Econometric Theory, **14**, 1998, 326–338. Printed in the United States of America.

ON ESTIMATING AN ARMA MODEL WITH AN MA UNIT ROOT

B.P.M. McCabe
University of British Columbia

S.J. Leybourne
University of Nottingham

This paper investigates the behavior of the maximum likelihood estimator of a Gaussian autoregressive moving average model with a unit root in the moving average polynomial. The results are primarily of interest in testing hypotheses that involve moving average unit roots as, for example, when testing for stationarity of a series.

1. INTRODUCTION

It is well known that classical likelihood theory (i.e., $T^{1/2}$-consistency and asymptotic normality) does not apply to autoregressive (AR) models that contain a unit root. The fact that this is also true of moving average (MA) models with a unit root seems to be less well known in spite of the work, among others, of Cryer and Ledolter (1981), Sargan and Bhargava (1983), Tanaka and Satchell (1989), Potscher (1991), and Shephard (1993). In the MA(1) case when a unit root is present, the maximum likelihood estimator (MLE) of θ, the MA parameter, is T-consistent in contrast to the usual $T^{1/2}$ rate when $|\theta| < 1$. Recently, the asymptotic distribution of the MLE of θ has been rigorously derived by Davis, Chen, and Dunsmuir (1994) and Davis and Dunsmuir (1996) in the Gaussian case. In addition to very accurate approximations to the distribution they also provide critical values for the generalized likelihood ratio test and show it has good power properties for a broad range of values of θ. In the ARMA case the consistency of the MLE's of θ and the AR parameters has been established by Potscher in the presence of an MA unit root even when a Gaussian pseudolikelihood is used. The object of this paper is to establish the (vector) marginal distribution of the MLE of the AR parameters in the case where an ARMA$(p,1)$ model is fitted to observations from a stationary Gaussian ARMA$(p,1)$ process and the true value of θ, the MA parameter, is unity. It turns out that the marginal distribution of the MLE of the AR parameters, in this situation, is asymptotically the same as the distribution of parameter estimates (MLE's) in a pure AR(p) model fitted to a station-

We thank Paul Newbold, the associate editor, and a referee for helpful comments on an earlier draft of this paper. Address correspondence to: S.J. Leybourne, Department of Economics, University of Nottingham, Nottingham NG7 2RD, United Kingdom.

 © 1998 Cambridge University Press 0266-4666/98 $9.50

https://doi.org/10.1017/S0266466698143025 Published online by Cambridge University Press

ary pure AR(p) process. In addition, it is shown that the marginal asymptotic distribution of the MLE of the MA parameter is the same as if a pure MA(1) model was fitted to an MA(1) process with a unit root. Strictly speaking the results presented here refer to a consistent root of the score equation and are termed "local minimizers" by Davis and Dunsmuir (1996). In principle, the results presented here could be extended along the lines of Potscher but at the expense of considerable complications in the algebra.

These distributional results prove useful, for example, in the context of testing for stationarity of a series, x_t say, against a difference stationary alternative. This arises because testing the null of stationarity is, in essence, equivalent to testing whether $y_t = \Delta x_t$ contains an MA unit root.[1] In practice it is usually necessary to account for correlation in y_t in excess of that implied by the pure MA unit root and (as in the case of AR unit root processes) it often proves efficacious to account for this correlation by inclusion of an autoregressive component in the model for y_t. Thus we are led to consideration of an ARMA(p,1) model in such testing problems. Implementation of these tests requires that the order of the autoregressive component has to be chosen prior to conducting the test for the MA unit root. Our results allow this choice to be made by means of the familiar backward or forward t-test procedures.

In fact, the distribution theory generalizes to an ARMA(p,q) process, where the MA polynomial contains a single unit root. If an ARMA(p,q) model is fitted to such a process, the limiting distribution of MLE of the autoregressive parameter vector is the same as if it was estimated from an ARMA($p,(q-1)$) model fitted to an invertible ARMA($p,(q-1)$) process. It proves convenient to provide some background on estimating nonergodic models to facilitate the presentation of the proof of the foregoing results.

2. MAXIMUM LIKELIHOOD ESTIMATION OF NONERGODIC MODELS

Let $\psi = [\phi', \theta']'$ be a vector of parameters partitioned into the subvectors ϕ and θ. Let $l(\psi) = l(\mathbf{y}|\psi)$ be the log likelihood of the $T \times 1$ vector $\mathbf{y} = [y_1, y_2, \ldots, y_T]'$. We denote the vector of partial derivatives of l as l_ψ and the matrix of second derivatives as $l_{\psi\psi}$ where

$$l_{\psi\psi} = \begin{bmatrix} l_{\phi\phi} & l_{\phi\theta} \\ l'_{\phi\theta} & l_{\theta\theta} \end{bmatrix}.$$

Our treatment follows Barndorf-Nielson and Sorenson (1994), where regularity conditions and further details may be found. We assume there is a consistent root of the score equation, which we define to be the MLE of $\psi, \hat{\psi}$. Thus,

$$l_\psi(\hat{\psi}) = \mathbf{0} = l_\psi(\psi) + l_{\psi\psi}(\psi^*)(\hat{\psi} - \psi),$$

where the vector ψ^* is made up of convex linear combinations of the elements of $\hat{\psi}$ and ψ, the true value of the parameter. We suppress the subscript T for economy

https://doi.org/10.1017/S0266466698143025 Published online by Cambridge University Press

of notation. Scaling to allow "information" of possibly different orders for the elements of ψ, we have

$$-\mathbb{E}_D(\psi)^{-1/2} l_\psi(\psi) = \mathbb{J}(\psi,\psi^*)\mathbb{E}_D(\psi)^{1/2}(\hat{\psi} - \psi), \quad \textbf{(1)}$$

$$= \mathbb{J}(\psi,\psi)\mathbb{E}_D(\psi)^{1/2}(\hat{\psi} - \psi) + \hat{R}(\psi,\psi^*), \quad \textbf{(2)}$$

where $\hat{R}(\psi,\psi^*)$ is a remainder term

$$\hat{R}(\psi,\psi^*) = [\mathbb{J}(\psi,\psi^*) - \mathbb{J}(\psi,\psi)]\mathbb{E}_D(\psi)^{1/2}(\hat{\psi} - \psi)$$

and

$$\mathbb{J}(\psi,\psi^*) = \mathbb{E}_D(\psi)^{-1/2} l_{\psi\psi}(\psi^*)\mathbb{E}_D(\psi)^{-1/2}$$

and the definition of $\mathbb{J}(\psi,\psi)$ follows from that of $\mathbb{J}(\psi,\psi^*)$. Here $\mathbb{E}$ represents the information matrix, $\mathbb{E}(\psi) = E[l_\psi(\psi) l_\psi(\psi)']$, and $\mathbb{E}_D(\psi)$ is the block diagonal matrix

$$\mathbb{E}_D(\psi) = \begin{bmatrix} E[l_\phi l_\phi'] & \mathbf{0} \\ \mathbf{0} & E[l_\theta l_\theta'] \end{bmatrix}.$$

It is clear that the behavior of the MLE depends critically on the standardized score and the matrix $\mathbb{J}(\psi,\psi)$.

The most common case (assuming the remainder term is negligible asymptotically) is where $\mathbb{J}(\psi,\psi)$ converges in probability to a constant nonsingular matrix $\mathbf{V}(\psi)$, and such models are called ergodic. In this situation $\mathbb{E}_D(\psi)^{1/2}(\hat{\psi} - \psi)$ is asymptotically normal with covariance matrix $\mathbf{V}(\psi)^{-1}$. This result includes maximum likelihood (ML) under classical regularity and also the case where y_t is a stationary invertible autoregressive moving average (ARMA) process, for example.

A generalization is the case where $\mathbf{V}(\psi)$ is a random matrix (the nonergodic case) and the mode of convergence of $\mathbb{J}(\psi,\psi)$ is convergence in probability. The randomness of $\mathbf{V}(\psi)$ induces a limiting distribution for $\mathbb{E}_D(\psi)^{1/2}(\hat{\psi} - \psi)$ that is no longer normal. The limiting distribution is a variance mixture of normals and gives rise to the so-called locally asymptotic mixed normal (LAMN) family.

Yet another case arises when $\mathbb{J}(\psi,\psi)$ converges in distribution, as opposed to converging in probability as was previously assumed. In this situation nonnormal limiting distributions for the standardized MLE arise. The most familiar example is the MLE of ρ, the parameter in an AR(1) model when the true value of ρ is unity. This category also includes the case of an MA unit root.

Now, consider the problem of estimating ϕ *assuming the true value of* θ *is known*. The MLE of ϕ, $\tilde{\phi}$ say, is defined by

$$l_\phi(\tilde{\phi},\theta) = \mathbf{0} = l_\phi(\phi,\theta) + l_{\phi\phi}(\phi^+,\theta)(\tilde{\phi} - \phi),$$

https://doi.org/10.1017/S0266466698143025 Published online by Cambridge University Press

where ϕ^+ is a vector of convex linear combinations of the elements of $\tilde{\phi}$ and ϕ. Similarly, the MLE of θ when ϕ is known, $\tilde{\theta}$ say, is defined by

$$l_\theta(\phi,\tilde{\theta}) = \mathbf{0} = l_\theta(\phi,\theta) + l_{\theta\theta}(\phi,\theta^+)(\tilde{\theta} - \theta),$$

and θ^+ is defined in the usual way. Combining these two expressions into a vector equation and standardizing, we may write

$$-\mathbb{E}_D(\psi)^{-1/2} l_\psi(\psi) = \mathbb{J}_D(\psi,\psi)\mathbb{E}_D(\psi)^{1/2}(\tilde{\psi} - \psi) + \tilde{R}(\psi,\psi^+), \tag{3}$$

where

$$\tilde{R}(\psi,\psi^+) = [\mathbb{J}_D(\psi,\psi^+) - \mathbb{J}_D(\psi,\psi)]\mathbb{E}_D(\psi)^{1/2}(\tilde{\psi} - \psi),$$

$$\mathbb{J}_D(\psi,\psi^+) = \mathbb{E}_D(\psi)^{-1/2} l_{\psi\psi D}(\psi^+)\mathbb{E}_D(\psi)^{-1/2},$$

$l_{\psi\psi D}(\psi^+)$ is the block diagonal matrix

$$l_{\psi\psi D}(\psi^+) = \begin{bmatrix} l_{\phi\phi}(\phi^+,\theta) & \mathbf{0} \\ \mathbf{0} & l_{\theta\theta}(\phi,\theta^+) \end{bmatrix},$$

and the definition of $\mathbb{J}_D(\psi,\psi)$ follows from that of the block diagonal matrix $\mathbb{J}_D(\psi,\psi^+)$. Combining equations (2) and (3) yields

$$\mathbb{E}_D(\psi)^{1/2}(\tilde{\psi} - \psi) = \mathbb{J}_D(\psi,\psi)^{-1}\mathbb{J}(\psi,\psi)\mathbb{E}_D(\psi)^{1/2}(\hat{\psi} - \psi) + R_c, \tag{4}$$

where R_c is the (scaled) compound remainder term

$$R_c = \mathbb{J}_D(\psi,\psi)^{-1}[\tilde{R}(\psi,\psi^+) - \hat{R}(\psi,\psi^*)].$$

Regarding (4), in the context of the present paper it will be apparent from the preceding paragraphs that neither $\mathbb{J}_D(\psi,\psi)$ nor $\mathbb{J}(\psi,\psi)$ can converge in probability to either a fixed or random matrix because this would contradict the nonnormality known to be associated with the MA unit root. Nevertheless, it may be possible to show that the product $\mathbb{J}_D(\psi,\psi)^{-1}\mathbb{J}(\psi,\psi)$ converges in probability (distribution) to the identity matrix. If so, equation (4) will show that the (standardized) full MLE's and the MLE's with partially known parameters have the same asymptotic distribution if, in addition, R_c can be shown to be $\mathbf{o}_p(1)$. The following section implements this approach for the ARMA(p,1) model with an MA unit root.

3. THE ARMA(p,1) MODEL WITH AN MA UNIT ROOT

Consider an ARMA(p,1) model for y_t, $t = p + 1,\ldots,T$, given by

$$\Phi(L)y_t = (1 - \theta L)\varepsilon_t. \tag{5}$$

Here $\Phi(L) = 1 - \phi_1 L - \cdots - \phi_p L^p$ is a pth-order autoregressive polynomial in the lag operator L, with roots outside the unit circle and the ε_t are assumed to be

https://doi.org/10.1017/S0266466698143025 Published online by Cambridge University Press

normally and independently distributed with mean zero and variance σ^2. We condition on the starting values $y_1,\ldots,y_p$ throughout.

In conformation with the notation used in Section 2, we define $\psi = [\phi',\theta]'$ where $\phi = [\phi_1,\ldots,\phi_p]'$. Also, $\hat{\psi} = [\hat{\phi}',\hat{\theta}]'$ denotes the vector of full MLE's of ϕ and θ from (5), and $\tilde{\psi} = [\tilde{\phi}',\tilde{\theta}]'$ contains the MLE of $\phi(\theta)$ from (5). It is assumed that $\theta(\phi)$ is known. As a general rule in this section most of the details have been relegated to the Appendix.

Our aim is to establish the behavior of $\hat{\psi}$ in the case where $\theta = 1$. Define the process $z_t(\theta)$, valid for $\theta = 1$, by

$$\Phi(L)z_t(\theta) = \varepsilon_t, \tag{6}$$

where the disturbance term ε_t is exactly the same as that in (5). Thus, assuming θ is known and not being estimated, the MLE of the autoregressive parameter vector ϕ in (5), $\tilde{\phi}$, is the same as that of the MLE of ϕ from the *pure* AR(p) process (6) in z_t, even when $\theta = 1$. Similarly, we may define another process $w_t(\phi) = \Phi(L)y_t$, such that

$$w_t(\phi) = \varepsilon_t - \theta\varepsilon_{t-1} \tag{7}$$

and using analogous arguments, when ϕ is known, the MLE of θ in (5), $\tilde{\theta}$, is the same as the MLE of θ from the *pure* MA(1) process (7) in w_t, even when $\theta = 1$.

Given these relationships, then (4) specialized to an ARMA(p,1) model with an MA unit root will establish that the MLE of ϕ in the ARMA(p,1) model (5), that is, $\hat{\phi}$, has the same asymptotic distribution as the MLE of ϕ in the corresponding pure AR(p) model (6) ($\tilde{\phi}$) if we can show that $\mathbb{J}_D(\psi,\psi)^{-1}\mathbb{J}(\psi,\psi)$ converges to the identity matrix, and that R_c converges to zero. At the same time, this will also establish the equivalence of the asymptotic distributions of the MLE of θ in the ARMA(p,1) model (5) ($\hat{\theta}$) and that of θ in the pure MA(1) model (7) ($\tilde{\theta}$).

The behavior of $\mathbb{J}(\psi,\psi)$ is dependent on the first and second derivatives of the log-likelihood function of the ARMA(p,1) model (5). Appendix A establishes the following results when the derivatives are evaluated at the true values of the ARMA(p,1) parameters $\psi = \psi^0 = [\phi^{0\prime},\theta^0]' = [\phi^{0\prime},1]'$:

$$T^{-1}E[l_\phi(\psi)l_\phi(\psi)'] = \mathbf{V}_1 + \mathbf{o}(1), \tag{8}$$

$$T^{-1}l_{\phi\phi}(\psi) = -\mathbf{V}_1 + \mathbf{o}_p(1), \tag{9}$$

$$T^{-2}E[l_\theta(\psi)^2] = v_2 + o(1), \tag{10}$$

$$T^{-2}l_{\theta\theta}(\psi) = -v_2 + O_p(1), \tag{11}$$

$$T^{-1}l_{\phi\theta}(\psi) = \mathbf{O}_p(1). \tag{12}$$

https://doi.org/10.1017/S0266466698143025 Published online by Cambridge University Press

Here $\mathbf{V}_1 = \sigma^{-2}E[\mathbf{z}_t\mathbf{z}_t']$, $\mathbf{z}_t = [z_t, \ldots, z_{t-p+1}]'$, and z_t is defined as in (6). Also, $v_2 = E[\int w(r)^2\,dr] = \int E[w(r)^2]\,dr = \int r\,dr = \frac{1}{2}$, where $w(r)$ is a Brownian motion process. Hence,

$$\mathbb{J}(\psi,\psi) = \mathbb{E}_D(\psi)^{-1/2} l_{\psi\psi}(\psi)\mathbb{E}_D(\psi)^{-1/2} = \begin{bmatrix} -\mathbf{I}_p + \mathbf{o}_p(1) & \mathbf{O}_p(T^{-1/2}) \\ \mathbf{O}_p(T^{-1/2}) & -1 + O_p(1) \end{bmatrix}. \quad \textbf{(13)}$$

Thus, the $(p+1, p+1)$ element of $\mathbb{J}(\psi,\psi)$ converges in distribution to a random variable, whereas all the other blocks converge in distribution to constants. The blocks of $\mathbb{J}_D(\psi,\psi)$ are simply the diagonal blocks of $\mathbb{J}(\psi,\psi)$, and so $\mathbb{J}_D(\psi,\psi)^{-1}\mathbb{J}(\psi,\psi) \to_d \mathbf{I}_{p+1}$ by the continuous mapping theorem because the off diagonal blocks of $\mathbb{J}(\psi,\psi)$ converge in distribution to zero.

Turning to R_c we have, from Appendix B,

$$\tilde{R}(\psi,\psi^+) = \mathbf{o}_p(1),$$

$$\hat{R}(\psi,\psi^*) = \mathbf{o}_p(1),$$

and because from (13) it follows that $\mathbb{J}_D(\psi,\psi)^{-1} = \mathbf{O}_p(1)$ then

$$R_c = \mathbb{J}_D(\psi,\psi)^{-1}[\tilde{R}(\psi,\psi^+) - \hat{R}(\psi,\psi^*)]$$
$$= \mathbf{o}_p(1).$$

Finally then, taking all these results together, (4) shows that the full MLE's have the asymptotic distributions given in the paragraph following (7).

4. A NUMERICAL ILLUSTRATION

We illustrate the result derived previously for the case of the estimated AR parameter in an ARMA(1,1) model with an MA unit root. The following AR(1) model was simulated

$$x_t = \phi_1 x_{t-1} + \varepsilon_t, \qquad t = 1, \ldots, T+40,$$
$$x_0 = 0$$

with ε_t generated as independent and identically distributed (i.i.d.) standard normal. Next, $y_t = \Delta x_t$ was constructed for the T observations $t = 41, \ldots, T+40$ (the first 40 observations being discarded to eliminate the influence of starting values). Thus, y_t follows an ARMA(1,1) process with an MA unit root and AR parameter ϕ_1. For each replication, an ARMA(1,1) model was fitted to y_t using the GAUSS ML algorithm. According to our results, when $|\phi_1| < 1$ the estimated AR parameter in this model, $\hat{\phi}_1$, behaves asymptotically as if it was estimated from a *pure* AR(1) process with parameter ϕ_1; that is, $T^{1/2}(\hat{\phi}_1 - \phi_1) \to_d N(0,(1-\phi_1^2))$. We considered the standardized variable

$$Z_{1T} = T^{1/2}(\hat{\phi}_1 - \phi_1)/(1-\hat{\phi}_1^2)^{1/2},$$

https://doi.org/10.1017/S0266466698143025 Published online by Cambridge University Press

https://doi.org/10.1017/S0266466698143025 Published online by Cambridge University Press

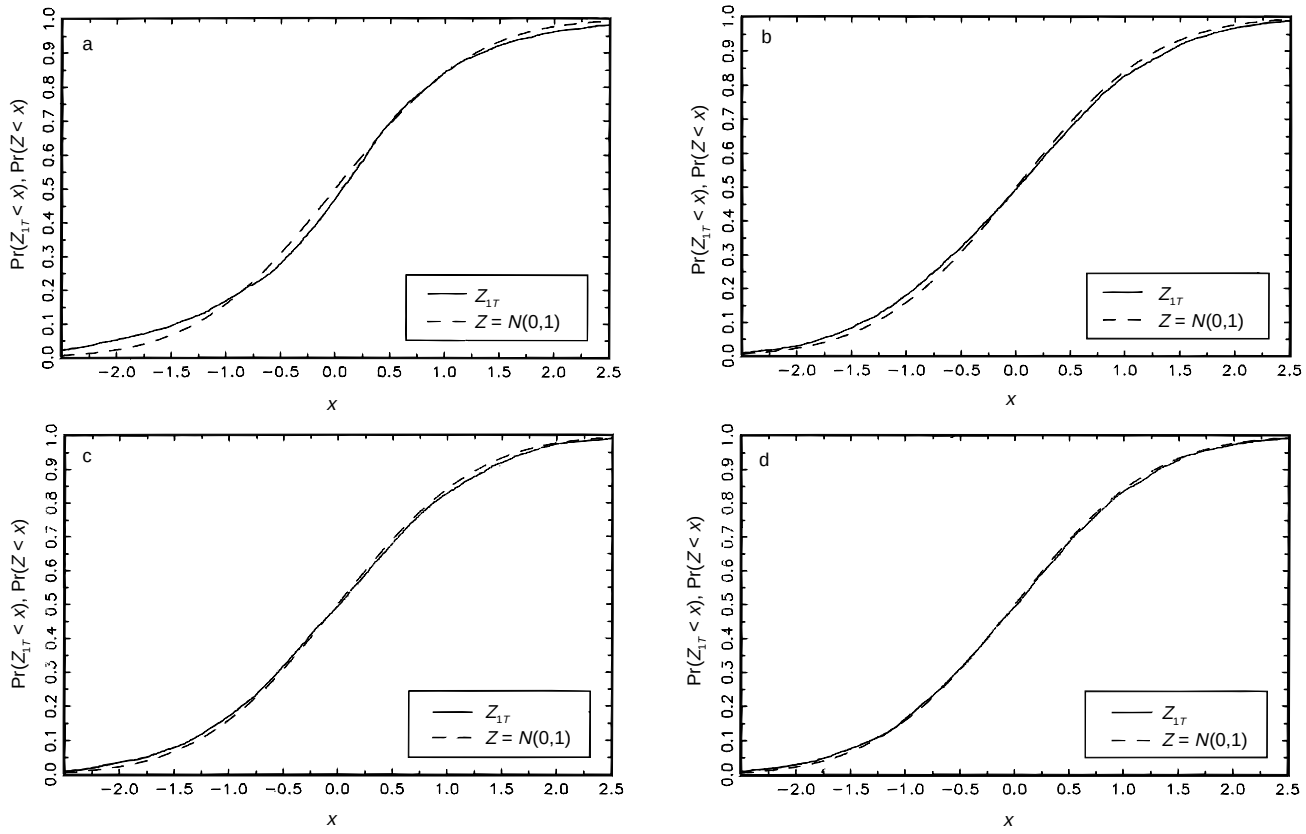


FIGURE 1. Empirical c.d.f. of Z_{1T}: (a) $\phi_1 = 0.4$, $T = 50$; (b) $\phi_1 = 0.4$, $T = 100$; (c) $\phi_1 = 0.4$, $T = 200$; (d) $\phi_1 = 0.4$, $T = 400$.

https://doi.org/10.1017/S0266466698143025 Published online by Cambridge University Press

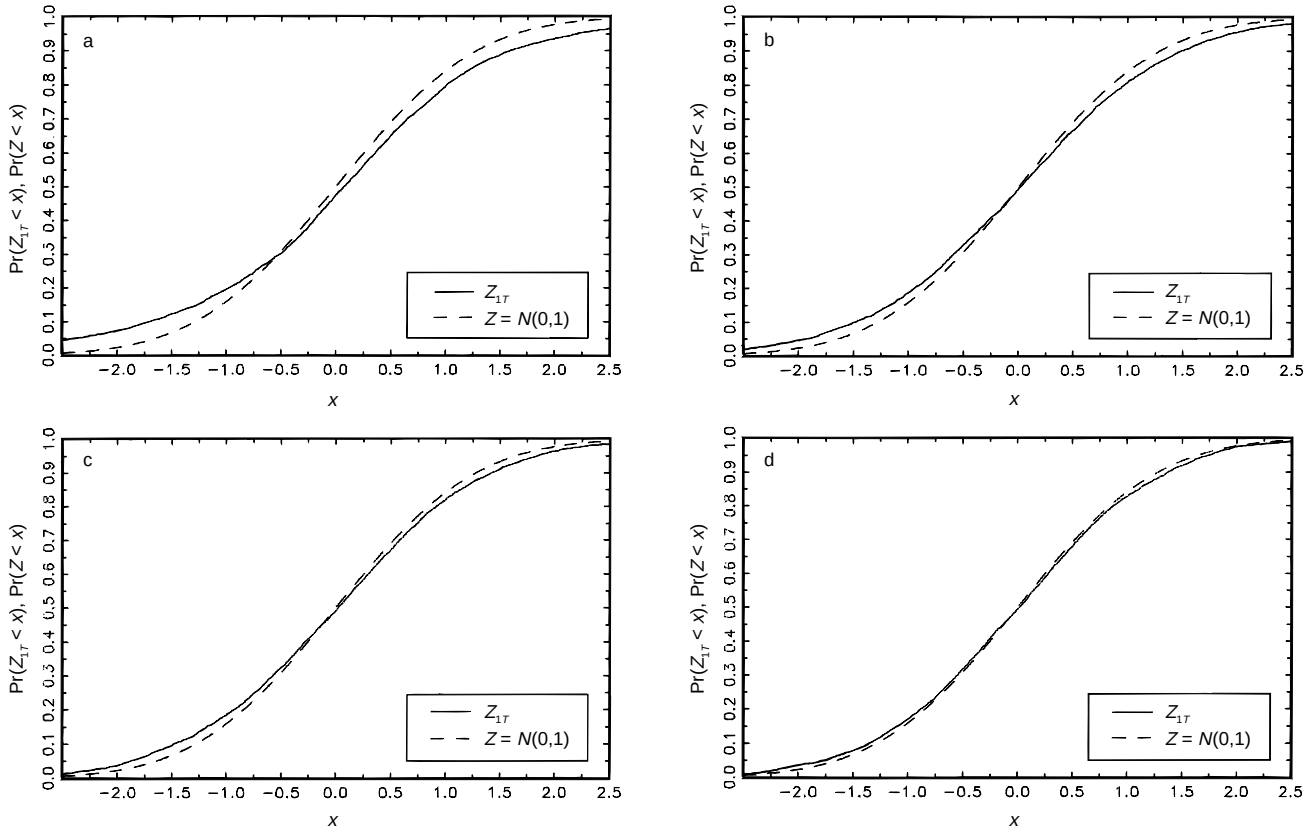


FIGURE 2. Empirical c.d.f. of Z_{1T}: (a) $\phi_1 = 0.8$, $T = 100$; (b) $\phi_1 = 0.8$, $T = 200$; (c) $\phi_1 = 0.8$, $T = 400$; (d) $\phi_1 = 0.8$, $T = 800$.

which ought to possess a standard normal limiting distribution. Figures 1a–d show the empirical cumulative distribution function (c.d.f.) of Z_{1T} when $\phi_1 = 0.4$, based on 4,000 replications. Also shown is the exact c.d.f. for a standard normal variate, denoted Z. It is clear that even when the sample size is as small as $T = 50$, the distribution of Z_{1T} is reasonably close to that of Z. Moreover, as the sample size increases the two distributions rapidly converge, being virtually indistinguishable when $T = 400$. Results in the case where $\phi_1 = 0.8$ are shown in Figures 2a–d. Because this ARMA(1,1) generating process implies a near common factor in the AR and MA polynomials and this causes convergence problems for the ML algorithm in small samples, we start the simulations with $T = 100$. Convergence of Z_{1T} to the standard normal is once more evident although it occurs more slowly than for $\phi_1 = 0.4$. This is only to be expected, however, in the presence of a near common factor in the generating process.

NOTE

1. Test procedures based on this principle have been suggested by, inter alia, Tanaka (1990), Tsay (1993), and Leybourne and McCabe (1994, 1996).

REFERENCES

Barndorf-Neilson, O.D. & M. Sorenson (1994) A review of some aspects of asymptotic likelihood theory for stochastic processes. *International Statistical Review* 42, 133–165.

Chan, N.H. & R.S. Tsay (1996) Asymptotic inference for non-invertible moving-average time series. *Journal of Time Series Analysis* 17, 1–17.

Cryer, J.D. & J. Ledolter (1981) Small-sample properties of the maximum likelihood estimator in the first-order moving average model. *Biometrika* 68, 691–694.

Davis, R.A., M. Chen, & W.T.M. Dunsmuir (1994) Inference for MA(1) processes with a unit root on or near the unit circle. *Probability and Mathematical Statistics* 15, 227–242.

Davis, R.A. & W.T.M. Dunsmuir (1996) Maximum-likelihood estimation for MA(1) processes with a unit root on or near the unit circle. *Econometric Theory* 12, 1–29.

Leybourne, S.J. & B.P.M. McCabe (1994) A consistent test for a unit root. *Journal of Business and Economic Statistics* 12, 157–166.

Leybourne, S.J. & B.P.M. McCabe (1996) Modified Stationarity Tests with Data Dependent Model Selection Rules. Mimeo, University of Nottingham.

Potscher, B.M. (1991) Noninvertibility and pseudo-maximum likelihood estimation of ARMA models. *Econometric Theory* 7, 435–449.

Sargan, J.D. & A. Bhargava (1983) Maximum-likelihood estimation of regression models with first order moving average errors when the root lies on the unit circle. *Econometrica* 51, 799–820.

Shephard, N. (1993) Distribution of the ML estimator of an MA(1) and local level model. *Econometric Theory* 9, 377–401.

Tanaka, K. (1990) Testing for a moving average unit root. *Econometric Theory* 6 433–444.

Tanaka, K. & S.E. Satchell (1989) Asymptotic properties of the maximum likelihood and nonlinear least squares estimators for noninvertible moving average models. *Econometric Theory* 5, 333–353.

Tsay, R.S. (1993) Testing for noninvertible models with applications. *Journal of Business and Economic Statistics* 11, 225–233.

https://doi.org/10.1017/S0266466698143025 Published online by Cambridge University Press

APPENDIX A

We suppose that we observe a sample y_t, $t = 1,\ldots,T$, generated from (5) of the main text. We initialize by defining

$$z_1(\theta) = y_1, \qquad z_2(\theta) = \theta z_1(\theta) + y_2, \ldots, z_p(\theta) = \theta z_{p-1}(\theta) + y_p,$$

and so conditioning on the y starting values is the same as conditioning on those of z. It follows that

$$\begin{aligned}(1 - \theta L)\Phi(L)z_t(\theta) &= (1 - \theta L)\varepsilon_t \\ &= \Phi(L)y_t,\end{aligned}$$

and on canceling the $\Phi(L)$ we have for $t = 2,\ldots,T$

$$z_t(\theta) = \theta z_{t-1}(\theta) + y_t. \tag{A.1}$$

Notice that the "disturbance" in the AR(1)-style process (A.1) is not noise but is the series y_t itself. Moreover, (A.1) is valid even for $\theta = 1$. A similar sort of derived process was used by Chan and Tsay (1996) in the pure MA case.

From (A.1) it is straightforward to demonstrate that the Jacobian of the transformation between $\mathbf{y}$ and $\mathbf{z}$ is unity. Hence, the log likelihood of the y_t may be expressed in terms of the z_t. We deduce that the log likelihood of the observations y_t, $t = p + 1,\ldots,T$, $l(\psi)$, under model (5) is, in view of (6),

$$l(\psi) = -0.5(T - p)\ln(2\pi\sigma^2) - 0.5\sigma^{-2}\sum_{p+1}^{T}[\Phi(L)z_t(\theta)]^2.$$

The derivatives of l with respect to ϕ and θ are

$$l_\phi = \sigma^{-2}\sum_{t=p+1}^{T}\Phi(L)z_t \times \mathbf{z}_{t-1}, \tag{A.2}$$

$$l_{\phi\phi} = -\sigma^{-2}\sum_{t=p+1}^{T}\mathbf{z}_{t-1}\mathbf{z}'_{t-1}, \tag{A.3}$$

$$l_\theta = -\sigma^{-2}\sum_{t=p+1}^{T}\Phi(L)z_t \times \Phi(L)dz_t/d\theta, \tag{A.4}$$

$$l_{\theta\theta} = -\sigma^{-2}\sum_{t=p+1}^{T}[\Phi(L)dz_t/d\theta]^2 - \sigma^{-2}\sum_{t=p+1}^{T}\Phi(L)z_t \times \Phi(L)d^2z_t/d\theta^2, \tag{A.5}$$

$$l_{\phi\theta} = \sigma^{-2}\sum_{t=p+1}^{T}\mathbf{z}_{t-1} \times \Phi(L)dz_t/d\theta + \sigma^{-2}\sum_{t=p+1}^{T}\Phi(L)z_t \times d\mathbf{z}_{t-1}/d\theta, \tag{A.6}$$

where $\mathbf{z}_t = [z_t,\ldots,z_{t-p+1}]'$. For the true values of the ARMA parameters $\psi = \psi^0 = [\phi^{0\prime}, \theta^0]'$, $\Phi(L)z_t(\theta^0) = \varepsilon_t$, regardless of whether $\theta^0 = 1$, and (8) and (9) follow straightforwardly from (A.2) and (A.3).

https://doi.org/10.1017/S0266466698143025 Published online by Cambridge University Press

Now, (A.4)–(A.6) require derivatives of $z_t(\theta)$. From (A.1) for $t \geq 2$ we obtain

$$dz_t/d\theta = \theta dz_{t-1}/d\theta + z_{t-1}, \tag{A.7}$$

$$d^2z_t/d\theta^2 = \theta d^2z_{t-1}/d\theta^2 + 2dz_{t-1}/d\theta, \tag{A.8}$$

where $dz_1/d\theta = 0$ and $d^2z_1/d\theta^2 = 0$ by virtue of the definition of z_1 (incidentally $d^2z_2/d\theta^2 = 0$ also holds). On setting $\theta^0 = 1$ we have

$$dz_t/d\theta = \sum_{m=1}^{t-1} z_m \qquad t \geq 2,$$

$$d\mathbf{z}_t/d\theta = \left[\sum_{m=1}^{t-1} z_m, \ldots, \sum_{m=1}^{t-p} z_m\right]' \qquad t \geq (p+1),$$

$$d^2z_t/d\theta^2 = 2\sum_{m=2}^{t-1}\sum_{n=1}^{m-1} z_n \qquad t \geq 3,$$

with $dz_1/d\theta = d^2z_1/d\theta^2 = d^2z_2/d\theta^2 = 0$. We require expressions for $\Phi(L)(dz_t/d\theta)$ when $t \geq (p+1)$. For $t = (p+1)$ we may show $\Phi(L)(dz_t/d\theta) = f_p$ where f_p depends only on $z_1, \ldots, z_p$. For $t > (p+1)$ from (A.7) with $\theta^0 = 1$ we obtain

$$\begin{aligned}\Phi(L)dz_t/d\theta &= \Phi(L)dz_{t-1}/d\theta + \Phi(L)z_{t-1} \\ &= \Phi(L)dz_{p+1}/d\theta + \sum_{m=p+1}^{t-1}\Phi(L)z_m \\ &= f_p + \sum_{m=p+1}^{t-1}\varepsilon_m. \end{aligned} \tag{A.9}$$

Similarly, we need $\Phi(L)d^2z_t/d\theta^2$ for $t \geq (p+1)$. We may show that $\Phi(L)d^2z_{p+1}/d\theta^2 = g_{p-1}$, say, where g_{p-1} only depends on $z_1, \ldots, z_{p-1}$ and also that $\Phi(L)d^2z_{p+2}/d\theta^2 = h_p$ where h_p only depends on $z_1, \ldots, z_p$. For $t > (p+2)$, from (A.8) with $\theta^0 = 1$ we have

$$\begin{aligned}\Phi(L)d^2z_t/d\theta^2 &= \Phi(L)d^2z_{t-1}/d\theta^2 + 2\Phi(L)dz_{t-1}/d\theta \\ &= \Phi(L)d^2z_{p+2}/d\theta^2 + 2\sum_{m=p+2}^{t-1}\Phi(L)dz_m/d\theta \\ &= h_p + 2\sum_{m=p+2}^{t-1}\left(f_p + \sum_{n=p+1}^{m-1}\varepsilon_n\right).\end{aligned}$$

On substituting, we obtain

$$\begin{aligned}T^{-2}l_{\theta\theta} = &-\sigma^{-2}T^{-2}\sum_{t=p+2}^{T}\left(f_p + \sum_{m=p+1}^{t-1}\varepsilon_m\right)^2 \\ &-\sigma^{-2}T^{-2}\sum_{t=p+3}^{T}\varepsilon_t\left[h_p + 2\sum_{m=p+2}^{t-1}\left(f_p + \sum_{n=p+1}^{m-1}\varepsilon_n\right)\right] + o_p(1),\end{aligned}$$

https://doi.org/10.1017/S0266466698143025 Published online by Cambridge University Press

where the various end-effect terms for $t = p + 1, p + 2$ have been subsumed into the $o_p(1)$ term. Further calculation of the relevant orders then establishes that

$$T^{-2} l_{\theta\theta} = -\sigma^{-2} T^{-2} \sum_{t=p+2}^{T} \left(\sum_{m=p+1}^{t-1} \varepsilon_m \right)^2 - \sigma^{-2} 2T^{-2} \sum_{t=p+3}^{T} \varepsilon_t \left[\sum_{m=p+2}^{t-1} \left(\sum_{n=p+1}^{m-1} \varepsilon_n \right) \right] + o_p(1). \quad \textbf{(A.10)}$$

A very similar analysis also shows that

$$T^{-1} l_{\phi\theta} = \sigma^{-2} T^{-1} \sum_{t=p+2}^{T} \mathbf{z}_{t-1} \sum_{m=p+1}^{t-1} \varepsilon_m + \sigma^{-2} T^{-1} \sum_{t=p+2}^{T} \varepsilon_t \mathbf{s}_{t-2} + \mathbf{o}_p(1), \quad \textbf{(A.11)}$$

where $\mathbf{s}_t = [\sum_{m=1}^{t} z_m, \ldots, \sum_{m=1}^{t-p+1} z_m]'$.

Now, using (A.4), (6), and (A.9) we have

$$T^{-1} l_{\theta} = -\sigma^{-2} T^{-1} \sum_{t=p+2}^{T} \varepsilon_t \sum_{m=p+1}^{t-1} \varepsilon_m + o_p(1). \quad \textbf{(A.12)}$$

Squaring (A.12) and taking expectations establishes (10) whereas (11) follows from (A.10). Finally, (12) is immediate from (A.11).

APPENDIX B

That the first p elements of $\tilde{R}(\psi^0, \psi^+)$ are $\mathbf{o}_p(1)$ follows by virtue of standard arguments in the case of the AR(p) component.

As regards $\hat{R}(\psi^0, \psi^*)$, for $\psi = \psi^0$ we have

$$\hat{R}(\psi^0, \psi^*) = [\mathbb{J}(\psi^0, \psi^*) - \mathbb{J}(\psi^0, \psi^0)] \mathbb{E}_D(\psi^0)^{1/2} (\hat{\psi} - \psi^0). \quad \textbf{(A.13)}$$

The term in square brackets on the right-hand side of (A.13) is

$$\mathbb{E}_D(\psi^0)^{-1/2} [l_{\psi\psi}(\psi^*) - l_{\psi\psi}(\psi^0)] \mathbb{E}_D(\psi^0)^{-1/2}. \quad \textbf{(A.14)}$$

At this point, we examine (A.14) only for the case $p = 1$. The following results generalize to the ARMA(p,1) case but at the expense of substantial complication in the notation.

Consider first $l_{\phi\phi}(\phi^*, \theta^*)$, element (1,1) of $l_{\psi\psi}(\psi^*)$. In view of (A.3) this does not depend on ϕ (recall that z_t depends only on the true value ϕ^0), and so expanding around the true value θ^0 we have

$$T^{-1} l_{\phi\phi}(\phi^*, \theta^*) = T^{-1} l_{\phi\phi}(\phi^0, \theta^*) = T^{-1} l_{\phi\phi}(\phi^0, \theta^0) + T^{-1} l_{\phi\phi\theta}(\phi^0, \theta^-)(\theta^* - \theta^0). \quad \textbf{(A.15)}$$

From (A.3) (with $p = 1$),

$$T^{-1} l_{\phi\phi\theta} = -\sigma^2 T^{-1} \sum_{t=p+1}^{T} 2 z_{t-1} dz_{t-1}/d\theta. \quad \textbf{(A.16)}$$

https://doi.org/10.1017/S0266466698143025 Published online by Cambridge University Press

Evaluated at $\theta^- < 1$, z_{t-1} and $dz_{t-1}/d\theta$ are stationary in view of (A.1) (with y_t from (5), the true process) and (A.7). Thus, (A.16) converges to a constant. Now

$$E[z_{t-1}dz_{t-1}/d\theta] = E[z_{t-1}(\theta dz_{t-2}/d\theta + z_{t-2})]$$
$$= E[z_{t-1}z_{t-2} + \theta z_{t-1}z_{t-3} + \theta^2 z_{t-1}z_{t-4} + \cdots + \theta^{t-3}z_{t-1}z_1],$$

and so it is seen that the limiting constant for (A.16) is a function of the covariance process, γ_s, of z_t. By writing

$$(1 - \phi L)(1 - \theta^- L)z_t = (1 - L)\varepsilon_t,$$

it is seen that z_t is a stationary ARMA process for all values of θ^- in the range of interest, that is, $0 \leq \theta^- \leq 1$. Thus γ_s is a continuous function of θ^- as is the mean $E[z_{t-1}dz_{t-1}/d\theta]$. So, the limit of (A.16), being a continuous function on the closed interval $0 \leq \theta^- \leq 1$, is bounded. Hence, a bound has been established for Lim $T^{-1}l_{\phi\phi\theta}(\phi^0,\theta^-)$ when $\theta^- < 1$. Now, for $\theta^- = 1$, z_{t-1} is stationary because this is the true value of the MA(1) parameter and (6) holds; $dz_{t-1}/d\theta$ is then $I(1)$ and (A.16) is $O_p(1)$. So, (A.16) is $O_p(1)$ at most. Also, $(\theta^* - \theta^0) = o_p(1)$ because $(\hat{\theta} - \theta^0) = o_p(1)$ as shown in Potscher (1991). Hence, the second right-hand-side term in (A.15) is $o_p(1)$ and

$$T^{-1}l_{\phi\phi}(\phi^*,\theta^*) - T^{-1}\, l_{\phi\phi}(\phi^0,\theta^0) \to_d 0.$$

For the (2,2) element of $l_{\psi\psi}(\psi^*)$, expanding around the true value ϕ^0 gives

$$T^{-2}l_{\theta\theta}(\phi^*,\theta^*) = T^{-2}l_{\theta\theta}(\phi^0,\theta^*) + T^{-2}l_{\theta\theta\phi}(\phi^-,\theta^*)(\phi^* - \phi^0). \quad \textbf{(A.17)}$$

Analogous arguments to those made previously can be used to show that $T^{-2}l_{\theta\theta\phi}(\phi^-,\theta^*)$ is $O_p(1)$ at most. Again, from Potscher (1991) $(\phi^* - \phi^0) = o_p(1)$ so that the second right-hand-side term of (A.17) is $o_p(1)$. The first right-hand-side term of (A.17) behaves exactly like its counterpart from a pure MA(1) process for any true value of the AR(1) parameter ϕ^0. We may appeal to Davis and Dunsmuir (1996) to show

$$T^{-2}l_{\theta\theta}(\phi^0,\theta^*) - T^{-2}l_{\theta\theta}(\phi^0,\theta^0) = Op(1),$$

yet

$$T^{-2}l_{\theta\theta}(\phi^*,\theta^*) - T^{-2}l_{\theta\theta}(\phi^0,\theta^+) \to_d 0.$$

In the case of element (1,2) of $l_{\psi\psi}(\psi^*)$, we find from a similar analysis that

$$T^{-3/2}l_{\phi\theta}(\phi^*,\theta^*) - T^{-3/2}l_{\phi\theta}(\phi^0,\theta^+) \to_d 0.$$

These results (when generalized to the ARMA$(p,1)$ case) combined with the known orders of $\mathbb{E}_D(\psi^0)$ in (A.14) establish that

$$\mathbb{J}(\psi^0,\psi^*) - \mathbb{J}(\psi^0,\psi^+) \to_d \mathbf{0}. \quad \textbf{(A.18)}$$

Finally, as $\mathbb{J}(\psi^0,\psi^0) = \mathbf{O}_p(1)$, (1) and (A.18) imply $\mathbb{E}_D(\psi^0)^{1/2}(\hat{\psi} - \psi^0) = \mathbf{O}_p(1)$. Thus, $R_c = \mathbf{o}_p(1)$.

https://doi.org/10.1017/S0266466698143025 Published online by Cambridge University Press