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

Survey Methodology

Replication variance estimation after sample-based calibration

by Jean D. Opsomer and Andreea L. Erciulescu

Release date: January 6, 2022

Statistics Canada Statistique Canada

Canada

How to obtain more information

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

You can also contact us by

Email at infostats@statcan.gc.ca

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

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

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 "Contact us" > "Standards of service to the public."

Note of appreciation

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

An HTML version is also available.

Cette publication est aussi disponible en français.

Survey Methodology, December 2021
Vol. 47, No. 2, pp. 265-277
Statistics Canada, Catalogue No. 12-001-X

Replication variance estimation after sample-based calibration

Jean D. Opsomer and Andreea L. Erciulescu[1]

Abstract

Sample-based calibration occurs when the weights of a survey are calibrated to control totals that are random, instead of representing fixed population-level totals. Control totals may be estimated from different phases of the same survey or from another survey. Under sample-based calibration, valid variance estimation requires that the error contribution due to estimating the control totals be accounted for. We propose a new variance estimation method that directly uses the replicate weights from two surveys, one survey being used to provide control totals for calibration of the other survey weights. No restrictions are set on the nature of the two replication methods and no variance-covariance estimates need to be computed, making the proposed method straightforward to implement in practice. A general description of the method for surveys with two arbitrary replication methods with different numbers of replicates is provided. It is shown that the resulting variance estimator is consistent for the asymptotic variance of the calibrated estimator, when calibration is done using regression estimation or raking. The method is illustrated in a real-world application, in which the demographic composition of two surveys needs to be harmonized to improve the comparability of the survey estimates.

Key Words: Fishing; Hunting and wildlife watching surveys; Raking; Regression estimation; Replicate construction.

1. Introduction

Variance estimation methods for complex surveys include linearization and replication methods. Some of the practical advantages of replication methods include the facts that multiple weight adjustments such as nonresponse adjustments and calibration are readily incorporated into the estimates, that detailed design information does not need to be released in the public-use datasets, and that data users can readily obtain variance estimates for wide classes of estimators without the need for derivations. There are numerous replication methods in use, with the appropriate choice of method dictated by the sampling design and the estimation objectives of the survey. We refer to Wolter (2007) for an overview of the types of variance estimation replication methods.

The problem we are addressing in this article is how to incorporate calibration into replication variance estimation, when the calibration control totals are themselves random and their variance is also estimated by a replication method. This problem occurred because we (the authors) were working with two surveys on the same topic and for the same target population, for which we were tasked with producing a unified set of estimates.

The first survey is the 2016 National Survey of Fishing, Hunting, and Wildlife-Associated Recreation (FHWAR). This survey, conducted by the U.S. Census Bureau, used successive difference replication (SDR), which is a variant of balanced repeated replication (BRR). SDR was originally proposed in Fay

1. Jean D. Opsomer, Westat, Inc., 1600 Research Blvd., Rockville, MD, USA, 20850. E-mail: JeanOpsomer@westat.com; Andreea L. Erciulescu, Westat, Inc., 1600 Research Blvd., Rockville, MD, USA, 20850. E-mail: AndreeaErciulescu@westat.com.

and Train (1995) and is frequently used for Census Bureau surveys. The second survey is the 2016 50-state Survey of FHWAR, conducted by the Rockville Institute, the nonprofit affiliate of Westat. This survey used Delete-A-Group Jackknife (DAGJK) as the replication method (Kott, 2001).

The two 2016 FHWAR surveys were fielded concurrently using different modes of data collection, specifically to allow for comparison between the two and for subsequent reconciliation of the estimates. The National survey used a combination of telephone and in-person data collection and had a sample size sufficient to produce estimates at the census division level. The 50-state survey was a mail-based survey and, as its name implies, had a sample size sufficient to produce estimates at the state level. However, these differences in mode, together with further differences including other survey implementation aspects, subsampling strategies and estimation methods, led to substantial and often statistically significant differences in the estimates, with typically higher estimates in the 50-State Survey than in the National Survey. See Fish and Wildlife Service and Census Bureau (2018) and Rockville Institute (2018) for more details about the two FHWAR surveys.

As noted above, we were responsible for developing a calibration approach to "align" the estimates from the two surveys, in the sense of producing estimates at the state level based on the 50-state survey but compatible with those obtained from the National Survey. This, in turn, would make it possible to compare the 2016 state-level estimates to those from prior iterations of the National survey, which has been conducted since 1955 and with survey results that are directly comparable since 1991. One of the key steps in reconciling the estimates involved calibrating the demographic composition of the 50-state survey to that of the National survey, given that the latter was considered the "gold standard" in this application. To this end, a set of demographic estimates from the National survey were used as control totals for calibration of the 50-State survey. Because these control totals are themselves estimates, however, it was necessary to make sure that their variability is reflected in the variance estimates of the calibrated 50-State Survey estimates. This is an application of *sample-based calibration* (calibrating to random control totals). Sample-based calibration is typically seen in multi-phase surveys, in which the samples and the estimation methods can be coordinated. In the current setting, the two surveys are independent and have two sets of replicates created using different replication methods.

There is a limited literature on how to account for sample-based calibration in replicate variance estimation. Fuller (1998) developed a replication variance estimator for two-phase samples, in which the phase two estimates are calibrated to phase one control totals. In this approach, the phase two replicates are modified by adjustments derived from the spectral decomposition of the phase one estimated variance-covariance matrix of the control totals. Dever and Valliant (2010) and Dever and Valliant (2016) studied weight calibration to estimated control totals under a scenario where a (benchmark) survey is used to calibrate another (analytic) survey, which is more closely related to our setting. In the latter article, their simulation studies were developed for a generalized regression estimator, and linearization and jackknife replication variance estimation methods were compared. For the jackknife replication, the authors compared the performance of the Fuller (1998) adjustment and two adjustments based on draws from a multivariate normal distribution: one using the full variance-covariance matrix of the control totals, and one using only the diagonal of this matrix. The latter approach had been proposed by Nadimpalli, Judkins and Chu (2004), but no theoretical justification was provided. The method was motivated by considering the asymptotic distribution of the estimated control totals, which is then used to generate "synthetic" versions of these estimates for use as replicate control totals.

In this paper, we describe an approach to modify the replicates of the survey to be calibrated by using the replicates from the control survey directly. We show how this method can be used even when the replication methods and/or the number of replicates differ between the two surveys. Interestingly, Kott (2005) already made a brief mention of an approach that likewise uses the replicates directly, in the special case of both surveys using DAGJK with the same number of replicates. Unlike the methods in Fuller (1998) and Nadimpalli et al. (2004), these approaches do not require explicit calculation of the variance-covariance matrix of the control survey, greatly simplifying implementation in practice. In addition, they use valid calibrated totals, unlike the methods relying on draws from a normal distribution which can result in unstable or even unfeasible calibrated totals.

More generally, methods for harmonizing estimates from two surveys can be viewed as an application of *statistical data integration* (SDI), (Lahiri, 2020), a set of methods used to combine multiple data sources to create improved or new estimates compared to what can be obtained from the separate datasets. While they did not use the term SDI, Lohr and Raghunathan (2017) give an overview of the state-of-the-art tools available to perform most of the commonly encountered SDI activities. In a typical SDI application, the goal is the optimal combination of the information in the multiple data sources, which almost always involves creating an estimator that is different from those that are obtained from the separate sources. Methods to achieve this can be design-based, as in multi-frame estimation (Lohr and Rao, 2006) and composite regression estimation (Merkouris, 2004), or model-based (e.g., Raghunathan, Xie, Schenker, Parsons, Davis, Dodd and Feuer, 2007). Sample-based calibration falls in the design-based category, but also aims to reproduce the estimates from one of the data sources exactly.

The remainder of the paper is as follows. The proposed method is developed under the setting of regression estimation in Section 2. Raking is another common calibration method and the one used for the two surveys of interest, so we extend the results to this setting in Section 3. In Section 4 we illustrate both the Fuller (1998) method and the proposed method using data from the two 2016 surveys of FHWAR. Section 5 provides overall conclusions.

2. Sample-based regression calibration

We consider a survey of a population U with sample s, weights w_i, target variables y_i. For a given survey estimator $\hat{\theta}$ constructed using the weights w_i, inference is conducted by replication, implemented through the provision of R sets of replicate weights $w_i^{(r)}, r = 1, \ldots, R,$ and variance estimation formula

$$\hat{V}(\hat{\theta}) = A\sum_{r=1}^{R}(\hat{\theta}^{(r)} - \hat{\theta})^2, \tag{2.1}$$

where the $\hat{\theta}^{(r)}$ are computed in the same manner as $\hat{\theta}$ but replacing w_i by the $w_i^{(r)}$. The constant A depends on the replication method. For simplicity, we focus in what follows on the Horvitz-Thompson estimator of $t_y = \sum_U y_i$, denoted by $\hat{t}_y = \sum_s y_i/\pi_i$. In this case, many replication methods of the form (2.1) lead to a design consistent estimator of $\mathrm{Var}(\hat{t}_y)$. We will refer to this survey as the "primary survey".

We are interested in creating adjusted weights w_i^* that are calibrated to a set of control totals from a secondary survey of U with sample s_C, weights w_{Ci}. An estimator from this survey is denoted by $\hat{\theta}_C$. For the second survey, a replication-based variance estimator is also provided,

$$\hat{V}_C(\hat{\theta}_C) = A_C\sum_{r=1}^{R_C}(\hat{\theta}_C^{(r)} - \hat{\theta}_C)^2,$$

with replicate weights $w_{Ci}^{(r)}, r = 1, \ldots, R_C$, and replication-specific constant A_C. The control variables will be denoted by $\mathbf{x}_i$, with estimated totals $\hat{\mathbf{t}}_{Cx}$. Using regression estimation as a framework for calibration, the adjusted estimator is

$$\hat{t}_{y,\,\mathrm{reg}} = \hat{t}_y + (\hat{\mathbf{t}}_{Cx} - \hat{\mathbf{t}}_x)^T\hat{\boldsymbol{\beta}} = \sum_s w_i^* y_i \tag{2.2}$$

where $\hat{\boldsymbol{\beta}} = (\mathbf{X}_s^T\mathbf{W}_s\mathbf{X}_s)^{-1}\mathbf{X}_s^T\mathbf{W}_s\mathbf{Y}_s$ with $\mathbf{X}_s$ a matrix with i^{th} row equal to $\mathbf{x}_i^T$, $\mathbf{W}_s$ a diagonal matrix with i^{th} entry w_i and $\mathbf{Y}_s$ a vector containing the $y_i, i \in s$. Hence, the calibrated weights can be written as

$$w_i^* = w_i\left(1 + (\hat{\mathbf{t}}_{Cx} - \hat{\mathbf{t}}_x)^T(\mathbf{X}_s^T\mathbf{W}_s\mathbf{X}_s)^{-1}\mathbf{x}_i\right). \tag{2.3}$$

We note that post-stratification is a special case of regression estimation, see Särndal, Swensson and Wretman (1992, Chapter 7.6).

To obtain a variance estimator, we follow the traditional linearization approach for regression estimators with respect to the sampling design (see e.g., Särndal et al., 1992, Chapter 5.5). Under mild regularity conditions (such as design consistency of Horvitz-Thompson estimators and invertibility of required matrices), the linearized version of the regression estimator (2.2) is equal to the difference estimator,

$$\hat{t}_{y,\,\mathrm{diff}} = \hat{t}_y + (\hat{\mathbf{t}}_{Cx} - \hat{\mathbf{t}}_x)^T\boldsymbol{\beta}_N = \hat{\mathbf{t}}_{Cx}^T\boldsymbol{\beta}_N + \hat{t}_e \tag{2.4}$$

where $\boldsymbol{\beta}_N = (\mathbf{X}_U^T\mathbf{X}_U)^{-1}\mathbf{X}_U^T\mathbf{Y}_U$ is the population target of $\hat{\boldsymbol{\beta}}$ and $\hat{t}_e = \sum_s w_i(y_i - \mathbf{x}_i^T\boldsymbol{\beta}_N)$. The variance of $\hat{t}_{y,\,\mathrm{diff}}$ is equal to

$$\mathrm{Var}(\hat{t}_{y,\mathrm{diff}}) = \mathrm{Var}(\hat{t}_e) + \boldsymbol{\beta}_N^T\,\mathrm{Var}(\hat{\mathbf{t}}_{Cx})\,\boldsymbol{\beta}_N, \tag{2.5}$$

since the two surveys are independent. This "linearized variance" is the variance of the asymptotic distribution of the regression estimator (2.2). In expression (2.5), the first variance term can be estimated using the replicates from the primary survey and the variance-covariance of the control totals in the second term can be estimated using the replicates from the secondary survey. Hence, the plug-in variance estimator

$$\tilde{V}(\hat{t}_{y,\mathrm{reg}}) = V(\hat{t}_{\hat{e}}) + \hat{\boldsymbol{\beta}}^T \hat{V}_C(\hat{\mathbf{t}}_{Cx})\,\hat{\boldsymbol{\beta}}, \tag{2.6}$$

where $\hat{t}_{\hat{e}} = \sum_s w_i(y_i - \mathbf{x}_i^T\hat{\boldsymbol{\beta}}),$ can be used for asymptotically valid inference for $\hat{t}_{y,\mathrm{reg}}.$

However, it is often not practical to maintain the two datasets and associated sets of replicates for variance estimation purposes. In the context of survey calibration, the organization in charge of creating the adjusted weights for the primary survey would often prefer to continue providing their dataset unchanged except for the new calibrated weights and associated replicate weights, so that data users can perform their analyses using traditional survey tools. Hence, it is of interest to create a single set of replicates for the primary survey that can be used to estimate the variance, while accounting for the fact that the control totals are themselves estimated from a different survey.

We therefore propose to construct new replicates for the primary survey to estimate (2.5). Assume for now that $R_C \le R.$ Starting from the replicate weights $w_i^{(r)}$ for the primary survey variance estimator, a replicate variance estimator of the first term in (2.6) is obtained by using the calibrated replicate weights

$$w_{1i}^{*(r)} = w_i^{(r)}\left(1 + (\hat{\mathbf{t}}_{Cx} - \hat{\mathbf{t}}_x^{(r)})^T (\mathbf{X}_s^T\,\mathbf{W}_s^{(r)}\mathbf{X}_s)^{-1}\mathbf{x}_i\right). \tag{2.7}$$

These replicate weights are obtained by repeating the calibration for each of the replicate weights $w_i^{(r)}$ and lead to consistent variance estimation for regression estimators, as discussed for the general case in Fuller (2009, Chapter 4). See also Valliant (1993) for the special case of post-stratification.

The replicate weights $w_{1i}^{*(r)}$ can be further modified to capture the second term in (2.6) as follows:

$$w_i^{*(r)} = w_{1i}^{*(r)} + a_r w_i^{(r)}\,(\hat{\mathbf{t}}_{Cx}^{(r)} - \hat{\mathbf{t}}_{Cx})^T (\mathbf{X}_s^T\;\mathbf{W}_s^{(r)}\mathbf{X}_s)^{-1}\mathbf{x}_i, \tag{2.8}$$

with the constants a_r to be further defined below. Combining (2.7) and (2.8), the resulting replicate weights are

$$w_i^{*(r)} = w_i^{(r)}\left(1 + (\hat{\mathbf{t}}_{Cx} + a_r\,(\hat{\mathbf{t}}_{Cx}^{(r)} - \hat{\mathbf{t}}_{Cx}) - \hat{\mathbf{t}}_x^{(r)})^T (\mathbf{X}_s^T\;\mathbf{W}_s^{(r)}\;\mathbf{X}_s)^{-1}\,\mathbf{x}_i\right). \tag{2.9}$$

These weights are again obtained by applying the same calibration as for the original weights to each of the replicates, but with replicate control totals $\hat{\mathbf{t}}_{Cx}^{*(r)} = \hat{\mathbf{t}}_{Cx} + a_r\,(\hat{\mathbf{t}}_{Cx}^{(r)} - \hat{\mathbf{t}}_{Cx}).$ The resulting replicate estimates are

$$\hat{t}_{y,\mathrm{reg}}^{(r)} = \sum_s w_i^{*(r)} y_i = \hat{t}_y^{(r)} + (\hat{\mathbf{t}}_{Cx} - \hat{\mathbf{t}}_x^{(r)})^T\,\hat{\boldsymbol{\beta}}^{(r)} + a_r\,(\hat{\mathbf{t}}_{Cx}^{(r)} - \hat{\mathbf{t}}_{Cx})^T\,\hat{\boldsymbol{\beta}}^{(r)}.$$

The constants a_r are chosen to account for the difference between the primary and control replication methods, in particular between R_C and R and A_C and A, by letting

$$a_r = \begin{cases} \sqrt{\dfrac{A_C}{A}} & r = 1, \ldots, R_C \\ 0 & r = R_C + 1, \ldots, R. \end{cases} \tag{2.10}$$

This implies that for $r > R_C$, the replicate weights $w_i^{*(r)} = w_{1i}^{*(r)}$ in (2.8), i.e. the unadjusted control totals are used to calibrate the replicate weights. While the a_r are written with the first R_C values non-zero, this is for notational convenience only. The assignment of the replicates from the control survey to those of the primary survey should be randomized, to ensure that estimators and replicate estimators from both surveys remain independent regardless of the replication methods.

Using the replicate weights (2.9) with constants (2.10), the replicate variance estimator (2.1) becomes

$$\hat{V}(\hat{t}_{y,\text{reg}}) = A \sum_{r=1}^{R} (\hat{t}_{y,\text{reg}}^{(r)} - \hat{t}_{y,\text{reg}})^2, \tag{2.11}$$

Ignoring terms of smaller order as well as those with $a_r = 0$, this is approximately equal to

$$\begin{aligned} \hat{V}(\hat{t}_{y,\text{reg}}) \approx\ & A \sum_{r=1}^{R} (\hat{t}_{\hat{e}}^{(r)} - \hat{t}_{\hat{e}})^2 + \hat{\boldsymbol{\beta}}^T A_C \sum_{r=1}^{R_C} (\hat{\mathbf{t}}_{Cx}^{(r)} - \hat{\mathbf{t}}_{Cx})(\hat{\mathbf{t}}_{Cx}^{(r)} - \hat{\mathbf{t}}_{Cx})^T \hat{\boldsymbol{\beta}} \\ & + A \sum_{r=1}^{R} a_r (\hat{t}_{\hat{e}}^{(r)} - \hat{t}_{\hat{e}})(\hat{\mathbf{t}}_{Cx}^{(r)} - \hat{\mathbf{t}}_{Cx})^T \hat{\boldsymbol{\beta}}. \end{aligned}$$

The cross-term is likewise of smaller order because of the independence of the two surveys and the fact that $\sum_{r=1}^{R} \hat{t}_{\hat{e}}^{(r)} \big/ R \approx \hat{t}_{\hat{e}}$ and $\sum_{r=1}^{R_C} \hat{\mathbf{t}}_{Cx}^{(r)} \big/ R_C \approx \hat{\mathbf{t}}_{Cx}$. Hence, the replicate variance estimator (2.11) inherits the design consistency of the original replication methods for both surveys and is design consistent for the linearized variance (2.5).

Finally, we discuss the case when $R_C > R$. The above approach is readily extended to this case by repeating the R replicates of the primary survey K times, such that $R_C \le KR$ with K the smallest positive integer for which this inequality is satisfied. The resulting replicate variance estimator is of the same form as (2.1) but with R replaced by KR and A is replaced by A/K. Then, the method discussed above applies directly to this new replicate variance estimator for the primary survey. For instance, if $R = 120$ and $R_C = 150$, each replicate in the primary survey will be repeated $K = 2$ times, leading to 240 replicates for the primary survey of which 150 will be modified.

3. Sample-based raking calibration

In the application to the two 2016 FHWAR surveys, we used raking rather than regression estimation for calibration. As for regression estimation, the goal is to create new raking controls for each of the replicates, so that the replicate variance estimator for the primary survey accounts for the variability of the control totals from the secondary survey. The above results for regression estimation do not apply directly, but we can apply the same reasoning as in Deville and Särndal (1992) to show that they continue to hold for raking. Instead of relying on this general result, we will derive it here explicitly to show how to obtain the adjusted control totals for the replicates.

For simplicity, we describe here the case in which we are controlling for the marginal counts in domains defined by the levels of 2 categorical variables, denoted a and b, having K and L levels, respectively. In the primary survey, the estimated counts in the domains defined by the intersections of the two variables are $\hat{N}_{a_k b_l} = \sum_s w_i \delta_i (a_k, b_l)$, with $\delta_i (a_k, b_l) = 1$ if element i is in the domain defined by the intersection of a_k and b_l and 0 otherwise. The marginal estimated counts are defined analogously, $\hat{N}_{a_k} = \sum_s w_i \delta_i (a_k, \cdot)$ and $\hat{N}_{b_l} = \sum_s w_i \delta_i (\cdot, b_l)$. We write $\boldsymbol{\delta}_i = \left(\delta_i (a_1), \ldots, \delta_i (a_K), \delta_i (b_1), \ldots, \delta_i (b_L)\right)^T$ for the vector of indicators for the marginal domains for element i. The estimated marginal counts in the primary survey are $\hat{\mathbf{N}} = \sum_s w_i \boldsymbol{\delta}_i = (\hat{N}_{a_1}, \ldots, \hat{N}_{a_K}, \hat{N}_{b_1}, \ldots, \hat{N}_{b_L})^T$ and the corresponding control totals from the secondary survey are $\hat{\mathbf{N}}_C = \sum_{s_C} w_{Ci} \, \boldsymbol{\delta}_i = (\hat{N}_{Ca_1}, \ldots, \hat{N}_{Ca_K}, \hat{N}_{Cb_1}, \ldots, \hat{N}_{Cb_L})^T$. Using the classical raking ratio algorithm of Deming and Stephan (1940) until convergence, the raked weights for the primary survey can be written as

$$\begin{aligned} w_i^* &= w_i \exp(\hat{u}_{a_k} + \hat{u}_{b_l}) \quad \text{for} \quad \delta_i (a_k, b_l) = 1 \\ &= w_i \exp(\hat{\mathbf{u}}^T \boldsymbol{\delta}_i) \end{aligned} \tag{3.1}$$

where $\hat{\mathbf{u}} = (u_{a_1}, \ldots, u_{a_K}, u_{b_1}, \ldots, u_{b_L})^T$ is a solution to the system of $K + L$ equations

$$\begin{aligned} \sum_{l=1}^{L} \hat{N}_{a_k b_l} \exp(u_{a_k} + u_{b_l}) &= \hat{N}_{Ca_k} \quad (k = 1, \ldots, K) \\ \sum_{k=1}^{K} \hat{N}_{a_k b_l} \exp(u_{a_k} + u_{b_l}) &= \hat{N}_{Cb_l} \quad (l = 1, \ldots, L). \end{aligned} \tag{3.2}$$

The solution to these equations is not unique, so one of the unknowns can be set to 0 and an equation removed. This does not affect the values of $\exp(u_{a_k} + u_{b_l})$, and we will set $v_{b_L} = 0$ and remove the last equation in what follows.

There is no explicit expression for the solution to (3.2), but it can be approximated by using a linearization argument. Under the usual survey asymptotic framework that ensures design consistency of Horvitz-Thompson estimators, the $\hat{u}_{a_k}$ and $\hat{u}_{b_l}$ converge to 0 as the sample sizes of the two surveys increase, so that expansion around 0 is valid. Doing so for the equations in (3.2), we approximate the reduced set of $K + L - 1$ equations by

$$\begin{aligned} \sum_{l=1}^{L} \hat{N}_{a_k b_l} (1 + u_{a_k} + u_{b_l} + o_p(1)) &= \hat{N}_{Ca_k} \quad (k = 1, \ldots, K) \\ \sum_{k=1}^{K} \hat{N}_{a_k b_l} (1 + u_{a_k} + u_{b_l} + o_p(1)) &= \hat{N}_{Cb_l} \quad (l = 1, \ldots, L - 1), \end{aligned}$$

which can be rewritten as

$$\begin{aligned} \hat{N}_{a_k} u_{a_k} + \sum_{l=1}^{L} \hat{N}_{a_k b_l} u_{b_l} &= (\hat{N}_{Ca_k} - \hat{N}_{a_k})(1 + o_p(1)) \quad (k = 1, \ldots, K) \\ \hat{N}_{b_l} u_{b_l} + \sum_{k=1}^{K} \hat{N}_{a_k b_l} u_{a_k} &= (\hat{N}_{Cb_l} - \hat{N}_{b_l})(1 + o_p(1)) \quad (l = 1, \ldots, L - 1). \end{aligned} \tag{3.3}$$

Ignoring the smaller order remainders, the solution to this system of linear equations can be written in the form $\hat{\mathbf{u}} = \hat{\mathbf{J}}^{-1}(\hat{\mathbf{N}}_C - \hat{\mathbf{N}})$, where $\hat{\mathbf{J}}$ is a symmetric $(K + L - 1) \times (K + L - 1)$ matrix containing estimated domain counts readily obtained from the left-hand side of (3.3). Note that the resulting $\hat{\mathbf{u}}$ is not a linear estimator, because in the linearization we conditioned on $\hat{\mathbf{N}}$. Finally, plugging $\hat{\mathbf{u}}$ into the expression (3.1) and linearizing again, we obtain

$$w_i^* \approx w_i \left(1 + (\hat{\mathbf{N}}_C - \hat{\mathbf{N}})^T \hat{\mathbf{J}}^{-1} \boldsymbol{\delta}_i\right) \tag{3.4}$$

and the estimator after raking is

$$\hat{t}_{y,\text{rak}} = \sum_s w_i^* y_i \approx \sum_s w_i y_i + (\hat{\mathbf{N}}_C - \hat{\mathbf{N}})^T \hat{\mathbf{J}}^{-1} \boldsymbol{\delta}_s^T \mathbf{W}_s \mathbf{Y}_s. \tag{3.5}$$

Note that the size of the control variable $\boldsymbol{\delta}_i$ and associated estimates is now $K + L - 1$, but we maintain the prior notation for simplicity. In (3.5), $\hat{\mathbf{J}}^{-1} \boldsymbol{\delta}_s^T \mathbf{W}_s \mathbf{Y}_s$ corresponds to $\hat{\boldsymbol{\beta}}$ in the regression estimator (2.2). This asymptotic equivalence between the raking estimator and the regression estimator with the same control totals was established by Deville and Särndal (1992). In particular, they provide sufficient conditions under which the asymptotic variance of the equivalent regression estimator can also be used for inference for the raking estimator. Hence, a variance estimator of the form (2.6) can be constructed, with $\hat{t}_{\hat{e}} = \sum_s w_i (y_i - \boldsymbol{\delta}_i^T \hat{\mathbf{J}}^{-1} \boldsymbol{\delta}_s^T \mathbf{W}_s \mathbf{Y}_s)$.

We now consider the construction of replicate weights for the primary survey that estimate the asymptotic variance of the raking estimator. As before, we construct new replicate control totals $\hat{\mathbf{N}}_C + a_r (\hat{\mathbf{N}}_C^{(r)} - \hat{\mathbf{N}}_C)$ using the replicate estimates $\hat{\mathbf{N}}_C^{(r)}$ from the secondary survey. Each of the sets of replicate weights $w_i^{(r)}$ of the primary survey are adjusted by raking to its corresponding set of control totals, to obtain the $w_i^{*(r)}$ and the raked replicate estimates $\hat{t}_{y,\text{rak}}^{(r)}$. Using the approximation in (3.5) for each replicate, we obtain that the resulting replication variance estimator is consistent for the asymptotic variance of the raking estimator.

4. Application

We return to the calibration problem encountered while bridging the two 2016 FHWAR surveys. For both surveys, the population is defined as individuals of ages 16 and older, living in U.S. households. The main data sources for this application are record-level data files, containing weights and replicate weights for both surveys. Using these datasets, we conducted an initial analysis and identified discrepancies in the demographics, which we adjusted by sample-based raking. Estimated population totals constructed using the record-level data from the National survey were considered as random controls, for the crosstabs of census divisions (nine categories) and each of the following demographic variables:

- residency: two categories corresponding to urban and rural classification,
- age: eight categories corresponding to age ranges 16-17; 18-24; 25-34; 35-44; 45-54; 55-64; 65-74, and 75+,
- sex: two categories corresponding to male and female,
- race-ethnicity: four categories corresponding to Hispanic, non-Hispanic White, non-Hispanic African American, and non-Hispanic all other,
- annual income: nine categories corresponding to income ranges -$20,000; $20,000-$29,999; $30,000-$39,999; $40,000-$49,999; $50,000-$74,999; $75,000-$99,999; $100,000-$149,999; $150,000+, and not reported.

For the application in this article, we use the 50-State survey public use file, which does not contain information on income. Therefore, we illustrate the proposed method in a slightly simplified setting here, using the crosstabs of census divisions and residency, age, sex, and race-ethnicity as the raking dimensions. We implemented both the Fuller (1998) method and the proposed calibration method described in Section 2 using the public-use data files available for both surveys. For comparison, we also show the results of calibrating without adjusting the variance estimates for the random controls, referred to below as the “naive” method because it ignores the variability of the controls in the variance estimates. To compare the variance estimation methods for survey variables that are not control variables, we will also show estimates for domains defined by crosstabs of residency and sex.

While the replication methods of the two FHWAR surveys are different, they both use $R = R_C = 160$ replicates. Referring to expression (2.1), the replication constant for the DAGJK method of the 50-State survey is $A = 159/160$ and the corresponding constant for the SDR method of the National survey is $A_C = 4/160,$ both available from their respective survey documentation. Hence, the replication adjustment constants a_r in (2.10) are equal to $2/\sqrt{159}$ for $r = 1, \ldots, R.$

The estimates we will consider are all estimated domain counts, so we define the target variable $y_i = I_{\{i \in U_d\}}$ for a domain of interest $U_d.$ For the 144 domains defined by the raking dimensions, we write the estimated domain counts as $\hat{t}_k = \sum_s w_i I_{\{i \in U_k\}}, k = 1, \ldots, 144.$ Likewise, the control totals are estimated domain counts from the National survey, so the auxiliary variable vector is $\mathbf{x}_i = \mathbf{I}_i,$ a vector of length 144 containing the indicators for inclusion of respondent i in the control domains $U_k, k = 1, \ldots, 144,$ and let $\hat{t}_{C,k} = \sum_{s_C} w_{Ci} I_{\{i \in U_k\}}.$ We denote the vector of control totals as $\hat{\mathbf{t}}_C = (\hat{t}_{C,1}, \ldots, \hat{t}_{C,144})^T$ and the adjusted replicate control totals are $\hat{\mathbf{t}}_C^{*(r)} = \hat{\mathbf{t}}_C + 2/\sqrt{159}(\hat{\mathbf{t}}_C^{(r)} - \hat{\mathbf{t}}_C).$

In order to implement the Fuller (1998) method, we estimated the variance-covariance matrix of the control totals $\hat{V}_C(\hat{\mathbf{t}}_C)$ using the National survey replicate weights. The spectral decomposition of this matrix resulted in a set of 144 eigenvectors $\mathbf{q}_i$ and associated eigenvalues λ_i, for $i = 1, \ldots, 144$. Following Fuller (1998), we obtain a set of 144 vectors $\mathbf{v}_i$ satisfying

$$\hat{V}_C(\hat{t}_C) = \sum_{i=1}^{144} \mathbf{v}_i' \mathbf{v}_i,$$

where $\mathbf{v}_i = \sqrt{\lambda_i}\, \mathbf{q}_i$, for $i = 1, \ldots, 144$. Finally, the adjusted replicate controls are $\hat{\mathbf{t}}_C^{*(r)} = \hat{\mathbf{t}}_C + \frac{\sqrt{160}}{2} \mathbf{v}_r$ for $r = 1, \ldots, 144$ and $\hat{\mathbf{t}}_C^{*(r)} = \hat{\mathbf{t}}_C$ for $r = 145, \ldots, 160$. This points to a drawback of the Fuller (1998) method: while our approach perturbs the control totals of all 160 replicates, this method only perturbs a fraction of them in this case. In addition, 30 of the 144 eigenvalues were nearly zero, 18 of which less than zero due to floating point error. We truncated the 18 negative eigenvalues to zero, and left the rest unchanged. Hence, to the extent that not all replicates contribute to variance estimates for some survey estimates (e.g., domain totals), there is a risk that the sample-based calibration will be imperfectly reflected in the variance estimates. In general, we expect that a larger number of replicates will be perturbed using our approach, since the estimated variance-covariance matrix of the control totals can only be reliably estimated if its dimension is suitably smaller than R_C.

Tables 4.1 and 4.2 show the estimates and standard errors, respectively, for domains defined by residency and sex, before and after calibration. The first four rows contain the results for marginal totals for raking variables, which are exactly calibrated, while the last four are totals that correspond to the intersection of raking dimensions and are therefore not exactly calibrated.

Both surveys are representative of the same target population, but the estimates and associated standard errors differ, reflecting both sampling variability as well as different calibration approaches applied by the two survey organizations. As Table 4.1 confirms, after the 50-state survey is raked to the National survey, the estimated totals for domains defined as exact calibration domains indeed match exactly between both surveys. For the domains defined by the crosstabulation of residency and sex, the raked estimates for the 50-State survey are close but not identical to those of the National survey.

Table 4.1
Population estimates before and after calibration, rounded to the nearest integer, after scaling by 10^3

Domain		Before Calibration		After Calibration
		50-State	National	
Residency:	Urban	203,445	208,695	208,695
	Rural	51,511	45,991	45,991
Sex:	Male	128,276	121,775	121,775
	Female	126,680	132,911	132,911
Rural:	Male	99,547	98,511	98,089
	Female	103,898	110,184	110,607
Urban:	Male	28,729	23,264	23,686
	Female	22,782	22,727	22,305

Table 4.2 shows the standard errors obtained by the two replication methods with adjusted control totals and by the naive method, which does not account for the randomness of the control totals. By construction, the proposed replication-based adjustment method and the Fuller (1998) method lead to identical variance estimates for domains that are used in the calibration. These variance estimates are also equal to those from the control survey in this case. This reflects the fact that the variance component corresponding to the first term in (2.5) is set to 0 for the control totals, while the variance component for the second term is exactly equal to the control survey variance estimate in the case of raking. Because that variance component is ignored in the naive method, the variance estimates are equal to zero. For the estimated totals for domains defined as the crosstabulation of residency and sex, the variance estimates of the two methods are not identical but close (within 8% of each other), reflecting the fact that both are consistent for the asymptotic variance (2.5). The variance estimates under the naive method are smaller than the variance estimates under the other two calibration methods, leading to an obviously incorrect result due to not accounting for the variance in the random control totals. For other variables, the variance is still expected to be underestimated under the naive method, due to the fact that the second term in the asymptotic variance (2.5) is ignored.

Table 4.2
Standard errors of population estimates before and after calibration, rounded to the nearest integer, after scaling by 10^3

Domain		Before Calibration		After Calibration		
		50-State	National	Naive	Fuller	Proposed
Residency:	Urban	1,922	2,664	0	2,664	2,664
	Rural	1,922	2,598	0	2,598	2,598
Sex:	Male	2,117	1,074	0	1,074	1,074
	Female	2,117	1,112	0	1,112	1,112
Rural:	Male	2,118	1,399	853	1,658	1,533
	Female	2,514	1,797	853	1,964	1,970
Urban:	Male	1,595	1,449	853	1,709	1,641
	Female	979	1,271	853	1,470	1,547

5. Conclusions

We have proposed an approach to account for sample-based calibration in the variance estimates. The approach applies to situations in which both the survey being calibrated and the survey providing the calibration controls use replicate variance estimation, as is often the case in large-scale government surveys. The replication methods in each are arbitrary, as long as they are both valid for their specific surveys. We described the approach for the cases of calibration by regression estimation (including post-stratification) and raking, two commonly used methods in practice, and we anticipate it would work similarly for other types, such as the general calibration estimators of Deville and Särndal (1992).

The main alternative to the proposed method is that of Fuller (1998). Relative to that method, an important advantage of our approach is that it does not require computation of the estimated variance-covariance matrix of the control totals, so that it is very straightforward to implement. In the typical application in which the number of control totals is smaller than the number of replicates, another potential advantage of the proposed method is that the perturbations will be applied across a larger fraction of the replicates. This reduces the risk of computing replicate variance estimates that do not fully reflect the variability of the control totals. For instance, this can occur when only a subset of the replicates contributes to the variance estimate of a domain mean. If these replicates are mostly unperturbed, the resulting variance estimate can underestimate the variance. Further investigation of the performance of the proposed method when the number of replicates of the two surveys are different appears warranted.

References

Deming, W., and Stephan, F. (1940). On a least squares adjustment of a sampled frequency table when the expected marginal totals are known. *Annals of Mathematical Statistics*, 11, 427-444.

Dever, J.A., and Valliant, R. (2010). A comparison of variance estimators for poststratification to estimated control totals. *Survey Methodology*, 36, 1, 45-56. Paper available at https://www150.statcan.gc.ca/n1/en/pub/12-001-x/2010001/article/11251-eng.pdf.

Dever, J., and Valliant, R. (2016). General regression estimation adjusted for undercoverage and estimated control totals. *Journal of Survey Statistics and Methodology*, 4, 289-318.

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

Fay, R.E., and Train, G. (1995). Aspects of survey and model-based postcensal estimation of income and poverty characteristics for states and counties. In *Proceedings of the Section on Survey Research Methods*, American Statistical Association, Alexandria, VA, 154-159.

Fish and Wildlife Service and Census Bureau (2018). *2016 National Survey of Fishing, Hunting, and Wildlife-Related Recreation*. Methodology Report.

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

Fuller, W.A. (2009). *Sampling Statistics*. Hoboken, NJ: John Wiley & Sons, Inc.

Kott, P.S. (2001). The delete-a-group jackknife. *Journal of Official Statistics*, 17, 521-526.

Kott, P.S. (2005). Using the delete-a-group jackknife variance estimator in an economic survey of US farms. In *Proceedings of the 55th Session of the World Statistics Congress*, Sydney. International Statistical Institute.

Lahiri, P. (2020). Preface to special issue on statistical data integration. *Statistics in Transition*, 21, III-VI.

Lohr, S.L., and Raghunathan, T.E. (2017). Combining survey data with other data sources. *Statistical Science*, 32, 293-312.

Lohr, S., and Rao, J.N.K. (2006). Estimation in multiple-frame surveys. *Journal of the American Statistical Association*, 101(475), 1019-1030.

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

Nadimpalli, V., Judkins, D. and Chu, A. (2004). Survey calibration to CPS household statistics. In *Proceedings of the Section on Survey Research Methods*, American Statistical Association, Alexandria, VA, 4090-4094.

Raghunathan, T.E., Xie, D., Schenker, N., Parsons, V.L., Davis, W.W., Dodd, K.W. and Feuer, E.J. (2007). Combining information from two surveys toestimate county-level prevalence rates of cancer risk factors and screening. *Journal of the American Statistical Association*, 102, 474-486.

Rockville Institute (2018). *2016 50-State Survey of Fishing, Hunting, and Wildlife-Related Recreation*. Methodology Report.

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

Valliant, R. (1993). Poststratification and conditional variance estimation. *Journal of the American Statistical Association*, 88, 89-96.

Wolter, K.M. (2007). *Introduction to Variance Estimation* (2 Ed.). New York: Springer-Verlag Inc.