EUI WORKING PAPERS IN ECONOMICS

EUI Working Paper ECO No. 93/7

Initializing the Kalman Filter with Incompletely Specified Initial Conditions

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. 93/7

Initializing the Kalman Filter with Incompletely Specified Initial Conditions

VÍCTOR GÓMEZ
and
AGUSTÍN MARAVALL

BADIA FIESOLANA, SAN DOMENICO (FI)

Printed in Italy in February 1993
European University Institute
Badia Fiesolana
I – 50016 San Domenico (FI)
Italy

INITIALIZING THE KALMAN FILTER WITH INCOMPLETELY SPECIFIED INITIAL CONDITIONS

Víctor Gómez, Instituto Nacional de Estadistica, Madrid
Agustín Maravall, European University Institute, Badia Fiesolana, I-50016, S. Domenico di Fiesole (FI), Italy

Abstract

We review different approaches to Kalman filtering with incompletely specified initial conditions, appropiate for example when dealing with nonstationarity. We compare in detail the transformation approach and modified Kalman Filter (KF) of Ansley and Kohn, the diffuse likelihood and diffuse KF of de Jong, the approach of Bell and Hillmer, whereby the transformation approach applied to an initial stretch of the data yields initial conditions for the KF, and the approach of Gòmez and Maravall, which uses a conditional distribution on initial observations to obtain initial conditions for the KF. It is concluded that the later approach yields a substantially simpler solution to the problem of optimal estimation, forecasting and interpolation for a fairly general class of models.

1. Introduction

We consider observations generated by a discrete time state space model (SSM) such that the initial state vector x_0 has a distribution which is unspecified. We will further allow for unknown regression type parameters. Examples are non-stationary time series which follow an ARIMA model, regression models with ARIMA disturbances, structural models (as in [8]) and ARIMA component models, among others. In all these cases it is not possible to initialize the Kalman Filter (KF) as usual, by means of the first two moments of the distribution of x_0, because they are not well defined. Therefore, it is necessary to incorporate new assumptions in order to deal with this initialization problem.

Among the different alternatives that have been proposed in the literature, we will focus on the transformation approach of Kohn and Ansley, the diffuse Kalman filter (DKF) of de Jong, the initialization procedure of Bell and Hillmer and the approach of Gómez and Maravall, based on a trivial extension of the KF, to be denoted the Extended Kalman Filter (XKF), with a distribution defined conditionally on the initial observations. There are other approaches as well, like the so called "big k" method (see, for example,[5] and [7]). This method uses a matrix of the form kI to initialize the state covariance matrix, where k is large to reflect uncertainty regarding the initial state. The big k method is not only numerically dangerous, it is also inexact. An alternative to the big k method is to use the information filter (see [1]). However, as seen in [2], the information filter breaks down in many important cases, including ARMA models.

The paper is structured as follows. In Section 2 we will define the SSM and consider some illustrative examples. In Section 3 we suppose that the initial state vector x_0 is fixed, define the likelihood and show how the XKF and the DKF can be used to evaluate it. In Section 4 we will deal with the different approaches to define and evaluate the likelihood of the SSM in the case when there are no regression type parameters and the initial state vector has an unspecified distribution. In Section 5 we will extend these results to include regression type parameters.

2. State Space Model

DEFINITION 1. A vectorial time series $\nu = (\nu_1', \ldots, \nu_N')'$ is said to be generated by the State Space Model (SSM) if, for $t = 1, \ldots, N$,

$$\nu_t = X_t\beta + C_t x_t + Z_t \xi_t, \qquad x_t = W_{t-1}\beta + A_{t-1}x_{t-1} + H_{t-1}\xi_{t-1}, \tag{1}$$

where $x_0 = B\delta$, $\xi_t \sim Niid(0, \sigma^2 I)$, $t = 0, \ldots, N$, $\delta \sim N(c, \sigma^2 C)$ with C nonsingular or $C = 0$, δ and $\xi = (\xi_0', \ldots, \xi_N')'$ are independent, B is of full column rank and β is a vector of fixed regression parameters. Also, $\mathrm{Var}(\nu)$ is nonsingular if $C = 0$.

This definition is similar to the one in [13]; the vector δ models uncertainty with respect to the initial conditions. Following [13], we will say that δ is diffuse if C^{-1} is arbitrarily close to 0 in the euclidean norm, denoted $C \to \infty$. Contrary to de Jong, we will always suppose that β, the vector of regression parameters, is fixed; considering β diffuse introduces confusion as to what likelihood should be used and it affects neither the equations nor the

computations with the DKF, to be defined below.

The formulation we use for the SSM has the virtue of explicitly separating the time-invariant "mean" effect β from the state vector x_t, keeping its dimension to a minimum. Choosing adequately the matrices X_t, W_t, H_t and Z_t, appropiate components of β and ξ_t can be excluded from or included in each equation. Thus, the specification covers the case where the mean and disturbance effects in each equation are distinct.Two simple examples will illustrate the definition.

EXAMPLE 1. Suppose a regression model with random walk disturbance and scalar v_t,

$$\nabla(v_t - y_t'\beta) = a_t, \tag{2}$$

where $\nabla = 1 - L$, L is the lag operator $(L(v_t) = v_{t-1})$, and the $a_t \sim N(0, \sigma^2)$ are independent. Model (2) can be put into state space form by defining $X_t = y_t'$, $C_t = 1$, $Z_t = 0$, $W_t = 0$, $A_t = 1$, $H_t = 1$, $x_t = v_t - y_t'\beta$ and $\xi_{t-1} = a_t$. That is,

$$x_t = x_{t-1} + a_t \tag{3a}$$

$$v_t = y_t'\beta + x_t \tag{3b}$$

For initialization, we make $A_0 = 1$, $H_0 = 1$, $W_0 = 0$, $B = 1$ and $x_0 = \delta$. Therefore, the first state is $x_1 = \delta + a_1$ and δ is in this case equal to the initial state. Because $\{x_t\}$ follows the non-stationary model (3a), the distribution of δ is unspecified. □

EXAMPLE 2. Suppose Example 1, but with ∇ replaced by $1 - \rho L$, where $|\rho| < 1$. Then, we have a regression model with AR(1) disturbances. The SSM is

$$x_t = \rho x_{t-1} + a_t \tag{4}$$

and (3b). For initialization, we make $A_0 = 1$, $H_0 = 1/\sqrt{1-\rho^2}$, $W_0 = 0$, $B = 1$ and $x_0 = 0$ $(c = 0, C = 0)$. In this case, $\{x_t\}$ follows the stationary model (4) and we can use the first two moments of x_t, namely $E(x_t) = 0$ and $Var(x_t) = \sigma^2/(1-\rho^2)$, to set up the initial conditions. The first state is $x_1 = (1/\sqrt{1-\rho^2})a_1$. □

A representation which will be very useful in what follows is given by the next theorem.

THEOREM 1. *If* $v = (v_1', \ldots, v_N')'$ *is generated by the SSM* (1), *then* $v = R\delta + S\beta + \varepsilon$, *where the rows of S are*

$$S_1 = X_1 + C_1 W_0$$
$$S_2 = X_2 + C_2(W_1 + A_1 W_0)$$
$$\ldots\ldots$$
$$S_N = X_N + C_N\{W_{N-1} + A_{N-1}W_{N-2} + \ldots + (A_{N-1}\ldots A_1)W_0\}$$

and those of R are

$$R_i = C_i A_{i-1} \dots A_0 B, \quad i = 1, \dots, N.$$

Besides, $\varepsilon \sim N(0, \sigma^2 \Sigma)$ *with* Σ *nonsingular and* $Cov(\delta, \varepsilon) = 0$.

PROOF. The expressions for S_i and R_i are obtained by repeated substitution using (1). The vectors ε_i are linear combinations of $\xi_0, \xi_1, \dots, \xi_i$, $i = 1, \dots, N$. □

3. Fixed Initial State

If δ is fixed ($C = 0$), then $\delta = c$ and the representation of Theorem 1, $\nu = R\delta + S\beta + \varepsilon$, constitutes a regression model where the distribution of ε is known. If we define $X = (R, S)$ and $\gamma = (\delta', \beta')'$, then the log-likelihood of this model, based on ν , is (throughout the paper all log-likelihoods will be defined up to an additive constant)

$$\lambda(\nu) = -\frac{1}{2}\{M \ln(\sigma^2) + \ln|\Sigma| + (\nu - X\gamma)'\Sigma^{-1}(\nu - X\gamma)/\sigma^2\},$$

where $\mathrm{Var}(\nu) = \sigma^2\Sigma$ and M denotes the number of components in ν , the vector of stacked observations. The maximum likelihood estimator of σ^2 is

$$\hat{\sigma}^2 = (\nu - X\gamma)'\Sigma^{-1}(\nu - X\gamma)/M \ . \tag{5}$$

Substituting $\hat{\sigma}^2$ back in $\lambda(\nu)$ yields the σ^2 -maximized log-likelihood :

$$l(\nu) = -\frac{1}{2}\{M \ln(\hat{\sigma}^2) + \ln|\Sigma|\} \ .$$

It turns out that we can evaluate $l(\nu)$ efficiently using the KF.

DEFINITION 2. The Kalman Filter (KF) is the set of recursions

$$\begin{aligned} &e_t = \nu_t - X_t\beta - C_t\hat{x}_{t,t-1}, \quad D_t = C_t P_{t,t-1} C_t' + Z_t Z_t', \\ &G_t = (A_t P_{t,t-1} C_t' + H_t Z_t') D_t^{-1}, \quad \hat{x}_{t+1,t} = W_t\beta + A_t\hat{x}_{t,t-1} + G_t e_t, \\ &P_{t+1,t} = (A_t - G_t C_t) P_{t,t-1} A_t' + (H_t - G_t Z_t) H_t', \end{aligned} \tag{6}$$

with starting conditions $\hat{x}_{1,0} = W_0\beta + A_0 B\delta$ and $P_{1,0} = H_0 H_0'$.

Here $\hat{x}_{t,t-1}$ is the predictor of x_t using $(\nu_1', \dots, \nu_{t-1}')'$ and $\mathrm{Var}(\hat{x}_{t,t-1} - x_{t,t-1}) = P_{t,t-1}$. The e_t are the errors of predicting ν_t using $(\nu_1', \dots, \nu_{t-1}')'$. They constitute an orhogonal sequence with $E(e_t) = 0$ and $\mathrm{Var}(e_t) = D_t$, as given in (6). Note that we have supposed $\sigma^2 = 1$ in the equations (6) because we will estimate σ^2 using (5). It can be shown (see, for example, [13]) that if $e = (e_1', \dots, e_N')'$, then there exists a lower triangular matrix K with ones in the main diagonal such that $e = K(\nu - X\gamma)$ and $K\Sigma K' = D = \mathrm{diag}(D_1, D_2, \dots, D_N)$. Therefore, $\Sigma^{-1} = K' D^{-1} K$ and $\hat{\sigma}^2 = e' D^{-1} e / M$, $\ln|\Sigma| = \ln|D_1| + \ln|D_2| + \dots + \ln|D_N|$. In the case of scalar ν_t , the D_t are also scalar and we can obtain a "square root" of Σ^{-1} by

putting $\Sigma^{-1/2} = D^{-1/2}K$. Then, we can use a vector of standardized residuals $\tilde{e} = D^{-1/2}e$ such that $\hat{\sigma}^2 = \tilde{e}'\tilde{e}/M$ and maximizing $l(\nu)$ becomes equivalent to minimizing the non-linear sum of squares

$$S(\nu,\gamma) = (\prod_{t=1}^{N}|D_t^{1/2}|)^{1/M}\,\tilde{e}'\tilde{e}\,(\prod_{t=1}^{N}|D_t^{1/2}|)^{1/M} \quad .$$

For vectorial ν_t , suppose that in each step of the KF, in addition to D_t , we also obtain, by means of its Cholesky decomposition, a "square root" $D_t^{1/2}$ of D_t , $t = 1,\ldots,N$. Then, the matrix $D^{1/2} = \mathrm{diag}(D_1^{1/2}, D_2^{1/2}, \ldots, D_N^{1/2})$ is a "square root" of D and we can proceed as in the scalar case to evaluate $S(\nu,\gamma)$.

EXAMPLE 2 (Continued). In this case $\delta = 0$, so that, by Theorem 1, we have $\nu = S\beta + \varepsilon$. The initial conditions for the KF are $\hat{x}_{1,0} = 0$ and $P_{1,0} = 1/(1-\rho^2)$. Then, the KF gives

$$e_t = \nu_t - X_t\beta - \hat{x}_{t,t-1}, \qquad D_1 = 1/(1-\rho^2), \qquad D_t = 1,\ t>1,$$

$$G_t = \rho, \qquad \hat{x}_{t+1,t} = \rho\,\hat{x}_{t,t-1} + \rho\,e_t,$$

$$P_{1,0} = 1/(1-\rho^2), \qquad P_{t+1,t} = 1,\ t>0 \quad .$$

The vector of residuals is $e = K(\nu - S\beta)$, where

$$K = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ -\rho & 1 & 0 & \ldots & 0 & 0 \\ 0 & -\rho & 1 & \ldots & 0 & 0 \\ . & . & . & \ldots & . & . \\ 0 & 0 & 0 & \ldots & -\rho & 1 \end{bmatrix},$$

and the vector of standardized residuals is $\tilde{e} = D^{-1/2}e$, with $\tilde{e}_1 = (\nu_1 - y_1'\beta)\sqrt{1-\rho^2}$ and $\tilde{e}_t = (\nu_t - y_t'\beta) - \rho\,(\nu_{t-1} - y_{t-1}'\beta)$, $t>1$. The nonlinear sum of squares is

$$S(\nu,\gamma) = (1/\sqrt{1-\rho^2}\,)^{1/N}\,\tilde{e}'\tilde{e}\,(1/\sqrt{1-\rho^2}\,)^{1/N} \quad . \qquad \square$$

We can concentrate γ out of $S(\nu,\gamma)$ if we replace in $S(\nu,\gamma)$ γ by its maximum likelihood estimator $\hat{\gamma}$, which is the generalized least squares (GLS) estimator of the model

$$\nu = X\gamma + \varepsilon \quad . \tag{7}$$

We now show how to obtain $\hat{\gamma}$ by means of the XKF. From what we have just seen, it is clear that the KF can be seen as an algorithm that, applied to a vector v of the same dimension than ν , yields Kv and D. The algorithm can be trivially extended to compute also $D^{1/2}$. Therefore, if we apply this extended algorithm in model (7) to the data ν and to the columns of the X matrix, we obtain $D^{-1/2}K\nu = D^{-1/2}KX\gamma + D^{-1/2}K\varepsilon$, where $\mathrm{Var}(D^{-1/2}K\varepsilon) = \sigma^2 I_M$, and we have transformed a GLS regression model (7) into an

ordinary least squares (OLS) one. The estimator $\hat{\gamma}$ can now be efficiently and accurately obtained using the QR algorithm. Supposing X of full column rank, if p is the number of components in γ, this last algorithm premultiplies both $D^{-1/2}K\nu$ and $D^{-1/2}KX$ by an orthogonal matrix Q to obtain $\omega = QD^{-1/2}K\nu$ and $(U', 0')' = QD^{-1/2}KX$, where U is a nonsingular $p \times p$ upper triangular matrix. Then $\hat{\gamma} = U^{-1}\omega_1$, where ω_1 consists of the first p elements of ω, and we can evaluate

$$S(\nu, \hat{\gamma}) = (\prod_{t=1}^{N} |D_t^{1/2}|)^{1/M} \, \omega_2'\omega_2 \, (\prod_{t=1}^{N} |D_t^{1/2}|)^{1/M},$$

where ω_2 consists of the last $M - p$ elements of ω.

DEFINITION 3. The Extended Kalman Filter (XKF) is the KF (6) with the equations for e_t and $\hat{x}_{t+1,t}$, respectively, replaced by

$$E_t = (\nu_t, 0, X_t) - C_t\hat{X}_{t,t-1}, \qquad \hat{X}_{t+1,t} = (0, 0, -W_t) + A_t\hat{X}_{t,t-1} + G_tE_t,$$

with starting condition $\hat{X}_{1,0} = (0, -A_0B, -W_0)$. Also, $D_t^{1/2}$ is computed along with D_t.

The columns of the matrix $\hat{X}_{t,t-1}$ contain the state estimates, and those of E_t the prediction errors, corresponding to the data and to the columns of the X matrix, respectively. The XKF has been suggested in [15] and [19]; it has been generalized to the case of a rank defficient X matrix in [6].

EXAMPLE 2 (Continued). Applying the XKF with starting condition $\hat{X}_{1,0} = (0, 0)$, we get

$$E_t = (\nu_t, y_t') - \hat{X}_{t,t-1}, \quad \tilde{E}_t = D_t^{-1/2}E_t, \quad \hat{X}_{t+1,t} = \rho\hat{X}_{t,t-1} + \rho E_t.$$

This implies $E_t = (\nu_t - \rho\nu_{t-1}, y_t' - \rho y_{t-1}')$, $t > 1$, and $E_1 = (\nu_1, y_1')$. The GLS model $\nu = S\beta + \varepsilon$ has been transformed into the OLS model

$$\begin{bmatrix} \nu_1\sqrt{1-\rho^2} \\ \nu_2 - \rho\nu_1 \\ \cdots \\ \nu_N - \rho\nu_{N-1} \end{bmatrix} = \begin{bmatrix} y_1'\sqrt{1-\rho^2} \\ y_2' - \rho y_1' \\ \cdots \\ y_N' - \rho y_{N-1} \end{bmatrix} \beta + \tilde{\varepsilon}, \qquad \tilde{\varepsilon} \sim N(0, \sigma^2 I). \qquad \square$$

Consider now predicting the state x_t using $(\nu_1', \nu_2', \ldots, \nu_{t-1}')'$. This is equivalent to first predicting x_t using $(\gamma', \nu_1', \nu_2', \ldots, \nu_{t-1}')'$, and then predicting this predictor using $(\nu_1', \nu_2', \ldots, \nu_{t-1}')'$. The first mentioned predictor is $\hat{x}_{t,t-1}$, as given in (6). It is easy to check that $\hat{X}_{t,t-1}(1, -\gamma')'$, where $\hat{X}_{t,t-1}$ is given by the XKF, verifies the same recursion and starting condition as $\hat{x}_{t,t-1}$ and hence $\hat{x}_{t,t-1} = \hat{X}_{t,t-1}(1, -\gamma')'$. Thus, $\tilde{x}_{t,t-1} = \hat{X}_{t,t-1}(1, -\hat{\gamma}')'$ is the predictor we are looking for. Its mean squared error (Mse) is

$$\begin{aligned}\mathrm{Mse}(\tilde{x}_{t,t-1}) &= \mathrm{Var}(x_t - \hat{x}_{t,t-1} + \hat{x}_{t,t-1} - \tilde{x}_{t,t-1}) = \mathrm{Var}(x_t - \hat{x}_{t,t-1}) + \mathrm{Var}(\hat{x}_{t,t-1} - \tilde{x}_{t,t-1}) \\ &= \sigma^2 P_{t,t-1} + \mathrm{Var}\left\{ \hat{X}_{t,t-1}(0, (\hat{\gamma} - \gamma)')' \right\} \\ &= \sigma^2 P_{t,t-1} + \hat{X}(\gamma)_{t,t-1} \mathrm{Mse}(\hat{\gamma}) \hat{X}(\gamma)'_{t,t-1},\end{aligned}$$

where $\hat{X}(\gamma)_{t,t-1}$ is the submatrix of $\hat{X}_{t,t-1}$ formed by all its columns except the first, and $\mathrm{Mse}(\hat{\gamma}) = \sigma^2 (U'U)^{-1}$. If $\tilde{v}_t$ is the predictor of v_t using $(v'_1, v'_2, \ldots, v'_{t-1})'$, then it can be shown analogously that $v_t - \tilde{v}_t = E_t(1, -\hat{\gamma}')'$, $\mathrm{Mse}(\tilde{v}_t) = \sigma^2 D_t + E(\gamma)_t \mathrm{Mse}(\hat{\gamma}) E(\gamma)'_t$, where $E(\gamma)_t$ is the submatrix of E_t formed by all its columns except the first.

The DKF of de Jong can also be used for likelihood evaluation when $C = 0$, although, as we will see, it has other uses as well.

DEFINITION 4. The Diffuse Kalman Filter (DKF) is the XKF without the computation of $D_t^{1/2}$ and with the added recursion $Q_{t+1} = Q_t + E'_t D_t^{-1} E_t$, where $Q_1 = 0$.

Given that $(v - X\hat{\gamma})'\Sigma^{-1}(v - X\hat{\gamma}) = q - s'S^{-1}s$, where $q = v'\Sigma^{-1}v$, $s = X'\Sigma^{-1}v$, and $S = X'\Sigma^{-1}X$, the Q_t matrix accumulates the partial squares and cross products and

$$Q_{N+1} = \begin{bmatrix} q & s' \\ s & S \end{bmatrix}. \tag{8}$$

Therefore, the DKF allows us to evaluate the (σ^2, γ) -maximized log-likelihood, given by

$$-\frac{1}{2}\{M \ln((q - s'S^{-1}s)/M) + \sum_{t=1}^{N} \ln|D_t|\}.$$

EXAMPLE 2 (Continued). The DKF gives, besides E_t and $\hat{X}_{t+1,t}$, computed as in the XKF, the quantities Q_t. In this case,

$$Q_{N+1} = \begin{bmatrix} (1-\rho^2)v_1^2 + \sum_{t=2}^{N} (v_t - \rho v_{t-1})^2 & * \\ (1-\rho^2)v_1 y_1 + \sum_{t=2}^{N} (v_t - \rho v_{t-1})(y_t - \rho y_{t-1}) & (1-\rho^2)y_1 y_1' + \sum_{t=2}^{N} (y_t - \rho y_{t-1})(y_t - \rho y_{t-1})' \end{bmatrix}.$$

The estimator $\hat{\gamma}$ is obtained by solving the normal equations of the regression, $S\gamma = sv$. However, to solve the normal equations this way can lead to numerical difficulties because what we are doing is basically squaring a number and then taking its square root. It is numerically more efficient to use a device such as the *QR* algorithm or the singular value decomposition, once the XKF has been applied. Another alternative, but computationally more expensive, is to use a square root filter version of the DKF.

4. Initial State with an unspecified distribution. No regression parameters

In this Section, we suppose that δ in $x_0 = B\delta$ has an unspecified distribution, that is $\delta \sim N(c, \sigma^2 C)$ with C nonsingular. We also suppose that there are no regression parameters and, therefore, $W_t = 0$, and $X_t = 0$. Then, Theorem 1 implies

$$v = R\delta + \varepsilon \ . \tag{9}$$

Ansley and Kohn [2], hereafter AK, define the likelihood of (9) by means of a transformation of the data that eliminates dependence on initial conditions. Let J be a matrix with $|J| = 1$ such that JR has exactly rank(R) rows different from zero. Such a matrix always exists. Let J_1 consist of those rank(R) rows of J corresponding to the nonzero rows of JR and let J_2 consist of the other rows of J so that $J_2 R = 0$. AK define the likelihood of (9) as the density of $J_2 v$. We will show later that, under an extra assumption, this definition does not depend on the matrix J. To evaluate the likelihood, however, and merely for algorithmic purposes, given that the transformation usually destroys the covariance structure of the data, they use an equivalent definition of the likelihood and develop what they call "modified Kalman Filter" and "modified Fixed Point Smoother" algorithms. The modified Kalman Filter is of considerable complexity, difficult to program and is less computationally efficient than the procedure in [6], when applicable, or the DKF. Also, it does not explicitly handle fixed effects and requires specialized assumptions regarding the SSM (see [14]).

Another approach to defining the likelihood of (9) is that of de Jong [13], where δ is considered diffuse by letting $C \to \infty$. In order to take this limit we need the following theorem.

THEOREM 2. *Let* $\delta \sim N(c, \sigma^2 C)$ *with* C *nonsingular. Then, the log-likelihood of* v *is*

$$\lambda(v) = -\frac{1}{2}\{\ln|C| + \ln|\sigma^2 \Sigma| + \ln|C^{-1} + R'\Sigma^{-1}R| + \left\{(\delta - c)'C^{-1}(\delta - c) + (v - R\delta)'\Sigma^{-1}(v - R\delta)\right\}/\sigma^2\},$$

where $\tilde{\delta} = (C^{-1} + R'\Sigma^{-1}R)^{-1}(C^{-1}c + R'\Sigma^{-1}v)$ *and* $\tilde{\delta}$ *coincides with the conditional expectation* $E(\delta|v)$. *Also,* $\mathrm{Mse}(\tilde{\delta}) = \mathrm{Var}(\delta|v) = \sigma^2(C^{-1} + R'\Sigma^{-1}R)^{-1}$.

PROOF. The density $p(v)$ verifies $p(\delta|v)p(v) = p(v|\delta)p(\delta)$, where the vertical bar denotes conditional distribution. The maximum likelihood estimator $\tilde{\delta}$ of δ in the left hand side of this equation must be equal to the one in the right hand side. Given that the equality between densities implies

$$(\delta - E(\delta|v))'\Omega_{\delta|v}^{-1}(\delta - E(\delta|v)) + (v - Rc)'\Omega_v^{-1}(v - Rc) = (\delta - c)'C^{-1}(\delta - c) + (v - R\delta)'\Sigma^{-1}(v - R\delta) \ ,$$

where $\Omega_{\delta|v}$ and Ω_v are the covariance matrices of $p(\delta|v)$ and $p(v)$, respectively, the left hand side is minimized for $\tilde{\delta} = E(\delta|v)$. To minimize the right hand side, consider the regression model $(c', v')' = (I, R')'\delta + v$, $v \sim N(0, \mathrm{diag}(C, \Sigma))$. Then $\tilde{\delta}$ is as asserted and $\mathrm{Var}(\delta|v) = \mathrm{Var}(\tilde{\delta}) = \sigma^2(C^{-1} + R'\Sigma^{-1}R)^{-1}$. □

THEOREM 3. *With the notation and assumption of Theorem 2, if* $R'\Sigma^{-1}R$ *is nonsingular, then, letting* $C \to \infty$ *, we have*

$$\lambda(\nu) + \frac{1}{2}\ln|C| \to -\frac{1}{2}\left\{\ln|\sigma^2\Sigma| + \ln|R'\Sigma^{-1}R| + (\nu - R\hat{\delta})'\Sigma^{-1}(\nu - R\hat{\delta})/\sigma^2\right\},$$
$$\delta \to \hat{\delta} = (R'\Sigma^{-1}R)^{-1}R'\Sigma^{-1}\nu, \quad \mathrm{Mse}(\delta) \to \mathrm{Mse}(\hat{\delta}) = \sigma^2(R'\Sigma^{-1}R)^{-1}.$$

PROOF. It is an immediate consequence of Theorem 2. □

It is shown in [13] that $\lambda(\nu) + (1/2)\ln|C|$ tends to a proper log-likelihood, called the diffuse log-likelihood. By Theorem 3, in order to compute it, all we have to do is to consider δ fixed in (9) and apply the methodology of Section 3. If the XKF is used, then, with the notation of Section 3, the results of Theorem 3 can be rewritten

$$\lambda(\nu) + \frac{1}{2}\ln|C| \to -\frac{1}{2}\left[M\ln(\sigma^2) + 2\{\sum_{t=1}^{N}\ln|D_t^{1/2}| + \ln|U|\} + w_2'w_2/\sigma^2\right],$$
$$\delta \to \hat{\delta} = U^{-1}w_1, \quad \mathrm{Mse}(\delta) \to \mathrm{Mse}(\hat{\delta}) = \sigma^2(U'U)^{-1},$$

whereas if the DKF is used, then, with the notation of (8), we obtain

$$\lambda(\nu) + \frac{1}{2}\ln|C| \to -\frac{1}{2}\left[M\ln(\sigma^2) + \sum_{t=1}^{N}\ln|D_t| + \ln|S| + (q - s'S^{-1}s)/\sigma^2\right],$$
$$\delta \to \hat{\delta} = S^{-1}s, \quad \mathrm{Mse}(\delta) \to \mathrm{Mse}(\hat{\delta}) = \sigma^2S^{-1}.$$

If S in (8) is singular, de Jong leaves the diffuse log-likelihood undefined. In order to define the limiting expressions of Theorem 3 when S is singular, we have to consider model (9) with an R matrix that is not of full column rank. Let K be a selector matrix formed by zeros and ones such that KSK' has rank equal to rank(R) and replace model (9) by

$$\nu = RK'\delta_1 + \varepsilon, \tag{10}$$

where $\delta_1 \sim N(c, \sigma^2C)$, with C nonsingular and δ_1 is the vector formed by choosing those components in δ corresponding to the selected columns RK'. This amounts to making the assumption that the other components in δ cannot be estimated from the data without further information and are assigned value zero with probability one. The next theorem generalizes the results of Theorem 2 to the case of a possibly singular S matrix.

THEOREM 4. *Suppose model (10) with the convention that if R is of full column rank, then matrix K is the identity matrix and* $\delta_1 = \delta$ *. Then, with the notation and assumptions of Theorem 3, letting* $C \to \infty$ *, we have*

$$\lambda(\nu) + \frac{1}{2}\ln|C| \to -\frac{1}{2}\left\{\ln|\sigma^2\Sigma| + \ln|KR'\Sigma^{-1}RK'| + (\nu - R\hat{\delta})'\Sigma^{-1}(\nu - R\hat{\delta})/\sigma^2\right\},$$
$$\delta \to \hat{\delta} = (R'\Sigma^{-1}R)^{-}R'\Sigma^{-1}\nu, \quad \mathrm{Mse}(\delta) \to \mathrm{Mse}(\hat{\delta}) = (R'\Sigma^{-1}R)^{-}.$$

where $(R'\Sigma^{-1}R)^{-} = K'(K(R'\Sigma^{-1}R)K')^{-1}K$ *and* $\tilde{\delta}$ *and* $\hat{\delta}$ *are interpreted as the particular maximizers obtained by making zero the elements not in* δ_1 *and* $\hat{\delta}_1$ *, respectively.*

PROOF. The only thing that needs to be proved is that $|KR'\Sigma^{-1}RK'|$ does not depend on K. This can be seen in [18, p 527]. □

The next theorem shows the relationship between the likelihood of AK and the diffuse likelihood of de Jong. When S is singular, we take as diffuse log-likelihood the one given by Theorem 4.

THEOREM 5. *Let* J *be a matrix with* $|J| = 1$ *like those used by AK to define their likelihood, and let* J_1 *and* J_2 *be the corresponding submatrices such that* $J_1R \neq 0$ *and* $J_2R = 0$ *. If* $p(\nu)$ *is the density of* ν *when* C *is nonsingular, as given by Theorem 2, and* $p(J_2\nu)$ *is the AK likelihood, then, letting* $C \to \infty$ *, we have*

$$|\sigma^2 C|^{+/2} p(\nu) \to \left\{ \Pi_+ / (2\pi)^{d/2} \right\} p(J_2\nu) \ ,$$

where Π_+ *is the product of the nonzero eigenvalues of the matrix* $R'J_1'J_1R$ *and* d *is the number of columns of* R*,* $\text{rank}(R) \leq d$ *.*

PROOF. Let J be as specified in the theorem. Then, $p(\nu) = P(J\nu)$ because $|J| = 1$. Permuting the rows of JR if necessary, we can always suppose that J_1R are the first rows of JR. This amounts to premultiply JR by a matrix P obtained from the unit matrix by performing the same permutations. Given that P is orthogonal, we can take PJ instead of J. Let K be a selector $r \times d$ matrix, where $r = \text{rank}(R)$, and consider model (10). If R is of full column rank, then $K = I_d$ and $r = d$. That the determinant $|K'R'J_1'J_1RK'|$ is equal to the product of the nonzero eigenvalues of $R'J_1'J_1R$ can be seen, for example, in [18, p 527]. Let $J_1RK' = M$. If $\Sigma_J = J\Sigma J'$ and we partition $\Sigma_J = (\Omega_{ij})$, $\Sigma_J^{-1} = (\Omega^{ij})$, $i,j = 1,2$, conforming to $J = (J_1', J_2')'$, then, by Theorem 4, the log-likelihood of ν verifies

$$\lambda(\nu) + \frac{1}{2}\ln|C| \to -\frac{1}{2}\left\{ \ln|\sigma^2 \Omega_{22}(\Omega^{11})^{-1}| + \ln|M'\Omega^{11}M| + (J_2\nu)'\Omega_{22}^{-1}(J_2\nu)/\sigma^2 \right\} \ .$$

□

Ansley and Kohn [2], make the following assumption.

Assumption A. Matrix R in (9) and (10) does not depend on the model parameters.

This assumption holds in many practical situations, including the examples of Section 2.

COROLLARY 1. *If Assumption A holds, then the AK likelihood does not depend on the matrix* J*.*

PROOF. It is an immediate consequence of Theorem 5. □

Even if Assumption A holds, Theorem 5 shows that the diffuse log-likelihood and the AK log-likelihood, when maximized with respect to σ^2 , do not give the same results. The

difference lies in the term $M\ln(\hat{\sigma}^2)$ in the σ^2 -maximized diffuse log-likelihood versus $(M-d)\ln(\tilde{\sigma}^2)$ in the AK log-likelihood, where $\hat{\sigma}^2 = (1/M)(q - s'S^{-}s)$ and $\tilde{\sigma}^2 = (1/(M-d))(q - s'S^{-}s)$. This is a consequence of de Jong taking the limit of $\lambda(v) + (1/2)\ln|C|$ instead of $\lambda(v) + (1/2)\ln|\sigma^2 C|$ in Theorem 4. We think that it would have been more appropiate to do the latter than the former. For instance, when dealing with an ARIMA model, the AK likelihood coincides with the usual Box-Jenkins likelihood (see [4]), whereas the diffuse likelihood does not.

We now consider predicting the state x_t and v_t using $(v_1', v_2', \dots, v_{t-1}')'$. The next two theorems give the details.

THEOREM 6. *Let* $\delta \sim N(c, \sigma^2 C)$ *with C nonsingular and let* $\hat{\delta}_t$, $\tilde{x}_t$ *and* $\tilde{v}_t$ *be the predictors of* δ, x_t *and* v_t *using* $(v_1', v_2', \dots, v_{t-1}')'$ *, respectively. Suppose the XKF or the DKF is applied and let* $\hat{X}(\delta)_{t,t-1}$ *and* $E(\delta)_t$ *be the submatrices formed by all but the first columns of* $\hat{X}_{t,t-1}$ *and* E_t *, respectively. Then*

$$\hat{\delta}_t = (C^{-1} + R_t'\Sigma_t^{-1}R_t)^{-1}(C^{-1}c + R_t'\Sigma_t^{-1}v_t) \ , \ \mathrm{Mse}(\hat{\delta}_t) = \sigma^2(C^{-1} + R_t'\Sigma_t^{-1}R_t)^{-1},$$

$$\tilde{x}_{t,t-1} = \hat{X}_{t,t-1}(1, -\hat{\delta}_t')' \ , \ \mathrm{Mse}(\tilde{x}_{t,t-1}) = \sigma^2 P_{t,t-1} + \hat{X}(\delta)_{t,t-1}\mathrm{Mse}(\hat{\delta}_t)\hat{X}(\delta)'_{t,t-1} \ ,$$

$$v_t - \tilde{v}_t = E_t(1, -\hat{\delta}_t')', \quad \mathrm{Mse}(\tilde{v}_t) = \sigma^2 D_t + E(\delta)_t\mathrm{Mse}(\hat{\delta}_t)E(\delta)'_t,$$

where R_t *is the submatrix formed by the first t rows of R and* $\Sigma_t = \sigma^{-2}\,Var((\varepsilon_1', \dots, \varepsilon_t')')$.

PROOF. The first two equalities are a consequence of Theorem 2. The other expressions can be proved as the corresponding ones for the case $C = 0$ in Section 3 . □

THEOREM 7. *With the notation and assumption of Theorems 6 and 4, if the rows of* $\hat{X}(\delta)_{t,t-1}$ *and* $E(\delta)_t$ *are in the space generated by the rows of* $(R_t'\Sigma_t^{-1}R_t)^{-}$ *, then, letting* $C \to \infty$ *,*

$$\hat{\delta}_t \to \hat{\delta}_t = (R_t'\Sigma_t^{-1}R_t)^{-}R_t'\Sigma_t^{-1}v_t \ , \ \mathrm{Mse}(\hat{\delta}_t) \to \mathrm{Mse}(\hat{\delta}_t) = \sigma^2(R_t'\Sigma_t^{-1}R_t)^{-},$$

$$\tilde{x}_{t,t-1} \to \hat{X}_{t,t-1}(1, -\hat{\delta}_t')' \ , \ \mathrm{Mse}(\tilde{x}_{t,t-1}) \to \sigma^2 P_{t,t-1} + \hat{X}(\delta)_{t,t-1}\mathrm{Mse}(\hat{\delta}_t)\hat{X}(\delta)'_{t,t-1} \ ,$$

$$\tilde{v}_t \to C_t\hat{X}_{t,t-1}(1, -\hat{\delta}_t')', \quad \mathrm{Mse}(\tilde{v}_t) \to \sigma^2 D_t + E(\delta)_t\mathrm{Mse}(\hat{\delta}_t)E(\delta)'_t.$$

PROOF. The first two limits are a consequence of Theorems 2 and 4. The other expressions are a direct consequence of Theorem 6. □

By Theorem 7, in order to get the desired predictors, we must consider the regression model (9) with δ fixed and apply GLS theory. We can use the results of Section 3 and, in order to get an afficient algorithm, we can apply the XKF or the DKF for likelihood evaluation or prediction. Note that the difficulties that may arise stem from the fact that the matrix R may be rank defficient. In this last case, we have to use generalized inverses throughout the process and neither all observations will be predictable, nor will all states be

estimable.

The next theorem states that the predictors obtained with the modified Kalman Filter coincide with those obtained by means of the XKF or the DKF.

THEOREM 8. *Let Assumption A hold. Then, the predictors of* x_t *and* v_t *obtained with the modified Kalman Filter and those obtained with the XKF or the DKF coincide. If the same estimator of* σ^2 *is used in both procedures, then the Mse errors also coincide.*

PROOF. Theorem 5.2 in [2] states that the AK predictors coincide with the diffuse predictors and the statement about the Mse follows trivially. □

We have seen that, in order to evaluate the AK log-likelihood, we can use the modified Kalman Filter of AK, although it is not the best procedure, or we may use the efficient XKF or DKF to evaluate the diffuse log-likelihood, which, by Theorem 5, differs from the AK log-likelihood only in a constant. This constant, under Assumption A, does not depend on model parameters. The XKF or the DKF should be applied to model (9) considering δ fixed (C = 0). It would be nice to employ the XKF or the DKF only for an initial stretch of the data, as short as possible, to construct an estimator of δ and, from then on, use the KF. When this occurs, one speaks of a collapse of the XKF or the DKF to the KF. Let rank(R) = r and suppose that the first r rows of R are linearly independent. Let R_I be the submatrix formed by the first r rows and let R_{II} consist of the other rows of R. Partition $v = (v_I', v_{II}')'$ and $\varepsilon = (\varepsilon_I', \varepsilon_{II}')'$ conforming to $R = (R_I', R_{II}')'$. Then, we can write

$$v_I = R_I\delta + \varepsilon_I \tag{11a}$$
$$v_{II} = R_{II}\delta + \varepsilon_{II} . \tag{11b}$$

The next theorem shows how to implement the collapsing of the XKF or DKF to the KF.

THEOREM 9. *Under Assumption A, let* J *with* $|J| = 1$ *be a matrix like those used by AK to define their likelihood, with corresponding submatrices* J_1 *and* J_2 *such that* $J_1R \neq 0$ *and* $J_2R = 0$, *and let* $p(v_{II}|v_I, \hat{\delta}_I)$ *be the density of* $v_{II} - E(v_{II}|v_I, \hat{\delta}_I)$, *where* $E(v_{II}|v_I, \hat{\delta}_I)$ *is the conditional expectation of* v_{II} *given* v_I *in model* (11a) *and* (11b), *considering* δ *fixed* (C = 0), *and* δ *replaced by its maximum likelihood estimator* $\hat{\delta}_I$ *in model* (11a). *Then,*

$$p(J_2 v) = p(v_{II}|v_I, \hat{\delta}_I) ,$$

where $p(J_2 v)$ *is the density of* $J_2 v$.

PROOF. For simplicity, consider that R_I is of full column rank. If not, we would use generalized inverses, but the proof would not be affected.From model (11a) and (11b), we have, considering δ fixed, $E(v_{II}|v_I) = R_{II}\delta + \Sigma_{21}\Sigma_{11}^{-1}(v_I - R_I\delta)$, where $\Sigma_{21} = \mathrm{Cov}(\varepsilon_{II}, \varepsilon_I)$ and $\Sigma_{11} = \mathrm{Var}(\varepsilon_I)$. Then, $E(v_{II}|v_I, \hat{\delta}_I) = R_{II}R_I^{-1}v_I$ and $v_{II} - E(v_{II}|v_I, \hat{\delta}_I) = v_{II} - R_{II}R_I^{-1}v_I$. Define the matrix $J = (J_1', J_2')'$ with $J_1 = (I, 0)$ and $J_2 = (-R_{II}R_I^{-1}, I)$. Then, J is a matrix of the type used by AK to define their likelihood and $v_{II} - E(v_{II}|v_I, \hat{\delta}_I) = J_2 v$. □

By Theorem 9, to evaluate the log-likelihood of $v_{II} - E(v_{II}|v_I, \delta_I) = J_2 v$, we can proceed as follows. First, use the XKF or the DKF in model (11a) to obtain the maximum likelihood estimator (mle) $\hat{\delta}_I$ and initial conditions for the KF

$$E(x_s) = \hat{X}_{s,s-1}(1, -\delta_I')', \quad Var(x_s) = \sigma^2 P_{s,s-1} + \hat{X}(\delta)_{s,s-1} Mse(\hat{\delta}_I) \hat{X}(\delta)'_{s,s-1},$$

where v_s is the first observation in (11b). Then, proceed with the KF, applied to the second stretch of the data v_{II}, to obtain the log-likelihood of $v_{II} - E(v_{II}|v_I, \delta_I) = J_2 v$ as in Section 3, but with no regression parameters. We have used the initial stretch v_I of the data to construct the mle $\hat{\delta}_I$ and the initial conditions for the KF. After that, the effect of δ has been absorbed into the estimator of the state vector and that is the reason why we can collapse the XKF or DKF to the KF.

We now give another interpretation to the result of Theorem 9. With the notation of Theorem 2, we can write

$$\lambda(v) + \ln|C|/2 = \{\lambda(v_I) + \ln|C|/2\} + \lambda(v_{II}|v_I), \tag{12}$$

where $\lambda(v_{II}|v_I)$ is the conditional log-likelihood of v_{II} given v_I. By Theorem 3, letting $C \to \infty$, the term in curly brackets tends to $\{M_I \ln(\sigma^2) + |R_I' R_I|\}/2$, where M_I is the number of components in v_I, whereas $\lambda(v_{II}|v_I)$ converges to the log-likelihood of $J_2 v$, which, by Theorem 9, is equal to the log-likelihood of $v_{II} - E(v_{II}|v_I, \delta_I)$. Thus, the diffuse log-likelihood of v is the sum of the diffuse log-likelihood of v_I and the log-likelihood of $J_2 v$. Note that the first term does not contribute to either the determinant or the sum of squares of the diffuse log-likelihood.

Bell and Hillmer [3] use a similar idea to construct initial conditions for the KF. Instead of employing the KF for the initial stretch of the data, they use the transformation approach of AK to construct the mle $\hat{\delta}_I$ and the initial conditions for the KF directly. Whether this approach is more advantageous than using the KF, is something that depends on the pecularities of the problem at hand. If it is easier to obtain the mle and the initial conditions directly, then it can be used. However, the KF approach to construct the mle and the initial conditions has the advantage that it is easy to implement, does not depend on ad hoc procedures and it imposes very little computational and/or programming burden.

The case we have been considering, where the submatrix R_I is formed by the first r rows of R is important because it happens often in practice. Examples of this are ARIMA models and ARIMA component models.

If in model (11a) and (11b) we have $v_I = \delta$, then Theorem 9 implies $p(J_2 v) = p(v_{II}|v_I)$ Also, if J is a matrix like those used by AK to define their likelihood and J is of the form $J = (J_1', J_2')'$ with $J_1 = (I, 0)$, then v_I is independent of $J_2 v$. This is the conditional likelihood approch used in [6] in the context of regression models with ARIMA disturbances and generalized to the case when there are missing observations. For ARIMA (p, d, q) models, the situation simplifies still further because it is not necessary to employ the XKF or the DKF for the initial stretch of the data $v_I = \delta$ to obtain initial conditions for the KF. The SSM can be redefined by simply translating forward the initial conditions d units in time, where d is the degree of the differencing operator.

Suppose there are missing observations in ν and that in (11a) the vector ν_1 contains a subvector ν_{1M} of missing observations. Let ν_{1O} be the subvector of ν_1 formed by the nonmissing observations and let ν_{II} be the subvector of ν containing the rest of the nonmissing observations. Then, by analogy with the result of Theorem 9, we can still consider $\nu_{II} - E(\nu_{II}|\nu_1, \delta_1)$, treat ν_{1M} as a vector of fixed parameters, and define the likelihood of $(\nu'_{1O}, \nu'_{II})'$ as that of the regression model

$$\eta_{II} = R_{II}U_M\nu_{1M} + \omega_{II}, \tag{13}$$

where $\eta_{II} = \nu_{II} - R_{II}U_O\nu_{1O}$, $\omega_{II} = \varepsilon_{II} - R_{II}R_1^{-1}\varepsilon_1$, and U_O and U_M are the submatrices of R_1^{-1} formed by the columns corresponding to ν_{1O} and ν_{1M}, respectively. Here we have supposed that R_1 is of full column rank. If not, we would use generalized inverses, but the main result would not be affected. Note that the vector ν_{1M} is considered as a vector of fixed parameters that have to be estimated along with the other parameters of the model. The next theorem shows that this definition of the likelihood is equivalent to the AK definition.

THEOREM 10. *Under Assumption A, the* (σ^2, ν_{1M}) *- maximized log-likelihood corresponding to* (13) *coincides, up to a constant, with the* σ^2 *- maximized AK log-likelihood.*

PROOF. Let $\mathrm{Var}(\omega_{II}) = \sigma^2\Omega$ and let $\Omega = LL'$ be the Cholesky decomposition of Ω. If we premultiply (13) by L^{-1}, we obtain the OLS model

$$L^{-1}\eta_{II} = L^{-1}R_{II}U_M\nu_{1M} + L^{-1}\omega_{II}.$$

The *QR* algorithm, applied to the $L^{-1}R_{II}U_M$ matrix, yields an upper triangular matrix S with nonzero elements in the main diagonal such that $Q'L^{-1}R_{II}U_M = (S', 0)'$, where Q is an orthogonal matrix. Then, we can write

$$Q'L^{-1}\eta_{II} = \begin{bmatrix} S \\ 0 \end{bmatrix}\nu_{1M} + Q'L^{-1}\omega_{II}.$$

The matrix L^{-1} will not have, in general, unit determinant. If we multiply L^{-1} by $\alpha = |L|^{1/M_{II}}$, where M_{II} is the number of components in ν_{II}, then αL^{-1} has unit determinant. Let $K = (K'_1, K'_2)'$ with $K_1 = (I, 0)$ and $K_2 = (-R_{II}U_O, I)$. and let $P = (P'_1, P'_2)'$ with $P_1 = (I, 0)$ and $P_2 = (0, \alpha Q'L^{-1})$. Partition $Q' = (Q_1, Q_2)'$ conforming to $Q'_1L^{-1}R_{II}U_M = S$ and $Q'_2L^{-1}R_{II}U_M = 0$. If $J = PK$, then J has unit determinant and

$$\begin{aligned} J(\nu'_{1O}, \nu'_{II})' &= (\nu'_{1O}, (\alpha Q'L^{-1}\eta_{II})')' \\ &= (R'_{1O}, (\alpha SR_{1M})', 0)'\delta + (\varepsilon'_{1O}, (\varepsilon_{1M} + \alpha Q'_1L^{-1}\omega_{II})', (\alpha Q'_2L^{-1}\omega_{II})')', \end{aligned}$$

where we have partitioned R_1 and ε_1 conforming to the partition of ν_1 into ν_{1O} and ν_{1M}. Given that $(R'_{1O}, (\alpha SR_{1M})')'$ has rank equal to that of R_1, the matrix J is of the AK type. Therefore, the AK likelihood is the density of $\alpha Q'_2L^{-1}\eta_{II}$ and the AK log-likelihood,

maximized with respect to σ^2 , is

$$-\frac{1}{2}\left\{(M_{II}-r_s)\ln(\hat{\sigma}^2)+\ln|L|^{2(M_{II}-r_s)/M_{II}}\right\}$$

where $\hat{\sigma}^2 = (1/(M_{II}-r_s))\eta'_{II}(L^{-1})'Q_2Q_2'L^{-1}\eta_{II}$ and r_s =rank(S). The log-likelihood of (13), maximized with respect to σ^2 and v_{1M} is $-\left\{M_{II}\ln(\tilde{\sigma}^2)+\ln|L|^2\right\}/2$, where

$$\tilde{\sigma}^2 = (1/M_{II})\eta'_{II}(L^{-1})'Q_2Q_2'L^{-1}\eta_{II} \ . \qquad \square$$

Theorem 10 generalizes the result obtained in [6] for ARIMA models with missing data. This approach is useful when the matrix R_1 corresponding to the first observations v_1 (included the missing ones) is of full column rank.

We now suppose that in model (9) the first r rows of R do not, in general, constitute a submatrix of R of rank r. Let R_1 be the first submatrix of R formed adjoining consecutive rows to the first row, such that it has full column rank and let R_{II} consist of the other rows of R. Partition $v = (v_1', v_{II}')'$ and $\varepsilon = (\varepsilon_1', \varepsilon_{II}')'$ conforming to $R = (R_1', R_{II}')'$. In the rest of the section, whenever we refer to models (11a) and (11b), we will refer to this partition. Consider the decomposition given by (12). Then, letting $C \to \infty$ as before, the term in curly brackets tends to

$$-\frac{1}{2}\{M_1\ln(\sigma^2)+|\Sigma_{11}|+|R_1'\Sigma_{11}^{-1}R_1|+(v_1-R_1\delta_1)'\Sigma_{11}^{-1}(v_1-R_1\delta_1)/\sigma^2\},$$

where $\mathrm{Var}(\varepsilon_1)=\sigma^2\Sigma_{11}$ and $\hat{\delta}_1=(R_1'\Sigma_{11}^{-1}R_1)^{-1}R_1'\Sigma_{11}^{-1}v_1$. The conditional log-likelihood $\lambda(v_{II}|v_1)$ converges to the log-likelihood of J_2v , where $J_2=(-R_{II}S_1^{-1}T_1, I)$, $S_1=(R_1'\Sigma_{11}^{-1}R_1)^{-1}$ and $T_1=R_1'\Sigma_{11}^{-1}$. To see this, define $J=(J_1', J_2')'$ with $J_1=(I,0)$. Then, $\lambda(v)=\lambda(Jv)$ because J has unit determinant and

$$\lambda(v)+\frac{1}{2}\ln|C| = \{\lambda(v_1)+\frac{1}{2}\ln|C|\}+\lambda(J_2v|v_1).$$

Note that now J is not a matrix of the AK type.

THEOREM 11. *Let J be the matrix we have just defined , with corresponding submatrices J_1 and J_2 , and let $p(v_{II}|v_1,\delta_1)$ be the density of $v_{II}-E(v_{II}|v_1,\delta_{II})$, where $E(v_{II}|v_1,\delta_1)$ is the conditional expectation of v_{II} given v_1 in model* (11a) *and* (11b), *considering δ fixed* ($C=0$) *and replaced by its maximum likelihood estimator $\hat{\delta}_1$ in model* (11a). *Then,*

$$p(J_2v) = p(v_{II}|v_1,\hat{\delta}_1) \ ,$$

where $p(J_2v)$ is the density of J_2v .

PROOF. The proof is analogous to that of Theorem 9. $\square$

Thus, to evaluate the AK log-likelihood or the diffuse log-likelihood, we can still use the XKF or the DKF as before, until we have processed a stretch of observations such that the corresponding submatrix of R has full column rank, and then collapse to the KF. The

likelihood is evaluated as the sum of two terms. One corresponding to the stretch v_I and the other corresponding to v_{II} . More specifically, the XKF or the DKF applied to model (11a) yields

$$|\Sigma_{11}| \ , \ |R_I'\Sigma_{11}^{-1}R_I| \text{ and } (v_I - R_I\hat{\delta}_I)'\Sigma_{11}^{-1}(v_I - R_I\hat{\delta}_I) \ , \tag{14}$$

where $\hat{\delta}_I = (R_I'\Sigma_{11}^{-1}R_I)^{-1}R_I'\Sigma_{11}^{-1}v_I$. These three terms will be needed for the computation of the likelihood because now there will be no cancelation of terms. With the notation of Section 3, if the XKF is used, the expressions in (14) are

$$\ln|\Sigma_{11}| = 2\sum_{t=1}^{N_I} \ln|D_t^{1/2}| \ , \ |R_I'\Sigma_{11}^{-1}R_I| = |U_I|^2 \text{ and } (v_I - R_I\hat{\delta}_I)'\Sigma_{11}^{-1}(v_I - R_I\hat{\delta}_I) = w_{I,2}'w_{I,2} \ ,$$

whereas, if the DKF is used, they are

$$\ln|\Sigma_{11}| = \sum_{t=1}^{N_I} \ln|D_t| \ , \ |R_I'\Sigma_{11}^{-1}R_I| = |S_I| \text{ and } (v_I - R_I\hat{\delta}_I)'\Sigma_{11}^{-1}(v_I - R_I\hat{\delta}_I) = q_I - s_I'S_I^{-1}s_I \ .$$

The initialization for the KF, to be used with the second stretch of the data v_{II} , is

$$E(x_s) = \hat{X}_{s,s-1}(1, -\hat{\delta}_I')' \ , \quad Var(x_s) = \sigma^2 P_{s,s-1} + \hat{X}(\delta)_{s,s-1}Mse(\hat{\delta}_I)\hat{X}(\delta)'_{s,s-1} \ ,$$

where, as before, v_s is the first observation in (11b). Once the run of the KF is completed, we have to add up the terms in (14) to the corresponding terms obtained with the KF, $|Var(J_2v)|$ and $(J_2v)'(Var(J_2v))^{-1}(J_2v)$.

The fact that we don't know for how long we will have to use the XKF or the DKF before we make the transition to the KF may make collapsing unattractive. There is an alternative procedure to evaluate the AK log-likelihood or the diffuse log-likelihood that might be of interest in some cases. It consists essentially of reshuffling the observations in such a way that again the first r rows of R are linearly independent. An algorithm to achieve this is the following. Apply the XKF or the DKF to model (9) and, at the same time, obtain the row echelon form of the R matrix. Each time a new observation v_t is being incorporated, we check whether its corresponding row vector R_t is a linear combination of the rows already processed. If it is, we skip this observation as if it were missing (see [10]). Otherwise, we process the observation as part of the initial stretch of the data v_I . Proceeding in this way, after some time we will have processed a stretch of the data v_I for which the corresponding submatrix R_I of R will be formed by a maximal set of linearly independent row vectors. Let v_{II} consist of the other observations and let v_s be the first observation that we skip as if it were missing. This will be the first observation of v_{II} . Suppose the Fixed Point Smoother (FPS) corresponding to v_s is applied, along with the XKF or the DKF, to all the columns of $\hat{X}_{t+1,t}$. Then, after processing v_I , we can set up as initial conditions for the KF, to be applied to v_{II} , the following

$$E(x_s) = \hat{X}_{s,I}(1, -\hat{\delta}_I')' \ , \quad Var(x_s) = \sigma^2 P_{s,I} + \hat{X}(\delta)_{s,I}Mse(\hat{\delta}_I)\hat{X}(\delta)'_{s,I} \ ,$$

where $\hat{X}_{s,I}$, $P_{s,I}$ and $\hat{X}(\delta)_{s,I}$ are the quantities obtained with the FPS, and $\hat{\delta}_I$ is the mle corresponding to v_I . Note that the advantage of using only the KF for likelihood evaluation

comes at the expense of an increase in the computations.

EXAMPLE 3. Consider the following ARIMA (1,1,0) model

$$(1+\phi L)\nabla v_t = a_t \ ,$$

where the notation is as in Example 1. To obtain a SSM formulation, we define $X_t = 0$, $C_t = C = (1, 0)$, $Z_t = 0$, $W_t = 0$,

$$A_t = \begin{bmatrix} 0 & 1 \\ \phi & 1-\phi \end{bmatrix} = A \ ,$$

$H_t = H$, with $H_1 = 1$, $H_2 = 1-\phi$, $x_{1,t} = v_t$, $x_{2,t} = v_{t+1} - a_{t+1}$ and $\xi_{t-1} = a_t$. Then, we can write

$$x_t = Ax_{t-1} + Ha_t \ ; \quad v_t = Cx_t \ .$$

To initialize, we consider that $(1-L)v_t = u_t$ is stationary and follows the model $(1+\phi L)u_t = a_t$. Then,

$$x_1 = \begin{bmatrix} 1 \\ 1 \end{bmatrix} v_0 + \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -\phi \end{bmatrix} u_1 \ ,$$

and we can choose $A_0 = I$, $B = (1,1)'$, $x_0 = B\delta$, $\delta = v_0$ and

$$H_0 = \begin{bmatrix} 1 & 0 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1 \\ -\phi \end{bmatrix} 1/\sqrt{1-\phi^2} \ .$$

The first state is $x_1 = B\delta + H_0 a_1$. Model (9) specializes to $R = (1, 1, \ldots, 1)'$ and $\varepsilon_t = u_1 + \ldots + u_t$, $t = 1, \ldots, N$. The AK likelihood can be obtained as the density of the differenced data. This is equivalent to multiply v by the matrix

$$J = \begin{bmatrix} 1 & 0 & 0 & \ldots & 0 & 0 \\ -1 & 1 & 0 & \ldots & 0 & 0 \\ 0 & -1 & 1 & \ldots & 0 & 0 \\ . & . & . & \ldots & . & . \\ 0 & 0 & 0 & \ldots & -1 & 1 \end{bmatrix} ,$$

define $J_1 = (1, 0, \ldots, 0)$ and J_2 such that $J = (J_1', J_2')'$, and take as AK density the density of $J_2 v$. Note that $JR = (1, 0, \ldots, 0)'$ and $J\varepsilon = (u_1, u_2, \ldots, u_N)'$. The XKF or the DKF

produces $D_1 = 1/(1-\phi^2)$, $D_t = 1$, $t = 2,\ldots,N$,

$$\ln|\Sigma| = \ln|J\Sigma J'| = \ln|D_1| + \ldots + \ln|D_N| = \ln(1/(1-\phi^2)),$$

$$\ln|R'\Sigma^{-1}R| = \ln|R'J'(J\Sigma J')^{-1}JR| = 0 \ , \ \delta = v_1,$$

$$(v - R\delta)'\Sigma^{-1}(v - R\delta) = (v_2 + \phi v_1 - v_1)^2 + \sum_{t=3}^{N}\left[(v_t + \phi v_{t-1}) - (v_{t-1} + \phi v_{t-2})\right]^2 .$$

In case the DKF is applied,

$$Q_{N+1} = \begin{bmatrix} (1-\phi^2)v_1^2 + (v_2 + \phi v_1 - v_1)^2 + \sum_{t=3}^{N}\left[(v_t + \phi v_{t-1}) - (v_{t-1} + \phi v_{t-2})\right]^2 & (1-\phi^2)v_1' \\ (1-\phi^2)v_1 & (1-\phi^2) \end{bmatrix},$$

Model (11a) becomes the first equation of (9), $v_1 = \delta + u_1$, where $v_1 = v_1$, $R_1 = 1$ and $\varepsilon_1 = u_1$. Model (11b) consists of the rest of the equations. Suppose we use the XKF or the DKF in (11a) to obtain $\hat{\delta}_1$ and initial conditions for the KF, that we will apply later to model (11b). Then,

$$\hat{X}_{1,0} = \begin{bmatrix} 0 & -1 \\ 0 & -1 \end{bmatrix}, \quad E_1 = (v_1, 1), \quad P_{1,0} = \frac{1}{1-\phi^2}\begin{bmatrix} 1 & 1-\phi \\ 1-\phi & (1-\phi)^2 \end{bmatrix}, \quad G_1 = \begin{bmatrix} 1-\phi \\ \phi + (1-\phi)^2 \end{bmatrix},$$

$$D_1 = \frac{1}{1-\phi^2}, \quad \hat{X}_{2,1} = \begin{bmatrix} (1-\phi)v_1 & -\phi \\ (1-\phi+\phi^2)v_1 & \phi(\phi-1) \end{bmatrix}, \quad P_{2,1} = \begin{bmatrix} 1 & 1-\phi \\ 1-\phi & (1-\phi)^2 \end{bmatrix},$$

Given that $\delta = v_1$, we have $\mathrm{Mse}(\hat{\delta}_1) = 1/(1-\phi^2)$ and the initial conditions for the KF are

$$E(x_2) = \hat{X}_{2,1}\begin{bmatrix} 1 \\ -\hat{\delta}_1 \end{bmatrix} = \begin{bmatrix} v_1 \\ v_1 \end{bmatrix}, \quad \mathrm{Var}(x_2) = P_{2,1} + \hat{X}_{2,1}(\delta)\mathrm{Mse}(\hat{\delta}_1)\hat{X}_{2,1}'(\delta) = \frac{1}{1-\phi^2}\begin{bmatrix} 1 & 1-\phi \\ 1-\phi & (1-\phi)^2 \end{bmatrix}.$$

Therefore, using the XKF or the DKF in (11a) to estimate δ and to compute initial conditions for the KF yields the same starting values, but shifted ahead one period of time. This is an example where we can redefine the SSM, taking $v_1 = \delta$ and translate the initial conditions forward one unit of time. This is true for all ARIMA (p, d, q) models (see [6]).

5. Initial State with an unspecified distribution. The general case

In this Section we consider a more general SSM than that of Section 3. Besides making the assumption that δ in $x_0 = B\delta$ has an unspecified distribution, $\delta \sim N(c, \sigma^2 C)$, with C

nonsingular, we allow for regression parameters. That is, we consider β fixed but unknown. By Theorem 1, we have $v = R\delta + S\beta + \varepsilon$ .Defining $X = (R, S)$ and $\gamma = (\delta', \beta')'$, we can write the model more concisely as

$$v = X\gamma + \varepsilon \ . \tag{15}$$

To define the AK likelihood, consider a matrix J of the type used by AK when there are no regression parameters and let J_1 and J_2 be the corresponding submatrices such that $J_1 R \neq 0$ and $J_2 R = 0$. Then, the AK likelihood is the density of $J_2(v - S\beta)$. In order to efficiently evaluate the likelihood and predict and interpolate unobserved $v_t's$, they use their modified KF and modified FPS, applying them also to the columns of the regression matrix, as outlined in Section 3. The reader is referred to [15] and [16] for details. For the reasons mentioned in Section 4, we consider the modified KF and modified FPS computationally less efficient and conceptually more complex than the XKF or the DKF.

To compute the diffuse log-likelihood of (15) we have to consider that δ is diffuse, $C \to \infty$, and β is fixed. De Jong does not consider explicitly this case, although it is a case that is often encountered in practice. Proceeding as in Theorems 2 and 3 of Section 4, replacing v by $v - S\beta$, and letting $C \to \infty$, we have

$$\lambda(v) + \frac{1}{2}\ln|C| \to -\frac{1}{2}\left\{\ln|\sigma^2\Sigma| + \ln|R'\Sigma^{-1}R| + (v - S\beta - R\hat{\delta})'\Sigma^{-1}(v - S\beta - R\hat{\delta})/\sigma^2\right\},$$
$$\delta \to \hat{\delta} = (R'\Sigma^{-1}R)^{-1}R'\Sigma^{-1}(v - S\beta).$$

Minimizing this diffuse log-likelihood with respect to β yields an estimator $\hat{\beta}$ which minimizes $(v - S\beta)'P'\Sigma^{-1}P(v - S\beta)$, where $P = I - R(R'\Sigma^{-1}R)^{-1}R'\Sigma^{-1}$. It can be shown that the estimators $\hat{\delta}$ and $\hat{\beta}$ obtained in this way can be obtained in a single stage as the GLS estimator $\hat{\gamma} = (\hat{\delta}', \hat{\beta}')'$ of model (15). Thus, the XKF or the DKF can be used to compute the (σ^2, γ)- maximized diffuse log-likelihood, given by

$$-\frac{1}{2}\left\{M\ln(\hat{\sigma}^2) + \ln|\Sigma| + \ln|R'\Sigma^{-1}R|\right\} \ ,$$

where M is the number of components in v and $\hat{\sigma}^2 = (1/M)(v - X\hat{\gamma})'\Sigma^{-1}(v - X\hat{\gamma})$. Under Assumption A of Section 4, the AK (σ^2, γ)- maximized log-likelihood differs from the (σ^2, γ)- maximized diffuse log-likelihood only in a constant. As in Section 4, it is possible to employ the XKF or the DKF for an initial stretch of the data to construct an estimator of δ . However, it will not be possible now to collapse to the KF because we will still have to estimate the β parameters. The most we can do is to collapse to a reduced dimension XKF or DKF. More specifically, let rank$(R) = r$ and suppose that the first r rows of R are linearly independent. Let R_I be the submatrix formed by the first r rows and let R_{II} consist of the other rows of R. Partition $v = (v_I', v_{II}')'$, $S = (S_I', S_{II}')'$ and $\varepsilon = (\varepsilon_I', \varepsilon_{II}')'$ conforming to $R = (R_I', R_{II}')'$. Then, we can write

$$v_I = R_I\delta + S_I\beta + \varepsilon_I \tag{16a}$$

$$v_{II} = R_{II}\delta + S_{II}\beta + \varepsilon_{II}. \tag{16b}$$

Suppose that R_l has full column rank. If not, we would use generalized inverses instead of true inverses but the main result would not be affected. As in Section 4, we will apply first the XKF or the DKF to (16a) to obtain a GLS estimator $\hat{\delta}_l$ of δ. However it will not be possible now to absorb both δ and β into the state estimator. Only δ will be absorbed. In this way, the XKF or the DKF will only be simplified, not collapsed to the KF, when we apply it to (16b) in the second step of the procedure. The number of states of the XKF will be reduced by a number equal to the number of components in δ. Let ν_s be the first observation in (16b). We showed in Section 3 that, if δ and β are known, then the estimator of the state x_s using $(\gamma', \nu_1', \nu_2', \ldots, \nu_{s-1}')'$ is

$$\hat{x}_{s,s-1} = \hat{X}_{s,s-1}(1, -\gamma')' = \hat{X}(\nu)_{s,s-1} - \hat{X}(\delta)_{s,s-1}\delta - \hat{X}(\beta)_{s,s-1}\beta \quad , \tag{17}$$

where $\hat{X}(\nu)_{s,s-1}$, $\hat{X}(\delta)_{s,s-1}$ and $\hat{X}(\beta)_{s,s-1}$ are the columns of $\hat{X}_{s,s-1}$ corresponding to ν_s, δ and β, respectively. The GLS estimator $\hat{\delta}_l$ of δ obtained from (16a) is

$$\hat{\delta}_l = S^{-1}T(\nu_l - S_l\beta) \quad ,$$

where $S = R_l'\Omega_l^{-1}R_l$, $T = R_l'\Omega_l^{-1}$ and $\mathrm{Var}(\varepsilon_l) = \sigma^2\Omega_l$. Substituting $\hat{\delta}_l$ back in (17) yields

$$\begin{aligned}\tilde{x}_{s,s-1} &= \hat{X}(\nu)_{s,s-1} - \hat{X}(\delta)_{s,s-1}S^{-1}T\nu_l - (\hat{X}(\beta)_{s,s-1} - \hat{X}(\delta)_{s,s-1}S^{-1}TS_l)\beta \\ &= \tilde{X}(\nu)_{s,s-1} - \tilde{X}(\beta)_{s,s-1}\beta \,,\end{aligned}$$

where $\tilde{x}_{s,s-1}$ is the estimator of x_s using $(\beta', \nu_1', \nu_2', \ldots, \nu_{s-1}')'$ and $(\tilde{X}(\nu)_{s,s-1}, \tilde{X}(\beta)_{s,s-1})$ are the estimators, respectively, of the states corresponding to the data and the β parameters. Given that

$$\mathrm{Mse}(\tilde{x}_{s,s-1}) = \mathrm{Var}(x_s - \hat{x}_{s,s-1} + \hat{x}_{s,s-1} - \tilde{x}_{s,s-1}) = \mathrm{Var}(x_s - \hat{x}_{s,s-1}) + \mathrm{Var}(\hat{x}_{s,s-1} - \tilde{x}_{s,s-1}) \quad ,$$

we have $\mathrm{Mse}(\tilde{x}_{s,s-1}) = \sigma^2 P_{s,s-1} + \hat{X}(\delta)_{s,s-1}\mathrm{Mse}(\hat{\delta}_l)\hat{X}(\delta)'_{s,s-1}$. By Theorem 9 of Section 4, the XKF, to be applied to (16b), can then be initialized with $\mathrm{Var}(x_s) = \mathrm{Mse}(\tilde{x}_{s,s-1})$ and $\hat{X}_{s,s-1} = (\tilde{X}(\nu)_{s,s-1}, \tilde{X}(\beta)_{s,s-1})$. If the DKF is to be employed, the initialization for the Q matrix would be

$$\begin{bmatrix} Q_{11} & Q_{13} \\ Q_{31} & Q_{33} \end{bmatrix} - \begin{bmatrix} Q_{12} \\ Q_{32} \end{bmatrix} Q_{22}^{-1} \begin{bmatrix} Q_{21} & Q_{23} \end{bmatrix} \quad ,$$

where $Q_s = (Q_{ij})$, $i, j = 1, 2, 3$. This can be seen considering that, after estimating δ, the sum of squares is $(\nu_l - S_l\beta)'P'\Sigma_l^{-1}P(\nu_l - S_l\beta)$, with $P = I - R_l(R_l'\Sigma_l^{-1}R_l)^{-1}R_l'\Sigma_l^{-1}$. If the first r rows of R do not constitute a submatrix of R of rank r, we would proceed as in the last part of Section 4.

REFERENCES

1. Anderson, B., and J. Moore : *Optimal Filtering*. Prentice Hall, New Jersey, 1979.
2. Ansley, C. F., and R. Kohn : Estimation, Filtering and Smoothing in State Space Models with Incompetely specified Initial Conditions, *Annals of Statistics*, 13 (1985), 1286-1316.
3. Bell, W. and S. C. Hillmer : Initializing the Kalman Filter for Nonstationary Time Series Models, *Journal of Time Series Analysys*, 12 (1991), 283-300.
4. Box, G. E. P., and G. M. Jenkins : *Time Series Analysis, Forecasting and Control*, Holden - Day, San Francisco 1976.
5. Burridge, P. and K. F. Wallis : Calculating the variance of seasonally adjusted series, *Journal of the American Statistical Association*, 80 (1985), 541-552.
6. Gómez, V. and A. Maravall : Estimation, Prediction and Interpolation for Nonstationary Series with the Kalman Filter, EUI Working Paper ECO 92/80 (under revision for the *Journal of the American Statistical Association).*
7. Harvey, A. C., and G. D. A. Phillips : Maximum Likelihood Estimation of Regression Models with Autoregressive-Moving Average Disturbances, *Biometrika*, 66 (1979),49-58.
8. Harvey, A. C., and R. G. Pierse : Estimating Missing Observations in Economic Time Series, *Journal of the American Statistical Association*, 79 (1984), 125-131.
9. Harvey, A. C. : *Forecasting, Structural Time Series Models and the Kalman Filter* Cambridge University Press, Cambridge 1989
10. Jones, R. : Maximum Likelihood Fitting of ARMA Models to Time Series With Missing Observations, *Technometrics*, 22 (1980), 389-395.
11. Jong, Piet de : The Likelihood for the State Space Model, *Biometrika*, 75 (1988), 165-169.
12. Jong, Piet de : Smoothing and Interpolation With the State Space Model,*Journal of the American Statistical Association*, 84 (1989), 408-409.
13. Jong, Piet de : The Diffuse Kalman Filter, *Annals of Statistics*, 19 (1991), 1073-1083.
14. Jong, Piet de : Stable Algorithms for the State Space Model, *Journal of Time Series Analysis*, 12 (1991), 143-156.
15. Kohn, R., and C. F. Ansley : Efficient Estimation and Prediction in Time Series Regression Models, *Biometrika*, 72 (1985), 694-697.
16. Kohn, R., and C. F. Ansley : Estimation, Prediction and Interpolation for ARIMA Models with Missing Data," technical report, Graduate School of Business, University of Chicago, 1984.
17. Kohn, R., and C. F. Ansley : Estimation, Prediction and Interpolation for ARIMA Models With Missing Data, *Journal of the American Statistical Association*, 81 (1986), 751-761.
18. Rao, C. : *Linear Statistical Inference and its Applications*, John Wiley & Sons, New York 1973.
19. Wecker, W., and C. F. Ansley : The Signal Extraction Approach to Nonlinear Regression and Spline Smoothing, *Journal of the American Statistical Association*, 78 (1983), 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

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

From Name . . . . . .
Address. . . . . .
. . . . . .
. . . . . .
. . . . . .
. . . . . .

❒ 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 1993/94
❒ Please send me the EUI Research Report

Please send me the following EUI 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

* Working Paper out of print

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 Estimates 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

* Working Paper out of print

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

* Working Paper out of print

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

ECO No. 92/81
Víctor GÓMEZ and Agustín MARAVALL
Time Series Regression with ARIMA Noise and Missing Observations Program TRAM

ECO No. 92/82
J. Bradford DE LONG/ Marco BECHT
"Excess Volatility" and the German Stock Market, 1876-1990

ECO No. 92/83
Alan KIRMAN/Louis PHLIPS
Exchange Rate Pass-Through and Market Structure

ECO No. 92/84
Christian DUSTMANN
Migration, Savings and Uncertainty

ECO No. 92/85
J. Bradford DE LONG
Productivity Growth and Machinery Investment: A Long-Run Look, 1870-1980

ECO NO. 92/86
Robert B. BARSKY and J. Bradford DE LONG
Why Does the Stock Market Fluctuate?

ECO No. 92/87
Anthony B. ATKINSON/John MICKLEWRIGHT
The Distribution of Income in Eastern Europe

ECO No.92/88
Agustín MARAVALL/Alexandre MATHIS
Encompassing Unvariate Models in Multivariate Time Series: A Case Study

ECO No. 92/89
Peter J. HAMMOND
Aspects of Rationalizable Behaviour

ECO 92/90
Alan P. KIRMAN/Robert J. WALDMANN
I Quit

ECO No. 92/91
Tilman EHRBECK
Rejecting Rational Expectations in Panel Data: Some New Evidence

ECO No. 92/92
Djordje Suvakovic OLGIN
Simulating Codetermination in a Cooperative Economy

ECO No. 92/93
Djordje Suvakovic OLGIN
On Rational Wage Maximisers

ECO No. 92/94
Christian DUSTMANN
Do We Stay or Not? Return Intentions of Temporary Migrants

ECO No. 92/95
Djordje Suvakovic OLGIN
A Case for a Well-Defined Negative Marxian Exploitation

ECO No. 92/96
Sarah J. JARVIS/John MICKLEWRIGHT
The Targeting of Family Allowance in Hungary

ECO No. 92/97
Agustín MARAVALL/Daniel PEÑA
Missing Observations and Additive Outliers in Time Series Models

ECO No. 92/98
Marco BECHT
Theory and Estimation of Individual and Social Welfare Measures: A Critical Survey

ECO No. 92/99
Louis PHLIPS and Ireneo M'guel MORAS
The AKZO Decision: A Case of Predatory Pricing?

ECO No. 92/100
Stephen MARTIN
Oligopoly Limit Pricing With Firm-Specific Cost Uncertainty

* Working Paper out of print

ECO No. 92/101
Fabio CANOVA/Eric GHYSELS
Changes in Seasonal Patterns: Are They Cyclical?

ECO No. 92/102
Fabio CANOVA
Price Smoothing Policies: A Welfare Analysis

***

ECO No. 93/1
Carlo GRILLENZONI
Forecasting Unstable and Non-Stationary Time Series

ECO No. 93/2
Carlo GRILLENZONI
Multilinear Models for Nonlinear Time Series

ECO No. 93/3
Ronald M. HARSTAD/Louis PHLIPS
Futures Market Contracting When You Don't Know Who the Optimists Are

ECO No. 93/4
Alan KIRMAN/Louis PHLIPS
Empirical Studies of Product Markets

ECO No. 93/5
Grayham E. MIZON
Empirical Analysis of Time Series: Illustrations with Simulated Data

ECO No. 93/6
Tilman EHRBECK
Optimally Combining Individual Forecasts From Panel Data

ECO NO. 93/7
Víctor GÓMEZ/Agustín MARAVALL
Initializing the Kalman Filter with Incompletely Specified Initial Conditions

* Working Paper out of print