MEAN SQUARE ERROR ESTIMATION OF KALMAN FILTER ESTIMATES FROM AGGREGATE SURVEY DATA

Stephen M. Miller, Bureau of Labor Statistics
441 G St. N.W. Room 2126, Washington D.C. 20212

1. Introduction and Derivation of Mean Square Error

The use of time series methods applied to survey data is a topic that is receiving renewed interest among statisticians. Early papers by Scott and Smith (1974) and Scott, Smith and Jones (1977) demonstrated how times series methods could be applied to survey data. More recent papers such as Tam (1987), Bell and Hillmer (1987) and (1989), Eltinge and Fuller (1989), and Tiller (1989) have expanded on these ideas and stressed the use of state space model techniques as well as the Kalman filter.

We will consider Kalman filter estimation as applied to the following situation. Assume that a survey has been conducted at times $t = 1, \ldots, T$, and for each time period t an estimate Y_t of a finite population quantity y_t is computed. Define

$$Y_t = y_t + w_t \tag{1}$$

and call w_t the sampling error at time t . We will assume that the sampling errors are independent of the fixed sequence of values y_t , and that the w_t have expectation zero for all t . We will also assume that for each time period t , a set of fixed auxiliary variables are available which are contained in the $1 \times p$ dimensional row vector $\underset{\sim}{X}_t$. It will always be assumed that the constant 1 appears as the first element of $\underset{\sim}{X}_t$ for all t . In practice auxillary variables will be selected which are correlated with the true value y_t , but we will not assume the existence of any particular relationship between y_t and $\underset{\sim}{X}_t$ as is often done in model-based sampling.

We assume that the primary interest of the survey is to estimate the y_t , and we will consider estimators of the form

$$\hat{y}_t = \underset{\sim}{X}_t \hat{\underset{\sim}{\beta}}_t \quad , \text{ for } t = 1, \ldots, T, \tag{2}$$

where $\hat{\underset{\sim}{\beta}}_t$ is constructed based upon the following iterative formula

$$\hat{\underset{\sim}{\beta}}_t = \hat{\underset{\sim}{\beta}}_{t-1} + \underset{\sim}{L}_t(Y_t - \underset{\sim}{X}_t\hat{\underset{\sim}{\beta}}_{t-1}) , \tag{3}$$

where the $\underset{\sim}{L}_t$ are functions of the auxillary vectors and are thus considered to be a fixed sequence of vectors with respect to sampling variability. Notice that we can write

$$\hat{y}_t = (\underset{\sim}{X}_t\underset{\sim}{L}_t)Y_t + (1 - \underset{\sim}{X}_t\underset{\sim}{L}_t)\underset{\sim}{X}_t\hat{\underset{\sim}{\beta}}_{t-1} . \tag{4}$$

This implies that $\hat{y}_t$ is a weighted combination of the aggregate estimate Y_t and a *one-step-ahead* prediction $\underset{\sim}{X}_t\hat{\underset{\sim}{\beta}}_{t-1}$, where the weights add to 1 . The recursive estimation procedure in (3) is a distinctive feature of the Kalman filter method of estimation, and the $\underset{\sim}{L}_t$ is often referred to as the Kalman gain matrix. There are many ways of generating the sequence of matrices $\underset{\sim}{L}_t$ as well as choosing the initial value $\hat{\underset{\sim}{\beta}}_0$, and each of these methods produces its own particular estimator. The methods of ordinary least squares, weighted least squares, and variable coefficient models all fit within this iterative framework.

We are interested in obtaining an expression for the mean square error of the estimator in (2) . We begin by writing the prediction error as

$$\begin{aligned} y_t - \hat{y}_t &= y_t - \underset{\sim}{X}_t\hat{\underset{\sim}{\beta}}_t \\ &= (Y_t - w_t) - \underset{\sim}{X}_t\hat{\underset{\sim}{\beta}}_{t-1} - \underset{\sim}{X}_t\underset{\sim}{L}_t(Y_t - \underset{\sim}{X}_t\hat{\underset{\sim}{\beta}}_{t-1}) \\ &= -w_t + (1 - \underset{\sim}{X}_t\underset{\sim}{L}_t)(Y_t - \underset{\sim}{X}_t\hat{\underset{\sim}{\beta}}_{t-1}) . \end{aligned} \tag{5}$$

Notice that an important component of the prediction error is the quantity $(Y_t - \underset{\sim}{X}_t\hat{\underset{\sim}{\beta}}_{t-1})$ which we will call the *one-step-ahead prediction error.* We can write the mean square error of $\hat{y}_t$ as

$$\begin{aligned} E\{(y_t - \hat{y}_t)^2\} = {} & \sigma^2_{wt} + \\ & (1 - \underset{\sim}{X}_t\underset{\sim}{L}_t)^2 E\{(Y_t - \underset{\sim}{X}_t\hat{\underset{\sim}{\beta}}_{t-1})^2\} \\ & -2(1 - \underset{\sim}{X}_t\underset{\sim}{L}_t)\sigma^2_{wt} + 2(1 - \underset{\sim}{X}_t\underset{\sim}{L}_t)E\{w_t\underset{\sim}{X}_t\hat{\underset{\sim}{\beta}}_{t-1}\} , \end{aligned} \tag{6}$$

where σ^2_{wt} is defined as the variance of w_t . Throughout this paper the symbol E will denote expectation with respect to the sample design. Notice that by using equation (3) iteratively we can write

$$\begin{aligned} \hat{\underset{\sim}{\beta}}_t &= \underset{\sim}{L}_t Y_t + (\underset{\sim}{I}_p - \underset{\sim}{L}_t\underset{\sim}{X}_t)\hat{\underset{\sim}{\beta}}_{t-1} \\ &= \underset{\sim}{L}_t Y_t + (\underset{\sim}{I}_p - \underset{\sim}{L}_t\underset{\sim}{X}_t)\underset{\sim}{L}_{t-1}Y_{t-1} \\ &\quad + (\underset{\sim}{I}_p - \underset{\sim}{L}_t\underset{\sim}{X}_t)(\underset{\sim}{I}_p - \underset{\sim}{L}_{t-1}\underset{\sim}{X}_{t-1})\hat{\underset{\sim}{\beta}}_{t-2} \\ &= \sum_{j=0}^{t-1} \{\prod_{i=1}^{j} (\underset{\sim}{I}_p - \underset{\sim}{L}_{t-i+1}\underset{\sim}{X}_{t-i+1})\} \underset{\sim}{L}_{t-j}Y_{t-j} \\ &\quad + \{\prod_{i=1}^{t} (\underset{\sim}{I}_p - \underset{\sim}{L}_{t-i+1}\underset{\sim}{X}_{t-i+1})\}\hat{\underset{\sim}{\beta}}_0 , \end{aligned} \tag{7}$$

where the products above are defined to be $\underset{\sim}{I}_p$ when the upper limit is 0 , and $\underset{\sim}{I}_p$ stands for the identity matrix of dimension p . If we assume that the w_t is uncorrelated with the initial vector $\hat{\underset{\sim}{\beta}}_0$ then we can write

$$E\{w_t\underset{\sim}{X}_t\hat{\underset{\sim}{\beta}}_{t-1}\} = \sum_{j=0}^{t-2} a_{tj}E(w_t w_{t-1-j}) \tag{8}$$

where for $j = 1, \ldots, t-2$

$$a_{tj} = \underset{\sim}{X}_t[\prod_{i=1}^{j} (\underset{\sim}{L}_p - \underset{\sim}{L}_{t-i}\underset{\sim}{X}_{t-i})]\, \underset{\sim}{L}_{t-1-j} \tag{9}$$

$$a_{t0} = \underset{\sim}{X}_t\, \underset{\sim}{L}_{t-1}\,.$$

Therefore we can write the mean square error as

$$\begin{aligned} E\{(y_t - \hat{y}_t)^2\} &= (2\, \underset{\sim}{X}_t\underset{\sim}{L}_t - 1)\sigma^2_{wt} \\ &+ (1 - \underset{\sim}{X}_t\underset{\sim}{L}_t)^2 E\{(Y_t - \underset{\sim}{X}_t\hat{\underset{\sim}{\beta}}_{t-1})^2\} \\ &+ 2(1 - \underset{\sim}{X}_t\underset{\sim}{L}_t)\sum_{j=0}^{t-2} a_{tj}E(w_t w_{t-1-j})\,. \end{aligned} \tag{10}$$

Notice that the second moment of the one-step-ahead prediction error is a major component of the mean square error. We will next derive two expressions for $E\{(Y_t - \underset{\sim}{X}_t\hat{\underset{\sim}{\beta}}_{t-1})^2\}$. We begin by writing

$$\begin{aligned} Y_t - \underset{\sim}{X}_t\underset{\sim}{H}_t\hat{\underset{\sim}{\beta}}_{t-1} &= w_t + y_t - \underset{\sim}{X}_t\hat{\underset{\sim}{\beta}}_{t-1} \\ &= w_t + q_t - \underset{\sim}{X}_t(\hat{\underset{\sim}{\beta}}_{t-1} - \underset{\sim}{\beta}_{t-1})\,, \end{aligned} \tag{11}$$

where $\underset{\sim}{\beta}_{t-1} \equiv E\{ \hat{\underset{\sim}{\beta}}_{t-1} \}$ for all t, and

$$q_t = y_t - \underset{\sim}{X}_t\underset{\sim}{\beta}_{t-1}\,, \text{ for } t = 1, \ldots, T. \tag{12}$$

The quantity q_t is the expected value of the one-step-ahead prediction error, and is a function of the true y_t values as well as the estimation method use to construct the $\hat{\underset{\sim}{\beta}}_t$. We can write

$$\begin{aligned} E\{(Y_t - \underset{\sim}{X}_t\hat{\underset{\sim}{\beta}}_{t-1})^2\} &= \sigma^2_{wt} + q_t^2 \\ &+ \underset{\sim}{X}_t V\{ \hat{\underset{\sim}{\beta}}_{t-1} \}\underset{\sim}{X}_t' - 2E\{w_t\underset{\sim}{X}_t(\hat{\underset{\sim}{\beta}}_{t-1} - \underset{\sim}{\beta}_{t-1})\} \end{aligned} \tag{13}$$

where $V\{ \hat{\underset{\sim}{\beta}}_{t-1} \}$ is the covariance matrix of $\hat{\underset{\sim}{\beta}}_{t-1}$, defined as

$$V\{ \hat{\underset{\sim}{\beta}}_{t-1} \} = E\{(\hat{\underset{\sim}{\beta}}_{t-1} - \underset{\sim}{\beta}_{t-1})(\hat{\underset{\sim}{\beta}}_{t-1} - \underset{\sim}{\beta}_{t-1})'\}\,. \tag{14}$$

Thus we can write

$$\begin{aligned} E\{(Y_t - \underset{\sim}{X}_t\hat{\underset{\sim}{\beta}}_{t-1})^2\} &= \sigma^2_{wt} + q_t^2 \\ &+ (\underset{\sim}{X}_t\otimes\underset{\sim}{X}_t)\underset{\sim}{\Phi}_p vech(V\{ \hat{\underset{\sim}{\beta}}_{t-1} \}) \\ &- 2\sum_{j=0}^{t-2} a_{tj}E(w_t w_{t-1-j})\,, \end{aligned} \tag{15}$$

where $\otimes$ stands for Kronecker product and the notation $\underset{\sim}{\Phi}_p$ and *vech* are defined in Fuller(1987). The vector $vech(V\{ \hat{\underset{\sim}{\beta}}_{t-1} \})$, by definition, contains all of the unique elements of the matrix $V\{ \hat{\underset{\sim}{\beta}}_{t-1} \}$ ordered by columns, and is of dimension $\frac{1}{2}p(1 + p) \times 1$.

Thus we can write the mean square error as

$$\begin{aligned} E\{(y_t - \hat{y}_t)^2\} &= (\underset{\sim}{X}_t\underset{\sim}{L}_t)^2\sigma^2_{wt} \\ &+ (1 - \underset{\sim}{X}_t\underset{\sim}{L}_t)^2[q_t^2 + (\underset{\sim}{X}_t\otimes\underset{\sim}{X}_t)\underset{\sim}{\Phi}_p vech(V\{ \hat{\underset{\sim}{\beta}}_{t-1} \})] \\ &+ 2(\underset{\sim}{X}_t\underset{\sim}{L}_t)(1 - \underset{\sim}{X}_t\underset{\sim}{L}_t)\sum_{j=0}^{t-2} a_{tj}E(w_t w_{t-1-j})\,. \end{aligned} \tag{16}$$

The term

$$q_t^2 + (\underset{\sim}{X}_t\otimes\underset{\sim}{X}_t)\underset{\sim}{\Phi}_p vech(V\{ \hat{\underset{\sim}{\beta}}_{t-1} \}) \tag{17}$$

in the mean square error expression is interesting. If ordinary least squares estimation is used then under general conditions $V\{ \hat{\underset{\sim}{\beta}}_{t-1} \} = O(t^{-1})$ as $t \to \infty$, which says that (17) will be dominated by q_t^2 for large values of t. If a variable coefficient model is used, then under general conditions $V\{ \hat{\underset{\sim}{\beta}}_{t-1} \} = O(1)$ as $t \to \infty$ but it is likely that q_t^2 will be smaller than in the ordinary least squares case. This refects the inherent trade-off between bias and variance when using an estimation scheme which allows the coefficient vectors to change over time (i.e. the variable coeffients models) as opposed to ordinary least squares which treats the coefficients as fixed over time.

We next derive an alternative expression for $E\{(Y_t - \underset{\sim}{X}_t\hat{\underset{\sim}{\beta}}_{t-1})^2\}$. We begin by noting that

$$\begin{aligned} Y_t - \underset{\sim}{X}_t\hat{\underset{\sim}{\beta}}_{t-1} &= w_t + q_t - \underset{\sim}{X}_t(\hat{\underset{\sim}{\beta}}_{t-1} - \underset{\sim}{\beta}_{t-1}) \\ &= q_t + w_t - \sum_{j=0}^{t-2} a_{tj}w_{t-1-j}\,. \end{aligned} \tag{18}$$

Thus the one-step-ahead prediction error can be written as the sum of two independent terms, the first term being q_t and the second term coming from a linear combination of the sampling errors across time. We can write the second moment of the one-step-ahead prediction error as

$$\begin{aligned} E\{(Y_t - \underset{\sim}{X}_t\hat{\underset{\sim}{\beta}}_{t-1})^2\} &= q_t^2 + \sigma^2_{wt} \\ &- 2\sum_{j=0}^{t-2} a_{tj}E(w_t w_{t-1-j}) \\ &+ \sum_{j=0}^{t-2}\sum_{k=0}^{t-2} a_{tj}a_{tk}E(w_{t-1-k}w_{t-1-j}). \end{aligned} \tag{19}$$

Thus we can write the mean square error as

$$\begin{aligned} E\{(y_t - \hat{y}_t)^2\} &= (\underset{\sim}{X}_t\underset{\sim}{L}_t)^2\sigma^2_{wt} \\ &+ (1 - \underset{\sim}{X}_t\underset{\sim}{L}_t)^2[q_t^2 + \sum_{j=0}^{t-2}\sum_{k=0}^{t-2} a_{tj}a_{tk}E(w_{t-1-k}w_{t-1-j})] \\ &+ 2(\underset{\sim}{X}_t\underset{\sim}{L}_t)(1 - \underset{\sim}{X}_t\underset{\sim}{L}_t)\sum_{j=0}^{t-2} a_{tj}E(w_t w_{t-1-j})\,. \end{aligned} \tag{20}$$

In this section we have concentrated on deriving expressions for the mean square error of $\hat{y}_t$, but we can also write an expression for its bias. Namely

$$E\{(\hat{y}_t - y_t)\} = -(1 - \underset{\sim}{X}_t\underset{\sim}{L}_t)q_t\,. \tag{21}$$

Therefore the bias of $\hat{y}_t$ is a known negative multiple of the expected value of the one-step-ahead prediction error.

In the next section we examine estimating the mean square error expressions we have derived in this section, but before doing that we will first derive lower bounds for the mean square error based upon expressions (10) and (16). Notice that the second term is positive in both expressions (10) and (16), therefore it is natural to define the lower bounds $B_1(\hat{y}_t)$ and $B_2(\hat{y}_t)$ as follows.

$$B_1(\hat{y}_t) = (2\,\utilde{X}_t\utilde{L}_t - 1)\sigma^2_{wt} + 2(1 - \utilde{X}_t\utilde{L}_t)\sum_{j=0}^{t-2} a_{tj}E(w_t w_{t-1-j}) \,, \tag{22}$$

$$B_2(\hat{y}_t) = (\utilde{X}_t\utilde{L}_t)^2\sigma^2_{wt} + 2(\utilde{X}_t\utilde{L}_t)(1 - \utilde{X}_t\utilde{L}_t)\sum_{j=0}^{t-2} a_{tj}E(w_t w_{t-1-j}) \,,$$

where $B_1(\hat{y}_t)$ was obtained by setting the second term of expression (10) equal to zero, and $B_2(\hat{y}_t)$ was obtained by setting the second term of expression (16) equal to zero. Notice that the bounds are functions of the covariance structure of the sampling errors w_t , and the auxiliary vectors $\utilde{X}_t$. Finally we can define an overall lower bound expression $B(\hat{y}_t)$ as

$$B(\hat{y}_t) \equiv max\{B(\hat{y}_t),\ B(\hat{y}_t),\ 0\} \,, \tag{23}$$

which implies that the mean square error of $\hat{y}_t$ is greater than or equal to $B(\hat{y}_t)$ for all t .

2. Estimation of Mean Square Error

In this section we present methods of estimating the mean square error expressions we calculated in the previous section. The two main mean square error expressions we calculated were given in (10) and (16). The estimation of expression (20) will be the subject of future research. Both expressions (10) and (16) involve variances and covariances of the sampling errors. For most surveys we can expect to have a survey-based estimate s^2_{wt} of σ^2_{wt} for each time period $t = 1, \ldots, T$. For some surveys we may also have a survey-based estimate $s_{wt}(j+1)$ of the covariance $E(w_t w_{t-1-j})$ for all t and $j = 0, \ldots, t-2$. If these variance and covariance estimates are unbiased we can construct an unbiased estimator of the mean square error from (10) as follows

$$m_1(\hat{y}_t) = (1 - \utilde{X}_t\utilde{L}_t)^2(Y_t - \utilde{X}_t\hat{\utilde{\beta}}_{t-1})^2 + (2\,\utilde{X}_t\utilde{L}_t - 1)s^2_{wt} + 2(1 - \utilde{X}_t\utilde{L}_t)\sum_{j=0}^{t-2} a_{tj}s_{wt}(j+1) \,. \tag{24}$$

While $m_1(\hat{y}_t)$ is unbiased we would expect it to have an unacceptably high variance since we are using $(Y_t - \utilde{X}_t\hat{\utilde{\beta}}_{t-1})^2$ as an estimator of its own expectation. This points out the fact that the main problem in estimating the mean square error is accurately estimating $E\{(Y_t - \utilde{X}_t\hat{\utilde{\beta}}_{t-1})^2\}$. In practice we would use the estimator $m_1^+(\hat{y}_t) = max\{m_1(\hat{y}_t),\ b(\hat{y}_t)\}$ where $b(\hat{y}_t)$ is the estimated lower bound in (23) .

Alternatively we could try estimating (16) but this involves the unknown vector $vech(\ V\{\ \hat{\utilde{\beta}}_{t-1}\ \}\ )$ as well as q_t . In general we would never expect to have survey-based estimates of $vech(\ V\{\ \hat{\utilde{\beta}}_{t-1}\ \}\ )$ and q_t as we did for the variances and covariances of the sampling errors above. For notational convenience we define the vector $\utilde{\theta}_t$ to be equal to $vech(\ V\{\ \hat{\utilde{\beta}}_{t-1}\ \}\ )$ for all components except the first, and define the first component to be the first component of $vech(\ V\{\ \hat{\utilde{\beta}}_{t-1}\ \}\ )$ plus q_t^2 . Thus, if we were able to construct an estimator $\hat{\utilde{\theta}}_t$ of $\utilde{\theta}_t$ then we could consider mean square error estimators of the form

$$m_2(\hat{y}_t) = (1 - \utilde{X}_t\utilde{L}_t)^2(\utilde{X}_t\otimes\utilde{X}_t)\utilde{\Phi}_p\hat{\utilde{\theta}}_t + (\utilde{X}_t\utilde{L}_t)^2 s^2_{wt} + 2(\utilde{X}_t\utilde{L}_t)(1 - \utilde{X}_t\utilde{L}_t)\sum_{j=0}^{t-2} a_{tj}s_{wt}(j+1) \,. \tag{25}$$

One method of constructing such an estimator $\hat{\utilde{\theta}}_t$ is as follows. Form the random variable Z_{1t} by

$$Z_{1t} = (Y_t - \utilde{X}_t\hat{\utilde{\beta}}_{t-1})^2 - s^2_{wt} + 2\sum_{j=0}^{t-2} a_{tj}s_{wt}(j+1) \,, \tag{26}$$

for all t . If we again assume that the variance and covariance estimates s^2_{wt} and $s_{wt}(j+1)$ are unbiased, then by using (15) we can write

$$Z_{1t} = (\utilde{X}_t\otimes\utilde{X}_t)\utilde{\Phi}_p\utilde{\theta}_t + \eta_{1t} \,, \quad \text{for } t = 1, \ldots, T \tag{27}$$

where $E(\eta_{1t}) = 0$ for all t . Thus, estimating $\utilde{\theta}_t$ is equivalent estimating the regression problem defined by (27) in which the regression coefficients change over time. There are many different methods of estimating the coefficients for this problem, and the performance of any particular method depends upon the behavior of the sequence of $\utilde{\theta}_t$ vectors. One simple solution would be to use the ordinary least squares estimate $\hat{\utilde{\theta}}_T$ from the regression of Z_{1t} on $(\utilde{X}_t\otimes\utilde{X}_t)\utilde{\Phi}_p$ for $t = 1, \ldots, T$. In the next section were we work an example with data taken from the Bureau of Labor Statistics, we actually use weighted least squares instead of ordinary least squares. In practice we would use the estimator $m_2^+(\hat{y}_t) = max\{m_2(\hat{y}_t),\ b(\hat{y}_t)\}$ where $b(\hat{y}_t)$ is the estimated lower bound in (23) .

Up to this point we have assumed that estimates of the covariances as well as the variances of the sampling errors are available. For many surveys, only variance estimates s^2_{wt} will be available. Unless the covariance are known to be zero (as in the case of independent sampling) additional assumptions will have to be made in order to estimate the mean square error. We will assume there exists a constant l such that

$$E(w_t w_{t-s}) = 0 \quad \text{for } s > l, \tag{28}$$

for all t. In general the other covariances will be assumed to be unknown. Such a covariance assumption could, for instance, be a reasonable approximation to the covariance structure of a rotating panel design in which panels of sample units enter a sample together and stay in for exactly l time periods before being removed from the sample. With the assumption in (28) it will be convenient to define the l-dimensional vector $\underset{\sim}{\gamma}_t$ as follows

$$\underset{\sim}{\gamma}_t = (\gamma_t(1), \gamma_t(2), \ldots, \gamma_t(l))', \tag{29}$$

where

$$\gamma_t(s) = E(w_t w_{t-s}),$$

for all t. Also define the l-dimensional vector $\underset{\sim}{A}_t$ as follows

$$\underset{\sim}{A}_t = (a_{t0}, a_{t1}, \ldots, a_{tl-1}), \tag{30}$$

where the a_{tj} were defined in (9). Thus, if we were able to construct an estimator $\hat{\underset{\sim}{\theta}}_t$ of $\underset{\sim}{\theta}_t$ and an estimator $\hat{\underset{\sim}{\gamma}}_t$ of $\underset{\sim}{\gamma}_t$ then we could consider mean square error estimators of the form

$$m_3(\hat{y}_t) = (1 - \underset{\sim}{X}_t\underset{\sim}{L}_t)^2(\underset{\sim}{X}_t \otimes \underset{\sim}{X}_t)\underset{\sim}{\Phi}_p\hat{\underset{\sim}{\theta}}_t + \tag{31}$$

$$+ (\underset{\sim}{X}_t\underset{\sim}{L}_t)^2 s_{wt}^2 + 2(\underset{\sim}{X}_t\underset{\sim}{L}_t)(1 - \underset{\sim}{X}_t\underset{\sim}{L}_t)\underset{\sim}{A}_t\hat{\underset{\sim}{\gamma}}_t.$$

One method of constructing such estimators $\hat{\underset{\sim}{\theta}}_t$ and $\hat{\underset{\sim}{\gamma}}_t$ is as follows. Form the random Z_{2t} by

$$Z_{2t} = (Y_t - \underset{\sim}{X}_t\hat{\underset{\sim}{\beta}}_{t-1})^2 - s_{wt}^2, \tag{32}$$

for all t. If we again assume that the variance estimate s_{wt}^2 is unbiased, then by using (15) we can write for $t = 1, \ldots, T$

$$Z_{2t} = (\underset{\sim}{X}_t \otimes \underset{\sim}{X}_t)\underset{\sim}{\Phi}_p\underset{\sim}{\theta}_t - 2\underset{\sim}{A}_t\underset{\sim}{\gamma}_t + \eta_{2t}, \tag{33}$$

where $E(\eta_{2t}) = 0$ for all t. Thus, estimating $\underset{\sim}{\theta}_t$ and $\underset{\sim}{\gamma}_t$ is equivalent to estimating the regression coefficients in the regression of Z_{2t} on $[(\underset{\sim}{X}_t \otimes \underset{\sim}{X}_t)\underset{\sim}{\Phi}_p, -2\underset{\sim}{A}_t]$ where the coefficients are allowed to change with t. While there are many methods of estimating the coefficient vector for this problem one simple solution would be to use ordinary least squares. In the next section we work an example where weighted least squares is used instead of ordinary least squares. In practice we would use the estimator $m_3^+(\hat{y}_t) = max\{m_3(\hat{y}_t), b(\hat{y}_t)\}$ where $b(\hat{y}_t)$ is the estimated lower bound in (23).

Note that given the estimator $\hat{\underset{\sim}{\gamma}}_T$ we can construct an additional estimator of the mean square error, namely

$$m_4(\hat{y}_t) = (1 - \underset{\sim}{X}_t\underset{\sim}{L}_t)^2(Y_t - \underset{\sim}{X}_t\hat{\underset{\sim}{\beta}}_{t-1})^2 \tag{34}$$

$$+ (2\underset{\sim}{X}_t\underset{\sim}{L}_t - 1)s_{wt}^2 + 2(1 - \underset{\sim}{X}_t\underset{\sim}{L}_t)\underset{\sim}{A}_t\hat{\underset{\sim}{\gamma}}_t.$$

In practice we would use the estimator $m_4^+(\hat{y}_t) = max\{m_4(\hat{y}_t), b(\hat{y}_t)\}$ where $b(\hat{y}_t)$ is the estimated lower bound in (23).

We next define one last pair of estimators, but in order to do so we need to define some notation. Define the l-dimensional column vector

$$\underset{\sim}{\rho}_t = (\rho_t(1), \rho_t(2), \ldots, \rho_t(l))', \tag{35}$$

where

$$\rho_t(s) = Corr(w_t, w_{t-s})$$

$$= \gamma_t(s)\,(\sigma_{wt}^2\sigma_{wt-s}^2)^{-1/2},$$

for all t. Also define the l-dimension row vector

$$\underset{\sim}{C}_t = [a_{t0}(\sigma_{wt}^2\sigma_{wt-1}^2)^{1/2}, \ldots, a_{tl-1}(\sigma_{wt}^2\sigma_{wt-l}^2)^{1/2}], \tag{36}$$

for all t. Thus, if we were able to construct estimators $\hat{\underset{\sim}{\theta}}_t$, $\hat{\underset{\sim}{\rho}}_t$, and $\hat{\underset{\sim}{C}}_t$ then we could consider mean square error estimators of the form

$$m_5(\hat{y}_t) = (1 - \underset{\sim}{X}_t\underset{\sim}{L}_t)^2(\underset{\sim}{X}_t \otimes \underset{\sim}{X}_t)\underset{\sim}{\Phi}_p\hat{\underset{\sim}{\theta}}_t + \tag{37}$$

$$+ (\underset{\sim}{X}_t\underset{\sim}{L}_t)^2 s_{wt}^2 + 2(\underset{\sim}{X}_t\underset{\sim}{L}_t)(1 - \underset{\sim}{X}_t\underset{\sim}{L}_t)\hat{\underset{\sim}{C}}_t\hat{\underset{\sim}{\rho}}_t.$$

The estimator $\hat{\underset{\sim}{C}}_t$ we will use is the one which replaces each of the σ_{wj}^2 in expression (36) with the sample estimate s_{wj}^2 for all j. In order to obtain estimates of $\underset{\sim}{\theta}_t$ and $\underset{\sim}{\rho}_t$ we recommend using weighted least squares, regressing Z_{2t} on $[(\underset{\sim}{X}_t \otimes \underset{\sim}{X}_t)\underset{\sim}{\Phi}_p, -2\hat{\underset{\sim}{C}}_t]$ for all t. One could also consider using an errors-in-variables regression procedure which accounts for the fact that we are using $\hat{\underset{\sim}{C}}_t$ in the regression instead of $\underset{\sim}{C}_t$ which appears in the mean of Z_{2t}, namely

$$E\{Z_{2t}\} = (\underset{\sim}{X}_t \otimes \underset{\sim}{X}_t)\underset{\sim}{\Phi}_p\underset{\sim}{\theta}_t - 2\underset{\sim}{C}_t\underset{\sim}{\rho}_t, \quad \text{for } t = 1, \ldots, T.$$

In practice we would use the estimator $m_5^+(\hat{y}_t) = max\{m_5(\hat{y}_t), b(\hat{y}_t)\}$ where $b(\hat{y}_t)$ is the estimated lower bound in (23).

Finally, given an estimator $\hat{\underset{\sim}{\rho}}_t$ we can construct the estimator

$$m_6(\hat{y}_t) = (1 - \underset{\sim}{X}_t\underset{\sim}{L}_t)^2(Y_t - \underset{\sim}{X}_t\hat{\underset{\sim}{\beta}}_{t-1})^2 \tag{38}$$

$$+ (2\underset{\sim}{X}_t\underset{\sim}{L}_t - 1)s_{wt}^2 + 2(1 - \underset{\sim}{X}_t\underset{\sim}{L}_t)\hat{\underset{\sim}{C}}_t\hat{\underset{\sim}{\rho}}_t.$$

In practice we would use the estimator $m_6^+(\hat{y}_t) = max\{m_6(\hat{y}_t), b(\hat{y}_t)\}$ where $b(\hat{y}_t)$ is the estimated lower bound in (23). In the next section we demonstrate the use of the estimators presnted in this section.

3. An Example

Each month the Bureau of Labor Statistics (BLS) produces monthly estimates of the unemployment rate for each of the fifty states and the District of Coulmbia. For simplicity, we will refer to these as *state level* estimates. The primary source of data for these unemployment rate estimates is the Current Population Survey (CPS). The CPS is a national probability sample of approximately 59,000 households, and is conducted by the Census Bureau for the Bureau of Labor Statistics. It is possible to compute estimates of the monthly unemployment rate at the state level directly from the CPS data, but the reliability of these estimates can differ greatly because the CPS sample is not distributed equally among the fifty states and the District of Columbia. As of January 1989 BLS publishes the direct CPS unemployment rate estimates for eleven states, while for the remaining 39 states and the District of Columbia it uses estimates based upon the Kalman filter which incorporates CPS data along with auxiliary information. We next describe an estimator which is closely related to one which BLS uses for the monthly unemployment rate estimates for one of these states, and we demonstrate the calculation of the mean square error expressions derived in Section 2 for this estimator.

We will be using monthly data for a particular state (which will remain unidentified) covering the period of time from January 1976 to February 1989, which comprises 158 months. In terms of the notation defined in Section 1 we define the problem for month t as follows

Y_t = Direct CPS monthly state unemployment rate estimate,

y_t = True population monthly state unemployment rate,

w_t = CPS sampling error in the direct estimate,

where $t = 1, \ldots, 158$. The goal is to estimate the true, but unobservable, monthly unemployment rate y_t . The estimator we will describe makes use of three auxillary variables for month t

X_{2t} = CES state employment to population ratio,

X_{3t} = National CPS entrant rate,

X_{4t} = Unemployment Insurance claims rate for the state ,

where by definition $X_{1t} \equiv 1$ for all t , and $\underset{\sim}{X}_t \equiv (1, X_{2t}, X_{3t}, X_{4t})$. Tiller (1989) gives an in-depth description of these variables, as well as the estimation methodology used by BLS.

The monthly estimator we will examine $\hat{y}_t$ is given by

$$\hat{y}_t = \underset{\sim}{X}_t \hat{\underset{\sim}{\beta}}_t \quad \text{for} \quad t = 1, \ldots, 158$$

where

$$\hat{\underset{\sim}{\beta}}_t = \hat{\underset{\sim}{\beta}}_{t-1} + \underset{\sim}{L}_t (Y_t - \underset{\sim}{X}_t \hat{\underset{\sim}{\beta}}_{t-1})$$

$$\underset{\sim}{L}_t = (\underset{\sim}{P}_{t-1} + \underset{\sim}{Q}) \underset{\sim}{X}_t' f_t^{-1}$$

$$f_t = 1 + \underset{\sim}{X}_t (\underset{\sim}{P}_{t-1} + \underset{\sim}{Q}) \underset{\sim}{X}_t'$$

$$\underset{\sim}{P}_t = \underset{\sim}{P}_{t-1} + \underset{\sim}{Q} - \underset{\sim}{L}_t \underset{\sim}{L}_t' f_t$$

where $\hat{\underset{\sim}{\beta}}_0 \equiv \underset{\sim}{0}$ and $\underset{\sim}{P}_0 \equiv \kappa \underset{\sim}{I}$, where κ is a large positive number. The matrix $\underset{\sim}{Q}$ is a diagonal matrix with diagonal elements given by

$$q_{11} = 0.00003612, \qquad q_{22} = 0.00003889,$$

$$q_{33} = 0.00003613, \qquad q_{44} = 0.00003634.$$

The estimator described above is the result of fitting a *variable coefficients model* to the data where the regression parameters are assumed to follow a random walk. The elements of the matrix $\underset{\sim}{Q}$ were actually derived by the method of maximum likelihood under the variable coefficients model specification and the assumption of normality. Notice that the matrix $\underset{\sim}{Q}$ remains fixed, and does not depend on t .

Along with the direct CPS unemployment rate estimates Y_t over time we also have estimates s_{wt}^2 of their variance over time. These variance estimates are actually calculated from generalized variance functions, and even though they are not exactly unbiased estimates we will assume they are unbiased here for the purposes of illustration. We next demonstrate the calculation of the six mean square error estimators given in Section 2. Each of these estimators will make use of the variance estimates s_{wt}^2 .

Unfortunately we do not have direct estimates of the covariances of the CPS sampling errors w_t over time. In order to demonstrate the calculation of estimators m_1 and m_2 we calculated indirect estimates of the sampling covariances over time in the following way. Train, Cahoon and Makens (1978), among other things, calculated average correlations over time for the CPS estimates of total national unemployment. As a rough approximation we decided to use their estimated correlations for the first four lags for our state data, namely

Lag-1 correlation = 0.45 , Lag-2 correlation = 0.28 ,

Lag-3 correlation = 0.17 , Lag-4 correlation = 0.08 , and assume that the correlations for higher lags for the state were zero. We then estimated the sampling covariances over time as

$$s_{wt}(1) = (0.45)(s_{wt}^2 s_{wt-1}^2)^{1/2} ,$$

$$s_{wt}(2) = (0.28)(s_{wt}^2 s_{wt-2}^2)^{1/2} ,$$

$$s_{wt}(3) = (0.17)(s_{wt}^2 s_{wt-3}^2)^{1/2} ,$$

$$s_{wt}(4) = (0.08)(s_{wt}^2 s_{wt-4}^2)^{1/2} ,$$

and the the higher order covariances were set equal to zero. With these approximations we estimated m_1 and m_2. Estimator m_2 was estimated using weighted least squares where the weight for the *t-th* observation is t. As would be expected estimator m_1 appeared more erratic than m_2 with the distinction being most evident after about 100 months. On the average, both m_1 and m_2 seem to indicate that $\hat{y}_t$ has a slightly smaller mean square error over time than the CPS direct estimate Y_t.

Estimators m_3 and m_4 were computed using weighted least squares, where the *t-th* observation received a weight of t. In order to to compute these estimators we made the assumption that $l = 4$, which is consistent with the correlation assumption made in calculating m_1 and m_2. We found that m_3 was more variable than m_1, and that m_4 was more variable than m_2. One reason for this is that the estimated lower bound $b(\hat{y}_t)$ used for both m_3 and m_4 often takes on the value zero. This occurred because the $\hat{\underset{\sim}{\gamma}}_t$ estimates tended to be unstable, often taking on negative values large enough in magnitude to drive the estimated lower bound to zero. Because of this problem we do not recommend using estimators m_3 and m_4, but instead we recommend using m_5 and m_6 which we describe next.

Estimators m_5 and m_6 were computed using nonlinear weighted least squares, where the *t-th* observation received a weight of t. The NLIN procedure of PC SAS was used to compute the estimates $\hat{\underset{\sim}{\theta}}_t$ and $\hat{\underset{\sim}{\rho}}_t$. The NLIN procedure allows the user to place inequality restrictions on the parameters, and we imposed the following restrictions

$$\theta_{tj} > 0 \ , \ \text{for } j = 1, 5, 8, 10$$

where $\underset{\sim}{\theta}_t = (\theta_{t1}, \theta_{t2}, \ldots, \theta_{t10})'$ and

$$0 \leq \rho_{tj} \leq 1 \ , \ \text{for } j = 1, 2, 3, 4$$

where $\underset{\sim}{\rho}_t = (\rho_{t1}, \rho_{t2}, \rho_{t3}, \rho_{t4})'$. The theta restrictions may look strange, but these theta parameters correspond to the diagonal elements of $V\{\hat{\underset{\sim}{\beta}}_{t-1}\}$ defined in (14) which are positive by definition. We restricted the elements of $\underset{\sim}{\rho}_t$ to be in the parameter space of positive correlations because based on the paper of Train, Cahoon and Makens (1978), as well as intuition resulting from an understanding of the CPS sample design and estimation methodology, we would expect the sampling errors w_t to be positively correlated rather than negatively correlated. The estimates m_5 and m_6 are much more well behaved than the corresponding estimates m_3 and m_4. It was also observed that m_6 is very similar to m_1 and m_5 is very similar to m_2, with the main difference being that m_5 and m_6 were alightly larger than m_1 and m_2.

This ends our analysis of the state data. Our primary interest was demonstrate the calculation of the six mean square error estimators presented in Section 2 and not to draw conclusions about either the behavior of the unemployment rate for the state, or to assess the quality of the estimator $\hat{y}_t$. Our overall finding is that m_5 and m_6 are superior to m_3 and m_4, because by bounding some of the estimated parameters in the nonlinear regression we were able to reduce the erratic behavior in the final mean square error estimates.

Remarks

Any opinions expressed are those of the author and do not reflect policy of the Bureau of Labor Statistics.

References

Bell, William R., and Hillmer, Steven C. (1987). Time series methods for survey estimation. United States Bureau of the Census SRD Research Report Number CENSUS/SRD/RR-87/20.

Bell, William R., and Hillmer, Steven C. (1989). Modeling time series subject to sampling error. United States Bureau of the Census SRD Research Report Number CENSUS/SRD/RR-89/01.

Eltinge, John L., and Fuller, Wayne A. (1989). Time series random component models for sample surveys. Winter Meeting of the American Statistical Association.

Fuller, Wayne A. (1987). *Measurement Error Models.*

Scott, A. J., and Smith T. M. F. (1974). Analysis of repeated surveys using time series methods. *Journal of the American Statistical Association* 69 674-678.

Scott, A. J., Smith, T. M. F., and Jones, R. G. (1977). The application of time series methods to the analysis of repeated surveys. *International Statistical Review* 45 13-28.

Tam, S. M. (1987). Analysis of repeated surveys using a dynamic linear model. *International Statistical Review* 44 63-73.

Tiller, Richard (1989). A Kalman filter approach to labor force estimation using Current Population Survey data. Annual Meeting of the American Statistical Association, Washington D. C.

Train, G. L., Cahoon, and Makens, P. (1978). The Current Population Survey variances, inter-relationships, and design effects. *Proc. Survey Research Sec.* 443-448. Amer. Statist. Assoc., Washington D.C.