A Regression Composite Estimator with Application to the Canadian Labour Force Survey

Wayne A. Fuller and J.N.K. Rao

Abstract

The Canadian Labour Force Survey is a monthly survey of households selected according to a stratified multistage design. The sample of households is divided into six panels (rotation groups). A panel remains in the sample for six consecutive months and then is dropped from the sample. In the past, a generalized regression estimator, based only on the current month’s data, has been implemented with a regression weights program. In this paper, we study regression composite estimation procedures that make use of sample information from previous periods and that can be implemented with a regression weights program. Singh (1996) proposed a composite estimator, called MR2, which can be computed by adding \(x\)-variables to the current regression weights program. Singh’s estimator is considerably more efficient than the generalized regression estimator for one-period change, but not for current level. Also, the estimator of level can deviate from that of the generalized regression estimator by a substantial amount and this deviation can persist over a long period. We propose a “compromise” estimator, using a regression weights program and the same number of \(x\)-variables as MR2, that is more efficient for both level and change than the generalized regression estimator based only on the current month data. The proposed estimator also addresses the drift problem and is applicable to other surveys that employ rotation sampling.

Key Words: Survey sampling; Rotating samples; Combining estimators.

1. Introduction

Composite estimation is a term used in survey sampling to describe estimators for a current period that use information from previous periods of a periodic survey with a rotating design. When some units are observed in some of the periods, but not in all periods, it is possible to use this fact to improve estimates for all time periods.

Statistics Canada, U.S. Bureau of the Census and some other statistical agencies use a rotating design for labour force surveys. The current Canadian Labour Force Survey (LFS) is a monthly survey of about 59,000 households, which are selected according to a stratified multistage sampling design. The ultimate sampling unit is the household and a sample of households is divided into six panels (rotation groups). A rotation group remains in the sample for six consecutive months and is then dropped from the sample completely. Thus five-sixths of the sample of households is common between two consecutive months.

Singh, Drew, Gambino and Mayda (1990) and Gambino, Singh, Dufour, Kennedy and Lindeyer (1998) contain detailed descriptions of the LFS design. In the U.S. Current Population Survey (CPS), the sample is composed of eight rotation groups. A rotation group stays in the sample for four consecutive months, leaves the sample for the succeeding eight months, and then returns for another four consecutive months. It is then dropped from the sample completely. Thus there is a 75 percent month-to-month sample overlap and a 50 percent year-to-year sample overlap (Hansen, Hurwitz, Nisselson and Steinberg 1955).

Patterson (1950), following the initial work by Jessen (1942), provided the theoretical foundations for design and estimation for repeated surveys, using generalized least squares procedures. For the CPS, Hansen et al. (1955) proposed a simpler estimator, called the \(K\)-composite estimator. Gurney and Daly (1965) presented an improvement to the \(K\)-composite estimator, called the \(AK\)-composite estimator with two weighting factors \(A\) and \(K\). Breau and Ernst (1983) compared alternative estimators to the \(K\)-composite estimator for the CPS. Rao and Graham (1964) studied optimal replacement schemes for the \(K\)-composite estimator. Eckler (1955) and Wolter (1979) studied two-level rotation schemes such as the one used in the U.S. Retail Trade Survey. Yansaneh and Fuller (1998) studied optimal recursive estimation for repeated surveys. Fuller (1990) and Lent, Miller, Cantwell and Duff (1999) developed the method of composite weights for the CPS. The composite weights are obtained by raking the design weights to specified control totals that included population totals of auxiliary variables and \(K\)-composite estimates for characteristics of interest, \(y\). Using the composite weights, users can generate estimates from microdata files for the current month without recourse to data from previous months.

The above authors used the traditional design-based approach, assuming the unknown totals on each occasion to be fixed parameters. Other authors (Scott, Smith and Jones 1977; Jones 1980; Binder and Dick 1989; Bell and Hillmer 1990; Tiller 1989 and Pfeffermann 1991) developed estimates for repeated surveys under the assumption that the underlying true values constitute a realization of a time series.

Statistics Canada considered \(K\) and \(AK\) composite estimation for the Labour Force Survey at several times during the past 25 years (Kumar and Lee 1983), but did not...
adopt composite estimation. Instead, a generalized regression estimator, based only on the current months data, has been computed with a regression weights program. When composite estimation was considered in the 1990’s, there was strong pressure to develop a composite estimation procedure that used the existing estimation program. Singh (1996) proposed an ingenious method, called Modified Regression (MR), to address this issue. This method leads to

\[ \hat{N}_t = \sum_{i \in A} w_i + \sum_{i \in B} w_i = \hat{N}_t = \text{estimated population total}. \]

Let \( \theta_t \) be the fraction of the sample in the overlap at time \( t \):

\[ \theta_t = \hat{N}_t^{-1} \sum_{i \in A} w_i. \]

In the Labour Force Survey \( \theta_t \) is about 5/6 and is nearly constant over time. We will frequently omit the subscript \( t \) on \( A_t, B_t \) and \( \theta_t \), for simplicity.

### 2.1 Estimator

Singh’s (1996) MR2 estimator uses the control variable

\[ x_{ti} = \theta_t^{-1}(y_{t-1,i} - y_{ti}) + y_{ti} \quad \text{if} \quad i \in A_t \]

\[ = y_{ti} \quad \text{if} \quad i \in B_t, \]

in the regression program, where \( y_{ti} \) is the value of a characteristics of interest, \( y_t \) for element \( i \) at time \( t \). Because of nonresponse in the LFS, Singh’s original proposal used imputation for missing data and set \( \theta = 5/6 \), after imputation for missing data. In our initial discussion we use the \( \theta_1 \) as defined in (2.1), assuming no nonresponse so that imputation is not required. Note that “micromatching” of individual data files at \( t-1 \) and \( t \) is needed to calculate \( x_{ti} \), and the resulting MR2 estimator. Additional control variables of the form (2.2) associated with other \( y \)-variables as well as auxiliary variables with known population totals are also included in the regression estimation. The auxiliary variables in the LFS include demographic variables such as age, sex and location.

The particular \( x \)-variables in (2.2) is designed such that the estimated total of \( x_i \) is an estimator of the previous period total of \( y \). Thus, the control total for \( x_i \) in the regression procedure is the previous period estimator of the total of \( y \).

Let \( \hat{\mu}_{t-1} \) be the estimator of the mean of \( y \) for period \( t-1 \), let \( \bar{y}_{m,t-1} \) and \( \bar{y}_{mt} \) be the means of the matched panels at time \( t-1 \) and \( t \) respectively, let \( \bar{y}_{t,i} \) be the grand mean of all sample panels at time \( t \) and let \( \bar{y}_{bi} \) be the mean of the birth panel at time \( t \). Assume the sample of size \( n \) is divided into \( g \) panels of equal size and denote the matched sampling fraction by \( \theta \). To simplify the discussion we consider a single \( y \)-variable. Then Singh’s (1996) MR2 estimator at time \( t \), constructed with \( x_{ti} \), can be written in a regression estimator form as

\[ \hat{\mu}_t = \bar{y} + (x_{CNt} - \bar{x}_{Ct})b_{Ct} + \left[ \hat{\mu}_{t-1} - (\bar{y}_{m,t-1} - \bar{y}_{mt} + \bar{y}) \right]b_i, \]

where \( x_{CNt} \) is the population mean of the vector of auxiliary variables, such as age and sex, at time \( t \), \( \bar{x}_{Ct} \) is the weighted sample mean of the auxiliary variables, and \( (b_{Ct}, b_i) \) is the vector of regression coefficients for the regression of \( y_t \) on \( (x_{Ct}, x_{ti}) \).

One can write

\[ y_{t,i} = \hat{y}_{t,i} + d_{t,i}, \]

Statistics Canada, Catalogue No. 12-001
where \( \hat{y}_{t,i} \) is the predicted value in the regression of \( y_{t,i} \) on \( x_{Ci} \) and \( d_{t,i} \) is the deviation from the regression predicted value. Then

\[
x_{it} = \theta^{-1}(\hat{y}_{t-1,i} - \hat{d}_{t,i} - d_{t,i}) + \hat{d}_{t,i}
\]

Thus the partial regression coefficient \( b_j \) is close to the regression coefficient for the regression of \( d_{t,i} \) on \( x_{Ci} \) and the value depends on the correlation between \( d_{t,i} \) and \( d_{t-1,i} \). A simple model for \( d_{t,i} \) that has been used in the past, and the one we adopt in our analysis, is the assumption that the \( d_{t,i} \) is the sum of a fixed \( \mu_i \) and an error that is a first order autoregression with parameter \( \rho \).

To simplify the presentation, we discuss the simple random sampling model without \( x_{Ci} \). The results extend to the general case by considering the parameter \( \rho \) to be the partial correlation between \( y_t \) and \( y_{t-1} \) after adjusting for \( x_{Ci} \).

Under the autoregressive model with fixed \( \rho \), an intercept and no other \( x_{Ci} \) in the model, it can be shown that \( b_0 \) converges in probability to

\[
b_0 = \rho \lim_{n \to \infty} \lambda_1(n) \rho - (1 - \rho) \sigma_x^2 \Delta^2 \]

where \( \Delta^2 = (\mu_i - \mu_{-i})^2 \). Assuming \( \sigma_x^2 \Delta^2 \) is small relative to the other terms we get

\[
b_0 = \rho \lim_{n \to \infty} \lambda_1(n) \rho - (1 - \rho)^2 \rho \]  \( \text{(2.4)} \)

Thus the partial correlation coefficient \( b_j \) is close to the regression coefficient for the regression of \( d_{t,i} \) on \( x_{Ci} \) and the value depends on the correlation between \( d_{t,i} \) and \( d_{t-1,i} \). A simple model for \( d_{t,i} \) that has been used in the past, and the one we adopt in our analysis, is the assumption that the \( d_{t,i} \) is the sum of a fixed \( \mu_i \) and an error that is a first order autoregression with parameter \( \rho \).

For the LFS, \( b_0 = (7 - 2\rho)^{-1} \) \( \sigma_x \). Alternative representations for the estimator \( \hat{\mu}_i \), omitting \( x_{Ci} \), are obtained using the formula \( \overline{y}_t = \theta \overline{y}_{mt} + (1 - \theta) \overline{y}_{B,t} \). Thus

\[
\hat{\mu}_i = -(1 - b) \overline{y}_t + \left[ \hat{\mu}_{-i} + (\overline{y}_{mt} - \overline{y}_{m,t-1}) \right] \]

The first expression on the right of the equality of (2.5) gives the MR2 estimator as a linear combination of the direct estimator \( \overline{y}_t \) and the difference estimator \( \hat{\mu}_{t-1} + (\overline{y}_{mt} - \overline{y}_{m,t-1}) \) \textit{i.e.}, in the form of a composite estimator. The final expression of (2.5) gives the estimator as a linear combination of a “regression-type” estimator based on the overlap panels and the mean of the birth panels.

2.2 An Alternative Estimator

It is possible to define alternative regression variables to use in regression composite estimation. We present a particular regression variable in this subsection. The associated regression estimator is not suggested as the ultimate estimator, but the estimator is a member of a class for which efficiency calculations are given. An alternative to Singh’s (1996) MR2 estimator is outlined in section 5.

Define a variable to be equal to the previous period value if the individual is in the overlap sample and to be equal to the estimated mean for the previous period if the individual is in the birth sample. The regression variable is

\[
x_{t,0} = y_{t-1,i} \quad \text{if} \quad i \in A_i \]

The regression estimator constructed with \( x_{t,0} \) and recall that the control mean of \( y_{x,t} \) is \( \hat{\mu}_{i-1} \). The regression estimator using \( x_{t,0} \) can be written

\[
\hat{\mu}_{reg,t} = \overline{y}_t + (\hat{\mu}_{t-1} - \overline{y}_{x,t}) \hat{\beta}_t \]  \( \text{(2.8)} \)

where \( \hat{\beta}_t \) is the regression coefficient for the regression of \( y_t \) on \( x_{t,0} \) (subscript \( i \) is dropped on \( \hat{\beta}_i \) for simplicity), \( \overline{y}_t \) is the sample mean of \( y \) at time \( t \), and \( \overline{y}_{x,t} \) is the sample mean of \( x_{t,0} \) for all sample panels at time \( t - 1 \). The regression coefficient \( \hat{\beta}_t \) is, approximately, the regression of \( y_t \) on \( x_{t,0} \) in the set \( A \). The coefficient is not exactly the regression coefficient for the set \( A \) because \( \overline{y}_{m,t-1} \) is not equal to \( \hat{\mu}_{i-1} \), but the difference between the two estimators will usually be small. Singh (1996) called the regression estimator constructed with \( \overline{y}_{x,t} \), the MR1 estimator.

Using \( \overline{y}_t = \theta \overline{y}_{mt} + (1 - \theta) \overline{y}_{B,t} \) the regression estimator of \( \mu_t \) using \( x_{t,0} \) as a control variable is given by

\[
\hat{\mu}_t = (1 - \theta) \overline{y}_{B,t} + \theta \left[ \overline{y}_{mt} + (\hat{\mu}_{t-1} - \overline{y}_{m,t-1}) \hat{\beta}_t \right] \]  \( \text{(2.9)} \)
The expression within curly brackets in (2.9) is the regression estimator of \( \mu_t \) using the estimator \( \hat{\mu}_{t-1} \) and only the data from the matched sample \( A \). Note that the regression estimator

\[
\hat{\mu}_{m,t} = \bar{y}_{m,t} + (\hat{\mu}_{t-1} - \bar{y}_{m,t-1}) \hat{\beta},
\tag{2.10}
\]

where \( \hat{\beta} \) is the regression of \( y_t \) on \( y_{t-1} \) in the set \( A \), is the optimal estimator for \( \mu_t \) based on \( \hat{\mu}_{t-1} \) and the data of set \( A \). Note that \( \beta = \rho \) if the variances are the same at the two time periods. Hereafter, we often set \( \beta = \rho \).

Using the variable \( x_2 \), gives the optimal estimator, \( \hat{\mu}_{mt} \), based on data in set \( A \), but it does not combine that estimator with the mean of set \( B \) in an optimal way. As can be seen in (2.10), the weight given to the mean of set \( B \) is \( 1 - \theta \). In general, this weight is too large because the variance of the regression estimator is less than the variance of the simple mean.

### 3. Optimal Estimation

The way in which one chooses to combine the regression estimator for set \( A \) with the mean of set \( B \) depends on one’s objective function and on the variance of \( \hat{\mu}_{t-1} \). We give some illustrative calculations based on some simplifying assumptions. For convenience let \( V(\hat{\mu}_{t-1}) \) be expressed as a multiple of the variance of the birth panel,

\[
V(\hat{\mu}_{t-1}) = q_{r-1} V(\bar{y}_{B,t}).
\tag{3.1}
\]

Assume

\[
V(\bar{y}_{t}) = g^{-1} V(\bar{y}_{B,t}),
\tag{3.2}
\]

\[
\text{Cov}(\hat{\mu}_{t-1}, (\bar{y}_{m,t} - \bar{y}_{m,t-1}) \hat{\beta}) = 0,
\tag{3.3}
\]

\[
\text{Cov}(\hat{\mu}_{t-1}, \bar{y}_{B,t}) = 0,
\tag{3.4}
\]

and

\[
\text{Cov}(\bar{y}_{B,t}, (\bar{y}_{m,t} - \bar{y}_{m,t-1}) \hat{\beta}) = 0,
\tag{3.5}
\]

where \( g \) is the number of rotation groups (panels). Assumption (3.1) is reasonable if the original panels have a covariance function well approximated by that of a first order autoregressive process. For the LFS, the zero covariances in (3.4) and (3.5) and assumption (3.2) are only approximations because \( \bar{y}_{B,t} \) is not based on an entirely independent sample.

We write the regression estimator based on the overlap as

\[
\hat{\mu}_{m,t} = \bar{y}_{m,t} + (\hat{\mu}_{t-1} - \bar{y}_{m,t-1}) \hat{\beta},
\]

and, with the assumptions, obtain

\[
V(\hat{\mu}_{m,t}) = [g^{-1} \theta^{-1} (1 - \rho^2) + q_{r-1} \rho^2] V(\bar{y}_{B,t}).
\tag{3.6}
\]

For the LFS, \( g = 6 \) is the number of panels. Now consider an estimator that is a linear combination of \( \hat{\mu}_{m,t} \) and \( \bar{y}_{B,t} \),

\[
\hat{\mu} = \lambda \hat{\mu}_{m,t} + (1 - \lambda) \bar{y}_{B,t},
\tag{3.7}
\]

where \( 0 \leq \lambda \leq 1 \) is to be determined. To minimize the variance of current level, given \( \hat{\mu}_{t-1} \) with variance \( q_{r-1} V(\bar{y}_{B,t}) \), one would minimize

\[
V(\hat{\mu}) = V(\lambda \hat{\mu}_{m,t} + (1 - \lambda) \bar{y}_{B,t}) = \lambda^2 V(\hat{\mu}_{m,t}) + (1 - \lambda)^2 V(\bar{y}_{B,t}),
\tag{3.8}
\]

with respect to \( \lambda \). Under the assumptions (3.3), (3.4) and (3.5), the optimum \( \lambda \) for current level is

\[
\lambda_{\text{opt}} = [g^{-1} \theta^{-1} (1 - \rho^2) + q_{r-1} \rho^2 + 1]^{-1}.
\]

However, if one is planning on using the estimator for a long period of time, one must realize that only certain values of \( q_{r-1} \) are possible in the long run. The value of \( \lambda \) chosen to estimate \( \mu_t \) determines the variance of \( \hat{\mu} \), and hence, determines the variance that will go into the estimator of \( \mu_{t+1} \). Assuming \( \beta = \rho \), we have

\[
V(\hat{\mu}_t) = \{g^{-1} \theta^{-1} (1 - \rho^2) + q_{r-1} \rho^2 + (1 - \lambda)^2\} V(\bar{y}_{B,t})
\]

or

\[
q_{r-1} = g^{-1} \theta^{-1} \lambda^2 (1 - \rho^2) + (1 - \lambda)^2 + \lambda^2 \rho^2 q_{r-1}^{-1}.
\tag{3.9}
\]

Thus, for a given \( \lambda \), the limiting value for \( q_{r-1} \) is

\[
\lim_{t \to \infty} q_{r-1} = \left(1 - \lambda \rho^2\right)^{-1} \left[ g^{-1} \theta^{-1} \lambda^2 (1 - \rho^2) + (1 - \lambda)^2 \right].
\tag{3.10}
\]

This result is equivalent to that given by Cochran (1977), page 352 equation (12.86).

Table 1 contains values of the limit variances as the number of periods becomes large, for selected values of \( \rho \) and \( \lambda \), where \( \theta = 5/6 \) and \( g \theta = 5 \) for the LFS. The variances are standardized so that the variance of the direct estimator based on the mean of six panels is 1.00. Thus, the entries are six times the limiting value in (3.10). If the correlation is 0.95 and \( \lambda \) is set equal to 0.96, the long run variance of current level is 70% of that of the direct estimator. If \( \lambda \) is set equal to 0.90, the long run variance is 58% of that of the direct estimator when \( \rho = 0.95 \).

The first line in Table 1 is for \( \lambda = 5/6 \). This is the \( \lambda \) corresponding to the use of \( x_{2i} \) in a regression estimator. The variance with \( \lambda = 5/6 \) is always smaller than that of the direct estimator because of the improvement associated
with the use of the regression estimator \( \hat{\mu}_{m,t} \). Thus, if \( \rho \neq 0 \), the regression estimator with \( x_{2t} \) leads to significant reduction in variance over the direct estimator, \( \bar{y}_t \), that uses current data only.

The optimal \( \lambda \) is a function of \( \rho \) and increases slowly as \( \rho \) increases. For \( \rho = 0.0 \), the optimal \( \lambda \) is 0.833, for \( \rho = 0.7 \) the optimal \( \lambda \) is about 0.85, for \( \rho = 0.95 \) the optimal \( \lambda \) is about 0.91 and for \( \rho = 0.98 \) the optimal \( \lambda \) is about 0.93.

We now turn to the MR2 estimator (2.3) which can be written as

\[
\hat{m}_t = \lambda \left[ \bar{y}_{m,t} + (\hat{\mu}_{t-1} - \bar{y}_{m,t-1}) b^* \right] + (1 - \lambda) \bar{y}_{B,t},
\]

where \( \lambda_d \) and \( b^* \) are defined in (2.6). While the MR2 estimator is not a member of the class (3.7), to the degree that \( b^* \) is “close to” \( \rho \), it is “close to” a member of the class. For example if \( \rho = 0.95 \), then \( b_0 = 0.9314 \) and \( b^* = 0.9422 \). If \( \rho = 0.90 \), then \( b_0 = 0.8659 \) and \( b^* = 0.8853 \).

Using the limiting value \( b_0 \) of \( b \), we have \( (1 - \lambda_d) = (1 - \lambda)(1 - b_0) \), where \( b_0 \) is given by (2.4). Then \( \lambda_d = 0.9375 \), 0.9586, 0.9776, 0.9886, and 0.9954 for \( \rho = 0.70 \), 0.80, 0.90, 0.95 and 0.98, respectively. From Table 1, the standardized variances of \( \hat{m}_t \) for these values of \( \lambda_d \) are 0.986, 0.982, 0.978, 0.976, and 0.975, for \( \rho = 0.70 \), 0.80, 0.90, 0.95, and 0.98, respectively. Thus, the MR2 estimator for current level has an efficiency for level that is essentially the same as that of the direct estimator, \( \bar{y}_t \). The efficiency of the MR1 estimator is that for \( \lambda = 0.833 \) in Table 1 and is always superior to that of \( \bar{y}_t \).

4. Variance of One-Period Change

Given \( \hat{\mu}_{t-1}, \bar{y}_{m,t-1}, \bar{y}_{m,t} \), and \( \bar{y}_{B,t} \), the optimal estimator of \( \mu_{t-1} \) is no longer \( \hat{\mu}_{t-1} \), because \( \bar{y}_{m,t} \) contains information about \( \mu_{t-1} \). However, it is not customary practice to revise the estimator of \( \mu_{t-1} \). Given no revision, the estimator of change is \( \hat{\mu}_t - \hat{\mu}_{t-1} \).

Under no revision in \( \hat{\mu}_{t-1} \) and conditions (3.2) through (3.5), the variance of \( \hat{\mu}_t - \hat{\mu}_{t-1} \), where \( \hat{\mu}_t \) is defined in (3.7), is

\[
V(\hat{\mu}_t - \hat{\mu}_{t-1}) = V(\lambda \bar{y}_t + (\hat{\mu}_{t-1} - \bar{y}_{m,t-1}) \rho]
\]

\[+ (1 - \lambda) \bar{y}_{B,t} - \hat{\mu}_{t-1})
\]

\[= [g^{-1} \theta^{-1} (1 - \rho^2) + (1 - \lambda)^2
\]

\[+ \rho \lambda - 1)^2 q_t^{-1} V(\bar{y}_{B,t}). \tag{4.1}
\]

Table 2 contains standardized limit variances of the estimated change, \( \hat{\mu}_t - \hat{\mu}_{t-1} \), for selected values of \( g \) and \( \lambda \), with \( g = 0.5 \). The entries in the table are the limiting variances of estimated change divided by the variance of change based on the common elements, \( V(\bar{y}_{m,t} - \bar{y}_{m,t-1}) \). The variance of change based on the common elements is \( 20^{-1} (1 - \theta) (1 - \rho) V(\bar{y}_{B,t}) \). Tables 1 and 2 make clear the cost of not revising the estimate of \( \hat{\mu}_{t-1} \). For example, if \( \rho = 0.95 \), the variance of no-revision one period change is minimized with \( \lambda = 0.99 \), but the variance of level is minimized with \( \lambda = 0.91 \). A compromise value of \( \lambda = 0.95 \) gives a variance of level that is about 14% larger than optimal and a variance of change that is about 7% larger than the smallest variance of Table 2.

Using the values of \( \lambda_d \) associated with the MR2 estimator, the entries in Table 2 are 0.940, 0.960, 0.979, 0.989, and 0.996 for \( \rho = 0.70 \), 0.80, 0.90, 0.95 and 0.98, respectively. Thus, given no revision, and ignoring the difference between \( b_0 \) and \( \rho \), the MR2 estimator is nearly optimal as an estimator of change, unlike the MR1 estimator, where the MR1 estimator corresponds to \( \lambda = 0.833 \) in Table 2.
5. A Compromise Estimator

On the basis of Table 2, the efficiency of the MR2 estimator of change for the LFS based on \( x_{2t} \), for the no-revision case, is quite good. The MR1 no-revision estimator of change based on \( x_{2t} \) has relatively poor efficiency because it is a member of the class (3.7) with \( \lambda = 0 \) = 0.8333. On the other hand, the MR1 estimator of level based on \( x_{3t} \) is superior to the MR2 estimator based on \( x_{3t} \), and there are members of the class (3.7) that are much superior to the MR2 estimator of level.

Because the \( \lambda \) in the MR2 estimator is relatively large and the \( \lambda \) for the MR1 estimator is relatively small, we can create approximations to most interesting members of the class (3.7) as linear combinations of (2.10) and (2.5). Let

\[
x_{3t} = \alpha x_{1t} + (1 - \alpha) x_{2t}, \tag{5.1}
\]

where \( 0 \leq \alpha \leq 1 \) is a fixed number. The regression estimator based on \( x_{3t} \) gives an approximation to a member of the class (3.7) with

\[
\lambda = \alpha \lambda_d + (1 - \alpha) \theta, \tag{5.2}
\]

where \( \lambda_d \) is defined in (2.6). Thus, if \( \alpha = 0.95 \),

\[
\lambda = \alpha(0.9886) + (1 - \alpha)(5/6),
\]

for the LFS rotation pattern with \( \theta = 5/6 \) and \( b_0 = (7 - 2\rho)^{-1} \). \( \lambda = 0.95 \) if \( \alpha = 0.75 \).

We choose \( \alpha \) to give the desired combination of \( \bar{y}_{b,1} \) and the “regression estimator” based on observations in set \( A \). If one does not revise the estimator of \( \mu_{1t} \), the preferred combination depends on the relative importance assigned to the variance of level and to the variance of change.

Table 3 gives the variance of the MR2 estimator \( (\alpha = 1) \) relative to the variance of the estimator constructed using \( \alpha = 0.75 \) and the variance of the estimator constructed using \( \alpha = 0.65 \). An entry in Table 3 for \( \mu_{1t} \) is expression (3.10) evaluated at \( \lambda_d \) of (2.6) and \( \rho \), divided by (3.10) evaluated at \( \lambda \) of (5.2) and \( \lambda \). An entry for \( \hat{\mu}_{1t} - \hat{\mu}_{1t-1} \) is expression (4.1) evaluated at \( \lambda_d \) of (2.6) and \( \rho \), divided by (4.1) evaluated at \( \lambda \) of (5.2) and \( \lambda \). These are approximations to actual efficiencies because \( \rho \) is used for the coefficient of \( x_{3t} \). It is clear from Table 3 that the compromise estimator is slightly inferior to the MR2 estimator for one-period change, but is much superior to the MR2 estimator for level. For example, with \( \rho = 0.95 \) and \( \alpha = 0.65 \), the relative efficiency of the compromise estimator is 1.62 for level and 0.87 for one-period change.

For larger values of \( \rho \), the variance of change is much smaller than the variance of level. Thus, for \( \rho = 0.95 \), the variance of level and of change for \( \alpha = 1.00 \) are about 1.00 and 0.12, respectively, while the variance of level and of change for \( \alpha = 0.75 \) are about 0.67 and 0.13, respectively, when expressed in common units.

The smaller \( \alpha \) has the advantage that the composite estimator will be closer to the direct estimator. Thus, potential biases associated with the composite estimator should be smaller with the smaller \( \alpha \).

<table>
<thead>
<tr>
<th>( \rho )</th>
<th>( b_0 )</th>
<th>( 1 - \lambda_d )</th>
<th>( \hat{\mu}_t )</th>
<th>( \hat{\mu}<em>t - \hat{\mu}</em>{t-1} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.70</td>
<td>0.625</td>
<td>0.0625</td>
<td>1.052</td>
<td>0.999</td>
</tr>
<tr>
<td>0.80</td>
<td>0.741</td>
<td>0.0452</td>
<td>1.099</td>
<td>0.994</td>
</tr>
<tr>
<td>0.90</td>
<td>0.865</td>
<td>0.0224</td>
<td>1.238</td>
<td>0.975</td>
</tr>
<tr>
<td>0.95</td>
<td>0.931</td>
<td>0.0114</td>
<td>1.502</td>
<td>0.936</td>
</tr>
<tr>
<td>0.98</td>
<td>0.972</td>
<td>0.0046</td>
<td>2.177</td>
<td>0.833</td>
</tr>
</tbody>
</table>

6. Drift Problem

As noted in the Introduction, the MR2 estimator could deviate from the direct estimator by a substantial amount and this deviation could extend over a long period. We now illustrate the basis for this phenomenon. We can express the deviation of the compromise regression estimator \( \hat{\mu}_{1t} \) based on \( x_{3t} \), from the true mean \( \mu_t \), as

\[
\hat{\mu}_{1t} - \mu_t = (\lambda \rho)^l (\hat{\mu}_0 - \mu_0) + \sum_{j=0}^{l-1} (\lambda \rho)^j [\lambda \bar{y}_{m,t-j} + (1 - \lambda)(\bar{y}_{m,t-j} - \mu_{t-j})], \tag{6.1}
\]

where \( \mu_0 \) is the mean at the initiation of the process and \( \bar{y}_{m,t} = \bar{y}_{m,t} - \rho(\bar{y}_{m,t-1} - \mu_{t-1}) \).

If \( \rho \) is close to one and we use \( \lambda = 1 \), then the error \( \hat{\mu}_{1t} - \mu_{1t} \) behaves roughly as a random walk which can lead to long periods in which \( \hat{\mu}_{1t} - \mu_{1t} \) has the same sign. On the other hand, if \( \alpha < 1 \) and \( \rho = 1 \), then \( \lambda < 1 \) and the error \( \hat{\mu}_{1t} - \mu_{1t} \) exhibits less drift. For example, if \( \alpha = 0.70 \), the correlation between adjacent errors \( \hat{\mu}_{1t} - \mu_{1t} \) will be no greater than 0.95 under assumption (3.2)-(3.5). For the MR2 estimator, \( \lambda \to 1 \) as \( \rho \to 1 \) and hence the MR2 estimator can exhibit drift for \( \rho \) close to one.

7. Concluding Remarks

For simplicity, we often assumed simple random sampling to obtain theoretical results. Similar results hold for complex designs and additional auxiliary variables, by considering \( \rho \) to be a partial autocorrelation. Also, we used \( x_{3t} \)-variables corresponding to only one variable \( y \), but several \( y \)-variables can be used in constructing the corresponding \( x \)-variables for use in regression estimation. Gambino, Kennedy and Singh (2001) conducted an empirical study with LFS data using several \( x_{3t} \)-variables with a common \( \alpha \), and arrived at a compromise \( \alpha \) for use in the LFS.
In section 2.1, we assumed no nonresponse so that imputation is not required. But in the LFS, nonresponse on an item may occur either at time \( t - 1 \) or a time \( t \) or at both time points. Gambino, Kennedy and Singh (2001) provide details of the imputation methods actually used in the LFS.

Acknowledgements

The research of Wayne Fuller was partly supported by Cooperative Agreement 43-3AEU-3-80088 between Iowa State University, the National Agricultural Statistics Service and the U.S. Bureau of the Census. We thank Harold Mantel for a careful reading of the manuscript that led to improvements.

References
