Some Thoughts on Calibration Applications for Multipurpose Estimations

Yan K. Liu[1], Yuqi Liu[2]
[1]Statistics of Income/IRS, 1111 Constitution Avenue, NW, Washington, DC 20224
[2]The George Washington University, 801 22nd Street. NW, Washington, D.C. 20052

Abstract
The linear calibration approach is commonly used to adjust sample weights for multipurpose estimations. Although the linear calibration estimator has many good properties, it has two major limitations. One is the efficiency issue and the other is the convergence issue. Many modified methods have been suggested in the literature, but they do not address both issues at the same time. The modified methods intended for efficiency gain only apply to a single study variable, while modified methods intended for stable estimations for multiple study variables do not offer efficiency gains. In this paper, we first review some modified calibration methods and offer some thoughts about calibration applications to balance efficiency gain and multipurpose use. In particular, we propose a composite calibration weighting method and some modifications of the regular linear calibration method. The proposed methods aim at improving the efficiency for some 'priority' study variables, while still keeping the multipurpose property. Results of a limited simulation study are presented as well.

Key Words: Auxiliary variable, Composite calibration estimation, Generalized linear regression, Multipurpose estimation, Sample weight, Weight calibration.

1. Introduction

The calibration method is commonly used to adjust sample weights when auxiliary variables are available. The calibration weights are adjusted weights such that weighted sample estimates of auxiliary variables conform to the population values. We follow the conventional notations. Let s be the probability sample from the finite population $U = \{1, 2, \cdots, k, \cdots, N\}$ and d_k and w_k be the base weight and the calibration weight of unit k. The objective is to estimate the population total of the study variable: $Y = \sum_{k \in U} y_k$ or population totals of study variables: $Y_j = \sum_{k \in U} y_{jk}$ ($j = 1,2,\cdots, J$) if there are J study variables. The calibration estimator of Y is $\hat{Y}_{CAL} = \sum_{k \in s} w_k y_k$, where calibration weights w_k are calculated through the use of auxiliary information to satisfy a set of benchmark constraints (BC) or calibration equations:

$$\sum_s w_k \mathbf{x}_k = \sum_U \mathbf{x}_k \ , \tag{1.1}$$

where N and $\sum_U \mathbf{x}_k$ are known control totals. Suppose there are p auxiliary variables, then $\mathbf{x}_k$ is a column vector such as $\mathbf{x}_k' = \left(1, x_{1k}, \ldots, x_{pk}\right)$. The '1' is included in $\mathbf{x}_k$ to

ensure that $\sum_s w_k = N$ holds. There is one calibration equation for each of p auxiliary variables. There are sets of calibration weights $\{w_k\}$ that satisfy calibration equations (1.1). To choose among the sets, a distance function of d_k is specified so that a unique set of calibration weights is achieved by minimizing the distance function subject to calibration equations. There are a variety of alternative distance functions, as described in Deville and Särndal (1992). The most commonly used distance function is the chi-squared distance function:

$$D(w_1,\cdots,w_n) = \sum_s (w_k - d_k)^2 / q_k\, d_k \tag{1.2}$$

where q_k is a scale factor. The choice of q_k has some but often limited impact on the accuracy of calibration estimators. The standard choice is q_k=1, which we use in this paper. Alternative distance functions generate asymptotically equivalent calibration estimators. The chi-squared distance function with q_k=1 creates the typical linear calibration estimator $\hat{Y}_{CAL}$ and calibration weights w_k:

$$\hat{Y}_{CAL} = \sum_s w_k y_k, \tag{1.3}$$

where $w_k = d_k g_k$ and $g_k = 1 + \left(\sum_U \mathbf{x}_k - \sum_s d_k \mathbf{x}_k\right)' \left(\sum_s d_k \mathbf{x}_k \mathbf{x}_k'\right)^{-1} \mathbf{x}_k$. This calibration estimator under the chi-squared distance function happens to be identical to the linear GREG estimator that is generated by a linear assisting model, but is derived using different reasoning (Särndal, 2007):

$$\hat{Y}_{GREG} = \sum_s d_k y_k + \left(\sum_U \mathbf{x}_k - \sum_s d_k \mathbf{x}_k\right)' \hat{\mathbf{B}}, \tag{1.4}$$

where $\hat{\mathbf{B}} = \left(\sum_s d_k \mathbf{x}_k \mathbf{x}_k'\right)^{-1} \left(\sum_s d_k \mathbf{x}_k y_k\right)$.

Although there is no explicit model, the linear calibration estimator is not 'model-free.' In fact, a linear assisting model, called a calibration model, is implied in the calculation. If there are J study variables, then for each study variable y_j $(j = 1,2,\cdots,J)$, a corresponding linear calibration model is implied:

$$y_{jk} = \mathbf{x}_k \mathbf{B}_j + \varepsilon_{jk},\ (j = 1,2,\cdots,J;\ \ k = 1,2,\cdots,N). \tag{1.5}$$

Note that vector $\mathbf{x}_k$ is common to each study variable y_j; but vector $\mathbf{B}_j$ is different for each y_j. The common set $\mathbf{x}_k$ allows the creation of a common set of calibration weights for all study variables (Kott, 2009).

Calibration estimators have several advantages. First, calibration estimates of auxiliary variables are consistent with known population totals, that is, $\sum_{k\in s} w_k x_{ik} = \sum_{k\in U} x_{ik}$ $(i = 1,2,\cdots,p)$. Second, calibration weights can be applied to all *y*-variables. This multipurpose use is an important property in survey practice. In

addition, the calculation of calibration weights does not require the population information on auxiliary variables at the unit level, as long as control totals are available.

Calibration estimators have two limitations: efficiency issue and convergence issue. Regarding efficiency, calibration estimators should be considered efficient when the calibration model in the equation (1.5) can explain all study variables well. But, it may not be reasonable to expect a good linear model with the same set of auxiliary variables for each study variable. Therefore, the multipurpose property of calibration weights comes at a price of increased variance. For the convergence issue, iterative process of calculating calibration weights may not always converge, especially when there are many auxiliary variables in the calibration model. Around those issues, methods have been proposed in the literature. Some methods are intended to improve the efficiency of calibration estimators through calibration model fitting and auxiliary variable selection. But they only apply to a single study variable. Others are intended to ensure calibration convergence and stable estimations for multipurpose use, but do not aim at improving efficiency. Is there a way to have both efficiency gain and multipurpose property? In this paper, we first review some modified calibration methods in the literature and offer some thoughts about balancing efficiency gain and multipurpose use. In particular, we propose a composite calibration weighting method and some modifications of the linear calibration method. They are intended to improve efficiency for a 'priority' study variable and keep the multipurpose property.

We divide this paper into six sections. Section 1, this section, serves as the introduction section serves as Section 1. In Section 2, we review a number of methods that have been suggested in the literature to improve the efficiency of calibration estimators for a single study variable. In Section 3, we look at modified calibration methods in literature intended for multipurpose use. In Section 4, we propose a composite calibration estimation method and some modifications of linear calibration estimator to improve the efficiency of the estimate for most important variable, while keeping the multipurpose property. In Section 5, we present results from a limited simulation study. In Section 6, we wrap up with the summary.

2. Modified Calibration Methods for a Single Study Variable

There are two basic components that determine calibration estimators: the distance function and the calibration model. The choice of the distance function is less important as different distance functions give asymptotically equivalent calibration estimators (Deville and Särndal, 1992). However, a good calibration model may greatly improve the efficiency of the calibration estimator. As indicated in Särndal (2007), the linear calibration estimator (or the linear GREG estimator) is bias-robust in the sense that they are nearly unbiased even if the assisting model falls short of 'correct'. However, they are not variance-robust in the sense that model-dependent alternatives may have considerably smaller variance. The limitation is that an explicit model can be specified only for a single study variable. In this section, we review some methods in the literature that aim to improve the efficiency of calibration estimators for a single study variable. These methods use the refined calibration model fitting of either a nonlinear or linear model with a selected set of auxiliary variables.

2.1 Model-Calibration Approach

As proposed by Wu and Sitter (2001), the idea of the model-calibration approach is to make more effective use of the known auxiliary information in fitting the assisting model.

The calibration model is a non-linear model $E(y_k/\mathbf{x}_k)=u(\mathbf{x}_k,\theta)$ that captures the relationship between y_k and $\mathbf{x}_k$. The fitted value $\hat{u}_k=u(\mathbf{x}_k,\hat{\theta})$, $k=1,2,\ldots,N$ is known for each unit in the population. Weights are calibrated to the population total of fitted values, not totals of values on auxiliary variables. Benchmark constraints (BC) or calibration equations are therefore

$$\begin{aligned}\sum_s w_k &= N \\ \sum_s w_k\hat{u}_k &= \sum_U \hat{u}_k\end{aligned}. \tag{2.1}$$

Note that there is only one equation in (2.1) involving the auxiliary variables through $u(\mathbf{x}_k,\hat{\theta})$, while there are *p* equations for *p* auxiliary variables in the linear calibration equations (1.1). The model-calibration estimator is calculated by minimizing the chi-squared distance function subject to calibration equations (2.1):

$$\hat{Y}_{MCAL}=\sum_s w_k y_k=\sum_s d_k y_k+\left(\sum_U \hat{u}_k-\sum_s d_k\hat{u}_k\right)\hat{B}_M, \tag{2.2}$$

Where $\hat{B}_M=\sum_s d_k(\hat{u}_k-\bar{u})(\hat{y}_k-\bar{y})\Big/\sum_s d_k(\hat{u}_k-\bar{u})^2$, $\bar{u}=\sum_s d_k\hat{u}_k\Big/\sum_s d_k$ and $\bar{y}=\sum_s d_k\hat{y}_k\Big/\sum_s d_k$.

Compared to the linear calibration estimator, the model-calibration estimator has a considerable variance advantage because of the refined model. However, model-calibration weights are calibrated to $\sum_U \hat{u}_k$ and are not independent of the *y*-variable, resulting in the loss of the practical advantages of consistency with known population totals $\sum_U x_k$ and the multipurpose use of calibration weights to all *y*-variables.

2.2 Variable Selection Approach

In the variable selection approach, the calibration model is still a linear model. However, the set of auxiliary variables is selected carefully based on the mean squared error of the calibration estimator for a particular study variable, calculated from the sample data. The idea is that using too many auxiliary variables generally reduces the bias of the calibration estimate, but may increase the variance a lot, and therefore, actually increases the mean squared error. Properly chosen calibration variables give a more efficient calibration estimate than calibrating on all available variables.

Silva and Skinner (1997) explored the selection of auxiliary variables based on a model-assisted framework and a simple random sample setting. Their approach was to identify a subset of auxiliary variables under which the linear calibration estimator (or GREG estimator) of a specific study variable has the smallest mean squared error (MSE) estimate. With $d_k=1$ under the simple random sample setting, the linear calibration estimator (or GREG estimator) from equations (1.3) or (1.4) is: $\hat{Y}_{CAL}=\hat{Y}_{GREG}=\sum_s y_k+\left(\sum_U \mathbf{x}_k-\sum_s \mathbf{x}_k\right)'\hat{\mathbf{B}}$, where $\hat{\mathbf{B}}=\left(\sum_s \mathbf{x}_k\mathbf{x}_k'\right)^{-1}\left(\sum_s \mathbf{x}_k y_k\right)$. Silva and Skinner considered different variable selection procedures. Their 'all subset' approach involved computing the estimated MSE for all 2^p possible subsets of *p*

auxiliary variables and choosing the subset corresponding to the smallest MSE estimate. Their forward selection approach started with a sample mean as an initial estimator and then adds one variable at a time until the MSE estimate started to increase.

Clark and Chambers (2008) extended the variable selection approach to the model-based framework. They called the method 'adaptive calibration' because the set of variables is chosen adaptively from sample data. Under the model assumption $E(y_k) = \mathbf{x}_k' \hat{\boldsymbol{\beta}}$ and $\text{var}(y_k) = v_k \sigma^2$, they considered the best linear unbiased predictor (BLUP): $\hat{Y}_{BLUP} = \sum_s y_k + \sum_r \mathbf{x}_k' \hat{\boldsymbol{\beta}}$, where $r = U - s$ is the set of non-sample units and $\hat{\boldsymbol{\beta}} = \left(\sum_s v_k^{-1} \mathbf{x}_k \mathbf{x}_k'\right)^{-1} \sum_s v_k^{-1} \mathbf{x}_k y_k$. Similar to equation (1.3), $\hat{Y}_{BLUP}$ is rewritten as

$$\hat{Y}_{BLUP} = \sum_s w_k^* y_k \,, \tag{2.3}$$

where $w_k^* = 1 + \left(\sum_r \mathbf{x}_k\right)' \left(\sum_s v_k^{-1} \mathbf{x}_k \mathbf{x}_k'\right)^{-1} v_k^{-1} \mathbf{x}_k$. It is straightforward to show that $\sum_s w_k^* \mathbf{x}_k = \sum_u \mathbf{x}_k$. Therefore, $\hat{Y}_{BLUP}$ is a predictor or calibration estimator of *Y*. The decision whether to choose a variable at each step was based on minimizing $\text{MSEP}\left(\hat{Y}_{BLUP}\right)$, the mean squared error of prediction for $\hat{Y}_{BLUP}$. Clark and Chamber's approach was to choose between $\hat{Y}_A$, the predictor based on all auxiliary variables and $\hat{Y}_B$, the predictor based on a subset of auxiliary variables. They developed estimates of $\Delta = \text{MSEP}(\hat{T}_A) - \text{MSEP}(\hat{T}_B)$ for single stage sampling and multi-stage sampling. If $\hat{\Delta} < 0$, then $\hat{Y}_A$ is preferred. If $\hat{\Delta} > 0$, then $\hat{Y}_B$ is selected. Note that $\hat{Y}_{BLUP}$ in equation (2.3) is a model-based calibration estimator, while $\hat{Y}_{CAL}$ in equation (1.3) is a model-assisted calibration estimator.

The work by Silva and Skinner applies to the linear GREG estimator (or the linear calibration estimator) that is based on the model-assisted framework, while the work by Clark and Chambers applies to the BLUP estimator that is based on the model-based framework. The GREG estimator is approximately bias robust, but less efficient even when the model is correct. The BLUP is sensitive to the model assumption, but efficient when the model is correct. Both are variable-specific and sample-specific.

3. Modified Calibration Methods for Multiple Study Variables

The methods in Section 2 aim to improve the efficiency of the calibration estimator for a variable of interest. The resulting calibration weights are efficient for a single study variable, but generally not efficient for other study variables. Now, we look at two methods that have been proposed to improve the calibration estimation for multiple study variables. One is the auxiliary variable reduction approach and other is the ridge regression approach. Both approaches do not require a study variable and aim to provide a single set of calibration weights for all survey variables. They do not guarantee a gain in efficiency.

3.1 Auxiliary Variable Reduction Approach

This approach provides some numerical analysis on auxiliary variables for linear calibration. The idea is that linearly dependent and nearly linearly dependent constraints from the matrix, $\left(\sum_s d_k \mathbf{x}_k \mathbf{x}_k'\right)$, should be discarded to ensure calibration convergence and a stable estimation since the inverse of $\left(\sum_s d_k \mathbf{x}_k \mathbf{x}_k'\right)$ involves in linear calibration calculation (see equations 1.3 and 1.4). The linearly dependent columns or singularity is checked by looking at the Hermite canonical form of $\left(\sum_s d_k \mathbf{x}_k \mathbf{x}_k'\right)$, as introduced in Silva and Skinner (1997), except they had a the simple random sampling setting and d_k=1 (for all k). Any zero in the diagonal of the Hermite canonical form indicates the corresponding auxiliary variables should be dropped.

To identify nearly linearly dependent constraints, Bankier *et al* (1997) used the Conditional Number (CN) method. $\mathrm{CN} = \lambda_{\max} / \lambda_{\min}$, where $\lambda_{\max}$ and $\lambda_{\min}$ are the largest and smallest eigenvalues of $\left(\sum_s d_k \mathbf{x}_k \mathbf{x}_k'\right)$. If CN is larger than some specified number, the corresponding auxiliary variable is removed. Since eigenvalues depend on the units of auxiliary variables, the conditional number method may be limited to a setting where all auxiliary variables have the same scale such as counts. Silva and Skinner (1997) suggested standardizing auxiliary variables before calculating CN.

It is worth noting that SAS performs collinearity analysis on the auxiliary variable matrix $\mathbf{X}$. With the COLLIN option, PROC REG calculates the condition number of the scaled $\mathbf{X}$ matrix. First, $\mathbf{X}'\mathbf{X}$ is scaled to have 1s on the diagonal. Then the eigenvalues and condition indices are calculated. The condition indices are the square roots of the ratio of the largest eigenvalue to each individual eigenvalue. The largest condition index is the condition number of the scaled matrix. For each x-variable, the PROC REG calculates the proportion of the variance of the estimate accounted for by each principal component. A collinearity problem occurs when a component associated with a high condition index contributes strongly (variance proportion greater than about 0.5) to the variance of two or more variables. In addition, the PROC REG provides the variance inflation factors (VIF) using the VIF option. These factors measure the inflation in the variances of the parameter estimates due to collinearities that exist among the x-variables. There is no formal cutoff of VIF value for determining the presence of collinearity. But a large VIF value (greater than 10) indicates a possible collinearity problem. More details and examples can be found in the SAS/STAT 9.2 User's Guide: The REG Procedure.

3.2 Ridge-Shrinkage Calibration Approach

A disadvantage of linear calibration is that weights can be less than one or extremely large. It is common to put Range Restrictions (RR) on final calibration weights, that is, pre-specified lower and upper bounds on calibration weights. As indicated in Rao and Singh (1997), benchmark constraints (BC) are need for the efficiency due to correlated auxiliary information, and make estimates consistent with auxiliary population totals, while RR are needed to avoid extreme weights. However, simultaneously satisfying both BC and RR may cause convergence problems in linear calibration. On the other hand, hitting benchmarks is not the goal in practice, while relaxing BC can introduce much flexibility in calibration. Therefore, several modifications to the linear calibration have been proposed in the literature to balance RR and BC, one of which is the ridge-shrinkage method (Rao and Singh 1997). The strategy of this iterative method is to let BC relax minimally so as to satisfy RR, while minimizing the linear distance function at the same

time. For a given RR, the set of benchmark constraints $\sum_s w_k \mathbf{x}_k - \sum_U \mathbf{x}_k = 0$ in equation (1.1) is replaced by a different set of benchmark constraints:

$$\left| \sum_s w_k \mathbf{x}_k - \sum_U \mathbf{x}_k \right| < \boldsymbol{\delta} . \tag{3.1}$$

The tolerance levels of BC, $\boldsymbol{\delta}$, start as 0 and are revised adaptively (in small increments if convergence is not achieved) only when necessary in the interest of efficiency and consistency. The R function 'CALFUN' can easily calculate the ridge-shrinkage weights.

The ridge-shrinkage calibration is a very useful method of multipurpose estimations in practice. After all, it is not necessary to satisfy auxiliary benchmark constraints perfectly, while a small discrepancy between auxiliary population benchmarks and their estimates can trade the calibration convergence. Prior to the calibration, it is helpful to perform some analysis to identify and remove linearly dependent and nearly linearly dependent constraints as described in Section 3.1.

4. Some Thoughts on the Calibration Application for Multipurpose Estimations

The linear calibration estimator is bias-robust or nearly unbiased even when the assisting model is not correct, but not variance-robust (Beaumont and Alavi 2004, Sarndal 2007). In literature, modified methods suggested for efficiency gain only apply to a single specific study variable, while the modified methods intended for multipurpose use do not offer efficiency gain. Is there a way to have both efficiency gain and multipurpose property? For this question, we propose two modified methods, one is the composite calibration estimation and the other is the modified nonlinear GREG.

4.1 Composite Calibration Estimation

In some surveys, multiple study variables have different levels of importance. For example, in the IRS Statistics of Income's Sales and Capital Assets study, there are 16 variables of interest, but some are more important than others if they need to be put in order. To balance between the efficiency of some 'priority' estimates and the multipurpose use of weights, we consider composite calibration weights that combine variable-specific weights and the multipurpose weights.

4.1.1 Composite Calibration Weights and Composite Calibration Estimator

Suppose that $y_1, y_2, \ldots, y_J$ are J study variables in the order of the importance. Our goal is to find a set of weights for estimating totals of J study variables and provide good efficiency especially for estimating the total of y_1. One option is to use composite calibration weights:

$$w_k^C = \alpha \; w_k^1 + (1-\alpha) w_k^0, \quad 0 \le \alpha \le 1, \tag{4.1}$$

where w_k^1 is the y_1-specific calibration weight and w_k^0 is the multipurpose linear calibration weight for unit k. We may call w_k^1 the 'y_1-priority' weight since it aims at improving the efficiency of variable y_1. The y_1-specific weight w_k^1 can be either linear or non-linear calibration weight, but only focus on the efficiency of the estimate of y_1. The

multipurpose weight w_k^0 is the linear calibration weight based on all x-variables or a subset of x-variables. The composite weight w_k^C should still be multipurpose, but offers a higher efficiency for estimating the total of y_1 than w_k^0, at a price of the possibly lower efficiency for estimates of other variables $y_2, \ldots, y_J$.

The corresponding composite calibration estimator of Y_j, the total for the study variable y_j $(j = 1, 2, \cdots, J)$, is

$$\hat{Y}_j^C = \sum_{k \in s} w_k^C y_{jk} = \alpha \hat{Y}_j^1 + (1-\alpha)\hat{Y}_j^0, \quad 0 \le \alpha \le 1, \tag{4.2}$$

where $\hat{Y}_j^1 = \sum_{k \in s} w_k^1 y_{jk}$ and $\hat{Y}_j^0 = \sum_{k \in s} w_k^0 y_{jk}$ are estimators of Y_j from weights w_k^1 and w_k^0 separately. The composite calibration estimator $\hat{Y}_j^C$ is approximately design-unbiased, as the two components $\hat{Y}_j^1$ and $\hat{Y}_j^0$ are approximately design-unbiased. With this composite calibration estimator, we can focus on the priority variable y_1 more or less by choosing a value for α. When $\alpha = 0$, the composite estimator reduces to the linear calibration estimator.

4.1.2 Mean Squared Error Reduction in the Composite Calibration Estimator

To show the efficiency change of the composite calibration estimator $\hat{Y}_j^C$ compared to the linear calibration estimator $\hat{Y}_j^0$, we compare the mean squared errors (MSE) of two estimators. Assume the two components in equation (4.2) $\hat{Y}_j^1$ and $\hat{Y}_j^0$ are approximately independent. At least, this can be done by choosing the set of *x*-variables involving in $\hat{Y}_j^1$ and the set of *x*-variables involving $\hat{Y}_j^0$ non-overlap. The mean squared error of the composite calibration cstimator is:

$$MSE(\hat{Y}_j^C) \approx \alpha^2 MSE(\hat{Y}_j^1) + (1-\alpha)^2 MSE(\hat{Y}_j^0). \tag{4.3}$$

Equivalently,

$$\frac{MSE(\hat{Y}_j^C)}{MSE(\hat{Y}_j^0)} \approx \alpha^2 r_j + (1-\alpha)^2, \tag{4.4}$$

where $r_j = \dfrac{MSE(\hat{Y}_j^1)}{MSE(\hat{Y}_j^0)}$ for the notation simplification. Equation (4.4) gives the ratio of the MSE of the composite calibration estimator to the MSE of the multipurpose linear calibration estimator as a function of α. To give an idea of possible MSE reduction of the composite calibration estimator, we use the graph illustrations similar to those in the small area estimation (Schaible 1978). The following graphs in Figure 1 show the

relationship between $\frac{MSE(\hat{Y}_j^C)}{MSE(\hat{Y}_j^0)}$ and α for a few scenarios. The first scenario is that for $r_j = 0.5$, or that $MSE(\hat{Y}_j^1)$ is half of $MSE(\hat{Y}_j^0)$, the composite calibration estimator has a smaller MSE than the linear multipurpose calibration estimator and is the lowest for α around 0.65. The last scenario is when $r_j = 4$. It shows that the MSE of the composite calibration estimator can be much larger than that of the linear calibration estimator when $MSE(\hat{Y}_j^1)$ is much larger than $MSE(\hat{Y}_j^0)$, which can be the case for 'non-priority' study variables. Note that $r_1 < 1$ by default because the calibration estimator for the 'priority' study variable y_1 is designed to have a better efficiency than the linear multipurpose calibration estimator. For other 'non-priority' study variables y_j $(j \neq 1)$, r_j can still be less than 1, but is more likely to be larger than 1. The point here is that the composite estimator can be more efficient than the linear calibration estimator for some study variables, but less efficient for other study variables within a tolerable range.

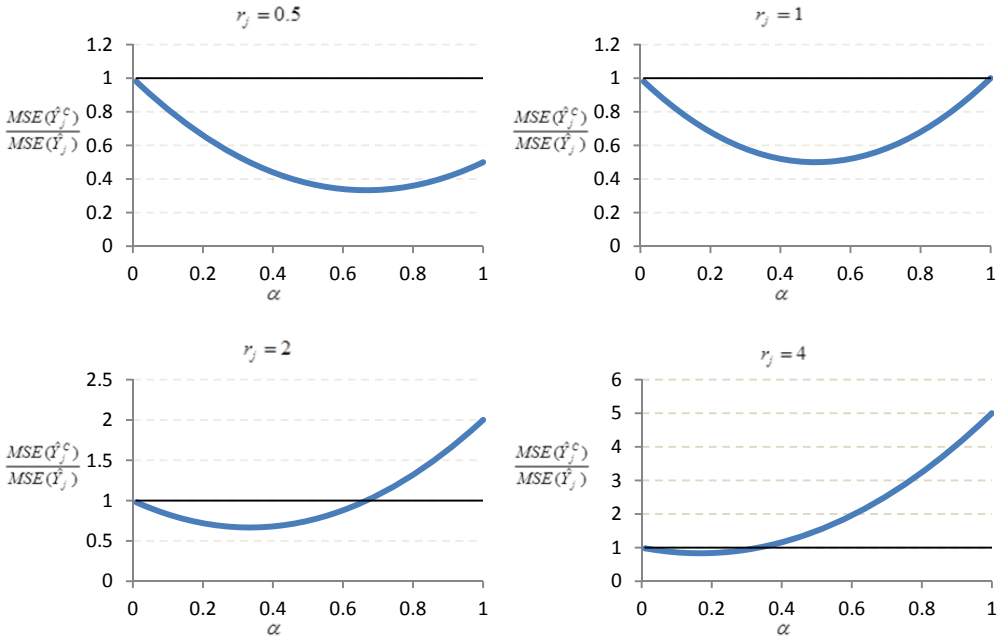

Figure 1: Mean Square Error of the Composite Calibration Estimator Relative to Mean Square Error of the Linear Calibration Estimator as a function of α

4.1.3 Choosing an Value for α

It is important to choose a proper value for α to properly weight w_k^1 and w_k^0 to balance the MSE deduction for the estimate of the 'priority' study variable and the possible variance increase for the estimates of other study variables. We may consider two measures in choosing a proper value for α:

(1) The relative error of the estimate of each auxiliary variable x_i $(i = 1, 2, \ldots, p)$

$$RE_i = \left| \hat{X}_i^C / X_i - 1 \right|, \quad i = 1, 2, \ldots, p, \tag{4.5}$$

where $\hat{X}_i^C = \sum_{k \in s} w_k^C x_{ik} = \sum_{k \in s} [\alpha w_k^1 + (1-\alpha) w_k^0] x_{ik}$ and X_i is the known population total.

(2) The relative efficiency of the composite calibration estimator for each study variable y_j ($j = 1, 2, \ldots, J$)

$$\frac{MSE(\hat{Y}_j^C)}{\hat{Y}_j^0}, \quad j = 1, 2, \ldots, J, \tag{4.6}$$

where $MSE(\hat{Y}_j^C)$ is often replaced by the variance estimate in practice.

The first measure allows us to see how much benchmark constraints (BC) $\boldsymbol{\delta}$ should be relaxed for a given value of α; or what values α can take for certain tolerance level $\boldsymbol{\delta}$. For the given weights w_k^1 and w_k^0, the relative error $\left| \hat{X}_i^C / X_i - 1 \right|$ is a function of α. The second measure is the relative efficiency of the composite calibration estimator for each study variable. The efficiency may be good for some study variables, but not so good for others. We have to find a balance by properly choosing a value for α. The first measure is not that important if consistency on the auxiliary variables is not of interest.

The ad hoc method of choosing a value for α is through plots. From plots of the relative error against α for auxiliary variables x_i ($i = 1, 2, \ldots, p$), we can choose a value range of α such that benchmark constraints $\boldsymbol{\delta}$ are still within our tolerance level. Then from plots of relative efficiency against α for study variables y_j ($j = 1, 2, \ldots, J$), we can narrow down a value or a value range of α that gives reasonable efficiency levels for all study variables.

4.2 Modified Linear Calibration

It is a simple modification over the model-based calibration for a single variable and the linear calibration. Recall that the model-based calibration weights are calibrated to the total of fitted values such that $\sum_s w_k = N$ and $\sum_s w_k \hat{u}_k = \sum_U \hat{u}_k$ (equation 2.1), while linear calibration weights are calibrated to totals of auxiliary variables such that $\sum_s w_k = N$ and $\sum_{k \in s} w_k x_{ik} = \sum_{k \in U} x_{ik}$ $(i = 1, 2, \cdots, p)$ (equation 1.1). The modified linear calibration is simply the mix of these two. First, the set of auxiliary variables is divided into two sets, one for working model of the 'priority' study variable y_1 and the other for linear calibration models of 'non-priority' study variables y_j $(j \neq 1)$. Let $\mathbf{x}_k = (\mathbf{x}^1_k, \mathbf{x}^0_k)$. The calibration equations are:

$$\begin{aligned} &\sum_s w_k \mathbf{x}^0_k = \sum_U \mathbf{x}^0_k \\ &\sum_s w_k \hat{u}(\mathbf{x}^1_k) = \sum_U \hat{u}(\mathbf{x}^1_k) \end{aligned}. \tag{4.7}$$

This method is supposed to reduce the variance of the estimate for the 'priority' study variable under the condition that the working model for fitted values is good enough. However, it can also increase the variances of estimates for other study variables $y_j\ (j \neq 1)$.

5. Simulation Study

The simulation population includes 3000 units, 8 study variables $y_1, y_2, \cdots, y_8$ and 8 auxiliary variables $x_1, x_2, \cdots, x_8$. Let y_1 be our priority study variable and closely related to e^{x_1}. The correlation coefficients among variables are given in Table 1.

Table 1: Correlation Coefficient Matrix

	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	e^{x_1}
y_1	0.16	0.08	0.35	0.14	-0.04	0.04	0.08	0.04	0.93
y_2	-0.02	0.89	0.22	0.08	0.00	0.00	0.03	-0.05	0.10
y_3	0.06	0.19	0.89	0.24	-0.04	-0.02	0.16	0.10	0.32
y_4	0.04	0.05	0.21	0.92	-0.01	0.00	0.05	0.06	0.13
y_5	-0.02	0.00	-0.05	-0.01	0.88	-0.07	-0.02	-0.01	-0.02
y_6	0.07	0.02	-0.01	0.01	-0.08	0.91	-0.01	-0.02	0.04
y_7	0.01	0.00	0.20	0.05	-0.04	0.05	0.68	0.23	0.06
y_8	0.02	-0.04	0.17	0.06	-0.02	0.00	0.16	0.87	0.04

The simulation population is divided into five strata based on a design variable. B=200 stratified simple random samples are selected. For the 'priority' variable y_1, five estimates are calculated:

(1) $\hat{Y}_1^1$ - the linear calibration on $x_1, x_2, \cdots, x_8$;
(2) $\hat{Y}_1^2$ - the model-based calibration on $\hat{u} = \hat{\beta}_0 + \hat{\beta}_1 z_1$, where $z_1 = e^{x_1}$;
(3) $\hat{Y}_1^3$ - the linear calibration on $x_1, x_2, \cdots, x_8$ and z_1, where $z_1 = e^{x_1}$;
(4) $\hat{Y}_1^4$ - the linear calibration on z_1 only, where $z_1 = e^{x_1}$;
(5) $\hat{Y}_1^C$ - Composite calibration estimate, $\hat{Y}_1^C = \alpha \hat{Y}_1^1 + (1-\alpha)\hat{Y}_1^2$

Methods (3) and (4) are the modified linear calibration outlined in section 4.2. Another modified linear calibration estimation could be the linear calibration on $x_1, x_2, \cdots, x_8$ and $\hat{u}$, also outlined in section 4.2. This method makes sense only if method (2) performs well, which is not the case in the simulation. So we do not present it here. The relative efficiency is $\sqrt{MSE(\hat{Y}_1)}\big/Y_1$, where $MSE(\hat{Y}_1) = 100 \times B^{-1} \sum_{i=1}^{B} (\hat{Y}_i - Y)^2$. Figure 2 shows the relative efficiency of the composite calibration estimate of Y_1 vs. α, along with reference lines of relative efficiency from the other four methods. The green line is the composite calibration estimate. As shown in Figure 2, the regular linear calibration estimate on all *x*-variables has the worst efficiency and the model-based calibration on

$\hat{u} = \hat{\beta}_0 + \hat{\beta}_1 e^{x_1}$ also performs poorly. The linear calibration on e^{x_1} only has the best efficiency, while the modified linear calibration on all *x*-variables and e^{x_1} is also a good candidate. The composite calibration estimator performs well when α is small. However, as shown in Figure 3, the composite calibration estimator would give bad efficiency for other study variables when α is too small.

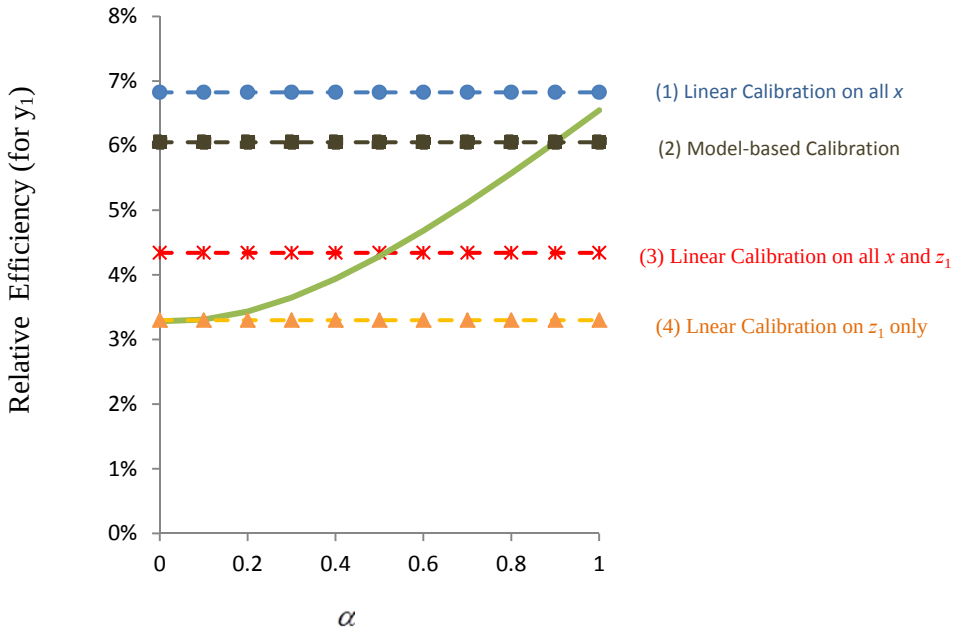

Figure 2: Relative Efficiency of the Composite Calibration Estimator $\hat{Y}_1^C$ vs. α

The relative efficiency for the estimates of two of other study variables y_2 and y_3 are shown in Figure 3. Since methods (2) and (4) perform very poorly, we only compare methods (1), (3) and (5) in Figure 3. As expected, the relative efficiency of the composite calibration estimator gets better as the value of α gets larger. Small values of α that result in good efficiency for $\hat{Y}_1$ would produce poor efficiency for $\hat{Y}_2$ and $\hat{Y}_3$. From the plots in Figure 2 and Figure 3, $\alpha \approx 0.5$ seems to be a reasonable ad hoc choice. But then, method (3), the linear calibration on all *x*-variables and e^{x_1} outperforms the composite calibration when $\alpha \approx 0.5$ because it gives similar efficiency for $\hat{Y}_1$, but better efficiency for $\hat{Y}_2$ and $\hat{Y}_3$.

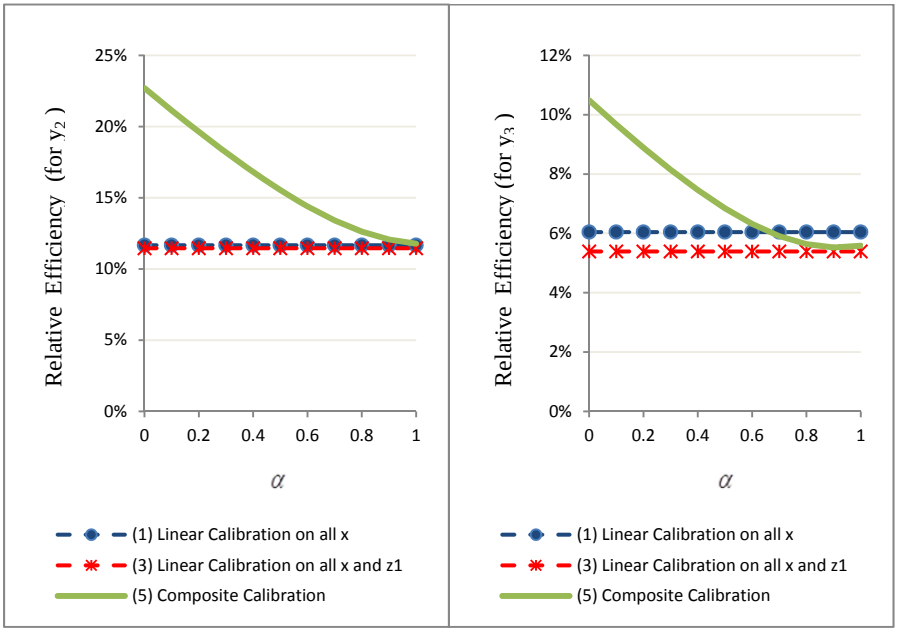

Figure 3: Relative Efficiency of Composite Calibration Estimators $\hat{Y}_2^C$ and $\hat{Y}_3^C$ vs. α

6. Summary

The composite calibration estimator is the weighted form of y_1-specific calibration estimator and a general linear calibration estimator. Here, y_1 is the 'priority' variable for which an efficient estimate is desired. This composite calibration estimator is approximately design-unbiased and multipurpose. It is more efficient for the study variable y_1 than the regular linear calibration estimator, but less efficient for some other study variables. The y_1-specific calibration estimator can be a model-based calibration estimator or linear calibration estimator on a set of selected auxiliary variables. The property of consistency to the known population x-totals may be lost, but this may not be a concern because our goal is not estimating for auxiliary variables.

The choice of the value of α is ad hoc. It can be based on sample data or prior surveys. The α-value balances the efficiency gain in the estimation for the 'priority variable y_1 and the efficiency loss in the estimations for some other study variables. The composite calibration estimator can be extended to the case of more than one 'priority' variable. This is interesting as an academic exercise. In practice, it is probably better off to just use the modified linear calibration estimation such as mothod (3) in section 5.

Based on the limited simulation study, the composite linear calibration has the advantage when a high efficiency for $\hat{Y}_1$ is preferred even though the efficiency of estimates for other variables may not sacrificed at a level we can accept. This decision depends on the practical situation. A modified linear calibration estimator such as method (3) in the simulation shows a very good potential.

References

Bankier, M. D., Rathwell, S. and Majkowski, M. (1992). "Two step generalized least squares estimation in the 1991 Canadian Census," *Proceedings of Survey Research Methods Section*, American Statistical Association, pp. 764–769.

Bankier, M., Houle, A.M. and Luc, M. (1997). "Calibration estimation in the 1991 and1996 Canadian censuses," *Proceedings*, *Section on Survey Research Methods*, American Statistical Association, p. 6675.

Beaumont J.F. and Alavi A. (2004). "Robust generalized regression estimation," *Survey Methodology*, 30, pp. 195-208.

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

Clark, R. G. & Chambers, R. L. (2008). "Adaptive Calibration for Prediction of Finite Population Totals," *Survey Methodology*, 34 (2), pp. 163-172.

Kott, P. (2009), "Calibration weighting: combining probability samples and linear prediction models," in D. Pfeffermann and C. R. Rao (Eds.), *Handbook of Statistics*, *Sample Surveys: Design, Methods and Application,* 29B, Amsterdam: Elsevier BV.

Silva, P.L.D. and Skinner, C.J. (1997). "Variable selection for regression estimation in finite populations," *Survey Methodology, 23,* pp. 23-32.

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

Särndal, C.E., Swensson, B. and Wretman, J. (1992). "The weighted residual technique for estimating the variance of the general regression estimator of the finite population total," *Biometrika*, 76 (3): pp. 527-537.

Särndal, C.E. (2007). "The Calibration Approach in Survey Theory and Practice," *Survey Methodology*, 33, pp. 99–119.

Schaible, W.L. (1978). "Choosing weights for composite estimators for small area statistics," *Proceedings of the Section on Survey Research Methods,* American Statistical Association, Washington, D.C., pp. 741—746.

Wu, C., and Sitter, R.R. (2001). "A model calibration approach to using complete auxiliary information from survey data," *Journal of the American Statistical Association*, 96, pp. 185-193.

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