EUI WORKING PAPERS IN ECONOMICS

**Estimation, Prediction and Interpolation
for Nonstationary Series with the Kalman Filter**

VÍCTOR GÓMEZ
and
AGUSTÍN MARAVALL

European University Institute, Florence

Please note
As from January 1990 the EUI Working Paper Series is divided into six sub-series, each sub-series is numbered individually (e.g. EUI Working Paper LAW No. 90/1).

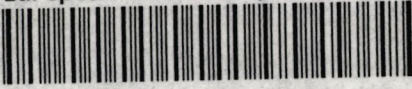EUROPEAN UNIVERSITY INSTITUTE, FLORENCE

ECONOMICS DEPARTMENT

EUI Working Paper ECO No. 92/80

**Estimation, Prediction and Interpolation
for Nonstationary Series with the Kalman Filter**

VÍCTOR GÓMEZ
and
AGUSTÍN MARAVALL

BADIA FIESOLANA, SAN DOMENICO (FI)

ESTIMATION, PREDICTION AND INTERPOLATION FOR NONSTATIONARY SERIES WITH THE KALMAN FILTER

Víctor Gómez
Instituto Nacional de Estadística
28046 Madrid
Spain
Tel:(34-1-) 5839229
Fax:(34-1-) 5839484

Agustín Maravall[1]
European University Institute
Badia Fiesolana
I - 50016 S. Domenico di
Fiesole (FI)
Italy
Tel:(39-55-) 5092347
Fax:(39-55-) 5092202

ABSTRACT

The problem of estimating any sequence of missing observations in series with a nonstationary ARIMA model representation was solved by Kohn and Ansley (1986). In their approach, the likelihood is defined first by means of a transformation of the data; then, in order to obtain an efficient estimation procedure, a modified Kalman filter and a modified fixed point smoothing algorithm are used. In this paper we show how an alternative definition of the likelihood, based on the usual assumptions made in estimation of and forecasting with ARIMA models, permits a direct and standard state space representation of the nonstationary (original) data, so that the ordinary Kalman filter and fixed point smoother can be efficiently used for estimation, forecasting and interpolation. Our approach, like that of Kohn and Ansley (1986), can handle any arbitrary pattern of missing data and we show that the same results are obtained with both approaches. In this way, the problem of estimating missing values in nonstationary series is considerably simplified.

When the available observations do not permit estimation of some of the missing values, the method indicates which are these values, and the forecasts that might be affected. Moreover, if linear combinations of the unestimable missing observations are estimable, the estimates are readily obtained. The method is illustrated using the same examples of Kohn and Ansley (1986), and an additional one for the case of unestimable missing values with estimable linear combinations thereof.

It is shown that our likelihood is equal to that of Kohn and Ansley (1986); it also coincides with that of Harvey and Pierse (1984) when applicable, and to that of Box and Jenkins (1976) when no observation is missing. The results are extended to regression models with ARIMA errors, and a computer program, written in Fortran for MSDOS computers, is available from the authors.

KEY WORDS: Time Series, Missing Observations, Nonstationarity, Arima Models, Likelihood Function.

---

[1](Please, address correspondence to the second author)

Introduction and Summary

The Kalman filter provides a well-established procedure to compute the likelihood of a time series which is the outcome of a stationary Autoregressive Moving Average (ARMA) process; see, for example, Harvey and Phillips (1979), or Pearlman (1980). Further, Jones (1980) extended the procedure for the case of missing observations in the series. However, when extending the method to a nonstationary ARIMA process, the likelihood cannot be defined in the usual sense. The main difficulty lies in the specification of the starting conditions to initialize the filter. One cannot use, as in the stationary case, the distribution of the initial state vector, since in the nonstationary case this distribution is not properly defined. Besides, when there are observations missing it is clearly not possible to use the likelihood of the differenced series (i.e., of its stationary transformation).

There have been several attempts to overcome the difficulty (see, for example, Harvey and Pierse, 1984; De Jong, 1988; Harvey and Peters, 1990; Kohn and Ansley, 1986;and, more recently, Bell and Hillmer, 1991). The path-breaking contribution of Harvey and Pierse, extending the state space methodology to ARIMA models with missing observations, presented two limitations. First, it could not be applied to series with missing values near the start or the end of the series; second, the chosen state space representation was not minimal. Possibly, the present state of the art is the powerful methodology developed by Ansley and Kohn in a sequence of papers: In order to define the likelihood, the data is transformed to eliminate dependence on the starting values. Then, in order to obtain an efficient procedure, a modified Kalman filter is used to compute the likelihood, and a modified fixed-point smoothing algorithm interpolates the missing observations. Both are generalizations of the ordinary Kalman filter and the ordinary fixed-point smoothing algorithms for handling a partially diffuse initial state vector; see Kohn and Ansley (1986).

In this paper we show how an alternative definition of the likelihood, based on the usual hypothesis made in estimation (Box and Jenkins, 1976, chapter 7) and prediction (Brockwell and Davis, 1987, pp. 304-307) of ARIMA models, permits a standard state space representation of the nonstationary series, easy to program, that does not require any transformation of the data, and provides a convenient structural interpretation of the state variable. As a consequence, the ordinary Kalman filter and the ordinary fixed-point smoothing algorithms can be efficiently used, without any modification, for estimation, forecasting, and interpolation. A notable feature of our approach is that it can be applied to any pattern of missing data because it doesn't destroy the covariance structure of the data. In this way, the problem of missing observations in nonstationary series can be considerably simplified. The results are extended to regression models with ARIMA errors.

It is seen how our likelihood coincides with that of Harvey and Pierse (1984), when the latter is applicable, and, when no observation is missing, is of course the same as that of Box and Jenkins (1976). It is shown also how the results obtained with our approach for estimation, prediction and interpolation are equal to those of Kohn and Ansley (1986).

Since most of the results for the stationary case will be valid for the nonstationary one, we begin by briefly reviewing the use of the Kalman filter for stationary series. This is done in Section 1. We proceed then to analyze a nonstationary series that follows a general ARIMA model. First, the definition of the likelihood function is considered in Section 2.1, and Section 2.2 develops the state space representation of the series, the Kalman filter, and the appropriate starting conditions. Estimation of the model is explained in Section 2.3, and Section 2.4 deals with interpolation and prediction of the series. Special attention is paid to

Digitised version produced by the EUI Library in 2020. Available Open Access on Cadmus, European University Institute Research Repository.

© The Author(s). European University Institute.

the difficulties that may arise when estimating missing observations at the beguining of the series, and the effect thereof on interpolation and prediction.

Section 2.5 analyzes in detail the relationship between our approach and that of Kohn and Ansley; Section 2.6 extends the methodology to regression models with ARIMA erros and missing obervations; finally, Section 2.8 contains a numerical application consisting of the four data sets considered by Kohn and Ansley (1986), with an additional one that illustrates a type of difficulty not discussed by them. The Appendix contains an example to illustrate the several steps in our approach; this example is, again, the same as that used by Kohn and Ansley (1986).

1 Stationary Series, ARMA Model

1.1 Prediction Error Decomposition

Let the observed series $z = (z(1), z(2), \ldots, z(N))'$ be the outcome of the ARMA model:

$$\phi(B)z(t) = \theta(B)a(t) , \tag{1.1}$$

where $\phi(B) = 1 + \phi_1 B + \ldots + \phi_p B^p$ and $\theta(B) = 1 + \theta_1 B + \ldots + \theta_q B^q$ are finite polynomials in the lag operator B, of orders p and q respectively, and $\{a(t)\}$ is a sequence of independent $N(0,\sigma^2)$ variables. The model is assumed stationary, that is, all roots of the polynomial $\phi(B)$ lie outside the unit circle. Using the prediction error decomposition, the likelihood can be written as

$$L(z(1), z(2), \ldots, z(N)) = L(z(N) \mid z(N-1), \ldots, z(1)) \ldots L(z(2) \mid z(1)) L(z(1)) , \tag{1.2}$$

where the vertical bar denotes conditional distribution. Defining, for $t = 2, \ldots, N$,

$$\hat{z}(t \mid t-1) = E(z(t) \mid z(t-1), \ldots, z(1)) , \tag{1.3a}$$

$$\sigma^2(t \mid t-1) = \frac{1}{\sigma^2} V(z(t) \mid z(t-1), \ldots, z(1)) = \frac{1}{\sigma^2} E(z(t) - \hat{z}(t \mid t-1))^2 , \tag{1.3b}$$

and using the marginal distribution of z to set the starting conditions:

$$\hat{z}(1 \mid 0) = E(z(1)) = 0 ,$$

$$\sigma^2(1 \mid 0) = \frac{1}{\sigma^2} V(z(1)) ,$$

the likelihood can be written as

$$Lkd = (2\pi\sigma^2)^{-N/2}(\sigma^2(1 \mid 0) \, \sigma^2(2 \mid 1) \, \ldots \, \sigma^2(N \mid N-1))^{-1/2} \exp\left(-1/(2\sigma^2) \sum_{t=1}^{N} \left((z(t) - \hat{z}(t \mid t-1)) / \sigma(t \mid t-1) \right)^2 \right)$$

Let $e(t)$, $t = 1, \ldots, N$, denote the sequence of standardized one-period-ahead forecast errors

$$e(t) = \frac{z(t) - \hat{z}(t \mid t-1)}{\sigma(t \mid t-1)} , \tag{1.4}$$

and define the vector $e = (e(1), \ldots, e(N))'$. Then, the log-likelihood can be expressed as

$$l = \text{const} - \frac{N}{2} \log\sigma^2 - \log(\sigma(1 \mid 0)\sigma(2 \mid 1) \, \ldots \, \sigma(N \mid N-1)) - \frac{1}{2\sigma^2} e'e .$$

Assuming known model parameters, this function will be maximized with respect to σ^2 when

$$\hat{\sigma}^2 = \frac{1}{N} e'e. \tag{1.5}$$

Therefore, σ^2 can be concentrated out of the function l, yielding the concentrated log-likelihood

$$l^* = \text{const} - \frac{N}{2} \log(e'e) - \log(\sigma(1 \mid 0)\sigma(2 \mid 1) \, \ldots \, \sigma(N \mid N-1)) = \text{const} - \frac{N}{2} \log S,$$

3

where

$$S = (\sigma(1|0)\sigma(2|1) \ \ldots \ \sigma(N|N-1))^{1/N} \, e'e \, (\sigma(1|0)\sigma(2|1) \ \ldots \ \sigma(N|N-1))^{1/N}. \qquad (1.6)$$

Exact Maximum Likelihood (ML) estimation of the model parameters minimizes, thus, the nonlinear sum of squares S, for which Marquardt's method (see, for example, Fletcher, 1987, ch. 5) provides a robust and dependable procedure. (When S is replaced by $\tilde{S} = e'e$, the method is often called unconditional least squares.)

It can be shown (Wecker and Ansley, 1983), that the vector e contains the sequence of orthonormal variables obtained from the Gram-Schmidt orthonormalization process applied to the series z. Hence, the random variables $e(t)$, $t = 1,\ldots,N$, are independent $N(0,\sigma^2)$ variables. Furthermore, if the covariance matrix of z is $\sigma^2\Omega$, and $\Omega = LL'$ is the Cholesky decomposition of Ω with L lower triangular, then it is straightforward to verify that

$$e = L^{-1}z,$$
$$|\Omega| = |L|^2,$$
$$|L| = \sigma(1|0)\sigma(2|1)\ldots\sigma(N|N-1),$$

so that S can be written more compactly as

$$S = |L|^{1/N}e'e|L|^{1/N}.$$

1.2 State Space Representation and the Kalman Filter

Among the several state space representations of ARMA models, we select that of Jones (1980), originally proposed by Akaike(1973, 1974, 1975). It provides a minimal representation, easy to program with the Kalman filter, where the state vector has a convenient structural interpretation. (It is also found in some statistical/econometrics packages, such as SAS.)

If $\{z(t)\}$ follows the ARMA process given by (1.1), letting $r = \max\{ p, q+1 \}$, and defining $\phi_i = 0$ when $i > p$, the state space representation is given by

$$x(t) = \begin{bmatrix} z(t) \\ z(t+1|t) \\ \cdot \\ z(t+r-2|t) \\ z(t+r-1|t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \cdot & \cdot & \cdot & \ldots & \cdot \\ 0 & 0 & 0 & \ldots & 1 \\ -\phi_r & -\phi_{r-1} & -\phi_{r-2} & \ldots & -\phi_1 \end{bmatrix} \begin{bmatrix} z(t-1) \\ z(t|t-1) \\ \cdot \\ z(t+r-3|t-1) \\ z(t+r-2|t-1) \end{bmatrix} + \begin{bmatrix} 1 \\ \psi_1 \\ \cdot \\ \psi_{r-2} \\ \psi_{r-1} \end{bmatrix} a(t) \ , \qquad (1.7a)$$

$$z(t) = (1,0,\ldots,0)x(t) \ , \qquad (1.7b)$$

where the ψ - weights are the coefficients of the power series expansion in B

$$\psi(B) = \sum_{i=0}^{\infty} \psi_i \, B^i = \theta(B) / \phi(B) \ .$$

4

The expression $z(t+j|t)$ is the orthogonal projection of $z(t+j)$ on the subspace generated by $\{z(s): s \le t\}$, and coincides with the conditional expectation $E(z(t+j)|z(s): s \le t)$. The state vector $x(t)$ contains, thus, the series $z(t)$ and its $(r-1)$-periods-ahead forecast function with respect to the semi-infinite sample $\{z(s): s \le t\}$. Note that when the projection is on the finite sample $\{z(s): 1 \le s \le t\}$, we represent it with a hat, as in (1.3a).

In obvious matrix notation, the system (1.7) can be written compactly as

$$x(t) = Fx(t-1) + Ga(t) ,$$ (1.8a)

$$z(t) = H'x(t) .$$ (1.8b)

Since σ^2 can be concentrated out of the likelihood, without loss of generality, we can set $\sigma^2 = 1$. (Once the parameters have been estimated, $\hat{\sigma}^2$ will be obtained with (1.5)). The Kalman filter consists of the following equations. First, the starting conditions:

$$x(0|0) = E(x(0)) = 0 ,$$
$$\Sigma(0|0) = E\left[(x(0) - x(0|0))(x(0) - x(0|0))'\right] = V(x(0)) ,$$

and, then, the recursions:

$$x(t|t-1) = F\, x(t-1|t-1) ,$$
$$\Sigma(t|t-1) = F\, \Sigma(t-1|t-1)\, F' + Q ,$$
$$K(t) = \Sigma(t|t-1)\, H \left(H'\, \Sigma(t|t-1)\, H \right)^{-1} ,$$ (1.9)
$$\Sigma(t|t) = \left(I - K(t)\, H' \right) \Sigma(t|t-1) ,$$
$$x(t|t) = x(t|t-1) + K(t) \left(z(t) - H'\, x(t|t-1) \right) , \quad t = 1,2,\ldots,N,$$

where $Q = GG'$, and

$$x(t|T) = E\left(x(t)|z(1),\ldots,z(T)\right),$$ (1.10a)

$$\Sigma(t|T) = V(x(t)|z(1),\ldots,z(T)), \quad 1 \le t \le N, 1 \le T \le N.$$ (1.10b)

The filter can also be initialized with

$$x(1|0) = E(x(1)) = 0 ,$$
$$\Sigma(1|0) = \Sigma(0|0) ,$$

since stationarity implies $\Sigma(0|0) = F\, \Sigma(0|0)\, F' + Q$. An efficient procedure to compute the initial covariance matrix $\Sigma(1|0)$ can be found in Jones (1980).

In order to maximize the likelihood, at each iteration it is necessary to obtain the values

$$\hat{z}(t|t-1) = H'\, x(t|t-1) ,$$
$$\sigma^2(t|t-1) = H'\, \Sigma(t|t-1)\, H ,$$

and the residual $e(t)$ given by (1.4). Once the last iteration $(t=N)$ has been completed, the objective function S is obtained with (1.6).

The vector $x(N|N)$ directly provides the $(r-1)$-periods-ahead forecast function of the

variable. If more forecasts are needed, they can simply be obtained with additional runs of the filter.

When $q \geq p$, an alternative representation, which yields a state space vector of minimal dimension, can be obtained by eliminating the first element in both the state vector and the G vector and by expanding the observation equation (1.8b) with

$$z(t) = H'x(t-1) + a(t).$$

The new matrix F in $x(t) = Fx(t-1) + Ga(t)$ would be the one of (1.8a) without the first row and the first column, the new matrix G would be the old one without the first element, and the new matrix H would be the old one without the last zero. In the rest of the paper we shall always refer to the representation (1.7).

1.3 ARMA Model with Missing Observations

Assume, in all generality, that only the values $\{z(t_1), z(t_2), \dots, z(t_M)\}$, $1 \leq t_1 < \dots < t_M$, are observed. To obtain the prediction error decomposition, the observation equation (1.8b) can be replaced by

$$z(t) = H'(t)x(t) + \alpha(t)W(t) , \quad t = 1,\dots,N, \tag{1.11}$$

where $H'(t) = (1,0,\dots,0)$, $\alpha(t) = 0$ if $z(t)$ is observed, and $H'(t) = (0,0,\dots,0)$, $\alpha(t) = 1$ if $z(t)$ is missing (see Brockwell and Davis 1987, p. 494). The variable $W(t)$ represents a $N(0,1)$ sample, independent of the observations $\{z(t_1),\dots,z(t_M)\}$. Thus, when $z(t)$ is missing, $x(t|t) = x(t|t-1)$, $\Sigma(t|t) = \Sigma(t|t-1)$, and both the residual and the standard error corresponding to a missing value are ignored when computing the function S of (1.6); see Jones(1980).

Having obtained parameter estimates by minimizing the appropriate function S, estimators of the missing values can be obtained through the Fixed-Point-Smoother (FPS); see Anderson and Moore (1979). Assume the j-th observation is missing, and define the starting condition

$$\Sigma^a(j|\ j-1) = \Sigma(j|\ j-1) .$$

Then, letting $k = j+1 , \dots , N$, the equations of the FPS are given by

$$\Sigma^a(k+1|k) = \Sigma^a(k|k-1) \left(F - FK(k)H'(k) \right)' ,$$
$$K^a(k) = \Sigma^a(k|k-1) H(k) \left(H'(k)\Sigma(k|k-1)H(k) + \alpha^2(k)\sigma_w^2 \right)^{-1} , \tag{1.12}$$
$$x(j|k) = x(j|k-1) + K^a(k) \left(z(k) - H'(k)x(k|k-1) \right) ,$$
$$\Sigma(j|k) = \Sigma(j|k-1) - \Sigma^a(k|k-1) H(k) \left(K^a(k) \right)' ,$$

where $H(k)$ and $\alpha(k)$ are as in (1.11), and $x(j|k)$, $\Sigma(j|k)$ and $K(k)$ are as in (1.9) and (1.10). For ARMA models, the set of equations (1.12) can be simplified by noticing that only the first element of $x(j|k)$ and the (1,1) element of $\Sigma(j|k)$, namely $\hat{z}(j|k)$ and $\sigma^2(j|k)$, respectively, are of interest. Defining

$$v'(j|k) = H'\Sigma^a(k|k-1) ,$$
$$b(k) = H'K^a(k) , \quad k \geq j ,$$

the FPS simplifies into

$$v'(j|j) = H'\Sigma^a(j|j-1) = H'\Sigma(j|j-1),$$
$$v'(j|k+1) = v'(j|k)(F - FK(k)H'(k))',$$
$$b(k) = v'(j|k) H(k) \left(H(k)' \Sigma(k|k-1) H(k) + \alpha^2(k)\sigma_w^2 \right)^{-1},$$
$$\hat{z}(j|k) = \hat{z}(j|k-1) + b(k) \left(z(k) - H(k)' x(k|k-1) \right),$$
$$\sigma^2(j|k) = \sigma^2(j|k-1) - v'(j|k) H(k) b(k), \qquad k \geq j,$$

which only requires storage of the vector $v(j|k)$ and of the scalars $\hat{z}(j|k)$ and $\sigma^2(j|k)$.

1.4 Regression Model with Missing Observations and ARMA Errors

Consider the regression model

$$z(t) = y'(t)\beta + v(t), \tag{1.13a}$$

where $\beta = \left(\beta_1, \ldots, \beta_h\right)'$ is a vector of parameters, $y'(t)$ is a vector of h independent variables, $z(t)$ is the dependent variable, and $v(t)$ is assumed to follow the ARMA model given by (1.1). If, as before, $\{z(t_1), \ldots, z(t_M)\}$ denote the observed values, defining the vectors $z = \left(z(t_1), \ldots, z(t_M)\right)'$, $v = \left(v(t_1), \ldots, v(t_M)\right)'$, and the $(M \times h)$ matrix Y with the vectors $y'(t)$, $t = t_1, \ldots, t_M$, as rows, we can write

$$z = Y\beta + v, \tag{1.13b}$$

where the matrix Y is assumed of rank h. Denoting by $\sigma^2\Omega$ the covariance matrix of v, the likelihood corresponding to (1.13b) is given by

$$Lkd = (2\pi\sigma^2)^{-M/2}|\Omega|^{-1/2} \exp\left(-\frac{1}{2\sigma^2} \left(z - Y\beta \right)' \Omega^{-1} \left(z - Y\beta \right)\right).$$

The β parameters and those of the ARMA model for $v(t)$ can be jointly estimated using the Kalman filter as described above by simply defining the state vector $x(t) = \left(v(t), v(t+1|t), \ldots, v(t+r-1|t)\right)'$ and using equations (1.8a) and

$$z(t) = y'(t)\beta + H'(t)x(t) + \alpha(t)W(t), \tag{1.14}$$

instead of (1.11). As before, the only residuals included in the computation of the likelihood are those for $t = t_1, \ldots, t_M$, given by

$$e(t) = (z(t) - y'(t)\beta - \hat{v}(t|t-1))/\sigma(t|t-1),$$

and the estimators of the missing values are obtained with the FPS. When this procedure is followed, Mean Squared Error (MSE) estimators obtained in subsequent smoothing or forecasting operations will all be conditional on β (see Harvey and Pierse, 1984).

One way to overcome this limitation, which at the same time yields a more efficient computational procedure, is to use the approach of Kohn and Ansley (1985), concentrating β and σ^2 out of the likelihood function. The BLUE estimator of β can be obtained by GLS, and then the Kalman filter and the FPS can be used to compute minimum MSE missing observations estimators and forecasts, as well as their MSE. More specifically, let $\Omega = LL'$ be

the Cholesky decomposition of Ω, with L lower triangular. Expression (1.13b) can be rewritten as

$$L^{-1}z = L^{-1}Y\beta + L^{-1}v. \tag{1.15}$$

Replacing z with v, the Kalman filter associated with equation (1.7a) and the observation equation (1.11) can be seen as an algorithm to obtain the residuals $e = L^{-1}z$ and the determinant $|L|$. The same algorithm applied to the columns of Y provides simultaneously $e = L^{-1}z$, $L^{-1}Y$ and $|L|$ (see Wecker and Ansley, 1983, and Kohn and Ansley, 1985). Therefore, applying the Kalman filter in this way, we can move from (1.13b) to (1.15).

Using the QR algorithm on the matrix $L^{-1}Y$, an orthogonal (Mxh) matrix Q is obtained, such that

$$Q'L^{-1} = (R', 0')',$$

where R is an upper triangular (hxh) matrix with nonzero elements in the main diagonal. Let Q'_1 and Q'_2 be the submatrices formed with the rows of Q' such that

$$Q'_1 L^{-1}Y = R \qquad \text{and} \qquad Q'_2 L^{-1}Y = 0.$$

Premultiplying (1.15) by Q'_1, it is found that

$$Q'_1 L^{-1}z = R\beta + Q'_1 L^{-1}v,$$

and hence the GLS estimator of β is given by

$$\hat{\beta} = R^{-1}Q'_1 L^{-1}z,$$

from which the estimator of σ^2 is also obtained:

$$\hat{\sigma}^2 = \frac{1}{M}\left(z - Y\hat{\beta}\right)'\Omega^{-1}\left(z - Y\hat{\beta}\right) = \frac{1}{M}\left(Q'L^{-1}z - Q'L^{-1}Y\hat{\beta}\right)'\left(Q'L^{-1}z - Q'L^{-1}Y\hat{\beta}\right) =$$

$$= \frac{1}{M} z'(L^{-1})' Q_2 Q'_2 L^{-1} z.$$

The function to be minimized becomes

$$S = |L|^{1/M} z'(L^{-1})' Q_2 Q'_2 L^{-1} z |L|^{1/M}.$$

For more details on how to obtain the missing observation estimators and forecasts, as well as their MSE, see Kohn and Ansley (1985).

2 Nonstationary Series, ARIMA Model

2.1 The Likelihood Function

Let $\{z(t)\}$ be a nonstationary process such that the transformation $u(t) = \delta(B)z(t)$ renders it stationary and let $\{u(t)\}$ follow the ARMA model (1.1). Then, $\{z(t)\}$ follows the nonstationary model

8

$$\phi(B)\delta(B)z(t) = \theta(B)a(t), \tag{2.1}$$

where $\delta(B) = 1+\delta_1 B+\ldots+\delta_d B^d$ denotes a polynomial in B with all roots on the unit circle. Typically, $\delta(B)$ will contain regular and/or seasonal differences. Let $z = (z(1),z(2),\ldots,z(N))'$ be the observed series. The nonstationarity of $\{ z(t) \}$ prevents us from using the prediction error decomposition (1.2), since the distribution of $z(1)$ is not well defined. New assumptions are required in order to define the likelihood properly; we shall use the following ones:

Assumption A: The variables $\{ z(1),\ldots,z(d) \}$ are independent of the variables $\{ u(t) \}$.

Assumption B: The variables $\{ z(1),\ldots,z(d) \}$ are jointly normally distributed.

The first assumption is a standard one when forecasting with ARIMA models (see Brockwell and Davis, 1987, pp. 304 - 307).

The likelihood of ARIMA models is usually defined as the likelihood of the differenced series $L(u(d+1),\ldots,u(N))$; see Box and Jenkins, chapter 7. However, an expression in terms of the original series itself could be very useful, since the Kalman filter could then be used directly to estimate missing observations in nonstationary series. Accordingly, we define as our likelihood the density

$$Lkd = L(z(d+1),\ldots,z(N) \mid z(1),\ldots,z(d)). \tag{2.2}$$

This is a well defined likelihood since, following Bell (1984, p. 650) the variable $z(t)$ can be expressed as

$$z(t) = A'(t)z_\bullet + \sum_{i=0}^{t-d-1} \xi_i\, u(t-i) , \quad t>d, \tag{2.3}$$

where $z_\bullet = (z(1),\ldots,z(d))'$, $\xi(B) = 1/\delta(B) = \sum_{i=0}^{\infty} \xi_i B^i$, $A(t) = (A_1(t),\ldots,A_d(t))'$, and the coefficients $A_i(t)$ are obtained recursively from

$$A_i(t) = -\delta_1 A_i(t-1) -\ldots -\delta_d A_i(t-d) , \quad t>d, \; i = 1,\ldots,d,$$
$$A_i(j) = \begin{cases} 1 & \text{if} \quad i = j = 1,\ldots,d \; ; \\ 0 & \text{if} \quad i,j = 1,\ldots,d; \; i \neq j \; . \end{cases} \tag{2.4}$$

Let Ξ be the lower triangular $(N\text{-}d)\times(N\text{-}d)$ matrix with rows the vectors $\left(\xi_{j-1},\xi_{j-2},\ldots,1,0,\ldots,0\right)$, $j = 1,\ldots,N-d$, and let A be the $(N\text{-}d)\times d$ matrix with rows the vectors $A'(t)$, $t = d+1,\ldots,N$. Define $J=(J_1',J_2')'$, where J_1 and J_2 are, respectively, the $d\times N$ and $(N - d)\times N$ submatrices such that $J_1 = (I_d , 0)$ and $J_2 = (A, \Xi)$. Then, the transformation

$$\left[z(1),\ldots z(d),z(d+1),\ldots z(N)\right]' = J \left[z(1),\ldots,z(d),u(d+1),\ldots,u(N)\right]' , \tag{2.5}$$

has determinant equal to one. Therefore, the densities will satisfy, under Assumptions A and B,

$$L(z(1),\ldots,z(N)) = L(z(1),\ldots,z(d),u(d+1),\ldots,u(N)) = L(z(1),\ldots,z(d))L(u(d+1),\ldots,u(N)),$$

9

from which the following result is obtained:

LEMMA 1. $L(z(d+1),\ldots,z(N)\,|\,z(1),\ldots,z(d)) = L(u(d+1),\ldots,u(N))$.

Lemma 1 establishes the equality between the classical definition of the likelihood in ARIMA models and our definition (2.2). The prediction error decomposition associated with the latter is given by

$$L(z(d+1),\ldots,z(N)\,|\,z(1),\ldots,z(d)) = L(z(N)\,|\,z(N-1),\ldots,z(1))\ldots \qquad (2.6)$$
$$L(z(d+1)\,|\,z(d),\ldots,z(1)).$$

For $t = d+2,\ldots,N$, define $\hat{z}(t|t-1)$ and $\sigma^2(t\,|\,t-1)$ as in (1.3) and let for $t = d+1$,

$$\hat{z}(d+1\,|\,d) = E(z(d+1)\,|\,z(d),\ldots,z(1)) = A'(d+1)z_{\bullet},$$
$$\sigma^2(d+1\,|\,d) = \frac{1}{\sigma^2}E\left[\left(z(d+1) - \hat{z}(d+1\,|\,d)\right)^2\right] = \frac{1}{\sigma^2}E\left(u^2(d+1)\right) = \frac{1}{\sigma^2}V(u(d+1)).$$

Then, the residuals corresponding to the prediction error decomposition (2.6) are

$$e(t) = \left(z(t) - \hat{z}(t|t-1)\right)/\sigma(t|t-1), \qquad t = d+1,\ldots,N.$$

Defining for $t = d+2,\ldots,N$,

$$\hat{u}(t|t-1) = E(u(t)\,|\,u(t-1),\ldots,u(d+1),z(d),\ldots,z(1)) = E(u(t)\,|\,u(t-1),\ldots,u(d+1)),$$

the following lemma shows the equality between the residuals corresponding to (2.6) and those obtained in the Box and Jenkins likelihood.

LEMMA 2. *Under Assumptions A and B*,

$$z(t) - \hat{z}(t|t-1) = u(t) - \hat{u}(t|t-1),$$
$$\sigma^2(t|t-1) = \frac{1}{\sigma^2}E\left[\left(u(t) - \hat{u}(t|t-1)\right)^2\right] = \frac{1}{\sigma^2}V\left(u(t)\,|\,u(t-1),\ldots,u(d+1)\right), \qquad t \geq d+1,$$

where

$$\hat{u}(d+1\,|\,d) = E\left(u(d+1)\right) = 0 \quad and \quad \sigma^2(d+1\,|\,d) = \frac{1}{\sigma^2}V\left(u(d+1)\right).$$

Proof. From (2.5) it is immediately seen that the subspace generated by $\{\,z(1),\ldots,z(d),$ $z(d+1),\ldots,z(t)\,\}$ and by $\{\,z(1),\ldots,z(d),u(d+1),\ldots,u(t)\,\}$ are the same for $t \geq d+1$. Thus,

$$\hat{z}(t|t-1) = E\left(z(t)\,|\,z(t-1),\ldots,z(d+1),z(d),\ldots,z(1)\right) = E\left(z(t)\,|\,u(t-1),\ldots u(d+1),z(d),\ldots,z(1)\right),$$

and , considering (2.3),

$$\hat{z}(t|t-1) = A'(t)z_{\bullet} + \sum_{i=0}^{t-d-1} \xi_i\, E\left(u(t-i)\,|\,u(t-1),\ldots,u(d+1),z(d),\ldots,z(1)\right) =$$
$$= A'(t)z_{\bullet} + E\left(u(t)\,|\,u(t-1),\ldots,u(d+1)\right) + \sum_{i=1}^{t-d-1} \xi_i\, u(t-i) =$$
$$= A'(t)z_{\bullet} + \hat{u}(t|t-1) + \sum_{i=1}^{t-d-1} \xi_i\, u(t-i).$$

10

Hence, $z(t) - \hat{z}(t\,|\,t-1) = u(t) - \hat{u}(t\,|\,t-1)$. ∎

2.2 State Space Representation of a Nonstationary Series and the Kalman Filter

It is proved in Bell (1984, pp. 649-651) that there is a one-to-one correspondence between $\{z(t)\}$ and $\{z_*, \{u(t)\}\}$. Using Bell's backward representation as well as (2.3), it is not difficult to check that, for any $t \geq d+1$, the subspaces generated by $\{z(s): s \leq t\}$ and $\{z_*, \{u(s): s \leq t\}\}$ coincide. Thefore, under assumptions A and B, if we define $z(s\,|\,t) = E(z(s)\,|\,z(v): v \leq t)$ and $u(s\,|\,t) = E(u(s)\,|\,z(v): v \leq t)$ for $s \geq t \geq d+1$, we have

$$z(s\,|\,t) = E\left(z(s)\,|\,z(v): v \leq t\right) = E\left(z(s)\,|\,z_*, \{u(v): v \leq t\}\right), \tag{2.7a}$$

$$u(s\,|\,t) = E\left(u(s)\,|\,z(v): v \leq t\right) = E\left(u(s)\,|\,u(v): v \leq t\right). \tag{2.7b}$$

The following lemma will allow us to preserve the same state space representation as in the stationary case. Consider the ARIMA model (2.1). Let $\phi^*(B) = \phi(B)\delta(B)$, $\psi^*(B) = \theta(B)/\phi^*(B) = \sum_{i=0}^{\infty} \psi_i^* B^i$ and $\phi_i^* = 0$ when $i > p+d$. Then ,

LEMMA 3.

$$z(t+r-1\,|\,t) = -\phi_r^* z(t-1) - \phi_{r-1}^* z(t\,|\,t-1) - \ldots - \phi_1^* z(t+r-2\,|\,t-1) + \psi_{r-1}^* a(t),$$

where $r = \max\{p+d, q+1\}$.

Proof. Using (2.7) in (2.3), it is obtained that

$$z(t+r-1) - z(t+r-1\,|\,t) = \xi_0\left(u(t+r-1) - u(t+r-1\,|\,t)\right) + \ldots + \xi_{t+r-d-2}\left(u(t+1) - u(t+1\,|\,t)\right).$$

Since $\{u(t)\}$ follows the model (1.1), we have

$$u(t+j) - u(t+j\,|\,t) = \sum_{i=0}^{j-1} \psi_i\, a(t+j-i), \quad j = 1, \ldots, r-1,$$

so that

$$z(t+r-1) - z(t+r-1\,|\,t) = \xi_0 \psi_0 a(t+r-1) + \ldots + \left(\xi_0 \psi_{r-2} + \ldots + \xi_{r-2} \psi_0\right) a(t+1) =$$
$$= \psi_0^* a(t+r-1) + \ldots + \psi_{r-2}^* a(t+1),$$

given that $\psi^*(B) = \xi(B)\psi(B)$. Therefore,

$$z(t+r-1\,|\,t) = z(t+r-1\,|\,t-1) + \psi_{r-1}^* a(t),$$

and considering that

$$z(t+r-1\,|\,t-1) = -\sum_{i=1}^{r-1} \phi_i^* z(t+i-1\,|\,t-1) - \phi_r^* z(t-1),$$

the lemma follows directly. ∎

11

As a consequence, for the nonstationary case, the state space representation is also given by (1.7), with $r = \max\{p+d, q+1\}$, and the ϕ and ψ coefficients replaced by the ϕ^* and ψ^* ones; that is

$$x(t) = \begin{bmatrix} z(t) \\ z(t+1|t) \\ . \\ . \\ z(t+r-2|t) \\ z(t+r-1|t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ . & . & . & \dots & . \\ 0 & 0 & 0 & \dots & 1 \\ -\phi_r^* & -\phi_{r-1}^* & -\phi_{r-2}^* & \dots & -\phi_1^* \end{bmatrix} \begin{bmatrix} z(t-1) \\ z(t|t-1) \\ . \\ . \\ z(t+r-3|t-1) \\ z(t+r-2|t-1) \end{bmatrix} + \begin{bmatrix} 1 \\ \psi_1^* \\ . \\ . \\ \psi_{r-2}^* \\ \psi_{r-1}^* \end{bmatrix} a(t) , \tag{2.9a}$$

$$z(t) = (1,0,\dots,0)x(t) . \tag{2.9b}$$

The Kalman filter can then be applied in an identical manner to compute the likelihood through the prediction error decomposition (2.6). The starting conditions however will be different. Considering that

$$x(d+1) = \begin{bmatrix} A'(d+1)z_* \\ A'(d+2)z_* \\ . \\ . \\ A'(d+r)z_* \end{bmatrix} + \begin{bmatrix} 1 & 0 & \dots & 0 \\ \xi_1 & 1 & \dots & 0 \\ . & . & \dots & . \\ . & . & \dots & . \\ \xi_{r-1} & \xi_{r-2} & \dots & 1 \end{bmatrix} \begin{bmatrix} u(d+1) \\ u(d+2|d+1) \\ . \\ . \\ u(d+r|d+1) \end{bmatrix} = A_* z_* + \Xi U,$$

where A_*, Ξ and U are the obvious matrices and vectors, respectively, the starting conditions for the nonstationary case are given by

$$x(d+1|d) = A_* z_* , \tag{2.10a}$$

$$\Sigma(d+1|d) = \Xi E(UU')\Xi' = \Xi \bar{\Sigma}(1|0) \Xi' , \tag{2.10b}$$

where $\bar{\Sigma}(1|0) = E(UU')$ can be computed from the stationary process $\{u(t)\}$, which follows model (1.1), as in Section 1.2. The dimension of the state space representation (2.9) is minimal, and hence smaller than that of Harvey and Pierse (1984), and equal to the one in Kohn and Ansley (1986). Our representation has the appeal of its simplicity, of the direct interpretation of the state vector (i.e., the nonstationary series and its $(r-1)$-periods-ahead forecast function), and of the easiness in computing the starting conditions. A simple example to illustrate this and the following sections is discussed in Appendix A.

2.3 Missing Observations

As in the two previous sections, let $\{z(t)\}$ follow the ARIMA process (2.1) with Assumptions A and B holding. Let the available observations be $\{z(t_1), z(t_2), \dots, z(t_M)\}$, with $1 \le t_1 < t_2 < \dots < t_M$. If there are no missing observations among the first d values of the series, one could proceed as in the stationary case, with the fixed point smoother and the observation equation (1.11), using (2.10) as starting conditions. This is the case considered by Harvey and Pierse (1984). Given that their starting conditions and equations are the same as

ours, the two likelihoods coincide.

For the general case, assume that $z_I = \left(z(t_1),...,z(t_k)\right)'$ is the vector of observations in $z_* = \left(z(1),...,z(d)\right)'$. Letting z_J denote the vector of missing observations in z_*, then the observed values in z can be expressed as

$$z(t_i) = A'(t_i)z_* = C'(t_i)z_J = z(t_i), \qquad\qquad i = 1,...,k,$$
$$z(t_i) = A'(t_i)z_* + \bar{u}(t_i) = B'(t_i)z_J + C'(t_i)z_J + \bar{u}(t_i), \quad i = k+1,...,M,$$

where $\bar{u}(t) = \sum\limits_{j=0}^{t-d-1} \xi_j u(t-j)$, $t > d$, the vectors $A'(s)$ have been defined in (2.3), (2.4) and $B'(s)$ and $C'(s)$ are the appopriate subvectors of $A'(s)$. Let z_{II} and $\bar{u}$ denote the vectors $\left(z(t_{k+1}),...,z(t_M)\right)'$ and $\left(\bar{u}(t_{k+1}),...,\bar{u}(t_M)\right)'$, respectively, and A, B and C denote the matrices with rows $A'(s)$, $s = t_1,...,t_M$, $B'(s)$ and $C'(s)$, $s = t_{k+1},...,t_M$, respectively. Define $D = (D_1', D_2')'$, where D_1 and D_2 are the $k \times d$ and $(M-k) \times d$ submatrices, respectively, such that $D_1 = (I_k, 0)$ and $D_2 = (B, C)$. Then, the above equations can be rewritten as

$$\left[z_I', z_{II}'\right]' = Az_* + \left[0', \bar{u}'\right]' = D\left[z_I', z_J'\right]' + \left[0', \bar{u}'\right]'. \tag{2.11}$$

Defining $y = z_{II} - Bz_I$, (2.11) implies

$$y = Cz_J + \bar{u}. \tag{2.12}$$

A natural way of extending our likelihood (2.2) to the case of missing observations is to consider the likelihood of the observations z_{II} conditional on z_*, and to treat z_J as additional parameters. This definition of the likelihood is equivalent to considering (2.12) as a regression model whose errors $\bar{u}$ have a known covariance matrix $\sigma^2\Delta$. We define as our likelihood that associated with (2.12), that is

$$Lkd = (2\pi\sigma^2)^{-(M-k)/2} |\Delta|^{-1/2} \exp\left(-\frac{1}{2\sigma^2}\left(y - Cz_J\right)' \Delta^{-1}\left(y - Cz_J\right)\right), \tag{2.13}$$

where the unknown parameters are σ^2, z_J and the coefficients $(\phi, \theta) = (\phi_1,...,\phi_p, \theta_1,...,\theta_q)$ of model (1.1). Using the prediction error decomposition, and concentrating σ^2 out of the likelihood, it is seen that maximizing (2.13) over (ϕ, θ) and z_J is equivalent to minimizing

$$S = |L|^{1/(M-k)} e'e |L|^{1/(M-k)} = |L|^{1/(M-k)}\left(L^{-1}y - L^{-1}Cz_J\right)'\left(L^{-1}y - L^{-1}Cz_J\right)|L|^{1/(M-k)},$$

over (ϕ, θ) and z_J, where L is the lower triangular matrix such that $\Delta = LL'$ is the Cholesky decomposition of Δ, and $e = L^{-1}\left(y - Cz_J\right) = \left(e(t_{k+1}),...,e(t_M)\right)'$ is the vector of standardized residuals. For given values of (ϕ, θ) and z_J, the function S can be computed with the Kalman filter. The equations to use are (2.9a) and (1.11), with starting conditions (2.10). In this way, we can jointly estimate the unknown parameters (ϕ, θ) and z_J. As starting values for z_J, linear combinations of adjacent observations can be used.

However, as stated in Section 1.4, a more efficient computational procedure is obtained by using the approach of Kohn and Ansley (1985), whereby σ^2 and z_J can be

13

concentrated out of the likelihood. This approach, not only will allow us to compute MSE not conditional on z_I, but will also place on a sound theoretical basis the discussion of the difficulties that may arise when estimating z_I, and their implications for prediction and interpolation. To see how this approach can be used in the present context, let the Kalman filter be applied to the model

$$y = z_{II} - B z_I = \bar{u}.$$

This allows us to compute $L^{-1} y$ and $|L|$. The equations would again be (2.9a) and (1.11); the starting conditions,

$$x(d+1|d) = \left[B'(d+1) z_I, \dots, B'(d+r) z_I \right]'$$

and (2.10b). Note that it is not necessary to compute the entire matrix B and that the Kalman filter is applied to the vector of observations z_{II}. The same algorithm applied to the columns of the matrix C, with starting conditions $x(d+1|d) = 0$ and (2.10b) also permits us to compute $L^{-1} C$. We consider thus the model

$$L^{-1} y = L^{-1} C z_I + L^{-1} \bar{u}. \tag{2.14}$$

Let the rank of C be $r_c \leq d-k$. Then, the matrix $L^{-1} C$ has also rank r_c and two cases need to be distinguished:

a) rank$(C) = r_c = d-k$. Then, the QR algorithm applied to $L^{-1} C$ yields an orthogonal $(M-k) \times (M-k)$ matrix Q such that

$$Q' L^{-1} C = (R', 0')',$$

where R is an upper triangular $(d-k) \times (d-k)$ matrix with nonzero elements in the main diagonal.

b) rank$(C) = r_c < d-k$. A slight modification of the QR algorithm yields an orthogonal $(M-k) \times (M-k)$ matrix Q such that

$$Q' L^{-1} C = (E', 0')',$$

where E is an $r_c \times (d-k)$ matrix in echelon form. Permuting the variables in z_I if necessary, which implies permuting the columns of E, we can always assume without loss of generality that the matrix E has the form

$$E = (R, S),$$

where R is an $r_c \times r_c$ upper triangular matrix with nonzero elements in the main diagonal.

Thus, in general, let the QR algorithm when $r_c = d-k$, or its slight modification when $r_c < d-k$, be applied to the matrix $L^{-1} C$. Then, an orthogonal $(M-k) \times (M-k)$ matrix Q is obtained such that

$$Q' L^{-1} C = (E', 0')',$$

where $E = R$ if $r_c = d-k$ or $E = (R, S)$ if $r_c < d-k$. Let Q'_1 and Q'_2 be the submatrices of Q' formed with the rows of Q', so that $Q'_1 L^{-1} C$ consists of the first r_c nonzero rows of $Q' L^{-1} C$, , and $Q'_2 L^{-1} C = 0$. If (2.14) is premultiplied by Q', it is obtained that

14

$$Q_1' L^{-1} y = E z_j + Q_1' L^{-1} \tilde{u}, \tag{2.15a}$$
$$Q_2' L^{-1} y = \qquad Q_2' L^{-1} \tilde{u}. \tag{2.15b}$$

The normal equations corresponding to (2.15) are

$$E' E z_j = E' Q_1' L^{-1} y.$$

This system of linear equations is always consistent and its general solution is (see Rao, pp. 222-223)

$$\hat{z}_j = (E'E)^- E' Q_1' L^{-1} y + \left(I_{d-k} - (E'E)^- E'E \right) x, \tag{2.16a}$$

where x is an arbitrary vector of dimension $d - k$ and $(E'E)^-$ is a generalized inverse of $E'E$. If $r_c = d-k$, then $E = R$, an upper triangular matrix, and (2.16a) becomes

$$\hat{z}_j = R^{-1} Q_1' L^{-1} y . \tag{2.16b}$$

Therefore, instead of maximizing the likelihood (2.13) (or, equivalently, minimizing the function S) over (ϕ, θ), z_j and σ^2, we can concentrate σ^2 and z_j out of the likelihood by using the regression (2.16) to obtain $\hat{z}_j$, and

$$\hat{\sigma}^2 = \frac{1}{M-k} \left(L^{-1} y - L^{-1} C \hat{z}_j \right)' \left(L^{-1} y - L^{-1} C \hat{z}_j \right).$$

In this case, we seek to minimize the function

$$S^\bullet = |L|^{1/(M-k)} \left(L^{-1} y - L^{-1} C \hat{z}_j \right)' \left(L^{-1} y - L^{-1} C \hat{z}_j \right) |L|^{1/(M-k)}$$

over (ϕ, θ) to obtain $(\hat{\phi}, \hat{\theta})$; it is easy to verify that

$$S^\bullet = |L|^{1/(M-k)} \left(L^{-1} y - L^{-1} C \hat{z}_j \right)' Q Q' \left(L^{-1} y - L^{-1} C \hat{z}_j \right) |L|^{1/(M-k)} =$$
$$= |L|^{1/(M-k)} \left[\left(Q_1' L^{-1} y - E \hat{z}_j \right)' \left(Q_1' L^{-1} y - E \hat{z}_j \right) + y' (L^{-1})' Q_2 Q_2' L^{-1} y \right] |L|^{1/(M-k)}.$$

Since the matrix $E(E'E)^- E'$ is symmetric, idempotent and has rank r_c (see Rao, 1973, p. 25), there exists thus, an orthogonal $r_c \times r_c$ matrix Q_E such that

$$Q_E E (E'E)^- E' Q_E' = I_{r_c},$$

and, therefore $E(E'E)^- E' = I_{r_c}$. This implies

$$Q_1' L^{-1} y - E \hat{z}_j = Q_1' L^{-1} y - E(E'E)^- E' Q_1' L^{-1} y + \left(E - E(E'E)^- E'E \right) x = 0$$

and $S^\bullet$ can be rewritten as

$$S^\bullet = |L|^{1/(M-k)} y' (L^{-1})' Q_2 Q_2' L^{-1} y |L|^{1/(M-k)}. \tag{2.17}$$

The estimator $\hat{\sigma}^2$ becomes

$$\hat{\sigma}^2 = \frac{1}{M-k} y' (L^{-1})' Q_2 Q_2' L^{-1} y. \tag{2.18}$$

In summary, to estimate model (2.1) concentrating z_j and σ^2 out of the likelihood, we apply the Kalman filter with equations (2.9a) and (1.11) and starting conditions $x(d+1|d) = \left[B'(d+1), \dots, B'(d+r) \right]' z_j$ and (2.10b) to the vector of observations z_{jj} to obtain $L^{-1} y$ and $|L|$. We apply the same algorithm with starting conditions $x(d+1|d) = 0$ and (2.10b) to the columns of the matrix C to obtain $L^{-1} C$. Then, the QR algorithm, or a slight

15

modification of it, is used to obtain $Q'L^{-1}C = (E',0')'$. Finally, we compute $Q'_2 L^{-1}y$ and S^* in (2.17). Minimizing S^* yields the estimators $(\hat{\phi}, \hat{\theta})$, then we estimate z_j and σ^2 by (2.16) and (2.18), respectively.

2.4 Interpolation and Prediction

Suppose first we concentrate σ^2 and z_j out of the likelihood. For interpolation and prediction, we need the estimator (2.16) in (2.10a) to start the Kalman filter and the FPS. When $r_c = d-k$, we saw that $\hat{z}_j$ could be obtained with (2.16b).

If $r_c < d-k$, then, as stated in the previous section, by permuting the variables in z_j if necessary, we can assume without loss of generality that $E = (R, S)$, where R is an upper triangular matrix. It is easy to check that a generalized inverse of $E'E$ is the $(d-k)\times(d-k)$ matrix $(E'E)^- = (T'_1, T'_2)'$, where T_1 and T_2 are the $r_c \times (d-k)$ and $(d-k-r_c) \times (d-k)$ submatrices, respectively, such that $T_1 = ((R'R)^{-1}, 0)$ and $T_2 = (0,0)$. If $z_j = \left(z_j^{*'}, z_j^{**'} \right)'$, where z_j^* and z_j^{**} are the subvectors of z_j of dimension r_c and $d-k-r_c$, respectively, then (2.16a) implies

$$\hat{z}_j^* = R^{-1}Q'_1 L^{-1}y - R^{-1}S\tilde{x} \quad \text{and} \quad \hat{z}_j^{**} = \tilde{x},$$

where $\tilde{x}$ is an arbitrary vector of dimension $d-k-r_c$. It is interesting to note that the set $Rz_j^* + Sz_j^{**}$ is a maximal set of independent linear combinations that can be estimated from the data y without dependence on $\tilde{x}$.

The covariance matrix $V(\hat{z}_j)$ of $\hat{z}_j$ is obtained by GLS. Let $1 \le t < t_k$, and $z(t)$ be an unobserved value that we want to estimate. Since $z(t) = A'(t)z_* = C'(t)z_j$, we have

$$\hat{z}(t) = C'(t)\hat{z}_j. \tag{2.19a}$$

The MSE of $\hat{z}(t)$ is $C'(t)V(\hat{z}_j)C(t)$. Let now

$$z(t) = A'(t)z_* + \tilde{u}(t) = B'(t)z_j + C'(t)z_j + \tilde{u}(t)$$

be an unobserved value that we want to estimate with $t > t_k$. If $t < t_M$ we are interpolating and if $t > t_M$ we are predicting. Following Kohn and Ansley (1985), we have

$$\hat{z}(t) = B'(t)z_j + C'(t)\hat{z}_j + P(t)\left(y - C\hat{z}_j \right), \tag{2.19b}$$

where $E(\tilde{u}(t)|\tilde{u}) = P(t)\tilde{u} = Cov(\tilde{u}(t), \tilde{u})V^{-1}(\tilde{u})\tilde{u}$. The MSE of $\hat{z}(t)$ is given by

$$E\left(z(t) - \hat{z}(t) \right)^2 = \left(P(t)C - C'(t) \right)V(\hat{z}_j)\left(P(t)C - C'(t) \right)' + S(t), \tag{2.20}$$

where $S(t) = E\left(\tilde{u}(t) - P(t)\tilde{u} \right)^2$. The Kalman filter with equations (2.9a) and (1.11) and starting conditions (2.10) if $t > t_M$, or the FPS if $t < t_M$, can be used to compute both $P(t)\left(y - C\hat{z}_j \right)$ and $S(t)$. The same algorithm applied to the columns of C yields $P(t)C$.

If $r_c < d-k$, it may happen that $C'(t)\hat{z}_j = C'(t)\left(\hat{z}_j^{*'}, \hat{z}_j^{**'} \right)' = C'(t)\left(\hat{z}_j^{*'}, \tilde{x}' \right)'$ in (2.19a) or (2.19b) will depend on the arbitrary vector $\tilde{x}$. The vector $C'(t)$ lies in the space generated by the rows of the matrix (R, S), which is the same as that generated by the rows

16

of the matrix C, if, and only if, there exists a vector $c'(t)$ such that $C'(t) = c'(t)(R, S)$. Therefore, the vector $C'(t)\hat{z}_j = C'(t)\left[\hat{z}_j^{*\prime}, \bar{x}'\right]' = C'(t)\left[(R^{-1}Q_1'L^{-1}y)', 0'\right]' + C'(t)\left[(-R^{-1}S)', I'_{d-k-r_c}\right]'\bar{x}$ will not depend on $\bar{x}$ if, and only if, $C'(t)$ lies in the space generated by the rows of the matrix (R, S). Note that if $r_c < d-k$, then the matrix $V(\hat{z}_j)$ should be replaced by $\hat{\sigma}^2(E'E)^-$ and that this makes sense only if $C'(t)\hat{z}_j$ does not depend on $\bar{x}$.

Suppose now that we concentrate only σ^2 out of the likelihood and that we jointly estimate z_j and (ϕ, θ). We would use the estimator $\hat{z}(t)$ thus obtained in (2.19). However, in (2.20) we would only obtain $S(t)$ and thus the MSE of $\hat{z}(t)$ would be conditional on z_j.

2.5   The Relationship Between Our Approach and that of Kohn and Ansley

The following theorem specifies the precise relationship between our likelihood and that of Kohn and Ansley (1986).

THEOREM 1. *Concentrating σ^2 and z_j out of our likelihood (2.13), the same function to maximize is obtained as when σ^2 is concentrated out of the likelihood defined by Kohn and Ansley.*

Proof. As we showed in Section 2.3, maximizing (2.13) concentrated with respect to z_j and σ^2 is the same as minimizing (2.17). Kohn and Ansley (1986) define a likelihood which coincides with that of Box and Jenkins when there are no missing observations and with that of Harvey and Pierse (1984) when there are no missing observations among the first d values of the series. In these cases, thus, their definition coincides with ours. For the general case, Kohn and Ansley define $\eta = (z(1-d), ..., z(0))'$ and consider

$$z(t_j) = F'(t_j)\eta + w(t_j), \quad j = 1, ..., M,$$

where $w(s) = \sum_{j=0}^{t-1} \xi_j u(t-j)$ and the $F'(s)$, $s \geq 1$, are generated similarly to the $A'(s)$ of (2.3) and (2.4). If $z = (z(t_1), ..., z(t_M))'$, $w = (w(t_1), ..., w(t_M))'$ and F is the matrix having as rows the vectors $F'(s)$, $s = t_1, ..., t_M$, the previous equations can be written as

$$z = \left[z_I', z_{II}'\right]' = F\eta + w. \tag{2.21}$$

In order to see the relationship between expressions (2.21) and (2.11), let F_* be the $d \times d$ matrix having as rows the vectors $F'(s)$, $s = 1, ..., d$, and let $w_* = (w(1), ..., w(d))'$. By definition, equality $z_* = F_*\eta + w_*$ holds. Substituting in (2.11), it is obtained that

$$z = \left[z_I', z_{II}'\right]' = AF_*\eta + Aw_* + \left[0', \bar{u}'\right]'. \tag{2.22a}$$

Therefore, considering (2.21), the following equalities are found

$$F = AF_*, \qquad Aw_* + \left[0', \bar{u}'\right]' = w. \tag{2.22b}$$

Further, in (2.11), matrix D can be obtained from matrix A by permuting some of its columns.

17

Thus, the following equality also holds

$$D = AT_{A'}$$ (2.23a)

where T_A is the nonsingular matrix obtained from the unit matrix by permuting the columns in the same way, and we can rewrite equation (2.22a) as

$$z = [z_I', z_{II}']' = DT_A^{-1}F_\cdot\eta + DT_A^{-1}w_\cdot + [0', \bar{u}']'.$$ (2.23b)

The transformation that permits us to move from (2.12) to (2.15) will not have unit determinant because $|L^{-1}|$, in general, is different from 1. Multiplying L^{-1} by a positive number α, such that $|\alpha L^{-1}| = 1$, it is immediately seen that when $\alpha = |L|^{1/(M-k)}$, then the matrix $\alpha Q'L^{-1}$ has unit determinant. Premultiplying both sides of (2.23b) by

$$J = \begin{bmatrix} I_k & 0 \\ -\alpha Q'L^{-1}B & \alpha Q'L^{-1} \end{bmatrix}$$

which also has unit determinant, it is obtained that

$$J\begin{bmatrix} z_I \\ z_{II} \end{bmatrix} = \begin{bmatrix} I_k & 0 \\ 0 & \alpha E \\ 0 & 0 \end{bmatrix} T_A^{-1}F_\cdot\eta + \begin{bmatrix} I_k & 0 \\ 0 & \alpha E \\ 0 & 0 \end{bmatrix} T_A^{-1}w_\cdot + \begin{bmatrix} 0 \\ \alpha Q_1'L^{-1}\bar{u} \\ \alpha Q_2'L^{-1}\bar{u} \end{bmatrix},$$ (2.24)

and premultiplying (2.21) by J, yields

$$J\begin{bmatrix} z_I \\ z_{II} \end{bmatrix} = \begin{bmatrix} I_k & 0 \\ 0 & \alpha E \\ 0 & 0 \end{bmatrix} T_A^{-1}F_\cdot\eta + J w.$$ (2.25)

Define $J_1 = (-\alpha Q_1'L^{-1}B, \alpha Q_1'L^{-1})$ and $J_2 = (-\alpha Q_2'L^{-1}B, \alpha Q_2'L^{-1})$. Equating (2.24) and (2.25), it follows that

$$J_2[z_I', z_{II}']' = \alpha Q_2'L^{-1}y = \alpha Q_2'L^{-1}\bar{u} = J_2 w.$$ (2.26)

Since J is a transformation matrix of the type defined by Kohn and Ansley (1986), their likelihood is the density of $J_2 w$, which, from (2.26), is equal to that of $\alpha Q_2'L^{-1}\bar{u}$. This latter likelihood is given by

$$Lkd = (2\pi\sigma^2)^{-(M-k-r_c)/2}|\alpha|^{-(M-k-r_c)}\exp\left(-\frac{1}{2\sigma^2}y'(L^{-1})'Q_2Q_2'L^{-1}y\right) =$$

$$= (2\pi\sigma^2)^{-(M-k-r_c)/2}|L|^{-(M-k-r_c)/(M-k)}\exp\left(-\frac{1}{2\sigma^2}y'(L^{-1})'Q_2Q_2'L^{-1}y\right).$$ (2.27)

Concentrating σ^2 out of this likelihood using

$$\hat{\sigma}^2 = \frac{1}{M-k-r_c}y'(L^{-1})'Q_2Q_2'L^{-1}y,$$ (2.28)

it is easy to check that maximizing (2.27) is equivalent to minimizing

18

$$|L|^{1/(M-k)} \, y' (L^{-1})' Q_2 Q_2' L^{-1} y \, |L|^{1/(M-k)},$$

which is identical to (2.17). □

Note that the only difference between our estimator (2.18) and that of Kohn and Ansley (2.28) lies in the denominator. If we divide by $M-k-r_C$, then, by Lemma 4 below, both estimators coincide.

LEMMA 4. *Using the same notation as in the proof of Theorem 1, the matrix $F_\bullet$ is nonsingular and*

$$\mathrm{rk}(D) = \mathrm{rk}(F) = \mathrm{rk}(A) = k + r_C.$$

Proof. It is easy to check that $\mathrm{rank}(C) = k + r_C$. As we saw in the proof of Theorem 1, $F = AF_\bullet$ and $D = AT_A$. Hence, we only have to show that $F_\bullet$ is nonsingular. The rows $F'(t)$, $t = 1,\dots,d$, of $F_\bullet$ satisfy

$$F'(t) = -\delta_1 F'(t-1) - \dots - \delta_d F'(t-d),$$

Letting $t = d$,

$$F'(d) = \delta_1 F'(d-1) + \dots + \delta_{d-1} F'(1) + \delta_d F'(0),$$

and

$$\delta_d \det(F_\bullet) = \delta_d \det[F'(1),\dots,F'(d-1),F'(0)]'.$$

Letting $t = d-1$,

$$F'(d-1) = \delta_1 F'(d-2) + \dots + \delta_{d-1} F'(0) + \delta_d F'(-1),$$

and

$$\delta_d^2 \det(F_\bullet) = \delta_d^2 \det[F'(1),\dots,F'(d-2),F'(-1),F'(0)]'.$$

Proceeding in a similar manner, it is obtained that

$$\delta_d^d \det(F_\bullet) = \delta_d^d \det[F'(1-d),\dots,F'(-1),F'(0)]' = \delta_d^d \neq 0,$$

since $|\delta_d| = 1$. □

The next two Theorems show that the estimators of unobserved values, as well as their MSE, obtained by the approach of Kohn and Ansley (1986) and those obtained by our approach coincide.

THEOREM 2. *If $z(t)$ is an unobserved value, then $z(t)$ is estimable in the sense of Kohn and Ansley (1986) if, and only if, our estimator $\hat{z}(t)$ does not depend on the arbitrary vector $\hat{x}$.*

Proof. Using the same notation as in the proof of theorem 1, let

$$z(t) = F'(t)\eta + w(t) = A'(t)z_\bullet + \bar{u}(t) = B'(t)z_I + C'(t)z_J + \bar{u}(t), \qquad t \geq 1, \qquad (2.29)$$

where $\bar{u}(t) = 0$ if $t < t_k$, be an unobserved value that we want to estimate. Kohn and Ansley define $z(t)$ as estimable if the vector $F'(t)$ lies in the space generated by the rows of the matrix F. This in turn happens if, and only if, there exists a vector $d'(t)$ such that $F'(t) = d'(t)F$.

We will now show that $F'(t)$ lies in the space generated by the rows of the matrix F if, and only if, $C'(t)$ lies in the space generated by the rows of the matrix C. By definition, $z_{\bullet} = F_{\bullet}\eta + w_{\bullet}$, and, by Lemma 4, this implies $\eta = F_{\bullet}^{-1}z_{\bullet} - F_{\bullet}^{-1}w_{\bullet}$. Substituting in (2.29), we have

$$F'(t)\eta + w(t) = F'(t)F_{\bullet}^{-1}z_{\bullet} + w(t) - F_{\bullet}^{-1}w_{\bullet} = A'(t)z_{\bullet} + \bar{u}(t),$$

and, therefore $F'(t)F_{\bullet}^{-1} = A'(t)$. Given that the vector $(B'(t), C'(t))$ is obtained from $A'(t)$ by a permutation, the following equality holds

$$F'(t)F_{\bullet}^{-1}T_A = (B'(t), C'(t)). \tag{2.30}$$

Suppose there exists a vector $d'(t)$ such that $F'(t) = d'(t)F$. By (2.22b), (2.23a) and (2.30), this implies $(B'(t), C'(t)) = d'(t)D$. If we define $d'(t) = (b'(t), c'(t))$, where $b'(t)$ and $c'(t)$ are the subvectors of $d'(t)$ of dimension $1 \times k$ and $1 \times (M-k)$, respectively, we finally obtain $C'(t) = c'(t)C$.

To prove the " if " part, assume there exists a vector $c'(t)$ such that $C'(t) = c'(t)C$, and define $b'(t) = B'(t) - c'(t)B$ and $d'(t) = (b'(t), c'(t))$. It is easy to check that $d'(t)D = (B'(t), C'(t))$ and, therefore, by (2.30), (2.22b) and (2.23a), we have $F'(t) = d'(t)F$. Since we showed in Section 2.4 that the vector $C'(t)\hat{z}_J$ does not depend on the arbitrary vector $\bar{x}$ if, and only if, $C'(t)$ lies in the space generated by the rows of C, the theorem is proved. □

THEOREM 3. *If $z(t)$ is an unobserved value estimable in the sense of Kohn and Ansley (1986), then the estimator $\bar{z}(t)$ obtained by the method of Kohn and Ansley (1986) and our estimator coincide. If the same estimator of σ^2 is employed, then the MSE are also equal.*

Proof. The estimator of Kohn and Ansley (1986) can be expressed as

$$\bar{z}(t) = d'(t)z + E\left[w(t) - d'(t)w \mid J_2w\right],$$

where w and J_2w are those of (2.21) and (2.26) and $d'(t)$ is a vector such that $F'(t) = d'(t)F$. This vector exists because $z(t)$ is estimable. As we showed in the proof of Theorem 2, $d'(t)$ also satisfies $(B'(t), C'(t)) = d'(t)D$. This, by (2.21) and (2.11), implies

$$z(t) = d'(t)z + w(t) - d'(t)w = d'(t)z + \bar{u}(t) - d'(t)\left[0', \bar{u}'\right]', \tag{2.31}$$

and therefore

$$w(t) - d'(t)w = \bar{u}(t) - d'(t)\left[0', \bar{u}'\right]'. \tag{2.32}$$

As stated previously, by permuting the variables in z_J if necessary, we can assume without loss of generality that $E = (R, S)$ in equation (2.15a), where R is un upper triangular matrix with nonzero elements in the main diagonal. If $z_J = \left(z_J^{\bullet'}, z_J^{\bullet\bullet'}\right)'$, as in Section 2.4, then from equation (2.15a) we have

$$z_J^{\bullet} = R^{-1}Q_1'L^{-1}y - R^{-1}Sz_J^{\bullet\bullet} - R^{-1}Q_1'L^{-1}\bar{u}.$$

Substituting in equation (2.29), it is obtained that

20

$$z(t) = B'(t)z_t + C'(t)\left[\left[\left(R^{-1}Q_1'L^{-1}y\right)',0'\right]' + C'(t)\left[-(R^{-1}S)',I\right]'z_j^{\bullet\bullet} + \\ + \bar{u}(t) - C'(t)\left[\left(R^{-1}Q_1'L^{-1}\bar{u}\right)',0'\right]'.$$

The third term on the right is zero because $Q'L^{-1}C = \left((R,S)',0'\right)'$ and, as we showed in the proof of Theorem 2, if $d'(t) = (b'(t), c'(t))$, then $C'(t) = c'(t)C$. Therefore, equating this last equation and (2.31), yields

$$d'(t)z = B'(t)z_t + C'(t)\left[\left(R^{-1}Q_1'L^{-1}y\right)',0'\right]', \tag{2.33a}$$

$$\bar{u}(t) - d'(t)\left[0',\bar{u}'\right]' = \bar{u}(t) - C'(t)\left[\left(R^{-1}Q_1'L^{-1}\bar{u}\right)',0'\right]'. \tag{2.33b}$$

As we showed in Section 2.4, our estimator of $z(t)$ is $\hat{z}(t)$ in (2.19), where $C'(t)\hat{z}_j = C'(t)$ $[(R^{-1}Q_1'L^{-1}y)',0']'$ does not depend on the arbitrary vector $\hat{x}$ (because $C'(t) = c'(t)C$) and $E(\bar{u}(t)|\bar{u}) = P(t)\bar{u} = Cov(\bar{u}(t),\bar{u})V^{-1}(\bar{u})\bar{u}$. Note that if $1 \le t < t_k$, then $\bar{u}(t) = 0$ and $P(t) = 0$. Now, (2.19), (2.32) and (2.33) imply

$$\hat{z}(t) = d'(t)z + P(t)\left(y - C\hat{z}_j\right)$$

and

$$\bar{z}(t) = d'(t)z + E\left[\bar{u}(t) - C'(t)[(R^{-1}Q_1'L^{-1}\bar{u})',0']'\,|\,J_2w\right].$$

It remains to show the equality of the second terms in the last two equations. Given that $Q_1'L^{-1}\bar{u}$ and $Q_2'L^{-1}\bar{u}$ are two orthogonal sets of variables and that, by (2.26), $J_2w = \alpha Q_2'L^{-1}\bar{u}$, we have

$$E\left[\bar{u}(t) - C'(t)[(R^{-1}Q_1'L^{-1}\bar{u})',0']'\,|\,J_2w\right] = E\left[\bar{u}(t)\,|\,J_2w\right].$$

By definition of $P(t)$, the following equalities hold

$$P(t)\left(y - C\hat{z}_j\right) = cov\left(\bar{u}(t),\bar{u}\right)V^{-1}(\bar{u})\left(y - C\hat{z}_j\right) = \\
= E\left(\bar{u}(t)\bar{u}'\right)(L^{-1})'QQ'L^{-1}\left(y - C\hat{z}_j\right) = \\
= E\left(\bar{u}(t)\bar{u}'\right)\left((L^{-1})'Q_1,(L^{-1})'Q_2\right)\left(0',(Q_2'L^{-1}y)'\right)' = \\
= E\left(\bar{u}(t)(Q_2'L^{-1}\bar{u})'\right)Q_2'L^{-1}y = \\
= E\left(\bar{u}(t)(J_2w)'\right)V^{-1}(J_2w)J_2w = \\
= E\left[\bar{u}(t)\,|\,J_2w\right],$$

where we have used the fact that $Q_1'L^{-1}\left(y - C\hat{z}_j\right) = 0$ and $Q_2'L^{-1}\left(y - C\hat{z}_j\right) = Q_2'L^{-1}y$. Since $z(t) - \hat{z}(t) = z(t) - \bar{z}(t)$, if the same estimator of σ^2 is used, it follows trivially that the MSE are equal.

□

2.6   Regression Model with ARIMA errors and Missing Values

Consider the regression model (1.13a), where the vectors β and $y(t)$ are as in Section 1.4 and the residuals $\{v(t)\}$ follow the ARIMA model (2.1) with $z(t)$ replaced by $v(t)$. Defining the vectors z, v and the matrix Y as in Section 1.4, equation (1.13b) still holds and, similarly to the stationary case, we can proceed in two ways. First, the β parameters, the

initial missing values z_j and the parameters of the ARIMA model can be jointly estimated. The state space representation would be given by (1.14) adapted to the nonstationary case (i.e., equation (2.9a) with z replaced by v, and equation (1.14b)). The starting conditions would be (2.10b) and

$$ x(d+1\,|d) = \left[A'(d+1)\!\left(z_{\bullet\bullet} - Y_{\bullet}\beta\right), A'(d+2)\!\left(z_{\bullet\bullet} - Y_{\bullet}\beta\right), ..., A'(d+r)\!\left(z_{\bullet\bullet} - Y_{\bullet}\beta\right) \right]', $$

where $z_{\bullet\bullet}$ is the vector formed by the first k observed values (z_I) and some starting values for the missing observations in the first d values of the series (z_J); $Y_{\bullet}$ denotes the $d{\times}h$ matrix formed by the rows $y'(t)$, $t = 1,...,d$. The second procedure concentrates σ^2, β and z_J out of the likelihood function, and is analogous to that described in Section 1.4. Since $\{v(t)\}$ follows the ARIMA model (2.1), with the notation of Section 2.3, we can write

$$ v_{II} = Bv_I + Cv_J + \bar{u}, $$

where v_{II}, v_I, and v_J are the vectors of errors corresponding to the vectors of observations z_{II}, z_I, and z_J, defined at the beguining of Section 2.3. Let Y_I, Y_{II} and Y_J be the matrices with rows the vectors $y'(t)$ corresponding to the vectors v_I, v_{II} and v_J, respectively. Replacing v_I by $z_I - Y_I\beta$ and v_J by $z_J - Y_J\beta$ in the above expression, the following regression model is obtained

$$ z_{II} = Bz_I + Cz_J + Y_{II}\beta - BY_I\beta - CY_J\beta + \bar{u}, $$

where the regression parameters are z_J and β. Letting $y = z_{II} - Bz_I$, it can be rewritten as

$$ y = \left[C, Y_{II} - BY_I - CY_J \right]\!\left[z_J', \beta' \right]' + \bar{u} = $$
$$ = \left[C, Y_{II} - A_{II}Y_{\bullet} \right]\!\left[z_J', \beta' \right]' + \bar{u}, $$

where $Y_{\bullet}$ is the matrix with the rows given by the vectors $y'(t)$ corresponding to $z_{\bullet}$, and A_{II} is the matrix defined by $BY_I + CY_J = A_{II}Y_{\bullet}$. The Kalman filter applied to the model $y = z_{II} - Bz_I = \bar{u}$ yields $L^{-1}y$ and $|L|$, where L is as in Section 2.3. The starting conditions would be (2.10b) and

$$ x(d+1\,|d) = \left[B'(d+1)z_I, ..., B'(d+r)z_I \right]'. $$

Note that, as in Section 2.3, it is not necessary to compute the vector y and that the Kalman filter is applied to the vector of observations z_{II}. The same algorithm applied to the columns of the matrix $\left[C, Y_{II} - A_{II}Y_{\bullet} \right]$ allows us to compute the product of L^{-1} by this matrix. The starting conditions would be (2.10b) and $x(d+1\,|d) = 0$. Then, the QR algorithm can be applied to the transformed model

$$ L^{-1}y = L^{-1}\!\left[C, Y_{II} - A_{II}Y_{\bullet} \right]\!\left[z_J', \beta' \right]' + L^{-1}\bar{u}, $$

and we can proceed as in Section 1.4.

2.8   Application

We have written a program in Fortran, available for mainframes and PCs under MSDOS, which can be obtained directly from the authors. The program performs estimation, forecasting and interpolation of regression models with missing observations and ARIMA errors.

The regression variables can be inputed by the user or generated by the program. The variables that can be generated are Trading Day and Easter Effect (see Hillmer, Bell and Tiao, 1983) and certain types of intervention variables (see Box and Tiao 1975).

Estimation of the regression parameters (including the missing observations among the first d values of the series) plus the ARIMA model parameters, can be made by concentrating them out of the likelihood, or by joint estimation, as previously described. Several algorithms can be used for computing the likelihood, or more precisely, the nonlinear sum of squares to be minimized. When the differenced series can be used, we employ the algorithm of Morf, Sidhu and Kailath (1974), as discussed by Pearlman (1980) and improved by Mélard (1984).

For the nondifferenced series, it is possible to use the ordinary Kalman filter, as described in this paper (default option), or its square root version (see Anderson and Moore, 1979). The latter is adequate when numerical difficulties arise; however it is markedly slower and does not permit (at present) to concentrate the regression parameters out of the likelihood. By default, the exact maximum likelihood method is employed, and the unconditional least squares method is available as an option. Nonlinear maximization of the likelihood function and computation of the parameter estimates standard errors is made using Marquardt's method and first numerical derivatives.

For forecasting and interpolation, the ordinary Kalman filter or the square root filter options are available. Interpolation of missing values is made with the simplified Fixed Point Smoother, as described in the paper. When concentrating z_j and β out of the likelihood, mean squared errors of the forecasts and interpolations are obtained following the approach of Kohn and Ansley (1985), as described in the paper. If the rank of the C matrix is smaller than $d-k$, the program indicates which initial missing values are free parameters (the program flags the elements of the vector $\hat{z}_j^{**} = \bar{x}$ in Section 2.4) and also which forecasts or interpolations will depend on the arbitrary vector $\bar{x}$.

The user can then assign arbitrary values (typically, very large or very small) to the free parameters and rerun the program. Proceeding in this way, all parameters of the ARIMA model can be estimated because, as seen in Sections 2.3 and 2.4, the function to minimize does not depend on the free parameters. Moreover, it will be evident which forecasts and interpolations are affected by these arbitrary values because they will strongly deviate from the rest of the estimates. However, if all unknown parameters are jointly estimated, the program may not flag all free parameters. It may happen, as in Data Set 5 below, that there is convergence to a valid arbitrary set of solutions for z_j (i.e., that some linear combinations of the initial missing observations, including the free parameters, are estimable.)

Following the tradition set up by Harvey and Pierse (1984) and Kohn and Ansley (1986), we apply our procedure to the series of 144 monthly observations on international airline passengers (January 1949 - December 1960), given in and analyzed by Box and Jenkins (1976, chap. 9). The model indentified in all three cases is the multiplicative ARIMA $(0,1,1) \times (0,1,1)$ model:

$$\nabla_{12}\nabla z_t = (1 + \theta B)(1 + \Theta B^{12})a_t,$$

applied to the logs of the data. We consider the same four data sets of Kohn and Ansley (1986) and add a fifth data set:

Data Set 1. All 144 observations.

Data Set 2. The 78 observations that result from treating January through November in the last 6 years as missing.

Data Set 3. The 139 observations that result from treating July 1949, June, July and August 1957, and July 1960, as missing.

Data Set 4. The 130 observations that result from treating all July values, and June and August 1957, as missing.

Data Set 5. The 130 observations that result from treating all January values, and February 1951 and February 1954, as missing.

In order to facilitate the comparison of our results with those of Kohn and Ansley (1986), we use as estimator of σ^2 that of Ansley and Newbold (1981), which coincides with (2.28). Note that Data Set 2 is the example considered by Harvey and Pierse (1984), and that, since there are not 13 consecutive observations available at either the start or the end of series in the last three data sets, the methodology of Harvey and Pierse cannot be applied in these three cases. In Data Set 3, only one initial observation (number 7) is missing ($k = 1$) and the C matrix has rank one. Thus, all missing observations are estimable. In Data Set 4, there is also one initial observation (number 7) missing ($k = 1$) and the C matrix has rank 0. Therefore, the initial missing observation will be a free parameter. Finally, in Data Set 5, there are two initial missing observations (numbers 1 and 13, $k = 2$) and the C matrix has rank 1. This happens because the $C'(t)$ vectors corresponding to the observed values are all multiples of (-1,1) for $t > 13$. Thus, although $z(1)$ and $z(13)$ are not estimable, the linear combination $z(13) - z(1)$ can be estimated from the available observations. Therefore, if we assign (for example) the value 0 to $z(1)$, and rerun the program, the interpolation of $z(13)$ will be the minimum MSE estimator of the annual difference $(z(13) - z(1))$. Note that the approach of Kohn and Ansley (1986) would tell us that observations 1 and 13 are not estimable.

Tables 1 to 5 are equivalent to Tables 1 to 5 in Kohn and Ansley (1986), and present the results obtained with our program; Table 6 displays the results for Data Set 5. The differenced series was used for Data Set 1, and the ordinary Kalman filter and simplified FPS were applied to the nonstationary levels (as described in the paper) for Data Sets 2, 3, 4 and 5. The estimation method was always exact maximum likelihood, and in data sets 3 and 5, the initial missing values that are not free parameters (i.e., observation 7 in Data Set 3 and observation 1 in Data Set 5) were concentrated out of the likelihood function . As for the free parameters, we assigned value 10^{-9} to observation 7 in Data Set 4 and, as stated above, value 1 to observation 13 in Data Set 5. Note that no adjustment in the denominator of σ^2 is needed since we increase the number of observations by one and, at the same time, increase k (the number of nonmissing initial values) also by one.

24

The results obtained with joint estimation of all parameters were practically identical (although CPU time increased by close to 15 %) except for Data Set 5. In this case the program did not flag any free parameter and there was convergence to a valid set of solutions for the initial missing values. Note that two valid sets of solutions are equal up to an additive constant (the free parameter in this case).

Table 1 presents the estimates of the ARIMA model parameters and their standard errors. The estimate of θ is the same as the one in Kohn and Ansley (1986), except for a small difference in Data Set 3 (-.405 vs -.408). The estimate of Θ is identical in all cases. Table 2 displays the one to twelve months ahead forecasts (and their root mean squared errors), for the five data sets. The results are practically identical to those obtained by Kohn and Ansley (1986). The only noticeable difference is the forecast for September in Data Set 4 (6.326 vs 6.333). Given that the forecasts for September are 6.325 for data sets 1, 3 and 5, it is possible that the second forecast contains a typo. There is a discrepancy of .001 in the root mean squared errors of three forecasts

Table 1. Parameter Estimates and Standard Errors

Data set	θ	Θ	σ
1.(N = 144)	-.402	-.557	.037
	(.080)	(.084)	
2.(N = 78)	-.457	-.758	.042
	(.096)	(.227)	
3.(N = 139)	-.405	-.566	.037
	(.081)	(.083)	
4.(N = 130)	-.430	-.573	.037
	(.081)	(.085)	
5.(N = 130)	-.401	-.563	.037
	(.082)	(.089)	

(March for Data Set 1, October for Data Set 2 and October for Data Set 3) that could be attributed to the use by Kohn and Ansley of N-14 instead of N-15 as denominator to compute their residual variance in data sets 1, 2 and 3 (although the rank of their A matrix is 13). Table 3 contains the 11 values interpolated for the months January to November 1957 for Data Set 2. Table 4 presents the estimates for the 5 missing values in Data Set 3. Table 5 shows the estimates of the missing values in Data Set 4 that do not depend on the free parameter.

Finally, Table 6 exhibits the estimates of the missing values in Data Set 5 that do not depend on the free parameter and also the estimate of $z_{13} - z_1$. The results in Tables 3,

Table 2. 1961 Forecasts and Root Mean Squared Errors for Logged Data

Data Set	Jan.	Feb.	Mar.	April	May	June	July	Aug.	Sep.	Oct.	Nov.	Dec.
1.(N = 144)	6.110	6.054	6.172	6.199	6.233	6.369	6.507	6.503	6.325	6.209	6.064	6.168
	(.037)	(.043)	(.049)	(.053)	(.058)	(.062)	(.066)	(.069)	(.073)	(.076)	(.079)	(.082)
2.(N = 78)	6.084	6.091	6.247	6.205	6.199	6.308	6.409	6.414	6.299	6.174	6.043	6.174
	(.053)	(.059)	(.064)	(.069)	(.073)	(.077)	(.080)	(.083)	(.086)	(.089)	(.091)	(.087)
3.(N = 139)	6.110	6.054	6.173	6.199	6.232	6.367	6.497	6.503	6.325	6.209	6.064	6.168
	(.038)	(.044)	(.049)	(.054)	(.058)	(.062)	(.068)	(.070)	(.073)	(.077)	(.080)	(.083)
4.(N = 130)	6.111	6.055	6.174	6.200	6.233	6.368	*	6.503	6.326	6.209	6.064	6.169
	(.037)	(.043)	(.048)	(.053)	(.057)	(.061)	(.080)	(.068)	(.071)	(.074)	(.077)	(.080)
5.(N = 130)	*	6.055	6.172	6.199	6.232	6.369	6.507	6.503	6.325	6.209	6.064	6.168
	(.058)	(.043)	(.048)	(.053)	(.057)	(.061)	(.065)	(.069)	(.072)	(.075)	(.078)	(.081)

NOTE: Root Mean Squared errors are given in parentheses.
* Value depends on a free parameter.

25

Table 3. Data Set 2 (N = 78) Interpolation for January-November 1957: Estimates and Root Mean Squared Errors for Logarithms of Data

	Jan.	Feb.	Mar.	April	May	June	July	Aug.	Sep.	Oct.	Nov.
Estimate	5.733	5.738	5.893	5.850	5.843	5.951	6.051	6.055	5.938	5.812	5.680
	(.046)	(.050)	(.053)	(.055)	(.056)	(.056)	(.056)	(.055)	(.053)	(.050)	(.046)
Actual	5.753	5.707	5.875	5.852	5.872	6.045	6.146	6.146	6.001	5.849	5.720

NOTE: Root mean squared errors are given in parentheses.

Table 4. Data Set 3 (N = 139): Estimates and Root Mean Squared Errors for Logarithms of Data

	July	June	July	Aug.	July
	1949	1957	1957	1957	1960
Estimate	5.013	6.024	6.147	6.148	6.409
	(.031)	(.030)	(.031)	(.030)	(.032)
Actual	4.997	6.045	6.146	6.146	6.433

NOTE: Root mean squared errors are given in parentheses

Table 5. Data Set 4 (N = 130): Estimates and Root Mean Squared Errors for Logarithms of Data

	June 1957	Aug. 1957
Estimate	6.023	6.147
	(.030)	(.030)
Actual	6.045	6.146

NOTE: Root mean squared errors are given in parentheses.

4 and 5 are identical to those obtained by Kohn and Ansley (1986).

For data sets 4 and 5, as mentioned above, to know whether the estimator of an unobserved value depends on a free parameter or not, it may help to notice that the estimated value is very different from the other estimated values. However, the precise condition to check is whether its $C'(t)$ vector lies in the space generated by the C matrix. For example, we cannot predict January 1961 in Data Set 5 because its $C'(t)$ vector is (-11, 12), which does not lie in the space generated by the (-1, 1) vector.

Table 6. Data Set 5 (N = 130): Estimates and Root Mean Squared Errors for Logarithms of Data

	Jan. 1950 - Jan. 1949	Feb. 1951	Feb. 1954
Estimate	.068	5.020	5.327
	(.040)	(.029)	(.028)
Actual	.026	5.011	5.236

NOTE: Root mean squared errors are given in parentheses.

APPENDIX A

To illustrate our approach, we will use the same example of Kohn and Ansley (1986), namely the model

$$z(t) = z(t-4) + a(t) + \theta a(t-1). \tag{A.1}$$

With the notation of Section 2.1, we have $d = 4$, $p = 0$, $q = 1$, $u(t) = \delta(B)z(t)$, $\delta(B) = 1 - B^4$, $\theta(B) = 1 + \theta B$ and $\xi(B) = 1/\delta(B) = 1 + B^4 + B^8 + \dots$. Assume we observe $z = (z(1), z(4), z(5), z(6), z(8), z(9), z(10), z(12))'$. The $A'(t) = (A_1(t), A_2(t), A_3(t), A_4(t))$ vectors of (2.3)

26

and (2.4) are generated from the recursions

$$A_i(t) = A_i(t-4) , \quad t > 4, \ i = 1, 2, 3, 4,$$

$$A_i(j) = \begin{cases} 1 & \text{if} \quad i = j = 1, \dots, 4 \, ; \\ 0 & \text{if} \quad i, j = 1, \dots, 4 \, ; \quad i \neq j \, . \end{cases}$$

To obtain the state space representation, we need $\psi^*(B) = \theta(B)/\phi^*(B) = \theta(B)/\delta(B) = 1 + \theta B + B^4 + \dots$. Equations (2.9a) and (1.11) in this case become

$$x(t) = \begin{bmatrix} z(t) \\ z(t+1|t) \\ z(t+2|t) \\ z(t+3|t) \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} z(t-1) \\ z(t|t-1) \\ z(t+1|t-1) \\ z(t+2|t-1) \end{bmatrix} + \begin{bmatrix} 1 \\ \theta \\ 0 \\ 0 \end{bmatrix} a(t) , \qquad (A.2)$$

$$z(t) = (1,0,0,0)x(t) + \alpha(t)W(t).$$

Since there are missing values, with the notation of Section 2.3, we have $z_* = (z(1), z(2), z(3), z(4))'$, $z_I = (z(2), z(3))'$, $z_1 = (z(1), z(4))', M = 8, k = 2$, $\tilde{u}(t) = \sum_{j=0}^{t-5} \xi_j u(t-j), \ t > 4$, $z_{II} = (z(5), z(6), z(8), z(9), z(10), z(12))'$ and $\tilde{u} = (\tilde{u}(5), \tilde{u}(6), \tilde{u}(8), \tilde{u}(9), \tilde{u}(10), \tilde{u}(12))'$.
The matrices A, B, C and D corresponding to equation (2.11) are in this case

$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1 & 0 \\ 0 & 0 \\ 0 & 1 \\ 1 & 0 \\ 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad C = \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 0 \\ 0 & 0 \\ 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad D = \begin{bmatrix} I_2 & 0 \\ B & C \end{bmatrix}.$$

It is easy to check that matrix D is the result of first interchanging columns 2 and 4 and the interchanging columns 3 and 4, in matrix A. From the form of matrix C, it is immediate that rank(C) = 1 and that $z(3)$ is a free parameter. Let the covariance matrix of $\tilde{u}$ be $V(\tilde{u}) = \sigma^2 \Delta$ and let $\Delta = LL'$ be the Cholesky decomposition of Δ. In order to compute the function S^* in (2.17), we consider the regression model

$$y = C z_I + \tilde{u},$$

where $y = z_{II} - B z_I$, and proceed as in Section 2.3. First, we apply the Kalman filter to the model $y = z_{II} - B z_I = \tilde{u}$ to obtain $L^{-1}y$ and $|L|$. The equations to use are (A.2). The starting conditions for the Kalman filter are $x(5|4) = [z(1), 0, 0, z(4)]'$ and

$$\Sigma(5\,|\,4) = \Xi\tilde{\Sigma}\Xi' = I_4\tilde{\Sigma}I_4 = \tilde{\Sigma} = \begin{bmatrix} 1+\theta^2 & \theta & 0 & 0 \\ \theta & \theta^2 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix}. \tag{A.3}$$

Note that the Kalman filter is applied to the vector of observations z_{II} and that it is not necessary to compute the entire matrix B. Then, the same algorithm, applied to the columns of matrix C, permits us to compute $L^{-1}C$. The starting conditions are $x(5\,|\,4) = 0$ and (A.3). Therefore, we can consider the model

$$L^{-1}y = L^{-1}Cz_j + L^{-1}\bar{u}. \tag{A.4}$$

Since the second column of matrix C is zero, so will be the second column of $L^{-1}C$. Therefore, if we denote the first and second columns of matrix C by $C(1)$ and $C(2)$, respectively, we may write

$$\begin{aligned} L^{-1}y &= L^{-1}C(1)z(2) + L^{-1}C(2)z(3) + L^{-1}\bar{u} = \\ &= L^{-1}C(1)z(2) + L^{-1}\bar{u}, \end{aligned} \tag{A.4}$$

and $z(3)$ is a free parameter. We can apply the QR algorithm to the vector $L^{-1}C(1)$ to obtain an orthogonal matrix Q such that

$$Q'L^{-1}C(1) = \left[R, 0'\right]',$$

where R is a nonzero scalar. Let Q_1' denote the first row of Q' and let Q_2' denote the submatrix of Q' formed by the other three rows. Then, we can rewrite equations (A.4) in the form

$$\begin{aligned} Q_1'L^{-1}y &= (R, 0')z_j + Q_1'L^{-1}\bar{u}, \\ Q_2'L^{-1}y &= \qquad\qquad Q_2'L^{-1}\bar{u}. \end{aligned}$$

From this, we obtain $\hat{z}(2) = R^{-1}Q_1'L^{-1}y$, $\hat{z}(3) = \bar{x}$ (an arbitrary value) and

$$S^* = |L|^{1/6}y'(L^{-1})'Q_2Q_2'L^{-1}y|L|^{1/6}.$$

We now consider the numerical example of Kohn and Ansley. That is, we suppose $\theta = -.5$, $\sigma^2 = 1$ and $z = (1.2, -1.3, 2.1, 3.2, .5, .8, -.4, 1.2)'$. In table A.1, we give output from the Kalman filter applied as described above. For each time index, we present values for the covariance matrix $\Sigma(t\,|\,t-1)$, the vectors $x(t\,|\,t-1)$, and $c(t\,|\,t-1)$ (for column $C(1)$), the Kalman gain $K(t)$ and the corresponding elements of the vectors of standardized residuals $L^{-1}y$ and $L^{-1}C(1)$. It is easy to check that if Q' is the matrix obtained from the unit matrix by interchanging its first and second rows, then $Q'L^{-1}C(1) = (.976, 0')' = (R, 0')'$ and $Q'L^{-1}y = (3.474, .805, 1.610, -.566, -3.853, .626)'$. From this, we obtain $\hat{z}_2 = 3.56$ with MSE 2.222, $y'(L^{-1})'Q_2Q_2'L^{-1}y = 18.8$ and finally $S^* = 21.406$.

As for interpolation and prediction, we have just seen, $\hat{z}(2) = 3.56$ and $z(3)$ is a free parameter. To see whether the interpolators for $z(7)$ and $z(11)$ depend on the free parameter, we have to examine its $C'(t)$ vectors. We have $C'(7) = (0, 1)$ and $C'(11) = (0, 1)$. Therefore, both interpolators will depend on the free parameter. As for the forecasts of $z(13)$, $z(14)$ and $z(15)$, we have $C'(13) = (0, 0)$, $C'(14) = (1, 0)$ and $C'(15) = (0, 1)$. Therefore, only the forecast of $z(15)$ will depend on the free parameter. To obtain the forecasts, we put $z(2) = 3.56$ and $z(3) = 9999$ (an arbitrary value) and apply the Kalman filter to the model $z_{II} - Az_* = \bar{u}$. The starting conditions are (see Sections 2.4 and 2.2) $x(6\,|\,5) = (1.2, 3.56, 9999, -1.3)'$ and (A.3). In table (A.2) we present output from the Kalman filter

applied to $z_{II} - Az_{\bullet} = \tilde{u}$. This time, we only give for each time index values for $x(t\,|\,t-1)$, $L^{-1}(z_{II} - Az_{\bullet})$ and the MSE $\sigma^2(t\,|\,t-1)$ (the element in the first row and the first column of $\Sigma(t\,|\,t-1)$), the other values being the same as in table A.1 for $t = 1, \ldots , 12$.

Table A.1. Details of Kalman Filter for example A.1

t	x(t|t-1)	c(t|t-1)	Σ(t|t-1)				K(t)	L⁻¹y	L⁻¹C(1)
5	1.2	0	1.25	-.5	0	0	1	.805	0
	0	0	-.5	.25	0	0	-.4		
	0	0	0	0	0	0	0		
	-1.3	0	0	0	0	0	0		
6	-.36	0	1.05	-.5	0	0	1	3.474	.976
	0	0	-.5	.25	0	0	-.476		
	-1.3	0	0	0	0	0	0		
	2.1	0	0	0	0	0	0		
7	-1.695	-.476	1.012	-.5	0	0	0	-	-
	-1.3	0	-.5	.25	0	0	0		
	2.1	0	0	0	0	0	0		
	3.2	1	0	0	0	0	0		
8	-1.3	0	1.25	-.5	0	-.5	1	1.610	0
	2.1	0	-.5	.25	0	0	-.4		
	3.2	1	0	0	0	0	0		
	-1.695	-.476	-.5	0	0	1.012	-.4		
9	1.38	0	1.05	-.5	-.2	0	1	-.566	0
	3.2	1	-.5	.25	0	0	-.476		
	-2.415	-.476	-.2	0	.812	0	-.190		
	.5	0	0	0	0	0	0		
10	3.476	1	1.012	-.595	0	0	1	-3.853	0
	-2.305	-.476	-.595	1.024	0	0	-.588		
	.5	0	0	0	0	0	0		
	.8	0	0	0	0	0	0		
11	-.025	-.476	1.674	-.5	0	0	0	-	-
	.5	0	-.5	.25	0	0	0		
	.8	0	0	0	0	0	0		
	-.4	1	0	0	0	0	0		
12	.5	0	1.25	-.5	0	-.5	1	.626	0
	.8	0	-.5	.25	0	0	-.4		
	-.4	1	0	0	0	0	0		
	-.025	-.476	-.5	0	0	1.674	-.4		

Table A.2. Forecasting Details for example A.1

t	x(t|t-1)	L⁻¹(z_{II}-z_*)	σ²(t|t-1)
5	1.2	.805	1.25
	3.56		
	9999		
	-1.3		
6	3.2	0	1.05
	9999		
	-1.3		
	2.1		
7	9999	-	1.012
	-1.3		
	2.1		
	3.2		
8	-1.3	1.610	1.25
	2.1		
	3.2		
	9999		
9	1.38	-.566	1.05
	3.2		
	9998.28		
	.5		
10	3.476	-3.853	1.012
	9998.39		
	.5		
	.8		
11	10000.67	-	1.674
	.5		
	.8		
	-.4		
12	.5	.626	1.25
	.8		
	-.4		
	10000.67		
13	.52	-	1.05
	-.4		
	10000.39		
	1.2		
14	-.4	-	1.25
	10000.39		
	1.2		
	.52		
15	10000.39	-	2.724

REFERENCES

Akaike, H. (1973), "Maximum Likelihood Identification of Gaussian Autoregressive Moving Average Models,", *Biometrika*, 60, 255-265.

Akaike, H. (1974), "Markovian Representation of Stochastic Processes and its Application to the Analysis of Autoregressive Moving Average Processes," *Annals of the Institute of Statistical Mathematics*, 26, 363-387.

Akaike, H. (1975), "Markovian Representation of Stochastic Processes by Canonical Variables," *SIAM, J. Control*, 13, 162-173.

Anderson, B., and Moore, J. (1979), *Optimal Filtering*, New Jersey: Prentice Hall.

Ansley, C. F., and Kohn, R. (1990), "Filtering and Smoothing in State Space Models with Partially Diffuse Initial Conditions, " *Journal of Time Series Analysis*, 11, 275-293.

Ansley, C. F., and Newbold, P. (1981), "On the Bias in Estimates of Forecast Mean Squared Error, " *Journal of the American Statistical Association*, 76, 569-578.

Bell, W., (1984), "Signal Extraction for Nonstationary Series," *The Annals of Statistics*, 12, 646-664.

Bell, W., and Hillmer S. C. (1991), "Initializing the Kalman Filter for Nonstationary Time Series Models," *Journal of Time Series Analysis*, 12, 283-300.

Box, G. E. P., and Jenkins, G. M. (1976), *Time Series Analysis, Forecasting and Control*, San Francisco: Holden - Day.

Box, G. E. P., and Tiao, G. C. (1975), "Intervention Analysis with Applications to Economic and Environmental Problems," *Journal of the American Statistical Association*, 70, 70 - 79.

Brockwell, P., and Davis, R. (1987), *Time Series: Theory and Methods*, Berlin: Springer - Verlag.

Fletcher, R. (1987), *Practical Methods of Optimization*, New York: John Wiley & Sons.

Harvey, A. C., and Phillips, G. D. A. (1979), "Maximum Likelihood Estimation of Regression Models with Autoregressive - Moving Average Disturbances," *Biometrika*, 66, 49 - 58.

Harvey, A. C., and Pierse, R. G. (1984), "Estimating Missing Observations in Economic Time Series, " *Journal of the American Statistical Association*, 79, 125 - 131.

Harvey, A. C., and Peters, S. (1990), "Estimation Procedures for Structural Time Series Models, " *Journal of Forecasting*, 9, 89 - 108.

Hillmer, S. C., Bell, W. R., and Tiao, G. E. (1983), "Modeling Considerations in the Seasonal Adjustment of Economic Time Series, " in A. Zellner (ed.), *Applied Time Series Analysis of Economic Data*, Washington, D. C.: Bureau of the Census.

Jones, R. (1980), "Maximum Likelihood Fitting of ARMA Models to Time Series With Missing Observations, " *Technometrics*, 22, 389 - 395.

Jong, Piet de (1988), "The Likelihood for the State Space Model, " *Biometrika*, 75, 165 - 169.

Jong, Piet de (1989), "Smoothing and Interpolation With the State Space Model," *Journal of the American Statistical Association*, 84, 408 - 409.

Kohn, R., and Ansley, C. F. (1985), "Efficient Estimation and Prediction in Time Series Regression Models, " *Biometrika*, 72, 694 - 697.

Kohn, R., and Ansley, C. F. (1986), "Estimation, Prediction and Interpolation for ARIMA Models With Missing Data," *Journal of the American Statistical Association*, 81, 751 - 761.

Mélard, G. (1984), "A Fast Algorithm for the Exact Likelihood of Autoregressive - Moving Average Models," *Applied Statistics*, 35, 104 - 114.

Morf, M., Sidhu, G. S., and Kailath, T. (1974), "Some New Algorithms for Recursive Estimation on Constant, Linear, Discrete - Time Systems," *IEEE Transactions on Automatic Control*, AC - 19, 315 - 323.

Pearlman, J. G. (1980), "An Algorithm for the Exact Likelihood of a High - Order - Autoregressive - Moving Average Process," *Biometrika*, 67, 232 - 233.

Rao, C. (1973), *Linear Statistical Inference and its Applications*, New York: John Wiley & Sons.

Wecker, W., and Ansley, C. F. (1983), "The Signal Extraction Approach to Nonlinear Regression and Spline Smoothing," *Journal of the American Statistical Association*, 78, 81 - 89.

EUI
WORKING
PAPERS

EUI Working Papers are published and distributed by the
European University Institute, Florence

Copies can be obtained free of charge – depending on the availability of
stocks – from:

The Publications Officer
European University Institute
Badia Fiesolana
I-50016 San Domenico di Fiesole (FI)
Italy

Please use order form overleaf

Publications of the European University Institute

Economics Department Working Paper Series

To   Economics Department **WP**
European University Institute
Badia Fiesolana
I-50016 San Domenico di Fiesole (FI)
Italy

From   Name . . . .

Address. . . . .

. . . .

. . . .

. . . .

. . . .

(Please print)

❏ Please enter/confirm my name on EUI Economics Dept. Mailing List
❏ Please send me a complete list of EUI Working Papers
❏ Please send me a complete list of EUI book publications
❏ Please send me the EUI brochure Academic Year 1992/93

Please send me the following EUI **ECO** Working Paper(s):

No, Author   . . . .
Title: . . . .
No, Author   . . . .
Title: . . . .
No, Author   . . . .
Title: . . . .
No, Author   . . . .
Title: . . . .

Date . . . . Signature . . . .

Working Papers of the Department of Economics
Published since 1990

ECO No. 90/1
Tamer BASAR and Mark SALMON
Credibility and the Value of Information
Transmission in a Model of Monetary
Policy and Inflation

ECO No. 90/2
Horst UNGERER
The EMS – The First Ten Years
Policies – Developments – Evolution

ECO No. 90/3
Peter J. HAMMOND
Interpersonal Comparisons of Utility:
Why and how they are and should be
made

ECO No. 90/4
Peter J. HAMMOND
A Revelation Principle for (Boundedly)
Bayesian Rationalizable Strategies

ECO No. 90/5
Peter J. HAMMOND
Independence of Irrelevant Interpersonal
Comparisons

ECO No. 90/6
Hal R. VARIAN
A Solution to the Problem of
Externalities and Public Goods when
Agents are Well-Informed

ECO No. 90/7
Hal R. VARIAN
Sequential Provision of Public Goods

ECO No. 90/8
T. BRIANZA, L. PHLIPS and J.F.
RICHARD
Futures Markets, Speculation and
Monopoly Pricing

ECO No. 90/9
Anthony B. ATKINSON/ John
MICKLEWRIGHT
Unemployment Compensation and
Labour Market Transition: A Critical
Review

ECO No. 90/10
Peter J. HAMMOND
The Role of Information in Economics

ECO No. 90/11
Nicos M. CHRISTODOULAKIS
Debt Dynamics in a Small Open
Economy

ECO No. 90/12
Stephen C. SMITH
On the Economic Rationale for
Codetermination Law

ECO No. 90/13
Elettra AGLIARDI
Learning by Doing and Market Structures

ECO No. 90/14
Peter J. HAMMOND
Intertemporal Objectives

ECO No. 90/15
Andrew EVANS/Stephen MARTIN
Socially Acceptable Distortion of
Competition: EC Policy on State Aid

ECO No. 90/16
Stephen MARTIN
Fringe Size and Cartel Stability

ECO No. 90/17
John MICKLEWRIGHT
Why Do Less Than a Quarter of the
Unemployed in Britain Receive
Unemployment Insurance?

ECO No. 90/18
Mrudula A. PATEL
Optimal Life Cycle Saving With
Borrowing Constraints:
A Graphical Solution

ECO No. 90/19
Peter J. HAMMOND
Money Metric Measures of Individual
and Social Welfare Allowing for
Environmental Externalities

ECO No. 90/20
Louis PHLIPS/
Ronald M. HARSTAD
Oligopolistic Manipulation of Spot
Markets and the Timing of Futures
Market Speculation

ECO No. 90/21
Christian DUSTMANN
Earnings Adjustment of Temporary
Migrants

ECO No. 90/22
John MICKLEWRIGHT
The Reform of Unemployment
Compensation:
Choices for East and West

ECO No. 90/23
Joerg MAYER
U. S. Dollar and Deutschmark as
Reserve Assets

ECO No. 90/24
Sheila MARNIE
Labour Market Reform in the USSR:
Fact or Fiction?

ECO No. 90/25
Peter JENSEN/
Niels WESTERGÅRD-NIELSEN
Temporary Layoffs and the Duration of
Unemployment: An Empirical Analysis

ECO No. 90/26
Stephan L. KALB
Market-Led Approaches to European
Monetary Union in the Light of a Legal
Restrictions Theory of Money

ECO No. 90/27
Robert J. WALDMANN
Implausible Results or Implausible Data?
Anomalies in the Construction of Value
Added Data and Implications for Esti-
mates of Price-Cost Markups

ECO No. 90/28
Stephen MARTIN
Periodic Model Changes in Oligopoly

ECO No. 90/29
Nicos CHRISTODOULAKIS/
Martin WEALE
Imperfect Competition in an Open
Economy

* * *

ECO No. 91/30
Steve ALPERN/Dennis J. SNOWER
Unemployment Through 'Learning From
Experience'

ECO No. 91/31
David M. PRESCOTT/Thanasis
STENGOS
Testing for Forecastible Nonlinear
Dependence in Weekly Gold Rates of
Return

ECO No. 91/32
Peter J. HAMMOND
Harsanyi's Utilitarian Theorem:
A Simpler Proof and Some Ethical
Connotations

ECO No. 91/33
Anthony B. ATKINSON/
John MICKLEWRIGHT
Economic Transformation in Eastern
Europe and the Distribution of Income

ECO No. 91/34
Svend ALBAEK
On Nash and Stackelberg Equilibria
when Costs are Private Information

ECO No. 91/35
Stephen MARTIN
Private and Social Incentives
to Form R & D Joint Ventures

ECO No. 91/36
Louis PHLIPS
Manipulation of Crude Oil Futures

ECO No. 91/37
Xavier CALSAMIGLIA/Alan KIRMAN
A Unique Informationally Efficient and
Decentralized Mechanism With Fair
Outcomes

ECO No. 91/38
George S. ALOGOSKOUFIS/
Thanasis STENGOS
Testing for Nonlinear Dynamics in
Historical Unemployment Series

ECO No. 91/39
Peter J. HAMMOND
The Moral Status of Profits and Other
Rewards:
A Perspective From Modern Welfare
Economics

ECO No. 91/40
Vincent BROUSSEAU/Alan KIRMAN
The Dynamics of Learning in Mis-
Specified Models

ECO No. 91/41
Robert James WALDMANN
Assessing the Relative Sizes of Industry-
and Nation Specific Shocks to Output

ECO No. 91/42
Thorsten HENS/Alan KIRMAN/Louis
PHLIPS
Exchange Rates and Oligopoly

ECO No. 91/43
Peter J. HAMMOND
Consequentialist Decision Theory and
Utilitarian Ethics

ECO No. 91/44
Stephen MARTIN
Endogenous Firm Efficiency in a Cournot
Principal-Agent Model

ECO No. 91/45
Svend ALBAEK
Upstream or Downstream Information
Sharing?

ECO No. 91/46
Thomas H. McCURDY/
Thanasis STENGOS
A Comparison of Risk-Premium
Forecasts Implied by Parametric Versus
Nonparametric Conditional Mean
Estimators

ECO No. 91/47
Christian DUSTMANN
Temporary Migration and the Investment
into Human Capital

ECO No. 91/48
Jean-Daniel GUIGOU
Should Bankruptcy Proceedings be
Initiated by a Mixed
Creditor/Shareholder?

ECO No. 91/49
Nick VRIEND
Market-Making and Decentralized Trade

ECO No. 91/50
Jeffrey L. COLES/Peter J. HAMMOND
Walrasian Equilibrium without Survival:
Existence, Efficiency, and Remedial
Policy

ECO No. 91/51
Frank CRITCHLEY/Paul MARRIOTT/
Mark SALMON
Preferred Point Geometry and Statistical
Manifolds

ECO No. 91/52
Costanza TORRICELLI
The Influence of Futures on Spot Price
Volatility in a Model for a Storable
Commodity

ECO No. 91/53
Frank CRITCHLEY/Paul MARRIOTT/
Mark SALMON
Preferred Point Geometry and the Local
Differential Geometry of the Kullback-
Leibler Divergence

ECO No. 91/54
Peter MØLLGAARD/
Louis PHLIPS
Oil Futures and Strategic
Stocks at Sea

ECO No. 91/55
Christian DUSTMANN/
John MICKLEWRIGHT
Benefits, Incentives and Uncertainty

ECO No. 91/56
John MICKLEWRIGHT/
Gianna GIANNELLI
Why do Women Married to Unemployed
Men have Low Participation Rates?

ECO No. 91/57
John MICKLEWRIGHT
Income Support for the Unemployed in
Hungary

ECO No. 91/58
Fabio CANOVA
Detrending and Business Cycle Facts

ECO No. 91/59
Fabio CANOVA/
Jane MARRINAN
Reconciling the Term Structure of
Interest Rates with the Consumption
Based ICAP Model

ECO No. 91/60
John FINGLETON
Inventory Holdings by a Monopolist
Middleman

* Working Paper out of print

ECO No. 92/61
Sara CONNOLLY/John
MICKLEWRIGHT/Stephen NICKELL
The Occupational Success of Young Men
Who Left School at Sixteen

ECO No. 92/62
Pier Luigi SACCO
Noise Traders Permanence in Stock
Markets: A Tâtonnement Approach.
I: Informational Dynamics for the Two-
Dimensional Case

ECO No. 92/63
Robert J. WALDMANN
Asymmetric Oligopolies

ECO No. 92/64
Robert J. WALDMANN /Stephen
C. SMITH
A Partial Solution to the Financial Risk
and Perverse Response Problems of
Labour-Managed Firms: Industry-
Average Performance Bonds

ECO No. 92/65
Agustín MARAVALL/Víctor GÓMEZ
Signal Extraction in ARIMA Time Series
Program SEATS

ECO No. 92/66
Luigi BRIGHI
A Note on the Demand Theory of the
Weak Axioms

ECO No. 92/67
Nikolaos GEORGANTZIS
The Effect of Mergers on Potential
Competition under Economies or
Diseconomies of Joint Production

ECO No. 92/68
Robert J. WALDMANN/
J. Bradford DE LONG
Interpreting Procyclical Productivity:
Evidence from a Cross-Nation Cross-
Industry Panel

ECO No. 92/69
Christian DUSTMANN/John
MICKLEWRIGHT
Means-Tested Unemployment Benefit
and Family Labour Supply: A Dynamic
Analysis

ECO No. 92/70
Fabio CANOVA/Bruce E. HANSEN
Are Seasonal Patterns Constant Over
Time? A Test for Seasonal Stability

ECO No. 92/71
Alessandra PELLONI
Long-Run Consequences of Finite
Exchange Rate Bubbles

ECO No. 92/72
Jane MARRINAN
The Effects of Government Spending on
Saving and Investment in an Open
Economy

ECO No. 92/73
Fabio CANOVA and Jane MARRINAN
Profits, Risk and Uncertainty in Foreign
Exchange Markets

ECO No. 92/74
Louis PHLIPS
Basing Point Pricing, Competition and
Market Integration

ECO No. 92/75
Stephen MARTIN
Economic Efficiency and Concentration:
Are Mergers a Fitting Response?

ECO No. 92/76
Luisa ZANCHI
The Inter-Industry Wage Structure:
Empirical Evidence for Germany and a
Comparison With the U.S. and Sweden

ECO NO. 92/77
Agustín MARAVALL
Stochastic Linear Trends: Models and
Estimators

ECO No. 92/78
Fabio CANOVA
Three Tests for the Existence of Cycles
in Time Series

ECO No. 92/79
Peter J. HAMMOND/Jaime SEMPERE
Limits to the Potential Gains from Market
Integration and Other Supply-Side
Policies

ECO No. 92/80
Víctor GÓMEZ and Agustín
MARAVALL
Estimation, Prediction and Interpolation
for Nonstationary Series with the
Kalman Filter

* Working Paper out of print