

NBER WORKING PAPER SERIES

ON THE IDENTIFICATION OF  
TIME VARYING STRUCTURES

Thomas F. Cooley\*  
Kent D. Wall\*\*

Working Paper No. 85

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

May 1975

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 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 identifiability of reduced form econometric models with variable coefficients is investigated using the control theoretic concepts of uniform complete observability and uniform complete controllability. First, a variant of the state space representation of the traditional reduced form is introduced which transcribes the underlying non-stationary estimation problem into one particularly suited to a Kalman filtering solution. Using such a formulation, observability and controllability can be called upon to obtain a necessary and sufficient condition for identification of the specific parameterization. The results so obtained are completely analogous to those already established in the econometric literature, namely, that the parameters of the reduced form are always identified subject to the absence of multicollinearity (referred to as "persistent excitation" in the control literature). However, now the multicollinearity condition is seen to depend on the structure of the parameter variations as well as the statistical nature of the explanatory variables. The verification of identifiability thus reduces to a check for uniform complete observability which can always be affected in econometric applications. Some consistency results are also presented which derive from the above approach.

Contents

1. Introduction . . . . .	1
2. The Estimation Theory for Time Varying Structures . . . . .	4
3. Observability and Controllability . . . . .	11
4. Econometric Identification and Consistency . . . . .	21
5. Summary and Conclusions . . . . .	31
Bibliography . . . . .	37

## 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 themselves with establishing the conditions for the identifiability of structures whose parameters are assumed to be constant. In this paper we address the seemingly more complex issue of the identifiability of structures when the parameters themselves are varying systematically or 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. Although estimation methods have been suggested by several authors, little attention has been paid to the space 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, as we shall elaborate, is somewhat 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.<sup>1</sup> 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 structure. We characterize the problem in terms of a regression relation (or observation equation) and a "state" equation which describes the evolution of the parameters over time:

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

$$(1.2) \quad \beta_{t+1} = \phi \beta_t + u_t$$

The variables  $y$  and  $X$  represent the observables of the system,  $\phi$  is a  $(k \times k)$  matrix which governs the transitions of the  $k$  component parameter

$\beta_t$ , and  $e_t$  and  $u_t$  are independently and identically distributed random variables with mean zero and covariance matrixes  $\sigma^2$  and  $Q$  respectively. Intuitively it is seen that identification is more complex in this context because the underlying theoretical structure is itself a stochastic process. Since the process is also dependent upon initial conditions we must establish the conditions for the existence of a unique set of  $k \times (T+1)$  parameter vectors. In fact, however, the problem is not as difficult as it might appear because we have imposed additional structure on the problem through equation (1.2). Nevertheless, the information matrix conditions of Rothenberg and Bowden must be amended to establish identification in such a context.

The plan of the paper is as follows. In Section 2 we formulate the general estimation problem for time varying structures and present the recursive (Kalman Filtering) solution. Our purpose here is partly pedagogical because, while the Kalman Filtering solution to the time varying estimation problem has appeared elsewhere in the literature,<sup>2</sup> it provides a convenient vehicle for illustration of the identifiability conditions. The third section of the paper introduces concepts of uniform complete observability and uniform complete controllability and relates these to the performance of the Kalman Filter.

In Section 4 the observability matrix and controllability matrix are related to the information matrix of the constant parameter case considered by Rothenberg. In addition, these concepts are used to derive some interesting asymptotic results for the time varying problem. The final section discusses the results and suggests directions for further research.

## 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 parameters 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 the structural forms of simultaneous equation systems has not yet been developed.<sup>3</sup> Consequently, the most general regression structures we can deal with are of the form

$$(2.1) \quad y_t = A_1(t)y_{t-1} + \dots + A_p(t)y_{t-p} + B_0(t)W_t + \dots + B_q(t)W_{t-q} + e_t$$

where  $y_t$  is an  $l \times 1$  column vector of outputs (endogenous variables),  $W_{t-j}$  are vectors of exogenous variables, and  $e_t$  is an  $l \times 1$  vector of observation errors. This system of equations can be represented more compactly as

$$(2.2) \quad y_t = X_t \beta_t + e_t.$$

We also want to consider generalizations of the process which governs the evolution of the parameters. Our original characterization of the

parameter evolution as a first order Markov process, or, more generally, as an autoregressive or moving average process of low order has considerable appeal. Not only is it a convenient characterization but it is a natural one in a time series context in that such processes can capture well the evolution of the parameters.<sup>4</sup>

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

$$(2.3) \quad \beta_t = \phi \beta_{t-1} + \Gamma z_t + u_t,$$

which admits variation of all three types. If  $u_t$  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  $z_t$  is a unit vector and  $u_t$  is non-zero 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  $\Gamma$  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 (2.3) is a general one which encompasses many possibilities.<sup>5</sup> In any given context



prior restrictions will be placed on  $\phi$  and  $\Gamma$  (and also the covariance of the  $u_t$ ) by one's view of the particular problem.

Models like the one described by equation (2.2) and (2.3) have been extensively explored in the engineering literature following the work of Kalman [1960] 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] 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.

We begin by assuming that  $e_t$  and  $u_t$  from equations (2.2) and (2.3) are uncorrelated Gaussian sequences with

$$\begin{aligned} E[e_i] &= 0 & E[u_i] &= 0 \\ E[e_i e_j'] &= R\delta_{ij} & E[u_i u_j'] &= Q\delta_{ij} \end{aligned}$$

where  $Q$  and  $R$  are at least positive semi definite matrices and  $\delta$  is the Kronecker delta. The estimation problem is to obtain estimates of the  $\beta_t$  based on the observations  $[y_1 \dots y_T]$ . If we let  $b_{t/t^*}$  be an estimate of  $\beta_t$  based on observations  $[y_1 \dots y_{t^*}]$  where  $t^* \leq t$  and define the error covariance matrix of the estimated coefficients as

$$(2.4) \quad P_{t/t^*} = E[(\beta_t - b_{t/t^*})(\beta_t - b_{t/t^*})']$$

then the solution is easily obtained when  $\beta_0$ ,  $\phi$ ,  $Q$ ,  $R$  and  $\Gamma$  are known. The form of the solution is known as the Kalman filter and is represented

as

$$(2.5) \quad b_{t+1/t} = \phi b_{t/t} + \Gamma z_{t+1}$$

$$(2.6) \quad b_{t+1/t+1} = b_{t/t} + K_t \mu_t$$

$$(2.7) \quad \mu_t = Y_t - Y_t b_{t/t-1} - X_t \Gamma z_t$$

$$(2.8) \quad S_t = X_t P_{t/t-1} X_t' + R$$

$$(2.9) \quad K_t = P_{t/t-1} X_t' S_t^{-1}$$

$$(2.10) \quad P_{t+1/t} = \phi P_{t/t} \phi' + Q$$

$$(2.11) \quad P_{t/t} = (I - K_t X_t) P_{t/t-1}$$

Although the Kalman Filter has appeared many other places in the literature a brief interpretation may be useful. Equation (2.5) represents the one step ahead prediction of the parameters based on observations through period  $t$  when  $t^* = t-1$ . The quantity  $\mu_t$ , 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  $S_t$  is the covariance matrix of the innovations. In this light it is easy to see that the gain of filter is simply the optimal prediction correction factor.

It is obvious that  $\beta_0$ ,  $P_{0/0}$ ,  $Q$  and  $R$  will not be known in most applications. The log likelihood of the system represented by (2.5) - (2.11), however, is (see Mehra [1972]);

$$(2.12) \quad L(\beta_0, P_{0/0}, \theta) = -\frac{1}{2} \sum_{t=1}^T [\log |S_t| + \mu_t' S_t^{-1} \mu_t],$$

where  $S_t$  is the covariance matrix of the innovations and  $\theta = (R, Q, \Gamma, \Phi)$ . Thus, estimation proceeds by selecting initial values of  $\beta_0$ ,  $P_{0/0}$ ,  $\theta$  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  $\beta_0$ . Most of the literature assumes that  $\beta_0$  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 is the case in most engineering applications) because it is easily shown that under the appropriate conditions<sup>7</sup> 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 problems was proposed by Rosenberg [1968] and later generalized by him [1973b]. The solution involves concentration of the likelihood function with respect to the initial parameter vector  $\beta_0$ . This permits maximum likelihood estimation of  $\beta_0$  conditional on  $R, Q, \Phi$  and  $\Gamma$ . The recursive equations for  $\beta_0$  are<sup>8</sup>

$$(2.13) \quad \Omega_{0/0} = I$$

$$(2.14) \quad \Omega_{t+1/t} = \Phi \Omega_{t/t}$$

$$(2.15) \quad \Omega_{t+1/t+1} = \Omega_{t/t} - K_{t+1} X_{t+1} \Omega_{t+1/t}$$

$$(2.16) \quad H_t = (X_t \Omega_{t/t-1})' S_t^{-1} (X_t \Omega_{t/t-1})$$

$$(2.17) \quad h_t = (X_t \Omega_{t/t-1})' S_t^{-1} \mu_t$$

$$(2.18) \quad \hat{\beta}_0 = \left( \sum_{t=1}^T H_t \right)^{-1} \sum_{t=1}^T h_t$$

where  $K_t$ ,  $\mu_t$  and  $S_t$  are as defined in equations (2.5 - 2.11). The matrix  $\Omega$  then is simply a function of the transition matrix which extrapolates the initial parameter vector into the future.

Given an estimate of the initial parameter vector, say  $\hat{\beta}_0$ , estimation of any realization  $b_{t/t}$  is straightforward given equations (2.5 - 2.11). It is worth noting that, in econometric applications, we will be most interested in obtaining smoothed estimates of given realizations of the parameter trajectory ( $B_{t/T}$ ), that is, estimates which use all of the information in the sample. Smoothing equations are presented in Meditch [1969].

This review of estimation methods for time varying structures highlights one of the important features that is useful in the discussion of the identifiability of such structures. It is that estimation that is a two-step procedure. In the first step the initial parameter vector  $\hat{\beta}_0$  and the unknown covariance elements are estimated. The second step consists of estimating realizations of the parameter process ( $b_{t/t}$  or  $b_{t/T}$ ) conditional on the estimates obtained in the first step. Thus,

estimation of the parameter trajectories is essentially an empirical Bayesian procedure. Consequently, the identifiability criteria for such structures may be viewed as having two parts. The first, which obviously is the crucial part, establishes conditions for the existence of a unique underlying initial parameter vector. The second part simply involves the conditions for the existence of a trajectory conditional on  $b_0$ .

As we shall see subsequently there is a direct, but by no means simple, relationship between the classical identifiability criteria for constant parameter models and the identifiability criteria for models with time varying structure. Before deriving this relationship, however, we introduce some concepts from the control theory literature which will be useful in the subsequent analysis.

### 3. Observability and Controllability

#### 3.1 Definitions

The concepts which we shall find useful in studying the identifiability of time varying structures have evolved in the control theory literature from an essentially different, but parallel set of concerns. The rapid development of optimal control theory in the early 1960's led to a concern for the qualitative aspects of the optimal control problem. Naturally, of first concern were questions regarding the existence and uniqueness of optimal controls. This led to the consideration of whether or not it was possible to arbitrarily alter the state of a model solely by manipulation of the instruments (inputs). If a model possessed this ability it was said to be completely controllable, i.e., there existed a coupling between the inputs and all of the states. In relation to the usual state space form representation<sup>9</sup>

$$(3.1) \quad \begin{aligned} \xi_{t+1} &= F_t \xi_t + G_t x_t, \\ y_t &= H_t \xi_t + D_t x_t, \end{aligned}$$

the concept of controllability can be captured in the following definition:<sup>10</sup>

Definition 1. The model (3.1) is said to be uniformly completely controllable with, respect to the input  $x_t$ , if and only if there exists an integer  $N > 0$  and constants  $c_1, c_2 > 0$  such that<sup>11</sup>

$$(3.2) \quad 0 < c_1 I \leq C(t, t-N) \leq c_2 I$$

for all  $t \geq N$ , where the controllability matrix  $C(t_1, t_0)$  is defined by

$$(3.3) \quad C(t_1, t_0) = \sum_{\tau=t_0}^{t_1-1} \Phi(t_1, \tau) G_{\tau} G_{\tau}' \Phi'(t_1, \tau).$$

Uniform complete controllability (UCC) implies two things. First, when the lower positive bound,  $c_1 I$ , obtains, every mode of the dynamic model for  $\xi_t$  will be excited by  $x_t$ .<sup>12</sup> In other words, if  $x_t$  is interpreted as a stochastic process, then a random component will enter every element of  $\xi_t$  thus insuring that some uncertainty is present in each state. This characteristic is important when specializing these concepts to the estimation problem, and will be treated again below. Second, when the finite upper bound,  $c_2 I$ , exists and  $x_t$  is given a stochastic interpretation, then the effects of this random input on  $\xi_t$  will remain bounded in a mean-square sense. This implication also has an important role to play in the TVP estimation problem, and will be more thoroughly discussed below.

A second qualitative consideration arose from the feedback nature of many optimal control schemes. Since a feedback control required the state for its implementation, and normally only the endogenous variables (outputs) were available for measurement, it became increasingly important to ascertain whether information about the state of the system could be extracted solely from observations made on the outputs. If a model possessed this characteristic it was said to be observable, i.e., there existed a coupling between all of the states and the outputs. In relation to the representation (3.1), the concept of observability is best described by the following definition:<sup>13</sup>

Definition 2. The model (3.1) is said to be uniformly completely observable if, and only if, there exist an integer  $N > 0$  and constants  $c_1, c_2 > 0$  such that,

$$(3.4) \quad 0 < c_1 I \leq O(t, t-N) \leq c_2 I$$

for all  $t \geq N$ , where the observability matrix  $O(t_1, t_0)$  is defined by

$$(3.5) \quad O(t_1, t_0) = \sum_{\tau=t_0}^{t_1} \phi'(\tau, t_1) H_\tau' H_\tau \phi(\tau, t_1).$$

Uniform complete observability (UCO) implies that given enough observations  $\{y_\tau; t-N \leq \tau \leq t\}$  it is possible to solve exactly for  $\xi_t$ . In its original (purely deterministic) sense, with  $x_t$  interpreted as a known function of time, this ability to recover  $\xi_t$  also implies the exact recovery of  $\xi_{t-N}$ . In fact, for a linear deterministic system such as (3.1), the observability of the state at any one time  $t$  implies the observability at any other time.<sup>16</sup> When a stochastic environment is encountered such that  $x_t$  is random or measurement errors are present, then  $\xi_t$  cannot be determined exactly from a finite set of observations. However, uniform complete observability will be useful in that it can establish the possibility of estimating  $\xi_t$  exactly given an infinite set of data. This is akin to consistency so that observability has an important role to play in estimation.



### 3.2 Observability, Controllability, and Estimation

In order to establish the usefulness of UCO and UCC, they must be related to estimation and identification. To this end, two lemmas are presented which explicitly display the fundamental role played by these concepts. But before this is done, the definitions given previously are specialized to TVP estimation problems characterized by (2.2) - (2.3).

Definition 3. The model (2.2) - (2.3) is said to be uniformly completely controllable with respect to the noise,  $u_t$ , if and only if there exists an integer  $N > 0$  and constants  $c_1, c_2 > 0$  such that (3.2) holds for all  $t \geq N$ , where the controllability matrix is now defined as

$$(3.6) \quad C(t_1, t_0) = \sum_{\tau=t_0}^{t_1-1} \phi^{t_1-\tau-1} G G' (\phi^{t_1-\tau-1})'$$

where  $G = Q^{\frac{1}{2}}$ .<sup>16</sup>

Definition 4. The model (2.2) - (2.3) is said to be uniformly completely observable if and only if there exist an integer  $N > 0$  and constants  $c_1, c_2$  such that (3.4) obtains, where the observability matrix is now defined as

$$(3.7) \quad O(t_1, t_0) = \sum_{\tau=t_0}^{t_1} (\phi^{\tau-t_1})' X' X_{\tau} \phi^{\tau-t_1}$$

These are simply re-statements of Definitions 1 and 2 with the special structure imposed by (2.2) - (2.3) taken into account.

The following lemma exhibits the existence of a priori bounds on  $P_{t/t}$  in terms of  $C(t, t-N)$  and a slightly modified  $O(t, t-N)$ . Proof of

this lemma is omitted since it may be found in Jazwinski [1970, Chapt. 7, pp. 234-239], Aoki [1967, Chpt. 6, pp. 214-221], Bucy and Joseph [1968, Chapt. 6, pp. 70-71] or McGarty [1974, Appendix C, pp. 363-377].

Lemma 1. Let the system (2.2) - (2.3) be uniformly completely observable and uniformly completely controllable, and let  $P_{o/o} \geq 0$ , then  $P_{t/t}$  is uniformly bounded for all  $t \geq N$  according to

$$(3.8) \quad 0 < [O_R(t, t-N) + C^{-1}(t, t-N)]^{-1} \leq P_{t/t} \leq O_R^{-1}(t, t-N) + C(t, t-N) < \infty$$

where

$$(3.9) \quad O_R(t, t-N) = \sum_{\tau=N}^t (\phi^{\tau-t})' X_t' R^{-1} X_\tau \phi^{\tau-t}.$$

Thus UCC and UCO are seen to guarantee a meaningful problem in the sense that a positive definite  $P_{t/t}$  is assured for every  $t \geq N$ . Upon examination, (3.8) reveals that UCO is crucial in establishing a finite upper bound, while UCC is vital in establishing a positive definite lower bound (since  $[O_R + C^{-1}]^{-1} = [CO_R + I]^{-1}C$ ).

The UCC condition of Lemma 1 is somewhat restrictive in that there are three important cases in econometrics where UCC fails to obtain:

(1) constant parameter estimation; (2) purely deterministic (or systematic) parameter estimation; and (3) a mixture of systematic and stochastic parameter variation. In the first two instances  $Q \equiv 0$  forces  $C(t, t-N) \equiv 0$ , whereas in the third instance  $Q \geq 0$  together with certain  $\phi$  can lead to  $C(t, t-N) \geq 0$ . In these circumstances no positive definite lower bound on  $P_{t/t}$  exists by the hypotheses of Lemma 1. Fortunately this dilemma can be easily overcome by a slight modification of Lemma 1:

Lemma 2. Let the system (2.2) - (2.3) be UCO but not necessarily UCC,  $GG' \geq 0$  and  $P_{o/o} > 0$ , then  $P_{t/t}$  is uniformly bounded for all  $t \geq N$  according to

$$(3.10) \quad 0 < P_{t/t} \leq O_R^{-1}(t, t-N) + C(t, t-N) < \infty.$$

The key requirement of this lemma is that  $P_{o/o} > 0$ . If estimation is initiated with a proper prior on  $\beta_o$  then  $P_{t/t}$  will be positive definite whether or not controllability obtains. It appears quite reasonable in practice to expect  $P_{o/o} > 0$  when estimating constant or nonstochastic parameters, otherwise the nonstochastic elements of  $\beta_t$  would be known exactly at  $t=0$  and it would be senseless to attempt to estimate them! Note, however, that even when some  $\beta_t$  elements are nonstochastic ( $GG' \geq 0$ ),  $\phi$  may be such that  $C(t, t-N)$  is positive definite for some  $t$ . This situation would be a manifestation of the controllability property as discussed after (3.2): Namely, that the particular structure in  $\phi$  would eventually result in the "scattering" or distribution of randomness due to  $u_t$  working its way through the system into all the components of  $\beta_t$ . Eventually all of the elements of the parameter vector would contain some uncertainty so that  $P_{t/t} > 0$  must be satisfied.

The results contained in Lemmas 1 and 2 should not be interpreted as restricted solely to the Kalman filter method of parameter estimation. The Kalman Filter is a member of the BLUE class so that any and all estimators that are BLUE must yield the same  $P_{t/t}$  matrix. Thus the bounds presented above remain independent of the estimation method. Indeed, both  $O(t, t-N)$  and  $C(t, t-N)$  are defined independently of the estimator.

### 3.3 Observability, Controllability, and Identification

The main import of Lemmas 1 and 2 is the uniform boundedness of  $P_{t/t}$ , its definiteness, and its invertibility. Whenever  $0 < P_{t/t} < \infty$ ,  $P_{t/t}^{-1}$  exists and is positive definite. This is extremely important because it is intuitively clear that uncertainty (as represented by  $P_{t/t}$ ) and information are inversely proportional. Thus in obtaining a priori conditions which guarantee the positive definiteness of  $P_{t/t}$ , it is possible to insure, a priori, the identifiability of the parameters. The relationship between observability, controllability, and identifiability can be put on a more rigorous footing by examining the special case of systematic parameter variation.

Consider the identification problem associated with the estimation of  $\beta_t$  in (2.2) - (2.3) given only  $Y_t = \{y_{t-N}, \dots, y_t\}$  when  $Q=0$  and  $R=I$ . The classical approach to the identification problem would be to view  $\beta_t$  as the only vector to be estimated so that all identification questions focus on it and neglect the rest of the parameter (vector) trajectory. With this interpretation, identification would be determined by examination of the singularity of the Fisher Information Matrix or the definiteness of the Hessian of the Information Integral.<sup>17</sup> In the present situation both approaches are equivalent because the relevant probability densities are continuous in  $\beta_t$ . Therefore the Information Matrix,

$$(3.11) \quad I(t, t-N) = -E \left\{ \frac{\partial}{\partial b_t} \ln p(Y_t; b_t) \cdot \frac{\partial}{\partial b_t} \ln p(Y_t; b_t) \right\} \Big|_{b_t = \beta_t}$$

is constructed directly. From (2.2) - (2.3) it is known that the sequence  $\{y_{t-N}, \dots, y_t\}$  are independent, Gaussian, and have mean  $X_t b_t$  and variance-covariance  $I$ . Thus,

$$p(y_{\tau}; b_{\tau}) = \frac{1}{(2\pi)^{N/2} \det^{1/2}(I)} \exp \{-\frac{1}{2}[y_{\tau} - X_{\tau} b_{\tau}]' [y_{\tau} - X_{\tau} b_{\tau}]\}$$

$$p(Y_t; b_t) = \prod_{\tau=t-N}^t p(y_{\tau}; b_{\tau})$$

$$\begin{aligned} \ln p(Y_t; b_t) &= \sum_{\tau=t-N}^t \ln p(y_{\tau}; b_{\tau}) \\ &= \sum_{\tau=t-N}^t \ln p(y_{\tau}; \phi(\tau, t) b_t - \sum_{j=\tau}^{t-1} \phi(\tau, j+1) \Gamma z_j) \\ &= \text{const.} - \frac{1}{2} \sum_{\tau=t-N}^t \{ [y_{\tau} - X_{\tau} \phi(\tau, t) b_t + X_{\tau} \sum_{j=\tau}^{t-1} \phi(\tau, j+1) \Gamma z_j]' \\ &\quad \times [y_{\tau} - X_{\tau} \phi(\tau, t) b_t + X_{\tau} \sum_{j=\tau}^{t-1} \phi(\tau, j+1) \Gamma z_j] \} \end{aligned}$$

where the last line follows from the constraints placed on the parameter evolution, (2.3). The differentiation of the above expression with to  $b_t$  is now straightforward,

$$\begin{aligned} \frac{\partial}{\partial b_t} \ln p(Y_t; b_t) &= \sum_{\tau=t-N}^t \phi'(\tau, t) X_{\tau}' [y_{\tau} - X_{\tau} \phi(\tau, t) b_t \\ &\quad + X_{\tau} \sum_{j=\tau}^{t-1} \phi(\tau, j+1) \Gamma z_j]. \end{aligned}$$

Insertion of this expression into (3.11) yields

$$\begin{aligned}
 I(t, t-N) &= -E\left\{ \sum_{\tau=t-N}^t \phi'(\tau, t) X_{\tau}' [\cdot] \right. \\
 &\quad \left. X_{\tau} \phi(\tau, t) \right\} \Big|_{b_t = \beta_t} \\
 &= - \sum_{\tau=t-N}^t \left[ \phi'(\tau, t) X_{\tau}' E\{ [\cdot] [\cdot]' \} \right. \\
 &\quad \left. X_{\tau} \phi(\tau, t) \right] \Big|_{b_t = \beta_t} \\
 (3.12) \quad &= - \sum_{\tau=t-N}^t \phi'(\tau, t) X_{\tau}' X_{\tau} \phi(\tau, t)
 \end{aligned}$$

$$\text{where } [\cdot] = [y_{\tau} - X_{\tau} \phi(\tau, t) b_t + X_{\tau} \sum_{j=\tau}^{t-1} \phi(\tau, j+1) \Gamma z_j].$$

The quantity inside the brackets,  $[\cdot]$ , is just the observation error,  $e_t$ , and  $E\{e_t e_t'\} = I$ . Thus the information matrix and the observability matrix are identical.<sup>18</sup> It is clear that this result also holds for the constant parameter case ( $\phi = I$ ,  $Q = 0$ , and  $\Gamma z_t = 0$ ). When the observation errors are contemporaneously correlated,  $R \neq I$ , then it is easily shown that

$$\begin{aligned}
 (3.13) \quad I(t, t-N) &= \sum_{\tau=t-N}^t \phi'(\tau, t) X_{\tau}' R^{-1} X_{\tau} \phi(\tau, t) \\
 &= O_R(t, t-N)
 \end{aligned}$$

In realistic applications there are measurement errors in every element of  $y_t$  so that  $R$  and  $R^{-1}$  are full rank, positive definite matrices. Thus

it is easy to see that for any variance-covariance matrix  $R > 0$ , the condition of uniform complete observability is both necessary and sufficient for classical identification to obtain. These observations can be summarized by the following theorem:

Theorem 1. The TVP model (2.2) - (2.3) with  $Q = 0$ , (nonstochastic parameter variation, is completely identified) if and only if it is uniformly completely observable.

The feeling that UCO and identifiability were intimately related has now been borne out; in the case of systematic parameter, identifiability and observability are equivalent.

#### 4. Econometric Identification and Consistency

The results of the previous section are well known in the engineering literature although they have not, to the best of our knowledge been derived elsewhere in the context of the time varying parameter estimation problem. Most important, however is the fact that the equivalence of the observability and the information matrix reinforces the generic link between the econometric identification problem and the properties of the Kalman Filter. This link provides some insight as to how the criteria for the identification of time varying models differ from those of the constant parameter problem. In addition, we can more easily explore some of the asymptotic properties of time varying parameter estimators.

##### 4.1 Identification of Constant and Systematically Varying Parameterizations

The constant parameter model is obtained from the general representation (2.2)-(2.3) by specifying  $\Phi = I$ ,  $\Gamma = 0$  and  $Q = 0$  so that

$$(4.1) \quad \begin{aligned} \beta_{t+1} &= \beta_t \\ y_t &= X_t \beta_t + \varepsilon_t \end{aligned} \quad .$$

This model is identified if and only if

$$(4.2) \quad c_1 I \leq O(t, t-n) = \sum_{\tau=t-n}^t X'X_{\tau} \leq c_2 I$$

is satisfied for some  $c_1, c_2, N > 0$  and every  $t \geq N$ . Since  $\beta$  does not change it is obvious that  $t = T$  and  $N = T$  and that (4.2) is simply the well known multicollinearity condition or the requirement in the engineering literature that there be "persistently exciting inputs"(Astrom and Bohlin [1966]).

The systematically varying model is obtained from (2.2)-(2.3) by specifying  $Q = 0$  so that



$$(4.3) \quad \begin{aligned} \beta_{t+1} &= \Phi \beta_t + \Gamma Z_t \\ Y_t &= X_t \beta_t + \varepsilon_t \end{aligned} ,$$

where  $\Phi$  is a non-singular transition matrix, and  $\Gamma Z_t$  may or may not be zero. Theorem 1 states that this model is identified if and only if

$$(4.4) \quad c_1 I \leq O(t_1, t-N) = \sum_{\tau=t-N}^t (\Phi^{\tau-t})' X_{\tau}' X_{\tau} (\Phi^{\tau-t}) \leq c_1 I$$

is satisfied from some  $c_1, c_2, N > 0$  and  $t \geq N$ . The implication of systematic parameter variation is now clear. Since we have imposed additional structure on the problem by specification of the parameter transition process, identification is no longer solely dependent on the properties of the exogenous variables, it now also depends on the specified form of the parameter variation.

It is worth exploring how the specification of the parameter process can alter the standard conditions for identification. To illustrate this let us replace the summation in (4.4) by a matrix inner product:

$$O(t, t-N) = L' L$$

where,

$$L' = [(\Phi^{-N})' X_{t-N}' \mid (\Phi^{1-N})' X_{t-N+1}' \mid \dots \mid (\Phi^{-1})' X_{t-1}' \mid X_t'] .$$

$O(t, t-N)$  will be positive definite if  $L$  is of full rank. But, since  $\Phi$  is nonsingular, the rank of  $L$  is equivalent to the rank of

$$(4.5) \quad \tilde{L}' = [X_{t-N}' \mid \Phi' X_{t-N+1}' \mid \dots \mid (\Phi^{N-1})' X_{t-1}' \mid (\Phi^N)' X_t'] .^{19}$$

In the constant parameter case identification was concerned only with the linear independence of the columns of the various  $\{X_t (t-N \leq \tau \leq t)\}$ . Now, however, consideration must be given to the linear independence of the columns of  $X_{\tau}$  under the nonsingular transformation  $\Phi^{\tau-t}$ .

One of the consequences of Theorem 1 and the above observations is that the effects of multicollinearity may be neutralized by specification of a time-varying structure. Consider the following idealized example of the most extreme form of multicollinearity where an explanator is proportional to another:

$$X_t = [x_{1t} \ x_{2t}] = [x_{1t} \ \alpha x_{1t}]$$

Under the assumption of constant parameterization the observability criterion (4.2) reduces to a test of the rank of  $\tilde{L}'_C$  where

$$\tilde{L}'_C = \begin{bmatrix} x_{1,t-N} & | & x_{1,t-N+1} & | & \dots & | & x_{1,t} \\ \alpha x_{1,t-N} & | & \alpha x_{1,t-N+1} & | & & | & \alpha x_{1,t} \end{bmatrix}$$

obviously rank  $\tilde{L}'_C = 1$ . Now assume that  $\Phi \neq I$ , i.e.

$$\Phi = \begin{pmatrix} a & b \\ c & d \end{pmatrix}.$$

The rank test now must be applied to  $\tilde{L}'$  where,

$$\tilde{L}' = \begin{bmatrix} x_{1,t-N} & | & ax_{1,t-N+1} + \alpha x_{1,t-N+1} & | & \dots & | \\ \alpha x_{1,t-N} & | & bx_{1,t-N+1} + d\alpha x_{1,t-N+1} & | & & | \end{bmatrix}$$

If  $a \neq d$  and  $b \neq 0$ ,  $c = 0$  it is quite possible to find two linearly independent columns from among just the first two in  $\tilde{L}'$ . Therefore, the imposition of additional prior information in the form of  $\Phi$ , may serve to identify an otherwise unidentified model. It must be remembered, however, that the converse is also possible, namely a certain specification of  $\Phi$  when coupled with a linearly independent set  $\{X_t; 1 \leq t \leq T\}$  may be unidentified. Finally, note that the specification of any diagonally structured  $\Phi$  will play a benign role in regard to multicollinearity problems and identification.

As mentioned in Section 2, the econometric time-varying estimation problem is characterized by a two-step procedure, the first step dealing with the estimation of the unknown initial parameter vector  $\beta_0$ , and the second step dealing with the estimation of the trajectory starting from  $\beta_0$ . Now that time-variation has been introduced ( $\beta_0 \neq \beta_t \neq \beta_{t-}$ ), it is of interest to ask whether identification of  $\beta_t$  for some  $t$  implies identification of  $\beta_0$ : Do separate identification problems arise for each  $\beta_t$   $0 \leq t \leq T$ ? If  $\beta_t$  is identified for some  $t$  via satisfaction of (4.4), then  $\beta_0$  can always be recovered by solving (4.3a) backwards using  $\beta_t$  as a terminal condition. The identifiability of  $\beta_0$  given that of  $\beta_t$  can also be deduced from a rederivation of either the observability or classical information matrix in the special case where  $\beta_0$  is to be estimated instead of  $\beta_t$  given data over  $0 \leq t \leq T$ :

$$O(T,0) = \sum_{\tau=0}^T (\Phi^{T-\tau})' X_{\tau}' X_{\tau} \Phi^{T-\tau}$$

or

$$I(T,0) = \sum_{\tau=0}^T (\Phi^{T-\tau})' X_{\tau}' R^{-1} X_{\tau} \Phi^{T-\tau} .$$

Since  $I(T,0)$  will be full rank if  $O(T,0)$  is, identification of  $\beta_0$  reduces once again to a rank test on  $O(T,0)$ . But the observability matrix can be written as,

$$O(T,0) = M' M$$

where

$$(4.6) \quad M' = [X_0' \mid \Phi X_1' \mid \dots \mid (\Phi^N)' X_N'] .$$

Comparison with (4.5) reveals that the two criteria are identical. Thus, the identification of the initial parameter vector,  $\beta_0$ , and the identification of  $\beta_t$  are synonymous.

#### 4.2 Identification of Stochastically Varying Models

The introduction of stochastic parameter variation somewhat complicates the derivation of identifiability conditions. The complication is due to the presence of  $u_t$  in (2.3) which leads to a correlated error structure in the observation equation (2.2). This invalidates the construction used for Theorem 1 because now  $p(Y_t; b_t) \neq \prod_{\tau=t-N}^t p(y_\tau; b_t)$ , and a new expression must be derived for the Fisher Information matrix under these circumstances.

The fully stochastic nature of the model yields an equivalent observation equation, parameterized in  $\beta_t$ , of the following form

$$\begin{aligned}
 (4.7) \quad y_\tau &= X_\tau [\Phi(\tau, t)\beta_t + \sum_{j=t}^{\tau-1} \Phi(\tau, j+1)\Gamma z_j + \sum_{j=t}^{\tau-1} \Phi(\tau, j+1)G u_j] + \varepsilon_\tau \\
 &= X_\tau \Phi(\tau, t)\beta_t + X_\tau \sum_{j=0}^{\tau-1} \Phi(\tau, j+1)\Gamma z_j + \mu_\tau,
 \end{aligned}$$

where  $\mu_t$  is normally distributed with zero mean and variance-covariance matrix given by

$$(4.8) \quad \Omega_{ij} = E \{\mu_i \mu_j'\} = R\delta_{ij} + X_i \left[ \sum_{k=t}^{m_{ij}} \Phi(m, k+1) Q \Phi'(m, k+1) \right] X_j'$$

and  $m_{ij} = \min(k, j)$ . Thus the measurement errors,  $\mu_t$ , are no longer white -- each  $\mu_t$  is a moving-average process of order  $t$ .<sup>20</sup> In order to simplify the derivation,  $y_\tau$  is now replaced by  $\tilde{y}_\tau$  where

$$\tilde{y}_\tau = y_\tau - X_\tau \sum_{j=0}^{\tau-1} \Phi(\tau, j+1)\Gamma z_j. \quad .^{21}$$

Thus the model (2.2)-(2.3) can now be written as

$$(4.9) \quad \tilde{y}_\tau = X_\tau \Phi(\tau, t)\beta_t + \mu_t.$$

Both  $y_\tau$  and  $\tilde{y}_\tau$  are equivalent with regard to the identification problem for  $\beta_t$ .

The complicated expression for the sequential correlation of the  $\mu_t$ , (4.8), no longer permits the recursive development for the joint density required in the construction of  $I(t, t-N)$ . Therefore the joint process is formed directly, i.e.,

$$(4.10) \quad \tilde{Y}_t = Z_t b_t + \mu_t$$

where

$$\tilde{Y}_t = [\tilde{y}'_0 \mid \tilde{y}'_1 \mid \dots \mid \tilde{y}'_t]'$$

$$Z_t = [x'_0 \phi(0,t)' \mid x'_1 \phi(1,t)' \mid \dots \mid x'_t]'$$

$$U_t = [\mu'_0 \mid \mu'_1 \mid \dots \mid \mu'_t]'$$

The combined error vector still has zero mean but the associated variance-covariance matrix  $\underline{\Omega}$  is  $t\ell \times t\ell$

$$\underline{\Omega} = \begin{bmatrix} \Omega_{00} & \Omega_{01} & \dots & \Omega_{0t} \\ \Omega_{10} & \Omega_{11} & \dots & \Omega_{1t} \\ \vdots & \vdots & \ddots & \vdots \\ \Omega_{t0} & \Omega_{t1} & \dots & \Omega_{tt} \end{bmatrix} \quad ; \quad \Omega_{ij} = \Omega_{ji}$$

The joint process  $\tilde{Y}_t$  is normally distributed so that

$$p(\tilde{Y}_t; b_t) = \frac{1}{(2\pi)^{t\ell/2} \det^{\ell/2}(\underline{\Omega})} \exp \left\{ -\frac{1}{2} [\tilde{Y}_t - Z_t b_t]' \underline{\Omega}^{-1} [\tilde{Y}_t - Z_t b_t] \right\}$$

$$\ln p(\tilde{Y}_t; b_t) = -\frac{\ell t}{2} \ln(2\pi) - \frac{1}{2} \ln \det(\underline{\Omega}) - \frac{1}{2} U_t' \underline{\Omega}^{-1} U_t$$

$$(4.11) \quad \frac{\partial}{\partial b_t} \ln p(\cdot; \cdot) = Z_t' \underline{\Omega}^{-1} [\tilde{Y}_t - Z_t b_t]$$

Substitution of (4.11) into the definition (3.11) yields the result

$$I(t, t-N) = -Z_t' \underline{\Omega}^{-1} Z_t \Big|_{b_t = \beta_t}$$

$$(4.12) \quad I(t, t-N) = - [\phi'(0, t) x'_0 \mid \phi'(1, t) x'_1 \mid \dots \mid x'_t] \underline{\Omega}^{-1} \begin{bmatrix} x_0 \phi(0, t) \\ \hline x_1 \phi(1, t) \\ \hline \vdots \\ \hline x_t \end{bmatrix}$$

In order to have  $\beta_t$  identified, (4.12) must be a negative definite matrix. Since  $\underline{\Omega}$  is a variance-covariance matrix of a nonsingular stochastic process it is always positive definite, thus identification will obtain if and only if

$$[\phi'(0, t) x'_0 \mid \phi'(1, t) x'_1 \mid \dots \mid x'_t]$$

is full rank. But this is just the observability condition once again! Thus

Theorem 2. The fully stochastic TVP model (2.2)-(2.3), i.e. with  $Q \geq 0$ , is completely identified at  $t$  if and only if it is uniformly completely observable at time  $t$ .

Note that with  $Q \equiv 0$ ,  $\underline{\Omega}$  reduces to a block diagonal matrix in  $R^{-1}$  so that (4.12) can then be written as (3.13).

The identifiability of an empirically determined prior  $b_0$  can be investigated using Theorem 2 with  $t = 0$ . In this case, given data up through time  $t$ , the identifiability (observability) criterion reduces to a test for the full rank of

$$[x'_0 \mid \phi' x'_1 \mid \dots \mid (\phi^t)' x'_t] \quad .$$

But this is just  $(\phi^{-1}(0, t))' Z_t$  so once again, the identifiability of  $\beta_t$  implies the identifiability of  $\beta_\tau$  for any  $0 \leq \tau \leq t$ . This is the same result as obtained in the systematic variation case, although not so obvious.

It was noted in section 2 that estimation is a two-step procedure where the first step is the estimation of  $\beta_0$  and the covariance structures and the subsequent step the determination of particular realizations conditional on  $b_0$ . As the above result indicates, however, the identification of these realizations is a trivial matter, given  $b_0$ , because the state equation is an identifying function in the sense of Kadane [1974].

#### 4.3 Consistency in Time-Varying Parameter Estimation

The usefulness of the concepts of observability and controllability is not confined to questions of identification in time-varying models. The consistency of the Kalman filter estimates can also be examined using these concepts. In fact, uniform complete observability and controllability, together with Lemma 1, permit an almost trivial treatment.

Constant Parameters. Considering the behavior of the estimates based on all of the observations, Lemma 1 yields,

$$(4.13) \quad 0 \leq P_{t/t} \leq O_R^{-1}(t,0)^{22}$$

where  $t \geq N$ . Now it is possible to express  $O_R$  as the sum,

$$(4.14) \quad O_R(t,0) = O_R(N,0) + O_R(2N,N) + \dots$$

Given uniform complete observability,  $O_R(t,0)$  become the sum of positive definite matrices so  $O_R^{-1}(t,0) \rightarrow 0$  as  $t \rightarrow \infty$ . Thus (4.13) implies that uniform complete observability insures  $P_{t/t} \rightarrow 0$  as  $t \rightarrow \infty$ , i.e. consistency.

Systematic Parameter Variation. The same upper bound exhibited in (4.13) holds, but (4.14) must be replaced by

$$(4.15) \quad O_R(t,0) = (\Phi^{N-t})' [O_R(N,0) + O_R(2N,N) + \dots] \Phi^{N-t}$$

since, in general  $\Phi \neq I$ . Uniform complete observability still guarantees that the sum within the brackets grows without bound, but now there can exist  $\Phi$  such that  $\Phi^{N-t} \rightarrow 0$  as  $t \rightarrow \infty$ . Thus it is now possible for  $\Phi^{N-t}$  and  $\sum_{k=1}^{\infty} O_R(kN, Nk-N)$  to interact in such a way that (4.15) has a finite limit. More must be said of the structure of  $\Phi$  before (4.13) can be used to establish consistency. A simple sufficient condition resulting in consistent filtered estimates is that only  $\Phi$  with eigenvalues on or within the unit circle be considered in specifying the model structure. For such  $\Phi$ ,  $\Phi^{N-t}$  ( $t \geq N$ ) is an unstable matrix so all three factors in (4.15) grow without bound as  $t \rightarrow \infty$ . If  $\Phi = I$  then the above problem reduces to that of the constant parameter case, even though  $\Gamma z_t \neq 0$ .

Stochastic Parameter Variation. In the general stochastic case the upper bound from Lemma 1 must include  $C(t,0)$ :

$$(4.16) \quad 0 < P_{t/t} \leq O_R^{-1}(t,0) + C(t,0) .$$

Since  $Q \neq 0$  implies  $C(t,0) \neq 0$ , uniform complete observability with an appropriately chosen  $\Phi$  (i.e.,  $O_R^{-1}(t,0) \rightarrow 0$  as  $t \rightarrow \infty$ ) is no longer sufficient to guarantee consistency. Moreover, reference to (2.10) reveals that  $P_{t/t-1} \neq 0$  for any  $t$  so long as  $Q \neq 0$ . Thus at each observation  $P_{t/t}$  can never be zero because  $P_{t/t-1}$  is never zero. The best that can be hoped for is that some finite limiting distribution exist for  $\beta_{t/t}$ , and to achieve this (4.16) suggests that  $\Phi$  be specified such that  $C(t,0) \rightarrow \alpha < \infty$  as  $t \rightarrow \infty$ . In turn, the definition of  $C(t,0)$  reveals that, in order to obtain an *a priori* finite distribution on  $\beta_{t/t}$ ,  $\Phi$  must be a stable matrix (i.e., all eigenvalues within the unit circle). However, it should be noted that this is not necessary--it only implies that the bounding technique stemming from Lemma 1 loses its usefulness in such situations.



Thus, the stability of  $\Phi$  is again only a sufficient condition for the boundness of the variance-covariance matrix of the estimated coefficients. Whether one should consider unstable transition matrices would appear to depend on the properties of the inputs (the  $x_t$ ), because the observability of the system will depend upon these properties. If the system is observable then it is not clear that one should attach too much importance to the consistency property. In such instances, it seems that it is more important to worry about whether an estimator is efficient. Cooley and Prescott [1972] have shown that estimators with  $\Phi = I$  will be efficient.

The inability of the filter to produce consistent parameter estimates under stochastic variation suggests that the optimal "smoother" estimates might be a better alternative.<sup>15</sup> However this is not the case; inconsistency still persists so long as  $Q \neq 0$ . This claim can best be substantiated by viewing the optimal smoother as a combination of two optimal filters, one run forward in time from  $t=0$  to  $t$ , and the other run backwards in time from  $t=T$  to  $t$ , (See Fraser and Potter [1969]).

The smoothed estimation error variance-covariance matrix is given by

$$P_{t/T} = [\bar{P}_{t/t}^{-1} + \underline{P}_{t/t}^{-1}]^{-1}$$

where  $\bar{P}_{t/t}$  denotes the forward filter variance-covariance matrix and  $\underline{P}_{t/t}$  denotes the backward filter variance-covariance matrix. As  $T \rightarrow \infty$  only  $\underline{P}_{t/t}$  will change since only it depends on  $T$ . By increasing the data length,  $T$ , it might be hoped that the solution for  $\underline{P}_{t/t}$  could be initialized far enough into the future such that  $\underline{P}_{t/t} \rightarrow 0$ . But this can never happen. So long as  $Q \neq 0$ , the matrix Ricatti solution to the backward filter can never degenerate to zero, no matter how long in the filter its solution is initialized (Potter [1965]). Thus  $\underline{P}_{t/t}^{-1} < \infty$  and  $P_{t/T} > 0$  for all  $T \rightarrow \infty$ . Neither filtering or smoothing can yield consistent parameter estimates under stochastic parameter variation.

## 5. Summary and Conclusions

In the foregoing sections we have used the equivalence of the concepts of observability in control theory and information in statistics to establish the conditions for the identifiability of time varying structures in econometrics. Our intent has been to shed further light on the properties of time varying parameter estimation methods that have been introduced into econometrics in recent years. Several interesting albeit possibly obvious, conclusions emerge; the requirements for identification in time varying models are not in principle more stringent than in the classical constant parameter case and may be less so because of the additional structure imposed on the problem. The identifiability criteria and the asymptotic properties of the estimators suggest that it may not always make sense to consider processes with unstable transition structures. This latter caveat is not completely clear, however, because we have concentrated throughout on sufficient conditions for identification. Finally, the estimators for models subject to stochastic parameter variation will not be consistent in the usual sense. It has been shown, however, that a sufficient condition for them to have stable asymptotic distributions is that the transition matrix  $\Phi$  be stable (have eigenvalues within the unit circle). It should be noted that in the econometric environment where all observed data is to be processed "off-line" (i.e. in batch mode), it is always possible to verify the conditions for observability and controllability.

In the development of the results in this paper it has been assumed throughout that  $\phi$ , is known. It should be obvious that there is a deeper issue of identifiability that should be dealt with and that is the identifiability of  $\phi$  when it is unknown as it is in most applications. This is indeed a more complex problem which requires further investigation. It appears at the present time that a fruitful approach to the problem is to concentrate on the connection between specification errors in  $\phi$  and the serial correlation properties of the innovations process (the estimated residuals). In this approach one is essentially searching for the structure that is empirically best. This can be thought of as a generalization of the Box-Jenkins approach to modelling. Similar approaches have been taken in the engineering literature to studying the sensitivity of the Kalman Filter. See for example Mehra [1970], Boozer and McDaniel [1972], Martin and Stubberud [1974].

Finally, all of the current discussion has been confirmed to the reduced form models. A more difficult problem would be to look at the identification of time varying structural models. Before this can be done, however, the estimation methods must be extended to structural form models.

FOOTNOTES

1. For an interesting interpretation and presentation of Hannan's results see Preston and Wall [1974].
2. See Rosenberg [1968, 1973b], Pagan [1975], Sarris [1973].
3. The authors are currently working on the derivation of the appropriate recursive estimation procedures for structural form relations.
4. The reasons for expecting stochastic time variation have been extensively explored elsewhere. See for example Cooley [1971], Cooley and Prescott [1973], Lucas [1973] and Rosenberg [1968].
5. For an excellent survey of generic relations among models with non-constant coefficients see Rosenberg [1973a].
6. An exception is Aoki [1967] who did not, however, solve the problem correctly.
7. The conditions are that the system is uniformly completely observable and uniformly completely controllable. These concepts are explained in the next section.
8. This is a compact representation of the equations in Rosenberg [1973a].
9. The relationship between the state space representation and the other more traditional forms of econometric model representation is not too difficult to establish. For a simple translation from the reduced form see Chow [1972a], and from the structural form see Pindyck [1973].

For the purpose of this paper, assume  $\xi_t$  is a  $k \times 1$  vector of states for each  $t$ ,  $x_t$  an  $n \times 1$  vector of exogenous variables, and  $y_t$  is an  $l \times 1$  vector of endogenous variables. The first of equations (3.1) is a dynamic relationship which possesses a solution (see for example Ogata [1967], DeRusso et al. [1965], or Meditch [1969]) of the form

$$\xi_{t_1} = \phi(t_1, t_0) \xi_{t_0} + \sum_{\tau=t_0}^{t_1-1} \phi(t_1, \tau+1) G_{\tau} x_{\tau}$$

relating the state at  $t=t_1$  to the state at  $t=t_0$  given the value of the exogenous variables over the interval  $[t_0, t_1-1]$ . The matrix  $\phi(t_1, t_0)$  is called the fundamental matrix and is nonsingular, satisfying

$$\phi(t+1, t_0) = F_t \phi(t, t_0) ; \phi(t_0, t_0) = I,$$

$\phi(t, \tau) = \phi^{-1}(\tau, t)$ . For the time varying system of (3.1),

$$\phi(t, \tau) = F_t F_{t-1} \dots F_{\tau}.$$

For the time-invariant structure of (1.2) or (2.3) it is easily seen that  $\phi(t, \tau) = \phi^{t-\tau}$ . The solution of the state equation as given above plays a vital role in the derivation of the concepts of controllability and observability.

10. The derivation of the controllability criterion contained in the following definition is beyond the scope of this paper. The interested 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). The choice of uniform complete controllability in this work is principally motivated by its use in discussing the qualitative aspects of the Kalman filter.
11. Let  $x$  be any arbitrary  $p \times 1$  vector and  $A$  any  $p \times p$  matrix. Then  $\alpha I \leq A \leq \beta I$  is taken to mean  $x'x \leq x'Ax \leq \beta x'x$  where  $I$  is the  $p \times p$  identity matrix.
12. A complete treatment of the modal interpretation of controllability is beyond the scope of this paper. The reader is referred to DeRusso, et al. [1965, pp. 344-349 and pp. 429-431], Ogata [1967, pp. 424-425], and Kalman, Ho, and Narendra [1962].
13. Comments similar to those contained in Footnote 10 apply here: there are several definitions of observability, but only the one most useful with respect to time-varying parameter estimation and identification has been employed.
14. Since  $H_t$  is generally some time-varying matrix constructed of known time functions other than explanatory variables, the control literature emphasises how "controllability and observability are quantities

exhibited by the system and not the data". This is clearly not the case with the Kalman filter representation of the TVP estimation problem when  $X_t$  replaces  $H_t$ . The strict intrinsic interpretation of controllability does, however, remain valid since the data never enters into the definition of  $F_t$  or  $G_t$ .

15. The correct criterion for establishing the ability to determine  $\xi_t$  instead of  $\xi_{t-N}$  would involve  $O(t-N, t)$ . But this can be obtained from  $O(t, t-N)$  merely by exchanging the arguments in  $\phi$  in (3.5), i.e. by employing  $\phi(t_1, \tau)$ . Since  $\phi$  is nonsingular and  $\phi(\tau, t_1) = \phi^{-1}(t_1, \tau)$ , this can always be done.
16. Since  $Q$  is a variance-covariance matrix, it is symmetric and at least positive semi-definite. If it is positive definite it is always possible to find a unique factorization  $Q = Q^{\frac{1}{2}}(Q^{\frac{1}{2}})'$ . Then (2.3) is equivalent to

$$\beta_{t+1} = \phi \beta_t + G\mu_t \quad \mu_t \sim N(0, I)$$

If, say,  $k_1$  of the  $\beta$ 's are to be systematically varying, i.e. no stochastic components due to any element of  $\mu_t$ , then  $Q$  can only be positive semi-definite. A re-ordering of equations could then be employed to obtain

$$Q = \begin{bmatrix} Q_1 & | & 0 \\ \hline 0 & | & 0 \end{bmatrix}$$

where  $Q_1$  is a  $k_1 \times k_1$  positive definite matrix. The unique factorization of  $Q_1$  could then be effected and employed in the first  $k_1$  rows of  $G$  with the last  $k-k_1$  rows being replaced by a null matrix.

17. See Rothenberg [1971] and Bowden [1973].
18. This result is well known in the control literature, having been first stated by Kalman and Bucy [1960] and more recently by Jazwinski [1970].
19. An examination of the rank of  $\tilde{L}$  is preferred over that of  $L$  because no inverse of  $\phi$  is involved, making computation that much less demanding.

20. Note, however, that with  $Q \equiv 0$ ,  $\mu_t$  reduces to  $c_t$  and the previous development of the Information Matrix (Section 3) applies. The interested reader is referred to Sarris [1973] and Cooper [1973] who also use this method to reduce the TVP model to a standard regression model format.
21. This transformation can always be effected since it is assumed that  $\Gamma z_j$  is known exactly.
22. The exact relationship  $P_{t/t}^{-1} = P_{o/o}^{-1} + Q_R(t,0)$  holds in this case. However, in order to dispense with any consideration of the effects of the prior distribution, the relationship written as  $P_{t/t}^{-1} \geq Q_R(t,0)$ . See Jazwinski [1970, p. 231 - 236].
23. Both Jazwinski [1970, pp. 236] and Aoki [1967, pp. 215] note this, and state a sufficient condition which guarantees consistency: If  $\|Q_R(t,0)\|^{-1} \rightarrow 0$  faster than  $\|\phi(t,0)\| \rightarrow \infty$  then obviously  $P_{t/t} \rightarrow 0$ .

BIBLIOGRAPHY

- [1] Aoki, M. [1967], Optimization of Stochastic Systems, New York: Academic Press.
- [2] Aström, K.J. [1970], Introduction to Stochastic Control Theory, New York: Academic Press.
- [3] \_\_\_\_\_ and T. Bohlin [1966], "Numerical Identification of Linear Dynamic Sys. from Normal Oper. Records", in Theory of Self-Adaptive Control Systems (Ed. PhH. Hammond), New York: Plenum Press.
- [4] 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.
- [5] Boozer, D.D. and W.L. McDaniel [1972], "On Innovation Sequence Testing of the Kalman Filter", IEEE Trans. Auto. Control, Vol. AC-17, No. 1, pp. 158-160.
- [6] Bowden, R. [1973], "The Theory of Parametric Identification", Econometrica, Vol. 41, No. 6, pp. 1069-1074.
- [7] Bucy, R.S. and P.D. Joseph [1968], Filtering for Stochastic Processes with Applications to Guidance, New York: J. Wiley & Sons.
- [8] Chow, G.C. [1972], "Optimal Control of Linear Econometric Systems with Finite Time Horizon", IJR, Vol. 13, No. 1, (February), pp. 16-25.
- [9] Cooley, T.F. [1971], "Estimation in the Presence of Sequential Parameter Variation", Unpublished Ph.D. dissertation, Department of Economics, University of Pennsylvania.
- [10] Cooley, T.F. and E. Prescott [1973], "Varying Parameter Regression", "A Theory and Some Applications", Annals of Economic and Social Measurement, Vol. 3, October.
- [11] Cooper, J.P. [1973], "Time-Varying Regression Coefficients: A Mixed Estimation Approach and Operational Limitations of the General Markov Structure", Annals of Economic and Social Measurement, Vol. 2, No. 4, (October), pp. 525-530.
- [12] DeRusso, P.M., R.J. Roy, and C.M. Close [1965], State Variables for Engineers, New York: J. Wiley.
- [13] Fisher, F.M. [1966], The Identification Problem in Economics, New York: McGraw-Hill.
- [14] Hannan, E.J. [1970], Multiple Time Series, New York: Wiley.
- [15] \_\_\_\_\_ [1969], "The Identification of Vector Mixed Autogressive-Moving Average Systems", Biometrika, 57, pp. 223-225.
- [16] \_\_\_\_\_ [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. [1960], "A New Approach to Linear Filtering and Prediction Problems", Transactions of ASME, Series D, Journal of Basic Engineering, 82, pp. 35-45.
- [19] \_\_\_\_\_ 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.
- [20] \_\_\_\_\_ Y.C. Ho, and K.S. Narendra [1962], "Controllability of Linear Dynamical Systems"; Cont. Differential Equations, Vol. 1, pp. 189-213.
- [21] Koopmans, T.C. and O. Reiersol [1950], "The Identification of Structural Characteristics", Annals of Mathematical Statistics, Vol. 21, pp. 165-181.
- [22] \_\_\_\_\_ 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.
- [23] Martin, W.C. and A.R. Stubberud [1974], "An Additional Requirement for Innovations Testing in System Identification", IEEE Trans. Auto. Control, Vol. AC-19, No. 7, pp. 583-584.
- [24] McGarty, T.P. [1974], Stochastic Systems and State Estimation, New York: John Wiley & Sons.
- [25] Meditch, J.S. [1969], Optimal Linear Estimation and Control, New York: McGraw-Hill.
- [26] Mehra, R.K. [1970], "On the Identification of Variances and Adaptive Kalman Filtering", IEEE Transactions on Automatic Control, April, pp. 175-184.
- [27] \_\_\_\_\_ [1971], "On Line Identification of Linear Dynamic Systems with Applications to Kalman Filtering", IEEE Transactions on Automatic Control, AC-16, No. 1, February.
- [28] \_\_\_\_\_ [1972], "Approaches to Adaptive Filtering", IEEE Transactions on Automatic Control, October, pp. 693-698.
- [29] \_\_\_\_\_ [1974], "Identification in Control and Econometrics: Similarities and Differences", Annals of Economic and Social Measurement, Vol. 3, pp. 21-47.
- [30] Ogata, K. [1967], State Space Analysis of Control Systems, Englewood Cliffs: Prentice Hall.
- [31] Pagan, A.R. [1974], "A Note on the Extraction of Components from Time Series", Econometrica, 43, pp. 163-168.
- [32] Pindyck, R.S. [1973], Optimal Planning for Economic Stabilization, Amsterdam: North Holland.

- [33] Potter, J.E. [1965], "A Matrix Equation Arising in Statistical Filter Theory", NASA Contract Report, NASA CR-270.
- [34] Preston, A. and K.D. Wall [1973], "A Review of ARMAX System Identification Theory", Mimeo.
- [35] Rosenberg, B.M. [1968], "Varying-Parameter Estimation", Unpublished Ph.D. dissertation, Department of Economics, Harvard University.
- [36] \_\_\_\_\_ [1973a], "A Survey of Stochastic Parameter Regression", Annals of Economic and Social Measurement, Vol. 3, October.
- [37] \_\_\_\_\_ [1973b], "The Analysis of a Cross Section of Time Series by Stochastically Convergent Parameter Regression", Annals of Economic and Social Measurement, Vol. 3, October.
- [38] Rosenbrock, H.H. [1970], State-Space and Multivariable Theory, New York: John Wiley & Sons.
- [39] Rothenberg, T.J. [1971], "Identification in Parametric Models", Econometrica, 39, pp. 577.
- [40] Sarris, A. [1973], "A Bayesian Approach to Estimation of Time-Varying Regression Coefficients", Annals of Economic and Social Measurement, October, pp. 501-523.
- [41] Swamy, P.A.V.B. [1970], "Efficient Inference in a Random Coefficient Regression Model", Econometrica, 38, pp. 311-323.
- [42] Tse, E. and J. Anton [1972], "On the Identifiability of Parameters", IEEE Trans. Auto. Control, Vol. AC-17, No. 5, October.
- [43] Zadeh, L. and C.A. Desoer [1963], Linear Systems Theory--The State Space Approach, New York: McGraw-Hill.