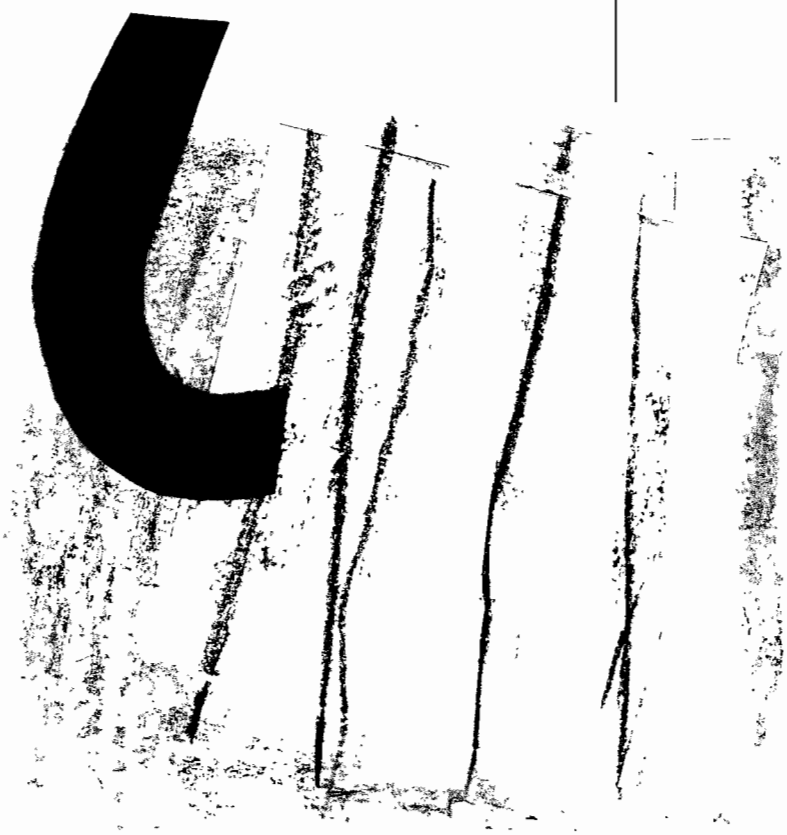
MISSING OBSERVATIONS IN ARIMA MODELS: SKIPPING STRATEGY VERSUS ADDITIVE OUTLIER APPROACH

Víctor Gómez, Agustín Maravall and Daniel Peña

97-15



WORKING PAPERS

Universidad Carlos III de Madrid

Working Paper 97-15
Statistics and Econometrics Series 08
February 1997

Departamento de Estadística y Econometría
Universidad Carlos III de Madrid
Calle Madrid, 126
28903 Getafe (Spain)
Fax (341) 624-9849

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

Víctor Gómez, Agustín Maravall and Daniel Peña*

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 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. Computationally efficient ways to apply both procedures, based on an Augmented Kalman Filter, are detailed. Finally, the two methods are compared through simulation, and their relative advantages assessed; the comparison also includes the AO method with the uncorrected likelihood.

Key Words

Time series, ARIMA models; missing observations; outliers; nonstationarity; likelihood; Kalman filter.

*Gómez, Ministerio de Economía y Hacienda; Maravall, Banco de España; Peña, Departamento de Estadística y Econometría, Universidad Carlos III de Madrid, e-mail: dpena@est-econ.uc3m.es.

Missing Observations in ARIMA Models: Skipping Strategy Versus Additive Outlier Approach

by

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

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 likihood (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 perfoming 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. Computationally efficient ways to apply both procedures, based on an Augmented Kalman Filter, are detailed. Finally, the two methods are compared through simulation, and their relative advantages assessed; the comparison also includes the AO method with the uncorrected likelihood.

KEY WORDS: Time Series, ARIMA Models, Missing Observations, Outliers, Nonstationarity, Likelihood, Kalman Filter.

* Ministerio de Economía y Hacienda, P. Castellana 162, E–28046 Madrid, Spain.
** Bank of Spain, Madrid.
*** Universidad Carlos III, Madrid

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

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 (2.2*b*) 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.

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

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

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

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

"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}).

$z_N \mid z_1, \ldots, z_d)$. 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, $Mse(\hat{\omega}) = Var(z_m \mid z_o) = \sigma^2(X'\Omega_z^{-1}X)^{-1}$.

$\hat{\omega}_{II} = \bar{z}_{IIm} - A_m z_I - E(z_{IIm} - A_m z_I \mid z_{IIo} - A_o z_I)$. *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.

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

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.

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 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 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 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

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).

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_{ν_0} 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 $x(t \mid t-1)$

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 $Y_{IIo} - A_oY_I$ 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 $X(d+1 \mid d) = (B_*z_{Io}, -C_*, -A_*Y_I)$, 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 $Y_{IIo} - A_oY_I$ 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

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

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

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 $v(s) = (\bar{u}(s), x^{*'}(s), y^{*'}(s))$, $r = max\{p, q+1\}$ 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,

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

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

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

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 $x(d+1 \mid d) = B_*\, z_{I0} + C_*\, \hat{z}_{Im}$, $\Sigma(d+1 \mid d) = \Xi_*\, \tilde{\Sigma}(d+1 \mid d)\, \Xi_*'$, 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

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

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.