MISSING OBSERVATIONS IN STOCHASTIC DIFFERENCE
EQUATION WITH ARMA ERRORS*

Pedro L. Valls Pereira**

ABSTRACT

This paper considers the estimation of stochastic difference equation with missing observations.

Time series data is available at different levels of aggregations. The most common is the annual-quarterly model, where for the first part of the time series data, some of the endogenous and/or exogenous variables are available on an annual basis and the second part on a quarterly basis.

A method of estimation, based on the Kalman Filter, enables us to obtain the exact maximum likelihood estimates of the model.

* An early version of this paper was presented at the V World Congress of the Econometric Society, Boston, and at the 5th Latin American Regional Meeting of the Econometric Society, Colombia. I would like to thank, without implicating, Bob Kohn, Andrew Harvey and Luiz Hotta for many helpful comments. This paper was partially supported through FAPESP grant nº 84/2043-4.

** Instituto de Pesquisas, INPES/IPEA.

KEYWORDS

Missing observations, Temporal Aggregation, Stochastic Difference Equation, ARMA Errors, Kalman Filter.

RESUMO

Este artigo apresenta um método de estimação de equações de diferença estocástica com observações faltando.

Dados de séries temporais podem estar disponíveis em diferentes níveis de agregação. O caso mais comum é o de modelos com dados anuais e trimestrais. Por exemplo, pode-se ter um modelo onde parte do período amostral, algumas das variáveis exógenas e/ou endógenas, são observadas anualmente e trimestralmente.

Um método de estimação que é baseado no Filtro de Kalman, permite obter estimadores de máxima verossimilhança exatos para estes modelos.

1. INTRODUCTION

This paper considers the estimation of stochastic difference equation with missing observations.

This problem is addressed to models where the time series data is available at different levels of aggregation. The economist is faced with situations where part of the series is available on an annual basis and another part of the series is available on a quarterly basis. If the quarterly data is large enough, the economist may estimate the model using only this part of the data set. But in general this is not possible, and if we discard the annual part of the series, information will be lost.

This paper presents a method of estimation, based on the Kalman Filter, which enables us to obtain the exact likelihood of the model.

The structure of the paper is as follows. In Section 2 the discussion is concentrated on missing observation on an endogenous variable, i.e., it is assumed that all exogenous variables are available for all time period but the endogenous variable is missing for part of the sample. It is assumed that the missing observations always occur at the beginning of the sample. Section 3, generalize the previous results in order to handle missing observations on exogenous variables but observing the endogenous variable for all time period. In order to solve this problem we make the exogenous variables endogenous by introducing an auxiliary equation which relates these variables to other exogenous variables. The resultant system is in triangular form and it is possible to solve this problem using similar arguments to those in Section 2.

Section 4 extends the results to missing observations on endogenous and exogenous variables.

Section 5 presents some computational aspects of the Kalman Filter recursions. Concluding remarks are made in Section 6.

2. MISSING OBSERVATIONS ON ENDOGENOUS VARIABLES

2.1 Introduction

Let the stochastic difference equation with ARMA errors be given by

$$\gamma(L)y_t = \beta'(L)\tilde{x}_t + u_t \quad t = 1, \dots, T \quad (2.1.1)$$

$$\phi(L)u_t = \theta(L)\varepsilon_t \quad (2.1.2)$$

where $\gamma(L) = (1 - \gamma_1 L - \dots - \gamma_r L^r)$.

$$\beta'(L) = (\beta_1(L), \dots, \beta_k(L)) \quad \text{and}$$

$$\beta_i(L) = (\beta_{0i} + \beta_{1i}L + \dots + \beta_{s_i i}L^{s_i}) \quad i = 1, \dots, k$$

$$\tilde{x}_t' = (x_{1t}, \dots, x_{kt})$$

$$\phi(L) = (1 - \phi_1 L - \dots - \phi_p L^p)$$

$$\theta(L) = (1 + \theta_1 L + \dots + \theta_q L^q)$$

ε_t is a white noise sequence with zero mean and variance σ^2 .

Model (2.1.1-2) encompasses most of the models used in the literature. For instance, if $\phi(L) = \theta(L) = 1$ - AD model (see Hendry & Richard (1983)), if $\phi(L) = 1$ - ARMAX model (see Hannan (1975)) etc.

It is assumed that the roots of the polynomials $\phi(L)$ and $\theta(L)$ lie outside the unit circle, but it is allowed for the roots of $\gamma(L)$ to be on the unit circle, i.e., $\gamma(L)$ could be factorized as $\gamma(L) = \Delta^d \Delta_s^p \gamma^*(L)$, where $r = r^* + d + sD$ where r^* is the order of the polynomial $\gamma^*(L)$.

The series x_t represent the exogenous variables which are available for the whole sample size and y_t , the endogenous variable, which for $t = 1, \dots, p_0 T$ ($0 < p_0 < 1$) is only observed at every m time period.

Two different patterns of missing observations will be considered. If y_t is a stock variable, for $t = 1, \dots, p_0 T$, y_t is observed at every m time period and for $t \not\equiv 0 \pmod{m}$ we have missing observations. If y_t is a flow variable, for $t = 1, \dots, p_0 T$, instead

$m-1$

of observing y_t , we observe $y_\tau^* = \sum_{j=0}^{m-1} y_{t-j}$, $\tau = mt$. In this case

the observed series y_τ^* is subject to temporal aggregation and for $\tau = 1, \dots, [\frac{p_0 T}{m}]$, system (2.1.1-2) will be called aggregated model, and m the period of aggregation.

Keeping in mind the distinction between stock and flow, we will call the model aggregated when $t = 1, \dots, p_0 T$ and disaggregated model when $t = p_0 T + 1, \dots, T$.

The estimation of system (2.1.1-2), when observations are available for all time period, has been studied by many authors (a good survey can be found in Pagan & Byron (1979)). When there are missing observations on the endogenous variable, most of the studies treated in the literature are special cases of system (2.1.1-2). For instance:

$$(i) \quad \gamma(L) = 1, \beta(L) = (\beta_{01}, \dots, \beta_{0k}), \phi(L) = 1, \theta(L) = 1 \quad -$$

Static regression models with uncorrelated errors - Affifi & Elashoff (1966, 1967).

(ii) $\gamma(L) = 1$, $\beta(L) = (\beta_{01}, \dots, \beta_{0k})$, $\phi(L) = 1 - \phi_1 L$, $\theta(L) = 1$ - Static regression models with AR(1) errors - Dagenais (1973).

(iii) $\gamma(L)$ unrestricted, $\beta(L) = (\beta_{01}, \dots, \beta_{0k})$, $\phi(L) = 1 - \phi_1 L$, $\theta(L) = 1$ - Dynamic regression models with AR(1) errors - Sargan & Drettakis (1974), Gilbert (1977), Harvey & Pereira (1980), Kmenta (1981).

(iv) $\gamma(L)$ unrestricted, $\beta(L)$ unrestricted, $\phi(L) = 1$ and $\theta(L)$ unrestricted - ARMAX models - Hannan (1971, 1976), Harvey & Pereira (1985), Palm & Nijman (1982-b).

When there are missing observations in the exogenous variables, the papers by Hsiao (1979), Kmenta (1981), Gourieroux and Monfort (1981), Palm and Nijman (1982-a), and Conniffe (1983) only consider special cases of (2.1.1-2). As far as we know, there is no solution in the literature for the estimation problem when $\gamma(L)$, $\beta(L)$, $\phi(L)$ and $\theta(L)$ are unrestricted, i.e., stochastic linear difference equation with ARMA errors and missing observations on endogenous and/or exogenous variables. When all observations are available, maximum likelihood estimation is given in Reinsel (1979-a).

To solve this problem we will use the Kalman Filter to obtain the log likelihood function of system (2.1.1-2). It is shown in Pereira (1983) that under the restrictions given by (iii) above, this likelihood is equivalent to Sargan and Drettakis (1974).

2.2 State Space Formulation

Define the state space vector $\alpha_t^+ = [y_t, \dots, y_{t-\Delta}; \alpha_t^*]'$ where α_t^* is the state space vector for the error u_t which follows an ARMA (p,q) process, $\Delta = \max\{r, m-1\}$ and m is the period of aggregation.

Then the state space representation for model (2.1.1-2) is given by equations (2.2.1) and (2.2.2) below. This representation generalize Harvey and Pierse (1984) and Harvey and McKenzie (1984), and it is preferable for two reasons. Harvey and McKenzie treat the

initial y's as fixed and, as pointed out in Harvey (1981 p. 271), this approach is similar to Cochrane-Orcutt transformation and has an adverse effect on small samples. This is avoided here since the 'first observations' are treated as random. The second reason is the introduction of lags in the exogenous variable implying the loss of some observations of this variable, which is avoided in the representation given below.

The transition equation is that given by

$$\begin{aligned}
 \alpha_t^+ = & \left[\begin{array}{c|c|c|c|c|c} \hline \gamma_1 \dots & \gamma_\Delta & \phi_1 & 1 & 0 & \dots & 0 \\ \hline & 0 & & & & & \\ & \vdots & & & & & \\ & \vdots & & & & & \\ & \vdots & & & & & \\ & 0 & & & & & \\ \hline & & \phi_1 & & & & \\ & & \vdots & & & & \\ & & \vdots & & & & \\ & & \vdots & & & & \\ & & \phi_s & & & & 0 \dots 0 \\ \hline \end{array} \right] \alpha_{t-1}^+ + \left[\begin{array}{c|c} \hline \beta(L)X_t & 1 \\ \hline 0 & 0 \\ \vdots & \vdots \\ \vdots & \vdots \\ \vdots & \vdots \\ 0 & 0 \\ \hline \end{array} \right] \epsilon_t \quad (2.2.1)
 \end{aligned}$$

t = Δ+1, ..., T

$$\alpha_t^+ = \left[\begin{array}{c|c} \hline \phi_1 & 1 & 0 & \dots & 0 \\ \hline T_Y & & & & \\ & & 0 & & \\ \hline 0 & & & T_U & \\ \hline \end{array} \right] \alpha_{t-1}^+ + x_t + \left[\begin{array}{c|c} \hline \Gamma & 1 \\ \hline 0 & 0 \\ \vdots & \vdots \\ \vdots & \vdots \\ 0 & 0 \\ \hline \end{array} \right] \epsilon_t$$

and the measurement equation is given by (2.2.2-a) if y_t is a stock variable and by (2.2.2-b) if y_t is a flow variable

$$y_t = [1 : 0'_{\Delta-1} ; 0'_s] \alpha_t^+ \quad t \equiv 0 \pmod{m} \quad \Delta + 1 \leq t \leq p_0 T, \\ \text{and } t \geq p_0 T + 1 \quad (2.2.2-a)$$

$$y_t = \begin{cases} [1'_m, 0'_{\Delta-m} ; 0'_s] \alpha_t^+ & \forall t \equiv 0 \pmod{m}, \Delta + 1 \leq t \leq p_0 T \\ [1 ; 0'_{\Delta-1} ; 0'_s] \alpha_t^+ & \forall t \geq p_0 T + 1 \end{cases} \quad (2.2.2-b)$$

When $\phi(L) = 1$, system (2.1.1-2) is called an ARMAX model (see Hannan (1975)) and the state space vector has dimension Δ instead of $\Delta + s$; the transition equation becomes

$$\alpha_t^+ = \begin{bmatrix} \gamma_1 & I_{\Delta-1} \\ \vdots & \\ \vdots & \\ \vdots & \\ \hline \gamma_{\Delta} & 0 \end{bmatrix} \alpha_{t-1}^+ + \begin{bmatrix} \beta(L)x_t \\ 0 \\ \vdots \\ \vdots \\ 0 \end{bmatrix} \begin{bmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{\Delta-1} \end{bmatrix} \epsilon_t \quad t = \Delta + 1, \dots, T \quad (2.2.3)$$

and the measurement equation is given by (2.2.4-a) if y_t is a stock variable and (2.2.4-b) if y_t is a flow variable.

$$y_t = [1, 0'_{\Delta-1}] \alpha_t^+ \quad t \equiv 0 \pmod{m}, \Delta < t \leq p_0 T \\ \text{and } t \geq p_0 T + 1 \quad (2.2.4-a)$$

$$y_t = \begin{cases} [1'_m ; 0'_{\Delta-m}] \alpha_t^+ & \forall t \equiv 0 \pmod{m}, \Delta < t \leq p_0 T \\ [1 ; 0'_{\Delta-1}] \alpha_t^+ & \forall t \geq p_0 T + 1 \end{cases} \quad (2.2.4-b)$$

2.3 Starting Values and Missing Observations as Parameters

When $y_1, \dots, y_{\Delta}$ are observed and taken to be fixed in repeated sample, the likelihood function is obtained from the prediction error decomposition by starting the Kalman Filter off with

$$a_{\Delta+1}^+ = [y_{\Delta+1}, \dots, y_1; 0_s'] \quad \text{and} \quad p_{\Delta+1}^+ = \begin{bmatrix} 0 & \cdot & 0 \\ \cdot & \cdot & \cdot \\ 0 & \cdot & P_{u,0} \end{bmatrix}$$

where $P_{u,0}$ is the solution of the matrix equation

$$(I - T_u \otimes T_u) \text{vec } P_{u,0} = (R_u \otimes R_u) \text{vec } Q_{\epsilon} \quad (2.3.1)$$

which is the usual starting value for an ARMA (p,q) process.

When $y_1, \dots, y_{\Delta}$ are not observed, three possible approaches are available:

- (i) make guesses for the initial missing observations $y_1, \dots, y_{\Delta+1}$;
- (ii) treat the missing observations $y_1, \dots, y_{\Delta+1}$ as parameter (Harvey and McKenzie (1984));
- (iii) treat all missing observations $y_1, \dots, y_{p_0 T}$ as parameter (Drettakis (1974)).

The first approach yields an approximation to the likelihood function where the 'estimates' of $y_1, \dots, y_{\Delta+1}$ are treated as though they were the true value. Hence, defining $a_{\Delta+1}^+ = [y_{\Delta+1}^+, \dots, y_1^+; 0_s']$

and $P_{\Delta+1}^+ = \begin{vmatrix} 0 & 0 \\ 0 & P_{u,o} \end{vmatrix}$ where $P_{u,o}$ is given by (2.3.1), the Kal

man Filter can be started off.

The computational considerations about the solution of the equation (2.3.1) are given in Section 5.

If the second approach is adopted, the best way to proceed is to concentrate the additional parameters, $y_1, \dots, y_{\Delta+1}$, out of the likelihood function by replacing them by their MMSE's. The full likelihood function is then computed by treating these estimates as true values. This approach is a three-step procedure.

In the first step, the Kalman Filter is started with $a_{\Delta+1}^+ = 0$ and $P_{\Delta+1}^+ = \begin{vmatrix} kI & 0 \\ 0 & P_{u,o} \end{vmatrix}$ where k is a finite large number and

$P_{u,o}$ is the same as in the first approach. With these initial conditions, system (2.1.1-2) is estimated yielding estimates of $\gamma(L)$, $\beta(L)$, $\phi(L)$ and $\theta(L)$.

The second step computes the MMSE's of the initial missing observations by a fixed-point smoothing algorithm, yielding $\bar{y}_1, \dots, \bar{y}_{\Delta+1}$.

The third step re-estimate model (2.1.1-2) with the starting values $a_{\Delta+1}^+ = [\bar{y}_{\Delta+1}, \dots, \bar{y}_1; 0_s]$ and $P_{\Delta+1}^+$ as in the first approach. An alternative to this three-step procedure is to use the algorithm proposed by Rosenberg (1973). An application can be found in Harvey & McKenzie (1984).

The third approach evaluates the likelihood function conditional on the parameters $\gamma(L)$, $\beta(L)$, $\phi(L)$ and $\theta(L)$ in two steps. In the first step the missing observations in the joint density function of $y_1, \dots, y_T$ are treated as though they were parameters, and the likelihood is maximised with respect to these parameters. In the second step, the missing observations are replaced by

their estimates obtained in the first step and the likelihood function is evaluated by treating these estimates as through they were true observations. But in this second step it is necessary to make an adjustment in the constant and determinant terms.

The adjustment in the determinant term is carried out by the basic lemma in Sargan and Drettakis (1974, pp. 40-41). It is shown in Pereira (1983) that the Sargan and Drettakis approach is equivalent to the second approach described above. The advantage of the second approach is twofold. The computation requirements for the second approach are simpler than Sargan and Drettakis, and the adjustment in the determinant term is not necessary. More precisely the full determinant term of the joint density function of $\underline{y}$, given by

$$\prod_{t=2}^T | \Pi F_t | = | V W^* |, \text{ see equation (3.8) Harvey and Mc-$$

Kenzie (1984), is not needed in the approach proposed in this paper.

The MMSE's of the missing observations for $\Delta+1 \leq t \leq p_0 T$ are obtained as a by-product of the second step where the fixed-point smoothing algorithm can be used to obtain estimates of y_t for $t = 1, \dots, p_0 T$.

From the equivalence between the second and third approaches, the former is preferable for obvious reasons. The computation considerations about the starting values for the Kalman Filter are given in Section 5.

3. MISSING OBSERVATIONS IN EXOGENOUS VARIABLES

3.1 Introduction

Consider system (2.1.1-2) of Section 2.2, i.e.

$$\gamma(L)y_t = \beta(L)x_t + u_t \quad t = 1, \dots, T \quad (2.1.1)$$

$$\phi(L)u_t = \theta(L)\varepsilon_t \quad (2.1.2)$$

where $\gamma(L)$, $\phi(L)$, $\theta(L)$ and $\beta(L)$ as defined in Section 2.2. Now X_t , the $k \times 1$ vector of exogenous variables, is not available for all t . y_t , the endogenous variable, is now available for all $t = 1, \dots, T$.

Since some of the exogenous variables are missing, the discussion of Section 2 must be modified in order to handle this case. One approach to this problem is to make the missing exogenous variables endogenous, by assuming that X_t is related to some exogenous variable Z_t , which is available for all t , by the following relationship

$$\xi(L)X_t = Z_t\omega + \eta(L)\varepsilon_t \quad t = 1, \dots, T \quad (3.1.1)$$

where $\xi(L)$ and $\eta(L)$ are matrices in the lag operator L of order k and ε_t is independently normal distributed with mean 0 and variance Σ and $E(\varepsilon_t \varepsilon_s) = 0 \quad \forall t, s$.

Therefore system (2.1.1-2) is replaced by

$$\gamma(L)y_t = \beta(L)X_t + u_t \quad (3.1.2-a)$$

$$\xi(L)X_t = Z_t\omega + \eta(L)\varepsilon_t \quad (3.1.2-b)$$

$$\phi(L)u_t = \theta(L)\varepsilon_t \quad (3.1.2-c)$$

For the sake of presentation it is assumed $x_t = x_t$, $z_t = z_t$, $\xi(L) = \xi_x(L)\xi_{s_x}(L^S)\Delta_d\Delta_s^D$ and $\eta(L) = \eta_x(L)\eta_{s_x}(L^S)$, i.e., x_t follows a seasonal ARMAX process.

There are two possible ways of estimating system (3.1.2). The first approach is to treat the missing observations on x_t as parameters, and since system (3.1.2) is a recursive system, estimate equation (3.1.2-b) using the method described in Section 2.2. Now replacing the missing observations by their MMSE's in equation (3.1.2-a), the resulting system is a dynamic regression with ARMA errors where all observations are available.

The second approach is to treat (3.1.2) as a system, i.e., write (3.1.2-a) and (3.1.2-b) in the following form

$$\begin{bmatrix} \gamma(L) & -\beta(L) \\ 0 & \xi(L) \end{bmatrix} \begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{bmatrix} 0 \\ z_t \omega \end{bmatrix} + \begin{bmatrix} u_t \\ u_{1t} \end{bmatrix} \quad t = 1, \dots, T \quad (3.1.3)$$

which in matrix notation is

$$A(L)\underline{y} = \beta \underline{z}_t + \underline{u}_t \quad (3.1.4-a)$$

Where the error $\underline{u}_t$ follows a multivariate ARMA process given by

$$\begin{bmatrix} \phi(L) & 0 \\ 0 & I \end{bmatrix} \begin{bmatrix} u_t \\ u_{1t} \end{bmatrix} = \begin{bmatrix} \theta(L) & 0 \\ 0 & \eta(L) \end{bmatrix} \begin{bmatrix} \varepsilon_t \\ e_t \end{bmatrix} \quad (3.1.4-b)$$

Then the estimation of system (3.1.2) when there are missing observations on the exogenous variable has been transformed to system (3.1.4), which is a simultaneous equation model with ARMA errors and missing observations on some of the endogenous variables.

The advantages of the second approach are twofold:

- (i) It takes simultaneity into consideration;
- (ii) The log likelihood function can be obtained using the methods described in the previous section.

The first approach yields a likelihood for the first equation (3.1.2-a) conditional on the MMSE's of the missing exogenous variable, and this is one of the main drawbacks of this procedure. Although the system is recursive when there are no missing observations, estimating the first equation in this way will be inefficient compared with estimating the system as a whole. Palm and Nijman (1982-a. p. 337) pointed out that in order to achieve efficiency, the variance and the regression coefficients in the vector of parameters have to be estimated jointly. Even though they assume that the exogenous variables are aggregated in the whole sample, their result can be generalised to what is done here.

where T_y and T_u are given in (2.2.2-a) and

$$\begin{aligned} \underline{f}' = & [\beta \xi_1 + \beta \delta_1, \beta; 0'_s x^{-2}; \delta_1^2 \beta + \delta_2 \beta; \delta_1 \delta_2 \beta + \\ & + \delta_3 \beta; \dots; \delta_1 \delta_{\Delta_1-1} \beta + \delta_{\Delta_1} \beta, \delta_1 \delta_{\Delta_1} \beta] \end{aligned}$$

and δ_j , $j = 1, \dots, \Delta_1$, are given by the identity $\Delta_{\Delta_1}^D = 1 - \sum_{j=1}^{\Delta_1} \delta_j L^j$.

The matrix T_x is the transition matrix for the seasonal ARMAX (3.1.2-b). The measurement equation, when x_t is a stock variable, is

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{cases} \begin{bmatrix} 1; 0'_{\Delta-1}; 0'_s; 0'_s x; 0'_{\Delta-1} \end{bmatrix} \alpha_t^+ & \text{for } t \equiv 0 \pmod{m}, \\ & t \leq p_0 T \text{ and} \\ \begin{bmatrix} 0; 0'_{\Delta-1}; 0'_s; d' \end{bmatrix} \alpha_t^+ & \forall t > p_0 T \end{cases}$$

$$\text{where } d' = [1; 0'_s x^{-1}; \delta_1, \dots, \delta_{\Delta}] \quad (3.2.1-b)$$

and when x_t is a flow variable, is given by

$$\begin{bmatrix} y_t \\ x_t \end{bmatrix} = \begin{cases} \begin{bmatrix} 1, 0'_{\Delta-1}; 0'_s; 0'_s x; 0'_{\Delta-1} \end{bmatrix} \alpha_t^+ & \text{for } t \equiv 0 \pmod{m} \\ \begin{bmatrix} 0, 0'_{\Delta-1}; 0'_s; g' \end{bmatrix} \alpha_t^+ & t \leq p_0 T \\ \begin{bmatrix} 1, 0'_{\Delta-1}; 0'_s; 0'_s x; 0'_{\Delta-1} \end{bmatrix} \alpha_t^+ & \text{for } t > p_0 T \\ \begin{bmatrix} 1, 0'_{\Delta-1}; 0'_s; d' \end{bmatrix} \alpha_t^+ & \end{cases}$$

$$\text{where } g' = [1, 0'_s x^{-1}, \delta_1 + 1, \dots, \delta_m + 1, \delta_{m+1}, \dots, \delta_{\Delta_1}]$$

3.3 Starting Values

The starting values in this case are similar to the second approach of Section 2.3. It is also a three-step procedure.

The first step the Kalman Filter is started with $a_{-\Delta_1}^+ + 1/\Delta_1 =$

$$= [y_{\Delta_1}, \dots, y_{\Delta_1 - \Delta}; \tilde{s}'_s; \tilde{s}'_x; 0_{\Delta_1}] \quad \text{and}$$

$$P_{\Delta_1+1/\Delta_1}^+ = \begin{vmatrix} 0 & 0 & \cdot & 0 \\ \cdot & \cdot & \cdot & \cdot \\ 0 & \cdot & P_u & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & kI \\ 0 & 0 & \cdot & \cdot \\ \cdot & \cdot & 0 & \cdot \\ \cdot & \cdot & \cdot & 0 \end{vmatrix} \quad (3.3.1)$$

where k is a large number and I is the identity matrix of order s_x

By the prediction error decomposition the log likelihood function is obtained and the estimates of the parameter are obtained by the maximization of the likelihood.

In the second step using the estimates obtained in the first step, a fixed point smoothing algorithm is applied to obtain the MMSE's of the initial missing observations i.e., $\tilde{x}_{\Delta_1+1}, \dots, \tilde{x}_1$.

Then, in the third step, the Kalman Filter is started with

$$a_{-\Delta_1+1/\Delta_1}^+ = [y_{\Delta_1}, \dots, y_{\Delta_1 - \Delta}; \tilde{s}'_s; \tilde{\omega}_{\Delta_1+1}, \tilde{s}'_{x-1}; \tilde{x}_{\Delta_1}, \dots, \tilde{x}_1]$$

where $\tilde{\omega}_{\Delta_1+1} = \Delta^d \Delta_s^D \tilde{x}_{\Delta_1+1}$ and the variance-covariance matrix is similar to (3.3.1) except that the bottom right block is now the zero matrix.

By the results in Pereira (1983), this approach is equivalent to Sargan and Drettakis (1973).

It should be pointed out that when there are no exogenous variables in equation (3.1.2-b), i.e. x_t follows a seasonal ARIMA model, the first step must be modified. Now the starting value for the variance-covariance matrix is

$$\xi(L)x_t = \omega z_t + \eta(L)\underline{e}_t \quad (4.1.1-c)$$

where $\xi(L)$, $\eta(L)$ and $\underline{e}_t$ are defined in Section 3.3.

System (4.1.1) above is similar to system 3.1.2 in Section 3.3 except that the endogenous variable is also missing.

There are two possible approaches to estimate system (4.1.1).

The first approach consists of a three-step procedure. The first step, (4.1.1-c), is estimated using the method of Section 2.2. Also using a smooth algorithm, the MMSE's of the missing x_t 's are obtained. Replacing the missing x 's in equation (4.1.1-a), the second step estimates (4.1.1-a-b) by the method of Section 2.2. The third step consists of re-estimating (4.1.1-a-b) by replacing the initial missing observations of the endogenous variables by their MMSE's given by applying a fixed-point smoothing algorithm after the second step.

This approach yields estimates of the parameters conditional on the MMSE's of the exogenous variable. This is a drawback in this method.

The second approach, which yields the full likelihood function, is based on the estimation of (4.1.1) as a system as in Section 3.3. This will be the approach adopted here.

4.2 State Space Formulation

The state space formulation is similar to Section 3.2.

4.3 Starting Values

It is similar to what was done in Section 3.3.

Also note that in this case, since the state space is started off at $\Delta_1 + 1$, the first $\Delta_1 + 1$ observations for both endogenous (y_t) and exogenous (x_t) variables will be assumed fixed. This was

not necessary in Section 3.3 because y_t 's was available for all t . Also note that when x_t is stationary it is legitimate to start with $a_{\Delta_1+1/\Delta_1}^+ = 0$ because y_t will be also stationary. In this case only one pass in the Kalman Filter is needed.

5. COMPUTATIONAL CONSIDERATION

5.1 Introduction

This section presents the recursions for the starting values of the variance-covariance matrix for the Kalman Filter.

Section 5.2 considers the starting values for the dynamic regression with missing observations in the endogenous variables. The recursions are obtained for the second approach of Section 2.3 i.e., when the missing observations are treated as parameters.

Section 5.3 considers the starting value for the dynamic regression with missing observations in the exogenous variables; similar to what will be done in Section 5.2, it is only concerned with the starting values for the approach when the missing observations are treated as parameters.

Section 5.4 considers the starting value for the dynamic regression with missing observations in the endogenous and exogenous variables.

5.2 Missing Observations in Endogenous Variables

In this case the starting value for the variance-covariance matrix is given by

$$P_{\Delta+1}^+ = \begin{vmatrix} kI & \vdots & 0 \\ \cdots & \vdots & \cdots \\ 0 & \vdots & P_{u,0} \end{vmatrix}$$

where $P_{u,0}$ is the solution of the matrix equation (2.3.1), i.e., the usual starting value for an ARMA (p,q) process.

This equation yields the following system

$$(I - \phi_1 T_u) P_{\Delta+1} = T_u P_{\Delta+2} + R_u \quad (5.2.1-a)$$

$$P_{\Delta+k} = \phi_k T_u P_{\Delta+1} + T_u P_{\Delta+k+1} + \theta_{k-1} R_u \quad k = 2, \dots, s-1 \quad (5.2.1-b)$$

$$P_{\Delta+s} = \phi_s T_u P_{\Delta+1} + \theta_{s-1} R_u \quad (5.2.1-c)$$

where $P_{\Delta+j}$, $j = 1, \dots, s$ are the last s components of the columns of $P_{\Delta+j}^+$ i.e. $P_{\Delta+j}^+ = [\tilde{0}_{\Delta}'; P_{\Delta+j}']'$.

The solution of system (5.2.1-a-b-c) is given by

$$(I - \phi_1 T_u - \dots - \phi_s T_u^s) P_{\Delta+1} = (I + \theta_1 T_u + \dots + \theta_{s-1} T_u^{s-1}) R_u \quad (5.2.2-a)$$

$$P_{\Delta+k} = \phi_k T_u P_{\Delta+1} + T_u P_{\Delta+k+1} + \theta_{k-1} R_u \quad k = 2, \dots, s-1 \quad (5.2.2-b)$$

$$P_{\Delta+s} = \phi_s T_u P_{\Delta+1} + \theta_{s-1} R_u \quad (5.2.2-c)$$

The solution of system (5.2.2.a-c) is given by

$$(I - \phi_1 T - \dots - \phi_s T^s) P_1 = [I + \theta_1 T + \dots + \theta_{s-1} T^{s-1}] R_u \quad (5.2.3-a)$$

$$P_k = \phi_k TP_1 + TP_{k+1} + \theta_{k-1} R_u \quad k = 2, \dots, s-1 \quad (5.2.3-b)$$

$$P_s = \phi_s TP_1 + \theta_{s-1} R_u \quad (5.2.3-c)$$

from (5.2.3-a-b-c) it might appear that the storage of all s power matrices of T are needed, but as was pointed out in Harvey and Pierse (1984), all these matrices can be defined in terms of the columns of T^s . Specifically if t_j is the j 'th column of T^s and e_j is the j 'th column of the identity matrix of order s then

$$T^k = [t_{s-k+1}; \dots; t_s; e_1, \dots, e_{s-k}] \quad (5.2.4)$$

Using this result it is possible to write $(I - \phi_1 T - \dots - \phi_s T^s)$ as a matrix $V = (v_1, \dots, v_s)$ which columns are given by

$$v_k = - \sum_{i=1}^k \phi_{i-1} e_{k-i+1} - \sum_{i=k}^s \phi_i t_{s+k-i} \quad (5.2.5)$$

where $\phi_0 = -1$.

The matrix T^s can be formed recursively from its last column $t_s = (\phi_1, \dots, \phi_s)'$ using the formulae

$$t_{i,k} = \phi_i t_{1,k+1} + t_{i+1,k+1} \quad i = 1, \dots, s-1 \quad (5.2.6-a)$$

$$t_{s,k} = \phi_s t_{1,k+1} \quad (5.2.6-b)$$

Using the results of (5.2.6-a-b) it is possible to write $(I + \theta_1 T + \dots + \theta_{s-1} T^{s-1})$ as a matrix $W = (w_1; \dots; w_s)$ whose columns are given by

$$w_k = \begin{cases} \sum_{i=1}^k \theta_{i-1} e_{k-i+1} + \sum_{i=k}^{s-1} \theta_i t_{s+k-i} & k < s \\ \sum_{i=1}^s \theta_{i-1} e_{k-i+1} & k = s \end{cases} \quad (5.2.7)$$

with $\theta_0 = 1$.

By (5.2.5), (5.2.6-a-b) and (5.2.7) together with (5.2.3-a), the first column of $P_{1/0}$ is obtained. Then using (5.2.3-c) the s 'th column of $P_{1/0}$ is obtained and using (5.2.3-b) backwards, i.e., $k = s - 1, \dots, 2$, the remaining columns of $P_{1/0}$ are obtained.

5.3 Missing Observations in the Exogenous Variable

When (3.3.1) is the starting value for the variance-covariance matrix, i.e., the exogenous variable follows an ARMAX process, all the results of (5.2) can be used in this case. This is so because

$$P_{\Delta+1/\Delta_1}^* = \begin{vmatrix} 0 & \vdots & 0 & \vdots & 0 & \dots \\ \dots & \vdots & \dots & \vdots & \dots & \dots \\ 0 & \vdots & P_u & \vdots & 0 & \dots \\ \dots & \vdots & \dots & \vdots & \dots & \dots \\ 0 & \vdots & 0 & \vdots & kI & 0 \\ \dots & \vdots & \dots & \vdots & \dots & \dots \\ 0 & \vdots & 0 & \vdots & 0 & 0 \end{vmatrix} \quad (5.3.1)$$

and P_u is the solution of (2.3.1).

Now if (3.3.2) is the starting value for the variance-covariance matrix, the procedure of section (5.2) is also appropriate with some changes. First observe that

$$P_{\Delta_1+1/\Delta_1}^* = \begin{vmatrix} 0 & \vdots & 0 & \vdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & P_u & & & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & P_w & & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 \end{vmatrix} \quad (5.3.2)$$

where P_u is given by the solution of (2.3.1) and P_w by the solution of the matrix equation

$$(I - T_w \otimes T_w) \text{vec } P_w = (R_w \otimes R_w) \text{vec } Q_w \quad (5.3.3)$$

where

$$T_w = \begin{vmatrix} \phi_1 & \vdots & & & \\ \vdots & \vdots & I_{s_x-1} & & \\ \vdots & \vdots & \vdots & \vdots & \\ \phi_{s_x} & \vdots & 0 & \dots & 0 \\ \vdots & \vdots & & & \end{vmatrix} \quad (5.3.4-a)$$

$$R_w = \begin{vmatrix} 1 \\ \theta_1 \\ \vdots \\ \theta_{s_x-1} \end{vmatrix} \quad (5.3.4-b)$$

$$Q_w = \sigma_x^2 \quad (5.3.4-c)$$

and ϕ_j , $j = 1, \dots, s_x$, and θ_k $k = 1, \dots, s_x - 1$ are obtained from the identities

$$\xi_x(L) \xi_{s_x}(L^s) = - \sum_{j=0}^{s_x} \phi_j L^j \quad (5.3.5-a)$$

$$\eta_x(L) \eta_{s_x}(L^s) = \sum_{j=0}^{s_x-1} \theta_j L^j \quad (5.3.5-b)$$

where $s_x = \max(p_x + s p_x, q_x + s q_x + 1)$, $\phi_0 = -1$ and $\theta_0 = 1$.

(5.3.3) can be solved using the results of the previous section.

5.4 Missing Observations in the Endogenous and Exogenous Variables

The starting value for the variance-covariance matrix in this case is given in (3.2). It is easy to see that the result of section (5.2) can be used here.

5.5 Smoothing Algorithm

The Kalman Filter described in the Appendix is a "forward filter". It operates on all the data before time t and produces the estimate at the state space vector $\alpha_{t+1/t}$. The smoothed estimate of the state space is based on all the information from 1 up to T . This estimate is denoted by $\alpha_{t/T}$ and it is obtained by the utilization of a "forward filter" together with a "backward filter" which operates on all the data after time t .

There are three types of smoothing algorithm. In fixed-interval smoothing the initial and final times 1 and T are fixed and the estimate $\alpha_{t/T}$ is sought, where t varies from 1 to T . In fixed-point smoothing, t is fixed and $\alpha_{t/T}$ is sought as T increases. In fixed-lag smoothing, $\alpha_{T-\Delta/T}$ is sought as T increases with Δ held fixed (see Anderson & Moore (1979), Chap. 7).

The purpose of the smoothing algorithm in this paper is to compute the MMSE's of the missing observations in the starting value for the state space vector and hence allowing the evaluation of the exact likelihood for nonstationary models.

As pointed out by Harvey & McKenzie (1982), the fixed-interval algorithm in the context of missing observations presents some difficulties. The main difficulty is the singularity of the covariance matrix for the smoother state space. The other inconvenience is the need to store the covariance matrices and the prediction of the state space vector for the forward filter for all t .

The alternative is the fixed-point smoothing algorithm. Assume that some of the elements of the state space vector are missing at a particular time, i.e.

$$y_{\tau} = z' \alpha_{\tau} \quad (5.5.1)$$

where z is constant for all τ . For instance, when dealing with time series model $z' = (1, 0, \dots, 0)$, since the first element of the state space vector corresponds to the missing observation at time $t = \tau$.

The smoothing recursions are given by

$$\tilde{y}_{\tau/t} = \tilde{y}_{\tau/t-1} + K'_{\tau/t-1} z'_t f_t^{-1} v_t \quad t = \tau, \dots, T \quad (5.5.2-a)$$

and

$$K_{\tau/t} = T(I - G_t z_t) K_{\tau/t-1} \quad t = \tau, \dots, T \quad (5.5.2-b)$$

where

$$G_t = P_{t/t-1} z'_t F_t^{-1} \quad (5.5.2-c)$$

the initial values are:

$$\tilde{y}_{\tau/\tau-1} = z' a_{\tau/\tau-1} \quad (5.5.3-a)$$

$$K_{\tau/\tau-1} = P_{\tau/\tau-1} z' \quad (5.5.3-b)$$

The quantities F_t , v_t , z_t and $P_{t/t-1}$ for all t are obtained from the forward filter using the Kalman Filter recursion given in section (2.2).

If there is a missing observation at time t , (5.5.2-a-b) collapse to $\tilde{y}_{t/t} = \tilde{y}_{t/t-1}$ and $K_{t/t} = TK_{t/t-1}$ respectively.

The MSE of $\tilde{y}_{T/T}$ is given by $\sigma^2 F_{T/T}$, where the covariance matrix is obtained from the recursions:

$$F_{t/t} = F_{t/t-1} - K'_{t/t-1} Z_t F_{t/t-1}^{-1} Z_t' K_{t/t-1} \quad t = \tau, \dots, T \quad (5.5.4-a)$$

with initial condition

$$F_{\tau/t-1} = Z' P_{\tau/t-1} Z \quad (5.5.4-b)$$

6. CONCLUDING REMARKS

This paper considered the estimation of stochastic difference equation with ARMA errors and missing observations on a stock endogenous (or exogenous) variable or temporal aggregation of a flow endogenous (or exogenous) variable. It is shown in Pereira (1983) that this method is equivalent to Sangan & Drettakis (1974) and should be preferable because it does not require the estimation of all the missing observations. As a by-product, the MMSE of the missing observations could be obtained as a by-pass of the fixed-point smoothing algorithm.

The estimation procedure was based on the Kalman Filter recursions. With these recursions it was possible to obtain the exact likelihood for this class of models.

Also, special attention was paid to the initialization of the Kalman Filter recursions.

The method proposed in this paper is one possible solution to the problem proposed. This solution is better than others in terms of computational cost and also because of its generality.

APPENDIX

The Kalman Filter Recursion are given by

(i) prediction equations

$$\text{state space: } \hat{a}_t|_{t-1} = T\hat{a}_{t-1} + \Gamma x_t \quad (1-a)$$

$$\text{covariance matrix: } P_t|_{t-1} = TP_{t-1}T' + RQR' \quad (1-b)$$

one-step-ahead prediction error:

$$v_t = y_t - z_t' \hat{a}_t|_{t-1} \quad (2-a)$$

variance of the prediction error:

$$F_t = z_t' P_t|_{t-1} z_t + H_t \quad (2-b)$$

(ii) updating equations

state space:

$$\hat{a}_t = \hat{a}_{t|t-1} + P_{t|t-1} Z_t' F_t^{-1} v_t \quad (3-a)$$

covariance matrix:

$$P_t = P_{t|t-1} - P_{t|t-1} Z_t' F_t^{-1} Z_t P_{t|t-1} \quad (3-b)$$

Now, if prior information is available, i.e.

$$\alpha_0 \sim N(\underline{a}_0, \sigma^2 P_0) \quad (4)$$

where $\underline{a}_0$ and P_0 are known, the Kalman Filter recursions will yield the MMSE of $\hat{\alpha}_t$ based on all T observations. Maximum likelihood estimates are obtained via the prediction error decomposition, see Harvey (1981).

REFERENCES

- AFFIFI, A.A. and ELASHOFF, R.M. "Missing Values in Multivariate Statistics I - Review of the Literature". Journal of the American Statistical Association, 61, pp. 595-604, 1966.
- AFFIFI, A.A. and ELASHOFF, R.M. "Missing Observations in Multivariate Statistics II - Point Estimation in Simple Linear Regression". Journal of the American Statistical Association, 62, pp. 10-29, 1967.
- ANDERSON, B.D.O. and MOORE, J.B. *Optimal Filtering*. Oxford: Prentice Hall, 1979.
- CONNIFIE, D. "Small Sample Properties of Estimators of Regression Coefficients Given a Common Pattern of Missing Data". The Review of Economic Studies, 50, pp. 111-120, 1983.
- DAGENAIS, M.G. "The use of Incomplete Observations in Multiple Regression Analysis". Journal of Econometrics, 1, pp. 317-328, 1973.
- ENGLE, R. & WATSON, M. "One Factor Multivariate Time Series Model of Metropolitan Wage Rates". Journal of the American Statistical Association, 76, pp. 774-781, 1981.
- GILBERT, C.L. "Regression using Mixed Annual and Quarterly Data". Journal of Econometrics, 5, pp. 221-239, 1977.
- GOURIEROUX, C. & MONFORT, A. "On the Problem of Missing Data in Linear Models". The Review of Economic Studies, 48, pp. 579-586, 1981.
- HANNAN, E.J. "The Identification Problem for Multiple Equation System with Moving Average Errors". Econometrica, 39, pp. 751-764, 1971.
- HANNAN, E.J. "ARMAX and State Space Systems and Recursion Calculations". In Frontiers of Quantitative Economics, (M. J. Intriligator Ed.) Vol. III-A, pp. 321-325, Amsterdam: North-Holland, 1975.
- HANNAN, E.J. "The Identification and Parametrization of ARMAX and State Space Form". Econometrica, 41, pp. 713-723, 1976.
- HARVEY, A.C. *Time Series Models*. Oxford: Philip Allan, 1981.
- HARVEY, A.C. & PEREIRA, P.L.V. "The Estimation of Dynamic Models with Missing Observations". Revista de Econometria, 5, pp. 81-95, 1985.

- HARVEY, A.C. & MCKENZIE, C.R. "Missing Observations in Dynamic Econometric Models: a Partial Synthesis". in Time Series Analysis of Irregularly Observed Data, (E. Parzen, Ed.) pp. 108-133. Springer-Verlag. New York, 1984.
- HARVEY, A.C. & PIERSE, R.G. "Estimating Missing Observations in Economic Time Series". Journal of The American Statistical Association, 79, 125-131, 1984.
- HENDRY, D.F. & RICHARD, J.F. "The Econometric Analysis of Economic Time Series". International Statistical Review, pp. 111-163, 1983.
- HSIAO, C. "Linear Regression Using Both Temporally Aggregated and Temporally Disaggregated Data". Journal of Econometrics, 10, pp. 243-252, 1979.
- KMENTA, J. "On the Problem of Missing Measurements in the Estimation of Economic Relationships". In Proceedings of the Econometric Society European Meeting 1979. (E. G. Charatsis, Ed.) pp. 233-257, Amsterdam: North-Holland, 1981.
- PAGAN, A. "A United Approach to Estimation and Inference For Stochastically Varying Coefficient Regression Models". Core Discussion Paper Nº 7814, 1978.
- PAGAN, A. & BYRON, R.P. "A Synthesis Approach to the Estimation of Models with Autocorrelated Disturbances". In Stability and Inflation. (A. R. Bergstrom, et. al., ed.) pp. 237-256, London: John Wiley, 1979.
- PALM, F.C. & NIJMAN, T.E. "Linear Regression Using Both Temporally Aggregated and Temporally Disaggregated Data". Journal of Econometrics, 19, pp.333-343, 1982a.
- PALM, F.C. & NIJMAN, T.E. "Missing Observations in Dynamic Regression Models". Paper presented at the European Meeting of the Econometric Society, Dublin, 1982b.
- PEREIRA, P.L.V. "Estimation of Time Series and Econometric Models with Missing Observations". Ph.D. Dissertation, London School of Economics, 1983.
- REINSEL, G. "Maximum Likelihood Estimates of Stochastic Difference Equations with Autoregressive - Moving Average Errors". Econometrica, 49, pp. 129-151, 1979.
- ROSENBERG, B. "Random Coefficients Models: The Analysis of a Cross-Section of Time Series by Stochastically Convergent Parameter Regressions". Annals of Economic and Social Measurement, 2, pp. 399-428, 1973.
- SARGAN, J.D. & DRETTAKIS, E.G. "Missing Data in an Autoregressive Model". International Economic Review, 15, pp. 39-58, 1974.