This PDF is a selection from an out-of-print volume from the National Bureau of Economic Research

Volume Title: Annals of Economic and Social Measurement, Volume 6, number 3

Volume Author/Editor: NBER

Volume Publisher:

Volume URL: http://www.nber.org/books/aesm77-3

Publication Date: July 1977

Chapter Title: Generalized Least Squares Applied to Time Varying Parameter Models

Chapter Author: Donald T. Sant

Chapter URL: http://www.nber.org/chapters/c10557

Chapter pages in book: (p. 301 - 314)

Annals of Economic and Social Measurement, 6/3, 1977

GENERALIZED LEAST SQUARES APPLIED TO TIME VARYING PARAMETER MODELS

BY DONALD T. SANT*

This paper shows the formal equivalence of Kalman filtering and smoothing techniques to generalized least squares. Smoothing and filtering equations are presented for the case where some of the parameters are constant. The paper further shows that generalized least squares will produce consistent estimates of those parameters that are not time varying.

When linear models have been used to model economic problems, it has been useful many times to allow for parameter variation across observations. Various statistical procedures have been developed to estimate and test this hypothesis of nonstable regression coefficients.[1] Recently, it has been recognized that a technique known as the Kalman filter has useful applications in estimating economic models with nonconstant coefficients.[2] The purpose of this paper is to show the formal equivalence of Kalman filtering and smoothing techniques with generalized least squares, to derive the Kalman filter and smoother without assuming all of the parameters are subject to stochastic variation, and to show that generalized least squares produces consistent estimates of those parameters which are not subject to stochastic change. An immediate use of this last result is in the model of Cooley and Prescott (1973, 1976). In their model (1973), only the intercept is subject to stochastic change, so generalized least squares will produce consistent estimates of all the slope coefficients.

The framework for presenting the filtering and smoothing techniques will be in a linear time-varying parameters model where the regression parameters follow a simple random walk.[3] Suppose the scalar y_t is generated by the model

$$y_t = x_t\beta_t + \epsilon_t \tag{1}$$

where x_t is a k-dimensional row vector of exogenous variables at time or observation t. It is also assumed that the k-dimensional column vector β_t evolves according to the structure

*Helpful comments of Gregory Chow and Roger Gordon are gratefully acknowledged.

[1]See the October 1973 issue of the *Annals of Economic and Social Measurement* for a collection of papers describing the different techniques and models that have been analyzed.

[2]For a description of the algorithm see Athans (1974) and for a use in testing hypothesis see Garbade (1975).

[3]More complex parameter variation can be analyzed, but it mainly serves to complicate the mathematics without substantially altering the results.

(2)
$$\beta_t = \beta_{t-1} + u_t$$

where ϵ_t and u_t are unobserved error terms with mean 0. There are T ordered observations on y_t and x_t, $t = 1, T$ and ϵ_t and u_t have variances-covariances described by

(3)
$$E(\epsilon_i \epsilon_j) = \delta_{ij}\sigma^2$$

(4)
$$E(u_i u_j') = \delta_{ij}\sigma^2 P$$

(5)
$$E(\epsilon_i u_j) = 0.$$

where δ_{ij} is the Kronecker delta and P is assumed known. The significance of the filtering and smoothing algorithms is that they give estimates of the β_t based on certain subsets of the T observations and what the relationships are between the different estimates.

If we let $\hat{\beta}_t(t')$ be an estimate of β_t using observations 1 through t' and $\sigma^2 R_t(t')$ be the covariance matrix of $\hat{\beta}_t(t')$, the Kalman filter is a sequential algorithm for estimating $\hat{\beta}_t(t)$ given by

(6)
$$\hat{\beta}_t(t) = \hat{\beta}_t(t-1) + K_t(y_t - x_t\hat{\beta}_t(t-1))$$

where

(7)
$$K_t = R_t(t-1)x_t'[x_t R_t(t-1)x_t' + 1]^{-1}$$

(8)
$$R_t(t) = R_t(t-1) - K_t x_t R_t(t-1)$$

(9)
$$\hat{\beta}_t(t-1) = \hat{\beta}_{t-1}(t-1)$$

(10)
$$R_t(t-1) = R_{t-1}(t-1) + P.$$ [4]

I. Equivalence with Generalized Least Squares.

If we stack the observations into a form amenable to the application of GLS (generalized least squares), the proof of equivalence follows from applying certain lemmas on matrix inverses given in the appendix. Let

$$Y_t = \begin{pmatrix} y_1 \\ \vdots \\ y_t \end{pmatrix} \qquad X_t = \begin{pmatrix} x_1 \\ \vdots \\ x_t \end{pmatrix}$$

$$E_t = \begin{pmatrix} \epsilon_1 \\ \vdots \\ \epsilon_t \end{pmatrix} \qquad U_t = \begin{pmatrix} u_2 \\ \vdots \\ u_t \end{pmatrix}$$

[4] Various algorithms are given in Sage and Melsa (1971).

$$A_t = \begin{bmatrix} x_1 & x_1 & & \cdots & x_1 \\ 0 & x_2 & & \cdots & x_2 \\ 0 & 0 & x_3 & \cdots & x_3 \\ & & & \ddots & \\ & & & & x_{t-1} \\ 0 & & \cdots & & 0 \end{bmatrix}$$

The relationship of the first t observations is now given by

(11) $$Y_t = X_t\beta_t + E_t - A_t U_t.$$

The covariance matrix of the error terms in equation 11 is

(12) $$E(E_t - A_t U_t)(E_t - A_t U_t)' = \sigma^2[I_t + A_t(I_{t-1} \otimes P)A_t']^5 = \sigma^2\Omega_t$$

Applying GLS to equation 11 gives us

(13) $$\hat{\beta}_t(t) = (X_t'\Omega_t^{-1}X_t)^{-1}X_t'\Omega_t^{-1}Y_t$$

and

(14) $$R_t(t) = (X_t'\Omega_t^{-1}X_t)^{-1}.$$

If one is estimating recursively, i.e., for the appropriate stacking procedure

(15) $$Y_{t-1} = X_{t-1}\beta_{t-1} + E_{t-1} - A_{t-1}U_{t-1}$$

a GLS estimate of β_{t-1} using the first $t - 1$ observations is

(16) $$\hat{\beta}_{t-1}(t-1) = (X_{t-1}'\Omega_{t-1}^{-1}X_{t-1})^{-1}X_{t-1}'\Omega_{t-1}^{-1}Y_{t-1}$$

with

(17) $$R_{t-1}(t-1) = (X_{t-1}'\Omega_{t-1}^{-1}X_{t-1})^{-1}$$

The presentation of the proof is to show that one can obtain the relationships given by equations 6 through 10 from the relationships given by equations 13 through 17.

Let G_t be the $(t - 1) \times t$ dimensional matrix, $G_t = [I_{t-1}\vdots 0]$, which removes the last row of a $t \times \ell$ matrix, so that

(18) $$\begin{aligned} G_t Y_t = Y_{t-1} &= G_t X_t \beta_t + G_t E_t - G_t A_t U_t \\ &= X_{t-1}\beta_t + E_{t-1} - G_t A_t U_t. \end{aligned}$$

GLS applied to 18 gives us

(19) $$\hat{\beta}_t(t-1) = (X_{t-1}'(G_t\Omega_t G_t')^{-1}X_{t-1})^{-1}X_{t-1}'(G_t\Omega_t G_t')^{-1}Y_{t-1}.$$

[5]The notation I_k will mean the identity matrix of dimension k.

Using the definition of Ω_t given in 12

(20) $$G_t\Omega_t G_t' = G_t G_t' + G_t A_t(I \otimes P)A_t' G_t'$$

$$= G_t G_t' + (A_{t-1} \vdots X_{t-1})(I \otimes P)\begin{pmatrix} A_{t-1}' \\ \cdots \\ X_{t-1}' \end{pmatrix}$$

$$= I_{t-1} + A_{t-1}(I_{t-2} \otimes P)A_{t-1}' + X_{t-1}PX_{t-1}'$$

(21) $$= \Omega_{t-1} + X_{t-1}PX_{t-1}'.$$

Using lemma 3 in the Appendix, equation 19 can be written as

$$\begin{aligned}\hat{\beta}_t(t-1) &= (R_{t-1}(t-1) + P)X_{t-1}'(\Omega_{t-1} + X_{t-1}PX_{t-1}')^{-1}Y_{t-1} \\ &= (R_{t-1}(t-1) + P) \\ &\quad (X_{t-1}'\Omega_{t-1}^{-1}Y_{t-1} - (R_{t-1}(t-1) + P)^{-1}PX_{t-1}'\Omega_{t-1}^{-1}Y_{t-1}) \\ &= \hat{\beta}_{t-1}(t-1)\end{aligned}$$

where lemma 1 in the Appendix has been used for $(\Omega_{t-1} + X_{t-1}PX_{t-1}')^{-1}$. This along with lemma 3 demonstrates equations 9 and 10 of the Kalman filter. For the rest of the derivation consider rewriting equation 11 as

(22) $$\begin{pmatrix} G_tY_t \\ \cdots \\ y_t \end{pmatrix} = \begin{pmatrix} G_tX_t \\ \cdots \\ x_t \end{pmatrix}\beta_t + \begin{pmatrix} G_tE_t \\ \cdots \\ \epsilon_t \end{pmatrix} - \begin{pmatrix} G_tA_tU_t \\ \cdots \\ 0 \end{pmatrix}$$

Then it follows that equation 13 is equivalent to

(23) $$\hat{\beta}_t(t) = (X_{t-1}'(G_t\Omega_tG_t')^{-1}X_{t-1} + x_t'x_t)^{-1}(X_{t-1}'(G_t\Omega_tG_t')^{-1}Y_{t-1} + x_t'y_t).$$

Using lemma 4, equation 23 becomes

$$\begin{aligned}&(R_t(t-1) - (1 + x_tR_t(t-1)x_t)^{-1}R_t(t-1)x_t'x_tR_t(t-1) \\ &\quad (X_{t-1}'(G_t\Omega_tG_t')^{-1}Y_{t-1} + x_t'y_t) = \hat{\beta}_t(t-1) - K_tx_t\hat{\beta}_t(t-1) \\ &\quad + (1 + x_tR_t(t-1)x_t')K_ty_t - K_tx_tR_t(t-1)x_t'y_t \\ &\quad = \hat{\beta}_t(t-1) + K_t(y_t - x_t\hat{\beta}_t(t-1)).\end{aligned}$$

This completes the derivation of equation 6 through 10 since equation 8 is just lemma 4.

To point out the major differences between this derivation and Duncan and Horn (1972) one should observe that it was not required for P to be nonsingular to derive the filtering equations so it permits some of the parameters to be constant. Further, a proper prior distribution for β_0 is not necessary to get estimating equations since the filter can be initialized at observation $k + 1$ by applying GLS to the first k observation and obtaining the estimate $\hat{\beta}_k(k)$.

II. Smoothing

The filtering algorithm does not use all the information available in T observations to estimate the parameters β_t, $t = 1, T$. Smoothing algorithms are available which, when given estimates of the form $\hat{\beta}_t(t)$ and observations taken at times $t + 1, T$, use all the relevant sample information in estimating each β_t. If we call estimates of β_t using all observations $1, T\ \hat{\beta}_t(T)$, and $\sigma^2 R_t(T)$ the covariance matrix of $\hat{\beta}_t(T)$, an algorithm for relating all smoothed estimates would be of the form

(24) $$\hat{\beta}_t(T) = \hat{\beta}_t(t) + H_t(\hat{\beta}_{t+1}(T) - \hat{\beta}_t(t))$$

(25) $$H_t = R_t(t)[R_t(t) + P]^{-1}$$

(26) $$R_t(T) = R_t(t) + H_t[R_{t+1}(T) - R_{t+1}(t)]H_t'.$$

These estimating equations are equivalent to GLS applied to all the observations.

If we let

$$Y_{t+1,T} = \begin{pmatrix} y_{t+1} \\ \vdots \\ y_T \end{pmatrix} \qquad X_{t+1,T} = \begin{pmatrix} x_{t+1} \\ \vdots \\ x_T \end{pmatrix}$$

$$E_{t+1,T} = \begin{pmatrix} \epsilon_{t+1} \\ \vdots \\ \epsilon_T \end{pmatrix} \qquad U_{t+1,T} = \begin{pmatrix} u_{t+1} \\ \vdots \\ u_T \end{pmatrix}$$

$$A_{t+1,T} = \begin{bmatrix} x_{t+1} & 0 & \cdots & & 0 \\ x_{t+2} & x_{t+2} & 0 & \cdots & 0 \\ \vdots & & & & \\ x_T & x_T & \cdots & & x_T \end{bmatrix}$$

we can relate future observations to β_t as

(27) $$Y_{t+1,T} = X_{t+1,T}\beta_t + E_{t+1,T} + A_{t+1,T}U_{t+1,T}$$

Applying generalized least squares to equation 27 would give us an estimate of β_t say $\hat{\beta}_t(t + 1, T)$ based on observations $t + 1$ through T. As in the last section a "backward" filter can be derived utilizing the structure of equation 27 to obtain the recursive estimating equations

$$\hat{\beta}_t(t+1, T) = \hat{\beta}_{t+1}(t+1, T) \tag{28}$$

$$V_t(t+1, T) = V_{t+1}(t+1, T) + P \tag{29}$$

where $\sigma^2 V_k(i, j)$ has the interpretation of being the covariance matrix of the GLS estimate of β_k using the consecutive observation i through j.

The generalized least squares estimate of β_t using all the observations can finally be obtained by combining $\hat{\beta}_t(t)$ and $\hat{\beta}_t(t+1, T)$ resulting in the estimate

$$\hat{\beta}_t(T) = (R_t^{-1}(t) + V_t^{-1}(t+1, T))^{-1}(R_t^{-1}(t)\hat{\beta}_t(t) + V_t^{-1}(t+1, T)\hat{\beta}_t(t+1, T)). \tag{30}$$

Equation 30 can be shown to be equivalent to the sequential procedure of equations 24 through 26 by combining the forward and "backward" filtering equations.

To simplify notation, let

$$B_1 = \hat{\beta}_t(t) = \hat{\beta}_{t+1}(t) \tag{31}$$

$$B_2 = \hat{\beta}_t(t+1, T) = \hat{\beta}_{t+1}(t+1, T) \tag{32}$$

$$R = R_t(t) \tag{33}$$

$$V = V_{t+1}(t+1, T). \tag{34}$$

Using the filtering equations we know that

$$R_{t+1}(t) = R + P \tag{35}$$

$$V_t(t+1, T) = V + P \tag{36}$$

so the *GLS* estimates of β_t and β_{t+1} using all observations are

$$\hat{\beta}_t(T) = [R^{-1} + (V+P)^{-1}]^{-1}[R^{-1}B_1 + (V+P)^{-1}B_2] \tag{37}$$

$$\hat{\beta}_{t+1}(T) = [(R+P)^{-1} + V^{-1}]^{-1}[(R+P)^{-1}B_1 + V^{-1}B_2] \tag{38}$$

Lemma:

If R and V are positive definite $k \times k$ matrices and P is a $k \times k$ positive semidefinite matrix then

$$[R^{-1} + (V+P)^{-1}]^{-1} = R + H\{[(R+P)^{-1} + V^{-1}]^{-1} - (R+P)\}H'$$

where $H = R(R+P)^{-1}$.

Proof

Apply lemma 4 (since $V + P$ is nonsingular) and obtain

$$
\begin{aligned}
[R^{-1} + (V + P)^{-1}]^{-1} &= R - R(V + P + R)^{-1}R \\
&= R - R(R + P)^{-1}(R + P)(V + P + R)^{-1} \\
&\quad (R + P)(R + P)^{-1}R \\
&= R - H(V(R + P)^{-1} + I)^{-1}(R + P)H' \\
&= R - H((R + P)^{-1} + V^{-1})^{-1}V^{-1}(R + P)H' \\
&= R + H\{[(R + P)^{-1} + V^{-1}]^{-1} - (R + P)\}H'
\end{aligned}
$$

by lemma 2.

Interpreting this lemma gives us equation 25 and 26 of the smoothing algorithm. Expanding equation 37 we have

$$
\begin{aligned}
\text{(39)} \quad \hat{\beta}_t(T) &= B_1 - HB_1 + H((R + P)^{-1} + V^{-1})^{-1}(R + P)^{-1}B_1 \\
&\quad + [R^{-1} + (V + P)^{-1}]^{-1}(V + P)^{-1}B_2 \\
&= B_1 - HB_1 + H((R + P)^{-1} + V^{-1})^{-1}(R + P)^{-1}B_1 \\
&\quad + R[R + V + P]^{-1}B_2 \\
&= B_1 - HB_1 + H((R + P)^{-1} + V^{-1})^{-1}(R + P)^{-1}B_1 \\
&\quad + R(R + P)^{-1}[(R + P)^{-1} + V^{-1}]^{-1}V^{-1}B_2 \\
&= B_1 + H[((R + P)^{-1} + V^{-1})^{-1}((R + P)^{-1}B_1 \\
&\quad + V^{-1}B_2) - B_1] \\
&= \hat{\beta}_t(t) + H[\hat{\beta}_{t+1}(T) - \hat{\beta}_t(t)].
\end{aligned}
$$

The importance of knowing P is now readily apparent from standard proofs of the properties of GLS estimators. Nice sampling distributions for use in hypothesis testing depend on knowing P and not having to estimate P. If P is not known, asymptotic properties of GLS estimators using a consistent estimate of P can be investigated, but in the general problem just presented, no one has yet demonstrated that P can be consistently estimated. In the situation where only one coefficient is stochastic, P and σ^2 can be consistently estimated, as was demonstrated by Cooley and Prescott (1973) for the case of a stochastic intercept and by Cooley and Prescott (1976) for a slope coefficient. The next section will show that the non-stochastic coefficients can be consistently estimated by GLS even though the time varying parameters cannot be estimated consistently. Combining these results will give us asymptotic distributions for the non-stochastic parameters which can be used as the basis for hypothesis testing.

III. Consistent Estimates

A useful result of the smoothing algorithm is that it permits us to obtain theorems regarding large sample properties of $\hat{\beta}_t$ without looking at $(X'\Omega_t^{-1}X)$. The following section will be devoted to showing that GLS will yield consistent estimates of those parameters that are not subject to stochastic variation.

In what follows assume that P is of the form

(40)
$$P = \begin{bmatrix} P_1 & 0 \\ 0 & 0 \end{bmatrix}$$

and

(41)
$$\beta_t' = (\beta_{1,t}', \beta')$$

where β is the $k_2 \times 1$ vector of coefficients that are constant across observations.

Proposition:

The GLS estimate $\hat{\beta}$ of β using all the observations 1 through T is invariant with respect to the parameterization regarding $\beta_{1,j}$ $1 \leq j \leq T$.

Proof

The proof consists of showing that from the smoothing algorithm

(42)
$$\hat{\beta}_{t+1}(T) - \hat{\beta}_t(T) = \begin{pmatrix} \hat{\xi}_t \\ 0 \end{pmatrix}$$

which is an equivalent statement to the proposition.

From equation 24

$$\hat{\beta}_t(T) = \hat{\beta}_t(t) + R_t(t)[R_t(t) + P]^{-1}(\hat{\beta}_{t+1}(T) - \hat{\beta}_t(t))$$

but $R_t(t)[R_t(t) + P]^{-1}$ is of the form

$$\begin{bmatrix} Q_1 & Q_2 \\ 0 & I_{k_2} \end{bmatrix}$$

which implies equation 42. This follows from theorems on inverses of partitioned matrices.

It follows from this proposition that the variance of β is invariant with respect to the parameterization and of the actual realization of Y and $\beta_{1,j}$.[6] That is the variance of the paramèters not subject to stochastic change do not depend on which $\beta_{1,j}$ $1 \leq j \leq T$ that is also estimated in the block GLS procedure. This observation gives us a proof of consistency at least in the situation where one considers that X is constant in repeated samples of size T.

Consider increasing the sample by observing

$$y_t = x_t\beta_t + \epsilon_t \quad \text{for } t = T + 1, 2T \tag{43}$$

but where we have set

$$x_t = x_{t-T} \quad t = T + 1, 2T. \tag{44}$$

We now have two samples and writing them in the form of equation 11 yields

$$Y_{2T} = \begin{pmatrix} X_T \\ X_T \end{pmatrix} \begin{pmatrix} \beta_{1,2T} \\ \beta \end{pmatrix} + E_{2T} - A_{2T}U_{2T}. \tag{45}$$

when they are stacked and both samples are parameterized using $\beta_{1,2T}$. Alternatively, if we don't stack them and parameterize them separately we have two equations of the form

$$Y_T = X_T \begin{pmatrix} \beta_{1,T} \\ \beta \end{pmatrix} + E_T - A_T U_T \tag{46}$$

$$Y_{T+1,2T} = X_T \begin{pmatrix} \beta_{1,2T} \\ \beta \end{pmatrix} + E_{T+1,2T} - A_T U_{T+2,2T}. \tag{47}$$

The matrices X_T and A_T are the same in equations 46 and 47 since the exogenous variables are the same. Applying GLS to equations 46 and 47 would give us two estimates of β having exactly the same covariance matrix. This follows from the fact that the exogenous variables are the same in the two samples and the covariance matrix of the residuals in the two particular parameterizations chosen are identical. Now having two estimates of β with the same covariance, one can obtain an estimate of β linear in the original $2T$ observations that has a covariance matrix equal to one half the covariance matrix of the estimator derived from using either sample. This estimator (the average of the two estimators which each use only half the total sample) must not have a smaller covariance matrix than the GLS estimator applied to equation 45 since the GLS esti-

[6]The invariance property of the variance besides being intuitively obvious is also easily proven using equation 26.

mator is best linear unbiased. In the limit then, the variance of the GLS estimator for β must converge to zero at least in the case of repeated exogenous variables. The case of repeated samples is not really that restrictive. The asymptotic results are used as approximations to the distributions we are actually interested in, and the case of repeated samples is just the easiest way to obtain these approximations. The case without repeated samples is not as intuitive and easy to understand. The coefficients which are subject to stochastic variations cannot be estimated consistently, so the standard procedure of finding conditions on X that insure the covergence of $(X'\Omega_t^{-1}X)/T$ does not apply. An orthogonality result like the previous proposition is needed to show the consistency of the estimators. An immediate use of the property that the non-varying coefficients can be consistently estimated is in deriving distribution theory in the model of Cooley and Prescott (1973, 1976). In their model (1973), only the intercept is varying randomly, but the appropriate variances are not assumed to be known. They show that these variances can be consistently estimated which when combined with the property that the slope coefficients can be consistently estimated, gives us the usual large sample approximations to the distributions of the slope coefficients. The large sample approximate distribution is the same as the true distribution when P is assumed to be known.[7]

APPENDIX

Lemma 1

Let Ω be a $t \times t$ non-singular positive definite matrix, P a $k \times k$ positive semidefinite matrix and X a $t \times k$ matrix. Then

$$(\Omega + XPX')^{-1} = \Omega^{-1} - \Omega^{-1}X(X'\Omega^{-1}X)^{-1}[(X'\Omega^{-1}X)^{-1} + P]^{-1}PX'\Omega^{-1}$$

Proof

$$\begin{aligned}
&\{\Omega + XPX'\}\{\Omega^{-1} - \Omega^{-1}X(X'\Omega^{-1}X)^{-1}[(X'\Omega^{-1}X)^{-1} + P]^{-1}PX'\Omega^{-1}\}\\
&= I_t - X(X'\Omega^{-1}X)^{-1}[(X'\Omega^{-1}X)^{-1} + P]^{-1}PX'\Omega^{-1} + XPX'\Omega^{-1}\\
&\quad - XP[(X'\Omega^{-1}X)^{-1} + P]^{-1}PX'\Omega^{-1}\\
&= I_t + X\{I_k - (X'\Omega^{-1}X)^{-1}[(X'\Omega^{-1}X)^{-1} + P]^{-1} - P[(X'\Omega^{-1}X)^{-1}\\
&\quad + P]^{-1}\}PX'\Omega^{-1}\\
&= I_t + X\{I_k - [(X'\Omega^{-1}X)^{-1} + P]\,[(X'\Omega^{-1}X)^{-1} + P]^{-1}\}PX'\Omega^{-1}\\
&= I_t
\end{aligned}$$

[7]A proof of this type of result is given in Amemiya (1973).

Lemma 2

For the same matrices as in lemma 1

$$[(X'\Omega^{-1}X)^{-1} + P]^{-1} = X'\Omega^{-1}X - [(X'\Omega^{-1}X)^{-1} + P]^{-1}PX'\Omega^{-1}X$$

Proof

Premultiply by $[(X'\Omega^{-1}X)^{-1} + P]$ and postmultiply by $(X'\Omega^{-1}X)^{-1}$

Lemma 3

For the same matrices as in lemma 1

$$X'(\Omega + XPX')^{-1}X = [(X'\Omega^{-1}X)^{-1} + P]^{-1}$$

Proof

From lemma 1

$$X'(\Omega + XPX')^{-1}X = X'\Omega^{-1}X - [(X'\Omega^{-1}X)^{-1} + P]^{-1}PX'\Omega^{-1}X$$
$$= [(X'\Omega^{-1}X)^{-1} + P]^{-1} \quad \text{by lemma 2.}$$

Lemma 4

For the same matrices as in lemma 1 but P non-singular then

$$(\Omega + XPX')^{-1} = \Omega^{-1} - \Omega^{-1}X(P^{-1} + X'\Omega^{-1}X)^{-1}X'\Omega^{-1}$$

Proof

See Duncan and Horn (1972) or it follows directly from Lemma 1.

References

Amemiya, Takeshi (1973). "Generalized Least Squares with an Estimated Autocovariance Matrix," *Econometrica* 41, 723–732.

Athans, Michael (1974). "The Importance of Kalman Filtering Methods for Economic Systems," *Annals of Economic and Social Measurement* 3, 49–64.

Cooley, T. F. and E. C. Prescott (1973). "The Adaptive Regression Model," *International Economic Review* XIV, 365–371.

——— (1976). "Estimation in the Presence of Stochastic Parameter Variation," *Econometrica* 44, 167–184.

Duncan, D. B. and S. D. Horn (1972). "Linear Dynamic Recursive Estimation from Viewpoint of Regression Analysis," *Journal of the American Statistical Association* 67, 815–821.

Garbade, Kenneth (1975). "Two Methods for Examining the Stability of Regression Coefficients," Econometric Research Program, Memorandum No. 186, Princeton University.

Sage, Andrew and James Melsa (1971). *Estimation Theory with Applications to Communications and Control.* McGraw-Hill, New York.

Princeton University

Annals of Economic and Social Measurement, 6/3, 1977

"GENERALIZED LEAST SQUARES APPLIED TO TIME VARYING PARAMETER MODELS: A COMMENT"

BY THOMAS COOLEY*

The paper by Donald Sant provides a useful service to the profession by showing clearly the formal equivalence of Kalman filtering and smoothing methods with generalized least squares. In addition, he derives the appropriate form of the filtering and smoothing equations for a model with both constant and time varying parameters. Although much of what is contained in this paper has either appeared elsewhere in the literature or is known to practitioners (and therefore is *assumed* to be obvious to others), much of the literature is somewhat inaccessible. The treatment of the varying parameter estimation problem as a generalized least squares problem in Cooley and Prescott (1973, 1976) was motivated, at least in part, by a desire to treat the problem in a way that is familiar to economists. Sant's paper does much to demystify the Kalman filtering approach and the purpose of this comment is to abet that process by making a few other useful references to the literature on this topic.

The derivation of the Kalman filter as a generalized least squares estimator is generalized from the paper of Duncan and Horn (1972) by the use of matrix relations which allow the variance covariance matrix of the states to be singular. This permits treatment of constant and varying parameters in the same model. This approach is also mentioned in an unpublished thesis by Rosenberg (1968) although the point is not made as explicitly. The smoothing equations are derived as a combination of a "forward" and "backward" filter. This approach first was proposed in an unpublished thesis by Fraser (1967) in the aforementioned thesis by Rosenberg and in a paper by Fraser and Potter (1967).

A recent paper by Cooley, Rosenberg and Wall (1976) derives the smoothing equations for a model with both constant and varying parameters as a combination of a backward and forward "information" filter. The information form has the advantage that it represents the filter in terms of the inverse of the covariance matrix of the states and thus eliminates the need to initialize the filter using a subset of observations as the author suggests. The initialization procedure proposed by Sant has also been suggested by Kaminski, Bryson and Schmidt (1974).

As a final comment I would like to point out that the problem of consistently estimating P (the variance covariance matrix of the states) in the

*University of California and N.B.E.R.

more general case has been addressed by Mehra (1970, 1972) and Cooley and Wall (1976) but it has not, to my knowledge, been unequivocally resolved.

References

Cooley, T. F. and E. C. Prescott (1973), "The Adaptive Regression Model," *International Economic Review*, XIV, 365–371.

———, (1976), "Estimation in the Presence of Stochastic Parameter Variation," *Econometrica* 44, 167–184.

Cooley, T. F., B. M. Rosenberg, K. D. Wall (1976), "A Note On Optimal Smoothing for Time Varying Parameter Problems," Forthcoming, *Annals of Economic and Social Measurement*.

Duncan, D. B. and S. D. Horn (1972), "Linear Dynamic Recursive Estimation from the Viewpoint of Regression Analysis," *Journal of the American Statistical Association*, 67, 815–821.

Fraser, D. C. (1967), "A New Method for the Optimal Smoothing of Data," Unpublished Sc.D. Thesis, Department of Aeronautics, Massachusetts Institute of Technology.

Fraser, D. C. and J. E. Potter (1967), "The Optimum Linear Smoother as a Combination of Two Optimum Linear Filters," *IEEE Transactions on Automatic Control*, Vol. AC-14, No. 4.

Kaminski, P. G., A. E. Bryson and S. F. Schmidt (1974), "Discrete Square Root Filtering: A Survey of Current Techniques," *IEEE Transactions on Automatic Control*, Vol. AC-18, No. 6.

Mehra, R. K., (1970), "On the Identification of Variances and Adaptive Kalman Filtering," *IEEE Transactions on Automatic Control*, Vol. AC-15, No. 2.

——— (1972), "Approaches to Adaptive Filtering," *IEEE Transactions on Automatic Control*, AC-17, No. 5.

Rosenberg, B. M. (1968), "Varying Parameter Estimation," Unpublished Ph.D. Thesis, Department of Economics, Harvard University.

Rosenberg, B. M. (1973), "A Survey of Stochastic Parameter Regression," *Annals of Economic and Social Measurement*, Vol. 2, No. 4, 381–397.