Model-assisted estimation of unemployment rates for longitudinal surveys with an application to the Current Population Survey

Daniel Bonnéry[*], Yang Cheng[†], Partha Lahiri[‡]

Disclaimer: Any views expressed are those of the authors and not necessarily those of the U.S. Census Bureau.

Abstract

The Current Population Survey (CPS), a household sample survey sponsored by the U.S. Bureau of Labor Statistics (BLS) and conducted by the U.S. Census Bureau, is the primary source of information on the U.S. employment and unemployment levels and rates. The Census Bureau has been using the so-called AK composite estimation technique for generating employment and unemployment levels and rates for the last several decades. The development of a regression composite estimation method by Singh et al. (2001) and Fuller & Rao (2001) and its subsequent adaptation by Statistics Canada for its production of labor force statistics encourage us to propose a new class of model-assisted estimators that covers commonly used composite estimators and is flexible enough to generate sensible estimators of unemployment rate and level using the CPS data. We compare different estimators of unemployment rate and level using a simulation study.

Key Words: Calibration; estimated controls; longitudinal survey; labor force statistics.

1. Introduction

In repeated surveys, different composite estimators that borrow strength over time have been proposed; see Yansaneh & Fuller (1998), Bell (2001), Singh et al. (2001), Fuller & Rao (2001) and others. Such composite estimators typically improve on the standard survey-weighted direct estimators in terms of design-based mean squared error (MSE) criterion and are commonly used by different government agencies for producing official labor force statistics. For example, to produce national employment and unemployment levels and rates, the U.S. Census Bureau uses the AK composite estimation technique developed using the ideas given in Gurney et al. (1965).

Motivated from a Statistics Canada application, Singh & Merkouris (1995) introduced an ingenious idea for generating a composite estimator that can be computed using Statistics Canada's existing software for computing generalized regression estimates. The key idea in Singh & Merkouris (1995) is to create a proxy (auxiliary) variable that uses information at the individual level as well as estimates at the population level from both previous and current periods. Using this proxy variable, Singh & Merkouris (1995) obtained a composite estimator, referred to as MR1 in the literature (MR stands for Modified Regression). However, Singh et al. (1997) noted that MR1 does not perform well in estimating changes in labor force statistics, which motivated them to propose a different composite estimator,

[*]JPSM and US Census Bureau Research Associate, JPSM, University of Maryland, 1218 LeFrak Hall, College Park, MD 20742, United States, US Census Bureau, 4600 Silver Hill Rd, Suitland, MD 20233, United States

[†]US Census Bureau, 4600 Silver Hill Rd, Suitland, MD 20233, United States

[‡]JPSM, University of Maryland, 1218 LeFrak Hall, College Park, MD 20742, United States

called MR2, using a new proxy variable. Singh et al. (2001) generalized the idea of MR1 and MR2 estimators by suggesting a general set of proxy variables.

Fuller & Rao (2001) noted that the regression composite estimator proposed by Singh et al. (1997) is subject to a drift problem and suggested a regression composite method to rectify it. Their method differs from the method of Singh et al. (2001) in two directions. First, the idea of rectifying the drift problem by a weighted combination of the two proxy variables used for MR1 and MR2 is new, although the weighted combination with known weight can be thought as a special case of the methodology proposed by Singh et al. (2001). Secondly, their final regression composite estimator involves estimation of the weight assigned to MR1 or MR2 control variable in the weighted combination — this idea was not discussed in Singh et al. (2001). In short, the Fuller-Rao regression composite estimator with estimated weight cannot be viewed as a special case of Singh et al. (2001) and vice versa.

Gambino et al. (2001) conducted an empirical study to evaluate the Fuller-Rao regression composite estimator, offered missing value treatment and listed several advantages (e.g, weighting procedure, consistency, efficiency gain, etc.) of the Fuller-Rao regression composite estimator over the AK estimator. Statistics Canada now uses the Fuller-Rao method for their official labor force statistics production. Salonen (2007) conducted an empirical study to compare the currently used Finnish labor force estimator with the Fuller-Rao's regression composite and other estimators. Bell (2001) applied the generalized regression technique to improve on the Best Linear Unbiased Estimator (BLUE) based on a fixed window of time points and compared his estimator with the AK composite estimator of Gurney et al. (1965) and the modified regression estimator of Singh et al. (1997), using data from the Australian Labour Force Survey. Beaumont & Bocci (2005) proposed a regression composite estimator with missing covariates defined using variables of interest from the previous month.

In Section 2, we introduce a few notations used in the paper. In Section 3, we review different classes of existing estimators : direct, AK composite, and regression composite. In Section 4, we describe a working model for regression composite estimation, and derive a new model-assisted estimator. In Section 5, we compare different estimators using simulation. In the appendix, we outline the relevant part of the CPS design and a R code used used to mimick the CPS design in our simulations.

2. Notations

2.1 Population

Consider a sequence of finite populations of individuals $(U_m)_{m\in\{1\ldots M\}}$ with sizes N_m, where U_m refers to the finite population for month m. Let $\mathsf{y}_{m,k} = (\mathsf{y}_{m,k,1}, \ldots, \mathsf{y}_{m,k,3})$ be the value of a categorical employment status variable with 3 levels for month m and the kth individual in population U_m. In other words, $\mathsf{y}_{m,k}$ takes on the following values

$$\mathsf{y}_{m,k} = \begin{cases} \text{status}_1 = (1,0,0) & \text{if individual } k \text{ is employed at time } m, \\ \text{status}_2 = (0,1,0) & \text{if individual } k \text{ is unemployed at time } m, \\ \text{status}_3 = (0,0,1) & \text{if individual } k \text{ is not in the labor force at time } m. \end{cases}$$

For $i \in \{1,\ldots,3\}$, we denote $\mathbf{y}_{m,i}$ the vector $\mathbf{y}_{m,i} = (\mathsf{y}_{m,k,i})_{k\in U_m}$. (Notice the difference with $\mathsf{y}_{m,k} = (\mathsf{y}_{m,k,i})_{i\in\{1,\ldots,3\}}$) and $\mathbf{y}_m = (\mathsf{y}_{m,k})_{k\in U_m}$.

Let $\mathsf{x}_{m,k}$ define a $p \times 1$ vector of known constants or estimates. For $m = 1,\cdots,M$, define $\mathbf{x}_m = (\mathsf{x}_{m,k})_{k\in U_m}$. Let $\mathsf{t}^{\text{adj}}_{\mathbf{x}_m}$ be the true or estimated (from another source) total of

$\mathbf{x}$. We also define $\mathbf{z}_m = (\mathsf{z}_{m,k})_{k\in U_m}$, which could contain variables from $\mathbf{x}_m$ and $\mathbf{y}_m$. The total of a vector is denoted by the operator sign t ; for example, $\mathrm{t}_{\mathbf{y}_m} = \sum_{k\in U_m} \mathsf{y}_{m,k}$. We define the function $\mathrm{UR} : (0, +\infty)^3 \to (0, 1], (a, b, c) \mapsto b/(a + b)$. The unemployment rate at time m is given by $\mathrm{ur}_m = \mathrm{UR}\left(\mathrm{t}_{\mathbf{y}_m}\right)$.

2.2 The design

The CPS monthly sample comprises of about 72,000 housing units and is collected for about 729 areas consisting of more than 1,000 counties covering every state and the District of Columbia. The CPS uses a 4-8-4 rotating panel and is conducted by the Census Bureau on a monthly basis. For any given month, the CPS sample can be grouped into eight subsamples corresponding to the eight rotation groups. All the units belonging to a particular rotating panel enter and leave the sample at the same time. A given rotating panel stays in the sample for four consecutive months, stays out of the sample during the eight succeeding months, and then returns for another four consecutive months. It is then dropped from the sample completely and is replaced by a group of nearby households. Of the two new rotation groups that are sampled each month, one is completely new (their rst appearance in the panel) and the other is a returning group, which has been out of the sample for eight months (their first appearance in the panel). Thus, in the CPS design, of the eight rotation groups, six are common between two consecutive months (i.e., 75% overlap) and four are common to the same month in consecutive years (i.e., 75% overlap); see Hansen et al. (1955). For month m, let S_m denote the sample of respondents. We write $S_m = \bigcup_{r=1}^{8} S_{m,r}$, where $S_{m,r}$ denotes the rth rotation group for month m. More details on the CPS design can be found in CPS Technical Report (2006).

3. Different estimators of unemployment rates and levels

3.1 Direct and month-in-sample estimators

Let $w_{m,k}$ denote the so-called second-stage weight for the kth indiviual in month m, which is obtained from the basic weight (or the reciprocal of the inclusion probability) after standard nonresponse and post-stratification adjustments (for more details, we refer to CPS Technical Report (2006)). The direct estimator of the total of y_m is given by $\hat{\mathrm{t}}_{\mathbf{y}_m}^{\mathrm{direct}} = \sum_{k\in S_m} w_{m,k}\mathsf{y}_{m,k}$. An estimator of $\mathrm{t}_{\mathbf{y}_m}$ based on $S_{m,r}$, the sample in the rth rotation group is given by $\hat{\mathrm{t}}_{\mathbf{y}_m}^{\mathrm{mis},r} = 8 \times \sum_{k\in S_{m,r}} w_{m,k}\mathsf{y}_{m,k}$. The direct estimator of ur_m is given by $\widehat{\mathrm{ur}}_m^{\mathrm{direct}} = \mathrm{UR}\left(\hat{\mathrm{t}}_{\mathbf{y}_m}^{\mathrm{direct}}\right)$.

3.2 The AK composite estimator

We define a general class of AK composite estimators. For $m = 1$, the AK estimator is set to the direct estimator:

$$\hat{\mathrm{t}}_{\mathbf{y}_1}^{\mathrm{AK}} = \hat{\mathrm{t}}_{\mathbf{y}_1}^{\mathrm{direct}}.$$

For $m \in 2, \ldots, M$, the AK-estimator of $\mathrm{t}_{\mathbf{y}_m}$ is defined recursively as:

$$\begin{aligned}\hat{\mathrm{t}}_{\mathbf{y}_m}^{\mathrm{AK}} \;&=\; K\,\hat{\mathrm{t}}_{\mathbf{y}_m}^{\mathrm{direct}} + (1-K)\left(\hat{\mathrm{t}}_{\mathbf{y}_{m-1}}^{\mathrm{AK}} + \frac{4}{3}\sum_{k\in S_m\cap S_{m-1}}\left(w_{m,k}\mathsf{y}_{m,k} - w_{m-1,k}\mathsf{y}_{m-1,k}\right)\right)\\ &\quad + A\left(\sum_{k\in S_m\setminus S_{m-1}} w_{m,k}\mathsf{y}_{m,k} - \frac{1}{3}\sum_{k\in S_m\cap S_{m-1}} w_{m,k}\mathsf{y}_{m,k}\right),\end{aligned}$$

where $A, K \in \mathbb{R}$, and $\setminus$ denotes the set difference operator. Then the corresponding unemployment rate estimator is obtained as: $\widehat{\mathrm{ur}}_m^{\mathrm{AK}} = \mathrm{UR}\left(\hat{\mathrm{t}}_{\mathbf{y}_m}^{\mathrm{AK}}\right)$. First two terms of the AK estimator is indeed a weighted average of the current month direct estimator and the previous month AK estimator suitably updated for the change. The last term of the AK estimator is correlated to the previous terms, and has 0 expectation with respect to the sample design. Gurney et al. (1965) explained the benefits of adding the third term in reducing the mean square error. The Census Bureau uses specific values of A and K, which were empirically determined in order to arrive at a compromise solution that worked reasonably well for both employment level and rate estimation (see. Lent et al. (1999)).

Note that the AK estimator can be written as a linear combination of the month-in-sample estimators:

$$\hat{\mathrm{t}}_{\mathbf{y}_m}^{\mathrm{AK}} = \sum_{m'=0}^{m} \sum_{r=1}^{8} c_{m,m',r} \hat{\mathrm{t}}_{\mathbf{y}_{m'}}^{\mathrm{mis},r},$$

where the coefficients $c_{m,m,r}$ are defined recursively as:

$$\forall r \in \{1, \ldots, 8\}, c_{0,0,r} = \frac{1}{8}$$

$$\forall m \in \{2, \ldots, M\}, \begin{cases} \forall r \in \{1,5\} & c_{m,m,r} & = (1-K)/8 + A \\ \forall r \in \{2,3,4,6,7,8\} & c_{m,m,r} & = (1-K)/8 + K/6 - A/3 \\ \forall r \in \{1,2,3,5,6,7\} & c_{m,m-1,r} & = c_{m-1,m-1,r} * K - K/6 \\ \forall r \in \{4,8\} & c_{m,m-1,r} & = c_{m-1,m-1,r} * K \\ \forall 1 \leq m' < m-1 & c_{m,m',r} & = c_{m-1,m',r} * K \end{cases}$$

$$\forall m' > m, r \in \{1, \ldots, 8\}, c_{m,m',r} = 0$$

When the covariance structure of the month-in-sample estimators is known, previous formulae allow us to express both the $c_{m,m',r}$ coefficients and the variance of AK estimators as polynomials of A and K and subsequently to obtain the best coefficients A and K that minimize the variance of the AK estimator.

3.3 Regression composite estimation

In this section we elaborate on the general definition of the class of regression composite estimators proposed by Fuller & Rao (2001). For $\alpha \in [0,1]$, the regression composite estimator of $\mathrm{t}_{\mathsf{y}_m}$ is the calibration estimator $\hat{\mathrm{t}}_{\mathsf{y}_m}^{\mathrm{RC},\alpha}$ defined recursively as follows:

For $m = 1$, define

$$\begin{aligned} \hat{\mathrm{t}}_{\mathbf{z}_m}^{\mathrm{RC},\alpha} &= \hat{\mathrm{t}}_{\mathbf{z}_m}^{\mathrm{direct}}, \\ w_{m,k}^{\mathrm{RC},\alpha} &= w_{m,k},\ \forall k \in S_m. \end{aligned}$$

For $m \in \{2, \ldots, M\}$, define

$$\mathsf{z}_{m,k}^{\star} = \begin{cases} \alpha\left(\tau_m^{-1}\left(\mathsf{z}_{m-1,k} - \mathsf{z}_{m,k}\right) + \mathsf{z}_{m,k}\right) + (1-\alpha)\, \mathsf{z}_{m-1,k} & \text{if } k \in S_m \cap S_{m-1}, \\ \alpha\, \mathsf{z}_{m,k} + (1-\alpha) \left(\sum_{k \in S_{m-1}} w_{m,k}^{\mathrm{RC},\alpha}\right)^{-1} \hat{\mathrm{t}}_{\mathbf{y}_{m-1}}^{\mathrm{c}} & \text{if } k \in S_m \setminus S_{m-1}, \end{cases}$$

where $\tau_m = \left(\sum_{k \in S_m \cap S_{m-1}} w_{m,k}\right)^{-1} \sum_{k \in S_m} w_{m,k}$.
Then the regression composite estimator of $\mathrm{t}_{\mathbf{y}_m}$ is given by

$$\hat{\mathrm{t}}_{\mathbf{y}_m}^{\mathrm{RC},\alpha} = \sum_{k \in S_m} w_{m,k}^{\mathrm{RC},\alpha} \mathsf{y}_{m,k},$$

where

$$\left(w_{m,k}^{\mathrm{RC},\alpha}\right)_{k\in S_m} = \operatorname{argmin}\left\{\sum_{k\in S_m}\left(w_k^\star - w_{m,k}\right)^2/w_{m,k}\;\middle|\; w^\star\in\mathbb{R}^{S_m},\;\begin{array}{l}\sum_{k\in S_m} w_k^\star \mathsf{z}_{m,k}^\star = \hat{\mathrm{t}}_{\mathsf{z}_{m-1}}^{\mathrm{RC},\alpha}\\ \sum_{k\in S_m} w_k^\star \mathsf{x}_{m,k} = \mathrm{t}_{\mathsf{x}_m}^{\mathsf{adj}}\end{array}\right\},$$

$$\hat{\mathrm{t}}_{\mathsf{z}_m}^{\mathrm{RC},\alpha} = \sum_{k\in S_m} w_{m,k}^{\mathrm{RC},\alpha}\mathsf{z}_{m,k}^\star.$$

The regression composite estimator of ur_m is given by $\widehat{\mathrm{ur}}_m^{\mathrm{RC},\alpha} = \mathrm{UR}\left(\hat{\mathrm{t}}_{\mathbf{y}_m}\right)$.

Choice of z and α

Fuller & Rao (2001) studied the properties of $\hat{\mathrm{t}}_{\mathbf{y}_{m,1}}^{\mathrm{RC},\alpha}$ for the choice $\mathbf{z}_m = \mathbf{y}_{m,1}$. As the employment rate is a function of $\mathbf{y}_{m,1}$ and $\mathbf{y}_{m,2}$, we studied the properties of regression composite estimator for the choice $\mathbf{z} = \mathbf{y}_m$.

Fuller & Rao (2001) proposed a method that allows an approximation of the optimal α coefficient for month-to-month change and level estimation, under a specific individual level superpopulation model for continuous variables. They proposed this superpopulation model to explain the drift problem of MR2 (regression composite estimator for $\alpha = 1$) and obtain the best coefficient α. Since we deal with a descrete multidimensional variable, the continuous superpopulation model assumed by Fuller & Rao (2001) is not appropriate in our situation. It will be interesting to propose an approach to estimate the best α in our situation. But for our preliminary study we examined a range of known α values in our simulations.

4. Model-assisted approach to regression composite estimation

4.1 Working model

In this section we develop new model assisted estimators of unemployment rate and level under the following semi parametric population working model on $(\mathbf{y}_{m-1}, \mathbf{y}_m)$:

$$\mathrm{E}\left[\mathsf{y}_{m,k} \mid \mathsf{y}_{m-1,k}\right] = \mathsf{y}_{m-1,k}\gamma_m,$$

where γ_m is a 3×3 transition matrix:

$$\gamma_m = \left[\mathrm{P}\left(\mathsf{y}_{m,k} = \mathrm{status}_{i'} \mid \mathsf{y}_{m-1,k} = \mathrm{status}_i\right)\right]_{i,i'=1,\ldots,3}.$$

If $\left(\mathsf{y}_{m-1,k}\right)_{k\in S_m}$ and $\mathrm{t}_{\mathbf{y}_{m-1}}$ were known, the generalized regression estimator of $\mathrm{t}_{\mathbf{y}_m}$ is given by:

$$\hat{\mathrm{t}}_{\mathbf{y}_m}^{\mathrm{direct}} + \left(\mathrm{t}_{\mathbf{y}_{m-1}} - \hat{\mathrm{t}}_{\mathbf{z}_m^*}^{\mathrm{direct}}\right)\hat{\gamma}_m,$$

where $\forall k\in S_m$, $\mathsf{z}_{m,k}^* = \mathsf{y}_{m-1,k}$, and $\hat{\gamma}_m$ is the survey weighted least squared estimator of γ_m, i.e., the coefficients of the weighted regression of $(\mathbf{y}_{m,k})_{k\in S_m}$ on $(\mathbf{z}_{m,k}^*)_{k\in S_m}$, with the weights $(w_{m,k})_{k\in S_m}$. When unknown, Singh et al. (2001) proposed to replace $\mathbf{z}_m^*$ by a proxy variable $\mathbf{z}^\star$ and $\mathrm{t}_{\mathbf{y}_{m-1}}$ by its estimator. Different proxy variables were proposed by Singh et al. (2001) for $\mathbf{z}_m^*$: For "MR1", (or "RC1") estimator:

$$\mathsf{z}_{m,k}^\star = \mathsf{z}_{m,k}^{MR1} = \begin{cases}\mathsf{y}_{m-1,k} & \text{if } k\in S_m\cap S_{m-1},\\ \hat{\mathsf{y}}_{k,m-1} & \text{if } k\in S_m\setminus S_{m-1},\end{cases}$$

where $\hat{\mathsf{y}}_{k,m-1} = N_{m-1}^{-1}\hat{\mathrm{t}}^{\mathrm{c}}_{\mathsf{y},m-1}$. For "MR2", (or "RC2"):

$$\mathsf{z}^{\star}_{m,k} = \mathsf{z}^{MR2}_{m,k} = \begin{cases} \left(\tau^{-1}[\mathsf{y}_{m-1,k} - \mathsf{y}_{m,k}] + \mathsf{y}_{m,k}\right) & \text{if } k \in S_m \cap S_{m-1}, \\ \mathsf{y}_{m,k} & \text{if } k \in S_m \setminus S_{m-1}. \end{cases}$$

In Fuller & Rao (2001),

$$\mathsf{z}^{\star}_{m,k} = \mathsf{z}^{\mathrm{RC},\alpha}_{m,k} = \alpha \mathsf{z}^{MR1}_{m,k} + (1-\alpha)\mathsf{z}^{MR2}_{m,k}.$$

We propose a new proxy variable, that is obtained from $\mathsf{z}^{\mathrm{RC},\alpha}_{m,k}$ when we replace $\hat{\mathsf{y}}_{m-1,k}$ in $\mathsf{z}^{MR1}_{m,k}$ by a better predictor $\hat{\mathsf{y}}^{\star}_{m-1,k} = \mathrm{E}\left[\mathsf{y}_{m-1,k} \mid \mathsf{y}_{m,k}, \ldots\right]$ of $\mathsf{y}_{m-1,k}$, and remove the range restriction (i.e. $[0,1]$ on α. We can expect that this variable provides better estimators. The resulting model-assisted estimator is given by $\hat{\mathrm{t}}^{\mathrm{mod.assist.}}_{\mathsf{y}}$.

5. Simulation Experiment

5.1 Description of Simulation Study

We conducted a simulation study to enhance our understanding of different composite estimators and to study their finite sample properties. The simulation procedure involved generation of a finite population of size 100,000, and different values for the same set of variables and categories given in the original CPS micro-data maintained by the U.S. Census Bureau for 85 months during the study period 2005-2012. In order to make the simulation experiment realistic, employment statuses were generated in a manner that attempt to capture the dynamics of the U.S. national employment rate direct estimates during the study period 2005-2012. Moreover, in order to assess the maximum gain from different composite estimators, the employment statuses between two consecutive months were made highly correlated subject to a constraint on the global employment rate evolution. The probability of month-to-month changes in employment statuses for an individual was assumed to be zero in case of no change in the corresponding direct estimates of national employment rates. Samples were selected according to a rotating design with systematic selection that mimics the CPS sample design. Since the number of possible samples is only 1000, we could compute the exact bias, variance and mean squared error of different estimators, and subsequently the best linear and AK estimators. Employment rate, total employed, and total unemployed series over the 85-month period were computed using the direct, AK and the Fuller-Rao composite regression methods. We then compared the best estimator in the class of regression composite estimators to the optimal AK and best model-assisted estimators.

5.2 Populations generation

We create a population of $N = 100.000$ individuals, indexed by $1, \ldots, N$. For each individual k of each population, we create a time series $(y_{m,k})_{m\in 1,\ldots,M}$, with value in $\{(1,0,0),(0,1,0),(0,0,1)\}$ (for unemployed, not in labor force, employed). Each individual belongs to a single household, which consists of $K = 5$ individuals. So the entire population contains 20000 housholds. The times series are created under certain constraints at the population level. The unemployment rates are the same as the direct estimates obtained from the CPS data, and the number of people who change status between two consecutive months is minimal.

5.3 Repeated design

We mimic the CPS design, described in Appendix A. For every month m, a sample S_m is the union of 8 mutually exclusive and exhaustive rotation groups. The creation of rotation groups is explained below. Rotation groups are made of $n_h = 20$ households, i.e. 100 individuals. So for month m, there are 800 individuals in the sample, and the inclusion probability of any unit is $1/125$. Rotation groups and samples are formed in the following way:

1. Draw an integer r between 1 and 1000, from a uniform distribution.
2. For $\ell \in 1,\ldots,M+15$, and $j \in \{1,\ldots,8\}$, create the rotation group
$$G_\ell = \{k_{i,j,\ell} \mid r = 1,\ldots,n_h, j = 1,\ldots,5\},$$
where $k_{i,j,\ell} = \mathrm{rem}\left((r-1+\ell-1)*K + \frac{N}{n_h} \times (i-1) + (j-1), N\right) + 1$, and $\mathrm{rem}(a,b)$ denotes the remainder of the Euclidean division of a by b.
3. For $m \in 1,\ldots,M$, create the sample
$$S_m = \bigcup_{\delta\in\{0,1,2,3,12,13,14,15\}} G_{m+\delta}.$$

(The R code used to draw all possible samples is provided in Appendix B)

5.4 Choice of the optimal estimator in each class

For each class $\mathscr{C}$ of unemployment rate estimators (AK, regression composite and model-assisted estimators), we obtain the parameters (A,K or α) that yield the optimal estimators with respect to three different criteria.

Best for level:
$$\operatorname{argmin}\left\{\sum_{m=1}^{M} \mathrm{Var}\left[\mathrm{UR}\left(\hat{\mathrm{t}}^{\star}_{\mathbf{y}_m}\right)\right] \middle| \hat{\mathrm{t}}^{\star}_{\mathbf{y}} \in \mathscr{C}\right\}.$$

Best for change:
$$\operatorname{argmin}\left\{\sum_{m=2}^{M} \mathrm{Var}\left[\mathrm{UR}\left(\hat{\mathrm{t}}^{\star}_{\mathbf{y}_m}\right) - \mathrm{UR}\left(\hat{\mathrm{t}}^{\star}_{\mathbf{y}_{m-1}}\right)\right] \middle| \hat{\mathrm{t}}^{\star}_{\mathbf{y}} \in \mathscr{C}\right\}.$$

Best compromise:
$$\operatorname{argmin}\left\{\sum_{m=1}^{M} \mathrm{Var}\left[\mathrm{UR}\left(\hat{\mathrm{t}}^{\star}_{\mathbf{y}_m}\right)\right] + \sum_{m=2}^{M} \mathrm{Var}\left[\mathrm{UR}\left(\hat{\mathrm{t}}^{\star}_{\mathbf{y}_m}\right) - \mathrm{UR}\left(\hat{\mathrm{t}}^{\star}_{\mathbf{y}_{m-1}}\right)\right] \middle| \hat{\mathrm{t}}^{\star}_{\mathbf{y}} \in \mathscr{C}\right\}.$$

For AK estimation, we we can express these objective functions as polynomial functions of A and K so we are able to obtain the optimal AK in the three different criteria. Table 1 displays the optimal values for A and K. We notice that for each population, the best set of coefficients for change, level and compromise are very close, and thus the optimal choice for level is also almost optimal for change. Table 2 displays the best coefficient α for the regression composite and model-assisted estimators.

	Level	Compromise	Change
(A, K)	$(0.028399, 0.88992)$	$(0.028308, 0.89011)$	$(0.028349, 0.88992)$

Table 1: Optimal (A, K) values for the different criteria.

	Level	Change	Compromise
Regression composite	0.8	1	0.8
Model-assisted	3	0.75	3

Table 2: Optimal α values for model-assisted and regression composite estimators under different criteria

5.5 Results

Figure 1(a) displays the relative mean squared errors for the different estimators of unemployment level, i.e. the times series :

$$\left(\frac{\mathrm{MSE}\left[\hat{\mathrm{ur}}_m^{\star}\right]}{\mathrm{MSE}\left[\hat{\mathrm{ur}}_m^{\mathrm{direct}}\right]}\right)_{m\in\{1,\ldots,M\}}, \text{for } \star \in \{\mathrm{direct}, \mathrm{AK}, \mathrm{RC}, \mathrm{mod.assist.}\},$$

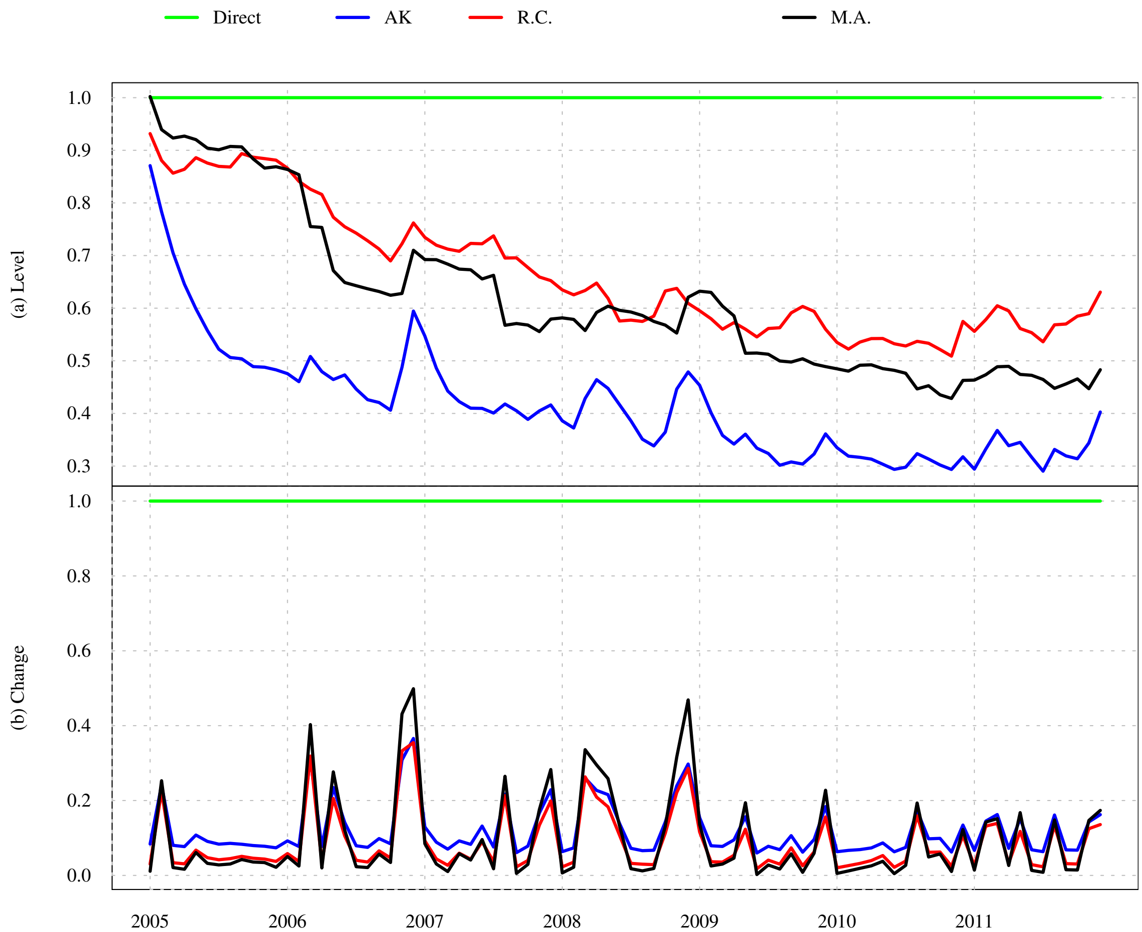
where the best estimators for compromise between month-to-month change and level estimation are chosen in each class.

Figure 1(b) displays the relative mean squared errors for the different estimators of unemployment month-to-month change, i.e. the times series :

$$\left(\frac{\mathrm{MSE}\left[\hat{\mathrm{ur}}_m^{\star}-\hat{\mathrm{ur}}_{m-1}^{\star}\right]}{\mathrm{MSE}\left[\hat{\mathrm{ur}}_m^{\mathrm{direct}}-\hat{\mathrm{ur}}_{m-1}^{\mathrm{direct}}\right]}\right)_{m\in\{2,\ldots,M\}}, \text{for } \star \in \{\mathrm{direct}, \mathrm{AK}, \mathrm{RC}, \mathrm{mod.assist.}\},$$

where the best estimators for compromise between month-to-month change and level estimation are chosen in each class.

Figure 1: Relative mean squared errors of different estimated series of unemployment level and change



Concluding Remarks

In this paper, we propose a general class of model-assisted estimators that covers commonly used composite estimators. While we used a simple working model to illustrate our extension of the standard generalized regression estimator to categorical response variable, auxiliary variables subject to missing values in the sample and unknown auxiliary variable population totals, we plan to include alternative models and conduct an extensive evalu-

	Level			Month-to-month change		
	AK	Regression composite	Model-Assisted	AK	Regression composite	Model-Assisted
0%	0.29	0.509	0.428	0.0598	0.0178	0.00229
25%	0.332	0.568	0.489	0.0733	0.0326	0.0189
50%	0.403	0.631	0.582	0.0849	0.0463	0.0351
75%	0.473	0.737	0.674	0.147	0.119	0.144
100%	1	1	1	0.366	0.355	0.499
Mean	0.42	0.669	0.62	0.119	0.0862	0.0962

Table 3: Quantiles and mean of the relative mean squared errors for different unemployment level estimators.

ation of the different competing models using the CPS historical data. The estimation of the α parameter, which we did not discuss in this paper, and the variance estimation of the resulting estimators of the unemployment level and rate are indeed challenging problems for our future research.

References

Beaumont, JF JeanFrançois, & Bocci, Cynthia. 2005. A Refinement of the Regression Composite Estimator in the Labour Force Survey for Change Estimates. SSC Annual Meeting, Proceedings of the Survey . . . , 1–6.

Bell, Philip. 2001. Comparison of alternative labour force survey estimators. Survey Methodology, **27**(1), 53–63.

CPS Technical Report. 2006. Current population survey: design and methodology. Tech. rept. 66. U.S. Census Bureau.

Fuller, Wayne A, & Rao, JNK N K. 2001. A regression composite estimator with application to the Canadian Labour Force Survey. Survey Methodology, **27**(1), 45–51.

Gambino, Jack, Kennedy, Brian, & Singh, MP Mangala P. 2001. Regression composite estimation for the Canadian labour force survey: Evaluation and implementation. Survey Methodology, **27**(1), 65–74.

Gurney, Daly, Gurney, M, & Daly, J F. 1965. A multivariate approach to estimation in periodic sample surveys. Page 257 of: Proceedings of the Social Statistics Section, American Statistical Association, vol. 242.

Hansen, MH Morris H, Hurwitz, William N, Nisselson, Harold, & Steinberg, Joseph. 1955. The redesign of the census current population survey. Journal of the . . . , **50**(271), 701–719.

Lent, Janice, Miller, Stephen M, Cantwell, Patrick J, & Duff, Martha. 1999. Effects of composite weights on some estimates from the current population survey. Journal of Official Statistics-Stockholm, **15**(1), 431–448.

Salonen, Riku. 2007. Regression Composite Estimation with Application to the Finnish Labour Force Survey. . . . 2.-7. June 2007, Kuusamo, Finland, **8**(3), 503–517.

Singh, AC C, & Merkouris, P. 1995. Composite estimation by modified regression for repeated surveys. Proceedings of the Survey Research Methods Section, 420–425.

Singh, AC C, Kennedy, B, Wu, S, & Brisebois, F. 1997. Composite estimation for the Canadian Labour Force Survey. Proceedings of the Survey Research . . . , 300–305.

Singh, Avinash C AC C, Kennedy, Brian, & Wu, Shiying. 2001. Regression composite estimation for the Canadian Labour Force Survey with a rotating panel design. Survey Methodology, **27**(1), 33–44.

Yansaneh, Ibrahim S, & Fuller, Wayne A. 1998. Optimal recursive estimation for repeated surveys. Survey Methodology, **24**, 31–40.

Acknowledgements

The research of the first and third authors has been supported by the U.S. Census Bureau Prime Contract No: YA1323-09-CQ-0054 (Subcontract No: 41-1016588).

A. The CPS design

We describe the CPS design in the notations of CPS Technical Report (2006), which are different from notations used in Sections 1 to 3. Let U be the intersection of a given basic primary sampling unit component (BPC) and one of the frames (see CPS Technical Report (2006)). The BPC is a set of clusters of about four housing units, where the clusters are the ultimate sampling units (USU). Let N be the number of clusters in U. The clusters in U are first sorted by geographical and demographic characteristics and then indexed by $k = 1 \ldots N$. In the sequence, we will a cluster by its index. Let $\mathrm{SI_w}$ be the adjusted within-PSU sampling interval, as defined in (CPS Technical Report, 2006, p. 3-11). Let $n = \lfloor (21 \times 8 * SI_w)^{-1} N \rfloor$, where $\lfloor . \rfloor$ is the floor function. The number n is the sample size for a sample rotation group. The drawing of the USU within the PSU consists in the generation of a random number X according to the uniform distribution on $[0, 1]$. For $i = 1 \ldots n$, $j = 1 \ldots 8$, $\ell = 85 \ldots (85 + 15)$, let $k_{i,j,\ell}$ denote the cluster $k_{i,j,\ell} = \lfloor (X + 8 \times (i-1) + j) \times SI_w + (\ell - 85) \rfloor$. Then, with the notations of CPS Technical Report (2006) for $\ell = 85 \ldots 100$, $j = 1 \ldots 8$, the rotation group j of sample A_ℓ is given by

$$A_{\ell,j} = \{k_{i,j,\ell} \mid i = 1 \ldots n\}.$$

For a given month, the sample consits of 8 rotation groups. For $t = 1 \ldots 120$, $j' \in \{1, \ldots, 8\}$, we compute $\ell_{t,j'}, j_{t,j'}$: $j_{t,j'} = t + j' - 1 - 8 \times \lfloor (t + j' - 2)/8 \rfloor$. If $j' \in \{1, \ldots, 4\}$, we have $\ell_{t,j'} = 85 + \lfloor (t + j' - 2)/8 \rfloor$. If $j' \in \{5, \ldots, 8\}$, we have $\ell_{t,j'} = 86 + \lfloor (t + j' - 2)/8 \rfloor$. The sample corresponding to the tth month, counting from November 2009, is

$$S_t = \bigcup_{j'=1}^{8} A_{\ell_{t,j'}, j_{t,j'}}.$$

For example, $t = 44$ corresponds to the sample of June 2013, counting from Novembre 2009. Then

$$\begin{aligned}
\ell_{t,1} &= 85 + \lfloor 43/8 \rfloor = 90, & j_{t,1} &= 44 - 8 \times \lfloor 43/8 \rfloor = 4 \\
\ell_{t,2} &= 85 + \lfloor 44/8 \rfloor = 90, & j_{t,2} &= 45 - 8 \times \lfloor 44/8 \rfloor = 5 \\
\ell_{t,3} &= 85 + \lfloor 45/8 \rfloor = 90, & j_{t,3} &= 46 - 8 \times \lfloor 45/8 \rfloor = 6 \\
\ell_{t,4} &= 85 + \lfloor 46/8 \rfloor = 90, & j_{t,4} &= 47 - 8 \times \lfloor 46/8 \rfloor = 7 \\
\ell_{t,5} &= 86 + \lfloor 47/8 \rfloor = 91, & j_{t,5} &= 48 - 8 \times \lfloor 47/8 \rfloor = 8 \\
\ell_{t,6} &= 86 + \lfloor 48/8 \rfloor = 92, & j_{t,6} &= 49 - 8 \times \lfloor 48/8 \rfloor = 1 \\
\ell_{t,7} &= 86 + \lfloor 49/8 \rfloor = 92, & j_{t,7} &= 50 - 8 \times \lfloor 49/8 \rfloor = 2 \\
\ell_{t,8} &= 86 + \lfloor 50/8 \rfloor = 92, & j_{t,8} &= 51 - 8 \times \lfloor 50/8 \rfloor = 3
\end{aligned}$$

We can check on the rotation chart that the sample of June 2013 consists of the 4th, 5th, 6th, 7th rotation groups of A90, of the 8th rotation group of A91, and of the 1st, 2d and 3rd rotation groups of A92:

$$S_{\text{June 2013}} = A_{90,4} \cup A_{90,5} \cup A_{90,6} \cup A_{90,7} \cup A_{91,8} \cup A_{92,1} \cup A_{92,2} \cup A_{92,3}.$$

B. R code for simulations

The R code to draw all the possible sequences of samples is the following:

```
T=85;K<-5;N<-100000;nh<-20;nb.samples<-T+15;H<-N/K;nrep <- 1000;
allsamples <- lapply(1:nrep,
 function(r){
  startp <- K*(r-1)
  samplei<-sapply(1:nb.samples,
   function(t){
     (startp-1+(rep((0:(nh-1))*(N/nh),each=K)+rep((t-1)*K+(1:K),nh)))%%N+1})
     sapply(1:(T+1),function(t){(samplei[,t+c(0,1,2,3,12,13,14,15)])})})
```