

REGRESSION AND KALMAN FILTER METHODS
FOR TIME-VARYING ECONOMETRIC MODELS

Carlo Carraro

Econometric Research Program
Research Memorandum No. 320

September 1985

The author is grateful to Gregory Chow, Whitney Newey and Peter Hartley for helpful comments on a previous version of this paper.

ABSTRACT

This paper analyses and generalizes two methods for estimating time-varying parameter models: the regression method and the Kalman filter-innovation correlation method. Identification conditions for the unknown parameters of a general time-varying model are provided and the two methods are shown to imply the same identification conditions. Furthermore, an explicit estimator is derived for all the parameters of the model, including the initial values of the time-varying parameters and the covariance matrices of the model. The properties of the estimates are also analysed. Finally, the relationship between observability and identifiability of the time-varying parameters is explored by using the smoothing equations.

INTRODUCTION

In recent years many papers have been written on the estimation of time varying linear econometric models.^{1/} With a few exceptions, these studies make use of strong prior information to solve identification and estimation problems. Furthermore, even when the identification and estimation of all the unknown parameters of the model are explicitly considered, different, apparently unrelated approaches have been proposed (see, for example, Swamy-Tinsley, 1980; Pagan, 1980). In particular, two general methods for estimating time-varying parameter models can be identified in the econometric literature: the regression method and the Kalman Filter method. A brief description of the two methods is useful in order to understand their differences and similarities.

The most general results on the estimation of time-varying parameter models based upon the regression method are provided by Swamy-Tinsley (1980). In that paper the following model is considered:

$$\begin{aligned}
 y_t &= x_t \beta_t \\
 \beta_t &= Z_t \delta + u_t & Z_t &= \text{diag}(z_{1t} \dots z_{kt}) \\
 u_t &\sim \text{ARMA}(r, s)
 \end{aligned}$$

where y_t is the dependent variable, x_t and $\{z_{it}; i = 1 \dots k\}$ are row vectors respectively of k and $\{\bar{p}_i; i = 1 \dots k\}$ fixed explanatory variables, β_t is a $(k \times 1)$ vector of time-varying parameters, and δ is a $(p \times 1)$ vector of constant parameters, where $p = \sum_{i=1}^k \bar{p}_i$. Furthermore, r and s are the known parameters defining the order of the ARMA process representing the structure of the stochastic process $\{u_t\}$. Identifiability conditions

and estimation equations for all the unknown parameters of the model are provided by Swamy-Tinsley (1980) by using the following approach: first apply OLS to:

$$y_t = x_t Z_t \delta + v_t \quad v_t = x_t u_t$$

then use the residuals obtained from this regression for estimating the structure of the ARMA process $\{u_t\}$ and re-estimate $y_t = x_t Z_t \delta + v_t$ by GLS. This is the basic framework of the regression method.

The Kalman Filter method, in contrast, considers the state-space representation of the econometric model and uses the Kalman Filter to derive the innovation sequence $\{m_t\}$ (see Appendix A) from which the innovation likelihood can be determined and then maximized in order to obtain estimates of the unknown parameters of the model (see Chow, 1983, for a detailed analysis of this approach). Identifiability conditions and estimation equations for all the unknown parameters of the model by using the Kalman Filter method are provided by Pagan (1980). He considers the following model

$$y_t = x_t \beta_t + \epsilon_t \quad \epsilon_t \sim \text{NID}(0, \sigma^2)$$

$$M(L)(\beta_t - \bar{\beta}) = u_t \quad u_t \sim \text{NID}(0, Q)$$

where $M(L) = \sum_{i=1}^r M_i L^i$ represents the autoregressive structure of the stationary stochastic process which is supposed to describe the dynamics of the time-varying parameter vector β_t . Furthermore, $\bar{\beta}$ is a vector of constant parameters and $\{\epsilon_t\}$ and $\{u_t\}$ are supposed white noise.

A particular version of the Kalman Filter approach was provided by Mehra(1970) where, instead of numerically maximizing the innovation likelihood, it is proposed to derive the estimates of the unknown parameters of the model from the autocovariance structure of the innovation sequence, which cannot be considered a white noise sequence when the covariance matrices and the transition matrix of the system are unknown. However, only time invariant state space form (and the state space form implied by Pagan's model is time invariant) are considered by Mehra(1970).

The brief outline of the two main methods for estimating time-varying parameter models that we have presented above gives us the possibility to provide a precise description of the results contained in this paper.

In particular:

(i) Both the regression method and the Kalman Filter method will be used to derive estimators of all the unknown parameters of the following time-varying linear econometric model:

$$(0.1) \quad y_t = x_t \beta_t + \varepsilon_t \quad \varepsilon_t \sim \text{NID}(0, \sigma^2)$$

$$(0.2) \quad \beta_t = M \beta_{t-1} + Z_t \delta + u_t \quad Z_t = \text{diag}(z_{1t} \dots z_{kt})$$

$$(0.3) \quad u_t = A(L)u_{t-1} + v_t \quad v_t \sim \text{NID}(0, Q)$$

where M is a $(k \times k)$ matrix representing the autoregressive structure of the time-varying vector β_t and $A(L) = \sum_{i=1}^s A_i L^{i-1}$ represents the autoregressive structure of the stochastic process $\{u_t\}$. It is possible to show that the model (0.1)-(0.3) includes, as particular cases, the models considered by Swamy-Tinsley(1980), Pagan(1980), Chow(1983), and it can also be considered general enough to include most applied econometric applications of time-varying parameter models.

Identifiability conditions and estimation equations will be provided for all the unknown parameters of the model: δ , $\{\beta_t; t=0, 1 \dots T\}$, σ^2 , Q , M , $(A_i; i=1 \dots s)$.

(ii) It will be shown that the identifiability conditions derived by applying the regression method to the model (0.1)-(0.3) coincide with the identifiability conditions derived by applying the Kalman Filter method. Therefore, Swamy-Tinsley's and Mehra's results will be generalized and compared.

(iii) The theoretical properties of the estimates obtained by using both methods will be explored. In particular, by using the results provided by Magnus(1978), it will be shown that the regression method provides consistent and asymptotically efficient estimates of β_0 and δ and consistent estimates of σ^2 , Q , when M and $\{A_i\}$ are supposed to be known. Otherwise, all the parameters of the model are consistently estimated. The same conclusion will be proved to hold when the Kalman filter method is used.

(iv) By using the regression method, an exact estimation equation for the initial conditions β_0^* and P_0 will be derived when β_0 is assumed to be $\beta_0 \sim N(\beta_0^*, P_0)$. Usually, either β_0^* and P_0 are estimated by using an approximate estimation equation (see, e.g. Chow, 1983) or the initial conditions are assumed to be known.

(v) By using the Kalman Filter method, the relationship between the observability of the system and the identifiability of the parameters included in the state vector (see Appendix A) will be examined, and it will be shown that the identification condition and the observability condition coincide only if the smoothing equations are used to estimate the state vector.

(vi) The innovation correlation method proposed by Mehra(1970) will be generalized to time-varying systems and then used, together with the Kalman Filter, for estimating the unknown parameters of the model.

The plan of the paper is the following: in the first part, regression methods will be analysed and the properties of the estimates obtained by performing a three stage least-squares algorithm will be examined. Furthermore, inequality constrained estimators will be proposed for the covariance matrices $\sigma^2 I$, P_0 , and Q in order to guarantee the positive semidefiniteness of those matrices.

The second part of the paper will consider the Kalman Filter method and will show how the parameters of the model (0.1)-(0.3) can be estimated by combining the Kalman Filter with the innovation correlation method. Furthermore, identifiability conditions and properties of the estimates will be analysed.

PART I: REGRESSION METHODS

This part of the paper deals with a simple algorithm, based on regression methods which is shown to be appropriate for estimating different types of time-varying parameter models. Restrictions for the covariance matrices of the model to be positive semidefinite will be provided and the properties of the estimates will be discussed. When possible, both the heteroscedasticity structure and the autocorrelation structure of the residuals will be used to identify the unknown elements of the covariance matrices of the model, the initial conditions and the transition matrix of the time-varying parameters, and the autocorrelation structure of the residuals.

A. A restricted model

We consider the general specification of the time-varying parameter model described in the introduction (eqs. (0.1)-(0.3)), but we assume, for the sake of simplicity, that the initial parameter vector β_0 is fixed so that $P_0 = 0$. Furthermore, we assume the matrix M to be known and the disturbance vector u_t to be independently and identically distributed.

Therefore, the model is:

$$(1.1.1) \quad y_t = x_t \beta_t + \varepsilon_t \quad \varepsilon_t \sim \text{NID}(0, \sigma^2)$$

$$(1.1.2) \quad \beta_t = M\beta_t + Z_t \delta + u_t \quad u_t \sim \text{NID}(0, Q)$$

The previous restrictive assumption will be relaxed in the next sections. The unknown parameters of the model are the coefficient vector β_0 and δ and the covariance matrices σ^2 and Q .

In order to apply the three stage regression method that will be described below, equations (1.1.1) and (1.1.2) have to be combined and then written as a function of the initial parameter vector β_0 and the time invariant parameter vector δ .

Stacking the T observation we have:

$$(1.2) \quad y = X\beta^* + \epsilon$$

where

$$\beta^* = \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_T \end{bmatrix} \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \vdots \\ \epsilon_T \end{bmatrix} \quad X = \begin{bmatrix} x_1 & & & 0 \\ & \ddots & & \\ 0 & & \ddots & \\ & & & x_T \end{bmatrix}$$

Furthermore, the stacked equations (1.1.2) can be written as:

$$\begin{aligned} \beta^* &= \begin{bmatrix} 0 & 0 & \dots & 0 \\ M & 0 & \dots & 0 \\ & & \ddots & \\ 0 & \dots & M & 0 \end{bmatrix} \beta^* + \begin{bmatrix} M \\ 0 \\ \vdots \\ 0 \end{bmatrix} \beta_0 + \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_T \end{bmatrix} \delta + u \\ &= \begin{bmatrix} I & 0 & \dots & 0 \\ M & I & & \\ & & \ddots & \\ 0 & & & M & I \end{bmatrix}^{-1} \begin{bmatrix} M \\ 0 \\ \vdots \\ 0 \end{bmatrix} \beta_0 + \begin{bmatrix} z_1 \\ \vdots \\ z_T \end{bmatrix} \delta + u \\ &= \begin{bmatrix} I & 0 & \dots & 0 \\ M & I & & \\ M^2 & M & I & \\ \vdots & \vdots & \vdots & \\ M^{t-1} & \dots & M & I \end{bmatrix} \begin{bmatrix} M \\ 0 \\ \vdots \\ 0 \end{bmatrix} \beta_0 + \begin{bmatrix} z_1 \\ \vdots \\ z_T \end{bmatrix} \delta + u \end{aligned}$$

$$= \begin{bmatrix} M \\ M^2 \\ \vdots \\ M^T \end{bmatrix} \beta_0 + \begin{bmatrix} z_1 \\ Mz_1 + z_2 \\ \vdots \\ M^{T-1}z_1 + \dots + Mz_{T-1} + z_T \end{bmatrix} \delta + \begin{bmatrix} u_1 \\ Mu_1 + u_2 \\ \vdots \\ M^{T-1}u_1 + \dots + Mu_{T-1} + u_T \end{bmatrix}$$

where $u' = [u_1 \dots u_T]$.

Define

$$M_0 = \begin{vmatrix} M \\ M^2 \\ \vdots \\ M^T \end{vmatrix} \quad F = \begin{vmatrix} I & 0 & \dots & 0 \\ M & I & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ M^{T-1} & \dots & M & I \end{vmatrix} \quad z = \begin{vmatrix} z_1 \\ \cdot \\ \cdot \\ \cdot \\ z_T \end{vmatrix}$$

Then, the equation describing the dynamics of the T time-varying parameters can be written as

$$\beta^* = M_0 \beta_0 + FZ\delta + Fu$$

which, substituted into (1.2), gives

$$(1.3) \quad y = XM_0 \beta_0 + XFZ\delta + XF u + \varepsilon$$

Suppose M is known (the case where M is unknown will be considered later); then, by applying OLS to (1.3) we obtain $\hat{\omega} = y - XM_0 \hat{\beta}_0 - XFZ\hat{\delta}$ which is a consistent estimate of ω under suitable conditions (see Theil, 1971, pp. 386-387). This is the first step of the three-stage regression algorithm.

In particular, (1.3) provides consistent estimates of δ , β_0 and ω , if the matrix $(1/T)W'W$ approaches a positive definite matrix as T goes to infinity, where $W = [XM_0 : XFZ]$. In order to determine the sum of squares to be minimized in the second step of the algorithm, the covariance matrix of the disturbance vector ω must be computed.

Since $\omega = XF_u + \varepsilon$, we have:

$$(1.4) \quad \Omega = \sigma^2 I + XF(I \otimes Q)F'X'$$

$$= \sigma^2 I + \begin{bmatrix} x_1 Q x_1' & x_1 Q M' x_2' & \dots & x_1 Q M^{T-1} x_T' \\ x_2 M Q x_1' & x_2 (M Q M' + Q) x_2' & \dots & x_2 M Q M^{T-1} x_T' + x_2 Q M^{T-2} x_T' \\ \vdots & \cdot & \cdot & \vdots \\ x_T M^{T-1} Q x_1' & \cdot & \cdot & x_T \left[\sum_{i=0}^{T-1} M^i Q (M^i)' \right] x_T' \end{bmatrix}$$

Let us therefore stack the $T(T+1)/2$ elements of the upper triangular part of Ω into a vector Ω_{ah} . Furthermore, define $M(i) = \sum_{j=0}^{i-1} (M^j \otimes M^j)$, $i = 1, 2, \dots, T$ and

$$MX(i) = \begin{bmatrix} x_i \\ x_{i+1} M \\ \vdots \\ x_T M^{T-i} \end{bmatrix} \quad i=1, \dots, T$$

where $MX(i)$ is a $(T-i+1) \times k$ matrix.

By an appropriate vectorization of the upper triangular elements of Ω , we have:

$$\Omega_{ah} = \begin{bmatrix} \ell_1 \\ \vdots \\ \ell_T \end{bmatrix} \sigma^2 + \begin{bmatrix} MX(1) & & 0 \\ & \cdot & \\ 0 & & MX(T) \end{bmatrix} \begin{bmatrix} I_k \otimes x_1 & & 0 \\ & \cdot & \\ 0 & & I_k \otimes x_T \end{bmatrix} \begin{bmatrix} M(1) \\ \vdots \\ M(T) \end{bmatrix} \Gamma Q^*$$

where $\Gamma Q^* = \text{vec}(Q)$, Γ is a known selection matrix which maps the $k(k+1)/2$ unknown elements of Q (defined by Q^*) into $\text{vec}(Q)$ and ℓ_i is a $(T-i+1)$ vector of which the first element is 1 and the remaining elements are zero.

By obvious definitions, the above equation can be rewritten as

$$(1.5) \quad \Omega_{ah} = \ell_0 \sigma^2 + M_X X_0 M^* \Gamma Q^* \equiv$$

$$\equiv \begin{bmatrix} \ell_0 \\ \vdots \\ M_X X_0 M^* \Gamma \end{bmatrix} \theta^* \equiv \Pi_X^{ah} \theta^* \quad \theta^* = \begin{bmatrix} \sigma^2 \\ \vdots \\ Q^* \end{bmatrix}$$

Therefore, after having obtained $\hat{\Omega} = \hat{\omega}\hat{\omega}'$, the following estimator can be derived:

$$(1.6) \quad \hat{\theta}^* = [(\Pi_x^{ah})' \Pi_x^{ah}]^{-1} (\Pi_x^{ah})' \hat{\Omega}_{ah}$$

Notice that, if the sample is not very large, the inequality constrained estimator must be computed in order to assure the positive semidefiniteness of the covariance matrices σ^2 and Q . If Q is assumed to be diagonal, the Dantzig-Cottle algorithm can be used. ^{2/}

Therefore, under the conditions stated by Liew(1976), (1.6) provides consistent estimates $\hat{\sigma}^2$ and \hat{Q} which can be used to replace σ^2 and Q into (1.4) so that $\hat{\Omega}$ can be obtained and the GLS estimator

$$(1.7) \quad \begin{bmatrix} \hat{\beta}_0 \\ \hat{\delta} \end{bmatrix} = (W' \hat{\Omega}^{-1} W)^{-1} W' \hat{\Omega}^{-1} y$$

can be computed in the third step of the algorithm.

Summarizing the previous analysis, the following three-step algorithm has been proposed:

Step 1: regress y on XM_0 and XFZ and compute $\hat{\omega} = y - XM_0 \hat{\beta}_0 - XFZ \hat{\delta}$.

Step 2: compute $\hat{\Omega} = \hat{\omega}\hat{\omega}'$ and regress $\hat{\Omega}_{ah}$ on Π_x^{ah} where $\hat{\Omega}_{ah}$ is a vector containing the elements of the upper triangular part of $\hat{\Omega}$.

Step 3: substitute $\hat{\sigma}^2$ and \hat{Q} obtained in Step 2 into (1.4) and compute the GLS estimator of β_0 and δ .

The properties of the estimates obtained by using the previous algorithm can easily be studied by applying the results presented in Magnus(1978). First of all, it is necessary to check the identifiability of the parameters of the model by analysing the structure of the information matrix defined by the log-likelihood function of the model (1.3). From Magnus(1978, Theorem 3) we have:

$$(1.8.1) \quad I = \begin{bmatrix} W' \Omega^{-1} W & 0 \\ 0 & I_{\theta\theta}^* \end{bmatrix}$$

where I denotes the information matrix and

$$(1.8.2) \quad I_{\theta\theta}^* = \frac{1}{2} \left[\frac{\partial \text{vec}(\Omega^{-1})}{\partial \theta^*} (\Omega \otimes \Omega) \frac{\partial \text{vec}(\Omega^{-1})'}{\partial \theta^*} \right]$$

Therefore, since the parameters in δ are independent from those in θ^* , a first identification condition is

$$(1.9) \quad \text{rank}(W) = k+p \quad T \geq k+p$$

where $p = \sum_{i=1}^k \bar{p}_i$ is the number of parameters contained in the vector δ (see Lemma 1 in Magnus, 1978).

Furthermore, Assumption 5 and Lemma 1 in Magnus (1978) imply that the vector θ^* is also identifiable if and only if the $k(k+1)/2 + 1$ vectors $\partial \text{vec} \Omega^{-1} / \partial \sigma^2$, $\{\partial \text{vec} \Omega^{-1} / \partial q_{ij}, i=1, \dots, k; j=1, \dots, k\}$, where $Q = \{q_{ij}\}$, are linearly independent.

By using Gabrielsen's theorem (see Gabrielsen, 1978), the above condition is satisfied if the following rank condition holds:

$$(1.10.1) \quad \text{rank}(\Pi_x^{\text{ah}}) = \text{rank}[\ell_0 : M_x' X_0 M^* \Gamma] = k(k+1)/2 + 1$$

Given the structure of the regression (1.5), (1.10.1) is indeed sufficient for $\hat{\theta}^*$ to be a consistent estimator of θ^* , thus implying the identifiability of θ^* . Notice that if the parameter vectors β_0, δ and θ^* are required to be identified for any T , the characteristic roots of the matrix M must be less than 1 in absolute value, so that the maximum eigenvalue of the covariance matrix Ω has an upper bound and the information matrix (1.8.1) is positive definite for any T (see Theorem 2 in Pagan, 1980).

The asymptotic properties of the final estimates of δ and β_0 can be derived by applying Lemma 2 in Magnus (1978). Since by (1.10.1) and the convergence in probability of $\hat{\omega}_t$ to ω_t the estimates of the covariance matrices $\sigma^2 I$ and Q are

consistent, then the iteration of the three-step algorithm previously described yields, upon convergence, a consistent root of the maximum likelihood equations derived from the model (1.3). This root is the unique maximum likelihood estimator (see Lemma 2 in Magnus, 1978). Since, under the usual assumptions, the ML estimator is consistent and asymptotically efficient, we can conclude that the final GLS estimates are also consistent and asymptotically efficient.

It must be noticed that the same conclusion could be obtained by using the block diagonality of the information matrix and theorems 1 and 3 in Pagan (1984). However, Magnus's results hold even when the information matrix is not diagonal, if the three-step algorithm is iterated until convergence.

The results obtained in this section can be compared with the results obtained by following different approaches. First of all, in the next section it will be shown that the identification condition (1.9) coincides with the observability of the system derived from the model (0.1)(0.2). Furthermore, two particular cases of the identification condition (1.10.1) can be considered.

Let us first consider the T diagonal elements of Ω , so that the heteroscedastic structure of the disturbance vector ω is used to estimate σ^2 and Q.

From equation (1.4), we have:

$$\begin{array}{c} \left| \begin{array}{c} \omega_{11} \\ \vdots \\ \omega_{TT} \end{array} \right| = i\sigma^2 + \left| \begin{array}{ccc} x_1 \otimes x_1 & & \\ & \ddots & \\ & & 0 \\ & & & \ddots & \\ & 0 & & & \\ & & & & x_T \otimes x_T \end{array} \right| \left| \begin{array}{c} M(1) \\ \vdots \\ M(T) \end{array} \right| \Gamma Q^* \end{array}$$

which, by obvious definition, can be written as:

$$\begin{aligned}
\Omega_h &= \sigma^2 + W_x^h M^* \Gamma Q^* \\
&\equiv [i : W_x^h M^* \Gamma] \theta^* \\
&\equiv \pi_x^h \theta^*
\end{aligned}$$

Hence, a sufficient identification condition is:

$$(1.10.2) \quad \text{rank} \left(\pi_x^h \right) = 1 + k(k+1)/2 \quad T \geq 1 + k(k+1)/2$$

which coincides with the condition derived by Swamy-Tinsley (1980) when $M = 0$ or $M = I$. Therefore, condition (1.10.2) generalizes Swamy-Tinsley's results to a more general model and condition (1.10.1) shows that Swamy-Tinsley's approach provides only a sufficient condition when $M \neq 0$.

Finally, we want to consider another particular case of equation (1.5). In the second part of this paper, we will derive identification conditions for σ^2 and Q by using the Kalman Filter and the innovation correlation method (see Mehra, 1970).

These identification conditions can be shown to coincide exactly with the conditions derived from the last column of (1.4), i.e. when only the T elements describing the autocovariance structure of the disturbance vector ω are used.

Let us therefore consider the last column of the covariance matrix Ω . By an appropriate vectorization, we obtain

$$\Omega_a = \begin{vmatrix} \omega_{1T} \\ \cdot \\ \cdot \\ \omega_{TT} \end{vmatrix} = \begin{vmatrix} 0 \\ 0 \\ \vdots \\ 1 \end{vmatrix} \sigma^2 + \begin{vmatrix} x_T^M \otimes x_1 & 0 \\ \cdot & \cdot \\ 0 & x_T^M \otimes x_{T-1} \\ \cdot & \cdot \\ x_T \otimes x_T \end{vmatrix} \begin{vmatrix} M(1) \\ \cdot \\ \cdot \\ n(T) \end{vmatrix} \text{vec}(Q) \equiv$$

$$\equiv [e_T : W_x^a M^* \Gamma] \begin{bmatrix} \sigma^2 \\ Q^* \end{bmatrix} \equiv \pi_x^a \theta^*$$

where the definitions are obvious.

Therefore, if the equation $\Omega_a = \pi_x^a \theta^*$ is used to estimate σ^2 and Q , the identification condition becomes

$$(1.10.3) \quad \text{rank}(\pi_x^a) = 1 + k(k+1)/2 \quad T \geq 1 + k(k+1)/2$$

which will be seen to coincide with the condition which is required when the innovation correlation method is used to estimate σ^2 and Q .

A final remark is related to the positive semidefiniteness of the covariance matrix Q . Even in this case, if the sample is not sufficiently large and Q is not diagonal, the second step of the algorithm must minimize either

$$(\hat{\Omega}_h - \pi_x^h \theta^*)' (\hat{\Omega}_h - \pi_x^h \theta^*) \quad \text{or} \quad (\hat{\Omega}_a - \pi_x^a \theta^*)' (\hat{\Omega}_a - \pi_x^a \theta^*)$$

under the constraints for Q to be positive semidefinite, which were described previously. If Q is supposed to be diagonal, the Dantzig-Cottle algorithm can be applied.

The consistent estimates of σ^2 and Q obtained in the second step are then used to obtain consistent and asymptotically efficient estimates of δ by applying GLS in the third step of the algorithm.

B. Unknown A Priori Distribution of β_0 .

If β_0 , instead of being fixed, is assumed to be

$$(1.11) \quad \beta_0 = \beta_0^* + v_0 \quad v_0 \sim N(0, P_0)$$

the estimation algorithm must be modified in the following way: substituting (1.11) into (1.3), we have:

$$(1.12) \quad y = XM_0\beta_0^* + XFZ\delta + \varepsilon + XFu + XM_0v_0$$

which provides the residuals of the first-stage OLS regression. The covariance matrix of the disturbance vector $\omega = \varepsilon + XFu + XM_0v_0$ is

$$(1.13) \quad \Omega = \sigma^2 I + XF(I \otimes Q)F'X' + XM_0P_0M_0'X'$$

which can be used to derive explicit estimators for the unknown elements of the covariance matrices σ^2 , Q and P_0 . Since we have

$$XM_0P_0M_0'X' = \begin{vmatrix} x_1M_0P_0M_0'x_1' & x_1M_0P_0(M^2)'x_2' & \dots & x_1M_0P_0(M^{T-1})'x_T' \\ \vdots & x_2M^2P_0(M^2)'x_2' & & \\ \vdots & & \ddots & \\ x_TM^TP_0M_0'x_1' & & & x_TM^TP_0(M^T)'x_T' \end{vmatrix}$$

an appropriate vectorization of the upper triangular part of Ω (and the results of the previous section) provide:

$$(1.14) \quad \Omega_{ah} = \ell_0\sigma^2 + M_X X_0 M^* \Gamma Q^* + M_X X_0 \bar{M} \Gamma_0 P_0^*$$

where Γ_0 is a known selection matrix which maps the vector P_0^* containing

the unknown elements of P_0 , into $\text{vec}(P_0)$; M_x , M^* and X_0 have been defined in the previous section and

$$\bar{M} = \begin{vmatrix} M \otimes M \\ M^2 \otimes M^2 \\ \vdots \\ M^T \otimes M^T \end{vmatrix}$$

Therefore, by obvious definitions, we have

$$(1.15) \quad \Omega_{ah} = \begin{bmatrix} l_0 & : & M_x X_0 M^{*T} & : & M_x X_0 \bar{M} \Gamma_0 \end{bmatrix} \begin{vmatrix} \sigma^2 \\ Q^* \\ P_0^* \end{vmatrix} \equiv \pi_0^{ah} \theta_0^*$$

which can be used to obtain consistent estimates of σ^2 , Q and P_0 .

The identification condition is

$$(1.16) \quad \text{rank}(\pi_0^{ah}) = 1 + k(k+1) \quad T \geq 1 + k(k+1)$$

After having obtained $\hat{\sigma}^2$, \hat{Q} and \hat{P}_0 from (1.15) where $\hat{\Omega}_{ah}$ has replaced Ω_{ah} , GLS can be applied to (1.12) in order to obtain, under the usual assumptions, consistent and asymptotically efficient estimates of δ and β_0^* .

Even in this case it is possible to select only T elements of the matrix Ω in order to obtain consistent estimates of σ^2 , Q and P_0 . If the diagonal elements (the heteroscedastic structure) of Ω are chosen, we have

$$\begin{aligned} \Omega_h &= i\sigma^2 + W_x^h M^* \Gamma Q^* + W_x^h \bar{M} \Gamma_0 P_0^* \equiv [i : W_x^h M^* \Gamma : W_x^h \bar{M} \Gamma_0] \theta_0^* \\ &\equiv \pi_0^h \theta_0^* \end{aligned}$$

The identification condition becomes

$$\text{rank } (\pi_0^h) = 1 + k(k + 1)$$

which is a sufficient condition for (1.16) to be satisfied. If the correlation structure of the disturbance vector ω is used, from (1.13), we have

$$\begin{aligned} \Omega_a &= e_T \sigma^2 + W_x^a M^* \Gamma Q^* + W_x^{a\bar{M}} \Gamma_0 P_0^* \equiv [i : W_x^a M^* \Gamma : W_x^{a\bar{M}} \Gamma_0] \theta_0^* \\ &\equiv \pi_0^a \theta_0^* \end{aligned}$$

which implies

$$\text{rank } (\pi_0^a) = 1 + k(k + 1)$$

which is not a sufficient condition for (1.16) to be satisfied, since $\text{rank } (W_x^{a\bar{M}}) \leq k$. Therefore, only if P_0 is assumed to be diagonal, does the correlation approach provide consistent estimates of all the unknown elements of P_0 .

We conclude this section by emphasizing that the regression approach that we have just proposed provides consistent estimates of β_0^* and P_0 (i.e. the a priori distribution of β_0) instead of requiring these parameters to be a priori known. This is important, since in the econometric literature no satisfactory solution for the initial condition problem has been provided.

C. The Transition Matrix M

All the previous analysis is based on the assumption that the transition matrix M is known. Therefore, in this section we provide a slightly more complicated algorithm which can be used to obtain consistent estimates of M . Suppose to start the three-step algorithm by assuming $M = M^0$ where M^0 is known and fixed. The goal of the following analysis is to derive an estimation equation for revising the starting matrix M^0 at the end of the first iteration of the three-step algorithm.

In the previous section we showed that $\omega_0 = \varepsilon + XFu + XM_0v_0$ is estimated at the end of the first step, θ^* at the end of the second step and

$$\hat{\omega}^1 = y - XM_0\hat{\beta}_0^* - XFZ\hat{\delta}$$

at the end of the third step, where $\hat{\beta}_0^*$ and $\hat{\delta}$ are GLS estimates and depend on the starting transition matrix M^0 . The fourth step of the algorithm is the following.

Step 4: Use $\hat{\omega}^1$ to compute a new estimate of Ω and use equation

$$(1.17) \quad \Omega = X(M_0P_0M_0' + F(I \otimes Q)F')X' + \sigma^2I$$

in order to derive an estimator for M . Equation (1.17) can be written as

$$\hat{\Omega}^1 = \hat{\sigma}^2I + \begin{bmatrix} x_1(M\hat{P}_0M' + \hat{Q})x_1' & \dots & x_1(M\hat{P}_0M' + \hat{Q})(M^{T-1})'x_T' \\ x_2M(M\hat{P}_0M' + \hat{Q})x_1' & & \vdots \\ \vdots & & \vdots \\ x_T M^{T-1}(M\hat{P}_0M' + \hat{Q})x_1' & \dots & x_T(M^T\hat{P}_0M^T + \sum_{i=0}^{T-1} M^i\hat{Q}M^{i'})x_T' \end{bmatrix}$$

where the previously obtained estimates of P_0 , Q and σ^2 were introduced so that the only unknown elements are the coefficients of M .

Let us consider the upper triangular part of $\hat{\Omega}^1$ without the main diagonal elements and define it as $\hat{\Omega}^u$. By stacking the rows of $\hat{\Omega}^u$ into a $T(T-1)/2$ vector $\hat{\Omega}_a^u$ and by vectorizing each element of $\hat{\Omega}^u$ we have:

$$(1.18) \quad \hat{\Omega}_a^u = \begin{bmatrix} N_x(1) & 0 \\ & \ddots \\ 0 & N_x(T-1) \end{bmatrix} \begin{bmatrix} I_k \otimes x_1 \hat{M}(1) \\ \vdots \\ I_k \otimes x_{T-1} \hat{M}(T-1) \end{bmatrix} \text{vec}(M')$$

where

$$N_x(i) = \begin{vmatrix} x_{i+1} \\ \vdots \\ x_{i+2}^M \\ \vdots \\ x_T^M \end{vmatrix} \quad i = 1, \dots, T-1$$

$$\text{and} \quad \hat{M}(i) = M^i P_0(M^i)' + \sum_{j=0}^{i-1} M^j \hat{Q}(M^j)' \quad i = 1, \dots, T-1$$

By obvious definitions, (1.18) can be rewritten as

$$(1.19.1) \quad \hat{\Omega}_a^u = (N_* M_*) \text{vec}(M') \equiv N^* \text{vec}(M')$$

so that the identification condition becomes

$$(1.19.2) \quad \text{rank}(N^*) = k^2 \quad T \geq k^2$$

Therefore M can be estimated by using equation (1.19.1) and the explicit estimator is given by

$$\text{vec}(\hat{M}) = (N^*{}'N^*)^{-1}N^*{}'\hat{\Omega}_a^u$$

Therefore, when M is unknown the algorithm is based on an interactive procedure: first, M is assumed known and all the other parameters of the model are estimated; then, given those estimates, the transition matrix M is estimated and the interactive procedure is iterated. Therefore, in this case, even the regression method requires a certain degree of a priori knowledge in the first iteration of the algorithm (see Todini, 1978, for a similar iterative procedure based upon the Kalman filter). How sensitive the speed of convergence is to a misspecified initial transition matrix is a matter to be investigated by Montecarlo simulations. Furthermore, since M is consistently estimated, Lemma 2 in Magnus (1978) can still be used to show that all the parameters of the model are consistently estimated by the four-step algorithm described in this section and that they are asymptotically normally distributed. However, when M is unknown, the information matrix becomes:

$$I = \begin{bmatrix} W'\Omega^{-1}W & W'\Omega^{-1}(\gamma' \otimes I)\partial W/\partial\theta \\ (\partial W/\partial\theta)'(\gamma \otimes I)\Omega^{-1}W & I_{\theta\theta} \end{bmatrix}$$

where

$$\gamma' = [\beta_0' \ ; \ \delta] \quad \theta' = [\theta_0^*{}' \ ; \ \text{vec}(M)']$$

and

$$I_{\theta\theta} = \frac{1}{2} \left(\frac{\partial \text{vec} \Omega^{-1}}{\partial \theta} \right) (\Omega \otimes \Omega) \left(\frac{\partial \text{vec} \Omega^{-1}}{\partial \theta} \right)' + \\ + \left(\frac{\partial W}{\partial \theta} \right)' (\gamma \otimes I) \Omega^{-1} (\gamma' \otimes I) \left(\frac{\partial W}{\partial \theta} \right)$$

Therefore, the information matrix is not block diagonal, so that the final estimates of the parameter vectors γ and θ can be proved to be consistent and asymptotically normally distributed, but not asymptotically efficient.

D. Autocorrelated Residuals

A similar situation arises when relaxing the assumption $u_t \sim \text{NID}(0, Q)$.

Assume, instead

$$(1.20) \quad u_t = A_1 u_{t-1} + A_2 u_{t-2} + \dots + A_s u_{t-s} + v_t \quad v_t \sim \text{NID}(0, Q)$$

where the $(k \times k)$ matrices $A_1 \dots A_s$ are unknown and must be estimated.

Equation (1.20) can be rewritten as

$$u_t^* = Au_{t-1}^* + v_t^*$$

where

$$u_t^* = \begin{vmatrix} u_t \\ \vdots \\ u_{t-s+1} \end{vmatrix} \quad A = \begin{vmatrix} A_1 & A_2 & \dots & A_s \\ I & 0 & \dots & \\ \vdots & & & \\ 0 & & & I & 0 \end{vmatrix} \quad v_t^* = \begin{vmatrix} v_t \\ 0 \\ \vdots \\ 0 \end{vmatrix}$$

and $u_t = Ju_t^*$ where $J \equiv [I \ 0 \ \dots \ 0]$.

Define:

$$u = \begin{vmatrix} u_s \\ \vdots \\ u_T \end{vmatrix} \quad v^* = \begin{vmatrix} J'v_s \\ \vdots \\ J'v_T \end{vmatrix} = (I \otimes J')v \quad v = \begin{vmatrix} v_s \\ \vdots \\ v_T \end{vmatrix}$$

Let us discuss the simple model presented in section B. We have:

$$w_t = \varepsilon_t + x_t u_t \quad t = s, \dots, T$$

and

$$\Omega = \sigma^2 I + X(I \otimes J)E(u^*u^{*'}) (I \otimes J')X'$$

Since

$$E(u^*u^{*'}) = \begin{vmatrix} I & 0 & \dots & 0 & J'QJ & 0 & I & A' & \dots & A^{T-1}, \\ A & I & & \vdots & & & 0 & & & \\ & & & \vdots & & & & & & A' \\ A^{T-1} & \dots & A & I & 0 & J'QJ & 0 & \dots & & I \end{vmatrix}$$

we can write

$$(1.21) \quad \Omega = \sigma^2 I + \begin{vmatrix} x_1' Q x_1' & x_1' Q A_1' x_2' & x_1' Q J (A^{T-1})' J' x_T' \\ x_2' J A J' Q x_1' & x_2' (A_1' Q A_1' + Q) x_2' & \vdots \\ \vdots & \vdots & \vdots \\ x_T' J A^{T-1} J' Q x_1' & \dots & x_T' \left(\sum_{i=0}^{T-1} J A^i J' Q J (A^i)' J' \right) x_T' \end{vmatrix}$$

Let us suppose to start the algorithm by assuming $A = A^0$. Then, the residuals \hat{w}_0 can be computed at the end of the first step and Ω can be estimated. The upper triangular part of $\hat{\Omega}$ can then be used to derive consistent estimates of σ^2 and Q , so that consistent estimates of δ can be obtained in the third step. Then, the residuals $\hat{w}_1 = y - W\hat{\delta}$ can be computed and the covariance matrix $\hat{\Omega}_1$ can be estimated, where $\hat{\Omega}_1$ has the form described by (1.21), but σ^2 and Q are replaced by $\hat{\sigma}^2$ and \hat{Q} . The upper triangular part of $\hat{\Omega}_1$ can then be vectorized in order to derive an estimation equation for A , so that the initial guessed value A^0 can be revised.

Since the algebra is particularly tedious and cumbersome and follows the derivation presented in the previous section for the transition matrix M , it will be omitted. However, it is important to emphasize that, as in the previous section, the upper triangular part of $\hat{\Omega}_1$ cannot be explicitly solved with respect to A , so that the identification conditions and the estimate of A depend on the prior matrix A^0 . The sensitivity of the final

estimates of the model to a misspecified starting matrix A^0 must be evaluated by Montecarlo experiments. Even if equation (1.21) cannot be explicitly solved for JA, it provides consistent estimates of the autocorrelation structure of the random disturbances of the model, and those estimates can be used to iterate the algorithm until convergence. A similar interactive procedure for estimating JA was suggested by Swamy-Tinsley (1980). 3/

Furthermore, even if it was not always explicitly stated, it is important to emphasize that when the vector θ^* is estimated (in the second step of the algorithm), the constraints for \hat{Q} to be positive semidefinite must always be imposed.

PART II: KALMAN FILTER METHODS

This part of the paper follows the engineering approach to the estimation of linear systems and provides identification conditions and estimation algorithms which are based on the Kalman filter and the innovation correlation approach. The structure of the time-varying parameter model coincides with the structure already examined in the first part of the paper, but a different estimation method is proposed. The properties of the estimates obtained by applying the Kalman filter will also be discussed at the end of this part of the paper.

A. Identification and Estimation

We start from the model (0.1)-(0.2), i.e.

$$(2.1) \quad y_t = x_t \beta_t + \varepsilon_t \quad \varepsilon_t \sim \text{NID}(0, \sigma^2)$$

$$(2.2) \quad \beta_t = M \beta_{t-1} + Z_t \delta + u_t \quad u_t \sim \text{NID}(0, Q)$$

of which the state-space form is

$$(2.3) \quad y_t = H_t b_t + \varepsilon_t$$

$$(2.4) \quad b_t = F_t b_{t-1} + G u_t$$

where

$$b_t = \begin{bmatrix} \beta_t \\ \delta_t \end{bmatrix} \quad H_t = [x_t \quad 0] \quad F_t = \begin{bmatrix} M & Z_t \\ 0 & I_p \end{bmatrix} \quad G = \begin{bmatrix} I_k \\ 0 \end{bmatrix}$$

Equations (2.3)-(2.4) can be interpreted as the state-space representation of a system with inputs u_t and output y_t where the structure of

the system is determined by the fixed matrices H_t , F_t , G for any time t .

The constancy of the coefficients δ is expressed by the equation $\delta_t = \delta_{t-1}$ while the $(n \times 1)$ coefficient vector b_t (where $n = k + p$) is known as a state vector.

The estimation method is composed of the following steps;

(1) Given some initial values M^0 , Q_0 , σ_0^2 , b_0 , P_0 where P_0 is the covariance matrix of the prior coefficient vector b_0 , the Kalman filter can be applied to obtain estimates of β_t and δ . Let us call I_t^0 the σ -algebra generated by the observation set $\{y_1 \dots y_t; x_1 \dots x_T; Z_1 \dots Z_T\}$. Then we let $b'_{t|j} = [\beta'_{t|j} \ ; \ \delta'_{t|j}]$ be an estimate, obtained by using the Kalman filter and conditional on I_j^0 , of β_t and δ , where $j \leq t$. The covariance matrix of these estimates is denoted by $P_{t|j}$.^{4/}

(2) The Kalman filter during its recursions generates the innovation sequence $\{y_t - H_t b_{t|t-1}; t = 1, \dots, T\}$ which is used to estimate the autocovariance function $C_{it} = E(m_t, m'_{t-i})$ where $m_t = y_t - H_t b_{t|t-1}$. When the Kalman filter is not optimal (i.e. when M , Q and σ^2 are unknown) we have:

$$C_{it} \neq 0 \quad i = 0, 1, \dots \quad \text{and} \quad t = 1, \dots, T.$$

Thus the sequence $\{C_{it}\}$ can be used to revise the initial values of the matrices M , Q , $\sigma^2 I$, P_0 . This approach is called the innovation correlation method and has been proposed by Mehra (1970) for constant systems.^{5/}

(3) Given revised initial values of the matrices Q , M , $\sigma^2 I$, P_0 and the vector b_0 , the previous steps are repeated until convergence, i.e. until the innovation sequence $\{m_t\}$ is a white-noise process.

Since the speed of convergence depends on the initial values b_0 , P_0 , Q_0 , M^0 , their choice becomes relevant. Furthermore, the final estimates of the coefficients should be independent of the initial conditions.

As far as b_0 and P_0 are concerned, it has been suggested (see Chow, 1981), to use the first $k + p$ observations for obtaining an approximate estimate of b_0 and P_0 . Furthermore, it can be shown that under appropriate conditions,^{6/} the effect of the initial values b_0 , P_0 vanishes asymptotically. However, when the sample is small, no result about the influence of the initial values on the final estimates of the parameters is available, so that this problem has to be investigated by numerical simulation (see Carraro-Sartore, 1984 for some preliminary results).

Let us consider first the estimation of the coefficient vector b_t . A sequence of estimated values $\{b_t|_t; t = 1, \dots, T\}$ can easily be obtained by applying to (2.3)-(2.4) the Kalman filter equations described in Appendix A. Furthermore, in Appendix B we show that these estimates coincide, either when the initial conditions are known or asymptotically, with the estimates obtained by applying to (2.3)-(2.4) the method of generalized least squares. This result generalizes Sant's (1977) proof to a dynamic system with time-varying transition matrix F_t . Furthermore, the proof provided in Appendix B gives us the possibility of studying the identifiability of b_t by analyzing the likelihood of the system (2.3)-(2.4). Therefore, the following analysis will use the likelihood proposed by Sant (1977) in order to show the coincidence between identifiability and obser-

vability of b_t (this result was first obtained by Cooley-Wall, 1975). However, it will be emphasized that a necessary and sufficient condition for b_t to be identified can be obtained only by using the smoothed estimate of b_t . This condition will be shown to coincide with the identification condition (1.9) derived in the first part of the paper.

In order to derive the likelihood function of b_t , let us write the system (2.3)-(2.4) in the following way:

$$y = xb_t + \varepsilon - Au \quad t = 2, 3, \dots, T$$

where

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_t \end{bmatrix} \quad x = \begin{bmatrix} H_1 M(1, t) \\ H_2 M(2, t) \\ \vdots \\ H_t \end{bmatrix} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_t \end{bmatrix} \quad u = \begin{bmatrix} u_2 \\ u_3 \\ \vdots \\ u_t \end{bmatrix}$$

$$A = \begin{bmatrix} H_1 M(1, 2)H & H_1 M(1, 3)G & \dots & H_1 M(1, t)G \\ 0 & H_2 M(2, 3)G & \dots & H_2 M(2, t)G \\ 0 & 0 & & \vdots \\ \vdots & \vdots & & \vdots \\ 0 & 0 & & H_{t-1} M(t-1, t)G \\ 0 & 0 & & 0 \end{bmatrix}$$

$$M(i, j) = \begin{cases} F_i F_{i-1} \dots F_{j+1} & i > j \\ I & i = j \\ F_{i+1}^{-1} F_{i+2}^{-1} \dots F_j^{-1} & i < j \end{cases}$$

The log-likelihood function of model (2.1) is

$$(2.5.1) \quad L = \text{const.} + \frac{1}{2} \log |S^{-1}| - \frac{1}{2} \text{tr}(y - xb_t)' S^{-1} (y - xb_t)$$

where

$$\begin{aligned} S &\equiv E(y - xb_t)(y - xb_t)' \equiv E(vv') \\ &= \sigma^2 I_t + A(I_{t-1} \otimes Q)A' \end{aligned}$$

Taking the first derivative of L with respect to b_t we have:

$$\frac{\partial L}{\partial b_t} = -x'S^{-1}(y - xb_t)$$

Therefore, it is easy to see that the coefficient vector b_t is globally identified if and only if

$$(2.5.2) \quad \text{rank}(x'S^{-1}x) = k + p = n$$

From Theorem 2 in Pagan(1980), we have:

$$(2.5.3) \quad 0 < \gamma_1 x'x \leq x'S^{-1}x \leq \gamma_m x'x$$

where γ_1 and γ_m are, respectively, the minimum and maximum eigenvalue of S^{-1} . Therefore, (2.5.2) and (2.5.3) imply $\gamma_1 > 0$, which is equivalent to saying that the maximum eigenvalue of the covariance matrix S must be finite. If (2.5.2) is required to hold for any $t \geq k + p$, then a sufficient condition for γ_1 to be greater than zero is $|\lambda_i| < 1$, where λ_i are the characteristic roots of the transition matrix M .

Furthermore, from (2.5.3), if $\gamma_1 > 0$, the condition for the identifiability of b_t becomes

$$(2.6.1) \quad \text{rank } (x) = n$$

where

$$x = \begin{bmatrix} x_1 M^{-t+1} & -x_1 M^{-1} (M^{-t+1} z_t + \dots + z_2) \\ \vdots & \vdots \\ x_{t-1} M^{-1} & -x_{t-1} M^{-1} z_t \\ x_t & 0 \end{bmatrix}$$

We prove now the correspondence between this condition and the observability of the system, thus connecting the econometric approach to the engineering literature.

The linear system (2.3)(2.4) is said to be completely observable at $i=1$ if and only if there exists some finite time t such that the (nxn) matrix

$$I(t,1) = \frac{1}{\sigma^2} \sum_{i=1}^t M'(i,t) H_i' H_i M(i,t) = \frac{1}{\sigma^2} x'x$$

is positive definite.

It follows that a necessary and sufficient condition for $I(t,1)$ to be positive definite is

$$\text{rank } (x) = k + p = n$$

which implies the order condition $t \geq k+p$.

This shows the coincidence between the identification condition (2.6.1) and the observability of the system (2.3)(2.4). However, (2.6.1) is only a sufficient condition for b_t to be identified since we proved in the first part of the paper that (1.9) was necessary and sufficient for the identification of b_t . Therefore, we want to prove two results: first, (1.9) coincides with $I(T,1) > 0$, where the symbol " > 0 " indicates that $I(T,1)$ is positive definite, and, secondly, we show that $I(T,1) > 0$ is equivalent to the condition for b_t to be identified when the smoothing equations are used for estimating the state of the system (2.3)(2.4).

The first result can easily be proved by observing that $R'I(T,1)R > 0$ if and only if $I(T,1) > 0$, where R is any square non-singular matrix of order $n=k+p$. Furthermore, $\text{rank}(x) = \text{rank}(xR)$. Therefore, let us choose $R = M(T,0)$, which is non-singular if $|M| \neq 0$. Then we have:

$$R'I(T,1)R = \frac{1}{\sigma^2} \sum_{i=1}^T M(i,0)' H_i^{-1} H_{i,m}(i,0) > 0$$

where we have used $M(i,T)M(T,0) = M(i,0)$, that implies

$$(2.6.2) \quad \text{rank}[xM(T,0)] = k + p$$

By using the definition of x given above, we can write (2.6.2) as:

$$\text{rank} \begin{bmatrix} x_1 M & x_1 Z_1 \\ x_2 M^2 & x_2 (MZ_1 + Z_2) \\ x_3 M^3 & \\ \vdots & \\ x_T M^T & x_T \left(\sum_{i=0}^{T-1} M^i Z_{t-i} \right) \end{bmatrix} = k+p$$

which coincides exactly with (1.9).

The second results, i.e. the coincidence between the identifiability of b_t from the smoothing equations, the positive definiteness of $I(T,1)$ and the necessary and sufficient condition (1.9), can be proved by writing the system (2.3) (2.4) in the following way:

$$(2.6.3) \quad \begin{bmatrix} y_1 \\ \vdots \\ y_t \\ \vdots \\ y_T \end{bmatrix} = \begin{bmatrix} H_1 M(1,t) \\ \vdots \\ H_t M(t,t) \\ \vdots \\ H_T M(T,t) \end{bmatrix} b_t + \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_t \\ \vdots \\ \varepsilon_T \end{bmatrix} + \begin{bmatrix} -A & 0 \\ 0 & B \end{bmatrix} \begin{bmatrix} u_2 \\ \vdots \\ u_t \\ \vdots \\ u_T \end{bmatrix}$$

where A was previously defined and

$$B = \begin{vmatrix} H_{t+1}G & 0 & \dots & 0 \\ H_{t+2}M(t+2,t+1)G & H_{t+2}G & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ H_T M(T,t+1)G & \dots & \dots & \dots & H_T G \end{vmatrix}$$

Since the smoothed estimate is equivalent to the estimate of b_t in the regression model (2.6.3) by Aitken's generalized least squares (see Chow, 1981, page 93), the identifiability condition for the state vector b_t , when all the T observations are used and $|\lambda_i| < 1$, becomes

$$(2.6.4) \quad \text{rank}(x^S) = \text{rank} \begin{bmatrix} H_1 M(1,t) \\ \vdots \\ H_{t-1} M(t-1,t) \\ H_t \\ H_{t+1} M(t+1,t) \\ \vdots \\ H_T M(T,t) \end{bmatrix} = k + p$$

The coincidence between (2.6.4) and (1.9) can be shown by post-multiplying x^S by $M(t,0)$. We have indeed $\text{rank}(x^S) = \text{rank}(x^S M(t,0))$, so that (2.6.4) can be written as:

$$(2.6.5) \quad \text{rank} \begin{bmatrix} x_1 M & Z_1 \\ x_2 M^2 & x_2 (MZ_1 + Z_2) \\ \vdots & \vdots \\ x_T M^T & x_T \left(\sum_{i=0}^{T-1} M^i Z_{T-i} \right) \end{bmatrix} = k + p$$

which coincides exactly with (1.9). Therefore, the identification condition derived by using the standard regression methods coincides with the identification condition derived by using the Kalman filter approach when the smoothed estimate of b_t is considered. Since we proved that (2.6.5) is satisfied if and only if $I(T,1) > 0$, the complete observability on the whole sample is a necessary and sufficient condition for the identifiability of the vector b_t for all

$t = 1, 2, \dots, T$. We now show which further conditions are necessary for identifying σ^2 , Q and M , and a comparison with the conditions derived in the first part of this paper will also be provided. In this way the relationship between the Kalman filter results and the econometric results previously obtained will be further explored and another link between the two approaches will be provided.

The traditional algorithm, based on the innovation likelihood function, for estimating σ^2 , Q and M can be described by the following steps:

- (1) given b_0 , P_0 , σ_0^2 , Q_0 , and M^0 , obtain $\{b_t|t; t = 1, \dots, T\}$ by applying the Kalman filter to the system (2.3)-(2.4).
- (2) Use the sequence of one-step-ahead prediction errors (innovations)

$$m_t = y_t - H_t b_t|t-1 \quad t = 1, \dots, T$$

to define the likelihood function:

$$\log L = \text{const.} - \frac{1}{2} \sum_{t=1}^T \log(C_{0t}) - \frac{1}{2} \sum_{t=1}^T m_t^2 C_{0t}^{-1}$$

- (3) Maximize the likelihood function with respect to the unknown parameters included in σ^2 , Q and M , by numerical optimization.

Therefore, since numerical optimization methods must be used, no explicit estimator can be derived for Q and M (however it is possible to concentrate the likelihood with respect to σ^2).

Another approach is still based on the innovation sequence $\{m_t\}$ but it analyzes explicitly its correlation structure in order to provide consistent estimates of Q and M .

This second approach, derived in Mehra (1970) for constant systems,^{7/} is called innovation correlation method and can also be described in the following way. Let us consider the standard Kalman filter described by equations A.1-A.7 (see Appendix A). It is easy to prove that the innovation sequence $\{m_t = y_t - H_t b_{t|t-1}\}$ is a Gaussian white noise sequence (see Mehra, 1970). In other words, if σ^2 , Q and M are known, the optimal choice of K_t , i.e. $K_t = P_{t|t-1} H_t' (H_t P_{t|t-1} H_t' + \sigma^2)^{-1}$ makes C_{it} vanish for all $i \neq 0$, where $\{C_{it}\}$ is the autocovariance function of the innovation sequence.

However, when Q , σ^2 and M are not known, the filter cannot be optimal and the autocorrelation function of the innovation process can be used to obtain consistent and asymptotically efficient estimates of Q , σ^2 and M .

Let us therefore compute the innovation correlation function for a suboptimal filter. Using equations (A.1)-(A.7) it is:^{8/}

$$\begin{aligned}
 (2.7.1) \quad \hat{C}_{0t} &= E(m_t m_t') = E\{[H_t(b_t - b_{t|t-1}) + \varepsilon_t][H_t(b_t - b_{t|t-1}) + \varepsilon_t]'\} \\
 &= H_t E(b_t - b_{t|t-1})(b_t - b_{t|t-1})' H_t' + E(\varepsilon_t \varepsilon_t') \\
 &= H_t P_{t|t-1} H_t' + \sigma^2
 \end{aligned}$$

$$\begin{aligned}
 (2.7.2) \quad \hat{C}_{1t} &= E(m_t m_{t-1}') \\
 &= E\{[H_t F_t (b_t - b_{t|t-1}) + \varepsilon_t][H_{t-1} (b_{t-1} - b_{t-1|t-2}) + \varepsilon_{t-1}]'\} \\
 &= E\{[H_t F_t (b_{t-1} - b_{t-1|t-2}) + H_t G u_t - H_t F_t K_{t-1} m_{t-1} + \varepsilon_t] \\
 &\quad \cdot [H_{t-1} (b_{t-1} - b_{t-1|t-2}) + \varepsilon_{t-1}]'\} \\
 &= H_t F_t P_{t-1|t-2} H_{t-1}' - H_t F_t K_{t-1} \hat{C}_{0t-1}
 \end{aligned}$$

$$= H_t M(t, t-2) P_{t-2} |_{t-2} F'_{t-1} H'_{t-1} + H_t M(t, t-1) G Q G' H'_{t-1} \\ - H_t F_t K_{t-1} \hat{C}_{0t-1}$$

In a similar way, but with fairly tedious algebra, we obtain:

$$(2.7.3) \quad \hat{C}_{2t} = H_t M(t, t-3) P_{t-3} |_{t-3} F'_{t-2} H'_{t-2} + H_t M(t, t-2) G Q G' H'_{t-2} \\ - H_t M(t, t-2) K_{t-2} \hat{C}_{0t-2} - H_t F_t K_{t-1} \hat{C}_{1t-1}$$

and \hat{C}_{it} , $i = 3, 4, \dots$

Define

$$A_{it} = H_t M(t, 0) P_0 M(t-i, 0) H'_{t-i} - H_t \left[\sum_{j=1}^{t-i-1} M(t, j) S_j M(t-i, j) \right] H'_{t-i}$$

$$B_{it} = H_t M(t, t-i) K_{t-i}$$

$$S_i = P_{i|i-1} H_i C_{0i}^{-1} H_i P_{i|i-1}$$

From the structure of the state-space representation of the model we have:

$$H_t = \begin{bmatrix} x_t & 0 \end{bmatrix} \quad G = \begin{bmatrix} I_k \\ 0 \end{bmatrix}$$

which implies

$$x_t M(t, t-i) G = x_t M^i$$

$$G' H'_{t-i} = x'_{t-i}$$

Therefore, equations (2.7) can be rewritten as

$$(2.8.1) \quad \hat{C}_{0t} = H_t P_{t|t-1} H' + \sigma^2 = A_{0,t} + x_t \left[\sum_{j=0}^{t-1} M^j Q (M^j)'\right] x_t' + \sigma^2$$

$$(2.8.2) \quad \hat{C}_{1t} = A_{1,t} - B_{1,t} C_{0,t-1} + x_t M \left[\sum_{j=0}^{t-2} M^j Q (M^j)'\right] x_{t-1}'$$

$$\vdots$$

$$(2.8.t) \quad C_{t-1,t} = A_{t-1,t} - B_{t-1,t} \hat{C}_{0,t-1} - \dots - B_{1,t} \hat{C}_{t-2,t-1} + x_t M^{t-1} Q x_1'$$

where we have used the matrix difference equation

$$(2.9) \quad P_{t+1|t} = F_t P_{t|t-1} F_t' + G Q G' - F_t P_{t|t-1} H_t' C_{0t}^{-1} H_t P_{t|t-1} F_t'$$

in order to obtain

$$P_{t|t} = M(t,0) P_0 M(t,0)' + \sum_{j=1}^t M(t,j) (G Q G' - S_j) M(t,j)'$$

which, substituted into (2.7), has provided equations (2.8).

We show now that an explicit estimator of σ^2 and Q can be derived from equations (2.8). Let $t = T$, since the estimates of σ^2 and Q can be computed on the full sample, and define

$$D_T = \begin{bmatrix} \hat{C}_{T-1,T} - A_{T-1,T} + B_{T-1,T} \hat{C}_{0,T-1} + \dots + B_{1T} \hat{C}_{T-2,T-1} \\ \vdots \\ C_{1,T} - A_{1,T} + B_{1,T} C_{0,T-1} \\ C_{0,T} - A_{0,T} \end{bmatrix}$$

By an appropriate vectorization of equations (2.8) and using the definition of $M(i)$ previously provided (see section C, part I), we have:

$$D_T = \begin{vmatrix} 0 & \sigma^2 & + & & & & \\ \cdot & & & x_T^{M^{T-1}} \otimes x_1 & & & \\ \cdot & & & & & & 0 \\ \cdot & & & & & & \\ \cdot & & & & & & \\ 1 & & & 0 & & & \\ & & & & & & x_T \otimes x_T \end{vmatrix} \begin{vmatrix} M(1) \\ \cdot \\ \cdot \\ \cdot \\ M(T) \end{vmatrix} \Gamma Q^*$$

which, by obvious definitions, can be rewritten as

$$D_T = e_T \sigma^2 + W_x^a M^* \Gamma Q^* \equiv [e_T : W_x^a M^* \Gamma] \theta^* \equiv \pi_x^a \theta^*$$

Therefore, if the sample is large enough, the innovation correlation method provides the following estimator,

$$\hat{\theta}^* = [(\pi_x^a)' \pi_x^a]^{-1} (\pi_x^a)' D_T$$

If the sample is small, the sum of squares $(D_T - \pi_x^a \theta^*) (D_T - \pi_x^a \theta^*)'$ must be minimized under the constraints for Q to be a positive semidefinite matrix. When Q is diagonal, the Dantzig-Cottle algorithm can be used.

The identification condition is

$$(2.10.1) \quad \text{rank} (\pi_x^a) = \text{rank} ([e_T : W_x^a M^* \Gamma]) = 1 + k(k+1)/2$$

which coincides exactly with the identification condition (1.10.3) derived in section C of part I, by using the residuals of the OLS regression.

Notice that the innovation correlation method provides only a sufficient condition for σ^2 and Q to be identified. All the $T(T+1)/2$ independent elements of the symmetric covariance matrix of the innovation sequence $\{m_t\}$ can indeed be used to identify σ^2 and Q .

By stacking the $T(T+1)/2$ elements of the upper triangular part of $\hat{C} = E(mm')$, when $m' = [m_1 \dots m_T]$ we have:

$$(2.10.2) \quad \begin{bmatrix} D_T \\ \vdots \\ D_1 \end{bmatrix} = \begin{bmatrix} e_T \\ \vdots \\ e_1 \end{bmatrix} \sigma^2 + \begin{bmatrix} L_x(T) \\ \vdots \\ L_x(1) \end{bmatrix} \Gamma Q^*$$

where

$$L_x(T) = W_x^a M^*$$

$$L_x(i) = \begin{bmatrix} (x_i^{M^{i-1}} \otimes x_1)M(1) \\ \vdots \\ (x_i^M \otimes x_{i-1})M(i-1) \\ (x_i \otimes x_i)M(i) \end{bmatrix} \quad i = 1, \dots, T-1$$

$$D_{T-i} = \begin{bmatrix} C_{T-i-1, T-i} - A_{T-i-1, T-i} + B_{T-i-1, T-i} \hat{C}_{i, i+1} + \dots + B_{i+1, T-1} \hat{C}_{t-i+2, T-i} \\ \vdots \\ \hat{C}_{1, T-i} - A_{1, T-i} + B_{1, T-i} \hat{C}_{0, T-i-1} \\ \hat{C}_{0, T-i} - A_{0, T-i} \end{bmatrix}$$

$$i = 0, 2, \dots, T-1$$

and where the difference equation (2.9) has been used again to solve backwards the covariance matrix $P_{t+1|t}$. From (2.10.2) we have, by obvious definitions,

$$(2.10.3) \quad D = e\sigma^2 + L_x \Gamma Q^* \equiv [e : L_x \Gamma] \theta^* \equiv L^* \theta^*$$

Therefore, the necessary and sufficient condition for σ^2 and Q to be identified is:

$$(2.10.4) \quad \text{rank } (L^*) = \text{rank } [e : L_x \Gamma] = 1 + k(k+1)/2$$

$$T \geq 1 + k(k+1)/2$$

which can be proved to coincide exactly with the identification condition (1.10.1) derived in section C of part 1 by appropriately changing the rows of the matrix L^* .

This result generalizes the conclusions contained in Mehra (1970) where it was shown that when the transition matrix of the system (2.3)-(2.4) is time invariant, then at most k elements of the covariance matrix Q can be identified. In contrast, when the time-varying system (2.3)-(2.4) is considered, if (2.10.3) is satisfied, then all the unknown elements of Q can be identified. Furthermore, not only the correlation structure of the innovation but all the elements of the covariance matrix of the innovation vector m can be used to estimate σ^2 and Q , so that the small sample efficiency can be increased.

Finally, let us consider the case in which the heteroscedasticity of the innovations is used to identify σ^2 and Q . The T diagonal elements of C can be written as:

$$\hat{C}_{0t} = \sigma^2 + A_{0t} + x_t Q x_t' + x_t M \left[\sum_{j=1}^{t-1} M(t-1, j) Q M(t-1, j)' \right] M' x_t'$$

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

where A_{0t} was previously defined. Therefore, by stacking the T variances \hat{C}_{0t} ; $t = 1, \dots, T$ and by appropriately vectorizing the right-hand side of

each equation, we obtain:

$$\begin{array}{c} \left| \begin{array}{c} \hat{C}_{01} - A_{01} \\ \vdots \\ \hat{C}_{0T} - A_{0T} \end{array} \right| = i\sigma^2 + \left| \begin{array}{ccc} x_1 \otimes x_1 & & \\ & \ddots & \\ & & 0 \\ & & & \ddots & \\ & & 0 & & \\ & & & & & \ddots & \\ & & & & & & x_T \otimes x_T \end{array} \right| \left| \begin{array}{c} M(1) \\ \vdots \\ M(T) \end{array} \right| \text{vec}(Q) \end{array}$$

Therefore, it is possible to write

$$\begin{aligned} D_0^{-1} &= i\sigma^2 + W_x^h M^* \Gamma Q^* \\ &\equiv [i : W_x^h M^* \Gamma] \theta^* \\ &\equiv \pi_x^h \theta^* \end{aligned}$$

where the definitions are obvious. The identification condition becomes

$$\text{rank}(\pi_x^h) = \text{rank}([i : W_x^h M^* \Gamma]) = 1 + k(k+1)/2$$

which coincides exactly with the identification condition (1.10.2).

This completes the proof of the coincidence between the identification conditions derived by using the regression approach described in part I and the identification conditions derived by using the Kalman filter approach described in this part of the paper.

Therefore, two goals have been achieved: first, we have shown that the identification conditions for the covariance matrices σ^2 and Q to be identifiable, obtained by using the innovation correlation method coincide with those obtained by using the covariance matrix of the OLS residuals; second-

dly, we have provided explicit estimators for all the unknown elements contained in σ^2 and Q .

Furthermore, another important result can be derived from equation (2.10.2). Define

$$P_t^* = \begin{bmatrix} H_t M(t,0) P_0 M(1,0)' H_1' \\ \vdots \\ H_t M(t,0) P_0 M(t,0)' H_t' \end{bmatrix}$$

and
$$D_t^* = D_t + P_t^* \quad t = 1, \dots, T$$

Therefore, (2.10.2) can be rewritten as

$$\begin{bmatrix} D_T^* \\ \vdots \\ D_1^* \end{bmatrix} = e\sigma^2 + L^* \Gamma Q^* + \begin{bmatrix} L_0(T) \\ \vdots \\ L_0(1) \end{bmatrix} \Gamma_0 P_0^*$$

where

$$L_0(t) = \text{vec}(P_t^*) = \begin{bmatrix} H_1 M(1,0) \otimes H_t M(t,0) \\ \vdots \\ H_t M(t,0) \otimes H_t M(t,0) \end{bmatrix} \text{vec}(P_0)$$

$$t = 1, \dots, T$$

and Γ_0 is a known selection matrix which maps the $(n^2 \times 1)$ vector $\text{vec}(P_0)$ into the $(n(n+1)/2 \times 1)$ vector P_0^* .

Therefore, by obvious definitions, (2.10.4) can be written as

$$(2.10.5) \quad D^* = \begin{bmatrix} e & : & L_x \Gamma & : & L_0 \Gamma_0 \\ \sigma^2 \\ Q^* \\ P_0^* \end{bmatrix} \equiv \begin{matrix} L_0^* \theta_0^* \\ 0 \\ 0 \end{matrix}$$

This equation can be used to estimate σ^2 , Q and P_0 if the identification condition

$$\text{rank}(L_0^*) = 1 + k(k+1)/2 + n(n+1)/2 \leq T$$

is satisfied.

This result provides an algorithm which can be used to revise the initial condition P_0 . Given a starting covariance matrix \hat{P}_0 (which may be obtained by performing a regression on the first $k+p$ observations (see Chow, 1981), the Kalman filter can be applied and the innovation sequence $\{m_t\}$ can be used to derive the vector D^* . Then, from (2.10.5) a revised estimate of P_0 (and of σ^2 and Q) can be obtained, and it can be used to iterate the Kalman filter on the same sample.

A simpler equation than (2.10.5) can be obtained if only the T diagonal elements of the covariance matrix C are used to estimate σ^2 , Q and P_0 .

A final estimation equation must be derived in order to provide an estimator to be used for revising the prior transition matrix M^0 . The first iteration of the Kalman filter which generates the innovation sequence $\{m_t\}$ depends indeed on the starting value M^0 which has been chosen for the transition matrix M . Therefore, the covariance matrix of the innovation vector m must also be used to derive an explicit estimator

for M . However, equation (2.10.2) is a very complex nonlinear function of M and it cannot be solved explicitly with respect to M . Let us therefore consider the upper triangular part of the covariance matrix C without the diagonal elements and define:

$$S_i = |I_{i-1} \ 0|$$

where S_i is an $i \times (i - 1)$ matrix which selects the first $(i - 1)$ rows of the matrix post-multiplying S_i . Furthermore, define

$$R_x(i) = \begin{bmatrix} x_i M^{i-2} \otimes x_1 \hat{M}(1) \\ \vdots \\ x_i \otimes x_{i-1} \hat{M}(i-1) \end{bmatrix}$$

and

$$\hat{M}(i) = \sum_{j=0}^{i-1} M^j \hat{Q}(M^j),$$

From equation (2.10.2) we have:

$$\begin{bmatrix} S_T D_T \\ \vdots \\ S_2 D_2 \end{bmatrix} = \begin{bmatrix} S_T e_T \\ \vdots \\ S_2 e_2 \end{bmatrix} \sigma^2 + \begin{bmatrix} R_x(T) \\ \vdots \\ R_x(2) \end{bmatrix} \text{vec}(M') = \begin{bmatrix} R_x(T) \\ \vdots \\ R_x(2) \end{bmatrix} \text{vec}(M')$$

Equivalently, we can write

$$(2.10.6) \quad D^S = R_x \text{vec}(M')$$

where the definitions are obvious.

If (2.10.6) is used to estimate M , then the identification condition is

$$\text{rank } (R_x) = k^2$$

which can be shown to coincide exactly with condition (1.19.2) by appropriately re-ordering the rows of R_x .

The revised estimate of the transition matrix M is provided by

$$\text{vec } (M') = (R_x' R_x)^{-1} R_x' D^s$$

and can be used to iterate the Kalman filter on the same sample.

After having determined consistent estimates of σ^2 , Q , P_0 and M , the algorithm described at the beginning of the second part of the paper can be iterated until convergence, where the innovation sequence $\{m_t\}$ can be used to define convergence. We have seen indeed that the innovation correlation method is based on the structure of the autocovariance function of $\{m_t\}$ when the filter is suboptimal. However, when the filter is optimal the stochastic process $\{m_t\}$ is white noise. Therefore, we can say that the algorithm has converged when the null hypothesis that the innovation sequence is white can be accepted. Having obtained the identification conditions for the unknown parameters of the model and having derived the relative estimators, we must now analyze the properties of the estimates obtained combining the Kalman filter and the innovation correlation method.

However, before introducing the statistical analysis of the previous estimates, we want to emphasize that a more general hypothesis on the residuals of the model can be introduced. As in the case of estimation procedures based on the least-squares method presented in the first part of the paper, it is possible to assume

$$u_t = A_1 u_{t-1} + \dots + A_s u_{t-s} + v_t \quad v_t \sim \text{NID}(0, Q)$$

Identification conditions for the coefficient matrices $A_1 \dots A_s$ are derived in Carraro (1984) by using again the innovation correlation approach, but will not be presented here since the algebra is particularly tedious.

B. Properties of the Estimates

The characteristics of optimality of the Kalman Filter can be used to analyse the properties of the estimates of the coefficient vector b_t , obtained by using the methodology presented in the previous section. It is well known indeed that the Kalman Filter provides unbiased and efficient estimates of the state vector b_t , when the initial conditions b_0 and P_0 and the matrices Q , M and σ^2 are supposed to be known.^{9/} Therefore, in this section, we want to analyse the properties of the Kalman Filter estimate of the state vector b_t under the assumption that σ^2 , Q and M are consistently estimated and the initial conditions b_0 and P_0 are unknown. Suppose first that σ^2 , Q and M are known: by minimizing the likelihood (2.5.1) it is possible to obtain the maximum likelihood estimate

$$\hat{b}_t = (x'S^{-1}x)^{-1}x'S^{-1}y$$

which can be shown to coincide asymptotically with the Kalman Filter estimate $b_{t|t}$ when b_0 and P_0 are unknown (see Carraro, 1985, App. B and Jazwinski, 1970, ch. 7). Therefore, since \hat{b}_t is asymptotically unbiased and efficient, $b_{t|t}$ also shares the same properties.

When σ^2 and Q are also unknown, the maximum likelihood estimate becomes $\hat{b}_t^* = (x'\hat{S}^{-1}x)^{-1}x'\hat{S}^{-1}y$ where $\hat{S} = \hat{\sigma}^2 I_t + A(I_{t-1} \otimes \hat{Q})A'$ was previously defined and $\hat{\sigma}^2$ and \hat{Q} are consistent estimates of σ^2 and Q . By replacing S with \hat{S} in Carraro (1985), App. B, it is possible to prove again the coincidence between $b_{t|t}$ and \hat{b}_t^* (asymptotically, if b_0 and P_0 are unknown). Therefore, it is sufficient to analyse the properties of \hat{b}_t^* .

Since \hat{b}_t^* is symmetrically distributed around b_t ,^{10/} Kakwani's results can be used to prove the unbiasedness of \hat{b}_t^* (see Kakwani, 1967). Further-

more, we have

$$(2.11) \quad E \left[\frac{\partial^2 L}{\partial b_t \partial \text{vec}(S)} \right] = E(v' \hat{S}^{-1} \otimes x' \hat{S}^{-1}) = 0$$

so that \hat{b}_t^* is also asymptotically efficient.

Therefore we can conclude that \hat{b}_t^* and hence $b_{t|t}$ are asymptotically unbiased and efficient when the covariance matrices σ^2 and Q are unknown.

Finally, let us consider the case where the transition matrix M is also unknown. The maximum likelihood estimator becomes $\hat{b}_t^{**} = (\hat{x}' \hat{S}^{-1} \hat{x})^{-1} \hat{x}' \hat{S}^{-1} y$ where \hat{x} is obtained by replacing M with \hat{M} into the matrices $M(i,t)$, $i = 1, \dots, t-1$, which were used to define x ; furthermore, \hat{M} is a consistent estimate of M . Again, by replacing M with \hat{M} into the proof provided by Carraro (1985), App. B, it is possible to show that \hat{b}_t^{**} and $b_{t|t}$ coincide asymptotically. However, $E[\partial^2 L / \partial b_t \partial \text{vec}(M)] \neq 0$ so that \hat{b}_t^{**} is asymptotically unbiased (Kakwani's proof can be applied again) but not asymptotically efficient.

A final result on the asymptotic properties of the Kalman Filter estimates can be derived by using the concepts of observability and controllability. Let $C(t,i)$ and $I(t,i)$ denote, respectively, the controllability and observability matrices of the system (2.3)-(2.4) where:

$$(2.12.1) \quad C(t,1) = \sum_{i=1}^t M(t,i) G Q G' M(t,i)$$

and

$$(2.12.2) \quad I(t,1) = \sum_{i=1}^t M(i,t)' H_1' H_1 M(i,t)$$

The following important result can be proved (see Jazwinski, 1970, ch. 7)

$$(2.13) \quad P_{t|t} \leq I^{-1}(t,1) + C(t,1) \quad t > 1$$

Furthermore, by using the definition of $M(i,j)$, G and H_i previously given, we have

$$(2.14) \quad C(t,1) = \text{Cov}(b_t) = \begin{vmatrix} \text{Cov}(\beta_t) & 0 \\ 0 & 0 \end{vmatrix}$$

where $\text{Cov}(\beta_t)$ is the covariance matrix of the stochastic time-varying coefficient vector β_t .

Previously, we proved that $b_{t|t}$ is an asymptotically unbiased estimate of b_t , i.e. that $E(b_{t|t} - b_t) = 0$ asymptotically. Now we want to prove that $\text{Cov}(b_{t|t})$ also converges to $\text{Cov}(b_t)$. By definition, we have $P_{t|t} = \text{Cov}(b_{t|t})$ so that (2.14) implies that we have to show

$$\lim_{t \rightarrow \infty} P_{t|t} = C(t,1)$$

By (2.13), this is equivalent to showing:

$$(2.15) \quad \lim_{t \rightarrow \infty} I^{-1}(t,1) = 0$$

In other words, if (2.15) can be proved, not only do we have

$$(2.16.1) \quad \lim_{t \rightarrow \infty} E(b_{t|t}) = E(b_t)$$

but also:

$$(2.16.2) \quad \lim_{t \rightarrow \infty} \text{Cov}(b_{t|t}) = \text{Cov}(b_t)$$

where, in fact, (2.16.1) holds for all $t \geq 1$. Furthermore, since $b'_{t|t} = [\beta'_{t|t} : \delta'_{t|t}]$ and $\delta_t = \delta_{t-1} = \delta$ is non-stochastic, we have, from equations (2.16):

$$\lim_{t \rightarrow \infty} E(\delta_t | t) = \delta \quad \lim_{t \rightarrow \infty} \text{Cov}(\delta_t | t) = 0$$

so that the consistency of the Kalman Filter estimate of the parameter vector δ is also proved by proving (2.15). Therefore, the following theorem has to be proved:

Theorem: If the system (2.3)-(2.4) is completely observable at any $t < \infty$, then the inverse of the observability matrix goes to zero as t goes to infinity.

Proof: The complete observability of the system implies (see Jazwinski, 1970):

$$(2.17) \quad I(t, t-N) \geq \alpha I > 0 \quad t > N$$

Furthermore, from Theorem 2 in Pagan(1980),

$$I^{-1}(t, 1) < \lambda_{mt} I$$

where λ_{mt} is the maximum eigenvalue of $I^{-1}(t, 1)$. It is also known that

$$\lambda_{mt} = \frac{1}{\gamma_{1t}}$$

where γ_{1t} is the minimum eigenvalue of $I(t, 1)$.

Since there exists a positive integer N^* such that

$$I(t, 1) = I(N^*, 1) + I(2N^*, N^* + 1) + I(3N^*, 2N^* + 1) + \dots + I(t, hN^* + 1) \quad \underline{12/}$$

then, assuming $t = N^*$, $2N^*$... and $N = N^* - 1$, it is easy to conclude from (2.17) that the complete observability of the system implies that

$I(t, 1)$ is the sum of symmetric positive definite matrices (the order condi-

tion is $N > k + p$) whose eigenvalues are all positive. Then γ_{lt} is greater or equal to a sum of positive numbers.

Hence, defining the minimum eigenvalue of $I(iN^*, (i-1)N^* + 1)$ by γ_{ilt} , we have

$$\lim_{t \rightarrow \infty} \gamma_{lt} \geq \lim_{h \rightarrow \infty} \sum_{i=1}^h \gamma_{ilt} \geq \lim_{h \rightarrow \infty} h \cdot \gamma_{il}^* = \infty$$

where $\gamma_{il}^* = \min_i \{\gamma_{ilt}\} > 0$.

Consequently, λ_{mt} converges to zero as t goes to infinity. Therefore:

$$0 \leq \lim_{t \rightarrow \infty} I^{-1}(t, 1) \leq \lim_{t \rightarrow \infty} \lambda_{mt} I = 0$$

This proves (2.15) and the theorem.

CONCLUSION.

Time varying parameter models can be estimated by using either the regression method or the Kalman filter-innovation correlation approach. This paper has compared the two methods and shown that they imply the same identification conditions for the parameters of the model. Furthermore, explicit estimators for all the unknown coefficients of the model have been derived and their properties have been studied.

However, the comparison between regression and Kalman filter methods for estimating time varying models should be further extended to the analysis of the actual performance of each method. Therefore, numerical simulations should be performed in order to test the sensitivity of the two methods to misspecified initial conditions and to the sample dimension. Furthermore, the speed of convergence and the precision of the two methods should also be examined.

Some preliminary results are contained in Carraro-Sartore(1984) and a full evaluation of the properties and characteristics of the regression and Kalman filter methods will be presented in a following paper.

Footnotes

¹See, for example, Belsley (1973), Cooley-Prescott (1973), Cooley-Wall (1975), Hatanaka (1978), Harvey-Phillips (1978, 1981), Pagan (1980), Rosemberg (1973), Sant (1977), Sarris (1973), Wall (1976), Brannas (1980), Swamy-Tinsley (1980).

²The inequality constrained estimator (see Liew, 1976) is the solution of the following problem:

$$\min (\Omega_{ah} - \Pi_x^{ah}\theta^*)'(\Omega_{ah} - \Pi_x^{ah}\theta^*)$$

subject to

$$S\theta^* \geq 0$$

where S is a known selection matrix, and can be obtained by applying the Dantzig-Cottle algorithm (see Dantzig-Cottle, 1967). It can be shown that, under the usual rank condition on the matrix Π_x^{ah} , the inequality constrained estimator is consistent and coincides with $\hat{\theta}^*$ when the sample is large (see Liew, 1976). However, when Q is not diagonal, some additional constraints must be added to $S\theta^* \geq 0$, where S is now a selection matrix which picks the diagonal elements of Q (see Carraro, 1985, part I), and the simple Dantzig-Cottle algorithm cannot be used. Therefore, a numerical optimization algorithm must be used in order to minimize the sum of squares $(\Omega_{ah} - \Pi_x^{ah}\theta^*)'(\Omega_{ah} - \Pi_x^{ah}\theta^*)$ subject to the constraints for σ^2 and Q to be positive definite.

³In Swamy-Tinsley (1980) and Swamy-Mehta (1977) the estimates of the parameters of the model obtained by using the interactive algorithm previously described are shown to be consistent and asymptotically normal. See Swamy-Tinsley (1980, pp. 124-125) for a discussion of this problem.

⁴Although the Kalman filter has appeared many other places in the literature, we give a brief description of this procedure in Appendix A.

⁵Since the method is based on autocovariance function C_{it} , it should be called innovation covariance method. However, the name correlation method has prevailed in the literature, and we will continue to use it to identify the method described above.

⁶It can be proved (see Jazwinski, 1970) that the effect of the initial conditions vanishes asymptotically if the system (2.3)-(2.4) is uniformly and completely observable and controllable

⁷Godbole (1974) extends Mehra's results to the case where the state and measurement noises have unknown means and are correlated to each other. However, he does not consider time-varying systems.

⁸We use $\{\hat{C}_{it}\}$ to indicate that the autocorrelation function is computed using t observations and, since σ^2 , Q , M are unknown, it differs from the true autocorrelation function $\{C_{it}\}$.

⁹See Wonham (1968) for the analysis of the efficiency of the Kalman Filter in the continuous case and Caines-Mayne (1971) for the extension to the discrete filter.

¹⁰This may be seen observing that the innovation sequence changes sign when $y - xb_t$ is replaced by $-(y - xb_t)$, but $\hat{C}_{0t}, \hat{C}_{1t}, \dots, \hat{C}_{qt}$ used to estimate Q and M remain unchanged for any t .

¹¹In this proof the initial conditions b_0 and P_0 are assumed known. When they are unknown $b_{t|t}$ and \hat{b}_t coincide asymptotically.

¹² h is the greatest integer less than t/N^* .

Appendix A

The Kalman filter is given by the following equations:

$$(A.1) \quad b_{t|t-1} = F_t b_{t-1|t-1}$$

$$(A.2) \quad P_{t|t-1} = F_t P_{t-1|t-1} F_t' + G Q G'$$

$$(A.3) \quad m_t = y_t - H_t b_{t|t-1}$$

$$(A.4) \quad C_{0t} = H_t P_{t|t-1} H_t' + \sigma^2$$

$$(A.5) \quad K_t = P_{t|t-1} H_t' (H_t P_{t|t-1} H_t' + \sigma^2)^{-1}$$

$$(A.6) \quad b_{t|t} = b_{t|t-1} + K_t (y_t - H_t b_{t|t-1})$$

$$(A.7) \quad P_{t|t} = (I - K_t H_t) P_{t|t-1}$$

Equations (A.1), (A.2) represent respectively the one-step-ahead predictor (based on the observations I_{t-1}^0) and its covariance matrix; m_t is the one-period prediction error for y_t and is called "innovation"; C_{0t} is the covariance matrix of the innovation at time t and K_t is called the gain of the Kalman filter. The last two equations give the revised estimates (based on I_t^0) of the coefficient vectors b_t and the relative covariance matrix.

References

- Abrahamse, A. P. J. and Koerts, J. (1969), "A Comparison between the Power of the Durbin-Watson Test and the Power of the BLUS Test," Journal of the American Statistical Association, 64, 938-952.
- Anderson, B. D. and Moore, J. B. (1979), Optimal Filtering, Prentice-Hall, Englewood Cliffs, New Jersey.
- Belsley, D. (1973), "On the Applicability of the Kalman Filter in the Determination of Systematic Parameter Variation," Annals of Economic and Social Measurement,
- Brännas, K. (1980), "On the Estimation of Econometric Systems in the Presence of Time-Varying Parameters," Ph.D. dissertation, University of Umea.
- Brogan, W. L. (1974), Modern Control Theory, Quantum Publications, New York.
- Caines, P. E. and Mayne, D. Q. (1971), "On the Discrete Time Matrix Riccati Equation of Optimal Control," International Journal of Control, 785-794.
- Carraro, C. (1984), "Identification and Estimation of Regression Models with Structural Variability of the Parameters," in T. Basar and L. F. Pau, eds., Dynamic Modelling and Control of National Economies, 1983, Pergamon Press, Oxford.
- Carraro, C. (1985), "New Methods for Macroeconomic Policy Analysis", Ph.D. Dissertation, Princeton University, Ch.1.
- Carraro, C. and Sartore, D. (1984), "Iterative Kalman Filter: Theory and Applications to Regression Models," Nota di Lavoro 8408, University of Venice.
- Chow, G. C. (1976), "A Note on the Derivation of Theil's BLUS Residuals," Econometrica, 44, 609-610.
- Chow, G. C. (1981), Econometric Analysis by Control Methods, John Wiley & Sons, Inc., New York.
- Chow, G. C. (1983), Econometrics, McGraw-Hill, New York.
- Cooley, T. F. and Prescott, E. C. (1973), "Varying-Parameter Regression: Theory and Some Applications," Annals of Economic and Social Measurement, , 463-473.
- Cooley, T. F. and Wall, K. D. (1975), "On the Identification of Time-Varying Structures," NBER Working Paper No. 85.
- Dantzig, G. B. and Cottle, R. W. (1967), "Positive (Semi)Definite Matrices and Mathematical Programming," in J. Abadie, ed., Nonlinear Programming, North-Holland, Amsterdam.

- Dhrymes, P. J. (1978), Mathematics for Econometrics, Springer-Verlag, New York.
- Duncan, D. B. and Horn, S. D. (1972), "Linear Dynamic Recursive Estimation from the View Point of Regression Analysis," Journal of the American Statistical Association, 815-821.
- Gabrielsen, A. (1978), "Consistency and Identifiability", Journal of Econometrics, 8, pp.261-263.
- Godbole, S. S. (1974), "Kalman Filtering with No A Priori Information About Noise--White Noise Case: Identification of Covariances," IEEE Transactions on Automatic Control, 561-563.
- Harvey, A. C. and Phillips, G. D. A. (1978), "The Estimation of Regression Models with Time-Varying Parameters," unpublished manuscript, University of Kent.
- Harvey, A. C. and Phillips, G. D. A. (1981), "The Estimation of Regression Models with Time-Varying Parameters," in Proceedings of the Symposium in Honour of O. Morgenstern.
- Hatanaka, M. (1978), "On the Subset of Constant and Variable Parameters of Regression Models in the Application of the Kalman Filter," Discussion Paper No. 99, Osaka University.
- Kailath, T. (1980), Linear Systems, Prentice-Hall, Englewood Cliffs, New Jersey.
- Kakwani, N. (1967), "The Unbiasedness of Zellner's Seemingly Unrelated Regression Equations Estimators," Journal of the American Statistical Association, 141-143.
- Kalman, R. E. (1960), "A New Approach to Linear Filtering and Prediction Problems," Journal of Basic Engineering.
- Kaminski, P. G., Bryson, A. E., and Schmidt, S. F. (1971), "Discrete Square Root Filtering: A Survey of Current Techniques," IEEE Transactions on Automatic Control, AC-16, 6, 727-735.
- Jazwinski, A. H. (1970), Stochastic Processes and Filtering Theory, Academic Press, New York.
- Jenkins, G. M. and Watts, D. G. (1968), Spectral Analysis and its Applications, Holden-Day Publishers, San Francisco.
- LaMotte, L. R. and McWhorther, A. Jr. (1978), "An Exact Test for the Presence of Random Walk Coefficients in a Linear Regression Model," Journal of the American Statistical Association, 73, 816-820.
- Liew, K. C. (1976), "Inequality Constrained Least Squares Estimation," Journal of the American Statistical Association, 71, 746-751.
- Magnus, J. R. (1978), "Maximum Likelihood Estimation of the GLS Model with Unknown Parameters in the Disturbance Covariance Matrix," Journal of Econometrics, 7, 281-312.
- Mehra, R. K. (1970), "On the Identification of Variances and Adaptive Kalman Filtering," IEEE Transactions on Automatic Control.

- Mehra, R. K. (1974), "Identification in Control and Econometrics: Similarities and Differences," Annals of Economic and Social Measurement, 3, 21-47.
- Pagan, A. (1984), "Two Stage and Related Estimators and their Applications," Australian National University, unpublished.
- Pagan, A. (1980), "Some Identification and Estimation Results for Regression Models with Stochastically Varying Coefficients," Journal of Econometrics, 341-363.
- Pollock, D. S. G. (1979), The Algebra of Econometrics, John Wiley & Sons, Inc., New York.
- Rosemberg, B. M. (1973), "The Analysis of a Cross-Section of Time Series by Stochastically Convergent Parameter Regression," Annals of Economic and Social Measurement, 399-450.
- Rothemberg, T. J. (1971), "Identification in Parametric Models," Econometrica, 577-591.
- Sant, D. T. (1977), "Generalized Least Squares Applied to Time Varying Parameter Models," Annals of Economic and Social Measurement, 301-311.
- Sarris, A. H. (1973), "A Bayesian Approach to Estimation of Time-Varying Regression Coefficients," Annals of Economic and Social Measurement, 501-523.
- Swamy, P. A. V. B. and Tinsley, P. A. (1980), "Linear Prediction and Estimation Methods for Regression Models with Stationary Stochastic Coefficients," Journal of Econometrics, 103-142.
- Swamy, P. A. and Mehta, J. S. (1977), "Estimation of Linear Models with Time and Cross-Sectionally Varying Coefficients," Journal of the American Statistical Association, 72, 890-898.
- Todini, E. (1978), "Mutually Interactive State/Parameter (MISP) Estimation," in Proceedings of the Conference on Application of Kalman Filtering to Hydrology, Hydraulics and Water Resources, Pittsburgh, 355-368.
- Theil, H. (1971), Principles of Econometrics, John Wiley & Sons, Inc., New York.
- Wall, K. D. (1976), "Time Varying Models in Econometrics: Identifiability and Estimation," IRIA Symposium on New Trends in System Theory, Paris.
- Wonham, W. M. (1968), "On the Matrix Riccati Equation of Optimal Control," SIAM Journal of Control, 681-698.