR}}$ and $\hat{t}_y^{\mathrm{ES}}$ estimators to the $\hat{t}_y^O$ and $\hat{t}_y^{\mathrm{PSO}}$ estimators, respectively. However, by Proposition 1 the equivalence of the latter two estimators, and hence of $\hat{t}_y^{\mathrm{GR}}$ and $\hat{t}_y^{\mathrm{ES}}$, does not hold in this sampling design.

8. Discussion

The described method of optimal and regression estimation for two-phase sampling involves a single-step calibration of the weigs of the combined first-and-second phase samples. Thus, using a single set of calibrated weigs that incorporate all the available information from the two phases, a substantially improved estimate of the total of a target variable can be obtained, as shown by the simulation study. These weigs could be used to calculate other weiged statistics, including means, ratios, quantiles and regression coefficients. The framework of the method is general enough to encompass complex designs with multiple stages and different stratification at the two phases, as well as various types of auxiliary variables known at the population or sample level – ten different cases of auxiliary information are identified in Estevao and Särndal (2002). Furthermore, the method may be extended to multi-phase sampling designs through the appropriate calibration setup.

Estimation of a total for any domain (subpopulation) of interest, $U_d \subset U$, can be carried out readily using the calibrated weigs and summing the weiged sample values of the variable of interest over U_d. For the resulting domain estimator to be optimal linear estimator, the domain estimates of $\mathbf{t}_{\mathbf{y}}$, $\mathbf{t}_{\mathbf{x}_1}$ and $\mathbf{t}_{\mathbf{x}_2}$ need to be combined linearly, by carrying out optimal calibration at the domain level with domain calibration totals and with the appropriate modification of the matrix $\mathcal{X}$. A number of calibration options, regarding the use of the available auxiliary information at the population, domain and two-phase sample levels, could be considered for the most efficient estimation of domain totals in any particular application. Related work in Merkouris (2010) would be helpful in this context.

The estimated approximate variances of the two-phase optimal estimator and the two-phase regression estimator, based on Taylor linearization, were given in Sections 4.1 and Section 5, respectively. For the two-phase regression estimator, replication methods of variance estimation, such as the jackknife method or the bootstrap method, could be alternatively applied, or would be the only option when first-phase or second-phase joint inclusion probabilities are not known. There is extensive literature on such replication methods for existing regression estimators in two-phase sampling. The single-step calibration feature of the proposed regression estimation method may be helpful in this direction; detailed study of this is beyond the scope of this paper.

Acknowledgements

The author thanks the Editor, the Associate Editor and the two referees for their comments and suggestions, which have led to a considerable improvement of the manuscript.

Appendix

Proof of Lemma 1

The symmetric matrix $\operatorname{Var}(\mathbf{w}_U^*)$ has the form of (3.8) but with $\operatorname{Cov}(\mathbf{w}_{1U}, \mathbf{w}_U)$ as off-diagonal block. The kl^{th} element of the matrix $\operatorname{Var}(\mathbf{w}_{1U})$ is

$$\operatorname{Cov}(w_{1U_k}, w_{1U_l}) = [E(I_{1k} I_{1l}) - E(I_{1k}) E(I_{1l})] / \pi_{1k} \pi_{1l} = (\pi_{1kl} - \pi_{1k}\pi_{1l}) / \pi_{1k} \pi_{1l}.$$

The kl^{th} element of the matrix $\operatorname{Var}(\mathbf{w}_U)$ is

$$\begin{aligned}\operatorname{Cov}(w_{U_k}, w_{U_l}) &= [E(I_{1k} I_{2k} I_{1l} I_{2l}) - E(I_{1k} I_{2k}) E(I_{1l} I_{2l})] / \pi_{1k} \pi_{2k} \pi_{1l} \pi_{2l} \\ &= [E_1(I_{1k} I_{1l} E_2(I_{2k} I_{2l})) - E_1(I_{1k} E_2(I_{2k})) E_1(I_{1l} E_2(I_{2l}))] / \pi_{1k} \pi_{2k} \pi_{1l} \pi_{2l} \\ &= [\pi_{1kl} \pi_{2kl} - \pi_{1k} \pi_{2k} \pi_{1l} \pi_{2l}] / \pi_{1k} \pi_{2k} \pi_{1l} \pi_{2l},\end{aligned}$$

where E_1 and E_2 denote expectation under first and second phase of sampling, respectively. Using similar arguments it follows that the kl^{th} element of the matrix $\operatorname{Cov}(\mathbf{w}_{1U}, \mathbf{w}_U)$ is

$$\mathrm{Cov}(w_{1U_k}, w_{U_l}) = [E(I_{1k}\, I_{1l}\, I_{2l}) - E(I_{1k})\, E(I_{1l}\, I_{2l})] \,/\, \pi_{1k}\, \pi_{1l}\, \pi_{2l} = (\pi_{1kl} - \pi_{1k}\, \pi_{1l}) \,/\, \pi_{1k}\, \pi_{1l}.$$

This shows that $\mathrm{Cov}(w_{1U_k}, w_{U_l}) = \mathrm{Cov}(w_{1U_k}, w_{1U_l})$ and thus $\mathrm{Cov}(\mathbf{w}_{1U}, \mathbf{w}_U) = \mathrm{Var}(\mathbf{w}_{1U})$, which completes the proof.

Proof of Theorem 1

Matrix $\boldsymbol{\Delta} = \mathrm{Var}(\mathbf{w}_U^*)$ is nonsingular if and only if $\mathrm{Var}(\mathbf{w}_U) - \mathrm{Var}(\mathbf{w}_{1U})$ is nonsingular. This follows from a general result on inverses of partitioned matrices (see Harville, 2008, page 98). But $\mathrm{Var}(\mathbf{w}_U) - \mathrm{Var}(\mathbf{w}_{1U}) = \mathrm{Var}(\mathbf{w}_{1U} - \mathbf{w}_U)$, because $\mathrm{Cov}(\mathbf{w}_{1U}, \mathbf{w}_U) = \mathrm{Var}(\mathbf{w}_{1U})$, and therefore $\mathrm{Var}(\mathbf{w}_U) - \mathrm{Var}(\mathbf{w}_{1U})$ is nonsingular, being a variance-covariance matrix. Next, to find the vector $\mathbf{c}_U^*$ that minimizes $(\mathbf{c}_U^* - \mathbf{w}_U^*)' \boldsymbol{\Delta}^{-1} (\mathbf{c}_U^* - \mathbf{w}_U^*)$ subject to the constraints $\boldsymbol{\mathcal{X}}_U' \mathbf{c}_U^* = \mathbf{t}_{\boldsymbol{\mathcal{X}}}$, consider the function $\mathbf{F} = (\mathbf{c}_U^* - \mathbf{w}_U^*)' \boldsymbol{\Delta}^{-1} (\mathbf{c}_U^* - \mathbf{w}_U^*) - \boldsymbol{\lambda}' \boldsymbol{\mathcal{X}}_U' \mathbf{c}_U^*$ where $\boldsymbol{\lambda}$ is a vector of Langrange multipliers. We then get the system of equations

$$\begin{aligned} \frac{\partial \mathbf{F}}{\partial \mathbf{c}_U^*} \;=\; 2\boldsymbol{\Delta}^{-1}(\mathbf{c}_U^* - \mathbf{w}_U^*) - \boldsymbol{\mathcal{X}}_U \boldsymbol{\lambda} &= \mathbf{0} \\ \boldsymbol{\mathcal{X}}_U' \mathbf{c}_U^* - \mathbf{t}_{\boldsymbol{\mathcal{X}}} &= \mathbf{0}. \end{aligned}$$

Multiplying the first equation by $\boldsymbol{\mathcal{X}}_U' \boldsymbol{\Delta}$, using $\boldsymbol{\mathcal{X}}_U' \mathbf{c}_U^* = \mathbf{t}_{\boldsymbol{\mathcal{X}}}$ and solving for $\boldsymbol{\lambda}$ gives $\boldsymbol{\lambda} = 2(\boldsymbol{\mathcal{X}}_U' \boldsymbol{\Delta} \boldsymbol{\mathcal{X}}_U)^{-1}\left(\mathbf{t}_{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}_U' \mathbf{w}_U^*\right)$. Inserting this into the first equation and solving for $\mathbf{c}_U^*$ gives $\mathbf{c}_U^* = \mathbf{w}_U^* + \boldsymbol{\Delta} \boldsymbol{\mathcal{X}}_U (\boldsymbol{\mathcal{X}}_U' \boldsymbol{\Delta} \boldsymbol{\mathcal{X}}_U)^{-1}\left(\mathbf{t}_{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}_U' \mathbf{w}_U^*\right)$.

Proof of Proposition 1

Clearly, the coefficients of $\hat{\mathbf{t}}_{\mathbf{x}} - \tilde{\mathbf{t}}_{\mathbf{x}}$ in (3.14) and (6.2) are identical if $\boldsymbol{\Delta}_1 = \delta \boldsymbol{\Delta}_2$. Next, using the partition $\mathbf{X}_U = (\mathbf{X}_{1U}, \mathbf{X}_{2U})$, the coefficient of $\mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1}$ in (6.2) is expressed as follows. First we obtain

$$\begin{aligned} \mathbf{Y}_U' \boldsymbol{\Delta}_2 \mathbf{X}_U \left(\mathbf{X}_U' \boldsymbol{\Delta}_2 \mathbf{X}_U\right)^{-1} \mathbf{X}_U' \boldsymbol{\Delta}_1 \mathbf{X}_{1U} &= \mathbf{Y}_U' \boldsymbol{\Delta}_2 (\mathbf{X}_{1U}, \mathbf{X}_{2U}) \begin{pmatrix} \mathbf{X}_{1U}' \boldsymbol{\Delta}_2 \mathbf{X}_{1U} & \mathbf{X}_{1U}' \boldsymbol{\Delta}_2 \mathbf{X}_{2U} \\ \mathbf{X}_{2U}' \boldsymbol{\Delta}_2 \mathbf{X}_{1U} & \mathbf{X}_{2U}' \boldsymbol{\Delta}_2 \mathbf{X}_{2U} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}_{1U}' \boldsymbol{\Delta}_1 \mathbf{X}_{1U} \\ \mathbf{X}_{2U}' \boldsymbol{\Delta}_1 \mathbf{X}_{1U} \end{pmatrix} \\ &= \mathbf{Y}_U' \boldsymbol{\Delta}_2 \mathbf{X}_{1U} \left[A_{11} (\mathbf{X}_{1U}' \boldsymbol{\Delta}_1 \mathbf{X}_{1U}) + A_{12} (\mathbf{X}_{2U}' \boldsymbol{\Delta}_1 \mathbf{X}_{1U})\right] \\ &\quad + \mathbf{Y}_U' \boldsymbol{\Delta}_2 \mathbf{X}_{2U} \left[A_{21} (\mathbf{X}_{1U}' \boldsymbol{\Delta}_1 \mathbf{X}_{1U}) + A_{22} (\mathbf{X}_{2U}' \boldsymbol{\Delta}_1 \mathbf{X}_{1U})\right], \end{aligned}$$

where A_{11}, A_{12}, A_{21}, A_{22} are derived by algebra of inverses of partitioned matrices. In particular,

$$A_{11} = (\mathbf{X}_{1U}' \boldsymbol{\Delta}_2 \mathbf{X}_{1U})^{-1} - A_{12} (\mathbf{X}_{2U}' \boldsymbol{\Delta}_2 \mathbf{X}_{1U}) (\mathbf{X}_{1U}' \boldsymbol{\Delta}_2 \mathbf{X}_{1U})^{-1}$$

and $A_{21} = -A_{22} (\mathbf{X}_{2U}' \boldsymbol{\Delta}_2 \mathbf{X}_{1U}) \times (\mathbf{X}_{1U}' \boldsymbol{\Delta}_2 \mathbf{X}_{1U})^{-1}$. Then,

$$\begin{aligned} A_{11} (\mathbf{X}_{1U}' \boldsymbol{\Delta}_1 \mathbf{X}_{1U}) + A_{12} (\mathbf{X}_{2U}' \boldsymbol{\Delta}_1 \mathbf{X}_{1U}) &= (\mathbf{X}_{1U}' \boldsymbol{\Delta}_2 \mathbf{X}_{1U})^{-1} \mathbf{X}_{1U}' \boldsymbol{\Delta}_1 \mathbf{X}_{1U} + A_{12} \mathbf{B} \\ A_{21} (\mathbf{X}_{1U}' \boldsymbol{\Delta}_1 \mathbf{X}_{1U}) + A_{22} (\mathbf{X}_{2U}' \boldsymbol{\Delta}_1 \mathbf{X}_{1U}) &= A_{22} \mathbf{B}, \end{aligned}$$

where

$$\mathbf{B} = \mathbf{X}'_{2U}\,\boldsymbol{\Delta}_1\,\mathbf{X}_{1U} - \mathbf{X}'_{2U}\,\boldsymbol{\Delta}_2\,\mathbf{X}_{1U}\,(\mathbf{X}'_{1U}\,\boldsymbol{\Delta}_2\,\mathbf{X}_{1U})^{-1}(\mathbf{X}'_{1U}\,\boldsymbol{\Delta}_1\,\mathbf{X}_{1U}).$$

It is then easy to verify that if $\boldsymbol{\Delta}_1 = \delta\boldsymbol{\Delta}_2$, we have $(\mathbf{X}'_{1U}\,\boldsymbol{\Delta}_2\,\mathbf{X}_{1U})^{-1}\mathbf{X}'_{1U}\,\boldsymbol{\Delta}_1\,\mathbf{X}_{1U} = \delta\mathbf{I}$ and $\mathbf{B} = \mathbf{0}$. It follows that $\mathbf{Y}'_U\,\boldsymbol{\Delta}_2\,\mathbf{X}_U\left(\mathbf{X}'_U\,\boldsymbol{\Delta}_2\,\mathbf{X}_U\right)^{-1}\mathbf{X}'_U\,\boldsymbol{\Delta}_1\,\mathbf{X}_{1U} = \mathbf{Y}'_U\,\boldsymbol{\Delta}_1\,\mathbf{X}_{1U}$, and thus the coefficients of $\mathbf{t}_{\mathbf{x}_1} - \hat{\mathbf{t}}_{\mathbf{x}_1}$ in (3.14) and (6.2) are also identical.

References

Australian Bureau of Statistics (2004). Estimation for the household income and expenditure survey. Research paper 1352.0.55.063.

Beaumont, J.-F., Beliveau, A. and Haziza, D. (2015). Clarifying some aspects of variance estimation in two-phase sampling. *Journal of Survey Statistics and Methodology*, 3, 524-542.

Brick, J.M., and Tourangeau, R. (2017). Responsive survey designs for reducing nonresponse bias. *Journal of Official Statistics*, 33, 735-752.

Chen, S., and Kim, J.K. (2014). Two-phase sampling experiment for propensity score estimation in self-selected samples. *The Annals of Applied Statistics*, 3, 1492-1515.

Chipperfield, J.O., and Steel, D.G. (2009). Design and estimation for split questionnaire surveys. *Journal of Official Statistics*, 25, 227-244.

Estevao, V.M., and Särndal, C.-E. (2002). The ten cases of auxiliary information for calibration in two-phase sampling. *Journal of Official Statistics*, 18, 233-255.

Estevao, V.M., and Särndal, C.-E. (2009). A new face on two-phase sampling with calibration estimators. *Survey Methodology*, 35, 1, 3-14. Paper available at tps://www150.statcan.gc.ca/n1/en/pub/12-001-x/2009001/article/10880-eng.pdf.

Fuller, W.A. (1990). Analysis of repeated surveys. *Survey Methodology*, 16, 2, 167-180. Paper available at tps://www150.statcan.gc.ca/n1/en/pub/12-001-x/1990002/article/14537-eng.pdf.

Fuller, W.A. (1998). Replication variance estimation for two-phase sampling. *Statistica Sinica*, 8, 1153-1164.

Fuller, W.A., and Isaki, C.T. (1981). Survey design under superpopulation models. In *Current Topics in Survey Sampling*, (Eds., D. Krewski, J.N.K. Rao and R. Platek), New York: Academic Press, 199-226.

Groves, R.M., and Heeringa, S.G. (2006). Responsive design for household surveys: Tools for actively controlling survey errors and costs. *Journal of the Royal Statistical Society: Series A*, 169, 439-457.

Harville, D.A. (2008). *Matrix Algebra from a Statistician's Perspective*. New York: Springer.

Haziza, D., Mecatti, F. and Rao, J.N.K. (2008). Evaluation of some approximate variance estimators under the Rao-Sampford unequal probability sampling design. *METRON-International Journal of Statistics*, vol LXVI, 91-108.

Hidiroglou, M.A. (2001). Double sampling. *Survey Methodology*, 27, 2, 143-154. Paper available at tps://www150.statcan.gc.ca/n1/en/pub/12-001-x/2001002/article/6091-eng.pdf.

Hidiroglou, M.A., and Särndal, C.-E. (1998). Use of auxiliary information for two-phase sampling. *Survey Methodology*, 24, 1, 11-20. Paper available at tps://www150.statcan.gc.ca/n1/en/pub/12-001-x/1998001/article/3905-eng.pdf.

Hidiroglou, M.A., Rao, J.N.K. and Haziza, D. (2008). Variance estimation in two-phase sampling. *Australian and New Zealand Journal of Statistics*, 51, 127-141.

Jones, R.G. (1980). Best linear unbiased estimators for repeated surveys. *Journal of the Royal Statistical Society, Ser. B*, 42, 221-226.

Kim, J.K., and Sitter, R.R. (2003). Efficient replication variance estimation for two-phase sampling. *Statistica Sinica*, 13, 641-653.

Kim, J.K., and Yu, C.L. (2011). Replication variance estimation under two-phase sampling. *Survey Methodology*, 37, 1, 67-74. Paper available at tps://www150.statcan.gc.ca/n1/en/pub/12-001-x/2011001/article/11448-eng.pdf.

Kim, J.K., Navarro, A. and Fuller, W.A. (2006). Replication variance estimation for two-phase stratified sampling. *Journal of the American Statistical Association*, 101, 311-320.

Merkouris, T. (2004). Combining independent regression estimators from multiple surveys. *Journal of the American Statistical Association*, 99, 1131-1139.

Merkouris, T. (2010). Combining information from multiple surveys by using regression for more efficient small domain estimation. *Journal of the Royal Statistical Society, Ser. B*, 72, 27-48.

Merkouris, T. (2015). An efficient estimation method for matrix survey sampling. *Survey Methodology*, 41, 1, 237-262. Paper available at tps://www150.statcan.gc.ca/n1/en/pub/12-001-x/2015001/article/14174-eng.pdf.

Montanari, G.E. (1987). Post-sampling efficient QR-prediction in large-scale surveys. *International Statistics Review*, 55, 191-202.

Rao, J.N.K. (1994). Estimating totals and distribution functions using auxiliary information at the estimation stage. *Journal of Official Statistics*, 10, 153-165.

Sitter, R.R. (1997). Variance estimation for the regression estimator in two-phase sampling. *Journal of the American Statistical Association*, 92, 780-787.

Särndal, C.-E., Swensson, B. and Wretman, J.H. (1992). *Model-Assisted Survey Sampling*. New York: Springer.

Turmelle, C., and Beaucage, Y. (2013). The integrated business statistics program: Using a two-phase design to produce reliable estimates. *Proceedings: Symposium 2013, Producing reliable estimates from imperfect frames.*

Wolter, K.M. (1979). Composite estimation in finite populations. *Journal of the American Statistical Association*, 74, 604-613.

Wu, C., and Luan, Y. (2003). Optimal calibration estimators under two-phase sampling. *Journal of Official Statistics*, 2, 119-131.