

# Computing and Using Residuals in Time Series Models

José Alberto Mauricio

Universidad Complutense de Madrid

---

**Abstract:** The most often used approaches to obtaining and using residuals in applied work with time series models, are unified and documented with both partially-known and new features. Specifically, three different types of residuals, namely “conditional residuals”, “unconditional residuals” and “innovations”, are considered with regard to (i) their precise definitions, (ii) their computation in practice after model estimation, (iii) their approximate distributional properties in finite samples, and (iv) potential applications of their properties in model diagnostic checking. The focus is on both conditional and unconditional residuals, whose properties have received very limited attention in the literature. However, innovations are also briefly considered in order to provide a comprehensive description of the various classes of residuals a time series analyst might find in applied work. Theoretical discussion is accompanied by practical examples, illustrating (a) that routine application of standard model-building procedures may lead to inaccurate models in cases of practical interest which are easy to come across, and (b) that such inaccuracies can be avoided by using some of the new results on conditional and unconditional residuals developed with regard to points (iii) and (iv) above. For ease and clarity of exposition, only stationary univariate autoregressive moving average models are considered in detail, although extensions to the multivariate case are briefly discussed as well.

**Keywords:** Autoregressive moving average model; Conditional residuals; Innovations; Normalized residuals; Unconditional residuals.

---

First published in *Computational Statistics and Data Analysis*, Vol. 52, No. 3, pp. 1746-1763

January 2008

## 1. INTRODUCTION

The approach to time series model building and forecasting introduced three decades ago by Box and Jenkins (1976) still represents one of the fundamental cornerstones in modern time series analysis. Since its original development, the usefulness of the so-called “Box-Jenkins approach” has been demonstrated through a massive number of publications dealing with both practical applications and theoretical contributions, many of which have been surveyed in such popular textbooks as Mills (1990), Harvey (1993), Box et al. (1994), Reinsel (1997), Franses (1998), Peña et al. (2001), Brockwell and Davis (2002), and Peña (2005).

This article contributes to the current state of the Box-Jenkins approach by presenting some theoretical and numerical properties of several classes of residuals in time series models which do not seem to have received much attention in the literature. Along with some partially-known results, new theoretical derivations are presented together with applications in model diagnostic checking, which is perhaps the most fundamental (although sometimes overlooked) stage in the Box-Jenkins iterative approach to building time series models. For ease and clarity of exposition, this article deals mainly with stationary, univariate autoregressive moving average (ARMA) models. Nonetheless, extensions to the case of stationary multivariate models are also considered, implying that similar results to those presented below can be shown to hold for any time series model which can be cast into a standard, stationary vector ARMA model, including, among many others, transfer function-noise models (Mauricio 1996) and partially nonstationary multivariate models for cointegrated processes (Mauricio 2006a).

Residuals constitute a critical piece of information at the diagnostic checking stage of a tentatively entertained model, where one seeks evidence that either assesses the adequacy of the model or provides directions along which it might be modified. The usual approach to deciding on these subjects consists of comparing patterns in computed residuals to those implied by their expected distributional properties under the assumption that the entertained model is adequate. Hence, in any given practical setting, it seems important to know which residuals are being used for model diagnostic checking (i.e., how such residuals have been actually computed), and which distributional properties should observed patterns in such residuals be compared to. In this respect, it is a standard practice at the diagnostic checking stage simply to compare residual patterns to those of Gaussian white noise (see, for example, Mills 1990, ch. 8; Box et al. 1994, ch. 8; Reinsel 1997, ch. 5; Franses 1998, ch. 3; Li 2004; and Peña 2005, ch. 11). However, it is at least partially known (see Harvey 1993, p. 76, for a general statement on this subject) that residuals from ARMA models do not have the statistical properties assumed on the random shocks of such models. Hence, the above standard practice, although firmly established, should not be recommended in general. A detailed simulation study supporting this point of view (especially for models with seasonal structure) was given nearly thirty years ago by Ansley and Newbold (1979), and, to some extent, new results developed in the present article provide a theoretical justification for the empirical findings of these authors.

Various types of residuals are currently available for being used at the diagnostic checking stage of a tentative model. Thus, as it often happens with parameter estimation (see Newbold et al. 1994), residual calculations usually differ among different computer programs for analyzing time series data. In fact,

there seems to exist some confusion in current practice as to which residuals are used in any given empirical work with time series models, how are they computed, and which are their theoretical distributional properties. Hence, the fundamental aim of the present article is to make practically accessible for time series analysts both partially-known and new results on residuals from time series models, by means of showing what might be found (and, to some extent, what should be done) in practice when dealing with residuals after estimation of any given model. It must be stressed that residuals are important because observed patterns in their plots and/or in their sample autocorrelations, are usually taken as an indication of (apart from possible outliers) a misspecified model structure which needs to be reconsidered. In order to use this idea in practice as effectively as possible, one should know as much as possible about the generating mechanism of residuals on which diagnostic checks are applied. Note that acceptance or reformulation of a tentatively entertained model might simply arise due to its inefficient estimation as well as due to a poor diagnostic checking procedure, regardless of whether the form of the model is adequate. In particular, forecasts coming from an inefficiently fitted and/or poorly diagnostically checked model, might be useless both on their own and as a reference point when assessing the performance of new forecasting methods.

In Section 2 of this article, three different types of residuals, namely “conditional residuals”, “unconditional residuals” and “innovations”, are precisely defined, and several explicit expressions are given for computing them in practice. In Section 3, it is shown (i) that both conditional and unconditional residuals follow approximate zero-mean distributions in finite samples, with covariance matrices for which explicit and easily computable expressions are given for the first time, and (ii) that invertibility plays a key role for establishing statistical convergence of residuals to white noise, implying that conditional and unconditional residual autocorrelations should be interpreted carefully in applied work with models whose parameter values lie on or near the boundary of the invertibility region. In Section 4, it is shown that a set of “normalized” (i.e., homoskedastic and uncorrelated) residuals can be obtained in any of several equivalent ways; according to previous work on the subject, using this set of residuals for diagnostic checking usually improves the chances of not rejecting a tentative model with autocorrelated conditional or unconditional residuals when the model is adequately specified. In Section 5, the potential practical advantages of using a proper set of residuals for diagnostic checking are illustrated through practical examples. Some guidelines on extending results to the case of multivariate time series models are given in Section 6. Finally, additional discussion and conclusions are provided in Section 7.

## 2. RESIDUAL DEFINITIONS AND COMPUTATIONS IN UNIVARIATE MODELS

Let an observed time series  $\mathbf{w} = [w_1, w_2, \dots, w_n]^T$  be generated by a stationary process  $\{W_t\}$  following the model

$$\phi(B)\tilde{W}_t = \theta(B)A_t, \quad (1)$$

where  $\phi(B) = 1 - \sum_{i=1}^p \phi_i B^i$  and  $\theta(B) = 1 - \sum_{i=1}^q \theta_i B^i$  are polynomials of degrees  $p$  and  $q$ ,  $B$  is the backshift operator,  $\tilde{W}_t = W_t - E[W_t]$ , and  $\{A_t\}$  is a white noise process with variance  $\sigma^2 > 0$ . For stationarity, the roots of  $\phi(x) = 0$  must lie outside the unit circle; a similar condition on  $\theta(x)$  ensures that the model is invertible. It is also assumed that  $\phi(x)$  and  $\theta(x)$  do not share any common factor.

Let  $\tilde{\mathbf{W}} = [\tilde{W}_1, \tilde{W}_2, \dots, \tilde{W}_n]^T$  ( $n \times 1$ ),  $\mathbf{A} = [A_1, A_2, \dots, A_n]^T$  ( $n \times 1$ ), and  $\mathbf{U}_* = [\tilde{W}_{1-p}, \dots, \tilde{W}_0, A_{1-q}, \dots, A_0]^T$   $[(p+q) \times 1]$ , and consider Eq. (1) for  $t = 1, 2, \dots, n$ . Then,  $\mathbf{w}$  can be regarded as a particular realization of a random vector  $\mathbf{W} = [W_1, W_2, \dots, W_n]^T$  following the model

$$\mathbf{D}_\phi \tilde{\mathbf{W}} = \mathbf{D}_\theta \mathbf{A} + \mathbf{V} \mathbf{U}_*, \quad (2)$$

where  $\mathbf{D}_\phi$  and  $\mathbf{D}_\theta$  are  $n \times n$  parameter matrices with 1's on the main diagonal and  $-\phi_j$  and  $-\theta_j$ , respectively, down the  $j$ th subdiagonal, and  $\mathbf{V}$  is an  $n \times (p+q)$  matrix with  $V_{ij} = \phi_{p+i-j}$  ( $i = 1, \dots, p; j = i, \dots, p$ ),  $V_{ij} = -\theta_{q+i-j+p}$  ( $i = 1, \dots, q; j = p+i, \dots, p+q$ ), and zeros elsewhere.

A useful approach to introducing different methods for computing residuals from estimated ARMA models, consists of considering which residuals arise naturally within different methods for computing the exact (unconditional) log-likelihood function for univariate models of the form (1) or (2). Under the assumption that  $\{W_t\}$  is a Gaussian (Normal) process, the exact log-likelihood computed at a given set of parameter estimates  $\hat{\boldsymbol{\beta}} = [\hat{\mu}, \hat{\phi}_1, \dots, \hat{\phi}_p, \hat{\theta}_1, \dots, \hat{\theta}_q]^T$  and  $\hat{\sigma}^2$ , can be written as

$$l(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2 | \mathbf{w}) = -\frac{1}{2} [n \log(2\pi\hat{\sigma}^2) + \log |\hat{\boldsymbol{\Sigma}}_{\mathbf{W}}| + \hat{\sigma}^{-2} \tilde{\mathbf{w}}^T \hat{\boldsymbol{\Sigma}}_{\mathbf{W}}^{-1} \tilde{\mathbf{w}}], \quad (3)$$

where  $\tilde{\mathbf{w}} = [\tilde{w}_1, \tilde{w}_2, \dots, \tilde{w}_n]^T$  with  $\tilde{w}_t = w_t - \hat{\mu}$  ( $t = 1, 2, \dots, n$ ),  $\hat{\mu}$  is an estimate of  $E[W_t]$ , and the  $n \times n$  matrix  $\hat{\boldsymbol{\Sigma}}_{\mathbf{W}}$  is an estimate of the theoretical ‘‘auto-covariance’’ matrix  $\boldsymbol{\Sigma}_{\mathbf{W}} = \sigma^{-2} E[\tilde{\mathbf{W}}\tilde{\mathbf{W}}^T]$ .

Noting (2), it can be seen that the auto-covariance matrix  $\boldsymbol{\Sigma}_{\mathbf{W}}$  is given by

$$\boldsymbol{\Sigma}_{\mathbf{W}} = \mathbf{D}_\phi^{-1} (\mathbf{D}_\theta \mathbf{D}_\theta^T + \mathbf{V} \boldsymbol{\Omega} \mathbf{V}^T) \mathbf{D}_\phi^{-1T} = \mathbf{K}^{-1} (\mathbf{I} + \mathbf{Z} \boldsymbol{\Omega} \mathbf{Z}^T) \mathbf{K}^{-1T}, \quad (4)$$

where  $\mathbf{K} = \mathbf{D}_\theta^{-1} \mathbf{D}_\phi$ ,  $\mathbf{Z} = -\mathbf{D}_\theta^{-1} \mathbf{V}$  and  $\boldsymbol{\Omega} = \sigma^{-2} E[\mathbf{U}_* \mathbf{U}_*^T]$  are  $n \times n$ ,  $n \times (p+q)$  and  $(p+q) \times (p+q)$ , respectively, parameter matrices, with  $\boldsymbol{\Omega}$  being readily expressible in terms of  $\phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$  as described, for example, in Ljung and Box (1979). Hence, using Eq. (4), the quadratic form  $\tilde{\mathbf{w}}^T \hat{\boldsymbol{\Sigma}}_{\mathbf{W}}^{-1} \tilde{\mathbf{w}}$  in Eq. (3) can be written as

$$\tilde{\mathbf{w}}^T \hat{\boldsymbol{\Sigma}}_{\mathbf{W}}^{-1} \tilde{\mathbf{w}} = \tilde{\mathbf{w}}^T \hat{\mathbf{K}}^T (\mathbf{I} + \hat{\mathbf{Z}} \hat{\boldsymbol{\Omega}} \hat{\mathbf{Z}}^T)^{-1} \hat{\mathbf{K}} \tilde{\mathbf{w}}, \quad (5)$$

where  $\hat{\mathbf{K}}$ ,  $\hat{\mathbf{Z}}$  and  $\hat{\boldsymbol{\Omega}}$  are estimates of the corresponding parameter matrices defined below (4). From Eq. (5), three different classes of residuals can be defined for a given set of parameter estimates as follows:

**Definition 1 – Conditional Residuals.** The ‘‘conditional residuals’’ associated with Eq. (5) are the elements of the  $n \times 1$  vector  $\hat{\mathbf{a}}^0 = \hat{\mathbf{K}} \tilde{\mathbf{w}}$ .  $\square$

**Definition 2 – Unconditional Residuals.** The ‘‘unconditional residuals’’ associated with Eq. (5) are the elements of the  $n \times 1$  vector  $\hat{\mathbf{a}} = (\mathbf{I} + \hat{\mathbf{Z}} \hat{\boldsymbol{\Omega}} \hat{\mathbf{Z}}^T)^{-1} \hat{\mathbf{K}} \tilde{\mathbf{w}} = \hat{\boldsymbol{\Sigma}}_0^{-1} \hat{\mathbf{a}}^0$ , where  $\hat{\boldsymbol{\Sigma}}_0$  is an estimate of the  $n \times n$  matrix  $\boldsymbol{\Sigma}_0 = \mathbf{I} + \mathbf{Z} \boldsymbol{\Omega} \mathbf{Z}^T$ , or, equivalently (e.g., Dhrymes 2000, p.44),  $\boldsymbol{\Sigma}_0 = [\mathbf{I} - \mathbf{Z}(\boldsymbol{\Omega}^{-1} + \mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T]^{-1}$ .  $\square$

**Definition 3 – Innovations.** The ‘‘innovations’’ associated with Eq. (5) are the elements of the  $n \times 1$  vector  $\hat{\mathbf{e}} = \hat{\mathbf{L}}^{-1} \tilde{\mathbf{w}} = (\hat{\mathbf{K}} \hat{\mathbf{L}})^{-1} \hat{\mathbf{a}}^0$ , where  $\hat{\mathbf{L}}$  is an estimate of the  $n \times n$  unit lower-triangular matrix  $\mathbf{L}$  in the factorization

$$\boldsymbol{\Sigma}_{\mathbf{W}} = \mathbf{L} \mathbf{F} \mathbf{L}^T = \begin{bmatrix} 1 & 0 & \dots & 0 \\ L_{21} & 1 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ L_{n1} & L_{n2} & \dots & 1 \end{bmatrix} \begin{bmatrix} F_1 & 0 & \dots & 0 \\ 0 & F_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & F_n \end{bmatrix} \begin{bmatrix} 1 & L_{21} & \dots & L_{n1} \\ 0 & 1 & \dots & L_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & 1 \end{bmatrix},$$

with  $F_t > 0$  for every  $t = 1, 2, \dots, n$ .  $\square$

Using Definitions 1 through 3, it can be seen that (5) can be written in three equivalent ways as  $\tilde{\mathbf{w}}^T \hat{\Sigma}_{\mathbf{W}}^{-1} \tilde{\mathbf{w}} = \hat{\mathbf{a}}^{0T} \hat{\Sigma}_0^{-1} \hat{\mathbf{a}}^0 = \hat{\mathbf{a}}^T \hat{\Sigma}_0 \hat{\mathbf{a}} = \hat{\mathbf{e}}^T \hat{\mathbf{F}}^{-1} \hat{\mathbf{e}}$ , where  $\hat{\mathbf{F}}$  is an estimate of the diagonal matrix  $\mathbf{F}$  in Definition 3. Note also, as a by-product, that  $|\hat{\Sigma}_{\mathbf{W}}|$  in (4) equals both  $|\hat{\Sigma}_0|$  and  $|\hat{\mathbf{F}}|$ . Hence, any out of the three types of residuals defined previously can be used to compute the exact log-likelihood given in (3). Some links between residuals defined thus far and usual ideas about residuals in classical time series analysis are considered in the following remarks.

**Remark 1.** The univariate ARMA model given in (2) can be written as  $\mathbf{A} = \mathbf{K}\tilde{\mathbf{W}} + \mathbf{Z}\mathbf{U}_*$ , where  $\mathbf{K}$  and  $\mathbf{Z}$  are defined below (4). Hence, for a given set of parameter estimates, the conditional residual vector (Definition 1) can be written as

$$\hat{\mathbf{a}}^0 = \hat{\mathbf{K}}\tilde{\mathbf{w}} = \hat{\mathbf{E}}[\mathbf{A} \mid \mathbf{W} = \mathbf{w}, \mathbf{U}_* = \mathbf{0}], \quad (6)$$

which represents the estimated expectation of the random-shock vector  $\mathbf{A}$  given an observed time series  $\mathbf{w}$ , under the condition that  $\mathbf{U}_* = \mathbf{0}$  (i.e., that  $\tilde{W}_{1-p} = \dots = \tilde{W}_0 = A_{1-q} = \dots = A_0 = 0$ ). On the other hand, the unconditional residual vector (Definition 2) can be written as

$$\begin{aligned} \hat{\mathbf{a}} &= (\mathbf{I} + \hat{\mathbf{Z}}\hat{\Omega}\hat{\mathbf{Z}}^T)^{-1} \hat{\mathbf{K}}\tilde{\mathbf{w}} = \left[ \mathbf{I} - \hat{\mathbf{Z}}(\hat{\Omega}^{-1} + \hat{\mathbf{Z}}^T\hat{\mathbf{Z}})^{-1} \hat{\mathbf{Z}}^T \right] \hat{\mathbf{K}}\tilde{\mathbf{w}} \\ &= \hat{\mathbf{K}}\tilde{\mathbf{w}} - \hat{\mathbf{Z}}(\hat{\Omega}^{-1} + \hat{\mathbf{Z}}^T\hat{\mathbf{Z}})^{-1} \hat{\mathbf{Z}}^T \hat{\mathbf{K}}\tilde{\mathbf{w}} = \hat{\mathbf{K}}\tilde{\mathbf{w}} + \hat{\mathbf{Z}}\hat{\mathbf{u}}_*, \end{aligned}$$

with  $\hat{\mathbf{u}}_*$  being defined as  $\hat{\mathbf{u}}_* = -(\hat{\Omega}^{-1} + \hat{\mathbf{Z}}^T\hat{\mathbf{Z}})^{-1} \hat{\mathbf{Z}}^T \hat{\mathbf{K}}\tilde{\mathbf{w}}$ . Additionally, it can be shown (see Box et al. 1994, pp. 292-294) that  $\hat{\mathbf{u}}_*$  so defined equals the estimated conditional expectation  $\hat{\mathbf{E}}[\mathbf{U}_* \mid \mathbf{W} = \mathbf{w}]$  (which is usually referred to as the “back-casted” value of the “pre-sample” vector  $\mathbf{U}_*$ ), implying in turn that

$$\hat{\mathbf{a}} = \hat{\mathbf{K}}\tilde{\mathbf{w}} + \hat{\mathbf{Z}}\hat{\mathbf{u}}_* = \hat{\mathbf{E}}[\mathbf{A} \mid \mathbf{W} = \mathbf{w}]. \quad (7)$$

In contrast to (6), Eq. (7) represents the estimated expectation of  $\mathbf{A}$  for an observed time series  $\mathbf{w}$  under no additional conditions.

**Remark 2.** Eq. (6) implies that  $\hat{\mathbf{D}}_\theta \hat{\mathbf{a}}^0 = \hat{\mathbf{D}}_\phi \tilde{\mathbf{w}}$  (recall from (4) that  $\mathbf{K} = \mathbf{D}_\theta^{-1} \mathbf{D}_\phi$ ), so that the elements of  $\hat{\mathbf{a}}^0$  can be computed recursively as  $\hat{a}_t^0 = w_t - [\hat{\mu} + \sum_{i=1}^p \hat{\phi}_i (w_{t-i} - \hat{\mu}) - \sum_{i=1}^q \hat{\theta}_i \hat{a}_{t-i}^0]$  ( $t = 1, \dots, n$ ), with  $w_j - \hat{\mu} = 0$  (i.e.,  $w_j = \hat{\mu}$ ) and  $\hat{a}_j^0 = 0$  for  $j < 1$ . On the other hand, Eq. (7) implies that  $\hat{\mathbf{D}}_\theta \hat{\mathbf{a}} = \hat{\mathbf{D}}_\phi \tilde{\mathbf{w}} - \hat{\mathbf{V}}\hat{\mathbf{u}}_*$  (recall from (4) that  $\mathbf{K} = \mathbf{D}_\theta^{-1} \mathbf{D}_\phi$  and  $\mathbf{Z} = -\mathbf{D}_\theta^{-1} \mathbf{V}$ ), so that the elements of  $\hat{\mathbf{a}}$  can be computed recursively as  $\hat{a}_t = w_t - [\hat{\mu} + \sum_{i=1}^p \hat{\phi}_i (w_{t-i} - \hat{\mu}) - \sum_{i=1}^q \hat{\theta}_i \hat{a}_{t-i}]$  ( $t = 1, \dots, n$ ), with values for  $w_j - \hat{\mu}$  ( $j = 1-p, \dots, 0$ ) and  $\hat{a}_j$  ( $j = 1-q, \dots, 0$ ) taken from the back-cast vector  $\hat{\mathbf{u}}_*$  given above (7) (note, however, that Definition 2 suggests a more direct method for computing  $\hat{\mathbf{a}}$ ). Hence, both of  $\hat{a}_t^0$  and  $\hat{a}_t$  are simply differences between an observed value  $w_t$  and a corresponding fitted value (i.e., they are one-step-ahead forecast errors), which means that both conditional and unconditional residuals are residuals in a fully usual sense.

**Remark 3.** The innovations introduced in Definition 3 arise naturally when considering the “innovations form” of the exact log-likelihood (3), described, for example, in Ansley (1979) and Box et al. (1994, pp. 275-279); see also Mélard (1984) for a description of the innovations form of (3) from the perspective of state-space methods. Despite the fact that innovations do not follow right from  $t = 1$  the recursive relations considered in Remark 2, they can still be interpreted as one-step-ahead forecast errors (see Brockwell and Davis 2002, pp. 100-108), so that innovations are also residuals in a fairly usual sense.

**Remark 4.** Other types of residuals have been considered in previous literature on ARMA model building. For example, a type of residuals which is frequently used in practice (especially with regard to conditional least-squares estimation of pure autoregressive models) can be defined through the recursions  $\hat{c}_t = w_t - [\hat{\mu} + \sum_{i=1}^p \hat{\phi}_i (w_{t-i} - \hat{\mu}) - \sum_{i=1}^q \hat{\theta}_i \hat{c}_{t-i}]$  for  $t = p+1, \dots, n$ , with  $\hat{c}_j$  for  $j \leq p$  either set to zero or estimated (back-casted) from the data. For this type of residuals, the  $p$  initial values  $\hat{c}_1, \hat{c}_2, \dots, \hat{c}_p$  are usually not reported, which may constitute an important loss of information (e.g., with regard to possible outliers) especially when  $p$  is large (e.g., when dealing with seasonal autoregressive structures; see Mauricio 1995, pp. 288-290, for an example). This type of residuals is not considered further here.

### 3. RESIDUAL PROPERTIES IN UNIVARIATE MODELS

The three types of residuals considered in Section 2 are all different from each other. This might explain, at least to some extent, why different computer programs usually generate different residuals from a given estimated model for a given time series, even when such programs had produced similar parameter estimates (see, however, Newbold et al. 1994, for an exercise showing that different computer programs usually calculate notably different parameter estimates, which makes things for the practitioner even more confusing). Nonetheless, some unifying sense can be gained from the fact that the three types of residuals considered in Section 2 share the following properties: (i) all of them are associated with the exact log-likelihood given in (3), (ii) all of them can be expressed in terms of the unconditional residual vector, as seen in Definitions 1 through 3, and (iii) all of them can be used to compute a unique set of “normalized” residuals in any of several equivalent ways. This last property is considered in Section 4 by means of using some of the theoretical properties of residuals to which we now turn.

**Theorem 1 – Properties of Conditional Residuals.** Let  $\hat{\mathbf{A}}^0 = \mathbf{K}\tilde{\mathbf{W}}$  be the random vector associated with the conditional residuals given in Definition 1, under the assumption that the true parameter values  $\mu, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$  of the stationary model (2) are known. Then:

(A)  $E[\hat{\mathbf{A}}^0] = \mathbf{0}$ ,  $\text{Var}[\hat{\mathbf{A}}^0] = \sigma^2 (\mathbf{I} + \mathbf{Z}\mathbf{\Omega}\mathbf{Z}^T)$ , and

(B)  $E[(\hat{\mathbf{A}}^0 - \mathbf{A})(\hat{\mathbf{A}}^0 - \mathbf{A})^T] = \sigma^2 \mathbf{Z}\mathbf{\Omega}\mathbf{Z}^T$ . □

**Proof.** Part (A) follows from  $\hat{\mathbf{A}}^0 = \mathbf{K}\tilde{\mathbf{W}}$  with  $E[\tilde{\mathbf{W}}] = \mathbf{0}$  and  $\text{Var}[\tilde{\mathbf{W}}] = \sigma^2 \mathbf{\Sigma}_{\tilde{\mathbf{W}}}$ , where  $\mathbf{\Sigma}_{\tilde{\mathbf{W}}}$  is given in (4). Part (B) follows from Remark 1 (i.e.,  $\hat{\mathbf{A}}^0 - \mathbf{A} = -\mathbf{Z}\mathbf{U}_*$ ) and the definition of  $\mathbf{\Omega}$  given below (4). ■

**Theorem 2 – Properties of Unconditional Residuals.** Let  $\hat{\mathbf{A}} = \mathbf{\Sigma}_0^{-1} \hat{\mathbf{A}}^0$  be the random vector associated with the unconditional residuals given in Definition 2, under the assumption that the true parameter values  $\mu, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$  of the stationary model (2) are known. Then:

(A)  $E[\hat{\mathbf{A}}] = \mathbf{0}$ ,  $\text{Var}[\hat{\mathbf{A}}] = \sigma^2 (\mathbf{I} + \mathbf{Z}\mathbf{\Omega}\mathbf{Z}^T)^{-1}$ , and

(B)  $E[(\hat{\mathbf{A}} - \mathbf{A})(\hat{\mathbf{A}} - \mathbf{A})^T] = \sigma^2 \mathbf{Z}(\mathbf{\Omega}^{-1} + \mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T$ . □

**Proof.** Part (A) follows from part (A) of Theorem 1, recalling from Definition 2 that  $\mathbf{\Sigma}_0 = \mathbf{I} + \mathbf{Z}\mathbf{\Omega}\mathbf{Z}^T = [\mathbf{I} - \mathbf{Z}(\mathbf{\Omega}^{-1} + \mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T]^{-1}$ . To prove part (B), note from Remark 1 that  $\hat{\mathbf{A}}^0 = \mathbf{A} - \mathbf{Z}\mathbf{U}_*$ . Hence,  $\hat{\mathbf{A}} - \mathbf{A} = \mathbf{\Sigma}_0^{-1} \hat{\mathbf{A}}^0 - \mathbf{A} = \mathbf{\Sigma}_0^{-1} (\mathbf{A} - \mathbf{Z}\mathbf{U}_*) - \mathbf{A} = -[(\mathbf{I} - \mathbf{\Sigma}_0^{-1})\mathbf{A} + \mathbf{\Sigma}_0^{-1} \mathbf{Z}\mathbf{U}_*]$ , and  $E[(\hat{\mathbf{A}} - \mathbf{A})(\hat{\mathbf{A}} - \mathbf{A})^T] = \sigma^2 [(\mathbf{I} - \mathbf{\Sigma}_0^{-1})(\mathbf{I} - \mathbf{\Sigma}_0^{-1}) + \mathbf{\Sigma}_0^{-1} \mathbf{Z}\mathbf{\Omega}\mathbf{Z}^T \mathbf{\Sigma}_0^{-1}]$ . Part (B) then follows from the two equivalent expressions for  $\mathbf{\Sigma}_0$  recalled above. ■

**Theorem 3 – Convergence of Conditional and Unconditional Residuals.** Under the assumptions of Theorems 1 and 2, invertibility of the stationary ARMA model (2) implies additionally that:

- (A)  $E[(\hat{A}_i^0 - A_i)^2] \rightarrow 0$ ,  $E[(\hat{A}_i^0)^2] \rightarrow \sigma^2$ , and  $E[\hat{A}_i^0 \hat{A}_j^0] \rightarrow 0$  ( $i \neq j$ ) for increasing  $i$  and fixed  $j$  (at an either small or large value);
- (B)  $E[(\hat{A}_i - A_i)^2] \rightarrow 0$ ,  $E[\hat{A}_i^2] \rightarrow \sigma^2$ , and  $E[\hat{A}_i \hat{A}_j] \rightarrow 0$  ( $i \neq j$ ) for increasing  $i$  and fixed  $j$  (at an either small or large value); and
- (C) (A) and (B) hold exactly for  $i \geq p + 1$  and  $j \geq 1$  when  $q = 0$ . □

**Proof.** Part (B) of Theorem 1 can be written, recalling from the beginning of Section 2 that  $\mathbf{Z} = -\mathbf{D}_\theta^{-1}\mathbf{V}$ , as

$$E[(\hat{\mathbf{A}}^0 - \mathbf{A})(\hat{\mathbf{A}}^0 - \mathbf{A})^T] = \sigma^2 \mathbf{D}_\theta^{-1} \begin{bmatrix} \mathbf{V}_1 \boldsymbol{\Omega} \mathbf{V}_1^T & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{D}_\theta^{-1T} = \sigma^2 \mathbf{H} \mathbf{V}_1 \boldsymbol{\Omega} \mathbf{V}_1^T \mathbf{H}^T, \quad (8)$$

where the  $g \times (p+q)$  matrix  $\mathbf{V}_1$  consists of the first  $g = \max\{p, q\}$  rows of  $\mathbf{V}$ , and the  $n \times g$  matrix  $\mathbf{H}$  consists of the first  $g$  columns of  $\mathbf{D}_\theta^{-1}$ . Similarly, part (B) of Theorem 2 can be written as

$$\begin{aligned} E[(\hat{\mathbf{A}} - \mathbf{A})(\hat{\mathbf{A}} - \mathbf{A})^T] &= \sigma^2 \mathbf{D}_\theta^{-1} \begin{bmatrix} \mathbf{V}_1 (\boldsymbol{\Omega}^{-1} + \mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{V}_1^T & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \mathbf{D}_\theta^{-1T} \\ &= \sigma^2 \mathbf{H} \mathbf{V}_1 (\boldsymbol{\Omega}^{-1} + \mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{V}_1^T \mathbf{H}^T. \end{aligned} \quad (9)$$

Additionally, Theorems 1 and 2 imply that

$$\text{Var}[\hat{\mathbf{A}}^0] = \sigma^2 \mathbf{I} + E[(\hat{\mathbf{A}}^0 - \mathbf{A})(\hat{\mathbf{A}}^0 - \mathbf{A})^T] \quad (10)$$

and

$$\text{Var}[\hat{\mathbf{A}}] = \sigma^2 \mathbf{I} - E[(\hat{\mathbf{A}} - \mathbf{A})(\hat{\mathbf{A}} - \mathbf{A})^T], \quad (11)$$

respectively. It is now noted (see, for example, Ljung and Box 1979) that  $\mathbf{D}_\theta^{-1}$  is an  $n \times n$  lower triangular parameter matrix with 1's on the main diagonal and  $\Xi_t$  down the  $t$ th subdiagonal, where the elements  $\Xi_t$  for  $t = 0, 1, \dots, n-1$  satisfy the difference equation  $\theta(B)\Xi_t = 0$ , with  $\Xi_0 = 1$  and  $\Xi_t = 0$  for  $t < 0$ . Hence, the  $(i, j)$ th element of  $\mathbf{H}$  in (8) and (9) is given by  $\Xi_{i-j}$  ( $i = 1, \dots, n$ ;  $j = 1, \dots, g$ ). Noting finally that invertibility implies that  $\Xi_t \rightarrow 0$  as  $t$  increases, it can be seen that the  $(i, j)$ th element of both (8) and (9) tends to zero, and that the  $(i, i)$ th element of both (10) and (11) tends to one, as  $i$  increases. This proves parts (A) and (B) of the theorem. Part (C) follows from the fact that  $\mathbf{H}$  in (8) and (9) has the form  $\mathbf{H} = [\mathbf{I}, \mathbf{0}]^T$  (i.e.,  $\Xi_t = 0$  for all  $t \neq 0$ ) when  $q = 0$  (i.e., in the case of pure autoregressive models). ■

The main results of Theorems 1 through 3 can be summarized as follows: Assuming that the true parameter values in the stationary model given in (1) or (3) are perfectly known, it follows that (i)  $\hat{\mathbf{A}}^0 = \mathbf{K}\tilde{\mathbf{W}} \sim (\mathbf{0}, \sigma^2 \boldsymbol{\Sigma}_0)$  with  $\boldsymbol{\Sigma}_0 = \mathbf{I} + \mathbf{Z}\boldsymbol{\Omega}\mathbf{Z}^T$ , and (ii)  $\hat{\mathbf{A}} = \boldsymbol{\Sigma}_0^{-1} \hat{\mathbf{A}}^0 \sim (\mathbf{0}, \sigma^2 \boldsymbol{\Sigma}_0^{-1})$  with  $\boldsymbol{\Sigma}_0^{-1} = \mathbf{I} - \mathbf{Z}(\boldsymbol{\Omega}^{-1} + \mathbf{Z}^T \mathbf{Z})^{-1} \mathbf{Z}^T$ ; in addition, if the model considered is invertible, then (iii) both conditional and unconditional residuals converge in mean square to the model white noise disturbances, (iv) both conditional and unconditional residuals tend to be uncorrelated, with  $\text{Var}[\hat{A}_t^0]$  converging from above and  $\text{Var}[\hat{A}_t]$  converging from below to  $\sigma^2$ , and (v) when  $q = 0$  (i.e., in the case of pure autoregressive models), the convergence results stated in (iii) and (iv) occur exactly at time  $t = p + 1$ . Some aspects of the practical relevance of these results are considered in the following remarks:

**Remark 5.** Theorems 1 through 3 imply, in the first place, that both conditional and unconditional residuals should not be expected in general to follow white noise patterns, even under perfect knowledge of the true model parameter values. If such values are replaced by consistent estimates (which is usually the case in applied analyses with non-experimental data), then Theorems 1 through 3 are expected to hold at least asymptotically (i.e., approximately in finite samples), implying that, in practice, observed patterns in residuals computed as in Definitions 1 and 2 might constitute a mere indication of their theoretical properties instead of model misspecification. From Theorem 3, this possibility seems more likely to occur (especially for small sample sizes) when a model which has to be diagnostically checked contains a moving average part with at least one root on or near the unit circle.

**Remark 6.** The results summarized above in point (iv) with regard to the convergence of  $\text{Var}[\hat{A}_t^0]$  (from above) and  $\text{Var}[\hat{A}_t]$  (from below) to  $\sigma^2$ , are eminently sensible, since the information contained in a given time series is used more efficiently when computing the unconditional residuals than when computing the conditional residuals (the unknown pre-sample values are set to zero for the conditional computation, whereas they are back-casted from available data for the unconditional computation; see Remarks 1 and 2). Hence, the unconditional computation is expected to give more accurate results. This property clearly suggests encouraging the use of the unconditional residuals against the conditional ones in model checking and forecasting.

Both unconditional residuals and innovations arise naturally when estimating scalar ARMA models through exact maximum likelihood (EML) (i.e., through numerical maximization of (3) with respect to  $\hat{\beta}$  and  $\hat{\sigma}^2$ ). On the other hand, conditional residuals arise naturally when estimating univariate ARMA models through conditional maximum likelihood (CML) (i.e., conditional least squares), which may provide poor parameter estimates in many cases (especially for small samples and/or nearly noninvertible parameter values; see, for example, Ansley and Newbold 1980; and Box et al. 1994, ch. 7). This fact, together with the loss of information implied by the conditions imposed for computing conditional residuals, suggest again that such residuals computed after CML estimation constitute a far-from-ideal tool for model diagnostic checking.

**Remark 7.** Theoretical properties of innovations under the assumption that the true parameter values  $\mu$ ,  $\phi_1, \dots, \phi_p$ ,  $\theta_1, \dots, \theta_q$  of the stationary model (3) are known, can be found, for example, in Box et al. (1994, pp. 275-279) and the references cited therein. Specifically, it follows trivially from Definition 3 that  $\hat{\mathbf{E}} = \mathbf{L}^{-1} \tilde{\mathbf{W}} \sim (\mathbf{0}, \sigma^2 \mathbf{F})$ , which implies that  $E[\hat{E}_t] = 0$ ,  $\text{Var}[\hat{E}_t] = \sigma^2 F_t$ , and  $\text{Cov}[\hat{E}_t, \hat{E}_{t+k}] = 0$  ( $t = 1, 2, \dots, n; k \neq 0$ ). Furthermore, the elements of  $\hat{\mathbf{E}}$  and  $\mathbf{F}$  can be described in terms of several recursive algorithms applied to the state-space representation of model (3) (see, for example, Gardner et al. 1980, and Mélard 1984), which, for an invertible model, can be shown to converge to a steady state, with  $F_t$  converging to one from above and  $\hat{E}_t$  converging to  $A_t$  in mean square; additionally, for pure autoregressive models these convergence results occur exactly at time  $t = p + 1$  (refer, for example, to Harvey 1993, ch. 4 for further details). Hence, the innovation vector  $\hat{\mathbf{E}}$  shows theoretical properties which are similar to those of Theorems 1 through 3 for  $\hat{\mathbf{A}}^0$  and  $\hat{\mathbf{A}}$ , in spite of their numerical values being computed in practice through quite different procedures. In fact, it is shown in the next section that any one of these three types of residuals can be used to compute a unique set of “normalized” residuals, which may prove useful for diagnostic checking of a tentatively entertained model.



#### 4. RESIDUAL DIAGNOSTIC CHECKING IN UNIVARIATE MODELS

The theoretical properties of conditional residuals, unconditional residuals and innovations considered in Section 3, suggest that none of these three types of residuals, computed after model estimation as described in Section 2, should be expected to follow white noise patterns (not even approximately), especially for small samples and moving average parameter estimates close to noninvertibility. Note, however, that letting  $\mathbf{P}$  represent a lower-triangular matrix such that  $\Sigma_0 = \mathbf{I} + \mathbf{Z}\Omega\mathbf{Z}^T = \mathbf{P}\mathbf{P}^T$ , it follows from Theorem 1 (A), Theorem 2 (A) and Remark 7, that the three random vectors  $\mathbf{P}^{-1}\hat{\mathbf{A}}^0$ ,  $\mathbf{P}^T\hat{\mathbf{A}}$  and  $\mathbf{F}^{-1/2}\hat{\mathbf{E}}$  follow  $(\mathbf{0}, \sigma^2\mathbf{I})$  distributions. In fact,

$$\mathbf{P}^T\hat{\mathbf{A}} = \mathbf{P}^T\Sigma_0^{-1}\hat{\mathbf{A}}^0 = \mathbf{P}^T(\mathbf{P}\mathbf{P}')^{-1}\hat{\mathbf{A}}^0 = \mathbf{P}^{-1}\hat{\mathbf{A}}^0,$$

and, noting from (5) and (10) that  $\Sigma_{\mathbf{W}} = \mathbf{K}^{-1}\mathbf{P}\mathbf{P}^T\mathbf{K}^{-1T} = \mathbf{L}\mathbf{F}^{1/2}\mathbf{F}^{1/2}\mathbf{L}^T$ , it can also be seen that

$$\mathbf{F}^{-1/2}\hat{\mathbf{E}} = \mathbf{F}^{-1/2}\mathbf{L}^{-1}\tilde{\mathbf{W}} = (\mathbf{L}\mathbf{F}^{1/2})^{-1}\mathbf{K}^{-1}\hat{\mathbf{A}}^0 = (\mathbf{K}^{-1}\mathbf{P})^{-1}\mathbf{K}^{-1}\hat{\mathbf{A}}^0 = \mathbf{P}^{-1}\hat{\mathbf{A}}^0.$$

Hence,

$$\mathbf{P}^{-1}\hat{\mathbf{A}}^0 = \mathbf{P}^T\hat{\mathbf{A}} = \mathbf{F}^{-1/2}\hat{\mathbf{E}}, \quad (12)$$

implying that, under the assumption that the true parameter values of the stationary model (3) are known, a vector of white noise “normalized” residuals can be defined in any of three equivalent ways using either  $\hat{\mathbf{A}}^0$  (conditional residuals),  $\hat{\mathbf{A}}$  (unconditional residuals) or  $\hat{\mathbf{E}}$  (innovations).

Furthermore, noting from (12) that  $\hat{\mathbf{A}}^0 = \mathbf{A} - \mathbf{Z}\mathbf{U}_*$ , it follows for the normalized residual vector given, for example, in the first part of (12), that  $\mathbf{P}^{-1}\hat{\mathbf{A}}^0 - \mathbf{A} = -(\mathbf{I} - \mathbf{P}^{-1})\mathbf{A} - \mathbf{P}^{-1}\mathbf{Z}\mathbf{U}_*$ , so that

$$E[(\mathbf{P}^{-1}\hat{\mathbf{A}}^0 - \mathbf{A})(\mathbf{P}^{-1}\hat{\mathbf{A}}^0 - \mathbf{A})^T] = \sigma^2 [(\mathbf{I} - \mathbf{P}^{-1}) + (\mathbf{I} - \mathbf{P}^{-1})^T]. \quad (13)$$

Recalling that for an invertible model (see the proof of Theorem 3) the  $(i, i)$ th element of  $\Sigma_0 = \mathbf{I} + \mathbf{Z}\Omega\mathbf{Z}^T = \mathbf{P}\mathbf{P}^T$  converges to one from above as  $i$  increases, it turns out that the  $(i, i)$ th element of  $\mathbf{P}^{-1}$  (which is strictly positive for all  $i$ ) converges to one from below as  $i$  increases, so (13) implies that the elements of (12) converge in mean square to the model white noise disturbances, with exact convergence occurring at  $t = p + 1$  if  $q = 0$ . Hence, the mean-square convergence property shared by  $\hat{\mathbf{A}}^0$ ,  $\hat{\mathbf{A}}$  and  $\hat{\mathbf{E}}$ , is preserved through appropriate linear transformations as those stated in (12).

In practice, these results mean that when the true model parameter values are replaced by consistent estimates, the elements of the computed normalized residual vector

$$\hat{\mathbf{v}} = \hat{\mathbf{P}}^{-1}\hat{\mathbf{a}}^0 = \hat{\mathbf{P}}^T\hat{\mathbf{a}} = \hat{\mathbf{F}}^{-1/2}\hat{\mathbf{e}} \quad (14)$$

should (at least approximately) follow white noise patterns and converge to the model unobservable random shocks, under the hypothesis that the entertained model is adequate. In particular, the use of  $\hat{\mathbf{v}}$  in model diagnostic checking instead of  $\hat{\mathbf{a}}^0$  or  $\hat{\mathbf{a}}$ , might help to avoid a possibly incorrect interpretation of residual autocorrelation (recall Remark 5 in Section 3), whose only source in the case of the elements of  $\hat{\mathbf{v}}$  is (at least approximately and apart from outliers) model misspecification. Furthermore, note that working with  $\hat{\mathbf{v}}$  solves also the theoretical heteroskedasticity associated with all of  $\hat{\mathbf{a}}^0$ ,  $\hat{\mathbf{a}}$  and  $\hat{\mathbf{e}}$ , which is expected to be present unless the estimated sequence  $\hat{\mathbf{z}}_0, \hat{\mathbf{z}}_1, \hat{\mathbf{z}}_2, \dots$  (recall the proof of Theorem 3) converges to zero very quickly. Some additional issues on computing and using the normalized residual

vector  $\hat{\mathbf{v}}$  given in (14), are considered in the following two remarks:

**Remark 8.** From a computational standpoint, the most convenient method for obtaining the elements of  $\hat{\mathbf{v}}$  seems to be that based on the innovation vector  $\hat{\mathbf{e}}$ , since, after model estimation, such method only requires  $n$  additional square roots and divisions (as opposed to a considerably larger number of operations required for computing  $\hat{\Sigma}_0 = \mathbf{I} + \hat{\mathbf{Z}}\hat{\Omega}\hat{\mathbf{Z}}^T$ , its Cholesky factor  $\hat{\mathbf{P}}$ , and any of the two matrix-vector products  $\hat{\mathbf{P}}^{-1}\hat{\mathbf{a}}^0$  or  $\hat{\mathbf{P}}^T\hat{\mathbf{a}}$ ). This fact, together with the apparent lack in previous literature of analytical expressions for both  $\text{Var}[\hat{\mathbf{A}}^0]$  and  $\text{Var}[\hat{\mathbf{A}}]$  (as opposed to well-known results on  $\text{Var}[\hat{\mathbf{E}}]$ ), might explain why computer programs for estimation of ARMA models based on the innovations approach (e.g., computer programs based on state-space methods) usually provide the practitioner with both the computed innovations  $\hat{e}_t$  and the corresponding normalized residuals  $\hat{v}_t = \hat{F}_t^{-1/2}\hat{e}_t$  (see, for example, Brockwell and Davis 2002, pp. 164-167), whereas computer programs based on other approaches usually give only the conditional residuals or the unconditional residuals. Note, however, that the calculations required for obtaining  $\hat{\mathbf{P}}^{-1}\hat{\mathbf{a}}^0$  or  $\hat{\mathbf{P}}^T\hat{\mathbf{a}}$  need to be carried out only after model estimation (i.e., they are not explicitly required for estimation purposes), and that, indeed, such calculations involve a negligible amount of computing time for most modern computers.

**Remark 9.** Ansley and Newbold (1979, pp. 551-553) have demonstrated through simulation experiments that the use of the normalized residual vector  $\hat{\mathbf{v}}$  instead of the unconditional residual vector  $\hat{\mathbf{a}}$  (especially for seasonal models), extends the range of cases for which statistics frequently used in model diagnostic checking (e.g., Ljung and Box 1978) can be usefully interpreted through the usual asymptotic significance levels. However, these authors suggest (i) a single way of computing  $\hat{\mathbf{v}}$  (Ansley 1979), and (ii) that the only reason for the superior sampling properties of tests based on  $\hat{\mathbf{v}}$  is that unconditional residuals can, in moderate sample sizes, have variance much smaller than  $\sigma^2$  (recall (11) and (9) above), whereas normalized residuals have the same variance as the model random shocks. Taking into account the results from Sections 3 and 4 of the present article, the conclusions derived by Ansley and Newbold (1979) can be expanded as follows: (i) the residual vector  $\hat{\mathbf{v}}$  can be computed in any of several equivalent ways (apart from the usual innovations approach), and (ii) the practical benefits from using  $\hat{\mathbf{v}}$  instead of  $\hat{\mathbf{a}}$  stem from the fact that (11) is not a diagonal matrix, so not only the unconditional residuals have variance smaller than  $\sigma^2$ , but (more importantly) they are also autocorrelated. Note that point (i) above is especially relevant in practice, since the computation of  $\hat{\mathbf{v}}$  through  $\hat{\mathbf{a}}$  and  $\hat{\mathbf{a}}^0$  is now possible by using the output generated by any computer program for estimation of ARMA models outside the innovations framework. As a final comment, it may be noted that the simulation results reported by Ansley and Newbold (1979) are so detailed that no additional simulation evidence seems to be required to demonstrate the expected practical benefits from using  $\hat{\mathbf{v}}$  instead of  $\hat{\mathbf{a}}$  (or  $\hat{\mathbf{a}}^0$  for that matter) in model diagnostic checking, especially when the model considered contains moving average roots on or near the unit circle and/or when the sample size  $n$  is small. Practical examples illustrating this point are considered in the next section.

## 5. PRACTICAL EXAMPLES

The examples developed in this section are focused on the specific point that unconditional residuals

showing significant autocorrelation patterns do not necessarily imply model misspecification. Similar examples might be developed for conditional residuals. However (recall Remark 6 above), the properties of such residuals, as well as their link to conditional maximum likelihood or least squares estimation, strongly suggest avoiding using them for model checking in practice. Hence, examples on conditional residual are not considered below.

### 5.1. Expected Patterns in Residual Sample Autocorrelations

In the first part of this section, we briefly describe which patterns are expected to arise in the sample autocorrelations of unconditional residuals for correctly specified models. In order to describe such patterns, we have carried out a comprehensive set of exercises consisting of the following operations: (i) set up several ARMA models for unspecified time series of various lengths, (ii) compute the theoretical residual covariance matrices for such models (which depend just on the model parameters, but not on any specific time series), and (iii) evaluate the expected sample ACF of the corresponding model residuals. In strict agreement with both the theoretical results of Section 3 and the simulation results of Ansley and Newbold (1979) (see Remark 9 above), we have found that unconditional residuals showing clear autocorrelation patterns are expected to arise in several cases, especially for models containing moving average roots on or near the unit circle.

To illustrate, consider the extremely popular  $MA(1) \times MA(1)_{12}$  model for a stationary, zero-mean process  $\{W_t\}$  which is observed on a monthly basis,

$$W_t = (1 - \theta_1 B)(1 - \Theta_1 B^{12})A_t,$$

where  $\{A_t\}$  is a white noise process with unit variance. For this multiplicative model, and for several values of  $\theta_1$ ,  $\Theta_1$  and  $n$  (the number of observations), we have computed the theoretical covariance matrix  $(\mathbf{I} + \mathbf{Z}\mathbf{\Omega}\mathbf{Z}^T)^{-1}$  of the unconditional residuals (see Theorem 2), along with the quantities

$$\rho_j = \frac{1}{n-j} \sum_{i=1}^{n-j} \delta_{i,i+j} \quad \text{for } j = 1, 2, \dots,$$

where  $\delta_{i,j}$  represents the  $(i, j)$ th element of  $(\mathbf{I} + \mathbf{Z}\mathbf{\Omega}\mathbf{Z}^T)^{-1}$ . The quantity  $\rho_j$  is an expected value of the  $j$ th order sample autocorrelation of the unconditional residuals. Hence, if the sequence  $\rho_1, \rho_2, \dots$  exhibits any clear pattern, then such pattern is expected to arise as well in the sample ACF of the computed unconditional residuals for the corresponding estimated model (as long as the model is adequate and irrespective of the specific time series that might have been used to estimate it). In this respect, Fig. 1 contains eight plots of the sequence  $\rho_1, \rho_2, \dots, \rho_{39}$  associated with eight different instances of the multiplicative  $MA(1) \times MA(1)_{12}$  model.

From Fig. 1, it can be seen that the most significant features of the expected ACF happen at the annual lags (12, 24 and 36); additionally, these features are more significant when the  $MA(1)_{12}$  operator is close to being noninvertible. The influence of the regular  $MA(1)$  operator on the appearance of the expected ACF is much less significant. Finally, the patterns in the expected ACF tend to diminish as the sample size increases. Hence, as a general rule, one should expect to find significant patterns in the

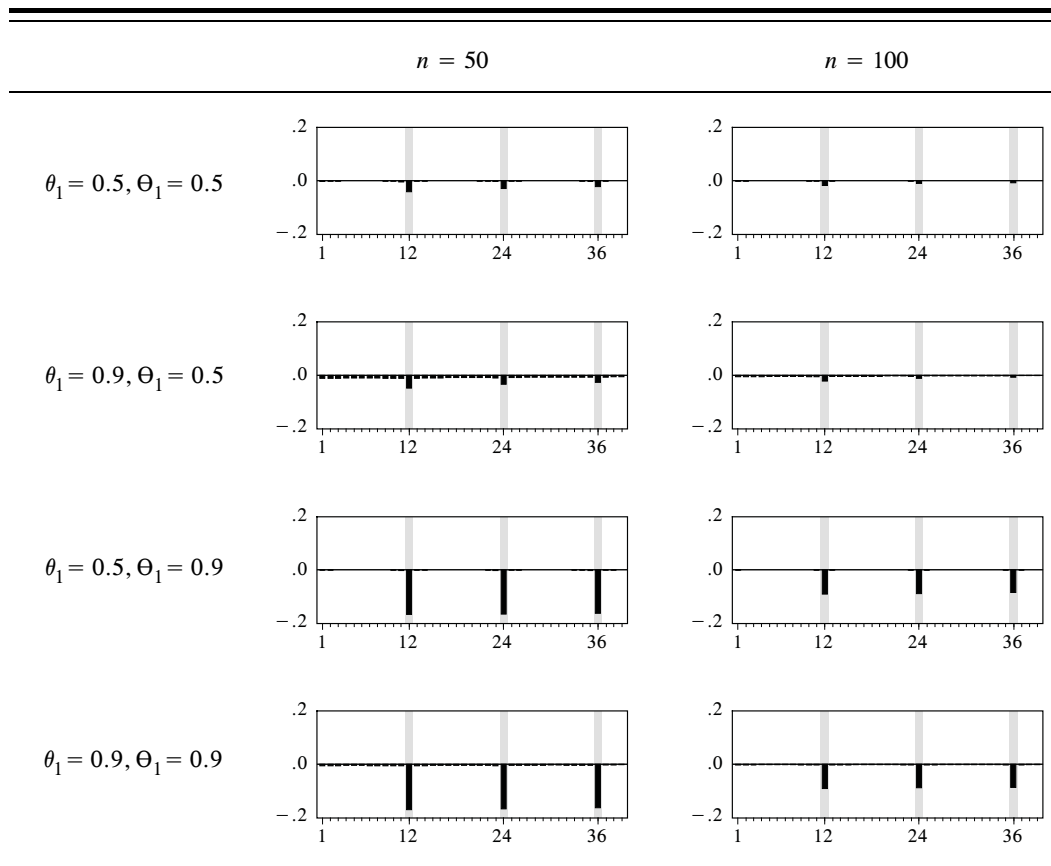


Fig. 1. Expected Sample Autocorrelations for the Unconditional Residuals Associated with the Multiplicative  $MA(1) \times MA(1)_{12}$  Model.

sample ACF of the unconditional residuals computed for any model with an estimated  $MA(1)_{12}$  term close to being noninvertible. This is an important result from a practical point of view, since many seasonal time series found in practice seem to require models with such terms to be well described. The second part of this section gives an example on this point using actual data.

For some other types of models (especially for pure autoregressive models), the results of computations analogous to those described for the  $MA(1) \times MA(1)_{12}$  model are not so significant. Nonetheless, such results do confirm without ambiguities both the theoretical results of Section 3 and the simulation results of Ansley and Newbold (1979), namely that autocorrelation patterns are expected to be present (with varying degrees of intensity) in the unconditional residuals for any estimated ARMA model.

## 5.2. An Example with Actual Data

All of the results presented so far suggest that using the normalized residuals for model checking, instead of the conditional or the unconditional residuals, might help to avoid a possibly incorrect interpretation of residual autocorrelation. The example developed below shows that unconditional residuals with significant autocorrelation patterns do not necessarily imply model misspecification.

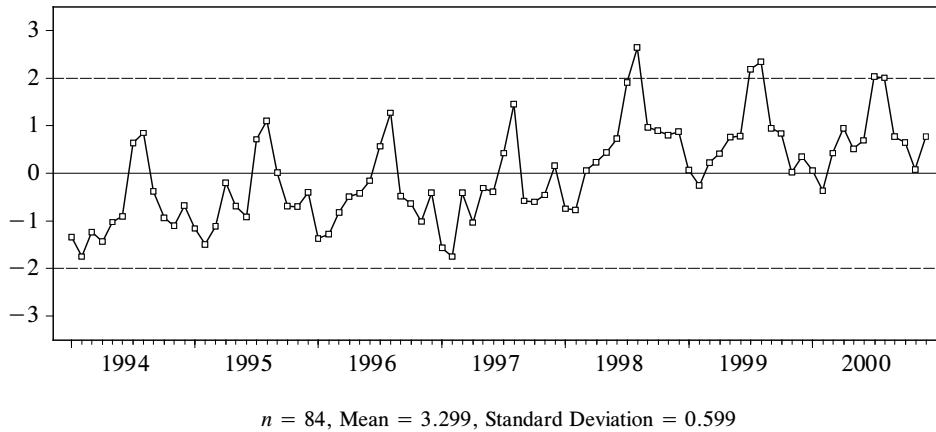


Fig. 2. Original Series:  $y_t$  (Traffic Accidents with Victims on Spain Roads, in thousands).

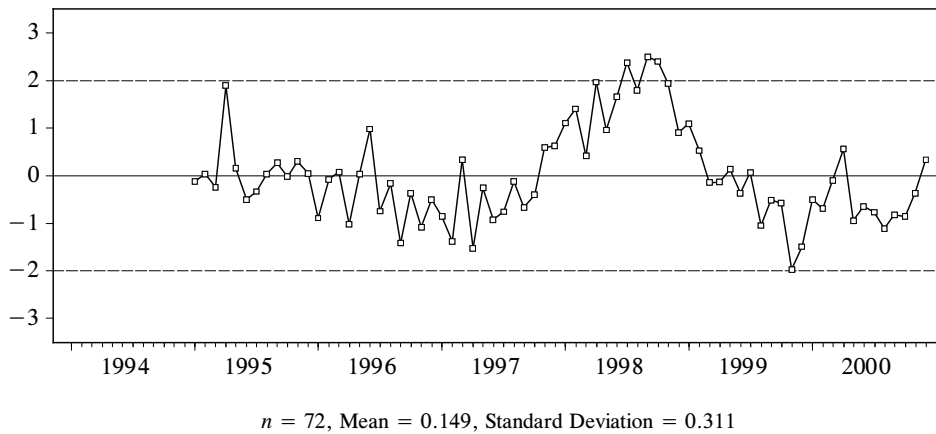


Fig. 3. Seasonally Differenced Series:  $\nabla_{12} y_t = (1 - B^{12})y_t$ .

Table 1. Unit Root Tests for the Series  $\nabla_{12} y_t$  Displayed in Fig. 3.

ADF-ERS Test [†] ( $H_0$ : Series has a unit root)	PP Test [†] ( $H_0$ : Series has a unit root)	KPSS Test [†] ( $H_0$ : Series is stationary)
$1\% < p\text{-value} < 5\%$	$p\text{-value} < 1\%$	$p\text{-value} > 10\%$

[†] ADF-ERS: Elliot-Rothenberg-Stock modification of Augmented Dickey-Fuller, PP: Phillips-Perron, KPSS: Kwiatkowski-Phillips-Schmidt-Shin. See Q.M.S. (2002, pp. 329-337) for the details. In all cases, the auxiliary regressions were run with an intercept and without a linear trend.

We consider the monthly series Traffic Accidents with Victims on Spain Roads, obtained from <http://www.ine.es> (Instituto Nacional de Estadística, Spain) on June 16, 2006. We use data for the period January 1994 through December 2000 for modelling purposes, and hold back data for January 2001 through December 2003 for forecasting evaluation. All of the time series plots are standardized for easier comparison.

The original data shown in Fig. 2 present a clear seasonal pattern, as well as a possible permanent

level shift around the end of 1997. Hence, we consider the seasonally differenced series shown in Fig. 3, which reflects two notable intervention events: a possible step occurring on November 1997, due to policy measures aimed at increasing new car registrations in Spain (the so-called “Plan Renove”), as well as a clear impulse on April 1995. That step event, along with the results displayed in Table 1 on unit root tests on the seasonally differenced series of Fig. 3, suggest that such series is possibly stationary. Hence, we estimate through EML the intervention model shown in (15) below, which includes the step,  $\xi_t^{S,1997:11}$ , and the impulse,  $\xi_t^{I,1995:04}$ , mentioned previously, as well as two seasonal ARMA terms: a second-order autoregressive term with its first coefficient set to zero (because it was clearly insignificant in a previous estimation run), and a first-order moving average term which turns out to be estimated noninvertible, suggesting deterministic seasonal patterns in the original series:

$$y_t = \frac{0.197}{(0.037)} \frac{\xi_t^{S,1997:11}}{1 - 0.704 B} + 0.326 \frac{\xi_t^{I,1995:04}}{(0.130)} + \hat{\varepsilon}_t, \quad (15)$$

$$(1 + 0.426 B^{24}) (\nabla_{12} \hat{\varepsilon}_t - 0.032) = (1 - 0.999 B^{12}) \hat{a}_t, \quad (15)$$

$$(0.131) \quad (0.017) \quad (0.117)$$

$$n = 72, \hat{\sigma} = 0.122, l_* = 31.835.$$

The standard errors (in parentheses) indicate that the model parameters are clearly significant, and the estimated standard deviation of the model disturbances ( $\hat{\sigma}$ ) as well as the logarithm of the exact log-likelihood ( $l_*$ ) suggest an adequate fit.

The unconditional residuals for model (15) are shown in Fig. 4, and they seem to be clearly stationary. However, their simple (ACF) and partial (PACF) sample autocorrelations indicate that they are significantly autocorrelated (especially at the seasonal lags). In this respect, the usual Ljung-Box statistic rejects the estimated model, even at the 1% significance level.

The picture changes notably when we look at the normalized residuals instead. The normalized residuals for model (15) are shown in Fig. 5. These residuals remain plausibly stationary and, as opposed to the unconditional residuals, they do not show any significant autocorrelation, even at the 20% significance level.

Hence, on the basis of the normalized residuals of Fig. 5, the intervention model (15) cannot be rejected, although it would have been discarded if the unconditional residuals of Fig. 4 had been used for model checking.

For the purpose of forecasting comparison, we have built an additional model for the time series displayed in Fig. 2. To this end, we apply an additional regular difference to the previously seasonally differenced series, and obtain the clearly stationary series shown in Fig. 6. We consider the same intervention events as before, and estimate through EML the intervention model shown in (16) below, which includes the step and the impulse considered previously, as well as three ARMA terms: the same two seasonal terms as in model (15) (although the moving average now seems invertible), plus an additional regular first-order moving average term. The values of both  $\hat{\sigma}$  and  $l_*$  suggest a slightly worse fit for this model than in the case of model (15).

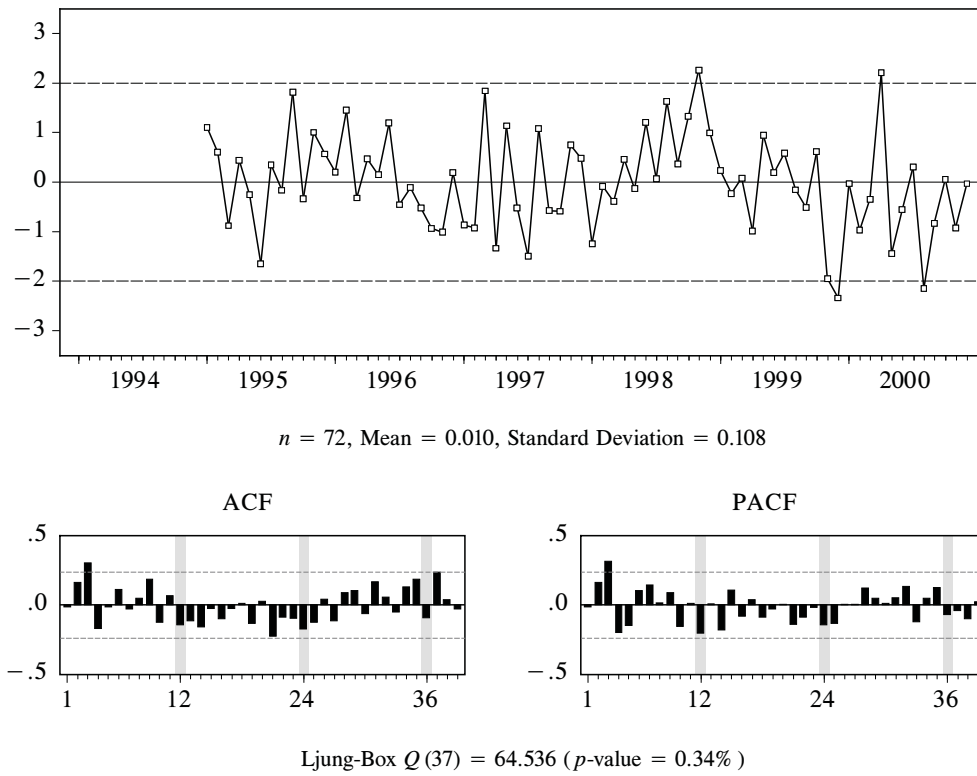


Fig. 4. Unconditional Residuals for Estimated Intervention Model (15).

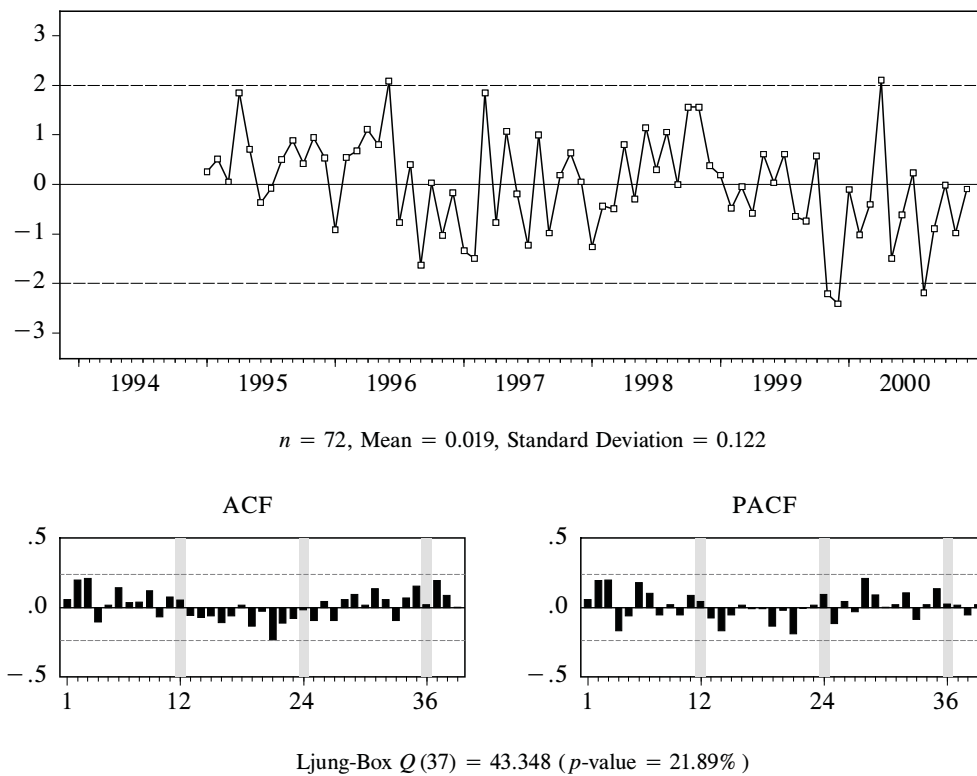


Fig. 5. Normalized Residuals for Estimated Intervention Model (15).

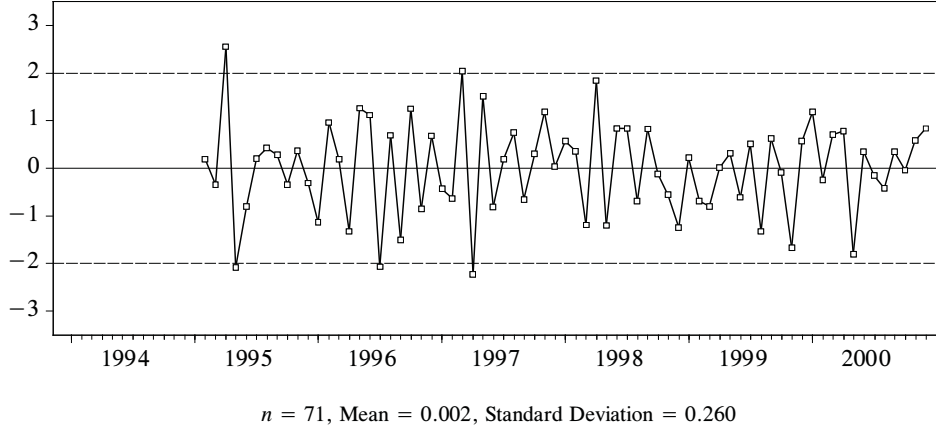


Fig. 6. Regularly and Seasonally Differenced Series:  $\nabla\nabla_{12} y_t = (1 - B)(1 - B^{12})y_t$ .

$$y_t = \frac{0.188}{(0.037)} \frac{\xi_t^{S,1997:11}}{1 - 0.766 B} + \frac{0.375}{(0.136)} \xi_t^{I,1995:04} + \hat{\varepsilon}_t,$$

(0.054)

$$(1 + \frac{0.378 B^{24}}{(0.142)}) \nabla\nabla_{12} \hat{\varepsilon}_t = (1 - \frac{0.830 B}{(0.066)})(1 - \frac{0.619 B^{12}}{(0.237)}) \hat{a}_t, \quad (16)$$

$$n = 71, \hat{\sigma} = 0.142, l_* = 30.827.$$

For this second model, the unconditional residuals of Fig. 7 are clearly stationary, and do not show any significant autocorrelation. Hence, as opposed to model (15), model (16) would not be discarded on the basis of the unconditional residuals. This is also the case for model (16) if the normalized residuals of Fig. 8 are used instead: the normalized residuals are also clearly stationary and do not show any significant autocorrelation either. In summary, as opposed to the case of model (15), on the basis of both types of residuals model (16) cannot be rejected.

To conclude this example, we have used both models (15) and (16) for computing out-of-sample forecasts for the period January 2001 through December 2003, for a total of 36 consecutive months. Fig. 9 represents the point forecasts (as well as the 95% confidence intervals) obtained with each model, along with the actually observed data for the time period considered. Looking at both the plots and the forecast accuracy measures displayed in Fig. 9, it can be seen that model (15) forecasts better than model (16) for the time period considered, most notably for the last of the three forecasting years. Recall now that model (15) (the one providing better forecasts) would have been discarded if the unconditional residuals of Fig. 4 had been used for diagnostic checking, whereas it was clearly not rejected by the normalized residuals of Fig. 5. Model (16) (the one providing worse forecasts) was not rejected in any case.

In summary, this example has shown that using unconditional residuals for model checking may lead to discarding adequate models (which may lead in turn to discarding useful forecasts), and that normalized residuals may help in solving this type of problem. This can be especially relevant within the context of automatic model building and forecasting systems, where user involvement and inspection are usually low.



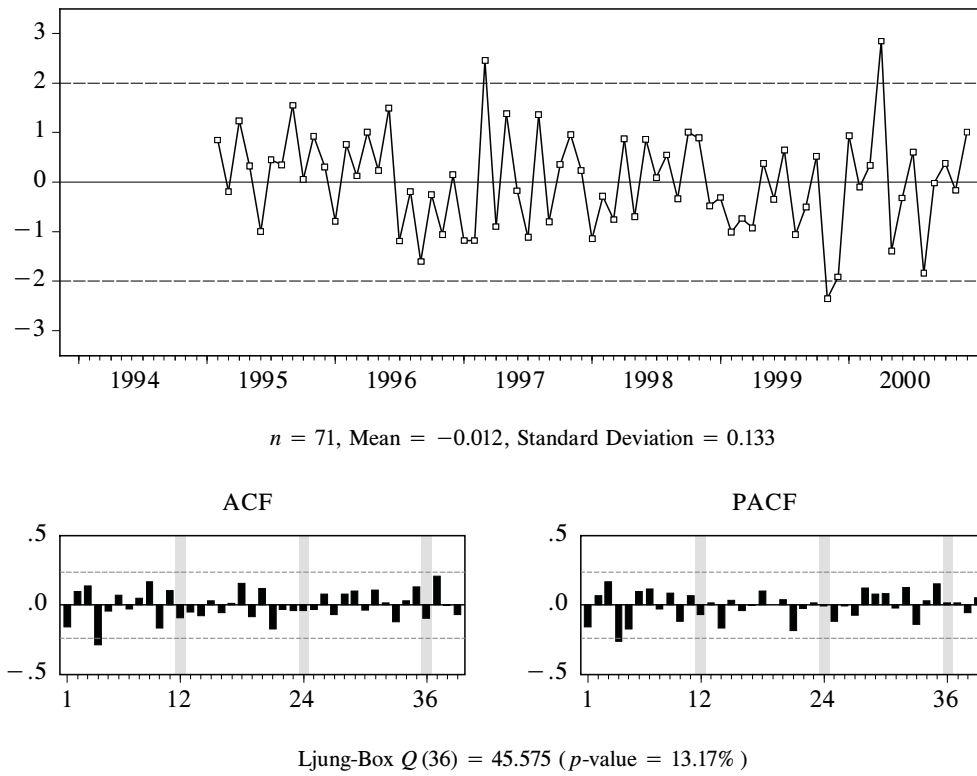


Fig. 7. Unconditional Residuals for Estimated Intervention Model (16).

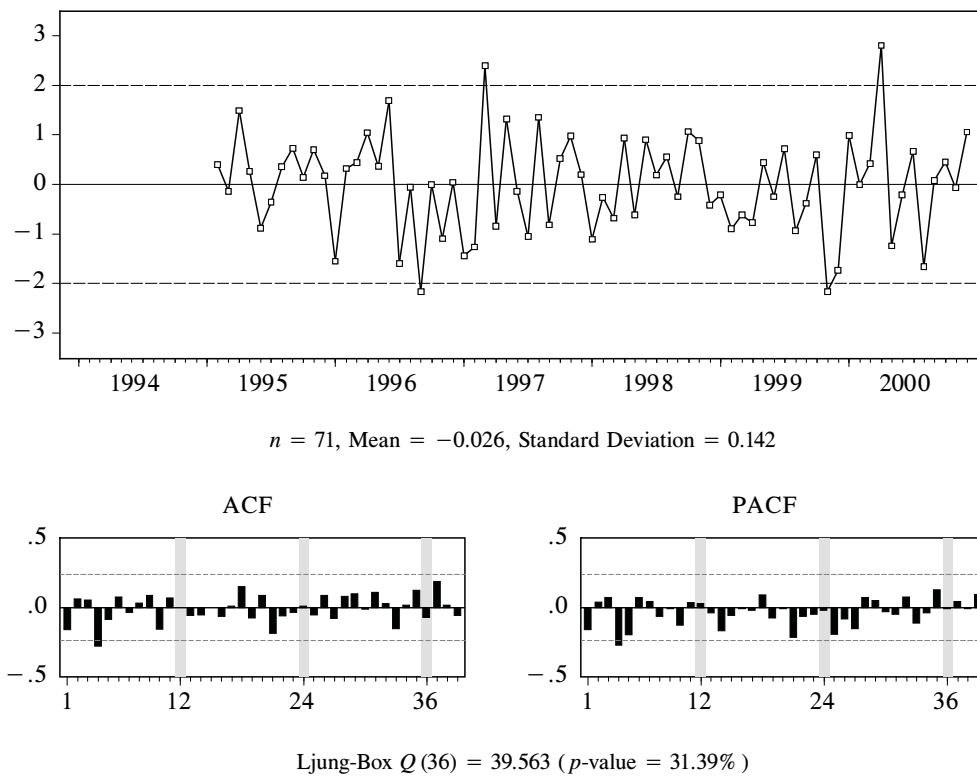


Fig. 8. Normalized Residuals for Estimated Intervention Model (16).

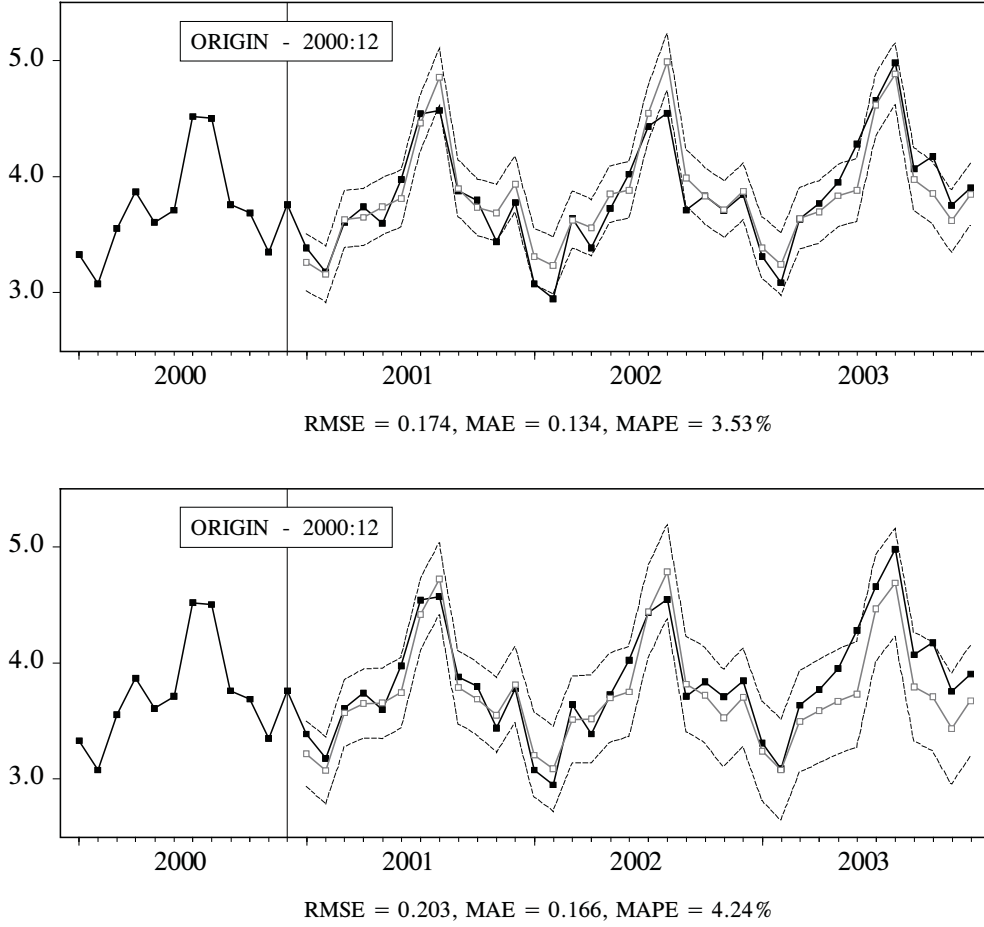


Fig. 9. Forecasts Obtained with Models (15) (upper panel) and (16) (lower panel). Actually observed data are represented with a darker color (starting one year before the forecasting origin), and forecasts are represented with a lighter color. 95% confidence bands are represented with dashed lines. Forecast accuracy measures are reported as RMSE (root mean squared error), MAE (mean absolute error) and MAPE (mean absolute percentage error).

### 6. RESIDUALS IN MULTIVARIATE MODELS

This section gives some details on extending the results of the previous sections to the case of stationary vector ARMA (VARMA) models. It may be noted that, with minor modifications, the results presented below hold also for any time series model which can be cast into a standard, stationary VARMA model, including, among many others, transfer function-noise models (Mauricio 1996) and partially nonstationary multivariate models with reduced rank structure for cointegrated processes (Mauricio 2006a).

Let an observed multiple time series  $\mathbf{w}_t$  ( $t = 1, 2, \dots, n$ ), with  $\mathbf{w}_t = [w_{t1}, \dots, w_{tm}]^T$  ( $m \geq 2$ ), be generated by a stationary vector time series process  $\{\mathbf{W}_t\}$ , with  $\mathbf{W}_t = [W_{t1}, \dots, W_{tm}]^T$ , following the VARMA model

$$\Phi(B)\tilde{\mathbf{W}}_t = \Theta(B)\mathbf{A}_t, \tag{17}$$

where  $\Phi(B) = \mathbf{I}_m - \sum_{i=1}^p \Phi_i B^i$  and  $\Theta(B) = \mathbf{I}_m - \sum_{i=1}^q \Theta_i B^i$  are matrix polynomials in the backshift

operator  $B$  of degrees  $p$  and  $q$ ,  $\Phi_1, \dots, \Phi_p$  and  $\Theta_1, \dots, \Theta_q$  are  $m \times m$  parameter matrices,  $\tilde{\mathbf{W}}_t = \mathbf{W}_t - E[\mathbf{W}_t]$ , and  $\{\mathbf{A}_t\}$  is a sequence of IID( $\mathbf{0}, \Sigma$ )  $m \times 1$  random vectors, with  $\Sigma$  ( $m \times m$ ) symmetric and positive definite. For stationarity it is required that the roots of  $|\Phi(x)| = 0$  lie outside the unit circle. A similar condition on  $|\Theta(x)|$  ensures that the model is invertible. Additional conditions on  $\Phi(B)$  and  $\Theta(B)$  for parameter uniqueness, such as those considered by Reinsel (1997, pp. 37-40), are also assumed.

Let  $\tilde{\mathbf{W}} = [\tilde{\mathbf{W}}_1^T, \dots, \tilde{\mathbf{W}}_n^T]^T$  ( $nm \times 1$ ),  $\mathbf{A} = [\mathbf{A}_1^T, \dots, \mathbf{A}_n^T]^T$  ( $nm \times 1$ ), and  $\mathbf{U}_* = [\tilde{\mathbf{W}}_{1-p}^T, \dots, \tilde{\mathbf{W}}_0^T, \mathbf{A}_{1-q}^T, \dots, \mathbf{A}_0^T]^T$   $[(p+q)m \times 1]$ , and consider Eq. (17) for  $t = 1, 2, \dots, n$ . Then, the multiple time series  $\mathbf{w} = [\mathbf{w}_1^T, \dots, \mathbf{w}_n^T]^T$  can be regarded as a particular realization of a random vector  $\mathbf{W} = [\mathbf{W}_1^T, \dots, \mathbf{W}_n^T]^T$  following the model

$$\mathbf{D}_\Phi \tilde{\mathbf{W}} = \mathbf{D}_\Theta \mathbf{A} + \mathbf{V} \mathbf{U}_*, \quad (18)$$

where  $\mathbf{D}_\Phi$  and  $\mathbf{D}_\Theta$  are  $nm \times nm$  parameter block-matrices with identity matrices of order  $m$  on the main diagonal and  $-\Phi_j$  and  $-\Theta_j$ , respectively, down the  $j$ th subdiagonal, and  $\mathbf{V}$  is an  $nm \times (p+q)m$  block-matrix with  $\mathbf{V}_{ij} = \Phi_{p+i-j}$  ( $i = 1, \dots, p; j = i, \dots, p$ ),  $\mathbf{V}_{ij} = -\Theta_{q+i-j+p}$  ( $i = 1, \dots, q; j = p+i, \dots, p+q$ ), and zeros elsewhere. Noting (18), it can be seen that

$$\begin{aligned} E[\tilde{\mathbf{W}}\tilde{\mathbf{W}}^T] &= \mathbf{D}_\Phi^{-1} \left[ \mathbf{D}_\Theta (\mathbf{I} \otimes \Sigma) \mathbf{D}_\Theta^T + \mathbf{V} \Omega \mathbf{V}^T \right] \mathbf{D}_\Phi^{-1T} \\ &= \mathbf{K}^{-1} \left[ (\mathbf{I} \otimes \Sigma) + \mathbf{Z} \Omega \mathbf{Z}^T \right] \mathbf{K}^{-1T}, \end{aligned}$$

where  $\mathbf{K} = \mathbf{D}_\Theta^{-1} \mathbf{D}_\Phi$ ,  $\mathbf{Z} = -\mathbf{D}_\Theta^{-1} \mathbf{V}$  and  $\Omega = E[\mathbf{U}_* \mathbf{U}_*^T]$  are parameter matrices of dimensions  $nm \times nm$ ,  $nm \times (p+q)m$  and  $(p+q)m \times (p+q)m$ , respectively, with  $\Omega$  being readily expressible in terms of  $\Phi_1, \dots, \Phi_p, \Theta_1, \dots, \Theta_q$  and  $\Sigma$  as described, for example, in Reinsel (1997, pp. 59-60).

Using hats ( $\hat{\cdot}$ ) to represent estimates and letting  $\tilde{\mathbf{w}} = [\tilde{\mathbf{w}}_1^T, \dots, \tilde{\mathbf{w}}_n^T]^T$ , with  $\tilde{\mathbf{w}}_t = \mathbf{w}_t - \hat{\boldsymbol{\mu}}$  and  $\hat{\boldsymbol{\mu}} = \hat{E}[\mathbf{W}_t]$  ( $t = 1, \dots, n$ ), the three types of residuals considered in Sections 2 and 3 can be defined for model (18) as follows:

**Definition M1.** The conditional residuals associated with model (18) are the  $n$  block-elements of the  $nm \times 1$  vector  $\hat{\mathbf{a}}^0 = \hat{\mathbf{K}} \tilde{\mathbf{w}}$ . The elements of  $\hat{\mathbf{a}}^0 = [\hat{\mathbf{a}}_1^{0T}, \hat{\mathbf{a}}_2^{0T}, \dots, \hat{\mathbf{a}}_n^{0T}]^T$  can be computed recursively for a given set of parameter estimates as

$$\hat{\mathbf{a}}_t^0 = \tilde{\mathbf{w}}_t - \sum_{j=1}^p \Phi_j \tilde{\mathbf{w}}_{t-j} + \sum_{j=1}^q \Theta_j \hat{\mathbf{a}}_{t-j}^0 \quad (t = 1, 2, \dots, n),$$

with every  $\tilde{\mathbf{w}}_i$  and  $\hat{\mathbf{a}}_i^0$  for  $i < 1$  replaced by  $\mathbf{0}$  (i.e., with every pre-sample quantity set to zero).  $\square$

**Definition M2.** The unconditional residuals associated with model (18) are the  $n$  block-elements of the  $nm \times 1$  vector  $\hat{\mathbf{a}} = (\mathbf{I} \otimes \hat{\Sigma}) \hat{\Sigma}_0^{-1} \hat{\mathbf{a}}^0$ , where  $\hat{\Sigma}_0$  is an estimate of  $\Sigma_0 = (\mathbf{I} \otimes \Sigma) + \mathbf{Z} \Omega \mathbf{Z}^T$ . The elements of  $\hat{\mathbf{a}} = [\hat{\mathbf{a}}_1^T, \hat{\mathbf{a}}_2^T, \dots, \hat{\mathbf{a}}_n^T]^T$  can be computed recursively for a given set of parameter estimates as

$$\hat{\mathbf{a}}_t = \tilde{\mathbf{w}}_t - \sum_{j=1}^p \Phi_j \tilde{\mathbf{w}}_{t-j} + \sum_{j=1}^q \Theta_j \hat{\mathbf{a}}_{t-j} \quad (t = 1, 2, \dots, n),$$

with  $\tilde{\mathbf{w}}_i$  ( $i = 0, \dots, 1-p$ ) and  $\hat{\mathbf{a}}_i$  ( $i = 0, \dots, 1-q$ ) replaced by the corresponding elements in the back-cast vector  $\hat{\mathbf{u}}_* = -[\hat{\Omega}^{-1} + \hat{\mathbf{Z}}^T (\mathbf{I} \otimes \hat{\Sigma}^{-1}) \hat{\mathbf{Z}}]^{-1} \hat{\mathbf{Z}}^T (\mathbf{I} \otimes \hat{\Sigma}^{-1}) \hat{\mathbf{a}}^0$ . Refer to Mauricio (1995) and Reinsel (1997, Sec. 5.3.1) for further details.  $\square$

**Definition M3.** The innovations associated with model (18) are the  $n$  block-elements of the  $nm \times 1$  vector  $\hat{\mathbf{e}} = \hat{\mathbf{L}}^{-1}\tilde{\mathbf{w}} = (\hat{\mathbf{K}}\hat{\mathbf{L}})^{-1}\hat{\mathbf{a}}^0$ , where  $\hat{\mathbf{L}}$  is an estimate of the  $nm \times nm$  unit lower-triangular block matrix  $\mathbf{L}$  in the factorization  $E[\tilde{\mathbf{W}}\tilde{\mathbf{W}}^T] = \mathbf{L}\mathbf{F}\mathbf{L}^T$ , with  $\mathbf{F}$  being a block-diagonal  $nm \times nm$  matrix with symmetric, positive definite diagonal blocks  $\mathbf{F}_t$  ( $t = 1, 2, \dots, n$ ). The elements of  $\hat{\mathbf{e}}$  (along with the corresponding estimates  $\hat{\mathbf{F}}_t$ ) are frequently computed through recursive algorithms of the Chandrasekhar type (see, for example, Shea 1989), although this is rarely the most efficient option (see Mauricio 2002).  $\square$

The results considered in Sections 3 and 4 can be easily extended to the case of stationary VARMA models. Theorems M1 through M4 below summarize the main results. Proofs are omitted because they are essentially identical to the ones provided in Sections 3 and 4 for the univariate case.

**Theorem M1.** Let  $\hat{\mathbf{A}}^0 = \mathbf{K}\tilde{\mathbf{W}}$  be the random vector associated with the conditional residuals given in Definition M1, under the assumption that the true parameter values of model (18) are known. Then,  $\hat{\mathbf{A}}^0 \sim [\mathbf{0}, (\mathbf{I} \otimes \Sigma) + \mathbf{Z}\Omega\mathbf{Z}^T]$  and  $E[(\hat{\mathbf{A}}^0 - \mathbf{A})(\hat{\mathbf{A}}^0 - \mathbf{A})^T] = \mathbf{Z}\Omega\mathbf{Z}^T$ .  $\square$

**Theorem M2.** Let  $\hat{\mathbf{A}} = (\mathbf{I} \otimes \Sigma)\Sigma_0^{-1}\hat{\mathbf{A}}^0$  be the random vector associated with the unconditional residuals given in Definition M2, under the assumption that the true parameter values of the stationary model (18) are known. Then,  $\hat{\mathbf{A}} \sim [\mathbf{0}, (\mathbf{I} \otimes \Sigma)\Sigma_0^{-1}(\mathbf{I} \otimes \Sigma)]$ , where  $\Sigma_0$  is given in Definition M2, and  $E[(\hat{\mathbf{A}} - \mathbf{A})(\hat{\mathbf{A}} - \mathbf{A})^T] = \mathbf{Z}[\Omega^{-1} + \mathbf{Z}^T(\mathbf{I} \otimes \Sigma^{-1})\mathbf{Z}]^{-1}\mathbf{Z}^T$ .  $\square$

**Theorem M3.** Under the assumptions of Theorems M1 and M2, invertibility of the stationary VARMA model (18) implies additionally that:

- (A)  $E[(\hat{\mathbf{A}}_i^0 - \mathbf{A}_i)(\hat{\mathbf{A}}_i^0 - \mathbf{A}_i)^T] \rightarrow \mathbf{0}$ ,  $E[\hat{\mathbf{A}}_i^0\hat{\mathbf{A}}_i^{0T}] \rightarrow \Sigma$ ,  $E[\hat{\mathbf{A}}_i^0\hat{\mathbf{A}}_j^{0T}] \rightarrow \mathbf{0}$  ( $i \neq j$ ) for increasing  $i$  and fixed  $j$  (at an either small or large value);
- (B)  $E[(\hat{\mathbf{A}}_i - \mathbf{A}_i)(\hat{\mathbf{A}}_i - \mathbf{A}_i)^T] \rightarrow \mathbf{0}$ ,  $E[\hat{\mathbf{A}}_i\hat{\mathbf{A}}_i^T] \rightarrow \Sigma$ ,  $E[\hat{\mathbf{A}}_i\hat{\mathbf{A}}_j^T] \rightarrow \mathbf{0}$  ( $i \neq j$ ) for increasing  $i$  and fixed  $j$  (at an either small or large value); and
- (C) (A) and (B) hold exactly for  $i \geq p + 1$  and  $j \geq 1$  when  $q = 0$ .  $\square$

Theoretical properties of innovations under the assumption that the true parameter values of model (18) are known, can be found, for example, in Reinsel (1997, pp. 231-232) and the references cited therein. Specifically, it follows from Definition M3 that  $\hat{\mathbf{E}} = \mathbf{L}^{-1}\tilde{\mathbf{W}} \sim (\mathbf{0}, \mathbf{F})$ . Additionally, the elements of  $\hat{\mathbf{E}}$  and  $\mathbf{F}$  can be described in terms of a recursive algorithm of the Chandrasekhar type (see, for example, Shea 1989), which, for an invertible model, can be shown to converge to a steady state, with  $\mathbf{F}_t$  converging to  $\Sigma$  and  $\hat{\mathbf{E}}_t$  converging to  $\mathbf{A}_t$  in mean square. Furthermore, for pure autoregressive models these convergence results occur exactly at time  $t = p + 1$ .

**Theorem M4.** Let  $\hat{\mathbf{A}}^0 = \mathbf{K}\tilde{\mathbf{W}}$ ,  $\hat{\mathbf{A}} = (\mathbf{I} \otimes \Sigma)\Sigma_0^{-1}\hat{\mathbf{A}}^0$  and  $\hat{\mathbf{E}} = \mathbf{L}^{-1}\tilde{\mathbf{W}}$  be the random vectors associated with the three types of residuals given in Definitions M1, M2 and M3, respectively, under the assumption that the true parameter values of the stationary model (18) are known. Let  $\mathbf{P}$  represent a lower-triangular matrix such that  $\Sigma_0 = (\mathbf{I} \otimes \Sigma) + \mathbf{Z}\Omega\mathbf{Z}^T = \mathbf{P}\mathbf{P}^T$ . Then, there exists a normalized residual vector  $\hat{\mathbf{V}} = \mathbf{P}^{-1}\hat{\mathbf{A}}^0 = \mathbf{P}^T(\mathbf{I} \otimes \Sigma^{-1})\hat{\mathbf{A}} = \mathbf{F}^{-1/2}\hat{\mathbf{E}}$ , with the property that  $\hat{\mathbf{V}} \sim (\mathbf{0}, \mathbf{I})$ . Additionally, invertibility of model (18) implies that the elements of  $\hat{\mathbf{V}}$  converge in mean square to the model disturbances, with exact convergence occurring at time  $t = p + 1$  when  $q = 0$ .  $\square$

As in the case of univariate models, the use of the normalized residuals in model diagnostic checking instead of the conditional or the unconditional residuals may help to avoid a possibly incorrect

interpretation of residual autocorrelation and cross-correlation, whose only source in the case of the normalized residuals is (at least approximately and apart from outliers) model misspecification.

## 7. CONCLUSIONS

Both the theoretical developments and the examples on computing and using residuals in time series models presented in this article have shown the following important points:

1. It is possible to characterize accurately the numerical and statistical properties of both conditional and unconditional residuals, which, in spite of having received very limited attention in the previous literature, are frequently used in practice when analyzing time series data.
2. In particular, both conditional and unconditional residuals follow approximate distributions with covariance matrices for which easy-to-compute expressions have been given for the first time.
3. Invertibility plays a key role for establishing statistical convergence of residuals to white noise, so that residual autocorrelation should be interpreted carefully when dealing with models with parameter values on or near the boundary of the invertibility region.
4. A set of uncorrelated and homoscedastic residuals can be obtained in any of several equivalent ways. Using this set of residuals for model checking can improve the chances of not discarding a tentative model with autocorrelated unconditional residuals when the model is adequately specified.

Routine and uncared-for application of standard residual-checking procedures is not uncommon in current time series analysis, mainly due to the widespread availability of “user-friendly” and “automatic-modelling” software packages implementing them. It is true that such procedures may work sufficiently well in many cases. However, it is also true that a slight additional effort on the applied researcher’s part, required by a few simple residual computations for model diagnostic checking, may give clear benefits in cases of practical interest which are not difficult to come across.

## Acknowledgements

I am grateful to the organizers and participants at the 17th COMPSTAT Symposium of ERS-IASC (Rome, 28 August - 1 September 2006) for their comments on developing significant extensions and updates of Mauricio (2006b). I am also grateful to Marcos Bujosa, Miguel Jerez, Sonia Sotoca, Arthur B. Treadway, the editors of CSDA and the referees, whose comments (especially with regard to the contents of Subsection 5.1) significantly improved the overall presentation of this article. Financial support from research projects SEJ2005-07388/ECON - Ministerio de Educación y Ciencia (Spain), and 940223 - Universidad Complutense de Madrid (Spain) is also acknowledged.

## References

Ansley, C.F., 1979. An Algorithm for the Exact Likelihood of a Mixed Autoregressive Moving Average Process.

- Biometrika 66, 59-65.
- Ansley, C.F., Newbold, P., 1979. On the Finite Sample Distribution of Residual Autocorrelations in Autoregressive Moving Average Models. *Biometrika* 66, 547-553.
- Ansley, C.F., Newbold, P., 1980. Finite Sample Properties of Estimators for Autoregressive Moving Average Models. *Journal of Econometrics* 13, 159-183.
- Box, G.E.P., Jenkins, G.M., 1976. *Time Series Analysis. Forecasting and Control*, revised ed. Holden Day, Oakland.
- Box, G.E.P., Jenkins, G.M., Reinsel, G.C., 1994. *Time Series Analysis. Forecasting and Control*, third ed. Prentice Hall, Englewood Cliffs.
- Brockwell, P.J., Davis, R.A., 2002. *Introduction to Time Series and Forecasting*, second ed. Springer-Verlag, New York.
- Dhrymes, P.J., 2000. *Mathematics for Econometrics*, third ed. Springer-Verlag, New York.
- Franses, P.H., 1998. *Time Series Models for Business and Economic Forecasting*. Cambridge University Press, Cambridge.
- Gardner, G., Harvey, A.C., Phillips, G.D.A., 1980. Algorithm AS 154: An Algorithm for Exact Maximum Likelihood Estimation of Autoregressive Moving Average Models by Means of Kalman Filtering. *Applied Statistics* 29, 311-322.
- Harvey, A.C., 1993. *Time Series Models*, second ed. The MIT Press, Cambridge.
- Li, W.K., 2004. *Diagnostic Checks in Time Series*. Chapman & Hall/CRC, Boca Raton.
- Ljung, G.M., Box, G.E.P., 1978. On a Measure of Lack of Fit in Time Series Models. *Biometrika* 65, 297-303.
- Ljung, G.M., Box, G.E.P., 1979. The Likelihood Function of Stationary Autoregressive Moving Average Models. *Biometrika* 66, 265-270.
- Mauricio, J.A., 1995. Exact Maximum Likelihood Estimation of Stationary Vector ARMA Models. *Journal of the American Statistical Association* 90, 282-291.
- Mauricio, J.A., 1996. Some Computational Aspects of Exact Maximum Likelihood Estimation of Time Series Models. In: Prat, A. (Ed.), *COMPSTAT 1996 Proceedings on Computational Statistics*. Physica-Verlag, Heidelberg, pp. 361-366.
- Mauricio, J.A., 2002. An Algorithm for the Exact Likelihood of a Stationary Vector Autoregressive Moving Average Model. *Journal of Time Series Analysis* 23, 473-486.
- Mauricio, J.A., 2006a. Exact Maximum Likelihood Estimation of Partially Nonstationary Vector ARMA Models. *Computational Statistics and Data Analysis* 50, 3644-3662.
- Mauricio, J.A., 2006b. Residuals in Time Series Models. In: Rizzi, A., Vichi, M. (Eds.), *COMPSTAT 2006 Proceedings on Computational Statistics*. Physica-Verlag, Heidelberg, pp. 1179-1185.
- Mélard, G., 1984. Algorithm AS 197: A Fast Algorithm for the Exact Likelihood Function of Autoregressive Moving Average Models. *Applied Statistics* 33, 104-114.
- Mills, T.C., 1990. *Time Series Techniques for Economists*. Cambridge University Press, Cambridge.
- Newbold, P., Agiakloglou, C., Miller, J., 1994. Adventures with ARIMA Software. *International Journal of Forecasting* 10, 573-581.
- Peña, D., 2005. *Análisis de Series Temporales*. Alianza Editorial, Madrid.
- Peña, D., Tiao, G.C., Tsay, R.S. (Eds.), 2001. *A Course in Time Series Analysis*. John Wiley, New York.
- Q.M.S., 2002. *EViews 4 User's Guide* (revised for EViews 4.1, march 11 2002). Quantitative Micro Software, Irvine.
- Reinsel, G.C., 1997. *Elements of Multivariate Time Series Analysis*, second ed. Springer-Verlag, New York.
- Shea, B.L., 1989. Algorithm 242: The Exact Likelihood of a Vector Autoregressive Moving Average Model. *Applied Statistics* 38, 161-204.