ALTERNATIVE ESTIMATORS FOR THE CURRENT POPULATION SURVEY.

Ibrahim S. Yansaneh and Wayne A. Fuller.
Ibrahim S. Yansaneh, Department of Statistics, Iowa State University, Ames, IA 50011.

KEY WORDS: Rotation Designs, Least Squares, Recursive Regression Estimation.

1. Introduction.

The Current Population Survey is a household survey conducted by the United States Census Bureau in cooperation with the Bureau of Labor Statistics. It is designed to generate estimates of labor force characteristics (such as employed, unemployed, and Civilian Labor Force), demographic characteristics, and other characteristics of the noninstitutionalised civilian population. The sample design of the Current Population Survey contains a rotation scheme that includes the replacement of a fraction of the households in the sample each month.

For any given month, the sample consists of eight time — in — sample panels or rotation groups, of which one is being interviewed for the first time, one is being interviewed for the second time, ..., and one is being interviewed for the eighth time. In other words, the interview scheme is balanced on time—in—sample. Households in a rotation group are interviewed for four consecutive months, dropped for the next eight succeeding months, and then interviewed for another four consecutive months. They are then dropped from the sample entirely. This system of interviewing is called the 4—8—4 rotation scheme, and is a special case of the scheme described by Rao and Graham (1964).

The Current Population Survey is a repeated survey, the design and analysis of which have received considerable attention in the literature. Theoretical foundations for the design and estimation for repeated surveys were laid down by Patterson (1950). Least squares procedures were considered further by Jones (1980), and Fuller (1990). Composite estimation is a procedure of estimation which makes use of observations from the current period and the preceding period, and the estimator of the previous period. See Rao and Graham (1964), and Kumar and Lee (1983). A comparison of alternative estimators for the Current Population Survey was undertaken by Huang and Ernst (1981) and Breau and Ernst (1983). Least squares estimation for a fairly general class of repeated surveys was considered by Yansaneh (1992).

2. Basic Assumptions

Assume that in each period of the survey, s rotation groups are introduced into the sample, where $s > 1$ is fixed. For computational convenience, the data obtained over p periods can be arranged in a $p \times s$ data matrix, denoted by M, in such a way that all of the observations on a rotation group appear in a single column. The total number of observations is $n = ps$, where n is the number of entries in M. We refer to the columns of M as "streams". Assume that:

1. A given rotation group is observed over a period of total length m and the observation pattern is fixed ($m = 16$ for the Current Population Survey).
2. The columns of M are independent.
3. The covariance structure of the observations in a stream is constant over time, and it is the same for all streams.

The computation of the variances of the alternative estimators of current level and change are based on the estimated covariance structure constructed by Adam (1992).

3. Alternative Estimators of Level and Change

3.1. The Present Composite Estimator.

Composite estimators combine the estimators of previous periods with data from the current period and immediately preceding period to form an estimate of the current period. With the Current Population Survey, six of the eight rotation groups observed at time t were observed at time $t - 1$. Let $y_{i,t}$ be the estimate of level obtained from the rotation group which is in the i—th stream at time t, $i = 1, ..., s$. The estimator is of the general form

$$\hat{\theta}_{t,c} = \sum_{i=1}^{s} \omega_{1,k(i,t)} y_{i,t} + \sum_{i=1}^{s} \omega_{2,k(i,t)} y_{i,t-1} + \phi_2 \hat{\theta}_{t-1,c}, \qquad (1)$$

where $k(i, t) = k$ defines time—in—sample k as a function of (i, t), $\lambda_1 = 1/8$, $\lambda_2 = -1/6$, $\lambda_3 = 1/3$, $\phi_1 = 0.6$, $\phi_2 = 1 - \phi_1 = 0.4$, $\phi_3 = 0.2$,

$$\omega_{1,k} = \begin{cases} \phi_1\lambda_1 - \phi_2\lambda_2 - \phi_3\lambda_1\lambda_3 \\ \text{for} \quad k = 2, 3, 4, 6, 7, 8 \\ \lambda_1(\phi_1 + \phi_3) \quad \text{for } k = 1, 5 \end{cases}$$

$$\omega_{2,k} = \phi_2\lambda_2 .$$

Let $p = (\omega_{1,k(1,t)}, \ldots, \omega_{1,k(s,t)})'$, $q = (\omega_{2,k(1,t)}, \ldots, \omega_{2,k(s,t)})'$, and $y_t = (y_{1,t}, \ldots, y_{s,t})'$. Then,

$$\hat{\theta}_{t,c} = p'y_t + q'y_{t-1} + \phi_2\hat{\theta}_{t-1}. \qquad (2)$$

Substituting in (2) recursively, we have, for an estimator initiated at time zero,

$$\hat{\theta}_{t,c} = p'y_t + \sum_{j=1}^{t} \phi_2^{j-1}(q + \phi_2 p)'y_{t-j}. \qquad (3)$$

Equation (3) is an expression for $\hat{\theta}_{t,c}$ as a linear function of current and past observations and, since $|\phi_2| < 1$, the weights decline as the distance from the current period increases.

3.2. The Best Linear Unbiased Estimator

Suppose s streams of data collected over p periods are available. Let $Y_i = (y_{i,1}, \ldots, y_{i,p})'$, $i = 1, \ldots, s$, and let $Y_p = (y_1', y_2', \ldots, y_s')'$ be the $n \times 1$ vector of observations, let $\theta_p = (\theta_1, \theta_2, \ldots, \theta_{p-1}, \theta_p)'$ be the $p \times 1$ vector of parameters of interest, let $X = J_{s\times 1} \otimes I_{p\times p}$ be the Kronecker product of $J_{s\times 1}$, the $s \times 1$ vector of ones, and $I_{p\times p}$ the $p \times p$ identity matrix.

The linear model with no time–in–sample effects is

$$Y_p = X\theta_p + \epsilon_p , \qquad (4)$$

where ϵ_p is the vector of error terms and we assume that $E\{\epsilon_p\} = 0$. Let V_p be the covariance matrix of Y_p. The best linear unbiased estimator of θ_p is

$$\hat{\theta}_p = (X'V_p^{-1}X)^{-1}X'V_p^{-1}Y_p , \qquad (5)$$

with covariance matrix $\Sigma_p = (X'V_p^{-1}X)^{-1}$.

3.3. The Recursive Regression Estimator

The recursive regression estimation procedure is a computationally efficient method of producing minimum variance estimators in repeated surveys. Instead of using all the available information in a large least squares computation, the recursive regression estimation procedure uses a linear combination of an appropriate set of initial estimates and the new observations at the current level to produce the best linear unbiased estimators of current level and change.

At the current time, denoted by c, where $c \geq m$, we desire an estimator of θ_c, the value of a particular characteristic. We have available:

1. m best linear unbiased estimators of the parameters for the previous m periods,
$$\hat{\theta}_{c-1(m)} = (\hat{\theta}_{c-m}, \ldots, \hat{\theta}_{c-1})',$$
2. the $m \times m$ covariance matrix $\Sigma_{11,c-1(m)}$ of $\hat{\theta}_{c-1(m)}$,
3. s observations on the eight streams at the current time.

The s elementary observations can be transformed so that they are uncorrelated with previous observations. Let the transformed observations be

$$s_{ic} = y_{i,c} - \sum_{j=1}^{m} b_{k(i,c),j}y_{i,c-j},$$

where the b_{ij}'s are constructed so that s_{ic} is uncorrelated with $y_{i,c-j}$ for all $j > 0$. The expected values of $s_{1c}, s_{2c}, \ldots, s_{sc}$ are θ_c, $\theta_c - b_{21}\theta_{c-1}, \ldots, \theta_c - \sum_{j=1}^{m} b_{sj}\theta_{c-j}$, respectively. A linear model in the data available at the current time is

$$Z_c = W\theta_{c(m+1)} + \epsilon_c, \qquad (6)$$

where

$$W = \begin{bmatrix} I_{m\times m} & 0 \\ X_{21} & J_{s\times 1} \end{bmatrix},$$

$$Z_c = \begin{bmatrix} \hat{\theta}_{c-1(m)} \\ s_c \end{bmatrix},$$

$$\theta_{c(m+1)} = (\theta_{c-m}, \ldots, \theta_{c-1}, \theta_c),$$

$$s'_c = (s_{1c}, s_{2c}, \ldots, s_{sc}),$$

and X_{21} is an $s \times m$ matrix whose entries are functions of the $b_{k(i,c),j}$'s, which are in turn functions of the autocorrelations. The covariance matrix of Z_c is:

$$V_c = \begin{bmatrix} \Sigma_{11,c-1(m)} & 0 \\ 0 & Q_{00} \end{bmatrix},$$

where $Q_{00} = \text{Var}\{s_c\} = \text{Diag}\{\sigma_1^2, \sigma_2^2, \ldots, \sigma_s^2\}$, and $\sigma_i^2 = \text{var}\{s_{ic}\}$, $i = 1, 2, \ldots, s$.

The ***recursive regression estimator*** of the vector $\theta_{c(m+1)}$ of parameters at time c is

$$\hat{\theta}_{c(m+1)} = (W'V_c^{-1}W)^{-1}W'V_c^{-1}Z_c, \qquad (7)$$

The covariance matrix of $\hat{\theta}_{c(m+1)}$ is

$$Q_R = (W'V_c^{-1}W)^{-1}.$$

It can be shown that the recursive regression estimator of current level θ_t is the best linear unbiased estimator of θ_t based on data for periods 1, 2, ..., t [Yansaneh (1992), § 4.4]. To update the recursive regression estimator for the next period, we drop the initial estimate $\hat{\theta}_{t-m}$ from the data vector, and drop θ_{t-m} from the parameter vector. The parameter θ_{t+1} is then added to the parameter vector. This way, the dimension of the estimation problem is kept constant over time.

It can be proven that the covariance matrix of the vector of recursive least squares estimators obtained in the recursive regression procedure converges to a positive definite matrix as the number of periods increases [Yansaneh (1992), § 4.4].

3.4. The First Order Composite Estimator

The first order composite estimator is a composite estimator of the present composite type, constructed to give approximately optimum estimates of current level under a first order autoregressive model.

The weights to be used in the construction of the first order composite estimator of current level are the least squares weights constructed using two periods of data. The procedure is described in Yansaneh (1992), Section 5.5.3.

4. Time–in–Sample Effects

A major problem with most periodic surveys is the presence of time–in–sample effects. This refers to the phenomenon by which estimates of current level for a given period obtained from different rotation groups have different expected values, depending on the length of time they have been included in the sample. The effects on the estimates of current level and change have been studied by Bailar (1975) and by Kumar and Lee (1983).

We shall now examine the effect of rotation group bias or time–in–sample effects on the least squares estimators of current level and change. The least squares procedures described in Section 3 can be modified to incorporate time–in–sample effects. Our discussion will focus on the 4–8–4 rotation scheme, but our procedure can be easily modified and applied to any rotation design.

4.1. Best Linear Unbiased Estimation

We proceed exactly as in Section 3.2, where we used elementary estimators in a linear model to produce best linear unbiased estimators. In the presence of time–in–sample effects, the components of the linear model (4) in Section 3.2 that change are the design matrix X and the parameter vector θ_p.

Suppose τ_k is the rotation group effect for time t associated with the rotation group which is in its k–th time–in–sample. Then, for each time t , we may write the model

$$y_{k,t} = \theta_t + \tau_k + \epsilon_{k,t} , \qquad (8)$$

where $y_{k,t}$ is $y_{i,t}$ when $k(i, t) = k$ and the covariance structure of the errors $\epsilon_{k,t}$ is the same as in Section 3.2. With an estimability restriction on τ_k , the best linear unbiased estimator of $\beta_p = (\tau, ..., \tau_{s-1}, \theta_1, ..., \theta_{p-1}, \theta_p)'$ can be constructed by the usual least squares procedures.

4.2. The Recursive Regression Procedure

To construct a recursive estimator in the presence of time–in–sample effects, we proceed as in Section 3.3 with appropriate modifications in the design matrix and parameter vector of the corresponding linear model. We assume that at time t , the following quantities are available:

(a) $\ell = m + s - 1$ initial estimates $\hat{\beta}_{t-1(\ell)} = (\hat{\tau}'_{t-1(s-1)}, \hat{\theta}'_{t-1(m)})'$

where $\hat{\tau}'_{t-1(s-1)} = [\hat{\tau}_1(t-1), ..., \hat{\tau}_{s-1}(t-1)]$ and $\hat{\theta}'_{t-1(m)} = [\hat{\theta}_{t-m}(t-1), ..., \hat{\theta}_{t-1}(t-1)]$, where $\hat{\theta}_{t-1(m)}$ is the best estimator of the vector using data through time $t - 1$.

(b) the covariance matrix of $\hat{\beta}_{t-1(\ell)}$,

(c) s independent observations, denoted by $s_t = (s_{1t}, ..., s_{st})'$.

Then, the estimator of $\beta_{t(\ell+1)}$ is constructed as in Section 3.3.

In the recursive procedure, the current estimates of the time–in–sample effects are in the data vector throughout the iteration process. It therefore follows that the variance of each of the time–in–sample effects will converge to zero as the number of periods increases.

One may be unwilling to assume that the time–in–sample effects are constant over a long period. One way of permitting the time–in–sample effects to change slowly over time is to do a kind of "exponential smoothing" by adjusting the covariance matrix of the estimated effects at time t , used to construct the estimator. One procedure is to multiply the covariance matrix of the initial estimates of the time–in–sample effects by a constant bigger than one. In this procedure, it is important to distinguish between the matrix used to define the estimator and the actual covariance matrix of the estimators. The calculated matrix used to construct the estimator will converge as the number of periods increases. Then, the matrix defining the recursive regression estimator will also converge. In the limit, we can write

$$\hat{\beta}_{t(\ell+1)} = P_\tau Z_{t,\tau}, \qquad (9)$$

where $Z_{t,\tau} = (\hat{\beta}'_{t-1(\ell)}, s'_t)$, and P_τ is the limit matrix of coefficients. Since $\hat{\beta}_{t(\ell+1)}$ is a function of preceding estimates, one can use the procedure outlined in Section 3.1 to calculate the coefficients of the observations that define the estimator.

5. Results and Discussion

The variances of the alternative estimation procedures relative to the variance of the basic estimator of current level, for both characteristics of interest, are presented in Table 1. We define the basic estimator of the current level as the simple mean of the elementary estimates obtained from the 8 rotation groups at the current period. In general, the best linear unbiased procedure becomes more statistically efficient as the number of periods increases. For both characteristics, the results reveal that the best linear unbiased procedure based on 24 periods is uniformly more efficient than the present composite estimator. The precision of the best linear unbiased procedure relative to the present composite estimator for current level is 30% for the best linear unbiased estimator for 24 periods, and 33% for the recursive regression estimator. For unemployed, the corresponding gain in precision is about 3% for all the estimators. These results are a reflection of the nature of the autocorrelation functions of the characteristics. The autocorrelation function for unemployed declines much faster than that for employed.

With the exception of one period change in employed, there is a substantial improvement in the efficiency of the estimation of change from using the alternative estimators instead of the present composite estimator. The gain in precision increases as the interval of change increases, reaching a maximum value at the five–period change for both characteristics. The gain then decreases slightly. In the case of the recursive regression estimator, the maximum gain in estimated change is 64% for employed and 11% for unemployed.

The results of the comparison of alternative estimators and rotation designs are given in Table 2. The performance of the alternative estimators of current level, change in level, and average level for multiple time periods under the intermittent 4–8–4 rotation design and two continuous rotation designs are compared. The best estimator used in these comparisons is the best linear unbiased estimator of current level based on 36 periods. The efficiency of the 36–period least squares estimator is virtually the same as that of the recursive regression estimator. The continuous rotation designs are the 6 continuous scheme and the 8 continuous rotation scheme. The 6 continuous scheme is the rotation scheme used in the Canadian Labor Force Survey conducted by Statistics Canada [Kumar and Lee (1983)]. For each period of the survey, the sample consists of six rotation groups. A given rotation group remains in the sample for six consecutive periods and then drops out of the sample for good. In the 8 continuous scheme, there are 8 rotation groups in the sample for each period, one rotation group in the its first time in sample, ..., and one rotation group in its eighth month in sample. A given rotation group remains in the sample for eight consecutive periods and then drops out of the sample for good.

For all rotation schemes under consideration, there is some improvement in the precision of the estimators of current level from using the best estimator relative to the present composite estimator. As seen in Table 2, the gain is highest for employed where, under the 4–8–4 rotation scheme, the variance of the best estimator of current level is only 92% of that of the present composite estimator. The relative precision of the best estimators of change relative to the present composite estimator depends on the rotation design. From Table 1 and 2, we see that under the 4–8–4 rotation scheme, there is some gain in precision, which increases as the interval of change increases. For employed, the variance of the best estimator is 85% of the variance of the present composite estimator in estimating one–period change, 61% of the variance of the present composite estimator in estimating six–period change, and 96% of the variance of the present composite estimator in estimating twelve–period change.

For estimating twelve–period averages, the present composite estimator is about 13% less efficient than the best estimator and for estimating change in twelve–period averages, it is about 53% less efficient. For unemployed, there are only modest gains in precision from using the best estimator relative to the present composite estimator.

It can be seen by comparing Tables 1 and 3 that for employed, there is some gain in precision for estimation of current level and change from using the best estimator under the continuous rotation designs instead of the present composite estimator. Again, the gains in precision are higher for employed.

For estimation of twelve–period change, twelve–period average and change in twelve–period averages, the best estimator, under both continuous rotation schemes, is less efficient than the present composite estimator for both characteristics, as can be seen by comparing the last three rows of Tables 2 and 3.

In Table 3, we compare the variances of the first order composite estimator for employed and unemployed under the various rotation designs. Under the 4–8–4 rotation scheme, there are modest gains in the precision of estimation of current level from using the first order composite estimator instead of the present composite estimator. The gain is 9% for employed and 1% for unemployed. However, for employed, the first order composite estimator of change under the 4–8–4 rotation scheme is clearly superior to the present composite estimator. In estimating current level and change up to twelve periods for unemployed under the 4–8–4 rotation scheme, the first order composite estimator has roughly the same efficiency as the present composite estimator. The results for the continuous rotation designs are similar.

As mentioned in Section 4, in the presence of time–in–sample effects, the alternative estimators of current level and change are biased relative to the mean of the basic estimator. The variances of the alternative estimators of current level and change over several periods in the presence of time–in–sample effects are presented in Table 4 for employed and unemployed. In all cases, the sum of the time–in–sample effects is restricted to be zero. That is, the estimator is restricted to have a mean equal to the mean of the eight elementary estimators. Under this restriction, there is an increase in variance of about 10% for current level

Table 1: Variances of alternative estimators relative to the variance of the basic estimator of current level

Parameter	Employed			Unemployed		
	Present composite	BLUE 24 periods	Recursive regression estimator	Present composite	BLUE 24 periods	Recursive regression estimator
Current level	0.862	0.661	0.650	0.947	0.918	0.918
1–period change	1.511	0.432	0.432	1.070	1.073	1.073
6–period change	1.390	0.853	0.856	1.708	1.628	1.628
12–period change	0.992	0.747	0.761	1.593	1.563	1.564

Table 2: Variances of the best estimator; variance of the basic estimator of current level equals one

Quantity Estimated	Employed			Unemployed		
	4–8–4	8 Cont.	6 Cont.	4–8–4	8 Cont.	6 Cont.
Current level	0.653	0.761	0.739	0.918	0.944	0.938
1–period change	0.432	0.398	0.434	1.073	1.003	1.051
6–period change	0.854	0.828	0.970	1.628	1.577	1.655
12–period change	0.758	1.046	1.186	1.564	1.696	1.737
12–period average	0.326	0.440	0.394	0.249	0.301	0.266
Change in 12–period averages	0.162	0.368	0.403	0.262	0.372	0.359

Table 3: Variances of the first order composite estimator relative to the variance of the basic estimator of current level

Quantity Estimated	Employed			Unemployed		
	4–8–4	8 Cont.	6 Cont.	4–8–4	8 Cont.	6 Cont.
Current level	0.790	0.826	0.800	0.942	0.961	0.952
1–period change	0.482	0.417	0.457	1.090	1.011	1.060
6–period change	1.127	0.989	1.133	1.702	1.635	1.729
12–period change	0.951	1.298	1.341	1.580	1.751	0.774
12–period average	0.380	0.454	0.388	0.250	0.301	0.261
Change in 12–period averages	0.263	0.473	0.450	0.264	0.388	0.363

Table 4: Variances of alternative estimators in the presence of time–in–sample effects[*]; variance of the basic estimator of current level equals one

Quantity Estimated	Employed		Unemployed	
	BLUE for 24 periods	Recursive regression estimator	BLUE for 24 periods	Recursive regression estimator
Current level	0.729	0.688	0.928	0.923
1–period change	0.449	0.438	1.089	1.078
6–period change	0.931	0.878	1.646	1.635
12–period change	0.914	0.807	1.589	1.573

[*]The estimator is constructed so that the expected value of the estimator equals the expected value of the average of the eight elementary estimators.

of employed and virtually no increase for unemployed.

Generally, the effect of including time–in–sample effects in the model is to increase the variance of the estimators. The increase in variance is a function of the type of restriction imposed and the length of the period used to estimate the time–in–sample effects.

ACKNOWLEDGEMENTS

This research was partly supported by Joint Statistical Agreement 91–21 with the U.S. Bureau of the Census.

REFERENCES

Adam, A. (1992), Covariance estimation for the characteristics of the Current Population Survey. *A Ph.D. Dissertation.* Iowa State University, Ames, Iowa.

Bailer, B.A. (1975), The effects of rotation group bias on estimates from panel surveys. *Journal of the American Statistical Association*, **70**, 23–29.

Breau, P. and Ernst, L. (1983), Alternative estimators to the current composite estimator. *Proceedings of the American Statistical Association, Section on Survey Research Methods*, 397–402.

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

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

Jones, R.G. (1980), Best linear unbiased estimators for repeated repeated surveys. *Journal of the Royal Statistical Society, Series B*, **42**, 221–226.

Kumar, S., and Lee, H. (1983), Evaluation of composite estimation for the Canadian LaborForce Survey. *Survey Methodology*, **9**, 1–24.

Patterson, H.D. (1950), Sampling on successive occasions with partial replacement of units. *Journal of the Royal Statistical Society, Series B* **12**, 241–255.

Rao, J. N. K., and Graham, J. E. (1964), Rotation Designs for Sampling on Repeated Occasions. *Journal of the American Statistical Association* **59**, 492–509.

Yansaneh, I.S. (1992), Least squares estimation for repeated surveys. *A Ph.D. Dissertation.* Iowa State University, Ames, Iowa.