A New Approach to Composite Estimation for Repeated Surveys with Rotating Panels

Takis Merkouris*

Abstract
Composite estimation in repeated surveys with rotating panels refers to methods of estimation which exploit correlations in the data in the sample overlap between survey times to improve the precision of current estimates. In this article a novel approach to composite estimation is proposed, in which composite regression estimators of current totals for a number of key variables are generated from a simultaneous calibration of the sampling weights of the overlapping samples of the current and previous survey time. In this procedure, in addition to the usual calibration to known population totals, differences of estimates for the key variables based on the full sample and the common sample from the two consecutive times are calibrated to each other. The resulting multivariate composite regression estimator is particularly efficient as the regression coefficients incorporate information from the samples of both survey times. Unlike other composite regression estimators, the proposed estimator does not require micro-matching of data in the common sample, and, therefore, is free of problems of estimation quality associated with it. It is also considerably more practical than other composite regression estimators and the traditional AK-composite estimator.

Key Words: Composite calibration, composite regression estimator, AK-composite estimation, MR-composite estimation, sample overlap.

1. Introduction

Some repeated surveys, typically the Labor Force Surveys, use a sampling design with rotating panels for operational and statistical efficiency. Such design with large overlapping sample between survey times can increase the precision of estimates, especially for those variables for which there is a strong correlation between the values reported by the same units in successive times. Composite estimation refers to estimation methods that use information from the previous times to improve the precision of both the point-in-time ("level") estimates and estimates of change between successive times, by exploiting correlations in the data of the overlapping sample.

The earliest composite estimation method, known as the "K-composite estimation", was introduced for the US Current Population Survey by Hansen et. al (1955), and extended later to the "AK-composite estimation" by Gurney and Daly (1965), and to the "AK-composite weighting"; see Fuller (1990), Cantwell and Ernst (1992), Lent, Miller and Cantwell (1994), Lent, Miller, Cantwell and Duff (1999).

Recently, a type of regression method of composite estimation, called modified regression (MR) composite estimation, and having certain advantages over the AK-method, was developed for the Canadian Labor Force Survey; see Singh and Merkouris (1995), Singh Kennedy and Wu (2001), Gambino, Kennedy and Singh (2001), Fuller and Rao (2001).

In this article, a novel approach to composite estimation is proposed, in which composite regression estimators of current totals for a number of key variables are generated from a simultaneous calibration of the sampling weights of the overlapping samples of the current and previous survey time. In this procedure, in addition to the usual calibration to known population totals of auxiliary variables, differences of estimates for the key variables

*Athens University of Business and Economics, 2 Trias Street, Athens 11362, Greece, merkouris@aueb.gr

based on the full sample and the common sample of the two consecutive surveys are calibrated to each other. The resulting multivariate composite regression estimators of current totals and changes are particularly efficient because the regression coefficients incorporate information from the samples of both current and previous survey time. Furthermore, the simultaneous calibration of the samples of consecutive survey times facilitates greatly variance estimation by resampling methods.

Unlike the MR-composite method, the proposed method of composite regression estimation does not require micro-matching of data in the common sample, and therefore is free of problems of estimation quality associated with it. It is also considerably more practical than the MR-composite estimation and the traditional AK-composite estimation. The merits of the proposed estimator are discussed in more detail in the last section.

2. Notation and Preliminaries

We consider the case of a repeated survey with a rotating panel design. Its sample is made up of a number (say r) of subsamples ("panels", or "rotation groups") of equal size, each one used for r consecutive survey times and then rotated out of the survey and replaced by a newly selected panel. Each panel is a representative sample of units (e.g., dwellings), and so can provide a separate estimate by a proper scaling up of its sampling weights. For any two consecutive survey times there is a partial sample overlap of $100(r-1)/r$ percent, defining the "matched sample". The samples at times t and $t-1$ are denoted by s_t and s_{t-1}, respectively, and the vector of sampling weights at current time t is denoted by $\mathbf{w}_t$.

Let $\mathbf{y}$ be a vector of q key variables to be used in composite estimation, with vector of current-time totals $\mathbf{t_y}$, and let $\mathbf{x}$ be a vector of p auxiliary variables used in calibration, with vector of current-time known totals $\mathbf{t_x}$. Denote then the sample matrix of $\mathbf{y}$, of dimension $n_t \times q$, where n_t is the sample size at time t, by $\mathbf{Y}_t$ partitioned into the unmatched and matched part of the sample, $\mathbf{Y}_{ut}$ and $\mathbf{Y}_{mt}$, respectively. Similar is the notation for the previous survey time $t-1$. The sample matrix of $\mathbf{x}$, of dimension $n_t \times p$, is denoted by $\mathbf{X}_t$.

The current-time Horvitz-Thompson (HT) estimator of the total $\mathbf{t_y}$, based on the full sample, is $\hat{\mathbf{Y}}_t = \mathbf{Y}_t'\mathbf{w}_t$, and $\hat{\mathbf{X}}_t = \mathbf{X}_t'\mathbf{w}_t$ is the HT-estimator of $\mathbf{t_x}$. The estimator of $\mathbf{t_y}$ based on the matched sample is $\hat{\mathbf{Y}}_{mt} = R\mathbf{Y}_{mt}'\mathbf{w}_{mt}$, where $R = r/(r-1)$ is the inverse proportion of matched sample size or a weighted version of it, and $\mathbf{w}_{mt}$ is the subvector of sampling weights of units in the matched sample.

The standard calibration of the weights of the current sample s_t to the population totals $\mathbf{t_x}$ generates the vector of calibrated weights

$$\mathbf{c}_{\mathbf{x}t} = \mathbf{w}_t + \mathbf{W}_t\mathbf{X}_t(\mathbf{X}_t'\mathbf{W}_t\mathbf{X}_t)^{-1}(\mathbf{t_x} - \mathbf{X}_t'\mathbf{w}_t), \quad (1)$$

where $\mathbf{W}_t$ is the diagonal "weighting" matrix with diagonal elements the weights in the vector $\mathbf{w}_t$. Then, the calibration estimator of $\mathbf{t_y}$ is given by $\mathbf{Y}'\mathbf{c_x}$, or in the form of (generalized) regression estimator $\hat{\mathbf{Y}}_t^R$ by

$$\hat{\mathbf{Y}}_t^R = \hat{\mathbf{Y}}_t + \hat{\mathbf{B}}(\mathbf{t_x} - \hat{\mathbf{X}}_t), \quad (2)$$

where $\hat{\mathbf{B}} = \mathbf{Y}_t'\mathbf{W}_t\mathbf{X}_t(\mathbf{X}_t'\mathbf{W}_t\mathbf{X}_t)^{-1}$ is the matrix of regression coefficients. By the calibration property, $\hat{\mathbf{X}}_t^R = \mathbf{t_x}$. The regression estimator of $\mathbf{t_y}$ based on the matched sample, to be used in composite estimation, is

$$\hat{\mathbf{Y}}_{mt}^R = \hat{\mathbf{Y}}_{mt} + \hat{\mathbf{B}}_m(\mathbf{t_x} - \hat{\mathbf{X}}_t),$$

where $\hat{\mathbf{B}}_m = R\mathbf{Y}_{mt}'\mathbf{W}_{mt}\mathbf{X}_{mt}(\mathbf{X}_t'\mathbf{W}_t\mathbf{X}_t)^{-1}$, with the obvious notation for $\mathbf{W}_{mt}$ and $\mathbf{X}_{mt}$.

3. Composite Estimation

3.1 Composite calibration

The proposed composite regression estimation employs an extended calibration scheme which involves the samples of both current and previous survey times. This scheme is specified by the augmented regression matrix, the associated calibration totals and the weights of both samples given, respectively, by

$$\boldsymbol{\mathcal{X}} = \begin{pmatrix} \mathbf{X}_t & \boldsymbol{\Psi}_t \\ \mathbf{0} & -\boldsymbol{\Psi}_{t-1} \end{pmatrix}, \quad \mathbf{t} = \begin{pmatrix} \mathbf{t}_{\mathbf{x}} \\ \mathbf{0} \end{pmatrix}, \quad \mathbf{w} = \begin{pmatrix} \mathbf{w}_t \\ \tilde{\mathbf{w}}_{t-1} \end{pmatrix}, \tag{3}$$

where $\tilde{\mathbf{w}}_{t-1}$ is the vector of composite calibration weights of time $t-1$, $\boldsymbol{\Psi}_t = (\mathbf{Y}'_{ut}, (1-R)\mathbf{Y}'_{mt})'$ and $\boldsymbol{\Psi}_{t-1} = (\mathbf{Y}'_{ut-1}, (1-R)\mathbf{Y}'_{mt-1})'$. The vector of calibrated weights for the combined sample $s_{t-1} \cup s_t$ is given then by

$$\mathbf{c} = \mathbf{w} + \boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{X}}(\boldsymbol{\mathcal{X}}'\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{X}})^{-1}(\mathbf{t} - \boldsymbol{\mathcal{X}}'\mathbf{w}), \tag{4}$$

where $\boldsymbol{\mathcal{W}}$ is the weighting matrix $\text{diag}(\mathbf{W}_t, \tilde{\mathbf{W}}_{t-1})$ associated with $\mathbf{w}$, and

$$\boldsymbol{\mathcal{X}}'\mathbf{w} = \begin{pmatrix} \mathbf{X}'_t\mathbf{w}_t \\ \boldsymbol{\Psi}'\mathbf{w}_t - \boldsymbol{\Psi}'_{t-1}\tilde{\mathbf{w}}_{t-1} \end{pmatrix} = \begin{pmatrix} \hat{\mathbf{X}}_t \\ \hat{\mathbf{Y}}_t - \hat{\mathbf{Y}}_{mt} - (\hat{\mathbf{Y}}^c_{t-1} - \hat{\mathbf{Y}}^c_{mt-1}) \end{pmatrix}.$$

Here $\hat{\mathbf{Y}}^c_{t-1} = \mathbf{Y}'_{t-1}\tilde{\mathbf{w}}_{t-1}$ and $\hat{\mathbf{Y}}^c_{mt-1} = \mathbf{Y}'_{mt-1}\tilde{\mathbf{w}}_{mt-1}$ are the full sample and matched sample composite regression estimates for time $t-1$. The vector $\mathbf{c} = (\mathbf{c}'_t, \mathbf{c}'_{t-1})'$ satisfies the calibration constraints $\boldsymbol{\mathcal{X}}'\mathbf{c} = \mathbf{t}$, i.e., $\mathbf{X}'_t\mathbf{c}_t = \mathbf{t}_{\mathbf{x}}$ and $\boldsymbol{\Psi}'_t\mathbf{c}_t = \boldsymbol{\Psi}'_{t-1}\mathbf{c}_{t-1}$; the second constraint means that the differences in full and matched sample estimates from previous and current time are equated. The first time of employing composite calibration, the estimates $\hat{\mathbf{Y}}^c_{t-1}$ and $\hat{\mathbf{Y}}^c_{mt-1}$ are just the regression estimates $\hat{\mathbf{Y}}^R_{t-1}$ and $\hat{\mathbf{Y}}^R_{mt-1}$ for previous time.

3.2 Estimates of Levels

Now set $\mathbf{Y}_{(t)} = (\mathbf{Y}'_t, \mathbf{0}')'$. Then the extended calibration generates the composite calibration estimator $\mathbf{Y}'_{(t)}\mathbf{c} = \mathbf{Y}'_t\mathbf{c}_t$ of $\mathbf{t}_{\mathbf{y}}$, written in composite regression form as

$$\hat{\mathbf{Y}}^{CR}_t = \hat{\mathbf{Y}}_t + \hat{\boldsymbol{\mathcal{B}}}\left(\mathbf{t} - \hat{\boldsymbol{\mathcal{X}}}\right), \tag{5}$$

where $\hat{\boldsymbol{\mathcal{X}}} = \boldsymbol{\mathcal{X}}'\mathbf{w}$ and $\hat{\boldsymbol{\mathcal{B}}} = \mathbf{Y}'_{(t)}\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{X}}(\boldsymbol{\mathcal{X}}'\boldsymbol{\mathcal{W}}\boldsymbol{\mathcal{X}})^{-1}$ is the $q \times (p+q)$ matrix of regression coefficients. This can be decomposed in the standard and additional regression terms as

$$\hat{\mathbf{Y}}^{CR}_t = \hat{\mathbf{Y}}_t + \hat{\mathbf{B}}_x(\mathbf{t}_{\mathbf{x}} - \hat{\mathbf{X}}_t) + \hat{\mathbf{B}}^c_t\left[\hat{\mathbf{Y}}^c_{t-1} - \hat{\mathbf{Y}}^c_{mt-1} - (\hat{\mathbf{Y}}_t - \hat{\mathbf{Y}}_{mt})\right], \tag{6}$$

where $\hat{\mathbf{B}}_x$ and $\hat{\mathbf{B}}^c_t$ are the partial regression coefficients, components of $\hat{\boldsymbol{\mathcal{B}}}$. Clearly, the estimator $\hat{\mathbf{Y}}^{CR}_t$ is recursive, carrying information from previous surveys to the current survey.

Partitioning the matrix $\boldsymbol{\mathcal{X}}$ by the column submatrices in (3) as $\boldsymbol{\mathcal{X}} = (\mathbf{X}, \boldsymbol{\Psi})$, the vector $\mathbf{c}$ can be decomposed (Merkouris 2004) as

$$\mathbf{c} = \mathbf{c}_{\mathbf{x}} + \mathbf{L}_{\mathbf{x}}\boldsymbol{\Psi}\left(\boldsymbol{\Psi}'\mathbf{L}_{\mathbf{x}}\boldsymbol{\Psi}\right)^{-1}\left(\mathbf{0} - \boldsymbol{\Psi}'\mathbf{c}_{\mathbf{x}}\right), \tag{7}$$

where $\mathbf{c}_{\mathbf{x}} = \mathbf{w} + \boldsymbol{\mathcal{W}}\mathbf{X}(\mathbf{X}'\boldsymbol{\mathcal{W}}\mathbf{X})^{-1}(\mathbf{t}_{\mathbf{x}} - \hat{\mathbf{X}}_t)$ is the vector of calibrated weights based on the regression matrix $\mathbf{X}$, so that $\mathbf{X}'\mathbf{c}_{\mathbf{x}} = \mathbf{t}_{\mathbf{x}}$, and $\mathbf{L}_{\mathbf{x}} = \boldsymbol{\mathcal{W}}(\mathbf{I} - \mathbf{P}_{\mathbf{x}})$, with $\mathbf{P}_{\mathbf{x}} =$

$\mathbf{X}(\mathbf{X}'\boldsymbol{\mathcal{W}}\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\mathcal{W}}$. Note that $\boldsymbol{\Psi}'\mathbf{c} = \mathbf{0}$, this being the partial calibration constraint associated with the differences in the last term of (6).

The vector $\mathbf{c_x}$ can be written analytically as

$$\mathbf{c_x} = \begin{pmatrix} \mathbf{c_x}_t \\ \mathbf{c_x}_{t-1} \end{pmatrix} = \begin{pmatrix} \mathbf{w}_t + \mathbf{W}_t\mathbf{X}_t\left(\mathbf{X}_t'\mathbf{W}_t\mathbf{X}_t\right)^{-1}\left(\mathbf{t_x} - \hat{\mathbf{X}}_t\right) \\ \tilde{\mathbf{w}}_{t-1} \end{pmatrix}. \qquad (8)$$

Then, with reference to (7), straightforward calculations give

$$\mathbf{Y}_{(t)}'\mathbf{c_x} = \hat{\mathbf{Y}}_t^R \qquad (9)$$

and

$$\boldsymbol{\Psi}'\mathbf{c_x} = \hat{\mathbf{Y}}_t^R - \hat{\mathbf{Y}}_{mt}^R - (\hat{\mathbf{Y}}_{t-1}^c - \hat{\mathbf{Y}}_{mt-1}^c), \qquad (10)$$

where $\hat{\mathbf{Y}}_t^R$ and $\hat{\mathbf{Y}}_{mt}^R$ are respectively the full sample and matched sample regression estimators defined in Section 2.

Now, using (7) and (8) the composite calibration estimator $\mathbf{Y}_{(t)}'\mathbf{c}$ can be expressed in the (alternative to (6)) composite regression form

$$\hat{\mathbf{Y}}_t^{CR} = \hat{\mathbf{Y}}_t^R + \hat{\mathbf{B}}_t^c\left(\hat{\mathbf{Y}}_{t-1}^c - \hat{\mathbf{Y}}_{mt-1}^c - (\hat{\mathbf{Y}}_t^R - \hat{\mathbf{Y}}_{mt}^R)\right), \qquad (11)$$

where $\hat{\mathbf{B}}_t^c = \mathbf{Y}_{(t)}'\mathbf{L_x}\boldsymbol{\Psi}\left(\boldsymbol{\Psi}'\mathbf{L_x}\boldsymbol{\Psi}\right)^{-1}$ is the partial regression coefficient in (6). A more explicit expression of $\hat{\mathbf{B}}_t^c$ is derived upon noting that

$$\mathbf{L_x} = \begin{pmatrix} \mathbf{L}_{\mathbf{x}_t} & \mathbf{0} \\ \mathbf{0} & \tilde{\mathbf{W}}_{t-1} \end{pmatrix}, \qquad (12)$$

where $\mathbf{L}_{\mathbf{x}_t} = \mathbf{W}_t(\mathbf{I} - \mathbf{P}_{\mathbf{x}_t})$, and $\mathbf{P}_{\mathbf{x}_t} = \mathbf{X}_t(\mathbf{X}'\mathbf{W}_t\mathbf{X}_t)^{-1}\mathbf{X}_t'\mathbf{W}_t$. It follows then that $\hat{\mathbf{B}}_t^c = \mathbf{Y}_t'\mathbf{L}_{\mathbf{x}_t}\boldsymbol{\Psi}_t\left(\boldsymbol{\Psi}_t'\mathbf{L}_{\mathbf{x}_t}\boldsymbol{\Psi}_t + \boldsymbol{\Psi}_{t-1}'\tilde{\mathbf{W}}_{t-1}\boldsymbol{\Psi}_{t-1}\right)^{-1}$.

Expression (11) of the composite regression estimator $\hat{\mathbf{Y}}_t^{CR}$ allows a direct comparison with the current-time regression estimator $\hat{\mathbf{Y}}_t^R$, separating the effect of incorporating previous-time information. We can write (11) alternatively, in a more interpretative form, as

$$\hat{\mathbf{Y}}_t^{CR} = (\mathbf{I} - \hat{\mathbf{B}}_t^c)\hat{\mathbf{Y}}_t^R + \hat{\mathbf{B}}_t^c\left(\hat{\mathbf{Y}}_{t-1}^c + \hat{\mathbf{Y}}_{mt}^R - \hat{\mathbf{Y}}_{mt-1}^c\right). \qquad (13)$$

Equation (13) shows that the composite regression estimator $\hat{\mathbf{Y}}_t^{CR}$ is a linear combination of the current-time regression estimator and the previous-time composite regression estimator updated with the change estimator based on the matched sample.

The simultaneous calibration of the two samples results also in an updated estimator for the previous time, incorporating information from current time. Setting $\mathbf{Y}_{(t-1)} = (\mathbf{0}', \mathbf{Y}_{t-1}')'$, we obtain the updated calibration estimator $\mathbf{Y}_{(t-1)}'\mathbf{c}$ in the composite regression form, similar to (13),

$$\hat{\mathbf{Y}}_{t-1}^{CR} = (\mathbf{I} - \hat{\mathbf{B}}_{t-1}^c)\hat{\mathbf{Y}}_{t-1}^c + \hat{\mathbf{B}}_{t-1}^c\left(\hat{\mathbf{Y}}_t^R - \left(\hat{\mathbf{Y}}_{mt}^R - \hat{\mathbf{Y}}_{mt-1}^c\right)\right), \qquad (14)$$

where $\hat{\mathbf{B}}_{t-1}^c = \mathbf{Y}_{t-1}'\tilde{\mathbf{W}}_{t-1}\boldsymbol{\Psi}_{t-1}\left(\boldsymbol{\Psi}_t'\mathbf{L}_{\mathbf{x}_t}\boldsymbol{\Psi}_t + \boldsymbol{\Psi}_{t-1}'\tilde{\mathbf{W}}_{t-1}\boldsymbol{\Psi}_{t-1}\right)^{-1}$. This shows that the updated composite regression estimator $\hat{\mathbf{Y}}_{t-1}^{CR}$ is a linear combination of the initial previous-time composite regression estimator and the current-time regression estimator reduced by the change estimator based on the matched sample.

It is worth emphasizing that in the simultaneous calibration of the previous and current time samples that generates the composite estimator, the differences in estimates based on full and matched samples from previous and current time (see 11) are calibrated to each other. This is due to the partial calibration constraint $\mathbf{\Psi}'\mathbf{c} = \mathbf{0}$ noted above. Writing $\mathbf{\Psi}'\mathbf{c} = \mathbf{\Psi}_t'\mathbf{c}_t - \mathbf{\Psi}_{t-1}'\mathbf{c}_{t-1}$ we easily verify that

$$\mathbf{\Psi}_t'\mathbf{c}_t = \mathbf{\Psi}_{t-1}'\mathbf{c}_{t-1} = (\mathbf{I} - \hat{\mathbf{B}}_d^c)\left(\hat{\mathbf{Y}}_t^R - \hat{\mathbf{Y}}_{mt}^R\right) + \hat{\mathbf{B}}_d^c\left(\hat{\mathbf{Y}}_{t-1}^c - \hat{\mathbf{Y}}_{mt-1}^c\right), \quad (15)$$

where $\hat{\mathbf{B}}_d^c = \mathbf{\Psi}_t'\mathbf{L}_{\mathbf{x}_t}\mathbf{\Psi}_t\left(\mathbf{\Psi}_t'\mathbf{L}_{\mathbf{x}_t}\mathbf{\Psi}_t + \mathbf{\Psi}_{t-1}'\tilde{\mathbf{W}}_{t-1}\mathbf{\Psi}_{t-1}\right)^{-1}$. This shows that the calibration equates both differences $\hat{\mathbf{Y}}_t^R - \hat{\mathbf{Y}}_{mt}^R$ and $\hat{\mathbf{Y}}_{t-1}^c - \hat{\mathbf{Y}}_{mt-1}^c$ to their combination in (15). Since $\mathbf{\Psi}_t'\mathbf{c}_t = \hat{\mathbf{Y}}_t^{CR} - \hat{\mathbf{Y}}_{mt}^{CR}$ and $\mathbf{\Psi}_{t-1}'\mathbf{c}_{t-1} = \hat{\mathbf{Y}}_{t-1}^{CR} - \hat{\mathbf{Y}}_{mt-1}^{CR}$, the composite calibration results in the equality

$$\hat{\mathbf{Y}}_t^{CR} - \hat{\mathbf{Y}}_{mt}^{CR} = \hat{\mathbf{Y}}_{t-1}^{CR} - \hat{\mathbf{Y}}_{mt-1}^{CR}. \quad (16)$$

3.3 Estimates of Change

It follows from (16) that

$$\hat{\mathbf{Y}}_t^{CR} - \hat{\mathbf{Y}}_{t-1}^{CR} = \hat{\mathbf{Y}}_{mt}^{CR} - \hat{\mathbf{Y}}_{mt-1}^{CR}, \quad (17)$$

which means that the estimate of change based on the full samples s_t, s_{t-1} is equal to the estimate of change based on the matched sample. Interestingly,

$$\hat{\mathbf{Y}}_t^{CR} = \hat{\mathbf{Y}}_{t-1}^{CR} + \hat{\mathbf{Y}}_{mt}^{CR} - \hat{\mathbf{Y}}_{mt-1}^{CR},$$

which shows that the composite regression estimate at time t is simply the updated composite regression estimate at time $t-1$ plus the change estimate based on the matched sample at times t and $t-1$.

The change estimate in (17) involves the updated previous-time estimates $\hat{\mathbf{Y}}_{t-1}^{CR}$ and $\hat{\mathbf{Y}}_{mt-1}^{CR}$. On the other hand, if the initial previous-time estimates $\hat{\mathbf{Y}}_{t-1}^c$ and $\hat{\mathbf{Y}}_{mt-1}^c$ are used, then it follows easily from (13) that the estimate of change $\hat{\mathbf{Y}}_t^{CR} - \hat{\mathbf{Y}}_{t-1}^c$ can be expressed as the combination of full-sample and matched-sample estimates of change

$$\hat{\mathbf{Y}}_t^{CR} - \hat{\mathbf{Y}}_{t-1}^c = \left(\mathbf{I} - \hat{\mathbf{B}}_t^c\right)\left(\hat{\mathbf{Y}}_t^R - \hat{\mathbf{Y}}_{t-1}^c\right) + \hat{\mathbf{B}}_t^c\left(\hat{\mathbf{Y}}_{mt}^R - \hat{\mathbf{Y}}_{mt-1}^c\right),$$

and conveniently obtained as $\mathbf{Y}_t'\mathbf{c}_t - \mathbf{Y}_{t-1}'\tilde{\mathbf{w}}_{t-1}$.

3.4 Special Cases of Composite Estimation

Expression (13) gives the composite regression estimator $\hat{\mathbf{Y}}_t^{CR}$ in multivariate form for all components of the vector $\mathbf{y}$. It follows easily from (13) that for any of the q components of $\mathbf{y}$, say y_g, the composite regression estimator of its total is

$$\begin{aligned}\hat{Y}_{gt}^{CR} = (1 - \hat{\beta}_g^c)\hat{Y}_{gt}^R \quad &+ \quad \hat{\beta}_g^c\left(\hat{Y}_{gt-1}^c + \hat{Y}_{gmt}^R - \hat{Y}_{gmt-1}^c\right) \\ &+ \quad \hat{\boldsymbol{\beta}}_{\bar{g}}^c\left(\hat{\mathbf{Y}}_{\bar{g}t-1}^c + \hat{\mathbf{Y}}_{\bar{g}mt}^R - \hat{\mathbf{Y}}_{\bar{g}mt-1}^c\right)\end{aligned}, \quad (18)$$

where $\hat{\beta}_g^c$ is the g-th diagonal element of $\hat{\mathbf{B}}_t^c$, $\hat{\boldsymbol{\beta}}_{\bar{g}}^c$ is the g-th row vector of $\hat{\mathbf{B}}_t^c$ without the g-th element, and the quantities in the last bracket of (18) are the indicated vector estimators for the other $q-1$ components of $\mathbf{y}$. Thus, although the composite estimator $\hat{Y}_{gt}^{CR}$ incorporates all information on y_g available in the two overlapping samples, in the manner indicated by the linear combination in the first two terms of (18), the additional third term suggests that $\hat{Y}_{gt}^{CR}$ may realize additional efficiency due to correlation of y_g with the rest of the components of $\mathbf{y}$. Of course, $\hat{Y}_{gt}^{CR}$ can be conveniently obtained as calibration estimator $\mathbf{Y}'_{g(t)}\mathbf{c}$, where $\mathbf{Y}'_{gt}$ is the g-th column of $\mathbf{Y}'_{(t)}$.

Let now z be any other single variable, with total t_z and current-sample matrix $\mathbf{Z}_t$ of dimension $n_t \times 1$. Setting $\mathbf{Z}_{(t)} = (\mathbf{Z}'_t, \mathbf{0}')'$ and using (7) we obtain the composite calibration estimator $\mathbf{Z}'_{(t)}\mathbf{c}$ of t_z in composite regression form, analogous to (11), as

$$\hat{Z}_t^{CR} = \hat{Z}_t^{R} + \hat{\mathbf{B}}_z^c \left(\hat{\mathbf{Y}}_{t-1}^c - \hat{\mathbf{Y}}_{mt-1}^c - (\hat{\mathbf{Y}}_t^R - \hat{\mathbf{Y}}_{mt}^R) \right), \tag{19}$$

where $\hat{\mathbf{B}}_z^c = \mathbf{Z}'_{(t)}\mathbf{L}_{\mathbf{x}}\mathbf{\Psi}\left(\mathbf{\Psi}'\mathbf{L}_{\mathbf{x}}\mathbf{\Psi}\right)^{-1} = \mathbf{Z}'_t\mathbf{L}_{\mathbf{x}_t}\mathbf{\Psi}_t\left(\mathbf{\Psi}'_t\mathbf{L}_{\mathbf{x}_t}\mathbf{\Psi}_t + \mathbf{\Psi}'_{t-1}\tilde{\mathbf{W}}_{t-1}\mathbf{\Psi}_{t-1}\right)$, and $\hat{Z}^R$ is the regression estimator of the form (2). It is clear that the efficiency of the composite regression estimator $\hat{Z}_t^{CR}$ relative to the standard regression estimator $\hat{Z}_t^R$ depends on the strength of correlation of z with $\mathbf{y}$.

4. Discussion

This section presents a summary of properties and relative merits of the proposed approach to composite estimation.

The proposed composite estimators for levels and changes derive their efficiency from the fact that the coefficient $\hat{\mathbf{B}}_t^c$ incorporates information from both previous-and-current time samples. Supporting this argument is the observation that the variance minimizing coefficient in (11) is a function of estimates from both survey times. Note that replacing in equation (5) the weighting matrix $\boldsymbol{\mathcal{W}}$ in $\hat{\boldsymbol{\mathcal{B}}}$ by the estimated variance $\widehat{\text{Var}}(\mathbf{w})$ gives the optimal (asymptotically variance minimizing) coefficient $\hat{\boldsymbol{\mathcal{B}}}^o = \widehat{\text{Cov}}(\hat{\mathbf{Y}}_t, \hat{\boldsymbol{\mathcal{X}}})[\widehat{\text{Var}}(\hat{\boldsymbol{\mathcal{X}}})]^{-1}$, which clearly is a function of data from both survey times. The regression coefficient $\hat{\boldsymbol{\mathcal{B}}}$ – and more specifically the partial regression coefficient $\hat{\mathbf{B}}_t^c$ – as approximation to the optimal coefficient is also a function of data from both survey times. In contrast, the MR-composite regression estimator is generated by a calibration in which current-time estimates are calibrated to previous-times estimates, the latter being treated as constants in calibration, and thus the regression coefficient incorporates information from current time only. An empirically chosen tuning constant $\alpha \in (0, 1)$, multiplying $\hat{\mathbf{B}}_t^c$, could provide a balance between the improvement of level and change estimates for important variables, by giving more weight to one of the two terms of (13).

The proposed composite estimation is free of problems with the matching of the sample between two consecutive times at the individual record level, as required in the MR-composite estimation. These problems arise when, for a given matched sample, data is available only for one survey time. This may occur due to nonresponse in either survey time or when a move or change in scope has taken place between the two consecutive survey times; see Gambino et. al. (2001). Biases that remain after the treatment of these problems, and which can be accumulated over time due to the recursive nature of the composite estimator, are avoided in the proposed estimation procedure. The proposed method is also free of operational complexities of the MR-composite estimation, which include the extra calibration of past-month data to the current-month population totals, and the cumbersome variance estimation by resampling methods due to the random calibration totals; see Statistics Canada (2017).

The composite regression estimator (11) has the form of the traditional K-composite estimator, with the time-dependent regression coefficient $\hat{\mathbf{B}}_t^c$ in place of the coefficient K. In K-composite and AK-composite estimation, values of A and K that are optimal over time in the sense of minimum variance of the estimator are empirically chosen for each variable of interest. In contrast, the proposed composite estimator, with the time-dependent matrix coefficient $\hat{\mathbf{B}}_t^c$, is multivariate and thus the efficiency of estimation for each of the components of $\mathbf{y}$ may be enhanced by the correlation with other components, as indicated by (18). Also unlike AK-composite estimation, in which only the most important estimates are true composite estimates, the proposed composite calibration generates composite regression estimates for any variable, as shown in (19). Operationally, unlike the AK-estimation, where weighting to satisfy known population totals and composite estimation are separate steps, calibration weighting in the proposed composite regression estimation is done in one step, i.e., simultaneously with weighting to satisfy the standard calibration constraints.

An extension of the composite regression estimator $\hat{\mathbf{Y}}_t^{CR}$, analogous to the extension of the K-composite to the AK-composite estimator, could involve the additional regression term $\hat{\mathbf{Y}}_{ut} - \hat{\mathbf{Y}}_{mt}$ in (6), where $\hat{\mathbf{Y}}_{ut} = r\mathbf{Y}_{ut}'\mathbf{w}_{ut}$ is the estimate of $\mathbf{t}_\mathbf{y}$ based on the unmatched ("birth") panel at time t. This is done by augmenting the matrix $\boldsymbol{\mathcal{X}}$ in (3) by the column $(\bar{\boldsymbol{\Psi}}_t', \mathbf{0}')'$, where $\bar{\boldsymbol{\Psi}}_t = -r\boldsymbol{\Psi}_t$, and using the vector of calibration totals $(\mathbf{t_x}', \mathbf{0}', \mathbf{0}')'$. The extended calibration corresponding to this extended regression estimation will result in $\hat{\mathbf{Y}}_{ut}^{CR} = \hat{\mathbf{Y}}_{mt}^{CR}$, which may help to reduce the birth rotation bias due to the usual difference of the birth panel from the other panels.

The performance of the proposed composite regression estimator can be assessed through an extensive empirical study using actual data from a repeated survey with rotating panels (e.g., data from a Labour Force Survey). This estimator should be evaluated for a number of survey characteristics using data over a sufficient period of time, and its advantages should be judged not only on its statistical efficiency but also on its impact on various time series, with respect to their stability and seasonal adjustment.

REFERENCES

Cantwell, P.J., and Ernst, L.R. (1992), "New developments composite estimation for the Current Population Survey," Proceedings: Symposium 92, Design and Analysis of Longitudinal Surveys, Statistics Canada, 121–130.

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

Fuller, W.A., and Rao, J.N.K. (2001), "A regression composite estimator with application to the Canadian Labour Force Survey," *Survey Methodology*, 27, 45–51.

Gambino, J., Kennedy, B., and Singh, M.P. (2001), "Regression composite estimation for the Canadian Labour Force Survey: Evaluation and implementation," *Survey Methodology*, 27, 65–74.

Hansen, M.H., Hurwitz, W.N., Nisselson H. and Steinberg, J. (1955), "The redesign of the census current population survey," *Journal of the American Statistical Association*, 50, 701–719

Lent, J., Miller, S. and Cantwell, P. (1994), "Composite weights for the Current Population Surveys," Proceedings of the Survey Research Methods Section, American Statistical Association, 867–872.

Lent, J., Miller, S.M., Cantwell, P.J. and Duff, M. (1999), "Effect of composite weights on some estimates from the Current Population Survey," *Journal of Official Statistics*, 14, 431–448.

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

Singh, A.C., and Merkouris, P. (1995), "Composite estimation by modified regression for repeated surveys," Proceedings of the Survey Research Methods Section, American Statistical Association, 420–425.

Singh, A.C., Kennedy, B. and Wu, S. (2001), "Regression composite estimation for the Canadian Labour Force Survey with a rotating panel design," *Survey Methodology*, 27, 33–44.

Statistics Canada (2017), "Methodology of the Canadian Labour Force Survey," Statistics Canada, Catalogue no. 71-526-X.