COMPSTAT

Proceedings in
Computational Statistics

2006

Edited by
Alfredo Rizzi
Maurizio Vichi

Physica-Verlag
A Springer Company

Residuals in Time Series Models

José Alberto Mauricio

Universidad Complutense de Madrid

Summary. Three types of residuals in time series models, namely "conditional residuals", "unconditional residuals" and "innovations", are considered with regard to (i) their precise definitions, (ii) their computation after model estimation, (iii) their approximate distributions in finite samples, and (iv) potential applications of their properties in model diagnostic checking. Both partially-known and new results are presented, showing what might be found (and, to some extent, what should be done) in practice when dealing with residuals after estimation of time series models.

Keywords. Diagnostic checking; Residuals; Time series models

1 Introduction

The methodological approach introduced three decades ago by Box and Jenkins (1976) still represents one of the fundamental cornerstones in modern time series analysis. This article contributes to such approach by presenting some properties of residuals in autoregressive moving average (ARMA) models that do not have received much attention in the past.

Residuals constitute an important 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 should be modified. The usual approach consists of comparing patterns in computed residuals to those implied by their distributional properties under the assumption that the entertained model is adequate. Hence, in any given practical setting, it is important to know which residuals are being used for model diagnostic checking (i.e., how such residuals have been actually computed), and which theoretical properties should their observed patterns be compared to. In this respect, it is a standard practice to compare residual patterns to those of Gaussian white-noise (see, for example, Box et al. 1994, Chap. 8; Franses 1998, Chap. 3; and Enders 2004, Chap. 2). However, it is at least partially known (see, for example, 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 was given nearly thirty years ago by Ansley and Newbold (1979), and, to some extent, new

COMPSTAT 2006 PROCEEDINGS IN COMPUTATIONAL STATISTICS
Part XV Time Series Analysis and Spatial Analysis (on CD-ROM), pp. 1179-1185
Edited by A. Rizzi and M. Vichi, Physica-Verlag Heidelberg 2006

results discussed in the present article provide a theoretical justification for the empirical findings of those 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 ARMA 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 ARMA 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 ARMA model. 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), (ii) that invertibility plays a key role for establishing statistical convergence of residuals to white noise, and (iii) that a set of "normalized" 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 when it is adequately specified. Additional discussion and conclusions are provided in Section 4.

For ease and brevity of exposition, results are presented only for stationary univariate ARMA models, although extensions to the case of stationary multivariate models are straightforward. Hence, 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 with reduced-rank structure (Mauricio 2006). Several supplements to this article are available upon request, including (i) proofs of the theorems given in Section 3, (ii) numerical examples, and (iii) extensions to the case of multivariate models.

2 Residual Definitions and Computations

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

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

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 in B of degrees p and q, B is the backshift operator, $\tilde{W}_t = W_t - \mathrm{E}[W_t]$, and $\{A_t\}$ is a white-noise process with variance σ^2. 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. It is also assumed that $\phi(x)$ and $\theta(x)$ do not share any common factor.

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

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

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

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

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

where $\tilde{\mathbf{w}} = [\tilde{w}_1, \tilde{w}_2, \ldots, \tilde{w}_n]^{\mathrm{T}}$ with $\tilde{w}_t = w_t - \hat{\mu}$ $(t = 1, 2, \ldots, n)$, $\hat{\mu}$ is an estimate of $\mathrm{E}[W_t]$, and the $n\times n$ matrix $\hat{\boldsymbol{\Sigma}}_{\mathbf{W}}$ is an estimate of the "auto-covariance" matrix $\boldsymbol{\Sigma}_{\mathbf{W}} = \sigma^{-2}\mathrm{E}[\tilde{\mathbf{W}}\tilde{\mathbf{W}}^{\mathrm{T}}]$. Noting Eq. (2), it can be seen that

$$\boldsymbol{\Sigma}_{\mathbf{W}} = \mathbf{D}_\phi^{-1}(\mathbf{D}_\theta\mathbf{D}_\theta^{\mathrm{T}} + \mathbf{V}\boldsymbol{\Omega}\mathbf{V}^{\mathrm{T}})\mathbf{D}_\phi^{-1\mathrm{T}} = \mathbf{K}^{-1}(\mathbf{I} + \mathbf{Z}\boldsymbol{\Omega}\mathbf{Z}')\mathbf{K}^{-1\mathrm{T}}, \tag{4}$$

where $\mathbf{K} = \mathbf{D}_\theta^{-1}\mathbf{D}_\phi$, $\mathbf{Z} = -\mathbf{D}_\theta^{-1}\mathbf{V}$ and $\boldsymbol{\Omega} = \sigma^{-2}\mathrm{E}[\mathbf{U}_*\mathbf{U}_*^{\mathrm{T}}]$ 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 $\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q$ as described, for example, in Ljung and Box (1979). Hence, using Eq. (4), the quadratic form $\tilde{\mathbf{w}}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_{\mathbf{W}}^{-1}\tilde{\mathbf{w}}$ in Eq. (3) can be written as

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

where $\hat{\mathbf{K}}$, $\hat{\mathbf{Z}}$ and $\hat{\boldsymbol{\Omega}}$ represent estimates of the corresponding parameter matrices defined below Eq. (4).

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

Definition 2. The "unconditional residuals" associated with Eq. (5) are the elements of the $n\times 1$ vector $\hat{\mathbf{a}} = (\mathbf{I} + \hat{\mathbf{Z}}\hat{\boldsymbol{\Omega}}\hat{\mathbf{Z}}^{\mathrm{T}})^{-1}\hat{\mathbf{K}}\tilde{\mathbf{w}} = \hat{\boldsymbol{\Sigma}}_0^{-1}\hat{\mathbf{a}}^0$, where $\hat{\boldsymbol{\Sigma}}_0$ is an estimate of the $n\times n$ matrix $\boldsymbol{\Sigma}_0 = \mathbf{I} + \mathbf{Z}\boldsymbol{\Omega}\mathbf{Z}^{\mathrm{T}} = [\mathbf{I} - \mathbf{Z}(\boldsymbol{\Omega}^{-1} + \mathbf{Z}^{\mathrm{T}}\mathbf{Z})^{-1}\mathbf{Z}^{\mathrm{T}}]^{-1}$.

Definition 3. 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 $\boldsymbol{\Sigma}_{\mathbf{W}} = \mathbf{L}\mathbf{F}\mathbf{L}^{\mathrm{T}}$, with $\mathbf{F}$ being a diagonal $n\times n$ matrix such that $F_t > 0$ $(t = 1, \ldots, n)$.

Using Definitions 1 through 3, it can be seen that equation (5) can be written in several equivalent ways as $\tilde{\mathbf{w}}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_{\mathbf{W}}^{-1}\tilde{\mathbf{w}} = \hat{\mathbf{a}}^{0\mathrm{T}}\hat{\boldsymbol{\Sigma}}_0^{-1}\hat{\mathbf{a}}^0 = \hat{\mathbf{a}}^{\mathrm{T}}\hat{\boldsymbol{\Sigma}}_0\hat{\mathbf{a}} = \hat{\mathbf{e}}^{\mathrm{T}}\hat{\mathbf{F}}^{-1}\hat{\mathbf{e}}$. Note also that $|\hat{\boldsymbol{\Sigma}}_{\mathbf{W}}|$ in (4) equals both $|\hat{\boldsymbol{\Sigma}}_0|$ and $|\hat{\mathbf{F}}|$. Hence, every type of residuals defined previously can

be used to compute the exact log-likelihood given in Eq. (3). Some links between residuals defined thus far and usual ideas about residuals in classic time series analysis are considered in the following three remarks.

Remark 1. The univariate ARMA model given in Eq. (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 Eq. (4). Hence, the conditional residual vector can be written as

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

which represents the estimated expectation of $\mathbf{A}$ given an observed time series $\mathbf{w}$, under the condition that $\mathbf{U}_* = \mathbf{0}$. On the other hand, the unconditional residual vector can be written as

$$\hat{\mathbf{a}} = \hat{\mathbf{K}}\tilde{\mathbf{w}} + \hat{\mathbf{Z}}\hat{\mathbf{u}}_* = \hat{\mathrm{E}}[\mathbf{A} \mid \mathbf{W} = \mathbf{w}], \tag{7}$$

where $\hat{\mathbf{u}}_* = \hat{\mathrm{E}}[\mathbf{U}_* \mid \mathbf{W} = \mathbf{w}] = -(\hat{\boldsymbol{\Omega}}^{-1} + \hat{\mathbf{Z}}^{\mathrm{T}}\hat{\mathbf{Z}})^{-1}\hat{\mathbf{Z}}^{\mathrm{T}}\hat{\mathbf{a}}^0$ is usually referred to as the "back-casted" value of the pre-sample vector $\mathbf{U}_*$ (see Box et al. 1994, Chap. 7). In contrast to Eq. (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 the elements of $\hat{\mathbf{a}}^0$ can be computed recursively as $\hat{a}_t^0 = w_t - [\,\hat{\mu} + \sum_{i=1}^{p}\hat{\phi}_i(w_{t-i} - \hat{\mu}) - \sum_{i=1}^{q}\hat{\theta}_i\hat{a}_{t-i}^0\,]$ $(t = 1, \ldots, n)$, 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 the elements of $\hat{\mathbf{a}}$ can be computed recursively as $\hat{a}_t = w_t - [\,\hat{\mu} + \sum_{i=1}^{p}\hat{\phi}_i(w_{t-i} - \hat{\mu}) - \sum_{i=1}^{q}\hat{\theta}_i\hat{a}_{t-i}\,]$ $(t = 1, \ldots, n)$, with values for $w_j - \hat{\mu}$ $(j = 1-p, \ldots, 0)$ and $\hat{a}_j$ $(j = 1-q, \ldots, 0)$ taken from the back-cast vector $\hat{\mathbf{u}}_*$ given below Eq. (7). 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; see Ljung and Box (1979) for further details.

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), Mélard (1984), and Box et al. (1994, pp. 275-279). 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.

3 Residual Properties and Model Checking

The three types of residuals considered in Section 2 are all different from each other, which might explain to some extent why different computer programs usually generate different residuals from a given estimated model for a given time series (additionally, see Newbold et al. 1994). However, every type of residuals considered in Section 2 can be used to compute a unique set of "normalized" residuals in any of several equivalent ways. This fact is stated precisely in Theorem 3 below, which is a direct implication of some of the theoretical properties of residuals to which we now turn.

Theorem 1. Let $\hat{\mathbf{A}}^0 = \mathbf{K}\tilde{\mathbf{W}}$ and $\hat{\mathbf{A}} = \boldsymbol{\Sigma}_0^{-1}\hat{\mathbf{A}}^0$ be the random vectors associated with the conditional and the unconditional residuals given in Definitions 1 and 2, respectively, under the assumption that the true parameter values of model (2) are known. Then, (i) $\hat{\mathbf{A}}^0 \sim (\mathbf{0}, \sigma^2\boldsymbol{\Sigma}_0)$, and (ii) $\hat{\mathbf{A}} \sim (\mathbf{0}, \sigma^2\boldsymbol{\Sigma}_0^{-1})$, where $\boldsymbol{\Sigma}_0$ is given in Definition 2.

Theorem 2. Under the assumptions of Theorem 1, invertibility of the univariate ARMA model (2) implies additionally that (i) both conditional and unconditional residuals converge in mean square to the model white-noise disturbances, (ii) both conditional and unconditional residuals tend to be uncorrelated, with $\mathrm{Var}[\hat{A}_t^0]$ converging from above and $\mathrm{Var}[\hat{A}_t]$ converging from below to σ^2, and (iii) when $q = 0$ (i.e., in the case of pure autoregressive models), the convergence results stated in points (i) and (ii) occur exactly at time $t = p + 1$.

Remark 4. Theorems 1 and 2 imply, in the first place, that both conditional and unconditional residuals should not be expected 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), then Theorems 1 and 2 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 2, this possibility seems more likely to occur (especially for small sample sizes) when a model contains a moving average part with one root on or near the unit circle.

Remark 5. Remarks 1 and 2, as well as point (ii) in Theorem 2, suggest that the loss of information implied by the conditions imposed for computing conditional residuals, render such residuals as a far from ideal tool for model diagnostic checking. This is especially true when conditional residuals are naturally computed after conditional maximum likelihood estimation of ARMA models (see, for example, Ansley and Newbold 1980; and Box et al. 1994, Chap. 7). Examples showing that conditional residuals often mask a notable lack of fit are not difficult to come across.

Remark 6. Theoretical properties of innovations under the assumption that the true parameter values of the stationary model (2) 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 $\hat{\mathbf{E}} = \mathbf{L}^{-1}\tilde{\mathbf{W}} \sim (\mathbf{0}, \sigma^2\mathbf{F})$. 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, 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; furthermore, for pure autoregressive models these convergence results occur exactly at time $t = p + 1$. Hence, the innovation vector $\hat{\mathbf{E}}$ shows theoretical properties which are similar to those of Theorems 1 and 2 for $\hat{\mathbf{A}}^0$ and $\hat{\mathbf{A}}$, in spite of their numerical values being computed in practice through quite different procedures.

Theorem 3. Let $\hat{\mathbf{A}}^0 = \mathbf{K}\tilde{\mathbf{W}}$, $\hat{\mathbf{A}} = \boldsymbol{\Sigma}_0^{-1}\hat{\mathbf{A}}^0$ and $\hat{\mathbf{E}} = \mathbf{L}^{-1}\tilde{\mathbf{W}}$ be the random vectors associated with the three types of residuals given in Definitions 1, 2 and 3, respectively, under the assumption that the true parameter values of model (2) are known. Let $\mathbf{P}$

represent a lower-triangular matrix such that $\boldsymbol{\Sigma}_0 = \mathbf{I} + \mathbf{Z}\boldsymbol{\Omega}\mathbf{Z}^{\mathrm{T}} = \mathbf{P}\mathbf{P}^{\mathrm{T}}$. Then, there exists a "normalized" residual vector $\hat{\mathbf{Z}}$ such that $\hat{\mathbf{Z}} = \mathbf{P}^{-1}\hat{\mathbf{A}}^0 = \mathbf{P}^{\mathrm{T}}\hat{\mathbf{A}} = \mathbf{F}^{-1/2}\hat{\mathbf{E}}$, with $\hat{\mathbf{Z}} \sim (\mathbf{0}, \sigma^2\mathbf{I})$. Additionally, invertibility of model (2) implies that the elements of $\hat{\mathbf{Z}}$ converge in mean square to the model white-noise disturbances, with exact convergence occurring at time $t = p + 1$ when $q = 0$.

Remark 7. Theorem 3 implies that when the true model parameter values are replaced by consistent estimates, the elements of the computed normalized residual vector, $\hat{\mathbf{z}} = \hat{\mathbf{P}}^{-1}\hat{\mathbf{a}}^0 = \hat{\mathbf{P}}^{\mathrm{T}}\hat{\mathbf{a}} = \hat{\mathbf{F}}^{-1/2}\hat{\mathbf{e}}$, 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. Hence, using $\hat{\mathbf{z}}$ 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 4), whose only source in the case of $\hat{\mathbf{z}}$ is (at least approximately and apart from outliers) model misspecification. Furthermore, working with $\hat{\mathbf{z}}$ 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 mean-square convergence occurs sufficiently fast in practice.

4 Discussion

Ansley and Newbold (1979, pp. 551-553) have demonstrated that using the normalized residual vector $\hat{\mathbf{z}}$ 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 can be usefully interpreted through the usual asymptotic significance levels. However, these authors suggest (i) a single way of computing $\hat{\mathbf{z}}$ (Ansley 1979), and (ii) that the only reason for the superior sampling properties of tests based on $\hat{\mathbf{z}}$ is that unconditional residuals can, in moderate sample sizes, have variance much smaller than σ^2, whereas normalized residuals have the same variance as the model random shocks. Recalling Theorems 2 and 3, the conclusions derived by Ansley and Newbold (1979) can be expanded as follows: (i) $\hat{\mathbf{z}}$ can be computed in any of several equivalent ways, and (ii) the practical benefits from using $\hat{\mathbf{z}}$ instead of $\hat{\mathbf{a}}$ are associated with the fact that $\mathrm{Var}[\hat{\mathbf{A}}] = \sigma^2\boldsymbol{\Sigma}_0^{-1}$ is not a diagonal matrix, so not only the unconditional residuals have variance smaller than σ^2, but (more importantly) they are also autocorrelated. The simulation results reported by Ansley and Newbold (1979) are so detailed that no additional evidence seems to be required to demonstrate the expected practical benefits from using $\hat{\mathbf{z}}$ instead of $\hat{\mathbf{a}}$ (or $\hat{\mathbf{a}}^0$ for that matter) in model diagnostic checking, especially when the model contains moving-average roots on or near the unit circle and/or when the sample size n is small.

From a computational standpoint, the most convenient method for obtaining $\hat{\mathbf{z}}$ 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 much larger number of operations required for computing $\hat{\boldsymbol{\Sigma}}_0 = \mathbf{I} + \hat{\mathbf{Z}}\hat{\boldsymbol{\Omega}}\hat{\mathbf{Z}}^{\mathrm{T}}$, its Cholesky factor $\hat{\mathbf{P}}$, and any of the products $\hat{\mathbf{P}}^{-1}\hat{\mathbf{a}}^0$ or $\hat{\mathbf{P}}^{\mathrm{T}}\hat{\mathbf{a}}$). This fact, together with the apparent lack in previous literature of analytical expressions for $\mathrm{Var}[\hat{\mathbf{A}}^0]$ and $\mathrm{Var}[\hat{\mathbf{A}}]$ (as opposed to well-known results on $\mathrm{Var}[\hat{\mathbf{E}}]$), might explain why computer programs for estimation of ARMA

models based on the innovations approach usually provide the practitioner with both the computed innovations $\hat{e}_t$ and the corresponding normalized residuals $\hat{z}_t = \hat{F}_t^{-1/2}\hat{e}_t$ (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. However, the calculations required for obtaining $\hat{\mathbf{P}}^{-1}\hat{\mathbf{a}}^0$ or $\hat{\mathbf{P}}^{\mathrm{T}}\hat{\mathbf{a}}$ need to be carried out only after model estimation, and, indeed, such calculations involve a negligible amount of computing time for most modern computers.

Acknowledgements. I am grateful to the referees for their comments. Financial support from research projects SEJ2005-07388/ECON - Ministerio de Educación y Ciencia (Spain), and 940223 - Universidad Complutense de Madrid (Spain), is also gratefully acknowledged.

References

Ansley CF (1979) An Algorithm for the Exact Likelihood of a Mixed Autoregressive Moving Average Process. Biometrika 66:59-65.

Ansley CF, Newbold P (1979) On the Finite Sample Distribution of Residual Autocorrelations in Autoregressive Moving Average Models. Biometrika 66:547–553.

Ansley CF, Newbold P (1980) Finite Sample Properties of Estimators for Autoregressive Moving Average Models. J Econometrics 13:159–183.

Box GEP, Jenkins GM (1976) Time Series Analysis. Forecasting and Control, revised edn. Holden Day, Oakland.

Box GEP, Jenkins GM, Reinsel GC (1994) Time Series Analysis. Forecasting and Control, 3rd edn. Prentice Hall, Englewood Ciffs.

Brockwell PJ, Davis RA (2002) Introduction to Time Series and Forecasting, 2nd edn. Springer, New York.

Enders WE (2004) Applied Econometric Time Series, 2nd edn. Wiley, New York.

Franses PH (1998) Time Series Models for Business and Economic Forecasting. Cambridge University Press, Cambridge.

Harvey AC (1993) Time Series Models, 2nd edn. The MIT Press, Cambridge.

Ljung GM, Box GEP (1979) The Likelihood Function of Stationary Autoregressive Moving Average Models. Biometrika 66:265–270.

Mauricio JA (1996) Some Computational Aspects of Exact Maximum Likelihood Estimation of Time Series Models. In: Prat A (ed) COMPSTAT 1996 Proceedings in Computational Statistics. Physica Verlag, Heidelberg, pp 361–366.

Mauricio JA (2006) Exact Maximum Likelihood Estimation of Partially Nonstationary Vector ARMA Models. Comput Stat Data An 50:3644-3662.

Mélard G (1984) A Fast Algorithm for the Exact Likelihood Function of Autoregressive Moving Average Models. J Roy Stat Soc C 46:104–114.

Newbold P, Agiakloglou C, Miller J (1994) Adventures with ARIMA Software. Int J Forecasting 10:573–581.