NBER WORKING PAPER SERIES

A GENERAL ALGORITHM FOR SIMULTANEOUS ESTIMATION OF CONSTANT AND RANDOMLY-VARYING PARAMETERS IN LINEAR RELATIONS

Alexander H. Sarris*

Working Paper No. 38

COMPUTER RESEARCH CENTER FOR ECONOMICS AND MANAGEMENT SCIENCE
National Bureau of Economic Research, Inc.
575 Technology Square
Cambridge, Massachusetts 02139

April 1974

Preliminary: not for quotation

NBER working papers are distributed informally and in limited numbers for comments only. They should not be quoted without written permission.

This report has not undergone the review accorded official NBER publications; in particular, it has not yet been submitted for approval by the Board of Directors.

* NBER Computer Research Center and Massachusetts Institute of Technology. Research supported in part by National Science Foundation Grant GJ-1154X3 to the National Bureau of Economic Research, Inc.

Abstract

A recursive algorithm for estimating linear models with both constant and time-varying parameters is derived by maximization of a likelihood function. Recursive formulas are also derived for derivatives of the likelihood function; the derivatives are needed for numerical evaluation of some parameters. Smoothing formulas are also derived. The estimation algorithm is compared with others for similar classes of models.

Contents

1. INTRODUCTION

In recent years many economists and statisticians have observed that the traditional linear model with constant coefficients is inadequate for the description of several phenomena. Individual effects across a population or systematically varying behaviour over time offer two examples. The literature on stochastic parameters has provided several arguments justifying the nonconstancy of parameters over different observations, the reader is referred to Rosenberg (1968) and Cooley (1971) among others. A survey of the existing literature can be found in Rosenberg (1973a). The state of the art in estimation techniques for several stochastic parameter models has been summarized in *Annals of Economic and Social Measurements*, Vol. 2, No. 4 (October 1973).

This paper presents an algorithm for estimating stochastic variations in a model that is more general than the most general model that has been analyzed before, that of Rosenberg (1973b).

In Section 2 the problem that is solved is stated, in Section 3 the recursive formulas for the maximum likelihood estimators are computed, Section 4 restates some well known results about smoothed estimates, and Section 5 gives a brief summary of the results.

2. PROBLEM STATEMENT

The model that will be analyzed is the following

$$y_t = A_t \alpha + B_t \beta_t + \varepsilon_t \tag{1}$$

$t = 1,2,\ldots, T$

where

y_t - $\ell \times 1$ vector of observations at time t

α - $r \times 1$ vector of constant parameters

β_t - $k \times 1$ vector of randomly varying parameters

A_t, B_t - $(\ell x r)$ and $(\ell x k)$ matrices of observations of explanatory variables at time t.

ε_t - $\ell \times 1$ gaussian vector of random terms with the following properties

$$E(\varepsilon_t) = 0 \quad t=1,2,\ldots, T \tag{2}$$

$$E(\varepsilon_i \varepsilon_j') = \sigma^2 Q \, \delta_{ij} \quad i,j = 1,2,\ldots,T \tag{3}$$ *

where Q is a symmetric $\ell x \ell$ positive definite matrix, σ^2 is a scale factor, and δ_{ij} is the Kronecker delta.

The random parameters will be assumed to obey a transition relation of the Markov kind

$$\beta_t = \phi \beta_{t-1} + u_{t-1} \tag{4}$$

where u_t are normal with $E(u_t) = 0 \quad t=1,2,\ldots, T$ (5)

---

* The apostrophe (') denotes transposition.

$$E(u_i u_j') = \sigma^2 R\delta_{ij} \qquad i,j=1,2,\ldots,T \qquad (6)$$

where R is a symmetric kxk positive definite matrix. The initial vector β_1 will be assumed constant but unknown.

A model like this can arise for example in the analysis of a cross-section of time series. Rosenberg (1973b) analyzed a model like this with the exception that he did not consider the constant term α. The recursive formulas that he presented do not hold when such a vector of constant parameters coexists with the vector of stochastic ones. A particular case of model (1) with $\ell=1$, $k=1$ and $B_t=1$ was analyzed by Cooley (1971) and Cooley and Prescott (1973). However, their procedures are not recursive and can quickly become unmanageable for large sample sizes. Sarris (1973) has analyzed a particular case with $A_t=0$ $\ell=1$ but his procedure is also not recursive.

The particular case of $A_t=0$ has also been analyzed extensively in the engineering literature under the name of Kalman filters (see e.g. Sage and Melsa (1971)). The difference between those models and the special case of the one considered here (with $A_t=o$) is that in our case the initial vector β_1 is constant but unknown, instead of having a well defined prior density. The case of constant β_1 gives rise to what is called the starting problem which proved quite troublesome until Rosenberg (1973b) solved it correctly.

We shall assume that the transition matrix ϕ depends on a vector of constant and unknown parameters θ_ϕ (whose location in ϕ is known).

Furthermore Q and R will be assumed to contain vectors of unknown parameters θ_q and θ_r respectively (with known locations in Q and R) We shall define

$$\theta \equiv (\theta_\phi' , \theta_q', \theta_r')' \tag{7}$$

It is important to emphasize the assumptions that Q and R are both positive definite. The most general model one would want to analyze in this framework would be the following

$$y_t = F_t \gamma_t + \varepsilon_t \tag{8}$$

$$\gamma_t = \phi \gamma_{t-1} + \mu_{t-1} \tag{9}$$

where all the previous assumptions on the random terms hold except that

$$E(\mu_i \mu_j') = \sigma^2 R \delta_{ij}$$

with R <u>positive semidefinite</u>. It is clear that (1) is a special case of this model. If γ_1 is given a proper prior density then the results of the Kalman filter literature cover this problem. However, when γ_1 is constant and unknown, a case which is most prevalent in statistics and econometerics, the Kalman algorithm does not hold and Rosenberg's (1973b) algorithm holds only if R is positive definite. The general problem of arbitrary positive semidefinite R has not as yet been solved.

The complete problem is now restated.

<u>Problem</u>. Estimate $\alpha, \beta_1, \sigma^2, \theta$ and $\beta_2\ \beta_3, \ldots, \beta_T$ given the data y_t, A_t, B_t $t=1,2,\ldots,T$

where

$$y_t = A_t \alpha + \beta_t \beta_t + \varepsilon_t \tag{10}$$

$$\beta_t = \phi\beta_{t-1} + \mu_{t-1} \tag{11}$$

3. MAXIMUM LIKELIHOOD ESTIMATION

In this section we shall derive a recursive formula for the likelihood function. Let us make the following definitions

$$y_i^j \equiv \{y_i, y_{i+1}, \ldots, y_j\} \tag{12}$$

$$\mu \equiv \{\alpha, \beta_1, \sigma^2, \theta\} \tag{13}$$

$$A_i^j \equiv \{A_i, A_{i+1}, \ldots, A_j\} \tag{14}$$

$$B_i^j \equiv \{B_i, B_{i+1}, \ldots, B_j\}$$

Then the likelihood of μ given the data is

$$L(\mu; y_1^T, A_1^T, B_1^T) = p(y_1^T|\mu) = \prod_{t=1}^{T} p(y_t \mid y_1^{t-1}, A_1^t, B_1^t, \mu) \tag{16}$$

where

$$p(y_1|y_1^{o}, A_1^1, B_1^1, \mu) = p\,(y_1|A_1, B_1, \mu) \tag{17}$$

3.1 Computation of $p(y_t|y_1^{t-1}, A_1^t, B_1^t, \mu)$

In this subsection we shall compute one of the factors of (16). In the process we shall also derive the Kalman filter for β_t.

$$p(y_t,\beta_t|y_1^{t-1},\ A_1^t,\ B_1^t,\mu) = p(y_t|\beta_t,y_1^{t-1},A_1^t,B_1^t,B_1^t,\mu)$$

$$p(\beta_t|y_1^{t-1},A_1^t,B_1^t,\mu) = p(\beta_t|y_1^t,A_1^t,B_1^t,\mu)$$

$$p(y_t|y_1^{t-1},A_1^t,B_1^t,\mu) \qquad (18)$$

Let us assume that the density $p(\beta_t|y_1^{t-1},A_1^t,B_1^t,\mu)$ is normal with mean $\beta_{t|t-1}$ and covariance matrix $M_{t|t-1}$. From (10) we obtain

$$p(y_t|\beta_t,y_1^{t-1},A_1^t,B_1^t,\mu)=\frac{1}{(2\pi)^{\ell/2}\sigma^{\ell}|Q|^{\frac{1}{2}}}\ \exp\ \{-\frac{1}{2\sigma^2}\ (y_t-A_t\alpha-B_t\beta_t)'Q^{-1}$$

$$(y_t-A_t\ \alpha-B_t\beta_t)\} \qquad (19)^*$$

We thus obtain that the left hand side of (18) is equal to

$$\frac{1}{(2\pi)^{\frac{k+\ell}{2}}\ \ \sigma^{k+\ell}|Q|^{\frac{1}{2}}|M_{t|t-1}|^{\frac{1}{2}}}\quad \exp\ \{-\frac{1}{2\sigma^2}\ [\ (y_t-A_t\alpha-B_t\beta_t)'Q^{-1}$$

$$(y_t-A_t\alpha-B_t\beta) + (\beta_t-\beta_{t|t-1})'M^{-1}_{t|t-1}\ (\beta_t-\beta_{t|t-1})]\} \qquad (20)$$

By rearrangement of the terms inside the brackets of the exponent in (20) we obtain the following equality.

$$(y_t-A_t\alpha-B_t\beta_t)'Q^{-1}(y_t-A_t\alpha-B_t\beta_t) + (\beta_t-\beta_{t|t-1})'M^{-1}_{t|t-1}\ (\beta_t-\beta_{t|t-1})$$

$$=(\beta_t-\beta_{t|t})'M^{-1}_{t|t}\ (\beta_t-\beta_{t|t}) + (y_t-A_t\alpha-B_t\beta_{t|t-1})'\cdot$$

$$(Q + B_tM_{t|t-1}B_t')^{-1}\ (y_t-A_t\alpha-B_t\beta_{t|t-1}) \qquad (21)$$

where

$$M_{t|t}=[M^{-1}_{t|t-1} + B_t'\ Q^{-1}B_t]^{-1}=M_{t|t-1}-M_{t|t-1}B_t'\ [Q+B_tM_{t|t-1}B_t']^{-1}$$

$$B_tM_{t|t-1} \qquad (22)$$

---

* Two bars | | denote determinant.

$$\beta_{t|t} = M_{t|t}[M_{t|t-1}^{-1}\,\beta_{t|t-1} + B_t'\,Q^{-1}(y_t - A_t\alpha.)] =$$

$$\beta_{t|t-1} + M_{t|t-1}\,B_t'\,(Q+B_tM_{t|t-1}B_t')^{-1}(y_t - A_t\alpha - B_t\beta_{t|t-1}) \qquad (23)$$

The formulas (22) - (23) are with minor variations the ones known as the Kalman filter in the engineering literature.

It is now easy to see from (18) and (21) that $p(\beta_t|y_1^t,A_1^t,B_1^t,\mu)$ is normal with mean $\beta_{t|t}$ and covariance matrix $\sigma^2 M_{t|t}$ given in (22) and (23) respectively. Consequently $p(y_t|y_1^{t-1},A_1^t,B_1^t,\mu)$ is also normal with mean $A_t\alpha + B_t\beta_{t|t-1}$ and covariance matrix equal to $\sigma^2(Q+B_tM_{t|t-1}B_t')$. along with the extrapolation formulas

$$\beta_{t+1|t} = \phi\beta_{t|t} \qquad (24)$$

$$M_{t+1|t} = M_{t|t} + R \qquad (25)$$

and the initial conditions

$$\beta_{1|0} = \beta_1 \qquad (26)$$

$$M_{1|0} = 0 \qquad (27)$$

3.2 Computation of the Likelihood Function

After the results of Subsection 3.1 we can apply (16) to obtain an expression for the likelihood function. We obtain

$$L\,(\mu;\, y_1^T,A_1^T,B_1^T) = \prod_{t=1}^{T} p(y_t|y_1^{t-1},A_1^t,B_1^t,\mu)$$

$$= \prod_{t=1}^{T} \frac{1}{(2n)^{\ell/2}\sigma^{\ell}\,|Q+B_tM_{t|t-1}B_t'|^{\frac{1}{2}}} \quad \exp\,\{-\frac{1}{2\sigma^2}\,(y_t - A_t - B_t\beta_{t|t-1})'$$

$$(Q+B_tM_{t|t-1}B_t')^{-1}\ (y_t-A_t\alpha-B_t\,\beta_{t|t-1})\}$$

$$=\frac{1}{(2n)^{T\ell/2}\sigma^{T\ell}\prod_{t=1}^{T}|Q+B_tM_{t|t-1}B_t'|^{\frac{1}{2}}}\exp\{-\frac{1}{2\sigma^2}\sum_{t=1}^{T}(y_t-A_t\alpha-B_t\beta_{t|t-1})'$$

$$(Q+B_tM_{t|t-1}B_t')^{-1}\ (y_t-A_t\alpha-B_t\beta_{t|t-1})\} \qquad (28).$$

For maximum likelihood estimation of μ, this quantity or its logarithm must be maximized with respect to α,β_1,σ^2 and θ.

3.3 Estimation of α,β_1,σ^2

In this section we present recursive estimators of α and β_1 and an estimator for σ^2. It can be easily seen from (22) - (27), that $M_{j|j-1}$ does not depend on β_1,α or σ^2, and that $\beta_{j|j-1}$ depends linearly on both β_1 and α. Let us thus denote

$$\beta_{s|k}=\mu_{s|k}+\Theta_{s|k}\,\alpha+\Xi_{s|k}\,\beta_1 \qquad (29).$$

We can immediately see from (26) that $\mu_{1|0}=0$, $\Theta_{1|0}=0$ $\Xi_{1|0}=I$ (30). (Note that the Θ's and Ξ's are kxr and kxk matrices respectively). Using the formulas (23) and (24) we can obtain recursive relations for μ, Θ, and Ξ. They are the following.

$$\mu_{t+1|t}=\phi\mu_{t|t} \qquad (31)$$

$$\Theta_{t+1|t}=\phi\,\Theta_{t|t} \qquad (32)$$

$$\Xi_{t+1|t}=\phi\Xi_{t|t} \qquad (33)$$

$$\mu_{t|t} = \mu_{t|t-1} + M_{t|t-1} B_t' (Q+B_t M_{t|t-1} B_t')^{-1} (y_t - B_t \mu_{t|t-1}) \quad (34).$$

$$\Theta_{t|t} = \Theta_{t|t-1} - M_{t|t-1} B_t' (Q+B_t M_{t|t-1} B_t')^{-1} (A_t + B_t \Theta_{t|t-1}) \quad (35).$$

$$\Xi_{t|t} = \Xi_{t|t-1} - M_{t\ t-1} B_t' (Q+B_t M_{t|t-1} B_t')^{-1} B_t \Xi_{t|t-1} \quad (36).$$

These formulas are identical to those of Rosenberg (1973b) with the addition of (32) and (35).

We rewrite the exponent of the likelihood function in (28) as follows

$$\sum_{t=1}^{T} (y_t - A_t\alpha - B_t\beta_{t|t-1})' (Q+B_t M_{t|t-1} B_t')^{-1} (y_t - A_t\alpha - B_t\beta_{t|t-1})$$

$$= \sum_{t=1}^{T} (y_t - A_t\alpha - B_t\mu_{t|t-1} - B_t \Theta_{t|t-1} \alpha - B_t \Xi_{t|t-1}\beta_1)' (Q+B_t M_{t|t-1} B_t)^{-1}$$

$$(y_t - A_t\alpha - B_t\mu_{t|t-1} - B_t \Theta_{t|t-1} \alpha - B_t \Xi_{t|t-1}\beta_1) =$$

$$= \sum_{t=1}^{T} (y_t - B_t\mu_{t|t-1})' (Q+B_t M_{t|t-1} B_t')^{-1} (y_t - B_t\mu_{t|t-1}) +$$

$$\alpha' F_T \alpha + \beta_1' H_T \beta_1 + 2\alpha' G_T \beta_1 - 2\alpha' f_T - 2\beta_1' h_T \quad (37).$$

where

$$F_T = \sum_{t=1}^{T} (A_t + B_t \Theta_{t|t-1})' (Q+B_t M_{t|t-1} B_t')^{-1} (A_t + B_t \Theta_{t|t-1}) \quad (38)$$

$$H_T = \sum_{t=1}^{T} \Xi_{t|t-1}' B_t' (Q+B_t M_{t|t-1} B_t')^{-1} B_t \Xi_{t|t-1} \quad (39).$$

$$G_T = \sum_{t=1}^{T} (A_t + B_t \Theta_{t|t-1})' (Q+B_t M_{t|t-1} B_t')^{-1} B_t \Xi_{t|t-1} \quad (40).$$

$$f_T = \sum_{t=1}^{T} (A_t + B_t \Theta_{t|t-1})' (Q + B_t M_{t|t-1} B_t')^{-1} (y_t - B_t \mu_{t|t-1}) \qquad (41).$$

$$h_T = \sum_{t=1}^{T} \Xi'_{t|t-1} B_t' (Q + B_t M_{t|t-1} B_t')^{-1} (y_t - B_t \mu_{t|t-1}) \qquad (42).$$

Maximization of the likelihood function with respect to α and β_1 is equivalent to minimization of the quadratic form in (37). We take the derivatives of that expression and set them equal to zero. We obtain the following two equations

$$F_T \alpha + G_T \beta_1 - f_T = 0 \qquad (43)$$

$$H_T \beta_1 + G_T' \alpha - h_T = 0 \qquad (44).$$

or

$$\begin{bmatrix} F_T & G_T \\ G_T' & H_T \end{bmatrix} \begin{bmatrix} \alpha \\ \beta_1 \end{bmatrix} = \begin{bmatrix} f_T \\ h_T \end{bmatrix} \qquad (45).$$

therefore

$$\begin{bmatrix} \alpha \\ \beta_1 \end{bmatrix} = \begin{bmatrix} F_T & G_T \\ G_T' & H_T \end{bmatrix}^{-1} \begin{bmatrix} f_T \\ h_T \end{bmatrix} \qquad (46).$$

or

$$\hat{\alpha} = (F_T - G_T H_T^{-1} G_T')^{-1} \quad (f_T - G_T H_T^{-1} h_T) \qquad (47).$$

$$\hat{\beta}_1 = (H_T - G_T' F_T^{-1} G_T)^{-1} (h_T - G_T' F_T^{-1} f_T) \qquad (48).$$

The only thing that needs to be examined is the invertibility of the matrix that premultiplies the parameters in (45). We now state and prove a theorem that will guarantee it. Before we do this we define the following two matrices.

$$X_1 \equiv \begin{bmatrix} A_1 \\ A_2 \\ \vdots \\ A_T \end{bmatrix} \tag{50}$$

$$X_2 \equiv \begin{bmatrix} B_1 \\ B_2\phi \\ B_3\phi^2 \\ \vdots \\ B_T\phi^{T-1} \end{bmatrix} \tag{51}$$

the theorem can now be stated.

Theorem 1 . A sufficient condition for the invertibility of the matrix $W \equiv \begin{bmatrix} F_T & G_T \\ G_T' & H_T \end{bmatrix}$ defined in (45) is that the matrix $\begin{bmatrix} X_1 & \vdots & X_2 \end{bmatrix}$ has full column rank.

Proof. Let us assume that the matrix $X \equiv \begin{bmatrix} X_1 & \vdots & X_2 \end{bmatrix}$ has full column rank. Then it will be true that the matrix $X'X$ is positive definite and hence invertible, and also that the matrix $X'\psi X$, where ψ is positive definite, is also positive definite and hence invertible.

We now consider the original data and substitute β_t as follows

$$\beta_t = \phi\beta_{t-1} + u_{t-1} = \phi^2\beta_{t-2} + \phi u_{t-2} + u_{t-1} = \cdots$$
$$= \phi^{t-1}\beta_1 + \sum_{j=1}^{t-1} \phi^{t-1-j} u_j \qquad (52).$$

We then obtain

$$y_1 = A_1\alpha + B_2\beta_1 + \varepsilon_1$$
$$y_2 = A_2\alpha + B_2\phi\beta_1 + B_2 u_1 + \varepsilon_2$$
$$\vdots$$
$$y_T = A_T\alpha + B_T\phi^{T-1}\beta_1 + B_T\left(\sum_{j=1}^{T-1} \phi^{T-1-j} u_j\right) + \varepsilon_T$$

or more compactly.

$$y = X_1\alpha + X_2\beta_1 + v \qquad (53).$$

Where y is the vector of left hand side observations and v is the vector of obviously non-spherical residuals. Let us denote the covariance matrix of v by V. Then it is well known that if the matrix $X \equiv \left[X_1 \vdots X_2\right]$ is of full column rank, the best linear estimates of α and β_1 are obtained by Aitken's estimator which in this case is

$$\begin{bmatrix} \hat{\alpha} \\ \hat{\beta}_1 \end{bmatrix} = (X'V^{-1}X)^{-1} X'V^{-1}y \qquad (54).$$

The same exact expression for $\hat{\alpha}$ $\hat{\beta}_1$ is obtained if we write the likelihood function of y from (53) and maximize it with respect to $\hat{\alpha}$ and $\hat{\beta}_1$.

However, this is exactly how we arrived at (45) only via a recursive formula. Therefore, the normal equations must give the same solution in both cases. We conclude that the matrix W must be equal to the product of a nonsingular matrix and $X'V^{-1}X$. Hence, since $X'V^{-1}X$ is invertible we are done. ||.

The maximum likelihood estimator of σ^2 is easy to find after the expressions for $\hat{\alpha}$ and $\hat{\beta}_1$ are substituted in (28). By differentiating the likelihood function and setting the derivative equal to zero we obtain:

$$\hat{\sigma} = \frac{1}{T\ell} \sum_{t=1}^{T} (y_t - A_t\hat{\alpha} - B_t\mu_{t|t-1} - B_t \;\Theta_{t|t-1}\, \hat{\alpha} - B_t \Xi_{t|t-1}\, \hat{\beta}_1)'$$

$$(Q + B_t M_{t|t-1} B_t')^{-1} \;(y_t - A_t\hat{\alpha} - B_t\mu_{t|t-1} - B_t \;\Theta_{\,t|t-1}\, \hat{\alpha} - B_t \Xi_{t|t-1}\hat{\beta}_1) =$$

$$= \frac{1}{T\ell} \left\{ \left[\sum_{t=1}^{T} (y_t - B_t\mu_{\,t|t-1})' \;(Q + B_t M_{t|t-1} B_t')^{-1} \;(y_t - B_t\mu_{t|t-1}) \right] - \right.$$

$$\left. - \;(f_T' \; h_T') \begin{bmatrix} F_T & G_T \\ G_T' & H_T \end{bmatrix}^{-1} \begin{bmatrix} f_T \\ h_T \end{bmatrix} \right\} \qquad (55).$$

3.4 Estimation of θ.

The estimators of $\beta_{t|t-1}$, $M_{t|t-1}$, α, β_1, and σ^2 all depend on the elements of the vector θ. The log-likelihood function for θ is found after substituting out the estimated values of σ^2, α, and β_1. It is given by the following expression.

$$\log L\,(\theta;\, y_1^T,\, A_1^T,\, B_1^T) = -\frac{T\ell}{2} \log \left\{ \sum_{t=1}^{T} (y_t - B_t\mu_{t|t-1})' \right.$$

$$(Q + B_t M_{t|t-1} B_t')^{-1}$$

$$(y_t - B_T\mu_{t|t-1}) - f_T' (F_T - G_T H_T^{-1} G_T')^{-1} \; f_T - h_T' (H_T - G_T' F_T G_T)^{-1} h_T -$$

$$- 2f_T' (F_T - G_T H_T^{-1} G_T')^{-1} G_T H_T^{-1} h_T \Big\} -$$

$$- \frac{1}{2} \sum_{t=1}^{T} \log |Q + B_t M_{t|t-1} B_t'| + \frac{T\ell}{2} (\log T\ell + \log 2\pi + 1) \quad (56)$$

The above expression is impossible to maximize analytically with respect to θ. Numerical maximization of it will require the derivatives of this expression with respect to the elements of θ. We now proceed to derive recursive formulas for these derivatives. The procedure will be to consider each term of (56) separately and obtain its derivatives with respect to each component of θ. We will bear in mind the decomposition of θ into unknown elements of ϕ, Q and R respectively given in (7).

The quantities whose derivatives will be needed are the following

$$\mu_{t|t-1}, \; Q + B_t M_{t|t-1} B_t', \; |Q + B_t M_{t|t-1} B_t'|, \; F_t, H_T, G_T, f_T, h_T$$

From (38)-(42) it can be seen that the derivatives of F_T, H_T, G_T, f_T and h_T will be determined by the derivatives of $H_{t|t-1}$, $\Xi_{t|t-1}$, $\mu_{t|t-1}$, $Q + B_t M_{t|t-1} B_t'$

So the quantities whose derivatives must be computed are

$$\mu_{t|t-1}, \; H_{t|t-1}, \; \Xi_{t|t-1}, \; Q + B_t M_{t|t-1} B_t' \text{ and } |Q + B_t M_{t|t-1} B_t'|$$

for all $1 \leq t \leq T$

Referring to formulas (31) and (34) the derivative of $\mu_{t|t-1}$ is easily found.

$$\frac{\partial \mu_{t+1|t}}{\partial \theta_{\phi_{ij}}} = \frac{\partial \phi}{\partial \theta_{\phi_{ij}}} \mu_{t|t} + \phi \frac{\partial \mu_{t|t}}{\partial \theta_{\phi_{ij}}} \qquad (57).$$

$$\frac{\partial \mu_{t|t}}{\partial \theta_{\phi_{ij}}} = \frac{\partial \mu_{t|t-1}}{\partial \theta_{\phi ij}} + \frac{\partial M_{t|t-1}}{\partial \theta_{\phi_{ij}}} B_t' (Q+B_t M_{t|t-1} B_t')^{-1} (y_t - B_t \mu_{t|t-1})$$

$$-M_{t|t-1} B_t' (Q+B_t M_{t|t-1} B_t')^{-1} B_t \frac{\partial M_{t|t-1}}{\partial \theta_{\phi_{ij}}} B_t' (Q+B_{t|t-1} B_t')^{-1}$$

$$(y_t - B_t \mu_{t|t-1}) - M_{t|t-1} B_t' (Q+B_t M_{t|t-1} B_t')^{-1} B_t \frac{\partial \mu_{t|t-1}}{\partial \theta_{\phi_{ij}}} \qquad (58)$$

The matrices $\frac{\partial M_{t|t-1}}{\partial \theta_{\phi_{ij}}}$ can be computed recursively by differentiating (25) and (22).

$$\frac{\partial \mu_{t+1|t}}{\partial \theta_{q_{ij}}} = \phi \frac{\partial \mu_{t|t}}{\partial \theta_{q_{ij}}} \qquad (59)$$

$$\frac{\partial \mu_{t|t}}{\partial \theta_{q_{ij}}} = \frac{\partial \mu_{t|t-1}}{\partial \theta_{q_{ij}}} + \frac{\partial M_{t|t-1}}{\partial \theta_{q_{ij}}} B_t'(Q+B_t M_{t|t-1} B_t')^{-1}(y_t - B_t \mu_{t|t-1}) -$$

$$-M_{t|t-1} B_t' (Q+B_t M_{t|t-1} B_t')^{-1} \left(\frac{\partial Q}{\partial \theta_{q_{ij}}} + B_t \frac{\partial M_{t|t-1}}{\partial \theta_{q_{ij}}} B_t' \right)$$

$$(Q+B_t M_{t|t-1} B_t)^{-1}$$

$$(y_t - B_t\mu_{t|t-1}) - M_{t|t-1}B_t' (Q+B_tM_{t|t-1}B_t')^{-1} \; B_t\frac{\partial\mu_{t|t-1}}{\partial\theta_{q_{ij}}} \qquad (60).$$

$\dfrac{\partial M_{t|t-1}}{\partial\theta_{q_{ij}}}$ is again generated recursively by differentiating (25) and (22).

$$\frac{\partial\mu_{t+1}}{\partial\theta_{r_{ij}}} = \phi \; \frac{\partial\mu_{t|t}}{\partial\theta_{r_{ij}}} \qquad (61).$$

$$\frac{\partial\mu_{t|t}}{\partial\theta_{r_{ij}}} = \frac{\partial\mu_{t|t-1}}{\partial\theta_{r_{ij}}} + \frac{\partial M_{t|t-1}}{\partial\theta_{r_{ij}}} \; B_t' \; (Q+B_tM_{t|t-1} \; B_t')^{-1}(y_t - B_t\mu_{t|t-1}) -$$

$$-M_{t|t-1}B_t'(Q+B_tM_{t|t-1}B_t')^{-1} \; B_t \; \frac{\partial M_{t|t-1}}{\partial\theta r_{ij}} \; B_t'(Q+B_tM_{t|t-1}B_t')^{-1}$$

$$(y_t - B_t\mu_{t|t-1}) - M_{t|t-1}B_t'(Q+B_tM_{t|t-1}B_t')^{-1} \; B_t \; \frac{\partial\mu_{t|t-1}}{\partial\theta_{r_{ij}}} \qquad (62).$$

$\dfrac{\partial M_{t|t-1}}{\partial\theta_{r_{ij}}}$ is recursively computed by (25) and (22). The initial conditions for (57)-(62) are all zero.

From (32), (33), (35) and (36) it can be seen that the derivatives of $\Theta_{t|t-1}$, $\Xi_{t|t-1}$, $Q+B_tM_{t|t-1}B_t'$ depend on the derivatives of $\mu_{t|t-1}$ and $M_{t|t-1}$ which we just analyzed.

So the last thing needed to complete the analysis is the computation of the derivatives of the determinant quantity. Let X denote a nxn matrix with rows $x_1, x_2, \ldots, x_n$. Then the following formula can be proved easily by the definition of the determinant function (η is a scalar)

$$\frac{\partial |X|}{\partial \eta} = \begin{vmatrix} \frac{\partial x_1}{\partial \eta} \\ x_2 \\ \vdots \\ x_n \end{vmatrix} + \begin{vmatrix} x_1 \\ \frac{\partial x_2}{\partial_\eta} \\ \vdots \\ x_n \end{vmatrix} + \dots + \begin{vmatrix} x_1 \\ x_2 \\ \vdots \\ \frac{\partial x_n}{\partial_\eta} \end{vmatrix} \tag{63}$$

If we denote the rows of B_t by $b^t_{i\cdot}$ $i=1,\dots \ell$ and the rows of $M_{t|t-1}$ by $m^t_{j\cdot}$ $j=1,\dots, k$, then the ij,th element of $B_t M_{t|t-1} B'_t$ is given by

$$\{B_t M_{t|t-1} B'_t\}_{ij} = \sum_{n=1}^{k} b_{in}\, m_n\, b'_j \tag{64}$$

It can thus be seen that the derivative of the determinant depends upon the derivatives of $M_{t|t-1}$ which we analyzed.

These recursive formulas for the derivatives are lengthier than the computation required to compute the numerical derivatives. The advantage, however, is that we obtain the exact derivatives as opposed to numerical approximations. The exact tradeoffs must be evaluated through numerical experiments which will be deferred to a later paper.

4. SMOOTHED ESTIMATES OF β_t

The procedure described in the previous section led to maximum likelihood estimates of all the unknown constants i.e. $\alpha, \beta_1, \sigma^2$, and θ. We also obtained the maximum likelihood estimator of β_T namely

$$\beta_{t|t}(\hat\beta_1, \hat\alpha, \hat\sigma^2, \hat\theta) = \mu_{T|T}(\hat\theta) + \theta'_{T|T}(\hat\theta)\,\hat\alpha + \Xi_{T|T}(\hat\theta)\,\hat\beta_1 \tag{65}$$

For the solution to be complete we need estimators for β_t $1 < t < T$ that are based on all the data from 1 to T and not only up to t as we have obtained. We thus need "smoothed" estimates of β_t, e. the quantities $\beta_{t|T}$, as opposed to "filtered" estimates $\beta_{t|t}$.

What we would like to obtain ideally would be the density

$p(\beta_2,\beta_3,\ldots,\beta_T|y_1^T)$. As we formulated the problem we have available

$$p(\beta_2^T|y_1^T, \beta_1,\alpha,\sigma^2,\theta) = \frac{p(\beta_2^T,y_1^T|\beta_1,\alpha,\sigma^2,\theta)}{p(y_1^T\;|\beta_1,\alpha,\sigma^2,\theta)} \qquad (66)$$

For the true values of the unknown conditioning parameters the posterior density if normal. We take the empirical Bayes view and estimate β_2^T as the posterior mode of (66) conditioned upon the estimated values of β_1,α,σ^2 and θ.

Sarris (1973) has shown formulas for the posterior mean and covariance matrix of β_2^T but those formulas are not recursive and hence of limited practical value. Recursive formulas have been computed by Rauch et. al. (1965) and are repeated here for completeness.

$$\beta_{t|T} = \beta_{t|t}+M_{t|t}\,\phi'(R+\phi M_{t|t}\,\phi')^{-1}\,(\beta_{t+1|T}-\phi\beta_{t|t}) \qquad (67)$$

$$M_{t|T} = M_{t|t} + M_{t|t}\phi'(R+\phi M_{t|t}\phi')^{-1}\,(M_{t+1|T}-M_{t+1|t})\,\phi\,M_{t|t}$$

$$(R+\phi M_{t|t}\phi')^{-1} \qquad (68)$$

The procedure starts at time T with $\beta_{T|T}$ and $M_{T|T}$ computed by the formulas of Section 3.

Another method that could be used would be to consider the density

$p(\beta_2^T,\, y_1^T \mid \beta_1,\alpha,\sigma^2,\, \theta)$

and maximize it with respect to $\beta_1,\alpha,\sigma^2,\theta$ as well as β_2^T. The conditional estimates of β_2^T conditioned on $\beta_1,\alpha,\sigma^2,\, \theta$ would be the same as obtained above. However, the estimates of $\alpha, \beta_1, \sigma^2, \theta$ would not be the same. This method has been used by Bar-Shalom (1973) and Masiello (1972). It is not clear at this point how these two different methods of estimating the unknown parameters compare.

5. CONCLUSIONS

The contribution of this paper is two fold. First it presents a recursive solution to the general problem of estimating a combination of constant and time varying parameters. Then it presents recursive formulas for the derivatives that are undoubtedly required to numerically maximize the likelihood function with respect to the parameters that appear nonlinearly.

It is hoped that this type of approach will help the more general problem of estimating combinations of time varying and constant parameters which obey some non-random restrictions.

REFERENCES

(1) Bar-Shalom, Y., "Optimal Simultaneous State Estimation and Parameter Identification in Linear Discrete-Time Systems," *IEEE Transactions on Automatic Control*, Vol. AC-17, No. 3, June 1973.

(2) Cooley, T.F., "Estimation in the Presence of Sequential Parameter Variation," Unpublished Ph.D. Dissertation, Department of Economics, University of Pennsylvania, 1971.

(3) Cooley, T.F., and E.C. Prescott, "The Adaptive Regression Model," *International Economic Review*, 14, June 1973.

(4) Masiello, R.D., "Adaptive Modelling and Control of Electric Power Systems," Ph.D. Thesis, Department of Electrical Engineering, Massachusetts Institute of Technology, October 1972.

(5) Rauch, H.E., F. Tung, and C.T. Striebel, "Maximum Likelihood Estimates of Linear Dynamic Systems," *Journal of the American Institute of Aeronautics and Astronautics*, 3, August 1965.

(6) Rosenberg, B., "Varying-Parameter Estimation," Unpublished Ph.D. Dissertation, Department of Economics, Harvard University, 1968.

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

(8) Rosenberg, B., "The Analysis of a Cross Section of Time Series by Stochastically Convergent Parameter Regression," *Annals of Economic and Social Measurement*, Vol. 2, No.4, October 1973.

(9) Sage, A.P. and J.L. Melsa, *Estimation Theory with Applications to Communications and Control*. New York: McGraw-Hill, 1971.

(10) Sarris, A.H., "A Bayesian Approach to Estimation of Time Varying Regression Coefficients," *Annals of Economic and Social Measurement*, Vol. 2, No.4, October 1973.