MISSING OBSERVATIONS IN ARIMA MODELS: SKIPPING STRATEGY VERSUS ADDITIVE OUTLIER APPROACH

Victor Gómez, Agustin Maravall
and Daniel Peña

Banco de España - Servicio de Estudios
Documento de Trabajo nº 9701

MISSING OBSERVATIONS IN ARIMA MODELS: SKIPPING STRATEGY VERSUS ADDITIVE OUTLIER APPROACH

Víctor Gómez*, Agustín Maravall**
and Daniel Peña***

(*) Ministerio de Economía y Hacienda, Madrid.
(**) Bank of Spain, Madrid.
(***) Universidad Carlos III, Madrid.

Banco de España - Servicio de Estudios
Documento de Trabajo nº 9701

1. INTRODUCTION AND SUMMARY

The problem of optimal estimation of missing observations in stationary Autoregressive Moving Average (ARMA) models was solved in Jones (1980). Extension of his approach to nonstationary integrated ARMA (i.e., ARIMA) models posed serious problems, having mostly to do with the specification of the starting conditions for the Kalman filter and the definition of a proper likelihood. Several solutions have been proposed, among them, the "transformation" approach of Kohn and Ansley (1986), the "diffuse prior" approach of De Jong (1991), and the "conditional likelihood" approach of Gómez and Maravall (1994). These solutions share the basic features of the approach in Jones: the use of (some version of) the Kalman Filter (KF) for likelihood evaluation, "skipping" in the computations the missing observations. Maximum likelihood estimation of the ARIMA parameters is then possible, and some smoothing algorithm, such as the Fixed Point Smoother (FPS), interpolates the missing values. We shall refer to this general approach as the "skipping approach". Since the Kohn–Ansley, De Jong, and Gómez–Maravall approaches are equivalent, due to its simplicity, we shall use the latter to represent the skipping approach method.

It is also well known (see, for example, Sargan and Drettakis, 1974) that a sensible alternative to the problem of missing observations estimation is through dummy variables that take a value of unity at the point where the observation is missing, and zero elsewhere; see Harvey (1989, pp. 145–146). Subject to a qualification to be made below, one could think of the following procedure: Fill, first, the holes corresponding to the missing values with arbitrary data, and then use maximum likelihood estimation of an ARIMA model with Additive Outliers; this procedure could be seen as a particular case of the intervention analysis of Box and Tiao (1975). The difference between the arbitrary value set by the user and its corresponding estimated parameter can be used as the missing observation estimator. In fact, when the model parameters are known, this difference coincides with the conditional expectation of the missing value given the observed data (see, Brubacher and Wilson, 1976).

We shall refer to this procedure as the "Additive Outlier" (AO) approach to missing observations estimation. Be that as it may, when the model parameters are not known and are to be estimated by maximum likelihood, the AO and the skipping approaches will differ, due to the fact that the determinantal term in the Gaussian likelihoods will be different. The determinantal term in the AO likelihood includes the effect of the filled–in values; that

of the skipping likelihood will ignore these effects. Since differences in likelihood produce differences in parameter estimates, if the AO likelihood is not corrected, the AO approach can only be seen as an approximate way to obtain the maximum likelihood estimators. The difference between the two likelihoods was pointed out by Peña (1987), in the context of autoregressive models, and, for stationary ARMA models, analysed by Ljung (1989), who went on to provide some insights into the nonstationary case. For this case, however, there was no attempt to define the likelihood of the nonstationary observed series. In this paper, we present a rigorous development of the AO approach to missing observations estimation in the general nonstationary case, which we shall denote the "corrected AO" approach. The paper further shows the equivalence of this and the skipping (plus smoothing) approach. Computationally efficient ways to perform both approaches are provided in detail, and it is further seen how the correction that needs to be applied to the AO likelihood is trivially obtained from KF computations for the usual AO likelihood. Results for the three (skipping, AO, and corrected AO) approaches are then compared through simulation for different models, different sample sizes, and different distributions of missing observations in the series.

One practical advantage of the standard AO approach, both in the stationary and nonstationary cases, is that it can be easily implemented with existing software if one is ready to accept the approximation implied by not correcting the determinantal term. In fact, this is the approach followed in the new X12ARIMA procedure (Findley et al., 1996). Assessing the influence of the determinantal correction is a by-product of the paper.

The last part of the paper contains a simulation exercise to assess the relative performance of the different approaches. It is concluded that there is a brief trade-off between both approaches. When the number of missing observations is small, the additive outlier approach can be easier and faster to implement. However, as the number of missing observations increases, it is clearly outperformed by the skipping approach.

The paper is structured as follows. Section 2 reviews briefly first the skipping approach in the stationary case, as suggested by Jones (1980), and then its generalization to the nonstationary case, following Gómez and Maravall (1994). In Section 3, we consider the additive outlier approach, and analyze in detail a nonstationary series that follows a general ARIMA model where all missing observations have been replaced by arbitrary values and a dummy variable has been specified for each of them. Section 4 presents the simulation exercise. Computational details to carry the estimation procedures efficiently, as well as

the proofs of the results mentioned in the main text, are presented in two Appendices.

2. SKIPPING APPROACH

2.1 Stationary Series, ARMA Model

Let the observed series $z_o = (z(t_1), z(t_2), ..., z(t_M))'$, $1 \leq t_1 < t_2 < ... < t_M \leq N$, be the outcome of the ARMA model:

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

where $\phi(B) = 1 + \phi_1 B + ... + \phi_p B^p$ and $\theta(B) = 1 + \theta_1 B + ... + \theta_q B^q$ are finite polynomials in the lag operator B, of degrees 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. To avoid unbounded standard errors of the interpolators, we further assume the model invertible, i.e. the roots of $\theta(B)$ lie outside the unit circle; see Maravall and Peña (1992). If there are no missing observations, letting $r = max\{p, q+1\}$ and defining $\phi_i = 0$ when $i > p$, one state space representation for this model is

$$x(t) = Fx(t-1) + Ga(t) \tag{2.2a}$$

$$z(t) = H'x(t), \tag{2.2b}$$

where $t = 1, \ldots, N$, $x(t) = (z(t), z(t+1 \mid t), ..., z(t+r-1 \mid t))'$, $G = (1, \psi_1, ..., \psi_{r-1})'$, $H = (1, 0, ..., 0)'$,

$$F = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \ldots & 1 \\ -\phi_r & -\phi_{r-1} & -\phi_{r-2} & \ldots & -\phi_1 \end{bmatrix},$$

and the ψ_i- weights are obtained from $\psi(B) = \theta(B)/\phi(B) = \sum_{i=0}^{\infty} \psi_i B^i$. The expression $z(t+i \mid t)$ is the orthogonal projection of $z(t+i)$ on the subspace generated by $\{z(s) : s \leq t\}$, and coincides with the conditional expectation $E(z(t+i) \mid z(s) : s \leq t)$, $\quad i = 1, \ldots, r-1$. 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 $\{z(s) : s \leq t\}$. The Kalman filter can then be applied to model (2.2) for prediction and likelihood evaluation. As starting conditions,

one takes the first two moments of the unconditional distribution of the initial state vector, $x(1)$.

For the general case, when some observations may be missing, the observation equation $(2.2b)$ is replaced with

$$z(t) = H'(t)x(t) + \alpha(t)W(t), \quad t = 1, \ldots, N,$$

where $H'(t) = (1, 0, \ldots, 0)$, $\alpha(t) = 0$ if $z(t)$ is observed, $H'(t) = (0, 0, \ldots, 0)$, $\alpha(t) = 1$ if $z(t)$ is missing (Brockwell and Davis 1987, p. 494). The variable $W(t)$ represents an $Niid(0, 1)$ variable, independent of $\{z(t_1), \ldots, z(t_M)\}$. Thus, when $z(t)$ is missing, in the Kalman filter equations, $x(t \mid t) = x(t \mid t-1)$, $\Sigma(t \mid t) = \Sigma(t \mid t-1)$, where $x(t \mid t+i) = E(x(t) \mid z(1), \ldots, z(t+i))$, $\Sigma(t \mid t+i) = Var(x(t) \mid z(1), \ldots, z(t+i))$, $1 \leq t \leq N$, $i = -1, 0$, and both the residual and the standard error corresponding to a missing value are ignored when evaluating the likelihood function; see Jones (1980).

Having obtained parameter estimates by maximizing the likelihood function using the prediction error decomposition, estimators of the missing values can be obtained through the simplified FPS of Gómez and Maravall (1994); see also Anderson and Moore (1979).

2.2 Nonstationary Series, ARIMA Model

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 (2.1). Then, $\{z(t)\}$ follows the nonstationary model

$$\phi(B)\delta(B)z(t) = \theta(B)a(t), \tag{2.3}$$

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.

Suppose first that there are no missing observations, and let $z = (z(1), z(2), \ldots, z(N))'$ and $u = (u(d+1), u(d+2), \ldots, u(N))'$ be the observed series and the differenced series, respectively. The nonstationarity of $\{z(t)\}$ prevents us from using the prediction error decomposition, since the distribution of $x(1)$ is not well defined. In order to define the likelihood, we proceed as in Gómez and Maravall (1994) and make the following assumptions:

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, $L(u)$; see Box and Jenkins, chap. 7. Letting $z_I = (z(1), \ldots, z(d))'$ and $z_{II} = (z(d+1), \ldots, z(N))'$, it is easily seen that differencing the data implies the transformation $[z_I', u']' = J\,[z_I', z_{II}']'$, where $J = [J_I', J_{II}']'$, $J_I = [I_d, 0]$, I_d is the identity matrix of rank d and

$$J_{II} = \begin{bmatrix} \delta_d & \ldots & \delta_1 & 1 & & \bigcirc \\ & \ddots & & & \ddots & \\ \bigcirc & & \delta_d & \ldots & \delta_1 & 1 \end{bmatrix}$$

If we partition $J_{II} = [J_1, J_2]$ conforming to z_I and z_{II}, one can write

$$\begin{bmatrix} z_I \\ z_{II} \end{bmatrix} = \begin{bmatrix} I_d & 0 \\ -\Xi J_1 & \Xi \end{bmatrix} \begin{bmatrix} z_I \\ u \end{bmatrix},$$

where $\Xi = J_2^{-1}$ is the lower triangular matrix

$$\Xi = \begin{bmatrix} 1 & & & \\ \xi_1 & 1 & & \\ \vdots & & \ddots & \\ \xi_{N-d-1} & \ldots & \xi_1 & 1 \end{bmatrix}.$$

The ξ_i-coefficients are obtained from

$$1/\delta(B) = \sum_{i=0}^{\infty} \xi_i B^i \tag{2.4}$$

and the rows of $-\Xi J_1$ can be obtained recursively as shown in Bell (1984). Specifically, letting $A_{ij} = \delta_{ij}$, $i, j = 1, \ldots, d$, where δ_{ij} is the Kronecker delta, the row vectors $A_t' = (A_{1t}, \ldots, A_{dt})$, $t = d+1, \ldots, N$, of the $(N-d) \times d$ matrix $A = -\Xi J_1$ can be obtained from the recursions

$$A_{it} = -\delta_1 A_{it-1} - \ldots - \delta_d A_{it-d}, \quad i = 1, \ldots, d, \quad t > d, \tag{2.5}$$

and the relation

$$z_{II} = A z_I + \Xi u \tag{2.6}$$

holds. Let $v = \Xi u$. Then, the likelihood $L(v)$ based on v coincides with the likelihood $L(u)$ based on u because Ξ has unit determinant. Given that $v = z_{II} - A z_I$, the log-likelihood

based on u is (throughout the paper all log-likelihoods will be defined up to an additive constant)

$$l(u) = -\frac{1}{2}\{(N-d)\ln(\sigma^2) + \ln \mid \Omega_v \mid + (z_{II} - Az_I)'\Omega_v^{-1}(z_{II} - Az_I)/\sigma^2\}, \qquad (2.7)$$

where $Var(v) = \sigma^2\Omega_v$, $\Omega_v = \Xi\Omega_u\Xi'$, and $Var(u) = \sigma^2\Omega_u$. Equation (2.7) constitutes an expression of the Box–Jenkins log–likelihood in terms of the original series. Another interpretation can be obtained if assumptions A and B hold. Given that the matrix $J = [J_I', J_{II}']'$ has unit determinant, the log–likelihood $l(z)$ of the observed series $z = [z_I', z_{II}']'$ verifies $l(z) = l(z_I, u) = l(z_I) + l(u)$. Therefore, under assumptions A and B, we have the result

LEMMA 1. $\quad l(u) = l(z_{II} \mid z_I)$.

That is, the Box–Jenkins log–likelihood is equal to the log–likelihood of z_{II} conditional on z_I. In order to use the Kalman filter with the original (not the differenced) series, we need a state space representation suitable for nonstationary series. One such representation is given also by (2.2), with the ϕ and ψ coefficients replaced with the ϕ^* and ψ^* ones, respectively, where $\phi^*(B) = \phi(B)\delta(B)$ and $\psi^*(B) = \theta(B)/\phi^*(B) = \sum_{i=0}^{\infty}\psi_i^*B^i$, $\phi_i^* = 0$ when $i > p + d$, and $r = max\{p + d, q + 1\}$. The elements of the state vector are now $z(t)$ and $z(t+i \mid t) = z(t+i) - \psi_0^*a(t+i) - \ldots - \psi_{i-1}^*a(t+1)$, $i = 1, \ldots, r-1$. The following lemma, whose proof is omitted, ensures that this state space representation is correct.

LEMMA 2. $\quad z(t+r-1 \mid t) = -\phi_r^*z(t-1) - \phi_{r-1}^*z(t \mid t-1) - \ldots - \phi_1^*z(t+r-2 \mid t-1) + \psi_{r-1}^*a(t)$.

The Kalman filter can then be applied to compute the conditional log–likelihood $l(z_{II} \mid z_I)$ through the prediction error decomposition. The starting conditions can be obtained from (2.6) as follows. If we consider the definition of the elements of the state vector $x(t)$, it can be seen that $x(d+1) = A_*z_I + \Xi_*U_*$, where A_* is the $r \times d$ submatrix of A formed by the first r rows, Ξ_* is the $r \times r$ submatrix of Ξ formed by the first r rows and the first r columns, $U_* = [u(d+1), u(d+2 \mid d+1), \ldots, u(d+r \mid d+1)]'$, and $u(d+i \mid d+1) = E(u(d+i) \mid u(t) : t \leq d+1)$, $i = 2, \ldots, r$. Therefore, we can take as starting conditions

$$x(d+1 \mid d) = E\,(x(d+1) \mid z(s) : 1 \leq s \leq d) = A_*z_I$$
$$\Sigma(d+1 \mid d) = Var\,(x(d+1) \mid z(s) : 1 \leq s \leq d) = \Xi_*\tilde{\Sigma}(d+1 \mid d)\Xi_*',$$

where $\tilde{\Sigma}(d+1 \mid d) = E(U_* U_*')$ can be computed from the stationary process $\{u(t)\}$, which follows model (2.1); see Jones (1980).

If there are missing observations and the observed series $z_o = (z(t_1), z(t_2), ..., z(t_M))'$, $1 \leq t_1 < t_2 < ... < t_M \leq N$, is a subvector of the complete series $z = [z_I', z_{II}']'$, we can proceed as follows. Let $z_{Io} = (z(t_1), \ldots, z(t_k))'$, $k \leq d$, and z_{Im} be the subvectors of z_I corresponding to the observed and missing values in z_I, respectively, and let z_{IIo} be the subvector of z_{II} formed with the observed values in z_{II}. Then, we can write

$$z_{IIo} = A_o z_I + v_o = B_o z_{Io} + C_o z_{Im} + v_o, \tag{2.8}$$

where A_o and v_o are the submatrix and subvector, respectively, of A and $v = \Xi u$ corresponding to the observed values in the series. Both Lemma 1 and (2.8) suggest a natural way to extend the log-likelihood (2.7) to the case of missing observations. We can consider in equation (2.8) z_{Im} as fixed and define the likelihood of the observed series as the likelihood of the generalized least squares (GLS) regression model

$$y_o = C_o z_{Im} + v_o, \tag{2.9}$$

where $y_o = z_{IIo} - B_o z_{Io}$. This is the definition of Gómez and Maravall (1994). Then, the log-likelihood when there are missing observations is

$$l(y_o) = -\frac{1}{2}\{(M-k)\ln(\sigma^2) + \ln \mid \Omega_{v_o} \mid + (y_o - C_o z_{Im})' \Omega_{v_o}^{-1} (y_o - C_o z_{Im})/\sigma^2\}, \tag{2.10}$$

where $Var(v_o) = \sigma^2 \Omega_{v_o}$. In order to evaluate the log-likelihood and interpolate missing values, we can now use the method of Gómez and Maravall (1994). The log-likelihood evaluation is made simpler by concentrating z_{Im} and σ^2 out of the log-likelihood (2.10). Given the parameters $(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$, of the ARMA model (2.1), this is done by replacing in (2.10) z_{Im} and σ^2 with their GLS estimators in model (2.9), which are also the maximum likelihood estimators.

It is important to mention that only the Kalman filter and the simplified fixed point smoother of Gómez and Maravall (1994) are necessary to perform the calculations. Compared with the modified versions of these algorithms of Kohn and Ansley (1986), this means a significant simplification both conceptually and in the programming burden. Also, as we will see in Section 4, it is not necessary to evaluate the vector y_o and the matrix C_o before applying the Kalman filter. The computations are done automatically by means of an

"Augmented Kalman filter" (AKF) algorithm, easy to program, and detailed in Appendix A.

2.3 Regression Model with ARIMA Errors

Consider the regression model

$$z(t) = y'(t)\beta + \nu(t), \tag{2.11}$$

where $\beta = (\beta_1, \ldots, \beta_h)'$ is a vector of parameters, $y'(t)$ is a vector of h independent variables, $z(t)$ is the dependent variable, and $\{\nu(t)\}$ is assumed to follow the ARIMA model given by (2.3). If, as in the previous Section, z_o denotes the observed series, defining the vector $\nu_o = (\nu(t_1), \ldots, \nu(t_M))'$ and the $M \times h$ matrix Y_o with the vectors $y'(t), t = t_1, \ldots, t_M$, as rows, we can write $z_o = Y_o\beta + \nu_o$, where the matrix Y_o is assumed of rank h. Since $\{\nu(t)\}$ follows the ARIMA model (2.3), similarly to (2.8), we can write $\nu_{IIo} = B_o\nu_{Io} + C_o\nu_{Im} + v_o$, where ν_{IIo}, ν_{Io} and ν_{Im} are the vectors of errors corresponding to the subvectors z_{IIo}, z_{Io} and z_{Im} of the complete series z, defined at the end of the previous section. Let Y_{Io}, Y_{IIo} and Y_{Im} be the matrices with rows the vectors $y'(t)$ corresponding to the vectors ν_{Io}, ν_{IIo} and ν_{Im}, respectively. Replacing ν_{IIo} with $z_{IIo} - Y_{IIo}\beta$, ν_{Io} with $z_{Io} - Y_{Io}\beta$ and ν_{Im} with $z_{Im} - Y_{Im}\beta$ in the above expression, the following regression model is obtained

$$z_{IIo} = B_o z_{Io} + C_o z_{Im} + Y_{IIo}\beta - B_o Y_{Io}\beta - C_o Y_{Im}\beta + v_o,$$

where the regression parameters are z_{Im} and β. Letting $y_o = z_{IIo} - B_o z_{Io}$, it can be rewritten as

$$\begin{aligned} y_o &= [C_o, Y_{IIo} - B_o Y_{Io} - C_o Y_{Im}]\,[z'_{Im}, \beta']' \\ &= [C_o, Y_{IIo} - A_o Y_I]\,[z'_{Im}, \beta']', \end{aligned} \tag{2.12}$$

where Y_I is the $d \times h$ matrix formed with the vectors $y'(t)$, $t = 1, \ldots, d$, as rows, and A_o is the matrix defined by $B_o Y_{Io} + C_o Y_{Im} = A_o Y_I$, which coincides with that of (2.8). The log-likelihood of the observed series is defined as that of the GLS model (2.12). The same algorithms of the previous section can now be used for prediction, interpolation and log-likelihood evaluation (the vector of regression parameters is now $[z'_{Im}, \beta']'$, instead of z_{Im}).

If we define the vector $x(t) = (\nu(t), \nu(t+1 \mid t), \ldots, \nu(t+r-1 \mid t))$, then the state space representation is given by (2.2a) and the observation equation $z(t) = y'(t)\beta + H'(t)x(t) + \alpha(t)W(t)$, where $H(t)$, $\alpha(t)$ and $W(t)$ are as in Section 2.1 and the elements of the state vector are as in Section 2.2 with z replaced with ν.

2.4 A Brief Remark on the Likelihood of Nonstationary Models and "The Conditional Likelihood" Approach

Let us assume, first, that there are no missing values, and the observed series is $z = (z_1, \ldots, z_N)'$. If d denotes the number of observations lost in differencing, letting $\gamma = (z_1, \ldots, z_d)$, we can write

$$z_t = A_t'\gamma + \tilde{u}_t, \qquad t = d+1, \ldots, N, \tag{2.13}$$

where $\tilde{u}_t = \sum_{i=0}^{t-d-1} \xi_i u_{t-i}$, u_t is the differenced series $\delta(B)z_t$, ξ_i is as in (2.4), and the A_ts are obtained through (2.5). The distribution of γ, the starting conditions, is unknown. If we consider, for instance, the state space representation of the equation $\delta(B)z_t = u_t$, given by

$$\begin{bmatrix} z_{t-d+1} \\ z_{t-d} \\ \cdots \\ z_t \end{bmatrix} = \begin{bmatrix} 0 & 1 & 0 & \ldots & 0 \\ 0 & 0 & 1 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ -\delta_d & -\delta_{d-1} & \ldots & \ldots & -\delta_1 \end{bmatrix} \begin{bmatrix} z_{t-d} \\ z_{t-d-1} \\ \cdots \\ z_{t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \cdots \\ u_t \end{bmatrix},$$

and iterate indefinitely (equivalent to making the starting conditions approach the infinitely remote past), given that the eigenvalues of F are all of unit modulus, the variance of $x(t)$ will go to ∞. It seems clear that γ should be seen as a vector of nuisance random variables, that should be removed in order to properly define the likelihood.

The classical Box–Jenkins approach, removes γ by differencing the series. This is the same as making a linear transformation of the data, $Jz = (\gamma, u)'$, with unit determinant, so that the likelihood can be factorized as

$$p(z) = p(Jz) = p(\gamma, u) = p(\gamma \mid u)p(u),$$

where $p(u)$ does not contain γ, and $p(\gamma \mid u)$ does not contain information on the model parameters. The likelihood is then defined as $p(u)$.

An alternative solution consists in assuming that γ is independent of the series $\{u_t\}$, condition on γ in (2.13), and define the likelihood using the conditional density $p(z_{d+1}, \ldots,$

$z_N \mid z_1, \ldots, z_d$). As Lemma 1 showed, the two likelihoods coincide. An advantage of the "conditional likelihood" approach is that it is easily extended to related models. Further, it is particularly adequate for algorithms that recursively update conditional expectations, such as the KF, and provides an easy solution to the problem of the starting values.

Thus, assume now that there are missing values, and the observations are $z = (z_{t_1}, \ldots, z_{t_M})'$ with $1 \leq t_1 < \ldots < t_M$. Expression (2.13) is still valid, although some of the missing values may be among the first d periods, and hence contained in γ. From the point of view of the conditional likelihood, however, this presence is of no relevance. We still assume that γ is independent of $\{u_t\}$, and condition on γ in expression (2.13); by doing so, γ becomes a fixed parameter. As a consequence, if there are missing values in γ, these become parameters in the likelihood, which is then defined as $p(z_{t_{k+1}}, \ldots, z_{t_M} \mid z_1, \ldots, z_d)$, where t_k is the largest integer in $(t_1, \ldots, t_M)$ which is $\leq d$. We next see how this conditional likelihood approach is straightforward to apply in the AO approach to missing observations estimation.

3. ADDITIVE OUTLIER APPROACH

3.1 Stationary Series, ARMA Model

Let the observed series z_o be that in Section 2.1 with the same assumptions holding, and let $z = (z(1), z(2), \ldots, z(N))'$ be the complete series, which includes the unobserved values. If $\bar{z}$ denotes the series obtained from z by replacing the missing values z_m with tentative values $\bar{z}_m$, the following theorem provides an expression for the log-likelihood $l(z_o)$ based on z_o, in terms of $\bar{z}$.

THEOREM 1. *Let $\omega = \bar{z}_m - z_m$. Then, the log-likelihood of the observed values z_o is*

$$l(z_o) = -\frac{1}{2}\{N \ln(\sigma^2) + \ln \mid \Omega_z \mid + \ln \mid X'\Omega_z^{-1}X \mid + (\bar{z} - X\hat{\omega})'\Omega_z^{-1}(\bar{z} - X\hat{\omega})/\sigma^2\},$$

where $Var(z) = \sigma^2\Omega_z$, X is the $N \times (N - M)$ matrix whose columns are unit vectors, such that the i-th column has a one in the position corresponding to the i-th missing value, $i = 1, 2, \ldots, N - M$, $\hat{\omega} = (X'\Omega_z^{-1}X)^{-1}X'\Omega_z^{-1}\bar{z}$ and $\hat{\omega} = \bar{z}_m - E(z_m \mid z_o)$. Also, $Mse(\hat{\omega}) = Var(z_m \mid z_o) = \sigma^2(X'\Omega_z^{-1}X)^{-1}$.

A similar result was first obtained by Peña (1987) for a first order autoregressive model, and was generalized to stationary ARMA models by Ljung (1989). Theorem 1 implies that, in order to evaluate the log–likelihood $l(z_o)$, all we have to do is, first, fill in the series z with tentative values $\bar{z}_m$ and then use a standard method to compute log–likelihoods for regression models with ARMA errors. Note, however, that the likelihood in Theorem 1 includes the determinantal term $\ln \mid X'\Omega_z^{-1}X \mid$. If this correction is not made, only an approximation to the exact log–likelihood is obtained. The interpolations of the missing values z_m are simply $\bar{z}_m - \hat{\omega}$. In Appendix B we describe in more detail the algorithms we use, simpler yet equivalent to those of Kohn and Ansley (1985). Note that the filled–in series is used for likelihood evaluation and, therefore, no skipping takes place. This allows for faster routines than the ones used with skipping. However, there is a computational burden implicit in the number of regression parameters.

3.2 Nonstationary Series, ARIMA Model

Let the observed series $z_o = (z(t_1), z(t_2), ..., z(t_M))'$, $1 \leq t_1 < t_2 < ... < t_M \leq N$, be a subvector of the complete series $z = [z_I', z_{II}']'$, with the assumptions and notation of Section 2.2 holding. Given the definition of the log–likelihood (2.10), we can proceed as in Section 3.1 because z_{Im} is considered fixed and the covariance structure of the error in model (2.9) is known. Let z_{IIm} be the subvector of z_{II} containing the missing values in z_{II}. Partition $z_{II} = Az_I + v = Bz_{Io} + Cz_{Im} + v$ conforming to z_{IIo} and z_{IIm}, such that (2.8) and $z_{IIm} = A_m z_I + v_m = B_m z_{Io} + C_m z_{Im} + v_m$ hold. If $\bar{z}_{II}$ denotes the series obtained from z_{II} replacing the unobserved values z_{IIm} with tentative values $\bar{z}_{IIm}$, the following theorem, analogous to Theorem 1, provides an expression for the log–likelihood $l(y_o)$ based on y_o, in terms of $[z_I', \bar{z}_{II}']'$.

THEOREM 2. *Let $\omega_{II} = \bar{z}_{IIm} - z_{IIm}$. Then, the log–likelihood based on y_o is*

$$l(y_o) = -\frac{1}{2}\{(M-k)\ln(\sigma^2) + \ln \mid \Omega_v \mid + \ln \mid X_{II}'\Omega_v^{-1}X_{II} \mid$$

$$+(\bar{z}_{II} - Az_I - X_{II}\hat{\omega}_{II})'\Omega_v^{-1}(\bar{z}_{II} - Az_I - X_{II}\hat{\omega}_{II})/\sigma^2\}, \qquad (3.1)$$

where X_{II} is the $(N-d) \times (N-M-d+k)$ matrix whose columns are unit vectors, such that the i–th column has a one in the position corresponding to the i–th missing value in z_{II}, $i = 1, 2, \ldots, N-M-d+k$, $\hat{\omega}_{II} = (X_{II}'\Omega_v^{-1}X_{II})^{-1}X_{II}'\Omega_v^{-1}(\bar{z}_{II} - Az_I)$, and

$\hat{\omega}_{II} = \bar{z}_{IIm} - A_m z_I - E(z_{IIm} - A_m z_I \mid z_{IIo} - A_o z_I)$. *Also,* $Mse(\hat{\omega}_{II}) = Var(z_{IIm} - A_m z_I \mid z_{IIo} - A_o z_I) = \sigma^2 (X'_{II} \Omega_v^{-1} X_{II})^{-1}$.

Note that in (3.1) the parameters to estimate are $(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$, σ^2 and z_{Im}. No tentative values have been assigned yet to the elements of z_{Im}. As we mentioned at the end of Section 2.2, replacing in (2.10) σ^2 and z_{Im} with the GLS estimators $\hat{\sigma}^2$ and $\hat{z}_{Im}$, respectively, of model (2.9), we can concentrate σ^2 and z_{Im} out of the log–likelihood. We will show later that the same concentrated log–likelihood can be obtained replacing also z_{Im} with tentative values $\bar{z}_{Im}$ and concentrating σ^2 and $\omega_I = \bar{z}_{Im} - z_{Im}$ out of the log–likelihood (3.1). But first we will give in the next corollary an alternative expression to (3.1) based on differencing $[z'_I, \bar{z}'_{II}]$ and the columns of $[0', X'_{II}]'$.

COROLLARY 1. *With the notation of theorem* 2, *let* $u^* = J_{II}[z'_I, \bar{z}'_{II}]'$ *and* $X^*_{II} = J_{II}[0', X'_{II}]'$, *where* J_{II} *is the matrix defined in Section* 2.2, *be the result of differencing* $[z'_I, \bar{z}'_{II}]'$ *and the columns of* $[0', X'_{II}]'$, *respectively. Then, the log–likelihood* (3.1) *can be expressed as*

$$l(y_o) = -\frac{1}{2}\{(M-k)\ln(\sigma^2) + \ln \mid \Omega_u \mid + \ln \mid X^{*'}_{II} \Omega_u^{-1} X^*_{II} \mid$$

$$+ (u^* - X^*_{II}\hat{\omega}_{II})' \Omega_u^{-1} (u^* - X^*_{II}\hat{\omega}_{II})/\sigma^2\}, \tag{3.2}$$

$\hat{\omega}_{II} = (X^{*'}_{II} \Omega_u^{-1} X^*_{II})^{-1} X^{*'}_{II} \Omega_u^{-1} u^*$ *and* $Mse(\hat{\omega}_{II}) = \sigma^2 (X^{*'}_{II} \Omega_u^{-1} X^*_{II})^{-1}$, *where, as in Section* 2.2 , $u = J_{II} z$ *is the differenced series and* $Var(u) = \sigma^2 \Omega_u$.

Suppose now that $\bar{z}_I$ denotes the vector obtained from z_I replacing the missing values z_{Im} with tentative values $\bar{z}_{Im}$ and let $\bar{z} = [\bar{z}'_I, \bar{z}'_{II}]'$ be the complete filled in series. Define $\omega_I = \bar{z}_{Im} - z_{Im}$ and $\omega = [\omega'_I, \omega'_{II}]'$. Then, we can write

$$\begin{bmatrix} z_I \\ z_{II} \end{bmatrix} = \begin{bmatrix} \bar{z}_I \\ \bar{z}_{II} \end{bmatrix} - \begin{bmatrix} X_I & 0 \\ 0 & X_{II} \end{bmatrix} \begin{bmatrix} \omega_I \\ \omega_{II} \end{bmatrix},$$

where X_I is the $d \times (d-k)$ matrix whose columns are unit vectors, such that the i–th column has a one in the position corresponding to the i–th missing value in z_I, $i = 1, 2, \ldots, d-k$, or, in obvious and more compact notation, $z = \bar{z} - X\omega$. The main result of this section is contained in the next theorem.

THEOREM 3. *Let* $\bar{u} = J_{II}\bar{z}$ *and* $X^* = J_{II}X$, *where* J_{II} *is the matrix defined in Section* 2.2, *be the result of differencing* $\bar{z}$ *and the columns of* X, *respectively. Then, maximizing the log-likelihood* (2.10) *or, equivalently,* (3.2) *with respect to* z_{Im} *yields the* z_{Im}*-maximized log-likelihood*

$$\lambda(y_o) = -\frac{1}{2}\{(M-k)\ln(\sigma^2) + \ln \mid \Omega_u \mid + \ln \mid X_{II}^{*'}\Omega_u^{-1}X_{II}^* \mid$$

$$+(\bar{u} - X^*\hat{\omega})'\Omega_u^{-1}(\bar{u} - X^*\hat{\omega})/\sigma^2\},$$

where $\hat{\omega}$ *is the GLS estimator of* ω *in the model* $\bar{u} = X^*\omega + u$, $X_{II}^* = J_{II}[0', X_{II}']'$, $u = J_{II}z$ *is, as in Section* 2.2, *the differenced series and* $Var(u) = \sigma^2\Omega_u$. *Therefore,* $\hat{\omega} = (X^{*'}\Omega_u^{-1}X^*)^{-1}X^{*'}\Omega_u^{-1}\bar{u}$ *and* $Mse(\hat{\omega}) = \sigma^2(X^{*'}\Omega_u^{-1}X^*)^{-1}$.

Note that, by Theorems 2 and 3, the interpolations $\hat{z}_{Im}$ of z_{Im} and $\hat{z}_{IIm}$ of z_{IIm} are simply $\bar{z}_{Im} - \hat{\omega}_I$ and $\bar{z}_{IIm} - \hat{\omega}_{II}$, respectively, where $\hat{\omega} = [\hat{\omega}_I', \hat{\omega}_{II}']'$ is given by Theorem 3. [Another expression for $\hat{z}_{IIm}$, is $\hat{z}_{IIm} = B_m z_{Io} + C_m\hat{z}_{Im} + P(y_o - C_o\hat{z}_{Im})$, where P is the matrix such that $E(z_{IIm} - A_m z_I \mid z_{IIo} - A_o z_I) = P(z_{IIo} - A_o z_I)$, $P = Cov(v_m, v_o)Var^{-1}(v_o)$.] Note also that in the z_{Im}-maximized log-likelihood $\lambda(y_o)$ of Theorem 3 the correction $\ln \mid X_{II}^{*'}\Omega_u^{-1}X_{II}^* \mid$ in the determinantal term involves only the missing values contained in z_{II} and not those contained in the data lost by differencing z_I.

In the stationary case, by Theorem 1, the interpolator in the AO approach, $\bar{z}_m - \hat{\omega}$, is equal to $E(z_m \mid z_o)$. Therefore it is identical to the one obtained in the skipping approach, since that conditional expectation is precisely what the KF, used as in Jones (1980), provides; see Section 2.1. This result extends to the nonstationary case, as stated in the following corollary.

COROLLARY 2. *The interpolator of the missing observations obtained with the skipping and the AO approaches are identical.*

As already mentioned, compared to the standard estimation of additive outliers in the series, the AO approach to interpolating missing values implies a correction in the likelihood. As stated in the following lemma, for a large enough number of observations, the effect of this correction becomes negligible.

LEMMA 3. *Let* M *denote the number of observations. If the location and the total number of missing values remain constant as* $M \to \infty$, *the determinantal correction vanishes.*

By Theorem 3, we can use the stationary series $\bar{u}$, obtained by differencing the filled in series $\bar{z}$, to evaluate the log-likelihood $\lambda(y_o)$. Hence, we can apply any of the fast algorithms existing in the literature to evaluate log–likelihoods of ARMA models. For example, the algorithm of Ansley (1979), the innovations algorithm of Brockwell and Davis (1987), or the Kalman filtering algorithm of Morf, Sidhu and Kailath (1974), as described by Pearlman (1980) and improved by Mélard (1984). We use an improved version of this last algorithm, detailed in Appendix A.

3.3 Regression Model with ARIMA Errors

Consider the regression model (2.11), where the vectors β and $y(t)$ are as in Section 2.3 and the residuals $\{\nu(t)\}$ follow the ARIMA model (2.3) with $z(t)$ replaced with $\nu(t)$. With the notation of the previous section, if we define the vector $\nu = (\nu(1), \ldots, \nu(N))'$ and the $N \times h$ matrix Y with the vectors $y'(t)$, $t = 1, \ldots, N$, as rows, we can write $\bar{z} = [X, Y][\omega', \beta']' + \nu$. Differencing this equation, we can proceed as in the previous section, the only difference being that the vector of regression parameters is now $[\omega', \beta']'$, instead of ω.

4. COMPUTATIONAL PERFORMANCE OF THE TWO APPROACHES

We have presented two approaches to the problem of optimal estimation of missing observations in possibly nonstationary time series. One uses first the Kalman filter for likelihood evaluation, skipping the missing observations, and applies then a smoothing algorithm to interpolate the unobserved values. This approach will be denoted the SK approach. The second approach fills the holes in the series with arbitrary numbers and treats them as additive outliers, with the likelihood function appropiately corrected. We shall refer to this as the AOC approach. Efficient and relatively simple ways to apply both approaches are detailed in Appendix A. It was seen how the two approaches are equivalent, so that they represent two alternative algorithms to compute the conditional expectation of the missing values given the available observations. While the SK approach avoids GLS estimation of the additive outlier parameters and requires less memory, the AOC approach uses a "complete" series so that differencing can take place and faster routines can be applied for likelihood evaluation. Thus, it is of interest to assess the relative performance

in practice of the two approaches. In the comparison, we shall include a third approach: the additive outlier approach without determinantal correction, to be denoted AON. As was seen in Section 3, this approach provides an asymptotic approximation and has the advantage that, since the likelihood considered is the standard additive outlier likelihood, it can be implemented with existing software.

We have run a simulation experiment on a 133 Pentium PC with series of length 100 generated from the following four models

$$\begin{array}{rrcl}
\text{AR(1)} & (1-.8B)z(t) & = & a(t) \\
\text{MA(1)} & z(t) & = & (1-.7B)a(t) \\
\text{ARIMA}(1,1,0) & (1-.8B)(1-B)z(t) & = & a(t) \\
\text{ARIMA(0,1,1)(0,1,1)} & (1-B)(1-B^{12})z(t) & = & (1-.4B)(1-.6B^{12})a(t),
\end{array}$$

where, $a(t) \sim N(0,1)$. To obtain each series, 600 observations were first generated using independent $N(0,1)$ deviates obtained by Box–Muller's method. Then, the first 500 observations of each series were discarded. We have considered three patterns of missing data: One missing observation (number 50), five consecutive missing observations (numbers $41 - -45$), and twenty missing observations (numbers $2,7,15,20,25$, $32,33,38,42,45,50,51,63$, $72,79,81,84$, $85,86,90$). The missing values have been obtained using three estimation procedures. The first corresponds to the SK approach, the second to the additive outlier approach with determinantal correction (AOC), and the third to the additive outlier approach without determinantal correction (AON). For each of the four models and for each possible combination of estimation procedure and pattern of missing data, we have performed 1000 simulations. All estimations have been made with the program TRAMO ("Time Series Regression with ARIMA Noise, Missing Observations, and Outliers "; Gómez and Maravall (1996), available from the first two authors upon request.)

Tables 1–4 below correspond to the four models. In each of these tables, we show the results of the three methods of estimation for each pattern of missing data. We have denoted by ME, RMSE, and TRMSE the mean error, root mean squared error and theoretical root mean squared error, respectively. The TRMSE's have been obtained by runnig the program with the specified models fixed and interpolating the missing values.

To facilitate interpretation of the tables, the following result is of help. For an infinite realization of an ARIMA series, the optimal interpolator of the missing values is a two-sided, convergent, filter. When there is only one missing observation, (at time t), the filter

is given by (see, for example, Brubacher and Wilson, 1976)

$$\nu(B,F) = -\sum_{k=1}^{\infty} \rho_k^{(i)}(B^k + F^k), \tag{4.1}$$

where $F = B^{-1}$, and $\rho_k^{(i)}$ is the k-lag autocorrelation of the inverse model of (2.3), namely

$$\theta(B)x(t) = \phi(B)\delta(B)a(t). \tag{4.2}$$

Further,

$$\text{RMSE}[\hat{z}(t)] = 1/\sigma_{(i)}, \tag{4.3}$$

where $\sigma_{(i)}$ is the standard deviation of $x(t)$ in model (4.2). In practice, this RMSE provides a lower bound for the RMSE of estimators in a finite sample. When close enough to the end of the series or to another missing value, the RMSE will, of course, be larger.

For the four models considered above, the ACF's of their inverses show that for the pure AR models (first and third model) convergence of (4.1) will occur with just one or two periods, respectively, at each side of t. The MA model and, in particular, the mixed model (i.e., models two and four) imply slower convergences, in accordance with the convergence properties of the expressions $(1-.7B)^{-1}$ and $(1-.6B^{12})^{-1}$. For the four models, expression (4.3) yields

AR(1):	$\text{RMSE}[(\hat{z}(t)] = .781,$
MA(1):	$\text{RMSE}[(\hat{z}(t)] = .714,$
ARIMA(1,1,0):	$\text{RMSE}[(\hat{z}(t)] = .453,$
ARIMA(0,1,1)(0,1,1):	$\text{RMSE}[(\hat{z}(t)] = .748.$

From Tables 1–4, it is seen that those (asymptotic) RMSE are identical to the TRMSE computed by the Kalman filter for the first three models when there is one missing observation. For the last model, the small discrepancy is caused by the fact that $(1-.6B^{12})^{-1}$ has not fully converged in 4 years. When there are 5 missing observations, the tables show the deterioration in RMSE caused by the presence of consecutive observations; this is particularly true for relatively simple models. When there are 20 missing values, Tables 1 and 3 show how for pure AR models, the filters converge fast, and the lower bound for the RMSE is often achieved. The MA model gets close on a few occasions, while the mixed model is always above. Comparing the four models, it is of some consolation however that for the case with RMSE systematically above the lower bound (the mixed model,) the deterioration due to consecutive missing values is markedly smaller.

Comparison of the SK and AOC columns reveals differences in the two alternative algorithms to compute the same conditional mean. Comparison of the AOC and the AON columns, in turn, shows the effect of the determinantal correction needed to compute the proper likelihood. The tables indicate that, for the pure AR models, the SK and the AOC approaches yield identical results, and that those of the AOC and AON approaches are nearly identical. For the MA and the mixed models, the SK and AOC approaches provide some differences, though small, and the same can be said of the AOC and AON approaches. Clearly, the results reflect the convergence properties of the "inverse" filters. The two equivalent algorithms yield identical results when the filters converge fast; when convergence is slow some differences may appear. In particular, for the more complex model (the mixed one,) the differences implied by not correcting the likelihood seem non-negligible.

The values of RMSE in the columns of Tables 1–4 are the averages over the 1000 replications of the RMSE provided as output of the Kalman filter. For the additive outlier approach without determinantal correction (the AON columns), the RMSE can be somewhat misleading because of the misspecification of the likelihood function implicit in this approach. Tables 5 and 6 present the Monte Carlo (MC) RMSE of the interpolators for all cases considered, computed, for each case, as the sample value over the 1000 replications . It is seen that the MC RMSE are always close to the theoretical RMSE. When the number of MO is small, the MC RMSE of the three methods (SK, AOC, AON) are practically identical; only for the more complex Airline model there is a slight difference and the AON approach tends to display slightly smaller precision when there are 5 missing values (the MC RMSE of the skipping approach is always smaller, except for the first observation, for which it is the same.)

When the number of missing observations increases to 20, it is seen that it still is true that for the simple AR models, the differences between the approaches are negligible. However, for the MA and mixed ARIMA model the skipping approach becomes noticeably better. For the MA model, 19 of the 20 missing values present larger MC RMSE (an average increase of about 6%); for the mixed ARIMA model, the 20 missing values are all better estimated with the skipping approach.

As for computational efficiency, Table 7 presents the elapsed times in seconds for the average of the simulations for all combinations: model for the series–patterns of missing data–estimation approach. In all cases, the AOC and AON approaches display negligible

differences. When there is only one missing observation, and if the model is small (any of the first three), there are practically no differences between the approaches. For the larger mixed model, the additive outlier approach is faster. When the number of missing observations increases to 5, for the small models the SK approach is slightly faster, while for the larger model, the additive outlier approach is still preferable. When the number of missing observations increases to 20, the SK approach is always much faster. (Notice that the fractions of seconds reported in Table 5, which include the printing of an output file, evidence the efficiency of the algorithms described in Appendix A).

Taken as whole, the results seem to indicate clearly the following. When there are few missing observations (1, even 5, in 100) the three approaches yield practically identical results, in terms of point estimators, their associated precision, and computational efficiency.

When the number of missing observations is large (20 in 100) the skipping approach becomes clearly preferable. It is considerably faster and yields more precise estimators. Further, from the precision point of view, enforcing the determinantal correction in the additive outlier approach may be important.

Table 1. *Model* $(1 - .8B)z(t) = a(t)$

observation	*SK*		*AOC*		*AON*		
number	*ME*	*RMSE*	*ME*	*RMSE*	*ME*	*RMSE*	*TRMSE*
			1 *missing observation*				
$n = 50$	.000	.767	.000	.767	.000	.767	.781
			5 *missing observations*				
$n = 41$	-.053	1.001	-.053	1.001	-.054	1.001	.979
42	-.039	1.219	-.039	1.219	-.039	1.219	1.211
43	-.007	1.278	-.007	1.278	-.007	1.278	1.274
44	-.017	1.216	-.017	1.216	-.017	1.216	1.211
45	.002	1.021	.002	1.021	.002	1.021	.979
			20 *missing observations*				
$n = 2$	.008	.776	.008	.776	.008	.775	.781
7	.004	.767	.004	.767	.004	.767	.781
15	-.046	.760	-.046	.760	-.045	.759	.781
20	.008	.765	.008	.765	.007	.765	.781
25	-.015	.793	-.015	.793	-.015	.792	.781
32	.004	.861	.004	.861	.004	.861	.895
33	-.025	.872	-.025	.872	-.026	.870	.895
38	-.010	.760	-.010	.760	-.011	.761	.781
42	-.010	.772	-.010	.772	-.010	.772	.781
45	.010	.804	.010	.804	.010	.803	.781
50	.015	.884	.015	.884	.013	.883	.895
51	.031	.909	.031	.909	.029	.907	.895
63	.016	.777	.016	.777	.017	.777	.781
72	.009	.745	.009	.745	.009	.745	.781
79	.000	.768	.000	.768	.000	.767	.781
81	.020	.787	.020	.787	.020	.788	.781
84	-.026	.943	-.026	.943	-.026	.941	.942
85	-.020	1.093	-.020	1.093	-.021	1.092	1.079
86	.005	.952	.005	.952	.004	.951	.942
90	-.044	.801	-.044	.801	-.045	.800	.781

SK: Skipping approach
AOC: Additive outlier approach with determinantal correction
AON: Additive outlier approach without determinantal correction

Table 2. Model $z(t) = (1 - .7B)a(t)$

observation	*SK*		*AOC*		*AON*		
number	*ME*	*RMSE*	*ME*	*RMSE*	*ME*	*RMSE*	*TRMSE*
			1 *missing observation*				
$n = 50$	-.003	.726	-.003	.726	-.003	.728	.714
			5 *missing observations*				
$n = 41$	-.050	1.033	-.050	1.033	-.049	1.033	1.000
42	.045	1.235	.045	1.235	.045	1.235	1.221
43	.026	1.200	.026	1.200	.026	1.200	1.221
44	-.023	1.206	-.023	1.206	-.023	1.206	1.221
45	.040	1.001	.041	1.001	.040	1.002	1.000
			20 *missing observations*				
$n = 2$	.014	.841	.019	.828	.018	.828	.828
7	-.018	.778	-.021	.753	-.021	.753	.726
15	-.015	.748	-.028	.728	-.028	.728	.726
20	.005	.744	-.007	.725	-.007	.725	.735
25	.011	.772	.014	.740	.014	.740	.727
32	.013	1.010	.007	1.003	.007	1.003	1.002
33	.001	.963	.001	.956	.001	.956	1.007
38	-.008	.774	-.003	.755	-.003	.755	.746
42	.018	.787	.021	.776	.021	.776	.781
45	.037	.794	.040	.784	.040	.784	.770
50	.002	1.008	.004	.995	.004	.995	1.007
51	-.017	1.008	-.021	.995	-.021	.995	1.000
63	-.013	.764	-.026	.725	-.026	.726	.715
72	.011	.757	.009	.717	.009	.717	.717
79	-.015	.823	-.009	.812	-.009	.812	.821
81	.028	.853	.032	.847	.032	.847	.860
84	-.036	1.049	-.034	1.046	-.034	1.046	1.033
85	.019	1.223	.019	1.223	.019	1.223	1.221
86	.007	1.029	.005	1.026	.005	1.026	1.016
90	-.017	.781	-.019	.748	-.019	.748	.736

SK: Skipping approach
AOC: Additive outlier approach with determinantal correction
AON: Additive outlier approach without determinantal correction

Table 3. *Model* $(1 - .8B)(1 - B)z(t) = a(t)$

observation	*SK*		*AOC*		*AON*		
number	*ME*	*RMSE*	*ME*	*RMSE*	*ME*	*RMSE*	*TRMSE*
			1 *missing observation*				
$n = 50$	-.008	.449	-.008	.449	-.008	.449	.453
			5 *missing observations*				
$n = 41$	-.034	.817	-.034	.817	-.034	.817	.801
42	-.044	1.286	-.044	1.286	-.044	1.286	1.298
43	-.018	1.453	-.018	1.453	-.018	1.453	1.476
44	-.008	1.273	-.008	1.273	-.008	1.272	1.298
45	.008	.797	.008	.797	.008	.797	.801
			20 *missing observations*				
$n = 2$	-.013	.483	-.013	.483	-.013	.482	.486
7	-.001	.438	-.001	.438	-.001	.438	.453
15	-.011	.429	-.011	.429	-.011	.429	.453
20	-.004	.446	-.004	.446	-.004	.446	.453
25	-.017	.459	-.017	.459	-.017	.459	.453
32	.016	.591	.016	.592	.017	.592	.605
33	.002	.597	.002	.597	.003	.597	.605
38	-.012	.457	-.012	.457	-.012	.457	.453
42	-.010	.451	-.010	.451	-.010	.451	.453
45	.011	.470	.011	.470	.011	.470	.453
50	-.009	.592	-.009	.592	-.009	.592	.605
51	-.001	.587	-.001	.587	-.002	.587	.605
63	.018	.443	.018	.443	.018	.444	.453
72	-.004	.443	-.004	.443	-.004	.443	.453
79	.008	.457	.008	.457	.008	.456	.459
81	.001	.457	.001	.457	.001	.458	.459
84	-.022	.679	-.022	.679	-.022	.678	.697
85	-.034	.897	-.034	.897	-.034	.897	.919
86	-.017	.679	-.017	.679	-.017	.679	.697
90	-.023	.453	-.023	.453	-.023	.453	.453

SK: Skipping approach
AOC: Additive outlier approach with determinantal correction
AON: Additive outlier approach without determinantal correction

Table 4. *Model* $(1-B)(1-B^{12})z(t) = (1-.4B)(1-.6B^{12})a(t)$

observation	*SK*		*AOC*		*AON*		
number	*ME*	*RMSE*	*ME*	*RMSE*	*ME*	*RMSE*	*TRMSE*
			1 *missing observation*				
$n = 50$	.002	.753	.001	.753	.001	.753	.751
			5 *missing observations*				
$n = 41$	-.041	.854	-.043	.854	-.043	.855	.837
42	-.041	.915	-.042	.915	-.042	.915	.905
43	-.025	.933	-.024	.934	-.024	.934	.927
44	-.056	.933	-.056	.933	-.056	.932	.905
45	-.029	.872	-.029	.872	-.029	.872	.837
			20 *missing observations*				
$n = 2$	-.020	.918	-.019	.917	-.019	.922	.884
7	-.002	.861	.000	.859	.003	.866	.849
15	-.051	.788	-.052	.786	-.050	.789	.792
20	.000	.834	-.004	.827	.003	.834	.814
25	-.014	.770	-.013	.770	-.020	.776	.772
32	-.009	.799	-.011	.795	-.004	.817	.826
33	-.032	.825	-.032	.824	-.029	.835	.818
38	-.013	.770	-.014	.770	-.016	.784	.788
42	-.013	.749	-.014	.749	-.015	.753	.759
45	-.009	.789	-.009	.788	-.006	.791	.780
50	.004	.822	.003	.820	.005	.832	.815
51	.022	.841	.019	.838	.022	.851	.810
63	.007	.771	.007	.771	.006	.776	.777
72	.018	.759	.018	.759	.019	.769	.786
79	-.010	.773	-.009	.772	-.009	.780	.790
81	.017	.805	.016	.804	.018	.815	.791
84	-.025	.875	-.025	.875	-.032	.881	.865
85	-.024	.906	-.023	.906	-.029	.913	.874
86	-.002	.881	-.002	.881	-.005	.888	.847
90	-.046	.888	-.047	.887	-.051	.894	.846

SK: Skipping approach
AOC: Additive outlier approach with determinantal correction
AON: Additive outlier approach without determinantal correction

Table 5. *MONTE CARLO MEAN SQUARE ERRORS*

observation	AR(1)			MA(1)		
number	*SK*	*AOC*	*AON*	*SK*	*AOC*	*AON*
			1 *missing observation*			
$n = 50$	.772	.772	.772	.715	.715	.716
			5 *missing observations*			
$n = 41$	.950	.950	.950	.965	.965	.966
42	1.155	1.155	1.155	1.229	1.229	1.229
43	1.251	1.251	1.251	1.212	1.212	1.212
44	1.205	1.205	1.205	1.194	1.194	1.194
45	1.016	1.016	1.016	.964	.964	.965
			20 *missing observations*			
$n = 2$	.761	.761	.760	.845	.874	.873
7	.788	.788	.786	.765	.840	.837
15	.762	.762	.762	.737	.835	.832
20	.799	.799	.799	.734	.823	.819
25	.781	.781	.781	.752	.821	.818
32	.912	.912	.911	1.021	1.053	1.048
33	.900	.900	.901	1.041	1.064	1.065
38	.762	.762	.761	.773	.817	.818
42	.773	.773	.774	.807	.882	.883
45	.792	.792	.792	.758	.818	.812
50	.884	.884	.884	1.007	1.030	1.029
51	.898	.898	.897	1.002	1.030	1.034
63	.757	.757	.758	.745	.810	.813
72	.805	.805	.804	.765	.852	.847
79	.815	.815	.816	.828	.870	.867
81	.753	.753	.753	.895	.940	.939
84	.951	.951	.952	1.032	1.051	1.050
85	1.066	1.066	1.066	1.269	1.269	1.269
86	.921	.921	.917	1.021	1.053	1.054
90	.788	.788	.789	.759	.836	.833

SK: Skipping approach
AOC: Additive outlier approach with determinantal correction
AON: Additive outlier approach without determinantal correction

Table 6. *MONTE CARLO MEAN SQUARE ERRORS*

observation	ARI(1,1)			ARIMA(0,1,1)(0,1,1)		
number	*SK*	*AOC*	*AON*	*SK*	*AOC*	*AON*
			1 *missing observation*			
$n = 50$	.451	.451	.451	.744	.744	.744
			5 *missing observations*			
$n = 41$	.787	.787	.788	.834	.834	.834
42	1.254	1.254	1.253	.886	.886	.887
43	1.427	1.427	1.426	.920	.920	.921
44	1.266	1.266	1.266	.895	.895	.896
45	.793	.793	.793	.851	.851	.853
			20 *missing observations*			
$n = 2$	.477	.477	.477	.924	.924	.936
7	.436	.436	.436	.853	.854	.869
15	.450	.450	.450	.793	.793	.802
20	.462	.462	.462	.838	.838	.850
25	.458	.458	.458	.769	.769	.778
32	.608	.608	.608	.846	.846	.855
33	.611	.611	.611	.819	.819	.830
38	.452	.452	.452	.781	.780	.784
42	.451	.451	.451	.755	.755	.770
45	.464	.464	.464	.788	.788	.793
50	.598	.598	.598	.809	.810	.822
51	.599	.599	.599	.822	.822	.839
63	.457	.457	.457	.768	.768	.783
72	.469	.469	.469	.825	.825	.834
79	.486	.486	.486	.826	.826	.832
81	.433	.433	.433	.781	.781	.789
84	.711	.711	.711	.909	.909	.924
85	.904	.904	.903	.896	.896	.909
86	.672	.672	.672	.852	.852	.868
90	.458	.458	.458	.861	.861	.875

SK: Skipping approach
AOC: Additive outlier approach with determinantal correction
AON: Additive outlier approach without determinantal correction

Table 7. Elapsed time in seconds (average)

approach	AR(1)	MA(1)	ARIMA(1,1,0)	ARIMA(0,1,1)(0,1,1)
		1 *missing observations*		
SK	.06	.07	.07	.52
AOC	.08	.08	.08	.17
AON	.09	.09	.08	.19
		5 *missing observations*		
SK	.06	.08	.07	.55
AOC	.10	.14	.11	.26
AON	.11	.12	.11	.33
		20 *missing observations*		
SK	.07	.07	.07	.65
AOC	.36	.33	.32	1.02
AON	.38	.33	.33	.95

SK: Skipping approach
AOC: Additive outlier approach with determinantal correction
AON: Additive outlier approach without determinantal correction

APPENDIX A: COMPUTATIONAL DETAILS

In Sections 2 and 3 we developed two equivalent approaches to the problem of maximum likelihood estimation of parameters and interpolation of missing values in general regression models with nonstationary ARIMA errors when some of the observations may be missing. Both procedures are based on the definition of a conditional likelihood, which is particularly well suited for efficient computation. In this appendix we provide the computational details of both procedures; they are implemented in the program TRAMO, mentioned in Section 4, and available from the first two authors upon request.

A.1 Skipping Approach and the Augmented Kalman Filter

In order to evaluate the log–likelihood (2.10), we use the state space representation

$$x(t) = Fx(t-1) + Ga(t)$$
$$z(t) = H'(t)x(t) + \alpha(t)W(t)$$

defined in Sections 2.1 and 2.2 and an "Augmented Kalman Filter" (AKF) algorithm which we now describe. The log–likelihood (2.10) is that of the GLS model (2.9), where $Var(v_o) = \sigma^2\Omega_{v_o}$. Given the parameters $(\phi,\theta) = (\phi_1,\ldots,\phi_p,\theta_1,\ldots,\theta_q)$, of the ARMA model (2.1), the log–likelihood (2.10) is maximized with respect to z_{Im} and σ^2 by replacing them with their maximum likelihood estimators $\hat{z}_{Im}$ and $\hat{\sigma}^2$, respectively, which coincide with the GLS estimators of model (2.9). Minus two times the (z_{Im},σ^2)–maximized log–likelihood is, apart from a constant, $S(y_o) = \mid \Omega_{v_o} \mid^{1/(M-k)}(y_o - C_o\hat{z}_{Im})'\Omega_{v_o}(y_o - C_o\hat{z}_{Im})$, and maximizing (2.10) is equivalent to minimizing $S(y_o)$. Let $\Omega_{v_o} = LL'$, where L is a lower triangular matrix, be the Cholesky decomposition of Ω_{v_o}. If we left–multiply (2.9) by the matrix L^{-1}, we obtain the ordinary least squares (OLS) model

$$L^{-1}y_o = L^{-1}C_o z_{Im} + L^{-1}v_o \qquad (A.1)$$

and the function $S(y_o)$ can be rewritten as the nonlinear sum of squares

$$S(y_o) = \tilde{e}'\tilde{e}, \qquad (A.2)$$

where $\tilde{e} = \mid L \mid^{1/(M-k)} L^{-1}(y_o - C_o\hat{z}_{Im})$. Therefore, if we can evaluate $\tilde{e}$, we can use a Gauss–Marquardt algorithm to minimize $S(y_o)$ with respect to (ϕ,θ).

If there are no missing observations among the first d values of the series, then $z_{Io} = z_I$, $y_o = z_{IIo} - A_o z_I$, model (2.9) becomes $y_o = v_o$ and the Kalman filter can be used to compute $L^{-1}y_o$ and $| L |$. The Kalman filter is the set of recursions

$$\left|\begin{array}{r} e(t) = z(t) - H'(t)x(t \mid t-1) \\ \sigma^2(t \mid t-1) = H'(t)\Sigma(t \mid t-1)H(t) + \alpha^2(t) \\ K(t) = F\Sigma(t \mid t-1)H(t)/\sigma^2(t \mid t-1) \\ x(t+1 \mid t) = Fx(t \mid t-1) + K(t)e(t) \\ \Sigma(t+1 \mid t) = (F - K(t)H'(t))\Sigma(t \mid t-1)F' + GG', \quad t = d+1, \ldots, N, \end{array}\right.$$

with starting conditions $x(d+1 \mid d) = A_* z_I$ and $\Sigma(d+1 \mid d) = \Xi_* \tilde{\Sigma}(d+1 \mid d)\Xi_*'$, as in Section 2.2. Here $x(t \mid t-1)$ is the predictor of $x(t)$ using $\{z(s) : 1 \leq t_1 \leq s \leq t-1\}$ and $Var(x(t) - x(t \mid t-1)) = \Sigma(t \mid t-1)$. If the series is stationary, then $x(t)$ is that of Section 2.1, $d = 0$, $x(d+1 \mid d) = E(x(1)) = 0$ and $\Sigma(d+1 \mid d) = Var(x(1))$. The $e(t)$ are the innovations or errors of predicting $z(t)$ using $\{z(s) : 1 \leq t_1 \leq s \leq t-1\}$; they constitute an orthogonal sequence with zero mean and $Var(e(t)) = \sigma^2(t \mid t-1)$, as given by the Kalman filter. Only those $e(t)$ and $\sigma^2(t \mid t-1)$ are computed for which the corresponding observation $z(t)$ is not missing. The elements $e(t)/\sigma(t \mid t-1)$ corresponding to the observed $z(t)$ form the vector $L^{-1}y_o$ and $| L |$ is simply the product of the $\sigma(t \mid t-1)$ for which $z(t)$ is not missing. Note that in the Kalman filter we suppose that $\sigma^2 = 1$ because σ^2 can be concentrated out of the log-likelihood.

For the general case, when there may be missing observations among the first d values of the series, Kohn and Ansley (1985) show that the Kalman filter can be applied to the vector y_o and the columns of the matrix C_o to obtain $L^{-1}y_o$ and $L^{-1}C_o$, respectively, as well as $| L |$. The AKF we apply uses a state vector for the data y_o and a state vector for each of the columns of the matrix C_o. Define $X(t)$ as the matrix whose columns are all those state vectors, in the order specified above, that is, first the state vector for the data y_o and then the state vectors for the columns of C_o. The AKF is the Kalman filter with the equations for $e(t)$ and $x(t \mid t-1)$, respectively, replaced with

$$E(t) = (z(t), 0, \ldots, 0) - H'(t)X(t \mid t-1), \quad X(t+1 \mid t) = FX(t \mid t-1) + K(t)E(t),$$

with the starting condition $X(d+1 \mid d) = (B_* z_{Io}, -C_*)$, where B_* and C_* are defined by the relation $A_* z_I = B_* z_{Io} + C_* z_{Im}$ and A_* is the matrix defined in Section 2.2. Note that the AKF automatically builds the vector $y_o = z_{IIo} - B_o z_{Io}$ and the matrix C_o and,

therefore, it is not necessary to compute them before it is applied. Once the AKF starts to run, if $z(t)$ is missing, then $E(t)$ is not computed and both $E(t)$ and $\sigma^2(t \mid t-1) = 1$ do not contribute to the calculation of either $(L^{-1}y_o, L^{-1}C_o)$ or $\mid L \mid$. The elements of $E(t)/\sigma(t \mid t-1)$ corresponding to the $z(t)$ actually observed constitute the rows of the matrix $(L^{-1}y_o, L^{-1}C_o)$. Similarly, $\mid L \mid$ is equal to the product of the $\sigma(t \mid t-1)$ corresponding to the observed values.

After computing the matrix $L^{-1}C_o$ with the AKF, the QR algorithm can be applied to obtain an orthogonal matrix Q such that $Q'L^{-1}C_o = [R', 0']'$, with R upper triangular. If we partition $Q = [Q_1', Q_2']'$ conforming to R and 0 in $[R', 0']'$, then, left-multiplying (A.1) by Q', it is obtained that

$$\begin{aligned} Q_1'L^{-1}y_o &= Rz_{Im} + Q_1'L^{-1}v_o \\ Q_2'L^{-1}y_o &= \qquad\quad Q_2'L^{-1}v_o. \end{aligned} \tag{A.3}$$

We assume in (A.3) that the matrix R has full rank, and refer the reader to Gómez and Maravall (1994) for the case in which there is a rank deficiency. From (A.3), the maximum likelihood estimators of z_{Im} and σ^2 can be obtained as $\hat{z}_{Im} = R^{-1}Q_1'L^{-1}y_o$ and $\hat{\sigma}^2 = (Q_2'L^{-1}y_o)'Q_2'L^{-1}y_o$. substituting in (A.2) yields $S(y_o) = \tilde{a}'\tilde{a}$, where $\tilde{a} = \mid L \mid^{1/(M-k)} Q_2'L^{-1}y_o$. Note that, if the ARMA model (2.1) is the true model, then $Q_2'L^{-1}y_o$ is distributed as $N(0, \sigma^2 I_{M-k})$ and this vector can be used as residuals to test model adequacy. If the series is stationary, then the AKF reduces to the ordinary Kalman filter, as in Jones (1980). It is worth noticing that the AKF we present is similar to the Diffuse Kalman Filter (DKF) of De Jong (1991). The difference lies in that the AKF does not use the recursion of the DKF that accumulates the partial sums of squares and crossproducts. We believe that it is numerically safer to use the Cholesky decomposition of Ω_{y_o} to move from (2.9) to (A.1), and then apply the QR algorithm in order to obtain the GLS estimators $\hat{z}_{Im}$ and σ^2.

Once the parameters (ϕ, θ) of the ARMA model (2.1) have been estimated, the AKF, and a simplification thereof, can be used to predict future values and to interpolate missing values, respectively. Specifically, let $z(t \mid N) = E(z(t) \mid z(t_1), \ldots, z(t_M))$ be the estimator of $z(t)$ using the observed series $z_o = (z(t_1), \ldots, z(t_M))'$. If $t < t_M = N$, we are interpolating and if $t > t_M$, we are predicting. Given the definition of the state vector $x(t)$ in Sections 2.1 and 2.2, it is easy to check that $z(t \mid N) = H'x(t \mid N)$, where $H' = (1, 0, \ldots, 0)$ and $x(t \mid N) = E(x(t) \mid z(t_1), \ldots, z(t_M))$.

Consider next predicting the state $x(t)$ using $\{z(s) : t_1 \leq s \leq t-1\}$ and let $x(t \mid t-1)$

be this predictor. Suppose first that z_{Im} is known. Then, it is not difficult to verify that $X(t \mid t-1)[1, -z'_{Im}]'$, where $X(t \mid t-1)$ is given by the AKF, satisfies the same recursion and starting condition as $x(t \mid t-1)$ in the ordinary Kalman filter. Hence, $x(t \mid t-1) = X(t \mid t-1)[1, -z'_{Im}]'$.

If z_{Im} is not known and we estimate it with $\hat{z}_{Im}$, obtained from (A.3), we can proceed in two steps. First, we predict $x(t)$ using $\{z_{Im}, \{z(s) : t_1 \leq s \leq t-1\}\}$, obtaining $\tilde{x}(t \mid t-1) = X(t \mid t-1)[1, -z'_{Im}]'$, and then we predict this predictor using $\{z(s) : t_1 \leq s \leq t-1\}$. Thus, $x(t \mid t-1) = X(t \mid t-1)[1, -\hat{z}'_{Im}]'$. Given the orthogonality of $x(t) - \tilde{x}(t \mid t-1)$ and $\tilde{x}(t \mid -1) - x(t \mid t-1)$, the mean squared error of $x(t \mid t-1)$ is

$$\begin{aligned} Mse(x(t \mid t-1)) &= Var(x(t) - \tilde{x}(t \mid t-1) + \tilde{x}(t \mid t-1) - x(t \mid t-1)) \\ &= Var(x(t) - \tilde{x}(t \mid t-1)) + Var(\tilde{x}(t \mid t-1) - x(t \mid t-1)) \\ &= \sigma^2 \Sigma(t \mid t-1) + Var\left(X(t \mid t-1)[0, (z_{Im} - \hat{z}_{Im})']'\right) \\ &= \sigma^2 \Sigma(t \mid t-1) + X_{Im}(t \mid t-1) Var(\hat{z}_{Im}) X'_{Im}(t \mid t-1), \end{aligned}$$

where $X_{Im}(t \mid t-1)$ is the submatrix of $X(t \mid t-1)$ formed with all its columns except the first.

If we are predicting, in order to obtain $z(N+i \mid N)$, $i = 1, 2, \ldots$, we use the recursions

$$X(N+i \mid N) = FX(N+i-1 \mid N), \quad \Sigma(N+i \mid N) = F\Sigma(N+i-1 \mid N)F' + GG',$$

initialized with $X(N+1 \mid N)$ and $\Sigma(N+1 \mid N)$, as given by the AKF. The estimator of $z(N+i)$ is $z(N+i \mid N) = H'x(N+i \mid N)$, where $x(N+i \mid N) = X(N+i \mid N)[1, -\hat{z}'_{Im}]'$. Its mean squared error is $Mse(z(N+i \mid N)) = H'Mse(x(N+i \mid N))H$, where $Mse(x(N+i \mid N)) = \sigma^2 \Sigma(N+i \mid N) + X_{Im}(N+i \mid N) Var(\hat{z}_{Im}) X'_{Im}(N+i \mid N)$ and $X_{Im}(N+i \mid N)$ is the submatrix of $X(N+i \mid N)$ formed with all its columns except the first.

If we are interpolating the missing observation $z(j)$, where $z(j)$ is not contained in z_{Im} or, equivalently, $j > d$, we can proceed as follows. Define first an Augmented FPS for $k \geq j$ by the equations

$$\Sigma^a(k+1 \mid k) = \Sigma^a(k \mid k-1)(F - K(k)H'(k))', \quad K^a(k) = \Sigma^a(k \mid k-1)H(k)\sigma^{-2}(k \mid k-1),$$

$$X(j \mid k) = X(j \mid k-1) + K^a(k)E(k), \quad \Sigma(j \mid k) = \Sigma(j \mid k-1) - \Sigma^a(k \mid k-1)H(k)(K^a(k))',$$

initialized with $\Sigma^a(j \mid j-1) = \Sigma(j \mid j-1)$. Here, $K(k)$, $\sigma^2(k \mid k-1)$, $E(k)$, $X(j \mid j-1)$ and $\Sigma(j \mid j-1)$ are those of the AKF. As in the case of prediction, it is easy to verify

that the interpolator $x(j \mid k)$ of the state $x(j)$ using $\{z(s) : t_1 \leq s \leq k\}$ is $x(j \mid k) = X(j \mid k)[1, -\hat{z}'_{Im}]'$. Also, $z(j \mid k) = H'x(j \mid k)$ and $Mse(z(j \mid k)) = H'Mse(x(j \mid k))H$. Hence, if we define $b(k) = H'K^a(k)$, $v'(j \mid k) = H'\Sigma^a(k \mid k-1)$ and $\sigma^2(j \mid k) = H'\Sigma(j \mid k)H$, we obtain, for $k = j, j+1, \ldots, N$, the simplified equations

$$v'(j \mid k+1) = v'(j \mid k)(F - K(k)H'(k))', \quad b(k) = v'(j \mid k)H(k)\sigma^{-2}(k \mid k-1),$$

$$H'X(j \mid k) = H'X(j \mid k-1) + b(k)E(k), \quad \sigma^2(j \mid k) = \sigma^2(j \mid k-1) - v'(j \mid k)H(k)b(k),$$

initialized with $v'(j \mid j) = H'\Sigma(j \mid j-1)$. These equations only require the storage of two scalars, $b(k)$ and $\sigma^2(j \mid k)$, and two vectors, $v'(j \mid k)$ and $H'X(j \mid k)$. When $k = N$, we obtain the interpolator of $z(j)$, which is $z(j \mid N) = H'x(j \mid N)$, where $x(j \mid N) = X(j \mid N)[1, -z'_{Im}]'$. Its mean squared error is $Mse(z(j \mid N)) = H'Mse(x(j \mid N))H = \sigma^2\{\sigma^2(j \mid N)\} + H'X_{Im}(j \mid N)Var(\hat{z}_{Im})X'_{Im}(j \mid N)H$ and $X_{Im}(j \mid N)$ is the submatrix of $X(j \mid N)$ formed with all its columns except the first.

Finally, we consider the regression model (2.11) with the state space representation defined in Section 2.3. The log-likelihood of this model was defined as that of the GLS model (2.12). To evaluate the log-likelihood, we use the AKF with a matrix $X(t)$ which includes h new columns added to the right of the existing ones, corresponding to states for the columns of the $Y_{II_o} - A_oY_I$ matrix. Also, the equation for $E(t)$ is replaced with $E(t) = (z(t), 0, y'(t)) - H'(t)X(t \mid t-1)$, where 0 is a $1 \times (d-k)$ vector corresponding to the columns of the C_o matrix, and the initialization for $X(d+1 \mid d)$ is replaced with $X(d+1 \mid d) = (B_*z_{I_o}, -C_*, -A_*Y_I)$, where Y_I is the matrix defined in Section 2.3 and A_* is the $r \times d$ matrix defined in Section 2.2. Note that it is not necessary to evaluate the $Y_{II_o} - A_oY_I$ matrix because the AKF builds it automatically. We can now proceed as before for log-likelihood evaluation, prediction or interpolation.

A.2 Additive Outlier Approach and Augmented Morf-Sidhu-Kailath Filter

With the notation of Section 3.3, consider the regression model $\bar{z} = [X, Y][\omega', \beta']' + \nu$. By Theorem 3, we can work with the differenced series and we showed in Section 2.2 that differencing a series is equivalent to multiplying it by the left by the matrix J_{II} defined in that section. Hence, we consider the model

$$\bar{u} = [X^*, Y^*][\omega', \beta']' + \nu^*, \tag{A.4}$$

where $\bar{u} = J_{II}\bar{z}$, $X^* = J_{II}X$, $Y^* = J_{II}Y$ and $\nu^* = J_{II}\,\nu$.

In Section A.1, we considered the GLS model (2.9) or, more generally, (2.12) and developed a procedure to estimate the parameters of the ARIMA model, the regression parameters and the variance of the innovations. Clearly, this methodology can be applied to any regression model with ARIMA errors for the same purposes; only the meaning of the regression parameters will change. In particular, we can apply it to (A.4), and use an augmented version of the Kalman filtering algorithm of Morf, Sidhu and Kailath (1974) (in short AMSK), because in (A.5) there are no missing values and the errors are stationary. The algorithm of Morf, Sidhu and Kailath (1974) is described in Pearlman (1980) and has been improved by Mélard (1984). We use a similar algorithm, suited to our state space representation (2.2), which includes both the improvement for $p > q$ of Mélard (1984) and the improvement for seasonal moving average models due to Ansley (1979), discussed, but not implemented, also in Mélard (1984).

We describe first the basic AMSK, independently of the state space representation, and explain later how we can improve this algorithm by using the state space representation (2.2). Let $X(t)$ be the matrix whose columns are the state vectors corresponding to the data $\bar{u}$ and the columns of the matrix $[X^*, Y^*]$, in that order. To simplify the notation, define $R(t) = H'\Sigma(t \mid t-1)H$, where $\Sigma(t \mid t-1)$ is that of the Kalman filter corresponding to the ARMA model (2.1) and $H = (1, 0, \ldots, 0)'$. Then, the AMSK is the set of recursions

$$\left\{\begin{array}{r} E(t) = (\bar{u}(t), x^{*'}(t), y^{*'}(t)) - H'X(t \mid t-1) \\ K(t+1) = K(t) - (H'L(t)/R(t))FL(t)/R(t) \\ L(t+1) = FL(t) - H'L(t)K(t) \\ R(t+1) = R(t) - (H'L(t))^2/R(t) \\ X(t+1 \mid t) = FX(t \mid t-1) + K(t)E(t), \quad t = d+1, \ldots, N, \end{array}\right.$$

with starting conditions $X(d+1 \mid d) = 0$, $L(d+1) = F\Sigma(d+1 \mid d)H$, $K(d+1) = L(d+1)/R(d+1)$ and $R(d+1) = H'\Sigma(d+1 \mid d)H = H'Var(x(d+1))H$. Here $x^{*'}(t)$ and $y^{*'}(t)$ are the rows of X^* and Y^*, respectively, $x(t)$ is the state vector corresponding to the ARMA model (2.1) with z replaced with ν^*, and $K(t)$ is that of the Kalman filter corresponding to the ARMA model (2.1).

By Theorem 3, we have to use a correction in the determinantal term for log-likelihood evaluation. The computation of the correction term $\mid X_{II}^{*'}\Omega^{-1}X_{II}^{*} \mid$, where $Var(\nu^*) = \sigma^2\Omega$, can be facilitated if we change the order of the elements of the vector of regression

parameters ω. Specifically, put $\omega = [\omega_{II}', \omega_I']'$,

$$X = \begin{bmatrix} X_{II} & 0 \\ 0 & X_I \end{bmatrix},$$

and $X^* = J_{II}X$. If $\Omega = LL'$, with L lower triangular, is the Cholesky decomposition of the matrix Ω, applying the QR algorithm to the matrix $L^{-1}[X^*, Y^*]$ yields an orthogonal matrix Q such that $Q'L^{-1}[X^*, Y^*] = [R', 0']'$, with R upper triangular. If we partition R conforming to $[\omega_{II}', \omega_I']'$ and R_{II} is the upper triangular submatrix of R corresponding to ω_{II} in the partition, then $\mid X_{II}^{*'}\Omega^{-1}X_{II}^* \mid = \mid R_{II}'R_{II} \mid$. Thus, the determinantal correction is a simple by-product of the computations in the standard additive outlier case.

For prediction, we can use the AMSK as in Section A.1 and for interpolation, by the results of Section 3, once we have estimated the regression parameters, the interpolations of the missing observations are simply the difference between the tentative values assigned by the user and the estimators $\hat{\omega}_i$, which are the elements of $\hat{\omega}$. The Mse of the interpolators are obtained by GLS.

The AMSK represents a significant improvement with respect to the AKF, since the matrix recursion for $\Sigma(t \mid t-1)$ in the AKF has been replaced with a vector recursion, that for $L(t)$, in the AMSK, but it can be improved still further if we pay attention to the covariance structure of the ARMA model (2.1). To see this, suppose the series is stationary, with no missing observations and no regression parameters. Then, there is no need for either differencing or filling in the series and the above regression model reduces to $z = \nu$, where $\{\nu(t)\}$ follows the ARMA model (2.1) and $z = (z(1), \ldots, z(N))'$ is the observed series. Hence, the AMSK becomes the Kalman filtering algorithm of Morf, Sidhu and Kailath (1974), which is the AMSK with the vector $E(t)$ and the matrix $X(t \mid t-1)$ replaced with the scalar $e(t)$ and the vector $x(t \mid t-1)$, respectively. As in the Kalman filter, the $e(t) = z(t) - H'x(t \mid t-1)$ are the innovations, where $x(t \mid t-1) = E(x(t) \mid z(s) : 1 \leq s \leq t-1)$ and $x(t)$ is the state vector defined in Section 2.1. The innovations $e(t)$ constitute an orthogonal sequence with $E(e(t)) = 0$ and $Var(e(t)) = R(t)$, as given by the AMSK.

We now show that if $e = (e(1), \ldots, e(N))'$, then there exists a lower triangular matrix K with ones in the main diagonal such that $z = Ke$ and $\Omega = KDK'$, where $Var(\nu) = \sigma^2\Omega$ and $D = \text{diag}\,(R(1), \ldots, R(N))$. From the observation equation $z(t+1) = H'x(t+1)$ and the relation $x(t+1 \mid t) = Fx(t \mid t-1) + K(t)e(t)$, given by the AMSK, it is obtained that

$z(t+1) = e(t+1) + H'K(t)e(t)$. Iterating, we have

$$\begin{aligned} z(t+1) =& e(t+1) + H'K(t)e(t) + H'FK(t)e(t-1) + \dots \\ &+ H'F^{t-2}K(2)e(2) + H'F^{t-1}e(1), \quad t \geq 1, \end{aligned}$$

with $z(1) = e(1)$. Therefore, K is the lower triangular matrix

$$K = \begin{bmatrix} 1 & & & & \\ H'K(1) & 1 & & & \\ H'FK(1) & H'K(2) & 1 & & \\ \vdots & \vdots & \vdots & \ddots & \\ H'F^{N-2}K(1) & H'F^{N-3}K(2) & \dots & H'K(N-1) & 1 \end{bmatrix}.$$

Given the definitions of the matrices H and F, it is straightforward to check that $H'K(s) = K_1(s)$, $H'FK(s) = K_2(s), \dots, H'F^{r-1}K(s) = K_r(s)$, where $r = max\{p, q+1\}$ and $K_i(s)$ is the i-th component of $K(s), i = 1, \dots, r$. The elements of the matrix K satisfy certain relations due to the covariance structure of the ARMA model, which can be made explicit if we use the transformation of Ansley (1979). In this transformation, the new variables $\omega(t)$ are defined by

$$\omega(t) = \begin{cases} z(t) & \text{if } t = 1, \dots, m, \\ \phi(B)z(t) & \text{if } t = m+1, \dots, N, \end{cases} \tag{A.5}$$

where $\phi(B)$ is the autoregressive polynomial in model 2.1 and $m = max\{p, q\}$. If $\omega = (\omega(1), \dots, \omega(N))'$, then we can write $\omega = Jz$, where $J = [J_1', J_2']'$, $J_1 = [I_m, 0]$, I_m is the identity matrix of rank m and

$$J_2 = \begin{bmatrix} \phi_m & \dots & \phi_1 & 1 & & 0 \\ & \ddots & & & \ddots & \\ 0 & & \phi_m & \dots & \phi_1 & 1 \end{bmatrix},$$

with the convention $\phi_i = 0$ if $i > p$. Let $L = JK$. Then, $Var(\omega) = \sigma^2 LDL'$ is a band matrix with maximum bandwith m for the first m rows and q thereafter and L is a lower triangular band matrix with ones in the main diagonal and bandwidths corresponding to those of $Var(\omega)$; see Ansley (1979). This implies the relations $0 = \phi_p K_{q-p+1}(t) + \dots + \phi_1 K_q(t) + K_{q+1}$, $t \geq 1$ if $q \geq p$ and

$$\begin{cases} 0 = \phi_p + \phi_{p-1}K_1(t) + \dots + \phi_1 K_{p-1}(t) + K_p(t,) & t \geq 1 \\ 0 = \phi_{p-1} + \phi_{p-2}K_1(t) + \dots + \phi_1 K_{p-2}(t) + K_{p-1}(t), & t \geq 2 \\ \quad \cdot \\ 0 = \phi_{q+1} + \phi_q K_1(t) + \dots + \phi_1 K_q(t) + K_{q+1}(t), & t \geq p - q \end{cases}$$

if $p > q$.

If we insert these relations into the recursions of the AMSK, it is not difficult to verify that the following simplification in the AMSK is obtained.

For $t = 1, \ldots, p-q$ use the AMSK (only if $p > q$).
For $t = p-q+1, p-q+2, \ldots$ if $p > q$ or $t = 1,2, \ldots$ if $p \leq q$
  For $i = 1, \ldots, q$
    Update, using the recursions of the AMSK, the elements $K_i(t)$, $L_i(t)$ of $K(t)$, $L(t)$, respectively, and the i-th row $X_i(t \mid t-1)$ of $X(t \mid t-1)$.
  For $i = q+1$
    Compute $L_{q+1}(t) = -\phi_q L_1(t) - \cdots - \phi_1 L_q(t)$ and $X_{q+1}(t \mid t-1) = -\phi_r v(t-r+q) - \cdots - \phi_{q+1} v(t-1) - \phi_q X_1(t \mid t-1) - \cdots - \phi_1 X_q(t \mid t-1)$, where $v(s) = (\bar{u}(s), x^{*\prime}(s), y^{*\prime}(s))$, $r = max\{p, q+1\}$ and $\phi_i = 0$ if $i > p$.

When $p > q$, we could still simplify the recursions for $t = 1, \ldots, p-q$ using the above relations, but we have not done so for simplicity, since the gain would be marginal. However, a notable improvement in the algorithm can be obtained in the case of a seasonal moving average process of the form $z(t) = \theta(B)\Theta(B^s)a(t)$, where B is the backshift operator, $\theta(B)$ is a polynomial in B of degree q, $\Theta(B^s)$ is a polynomial in B^s of degree Q and s is the length of the seasonal cycle, such that $q < (1/2)s$. For this model, Ansley's transformation (A.5) is the identity transformation and $F^{q+1} = 0$. Hence, with the above notation, we have $L = K$ and the elements l_{it} of L are related to the elements $K_i(t)$ of vectors $K(t)$ by $l_{it} = K_{i-t}(t)$, $max\{i-q, 1\} \leq t < i = 2, \ldots, N$. This implies, by Theorem 4.1 of Ansley (1979), $K_i(t) = 0$ for the two sets

$$\begin{aligned} i &= hs + l \qquad h = 0,1, \ldots;\ l = q+1, \ldots, s-1, \\ t &= Hs + k - i \quad H = 0,1, \ldots;\ k = 1, \ldots, s-q. \end{aligned}$$

APPENDIX B: PROOFS OF RESULTS

Proof of Theorem 1: The likelihood functions verify $L(z) = L(z_m \mid z_o)L(z_o)$, where the vertical bar denotes conditional distribution. Then,

$$z'\Omega_z^{-1}z = (z_m - E(z_m \mid z_o))'\Omega_{z_m|z_o}^{-1}(z_m - E(z_m \mid z_o)) + z_o'\Omega_{z_o}^{-1}z_o, \qquad (B.1)$$

where Ω_{z_o} and $\Omega_{z_m|z_o}$ are the covariance matrices, divided by σ^2, of $L(z_o)$ and $L(z_m \mid z_o)$, respectively. Given that $\bar{z} = X\omega + z$, replacing in (B.1) z with $\bar{z} - X\omega$, it is obtained that

$$(\bar{z} - X\omega)'\Omega_z^{-1}(\bar{z} - X\omega) = (\bar{z}_m - E(z_m \mid z_o) - \omega)'\Omega_{z_m|z_o}^{-1}(\bar{z}_m - E(z_m \mid z_o) - \omega) + z_o'\Omega_{z_o}^{-1}z_o. \tag{B.2}$$

The maximum likelihood estimator $\hat{\omega}$ of ω on the left hand side of $L(z) = L(z_m \mid z_o)L(z_o)$ must be equal to the one on the right hand side. Clearly, the right hand side of (B.2) is minimized for $\hat{\omega} = \bar{z}_m - E(z_m \mid z_o)$. To minimize the left hand side, consider the regression model $\bar{z} = X\omega + z$. Then, $\hat{\omega}$ is as asserted, $\hat{\omega} - \omega = z_m - E(z_m \mid z_o)$ and $Var(z_m \mid z_o) = Mse(\hat{\omega}) = \sigma^2(X'\Omega_z^{-1}X)^{-1}$. ∎

Proof of Theorem 2: Define $y = z_{II} - Bz_{Io}$ and $y_m = z_{IIm} - B_m z_{Io}$. Then, y, y_o and y_m are distributed as $N(Cz_{Im}, \sigma^2\Omega_v)$, $N(C_o z_{Im}, \sigma^2\Omega_{v_o})$ and $N(C_m z_{Im}, \sigma^2\Omega_{v_m})$, respectively, where $Var(v_m) = \sigma^2\Omega_{v_m}$, because z_{Im} is considered fixed in the definition of the likelihood (2.10). Therefore, the likelihood functions verify $L(y) = L(y_m \mid y_o)L(y_o)$ and the rest of the proof is similar to that of Theorem 1. ∎

Proof of Corollary 1: Put

$$\bar{z}_{II} - Az_I - X_{II}\hat{\omega}_{II} = [-A, I_{N-d}]\left[\begin{pmatrix} z_I \\ \bar{z}_{II} \end{pmatrix} - \begin{pmatrix} 0 \\ X_{II} \end{pmatrix}\hat{\omega}_{II}\right]$$

and consider that $\Omega_v = \Xi\Omega_u\Xi'$, where $\Xi = J_2^{-1}$, $J_{II} = [J_1, J_2]$ and $A = -\Xi J_1$. Then, we can write

$$(\bar{z}_{II} - Az_I - X_{II}\hat{\omega}_{II})'\Omega_v^{-1}(\bar{z}_{II} - Az_I - X_{II}\hat{\omega}_{II}) =$$

$$\left[\begin{pmatrix} z_I \\ \bar{z}_{II} \end{pmatrix} - \begin{pmatrix} 0 \\ X_{II} \end{pmatrix}\hat{\omega}_{II}\right]' [-A, I_{N-d}]' J_2'\Omega_u^{-1}J_2 [-A, I_{N-d}] \left[\begin{pmatrix} z_I \\ \bar{z}_{II} \end{pmatrix} - \begin{pmatrix} 0 \\ X_{II} \end{pmatrix}\hat{\omega}_{II}\right]$$

Finally, $|\Omega_v| = |\Omega_u|$ because Ξ has unit determinant. ∎

Proof of Theorem 3: Maximizing (2.10) with respect to z_{Im}, we obtain the maximum likelihood estimator $\hat{z}_{Im} = (C_o'\Omega_{v_o}^{-1}C_o)^{-1}C_o'\Omega_{v_o}^{-1}y_o$, whereas replacing in (2.10) z_{Im} with $\bar{z}_{Im} - \omega_I$ and maximizing with respect to ω_I yields the maximum likelihood estimator $\hat{\omega}_I = -(C_o'\Omega_{v_o}^{-1}C_o)^{-1}C_o'\Omega_{v_o}^{-1}(y_o - C_o\bar{z}_{Im}) = \bar{z}_{Im} - \hat{z}_{Im}$. Given that (2.10) and (3.2) coincide, maximizing also (3.2) with respect to ω_I yields the same estimator $\hat{\omega}_I$, and the estimator $\hat{\omega}_{II}$ of Corollary 1 becomes

$$\hat{\omega}_{II} = (X_{II}^{*\prime}\Omega_u^{-1}X_{II}^*)^{-1}X_{II}^{*\prime}\Omega_u^{-1}J_{II}\begin{bmatrix} \bar{z}_I - X_I\hat{\omega}_I \\ \bar{z}_{II} \end{bmatrix}.$$

The estimator $\hat{\omega}_{II}$ in Theorem 2 and Corollary 1 was obtained independently of the value of z_{Im} or, equivalently, ω_I, which was considered fixed. This means that $\hat{\omega}_{II}$ minimizes the sum of squares $(u^* - X^*_{II}\omega_{II})'\Omega_u^{-1}(u^* - X^*_{II}\omega_{II})$, where u^* and X^*_{II} are those of Corollary 1, with respect to ω_{II} for any fixed value of z_{Im} or, equivalently, ω_I. Therefore, minimizing the sum of squares

$$(u^* - X^*_{II}\omega_{II})'\Omega_u^{-1}(u^* - X^*_{II}\omega_{II}) =$$

$$(\bar{z} - X\omega)'\,[-A, I_{N-d}]'\,J_2'\Omega_u^{-1}J_2\,[-A, I_{N-d}]\,(\bar{z} - X\omega) \qquad (B.3)$$

in two steps, first with respect to ω_{II}, considering ω_I fixed, and then with respect to ω_I, is equivalent to minimizing it in one step with respect to both ω_I and ω_{II}, or $\omega = [\omega_I', \omega_{II}']'$. Finally, it is easy to verify that the estimator $\hat{\omega}$ that minimizes (B.3) is the GLS estimator of the model $\bar{u} = X^*\omega + u$, where $u = J_{II}z$ and $Var(u) = \sigma^2\Omega_u$. ∎

Proof of Corollary 2: As stated in the text, for the stationary case, the proof is trivial. When the series is nonstationary, by Theorems 2 and 3 we can write

$$\begin{array}{lll} \text{(a)} & \hat{z}_{Im} = & \bar{z}_{Im} - \hat{\omega}_I \\ \text{(b)} & \hat{z}_{IIm} = & B_m z_{I0} + C_m \hat{z}_{Im} + E[z_{IIm} - A_m z_I \mid z_{II0} - B_m z_{I0} - C_m \hat{z}_{Im}] \\ & = & \hat{z}_{IIm} - \hat{\omega}_{II}, \end{array}$$

where $(\hat{\omega}_I', \hat{\omega}_{II}')'$ is the GLS estimator of $(\omega_I', \omega_{II}')'$ in the model

$$J_{II}\begin{bmatrix} \bar{z}_I \\ \bar{z}_{II} \end{bmatrix} = J_{II}\begin{bmatrix} X_I & 0 \\ 0 & X_{II} \end{bmatrix}\begin{bmatrix} \omega_I \\ \omega_{II} \end{bmatrix} + J_{II}\begin{bmatrix} z_I \\ z_{II} \end{bmatrix},$$

or, $\bar{u} = X^*\omega + u$. In the skipping approach, $\hat{z}_{Im}$ is estimated by GLS in the model $y_0 = C_0\, z_{Im} + \nu_0$ (see Appendix A.1). By Theorems 2 and 3, $\hat{z}_{Im}$ in this model and (a) above coincide. Replacing z_{Im} by $\hat{z}_{Im}$, and running the KF with initial conditions $x(d+1 \mid d) = B_*\, z_{I0} + C_*\, \hat{z}_{Im}$, $\Sigma(d+1 \mid d) = \Xi_*\, \hat{\Sigma}(d+1 \mid d)\, \Xi_*'$, yields (b) above. ∎

Proof of Lemma 3: Suppose model (A.4) with $Var(\nu^*) = \sigma^2\Omega$. Let $\Omega = L\,L'$ be the Cholesky decomposition of Ω, with L lower triangular, and write

$$X = \begin{bmatrix} X_{II} & 0 \\ 0 & X_I \end{bmatrix}; \qquad X^* = J_{II}\,X.$$

The QR algorithm applied to $L^{-1}[X^*, Y^*]$ yields an orthogonal matrix Q, such that

$$Q'\,L^{-1}[X^*, Y^*] = \begin{bmatrix} R \\ 0 \end{bmatrix},$$

with R upper triangular. Partitioning R conforming to $[\omega'_{II},\omega'_{I}] = \omega$, and denoting by R_{II} the upper triangular submatrix of R that corresponds to ω_{II} in the partition, then it is straightforward to verify that minimizing the likelihood of Theorem 3 is the same as minimizing the nonlinear function

$$S = (\mid L \mid\mid R_{II} \mid)^{\frac{1}{M-k}}\, \bar{u}\, Q'_2\, Q_2\, \bar{u} (\mid L \mid\mid R_{II} \mid)^{\frac{1}{M-k}},$$

where $Q = [Q'_1, Q'_2]'$ and the partition is made conforming to R and 0 in $[R', 0']'$. The determinantal correction is $\mid R_{II} \mid^{\frac{1}{M-k}}$, where k is the dimension of the vector z_{Im} of missing values in z_I and z_I and z_{Im} were defined in Section 2.2. Since, under the assumptions of the lemma, $\mid R_{II} \mid$ remains constant as $M \rightarrow \infty$, $\mid R_{II} \mid^{\frac{1}{M-k}} \rightarrow 1$. ∎

References

1. Anderson, B.D.A. and Moore, J.B. (1979), *Optimal Filtering*, Prentice-Hall.
2. Ansley, C.F. (1979), "An algorithm for the exact likelihood of a mixed autoregressive-moving average process", *Biometrika*, **66**, 59-65.
3. Bell, W. (1984), "Signal Extraction for Nonstationary Series", *The Annals of Statistics*, **12**, 646-664.
4. Brockwell, P., and Davis, R. (1987), *Time Series: Theory and Methods*, Springer-Verlag, Berlin.
5. Box, G.E.P., and Jenkins, G.M. (1976), *Time Series Analysis, Forecasing and Control*, Holden-Day, San Francisco.
6. Box, G.E.P., and Tiao, G.C. (1975), "Intervention Analysis with Applications to Economic and Environmental Problems", *Jounal of the American Statistical Association*, **70**, 70-79.
7. Brubacher, S.R., and Wilson, G.T. (1976), "Interpolating Time Series with Application to the Estimation of Holiday Effects on Electricity Demand", *Applied Statistics*, **25**, 2, 107-116.
8. De Jong, Piet (1991), "The Diffuse Kalman Filter", *Annals of Statistics*, **19**, 1073-1083.
9. Findley, D.F., Monsell, B.C., Bell, W.R., Otto, M.C., and Chen, B.C. (1996), "New Capabilities and Methods of the X-12-ARIMA Seasonal Adjustment Program", forthcoming in the *Journal of Business and Economic Statistics.*
10. Gómez, V., and Maravall, A., (1994), "Estimation, Prediction and Interpolation for Nonstationary Series With the Kalman Filter", *Journal of the American Statistical Association*, **89**, 611-624.
11. Gómez, V., and Maravall, A., (1996), "Programs TRAMO and SEATS", Instructions for the User (Beta Version: September 1996), Working Paper No. 9628, Bank of Spain.
12. Harvey, A.C. (1989), *Forecasting, structural time series models and the Kalman filter*, Cambridge University Press, Cambridge.
13. Jones, R. (1980), "Maximum Likelihood Fitting of ARMA Models to Time Series With Missing Observations", *Technometrics*, **22**, 389-395.
14. Kohn, R., and Ansley, C.F. (1985), "Efficient Estimation and Prediction in Time Series Regression Models", *Biometrika*, **72**, 694-697.

15. 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.
16. Ljung, G.M. (1989), "A note on the estimation of missing values in time series," *Communications in Statistics, Simulation and Computation*, **18(2)**, 459–465.
17. Maravall, A., and Peña, D. (1992), "Missing Observations and Additive Outliers in Time Series Models", EUI Working Paper ECO No. 92/97, Department of Economics, European University Institute. Forthcoming in Mariano, R. (ed.), *Advances in Statistical Analysis and Statistical Computing* (vol. 3), JAI Press.
18. Mélard, G. (1984), "A Fast Algorithm for the Exact Likelihood of Autoregressive–Moving Average Models", *Applied Statistics*, **35**, 104–114.
19. 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.
20. Peña, D. (1987), "Measuring the importance of outliers in ARIMA models", in *New Perspectives in Theoretical and Applied Statistics*, eds. M.L. Puri et al., Wiley, 109–118.
21. Pearlman, J.G. (1980), "An Algorithm for the Exact Likelihood of a High–Order–Autoregressive–Moving Average Process", *Biometrika*, **67**, 232–233.
22. Sargan, J.D., and Drettakis, E.G. (1974), "Missing data in an autoregressive model", *International Economic Review*, **15**, 39–58.

WORKING PAPERS (1)

9406 **Juan José Dolado, José Manuel González-Páramo y José M.ª Roldán:** Convergencia económica entre las provincias españolas: evidencia empírica (1955-1989).

9407 **Ángel Estrada e Ignacio Hernando:** La inversión en España: un análisis desde el lado de la oferta.

9408 **Ángel Estrada García, M.ª Teresa Sastre de Miguel y Juan Luis Vega Croissier:** El mecanismo de transmisión de los tipos de interés: el caso español.

9409 **Pilar García Perea y Ramón Gómez:** Elaboración de series históricas de empleo a partir de la Encuesta de Población Activa (1964-1992).

9410 **F. J. Sáez Pérez de la Torre, J. M.ª Sánchez Sáez y M.ª T. Sastre de Miguel:** Los mercados de operaciones bancarias en España: especialización productiva y competencia.

9411 **Olympia Bover and Ángel Estrada:** Durable consumption and house purchases: Evidence from Spanish panel data.

9412 **José Viñals:** Building a Monetary Union in Europe: Is it worthwhile, where do we stand, and where are we going? (The Spanish original of this publication has the same number.)

9413 **Carlos Chuliá:** Los sistemas financieros nacionales y el espacio financiero europeo.

9414 **José Luis Escrivá and Andrew G. Haldane:** The interest rate transmission mechanism: Sectoral estimates for Spain. (The Spanish original of this publication has the same number.)

9415 **M.ª de los Llanos Matea y Ana Valentina Regil:** Métodos para la extracción de señales y para la trimestralización. Una aplicación: Trimestralización del deflactor del consumo privado nacional.

9416 **José Antonio Cuenca:** Variables para el estudio del sector monetario. Agregados monetarios y crediticios, y tipos de interés sintéticos.

9417 **Ángel Estrada y David López-Salido:** La relación entre el consumo y la renta en España: un modelo empírico con datos agregados.

9418 **José M. González Mínguez:** Una aplicación de los indicadores de discrecionalidad de la política fiscal a los países de la UE.

9419 **Juan Ayuso, María Pérez Jurado and Fernando Restoy:** Is exchange rate risk higher in the E.R.M. after the widening of fluctuation bands? (The Spanish original of this publication has the same number.)

9420 **Simon Milner and David Metcalf:** Spanish pay setting institutions and performance outcomes.

9421 **Javier Santillán:** El SME, los mercados de divisas y la transición hacia la Unión Monetaria.

9422 **Juan Luis Vega:** Is the ALP long-run demand function stable? (The Spanish original of this publication has the same number.)

9423 **Gabriel Quirós:** El mercado italiano de deuda pública.

9424 **Isabel Argimón, José Manuel González-Páramo y José María Roldán:** Inversión privada, gasto público y efecto expulsión: evidencia para el caso español.

9425 **Charles Goodhart and José Viñals:** Strategy and tactics of monetary policy: Examples from Europe and the Antipodes.

9426 **Carmen Melcón:** Estrategias de política monetaria basadas en el seguimiento directo de objetivos de inflación. Las experiencias de Nueva Zelanda, Canadá, Reino Unido y Suecia.

9427 **Olympia Bover and Manuel Arellano:** Female labour force participation in the 1980s: the case of Spain.

9428 **Juan María Peñalosa:** The Spanish catching-up process: General determinants and contribution of the manufacturing industry.

9429 **Susana Núñez:** Perspectivas de los sistemas de pagos: una reflexión crítica.

9430 **José Viñals:** ¿Es posible la convergencia en España?: En busca del tiempo perdido.

9501 **Jorge Blázquez y Miguel Sebastián:** Capital público y restricción presupuestaria gubernamental.

9502 **Ana Buisán:** Principales determinantes de los ingresos por turismo.

9503 **Ana Buisán y Esther Gordo:** La protección nominal como factor determinante de las importaciones de bienes.

9504 **Ricardo Mestre:** A macroeconomic evaluation of the Spanish monetary policy transmission mechanism.

9505 **Fernando Restoy and Ana Revenga:** Optimal exchange rate flexibility in an economy with intersectoral rigidities and nontraded goods.

9506 **Ángel Estrada and Javier Vallés:** Investment and financial costs: Spanish evidence with panel data. (The Spanish original of this publication has the same number.)

9507 **Francisco Alonso:** La modelización de la volatilidad del mercado bursátil español.

9508 **Francisco Alonso y Fernando Restoy:** La remuneración de la volatilidad en el mercado español de renta variable.

9509 **Fernando C. Ballabriga, Miguel Sebastián y Javier Vallés:** España en Europa: asimetrías reales y nominales.

9510 **Juan Carlos Casado, Juan Alberto Campoy y Carlos Chuliá:** La regulación financiera española desde la adhesión a la Unión Europea.

9511 **Juan Luis Díaz del Hoyo y A. Javier Prado Domínguez:** Los FRAs como guías de las expectativas del mercado sobre tipos de interés.

9512 **José M.ª Sánchez Sáez y Teresa Sastre de Miguel:** ¿Es el tamaño un factor explicativo de las diferencias entre entidades bancarias?

9513 **Juan Ayuso y Soledad Núñez:** ¿Desestabilizan los activos derivados el mercado al contado?: La experiencia española en el mercado de deuda pública.

9514 **M.ª Cruz Manzano Frías y M.ª Teresa Sastre de Miguel:** Factores relevantes en la determinación del margen de explotación de bancos y cajas de ahorros.

9515 **Fernando Restoy and Philippe Weil:** Approximate equilibrium asset prices.

9516 **Gabriel Quirós:** El mercado francés de deuda pública.

9517 **Ana L. Revenga and Samuel Bentolila:** What affects the employment rate intensity of growth?

9518 **Ignacio Iglesias Araúzo y Jaime Esteban Velasco:** Repos y operaciones simultáneas: estudio de la normativa.

9519 **Ignacio Fuentes:** Las instituciones bancarias españolas y el Mercado Único.

9520 **Ignacio Hernando:** Política monetaria y estructura financiera de las empresas.

9521 **Luis Julián Álvarez y Miguel Sebastián:** La inflación latente en España: una perspectiva macroeconómica.

9522 **Soledad Núñez Ramos:** Estimación de la estructura temporal de los tipos de interés en España: elección entre métodos alternativos.

9523 **Isabel Argimón, José M. González-Páramo y José M.ª Roldán Alegre:** Does public spending crowd out private investment? Evidence from a panel of 14 OECD countries.

9524 **Luis Julián Álvarez, Fernando C. Ballabriga y Javier Jareño:** Un modelo macroeconométrico trimestral para la economía española.

9525 **Aurora Alejano y Juan M.ª Peñalosa:** La integración financiera de la economía española: efectos sobre los mercados financieros y la política monetaria.

9526 **Ramón Gómez Salvador y Juan J. Dolado:** Creación y destrucción de empleo en España: un análisis descriptivo con datos de la CBBE.

9527 **Santiago Fernández de Lis y Javier Santillán:** Regímenes cambiarios e integración monetaria en Europa.

9528 **Gabriel Quirós:** Mercados financieros alemanes.

9529 **Juan Ayuso Huertas:** Is there a trade-off between exchange rate risk and interest rate risk? (The Spanish original of this publication has the same number.)

9530 **Fernando Restoy:** Determinantes de la curva de rendimientos: hipótesis expectacional y primas de riesgo.

9531 **Juan Ayuso and María Pérez Jurado:** Devaluations and depreciation expectations in the EMS.

9532 **Paul Schulstad and Ángel Serrat:** An Empirical Examination of a Multilateral Target Zone Model.

9601 **Juan Ayuso, Soledad Núñez and María Pérez-Jurado:** Volatility in Spanish financial markets: The recent experience.

9602 **Javier Andrés e Ignacio Hernando:** ¿Cómo afecta la inflación al crecimiento económico? Evidencia para los países de la OCDE.

9603 **Barbara Dluhosch:** On the fate of newcomers in the European Union: Lessons from the Spanish experience.

9604 **Santiago Fernández de Lis:** Classifications of Central Banks by Autonomy: A comparative analysis.

9605 **M.ª Cruz Manzano Frías y Sofía Galmés Belmonte:** Credit Institutions' Price Policies and Type of Customer: Impact on the Monetary Transmission Mechanism. (The Spanish original of this publication has the same number.)

9606 **Malte Krüger:** Speculation, Hedging and Intermediation in the Foreign Exchange Market.

9607 **Agustín Maravall:** Short-Term Analysis of Macroeconomic Time Series.

9608 **Agustín Maravall and Christophe Planas:** Estimation Error and the Specification of Unobserved Component Models.

9609 **Agustín Maravall:** Unobserved Components in Economic Time Series.

9610 **Matthew B. Canzoneri, Behzad Diba and Gwen Eudey:** Trends in European Productivity and Real Exchange Rates.

9611 **Francisco Alonso, Jorge Martínez Pagés y María Pérez Jurado:** Weighted Monetary Aggregates: an Empirical Approach. (The Spanish original of this publication has the same number.)

9612 **Agustín Maravall and Daniel Peña:** Missing Observations and Additive Outliers in Time Series Models.

9613 **Juan Ayuso and Juan L. Vega:** An empirical analysis of the peseta's exchange rate dynamics.

9614 **Juan Ayuso Huertas:** Un análisis empírico de los tipos de interés reales *ex-ante* en España.

9615 **Enrique Alberola Ila:** Optimal exchange rate targets and macroeconomic stabilization.

9616 **A. Jorge Padilla, Samuel Bentolila and Juan J. Dolado:** Wage bargaining in industries with market power.

9617 **Juan J. Dolado and Francesc Marmol:** Efficient estimation of cointegrating relationships among higher order and fractionally integrated processes.

9618 **Juan J. Dolado y Ramón Gómez:** La relación entre vacantes y desempleo en España: perturbaciones agregadas y de reasignación.

9619 **Alberto Cabrero and Juan Carlos Delrieu:** Construction of a composite indicator for predicting inflation in Spain. (The Spanish original of this publication has the same number.)

9620 **Una-Louise Bell:** Adjustment costs, uncertainty and employment inertia.

9621 **M.ª de los Llanos Matea y Ana Valentina Regil:** Indicadores de inflación a corto plazo.

9622 **James Conklin:** Computing value correspondences for repeated games with state variables.

9623 **James Conklin:** The theory of sovereign debt and Spain under Philip II.

9624 **José Viñals and Juan F. Jimeno:** Monetary Union and European unemployment.

9625 **María Jesús Nieto Carol:** Central and Eastern European Financial Systems: Towards integration in the European Union.

9626 **Matthew B. Canzoneri, Javier Vallés and José Viñals:** Do exchange rates move to address international macroeconomic imbalances?

9627 **Enrique Alberola Ila:** Integración económica y unión monetaria: el contraste entre Norteamérica y Europa.

9628 **Víctor Gómez and Agustín Maravall:** Programs TRAMO and SEATS.

9629 **Javier Andrés, Ricardo Mestre y Javier Vallés:** Un modelo estructural para el análisis del mecanismo de transmisión monetaria: el caso español.

9630 **Francisco Alonso y Juan Ayuso:** Una estimación de las primas de riesgo por inflación en el caso español.

9631 **Javier Santillán:** Política cambiaria y autonomía del Banco Central.

9632 **Marcial Suárez:** Vocábula (Notas sobre usos lingüísticos).

9633 **Juan Ayuso and J. David López-Salido:** What does consumption tell us about inflation expectations and real interest rates?

9701 **Víctor Gómez, Agustín Maravall and Daniel Peña:** Missing observations in ARIMA models: Skipping strategy versus outlier approach.

(1) Previously published Working Papers are listed in the Banco de España publications catalogue.

Queries should be addressed to: Banco de España
Sección de Publicaciones. Negociado de Distribución y Gestión
Telephone: 338 51 80
Alcalá, 50. 28014 Madrid