ESTIMATION ERROR COVARIANCE IN REGRESSION WITH SEQUENTIALLY VARYING PARAMETERS*

By BARR ROSENBERG

In regression with parameters that obey a first-order Markov process, it is useful to estimate the parameters' trajectory over time. Formulas to compute the time series of "smoothed" estimators of the stochastic parameters, and the estimation error variance at any time, are well known. This article derives the intertemporal estimation error covariance and thence the sampling distribution of estimated parameter shifts. Concomitant new intertemporal covariances are linked by the "join" matrices of "forward" and "backward" filters.

1. INTRODUCTION

This note completes the results of the preceding paper (Cooley, Rosenberg, and Wall, 1977 [CRW]), and the notation and definitions of that paper continues here. The covariance between estimation errors for parameter vectors \( \beta_s \) and \( \beta_t \) at times \( s \) and \( t \):

\[
P_{s,t} = E((\eta_1 - \beta_s)(\eta_2 - \beta_t)^\prime)
\]

has previously been reported only for the case where \( s = t \) (Rauch, 1963; Rauch, Tung, and Striebel, 1965; Meditch, 1967; and references in the preceding paper). Formulas for these intraperiod variances (denoted by \( \bar{P}_{s,t} \) rather than \( P_{s,t} \) to simplify notation) were given in CRW. In an earlier work (Rosenberg, 1968), the interperiod estimation error covariances were derived but in a relatively obscure form. The purpose of this note is to provide easily accessible and computationally convenient formulas.

Inference concerning the historical behavior of parameters generally requires both intraperiod error variances and interperiod covariances. For example, consider the parameter shift between periods \( s \) and \( t \), \( \beta_s - \beta_t \). The minimum mean square error linear unbiased estimator or MMSE (Rosenberg, 1973) is \( h_{s,t} - h_{t,t} \), that is, the difference between the smoothed estimators at these two time points. The mean square error of the estimator is:

\[
\text{MSE}(\beta_s - \beta_t) = E((h_{s,t} - h_{t,t} - (\beta_s - \beta_t))^2)
\]

Inference concerning the historical behavior of the parameters requires both intraperiod error variances and interperiod covariances. For example, consider the parameter shift between periods \( s \) and \( t \), \( \beta_s - \beta_t \). The minimum mean square error linear unbiased estimator or MMSE (Rosenberg, 1973) is \( h_{s,t} - h_{t,t} \), that is, the difference between the smoothed estimators at these two time points. The mean square error of the estimator is:

\[
\text{MSE}(\beta_s - \beta_t) = E((h_{s,t} - h_{t,t} - (\beta_s - \beta_t))^2)
\]

*This research was begun during a Woodrow Wilson Dissertation Fellowship at Harvard University and was completed with the support of National Science Foundation Grant GS-306. The helpful comments of Iaffa Yehudai are gratefully acknowledged.
Thus, the mean square error is a function of intraperiod error variances and interperiod covariances.

Interperiod estimation error covariances are also needed for inference concerning the values of parameters at two or more times. Let \( t_1, \ldots, t_p \) be \( p \) given times. The variance matrix of estimation errors for \( \beta_{t_1}, \ldots, \beta_{t_p} \) is made up of intraperiod variance and interperiod covariance submatrices: the submatrix in location \((i,j)\) is \( P_{i,j} \).

2. A RECURSION FOR THE ERROR VARIANCE MATRICES

The estimates for the parameter vectors, \( \beta_1, \ldots, \beta_t \), were given by CRW through the "gain" matrices of the "forward" and "backward filters":

(1) (forward) \[
K_t = (\Phi^t + H_{t,t} \Phi^{t-1} Q)^{-1}
\]

(2) (backward) \[
J_t = \Phi^t (I + G_{t,t} \Phi^t)^{-1}.
\]

The computational procedure involves recursive application of these filters to accumulate the forward information matrices \( H_{t,t} \), and filtered variables \( f_{1,t}, t = 1, \ldots, T \) and the corresponding backward terms \( G_{t,t} \), and \( r_{1,t}, t = T, \ldots, 1 \). Then the information from forward and backward filters is summed to obtain the smoothed estimates \( \hat{\beta}_{t} \) and their estimation variances \( P_{t,t} \).

The recursive formulas hide the simple relationship between successive estimation error covariances. In fact, the variance matrices are themselves lined by the recursion:

(3) \[
P_{t,t} = K_t P_{t+1,t} J_t^{-1}.
\]

To verify this equality, postmultiply by \( J_t \) and invert both sides of the equality to obtain:

\[
P_{t,t} = K_t P_{t+1,t} J_t^{-1} \iff J_t^{-1} P_{t,t} J_t^{-1} = P_{t+1,t} K_t^{-1}.
\]

Then substitute the expression (CRW.16) for the information matrices, taking note of the fact that \( H_{t+1,t} + G_{t+1,t} = H_{t,t} + G_{t,t} \).

\[
\iff J_t^{-1} (H_{t+1,t} + G_{t+1,t}) = (H_{t,t} + G_{t,t}) K_t^{-1}
\]

Next, express \( H_{t+1,t} \) in terms of \( H_t \), by (CRW.7), and \( G_{t,t} \) in terms of \( G_{t+1,t} \) by (CRW.2).

\[
\iff J_t^{-1} (H_{t,t} + J_t G_{t+1,t}) = (K_t H_t K_t^{-1} + G_{t,t}) K_t^{-1}.
\]
Then substitute (1) and (2) for the gain matrices.

$$I + G_{t+1} = \Phi I H_{t+1} + G_{t+1} \Psi$$

$$- (I + H_{t+1} \Phi^{t+1} Q)^{-1} H_{t+1} \Phi^{t+1} + G_{t+1} (I + Q \Phi^{t+1} H_{t+1}).$$

The terms in $G_{t+1}$ now cancel out.

$$\Phi^{t+1} H_{t+1} = (I + H_{t+1} \Phi^{t+1} Q)^{-1} H_{t+1} \Phi^{t+1} (I + Q \Phi^{t+1} H_{t+1}).$$

Premultiplication by $\Phi^{t+1} H_{t+1} \Phi^{t+1} Q$ leads to the desired equality.

$$\Phi^{t+1} H_{t+1} = (I + H_{t+1} \Phi^{t+1} Q)^{-1} H_{t+1} \Phi^{t+1} (I + Q \Phi^{t+1} H_{t+1}).$$

3. COVARIANCE BETWEEN ESTIMATION ERRORS IN DIFFERENT PERIODS

For any two time periods, $s$ and $t$, $s < t$, it is shown below that

$$P_{s,t} = K_{s} K'_{s} + \ldots + K_{t} K'_{t} = P_{s,t} J_{s} \ldots J_{t}.$$  

Equations (3) and (4) provide a complete recursion among all covariance matrices. For example, by recursive application of (3) to (4), every matrix may be expressed in terms of the variance matrix of the latest estimate:

$$P_{s,t} = K_{s} K'_{s} + \ldots + K_{t} K'_{t} = P_{s,t} J_{s} \ldots J_{t}$$

The derivation of (4) is rather tedious and will be presented here for the simplified case where $\Phi = I$. Let $H_{t} = x_{t} \sigma^{2}$ be the information matrix in period $t$, and let $h_{t} = x_{t} \sigma^{2}$. Also, let $K_{n} = K_{n+1} K_{n+2} \ldots K_{t}$ and $J_{n} = J_{n+1} J_{n+2} \ldots J_{t}$ be the sequential products of gain matrices over the interval $1 \leq t$. (Notice the complicated subscript convention; for $s = t$, $J_{n} = K_{n} = 1$.) The estimation error $(b_{s,t} - b_{t})$ can be expressed as a function of the stochastic terms in the model by collecting the recursive equations in CRW, as:

$$\begin{align*}
(b_{s,t} - b_{t}) &= P_{s,t} \left[ \sum_{t+1}^{s} K_{n} x_{t} / \sigma^{2} - \sum_{t+1}^{s} K_{n} H_{t} u_{t} \\
&\quad + \sum_{t+1}^{s} J_{n} x_{t} / \sigma^{2} + \sum_{t+1}^{s} J_{n} G_{t} u_{t} \right].
\end{align*}$$

Taking the expression (6) for two periods, $s$ and $t$, and evaluating the
product moment, one finds:

\[ P_{t+1} = P_t \left[ \left( \sum_{k=1}^{t-1} K_{k} H_{k} K_{k+1} - \sum_{k=1}^{t-1} K_{k} H_{-k} K_{k+1} \right) K_{-1} + \left( \sum_{k=1}^{t-1} K_{k} H_{k} K_{k+1} - \sum_{k=1}^{t-1} \sum_{i=1}^{t} J_{si} G_{i-t-i} Q H_{i} K_{k+1} \right) K_{-1} \right. \\
\left. + \sum_{k=1}^{t-1} J_{k} H_{k} J_{k+1} + \sum_{k=1}^{t-1} \sum_{i=1}^{t} J_{k} G_{i-t-i} Q G_{i-t-i} J_{k+1} \right] P_t. \]

The three expressions are the respective contributions to error from stochastic terms up to period \( t \), between \( s \) and \( t \), and beyond \( t \). The second summation in each expression can be rewritten as the negative of the first summation plus one or two additional terms. For example, since

\[ K_{s-1} = 1 + H_{s-1} Q \]

by (1), it follows that

\[ \sum_{k=1}^{t-1} K_{k} H_{s-1} Q H_{s-1} K_{s} = \sum_{k=1}^{t-1} K_{k} (K_{s-1} - 1) H_{s-1} K_{s} = \sum_{k=1}^{t-1} K_{s-1} (H_{s-1} - K_{s-1}) K_{s}. \]

Regrouping terms and adding and subtracting \( K_{s} H_{0} K_{s} \) yields:

\[ = -K_{s} H_{0} K_{s} + \sum_{k=1}^{t-1} K_{k} (H_{s-1} + H_{k}) K_{k} + K_{0} H_{s-1}. \]

Furthermore, since \( H_{0} = 0 \), \( K_{0} = I \), and \( H_{k} = H_{k+1} K_{k+1} K_{k} + H_{k} \), from (CRW, 10), the preceding expression reduces to:

\[ = - \sum_{k=1}^{t-1} K_{s} H_{k} K_{s} + H_{s-1}. \]

Similar simplifications for the second and third expressions yield:

\[ \sum_{j=1}^{t-1} J_{s} G_{j-t-j} Q H_{s-1} K_{s} = H_{s-1} K_{s} + \sum_{k=1}^{t-1} K_{s} H_{k} K_{s} - J_{s} H_{s}, \]

\[ \sum_{j=1}^{t-1} J_{s} G_{j-t-j} Q G_{j-t-j} J_{s} = G_{j-t-j} - \sum_{j=1}^{t-1} J_{s} H_{j} J_{s}. \]
Substitution of these expressions into (7) results in:

\[
\begin{align*}
P_{n+1} &= P_{n+1}(H_{n+1} + K_{n+1} + J_{n+1}H_{n+1}) \\
& \quad + \left(J_{n+1}G_{n+1}\right)P_{n+1} \\
& = P_{n+1}(H_{n+1} + G_{n+1})P_{n+1} = P_{n+1}J_{n+1}P_{n+1} = P_{n+1}J_n.
\end{align*}
\]

This is the second equality in (4). The first equality then follows by repeated application of (3) to this equation. Q.E.D.

4. MEAN SQUARE ESTIMATION ERROR FOR A PARAMETER SHIFT

It is often of interest to compare values of the parameter vector at two points in time by means of the estimated parameter shift \( (b_{11} - b_{10}) \). From (3) and (4), the mean square estimation error matrix is:

\[
\text{MSE}((b_{11} - b_{10}) - (d_1 - d_0)) = P_{1/1} - P_{1/1} - P_{1/1} + P_{1/1} - P_{1/1}K_{1/1} + P_{1/1}.
\]

5. COMPUTATION

These results show an elegant relationship among the estimation error covariances. The smoothing procedure proposed in CRW, which does not follow this pattern, involves \( 3T \) matrix inversions; \( 7 \) operations to compute the forward gain matrices, by formula (1) above; \( T \) to compute the backward gain matrices by formula (2) above; and \( T \) to compute the variance matrices by formula (16) in CRW. Once these matrices have been computed and stored, no further matrix inversions are required to compute all intertemporal covariance matrices via formula (4).

Simplifying formulas do exist that reduce computational difficulty by exploiting the structure implicit in equation (3). The procedure derived in Rosenberg (1968, chapters 4 and 5) entailed no matrix inversions in the filtering state and required only a single matrix inversion overall. Alternatively, recursive application of formula (3), as a substitute for (CRW, 17) reduces matrix inversions to \( 2T + 1 \). However, despite the gain in computational efficiency, these algorithms seem to be worthless because of numerical instability. The CRW filtering formulas are inherently stable because the errors in computation are attenuated due to subsequent multiplication by gain matrices with eigenvalues smaller than unity. Formula (3) uses the product of one matrix with eigenvalues smaller than unity, with another having eigenvalues greater than unity, to extrapolate information from one period to the next. The errors in such a procedure inevitably build up over a long sample series and, in our ex-
perience, have proved uncontrollable. The only practical usefulness of
formula (3) above is as an error check to validate an algorithm based on
(CRW. 16). Formula (4) is the efficient method to compute intertemporal
estimation error covariance.

Submitted: March 1977
Revised: June 1977

REFERENCES

Cooley, T. L., Rosenberg, B., and Wall, K. "A Note on Optimal Smoothing for Non
Varying Coefficient Problems." Annals of Economic and Social Measurement
(Russ Meditch, Jami S. 1967). "Orthogonal Projection and Discrete Optimal Linear Smoothing" -
IEEE Journal on Control 5 (February).

Ranich, H. E. 1963. "Solutions to the Linear Smoothing Problem." IEE Transactions on
Automatic Control AC-8 (October).


Journal 1 (August).

Rosenberg, B. 1968. "Varying-Coefficient Regression." Ph.D. dissertation, Department of
Economics, Harvard University.

Social Measurement 2 (October).