NBER WORKING PAPER SERIES

IDENTIFICATION THEORY FOR TIME-VARYING MODELS

Thomas F. Cooley*
Kent D. Wall**

Working Paper No. 127

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

March 1976

Preliminary: Not for quotation

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

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

Abstract

The identification of time-varying coefficient regression models is investigated using an analysis of the classical information matrix. The variable coefficients are characterized by autoregressive stochastic processes, allowing the entire model to be case in state space form. Thus the unknown stochastic specification parameters and priors can be interpreted in terms of the coefficient matrices and initial state vector. Concentration of the likelihood function on these quantities allows the identification of each to be considered separately. Suitable restriction of the form of the state space model, coupled with the concept of controllability, lead to sufficient conditions for the identification of the coefficient transition parameters. Partial identification of the variance-covariance matrix for the random disturbances on the coefficients is established in a like manner. Introducing the additional concept of observability then provides for necessary and sufficient conditions for identification of the unknown priors. The results so obtained are completely analogous to those already established in the econometric literature, namely, that the coefficients of the reduced form are always identified subject to the absence of multicollinearity. Some consistency results are also presented which derive from the above approach.

1. INTRODUCTION

Identification is an issue which arises in connection with all parametric statistical models. Simply stated, the issue is whether one can infer from observed samples the existence of a unique underlying theoretical structure. Econometricians have long concerned themsleves with establishing the conditions for the identifiability of structures whose parameters are assumed to be constant over time. In this paper we address the seemingly more complex issue of the identifiability of structures when the regression coefficients themselves are varying stochastically over time. This is a relevant problem because in recent years increasing attention has been focused on the problem of estimating time varying structures.[1] Although estimation methods have been suggested by several authors, little attention has been paid to the problem of identification or to the asymptotic theory for these estimators. Many of the issues we address in this paper have been investigated by others (Tse & Anton [1972] and Mehra [1974] for example) but the context and the results, as we shall elaborate, are quite different.

The identification problem for the traditional linear econometric model with uncorrelated errors was first recognized by Koopmans and Reiersol [1950] and solutions were provided by Koopmans _et al_. [1950]. This theory was later extended and elaborated upon by Fisher [1966] in his comprehensive book on the subject. Two important papers by Hannan [1969, 1971] generalize the earlier theory to encompass models with moving average error processes. Most of this prior theory concentrates on conditions which guarantee unique solutions to the set of equations which characterize the structural form parameters in terms of the reduced form parameters as manifest by Hannan's solution. Rothenberg [1971] takes a different approach in characterizing the identifiability criteria in terms of the information matrix of classical mathematical statistics. Rothenberg's approach has been nicely extended to a more general representation by Bowden [1973]. It is this latter approach which is most appropriate to problems we are considering because of its relative independence from concepts related to stationary stochastic process theory.

The problem we are addressing can best be illustrated by considering the state space representation of a model with stochastically varying coefficients. We characterize the problem in terms of a regression relation (or observation equation) and a "state" equation which describes the evolution of the coefficients over time:

(1.1) $$y_t = X_t' \beta_t + e_t$$

(1.2) $$\beta_{t+1} = \Phi\beta_t + w_t$$

The variables y and X represent the observables of the system, Φ is a (K×K) matrix which governs the transitions of the K component coefficient β_t, and e_t and w_t are independently and identically distributed random variables with mean zero and covariance matrices σ^2 and Q respectively. It is clear that the identification is quite complex in this context because we must establish the conditions for the existence of a unique stochastic characterization of the process governing the coefficients. Identification of the coefficients β_t depends on the identification of the transition matrix Φ, the covariance matrix Q and the initial conditions of the coefficient process or β_o. The literature on varying parameter estimation has focused on the problem of estimating the initial conditions, but there has been no discussion of the conditions under which the other parameters will be identified

In the following section we formulate the general estimation problem for time varying coefficients and present the recursive (Kalman filtering) solution. The initial condition problem is discussed and the likelihood function, concentrated with respect to the initial conditions, is presented to facilitate the derivation of the identification conditions. In Section 3 the Information Matrix is derived and analyzed to give simple sufficient conditions for the identification of Φ and Q. It is shown that there are restrictions on the forms of Φ and Q which can be identified.

Section 4 briefly states the conditions for the identifiability of β_o using the results of Section 3. Asymptotic properties of the estimators are also discussed. The final section summarizes the results and draws some conclusions.

2. THE ESTIMATION THEORY FOR TIME VARYING STRUCTURES

In the introduction we represented the problem of time varying structures in terms of a single equation regression relationship and an equation which characterizes the evolution of the coefficients as a first order Markov process. As a point of departure for this section let us consider how we might generalize this representation. Ideally, we would like to be able to consider general simultaneous equation regression relationships. In practice, however, we must restrict ourselves to the consideration of reduced form relationships because the estimation theory for time-varying structural forms of simultaneous equation systems has not yet been developed.

In many instances one might expect to observe variation that is systematic but non-stochastic, or variation that is purely random. To include these possibilities we can modify our state equation to the form

(2.1) $$\beta_{t+1} = \Phi \beta_t + D\xi_t + w_t ,$$

which admits variation of all three types. If w_t is equal to zero then the variation is purely systematic. Thus, if the parameters follow a time trend, a sinusoidal pattern, or are correlated with exogenous variables it can be represented in this fashion. Similar models have been considered by Belsley [1973]. If ξ_t is a unit vector, w_t is nonzero while $\Phi = 0$, then the formulation is equivalent to the random coefficients model considered by Swamy [1970] and others where the parameters are regarded as random drawings from a multivariate distribution with mean vector $D\xi$ in the above representation. Although this is not properly a state space formulation it can still be handled within this framework. Thus, the evolution of the state of the system represented by equation (2.1) is a general one which encompasses many possibilities.[2] In this paper we concentrate on stochastic coefficient variation because it is this which presents the most difficult problems of identification.[3]

We wish to extend the basic state space model to permit the coefficients to be characterized by more general stochastic processes. Let each of the β_{kt} obey an autoregressive process of order n_k, $k=1,2,\ldots K$. Thus

$$(2.2)\qquad \beta_{kt} = \phi_{k1}\beta_{k,t-1} + \phi_{k2}\beta_{k,t-2} + \ldots + \phi_{kn_k}\beta_{k,t-n_k} + \eta_{k,t-1}$$

where $\phi_{kn_k} \neq 0$ and η_{kt} is a normally distributed zero mean sequentially independent random process with $E\{\eta_{kt}\eta_{\ell t}\} = q_{k\ell}$. The model can then be represented compactly as

$$(2.3)\qquad y_t = X_t\beta_t + e_t$$

$$(2.4)\qquad \beta_t = Hz_t$$

$$(2.5a)\qquad z_t = \Phi z_{t-1} + \Lambda\eta_{t-1}$$

where z_t is the state vector of the model describing the evolution of β_t:

$$z_t = [(z_t^1)' \mid (z_t^2)' \mid \cdots \mid (z_t^K)']'$$

With z_t^k representing the n_k state variable (the "substate" vector) for β_k.

H is a K×n ($n = \sum^K n_k$) matrix of the form

$$H = \begin{bmatrix} h_1' \\ \hdashline h_2' \\ \hline \vdots \\ \hdashline h_K' \end{bmatrix} \qquad \Lambda = H'$$

and h_k' is a row vector of zeros except for a one in the $1 + n_1 + \ldots + n_{k-1}$ column. The matrix Φ is now assumed to be of the form

$$\Phi = \begin{bmatrix} \Phi_{11} & \Phi_{12} & \cdots & \Phi_{1k} \\ \Phi_{21} & \cdot & & \vdots \\ \vdots & & \cdot & \vdots \\ \Phi_{k1} & \cdots & \cdot & \Phi_{kk} \end{bmatrix}$$

where

$$\Phi_{kk} = \left[\begin{array}{c|c} \multicolumn{2}{c}{\phi_k'} \\ \hline I_{n_k-1} & 0 \end{array}\right]$$

$$\phi_k' = \begin{bmatrix} \phi_{1k} & \phi_{2k} & \cdots & \phi_{n_k} \end{bmatrix}$$

The assumed form of Φ_{kk} is a natural one given the autoregressive representation of the process governing each coefficient. In the following discussion we assume that the off-diagonal blocks are null matrices ($\Phi_{ij} = 0$; $i \neq j$) since this is a restriction which must be imposed to derive a sufficient condition for identification. In Section 3.3 we present a counter example showing that a model without this restriction is underidentified.

To further simplify derivation of the identifiability conditions we replace the stochastic term of the state equation, $\Lambda\eta_t$ by an equivalent term Γu_t where u_t is K×1 vector of stochastic elements such that

$$E[u_t] = E[\eta_t] = 0 \ , \quad E[u_t u_t'] = I$$

and

$$E[\eta_t\eta_t'] \equiv Q = E[(H\Lambda\eta_t)(H\Lambda\eta_t)'] = E[(H\Gamma u_t)(H\Gamma u)'].$$

Γ is an $n\times K$ matrix of the same structure as Λ with the exception that the nonzero rows contain the corresponding rows of the unique lower triangular factorization of Q. Henceforth, (2.5a) will be replaced by

(2.5) $$z_t = \Phi z_{t-1} + \Gamma u_{t-1}.$$

Models like the one described by equations (2.3) - (2.5) have been extensively explored in the engineering literature following the work of Kalman [1960b] and Kalman and Bucy [1961]. The first recognition of the applicability of state space representations and Kalman filtering solutions to the problem of estimating econometric relationships with time varying structure was by Rosenberg [1968]. Other approaches to estimating models similar to the one described above have been suggested by Cooley and Prescott [1973, 1976] and Sarris [1973]. Here, however, we shall briefly review only the optimal recursive estimation method because it is the most convenient for establishing the identifiability criteria.

The estimation problem is to obtain estimates of the states, z_t, based on the observations $[y_1 \cdots\cdots y_T]$. If we let $\hat{z}_{t/t*}$ be an estimate of z_t based on observations $[y_1 \cdots\cdots y_{t*}]$ where $t^* \leq t$ and define the error covariance matrix of the estimated states as

(2.6) $$P_{t/t*} = E[z_t - \hat{z}_{t/t*})(z_t - \hat{z}_{t/t*})\,]$$

then the solution is easily obtained when z_o, Φ, σ^2 and Γ are known. The form of the solution is known as the Kalman filter and is represented as

(2.7) $\hat{z}_{t/t-1} = \hat{z}_{t-1/t-1}$

(2.8) $P_{t/t-1} = \Phi P_{t-1/t-1} \Phi' + \Gamma\Gamma'$

(2.9) $\mu_t = y_t - \tilde{X}_t \hat{z}_{t/t-1}$

(2.10) $M_t = \tilde{X}_t P_{t/t-1} \tilde{X}_t' + \sigma^2$

(2.11) $K_t = P_{t/t-1} \tilde{X}_t' M_t^{-1}$

(2.12) $\hat{z}_{t/t} = \hat{z}_{t/t-1} + K_t \mu_t$

(2.13) $P_{t/t} = (I-K_t\tilde{X}_t) P_{t/t-1}$

where $\tilde{X}_t = X_t H$.

Although the Kalman Filter has appeared many other places in the literature a brief interpretation may be useful. Equation (2.7) represents the one step ahead prediction of the states based on observations through period t when $t^* = t-1$. The quantity μ_t, which is called the "innovations" series, is obviously the one period prediction error for the y_t. The quantity K_t is called the gain of the Kalman Filter and M_t is the covariance matrix of the innovations. In this light it is easy to see that the gain of the filter is simply the optimal prediction correction factor.

It is obvious that z_o, P_o, Φ, σ^2 and Γ will not be known in most applications. The log likelihood of the system represented by (2.7) - (2.13), however, is (see Mehra [1972]);

(2.14) $\ell(z_o,P_o,\Theta) = -1/2 \sum_{t=1}^{T}[\log|M_t| + \mu_t' M_t^{-1}\mu_t]$, where $\Theta = (\sigma^2, \Gamma, \Phi)$. Thus, estimation proceeds by selecting initial values of z_o, P_o, Θ and using the equations of the Kalman Filter to define the likelihood function. This process proceeds iteratively and is known in the engineering literature as "tuning the filter". The engineering literature, however, has not in general been sensitive to problems of estimating the initial state vector z_o. Most of the literature assumes that z_o has a proper prior distribution which eliminates the problem. That this is seldom the case, however, is not a serious problem in dealing with real time systems with many observations (as in most engineering applications) because it is easily shown that under the appropriate conditions[4] the discrete Kalman Filter is asymptotically stable and the effects of the initial conditions are ultimately forgotten (see Jazwinski [1970, pp. 240-243]). In econometrics, however, the situation is somewhat different in that we do not deal with real time systems, our observation intervals are often relatively short, and we are often primarily interested in how the structure of the system evolves over time. For all of these reasons it is particularly important to be sensitive to the starting problems. The first correct solution to the starting problem was proposed by Rosenberg [1968] and later generalized by him [1973b]. The solution involves concentration of the likelihood function with respect to the initial state vector z_o. This permits maximum likelihood estimation of z_o conditional on σ^2, Φ and Γ. The recursive equations for z_o are

(2.15) $\Psi_{o/o} = I$

(2.16) $\Psi_{t/t-1} = \Phi \, \Psi_{t-1/t-1}$

(2.17) $\Psi_{t/t} = \Psi_{t/t-1} - K_{t+1}\tilde{X}_{t+1}\Psi_{t/t-1}$

(2.18) $H_t = (\tilde{X}_t\Psi_{t/t-1})' M_t^{-1} (\tilde{X}_t \Psi_{t/t-1})$

(2.19) $h_t = (\tilde{X}_t \Psi_{t/t-1})' M_t^{-1} \mu_t$

(2.20) $z_o = (\sum_{t=1}^{T} H_t)^{-1} \sum_{t=1}^{T} h_t$

where K_t, μ_t and M_t are as defined in equations (2.7) - (2.13). The matrix Ψ then is simply a function of the transition matrix which extrapolates the initial parameter vector into the future.

Equations (2.15) - (2.20) show that the likelihood function of the system can be concentrated with respect to the initial state vector and thus, the identifiability of z_o simply requires the invertability of $(\sum_{t=1}^{T} H_t)$ which in turn depends on the identification of Φ and Γ and the properties of the $\tilde{X}_t$. Consequently we can approach the problem at hand by first looking at the conditions for the identification of Φ and Γ.

It is worth noting that the initial condition approach outlined above does not provide estimates of the initial covariance matrix P_o. The consequences of this have recently been discussed in a paper by Garbade [1975a]. In econometric applications one should be most interested in obtaining "smoothed" estimates of the states $(z_{t/T})$, that is, estimates which use all of the information in the sample. A smoothing algorithm which avoids all of the initial condition problems has been derived in Cooley and Wall [1976].

3. IDENTIFICATION CONDITIONS FOR Φ AND Γ

The identifiability of the unknown stochastic specification parameters can be determined through an investigation of the classical Information matrix. This approach has two advantages: First, it permits the identification problem to be studied within the general framework of statistical information theory - a point well emphasized by Bowden [1973]. Also, it provides a useful connection between certain concepts in control systems theory and mathematical statistics.

3.1 The Information Matrix

The classical Information Matrix of R. A. Fisher is defined as (see Rothenberg [1971] or Bowden [1971]):

$$I(\Psi) = -E\left\{\left(\frac{\partial \ln p}{\partial \psi}\right)\left(\frac{\partial \ln p}{\partial \psi}\right)' \Bigg|_{\psi=\psi^o}\right\}$$

where, ψ is the N×1 vector of unknown parameters with true value ψ^o; and $\ln p$ is the natural logarithm of the density function for the jointly observed outputs over the interval $1 \leq t \leq T$. Thus, the first step is to derive the density function for the jointly observed outputs.

Combining the state and output equations (2.3) - (2.5) permits the formation of an expression for y_t ($t = 1,2,\ldots,T$) explicitly in terms of the vector of unknown parameters.

$$\begin{aligned} y_t &= X_t[H\Phi \mid H\Gamma]\begin{bmatrix} z_{t-1} \\ \hline u_{t-1} \end{bmatrix} + e_t \\ &= X_t W_t \psi + e_t \end{aligned} \qquad (3.1)$$

where,

$$\psi = \left[\phi_1' \,\vdots\, \phi_2' \,\vdots\, \cdots \,\vdots\, \phi_K' \,\vdots\, \gamma_{11}, \gamma_{21}, \gamma_{22}, \gamma_{31}, \ldots, \gamma_{kk}\right]'$$

$$W_t = \left[\begin{array}{c|c|c|c|ccccccc}
(z^1_{t-1})' & 0 & \cdots & 0 & u_{1,t-1} & 0 & & & 0 & \cdots & 0 \\
\hline
0 & (z^2_{t-1})' & \cdots & 0 & 0 & u_{1,t-1} & u_{2,t-1} & & 0 & \cdots & 0 \\
\hline
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots & & \vdots & \ddots & \vdots \\
\hline
0 & 0 & \cdots & (z^K_{t-1})' & 0 & 0 & 0 & & 0 & & u'_{t-1}
\end{array}\right]$$

W_t is thus a K×N matrix with its first n columns exhibiting a block diagonal structure, the $(k,k)^{th}$ block having dimension $1 \times n_k$ and containing the sub-state vector associated with the k^{th} regression coefficient. The last K(K+1)/2 columns form a matrix in the elements of u_{t-1}, with the last K columns of the last row consisting of u'_{t-1}. The joint observation can now be written compactly in terms of (3.1):

$$(3.2) \qquad Y_T = \underline{X}_T W \psi + E_T$$

where,

$$Y_T = [y_1, y_2, \ldots, y_T]'$$

$$\underline{X}_T = \left[\begin{array}{c|c|c|c}
X_1 & 0 & \cdots & 0 \\
\hline
0 & X_2 & \cdots & 0 \\
\hline
\vdots & \vdots & \ddots & \vdots \\
\hline
0 & 0 & \cdots & X_T
\end{array}\right]$$

$$W = \begin{bmatrix} W_1 \\ \hline W_2 \\ \hline \vdots \\ \hline W_T \end{bmatrix}$$

$$E_T = [e_1, e_2, \ldots, e_T]' .$$

The e_t are independently and identically distributed normal random variables so the probability density function for Y_T conditional on $\underline{X}_T$ and ψ is;

$$(3.3) \qquad p(Y_T;\ \underline{X}_T,\ \psi) = \frac{1}{(2\pi)^{T/2}(\sigma^2)^{T/2}} \ \exp\{-\frac{1}{2\sigma^2}[Y_T-\underline{X}_T W\psi]'[Y_T-\underline{X}_T W\psi]\}$$

Taking natural logarithms and then partially differentiating with respect to ψ gives

$$(3.4) \qquad \frac{\partial\ \ln p(Y_T;\underline{X}_T\ \psi)}{\partial\psi} = -\frac{\partial}{\partial\psi}\ \{\frac{1}{2\sigma^2}[Y_T-\underline{X}_T W\psi]'[Y_T-\underline{X}_T W\psi]\} .$$

$$= -\frac{1}{\sigma^2} W'\underline{X}_T' [Y_T-\underline{X}_T W\psi] .$$

Finally, the above expression may be substituted into the definition for $I(\psi)$ to yield

$$(3.5) \qquad I(\psi) = -E\{(\frac{1}{\sigma^2} W'\underline{X}_T' E_T')(\frac{1}{\sigma^2} E_T \underline{X}_T W)\} .$$

The replacement of $Y_T-\underline{X}_T W\psi$ by E_T follows from the evaluation of $I(\psi)$ at $\psi=\psi^o$. The Information matrix is seen to depend on the expectation of a product of random matrices.

In order to facilitate the evaluation of the expectation operation, we resort to consideration of the $(i,j)^{th}$ element of $I(\psi)$:

(3.6) $$\{I(\psi)\}_{ij} = \frac{1}{\sigma^4} E\left\{\left[\sum_{r=1}^{T} \sum_{\ell=1}^{TK} e_r \underline{x}_{r\ell} w_{\ell i}\right]\left[\sum_{s=1}^{T} \sum_{m=1}^{TK} e_s \underline{x}_{sm} w_{mj}\right]\right\}$$

$$= \frac{1}{\sigma^4} E\left\{\left[\sum_{r=1}^{T} e_r \sum_{\ell=1}^{TK} \underline{x}_{r\ell} w_{\ell i}\right]\left[\sum_{s=1}^{T} e_s \sum_{m=1}^{TK} \underline{x}_{sm} w_{mj}\right]\right\}$$

$$= \frac{1}{\sigma^4} \left[\sum_{t=1}^{T} E\{e_t^2\} E\left\{\left(\sum_{\ell=1}^{TK} \underline{x}_{t\ell} w_{\ell i}\right)\left(\sum_{m=1}^{TK} \underline{x}_{tm} w_{mj}\right)\right\}\right]$$

$$= \frac{1}{\sigma^2} \sum_{t=1}^{T} E\{(\underline{x}_t' w_i)(\underline{x}_t' w_j)\}$$

$$= \frac{1}{\sigma^2} \sum_{t=1}^{T} E\{\underline{x}_t' w_i)(w_j' \underline{x}_t)\}$$

$$= \frac{1}{\sigma^2} \sum_{t=1}^{T} \underline{x}_t' E\{w_i w_j'\} \underline{x}_t .$$

Here $\underline{x}_t'$ denotes the t^{th} row of $\underline{X}_T$ and w_i with i^{th} column of W. It is now possible to construct the Information Matrix, element-by-element once the expectation of the outer product $w_i w_j'$ is computed.

Appendix A contains the details of the element-by-element construction of $I(\psi)$, along with some additional steps required to put $I(\psi)$ into a more useful form for analysis. The end result is:

(3.7) $$\boxed{I(\psi) = -\frac{1}{\sigma^2} \sum_{t=1}^{T} \underline{\Xi}_t \, \underline{\Omega}_t \, \underline{\Xi}_t'}$$

$\tilde{\Xi}_t = P^{-1} \Xi_t$ is a nonsingular elementary transformation of the data matrices:

$$\Xi_t = \left[\begin{array}{cccc|cccc} Z_1 & & & & & & & \\ & Z_2 & & & & 0 & & \\ & & \ddots & & & & & \\ & & & Z_K & & & & \\ \hline & & & & V_1 & & & \\ & & & & & V_2 & & \\ & 0 & & & & & \ddots & \\ & & & & & & & V_K \end{array}\right]$$

$$Z_k = x_{kt} I_{n_k}$$

$$V_R = x_{kt} I_k$$

$$\Omega_t = \left[\begin{array}{c|c} S_{t-1} & 0 \\ \hline 0 & D_K \end{array}\right]$$

The $n \times n$ matrix S_{t-1} is the generalized variance-covariance matrix for the state-space process z_{t-1}, i.e. $S_{t-1} = E\{z_{t-1} z'_{t-1}\}$ (See Bryson & Ho [1969] pp. 320-325). The $K(K+1)/2 \times K(K+1)/2$ matrix D_K has a block diagonal structure

$$D_K = \begin{bmatrix} \Delta_K & & & \\ & \Delta_{K-1} & & \\ & & \ddots & \\ & & & \Delta_1 \end{bmatrix} ,$$

where each Δ_k ($k = K, K-1, \ldots, 1$) is a $k \times k$ matrix with unity in every location. In view of (3.7), it is clear that the identification of the unknowns in Φ and Γ depends upon the rank (or, equivalently, the positive definiteness) of both S_{t-1} and D_k.

3.2 Identification Conditions

Two points are immediately evident from (3.7). First, D_K is never of full rank since each Δ_k has only one linearly independent column. Thus all the unknown elements in Γ can never be identified simultaneously, but K linear combinations of these elements are identified. Second, the identifiability of the Φ parameters depends on whether or not S_{t-1} is positive definite. If conditions can be found which establish this, then the unknown elements of Φ will be identified.

The question of identification of Φ can readily be resolved with the aid of the concept of controllability:[5]

> Definition 1. The state-space model (2.5) is said to be uniformly completely controllable (UCC) with respect to the disturbances, u_{t-1}, if and only if there exists an integer $N_1 > 0$ and constants c_1, $c_2 > 0$ such that[6]
>
> $$0 < c_1 I \leq C(t,t-N_1) \leq c_2 I < \infty$$
>
> for all $t > N_1$, where the controllability matrix $C(t,t-N_1)$ is defined by

$$C(t,t-N_1) = \sum_{\tau=t-N_1}^{t-1} (\Phi^{t-1-\tau})\Gamma\Gamma'(\Phi^{t-1-\tau})'. \tag{3.8}$$

Definition 1 and the restricted structure of Φ are all that is needed to prove the main result of this paper:

Theorem 1. If the time-varying coefficients of (2.3) have their transition relationships realized by (2.5) with $\Phi_{ij}=0(i\neq j)$, and if (2.5) is UCC then: (i) the unknown stochastic specification parameters in Φ are globally identified; and (ii) only K linear combinations of the unknowns in Γ are identified.

Proof: The identifiability result for Γ has already been established from our observations concerning the D_K matrix, so we shall concentrate on the proof of (i). From the state equations for z_t, the generalized variance-covariance matrices are seen to obey the equation

$$S_t = \Phi\, S_{t-1}\, \Phi' + \Gamma\Gamma' ,$$

which has a unique solution given by

$$S_t = \Phi^{t-1} S_1 (\Phi^{t-1})' + \sum_{\tau=0}^{t-1} \Phi^{t-1-\tau}\Gamma\, \Gamma'(\Phi^{t-1-\tau})'. \tag{3.9}$$

The second term on the righthand side of (3.9) is nothing more than the controllability matrix $C(t,0)$ defined in (3.8) with $N_1=t$. From the UCC of (2.5) there will always exist a $t=t_1$ such that $C(t,0) > 0$ for all $t\geq t_1$. Thus for all $t\geq t_1$ (3.9) will be positive definite and the identification of Φ is established.

3.3 Remarks

The identifiability of Φ relies almost exclusively on the special structure underlying the state-space model, with the principle condition being the block diagonal form. This results in an $\underline{\Omega}_t$ with its upper left-hand block identically equal to S_{t-1}. The controllability condition is then imposed to guarantee that $S_{t-1} > 0$ for all $t>t_1+1$. Controllability alone is not a sufficient condition for identification of Φ - it must be accompanied by appropriate structure in Ω_t. Actually, any (Φ,Γ) pair which yields this structure in Ω_t, has Y_T linear in ψ, and is UCC will give exactly the same conclusions as Theorem 1.

The controllability requirement may appear impossible to verify a priori since it is stated in terms of the unknowns. In practice, however, this is no real limitation since the block diagonaltiy of Φ permits (2.5) to be viewed as a grouping of K independent subsystems (see Luenberger [1967]. Each "subsystem" will be UCC if and only if $\phi_{knk} \neq 0$ and at least one nonzero element appears in the corresponding row of Γ. If each subsystem is UCC, then the overall state-space model will be UCC. The first requirement is met if the specified order, n_k, is less than or equal to the "true" autoregressive order, while the second is met if there is any trace of randomness in each coefficient. It is difficult to concieve of a realistic situation where such conditions will be absent.

In the case where all K coefficients obey first order autoregressive processes each lagged β_{kt} becomes an element of the state vector (i.e.,H=I), and our results regarding the lack of complete identification of Γ agree with the results of Mehra [1971] concerning the identifiability of Q. His other results are not generally comparable to ours because he considers only models with stationary regression relationships, i.e. X_t = constant for all t.

The results in the control theory literature (see Tse and Weinert [1975]) suggest that more general forms for Φ can be identified (specifically, block triangular Φ). The following counter example, however, demonstrates that this specification for Φ will not be identified in the time-varying coefficients problem.

Consider the special case where K=2, $n_1=n_2 = 1$ (which more closely resembles the control theory case), and let some intercoupling between coefficients be permitted through $\phi_{21}\neq 0$.

$$\Phi = \begin{bmatrix} \phi_{11} & 0 \\ \phi_{21} & \phi_{22} \end{bmatrix}$$

This lower (block) triangular Φ is not identified, as can be seen by constructing the associated $I(\psi)$ following the steps contained in the Appendix:

$$I(\psi) = -\frac{1}{\sigma^2}\sum_{t=1}^{T} \tilde{\Xi}_t \, \underline{\Omega}_t \, \tilde{\Xi}_t'$$

$$\tilde{\Xi}_t = \left[\begin{array}{c|cc|c|cc} x_{1t} & & & & & \\ \hline & x_{2t} & & & & \\ & & x_{2t} & & & \\ \hline & & & x_1 & & \\ \hline & & & & x_{2t} & \\ & & & & & x_{2t} \end{array}\right]$$

$$\underline{\Omega}_t = \left[\begin{array}{ccc|ccc} s_{11} & s_{11} & s_{12} & & & \\ s_{11} & s_{11} & s_{12} & & 0 & \\ s_{12} & s_{12} & s_{22} & & & \\ \hline & & & 1 & 1 & 0 \\ & 0 & & 1 & 1 & 0 \\ & & & 0 & 0 & 1 \end{array}\right]$$

Clearly the upper-left 3×3 block of $\underline{\Omega}_t$ is singular so that all ϕ_{ij} elements are not identified. Whereas the control theory state model has one substate-vector associated with each element of y_t, the time-varying coefficient model has one substate-vector associated with each β_t element.[7]

4. CONTROLLABILITY, OBSERVABILITY, AND CONSISTENCY[8]

In Section 2 it was shown that the likelihood function can be concentrated with respect to β_0, the initial vector. This allowed us to consider the identification of Φ and Γ separately. We now turn to the establishment of the identification conditions for β_0. Conditions which establish the identification of any β_t can easily be derived with the aid of certain qualitative concepts from control theory as in the previous section. In addition to controllability, the concept of Uniform Complete Observability (UCO) is helpful. It is introduced by a second definition:[9]

> Definition 2. The model (2.3) - (2.5) is said to be uniformly completely observable (UCO) with respect to the output, y_t, if and only if there exists an integer $N_2>0$ and constants $c_3,c_4>0$ such that

$$0 < c_3 I \leq \mathcal{O}(t,t-N_2) \leq c_4 I < \infty \tag{4.1}$$

> for all $t \geq N_2$ where the observability matrix $\mathcal{O}(t,t-N_2)$ is defined by

$$\mathcal{O}(t,t-N_2) = \sum_{\tau=t-N_2}^{t} (\Phi^{\tau-t})' \tilde{X}_\tau' \tilde{X}_\tau \Phi^{\tau-t}. \tag{4.2}$$

Taken together, controllability and observability imply the identification of each point on the trajectory for β_t. The main result is given by the following theorem

Theorem 2. If the system defined by (2.3) - (2.5) is both UCO and UCC, then β_t is completely identified.

Proof: Each z_τ can be expressed in terms of z_t via solution of the underlying state equations, i.e.,

$$z_\tau = \Phi^{\tau-t} z_t + \sum_{s=t}^{\tau-1} \Phi^{\tau-1-s} \Gamma u_s \; .$$

Substitution of this expression into that for the observed outputs yields,

$$\begin{aligned} y_\tau &= \tilde{X}_\tau \Phi^{\tau-t} z_t + \tilde{X}_\tau \sum_{s=t}^{\tau-1} \Phi^{\tau-1-s} \Gamma u_s + e_\tau \\ &= \Lambda_\tau z_t + v_\tau . \end{aligned}$$

The jointly observed process, with z_t as an unknown parameter vector, can now be represented as in regression theory: (let $N = \max\{N_1, N_2\}$)

$$Y = \Lambda z_t + V$$

where

$$Y = [y_{t-N}, y_{t-N+1}, \ldots, y_t]'$$

$$\Lambda = [\Lambda'_{t-N} \mid \Lambda'_{t-N+1} \mid \cdots \mid \Lambda'_t]'$$

$$V = [v_{t-N}, v_{t-N+1}, \ldots, v_t]'.$$

The standard conditions for z_t to be unique are that both

$$\Lambda'\Lambda = \sum_{\tau=t-N}^{t} (\Phi^{\tau-t})' \tilde{X}'_\tau \tilde{X}_\tau \Phi^{\tau-t} \; ,$$

and the variance-covariance matrix

$$E\{VV'\} = \begin{bmatrix} \Omega_{11} & \Omega_{12} & \cdots & \\ \Omega_{21} & \Omega_{22} & \cdots & \\ \vdots & \vdots & \ddots & \\ & & & \Omega_{TT} \end{bmatrix}$$

$$\Omega_{ij} = R\delta_{ij} + \tilde{X}_i \sum_{\ell=\tau}^{m_{ij}} \Phi^{m_{ij}-\ell-1}\Gamma\Gamma'(\Phi^{m_{ij}-\ell-1})']\tilde{X}_j$$

$$m_{ij} = \min\{i,j\}$$

$$\delta_{ij} = \text{Kronecker delta}$$

be nonsingular. The first of these is just UCO, while the second follows immediately from UCC. Finally, since $\beta_t = Hz_t$ and H is full rank, β_t is unique whenever z_t is identified.

Theorem 2 can immediately be specialized to the problem of estimating unknown priors. In such a situation, the observation interval runs from the point of interest, t=0, forward to T. Thus, by setting N=-T we find that β_o is identified if and only if

$$(4.3) \qquad 0 < \sum_{\tau=T+1}^{1} (\Phi^{\tau-1})'\tilde{X}_\tau'\tilde{X}_\tau\Phi^{\tau-1} < \infty.$$

The above condition is equivalent to requiring that the matrix

$$[\tilde{X}_1' \mid \Phi'\tilde{X}_2 \mid \ldots \mid (\Phi^{T-1})'\tilde{X}_T']$$

be of (full) rank K. The full rank interpretation of (4.4) can be interpreted as a generalized multicollinearity condition.

Observability and controllability are also quite useful in examining the consistency of time-varying coefficient estimates. Both $C(t,t-N)$ and $O(t,t-N)$ can be used to establish bounds on the estimation error variance-covariance matrix and thus to study the behavior of the error as $T \to \infty$. The essence of these steps is contained in the following theorem.

Theorem 3. Let the time-varying coefficient model (2.3) - (2.5) be both UCO and UCC. If $t \geq N = \max\{N_1, N_2\}$ then the best linear unbiased estimate of β_t is never consistent.

Proof: First, consider the behavior of the filter estimation error variance-covariance matrix $P_{t/t}$. Together UCO and UCC guarantee the existence, uniqueness, and stability of $P_{t/t}$ as t_o, the initial time, tends toward $-\infty$.[10] Furthermore, for any prior $P_o \geq 0$, UCC guarantees that $P_{t/t} > 0$. (See Jazwinski [1970], Lemma 7.3, p. 238-239). Thus the filtered estimate's error variance-covariance matrix can never decay to zero no matter how much data, up through time t, has been employed.

Next consider the behavior of the smoother estimation error variance-covariance matrix, $P_{t/T}$, which uses all data in the sample. Fraser and Potter [1969] have shown that

$$P_{t/T} = [(P^f_{t/t})^{-1} + (P^b_{t/t+1})^{-1}]^{-1}$$

where $P^f_{t/t}$ is the filtered error variance-covariance of a "forward" time filter begun at t = 1 and running forward to t = T, while $P^b_{t/t+1}$ is the one-step prediction error variance-covariance of a "backward" filter begun at t = T and running backward until t = 1.[11] Fixing t and letting $T \to \infty$ reveals that only $P^b_{t/t+1}$ will change since only it is dependent on T. Now, from the positive definite property of $P^b_{t+1/t+1}$, no matter how far "back" it was started (i.e. how large T is), it is clear that $P^b_{t/t+1} > 0$ and hence $P_{t/T} > 0$. The smoothed estimate will always be inconsistent

Specializing the theorem to the estimation of unknown priors, β_0, it is clear that inconsistency persists. For diffuse priors,

$$P_{0/T} = [0 + (P^b_{0/1})^{-1}]^{-1} = P^b_{0/1} ,$$

which reveals the lack of consistency under stochastic excitation for the β's.

5. SUMMARY AND CONCLUSIONS

The growing literature on the estimation of models with time varying regression coefficients has largely ignored the issue of the identifiability of such models and consequently has left in doubt the generality with which they can be specified. This paper has used the classical information matrix of statistics to establish sufficient conditions for identifiability. The main result of the paper shows that the parameter transition matrix Φ will be completely identified if it is in block diagonal form. This special form of the transition matrix permits generality in the specification of the process governing each coefficient in the regression relation but rules out the estimation of intercouplings among coefficients. This is an important restriction in that many theoretical considerations which lead one to expect stochastic variation in coefficients also suggest that the movements in the coefficients will be related. The restriction, however, does not preclude *a priori* specification of known off diagonal transition parameters, and it may often be the case that theory will suggest *a priori* values for parameters. The identifiability conditions also show that only K linear combinations of the elements of the variance-covariance matrix of the coefficients can be identified. Diagonal Q matrices will therefore always be identified.

The results of Section 3 are equally applicable to the multivariate output model where y_t is an L×1 vector. The observation error variance σ^2 is replaced by an L×L matrix R, which, as before, is always identified. In addition, the unknown priors are identified subject to exactly the same conditions as given in Section 4 (these conditions being obtained independent of the dimension of y_t). The only complication is in the development of the expression for $I(\psi)$. R^{-1} replaces the scalar $1/\sigma^2$ in the earlier stages of the algebraic manipulations and cannot be so easily "factored out" of the ensuing derivation. With the aid of the Kronecker product, however, an expression similar to (3.9) can be obtained which yields exactly the same conclusions as before:

$$(5.1) \qquad \boxed{I(\psi) = \sum_{t=1}^{T} \underline{\tilde{\tilde{\Xi}}}_t' \, [\underline{\Omega}_t \otimes R^{-1}] \, \underline{\tilde{\tilde{\Xi}}}_t \, .}$$

Since the Kronecker product of two positive definite matrices is itself positive definite, the conditions for identification once again derive from an analysis of $\underline{\Omega}_t$ defined in (3.9). The $\tilde{\Xi}_t$ matrices above have exactly the same form given in the appendix with the exception that the scalars x_{kt} are replaced by L×1 column vectors.

The unknown priors are always identified subject to the generalized multicollinearity condition introduced in Section 4. If Φ is known *a priori* then the identification of the K×1 prior β_o, and hence any point on the β_t trajectory, can be established by examining the rank of the associated Observability Matrix of (4.3). Note that with Φ known, this check can always be carried out before estimation is attempted. The consistency of the prior estimate cannot, however, be established. The analysis of the dynamic properties of the estimation error variance-covariance matrix reveals that random excitation of the coefficients always prevents the limiting distribution for β_t from attaining a zero dispersion. Given complete observability (identification) the most that can be achieved is an asymptotically finite error distribution for the estimates of the randomly excited coefficients.

APPENDIX

EXPECTATION EVALUATION AND THE FORMATION OF THE INFORMATION MATRIX

The Information Matrix construction presented in Section 3.1 reduces to the evaluation of $E\{w_i(w_j)'\}$ where w_i and w_j denote the i^{th} and j^{th} columns, respectively, of W_t (see equation (3.3)).[12] Since $1 \leq i,j \leq N$, where N is the total number of unknown stochastic specification parameters for the β_t process, this amounts to the evaluation of N^2 matrix expectations formed from various vector outer products. These evaluations are straightforward if care is taken to avoid the potential for confusion. This can be achieved by decomposing the evaluations into three parts, depending on the relative value of the subscripts i and j. To this end, let C_k denote the set of integers containing the column numbers of the states in W_t associated with the k^{th} coefficient. In other words, C_k contains the column numbers in which the n_k states $z^k_{j,t-1}$ are located. For example,

$$
\begin{aligned}
C_1 &= 1, 2, \ldots, n_1 \\
C_2 &= n_1+1, \ldots, n_1+n_2 \\
&\vdots \\
C_K &= n_{K-1}+1, \ldots, n_1+n_2+\ldots+n_K.
\end{aligned}
$$

First, consider the case in which $i,j \leq n$, i.e., in which w_i and w_j are both taken from the first n columns of W_t. Thus let $i \in C_\ell$ and $j \in C_m$, then

$$E\{w_i(w_j)'\} = E\left\{\begin{bmatrix} 0 \\ \vdots \\ 0 \\ z_{i,t-1} \\ 0 \\ \vdots \\ 0 \end{bmatrix} [0 \ldots 0 \; z_{j,t-1} \; 0 \ldots 0]\right\}$$

= K×K matrix of zeros with the $(\ell,m)^{th}$ position replaced by $E\{z_{i,t-1}\, z_{j,t-1}\}$.[13]

The expectation $E\{z_{i,t-1}\, z_{j,t-1}\}$ is nothing more than the $(i,j)^{th}$ element of the generalized variance - covariance function $S_{t-1} = E\{z_{t-1}\, z'_{t-1}\}$ associated with the complete state representation for the coefficient transitions defined by (2.5). If this element is denoted by s_{ij}^{t-1} then[14]

$$E\{w_i(w_j)'\} = \text{K×K matrix of zeroes with the } (\ell,m)^{th} \text{ element replaced by } s_{ij}^{t-1}. \tag{A.1}$$

Second, consider the case in which $i \in C_k$ while j is associated with any of the last K(K+1)/2 columns of W_t. Then it is easy to see that

$$E\{w_i(w_j)'\} = \text{K×K null matrix}, \tag{A.2}$$

since z_{t-1} and u_{t-1} are independent. The same result holds with the roles of i and j exchanged.

Finally, consider the case in which both i and j are taken from the last K(K+1)/2 columns in W_t. In general the matrix expectations become,

$$E\{w_i(w_j)'\} = E\left\{\begin{bmatrix}0\\ \vdots\\ u_{p,t-1}\\ \vdots\\ 0\end{bmatrix}[0\ldots u_{q,t-1}\ldots 0]\right\}.$$

Obviously if q ≠ q the end result is a K×K null matrix. If p = q then the expectation operation results in a K×K matrix of zeros with one element replaced by unity. The exact location of unity depends, of course, on the respective row positions of $u_{p,t-1}$ and $u_{q,t-1}$. A concise characterization analogous to (A.1) does not seem possible. This difficulty, however, is of little consequence to the final expression for $I(\psi)$ as will become apparent below when resort is made to the use of elementary row and column transformations.

The above results concerning the evaluation of the expectations can now be combined with the definition of the $(i,j)^{th}$ element of $I(\psi)$ to permit an element-by-element construction of $I(\psi)$. More specifically;

1. If i,j belong to the first n columns of W_t, then

$$I_{ij}(\psi) = -\frac{1}{\sigma^2}\sum_{t=1}^{T} x_{it}\, s_{ij}^{t-1} x_{jt}. \qquad (A.3)$$

2. If i or j belongs to the first n columns of W_t while the other is associated with the last K(K+L)/2 columns of W_t, then

$$I_{ij}(\psi) = 0. \qquad (A.4)$$

3. If i and j are both associated with the last K(K+1)/2 columns of W_t, then

$$I_{ij}(\psi) = \frac{1}{\sigma^2} \sum_{t=1}^{T} x_{it}\, E\{u_{p,t-1}\, u_{q,t-1}\}\, x_{jt} \qquad \text{(A.5)}$$

$$= \frac{1}{\sigma^2} \sum_{t=1}^{T} x_{it}\, x_{jt} \quad ; \quad p=q$$

$$= 0 \quad ; \quad p=q.$$

The element-by-element construction can finally be combined to give the complete Information Matrix:

$$I(\psi) = -\frac{1}{\sigma^2} \sum_{t=1}^{T} \Xi_t' \, \Omega_t \, \Xi_t \qquad \text{(A.6)}$$

where,

$$\Xi_t = \left[\begin{array}{cccc|cccc} Z_1 & & & & & & & \\ & Z_2 & & & & 0 & & \\ & & \ddots & & & & & \\ & & & Z_K & & & & \\ \hline & & & & V_1 & & & \\ & 0 & & & & V_2 & & \\ & & & & & & \ddots & \\ & & & & & & & V_K \end{array}\right]$$

$$Z_k = \begin{bmatrix} x_{kt} & & & 0 \\ & x_{kt} & & \\ & & \ddots & \\ 0 & & & x_{kt} \end{bmatrix} : n_k \times n_k$$

$$V_k = \begin{bmatrix} x_{kt} & & & 0 \\ & x_{kt} & & \\ & & \ddots & \\ 0 & & & x_{kt} \end{bmatrix} : k \times k$$

$$
\Omega_t = \left[\begin{array}{c|c} S_{t-1} & \\ \hline & T_k \end{array}\right] .
$$

Although Z_k and V_k contain the same elements on the diagonal, they have different dimensions. Thus Z_1 is an $n_1 \times n_1$ matrix whereas V_1 is merely a scalar (1×1 matrix). The $K(K+1)/2 \times K(K+1)/2$ matrix T_k is difficult to write down in general; it contains only zeroes and one's following the pattern set out below.

$$
T_k = \left[\begin{array}{cccccccccccc}
1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & \ldots\ldots \\
1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & \ldots\ldots \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & \ldots\ldots \\
1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & \ldots\ldots \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & \ldots\ldots \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & \ldots\ldots \\
1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & \ldots\ldots \\
0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & \ldots\ldots \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & \ldots\ldots \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & \ldots\ldots \\
1 & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & . \\
\vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \ddots
\end{array}\right.
$$

The Information matrix of (A.6) can be written in a more convenient form for analysis of identification by resorting to elementary row and column transformations. In particular, there will always exist nonsingular matrices P and Q, composed of elementary exchanges of rows and columns respectively, such that

$$P\Omega_t Q = \left[\begin{array}{c|c} S_{t-1} & 0 \\ \hline 0 & D_K \end{array}\right] = \underline{\Omega}_t$$

where

$$D_K = \begin{bmatrix} \Delta_K & & & \\ & \Delta_{K-1} & & \\ & & \ddots & \\ & & & \Delta_1 \end{bmatrix}$$

with Δ_k a k×k matrix with 1's everywhere. In addition, the structure of T_k reveals that for every row exchange required to bring T_k to D_k, there is a corresponding column exchange. Thus Q=P', and the final expression for the information matrix becomes

$$\boxed{I(\psi) = -\frac{1}{\sigma^2} \sum_{t=1}^{T} \Xi_t' \, P^{-1} \, \underline{\Omega}_t \, (P^{-1})' \, \Xi_t \, .} \qquad \text{(A.7)}$$

Since P is nonsingular the rank of $I(\psi)$, and hence its definiteness, depend on the rank and definiteness of $\underline{\Omega}_t$.

BIBLIOGRAPHY

[1] Aoki, M. [1967], Optimization of Stochastic Systems, New York: Academic Press.

[2] Belsley, David [1973], "On the Determination of Systematic Parameter Variation in the Linear Regression Model", Annals of Economic and Social Measurement, Vol. 3, October, 1973.

[3] Bowden, R. [1973], "The Theory of Parametric Identification", Econometrica, Vol. 41, No. 6, pp. 1069-1074.

[4] Bryson, A. E., and Y. C. Ho [1969], Applied Optimal Control, Waltham, Mass.: Ginn-Blaisdell.

[5] Cooley, T. F. and E. Prescott [1973], "Varying Parameter Regression", "A Theory and Some Applications", Annals of Economic and Social Measurement, Vol. 3, October.

[6] ———— and ———— [1976], "Estimation in the Presence of Stochastic Parameter Variation" Econometrica, Vol. 44, No. 1, January.

[7] ———— and K. D. Wall [1975], "On the Identification of Time Varying Structures", NBER, Working Paper No. 85, Mary.

[8] ———— and ———— [1976], "A Note On Optimal Smoothing In Varying Coefficient Regression", NBER, March.

[9] Deyst, J. J. [1973], Correction to "Conditions for Asymptotic Stability of the Discrete Minimum-Variance Linear Estimator", IEEE Trans. Auto. Control, Vol. AC-18, no. 5, October, pp. 562-563.

[10] ———— and C. R. Price [1968], "Conditions for Asymptotic Stability of the Discrete Minimum-Variance Linear Estimator", IEEE Trans. Auto. Control, Vol. AC-13, no. 6, December, pp. 702-705.

[11] Fisher, F. M. [1966], The Identification Problem in Economics, New York: McGraw-Hill.

[12] Fraser, D. C. and J. E. Potter [1969], "The Optimum Linear Smoother as a Combination of Two Optimum Linear Filters", IEEE Trans. Auto Control.

[13] Garbade, K. [1975a], "The Initialization Problem In Variable Parameter Regression", Working Paper, New York University, August.

[14] ———— [1975b], "Two Methods for Examining the Stability of Regression Coefficients", Research Memorandum No. 186, Econometrics Research Program, Princeton Univeristy, October.

[15] Hannan, E. J. [1969], "The Identification of Vector Mixed Autoregressive-Moving Average Systems", Biometrika, 57, pp. 223-225.

[16] Hannan, E. J. [1971], "The Identification Problem for Multiple Equation Systems with Moving Average Errors", Econometrica, 39, September, pp. 751-765.

[17] Jazwinski, A. H. [1970], Stochastic Processes and Filtering Theory, New York: Academic Press.

[18] Kalman, R. E. [1969a], "On the General Theory of Control Systems", Proceedings First Int'l Congress IFAC, Moscow, U.S.S.R.

[19] __________ [1960b], "A New Approach to Linear Filtering and Prediction Problems", Transactions of ASME, Series D, Journal of Basic Engineering, 82, pp. 35-45.

[20] __________ and R. S. Bucy [1961], "New Results in Linear Filtering and Prediction Theory", Transactions of ASME, Series D, Journal of Basic Engineering, 83, pp. 95-108.

[21] __________ [1963], "New Methods in Wiener Filtering Theory", Proc. Symp. Eng. Appl. Random Function Theory and Probability (Bogdanoff and Kozin, eds.) New York: Wiley.

[22] Koopmans, T. C. and O. Reiersol [1950], "The Identification of STructural Characteristics", Annals of Mathematical Statistics, Vol. 21, pp. 165-181.

[23] __________, H. Rubin and R. B. Leipnik [1950], "Measuring the Equation Systems of Dynamic Economics", Statistical Inference in Dynamic Economic Models, Cowles Commission Monograph, No. 10

[24] Luenberger, D. G. [1967], "Canomical Forms for Linear Multivariable Systems", IEEE Trans. Auto. Control, vol. AC-12, no. 3, June, pp. 290-293.

[25] Mehra, R. K. [1970], "On the Identification of Variances and Adaptive Kalman Filtering", IEE Transactions on Automatic Control, April, pp. 175-184.

[26] __________ [1971], "On Line Identification of Linear Dynamic Systems with Applications to Kalman Filtering", IEEE Transactions on Automatic Control, AC-16, No. 1, February.

[27] __________ [1972], "Approaches to Adaptive Filtering", IEEE Transactions on Automatic Control, October, pp. 693-698.

[28] __________ [1974], "Identification in Control and Econometrics: Similarities and Differences", Annals of Economic and Social Measurement, Vol. 3, pp. 21-47.

[29] Pagan, A. R. [1974], "A Note on the Extraction of Components from Time Series", Econometrica, 43, pp. 163-168.

[30] Rosenberg, B. M. [1968], "Varying-Parameter Estimation", Unpublished Ph.D. dissertation, Department of Economics, Harvard University.

[31] __________ [1973a], "A Survey of Stochastic Parameter Regression", Annals of Economic and Social Measurement, Vol. 3, October.

[32] __________ [1973b], "The Analysis of a Cross Section of Time Series by Stochastically Convergent Parameter Regression", Annals of Economic and Social Measurement, Vol. 3, October.

[33] Rosenbrock, H. H. [1970], State-Space and Multivariable Theory, New York: John Wiley & Sons.

[34] Rothenberg, T. J. [1971], "Identification in Parametric Models", Econometrica, 39, pp. 577.

[35] Sarris, A. [1973], "A Bayesian Approach to Estimation of Time-Varying Regression Coefficients", Annals of Economic and Social Measurement, October, pp. 501-523.

[36] Swamy, P. A. V. B. [1970], "Efficient Inference in a Random Coefficient Regression Model", Econometrica, 38, pp. 311-323.

[37] Tse, E. and J. Anton [1972], "On the Identifiability of Parameters", IEEE Trans. Auto. Control, Vol. AC-17, No. 5, October.

[38] ______, and H. Weinert [1975], "Structure Determination and Parameter Identification for Multivariable Stochastic Linear Systems", IEEE Trans. Auto. Control, vol. AC-20, no. 6, December, pp

FOOTNOTES

1. See, for example, Garbade [1975a, 1975b], Cooley [1974], Pagan [1974], Rosenberg [1973a, 1973b], Cooley and Prescott [1976].

2. For an excellent survey of generic relations among models with non-constant coefficients see Rosenberg [1973a].

3. An earlier version of this paper (Cooley and Wall [1975]) analyzed the differences between the sequential nonstochastic variation problem and the stochastic variation problem.

4. The conditions are that the system is uniformly completely observable and uniformly completely controllable.

5. The derivation of controllability contained in the folliwing definition is beyond the scope of this paper. The reader may consult any number of introductory texts such as Zadeh and Desoer [1963; pp. 505-509] for an excellent development. It should be noted that there are many definitions of controllability, each with its own subtle twist (see Rosenbrock [1970; Chpt. 5 & 6]).

6. Let x be any arbitrary p×1 vector and A any p×p matrix. Then $\alpha I \leq A \leq \beta I$ is taken to mean $\alpha x'x \leq x'Ax \leq \beta x'x$ where I is the p×p identify matrix.

7. There is a basic difference between the identification problem posed in this paper and that considered the control engineering literature. There is more information available in the latter case because each diagonal block in the transition matrix (i.e., each state subgroup, $z_t^{(i)}$) is associated with a different observed output (y_{it}).

8. The authors wish to acknowledge an anonymous referee whose helpful comments have greatly facilitated the presentation of Theorem 2.

9. Observability was another qualitative system property first defined by Kalman [1960a] in a purely deterministic framework. Stochastic versions of this concept were also introduced by Kalman [1963], Aoki [1967], and Jazwinski [1970]. The sole difference between the stochastic and deterministic versions is the insertion of R^{-1}, the variance-covariance of e_t, between $\tilde{X}_t'$ and $\tilde{X}_t$. This distinction is, however, immaterial so long as R is assumed positive definite.

10. This fundamental result was first obtained by Kalman [1963] for the continuous time case. The discrete-time case was first considered by Deyst and Price [1968] and Deyst [1973], with subsequent pedagogical presentation given by Jazwinski [1970]. The interested reader is referred to any of these for a proof.

11. Alternatively, the expression could have been written using $p^f_{t-1/t}$ and $p^b_{t/t}$ with equal validity. In a continuous-time frame work this idiosyncracy disappears, i.e. both filter variance-covariances are employed.

12. The time dependence of each column of W_t is suppressed in order to avoid the use of double subscripts which are reserved for elements of matrices.

13. From the definition of the overall state vector, z_t, given in Section 2, it is clear that $z^{\ell}_{p,t-1} = z_{i,t-1}$ if p is such that $z^{\ell}_{p,t-1}$ appears in the ith column of W_t. Likewise, $z^m_{q,t-1} = z_{j,t-1}$ if q is such that $z^m_{q,t-1}$ appears in the jth column of W^q_t.

14. The superscript t is used here to avoid triple subscripts while still retaining explicit indication of the time dependence of s_{ij}.