Optimal AK Composite Estimators in Current Population Survey

Jun Shao* Zhou Yu† Yang Cheng‡

Abstract
The Current Population Survey (CPS) is a monthly household sample survey consisting of eight rotation groups so that each selected household will be interviewed for 4 consecutive months and another 4 consecutive months after resting 8 consecutive months. A composite type estimator is adopted in the CPS for the estimation of the monthly population total, which combines sample information from the current month survey and previous months using the fact that 75% households have data for two consecutive months. There are two tuning parameters, A and K, in the composite estimator to decide how to combine the available information, and thus this estimator is called the AK composite estimator. However, the current choices of the tuning parameter values were determined by some empirical studies without theory support. In this paper, we derive a formula of the mean squared error of the AK composite estimator, and show that this formula is a quadratic form of A for each fixed K. Using this result, we propose an easy-to-use method of choosing the tuning parameters A and K. Our method is data-driven, i.e., we propose a method to estimate some population quantities in the mean squared error formula using observed data. Some numerical studies are conducted to illustrate the effectiveness of the proposed method.

Key Words: Composite Estimator, Current Population Survey, Mean Square Error, Quadratic form, Sample notation

1. Introduction

The Current Population Survey (CPS) is a household sample survey sponsored by the U.S. Bureau of Labor Statistics and conducted monthly by the U.S. Census Bureau to provide estimates of employment, unemployment, and other characteristics of the non-institutionalized civilian population 16 years of age and older. The CPS adopts a 4-8-4 rotation sample design that consists of a sample of eight rotation groups, approximately equal in size, partitioned in such a manner that for any given month, 1/8 of the sample is interviewed for the first time, 1/8 for the second time,..., and 1/8 for the eighth time. Households in a rotation group are interviewed for 4 consecutive months, dropped for the next 8 months, and then returned to the sample for the following 4 months before they retire from the sample. The rotation paradigm ensures a 75% month-to-month overlap, which makes it possible to increase the efficiency of the current month estimators using data from previous months.

Let Y_t be the unknown population total or mean of interest (e.g., total unemployed or unemployment rate) at month t. The current estimation procedure in the CPS can be described as follows. Based on the data in rotation group i and month t, let $\hat{Y}_{t,i}$ be a ratio, regression, or calibration estimator using some covariates such as age, sex, race, ethnicity, and other household characteristics. A simple estimator of Y_t is the average of $\hat{Y}_{t.i}$ over the

*University of Wisconsin-Madison, WI 53706
†U.S. Census Bureau, Washington, DC 20233
‡U.S. Census Bureau, Washington, DC 20233. This report is released to inform interested parties of research and to encourage discussion of work in progress. Any views expressed on statistical, methodological, or operational issues are those of the authors and not necessarily those of the U.S. Census Bureau.

8 rotation panels, i.e.,

$$\hat{Y}_t = \frac{1}{8}\sum_{i=1}^{8}\hat{Y}_{t,i}. \tag{1}$$

Using data from the 75% sampled units in month t having data for the two consecutive months t and $t-1$, we can estimate the month-to-month change $\Delta_t = Y_t - Y_{t-1}$ by

$$\hat{\Delta}_t = \frac{1}{6}\sum_{i\in s}(\hat{Y}_{t,i} - \hat{Y}_{t-1,i-1}),$$

where $s = \{2,3,4,6,7,8\}$. Note that units in group $i = 1$ or 5 do not have data for month $t-1$. This estimator of change together with the estimated monthly total for month $t-1$ provide an alternative estimator of Y_t:

$$\hat{Y}_{t-1} + \hat{\Delta}_t. \tag{2}$$

While the simple estimator in (1) is based on the data collected in month t only, the alternative estimator in (2) might be more efficient since it makes use of the data from month $t-1$ as well as data from month t in overlapping rotation groups in s. On the other hand, the estimator in (2) does not use data from month t and rotation groups 1 and 5 that are not in s. Thus, to combine the advantages of the estimators in (1) and (2), the following first generation composite estimator was used prior to 1985:

$$\hat{Y}_t' = (1-K)\hat{Y}_t + K(\hat{Y}_{t-1}' + \hat{\Delta}_t), \tag{3}$$

which is a convex combination of two estimators defined in (1) and (2) with a tuning value K between 0 and 1.

After a series of pioneer research studies (e.g., Gurney and Daly 1965; Huang and Ernst 1981), in 1985 a different composite estimator was introduced by adding another term to the composite estimator $\hat{Y}_t'$ in (3), which is the estimator of the net difference between the incoming and continuing parts of the current month's sample:

$$\hat{\zeta}_t = \frac{1}{8}\left(\sum_{i\notin s}\hat{Y}_{t,i} - \frac{1}{3}\sum_{i\in s}\hat{Y}_{t,i}\right).$$

The resulting second generation composite estimator, called the AK composite estimator, is

$$\hat{Y}_t'' = (1-K)\hat{Y}_t + K(\hat{Y}_{t-1}'' + \hat{\Delta}_t) + A\hat{\zeta}_t, \tag{4}$$

where A and K are tuning values. By assigning more weights to rotation groups that have been in the sample for the first and fifth time, the additional term $A\hat{\zeta}_t$ might reduce both the bias and variance of the composite estimators.

Prior to our study, empirical study was the only way to determine tuning values A and K in (4) or K in (3). In the current document for the CPS published in 2006, (A, K) was chosen as $(0.4, 0.7)$ for estimating the total employed or as $(0.3, 0.4)$ for estimating the total unemployed, which were based on some empirical results.

In this article, we derive an analytic formula for the mean squared error (MSE) of the AK composite estimator. This formula is quadratic in A for each fixed K. Based on this formula, the optimal tuning values that minimizes the MSE can be obtained in terms of some unknown population parameters. By substituting unknown parameters by appropriate sample estimators, we obtain a data-adaptive AK estimator which is approximately optimal in terms of the MSE. The proposed method is examined via some simulation studies.

2. The Optimal AK Composite Estimator

The biases of the composite estimator in (3) and the AK composite estimator in (4) were studied in Bailar (1975) and Huang and Ernst (1981), respectively, under the following condition:

(C1) $E(\hat{Y}_{t,i}) = Y_t + a_i$ for any month t, $i = 1, \ldots, 8$.

The biases a_i's in (C1) are mainly caused by the difference in data collection among different rotation groups. Using condition (C1) and the results in Huang and Ernst (1981), we write the bias of the AK composite estimator as a linear function of A:

$$\text{Bias}(\hat{Y}_t'') = E(\hat{Y}_t'') - Y_t = \frac{(\boldsymbol{\gamma}_0^T \mathbf{a})A + (\boldsymbol{\gamma}_1 - K\boldsymbol{\delta})^T \mathbf{a}}{1-K}, \tag{5}$$

where

$$\begin{aligned}
\mathbf{a} &= (a_1, a_2, a_3, a_4, a_5, a_6, a_7, a_8)^T, \\
\boldsymbol{\gamma}_0 &= (1/8, -1/24, -1/24, -1/24, 1/8, -1/24, -1/24, -1/24)^T, \\
\boldsymbol{\gamma}_1 &= \left(\frac{1-K}{8}, \frac{3+K}{24}, \frac{3+K}{24}, \frac{3+K}{24}, \frac{1-K}{8}, \frac{3+K}{24}, \frac{3+K}{24}, \frac{3+K}{24}\right)^T, \\
\boldsymbol{\delta} &= (1/6, 1/6, 1/6, 0, 1/6, 1/6, 1/6, 0)^T.
\end{aligned}$$

Next, we focus on the variance of the AK composite estimator. Huang and Ernst (1981) first gave an approximate variance formulae of AK composite estimator. We will re-derive the variance formula in terms of a quadratic function of A for each fixed K, along the development of Cantwell (1990). We also assume the same conditions as in Huang and Ernst (1981) and Cantwell (1990), which are listed as (C2) and (C3) as follows.

(C2) $\text{Var}(\hat{Y}_{t,i}) = \sigma^2$ for all t and i, and $\hat{Y}_{t,i}$ and $\hat{Y}_{s,j}$ are uncorrelated whenever they are based on groups with different sampled units;

(C3) Based on the structure of the rotation sample design, the following covariances are possibly not zero and we write them in terms of unknown σ^2 and correlation coefficients ρ_i's:

$$\begin{aligned}
\text{Cov}(\hat{Y}_{t,i+1}, \hat{Y}_{t-1,i}) &= \rho_1\sigma^2, \quad i = 1, 2, 3, 5, 6, 7; \\
\text{Cov}(\hat{Y}_{t,i+2}, \hat{Y}_{t-2,i}) &= \rho_2\sigma^2, \quad i = 1, 2, 5, 6; \\
\text{Cov}(\hat{Y}_{t,i+3}, \hat{Y}_{t-3,i}) &= \rho_3\sigma^2, \quad i = 1, 5; \\
\text{Cov}(\hat{Y}_{t,5}, \hat{Y}_{t-9,4}) &= \rho_9\sigma^2; \\
\text{Cov}(\hat{Y}_{t,i+2}, \hat{Y}_{t-10,i}) &= \rho_{10}\sigma^2, \quad i = 3, 4; \\
\text{Cov}(\hat{Y}_{t,i+3}, \hat{Y}_{t-11,i}) &= \rho_{11}\sigma^2, \quad i = 2, 3, 4; \\
\text{Cov}(\hat{Y}_{t,i+4}, \hat{Y}_{t-12,i}) &= \rho_{12}\sigma^2, \quad i = 1, 2, 3, 4; \\
\text{Cov}(\hat{Y}_{t,i+5}, \hat{Y}_{t-13,i}) &= \rho_{13}\sigma^2, \quad i = 1, 2, 3; \\
\text{Cov}(\hat{Y}_{t,i+6}, \hat{Y}_{t-14,i}) &= \rho_{14}\sigma^2, \quad i = 1, 2; \\
\text{Cov}(\hat{Y}_{t,8}, \hat{Y}_{t-15,1}) &= \rho_{15}\sigma^2.
\end{aligned}$$

Let $\mathbf{V}$ be the 8×8 matrix whose (i, j)th element is

$$\mathbf{V}_{i,j} = \begin{cases} K^{i-j}\rho_{i-j} & 1 \le j < i \le 4, \\ K^{i-j}\rho_{i-j} & 5 \le j < i \le 8, \\ K^{i+8-j}\rho_{i+8-j} & 5 \le i \le 8, \quad 1 \le j \le 4, \\ 0 & \text{otherwise.} \end{cases}$$

Also, let $\boldsymbol{\gamma} = A\boldsymbol{\gamma}_0 + \boldsymbol{\gamma}_1$. Following Theorem 1 of Cantwell (1990), we can obtain that

$$\mathrm{Var}(\hat{Y}_t'') = \frac{\sigma^2\{\boldsymbol{\gamma}^T\boldsymbol{\gamma} + K^2\boldsymbol{\delta}^T(\boldsymbol{\delta} - 2\boldsymbol{\gamma}) + 2(\boldsymbol{\gamma} - K^2\boldsymbol{\delta})^T\mathbf{V}(\boldsymbol{\gamma} - \boldsymbol{\delta})\}}{1 - K^2}.$$

By arranging terms as a quadratic form of A, we can write the variance of the AK estimator as

$$\mathrm{Var}(Y_t'') = a_v A^2 + b_v A + c_v, \tag{6}$$

where

$$\begin{aligned} a_v =& \frac{\sigma^2(\boldsymbol{\gamma}_0^T\boldsymbol{\gamma}_0 + 2\boldsymbol{\gamma}_0^T\mathbf{V}\boldsymbol{\gamma}_0)}{1 - K^2}, \\ b_v =& \frac{2\sigma^2(\boldsymbol{\gamma}_0^T\boldsymbol{\gamma}_1 - K^2\boldsymbol{\delta}^T\boldsymbol{\gamma}_0 - K^2\boldsymbol{\delta}^T\mathbf{V}\boldsymbol{\gamma}_0 - \boldsymbol{\gamma}_0^T\mathbf{V}\boldsymbol{\delta} + \boldsymbol{\gamma}_1^T\mathbf{V}\boldsymbol{\gamma}_0 + \boldsymbol{\gamma}_0^T\mathbf{V}\boldsymbol{\gamma}_0)}{1 - K^2}, \\ c_v =& \frac{\sigma^2(\boldsymbol{\gamma}_1^T\boldsymbol{\gamma}_1 + K^2\boldsymbol{\delta}^T\boldsymbol{\delta} - 2K^2\boldsymbol{\delta}^T\boldsymbol{\gamma}_1 + 2K^2\boldsymbol{\delta}^T\mathbf{V}\boldsymbol{\delta} - 2K^2\boldsymbol{\delta}^T\mathbf{V}\boldsymbol{\gamma}_1 - 2\boldsymbol{\gamma}_1'\mathbf{V}\boldsymbol{\delta} + 2\boldsymbol{\gamma}_1'\mathbf{V}\boldsymbol{\gamma}_1)}{1 - K^2} \end{aligned}$$

under conditions (C2) and (C3).

Combining results (5) and (6), we obtain a formula of the MSE of the AK estimator in terms of a quadratic form of A for each fixed K.

Theorem 1. *Assumed conditions* (C1) $-$ (C3)*. The mean square error of AK composite estimator is*

$$\mathrm{MSE}(\hat{Y}_t'') = E(\hat{Y}_t'' - Y_t)^2 = (a_v + a_b)A^2 + (b_v + b_b)A + (c_v + c_b), \tag{7}$$

where

$$a_b = \frac{(\boldsymbol{\gamma}_0^T\mathbf{a})^2}{(1 - K)^2}, \quad b_b = \frac{2(\boldsymbol{\gamma}_0^T\mathbf{a})\{(\boldsymbol{\gamma}_1 - K\boldsymbol{\delta})^T\mathbf{a}\}}{(1 - K)^2}, \quad c_b = \frac{\{(\boldsymbol{\gamma}_1 - K\boldsymbol{\delta})^T\mathbf{a}\}^2}{(1 - K)^2}.$$

Note that $a_b > 0$ and $a_v = (6 - \rho_1 K - 2\rho_2 K^2 - 3\rho_3 K^3)\sigma^2/\{144(1 - K^2)\} > 0$ for $K \in [0, 1)$. Hence, for each fixed K, the MSE in (7) is a quadratic form of A having a minimizer at $-(b_v + b_b)/\{2(a_v + a_b)\}$. Also, for each fixed K,

$$\min_A \left[(a_v + a_b)A^2 + (b_v + b_b)A + (c_v + c_b)\right] = (c_v + c_b) - \frac{(b_v + b_b)^2}{4(a_v + a_b)}.$$

Thus, if all population quantities in (C1)-(C3) are known, then the optimal K and A that minimize the MSE in (7) can be determined through the following algorithm.

Step 1. Find the optimal $K \in [0, 1)$ that minimizes $(c_v + c_b) - (b_v + b_b)^2/\{4(a_v + a_b)\}$ using some method; for example, a gird search.

Step 2. The optimal A is then chosen as $-(b_v + b_b)/\{2(a_v + a_b)\}$ with the K obtained in step 1.

3. Parameter Estimation

The results presented in the previous section are useful only when a_i's in (C1) and σ^2 and ρ_j's in (C2) and (C3) are all known. In practice, however, these values are usually unknown. We in this section address the issue of parameter estimation.

First, consider the estimation of the biases a_i, $i = 1, \dots, 8$. Unfortunately, a_i's are not estimable unless some conditions or constraints are imposed to them. The first results was established in Bailar (1975) under the following assumption:

(C4) $\sum_{i=1}^{8} a_i = 0$.

Under condition (C4),

$$E(\hat{Y}_{t,i} - \hat{Y}_t) = a_i - \frac{1}{8}\sum_{i=1}^{8} a_i = a_i.$$

Then, for a total T months, a_i can be estimated unbiasedly as

$$\hat{a}_i = \frac{1}{T}\sum_{t=1}^{T}(\hat{Y}_{t,i} - \hat{Y}_t).$$

This estimator converges to the true value of a_i as T increases to ∞ at the rate $1/\sqrt{T}$.

Condition (C4) means that the average of 8 rotation group estimators, $\hat{Y}_{t,1}$,...,$\hat{Y}_{t,8}$, is an unbiased estimator of the population total Y_t. Now we turn to the estimation of σ^2 and ρ_i's in (C2)-(C3). Under condition (C2), we obtain that, for any $i \neq j$ and t,

$$\begin{aligned} E(\hat{Y}_{t,i} - \hat{Y}_{t,j})^2 &= E\{(\hat{Y}_{t,i} - Y_t - a_i) - (\hat{Y}_{t,j} - Y_t - a_j) - (a_i - a_j)\}^2 \\ &= 2\sigma^2 + (a_i - a_j)^2. \end{aligned}$$

Hence,

$$\sigma^2 = \frac{E(\hat{Y}_{t,i} - \hat{Y}_{t,j})^2 - (a_i - a_j)^2}{2}.$$

Similarly, the following formulas can be derived under (C2) and (C3):

$$\begin{aligned}
\rho_1 =& E(\hat{Y}_{t+1,i+1} - Y_{t+1} - a_{i+1})(\hat{Y}_{t,i} - Y_t - a_i)/\sigma^2, \quad i = 1,2,3,5,6,7; \\
\rho_2 =& E(\hat{Y}_{t+2,i+2} - Y_{t+2} - a_{i+2})(\hat{Y}_{t,i} - Y_t - a_i)/\sigma^2, \quad i = 1,2,5,6; \\
\rho_3 =& E(\hat{Y}_{t+3,i+3} - Y_{t+3} - a_{i+3})(\hat{Y}_{t,i} - Y_t - a_i)/\sigma^2, \quad i = 1,5; \\
\rho_9 =& E(\hat{Y}_{t+9,i+2} - Y_5 - a_5)(\hat{Y}_{t,4} - Y_t - a_4)/\sigma^2; \\
\rho_{10} =& E(\hat{Y}_{t+10,i+2} - Y_{t+10} - a_{i+2})(\hat{Y}_{t,i} - Y_t - a_i)/\sigma^2, \quad i = 3,4; \\
\rho_{11} =& E(\hat{Y}_{t+11,i+3} - Y_{t+11} - a_{i+3})(\hat{Y}_{t,i} - Y_t - a_i)/\sigma^2, \quad i = 2,3,4; \\
\rho_{12} =& E(\hat{Y}_{t+12,i+4} - Y_{t+10} - a_{i+4})(\hat{Y}_{t,i} - Y_t - a_i)/\sigma^2, \quad i = 1,2,3,4; \\
\rho_{13} =& E(\hat{Y}_{t+13,i+5} - Y_{t+13} - a_{i+5})(\hat{Y}_{t,i} - Y_t - a_i)/\sigma^2, \quad i = 1,2,3; \\
\rho_{14} =& E(\hat{Y}_{t+14,i+6} - Y_{t+14} - a_{i+6})(\hat{Y}_{t,i} - Y_t - a_i)/\sigma^2, \quad i = 1,2; \\
\rho_{15} =& E(\hat{Y}_{t+15,8} - Y_{t+15} - a_8)(\hat{Y}_{t,i} - Y_t - a_1)/\sigma^2.
\end{aligned}$$

With the available unbiased estimator $\hat{a}_i$ for the rotation group bias under (C4), approximately unbiased estimators of σ^2 and ρ_i's based on data over T months can be constructed as follows:

$$\hat{\sigma}^2 = \frac{1}{56T} \sum_{t=1}^{T} \sum_{i<j} \{(\hat{Y}_{t,i} - \hat{Y}_{t,j})^2 - (\hat{a}_i - \hat{a}_j)^2\};$$

$$\hat{\rho}_1 = \frac{1}{6(T-1)\hat{\sigma}^2} \sum_{t=1}^{T-1} \sum_{i\in\{1,2,3,5,6,7\}} (\hat{Y}_{t+1,i+1} - \hat{Y}_{t+1} - \hat{a}_{i+1})(\hat{Y}_{t,i} - \hat{Y}_t - \hat{a}_i);$$

$$\hat{\rho}_2 = \frac{1}{4(T-2)\hat{\sigma}^2} \sum_{t=1}^{T-2} \sum_{i\in\{1,2,5,6\}} (\hat{Y}_{t+2,i+2} - \hat{Y}_{t+2} - \hat{a}_{i+2})(\hat{Y}_{t,i} - \hat{Y}_t - \hat{a}_i);$$

$$\hat{\rho}_3 = \frac{1}{2(T-3)\hat{\sigma}^2} \sum_{t=1}^{T-3} \sum_{i\in\{1,5\}} (\hat{Y}_{t+3,i+3} - \hat{Y}_{t+3} - \hat{a}_{i+3})(\hat{Y}_{t,i} - \hat{Y}_t - \hat{a}_i);$$

$$\hat{\rho}_9 = \frac{1}{(T-9)\hat{\sigma}^2} \sum_{t=1}^{T-9} (\hat{Y}_{t+9,5} - \hat{Y}_{t+9} - \hat{a}_5)(\hat{Y}_{t,4} - \hat{Y}_t - \hat{a}_4);$$

$$\hat{\rho}_{10} = \frac{1}{2(T-10)\hat{\sigma}^2} \sum_{t=1}^{T-10} \sum_{i\in\{3,4\}} (\hat{Y}_{t+10,i+2} - \hat{Y}_{t+10} - \hat{a}_{i+2})(\hat{Y}_{t,i} - \hat{Y}_t - \hat{a}_i);$$

$$\hat{\rho}_{11} = \frac{1}{3(T-11)\hat{\sigma}^2} \sum_{t=1}^{T-11} \sum_{i\in\{2,3,4\}} (\hat{Y}_{t+11,i+3} - \hat{Y}_{t+11} - \hat{a}_{i+3})(\hat{Y}_{t,i} - \hat{Y}_t - \hat{a}_i);$$

$$\hat{\rho}_{12} = \frac{1}{4(T-12)\hat{\sigma}^2} \sum_{t=1}^{T-12} \sum_{i\in\{1,2,3,4\}} (\hat{Y}_{t+12,i+4} - \hat{Y}_{t+12} - \hat{a}_{i+4})(\hat{Y}_{t,i} - \hat{Y}_t - \hat{a}_i);$$

$$\hat{\rho}_{13} = \frac{1}{3(T-13)\hat{\sigma}^2} \sum_{t=1}^{T-13} \sum_{i\in\{1,2,3\}} (\hat{Y}_{t+13,i+5} - \hat{Y}_{t+13} - \hat{a}_{i+5})(\hat{Y}_{t,i} - \hat{Y}_t - \hat{a}_i);$$

$$\hat{\rho}_{14} = \frac{1}{2(T-14)\hat{\sigma}^2} \sum_{t=1}^{T-14} \sum_{i\in\{1,2\}} (\hat{Y}_{t+14,i+6} - \hat{Y}_{t+14} - \hat{a}_{i+6})(\hat{Y}_{t,i} - \hat{Y}_t - \hat{a}_i);$$

$$\hat{\rho}_{15} = \frac{1}{(T-15)\hat{\sigma}^2} \sum_{t=1}^{T-15} (\hat{Y}_{t+15,8} - \hat{Y}_{t+15} - \hat{a}_8)(\hat{Y}_{t,1} - \hat{Y}_t - \hat{a}_1).$$

The next theorem confirms that these moment estimators are consistent with the convergence rate $\sqrt{T}$.

Theorem 2. *Assume conditions* (C1)-(C4). *Then, as $T \to \infty$,*

$$\sqrt{T}(\hat{\sigma}^2 - \sigma^2) = O_P(1),$$
$$\sqrt{T}(\hat{a}_i - a_i) = O_P(1),$$
$$\sqrt{T}(\hat{\rho}_j - \rho_j) = O_P(1)$$

for $i = 1, \ldots, 8$ and $j = 1, 2, 3, 9, \ldots, 15$, where $O_P(1)$ is a quantity bounded in probability.

The proof of this theorem is given in the Appendix.

4. Simulation Results

Our simulation study was based on a pseudo population containing the CPS micro data from January 2004 to December 2013. In each month, there are about $16,000$ individuals in each rotation group. We randomly sampled n individuals from each rotation group in month t_0, and kept them on track in the next fifteen month. We considered Y_t to be the employment rate at month t and obtained $\hat{Y}_{t_0,1}, \hat{Y}_{t_0+1,2}, \hat{Y}_{t_0+2,3}, \hat{Y}_{t_0+3,4}, \hat{Y}_{t_0+12,5}, \hat{Y}_{t_0+13,6}$, $\hat{Y}_{t_0+14,7}$ and $\hat{Y}_{t+15,8}$ through ratio estimation. Based on $\hat{Y}_{t,i}$'s with t ranging from January 2004 to December 2012, we estimated the optimal coefficients A and K according to the proposed procedure in Sections 2-3, and further obtained the optimal AK estimators for all months in 2013. The proposed estimator is denoted by $\hat{Y}''_{t,opt}$. For t ranging from January 2013 to December 2013, we compared $\hat{Y}''_{t,opt}$ with the AK estimator $\hat{Y}''_t$ currently used in the CPS, where $(A, K) = (0.4, 0.7)$.

Based on 500 repetitions, we approximated the estimation bias and standard deviation (SD) of $\hat{Y}''_{t,opt}$ and $\hat{Y}''_t$, and listed them in Table 1 for t ranging from Janunary to December, 2013. We also included in the tables the values of Y_t and the relative improvement in MSE by incorporating the proposed optimal AK composite estimator:

$$\text{IM} = \frac{\text{MSE}(\hat{Y}''_t) - \text{MSE}(\hat{Y}''_{t,opt})}{\text{MSE}(\hat{Y}''_t)}.$$

In general, the simulation results support the theory derived in Sections 2-3. From Table 1, in terms of the MSE, the approximate optimal AK composite estimator $\hat{Y}''_{t,opt}$ is better than the current AK composite estimator $\hat{Y}''_t$ except in two cases occurred in April 2013 when $n = 100$. Since $\hat{Y}''_{t,opt}$ is based on estimated parameters in (C1)-(C3), it is possible that it performs worse than $\hat{Y}''_t$. When $\hat{Y}''_{t,opt}$ improves $\hat{Y}''_t$, the relative improvement IM can be as high as 12.24% and the improvement is generally larger when n is smaller. The optimal AK composite estimator mainly improves the variance of the estimation, since the SD of $\hat{Y}''_{t,opt}$ is uniformly smaller than the SD of $\hat{Y}''_t$. It is also less biased in most of the cases under consideration. In fact, in cases where $\hat{Y}''_{t,opt}$ is more biased than $\hat{Y}''_t$, the improvements in MSE are small.

Based on our limited simulation results, updating the tuning coefficients A and K periodically is necessary. Also, the estimation of parameters in (C1)-(C3) is crucial for a good performance of $\hat{Y}''_{t,opt}$.

5. Discussions

Selecting the tuning coefficients in the AK composite estimators has been a longstanding issue in the CPS. We in this paper propose a novel method for choosing the optimal coefficients based on an analytic formula of the MSE of the AK composite estimator. Since this analytic formula can be expressed as a quadratic form of A for each fixed K, the optimal A and K can be easily obtained. We also propose some consistent estimators of the parameters in the optimal coefficients A and K so that a data-driven approximate AK composite estimator can be used in the CPS. This approach along with the parameters estimation procedure can be further developed for choosing optimal weighting coefficients of a class of generalized composite estimators (Park et al. 2001). In addition, combining with Theorem 2 in Cantwell (1990), our proposal can be extended to choose the optimal A and K for estimating the month-to-month change Δ_t.

The theoretical justifications and numerical results in this paper are based on some commonly used conditions in the CPS. Some recent studies suggest different correlation structures between two rotation groups. After 2003, a new grouping procedure between panel i and $i+4$ was introduced in the CPS ($i = 1, 2, 3, 4$). To further adapt such grouping procedure, we suggest modify assumptions (C1)-(C3) accordingly. Let $\hat{Z}_{t,i} = \hat{Y}_{t,i} + \hat{Y}_{t,i+4}$ for $i = 1, 2, 3, 4$. The assumptions (C1)-(C3) can be updated as

Table 1: Simulation results (in %) for employment rate under assumption $\sum_{i=1}^{8} a_i = 0$.

n	t	Y_t	Bias($\hat{Y}_t''$)	Bias($\hat{Y}_{t,opt}''$)	SD($\hat{Y}_t''$)	SD($\hat{Y}_{t,opt}''$)	IM
20	Jan 2013	92.09	-1.69	-1.48	5.10	4.81	12.24
	Feb 2013	92.44	-0.93	-0.84	4.71	4.47	10.15
	Mar 2013	92.85	-2.44	-2.43	4.80	4.65	5.49
	Apr 2013	93.37	-1.64	-1.69	4.54	4.42	3.69
	May 2013	93.33	-2.37	-2.23	4.23	4.03	9.67
	Jun 2013	92.85	-2.21	-2.12	5.05	4.75	10.94
	Jul 2013	92.91	-2.71	-2.65	5.79	5.44	10.64
	Aug 2013	93.30	-3.02	-3.02	4.32	4.24	2.64
	Sep 2013	93.63	-2.55	-2.45	4.27	4.11	7.48
	Oct 2013	93.59	-1.96	-1.99	4.16	4.01	5.15
	Nov 2013	93.94	-3.45	-3.49	4.39	4.16	5.64
	Dec 2013	93.96	-3.98	-3.92	5.02	4.94	3.10
50	Jan 2013	92.09	-1.43	-1.40	2.87	2.73	8.53
	Feb 2013	92.44	-1.68	-1.62	3.11	2.96	9.37
	Mar 2013	92.85	-1.94	-1.89	2.77	2.58	10.37
	Apr 2013	93.37	-1.80	-1.85	2.55	2.38	6.31
	May 2013	93.33	-2.16	-2.02	2.59	2.51	8.70
	Jun 2013	92.85	-1.77	-1.72	2.76	2.64	7.77
	Jul 2013	92.91	-1.96	-1.90	2.84	2.70	8.51
	Aug 2013	93.30	-1.34	-1.36	2.29	2.18	6.65
	Sep 2013	93.63	-2.32	-2.30	2.86	2.68	7.96
	Oct 2013	93.59	-2.13	-2.18	2.59	2.45	4.14
	Nov 2013	93.94	-2.87	-2.93	2.78	2.71	0.50
	Dec 2013	93.96	-3.42	-3.31	2.92	2.73	9.12
100	Jan 2013	92.09	-2.17	-2.15	2.24	2.14	4.45
	Feb 2013	92.44	-1.56	-1.52	2.12	2.03	6.73
	Mar 2013	92.85	-1.55	-1.58	1.90	1.79	4.84
	Apr 2013	93.37	-2.48	-2.57	1.88	1.77	-0.64
	May 2013	93.33	-2.18	-2.17	1.93	1.83	4.75
	Jun 2013	92.85	-2.58	-2.59	2.00	1.94	1.83
	Jul 2013	92.91	-2.68	-2.67	2.14	2.08	2.32
	Aug 2013	93.30	-1.99	-2.03	2.05	1.98	1.40
	Sep 2013	93.63	-2.10	-2.09	2.00	1.92	4.10
	Oct 2013	93.59	-2.44	-2.46	1.97	1.89	2.53
	Nov 2013	93.94	-3.11	-3.15	2.08	1.96	1.60
	Dec 2013	93.96	-3.29	-3.25	1.94	1.83	4.66

(C1') $\text{Var}(\hat{Z}_{t,i}) = \sigma_1^2$ for all t and i.

(C2') The covariances between different rotation groups in the same month are zero. That is $\text{Cov}(\hat{Z}_{t,i}, \hat{Z}_{t,j}) = 0$ for $1 \le i \ne j \le 4$.

(C3') All covariances between different rotation groups in different months are zero except for the following:

$$\begin{aligned}
\text{Cov}(\hat{Z}_{t,i+1}, \hat{Z}_{t-1,i}) &= \rho_1^*\sigma_1^2, \quad i = 1,2,3;\\
\text{Cov}(\hat{Z}_{t,i+2}, \hat{Z}_{t-2,i}) &= \rho_2^*\sigma_1^2, \quad i = 1,2;\\
\text{Cov}(\hat{Z}_{t,i+3}, \hat{Z}_{t-3,i}) &= \rho_3^*\sigma_1^2, \quad i = 1;\\
\text{Cov}(\hat{Z}_{t,1}, \hat{Z}_{t-9,4}) &= \rho_9^*\sigma_1^2;\\
\text{Cov}(\hat{Z}_{t,i-2}, \hat{Z}_{t-10,i}) &= \rho_{10}^*\sigma_1^2, \quad i = 3,4;\\
\text{Cov}(\hat{Z}_{t,i-1}, \hat{Z}_{t-11,i}) &= \rho_{11}^*\sigma_1^2, \quad i = 2,3,4;\\
\text{Cov}(\hat{Z}_{t,i}, \hat{Z}_{t-12,i}) &= \rho_{12}^*\sigma_1^2, \quad i = 1,2,3,4;\\
\text{Cov}(\hat{Z}_{t,i+1}, \hat{Z}_{t-13,i}) &= \rho_{13}^*\sigma_1^2, \quad i = 1,2,3;\\
\text{Cov}(\hat{Z}_{t,i+2}, \hat{Z}_{t-14,i}) &= \rho_{14}^*\sigma_1^2, \quad i = 1,2;\\
\text{Cov}(\hat{Z}_{t,4}, \hat{Z}_{t-15,1}) &= \rho_{15}^*\sigma_1^2.
\end{aligned}$$

In addition, the AK composite estimator based on $\hat{Z}_{t,i}$'s can be defined as

$$\hat{Z}_t = \frac{1-K}{8}\sum_{i=1}^{4}\hat{Z}_{t,i} + (1-K)\Big\{\hat{Z}_{t-1} + \sum_{i=2}^{4}(\hat{Z}_{t,i} - \hat{Z}_{t,i-1})/6\Big\} + A\Big\{\big(\sum_{i=2}^{4}\hat{Z}_{t,i}\big) - \hat{Z}_{t,1}\Big\}/8.$$

Then we can follow our proposed method to select optimal A and K under assumptions (C1')-(C3').

Acknowledgements

The research of the first two authors has been supported by the U.S. Census Bureau Prime Contract No: YA1323-09-CQ-0054.

Appendix

Proof of Theorem 2.
We only prove the results for $\hat{a}_i$'s and $\hat{\sigma}^2$ under (C1)-(C4). Proof for the asymptotic results under alternative assumption (C4') or for $\hat{\rho}_j$'s is similar and thus is omitted. We first deal with $\hat{a}_i$'s. Let $U_{t,i} = \hat{Y}_{t,i} - \sum_{j=1}^{8}\hat{Y}_{t,j}/8$. Then $\{U_{1,i}, \ldots, U_{T,i}\}$ is a stationary 16-dependent sequence. By Theorem 9.1 in DasGupta (2008), we have

$$\sqrt{T}(\hat{a}_i - a_i) = \frac{1}{\sqrt{T}}\sum_{t=1}^{T}\{U_{t,i} - E(U_{t,i})\} \longrightarrow N(0, \tau_i^2);$$

as T goes to infinity, where $\tau_i^2 = \text{Var}(U_{1,i}) + 2\sum_{j=2}^{16}\text{Cov}(U_{1,i}, U_{j,i})$ and $i = 1, \ldots, 8$.

Now we deal with $\hat{\sigma}^2$. Let $S_t = \sum_{i<j}\{(\hat{Y}_{t,i} - \hat{Y}_{t,j})^2 - (a_i - a_j)^2\}/56$. Then S_t is also a stationary 16-dependent sequence. Again by Theorem 9.1 in Dasgupta (2008), we can see that

$$\frac{1}{\sqrt{T}}\sum_{t=1}^{T}\{\hat{S}_t - E(S_t)\} \longrightarrow N(0, \tau_0^2);$$

as T goes to infinity, where $\tau_0^2 = \text{Var}(S_1) + 2\sum_{j=2}^{16} \text{Cov}(S_1, S_j)$. Then it is easy to see that $\hat{\sigma}^2 - \sigma^2 = O_P(T^{-1/2})$ since $\hat{a}_i - a_i = O_P(T^{-1/2})$. The proof is completed. $\square$

REFERENCES

Bailar, B.A. (1975). The effects of rotation group bias on estimates from panel surveys. *Journal of the American Statistical Association,* 70, 23-30.

Breau, P. and Ernst L.R. (1983) Alternative estimators to the current composite estimator. *Proceedings of the American Statistical Association, Section on Survey Methods Research*, 303-308.

Cantwell, P. J. (1990). Variance Formulae for Composite Estimators in One-and Multi-Level Rotation Designs. *Survey Methodology* 16(1), 153-163.

DasGupta, A. (2008). Asymptotic theory of statistics and probability. *Springer.*

Huang, E.T. and Ernst, L.R. (1981). Comparison of an alternative estimation to the current composite estimator in CPS. *Proceedings of the Section on Survey Research Methods, American Statistical Association,* 303-308.

Gurney, M. and and Daly, J.F. (1965). A multivariate approach to estimation in period- ic sample surveys. *Proceedings of the Social Statistics Section, American Statistical Association,* 242-257.

Park, S.P., Choi, J.W. and Kim, K.W. (2001). One-level rotation design balanced on time in monthly sample and in rotation group. *Journal of the American Statistical Association,* 96, 1483-1496.

U.S. Bureau of the Census (2006). Current Population Survey: Design and Methodology. *Technical Paper 66, Washington, DC: U.S. Govemment Printing Office (Dept. of Commerce).*