ELSEVIER

Journal of Econometrics 88 (1999) 341–363

JOURNAL OF Econometrics

Missing observations in ARIMA models: Skipping approach versus additive outlier approach

Víctor Gómez[b], Agustín Maravall[a,*], Daniel Peña[c]

[a] *Research Department, Bank of Spain, Alcala 50, 28014 Madrid, Spain*
[b] *Ministerio de Economía y Hacienda, P. de la Castellana 162, E-28046 Madrid, Spain*
[c] *Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, 28903 Getafe, Spain*

Received 1 January 1998; received in revised form 1 May 1998

Abstract

Optimal estimation of missing values in ARMA models is typically performed by using the Kalman filter for likelihood evaluation, 'skipping' in the computations the missing observations, obtaining the maximum likelihood (ML) estimators of the model parameters, and using some smoothing algorithm. The same type of procedure has been extended to nonstationary ARIMA models in Gómez and Maravall (1994). An alternative procedure suggests filling in the holes in the series with arbitrary values and then performing ML estimation of the ARIMA model with additive outliers (AO). When the model parameters are not known the two methods differ, since the AO likelihood is affected by the arbitrary values. We develop the proper likelihood for the AO approach in the general non-stationary case and show the equivalence of this and the skipping method. Finally, the two methods are compared through simulation, and their relative advantages assessed; the comparison also includes the AO method with the uncorrected likelihood.

JEL classification: C22

Keywords: Time series; ARIMA models; Missing observations; Outliers; Nonstationarity; Likelihood; Kalman filter

1. Introduction and summary

The problem of optimal estimation of missing observations in stationary autoregressive moving average (ARMA) models was solved in Jones (1980).

* Corresponding author. E-mail: maravall@bde.es.

0304-4076/99/$ – see front matter
PII: S0304-4076(98)00036-0

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' (a referee suggested the more informative term 'marginalizing the likelihood function approach'; for the sake of brevity, we shall use the term 'skipping'). Since the Kohn–Ansley et al., 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 to use dummy variables; see also Harvey ((1989), pp. 145–146). Subject to a qualification to be made below, one could 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. The difference between the arbitrary value set by the user and its corresponding estimated parameter, when the model parameters are known, 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.

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 this effect. 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, although no attempt was made 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. It is 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.

A word of caution, however, may be appropriate. Except perhaps for a few near-trivial cases, one may be tempted to conclude that the AO outlier approach might as well be dropped from consideration. Yet, it displays some other advantages, and an example of applied interest is the following. In the AO approach, the same algorithms for automatic model identification and automatic outlier detection can be used than in the case of no missing values. This fact simplifies enormously programming, because the skipping approach would require additional specialized (and difficult to program) routines.

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 analyse 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. Proofs of the results are presented in an appendix and the detailed computational algorithms in the working paper Gómez et al. (1997), available from the authors upon request.

2. Skipping approach

2.1. Stationary series, ARMA model

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

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

where $\phi(B) = 1 + \phi_1 B + \cdots + \phi_p B^p$ and $\theta(B) = 1 + \theta_1 B + \cdots + \theta_q B^q$ 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 (1996). 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 $t = 1, \ldots, N$, $x(t) = (z(t), z(t+1|t), \ldots, z(t+r-1|t))'$, $G = (1, \psi_1, \ldots, \psi_{r-1})'$, $H = (1, 0, \ldots, 0)'$,

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

and the ψ_i-weights are obtained from $\psi(B) = \theta(B)/\phi(B) = \sum_{i=0}^{\infty} \psi_i B^i$. The expression $z(t+i|t)$ is the orthogonal projection of $z(t+i)$ on the subspace generated by $\{z(s): s \leqslant t\}$, and coincides with the conditional expectation $E(z(t+i)|z(s): s \leqslant t)$, $i = 1, \ldots, r-1$. 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 $\{z(s): s \leqslant t\}$. 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 Eq. (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 iidN(0,1) variable, independent of $\{z(t_1), \ldots, z(t_M)\}$. Thus, when $z(t)$ is missing, in the Kalman filter equations, $x(t|t) = x(t|t-1)$, $\Sigma(t|t) = \Sigma(t|t-1)$, where $x(t|t+i) = E(x(t)|z(1), \ldots, z(t+i))$, $\Sigma(t|t+i) = \mathrm{Var}(x(t)|z(1), \ldots, z(t+i))$, $1 \leqslant$

$t \leqslant N$, $i = -1, 0$, 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 $\delta(B) = 1 + \delta_1 B + \cdots + \delta_d B^d$ 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 (1976), Chapter 7. Letting $z_{\mathrm{I}} = (z(1), \ldots, z(d))'$ and $z_{\mathrm{II}} = (z(d+1), \ldots, z(N))'$, it is easily seen that differencing the data implies the transformation $[z'_{\mathrm{I}}, u']' = J[z'_{\mathrm{I}}, z'_{\mathrm{II}}]'$, where $J = [J'_{\mathrm{I}}, J'_{\mathrm{II}}]'$, $J_{\mathrm{I}} = [I_d, 0]$, I_d is the identity matrix of rank d and

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

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

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

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

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

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

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

and the relation

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

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 $v = z_{\mathrm{II}} - A z_{\mathrm{I}}$, the log-likelihood based on u is (throughout the paper all log-likelihoods will be defined up to an additive constant)

$$l(u) = -\tfrac{1}{2}\{(N-d)\ln(\sigma^2) + \ln|\Omega_v| + (z_{\mathrm{II}} - A z_{\mathrm{I}})'\Omega_v^{-1}(z_{\mathrm{II}} - A z_{\mathrm{I}})/\sigma^2\}, \tag{2.7}$$

where $\mathrm{Var}(v) = \sigma^2\Omega_v$, $\Omega_v = \Xi\Omega_u\Xi'$, and $\mathrm{Var}(u) = \sigma^2\Omega_u$. Eq. (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 $J = [J_{\mathrm{I}}', J_{\mathrm{II}}']'$ has unit determinant, the log-likelihood $l(z)$ of the observed series $z = [z_{\mathrm{I}}', z_{\mathrm{II}}']'$ verifies $l(z) = l(z_{\mathrm{I}}, u) = l(z_{\mathrm{I}}) + l(u)$. Therefore, under assumptions A and B, we have the result

Lemma 1. $l(u) = l(z_{\mathrm{II}}|z_{\mathrm{I}})$.

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 Eqs. (2.2a) and (2.2b), 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 $z(t + i|t) = z(t + i) - \psi_0^* a(t + i) - \cdots - \psi_{i-1}^* a(t + 1)$, $i = 1, \ldots, r - 1$. The following lemma, whose proof is omitted, ensures that this state space representation is correct.

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

The Kalman filter can then be applied to compute the conditional log-likelihood $l(z_{\mathrm{II}}|z_{\mathrm{I}})$ through the prediction error decomposition. The starting conditions can be obtained from Eq. (2.6) as follows. If we consider the definition of the elements of the state vector $x(t)$, it can be seen that $x(d + 1) = A_* z_{\mathrm{I}} + \Xi_* U_*$, 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, $U_* = [u(d + 1), u(d + 2|d + 1), \ldots, u(d + r|d + 1)]'$, and $u(d + i|d + 1) = E(u(d + i)|u(t)\colon t \leqslant d + 1)$, $i = 2, \ldots, r$. Therefore, we can take as starting conditions

$$x(d + 1|d) = \mathrm{E}(x(d + 1)|z(s)\colon 1 \leqslant s \leqslant d) = A_* z_{\mathrm{I}},$$

$$\Sigma(d + 1|d) = \mathrm{Var}(x(d + 1)|z(s)\colon 1 \leqslant s \leqslant d) = \Xi_* \tilde{\Sigma}(d + 1|d)\Xi_*',$$

where $\tilde{\Sigma}(d + 1|d) = \mathrm{E}(U_* U_*')$ 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 $z_{\mathrm{o}} = (z(t_1), z(t_2), \ldots, z(t_M))'$, $1 \leqslant t_1 < t_2 < \cdots < t_M \leqslant N$, is a subvector of the complete series $z = [z_{\mathrm{I}}', z_{\mathrm{II}}']'$, we can proceed as follows. Let $z_{\mathrm{Io}} = (z(t_1), \ldots, z(t_k))'$, $k \leqslant d$, 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

$$z_{\mathrm{IIo}} = A_{\mathrm{o}} z_{\mathrm{I}} + v_{\mathrm{o}} = B_{\mathrm{o}} z_{\mathrm{Io}} + C_{\mathrm{o}} z_{\mathrm{Im}} + v_{\mathrm{o}}, \tag{2.8}$$

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 Eq. (2.8) suggest a natural way to extend the log-likelihood (2.7) to the case of missing observations. We can consider in Eq. (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

$$y_{\mathrm{o}} = C_{\mathrm{o}} z_{\mathrm{Im}} + v_{\mathrm{o}}, \tag{2.9}$$

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

$$l(y_{\mathrm{o}}) = -\tfrac{1}{2}\{(M-k)\ln(\sigma^2) + \ln|\Omega_{v_{\mathrm{o}}}| + (y_{\mathrm{o}} - C_{\mathrm{o}} z_{\mathrm{Im}})' \Omega_{v_{\mathrm{o}}}^{-1} (y_{\mathrm{o}} - C_{\mathrm{o}} z_{\mathrm{Im}})/\sigma^2\}, \tag{2.10}$$

where $\mathrm{Var}(v_{\mathrm{o}}) = \sigma^2 \Omega_{v_{\mathrm{o}}}$. 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 Eq. (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 Gómez et al. (1997).

2.3. Regression model with ARIMA errors

Consider the regression model

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

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 $\{v(t)\}$ is assumed to follow the

ARIMA model given by Eq. (2.3). If, as in the previous section, z_o denotes the observed series, defining the vector $v_o = (v(t_1), \ldots, v(t_M))'$ and the $M \times h$ matrix Y_o with the vectors $y'(t), t = t_1, \ldots, t_M$, as rows, we can write $z_o = Y_o\beta + v_o$, where the matrix Y_o is assumed of rank h. Since $\{v(t)\}$ follows the ARIMA model (2.3), similarly to Eq. (2.8), we can write $v_{IIo} = B_o v_{Io} + C_o v_{Im} + v_o$, where v_{IIo}, v_{Io} and v_{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 v_{Io}, v_{IIo} and v_{Im}, respectively. Replacing v_{IIo} with $z_{IIo} - Y_{IIo}\beta$, v_{Io} with $z_{Io} - Y_{Io}\beta$ and v_{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 Eq. (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 $x(t) = (v(t), v(t+1|t), \ldots, v(t+r-1|t))$, then the state space representation is given by Eq. (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 v.

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*

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

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

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 $\ln|X'\Omega_z^{-1}X|$. 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 Gómez et al. (1997) we describe in detail the algorithms, 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 $z_o = (z(t_1), z(t_2), \ldots, z(t_M))'$, $1 \leqslant t_1 < t_2 < \cdots < t_M \leqslant N$, 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*

$$l(y_o) = -\tfrac{1}{2}\{(M - k)\ln(\sigma^2) + \ln|\Omega_v| + \ln|X_{II}'\Omega_v^{-1}X_{II}| + (\bar{z}_{II} - Az_I - X_{II}\hat{\omega}_{II})'\Omega_v^{-1}(\bar{z}_{II} - Az_I - X_{II}\hat{\omega}_{II})/\sigma^2\}, \quad (3.1)$$

where X_{II} *is the* $(N-d)\times(N-M-d+k)$ *matrix whose columns are unit vectors, such that the ith column has a one in the position corresponding to the ith 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* $\hat{\omega}_{II}=\bar{z}_{IIm}-A_m z_I-E(z_{IIm}-A_m z_I|z_{IIo}-A_o z_I)$. *Also,* $Mse(\hat{\omega}_{II})=\mathrm{Var}(z_{IIm}-A_m z_I|z_{IIo}-A_o z_I)=\sigma^2(X_{II}'\Omega_v^{-1}X_{II})^{-1}$.

Note that in Eq. (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 Eq. (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 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*

$$l(y_o)=-\tfrac{1}{2}\{(M-k)\ln(\sigma^2)+\ln|\Omega_u|+\ln|X_{II}^{*\prime}\Omega_u^{-1}X_{II}^*| + (u^*-X_{II}^*\hat{\omega}_{II})'\Omega_u^{-1}(u^*-X_{II}^*\hat{\omega}_{II})/\sigma^2\}, \tag{3.2}$$

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

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 ith column has a one in the position corresponding to the ith 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*

$$\lambda(y_o) = -\tfrac{1}{2}\{(M-k)\ln(\sigma^2) + \ln|\Omega_u| + \ln|X_{\mathrm{II}}^{*\prime}\Omega_u^{-1}X_{\mathrm{II}}^{*}| + (\bar{u} - X^*\hat{\omega})'\Omega_u^{-1}(\bar{u} - X^*\hat{\omega})/\sigma^2\},$$

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

Note that, by Theorems 2 and 3, the interpolations $\hat{z}_{\mathrm{Im}}$ of z_{Im} and $\hat{z}_{\mathrm{IIm}}$ of z_{IIm} are simply $\bar{z}_{\mathrm{Im}} - \hat{\omega}_{\mathrm{I}}$ and $\bar{z}_{\mathrm{IIm}} - \hat{\omega}_{\mathrm{II}}$, respectively, where $\hat{\omega} = [\hat{\omega}_{\mathrm{I}}', \hat{\omega}_{\mathrm{II}}']'$ is given by Theorem 3. [Another expression for $\hat{z}_{\mathrm{IIm}}$, is $\hat{z}_{\mathrm{IIm}} = B_{\mathrm{m}}z_{\mathrm{Io}} + C_{\mathrm{m}}\hat{z}_{\mathrm{Im}} + P(y_{\mathrm{o}} - C_{\mathrm{o}}\hat{z}_{\mathrm{Im}})$, where P is the matrix such that $E(z_{\mathrm{IIm}} - A_{\mathrm{m}}z_{\mathrm{I}}|z_{\mathrm{IIo}} - A_{\mathrm{o}}z_{\mathrm{I}}) = P(z_{\mathrm{IIo}} - A_{\mathrm{o}}z_{\mathrm{I}})$, $P = \mathrm{Cov}(v_{\mathrm{m}},v_{\mathrm{o}})\,\mathrm{Var}^{-1}(v_{\mathrm{o}})$.] Note also that in the z_{Im}-maximized log-likelihood $\lambda(y_{\mathrm{o}})$ of Theorem 3 the correction $\ln|X_{\mathrm{II}}^{*\prime}\Omega_u^{-1}X_{\mathrm{II}}^{*}|$ 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, $\bar{z}_{\mathrm{m}} - \hat{\omega}$, is equal to $E(z_{\mathrm{m}}|z_{\mathrm{o}})$. 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 $\lambda(y_{\mathrm{o}})$. 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 et al. (1974), as described by Pearlman (1980) and improved by Mélard (1984). We use an improved version of this last algorithm, detailed in Gómez et al. (1997).

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 $\{v(t)\}$ follow the ARIMA model (2.3) with $z(t)$ replaced with $v(t)$. With the notation of the previous section, if we define the vector $v = (v(1), \ldots, v(N))'$ and the $N \times h$ matrix Y with the vectors $y'(t)$, $t = 1, \ldots, N$, as rows, we can write $\bar{z} = [X, Y][\omega', \beta']' + v$. 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. 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 three models

MA(1)	$z(t) = (1 - 0.7B)a(t)$
ARIMA(1,1,0)	$(1 - 0.8B)(1 - B)z(t) = a(t)$
ARIMA(0,1,1)(0,1,1)	$(1 - B)(1 - B^{12})z(t) = (1 - 0.4B)(1 - 0.6B^{12})a(t)$,

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 choice of models was based on the following consideration. The third model is the well-known 'Airline Model', popularized by Box and Jenkins (1976). It is a model that fits many series; a recent study by EUROSTAT (1996) on close to 15,000 series from the 15 EU countries, plus USA and Japan, found the model appropriate for about 50% of the series. The second model is a pure AR model of the type often used by macroeconomists to model, for example, the seasonally adjusted US GNP series. The first model pretends to introduce pure MA effects into the discussion. The choice of patterns for the missing values considers, first, historical estimation of a single missing observation. The second example attempts to capture the effect of consecutive missing values. The third example attempts to capture the effect of many, randomly ocurring, missing values.

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 three models and for each possible combination of estimation procedure and pattern of missing data, we have performed 5000 simulations. All estimations have been made with the program TRAMO ('Time Series Regression with ARIMA Noise, Missing Observations, and Outliers; Gómez and Maravall (1996).)

Tables 1–3 below correspond to the three 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 TRMSEs 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)

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

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

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

Table 1
Model $z(t) = (1 - 0.7B)a(t)$

Observation number	SK		AOC		AON		
	ME	RMSE	ME	RMSE	ME	RMSE	TRMSE
1 missing observation							
$n = 50$	0.000	0.741	0.000	0.741	0.001	0.744	0.714
5 missing observations							
$n = 41$	− 0.010	1.013	− 0.009	1.012	− 0.010	1.013	1.000
42	0.027	1.202	0.027	1.202	0.027	1.202	1.221
43	0.011	1.212	0.011	1.212	0.011	1.212	1.221
44	− 0.038	1.222	− 0.038	1.222	− 0.038	1.222	1.221
45	0.023	0.995	0.023	0.994	0.023	0.996	1.000
20 missing observations							
$n = 2$	0.002	0.844	− 0.007	0.879	− 0.007	0.876	0.828
7	0.000	0.765	0.000	0.828	0.000	0.826	0.726
15	− 0.002	0.754	− 0.009	0.812	− 0.008	0.808	0.726
20	0.009	0.750	0.007	0.813	0.008	0.810	0.735
25	0.000	0.752	0.005	0.824	0.008	0.821	0.727
32	− 0.005	1.000	− 0.010	1.022	− 0.009	1.021	1.002
33	− 0.018	1.003	− 0.020	1.027	− 0.020	1.025	1.007
38	0.004	0.752	0.006	0.820	0.006	0.815	0.746
42	0.019	0.792	0.025	0.838	0.024	0.836	0.781
45	0.007	0.779	0.011	0.827	0.012	0.824	0.770
50	− 0.011	1.022	− 0.016	1.046	− 0.014	1.044	1.007
51	0.014	1.021	0.006	1.043	0.007	1.042	1.000
63	0.005	0.774	0.011	0.841	0.010	0.838	0.715
72	− 0.016	0.751	− 0.015	0.822	− 0.014	0.819	0.717
79	− 0.014	0.826	− 0.018	0.874	− 0.020	0.872	0.821
81	− 0.010	0.860	− 0.014	0.890	− 0.014	0.888	0.860
84	0.006	1.030	0.006	1.044	0.006	1.044	1.033
85	− 0.019	1.220	− 0.019	1.220	− 0.019	1.220	1.221
86	− 0.001	1.019	0.002	1.044	0.001	1.042	1.016
90	− 0.007	0.765	− 0.004	0.821	− 0.004	0.820	0.736

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

Further,

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

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

Table 2
Model $(1 - 0.8B)(1 - B)z(t) = a(t)$

Observation number	SK		AOC		AON		
	ME	RMSE	ME	RMSE	ME	RMSE	TRMSE
1 missing observation							
$n = 50$	0.004	0.455	0.004	0.455	0.004	0.455	0.453
5 missing observations							
$n = 41$	− 0.016	0.797	− 0.016	0.797	− 0.016	0.797	0.801
42	− 0.015	1.295	− 0.015	1.295	− 0.015	1.295	1.298
43	0.006	1.463	0.006	1.463	0.006	1.463	1.476
44	− 0.002	1.299	− 0.002	1.299	− 0.002	1.299	1.298
45	− 0.001	0.799	− 0.001	0.799	− 0.001	0.799	0.801
20 missing observations							
$n = 2$	0.009	0.492	0.009	0.492	0.010	0.492	0.486
7	− 0.010	0.452	− 0.010	0.452	− 0.010	0.452	0.453
15	− 0.005	0.454	− 0.005	0.454	− 0.005	0.454	0.453
20	− 0.006	0.454	− 0.006	0.454	− 0.006	0.454	0.453
25	0.004	0.447	0.004	0.447	0.004	0.447	0.453
32	0.010	0.608	0.010	0.608	0.010	0.608	0.605
33	0.009	0.611	0.009	0.611	0.009	0.611	0.605
38	− 0.006	0.460	− 0.006	0.460	− 0.005	0.460	0.453
42	− 0.008	0.454	− 0.008	0.454	− 0.008	0.454	0.453
45	0.001	0.449	0.001	0.449	0.001	0.449	0.453
50	0.004	0.610	0.004	0.610	0.004	0.610	0.605
51	0.000	0.609	0.000	0.609	0.000	0.609	0.605
63	0.010	0.448	0.010	0.448	0.010	0.448	0.453
72	− 0.004	0.452	− 0.004	0.452	− 0.004	0.452	0.453
79	− 0.015	0.460	− 0.015	0.460	− 0.015	0.460	0.459
81	0.004	0.456	0.004	0.456	0.004	0.456	0.459
84	0.004	0.695	0.004	0.695	0.004	0.694	0.697
85	− 0.005	0.918	− 0.005	0.918	− 0.005	0.918	0.919
86	0.004	0.696	0.004	0.696	0.004	0.696	0.697
90	− 0.005	0.452	− 0.005	0.452	− 0.005	0.452	0.453

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

close enough to the end of the series or to another missing value, the RMSE will, of course, be larger.

For the three models considered above, the inverse ACF show that for the pure AR model convergence of Eq. (4.1) will occur in just two periods. The MA model and, in particular, the mixed model imply slower convergences, in accordance with the convergence properties of the expressions $(1 - 0.7B)^{-1}$ and $(1 - 0.6B^{12})^{-1}$. For the three models, expression (4.3) yields

$$\text{MA(1): RMSE}[(\hat{z}(t)] = 0.714,$$

Table 3
Model $(1-B)(1-B^{12})z(t) = (1-0.4B)(1-0.6B^{12})a(t)$

Observation number	SK		AOC		AON		
	ME	RMSE	ME	RMSE	ME	RMSE	TRMSE
1 missing observation							
$n = 50$	− 0.006	0.763	− 0.006	0.763	− 0.006	0.763	0.751
5 missing observations							
$n = 41$	− 0.021	0.841	− 0.021	0.841	− 0.021	0.842	0.837
42	− 0.003	0.918	− 0.003	0.918	− 0.003	0.919	0.905
43	0.012	0.934	0.011	0.934	0.011	0.934	0.927
44	− 0.018	0.903	− 0.018	0.903	− 0.018	0.903	0.905
45	− 0.002	0.832	− 0.001	0.832	− 0.001	0.832	0.837
20 missing observations							
$n = 2$	− 0.004	0.907	− 0.004	0.907	− 0.004	0.920	0.884
7	− 0.024	0.859	− 0.023	0.860	− 0.022	0.872	0.849
15	− 0.008	0.801	− 0.008	0.800	− 0.009	0.837	0.792
20	− 0.002	0.821	− 0.002	0.821	− 0.003	0.849	0.814
25	0.000	0.788	0.000	0.788	0.000	0.826	0.772
32	0.010	0.829	0.010	0.829	0.008	0.862	0.826
33	− 0.006	0.844	− 0.006	0.844	− 0.007	0.881	0.818
38	0.002	0.798	0.001	0.798	0.001	0.833	0.788
42	0.005	0.762	0.004	0.762	0.008	0.801	0.759
45	0.001	0.775	0.001	0.775	0.001	0.808	0.780
50	− 0.010	0.827	− 0.011	0.827	− 0.012	0.865	0.815
51	− 0.015	0.815	− 0.015	0.815	− 0.016	0.851	0.810
63	0.020	0.787	0.019	0.787	0.020	0.822	0.777
72	− 0.010	0.794	− 0.009	0.793	− 0.007	0.829	0.786
79	− 0.028	0.811	− 0.028	0.811	− 0.030	0.847	0.790
81	− 0.004	0.794	− 0.003	0.794	− 0.006	0.834	0.791
84	0.007	0.894	0.008	0.894	0.012	0.927	0.865
85	− 0.004	0.892	− 0.004	0.892	− 0.003	0.932	0.874
86	0.017	0.870	0.017	0.870	0.018	0.913	0.847
90	− 0.001	0.851	− 0.001	0.851	− 0.004	0.866	0.846

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

$$\text{ARIMA(1,1,0): RMSE}[(\hat{z}(t)] = 0.453,$$

$$\text{ARIMA(0,1,1)(0,1,1): RMSE}[(\hat{z}(t)] = 0.748.$$

From Tables 1–3, it is seen that those (asymptotic) RMSE are identical to the TRMSE computed by the Kalman filter for the first two models when there is one missing observation. For the last model, the small discrepancy is caused by the fact that $(1-0.6B^{12})^{-1}$ has not fully converged in 4 years. When there are

Table 4
Elapsed time in seconds (average)

Approach	MA(1)	ARIMA(1,1,0)	ARIMA(0,1,1)(0,1,1)
1 missing observations			
SK	0.07	0.07	0.52
AOC	0.08	0.08	0.17
AON	0.09	0.08	0.19
5 missing observations			
SK	0.08	0.07	0.55
AOC	0.14	0.11	0.26
AON	0.12	0.11	0.33
20 missing observations			
SK	0.07	0.07	0.65
AOC	0.33	0.32	1.02
AON	0.33	0.33	0.95

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

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 the pure AR model, the filter converges 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 three 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 model, 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 1–3 are the Monte Carlo (MC) RMSE of the interpolators for all cases considered, computed, for each

case, as the sample value over the 5000 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.

When the number of missing observations increases to 20, it is seen that it is still true that for the AR model, the differences between the approaches are negligible. However, for the MA model the skipping approach becomes noticeably better, since all of the 20 missing values are then better estimated. For the mixed ARIMA model the results obtained with the SK and the AOC approaches are practically identical. The MC RMSE given by the AOC and AON approaches are similar for the MA model, although slightly smaller in the case of the AON approach. For the mixed ARIMA model, the AON approach presents MC RMSE which are systematically bigger than those of the AOC and SK approaches.

The results obtained for the MA model in the case of 20 missing values are a little paradoxical and should be taken with care. After some investigation, we have come to the conclusion that the results for this model when there is a large number of missing values are more affected by the numerical performance of the algorithms than in the case of the other models. For example, we have noticed that setting the initial value for the search in the non-linear optimization algorithm equal to the theoretical parameter -0.7 can reverse the results and make the AOC approach perform slightly better than the SK approach.

As for computational efficiency, Table 4 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, 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.

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 preferable in terms of speed, yielding estimators which are slightly less precise in some cases than those given by the AOC approach. Finally, from the precision point of view, enforcing the determinantal correction in the additive outlier approach may be important.

Appendix A. Proofs of results

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

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

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

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

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

*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 $\mathrm{Var}(v_m) = \sigma^2\Omega_{v_m}$, because z_{Im} is considered fixed in the definition of the likelihood (2.10). Therefore, the likelihood functions verify $L(y) = L(y_m|y_o)L(y_o)$ 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

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

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 $\hat{z}_{\mathrm{Im}} = (C_{\mathrm{o}}'\Omega_{v_{\mathrm{o}}}^{-1}C_{\mathrm{o}})^{-1}C_{\mathrm{o}}'\Omega_{v_{\mathrm{o}}}^{-1}y_{\mathrm{o}}$, whereas replacing in Eq. (2.10) z_{Im} with $\bar{z}_{\mathrm{Im}} - \omega_{\mathrm{I}}$ and maximizing with respect to ω_{I} yields the maximum likelihood estimator $\hat{\omega}_{\mathrm{I}} = -(C_{\mathrm{o}}'\Omega_{v_{\mathrm{o}}}^{-1}C_{\mathrm{o}})^{-1}C_{\mathrm{o}}'\Omega_{v_{\mathrm{o}}}^{-1}(y_{\mathrm{o}} - C_{\mathrm{o}}\bar{z}_{\mathrm{Im}}) = \bar{z}_{\mathrm{Im}} - \hat{z}_{\mathrm{Im}}$. Given that Eqs. (2.10) and (3.2) coincide, maximizing also Eq. (3.2) with respect to ω_{I} yields the same estimator $\hat{\omega}_{\mathrm{I}}$, and the estimator $\hat{\omega}_{\mathrm{II}}$ of Corollary 1 becomes

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

The estimator $\hat{\omega}_{\mathrm{II}}$ 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 $\hat{\omega}_{\mathrm{II}}$ minimizes the sum of squares $(u^* - X_{\mathrm{II}}^{*}\omega_{\mathrm{II}})'\Omega_u^{-1}(u^* - X_{\mathrm{II}}^{*}\omega_{\mathrm{II}})$, 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

$$(u^* - X_{\mathrm{II}}^{*}\omega_{\mathrm{II}})'\Omega_u^{-1}(u^* - X_{\mathrm{II}}^{*}\omega_{\mathrm{II}}) = (\bar{z} - X\omega)'[-A, \mathrm{I}_{N-d}]'J_2'\Omega_u^{-1}J_2 \times [-A, \mathrm{I}_{N-d}](\bar{z} - X\omega) \tag{A.3}$$

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 $\omega = [\omega_{\mathrm{I}}', \omega_{\mathrm{II}}']'$. Finally, it is easy to verify that the estimator $\hat{\omega}$ that minimizes (A.3) is the GLS estimator of the model $\bar{u} = X^*\omega + u$, where $u = J_{\mathrm{II}}z$ and $\mathrm{Var}(u) = \sigma^2\Omega_u$. □

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

(a) $\hat{z}_{\mathrm{Im}} = \bar{z}_{\mathrm{Im}} - \hat{\omega}_{\mathrm{I}}$

(b) $\hat{z}_{\mathrm{IIm}} = B_{\mathrm{m}}z_{\mathrm{I0}} + C_{\mathrm{m}}\hat{z}_{\mathrm{Im}} + E[z_{\mathrm{IIm}} - A_{\mathrm{m}}z_{\mathrm{I}} | z_{\mathrm{II0}} - B_{\mathrm{m}}z_{\mathrm{I0}} - C_{\mathrm{m}}\hat{z}_{\mathrm{Im}}] = \bar{z}_{\mathrm{IIm}} - \hat{\omega}_{\mathrm{II}}$,

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

$$J_{\mathrm{II}}\begin{bmatrix}\bar{z}_{\mathrm{I}} \\ \bar{z}_{\mathrm{II}}\end{bmatrix} = J_{\mathrm{II}}\begin{bmatrix}X_{\mathrm{I}} & 0 \\ 0 & X_{\mathrm{II}}\end{bmatrix}\begin{bmatrix}\omega_{\mathrm{I}} \\ \omega_{\mathrm{II}}\end{bmatrix} + J_{\mathrm{II}}\begin{bmatrix}z_{\mathrm{I}} \\ z_{\mathrm{II}}\end{bmatrix},$$

or, $\bar{u} = X^*\omega + u$. In the skipping approach, $\hat{z}_{\mathrm{Im}}$ is estimated by GLS in the model $y_0 = C_0 z_{\mathrm{Im}} + v_0$ (see Section 2.2). By Theorems 2 and 3, $\hat{z}_{\mathrm{Im}}$ in this model and (a) above coincide. Replacing z_{Im} by $\hat{z}_{\mathrm{Im}}$, and running the KF with initial conditions $x(d+1|d) = B_* z_{\mathrm{I0}} + C_* \hat{z}_{\mathrm{Im}}$, $\Sigma(d+1|d) = \Xi_* \tilde{\Sigma}(d+1|d)\Xi'_*$, yields (b) above. □

Proof of Lemma 3. With the notation of Section 3.3, consider the regression model $\bar{z} = [X, Y][\omega', \beta']' + v$. 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

$$\bar{u} = [X^*, Y^*][\omega', \beta']' + v^*,$$

where $\bar{u} = J_{\mathrm{II}}\bar{z}$, $X^* = J_{\mathrm{II}}X$, $Y^* = J_{\mathrm{II}}Y$, $v^* = J_{\mathrm{II}}v$, and $\mathrm{Var}(v^*) = \sigma^2\Omega$. Let $\Omega = LL'$ be the Cholesky decomposition of Ω, with L lower triangular, and write

$$X = \begin{bmatrix} X_{\mathrm{II}} & 0 \\ 0 & X_{\mathrm{I}} \end{bmatrix}; \quad X^* = J_{\mathrm{II}}X.$$

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 $[\omega'_{\mathrm{II}}, \omega'_{\mathrm{I}}] = \omega$, 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

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

where $Q = [Q'_1, Q'_2]'$ and the partition is made conforming to R and 0 in $[R', 0']'$. The determinantal correction is $|R_{\mathrm{II}}|^{1/(M-k)}$, 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, $|R_{\mathrm{II}}|$ remains constant as $M \to \infty$, $|R_{\mathrm{II}}|^{1/(M-k)} \to 1$

References

Anderson, B.D.A., Moore, J.B., 1979. Optimal Filtering. Prentice-Hall, Englewood cliffs, NJ.

Ansley, C.F., 1979. An algorithm for the exact likelihood of a mixed autoregressive–moving average process. Biometrika 66, 59–65.

Bell, W., 1984. Signal extraction for nonstationary series. The Annals of Statistics 12, 646–664.
Brockwell, P., Davis, R., 1987. Time Series: Theory and Methods. Springer, Berlin.
Box, G.E.P., Jenkins, G.M., 1976. Time Series Analysis, Forecasting and Control. Holden-Day, San Francisco.
Brubacher, S.R., Wilson, G.T., 1976. Interpolating time series with application to the estimation of holiday effects on electricity demand. Applied Statistics 25 (2), 107–116.
De Jong, Piet, 1991. The diffuse kalman filter. Annals of statistics 19, 1073–1083.
EUROSTAT, 1996. Is the ARIMA Model Adapted to Economic Time Series? Informal Group on Seasonal Adjustment, EUROSTAT /D3/SA/04.
Findley, D.F., Monsell, B.C., Bell, W.R., Otto, M.C., Chen, B.C., 1996. New capabilities and methods of the X-12-ARIMA seasonal adjustment program. Journal of Business and Economic Statistics, forthcoming.
Gómez, V., Maravall, A., 1994. Estimation, prediction and interpolation for nonstationary series with the kalman filter. Journal of the American Statistical Association 89, 611–624.
Gómez, V., Maravall, A., 1996. Programs TRAMO and SEATS, Instructions for the User, Beta Version: September 1996. Working paper No. 9628, Bank of Spain.
Gómez, V., Maravall, A., Peña, D., 1997. Missing Observations in ARIMA Models. Working paper no. 9701, Bank of Spain.
Harvey, A.C., 1989. Forecasting Structural Time Series Models and the Kalman Filter. Cambridge University Press, Cambridge.
Jones, R., 1980. Maximum likelihood fitting of ARMA models to time series with missing observations. Technometrics 22, 389–395.
Kohn, R., Ansley, C.F., 1985. Efficient estimation and prediction in time series regression models. Biometrika 72, 694–697.
Kohn, R., Ansley, C.F., 1986. Estimation, prediction and interpolation for ARIMA models with missing data. Journal of the American Statistical Association 81, 751–761.
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.
Maravall, A., Peña, D., 1996. Missing observations and additive outliers in time series models. Working paper no. 9612, Bank of Spain. In: Mariano, R. (Ed.), Advances in Statistical Analysis and Statistical Computing, vol. 3, JAI Press, Greenwich, forthcoming.
Mélard, G., 1984. A fast algorithm for the exact likelihood of autoregressive-moving average models. Applied Statistics 35, 104–114.
Morf, M., Sidhu, G.S., Kailath, T., 1974. Some new algorithms for recursive estimation on constant, linear, discrete-time systems. IEEE Transactions on Automatic Control AC-19, 315–323.
Peña, D., 1987. Measuring the importance of outliers in ARIMA models. In: Puri et al. M.L. (Eds.), New Perspectives in Theoretical and Applied Statistics. Wiley, New York, pp. 109–118.
Pearlman, J.G., 1980. An Algorithm for the exact likelihood of a high - order – autoregressive – moving average process. Biometrika 67, 232–233.
Sargan, J.D., Drettakis, E.G., 1974. Missing data in an autoregressive model. International Economic Review 15, 39–58.