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

Survey Methodology 41-1

An efficient estimation method for matrix survey sampling

by Takis Merkouris

Release date: June 29, 2015

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 toll-free numbers:

- Statistical Information Service 1-800-263-1136
- National telecommunications device for the hearing impaired 1-800-363-7629
- Fax line 1-877-287-4369

Depository Services Program

- Inquiries line 1-800-635-7943
- Fax line 1-800-565-7757

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 "About us" > "The agency" > "Providing services to Canadians."

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.

Standard table symbols

The following symbols are used in Statistics Canada publications:

. not available for any reference period
.. not available for a specific reference period
... not applicable
0 true zero or a value rounded to zero
0^s value rounded to 0 (zero) where there is a meaningful distinction between true zero and the value that was rounded
^{p} preliminary
^{r} revised
x suppressed to meet the confidentiality requirements of the *Statistics Act*
^{E} use with caution
F too unreliable to be published
* significantly different from reference category ($p < 0.05$)

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, June 2015
Vol. 41, No. 1, pp. 237-262
Statistics Canada, Catalogue No. 12-001-X

An efficient estimation method for matrix survey sampling

Takis Merkouris[1]

Abstract

Matrix sampling, often referred to as split-questionnaire, is a sampling design that involves dividing a questionnaire into subsets of questions, possibly overlapping, and then administering each subset to one or more different random subsamples of an initial sample. This increasingly appealing design addresses concerns related to data collection costs, respondent burden and data quality, but reduces the number of sample units that are asked each question. A broadened concept of matrix design includes the integration of samples from separate surveys for the benefit of streamlined survey operations and consistency of outputs. For matrix survey sampling with overlapping subsets of questions, we propose an efficient estimation method that exploits correlations among items surveyed in the various subsamples in order to improve the precision of the survey estimates. The proposed method, based on the principle of best linear unbiased estimation, generates composite optimal regression estimators of population totals using a suitable calibration scheme for the sampling weights of the full sample. A variant of this calibration scheme, of more general use, produces composite generalized regression estimators that are also computationally very efficient.

Key Words: Best linear unbiased estimator; Calibration; Composite estimator; Generalized regression estimator; Non-nested matrix sampling; Split-questionnaire.

1 Introduction

Matrix sampling is a sampling design in which a long questionnaire is divided into subsets of questions (items), possibly overlapping, and each subset is then administered to one or more distinct random subsamples of an initial sample. In its various forms this design may serve a variety of purposes, such as reducing the length and cost of the survey process and addressing concerns related to respondent burden and data quality associated with a long questionnaire. Matrix sampling has been applied or explored in various fields, primarily in educational assessment and public health studies. A review of previous research on matrix sampling, with discussion of the issues arising in its implementation in surveys, is given in Gonzalez and Eltinge (2007). For recent work on design and estimation for matrix survey sampling, motivated by the potential benefits of such sampling schemes in large scale surveys, see Raghunathan and Grizzle (1995), Thomas, Raghunathan, Schenker, Katzoff and Johnson (2006), Gonzalez and Eltinge (2008), Chipperfield and Steel (2009, 2011), and references therein. Among the many matrix sampling designs explored in the literature, we distinguish the following four principal designs varying in the number of subsamples and the number of sub-questionnaires (overlapping or not) administered to each subsample.

(a) Different (non-overlapping) sets of questions are administered to different subsamples.
(b) An additional core set of questions is administered to all subsamples in design (a). There are several reasons for including a core set of items in all subsamples: High precision may be required for some items of special interest; some other items (e.g., demographic characteristics) define subpopulations and may be used in cross-tabulations of survey results; the correlation of the core items with the rest of items may be used to enhance the precision of estimates for all items.

1. Takis Merkouris, Department of Statistics, Athens University of Economics and Business, Patision 76, Athens 10434, Greece. E-mail: merkouris@aueb.gr.

(c) A variant of design (a) involving an additional subsample that receives the full questionnaire. It may be viewed as a generalization of two-phase sampling design. The motivation for this design is to allow for analysis of interaction between sets of questions, by having responses to all questions from the units of the additional sample, and to enable more efficient estimation.

(d) An extension of design (c), in which the core set of questions is administered to all subsamples. It embodies all features of the previous three designs.

A current trend in survey planning relates to a variant of matrix sampling in which a number of distinct surveys with overlapping content are integrated for the benefit of streamlined survey operations, harmonized survey content and data consistency, as well as improved estimation. In this nonstandard matrix sampling framework, the distinct surveys may use subsamples of a large master sample or independent samples from the same population. Such sampling schemes are actively being researched or implemented in various statistical agencies; see, for example, the integration of household surveys in the British Office of National Statistics (Smith 2009) and in the Australian Bureau of Statistics (2011). Although such integration may be viewed as a reverse process to splitting a questionnaire, the structure of the design with respect to the collection of different subsets of data items from different samples is essentially the same as in the standard framework. In the particular case where the samples from constituent surveys are independent, possibly with different sampling designs, the designs (b), (c) and (d) could be characterized as non-nested matrix sampling designs. It is to be noted that the advantages of matrix sampling are not always contingent on using subsamples (necessarily dependent) of an initial sample. It may be more practical in certain situations to use independent samples, notwithstanding the possibility of a negligible sample overlap.

In this paper we address the estimation problem in matrix sampling, namely the loss of precision of survey estimates due to not collecting all data items from all sample units. In the nonstandard matrix sampling of the preceding paragraph, the estimation problem is the improvement of the precision of estimates for each constituent survey. For matrix sampling designs (b), (c) and (d), involving overlapping subsets of questions, a dual estimation task is to combine data on common items from different subsamples for improved estimation, and to exploit correlations among items surveyed in different subsamples for more efficient estimation for all items. To this aim, estimation involving imputation of the missing values caused by the omitted items in each subquestionnaire has been explored in Raghunathan and Grizzle (1995) and Thomas et al. (2006). Estimation using a simple weighting adjustment that combines data on common items has been considered by Gonzalez and Eltinge (2008). In the particular case of non-nested design (b), the estimation problem of combining data from independent samples has also been dealt with in the literature; see, for example, Renssen and Nieuwenbroek (1997), Houbiers (2004), Merkouris (2004, 2010), Wu (2004) and Kim and Rao (2012). Non-nested design (d) has been considered in Renssen (1998). We propose an efficient estimation method, based on the principle of best linear unbiased estimation, which produces composite optimal regression estimators of totals by means of a suitable calibration procedure for the sampling weights of the combined sample, when the second-order sample inclusion probabilities are known. A variant of this calibration procedure of more general applicability produces composite generalized regression estimators, which for certain sampling settings are optimal regression estimators. The method exploits correlations of items across the subsamples to improve the efficiency of estimators even for items surveyed in all subsamples. It is also operationally very convenient, producing estimates for all items at population or domain level by means of a simple adaptation of the standard calibration system commonly used in statistical agencies. Introducing here the method, we study in detail the principal designs (c) and (d). Adaptations to more general designs are fairly straightforward.

In the following Section 2 and Section 3 we describe the proposed method for design (c). The application of the method to design (d) is described in Section 4. Domain estimation is dealt with in Section 5. A simulation study is presented in Section 6. We conclude with a discussion in Section 7.

2 Composite optimal regression estimation for design (c)

A general estimation method for matrix sampling is illustrated for design (c) through the simplest setting involving three samples S_1, S_2 and S_3 with arbitrary designs and sizes n_1, n_2, n_3, which may be subsamples of an initial sample of size $n = n_1 + n_2 + n_3$ from a population labeled $U = 1, \ldots, k, \ldots, N$, or may be drawn independently from U. A $p-$dimensional vector of variables $\mathbf{x}$ and a $q-$dimensional vector of variables $\mathbf{y}$ are surveyed in S_1 and S_2, respectively, and both vectors are surveyed in S_3. These two modes of matrix sampling, depicted in Figure 2.1, will henceforth be referred to as nested and non-nested matrix sampling, respectively, in analogy with the nested and non-nested two-phase sampling (Hidiroglou 2001).

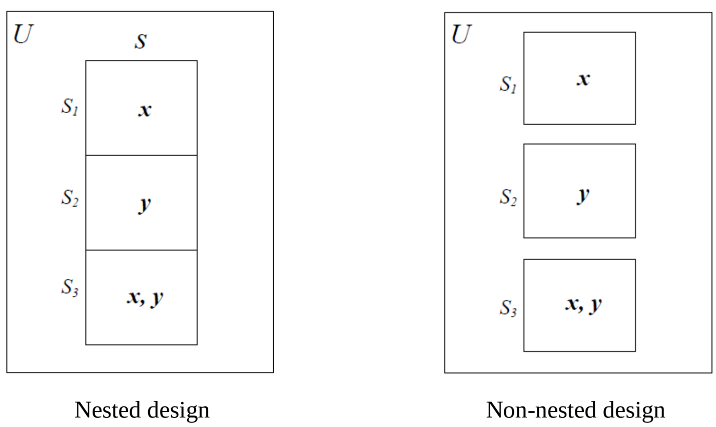

Figure 2.1 Nested and non-nested matrix sampling design (c)

We denote by $\mathbf{w}_i$ the vector of design weights for sample $S_i, i = 1, 2, 3,$ and by $\mathbf{X}_i$ and $\mathbf{Y}_i$ the sample matrices of $\mathbf{x}$ and $\mathbf{y}$, the subscripts indicating the sample. We obtain simple Horvitz-Thompson (HT) estimators $\hat{\mathbf{X}}_1 (= \mathbf{X}_1' \mathbf{w}_1)$ and $\hat{\mathbf{X}}_3$ of the population total $\mathbf{t}_{\mathbf{x}}$ of $\mathbf{x}$, using S_1 and S_3, respectively, and HT estimators $\hat{\mathbf{Y}}_2$ and $\hat{\mathbf{Y}}_3$ of the total $\mathbf{t}_{\mathbf{y}}$ of $\mathbf{y}$, using S_2 and S_3. For more efficient estimation of the totals $\mathbf{t}_{\mathbf{x}}$ and $\mathbf{t}_{\mathbf{y}}$ we seek composite estimators that combine all the available information on $\mathbf{x}$ and $\mathbf{y}$ in the three samples. Such composite estimators that are best linear unbiased estimators (BLUE), i.e., minimum-variance linear unbiased combinations of the four estimators $\hat{\mathbf{X}}_1, \hat{\mathbf{Y}}_2, \hat{\mathbf{X}}_3$ and $\hat{\mathbf{Y}}_3$, are denoted by $\hat{\mathbf{X}}^B$ and $\hat{\mathbf{Y}}^B$ and given in matrix form by

$$\begin{pmatrix} \hat{\mathbf{X}}^B \\ \hat{\mathbf{Y}}^B \end{pmatrix} = \mathcal{P} \begin{pmatrix} \hat{\mathbf{X}}_1 \\ \hat{\mathbf{Y}}_2 \\ \hat{\mathbf{X}}_3 \\ \hat{\mathbf{Y}}_3 \end{pmatrix}, \tag{2.1}$$

where $\mathcal{P} = (\mathbf{W}'\mathbf{V}^{-1}\mathbf{W})^{-1}\mathbf{W}'\mathbf{V}^{-1}$, the matrix $\mathbf{W}$ satisfies $E\left[(\hat{\mathbf{X}}_1', \hat{\mathbf{Y}}_2', \hat{\mathbf{X}}_3', \hat{\mathbf{Y}}_3')'\right] = \mathbf{W}(\mathbf{t}_\mathbf{x}', \mathbf{t}_\mathbf{y}')'$ and has entries 1's and 0's, and $\mathbf{V}$ is the variance-covariance matrix of $(\hat{\mathbf{X}}_1', \hat{\mathbf{Y}}_2', \hat{\mathbf{X}}_3', \hat{\mathbf{Y}}_3')'$. This estimation method was proposed by Chipperfield and Steel (2009), who provided analytical expressions of the BLUE for scalars x and y in non-nested matrix sampling, assuming simple random sampling and known $\mathbf{V}$. Such an approach to composite estimation has been explored also in a different context of survey sampling; see Wolter (1979), Jones (1980) and Fuller (1990). In general, computation of the BLUE given by (2.1) is not at all practical, as the computation of an estimated matrix $\mathbf{V}$ (and its inverse) in $\mathcal{P}$ would be quite laborious, especially if the number of variables or the sizes of the samples were large; it would be prohibitive if estimates for subpopulations were also required. Of course, the problem would become more difficult with more samples involved.

A more practical formulation of this estimation procedure is as follows. First, we express the composite estimators in (2.1) explicitly as linear combinations of the HT estimators $\hat{\mathbf{X}}_1, \hat{\mathbf{Y}}_2, \hat{\mathbf{X}}_3$ and $\hat{\mathbf{Y}}_3$, i.e.,

$$\begin{aligned} \hat{\mathbf{X}}^B &= \mathbf{B}_{1\mathbf{x}}\hat{\mathbf{X}}_1 + \mathbf{B}_{2\mathbf{x}}\hat{\mathbf{Y}}_2 + \mathbf{B}_{3\mathbf{x}}\hat{\mathbf{X}}_3 + \mathbf{B}_{4\mathbf{x}}\hat{\mathbf{Y}}_3 \\ \hat{\mathbf{Y}}^B &= \mathbf{B}_{1\mathbf{y}}\hat{\mathbf{X}}_1 + \mathbf{B}_{2\mathbf{y}}\hat{\mathbf{Y}}_2 + \mathbf{B}_{3\mathbf{y}}\hat{\mathbf{X}}_3 + \mathbf{B}_{4\mathbf{y}}\hat{\mathbf{Y}}_3. \end{aligned}$$

The condition of unbiasedness, $E(\hat{\mathbf{X}}^B) = \mathbf{t}_\mathbf{x}$ and $E(\hat{\mathbf{Y}}^B) = \mathbf{t}_\mathbf{y}$, implies that $\mathbf{B}_{3\mathbf{x}} = \mathbf{I} - \mathbf{B}_{1\mathbf{x}}$, $\mathbf{B}_{4\mathbf{x}} = -\mathbf{B}_{2\mathbf{x}}$ and $\mathbf{B}_{4\mathbf{y}} = \mathbf{I} - \mathbf{B}_{2\mathbf{y}}$, $\mathbf{B}_{3\mathbf{y}} = -\mathbf{B}_{1\mathbf{y}}$. Thus, $\mathcal{P}$ and $\mathbf{W}$ can be expressed as

$$\mathcal{P} = \begin{pmatrix} \mathbf{B}_{1\mathbf{x}} & \mathbf{B}_{2\mathbf{x}} & \mathbf{I} - \mathbf{B}_{1\mathbf{x}} & -\mathbf{B}_{2\mathbf{x}} \\ \mathbf{B}_{1\mathbf{y}} & \mathbf{B}_{2\mathbf{y}} & -\mathbf{B}_{1\mathbf{y}} & \mathbf{I} - \mathbf{B}_{2\mathbf{y}} \end{pmatrix}, \quad \mathbf{W}' = \begin{pmatrix} \mathbf{I} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{I} \end{pmatrix},$$

respectively, and the two composite estimators have necessarily the regression form

$$\begin{aligned} \hat{\mathbf{X}}^B &= \hat{\mathbf{X}}_3 + \mathbf{B}_{1\mathbf{x}}(\hat{\mathbf{X}}_1 - \hat{\mathbf{X}}_3) + \mathbf{B}_{2\mathbf{x}}(\hat{\mathbf{Y}}_2 - \hat{\mathbf{Y}}_3) \\ \hat{\mathbf{Y}}^B &= \hat{\mathbf{Y}}_3 + \mathbf{B}_{1\mathbf{y}}(\hat{\mathbf{X}}_1 - \hat{\mathbf{X}}_3) + \mathbf{B}_{2\mathbf{y}}(\hat{\mathbf{Y}}_2 - \hat{\mathbf{Y}}_3). \end{aligned} \tag{2.2}$$

Then writing $\mathcal{P} = (\mathcal{B}, \mathbf{I} - \mathcal{B})$, in obvious notation for matrix $\mathcal{B}$, we can express (2.1) as

$$\begin{pmatrix} \hat{\mathbf{X}}^B \\ \hat{\mathbf{Y}}^B \end{pmatrix} = \mathcal{B} \begin{pmatrix} \hat{\mathbf{X}}_1 \\ \hat{\mathbf{Y}}_2 \end{pmatrix} + (\mathbf{I} - \mathcal{B}) \begin{pmatrix} \hat{\mathbf{X}}_3 \\ \hat{\mathbf{Y}}_3 \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{X}}_3 \\ \hat{\mathbf{Y}}_3 \end{pmatrix} + \mathcal{B} \begin{pmatrix} \hat{\mathbf{X}}_1 - \hat{\mathbf{X}}_3 \\ \hat{\mathbf{Y}}_2 - \hat{\mathbf{Y}}_3 \end{pmatrix}, \tag{2.3}$$

the right-hand side of (2.3) being the matrix form of (2.2). The problem of finding the optimal (variance-minimizing) $\mathcal{P}$ of the BLUE in (2.1) reduces then to that of finding the optimal matrix $\mathcal{B}$ in (2.3). The estimated optimal $\hat{\mathcal{B}}^o$ is given by

$$\hat{\mathcal{B}}^o = -\widehat{\mathrm{Cov}}\left(\begin{pmatrix}\hat{\mathbf{X}}_3\\ \hat{\mathbf{Y}}_3\end{pmatrix}, \begin{pmatrix}\hat{\mathbf{X}}_1-\hat{\mathbf{X}}_3\\ \hat{\mathbf{Y}}_2-\hat{\mathbf{Y}}_3\end{pmatrix}\right)\left[\hat{V}\begin{pmatrix}\hat{\mathbf{X}}_1-\hat{\mathbf{X}}_3\\ \hat{\mathbf{Y}}_2-\hat{\mathbf{Y}}_3\end{pmatrix}\right]^{-1}, \tag{2.4}$$

and when the three samples are independent it reduces to

$$\hat{\mathcal{B}}^o = \hat{V}\begin{pmatrix}\hat{\mathbf{X}}_3\\ \hat{\mathbf{Y}}_3\end{pmatrix}\left[\hat{V}\begin{pmatrix}\hat{\mathbf{X}}_1\\ \hat{\mathbf{Y}}_2\end{pmatrix}+\hat{V}\begin{pmatrix}\hat{\mathbf{X}}_3\\ \hat{\mathbf{Y}}_3\end{pmatrix}\right]^{-1}. \tag{2.5}$$

In view of (2.3), with such optimal $\hat{\mathcal{B}}^o$ the estimated BLUE in (2.1) involving the estimated $\hat{\mathbf{V}}$, and with $\hat{\mathcal{P}} = \left(\hat{\mathcal{B}}^o, \mathbf{I}-\hat{\mathcal{B}}^o\right)$ is a special type of optimal multivariate regression estimator. For the form of the ordinary (single-sample) optimal regression estimator and relevant discussion, see Montanari (1987) and Rao (1994).

Expressing the estimated variance of the HT estimator of a total (see, for example, Särndal, Swensson and Wretman (1992), page 43) as a quadratic form with associated non-negative definite matrix $\mathbf{\Lambda}^0 = \{(\pi_{kl}-\pi_k\pi_l)/\pi_k\pi_l\pi_{kl}\}$, where π_k, π_{kl} are first-and-second order inclusion probabilities, it can be shown after some matrix algebra that

$$\hat{\mathcal{B}}^o = (\mathcal{X}_3'\mathbf{\Lambda}^0\mathcal{X})(\mathcal{X}'\mathbf{\Lambda}^0\mathcal{X})^{-1}, \tag{2.6}$$

where

$$\mathcal{X} = \begin{pmatrix}-\mathbf{X}_1 & \mathbf{0}\\ \mathbf{0} & -\mathbf{Y}_2\\ \mathbf{X}_3 & \mathbf{Y}_3\end{pmatrix} \tag{2.7}$$

is the $n\times(p+q)$ design matrix corresponding to the regression estimator (2.3), $\mathcal{X}_3$ is the matrix $\mathcal{X}$ with the first two rows set equal to zero, and $\mathbf{\Lambda}^0$ is associated with the combined sample $S = S_1\cup S_2\cup S_3$, reducing in the non-nested sampling to the block-diagonal matrix $\mathrm{diag}\{\mathbf{\Lambda}_i^0\}$ with $\mathbf{\Lambda}_i^0$ associated with the sample S_i. For the nested design, the probabilities defining $\mathbf{\Lambda}^0$ are products of the probabilities of inclusion in S and the conditional (on S) subsampling probabilities. With this estimated $\hat{\mathcal{B}}^o$, the estimated BLUE in (2.3), called composite optimal regression estimator (COR) and denoted by $\hat{\mathcal{X}}^{\mathrm{COR}}$, is written compactly as $\hat{\mathcal{X}}^{\mathrm{COR}} = \hat{\mathcal{X}}_3 - \hat{\mathcal{B}}^o\hat{\mathcal{X}}[= (\mathcal{X}_3 - \mathcal{X}\hat{\mathcal{B}}^{o'})'\mathbf{w}]$, where $\mathbf{w} = (\mathbf{w}_1', \mathbf{w}_2', \mathbf{w}_3')'$ is the vector of design weights of the combined sample S. It transpires that the COR estimator is in fact the sum of weighted sample regression residuals, and $\hat{\mathcal{B}}^o$ minimizes the quadratic form $\left(\mathcal{X}_3 - \mathcal{X}\hat{\mathcal{B}}^{o'}\right)'$ $\mathbf{\Lambda}^0\left(\mathcal{X}_3 - \mathcal{X}\hat{\mathcal{B}}^{o'}\right)$ in these residuals, which is the estimated approximate (large-sample) variance of $\hat{\mathcal{X}}^{\mathrm{COR}}$.

Now, upon writing $\hat{\mathcal{X}}^{\mathrm{COR}}$ as $\hat{\mathcal{X}}^{\mathrm{COR}} = \mathcal{X}_3'\left[\mathbf{w}+\mathbf{\Lambda}^0\mathcal{X}(\mathcal{X}'\mathbf{\Lambda}^0\mathcal{X})^{-1}(\mathbf{0}-\mathcal{X}'\mathbf{w})\right]$, it appears that the COR estimator has the form of a calibration estimator (with vector of calibration totals $\mathbf{0} = (\mathbf{0}', \mathbf{0}')'$ of dimension $(p+q))$, whose components satisfy the constraints $\hat{\mathbf{X}}_1^{\mathrm{COR}} = \hat{\mathbf{X}}_3^{\mathrm{COR}}$ and $\hat{\mathbf{Y}}_2^{\mathrm{COR}} = \hat{\mathbf{Y}}_3^{\mathrm{COR}}$, i.e., calibrated estimates of the same total from two different samples are equal. Indeed, the vector

$$\mathbf{c} = \mathbf{w} + \boldsymbol{\Lambda}^0 \boldsymbol{\mathcal{X}} (\boldsymbol{\mathcal{X}}' \boldsymbol{\Lambda}^0 \boldsymbol{\mathcal{X}})^{-1} (\mathbf{0} - \boldsymbol{\mathcal{X}}' \mathbf{w}), \tag{2.8}$$

is the vector of calibrated weights that minimizes the generalized least-squares distance $(\mathbf{c}-\mathbf{w})' (\boldsymbol{\Lambda}^0)^{-1} (\mathbf{c}-\mathbf{w})$ while satisfying the constraints $\mathbf{X}_1' \mathbf{c}_1 = \mathbf{X}_3' \mathbf{c}_3$ and $\mathbf{Y}_2' \mathbf{c}_2 = \mathbf{Y}_3' \mathbf{c}_3$, where the subcector $\mathbf{c}_i$ corresponds to sample S_i. This follows from a general result for the single-sample case, according to which calibration with the generalized least-squares distance measure may involve an arbitrary $n \times n$ positive definite matrix $\mathbf{R}$ instead of $\boldsymbol{\Lambda}^0$; see Andersson and Thorburn (2005).

We may now write the COR estimator formally as a calibration estimator, $\hat{\boldsymbol{\mathcal{X}}}^{\mathrm{COR}} = \boldsymbol{\mathcal{X}}_3' \mathbf{c}$, and using the subvector of calibrated weights $\mathbf{c}_3$, for sample S_3 only, we obtain the components of $\hat{\boldsymbol{\mathcal{X}}}^{\mathrm{COR}}$ directly in the simple linear forms

$$\hat{\mathbf{X}}^{\mathrm{COR}} = \mathbf{X}_3' \mathbf{c}_3 = \sum\nolimits_{S_3} c_k \mathbf{x}_k; \quad \hat{\mathbf{Y}}^{\mathrm{COR}} = \mathbf{Y}_3' \mathbf{c}_3 = \sum\nolimits_{S_3} c_k \mathbf{y}_k,$$

as in common survey practice. Yet, a decomposition of the vector $\mathbf{c}$ based on the following general lemma on calibration gives an analytic expression of $\hat{\mathbf{X}}^{\mathrm{COR}}$ and $\hat{\mathbf{Y}}^{\mathrm{COR}}$ of the form (2.2), which provides insight into the structure and the efficiency of the COR estimator. The proof of the lemma is given in the Appendix.

Lemma 1 *Let $\boldsymbol{\mathcal{X}}$ be a design matrix of dimension $n \times (p+q)$ and of full rank and written in partition form $(\boldsymbol{X}, \boldsymbol{\Psi})$, with corresponding vector of calibration totals $\mathbf{t}_{\boldsymbol{\mathcal{X}}} = (\mathbf{t}_{\boldsymbol{X}}', \mathbf{t}_{\boldsymbol{\Psi}}')'$, and let $\mathbf{R}$ be any positive definite matrix of dimension $n \times n$. Then the vector of calibrated weights $\mathbf{c} = \mathbf{w} + \mathbf{R} \boldsymbol{\mathcal{X}} (\boldsymbol{\mathcal{X}}' \mathbf{R} \boldsymbol{\mathcal{X}})^{-1}$ $(\mathbf{t}_{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}' \mathbf{w})$, obtained from the calibration procedure involving the distance measure $(\mathbf{c}-\mathbf{w})' \mathbf{R}^{-1}$ $(\mathbf{c}-\mathbf{w})$ and the constraint $\boldsymbol{\mathcal{X}}' \mathbf{c} = \mathbf{t}_{\boldsymbol{\mathcal{X}}}$, can be decomposed as*

$$\mathbf{c} = \mathbf{w} + \mathbf{L}_{\boldsymbol{\Psi}} \boldsymbol{X} (\boldsymbol{X}' \mathbf{L}_{\boldsymbol{\Psi}} \boldsymbol{X})^{-1} [\mathbf{t}_{\boldsymbol{X}} - \boldsymbol{X}' \mathbf{w}] + \mathbf{L}_{\boldsymbol{X}} \boldsymbol{\Psi} (\boldsymbol{\Psi}' \mathbf{L}_{\boldsymbol{X}} \boldsymbol{\Psi})^{-1} [\mathbf{t}_{\boldsymbol{\Psi}} - \boldsymbol{\Psi}' \mathbf{w}], \tag{2.9}$$

where $\mathbf{L}_{\boldsymbol{X}} = \mathbf{R}(\mathbf{I} - \mathbf{P}_{\boldsymbol{X}})$ with $\mathbf{P}_{\boldsymbol{X}} = \boldsymbol{X}(\boldsymbol{X}' \mathbf{R} \boldsymbol{X})^{-1} \boldsymbol{X}' \mathbf{R}$, and $\mathbf{L}_{\boldsymbol{\Psi}} = \mathbf{R}(\mathbf{I} - \mathbf{P}_{\boldsymbol{\Psi}})$ with $\mathbf{P}_{\boldsymbol{\Psi}} = \boldsymbol{\Psi}(\boldsymbol{\Psi}' \mathbf{R} \boldsymbol{\Psi})^{-1} \boldsymbol{\Psi}' \mathbf{R}$. The vector $\mathbf{c}$ can be written as

$$\mathbf{c} = \mathbf{c}_{\boldsymbol{\Psi}} + \mathbf{L}_{\boldsymbol{\Psi}} \boldsymbol{X} (\boldsymbol{X}' \mathbf{L}_{\boldsymbol{\Psi}} \boldsymbol{X})^{-1} [\mathbf{t}_{\boldsymbol{X}} - \boldsymbol{X}' \mathbf{c}_{\boldsymbol{\Psi}}], \tag{2.10}$$

where the vector

$$\mathbf{c}_{\boldsymbol{\Psi}} = \mathbf{w} + \mathbf{R} \boldsymbol{\Psi} (\boldsymbol{\Psi}' \mathbf{R} \boldsymbol{\Psi})^{-1} [\mathbf{t}_{\boldsymbol{\Psi}} - \boldsymbol{\Psi}' \mathbf{w}]$$

is generated by calibration of the design weights involving only $\boldsymbol{\Psi}$ and $\mathbf{t}_{\boldsymbol{\Psi}}$. By symmetry,

$$\mathbf{c} = \mathbf{c}_{\boldsymbol{X}} + \mathbf{L}_{\boldsymbol{X}} \boldsymbol{\Psi} (\boldsymbol{\Psi}' \mathbf{L}_{\boldsymbol{X}} \boldsymbol{\Psi})^{-1} [\mathbf{t}_{\boldsymbol{\Psi}} - \boldsymbol{\Psi}' \mathbf{c}_{\boldsymbol{X}}], \tag{2.11}$$

where

$$\mathbf{c}_{\mathcal{X}} = \mathbf{w} + \mathbf{R}\mathcal{X}(\mathcal{X}'\mathbf{R}\mathcal{X})^{-1}[\mathbf{t}_{\mathcal{X}} - \mathcal{X}'\mathbf{w}].$$

Now, if $\mathcal{X}$ is as in (2.7), with corresponding vector of calibration totals $\mathbf{t}_{\mathcal{X}} = (\mathbf{0}', \mathbf{0}')'$, and if $\mathbf{R} = \mathbf{\Lambda}^0$, then it follows from (2.9) that (2.8) can be written in the form

$$\mathbf{c} = \mathbf{w} + \mathbf{L}_{\Psi}\mathbf{X}(\mathbf{X}'\mathbf{L}_{\Psi}\mathbf{X})^{-1}[\hat{\mathbf{X}}_1 - \hat{\mathbf{X}}_3] + \mathbf{L}_{\mathbf{X}}\mathbf{\Psi}(\mathbf{\Psi}'\mathbf{L}_{\mathbf{X}}\mathbf{\Psi})^{-1}[\hat{\mathbf{Y}}_2 - \hat{\mathbf{Y}}_3],$$

and thus

$$\begin{aligned} \hat{\mathbf{X}}^{\mathrm{COR}} &= \mathbf{X}_3'\mathbf{c}_3 = \hat{\mathbf{X}}_3 + \hat{\mathbf{B}}_{1\mathbf{x}}^o(\hat{\mathbf{X}}_1 - \hat{\mathbf{X}}_3) + \hat{\mathbf{B}}_{2\mathbf{x}}^o(\hat{\mathbf{Y}}_2 - \hat{\mathbf{Y}}_3) \\ &= \hat{\mathbf{B}}_{1\mathbf{x}}^o\hat{\mathbf{X}}_1 + (\mathbf{I} - \hat{\mathbf{B}}_{1\mathbf{x}}^o)\hat{\mathbf{X}}_3 + \hat{\mathbf{B}}_{2\mathbf{x}}^o(\hat{\mathbf{Y}}_2 - \hat{\mathbf{Y}}_3), \end{aligned} \tag{2.12}$$

in obvious notation for $\hat{\mathbf{B}}_{1\mathbf{x}}^o$ and $\hat{\mathbf{B}}_{2\mathbf{x}}^o$. A similar expression is obtained for $\hat{\mathbf{Y}}^{\mathrm{COR}}$. It is seen from (2.12) that the COR estimator $\hat{\mathbf{X}}^{\mathrm{COR}}$ of $\mathbf{t}_{\mathbf{x}}$ is approximately (for large samples) unbiased, and derives its efficiency from combining the two elementary estimators $\hat{\mathbf{X}}_1$ and $\hat{\mathbf{X}}_3$ (pooling information from samples S_1 and S_3) and from borrowing strength from sample S_2 through the correlation between $\mathbf{x}$ and $\mathbf{y}$. In view of (2.10), the estimator $\hat{\mathbf{X}}^{\mathrm{COR}}$ takes the alternative forms

$$\begin{aligned} \hat{\mathbf{X}}^{\mathrm{COR}} &= \mathbf{X}_3'\mathbf{c}_{3\Psi} + \mathbf{X}_3'\mathbf{L}_{\Psi}\mathbf{X}(\mathbf{X}'\mathbf{L}_{\Psi}\mathbf{X})^{-1}[\mathbf{X}_1'\mathbf{c}_{1\Psi} - \mathbf{X}_3'\mathbf{c}_{3\Psi}] \\ &= \hat{\mathbf{X}}_3^{\mathrm{OR}} + \hat{\mathbf{B}}_{1\mathbf{x}}^o[\hat{\mathbf{X}}_1^{\mathrm{OR}} - \hat{\mathbf{X}}_3^{\mathrm{OR}}] \\ &= \hat{\mathbf{B}}_{1\mathbf{x}}^o\hat{\mathbf{X}}_1^{\mathrm{OR}} + (\mathbf{I} - \hat{\mathbf{B}}_{1\mathbf{x}}^o)\hat{\mathbf{X}}_3^{\mathrm{OR}}, \end{aligned} \tag{2.13}$$

where $\hat{\mathbf{X}}_i^{\mathrm{OR}} = \hat{\mathbf{X}}_i + \mathbf{X}_i'\mathbf{\Lambda}^0\mathbf{\Psi}(\mathbf{\Psi}'\mathbf{\Lambda}^0\mathbf{\Psi})^{-1}(\hat{\mathbf{Y}}_2 - \hat{\mathbf{Y}}_3)$ are optimal regression (OR) estimators incorporating the regression effect of the last term in (2.12).

In non-nested matrix sampling, $\mathbf{\Lambda}^0 = \mathrm{diag}\{\mathbf{\Lambda}_i^0\}$, $\hat{\mathbf{X}}_1^{\mathrm{OR}} = \hat{\mathbf{X}}_1$, $\hat{\mathbf{X}}_3^{\mathrm{OR}} = \hat{\mathbf{X}}_3 + \widehat{\mathrm{Cov}}(\hat{\mathbf{X}}_3, \hat{\mathbf{Y}}_3)[\hat{V}(\hat{\mathbf{Y}}_2) + \hat{V}(\hat{\mathbf{Y}}_3)]^{-1}[\hat{\mathbf{Y}}_2 - \hat{\mathbf{Y}}_3]$, having estimated approximate variance $\widehat{\mathrm{AV}}(\hat{\mathbf{X}}_3^{\mathrm{OR}}) = \hat{V}(\hat{\mathbf{X}}_3) - \widehat{\mathrm{Cov}}(\hat{\mathbf{X}}_3, \hat{\mathbf{Y}}_3)[\hat{V}(\hat{\mathbf{Y}}_2) + \hat{V}(\hat{\mathbf{Y}}_3)]^{-1}\widehat{\mathrm{Cov}}'(\hat{\mathbf{X}}_3, \hat{\mathbf{Y}}_3)$, and $\hat{\mathbf{B}}_{1\mathbf{x}}^o = \widehat{\mathrm{AV}}(\hat{\mathbf{X}}_3^{\mathrm{OR}})[\hat{V}(\hat{\mathbf{X}}_1) + \widehat{\mathrm{AV}}(\hat{\mathbf{X}}_3^{\mathrm{OR}})]^{-1}$ is the coefficient that minimizes the variance $\widehat{\mathrm{AV}}(\hat{\mathbf{X}}^{\mathrm{COR}})$. From the explicit form $\mathbf{I} - \hat{\mathbf{B}}_{1\mathbf{x}}^o = \hat{V}(\hat{\mathbf{X}}_1)\left[\hat{V}(\hat{\mathbf{X}}_1) + \hat{V}(\hat{\mathbf{X}}_3) - \widehat{\mathrm{Cov}}(\hat{\mathbf{X}}_3, \hat{\mathbf{Y}}_3) \times [\hat{V}(\hat{\mathbf{Y}}_2) + \hat{V}(\hat{\mathbf{Y}}_3)]^{-1}\widehat{\mathrm{Cov}}'(\hat{\mathbf{X}}_3, \hat{\mathbf{Y}}_3)\right]^{-1}$, it is then clear that the stronger the correlation between $\mathbf{x}$ and $\mathbf{y}$ the larger the $\mathbf{I} - \hat{\mathbf{B}}_{1\mathbf{x}}^o$ and more weight is given to the less variable component $\hat{\mathbf{X}}_3^{\mathrm{OR}}$. In this connection, it can be easily shown that $\widehat{\mathrm{AV}}(\hat{\mathbf{X}}^{\mathrm{COR}})$ satisfies

$$\widehat{\mathrm{AV}}(\hat{\mathbf{X}}^{\mathrm{COR}})[\hat{V}(\hat{\mathbf{X}}_1)]^{-1} = \hat{\mathbf{B}}_{1\mathbf{x}}^o < \mathbf{I}, \quad \widehat{\mathrm{AV}}(\hat{\mathbf{X}}^{\mathrm{COR}})\left[\widehat{\mathrm{AV}}(\hat{\mathbf{X}}_3^{\mathrm{OR}})\right]^{-1} = \mathbf{I} - \hat{\mathbf{B}}_{1\mathbf{x}}^o < \mathbf{I}.$$

These inequalities hold also for any linear combination of the components of each of the estimators involved. The optimal composite regression estimator $\hat{\mathbf{X}}^{\mathrm{COR}}$ is more efficient than each of its two components $\hat{\mathbf{X}}_1$ and $\hat{\mathbf{X}}_3^{\mathrm{OR}}$ by the shown quantities, with the efficiency depending on the strength of the correlation between $\mathbf{x}$ and $\mathbf{y}$. The estimator $\hat{\mathbf{X}}^{\mathrm{COR}}$ is also more efficient than the estimator

$\tilde{\mathbf{X}}^{\mathrm{COR}} = \tilde{\mathbf{B}}_{1\mathbf{x}}^{o}\hat{\mathbf{X}}_1 + (\mathbf{I} - \tilde{\mathbf{B}}_{1\mathbf{x}}^{o})\hat{\mathbf{X}}_3$, with $\tilde{\mathbf{B}}_{1\mathbf{x}}^{o} = \hat{V}(\hat{\mathbf{X}}_3)[\hat{V}(\hat{\mathbf{X}}_1) + \hat{V}(\hat{\mathbf{X}}_3)]^{-1}$, which does not incorporate the information on $\mathbf{y}$ (does not borrow strength from sample S_2) and has estimated variance $\widehat{\mathrm{AV}}(\tilde{\mathbf{X}}^{\mathrm{COR}}) = \hat{V}(\hat{\mathbf{X}}_1)[\hat{V}(\hat{\mathbf{X}}_1) + \hat{V}(\hat{\mathbf{X}}_3)]^{-1}\hat{V}(\hat{\mathbf{X}}_3)$. Indeed, writing the variance $\widehat{\mathrm{AV}}(\hat{\mathbf{X}}^{\mathrm{COR}}) = \hat{V}(\hat{\mathbf{X}}_1)\hat{\mathbf{B}}_{1\mathbf{x}}^{o}$ as $\widehat{\mathrm{AV}}(\hat{\mathbf{X}}^{\mathrm{COR}}) = \hat{V}(\hat{\mathbf{X}}_1)[\hat{V}(\hat{\mathbf{X}}_1) + \hat{V}(\hat{\mathbf{X}}_3)]^{-1}\hat{V}(\hat{\mathbf{X}}_3)\mathbf{E}$, where $\mathbf{E} = \mathbf{E}_1\mathbf{E}_2$ with $\mathbf{E}_1 = [\mathbf{I} - (\hat{V}(\hat{\mathbf{X}}_3))^{-1}$ $\widehat{\mathrm{Cov}}(\hat{\mathbf{X}}_3, \hat{\mathbf{Y}}_3)[\hat{V}(\hat{\mathbf{Y}}_2) + \hat{V}(\hat{\mathbf{Y}}_3)]^{-1}\widehat{\mathrm{Cov}}'(\hat{\mathbf{X}}_3, \hat{\mathbf{Y}}_3)]$ and $\mathbf{E}_2 = [\mathbf{I} - [\hat{V}(\hat{\mathbf{X}}_1) + \hat{V}(\hat{\mathbf{X}}_3)]^{-1}\widehat{\mathrm{Cov}}(\hat{\mathbf{X}}_3, \hat{\mathbf{Y}}_3)[\hat{V}(\hat{\mathbf{Y}}_2) +$ $\hat{V}(\hat{\mathbf{Y}}_3)]^{-1}\widehat{\mathrm{Cov}}'(\hat{\mathbf{X}}_3, \hat{\mathbf{Y}}_3)]^{-1}$, and noticing that $\mathbf{E} \le \mathbf{I}$, it follows that

$$\widehat{\mathrm{AV}}(\hat{\mathbf{X}}^{\mathrm{COR}})[\widehat{\mathrm{AV}}(\tilde{\mathbf{X}}^{\mathrm{COR}})]^{-1} = \mathbf{E} \le \mathbf{I},$$

that is, borrowing strength from S_2 reduces the variance of the composite estimator of $\mathbf{t}_{\mathbf{x}}$ by the factor $\mathbf{E}$, which depends on the strength of the correlation between $\mathbf{x}$ and $\mathbf{y}$. It can be easily verified that for two scalar variables x and y and simple random sampling this result reduces to the analogous analytical result on the efficiency of BLUE given in Chipperfield and Steel (2009, page 231). In this simple case $E = [n_1 + n_3][n_3 + n_2(1 - \rho^2)] / [(n_1 + n_3)(n_2 + n_3) - n_1 n_2 \rho^2]$, where ρ is the correlation between x and y. As an illustration, assuming equal sample sizes and correlation $\rho = 0.7$, the efficiency gain is 13.96%.

In nested matrix sampling, the two estimators in (2.13) are $\hat{\mathbf{X}}_i^{\mathrm{OR}} = \hat{\mathbf{X}}_i + \widehat{\mathrm{Cov}}(\hat{\mathbf{X}}_i, \hat{\mathbf{\Psi}})$ $[\hat{V}(\hat{\mathbf{\Psi}})]^{-1}[\hat{\mathbf{Y}}_2 - \hat{\mathbf{Y}}_3]$, and $\hat{\mathbf{B}}_{1x}^{o} = [\widehat{\mathrm{AV}}(\hat{\mathbf{X}}_3^{\mathrm{OR}}) - \widehat{\mathrm{AC}}(\hat{\mathbf{X}}_1^{\mathrm{OR}}, \hat{\mathbf{X}}_3^{\mathrm{OR}})][\widehat{\mathrm{AV}}(\hat{\mathbf{X}}_1^{\mathrm{OR}}) + \widehat{\mathrm{AV}}(\hat{\mathbf{X}}_3^{\mathrm{OR}}) - 2\widehat{\mathrm{AC}}(\hat{\mathbf{X}}_1^{\mathrm{OR}},$ $\hat{\mathbf{X}}_3^{\mathrm{OR}})]^{-1}$, where AC denotes approximate covariance. In this case, in addition to the correlation $\rho_{x3,y3}$ between $\hat{\mathbf{X}}_3$ and $\hat{\mathbf{Y}}_3$ in sample S_3, the efficiency of $\hat{\mathbf{X}}^{\mathrm{COR}}$ depends on the estimators' correlations $\rho_{x1,x3}, \rho_{y2,y3}, \rho_{y2,x3}$ due to the dependence of the subsamples. For univariate x and y and with the simplifying assumption of identical designs for the three subsamples (as in equal splitting of the full sample), we obtain some insight through the simple expressions $\widehat{\mathrm{AV}}(\hat{X}^{\mathrm{COR}}) =$ $V(\hat{X}_3)[2(1 - \rho_{x1,x3}^2)(1 - \rho_{y2,y3}) - (\rho_{x3,y3} - \rho_{y2,x3})^2] / [4(1 - \rho_{x1,x3})(1 - \rho_{y2,y3}) - (\rho_{x3,y3} - \rho_{y2,x3})^2]$, and $\widehat{\mathrm{AV}}(\tilde{X}^{\mathrm{COR}}) = V(\hat{X}_3)(1 + \rho_{x1,x3})/2$. Clearly, the estimator $\tilde{X}^{\mathrm{COR}}$, which ignores information on y, is more efficient than the simple average of single-sample estimators of t_x only when there is negative correlation $\rho_{x1,x3}$. The efficiency of $\hat{X}^{\mathrm{COR}}$ relative to $\tilde{X}^{\mathrm{COR}}$

$$\frac{\widehat{\mathrm{AV}}(\hat{X}^{\mathrm{COR}})}{\widehat{\mathrm{AV}}(\tilde{X}^{\mathrm{COR}})} = \frac{4(1 - \rho_{x1,x3}^2)(1 - \rho_{y2,y3}) - 2(\rho_{x3,y3} - \rho_{y2,x3})^2}{4(1 - \rho_{x1,x3}^2)(1 - \rho_{y2,y3}) - (1 + \rho_{x1,x3})(\rho_{x3,y3} - \rho_{y2,x3})^2}$$

depends on the sign and size of $\rho_{x1,x3}$ and the size of $|\rho_{x3,y3} - \rho_{y2,x3}|$.

Although the calibration procedure, with vector of calibrated weights (2.8), substantially facilitates the computation of the composite optimal regression estimator for any total of interest, the matrix $\mathbf{\Lambda}^0$ makes the calculations exceedingly demanding, particularly in nested sampling where the subsamples are dependent and thus $\mathbf{\Lambda}^0$ is not $\mathrm{diag}\{\mathbf{\Lambda}_i^0\}$. Besides, the probabilities π_{kl} are not known for most sampling designs. An alternative composite regression estimator that is computationally very efficient is developed in the next section.

3 Composite generalized regression estimation for design (c)

A computationally very convenient, but generally suboptimal, variant of $\hat{\mathcal{B}}^o$ in (2.6) is obtained by replacing the matrix $\mathbf{\Lambda}^0$ with the diagonal "weighting matrix" $\mathbf{\Lambda}$ having w_{ik}/q_{ik} as ik^{th} diagonal entry, where $\{w_{ik}\}$ are the design weights of S_i and $\{q_{ik}\}$ are positive constants. This gives the multivariate composite generalized regression (CGR) estimator of $(\mathbf{t}'_{\mathbf{x}}, \mathbf{t}'_{\mathbf{y}})'$

$$\begin{pmatrix} \hat{\mathbf{X}}^{\text{CGR}} \\ \hat{\mathbf{Y}}^{\text{CGR}} \end{pmatrix} = \hat{\mathcal{B}} \begin{pmatrix} \hat{\mathbf{X}}_1 \\ \hat{\mathbf{Y}}_2 \end{pmatrix} + (\mathbf{I} - \hat{\mathcal{B}}) \begin{pmatrix} \hat{\mathbf{X}}_3 \\ \hat{\mathbf{Y}}_3 \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{X}}_3 \\ \hat{\mathbf{Y}}_3 \end{pmatrix} + \hat{\mathcal{B}} \begin{pmatrix} \hat{\mathbf{X}}_1 - \hat{\mathbf{X}}_3 \\ \hat{\mathbf{Y}}_2 - \hat{\mathbf{Y}}_3 \end{pmatrix}, \tag{3.1}$$

where $\hat{\mathcal{B}} = (\mathcal{X}'_3 \mathbf{\Lambda} \mathcal{X})(\mathcal{X}' \mathbf{\Lambda} \mathcal{X})^{-1}$ is the associated matrix regression coefficient. For an extensive discussion of the generalized regression estimator in a single sample, see Särndal et al. (1992, Chapter 6). The CGR estimator may be compactly written as $\hat{\mathcal{X}}^{\text{CGR}} = \hat{\mathcal{X}}_3 - \hat{\mathcal{B}}\hat{\mathcal{X}} \left[= (\mathcal{X}_3 - \mathcal{X}\hat{\mathcal{B}}')' \, \mathbf{w} \right]$, i.e., as a sum of weighted sample regression residuals. The coefficient $\hat{\mathcal{B}}$ is optimal in the sense of generalized least squares, i.e., it minimizes the quadratic form $(\mathcal{X}_3 - \mathcal{X}\hat{\mathcal{B}}')' \, \mathbf{\Lambda} (\mathcal{X}_3 - \mathcal{X}\hat{\mathcal{B}}')$ in these residuals. Similarly to the COR estimator, the CGR estimator too can be obtained in calibration form as $\mathcal{X}'_3 \mathbf{c}$, where the vector $\mathbf{c} = \mathbf{w} + \mathbf{\Lambda}\mathcal{X}(\mathcal{X}'\mathbf{\Lambda}\mathcal{X})^{-1}(\mathbf{0} - \mathcal{X}'\mathbf{w})$ minimizes the generalized least-squares distance $(\mathbf{c} - \mathbf{w})' \, \mathbf{\Lambda}^{-1} (\mathbf{c} - \mathbf{w})$ and satisfies the constraints $\hat{\mathbf{X}}_1^{\text{CGR}} = \hat{\mathbf{X}}_3^{\text{CGR}}$ and $\hat{\mathbf{Y}}_2^{\text{CGR}} = \hat{\mathbf{Y}}_3^{\text{CGR}}$. This extends to the present context the well-known equivalence of generalized regression estimation and calibration estimation (Deville and Särndal 1992) for a single-sample setting. Now using the subvector of calibrated weights $\mathbf{c}_3$, for sample S_3 only, we obtain the composite estimators in (3.1) in the simple linear forms $\hat{\mathbf{X}}^{\text{CGR}} = \mathbf{X}'_3 \mathbf{c}_3$ and $\hat{\mathbf{Y}}^{\text{CGR}} = \mathbf{Y}'_3 \mathbf{c}_3$. Using Lemma 1 and the diagonal structure of $\mathbf{\Lambda}$, it works out that $\hat{\mathbf{X}}^{\text{CGR}}$ can be written as

$$\hat{\mathbf{X}}^{\text{CGR}} = \hat{\mathbf{B}}_{1\mathbf{x}} \hat{\mathbf{X}}_1 + (\mathbf{I} - \hat{\mathbf{B}}_{1\mathbf{x}}) \hat{\mathbf{X}}_3^{\text{GR}}, \tag{3.2}$$

where $\hat{\mathbf{X}}_3^{\text{GR}} = \hat{\mathbf{X}}_3 + \mathbf{X}'_3 \mathbf{\Lambda} \mathbf{\Psi} (\mathbf{\Psi}' \mathbf{\Lambda} \mathbf{\Psi})^{-1} (\hat{\mathbf{Y}}_2 - \hat{\mathbf{Y}}_3)$ is the generalized regression (GR) counterpart of $\hat{\mathbf{X}}_3^{\text{OR}}$. The matrix regression coefficient $\hat{\mathbf{B}}_{1\mathbf{x}}$ is written explicitly as $\hat{\mathbf{B}}_{1\mathbf{x}} = \mathbf{X}'_3 \mathbf{L}_{\mathbf{\Psi}} \mathsf{X} (\mathbf{X}'_1 \mathbf{\Lambda}_1 \mathbf{X}_1 + \mathbf{X}'_3 \mathbf{L}_{\mathbf{\Psi}} \mathsf{X})^{-1}$, where $\mathbf{X}'_3 \mathbf{L}_{\mathbf{\Psi}} \mathsf{X} = \mathbf{X}'_3 \mathbf{\Lambda}_3 \mathbf{X}_3 - \mathbf{X}'_3 \mathbf{\Lambda}_3 \mathbf{Y}_3 (\mathbf{Y}'_2 \mathbf{\Lambda}_2 \mathbf{Y}_2 + \mathbf{Y}'_3 \mathbf{\Lambda}_3 \mathbf{Y}_3)^{-1} \mathbf{Y}'_3 \mathbf{\Lambda}_3 \mathbf{X}_3$. If $\mathbf{x}$ and $\mathbf{y}$ were uncorrelated, or if information on $\mathbf{y}$ was not used in the estimation of $\mathbf{t}_{\mathbf{x}}$, then it would be $\hat{\mathbf{X}}_3^{\text{GR}} = \hat{\mathbf{X}}_3$ and $\hat{\mathbf{B}}_{1\mathbf{x}} = \mathbf{X}'_3 \mathbf{\Lambda}_3 \mathbf{X}_3 (\mathbf{X}'_1 \mathbf{\Lambda}_1 \mathbf{X}_1 + \mathbf{X}'_3 \mathbf{\Lambda}_3 \mathbf{X}_3)^{-1}$. But the GR estimator $\hat{\mathbf{X}}_3^{\text{GR}}$ is generally more efficient than the HT estimator $\hat{\mathbf{X}}_3$, and since $\mathbf{X}'_1 \mathbf{\Lambda}_1 \mathbf{X}_1 + \mathbf{X}'_3 \mathbf{L}_{\mathbf{\Psi}} \mathsf{X} < \mathbf{X}'_1 \mathbf{\Lambda}_1 \mathbf{X}_1 + \mathbf{X}'_3 \mathbf{\Lambda}_3 \mathbf{X}_3$ (in the partial ordering of non-negative definite matrices), it is clear that more weight is given to $\hat{\mathbf{X}}_3^{\text{GR}}$ in (3.2), through $\mathbf{I} - \hat{\mathbf{B}}_{1\mathbf{x}} = \mathbf{X}'_1 \mathbf{\Lambda}_1 \mathbf{X}_1 (\mathbf{X}'_1 \mathbf{\Lambda}_1 \mathbf{X}_1 + \mathbf{X}'_3 \mathbf{L}_{\mathbf{\Psi}} \mathsf{X})^{-1}$, than would have been given to the component estimator $\hat{\mathbf{X}}_3$ in the simple composite estimator involving only information on $\mathbf{x}$. This suggests that the CGR estimator in (3.2), incorporating information from sample S_2, is a more efficient estimator. Suggestive of the efficiency of $\hat{\mathbf{X}}^{\text{CGR}}$ is also its alternative expression, obtained using (2.11), $\hat{\mathbf{X}}^{\text{CGR}} = \tilde{\mathbf{X}}^{\text{CGR}} + \mathbf{X}'_3 \mathbf{L}_{\mathsf{X}} \mathbf{\Psi} (\mathbf{\Psi}' \mathbf{L}_{\mathsf{X}} \mathbf{\Psi})^{-1} [\hat{\mathbf{Y}}_2 - \hat{\mathbf{Y}}_3^{\text{GR}}]$, where $\tilde{\mathbf{X}}^{\text{CGR}} = \hat{\mathbf{X}}_3 + \mathbf{X}'_3 \mathbf{\Lambda} \mathsf{X} (\mathsf{X}' \mathbf{\Lambda} \mathsf{X})^{-1} (\hat{\mathbf{X}}_1 - \hat{\mathbf{X}}_3) = \tilde{\mathbf{B}}_{1\mathbf{x}} \hat{\mathbf{X}}_1 + (\mathbf{I} - \tilde{\mathbf{B}}_{1\mathbf{x}}) \hat{\mathbf{X}}_3$ is the composite regression estimator of $\mathbf{t}_{\mathbf{x}}$ using information on $\mathbf{x}$ from S_1 and S_3.

In general, the computationally simpler CGR estimator $(\hat{\mathbf{X}}^{\mathrm{CGR}}, \hat{\mathbf{Y}}^{\mathrm{CGR}})$, involving the coefficient $\hat{\mathscr{B}}$, is less efficient than the optimal composite regression estimator $(\hat{\mathbf{X}}^{\mathrm{COR}}, \hat{\mathbf{Y}}^{\mathrm{COR}})$ which involves the estimated optimal coefficient $\hat{\mathscr{B}}^o$ and has the same asymptotic variance as the BLUE in (2.3); the efficiency loss may be larger in nested matrix sampling, for which the matrix $\mathbf{\Lambda}^0$ is not block-diagonal. On the other hand, $(\hat{\mathbf{X}}^{\mathrm{COR}}, \hat{\mathbf{Y}}^{\mathrm{COR}})$ may be unstable in small samples, when there is a small number of degrees of freedom available for the estimation of $\hat{\mathscr{B}}^o$, which is particularly so in nested matrix sampling; for a discussion of the relative stability of the optimal versus the generalized regression estimator in the single-sample case see Rao (1994) or Montanari (1998). For certain sampling strategies, described in the following theorem, $\hat{\mathscr{B}} = \hat{\mathscr{B}}^o$ and the CGR estimator is the COR estimator, and asymptotically is BLUE; the proof is given in the Appendix.

Theorem 1 *Consider the following sampling strategies.*

Non-nested design

(a) *For all three samples S_1, S_2 and S_3 assume stratified simple random sampling without replacement (STRSRS) with sampling fraction $f_{ih} = n_{ih}/N_{ih}$ in stratum h of sample i, $h = 1, \ldots, H_i$ and N_{ih} denoting stratum size, and specify the constants q_{ik} in $\mathbf{\Lambda}_i$ as $q_{ik} = (n_{ih} - 1)/N_{ih}(1 - f_{ih})$ for all units of stratum h. Furthermore, assume that within each sample the units are sorted by stratum, and consider the augmented design matrix $\mathcal{Z} = (\mathcal{X}, \mathbf{D})$ in (2.7), where $\mathbf{D}$ is the block diagonal matrix $\operatorname{diag}\{\mathbf{D}_1, \mathbf{D}_2, \mathbf{D}_3\}$ and $\mathbf{D}_i$ is the diagonal matrix $\operatorname{diag}\{\mathbf{1}_{i1}, \ldots, \mathbf{1}_{ih}, \ldots, \mathbf{1}_{iH_i}\}$, with diagonal element $\mathbf{1}_{ih}$ being a vector of ones for all units of stratum h in sample S_i, and consider the corresponding augmented vector of calibration totals $\mathbf{t}_{\mathcal{Z}} = (\mathbf{0}', \mathbf{0}', \mathbf{N}_1', \mathbf{N}_2', \mathbf{N}_3')'$, where $\mathbf{N}_i$ is the vector of strata sizes for sample S_i.*

(b) *For all three samples S_1, S_2 and S_3 assume stratified Poisson sampling and specify the constants q_{ik} in the entries of $\mathbf{\Lambda}_i$ as $q_{ik} = \pi_{ihk}/(1 - \pi_{ihk})$ for the units of stratum h, where π_{ihk} is the inclusion probability of unit k in stratum h of the i^{th} survey.*

Nested design

(a') *Assume that an initial stratified simple random sample S is split by stratum into three simple random subsamples S_1, S_2 and S_3. Specify the sampling fractions f_{ih}, the constants q_{ik} in $\mathbf{\Lambda}_i$, the design matrix $\mathcal{Z} = (\mathcal{X}, \mathbf{D})$ and the vector of calibration totals $\mathbf{t}_{\mathcal{Z}}$ as in part (a).*

(b') *Assume that an initial stratified Poisson sample S is randomly split by stratum into three subsamples S_1, S_2 and S_3, with unequal inclusion probabilities for the units of each subsample. Specify the constants q_{ik} in $\mathbf{\Lambda}_i$ as $q_{ik} = \pi_{ihk}/(1 - \pi_{ihk})$ for the units of stratum h, where π_{ihk} is the marginal inclusion probability of unit k in stratum h of the i^{th} subsample.*

Under each of strategies (a) and (b), the calibration procedure with matrix $\mathbf{\Lambda}$ in the least-squares distance measure gives the CGR estimator in (3.1) with $\hat{\mathscr{B}} = \hat{\mathscr{B}}^o$, implying that the CGR estimator is the COR estimator. For (a') and (b'), this holds approximately when the strata sampling fractions are approximately zero.

Corollary 1 *The result of Theorem 1 holds also for the unstratified versions of all four designs. For simple random sampling without replacement (SRS), in particular, the matrix* $\mathbf{D}$ *reduces to the diagonal matrix* $\operatorname{diag}\{\mathbf{1}_1, \mathbf{1}_2, \mathbf{1}_3\}$ *having as its* i^{th} *diagonal element the* n_i*-dimensional unit vector* $\mathbf{1}_i$, *and the vector of calibration totals is then* $\mathbf{t}_{\mathcal{Z}} = (\mathbf{0}', \mathbf{0}', N, N, N)'$.

Corollary 2 *In non-nested sampling, when the sampling design for each of the three samples is one of the designs in* (*a*) *and* (*b*) *or one of their unstratified versions, but not the same for all samples, the result of Theorem 1 holds provided that the matrix* $\mathbf{D}$ *in* $\mathcal{Z}$ *and the vector* $\mathbf{t}_{\mathcal{Z}}$ *are reduced so as to correspond only to the samples for which SRS or STRSRS is used.*

The extended calibration scheme in Theorem $1(a, a')$ includes calibration to the stratum sizes (or to the population size in the SRS version), through the inclusion of an intercept for each stratum in the design matrix $\mathcal{X}$. No additional information is used beyond what is assumed in the sampling design in (a) and (a'), and the form of the resulting CGR estimator remains the same as in (3.1) because the HT estimates of the population and strata sizes are exact. The effect of this extended calibration (with the specified values of q_{ik}) is only to convert the CGR coefficient $\hat{\mathcal{B}}$ to the optimal coefficient $\hat{\mathcal{B}}^o$ and, thus, the CGR estimator to the COR estimator. The practical significance of this conversion lies in carrying out optimal composite regression estimation through the much simpler calibration procedure of generalized regression estimation.

Subsampling as in part (a'), with a priori fixed sample sizes, is a natural procedure in matrix sampling involving splitting a questionnaire. In contrast, in the subsampling scheme of part (b') n_i is the expected sample size of S_i, the actual size being random. Unequal subsampling probabilities may be determined adaptively for increased efficiency; see Gonzalez and Eltinge (2008).

The results of Theorem 1 could extend to other sampling designs, e.g., stratified two-stage simple random sampling in non-nested matrix sampling. However, the required adjustments in the matrices $\mathbf{\Lambda}_i$ would not be easier than using directly the matrices $\mathbf{\Lambda}_i^0$ in the calibration to obtain the optimal composite regression estimator.

For sampling designs other than those assumed in Theorem 1, the value of q_{ik} in the entries of $\mathbf{\Lambda}_i$ should be set to $q_{ik} = \tilde{n}_i / (\tilde{n}_1 + \tilde{n}_2 + \tilde{n}_3)$, where $\tilde{n}_i = n_i / d_i$, d_i denoting design effect, to take into account the differential in effective sample sizes among the three samples. If the same design is used for all samples, then $\tilde{n}_i = n_i$. The justification for this adjustment is based on the argument given in Merkouris (2010) for a similar problem of composite regression estimation.

4 Composite estimation for matrix sampling design (d)

4.1 Core set of variables with known totals

We discuss first a special case of the matrix sampling design (d) in which the variables that are common to the three samples have known totals. In this very realistic sampling setting, all samples collect also information on the same vector of auxiliary variables $\mathbf{z}$ for which the vector of population totals $\mathbf{t}_{\mathbf{z}}$ is known. For illustration we consider again three samples, as in Figure 2.1 (but with $\mathbf{z}$ added in all subsamples). Then, the CGR estimator $\hat{\mathbf{X}}^{\text{CGR}}$ in (3.1) may be augmented with the ordinary regression terms $\hat{\mathbf{B}}_{3\mathbf{x}}(\mathbf{t}_\mathbf{z} - \hat{\mathbf{Z}}_1) + \hat{\mathbf{B}}_{4\mathbf{x}}(\mathbf{t}_\mathbf{z} - \hat{\mathbf{Z}}_2) + \hat{\mathbf{B}}_{5\mathbf{x}}(\mathbf{t}_\mathbf{z} - \hat{\mathbf{Z}}_3),$ where $\hat{\mathbf{Z}}_i, i = 1,2,3$ is the HT estimator of $\mathbf{t}_\mathbf{z}$ based on sample S_i; similarly for $\hat{\mathbf{Y}}^{\mathrm{CGR}}$. This estimator has improved efficiency, as it incorporates additional information, and is generated by a calibration procedure that includes the additional three constraints $\hat{\mathbf{Z}}_i^{\mathrm{CGR}} = \mathbf{t}_\mathbf{z}$, and has the design matrix $\mathcal{X}$ in (2.7) augmented with the block-diagonal matrix $\mathbf{Z} = \mathrm{diag}\{\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3\}$. In the simplest case when the sample matrices $\mathbf{Z}_i$ reduce to the unit columns $\mathbf{1}_i$ (with corresponding total the size of the population), the calibration scheme is the one specified in Corollary 1 above. As shown in the proof of the next theorem, an application of Lemma 1 to the present calibration procedure, with partitioned design matrix $(\mathcal{X}, \mathbf{Z}), \mathbf{R} = \mathbf{\Lambda}$ and calibration totals $(\mathbf{0}', \mathbf{0}', \mathbf{t}'_\mathbf{z}, \mathbf{t}'_\mathbf{z}, \mathbf{t}'_\mathbf{z})'$, gives a modified CGR form of (3.1) with GR estimators incorporating information on $\mathbf{z}$ in place of HT estimators. This is compactly written as $\hat{\mathcal{X}}_3^{\mathrm{GR}} - \hat{\mathcal{B}}\hat{\mathcal{X}}^{\mathrm{GR}}$, where $\hat{\mathcal{X}}_3^{\mathrm{GR}} = \hat{\mathcal{X}}_3 + \mathcal{X}_3'\mathbf{\Lambda}\mathbf{Z}(\mathbf{Z}'\mathbf{\Lambda}\mathbf{Z})^{-1}(\mathbf{t}_{(\mathbf{z})} - \hat{\mathbf{Z}})$, with $\mathbf{t}_{(\mathbf{z})} = (\mathbf{t}'_\mathbf{z}, \mathbf{t}'_\mathbf{z}, \mathbf{t}'_\mathbf{z})'$, and $\hat{\mathcal{X}}^{\mathrm{GR}}$ expressed similarly, and where $\hat{\mathcal{B}} = [\mathcal{X}_3'\mathbf{\Lambda}(\mathbf{I} - \mathbf{P}_\mathbf{Z})\mathcal{X}][\mathcal{X}'\mathbf{\Lambda}(\mathbf{I} - \mathbf{P}_\mathbf{Z})\mathcal{X}]^{-1}$ with $\mathbf{P}_\mathbf{Z} = \mathbf{Z}(\mathbf{Z}'\mathbf{\Lambda}\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{\Lambda}$.

Replacing $\mathbf{\Lambda}$ by $\mathbf{\Lambda}^0$ in the calibration procedure gives the optimal composite regression estimator, compactly written as $\hat{\mathcal{X}}_3^{\mathrm{OR}} - \hat{\mathcal{B}}^o\hat{\mathcal{X}}^{\mathrm{OR}}$, with optimal regression estimators incorporating information on $\mathbf{z}$ in place of GR estimators, and with $\hat{\mathcal{B}}^o = [\mathcal{X}_3'\mathbf{\Lambda}^0(\mathbf{I} - \mathbf{P}_\mathbf{Z}^0)\mathcal{X}][\mathcal{X}'\mathbf{\Lambda}^0(\mathbf{I} - \mathbf{P}_\mathbf{Z}^0)\mathcal{X}]^{-1}$ where $\mathbf{P}_\mathbf{Z}^0 = \mathbf{Z}(\mathbf{Z}'\mathbf{\Lambda}^0\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{\Lambda}^0$. Noticing that $(\mathbf{I} - \mathbf{P}_\mathbf{Z}^0)\mathcal{X}_3$ is the matrix of residuals corresponding to $\hat{\mathcal{X}}_3^{\mathrm{OR}}$ and that $\mathcal{X}_3'\mathbf{\Lambda}^0(\mathbf{I} - \mathbf{P}_\mathbf{Z}^0)\mathcal{X} = \mathcal{X}_3'(\mathbf{I} - \mathbf{P}_\mathbf{Z}^0)'\mathbf{\Lambda}^0(\mathbf{I} - \mathbf{P}_\mathbf{Z}^0)\mathcal{X} = \widehat{\mathrm{AC}}(\hat{\mathcal{X}}_3^{\mathrm{OR}}, \hat{\mathcal{X}}^{\mathrm{OR}})$, and similarly for $\widehat{\mathrm{AV}}(\hat{\mathcal{X}}^{\mathrm{OR}})$, it follows that

$$\hat{\mathcal{B}}^o = -\widehat{\mathrm{AC}}\left[\begin{pmatrix}\hat{\mathbf{X}}_3^{\mathrm{OR}}\\ \hat{\mathbf{Y}}_3^{\mathrm{OR}}\end{pmatrix}, \begin{pmatrix}\hat{\mathbf{X}}_1^{\mathrm{OR}} - \hat{\mathbf{X}}_3^{\mathrm{OR}}\\ \hat{\mathbf{Y}}_2^{\mathrm{OR}} - \hat{\mathbf{Y}}_3^{\mathrm{OR}}\end{pmatrix}\right]\left[\widehat{\mathrm{AV}}\begin{pmatrix}\hat{\mathbf{X}}_1^{\mathrm{OR}} - \hat{\mathbf{X}}_3^{\mathrm{OR}}\\ \hat{\mathbf{Y}}_2^{\mathrm{OR}} - \hat{\mathbf{Y}}_3^{\mathrm{OR}}\end{pmatrix}\right]^{-1}, \tag{4.1}$$

in analogy with (2.4), or with (2.5) in non-nested sampling. Thus, $\hat{\mathcal{B}}^o$ is optimal in the sense of minimizing the approximate variance of the estimator $\hat{\mathcal{X}}_3^{\mathrm{OR}} - \hat{\mathcal{B}}^o\hat{\mathcal{X}}^{\mathrm{OR}}$, which is then asymptotically BLUE. An alternative estimator, of weaker optimality, has the form $\hat{\mathcal{X}}_3^{\mathrm{GR}} - \hat{\mathcal{B}}^{wo}\hat{\mathcal{X}}^{\mathrm{GR}}$, where the coefficient $\hat{\mathcal{B}}^{wo} = \left[\mathcal{X}_3'(\mathbf{I} - \mathbf{P}_\mathbf{Z})'\mathbf{\Lambda}^0(\mathbf{I} - \mathbf{P}_\mathbf{Z})\mathcal{X}\right]\left[\mathcal{X}'(\mathbf{I} - \mathbf{P}_\mathbf{Z})'\mathbf{\Lambda}^0(\mathbf{I} - \mathbf{P}_\mathbf{Z})\mathcal{X}\right]^{-1}$ has the form (4.1) but with GR estimators in place of OR estimators. This estimator, differing from the CGR only in the regression coefficient, is optimal in the restricted sense of being the composite of GR estimators incorporating information on $\mathbf{z}$ that has minimum approximate variance. In general, this later composite estimator cannot be obtained as a calibration estimator. The following theorem gives conditions under which the CGR estimator is optimal in one of the two senses in non-nested matrix sampling; the proof is given in the Appendix. The nested sampling version of the theorem, with subsampling schemes and proof as in Theorem 1, is omitted for brevity.

Theorem 2 *Consider the following sampling strategies.*

(*a*) *For all three samples S_1, S_2 and S_3 assume SRS with sampling fractions $f_i = n_i/N$, and specify all constants q_{ik} in $\mathbf{\Lambda}_i$ as $q_{ik} = (n_i - 1)/N(1 - f_i)$. Consider the augmented design matrix $\mathcal{Z} = (\mathcal{X}, \mathbf{Z})$ in (2.7), where $\mathbf{Z} = \mathrm{diag}\{\mathbf{Z}_1, \mathbf{Z}_2, \mathbf{Z}_3\}$, and with the corresponding augmented vector* *of calibration totals* $\mathbf{t}_{\mathcal{Z}} = (\mathbf{0}', \mathbf{0}', \mathbf{t}'_{\mathbf{z}}, \mathbf{t}'_{\mathbf{z}}, \mathbf{t}'_{\mathbf{z}})'$. *Further, suppose that* $\mathbf{Z}_i \mathbf{h}_i = \mathbf{1}$ *for constant vectors* $\mathbf{h}_i$.

Then, the calibration procedure gives the CGR as $\hat{\mathcal{X}}_3^{\mathrm{GR}} - \hat{\mathcal{B}}\hat{\mathcal{X}}^{\mathrm{GR}} = \hat{\mathcal{X}}_3^{\mathrm{GR}} - \hat{\mathcal{B}}^{wo}\hat{\mathcal{X}}^{\mathrm{GR}}$, *i.e., the CGR estimator is the optimal composite of GR estimators incorporating information on* $\mathbf{z}$.

(b) *For all three samples* S_1, S_2 *and* S_3 *assume STRSRS with sampling fraction* $f_{ih} = n_{ih} / N_{ih}$ *in stratum* h *of sample* $i, h = 1, \ldots, H_i$ *and* N_{ih} *denoting stratum size, and specify the constants in* $\mathbf{\Lambda}_i$ *as* $q_{ik} = (n_{ih} - 1) / N_h (1 - f_{ih})$ *for all units of stratum* h. *Further, assume that within each sample the units are sorted by stratum, and consider the augmented design matrix* $\mathcal{Z} = (\mathcal{X}, \mathbf{Z}, \mathbf{D})$ *in (2.7), with corresponding augmented vector of calibration totals* $\mathbf{t}_{\mathcal{Z}} = (\mathbf{0}', \mathbf{0}', \mathbf{t}'_{\mathbf{z}}, \mathbf{t}'_{\mathbf{z}}, \mathbf{t}'_{\mathbf{z}}, \mathbf{N}'_1, \mathbf{N}'_2, \mathbf{N}'_3)'$. *The definition of* $\mathbf{D}$ *and* $\mathbf{N}_i$ *is as before.*

Then, the calibration procedure gives the CGR as $\hat{\mathcal{X}}_3^{\mathrm{OR}} - \hat{\mathcal{B}}^{o}\hat{\mathcal{X}}^{\mathrm{OR}}$, *i.e., the CGR estimator is the optimal composite of optimal regression estimators incorporating information on* $\mathbf{z}$.

(c) *For all three samples* S_1, S_2 *and* S_3 *assume stratified Poisson sampling and specify the constants* q_{ik} *in the entries of* $\mathbf{\Lambda}_i$ *as* $q_{ik} = \pi_{ihk} / (1 - \pi_{ihk})$ *for the units of stratum* h.

Then, the calibration procedure, with $\mathcal{Z}$ *and* $\mathbf{t}_{\mathcal{Z}}$ *as in* (a), *gives the CGR as* $\hat{\mathcal{X}}_3^{\mathrm{GR}} - \hat{\mathcal{B}}\hat{\mathcal{X}}^{\mathrm{GR}} = \hat{\mathcal{X}}_3^{\mathrm{OR}} - \hat{\mathcal{B}}^{o}\hat{\mathcal{X}}^{\mathrm{OR}}$, *i.e., GR and OR estimators are identical, and the CGR estimator is the optimal composite of optimal regression estimators incorporating information on* $\mathbf{z}$.

The condition $\mathbf{Z}_i \mathbf{h}_i = \mathbf{1}$ in (a) of Theorem 2 is customarily satisfied when the vector $\mathbf{z}$ contains categorical variables. Results analogous to Corollaries 1 and 2 of the previous section hold also for parts (b) and (c) of Theorem 2. Here too, for sampling designs other than those assumed in Theorem 2, the value $q_{ik} = \tilde{n}_i / (\tilde{n}_1 + \tilde{n}_2 + \tilde{n}_3)$ in the entries of $\mathbf{\Lambda}$ should be used.

Finally, by analogy to (3.2), and with the appropriate decomposition of the vector of calibrated weights $\mathbf{c}$, the composite estimator $\hat{\mathbf{X}}^{\mathrm{CGR}}$ takes now the form

$$\hat{\mathbf{X}}^{\mathrm{CGR}} = \hat{\mathbf{B}}_{1\mathbf{x}} \hat{\mathbf{X}}_1^{\mathrm{GR}} + (\mathbf{I} - \hat{\mathbf{B}}_{1\mathbf{x}}) \hat{\mathbf{X}}_3^{\mathrm{GR}},$$

where $\hat{\mathbf{X}}_1^{\mathrm{GR}}$ and $\hat{\mathbf{X}}_3^{\mathrm{GR}}$ are GR estimators using information on $\mathbf{z}$ from S_1, and on $\mathbf{y}$ and $\mathbf{z}$ from S_2 and S_3, respectively, and $\hat{\mathbf{B}}_{1\mathbf{x}}$ is the corresponding matrix regression coefficient. Similar is the expression for $\hat{\mathbf{Y}}^{\mathrm{CGR}}$. Of course, $\hat{\mathbf{X}}^{\mathrm{CGR}}$ and $\hat{\mathbf{Y}}^{\mathrm{CGR}}$ can be obtained directly through this modified $\mathbf{c}$ in the simple linear forms $\hat{\mathbf{X}}^{\mathrm{CGR}} = \mathbf{X}'_3 \mathbf{c}_3$ and $\hat{\mathbf{Y}}^{\mathrm{CGR}} = \mathbf{Y}'_3 \mathbf{c}_3$.

4.2 Core set of variables with unknown totals

We turn now to the case of matrix sampling design (d) in which the variables $\mathbf{z}$ that are common to the three samples have unknown totals. Estimation in this setting includes the construction of a composite estimator of the vector of totals $\mathbf{t}_{\mathbf{z}}$. In line with the formulation of Section 2, composite estimators of $\mathbf{t}_{\mathbf{x}}, \mathbf{t}_{\mathbf{y}}$ and $\mathbf{t}_{\mathbf{z}}$ that are best linear unbiased combinations of the HT estimators $\hat{\mathbf{X}}_1, \hat{\mathbf{Z}}_1, \hat{\mathbf{Y}}_2, \hat{\mathbf{Z}}_2, \hat{\mathbf{X}}_3, \hat{\mathbf{Y}}_3, \hat{\mathbf{Z}}_3$ are given by

$$\begin{aligned}
\hat{\mathbf{X}}^{B} &= \mathbf{B}_{1\mathbf{x}}\hat{\mathbf{X}}_1 + (\mathbf{I}-\mathbf{B}_{1\mathbf{x}})\hat{\mathbf{X}}_3 + \mathbf{B}_{3\mathbf{x}}(\hat{\mathbf{Y}}_2-\hat{\mathbf{Y}}_3) + \mathbf{B}_{2\mathbf{x}}(\hat{\mathbf{Z}}_1-\hat{\mathbf{Z}}_3) + \mathbf{B}_{4\mathbf{x}}(\hat{\mathbf{Z}}_2-\hat{\mathbf{Z}}_3)\\
\hat{\mathbf{Y}}^{B} &= \mathbf{B}_{3\mathbf{y}}\hat{\mathbf{Y}}_2 + (\mathbf{I}-\mathbf{B}_{3\mathbf{y}})\hat{\mathbf{Y}}_3 + \mathbf{B}_{1\mathbf{y}}(\hat{\mathbf{X}}_1-\hat{\mathbf{X}}_3) + \mathbf{B}_{2\mathbf{y}}(\hat{\mathbf{Z}}_1-\hat{\mathbf{Z}}_3) + \mathbf{B}_{4\mathbf{y}}(\hat{\mathbf{Z}}_2-\hat{\mathbf{Z}}_3) \qquad (4.2)\\
\hat{\mathbf{Z}}^{B} &= \mathbf{B}_{2\mathbf{z}}\hat{\mathbf{Z}}_1 + \mathbf{B}_{4\mathbf{z}}\hat{\mathbf{Z}}_2 + (\mathbf{I}-\mathbf{B}_{2\mathbf{z}}-\mathbf{B}_{4\mathbf{z}})\hat{\mathbf{Z}}_3 + \mathbf{B}_{1\mathbf{z}}(\hat{\mathbf{X}}_1-\hat{\mathbf{X}}_3) + \mathbf{B}_{3\mathbf{z}}(\hat{\mathbf{Y}}_2-\hat{\mathbf{Y}}_3).
\end{aligned}$$

The estimators in (4.2) can be written in the matrix regression form

$$\begin{pmatrix}\hat{\mathbf{X}}^{B}\\ \hat{\mathbf{Y}}^{B}\\ \hat{\mathbf{Z}}^{B}\end{pmatrix} = \begin{pmatrix}\hat{\mathbf{X}}_3\\ \hat{\mathbf{Y}}_3\\ \hat{\mathbf{Z}}_3\end{pmatrix} + \mathcal{B}\begin{pmatrix}\hat{\mathbf{X}}_1-\hat{\mathbf{X}}_3\\ \hat{\mathbf{Z}}_1-\hat{\mathbf{Z}}_3\\ \hat{\mathbf{Y}}_2-\hat{\mathbf{Y}}_3\\ \hat{\mathbf{Z}}_2-\hat{\mathbf{Z}}_3\end{pmatrix}, \qquad (4.3)$$

with the variance-minimizing matrix of coefficients given by $\mathcal{B} = -\mathrm{Cov}(\mathbf{u}_3, \mathbf{u}_{12}-\mathbf{u}_3^{\star})[V(\mathbf{u}_{12}-\mathbf{u}_3^{\star})]^{-1}$, where $\mathbf{u}_3 = (\hat{\mathbf{X}}_3', \hat{\mathbf{Y}}_3', \hat{\mathbf{Z}}_3')'$, $\mathbf{u}_3^{\star} = (\hat{\mathbf{X}}_3', \hat{\mathbf{Z}}_3', \hat{\mathbf{Y}}_3', \hat{\mathbf{Z}}_3')'$, $\mathbf{u}_{12} = (\hat{\mathbf{X}}_1', \hat{\mathbf{Z}}_1', \hat{\mathbf{Y}}_2', \hat{\mathbf{Z}}_2')'$. With estimated covariance and variance matrices we obtain the estimated optimal matrix $\hat{\mathcal{B}}^{o}$, and (4.3) becomes then an optimal multivariate regression estimator. Then, proceeding as in Section 2, it can be shown that

$$\hat{\mathcal{B}}^{o} = (\mathcal{X}_{3-}'\mathbf{\Lambda}^{0}\mathcal{X})(\mathcal{X}'\mathbf{\Lambda}^{0}\mathcal{X})^{-1},$$

where

$$\mathcal{X} = \begin{pmatrix} -\mathbf{X}_1 & -\mathbf{Z}_1 & \mathbf{0} & \mathbf{0}\\ \mathbf{0} & \mathbf{0} & -\mathbf{Y}_2 & -\mathbf{Z}_2\\ \mathbf{X}_3 & \mathbf{Z}_3 & \mathbf{Y}_3 & \mathbf{Z}_3\end{pmatrix} \qquad (4.4)$$

is the design matrix corresponding to the regression estimator (4.3), $\mathcal{X}_{3-}$ is the matrix $\mathcal{X}$ with the second column eliminated and the first two rows set equal to zero, and $\mathbf{\Lambda}^{0}$ is as in Section 2.

Replacing the matrix $\mathbf{\Lambda}^{0}$ with the weighting matrix $\mathbf{\Lambda}$, gives the generalized regression coefficient $\hat{\mathcal{B}} = (\mathcal{X}_{3-}'\mathbf{\Lambda}\mathcal{X})(\mathcal{X}'\mathbf{\Lambda}\mathcal{X})^{-1}$, and (4.3) becomes the CGR estimator of $(\mathbf{t}_{\mathbf{x}}', \mathbf{t}_{\mathbf{y}}', \mathbf{t}_{\mathbf{z}}')'$

$$\begin{pmatrix}\hat{\mathbf{X}}^{\mathrm{CGR}}\\ \hat{\mathbf{Y}}^{\mathrm{CGR}}\\ \hat{\mathbf{Z}}^{\mathrm{CGR}}\end{pmatrix} = \begin{pmatrix}\hat{\mathbf{X}}_3\\ \hat{\mathbf{Y}}_3\\ \hat{\mathbf{Z}}_3\end{pmatrix} + \hat{\mathcal{B}}\begin{pmatrix}\hat{\mathbf{X}}_1-\hat{\mathbf{X}}_3\\ \hat{\mathbf{Z}}_1-\hat{\mathbf{Z}}_3\\ \hat{\mathbf{Y}}_2-\hat{\mathbf{Y}}_3\\ \hat{\mathbf{Z}}_2-\hat{\mathbf{Z}}_3\end{pmatrix}. \qquad (4.5)$$

The estimator (4.5) can be conveniently obtained through a calibration procedure that gives a vector of calibrated weights for the combined sample S having the form $\mathbf{c} = \mathbf{w} + \mathbf{\Lambda}\mathcal{X}(\mathcal{X}'\mathbf{\Lambda}\mathcal{X})^{-1}(\mathbf{0} - \mathcal{X}'\mathbf{w})$, as before, but now satisfying the additional constraint $\hat{\mathbf{Z}}_1^{\mathrm{CGR}} = \hat{\mathbf{Z}}_2^{\mathrm{CGR}} = \hat{\mathbf{Z}}_3^{\mathrm{CGR}}$. Expression (4.5) is then obtained simply as $\mathcal{X}_{3-}'\mathbf{c}$, based on sample S_3.

The explicit expression (4.2), different for the optimal regression and the generalized regression variants only in the form of the linear coefficients, shows that the composite estimators of $\mathbf{t}_{\mathbf{x}}$ and $\mathbf{t}_{\mathbf{y}}$ are more efficient than their counterparts in matrix sampling design (c), equation (2.2), because they incorporate information on the common variables $\mathbf{z}$, assuming non-zero correlation with $\mathbf{x}$ and $\mathbf{y}$. Particularly remarkable is the expression for the composite estimator of $\mathbf{t}_{\mathbf{z}}$: it involves a linear combination of the three HT estimators of $\mathbf{t}_{\mathbf{z}}$ derived from the three samples, plus the two regression terms implying additional efficiency through the correlation of $\mathbf{z}$ with $\mathbf{x}$ and $\mathbf{y}$. One would expect the additional terms to be zero because an optimal combination of the three estimators should incorporate all information on $\mathbf{z}$ available in the three samples. In general, however, the associated coefficients are not zero. In non-nested sampling, conditions under which these coefficients are zero are given by the following proposition, the proof of which is given in the Appendix. The result should also hold in nested sampling.

Proposition 1 *The coefficients* $\mathbf{B}_{1\mathbf{z}}$ *and* $\mathbf{B}_{3\mathbf{z}}$ *in the estimator* $\hat{\mathbf{Z}}^B$ *in (4.2) are zero only if*

$$\begin{aligned} \left[V\left(\hat{\mathbf{Z}}_1\right)\right]^{-1}\operatorname{Cov}\left(\hat{\mathbf{X}}_1,\hat{\mathbf{Z}}_1\right) &= \left[V\left(\hat{\mathbf{Z}}_3\right)\right]^{-1}\operatorname{Cov}\left(\hat{\mathbf{X}}_3,\hat{\mathbf{Z}}_3\right) \\ \left[V\left(\hat{\mathbf{Z}}_2\right)\right]^{-1}\operatorname{Cov}\left(\hat{\mathbf{Y}}_2,\hat{\mathbf{Z}}_2\right) &= \left[V\left(\hat{\mathbf{Z}}_3\right)\right]^{-1}\operatorname{Cov}\left(\hat{\mathbf{Y}}_3,\hat{\mathbf{Z}}_3\right). \end{aligned} \tag{4.6}$$

This can happen only if the sampling designs for the three samples are identical, including equal sample sizes, or only if the sampling design across samples is the same design with equal inclusion probability for all units, but not necessarily with the same sample size.

Noticing that the quantities on each side of the equations (4.6) are regression coefficients, according to Proposition 1 the terms of the estimator $\hat{\mathbf{Z}}^B$ incorporating the correlation of $\mathbf{z}$ with $\mathbf{x}$ and $\mathbf{y}$ are zero only if the effect of the regression of $\mathbf{x}$ and $\mathbf{y}$ on $\mathbf{z}$ is identical in samples S_1 and S_3 and in samples S_2 and S_3, respectively. The essence of this finding is that estimation of $\mathbf{t}_{\mathbf{z}}$ using only information on $\mathbf{z}$ from the three samples, but ignoring information on $\mathbf{x}$ and $\mathbf{y}$, will be suboptimal when there is differential regression effect of $\mathbf{x}$ and $\mathbf{y}$ on $\mathbf{z}$ in the various samples. The efficiency of $\hat{\mathbf{Z}}^B$ relative to the composite estimator $\tilde{\mathbf{Z}}^B$ that uses only information on $\mathbf{z}$ was possible to gauge in the simple setting involving scalar x, y and z, simple random sampling for S_1 and S_3 and Bernoulli sampling for S_2, and equal sampling rates for all three samples. Then only the first equation of (4.6) holds. After much tedious algebra the efficiency of $\hat{\mathbf{Z}}^B$ relative to $\tilde{\mathbf{Z}}^B$ was derived to be $\left[V\left(\tilde{\mathbf{Z}}^B\right)-V\left(\hat{\mathbf{Z}}^B\right)/V\left(\tilde{\mathbf{Z}}^B\right)\right]=G/H$, with

$$\begin{aligned} G &= 2\left(r_{xz}^2-1\right)\left(r_{yz}cv_y-cv_z\right)^2 \\ H &= \left(cv_z^2+1\right)\left(\left(12-9r_{yz}^2\right)r_{xz}^2-3r_{xy}\left(2r_{yz}r_{xz}-1\right)+12\left(r_{yz}^2-1\right)\right)cv_z^2cv_y^2 \\ &+ 2\left(r_{xy}^2+r_{yz}^2\right)cv_y^2+8\left(r_{xz}^2-1\right)cv_y^2-4r_{yz}r_{xy}r_{xz}cv_y^2 \\ &+ 6\left(r_{xz}^2-1\right)cv_z\left(cv_z-2r_{yz}cv_y\right) \end{aligned}$$

where r_{xy}, r_{xz} and r_{yz} denote population correlation coefficients, and cv_y, cv_z denote coefficients of variation. Although in this setting the departure from the conditions of Proposition 1 is minimal, different configurations of admissible values for $r_{xy}, r_{xz}, r_{yz}, cv_y$ and cv_z show that the efficiency gain may be substantial, making up for the inefficiency of the HT estimator of $\mathbf{t}_{\mathbf{z}}$ based on the Bernoulli sample S_2. For example, when $r_{xy}=0.3, r_{xz}=0.3, r_{yz}=0.3$ and $cv_y=0.1, cv_z=0.6$, the efficiency gain is 23%. In the case of the composite optimal regression estimator $\hat{\mathbf{Z}}^{\text{COR}}$, with estimated coefficients $\hat{\mathbf{B}}_{1\mathbf{z}}^o$ and $\hat{\mathbf{B}}_{3\mathbf{z}}^o$, the regression coefficients in (4.6) are estimated, and thus the equalities in (4.6) would never hold exactly because of the sample differences. Likewise in the case of the CGR estimator $\hat{\mathbf{Z}}^{\text{CGR}}$, for which equations formally identical to (4.6) are given in terms of sample generalized regression coefficients.

Regarding the efficiency of the CGR estimator (4.5), an exact analogue of Theorem 1 holds in the present setting, with the same sampling strategies for which the CGR estimator is optimal regression estimator and asymptotically BLUE.

Composite estimation for a matrix sampling scheme involving a core set of variables with both known and unknown totals can be carried out using the obvious extended calibration scheme.

5 Domain estimation

Composite estimators for domains (subpopulations) of interest may be readily obtained using the calibrated weights derived in the previous sections, that is, by summing the weighted values of a variable over any domain $U_d \subset U$. For instance, letting $\mathbf{X}_{id}$ denote the matrix $\mathbf{X}_i$, for sample S_i, with the entries of the k^{th} row set equal to 0 if $k \notin U_d$, the CGR estimator of the domain total $\mathbf{t}_{\mathbf{x}d}$ based on the weights of S_3 calibrated with the scheme of design (c) (see Section 3) is given by

$$\hat{\mathbf{X}}_{3d}^{\text{CGR}} = \mathbf{X}'_{3d}\mathbf{c}_3 = \hat{\mathbf{X}}_{3d}^{\text{GR}} + \mathbf{X}'_{3d}\mathbf{L}_{\Psi}\mathcal{X}(\mathcal{X}'\mathbf{L}_{\Psi}\mathcal{X})^{-1}\left[\hat{\mathbf{X}}_1 - \hat{\mathbf{X}}_3^{\text{GR}}\right],$$

where $\hat{\mathbf{X}}_{3d}^{\text{GR}} = \hat{\mathbf{X}}_{3d} + \mathbf{X}'_{3d}\boldsymbol{\Lambda\Psi}(\boldsymbol{\Psi}'\boldsymbol{\Lambda\Psi})^{-1}\left(\hat{\mathbf{Y}}_2 - \hat{\mathbf{Y}}_3\right)$ and the subscript d indicates domain. The CGR estimator $\hat{\mathbf{X}}_{1d}^{\text{CGR}}$ based on sample S_1 is obtained in the same manner. However, unlike the population-level estimator (3.2), resulting from calibration of two estimators to each other at population level, the estimators $\hat{\mathbf{X}}_{1d}^{\text{CGR}}$ and $\hat{\mathbf{X}}_{3d}^{\text{CGR}}$ are not constructed as composites of two domain estimators, based on samples S_1 and S_3, and they are not identical. Moreover, although both $\hat{\mathbf{X}}_{1d}^{\text{CGR}}$ and $\hat{\mathbf{X}}_{3d}^{\text{CGR}}$ incorporate information on $\mathbf{x}$ from samples S_1 and S_3, their construction (non-customized at domain level) may entail some loss of efficiency.

A simple modification of the calibration procedure that leads to efficient composite estimation for all totals of interest involves the augmentation of the design matrix with columns defined at each domain level for the relevant variables. Thus, for design (c) estimation of the domain total $\mathbf{t}_{\mathbf{x}d}$ involves the augmentation of the design matrix $\mathcal{X}$ in (2.7) with the column $\left(-\mathbf{X}'_{1d}, \mathbf{0}', \mathbf{X}'_{3d}\right)'$. The resulting estimator, $\tilde{\mathbf{X}}_d^{\text{CGR}}$, may be written in the forms

$$\begin{aligned}\tilde{\mathbf{X}}_d^{\text{CGR}} &= \hat{\mathbf{X}}_{3d} + \hat{\mathbf{B}}_{1\mathbf{x}d}\left(\hat{\mathbf{X}}_1 - \hat{\mathbf{X}}_3\right) + \hat{\mathbf{B}}_{2\mathbf{x}d}\left(\hat{\mathbf{Y}}_2 - \hat{\mathbf{Y}}_3\right) + \hat{\mathbf{B}}_{3\mathbf{x}d}\left(\hat{\mathbf{X}}_{1d} - \hat{\mathbf{X}}_{3d}\right) \\ &= \hat{\mathbf{B}}_{1\mathbf{x}d}\tilde{\mathbf{X}}_{1d}^{\text{GR}} + \left(\mathbf{I} - \hat{\mathbf{B}}_{1\mathbf{x}d}\right)\tilde{\mathbf{X}}_{3d}^{\text{GR}},\end{aligned} \tag{5.1}$$

where $\tilde{\mathbf{X}}_{1d}^{\text{GR}}$ and $\tilde{\mathbf{X}}_{3d}^{\text{GR}}$ are now the GR domain estimators incorporating the regression effect of the second and third terms of (5.1). Adding another term in (5.1) involving the difference $\hat{\mathbf{Y}}_{2d} - \hat{\mathbf{Y}}_{3d}$ may not improve appreciably the efficiency of $\tilde{\mathbf{X}}_d^{\text{CGR}}$ but will be necessary if estimation of the domain total $\mathbf{t}_{\mathbf{y}d}$ is also required. In any particular situation, the augmentation of the design matrix $\mathcal{X}$ involves only those components of $\mathbf{x}$ or $\mathbf{y}$ for which domain estimates are needed. A possible drawback of this procedure is the additional computational burden, which increases with the number of domains and the variables for which domain estimation is required.

An alternative approach that may be more appropriate when the domain estimates of interest are numerous, involves the separate production of the domain estimates by carrying out the composite calibration only at the domain level. For the domain total $\mathbf{t}_{\mathbf{x}d}$, this would give the domain CGR estimator, in analogy with the population CGR estimator (3.2),

$$\breve{\mathbf{X}}_d^{\text{CGR}} = \breve{\mathbf{B}}_{1\mathbf{x}d}\hat{\mathbf{X}}_{1d} + \left(\mathbf{I} - \breve{\mathbf{B}}_{1\mathbf{x}d}\right)\breve{\mathbf{X}}_{3d}^{\text{GR}},$$

where $\breve{\mathbf{B}}_{1\mathbf{x}d} = \mathbf{X}'_{3d}\mathbf{L}_{\mathbf{\Psi}_d}\mathbf{X}_d\left(\mathbf{X}'_d\mathbf{L}_{\mathbf{\Psi}_d}\mathbf{X}\right)_d^{-1}$ and $\breve{\mathbf{X}}_{3d}^{\text{GR}} = \hat{\mathbf{X}}_{3d} + \mathbf{X}'_{3d}\mathbf{\Lambda}\mathbf{\Psi}_d\left(\mathbf{\Psi}'_d\mathbf{\Lambda}\mathbf{\Psi}_d\right)^{-1}\left(\hat{\mathbf{Y}}_{2d} - \hat{\mathbf{Y}}_{3d}\right)$. The efficiency of the joint estimator $\left(\breve{\mathbf{X}}_d^{\text{CGR}}, \breve{\mathbf{Y}}_d^{\text{CGR}}\right)$ over the estimator $\left(\hat{\mathbf{X}}_d^{\text{CGR}}, \hat{\mathbf{Y}}_d^{\text{CGR}}\right)$ can be verified under the conditions of the following proposition (its proof in the Appendix).

Proposition 2 *Under the sampling schemes of Theorem 1,*

$$\widehat{\text{AV}}\begin{pmatrix}\breve{\mathbf{X}}_{3d}^{\text{CGR}}\\ \breve{\mathbf{Y}}_{3d}^{\text{GGR}}\end{pmatrix} < \widehat{\text{AV}}\begin{pmatrix}\hat{\mathbf{X}}_{3d}^{\text{CGR}}\\ \hat{\mathbf{Y}}_{3d}^{\text{CGR}}\end{pmatrix}.$$

Notably, the drawback of a separate production of the domain estimates, through composite calibration at the domain level, is the loss of consistency among estimates at population level and domain level.

The above considerations extend to domain estimation for matrix sampling design (d).

6 A simulation study

We have conducted a simulation to study the relative performance of the various composite estimators for the nested version of the basic design (c). Values of correlated scalar variables x and y were generated from a bivariate log-normal distribution with mean and variance parameters $\left(\mu_x, \mu_y\right)$ and $\left(\sigma_x^2, \sigma_y^2\right)$. With fixed $\mu_x = 3$, $\mu_y = 5$, four combinations of variances $\left(\sigma_x^2, \sigma_y^2\right)$ (5 and 10) and three values of the correlation $\rho(x, y)$ (0.5, 0.7, 0.9) were considered. Variances $\sigma_x^2 = 5$, $\sigma_x^2 = 10$ imply skewness 2.65 and 4.33, respectively, while variances $\sigma_y^2 = 5$, $\sigma_y^2 = 10$ imply skewness 1.43 and 2.15. For each of these twelve settings, a population of size $N = 1{,}000{,}000$ was created. From each of the twelve populations a simple random sample S of size $n = 5{,}000$ was drawn without replacement, and split into three simple random subsamples (S_1, S_2, S_3) with two different allocations, namely, $(n_1 = 2{,}000, n_2 = 2{,}000, n_3 = 1{,}000)$ and $(n_1 = 1{,}500, n_2 = 1{,}500, n_3 = 2{,}000)$, the second allocation giving larger combined samples $S_1 \cup S_3$ and $S_2 \cup S_3$. Thus, a total of 24 simulation settings were created. For each such setting, we computed the HT estimators of the totals t_x and t_y using the full sample S, as well as the HT estimator of t_x using S_1 and S_3 and the HT estimator of t_y using S_2 and S_3. For the HT estimators based on two subsamples, we employed the simple method for combining two subsamples (Gonzales and Eltinge 2008) by a weighting adjustment involving the probability of selection of a population unit in S_1 or in S_3 and in S_2 or in S_3. In addition, for both t_x and t_y we computed the CGR and COR estimators. Each simulation sampling setting was repeated 10,000 times.

The simulated bias (in percent) of all estimators was smaller than 0.05%, with the exception of two settings involving $\sigma_x^2 = 10,$ with associated population skewness of 4.33, where the largest observed values 0.14% and 0.17% correspond to CGR and COR for $t_x,$ respectively, in the sample allocation (2,000, 2,000, 1,000), dropping to 0.10% and 0.13% in the more favorable allocation (1,500, 1,500, 2,000). Thus the relative efficiencies of the estimators are evaluated using their simulated design variances.

Table 6.1 shows the efficiency of the composite estimators CGR and COR relative to the HT estimators that use $S_1 \cup S_3$ and $S_2 \cup S_3.$ The measure of this relative efficiency is the percent relative difference of variances [V(CGR)-V(HT)]/V(HT) and [V(COR)-V(HT)]/V(HT). A negative value of this measure indicates the efficiency gain achieved by the two composite estimators. Not shown in Table 6.1, the simulated loss of efficiency of the HT estimators of both t_x and t_y due to not using the full sample S is very close to the nominal loss for SRS, that is, 66.8% for the allocation (2,000, 2,000, 1,000), and 43.1% for the allocation (1,500, 1,500, 2,000).

Table 6.1
Relative differences (in percent) of variances of CGR and COR to HT for x and y, based on 10,000 simulated samples with two different sample allocations.

(n1, n2, n3)	(2,000; 2,000; 1,000)				(1,500; 1,500; 2,000)			
	x		*y*		*x*		*y*	
	CGR	COR	CGR	COR	CGR	COR	CGR	COR
$\sigma_x^2 = 5$ $\sigma_y^2 = 5$								
$\rho = 0.5$	-2.24	-6.86	26.39	-6.23	-5.19	-6.29	12.59	-6.52
$\rho = 0.7$	-11.90	-14.75	10.21	-13.96	-12.78	-13.24	0.25	-13.13
$\rho = 0.9$	-24.89	-28.57	-12.49	-28.10	-21.55	-23.37	-14.55	-23.03
$\sigma_x^2 = 5$ $\sigma_y^2 = 10$								
$\rho = 0.5$	-0.27	-6.75	6.50	-6.26	-3.94	-6.60	0.50	-6.44
$\rho = 0.7$	-11.47	-14.56	-6.29	-14.04	-12.87	-13.51	-9.51	-13.10
$\rho = 0.9$	-28.14	-28.42	-25.74	-28.23	-23.70	-23.54	-22.07	-23.09
$\sigma_x^2 = 10$ $\sigma_y^2 = 5$								
$\rho = 0.5$	-4.57	-6.51	28.64	-6.17	-5.90	-5.98	17.57	-6.44
$\rho = 0.7$	-11.29	-14.37	16.08	-13.92	-11.66	-12.90	6.69	-13.00
$\rho = 0.9$	-20.32	-28.09	-2.46	-28.19	-18.46	-22.97	-6.97	-22.91
$\sigma_x^2 = 10$ $\sigma_y^2 = 10$								
$\rho = 0.5$	-4.79	-6.49	8.54	-6.13	-6.06	-6.22	3.41	-6.34
$\rho = 0.7$	-13.27	-14.28	-2.57	-13.95	-13.27	-13.15	-6.00	-12.93
$\rho = 0.9$	-26.01	-28.06	-20.37	-28.21	-22.18	-23.17	-18.48	-22.89

For the variable $x,$ using the CGR estimator at low correlation $\rho = 0.5$ and with allocation (2,000, 2,000, 1,000) leads to an efficiency gain that ranges from 0.27% to 4.79% at the four different variance settings; this gain reflects the amount of lost information recovered by the CGR estimator. Substantial gain is achieved at $\rho = 0.7,$ ranging from 11.29% to 13.27%, and more so at $\rho = 0.9,$ ranging from 20.32% to 28.14%. With sample allocation (1,500, 1,500, 2,000) the CGR estimator performs better at $\rho = 0.5,$ and $\rho = 0.7,$ but not at $\rho = 0.9.$ Additional gain is achieved by the COR estimator, which is more efficient than the CGR estimator in all but two settings (where the estimators are equally efficient, see column 7). The efficiency of the COR estimator relative to HT estimator is close to the nominal for SRS efficiency, which is 6.25, 13.92 and 28.12 at $\rho = 0.5,$ $\rho = 0.7,$ $\rho = 0.9,$ respectively, for allocation (2,000, 2,000, 1,000), and 6.417, 13.186 and 23.30 for allocation (1,500, 1,500, 2,000); see quantity E in Section 2, third last paragraph. As expected, the CGR estimator competes better with the COR estimator with increasing correlation and sample size.

For the variable $y,$ the CGR estimator is inferior to the HT estimator at correlation level $\rho = 0.5$ and in half of the simulated settings at $\rho = 0.7;$ see positive values in columns 4 and 8. This inefficiency of the CGR estimator ranges from 6.50% (at $\rho = 0.7)$ to 28.64% (at $\rho = 0.5)$ in the sample allocation (2,000, 2,000, 1,000), and reduces to 0.25% (at $\rho = 0.7)$ to 17.57% (at $\rho = 0.5)$ in the sample allocation (1,500, 1,500, 2,000). This is explained by the larger skewness of x (the x variable being used a auxiliary to y in the regression procedure); the lower levels of inefficiency are observed at $\sigma_y^2 = 10,$ when the differential in skewness between x and y is the smallest. On the other hand, at correlation $\rho = 0.9$ and with allocation (2,000, 2,000, 1,000), the efficiency gain of the CGR estimator relative to the HT estimator ranges from 2.46% (when the skewness differential is the largest) to 25.74% (when the skewness differential is the smallest), with similar efficiency levels displayed for allocation (1,500, 1,500, 2,000). The COR estimator is more efficient than the CGR estimator in all settings, the relative efficiency being close to the nominal one for SRS (same efficiency as with $x).$ For y too, the CGR estimator competes better with COR estimator with increasing correlation and sample size.

This limited empirical study, which essentially simulates the SRS version of Theorem $1(a'),$ confirms the theory on the efficiency of the optimal estimator COR, even for modest sample size, and shows the usefulness of the two composite estimators CGR and COR in partially recovering the information loss due to splitting the full questionnaire. It also shows that the practical CGR estimator is not always a good substitute of the COR estimator for small samples and low correlation between x and $y.$

7 Discussion

The proposed estimation method for matrix sampling involves a single-step calibration of the weights of the combined sample. Estimates of totals for all variables can be obtained by using only the units of sample S_3 and their calibrated weights which incorporate all the available information from all three samples. These weights could be used to calculate other weighted statistics, including means, ratios, quantiles and regression coefficients. When the second-order inclusion probabilities are known, including cross-sample inclusion probabilities in the nested case, the calibration procedure of Section 2 can produce composite optimal regression estimators and their variances, but with great computational difficulty. For general sampling settings, the much simpler calibration scheme of Section 3 generates readily composite generalized regression estimators, which for certain sampling strategies are optimal regression estimators.

Estimation of the variance of a CGR estimator may, in principle, be based on the method of Taylor linearization of the generalized regression estimator (see, e.g., Särndal et al. 1992, pages 235, 237). This approach requires calculations that may not be practical, or even feasible for complex sampling designs because the second-order inclusion probabilities are rarely known. Replication methods for variance estimation, such as the jackknife method or the bootstrap method (see, for example, Rust and Rao 1996), can be applied to the CGR estimators of the previous sections. For example, the jackknife method, customarily used in surveys with stratified multistage sampling design, could be used to replicate the calibration procedures that give rise to the CGR estimators. For the non-nested design, this requires applying the jackknife method to the combined sample, with the three independent samples treated as sample superstrata containing the sample strata. The replication procedure would involve then the combined sample sorted by sample and by strata within each sample, to produce replicates of the calibrated weights defined in the previous sections. The total number of strata used in the jackknife replication procedure is the total number of strata in the three samples, with each replicate involving all strata. Public-use microfiles may include the replicate calibrated weights for easy variance estimation by users. For this purpose too, replicate weights for S_3 only need to be included, bringing about substantial economy of data storage in such microfiles. The case of nested design is more complicated. Further investigation in this direction will be a topic of separate study.

The described estimation method may be readily adapted to matrix sampling designs with more than two subquestionnaires or more than three subsamples, making more evident the operational power of the calibration procedure. In each case, the crucial step is to determine the design matrix $\boldsymbol{\mathcal{X}}$. In such designs there may be more complex patterns with respect to the number of subquestionnaires administered to the various subsamples. All composite estimates can then be obtained using the weighted variable values only from the minimum number of subsamples that in combination contain all items.

Acknowledgements

The author thanks the Editor, Associate Editor and two referees for their comments and suggestions, which have substantially improved the paper.

Appendix

Proof of Lemma 1

For the partitioned matrix $\boldsymbol{\mathcal{X}} = (\mathbf{X}, \boldsymbol{\Psi})$, the vector $\mathbf{c} = \mathbf{w} + \mathbf{R}\boldsymbol{\mathcal{X}}(\boldsymbol{\mathcal{X}}'\mathbf{R}\boldsymbol{\mathcal{X}})^{-1}(\mathbf{t}_{\boldsymbol{\mathcal{X}}} - \boldsymbol{\mathcal{X}}'\mathbf{w})$ takes the form

$$\begin{aligned}\mathbf{c} &= \mathbf{w} + (\mathbf{RX}, \mathbf{R\Psi})\begin{pmatrix}\mathbf{X}'\mathbf{RX} & \mathbf{X}'\mathbf{R\Psi}\\ \boldsymbol{\Psi}'\mathbf{RX} & \boldsymbol{\Psi}'\mathbf{R\Psi}\end{pmatrix}^{-1}\begin{pmatrix}\mathbf{t}_{\mathbf{X}} - \mathbf{X}'\mathbf{w}\\ \mathbf{t}_{\boldsymbol{\Psi}} - \boldsymbol{\Psi}'\mathbf{w}\end{pmatrix}\\ &= \mathbf{w} + (\mathbf{RXA}_{11} + \mathbf{R\Psi A}_{21})(\mathbf{t}_{\mathbf{X}} - \mathbf{X}'\mathbf{w}) + (\mathbf{RXA}_{12} + \mathbf{R\Psi A}_{22})(\mathbf{t}_{\boldsymbol{\Psi}} - \boldsymbol{\Psi}'\mathbf{w}),\end{aligned}$$

where, from algebra of partitioned matrices, $\mathbf{A}_{11} = [\mathbf{X}'\mathbf{RX} - \mathbf{X}'\mathbf{R\Psi}(\boldsymbol{\Psi}'\mathbf{R\Psi})^{-1}\boldsymbol{\Psi}'\mathbf{RX}]^{-1} = [\mathbf{X}'\mathbf{R}(\mathbf{I} - \mathbf{P}_{\boldsymbol{\Psi}})\mathbf{X}]^{-1}$ with $\mathbf{P}_{\boldsymbol{\Psi}} = \boldsymbol{\Psi}(\boldsymbol{\Psi}'\mathbf{R\Psi})^{-1}\boldsymbol{\Psi}'\mathbf{R}$, $\mathbf{A}_{22} = [\boldsymbol{\Psi}'\mathbf{R}(\mathbf{I} - \mathbf{P}_{\mathbf{X}})\boldsymbol{\Psi}]^{-1}$ with $\mathbf{P}_{\mathbf{X}} = \mathbf{X}(\mathbf{X}'\mathbf{RX})^{-1}\mathbf{X}'\mathbf{R}$, $\mathbf{A}_{12} = -(\mathbf{X}'\mathbf{RX})^{-1}$

$(\mathbf{X}'\mathbf{R\Psi})\mathbf{A}_{22}$ and $\mathbf{A}_{21} = -(\mathbf{\Psi}'\mathbf{R\Psi})^{-1}(\mathbf{\Psi}'\mathbf{RX})\mathbf{A}_{11}$. Then, equation (2.9) follows without difficulty. To prove equation (2.10), we set $\mathbf{c}_{\Psi} = \mathbf{w} + \mathbf{R\Psi}(\mathbf{\Psi}'\mathbf{R\Psi})^{-1}(\mathbf{t}_{\Psi} - \mathbf{\Psi}'\mathbf{w})$, so that $(\mathbf{X}'\mathbf{R\Psi})(\mathbf{\Psi}'\mathbf{R\Psi})^{-1}(\mathbf{t}_{\Psi} - \mathbf{\Psi}'\mathbf{w}) = \mathbf{X}'\mathbf{c}_{\Psi} - \mathbf{X}'\mathbf{w}$, and use the alternative form $\mathbf{A}_{22} = (\mathbf{\Psi}'\mathbf{R\Psi})^{-1} + (\mathbf{\Psi}'\mathbf{R\Psi})^{-1}(\mathbf{\Psi}'\mathbf{RX})\mathbf{A}_{11}(\mathbf{X}'\mathbf{R\Psi})(\mathbf{\Psi}'\mathbf{R\Psi})^{-1}$ to write $\mathbf{c}$ above without the second term as

$$\begin{aligned}
\mathbf{w} \quad + &\quad \mathbf{R\Psi A}_{22}(\mathbf{t}_{\Psi} - \mathbf{\Psi}'\mathbf{w}) - \mathbf{RX}(\mathbf{X}'\mathbf{RX})^{-1}(\mathbf{X}'\mathbf{R\Psi})\mathbf{A}_{22}(\mathbf{t}_{\Psi} - \mathbf{\Psi}'\mathbf{w})\\
= &\quad \mathbf{w} + \left[\mathbf{R\Psi}(\mathbf{\Psi}'\mathbf{R\Psi})^{-1} + \mathbf{R\Psi}(\mathbf{\Psi}'\mathbf{R\Psi})^{-1}(\mathbf{\Psi}'\mathbf{RX})\mathbf{A}_{11}(\mathbf{X}'\mathbf{R\Psi})(\mathbf{\Psi}'\mathbf{R\Psi})^{-1}\right](\mathbf{t}_{\Psi} - \mathbf{\Psi}'\mathbf{w})\\
- &\quad \mathbf{RX}(\mathbf{X}'\mathbf{RX})^{-1}\left[\mathbf{I} + (\mathbf{X}'\mathbf{R\Psi})(\mathbf{\Psi}'\mathbf{R\Psi})^{-1}(\mathbf{\Psi}'\mathbf{RX})\mathbf{A}_{11}\right](\mathbf{X}'\mathbf{R\Psi})(\mathbf{\Psi}'\mathbf{R\Psi})^{-1}(\mathbf{t}_{\Psi} - \mathbf{\Psi}'\mathbf{w})\\
= &\quad \mathbf{c}_{\Psi} + \mathbf{R\Psi}(\mathbf{\Psi}'\mathbf{R\Psi})^{-1}(\mathbf{\Psi}'\mathbf{RX})\mathbf{A}_{11}(\mathbf{X}'\mathbf{c}_{\Psi} - \mathbf{X}'\mathbf{w})\\
- &\quad \mathbf{RX}(\mathbf{X}'\mathbf{RX})^{-1}\left[\mathbf{I} + (\mathbf{X}'\mathbf{R\Psi})(\mathbf{\Psi}'\mathbf{R\Psi})^{-1}(\mathbf{\Psi}'\mathbf{RX})\mathbf{A}_{11}\right](\mathbf{X}'\mathbf{c}_{\Psi} - \mathbf{X}'\mathbf{w})\\
= &\quad \mathbf{c}_{\Psi} + \mathbf{R\Psi}(\mathbf{\Psi}'\mathbf{R\Psi})^{-1}(\mathbf{\Psi}'\mathbf{RX})\mathbf{A}_{11}(\mathbf{X}'\mathbf{c}_{\Psi} - \mathbf{X}'\mathbf{w})\\
- &\quad \mathbf{RX}(\mathbf{X}'\mathbf{RX})^{-1}\left[\mathbf{I} + \left(\mathbf{X}'\mathbf{RX} - \mathbf{A}_{11}^{-1}\right)\mathbf{A}_{11}\right](\mathbf{X}'\mathbf{c}_{\Psi} - \mathbf{X}'\mathbf{w})\\
= &\quad \mathbf{c}_{\Psi} + \left[\mathbf{R\Psi}(\mathbf{\Psi}'\mathbf{R\Psi})^{-1}(\mathbf{\Psi}'\mathbf{RX}) - \mathbf{RX}\right]\mathbf{A}_{11}(\mathbf{X}'\mathbf{c}_{\Psi} - \mathbf{X}'\mathbf{w})\\
= &\quad \mathbf{c}_{\Psi} - \mathbf{R}(\mathbf{I} - \mathbf{P}_{\Psi})\mathbf{X}\left[\mathbf{X}'\mathbf{R}(\mathbf{I} - \mathbf{P}_{\Psi})\mathbf{X}\right]^{-1}(\mathbf{X}'\mathbf{c}_{\Psi} - \mathbf{X}'\mathbf{w}).
\end{aligned}$$

Adding to this the second term of $\mathbf{c}$ from (2.9) gives (2.10), in the explicit form

$$\mathbf{c}_{\Psi} + \mathbf{R}(\mathbf{I} - \mathbf{P}_{\Psi})\mathbf{X}\left[\mathbf{X}'\mathbf{R}(\mathbf{I} - \mathbf{P}_{\Psi})\mathbf{X}\right]^{-1}(\mathbf{t}_{\mathbf{X}} - \mathbf{X}'\mathbf{c}_{\Psi}).$$

Proof of Theorem 1

(a) Calibration with design matrix $\mathcal{Z} = (\mathcal{X}, \mathbf{D})$ and vector of totals $\mathbf{t}_{\mathcal{Z}} = (\mathbf{0}', \mathbf{N}')'$, with $\mathbf{0} = (\mathbf{0}', \mathbf{0}')'$, $\mathbf{N} = (\mathbf{N}_1', \mathbf{N}_2', \mathbf{N}_3')'$, gives the vector of calibrated weights $\mathbf{c} = \mathbf{w} + \mathbf{\Lambda}\mathcal{Z}(\mathcal{Z}'\mathbf{\Lambda}\mathcal{Z})^{-1}(\mathbf{t}_{\mathcal{Z}} - \mathcal{Z}'\mathbf{w})$, which by Lemma 1 is written as $\mathbf{c} = \mathbf{c}_{\mathbf{D}} + \mathbf{L}_{\mathbf{D}}\mathcal{X}(\mathcal{X}'\mathbf{L}_{\mathbf{D}}\mathcal{X})^{-1}(\mathbf{0} - \mathcal{X}'\mathbf{c}_{\mathbf{D}})$, where $\mathbf{c}_{\mathbf{D}} = \mathbf{w} + \mathbf{\Lambda D}(\mathbf{D}'\mathbf{\Lambda D})^{-1}(\mathbf{N} - \mathbf{D}'\mathbf{w})$ and $\mathbf{L}_{\mathbf{D}} = \mathbf{\Lambda}(\mathbf{I} - \mathbf{P}_{\mathbf{D}})$, with $\mathbf{P}_{\mathbf{D}} = \mathbf{D}(\mathbf{D}'\mathbf{\Lambda D})^{-1}\mathbf{D}'\mathbf{\Lambda}$. For STRSRS with $f_{ih} = n_{ih}/N_{ih}$, $\mathbf{D}'\mathbf{w} = \hat{\mathbf{N}} = \mathbf{N}$, and thus $\mathbf{c} = \mathbf{w} + \mathbf{L}_{\mathbf{D}}\mathcal{X}(\mathcal{X}'\mathbf{L}_{\mathbf{D}}\mathcal{X})^{-1}(\mathbf{0} - \mathcal{X}'\mathbf{w})$. Then, in view of (2.8), in order to show that $\hat{\mathcal{B}} = \hat{\mathcal{B}}^{o}$ it suffices to show that $\mathbf{L}_{\mathbf{D}} = \mathbf{\Lambda}^{0}$. For STRSRS it is easy to show that $\mathbf{\Lambda}^{0} = \operatorname{diag}\{\lambda_{ih}(\mathbf{I} - \mathbf{P}_{\mathbf{1}ih})\}$, where $\lambda_{ih} = N_{ih}^{2}(1 - f_{ih})/[n_{ih}(n_{ih} - 1)]$ and $\mathbf{P}_{\mathbf{1}ih} = \mathbf{1}_{ih}(\mathbf{1}_{ih}'\mathbf{1}_{ih})^{-1}\mathbf{1}_{ih}'$. Next, observe that the matrix $\mathbf{P}_{\mathbf{D}}$ is diagonal with ih^{th} entry $\mathbf{1}_{ih}(\mathbf{1}_{ih}'\mathbf{\Lambda}_{ih}\mathbf{1}_{ih})^{-1}\mathbf{1}_{ih}'\mathbf{\Lambda}_{ih} = \mathbf{P}_{\mathbf{1}ih}$, because the elements of $\mathbf{\Lambda}_{ih}$ are constant. Since this constant element is $w_{ik}/q_{ik} = (N_{ih}/n_{ih})[N_{ih}(1 - f_{ih})/(n_{ih} - 1)] = \lambda_{ih}$, we get $\mathbf{L}_{\mathbf{D}} = \operatorname{diag}\{\mathbf{\Lambda}_{ih}(\mathbf{I} - \mathbf{P}_{\mathbf{1}ih})\} = \mathbf{\Lambda}^{0}$, o.e.d.

(b) For Poisson sampling, $\mathbf{\Lambda}_{i}^{0} = \operatorname{diag}\{(1 - \pi_{ihk})/\pi_{ihk}^{2}\}, h = 1, \ldots, H_{i}$. The proof follows immediately upon observing that with the specified constants q_{ik} in the entries of $\mathbf{\Lambda}_{i}$ we have $\mathbf{\Lambda}_{i} = \mathbf{\Lambda}_{i}^{0}$.

(a’) For simplicity drop the stratum subscript. Simple random subsampling is done sequentially with fixed sizes n_1, n_2 and n_3. It can be shown that the first-and-second order marginal inclusion probabilities for S_i are $\pi_{ik} = n_i / N$ and $\pi_{ikl} = n_i(n_i - 1)/[N(N-1)]$, as if S_i was drawn directly from U. A combinatorial argument shows that the conditional (given S) second-order inclusion probability for S_i and S_j is $\pi_{ikjl|S} = n_i n_j / [n(n-1)]$ and thus the marginal inclusion probability is $\pi_{ikjl} = n_i n_j / [N(N-1)]$. For $k = l, \pi_{ikjk} = 0$. Then $\Delta_{kl} = \pi_{ikjl} - \pi_{ik}\pi_{jl} = n_i n_j / [N^2(N-1)]$ and $\Delta_{kk} = -n_i n_j / N^2$. Thus $\Delta_{kl} \approx 0$, for $k, l \in U$ when the sampling fractions are small, and then $\mathbf{\Lambda}^0 \approx \operatorname{diag}\{\mathbf{\Lambda}_i^0\}$. Optimality of the CGR then follows from Theorem 1 (a).

(*b*’) Randomly assigning the units of S to three subsamples, with fixed expected subsample size, implies that inclusion of the units is done independently within and between the subsamples. Since in Poisson sampling the units of U are also included in S independently, $\Delta_{kl} = \pi_{ikjl} - \pi_{ik}\pi_{jl} = 0$ and $\Delta_{kk} = -\pi_{ik}\pi_{jl}$. Δ_{kk} is approximately zero for small sampling fractions, and then $\mathbf{\Lambda}^0 \approx \operatorname{diag}\{\mathbf{\Lambda}_i^0\}$. Optimality of the CGR follows then from Theorem 1(*b*).

Proof of Theorem 2

We start with the expression of the CGR estimator. By Lemma 1, with partitioned design matrix $(\mathcal{X}, \mathbf{Z})$ and $\mathbf{R} = \mathbf{\Lambda}$, the calibrated weight vector $\mathbf{c}$ can be written as $\mathbf{c} = \mathbf{c}_\mathbf{Z} + \mathbf{L}_\mathbf{Z}\mathcal{X}(\mathcal{X}'\mathbf{L}_\mathbf{Z}\mathcal{X})^{-1}(\mathbf{0} - \mathcal{X}'\mathbf{c}_\mathbf{Z})$, where $\mathbf{c}_\mathbf{Z} = \mathbf{w} + \mathbf{\Lambda Z}(\mathbf{Z}'\mathbf{\Lambda Z})^{-1}(\mathbf{t}_{(\mathbf{z})} - \mathbf{Z}'\mathbf{w})$ and $\mathbf{L}_\mathbf{Z} = \mathbf{\Lambda}(\mathbf{I} - \mathbf{P}_\mathbf{Z})$. Then $\hat{\mathcal{X}}_3^{\mathrm{GR}} = \mathcal{X}_3'\mathbf{c}_\mathbf{Z} = \hat{\mathcal{X}}_3 + \mathcal{X}_3'\mathbf{\Lambda Z}(\mathbf{Z}'\mathbf{\Lambda Z})^{-1}(\mathbf{t}_{(\mathbf{z})} - \hat{\mathbf{Z}})$ and $\hat{\mathcal{X}}^{\mathrm{GR}} = \hat{\mathcal{X}} + \mathcal{X}'\mathbf{\Lambda Z}(\mathbf{Z}'\mathbf{\Lambda Z})^{-1}(\mathbf{t}_{(\mathbf{z})} - \hat{\mathbf{Z}})$. It follows that the CGR estimator is given by $\mathcal{X}_3'\mathbf{c} = \hat{\mathcal{X}}_3^{\mathrm{GR}} - \hat{\mathcal{B}}\hat{\mathcal{X}}^{\mathrm{GR}}$, where $\hat{\mathcal{B}} = [\mathcal{X}_3'\mathbf{\Lambda}(\mathbf{I} - \mathbf{P}_\mathbf{Z})\mathcal{X}][\mathcal{X}'\mathbf{\Lambda}(\mathbf{I} - \mathbf{P}_\mathbf{Z})\mathcal{X}]^{-1}$.

(a) Since $\mathbf{P}_\mathbf{Z} = \operatorname{diag}\{\mathbf{P}_{\mathbf{Z}_i}\}$ and, for SRS, $\mathbf{\Lambda}^0 = \operatorname{diag}\{\lambda_i(\mathbf{I} - \mathbf{P}_{\mathbf{1}i})\}$, where $\lambda_i = N^2(1 - f_i)/[n_i(n_i - 1)]$ and $\mathbf{P}_{\mathbf{1}i} = \mathbf{1}_i(\mathbf{1}_i'\mathbf{1}_i)^{-1}\mathbf{1}_i'$, we have $\mathbf{\Lambda}^0(\mathbf{I} - \mathbf{P}_\mathbf{Z}) = \operatorname{diag}\{\lambda_i(\mathbf{I} - \mathbf{P}_{\mathbf{1}i})(\mathbf{I} - \mathbf{P}_{\mathbf{Z}_i})\}$. Now, by assumption $\mathbf{1} = \mathbf{Z}_i\mathbf{h}_i$, so that $\mathbf{1}'\mathbf{P}_{\mathbf{Z}_i} = \mathbf{1}'$ and hence $\mathbf{P}_{\mathbf{1}i}(\mathbf{I} - \mathbf{P}_{\mathbf{Z}_i}) = \mathbf{0}$. It follows that $\mathbf{\Lambda}^0(\mathbf{I} - \mathbf{P}_\mathbf{Z}) = \operatorname{diag}\{\lambda_i(\mathbf{I} - \mathbf{P}_{\mathbf{Z}_i})\}$ and, since the matrices $\mathbf{I} - \mathbf{P}_{\mathbf{Z}_i}$ are idempotent, $(\mathbf{I} - \mathbf{P}_\mathbf{Z})'\mathbf{\Lambda}^0(\mathbf{I} - \mathbf{P}_\mathbf{Z}) = \operatorname{diag}\{\lambda_i(\mathbf{I} - \mathbf{P}_{\mathbf{Z}_i})\}$. But $\lambda_i = w_{ik}/q_{ik}$, where $w_{ik} = N/n_i$ and q_{ik} are the specified constants in the entries of $\mathbf{\Lambda}_i$. It follows that $(\mathbf{I} - \mathbf{P}_\mathbf{Z})'\mathbf{\Lambda}^0(\mathbf{I} - \mathbf{P}_\mathbf{Z}) = \operatorname{diag}\{\mathbf{\Lambda}_i(\mathbf{I} - \mathbf{P}_{\mathbf{Z}_i})\} = \mathbf{\Lambda}(\mathbf{I} - \mathbf{P}_\mathbf{Z})$ and thus $\hat{\mathcal{B}} = \hat{\mathcal{B}}^{wo}$, so that $\hat{\mathcal{X}}_3^{\mathrm{GR}} - \hat{\mathcal{B}}\hat{\mathcal{X}}^{\mathrm{GR}} = \hat{\mathcal{X}}_3^{\mathrm{GR}} - \hat{\mathcal{B}}^{wo}\hat{\mathcal{X}}^{\mathrm{GR}}$.

(b) By Lemma 1, with the partitioned design matrix $\mathcal{Z} = (\mathcal{X}, \mathbf{Z}, \mathbf{D})$ and vector of totals $\mathbf{t}_\mathcal{Z} = (\mathbf{0}', \mathbf{t}_{(\mathbf{z})}', \mathbf{N}')'$, the vector of calibrated weights $\mathbf{c} = \mathbf{w} + \mathbf{\Lambda}\mathcal{Z}(\mathcal{Z}'\mathbf{\Lambda}\mathcal{Z})^{-1}(\mathbf{t}_\mathcal{Z} - \mathcal{Z}'\mathbf{w})$ can be written as $\mathbf{c} = \mathbf{c}_\mathbf{D} + \mathbf{L}_\mathbf{D}(\mathcal{X}, \mathbf{Z})\left[(\mathcal{X}, \mathbf{Z})'\mathbf{L}_\mathbf{D}(\mathcal{X}, \mathbf{Z})\right]^{-1}\left[(\mathbf{0}', \mathbf{t}_{(\mathbf{z})}')' - (\mathcal{X}, \mathbf{Z})'\mathbf{c}_\mathbf{D}\right]$, where $\mathbf{c}_\mathbf{D} = \mathbf{w} + \mathbf{\Lambda D}(\mathbf{D}'\mathbf{\Lambda D})^{-1}(\mathbf{N} - \mathbf{D}'\mathbf{w})$ and $\mathbf{L}_\mathbf{D} = \mathbf{\Lambda}(\mathbf{I} - \mathbf{P}_\mathbf{D})$, with $\mathbf{P}_\mathbf{D} = \mathbf{D}(\mathbf{D}'\mathbf{\Lambda D})^{-1}\mathbf{D}'\mathbf{\Lambda}$. But, as shown in the proof of Theorem 1(a), $\mathbf{c}_\mathbf{D} = \mathbf{w}$ and $\mathbf{L}_\mathbf{D} = \mathbf{\Lambda}^0$. Thus, $\mathbf{c} = \mathbf{w} + \mathbf{\Lambda}^0(\mathcal{X}, \mathbf{Z})\left[(\mathcal{X}, \mathbf{Z})'\mathbf{\Lambda}^0(\mathcal{X}, \mathbf{Z})\right]^{-1}\left[(\mathbf{0}', \mathbf{t}_{(\mathbf{z})}')' - (\mathcal{X}, \mathbf{Z})'\mathbf{w}\right]$. Next, by applying again Lemma 1, now with $\mathbf{R} = \mathbf{\Lambda}^0$ and design matrix $(\mathcal{X}, \mathbf{Z})$, we get $\mathbf{c} = \mathbf{c}_\mathbf{Z} + \mathbf{L}_\mathbf{Z}^0\mathcal{X}(\mathcal{X}'\mathbf{L}_\mathbf{Z}^0\mathcal{X})^{-1}(\mathbf{0} - \mathcal{X}'\mathbf{c}_\mathbf{Z})$, where $\mathbf{c}_\mathbf{Z} = \mathbf{w} + \mathbf{\Lambda}^0\mathbf{Z}(\mathbf{Z}'\mathbf{\Lambda}^0\mathbf{Z})^{-1}$

$\left(\mathbf{t}_{(z)} - \mathbf{Z}'\mathbf{w}\right)$ and $\mathbf{L}_{\mathbf{Z}}^{0} = \boldsymbol{\Lambda}^{0}\left(\mathbf{I} - \mathbf{P}_{\mathbf{Z}}^{0}\right)$. Then it follows that the CGR estimator is $\mathcal{X}_3'\mathbf{c} = \mathcal{X}_3'\mathbf{c}_{\mathbf{Z}} - \mathcal{X}_3'\mathbf{L}_{\mathbf{Z}}^{0}\mathcal{X}\left(\mathcal{X}'\mathbf{L}_{\mathbf{Z}}^{0}\mathcal{X}\right)^{-1}\mathcal{X}'\mathbf{c}_{\mathbf{Z}} = \hat{\mathcal{X}}_3^{\mathrm{OR}} - \hat{\mathcal{B}}^{o}\hat{\mathcal{X}}^{\mathrm{OR}}$, in obvious expressions for $\hat{\mathcal{X}}_3^{\mathrm{OR}}, \hat{\mathcal{X}}^{\mathrm{OR}}$ and $\hat{\mathcal{B}}^{o}$.

(c) It was shown in the proof of Theorem 1 that $\boldsymbol{\Lambda} = \boldsymbol{\Lambda}^{0}$. Clearly then it holds that $\hat{\mathcal{X}}_3^{\mathrm{GR}} = \hat{\mathcal{X}}_3^{\mathrm{OR}}$, $\hat{\mathcal{X}}^{\mathrm{GR}} = \hat{\mathcal{X}}^{\mathrm{OR}}$ and $\hat{\mathcal{B}} = \hat{\mathcal{B}}^{o}$, and thus $\hat{\mathcal{X}}_3^{\mathrm{GR}} - \hat{\mathcal{B}}\hat{\mathcal{X}}^{\mathrm{GR}} = \hat{\mathcal{X}}_3^{\mathrm{OR}} - \hat{\mathcal{B}}^{o}\hat{\mathcal{X}}^{\mathrm{OR}}$.

Proof of Proposition 1

All matrices appearing in this proof are defined at the population level. Partitioning the matrix $\mathcal{X}$ in (4.4) as $(\mathsf{Z}, \boldsymbol{\Psi})$, where Z consists of the second and fourth columns, and $\boldsymbol{\Psi}$ of the rest, and applying Lemma 1 with $\mathbf{R} = \boldsymbol{\Lambda}^{0} = \{(\pi_{kl} - \pi_k\pi_l)/\pi_k\pi_l\}$, we obtain the vector of calibrated weights decomposed as

$$\mathbf{c} = \mathbf{w} + \mathbf{L}_{\boldsymbol{\Psi}}^{0}\mathsf{Z}\left(\mathsf{Z}'\mathbf{L}_{\boldsymbol{\Psi}}^{0}\mathsf{Z}\right)^{-1}[\mathbf{0} - \mathsf{Z}'\mathbf{w}] + \mathbf{L}_{\mathsf{Z}}^{0}\boldsymbol{\Psi}\left(\boldsymbol{\Psi}'\mathbf{L}_{\mathsf{Z}}^{0}\boldsymbol{\Psi}\right)^{-1}[\mathbf{0} - \boldsymbol{\Psi}'\mathbf{w}],$$

where $\mathbf{L}_{\mathsf{Z}}^{0} = \boldsymbol{\Lambda}^{0}\left(\mathbf{I} - \mathbf{P}_{\mathsf{Z}}^{0}\right)$ with $\mathbf{P}_{\mathsf{Z}}^{0} = \mathsf{Z}\left(\mathsf{Z}'\boldsymbol{\Lambda}^{0}\mathsf{Z}\right)^{-1}\mathsf{Z}'\boldsymbol{\Lambda}^{0}$. The estimator $\hat{\mathbf{Z}}^{B}$ in (4.2) is obtained as $\mathbf{Z}_{3-}'\mathbf{c}$, where $\mathbf{Z}_{3-} = (\mathbf{0}', \mathbf{0}', \mathbf{Z}_3')'$. The last two terms of (4.2) are consolidated in the term $\mathbf{Z}_{3-}'\mathbf{L}_{\mathsf{Z}}^{0}\boldsymbol{\Psi}\left(\boldsymbol{\Psi}'\mathbf{L}_{\mathsf{Z}}^{0}\boldsymbol{\Psi}\right)^{-1}[\mathbf{0} - \boldsymbol{\Psi}'\mathbf{w}]$. These two terms vanish only if $\mathbf{Z}_{3-}'\mathbf{L}_{\mathsf{Z}}^{0}\boldsymbol{\Psi}(= \mathbf{Z}_{3-}'\boldsymbol{\Lambda}^{0}\boldsymbol{\Psi} - \mathbf{Z}_{3-}'\boldsymbol{\Lambda}^{0}\mathsf{Z}(\mathsf{Z}'\boldsymbol{\Lambda}^{0}\mathsf{Z})^{-1}\mathsf{Z}'\boldsymbol{\Lambda}^{0}\boldsymbol{\Psi}) = \mathbf{0}$. First, we easily get $\mathbf{Z}_{3-}'\boldsymbol{\Lambda}^{0}\boldsymbol{\Psi} = \left(\mathbf{Z}_3'\boldsymbol{\Lambda}_3^{0}\mathbf{X}_3, \mathbf{Z}_3'\boldsymbol{\Lambda}_3^{0}\mathbf{Y}_3\right)$ and $\mathbf{Z}_{3-}'\boldsymbol{\Lambda}^{0}\mathsf{Z} = \mathbf{Z}_3'\boldsymbol{\Lambda}_3^{0}\mathbf{Z}_3(\mathbf{I}, \mathbf{I})$, as well as

$$\mathsf{Z}'\boldsymbol{\Lambda}^{0}\boldsymbol{\Psi} = \begin{pmatrix} \mathbf{Z}_1'\boldsymbol{\Lambda}_1^{0}\mathbf{X}_1 + \mathbf{Z}_3'\boldsymbol{\Lambda}_3^{0}\mathbf{X}_3 & \mathbf{Z}_3'\boldsymbol{\Lambda}_3^{0}\mathbf{Y}_3 \\ \mathbf{Z}_3'\boldsymbol{\Lambda}_3^{0}\mathbf{X}_3 & \mathbf{Z}_2'\boldsymbol{\Lambda}_2^{0}\mathbf{Y}_2 + \mathbf{Z}_3'\boldsymbol{\Lambda}_3^{0}\mathbf{Y}_3 \end{pmatrix},$$

and

$$\mathsf{Z}'\boldsymbol{\Lambda}^{0}\mathsf{Z} = \begin{pmatrix} \mathbf{Z}_1'\boldsymbol{\Lambda}_1^{0}\mathbf{Z}_1 + \mathbf{Z}_3'\boldsymbol{\Lambda}_3^{0}\mathbf{Z}_3 & \mathbf{Z}_3'\boldsymbol{\Lambda}_3^{0}\mathbf{Z}_3 \\ \mathbf{Z}_3'\boldsymbol{\Lambda}_3^{0}\mathbf{Z}_3 & \mathbf{Z}_2'\boldsymbol{\Lambda}_2^{0}\mathbf{Z}_2 + \mathbf{Z}_3'\boldsymbol{\Lambda}_3^{0}\mathbf{Z}_3 \end{pmatrix}.$$

Next we write

$$(\mathsf{Z}'\boldsymbol{\Lambda}^{0}\mathsf{Z})^{-1} = \begin{pmatrix} \mathbf{A} & \mathbf{B} \\ \mathbf{B}' & \mathbf{D} \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{A}^{-1} + \mathbf{F}\mathbf{E}^{-1}\mathbf{F}' & -\mathbf{F}\mathbf{E}^{-1} \\ -\mathbf{E}^{-1}\mathbf{F}' & \mathbf{E}^{-1} \end{pmatrix},$$

where $\mathbf{E} = \mathbf{D} - \mathbf{B}'\mathbf{A}^{-1}\mathbf{B}$ and $\mathbf{F} = \mathbf{A}^{-1}\mathbf{B}$. It follows then that $\mathbf{Z}_{3-}'\boldsymbol{\Lambda}^{0}\mathsf{Z}(\mathsf{Z}'\boldsymbol{\Lambda}^{0}\mathsf{Z})^{-1} = (\mathbf{B}\mathbf{A}^{-1} + \mathbf{B}\mathbf{F}\mathbf{E}^{-1}\mathbf{F}' - \mathbf{B}\mathbf{E}^{-1}\mathbf{F}', \mathbf{B}(\mathbf{I} - \mathbf{F})\mathbf{E}^{-1}) = ((\mathbf{D} - \mathbf{B})\mathbf{E}^{-1}\mathbf{F}', \mathbf{B}(\mathbf{I} - \mathbf{F})\mathbf{E}^{-1})$. Using the analytic expressions $\mathbf{B} = \mathbf{Z}_3'\boldsymbol{\Lambda}_3^{0}\mathbf{Z}_3$, $\mathbf{D} = \mathbf{Z}_2'\boldsymbol{\Lambda}_2^{0}\mathbf{Z}_2 + \mathbf{Z}_3'\boldsymbol{\Lambda}_3^{0}\mathbf{Z}_3$, $\mathbf{F} = \left(\mathbf{Z}_1'\boldsymbol{\Lambda}_1^{0}\mathbf{Z}_1 + \mathbf{Z}_3'\boldsymbol{\Lambda}_3^{0}\mathbf{Z}_3\right)^{-1}\mathbf{Z}_3'\boldsymbol{\Lambda}_3^{0}\mathbf{Z}_3$ and $\mathbf{E} = \mathbf{Z}_2'\boldsymbol{\Lambda}_2^{0}\mathbf{Z}_2 + \mathbf{Z}_1'\boldsymbol{\Lambda}_1^{0}\mathbf{Z}_1\mathbf{F}$, we obtain after some algebra

$$\mathbf{Z}_{3-}'\boldsymbol{\Lambda}^{0}\mathsf{Z}(\mathsf{Z}'\boldsymbol{\Lambda}^{0}\mathsf{Z})^{-1} = \mathbf{K}^{-1}\left[\left(\mathbf{Z}_1'\boldsymbol{\Lambda}_1^{0}\mathbf{Z}_1\right)^{-1}, \left(\mathbf{Z}_2'\boldsymbol{\Lambda}_2^{0}\mathbf{Z}_2\right)^{-1}\right],$$

where $\mathbf{K} = \left(\mathbf{Z}_1'\boldsymbol{\Lambda}_1^{0}\mathbf{Z}_1\right)^{-1} + \left(\mathbf{Z}_2'\boldsymbol{\Lambda}_2^{0}\mathbf{Z}_2\right)^{-1} + \left(\mathbf{Z}_3'\boldsymbol{\Lambda}_3^{0}\mathbf{Z}_3\right)^{-1}$. We can now obtain without much difficulty

$$\begin{aligned} \mathbf{Z}_{3-}'\mathbf{L}_{\mathsf{Z}}^{0}\boldsymbol{\Psi} &= \mathbf{Z}_{3-}'\boldsymbol{\Lambda}^{0}\boldsymbol{\Psi} - \mathbf{Z}_{3-}'\boldsymbol{\Lambda}^{0}\mathsf{Z}(\mathsf{Z}'\boldsymbol{\Lambda}^{0}\mathsf{Z})^{-1}\mathsf{Z}'\boldsymbol{\Lambda}^{0}\boldsymbol{\Psi} \\ &= \mathbf{K}^{-1}\Big[\left(\mathbf{Z}_3'\boldsymbol{\Lambda}_3^{0}\mathbf{Z}_3\right)^{-1}\mathbf{Z}_3'\boldsymbol{\Lambda}_3^{0}\mathbf{X}_3 - \left(\mathbf{Z}_1'\boldsymbol{\Lambda}_1^{0}\mathbf{Z}_1\right)^{-1}\mathbf{Z}_1'\boldsymbol{\Lambda}_1^{0}\mathbf{X}_1, \\ &\qquad \left(\mathbf{Z}_3'\boldsymbol{\Lambda}_3^{0}\mathbf{Z}_3\right)^{-1}\mathbf{Z}_3'\boldsymbol{\Lambda}_3^{0}\mathbf{Y}_3 - \left(\mathbf{Z}_2'\boldsymbol{\Lambda}_2^{0}\mathbf{Z}_2\right)^{-1}\mathbf{Z}_2'\boldsymbol{\Lambda}_2^{0}\mathbf{Y}_2\Big]. \end{aligned}$$

It follows that $\mathbf{Z}_{3-}'\mathbf{L}_{\mathbf{Z}}^{0}\boldsymbol{\Psi}=(\mathbf{0},\mathbf{0})$ only if $\left(\mathbf{Z}_3'\boldsymbol{\Lambda}_3^0\mathbf{Z}_3\right)^{-1}\mathbf{Z}_3'\boldsymbol{\Lambda}_3^0\mathbf{X}_3=\left(\mathbf{Z}_1'\boldsymbol{\Lambda}_1^0\mathbf{Z}_1\right)^{-1}\mathbf{Z}_1'\boldsymbol{\Lambda}_1^0\mathbf{X}_1$ and $\left(\mathbf{Z}_3'\boldsymbol{\Lambda}_3^0\mathbf{Z}_3\right)^{-1}$ $\mathbf{Z}_3'\boldsymbol{\Lambda}_3^0\mathbf{Y}_3=\left(\mathbf{Z}_2'\boldsymbol{\Lambda}_2^0\mathbf{Z}_2\right)^{-1}\mathbf{Z}_2'\boldsymbol{\Lambda}_2^0\mathbf{Y}_2$. But these two equations are identical to the equations in (4.6). Since all the matrices in $\left(\mathbf{Z}_i'\boldsymbol{\Lambda}_i^0\mathbf{Z}_i\right)^{-1}\mathbf{Z}_i'\boldsymbol{\Lambda}_i^0\mathbf{X}_i$ are defined at the population level, with the subscript $i=1,3$ indicating survey, this quantity is constant across surveys only if the design-specific matrix $\boldsymbol{\Lambda}_i^0$ is constant, or if $\boldsymbol{\Lambda}_i^0$ differs among surveys by a constant multiple (depending on the sample size). This holds true also for $\left(\mathbf{Z}_i'\boldsymbol{\Lambda}_i^0\mathbf{Z}_i\right)^{-1}\mathbf{Z}_i'\boldsymbol{\Lambda}_i^0\mathbf{Y}_i$, $i=2,3$. This completes the proof.

Proof of Proposition 2

Under the sampling scheme (a) of Theorem 1, composite calibration at population level with design matrix $\boldsymbol{\mathcal{Z}}=(\boldsymbol{\mathcal{X}},\mathbf{D})$ and vector of totals $\mathbf{t}_{\mathcal{Z}}=(\mathbf{0}',\mathbf{N}')'$ produces the joint CGR domain estimator of $\left(\mathbf{t}_{\mathbf{x}d}',\mathbf{t}_{\mathbf{y}d}'\right)'$ based on the weights of S_3 and written in the form $\hat{\boldsymbol{\mathcal{X}}}_{3d}^{\mathrm{CGR}}=\hat{\boldsymbol{\mathcal{X}}}_{3d}+\hat{\boldsymbol{\mathcal{B}}}_d\left(\mathbf{t}_{\mathcal{Z}}-\hat{\boldsymbol{\mathcal{Z}}}\right)$, where $\hat{\boldsymbol{\mathcal{B}}}_d=\boldsymbol{\mathcal{X}}_{3d}'\boldsymbol{\Lambda}\boldsymbol{\mathcal{Z}}(\boldsymbol{\mathcal{Z}}'\boldsymbol{\Lambda}\boldsymbol{\mathcal{Z}})^{-1}$. The associated matrix of regression residuals is $\boldsymbol{\mathcal{X}}_{3d}-\boldsymbol{\mathcal{Z}}\hat{\boldsymbol{\mathcal{B}}}_d'$, alternatively written as $(\mathbf{I}-\mathbf{P}_{\mathcal{Z}})\boldsymbol{\mathcal{X}}_{3d}$, with $\mathbf{P}_{\mathcal{Z}}=\boldsymbol{\mathcal{Z}}(\boldsymbol{\mathcal{Z}}'\boldsymbol{\Lambda}\boldsymbol{\mathcal{Z}})^{-1}\boldsymbol{\mathcal{Z}}'\boldsymbol{\Lambda}$. Then $\widehat{\mathrm{AV}}\left(\hat{\boldsymbol{\mathcal{X}}}_{3d}^{\mathrm{CGR}}\right)=\boldsymbol{\mathcal{X}}_{3d}'(\mathbf{I}-\mathbf{P}_{\mathcal{Z}})'$ $\boldsymbol{\Lambda}^0(\mathbf{I}-\mathbf{P}_{\mathcal{Z}})\boldsymbol{\mathcal{X}}_{3d}$. Next recall from the proof of Theorem 1 that $\boldsymbol{\Lambda}^0=\boldsymbol{\Lambda}(\mathbf{I}-\mathbf{P}_{\mathbf{D}})$, with $\mathbf{P}_{\mathbf{D}}=\mathbf{D}(\mathbf{D}'\boldsymbol{\Lambda}\mathbf{D})^{-1}\mathbf{D}'\boldsymbol{\Lambda}$, and notice that $\mathbf{D}=\boldsymbol{\mathcal{Z}}\mathbf{H}$ for a suitable constant matrix $\mathbf{H}$. It is easy to verify that $\mathbf{P}_{\mathbf{D}}\mathbf{P}_{\mathcal{Z}}=\mathbf{P}_{\mathbf{D}}$. It follows then that $\boldsymbol{\Lambda}^0(\mathbf{I}-\mathbf{P}_{\mathcal{Z}})=\boldsymbol{\Lambda}(\mathbf{I}-\mathbf{P}_{\mathcal{Z}})$ and $(\mathbf{I}-\mathbf{P}_{\mathcal{Z}})'\boldsymbol{\Lambda}^0(\mathbf{I}-\mathbf{P}_{\mathcal{Z}})=\boldsymbol{\Lambda}(\mathbf{I}-\mathbf{P}_{\mathcal{Z}})$. Thus $\widehat{\mathrm{AV}}\left(\hat{\boldsymbol{\mathcal{X}}}_{3d}^{\mathrm{CGR}}\right)=\boldsymbol{\mathcal{X}}_{3d}'\boldsymbol{\Lambda}(\mathbf{I}-\mathbf{P}_{\mathcal{Z}})\boldsymbol{\mathcal{X}}_{3d}$. Now, composite calibration at domain level involves the design matrix $\boldsymbol{\mathcal{Z}}_d=(\boldsymbol{\mathcal{X}}_d,\mathbf{D})$; no need to restrict $\mathbf{D}$ to the domain U_d. The resulting CGR estimator is $\breve{\boldsymbol{\mathcal{X}}}_{3d}^{\mathrm{CGR}}=\hat{\boldsymbol{\mathcal{X}}}_{3d}+\breve{\boldsymbol{\mathcal{B}}}_d\left(\mathbf{t}_{\mathcal{Z}_d}-\hat{\boldsymbol{\mathcal{Z}}}_d\right)$ where $\breve{\boldsymbol{\mathcal{B}}}_d=\boldsymbol{\mathcal{X}}_{3d}'\boldsymbol{\Lambda}\boldsymbol{\mathcal{Z}}_d(\boldsymbol{\mathcal{Z}}_d'\boldsymbol{\Lambda}\boldsymbol{\mathcal{Z}}_d)^{-1}$. As with $\hat{\boldsymbol{\mathcal{X}}}_{3d}^{\mathrm{CGR}}$ above, it can be shown that $\widehat{\mathrm{AV}}\left(\breve{\boldsymbol{\mathcal{X}}}_{3d}^{\mathrm{CGR}}\right)=\boldsymbol{\mathcal{X}}_{3d}'\boldsymbol{\Lambda}\left(\mathbf{I}-\mathbf{P}_{\mathcal{Z}_d}\right)\boldsymbol{\mathcal{X}}_{3d}$, where $\mathbf{P}_{\mathcal{Z}_d}=\boldsymbol{\mathcal{Z}}_d(\boldsymbol{\mathcal{Z}}_d'\boldsymbol{\Lambda}\boldsymbol{\mathcal{Z}})_d^{-1}\boldsymbol{\mathcal{Z}}_d'\boldsymbol{\Lambda}$. Then $\widehat{\mathrm{AV}}\left(\hat{\boldsymbol{\mathcal{X}}}_{3d}^{\mathrm{CGR}}\right)-$ $\widehat{\mathrm{AV}}\left(\breve{\boldsymbol{\mathcal{X}}}_{3d}^{\mathrm{CGR}}\right)=\boldsymbol{\mathcal{X}}_{3d}'\boldsymbol{\Lambda}\left(\mathbf{P}_{\mathcal{Z}_d}-\mathbf{P}_{\mathcal{Z}}\right)\boldsymbol{\mathcal{X}}_{3d}$. Noticing that $\boldsymbol{\mathcal{X}}_{3d}'\boldsymbol{\Lambda}\boldsymbol{\mathcal{Z}}=\boldsymbol{\mathcal{X}}_{3d}'\boldsymbol{\Lambda}\boldsymbol{\mathcal{Z}}_d$, we can write $\mathbf{P}_{\mathcal{Z}}=$ $\boldsymbol{\mathcal{Z}}_d(\boldsymbol{\mathcal{Z}}'\boldsymbol{\Lambda}\boldsymbol{\mathcal{Z}})^{-1}\boldsymbol{\mathcal{Z}}_d'\boldsymbol{\Lambda}$. It is trivial then to show that $\left(\mathbf{P}_{\mathcal{Z}_d}-\mathbf{P}_{\mathcal{Z}}\right)=\left(\mathbf{P}_{\mathcal{Z}_d}-\mathbf{P}_{\mathcal{Z}}\right)^2$, and since the matrix $\boldsymbol{\Lambda}$ is diagonal with positive entries, it follows that $\boldsymbol{\mathcal{X}}_{3d}'\boldsymbol{\Lambda}\left(\mathbf{P}_{\mathcal{Z}_d}-\mathbf{P}_{\mathcal{Z}}\right)\boldsymbol{\mathcal{X}}_{3d}>\mathbf{0}$ and hence $\widehat{\mathrm{AV}}\left(\breve{\boldsymbol{\mathcal{X}}}_{3d}^{\mathrm{CGR}}\right)<\widehat{\mathrm{AV}}\left(\hat{\boldsymbol{\mathcal{X}}}_{3d}^{\mathrm{CGR}}\right)$.

Under the conditions of part (b), $\boldsymbol{\Lambda}=\boldsymbol{\Lambda}^0$ and the CGR domain estimator is identical to the COR domain estimator $\hat{\boldsymbol{\mathcal{X}}}_{3d}^{\mathrm{COR}}=\hat{\boldsymbol{\mathcal{X}}}_{3d}-\hat{\boldsymbol{\mathcal{B}}}_d^0\hat{\boldsymbol{\mathcal{X}}}$, where $\hat{\boldsymbol{\mathcal{B}}}_d^0=\boldsymbol{\mathcal{X}}_{3d}'\boldsymbol{\Lambda}^0\boldsymbol{\mathcal{X}}(\boldsymbol{\mathcal{X}}'\boldsymbol{\Lambda}^0\boldsymbol{\mathcal{X}})^{-1}$. The associated matrix of regression residuals is $(\mathbf{I}-\mathbf{P}_{\mathcal{X}})\boldsymbol{\mathcal{X}}_{3d}$, with $\mathbf{P}_{\mathcal{X}}=\boldsymbol{\mathcal{X}}(\boldsymbol{\mathcal{X}}'\boldsymbol{\Lambda}^0\boldsymbol{\mathcal{X}})^{-1}\boldsymbol{\mathcal{X}}'\boldsymbol{\Lambda}^0$. Then $\widehat{\mathrm{AV}}\left(\hat{\boldsymbol{\mathcal{X}}}_{3d}^{\mathrm{COR}}\right)=$ $\boldsymbol{\mathcal{X}}_{3d}'(\mathbf{I}-\mathbf{P}_{\mathcal{X}})'\boldsymbol{\Lambda}^0(\mathbf{I}-\mathbf{P}_{\mathcal{X}})\boldsymbol{\mathcal{X}}_{3d}=\boldsymbol{\mathcal{X}}_{3d}'\boldsymbol{\Lambda}^0(\mathbf{I}-\mathbf{P}_{\mathcal{X}})\boldsymbol{\mathcal{X}}_{3d}$. On the other hand, for the estimator $\breve{\boldsymbol{\mathcal{X}}}_{3d}^{\mathrm{COR}}=\breve{\boldsymbol{\mathcal{X}}}_{3d}-\breve{\boldsymbol{\mathcal{B}}}_d^0\hat{\boldsymbol{\mathcal{X}}}$, where $\hat{\boldsymbol{\mathcal{B}}}_d^0=\boldsymbol{\mathcal{X}}_{3d}'\boldsymbol{\Lambda}^0\boldsymbol{\mathcal{X}}_d\left(\boldsymbol{\mathcal{X}}_d'\boldsymbol{\Lambda}^0\boldsymbol{\mathcal{X}}_d\right)^{-1}$ we have $\widehat{\mathrm{AV}}\left(\breve{\boldsymbol{\mathcal{X}}}_{3d}^{\mathrm{COR}}\right)=\boldsymbol{\mathcal{X}}_{3d}'\boldsymbol{\Lambda}^0$ $\left(\mathbf{I}-\mathbf{P}_{\mathcal{X}_d}\right)\boldsymbol{\mathcal{X}}_{3d}$, with $\mathbf{P}_{\mathcal{X}_d}=\boldsymbol{\mathcal{X}}_d\left(\boldsymbol{\mathcal{X}}_d'\boldsymbol{\Lambda}^0\boldsymbol{\mathcal{X}}_d\right)^{-1}\boldsymbol{\mathcal{X}}_d'\boldsymbol{\Lambda}^0$. Then $\widehat{\mathrm{AV}}(\hat{\boldsymbol{\mathcal{X}}}_{3d}^{\mathrm{COR}})-\widehat{\mathrm{AV}}(\breve{\boldsymbol{\mathcal{X}}}_{3d}^{\mathrm{COR}})=\boldsymbol{\mathcal{X}}_{3d}'\boldsymbol{\Lambda}^0(\mathbf{P}_{\mathcal{X}_d}-$ $\mathbf{P}_{\mathcal{X}})\boldsymbol{\mathcal{X}}_{3d}$. Notice that $\boldsymbol{\mathcal{X}}_{3d}'\boldsymbol{\Lambda}^0\boldsymbol{\mathcal{X}}_d=\boldsymbol{\mathcal{X}}_{3d}'\boldsymbol{\Lambda}^0\boldsymbol{\mathcal{X}}_{3d}$ and since $\boldsymbol{\Lambda}^0$ is diagonal $\boldsymbol{\mathcal{X}}_{3d}'\boldsymbol{\Lambda}^0\boldsymbol{\mathcal{X}}=\boldsymbol{\mathcal{X}}_{3d}'\boldsymbol{\Lambda}^0\boldsymbol{\mathcal{X}}_{3d}$. It follows that $\boldsymbol{\mathcal{X}}_{3d}'\boldsymbol{\Lambda}^0\left(\mathbf{P}_{\mathcal{X}_d}-\mathbf{P}_{\mathcal{X}}\right)\boldsymbol{\mathcal{X}}_{3d}=\boldsymbol{\mathcal{X}}_{3d}'\boldsymbol{\Lambda}^0\left(\mathbf{P}_{\mathcal{X}_d}-\mathbf{P}_{\mathcal{X}}\right)^2\boldsymbol{\mathcal{X}}_{3d}$ and hence $\widehat{\mathrm{AV}}\left(\breve{\boldsymbol{\mathcal{X}}}_{3d}^{\mathrm{COR}}\right)<\widehat{\mathrm{AV}}\left(\hat{\boldsymbol{\mathcal{X}}}_{3d}^{\mathrm{COR}}\right)$.

For parts (a') and (b'), the proof is the same as in (a) and (b), in view of the proof of Theorem 1.

References

Andersson, P.G., and Thorburn, D. (2005). An optimal calibration distance leading to the optimal regression estimator. *Survey Methodology,* 31, 1, 95-99.

Australian Bureau of Statistics (2011). Household Expenditure Survey and Survey of Income and Housing, User Guide, Australia, 2009-10 (cat. no. 6503.0).

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

Chipperfield, J.O., and Steel, D.G. (2011). Efficiency of split questionnaire surveys. *Journal of Statistical Planning and Inference,* 141, 1925-1932.

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

Fuller, W.A. (1990). Analysis of repeated surveys. *Survey Methodology,* 16, 2, 167-180.

Gonzalez, J.M., and Eltinge, J.L. (2007). Multiple matrix sampling: A review. *Proceedings of the Survey Research Methods Section,* American Statistical Association, 3069-3075.

Gonzalez, J.M., and Eltinge, J.L. (2008). Adaptive matrix sampling for the consumer expenditure quarterly interview survey. *Proceedings of the Survey Research Methods Section,* American Statistical Association, 3069-3075.

Hidiroglou, M.A. (2001). Double sampling. *Survey Methodology,* 27, 2, 143-154.

Houbiers, M. (2004). Towards a social statistical database on unified estimates at Statistics Netherlands. *Journal of Official Statistics,* 20, 55-75.

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

Kim, J.K., and Rao, J.N.K. (2012). Combining data from two independent surveys: A model-assisted approach. *Biometrika,* 99, 1, 85-100.

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, Serie B,* 72, 27-48.

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

Montanari, G.E. (1998). On regression estimation of finite population means. *Survey Methodology,* 24, 1, 69-77.

Raghunathan, T.E., and Grizzle, J.E. (1995). A split questionnaire survey design. *Journal of the American Statistical Association,* 90, 54-63.

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

Renssen, R.H. (1998). Use of statistical matching techniques in calibration estimation. *Survey Methodology*, 24, 2, 171-183.

Renssen, R.H., and Nieuwenbroek, N.J. (1997). Aligning estimates for common variables in two or more sample surveys. *Journal of the American Statistical Association*, 92, 368-375.

Rust, K.F., and Rao, J.N.K. (1996). Variance estimation for complex surveys using replication techniques. *Statistical Methods in Medical Research*, 5, 283-310.

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

Smith, P. (2009). Survey harmonization in official household surveys in the United Kingdom. *Proceedings of the ISI World Statistical Congresses*, Dublin.

Thomas, N., Raghunathan, T.E., Schenker, N., Katzoff, M.J. and Johnson, C.L. (2006). An evaluation of matrix sampling methods using data from the National Health and Nutrition Examination Survey. *Survey Methodology*, 32, 2, 217-231.

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

Wu, C. (2004). Combining information from multiple surveys through the empirical likelihood method. *Canadian Journal of Statistics*, 32, 15-26.