TESTING FOR AUTOCORRELATION IN THE AUTOREGRESSIVE MOVING AVERAGE ERROR MODEL

John FITTS*

University of Michigan, Ann Arbor, U.S.A.

1. Introduction

Failure to allow for autocorrelation of the disturbances in a regression model can lead to biased and inconsistent parameter estimates, particularly if the model is autoregressive. While consistent estimation methods are available which allow for autocorrelation, estimation is usually much easier when there is some assurance that autocorrelation is absent. In pursuit of such assurance the present paper develops methods of testing for autocorrelation in two common autoregressive linear regression models.

The first model considered is the autoregressive moving average error model (ARMA). It occurs when a first order (two period) moving average of structural disturbances generates the error term of an equation which is to be estimated, and this moving average autocorrelation is taken into account in the process of estimation. A test is developed for the presence of additional Markov autocorrelation of the structural disturbances. The second model considered is the standard autoregressive linear regression model, with a disturbance term which for the purposes of estimation is taken to be serially uncorrelated. Estimation is therefore performed by means of ordinary least squares. A test is developed for the presence of moving average autocorrelation in the disturbance term.

In keeping with tradition, both tests are performed by examining the serial correlation of the estimated disturbances. The tests derive from a

recent paper by J. Durbin (1970), in which he introduces a general method which can be used to obtain the asymptotic distribution of certain statistics based upon the estimated disturbances from maximum likelihood estimation of models such as those considered here. Durbin applies this method .to the problem of testing for the presence of Markov autocorrelation in the disturbance term of the standard autoregressive model. The present paper imitates his method and extends his result in two directions: first to a more general model and then to a different type of autocorrelation in the standard model.

The organization of the paper is as follows. Sect. 2 presents the the ARMA model, so specified as to allow for the possibility of additional Markov autocorrelation. It then obtains the asymptotic distribution of the derivative of the log-likelihood function of this model, a necessary preliminary to the application of Durbin's method. Sect. 3 discusses a frequently encountered economic context in which this model may arise; model transformation (derivation of an estimable reduced form from the structural model) induces moving average autocorrelation in addition to whatever autocorrelation governs the structural disturbance term. It then derives a test for such original autocorrelation. Finally, it discusses some special cases and a generalization of the model, and adds some remarks about the power of the new test. Sect. 4 develops a test for moving average autocorrelation analogous to Durbin's test for Markov autocorrelation. Sect. 5 comments on these results and provides a summary of the computational steps necessary to carry out the two tests.

2. The ARMA model

Consider the linear autoregressive regression model

$$y = \gamma L y + \beta x + u,$$

in which the error term u exhibits compound serial correlation formed from the convolution of first order moving average and Markov processes,

$$u_t = \sum_{\tau=0}^{\infty} \rho^{\tau}(\epsilon_{t-\tau} - \alpha \, \epsilon_{t-\tau-1}).$$

It is notationally convenient to employ the lag operator L (not a likelihood function). Observations and errors y, x, u and ϵ are T-element column vectors, and L may be thought of as the $T \times T$ matrix with ones immediately below the main diagonal and zeroes elsewhere. It is also convenient to adopt the notation $M_\delta = I - \delta L$ for any scalar δ, where I is the $T \times T$ identity matrix. Then the above error process may be rewritten

$$u = M_\alpha M_\rho^{-1} \epsilon;$$

the reader should see that this is equivalent.[1]

To ensure stability of the autoregressive and error processes it is assumed that $|\rho| < 1$, $|\alpha| < 1$ and $|\gamma| < 1$; also ϵ is assumed to be distributed $N(0, \sigma^2 I)$ and x and ϵ are assumed to be asymptotically uncorrelated, i.e.,

$$\operatorname*{plim}_{T \to \infty} (1/T) x' L^\tau \epsilon = 0; \qquad \text{for any } \tau.$$

The following expressions involving the autocovariances of x are used in the sequel and are assumed to exist:[2]

$$\lim_{T \to \infty} (1/\sigma^2 T) x' L^\tau x = r_\tau,$$

$$\sum_{\tau=1}^\infty \delta^\tau r_\tau = \lim_{T \to \infty} (1/\sigma^2 T) x' \delta L M_\delta^{-1} x = q_\delta; \qquad \text{for any } \delta, |\delta| < 1.$$

This is the model to which we wish to apply Durbin's method. First, however, we obtain the asymptotic distribution of the derivative of its log-likelihood function. That this distribution has the form given below,

---

[1] For a discussion of the lag operator L see Dhrymes (1971) ch. 2. When L is formally defined as a $T \times T$ matrix certain of the formulas in the text are only asymptotic approximations, since $L^T = 0$. This does not affect the results. Note also that $L^0 = I$, and that the transpose L' of L may be regarded as the lead operator.

[2] Incidentally, the reader should observe the similarity of the above autocovariance expressions in x to the correction factor which must be applied to the conventional formula for the variance of ordinary least squares coefficient estimates in the presence of serial correlation in the error term. Thus if such serial correlation takes the commonly assumed form $u = M_\rho^{-1} \epsilon$, then in the model $y = \beta x + u$,

$$\operatorname{var}(\beta) = (r_0 + 2q_\rho)/r_0^2.$$

in eq. (1) and the preceding statement, is one of several conditions given by Durbin as jointly sufficient to the applicability of his method. Also, use of the method to derive actual test statistics requires that (1) be evaluated in terms of the parameters of the model; this is done in (4) below. The other conditions listed by Durbin (1970) are not considered explicitly here; they are satisfied by the standard autoregressive model and can be shown to be satisfied as well by its generalization considered here.

To obtain the desired asymptotic distribution we proceed as follows. The likelihood function for y, conditional upon pre-sample values, is $L(\omega, \sigma^2) = (2\pi\sigma^2)^{-T/2} \exp(-\epsilon'\epsilon/2\sigma^2)$, where $\omega = (\rho, \alpha, \beta, \gamma)'$. Here ϵ should be thought of as expressed in terms of the observations. This is done by inverting the lag structure of the model, yielding $\epsilon = M_\rho M_\alpha^{-1} M_\gamma y - \beta M_\rho M_\alpha^{-1} x = y - M_\alpha^{-1} Z v$, where $Z = (z_1, z_2, z_3) = (M\rho Ly, M\rho x, Ly)$ and $v = (v_1, v_2, v_3)' = (\gamma, \beta, \rho-\alpha)'$. The partial derivatives of ϵ are

$$\partial\epsilon/\partial\gamma = -M_\alpha^{-1} z_1,$$

$$\partial\epsilon/\partial\beta = -M_\alpha^{-1} z_2,$$

$$\partial\epsilon/\partial\alpha = -M_\alpha^{-1} z_3 \, \partial v_3/\partial\alpha - (\partial M_\alpha^{-1}/\partial\alpha) Z v$$

$$= M_\alpha^{-1} Ly - M_\alpha^{-2} LZv$$

$$= M_\alpha^{-1} L\epsilon,$$

$$\partial\epsilon/\partial\rho = -M_\alpha^{-1} z_3 \, \partial v_3/\partial\rho - M_\alpha^{-1} (\partial Z/\partial\rho) v$$

$$= -M_\alpha^{-1} Ly + \gamma M_\alpha^{-1} L^2 y + \beta M_\alpha^{-1} Lx$$

$$= -M_\alpha^{-1} Lu,$$

and therefore

$$(1/\sqrt{T}) \partial \log L/\partial\omega = -(1/\sqrt{T\sigma^2}) (\partial\epsilon'/\partial\omega)\epsilon$$

$$= \frac{1}{\sqrt{T\sigma^2}} \begin{pmatrix} u'L' \\ -\epsilon'L' \\ z_2' \\ z_1' \end{pmatrix} M_\alpha'^{-1} \epsilon.$$

It follows from a variant of the central limit theorem that this vector has an asymptotic normal distribution with mean zero and covariance matrix Ω/σ^2, where

$$\Omega = \sigma^2 \operatorname{plim}(1/T)(-\partial^2 \log L/\partial\omega\partial\omega') \qquad [3] \tag{1}$$

$$= \operatorname{plim}(1/T)(\partial\epsilon'/\partial\omega)(\partial\epsilon/\partial\omega'). \tag{2}$$

The second equality is a consequence of the fact that $\operatorname{plim}(1/T) \times (\partial^2\epsilon'/\partial\rho\partial\omega)\epsilon$ and similar terms involving second partials with respect to α, β, and γ vanish.[4] Let $\omega^* = (\rho^*, \alpha^*, \beta^*, \gamma^*)'$ denote the maximum likelihood estimator of ω; it is well known that asymptotically

$$\sqrt{T}(\omega^* - \omega) \sim N(0, \sigma^2\Omega^{-1}). \tag{3}$$

It will be sufficient for subsequent developments and notationally less cumbersome to express Ω in terms of the parameters of the model under the assumption that $\rho = 0$; i.e., that there is no Markov autocorrelation in the error term. Then $u = M_\alpha\epsilon$, and

$$\Omega/\sigma^2 = \begin{bmatrix} 1 & -1 & 0 & 1 \\ -1 & 1/(1-\alpha^2) & 0 & -1/(1-\alpha\gamma) \\ 0 & 0 & \theta/(1-\alpha^2) & \psi/(1-\alpha^2) \\ 1 & -1/(1-\alpha\gamma) & \psi/(1-\alpha^2) & (1+\phi)/(1-\gamma^2) \end{bmatrix}, \tag{4}$$

[3] Actually there is a slight discrepancy between Durbin's condition (vi) and this statement, in that the latter omits consideration of $\partial \log \mathcal{L}/\partial\sigma^2$. That this does not invalidate the subsequent results, which imitate this omission, follows from the fact that in this case the information matrix is block diagonal, or in other words that the maximum likelihood estimator of σ^2 is uncorrelated with those of ρ, α, β, and γ. This can be shown by an argument analogous to that applied by Theil (1971) p. 419, to a model which is equivalent to the present model if ρ is known to be zero and $\alpha = \gamma$. The same circumstance occurs in Durbin (1970).

[4] The derivation in this paragraph closely imitates Dhrymes (1971): see the proof of his eq. (7.192). (Note that this equation inadvertently omits the factor $1/\sqrt{T}$ on the left hand side. Also Dhrymes' (7.189) is misprinted and should read $\Pr\{|w_t^N - w_t| > \delta_1\} < \delta_2$, $\Pr\{|v_{t1}^N - v_{t1}| > \delta_1\} < \delta_2$.) The central limit theorem referred to in the text is Dhrymes' theorem 4.3. In this case it guarantees that the covariance matrix of the limiting distribution of $(1/\sqrt{T})\partial \log \mathcal{L}/\partial\omega$ is the limit of the finite sample covariance matrix $-(1/T)\partial^2 \log \mathcal{L}/\partial\omega\partial\omega'$. This however is equal to the probability limit of $-(1/T)\partial^2 \log \mathcal{L}/\partial\omega\partial\omega'$ since the elements of the latter matrix have variances $0(1/T)$ and so converge in probability to their expectations. The hypothesis of the central limit theorem requires only that the elements of ϵ be identically and independently distributed random variables; consequently the test statistics derived in the sequel remain valid on this weaker assumption, although their efficiency is then no longer guaranteed.

where

$$\theta = r_0 + 2q_\alpha,$$

$$\psi = [\alpha\beta/(1-\alpha\gamma)]\ (r_0 + r_1/\alpha + q_\alpha + q_\gamma) + \beta(q_\alpha - q_\gamma)/(\alpha-\gamma),$$

$$\phi = [\beta^2/(1-\alpha^2)(1-\alpha\gamma)]\ [(1+\alpha\gamma)r_0 + 2\alpha(1-\gamma^2)q_\alpha/(\alpha-\gamma)$$
$$+ 2\gamma(1-\alpha^2)q_\gamma/(\gamma-\alpha)].[5]$$

3. Testing for Markov autocorrelation in the ARMA model

This section will develop a test of the null hypothesis $H^+ : \rho = 0$. The motivation for this test arises from certain economic models, such as the Houthakker–Taylor dynamic demand model, which in order to eliminate unobservable variables are subjected to a transformation which induces first order moving average serial correlation in addition to whatever autocorrelation may have been present in the original error term. One might then wish to test whether such autocorrelation had indeed originally been present. This problem is touched upon by Houthakker and Taylor (1970) p. 34, and by Dhrymes and Mitchell in a recent paper (1972) p. 7. The latter authors propose that the dynamic demand model should be estimated by maximum likelihood methods, and give Monte Carlo results indicating that this "... method of estimation is vastly superior...". Aigner (1971) reviews the literature on maximum likelihood estimation in the ARMA model and refers to a number of other Monte Carlo studies which draw generally similar conclusions. Dhrymes and Mitchell further conjecture that a test of H^+ could then be based on the Durbin–Watson statistic, or equivalently upon the statistic $\hat{\rho}$ defined below. The virtue of this procedure is that it is based on the vector of estimated disturbances which arises from the routine maximum likelihood estimation of the model when the possibility of underlying, non-transformation induced autocorrelation is ignored. Dhrymes and Mitchell do not derive the asymptotic distribution of $\hat{\rho}$; the present paper applies Durbin's method to do so.

The test of H^+ is based on an estimate of the parameter ρ. How may

[5] The evaluation of (2), which yields (4), may be obtained by writing to the author at the RSQE, University of Michigan. It is straightforward and is carried out by expanding the various terms of (2) by partial fractions in L (see for example formula (8) of the appendix).

such an estimate be obtained? If α, β, and γ were known, it would be possible to calculate exactly $M_\rho^{-1}\epsilon = M_\alpha^{-1}u$. Denote the estimated first order autocorrelation coefficient of this vector by $\tilde{\rho} = u'M_\alpha'^{-1}LM_\alpha^{-1}u/$ $u'M_\alpha'^{-1}M_\alpha^{-1}u$; $\tilde{\rho}$ is the maximum likelihood estimator of ρ given α, β, and γ and is asymptotically normally distributed. Since the estimator ρ is based on knowledge of α, β, and γ, the reciprocal of the asymptotic variance of $\sqrt{T}(\tilde{\rho} - \rho)$ can be obtained by specializing (2) to the case in which only the single parameter ρ is to be estimated. Then Ω/σ^2 is the scalar $\mathrm{plim}(1/\sigma^2 T)\,(u'M_\alpha'^{-1}L'LM_\alpha^{-1}u) = 1/(1-\rho^2)$ which on H^+ is equal to unity (the $(1,1)$ element of (4)).

In the absence of knowledge of $M_\rho^{-1}\epsilon$ we use instead ϵ^+, the estimate of ϵ which arises from maximum likelihood estimation of α, β, and γ, say by α^+, β^+, and γ^+, when the estimation is done under the assumption that $\rho = 0$. Let $\hat{\rho}$ denote the statistic which arises from applying the formula for $\tilde{\rho}$ to ϵ^+ instead of to $M_\rho^{-1}\epsilon$. According to Durbin the asymptotic variance of $\hat{\rho}$ is smaller on H^+ than that of $\tilde{\rho}$ by a factor $v = 1 - A^{-1}CB^{-1}C'$, where A, B, and C are submatrices of Ω/σ^2 of order (1×1), (3×3), and (1×3), respectively, such that

$$\Omega/\sigma^2 = \begin{bmatrix} A & | & C \\ -- & + & -- \\ C' & | & B \end{bmatrix}.$$

Let $\xi = \phi - (1-\gamma^2)\psi^2/(1-\alpha^2)\theta$. Application of the rule for the inverse of a partitioned matrix to B and a bit of algebra yields the result

$$v = \alpha^2 \left(\frac{\gamma^2(\alpha-\gamma)^2 + (1-\alpha\gamma)^2\xi}{(\alpha-\gamma)^2 + (1-\alpha\gamma)^2\xi} \right). \tag{5}$$

A consistent estimate v of $\hat{v}$ may be calculated by replacing α, β and γ in (5) by their maximum likelihood estimates α^+, β^+ and γ^+ and σ^2 by $\epsilon^+{}'\epsilon^+/(T-3)$ and by using sample autocovariances for x. An asymptotic test of H^+ may be performed by computing the quantity $\hat{\rho}\sqrt{T/\hat{v}}$ and testing it as a standard normal deviate.

The remainder of this section is devoted to a series of remarks about this result. First we examine some special cases in which the formula for v is independent of ξ, and then the general case when β is a vector. Finally we add some reservations about the power of the test just developed.

If $\alpha = \gamma$ and this is taken into account in the estimation then $v = 1$; $\hat{\rho}$

has the same asymptotic distribution as does $\tilde{\rho}$. Except for the lag structure in x this is the standard non-autoregressive linear model. If on the other hand $\alpha = \gamma$ but this is not known a priori then $v = \alpha^2$. The reader should not confuse the case $\alpha = \gamma$ with the dynamic demand model in which typically $\alpha > \gamma$ for durable goods and $\alpha < \gamma$ for habit-forming goods.

If $\alpha = 0$, the problem degenerates as then $v = 0$. This is because asymptotically the maximum likelihood estimators of α and ρ in the full model are perfectly correlated when both parameters are actually zero. The likelihood function does not have a unique maximum with respect to α and ρ but is asymptotically maximized anywhere along the line $\alpha = \rho$. In other words, one can't test simple versus compound autocorrelation when there is none at all. On the other hand, if it is known a priori that $\alpha = 0$ the model becomes precisely that considered by Durbin (1970).

If $\beta = 0$, whether or not this is taken into account in the estimation, we have $\xi = 0$, and $v = \alpha^2 \gamma^2$. This result is in accord with the results of Box and Pierce (1970), who derive the joint asymptotic distribution of all (not only the first order) residual autocorrelations in the ARMA model when there is no exogenous variable x.

Interestingly, even if it is known a priori that $\gamma = 0$, the problem does not become trivial, as it does in the case considered in sect. 4 below and in the case dealt with by Durbin (1970). If the last row and column are deleted from (4) before $1 - A^{-1}CB^{-1}C'$ is calculated, the variance correction factor becomes $v = \alpha^2 < 1$. Even when none of the regressors are lagged dependent variables and the true value for ρ is zero, the maximum likelihood estimators of α and ρ are asymptotically correlated, which is hardly surprising. This example shows that Durbin's method has a wider potential range of application then simply to problems arising from a lagged dependent variable.

The above results may easily be generalized to allow for several exogenous variables, so that β is a vector of coefficients and X a matrix of observations. The scalar term θ which appears in (4) becomes the matrix

$$\theta = R_0 + Q_\alpha + Q'_\alpha, \quad \text{and}$$

$$\psi = [\alpha(R_0 + R_1/\alpha + Q'_\alpha + Q_\gamma)/(1-\alpha\gamma) + (Q_\alpha - Q_\gamma)/(\alpha - \gamma)]\beta,$$

$$\phi = \beta'[(1+\alpha\gamma)R_0 + 2\{\alpha(1-\gamma^2)Q_\alpha - \gamma(1-\alpha^2)Q_\gamma\}/(\alpha-\gamma)]$$

$$\times \beta/(1-\alpha^2)(1-\alpha\gamma),$$

$$\xi = \phi - (1-\gamma^2)\psi'\theta^{-1}\psi/(1-\alpha^2).$$

Finally, it is likely that the power of a test based on $\hat{\rho}$ is low. This is a well known problem with the statistic derived by Durbin (1970). The trouble is due in part to the likelihood that plim $\hat{\rho} < \rho$ when $\rho > 0$, and vice-versa. Analysis of the conditions under which this is true in the full present model is difficult, but it is intuitively clear how it may occur. Consider the estimated residuals

$$\epsilon^+ = M_{\alpha^+}^{-1}(M_{\gamma^+}y - \beta^+ x)$$

$$= M_{\alpha^+}^{-1}M_\alpha M_\gamma^{-1}M_{\gamma^+}M_\rho^{-1}\epsilon + M_{\alpha^+}^{-1}(\beta M_\gamma^{-1}M_{\gamma^+} - \beta^+ I)x.$$

The estimates α^+, β^+, and γ^+ are chosen so as to minimize $\epsilon^{+\prime}\epsilon^+$. There will be a tendency to chose $\gamma^+ > \gamma$ and $\alpha^+ < \alpha$ so as to compensate for the factor M_ρ^{-1} in the first term after the second equality above. Then ϵ^+ may exhibit less serial correlation than $M_\rho^{-1}\epsilon$, which it is used to estimate in the calculation of $\hat{\rho}$; this depends, however, on the structure of the serial correlation in x. In the special case $\gamma = 0$ it can be shown that plim $\alpha^+ < \alpha$ and plim $\hat{\rho} < \rho$ when $\rho > 0$; see the appendix. In the special case $\alpha = 0$ the bias of plim γ^+ and plim $\hat{\rho}$ are well known. It is unlikely that the combined effect of these two biases in the full model is other than mutually reinforcing.

4. Testing for moving average autocorrelation

In the model of the previous sections, the presence of moving average autocorrelation was assumed, and further autocorrelation was tested. This section takes as its null hypothesis instead the standard autoregressive model with a well behaved error term, and develops a test for moving average autocorrelation. Additional layers of autocorrelation are assumed not to exist. Thus the test developed here is precisely analogous to that developed by Durbin (1970), except that it tests for a different type of simple autocorrelation. The problem may be stated in terms of the model of sect. 2: we test the null hypothesis $H^\dagger : \alpha = \rho = 0$ against the alternative hypothesis $H^+ : \rho = 0$.

The subsequent development makes use of the notational apparatus of sect. 2, but it is assumed throughout that $\rho = 0$ and all terms referring to ρ are deleted. The inverse of the asymptotic covariance matrix of the maximum likelihood estimators α^+, β^+, and γ^+ is given by deleting the first row and column of (4) (the entries pertaining to ρ); under $H^\dagger$ this becomes

$$\Omega/\sigma^2 = \begin{bmatrix} 1 & 0 & -1 \\ 0 & \theta & \psi \\ -1 & \psi & (1+\phi)/(1-\gamma^2) \end{bmatrix}. \tag{6}$$

Let $\tilde{\alpha}$ denote the maximum likelihood estimator of α given the true values of β and γ, i.e., given the true disturbance vector $u = M_\alpha \epsilon$. Calculation of α is done by iteration so as to minimize $u' M_{\tilde{\alpha}}'^{-1} M_{\tilde{\alpha}}^{-1} u$. As before the reciprocal of the asymptotic variance of $\sqrt{T}(\tilde{\alpha} - \alpha)$ is obtained by specializing (2) to a one parameter case. This yields the expression $1/(1 - \alpha^2)$ which on $H^\dagger$ is equal to unity (the (1,1) element of (6)).[6] Let $\epsilon^\dagger$ denote the residuals which arise from maximum likelihood estimation of β and γ given that $\alpha = \rho = 0$; in this case $\epsilon^\dagger$ is just the ordinary least squares residual vector. Let $\hat{\alpha}$ denote the statistic which arises from applying the formula (algorithm) for $\tilde{\alpha}$ to $\epsilon^\dagger$ instead of to $M_\alpha \epsilon$. The asymptotic variance of $\hat{\alpha}$ is smaller on $H^\dagger$ than that of $\tilde{\alpha}$ by the factor

$$w = \frac{\gamma^2 + \xi}{1 + \xi} = \frac{\gamma^2 r_0 + \beta^2 [(r_0 + q_\gamma)^2 - q_\gamma^2 - (1-\gamma^2)(r_1 + q_\gamma/\gamma)^2]}{r_0 + \beta^2 [(r_0 + q_\gamma)^2 - q_\gamma^2 - (1-\gamma^2)(r_1 + q_\gamma/\gamma)^2]}. \tag{7}$$

A consistent estimate $\hat{w}$ of w may be calculated by replacing β, γ, and σ^2 in (7) by their conventional least squares estimates and by using sample autocovariances for x. An asymptotic test of $H^\dagger$ may then be based on the standard normal deviate $\hat{\alpha}\sqrt{T/\hat{w}}$.

As before we append some remarks about this result. In the special case $\beta = 0$ we have $w = \gamma^2$, which corresponds exactly to the result derived by Durbin for the case of Markov autocorrelation. In the case $\gamma = 0$, if this is not known a priori we do not have $w = 1$, so we cannot omit the correction factor even if the coefficient on the lagged dependent variable is insignificant. (A similar remark applies to the conventional Durbin–Watson statistic.) The proper procedure is to re-run the regression without the lagged dependent variable; then no correction is needed. Finally the power of this test, like that of sect. 3, is probably low.

---

[6] The similarity of the formulas for the asymptotic variances of $\tilde{\alpha}$ and $\tilde{\rho}$ can be explained by referring to (2) and noting that $\partial\epsilon/\partial\rho = -M_\alpha^{-1}Lu = -M_\rho^{-1}L\epsilon$ while $\partial\epsilon/\partial\alpha = M_\alpha^{-1}L\epsilon$.

Table 1

Steps in computing $\hat{\rho}$ and $\hat{\alpha}$.

	$\hat{\rho}$	$\hat{\alpha}$
Stage 1		
Estimate the model	$M_\gamma y = \beta x + M_\rho^{-1} M_\alpha \epsilon$	$M_\gamma y = \beta x + M_\alpha \epsilon$
under the null hypothesis	$H^+ : \rho = 0$	$H^\dagger : \alpha = 0$
by minimizing	$\epsilon^{+\prime}\epsilon^+ = (T-3)\sigma^{2+} = (M_{\gamma+}y - \beta^+ x)' M_{\alpha+}'^{-1} M_{\alpha+}^{-1}(M_{\gamma+}y - \beta^+ x)$ a	$\epsilon^{\dagger\prime}\epsilon^\dagger = (T-2)\sigma^{2\dagger} = (M_{\gamma\dagger}y - \beta^\dagger x)'(M_{\gamma\dagger}y - \beta^\dagger x)$ b
Stage 2		
Use the stage 1 residuals	ϵ^+ to estimate $M^{-1}\epsilon$	$\epsilon^\dagger$ to estimate $M_\alpha \epsilon$
and minimize	$\epsilon^{+\prime} M_{\hat\rho}' M_{\hat\rho} \epsilon^+$ b	$\epsilon^{\dagger\prime} M_{\hat\alpha}^{-1} M_{\hat\alpha}^{-1} \epsilon^\dagger$ a
Stage 3		
Use sample autocovariances	$\hat{r}_\tau = x'L^\tau x/\sigma^{2+}(T-\tau)$	$\hat{r}_\tau = x'L^\tau x/\sigma^{2\dagger}(T-\tau)$
to estimate variance	$\hat{v}$ from eq. (5)	$\hat{w}$ from eq. (7)
Test standard normal deviate	$\hat{\rho}\sqrt{T/\hat{v}}$	$\hat{\alpha}\sqrt{T/\hat{w}}$

a Estimate by an iterative search for the extremal value of α.
b Estimate by ordinary least squares.

5. Summary and conclusion

Table 1 summarizes the steps involved in computing the two statistics derived above.

Both the statistics $\hat{\rho}$ and $\hat{\alpha}$ require the use of an iterative computing routine for least squares estimation of α; for a discussion of the basic idea behind such procedures see Theil (1971) sect. 8.8. There are three stages in the computation of $\hat{\rho}$ or $\hat{\alpha}$. Stage (1) estimates the model under the null hypothesis which is to be tested. Stage (2) drops this null hypothesis and estimates the suspected autocorrelation parameter from the estimated disturbances obtained in stage (1). Stage (3) calculates an estimate of the variance correction factor by substituting the stage (1) parameter estimates and the sample autocovariances of x into the appropriate formula as given in the text. In both cases the variance by which this factor is multiplied is unity. The resulting statistic is then tested as a standard normal deviate. In many applications an a priori assumption of non-negativity for ρ makes a positive one-tailed test appropriate for $\hat{\rho}$. On the other hand a simple moving average error process (non-transformation induced) cannot give rise to positive serial correlation of the sort observed in most economic data unless α is negative.

Both statistics are only asymptotically valid, and probably suffer from low power. When moving average autocorrelation is *known* to be present the statistic $\hat{\alpha}$ is not appropriate. For that matter it is little easier to compute $\hat{\alpha}$ than it is to estimate the moving average model as in stage (1) of the computation of $\hat{\rho}$. The statistic derived by Durbin (1970) is probably highly correlated with $\hat{\alpha}$ and is much easier to calculate. On the other hand $\hat{\rho}$ is quite easy to calculate (and indeed can only be calculated) once the decision has been made to estimate the stage (1) model using the maximum likelihood methods appropriate to the presence of (transformation induced) moving average autocorrelation. The algebra involved in evaluating $\hat{v}$ would be cumbersome by hand but is readily mechanizable. It is just as dangerous to ignore the possibility of underlying Markov autocorrelation in the ARMA model as in the standard autoregressive model. Should it be found, remedies are available, although at some additional computational expense. Perhaps the availability of this statistic adds another advantage to those adduced elsewhere of maximum likelihood estimation in the moving average error model.

Appendix

Proof of asymptotic bias of $\hat{\rho}$ when $\gamma = 0$

Let $\gamma = 0$. Let $\bar{\alpha}$ be an arbitrary estimate of α such that $|\bar{\alpha}| < 1$, and let $\bar{\beta}$ be the corresponding maximum of the concentrated likelihood function $\bar{\beta}(\bar{\alpha}) = y'M_{\bar{\alpha}}^{\prime-1}M_{\bar{\alpha}}^{-1}x/x'M_{\bar{\alpha}}^{\prime-1}M_{\bar{\alpha}}^{-1}x$. Obviously plim $\bar{\beta} = \beta$ for any $\bar{\alpha}$. Note that $\beta^+ = \bar{\beta}(\alpha^+)$; β^+ is found by least squares applied to adjusted data. Let $\bar{\epsilon}(\bar{\alpha}) = M_{\bar{\alpha}}^{-1}(y - \bar{\beta}(\bar{\alpha})x)$. Then α^+ is the value of $\bar{\alpha}$ which minimizes $\bar{\epsilon}'\bar{\epsilon}$. Further, plim α^+ is the value of $\bar{\alpha}$ which minimizes plim $\bar{\epsilon}'\bar{\epsilon}/T$.
Now

plim $(\bar{\epsilon}'\bar{\epsilon}/T)$

$$
= \text{plim}\,(1/T)\,[(\beta - \bar{\beta})x + M_\rho^{-1}M_\alpha \epsilon]'M_{\bar{\alpha}}^{\prime-1}M_{\bar{\alpha}}^{-1}[(\beta - \bar{\beta})x + M_\rho^{-1}M_\alpha \epsilon]
$$

$$
= \text{plim}\,(1/T)\,\epsilon'M_\alpha'M_{\bar{\alpha}}'M_\rho^{\prime-1}M_\rho^{-1}M_{\bar{\alpha}}^{-1}M_\alpha \epsilon
$$

$$
= \text{plim}\,(1/T)\epsilon'\left(\frac{(1+\alpha^2)I - (L'+L)}{(1-\bar{\alpha})^2\,(1-\bar{\alpha}\rho)\,(1-\rho^2)}\right)
$$

$$
\times\left(-(1+\bar{\alpha}\rho)I + \rho\,\frac{(1-\bar{\alpha}^2)}{\rho - \bar{\alpha}}\,(M_\rho^{\prime-1} + M_\rho^{-1})\right.
$$

$$
\left.+ \frac{\bar{\alpha}(1-\rho^2)}{\bar{\alpha}-\rho}\,(M_{\bar{\alpha}}^{\prime-1} + M_{\bar{\alpha}}^{-1})\right)\epsilon \tag{8}
$$

$$
= \frac{\sigma^2(1+\alpha^2)}{(1-\bar{\alpha}^2)(1-\bar{\alpha}\rho)(1-\rho^2)}\left(-(1+\bar{\alpha}\rho) + \frac{2\rho(1-\bar{\alpha}^2)}{\rho-\bar{\alpha}} + \frac{2\bar{\alpha}^2(1-\rho^2)}{\bar{\alpha}-\rho}\right)
$$

$$
- \frac{\sigma^2\alpha}{(1-\bar{\alpha}^2)(1-\bar{\alpha}\rho)(1-\rho^2)}\left(\frac{2\rho^2(1-\bar{\alpha}^2)}{\rho-\bar{\alpha}} + \frac{2\bar{\alpha}^2(1-\rho^2)}{\bar{\alpha}-\rho}\right)
$$

$$
= \sigma^2\left(\frac{(1-\alpha^2 - 2\alpha\rho) + \bar{\alpha}(\rho - 2\alpha + \alpha^2\rho)}{(1-\bar{\alpha}^2)(1-\bar{\alpha}\rho)(1-\rho^2)}\right). \tag{9}
$$

The substitution involved in reaching the expression (8) is derived by expanding $M_\alpha'M_{\bar{\alpha}}^{\prime-1}M_\rho^{\prime-1}M_\rho^{-1}M_{\bar{\alpha}}^{-1}M_\alpha$ by means of partial fractions in L and L', using the asymptotic approximation $L'L \simeq I$. The expression (9)

can easily be shown to be positive for parameter values in the assumed range. Further analysis shows it to have a unique minimizing value of $\bar{\alpha}$ in the interval $(-1, 1)$, as two of its three stationary points must lie outside this interval. Differentiation with respect to $\bar{\alpha}$ and substitution of the value $\bar{\alpha} = \alpha$ yields the expression $2\rho/(1-\alpha\rho)$. This is positive when $\rho > 0$, implying that the minimizing value of $\bar{\alpha}$ must lie to the left of α; the opposite is true when $\rho < 0$. Denote the minimizing value of $\bar{\alpha}$ by plim $\alpha^+ = \zeta$. Now, plim $\epsilon^{+\prime}\epsilon^+/T = $ plim $\bar{\epsilon}'(\alpha^+)\bar{\epsilon}(\alpha^+)/T = $ plim $\bar{\epsilon}'(\zeta)\bar{\epsilon}(\zeta)/T$. Similarly

$$\text{plim } \epsilon^{+\prime}L\epsilon^+/T = \text{plim } \epsilon'M_\zeta'^{-1}M_\alpha'M_\rho'^{-1}M_\rho^{-1}M_\alpha M_\zeta^{-1}L\epsilon/T$$

$$= \sigma^2 \left(\frac{(-\alpha+\rho+\alpha^2\rho-\alpha\rho^2)+\zeta(1-2\alpha\rho+\alpha^2)+\zeta^2(-\alpha+\alpha\rho^2)}{(1-\zeta^2)(1-\zeta\rho)(1-\rho^2)} \right),$$

and thus

$$\text{plim } \hat{\rho} - \rho = \text{plim } \epsilon^{+\prime}L\epsilon^+/\epsilon^{+\prime}\epsilon^+ - \rho$$

$$= \frac{(1-\rho^2)(\zeta-\alpha)(1-\alpha\zeta)}{(1+\alpha^2-2\alpha\rho)+\zeta(\rho-2\alpha+\alpha^2\rho)},$$

which has the same sign as does $\zeta-\alpha$. Thus, if $\rho > 0$, plim $\hat{\rho} < \rho$, and vice-versa: $\hat{\rho}$ is asymptotically biased toward zero.

References

Aigner, D.J., 1971, A compendium on estimation of the autoregressive moving average model from time series data, International Economic Review 12, no. 3, 348–371.

Box, G.E.P. and D.A. Pierce, 1970, Distribution of residual autocorrelations in autoregressive-integrated moving average time series models, Journal of the American Statistical Association 65, no. 332, 1509–1526.

Dhrymes, P., 1971, Distributed lags (Holden–Day, San Francisco).

Dhrymes, P. and B.M. Mitchell, 1972, The econometrics of dynamic demand models, Discussion Paper no. 230 (The Wharton School, Jan. 1972).

Durbin, J., 1970, Testing for serial correlation in least squares regression when some of the regressors are lagged dependent variables, Econometrica 38, 410–421.

Houthakker, H.S. and L.D. Taylor, Consumer demand in the United States, 2nd ed. (Harvard University Press, Cambridge).

Theil, J., 1971, Principles of econometrics (Wiley, New York).