Journal of Econometrics 47 (1991) 333–357. North-Holland

Estimation of a linear regression model with stationary ARMA(p, q) errors*

Victoria Zinde-Walsh and John W. Galbraith
McGill University, Montreal, Quebec H3A 2T7, Canada

Received October 1987, final version received December 1989

It is well known that consistent estimation of a linear regression model with a stationary Gaussian ARMA process in the errors can be carried out by maximum likelihood or, alternatively, by two-stage procedures involving estimation of the nuisance parameters followed by feasible generalized least squares for the model parameters. We show that the estimators coincide up to $O_p(T^{-3/2})$ and derive the variance to $O(T^{-2})$, which up to terms of this order is the same for both estimators. Considering the form of the error covariance matrix for an ARMA(p, q) process allows us to examine a computationally convenient algorithm for estimation of the parameters of the regression model. Finally we provide a Monte Carlo comparison of the small-sample properties of OLS and two versions of the proposed estimator.

1. Introduction

This paper examines the estimation of a linear regression model with a stationary Gaussian ARMA process in the errors. It is well known that consistent estimators can be obtained by maximum likelihood, or alternatively by two-stage procedures involving estimation of the nuisance parameters followed by feasible generalized least squares for the coefficients of the model. As long as the nuisance parameters are consistently estimated to $O_p(T^{-1/2})$, the two-stage estimator coincides with the maximum-likelihood estimator to $O_p(T^{-1})$ and is second-order efficient. We propose an estimation algorithm (comprising estimation of the nuisance parameters in the ARMA process *and* estimation of the parameters of the regression model) which

*Research support from the NSERC to V. Zinde-Walsh is gratefully acknowledged. For research assistance the first author thanks Rosanna Giordano, Marc Bilodeau, and especially Quan Wen who was most helpful in deriving the expansions, derivatives, and other formulae presented here. The authors are grateful to Aman Ullah, Robin Carter, Steve Ambler, and an anonymous referee for their helpful comments.

uses the structure of the ARMA process to facilitate computation without compromising second-order efficiency for the two-stage estimators, and which is applicable to the general ARMA(p, q) case.

The method which we propose has a number of advantages over those in the existing literature. First of all, it is applicable to the estimation of a regression model with any stationary ARMA(p, q) process in the errors, in contrast with many existing methods which are applicable only to one such process [e.g., AR(1)] or one subclass of processes [e.g., AR(p)]; see Judge et al. (1985) for a survey. Second, the method here is based upon the exact expression or $[\mathrm{E}(uu')]^{-1}$, where u is the vector of errors in the regression model. Because this expression was not available for the general ARMA(p, q) until recently,[1] a number of computational methods have been based upon approximations to it. For example, the Cochrane and Orcutt estimator (see, e.g., Judge et al.) is based upon an approximation whereby the first observation is in effect ignored. The Prais–Winsten (1954) estimator uses the exact expression, but as with the Cochrane–Orcutt method is of course applicable only to the AR(1) case.

Another aspect of the second point is that the method presented here avoids the application of some computationally inconvenient methods which have been used in the previous literature. For the general ARMA(p, q) process, minimisation of $u'[\mathrm{E}(uu')]^{-1}u$ has had to be carried out using approximations; maximum-likelihood estimation for this case has also often involved approximation to the true likelihood, as well as the application of numerical optimisation algorithms to the likelihood function. For example, Pierce (1971) provides an approximation to the true ML estimates which yields a nonlinear problem. Harvey and Phillips (1979) use a Kalman filtering algorithm to derive an exact maximum-likelihood estimator for the parameters of the regression model and those of the error process, which requires that the likelihood be maximised by a numerical algorithm. The latter authors also provide an algorithm for computation of GLS estimates of the model parameters given values for the parameters of the error process.

The method here requires no numerical optimisation routines. Initial estimates of the parameters of the error process require only a small number of iterations of a straightforward sequence of steps; once these estimates have been obtained, the feasible GLS estimator of the parameters of the regression model then follows immediately, without further iteration and without inversion of matrices of order higher than max(p, q); some methods proposed in the previous literature for the regression model with MA errors require inversion of matrices of order T, where T is the sample size.

We derive the expansions of the simplified estimators of the ARMA parameters and the leading terms by which they differ from the maximum

[1] Zinde-Walsh (1988); Wise (1955) (an early example of special cases which have been found) gives the corresponding transformation matrix for the special case of AR(p) errors.

likelihood estimators; the difference is of order $O_p(T^{-1})$. We also obtain expansions for the estimators of the regression parameters to $O_P(T^{-3/2})$ and for their variances to $O(T^{-2})$. All of our formulae involve matrices the elements of which are either well-known or explicitly provided here or in a related paper [Zinde-Walsh (1988)].

In the expansions of the estimators of the model parameters and their variances, our paper generalizes the work of Ullah et al. (1983) and Magee (1985), who considered AR(1) errors. The expansions are explicitly derived here, and the results are consistent with Rothenberg (1984), but our approach takes advantage of important features of the ARMA error structure. For the estimation of the nuisance parameters we derive a simplified equation identical to the one solved by Godolphin (1984) for ARMA random variables. He provides a simple algorithm which, again, generally requires few iterations and avoids inversion of matrices of high order. We show that applying this algorithm to OLS residuals from the regression allows us to construct efficient two-stage estimators.

The paper has the following structure. Section 2 provides definitions and assumptions and describes some useful results concerning the structure of ARMA covariance matrices from Zinde-Walsh (1988). Section 3 derives expansions for maximum-likelihood and two-stage estimators and for their variances. Section 4 describes an algorithm for estimation and a simulation study of some properties of the estimators considered. Finally appendix 1 gives an expansion for GLS residuals, appendix 2 provides the matrices of the derivatives of the error covariance matrix with respect to the parameters of the ARMA process, and appendix 3 contains the proofs of each lemma and of Theorem 1.

2. Model, definitions, and assumptions

Consider the regression model

$$y = X\beta + u, \tag{2.1}$$

where y is a $T \times 1$ vector of observable random variables, X is a $T \times k$ matrix, β a $k \times 1$ vector of parameters, and u a $T \times 1$ random vector. We assume that u is generated by the stationary ARMA(p,q) process

$$P(B)u_t = Q(B)\varepsilon_t, \tag{2.2}$$

where B is the backshift operator; we also assume that the autoregressive polynomial

$$P(B) = I - \rho_1 B - \cdots - \rho_p B^p \tag{2.3}$$

is such that the latent roots μ_i of $x^p - \rho_1 x^{p-1} - \cdots - \rho_p = 0$ satisfy $|\mu_i| < 1$

(roots of the autoregressive part are outside the unit circle), the roots of the moving-average polynomial,

$$Q(B) = I - \theta_1 B - \cdots - \theta_q B^q, \tag{2.4}$$

cannot equal ± 1 and, to streamline the exposition, are further assumed to be outside the unit circle; the vector ε is such that $\mathrm{E}(\varepsilon) = 0$, and $\mathrm{E}(\varepsilon\varepsilon') = \sigma^2 I$.

The set $\{\varepsilon_t\}_{-\infty}^{\infty}$ forms the basis of a Hilbert space H_∞ of zero-mean random variables with the scalar product defined by

$$(Z, Y) = \mathrm{E}(ZY) \tag{2.5}$$

and the norm

$$\|Z\| = \left[\mathrm{E}(Z^2)\right]^{1/2} \tag{2.6}$$

for any $Z, Y \in H_\infty$.

The stationarity condition implies that u_t is a linear function of $\{\varepsilon_l\}_{-\infty}^{t}$; hence $\mathrm{E}(u_t^2) < \infty$ and so $u_t \in H_\infty$. We note that any linear function of $\{u_t\}_1^T$ (such as an OLS estimator, for example) belongs to H_∞ and that an arbitrary function of $\{u_t\}_1^T$ is a function on the product H_∞^T. Limits for sequences of random variables in H_∞ are taken with respect to the norm defined in (2.6).

Following Zinde-Walsh (1988) we can define the covariance matrix $\Sigma_\infty = \mathrm{E}(uu')$ as the matrix of the canonical mapping S: $H_\infty \to H_\infty^*$ of the Hilbert space into its dual. The matrix elements of Σ_∞ in terms of the parameters of the ARMA process are given in the same paper. For u, the infinite column vector $(u_t)_{-\infty}^{+\infty}$, we can write $u = \Sigma_\infty^{1/2}\varepsilon$. Denote the vector of combined autoregressive and moving-average parameters by τ, $\tau = (\tau_1, \ldots, \tau_m)$, where $m = p + q$, and denote by Σ_i the matrix with elements given by the derivatives of the elements of Σ_∞ with respect to τ_i, $\Sigma_i = \partial\Sigma_\infty/\partial\tau_i$. The formulae for these derivatives, which appear in some of the expansions below, are found in appendix 2.

In Zinde-Walsh (1988) the relationship between Σ_∞ for the infinite model in H_∞ and the finite (observable) portion of it is established. Consider $\pi\Sigma_\infty\pi$, the covariance matrix of a finite ARMA process corresponding to the infinite one which has Σ_∞ as covariance matrix. The elements of $\pi\Sigma_\infty\pi$ coincide with the corresponding elements in Σ_∞. We will therefore omit the π where the dimension of the matrix can easily be inferred from conformability. On the other hand, the matrix Σ_∞^{-1}, which represents the covariances of the inverse process,[2] which is also ARMA, does not provide matrix elements of $(\pi\Sigma\pi)^{-1}$, the inverse matrix for the finite process. The matrix $(\pi\Sigma\pi)^{-1}$ can

[2] If the original process is $P(B)u_t = Q(B)\varepsilon_t$, the inverse process refers to $Q(B)u_t = P(B)\varepsilon_t$.

be represented as

$$(\pi\Sigma\pi)^{-1} = \pi\Sigma_{\infty}^{-1}\pi + V, \tag{2.7}$$

where the elements of $\pi\Sigma_{\infty}^{-1}\pi$ equal the corresponding elements of Σ_{∞}^{-1}; the correction matrix V appears as a result of truncation of the infinite vector and has rank which depends upon p and q, but not T. This inverse is explicitly derived in Zinde-Walsh (1988). Here we will write Σ_{∞}^{-1} whenever $\pi\Sigma_{\infty}^{-1}\pi$ is implied, and pay special attention to the instances where $\Sigma_T^{-1} = (\pi\Sigma\pi)^{-1}$ appears.

We will assume below that the conditions necessary for the existence of $\sqrt{T}$-consistent maximum-likelihood estimators are satisfied; for related conditions see Amemiya (1985).

3. Second-order efficient estimators

In this section we are concerned with two methods of estimation of model (2.1). One is maximum likelihood and the other is a two-step procedure, the first step of which involves estimation of the parameters of the error process from the matrix Σ_{∞}^{-1} on the basis of the OLS residuals of (2.1), while the second step obtains the feasible GLS estimator for β. We show that the parameters of the error process obtained in the first step differ by $O_p(T^{-1})$ from the maximum-likelihood estimators; moreover the two methods provide estimators for β which are equally efficient to $O(T^{-2})$. We derive the variance of the estimators to $O(T^{-2})$.

Here we ignore σ^2; extending the system of first-order conditions for maximizing the log-likelihood to include the equation for σ^2, $T\sigma^2 = u'\Sigma^{-1}u$, or explicitly concentrating out σ^2, will not affect our results.

If we now assume that the infinite vector ε is distributed $N(0, \sigma^2 I)$, we can write the log-likelihood function as

$$\log L = \text{const} - \tfrac{1}{2}\log[\det \Sigma_T] - \tfrac{1}{2}u'\Sigma_T^{-1}u. \tag{3.1}$$

The first-order conditions are

$$\frac{\partial \log L}{\partial \beta} = 0 \quad \text{and} \quad \frac{\partial \log L}{\partial \tau} = 0$$

and are equivalent to

$$\beta^{\mathrm{ML}} = \left(X'\Sigma_T^{-1}X\right)^{-1}X'\Sigma_T^{-1}y, \tag{3.2a}$$

$$-\tfrac{1}{2}\left[\frac{\partial \det \Sigma_T}{\partial \tau_i}\right]\left[\det \Sigma_T\right]^{-1} - \tfrac{1}{2}u'_{\mathrm{ML}}\frac{\partial \Sigma_T^{-1}}{\partial \tau_i}u_{\mathrm{ML}} = 0, \qquad i = 1,\dots,m, \tag{3.2b}$$

$$u_{\mathrm{ML}} = y - X\beta^{\mathrm{ML}}. \tag{3.2c}$$

Again, we assume that maximum likelihood provides $\sqrt{T}$-consistent estimators; in other words, $\sqrt{T}(\beta^{\mathrm{ML}} - \beta)$ has a finite limit. The likelihood equations (3.2) are equivalent to polynomial equations in the unknowns τ and β. Therefore the solution will be an arbitrarily smooth function of the random variables almost everywhere in H_∞^T, for every X. Convergence of linear estimators (such as the OLS or GLS, as these belong to H_∞) is in mean square according to (2.6), which implies convergence in probability. We use the notation for orders in probability here for comparison with expansions in the existing literature, but convergence results relating to estimators in H_∞ hold in *mean square*. The estimator β^{ML} can be represented as

$$\beta^{\mathrm{ML}} = \beta + \beta_{-1/2}^{\mathrm{ML}} + \beta_{-1}^{\mathrm{ML}} + \mathrm{O}_P(T^{-3/2}), \tag{3.3}$$

where the terms in the expansion have subscripts corresponding to their orders in probability; i.e., $\operatorname{plim} T^{-\alpha}\beta_\alpha^{\mathrm{ML}}$ is bounded. Similarly τ^{ML}, the maximum-likelihood estimator of τ, can be written as

$$\tau^{\mathrm{ML}} = \tau + \tau_{-1/2}^{\mathrm{ML}} + \tau_{-1}^{\mathrm{ML}} + \mathrm{O}_P(T^{-3/2}). \tag{3.4}$$

We omit subscripts on τ when the statements made apply to each of the coordinates or to the whole vector of coordinates, and the context precludes any ambiguity.

Recall the representation (2.7). From it we have

$$\frac{\partial \Sigma_T^{-1}}{\partial \tau} = \frac{\partial \Sigma_\infty^{-1}}{\partial \tau} + \frac{\partial V}{\partial \tau}, \tag{3.5}$$

with $\partial V/\partial \tau$ a symmetric matrix of rank independent of T.

We will now show that solving equations in the first-order conditions (3.2) is not much different from solving an equation which ignores the determinant of Σ_T, replaces maximum-likelihood residuals by OLS residuals, and also ignores the low-rank component of the covariance matrix and considers only the matrix of the covariances of the inverse process. In fact, the difference between the equations is of $O_P(T^{-1})$. We use the following lemma:

Lemma 1. Consider symmetric matrices A_T of finite rank r, with eigenvalues uniformly bounded for any T. Then $T^{-1}u'A_T u$ is $O_p(T^{-1})$.

Proof. See appendix 3.

Consider now the following components of (3.2b):

$$D = \left[\frac{\partial \det \Sigma_T}{\partial \tau}\right][\det \Sigma_T]^{-1}, \tag{3.6a}$$

$$E = u'_{\mathrm{ML}}\frac{\partial \Sigma_\infty^{-1}}{\partial \tau}u_{\mathrm{ML}}, \tag{3.6b}$$

$$F = u'_{\mathrm{ML}}\frac{\partial V}{\partial \tau}u_{\mathrm{ML}}. \tag{3.6c}$$

In addition, denote $X(X'X)^{-1}X'$ by P, with $M = I - P$. It is shown in appendix 1 that

$$u_{\mathrm{ML}} = Mu + P(I - \Sigma_T^{-1})(M + P\Sigma_T^{-1})^{-1}Mu + \tilde{u}, \tag{3.7}$$

where Mu is the OLS residual, u_{OLS}. The elements of $\tilde{u}$ are $O_p(T^{-1/2})$.

Finally, denote the matrix $P(I - \Sigma_T^{-1})(M + P\Sigma_T^{-1})^{-1}$ by L.

Lemma 2. For the quantities defined above,

$$\begin{aligned}
&T^{-1}D = O_p(T^{-1}),\\
&T^{-1}E = O_p(1),\\
&T^{-1}F = O_p(T^{-1}),\\
&T^{-1}(u'ML'LMu) = O_p(T^{-1}),\\
&T^{-1}\left[u'M\left(L'\frac{\partial \Sigma_T^{-1}}{\partial \tau} + \frac{\partial \Sigma_T^{-1}}{\partial \tau}L\right)Mu\right] = O_p(T^{-1}).
\end{aligned} \tag{3.8}$$

Proof. See appendix 3.

Next, combining (3.7) and Lemma 2, we find that the left-hand side of (3.2b) (divided by T) can be represented as

$$-(2T)^{-1}\left[\frac{\partial \det \Sigma}{\partial \tau}\right][\det \Sigma]^{-1} - (2T)^{-1}u'_{\mathrm{ML}}\frac{\partial \Sigma_T^{-1}}{\partial \tau}u_{\mathrm{ML}}$$

$$= -(2T)^{-1}u'_{\mathrm{OLS}}\frac{\partial \Sigma_\infty^{-1}}{\partial \tau}u_{\mathrm{OLS}} + \mathrm{O}_P(T^{-1}). \tag{3.9}$$

It follows from the lemma that the first term on the left-hand side of (3.9) is $\mathrm{O}_P(T^{-1})$, the second is $\mathrm{O}_P(1)$, and the term on the right-hand side is $\mathrm{O}_P(1)$. Next, denote by τ^{MOM} the solution to

$$u'_{\mathrm{OLS}}\frac{\partial \Sigma_T^{-1}}{\partial \tau}u_{\mathrm{OLS}} = 0. \tag{3.10}$$

Assuming the existence of a maximum-likelihood consistent estimator and of a solution to (3.10) in its neighbourhood is sufficient in this case for the assumptions of Robinson (1988) to hold, ensuring the applicability of Theorem 1 of that paper. We therefore find that the solutions differ by $\mathrm{O}_P(T^{-1})$:

$$\tau^{\mathrm{MOM}} = \tau^{\mathrm{ML}} + \mathrm{O}_P(T^{-1}),$$

and so

$$\tau^{\mathrm{MOM}} = \tau + \tau_{-1/2}^{\mathrm{MOM}} + \mathrm{O}_P(T^{-1}) \quad \text{with} \quad \tau_{-1/2}^{\mathrm{MOM}} = \tau_{-1/2}^{\mathrm{ML}}.$$

The notation τ^{MOM} for the solution to (3.10) is motivated by the fact that (3.10) represents a condition on the moments; thus τ^{MOM} can be seen as a method of moments (MOM) estimator. Note that in (3.10) the conditions on the sample covariances are equivalent to Yule–Walker equations for the autoregressive parameters; for the moving-average parameters they give equations unlike the ones currently used by econometricians but which could prove to be useful in yielding simple estimation techniques [see, e.g., Godolphin (1984)]. We will discuss (3.10) in more detail in section 4. We now state the main result of this section regarding the maximum-likelihood estimator β^{ML} and the two-stage estimator, obtained by GLS after solving (3.10), which we denote by β^{MOM}.

Theorem 1. The estimators β^{ML} and β^{MOM} are equal up to $\mathrm{O}_P(T^{-3/2})$. Both estimators are unbiased, and both are equally efficient to $\mathrm{O}(T^{-2})$, i.e., $\mathrm{E}(\beta^{\mathrm{ML}}-\beta)(\beta^{\mathrm{ML}}-\beta)' = \mathrm{E}(\beta^{\mathrm{MOM}}-\beta)(\beta^{\mathrm{MOM}}-\beta)' + \mathrm{o}(T^{-2})$.

Furthermore, for $\hat{\beta}$ equal to either of β^{ML} or β^{MOM},

$$\mathrm{E}(\hat{\beta}-\beta)(\hat{\beta}-\beta)' = \sigma^2\Omega + \frac{2\sigma^2}{T^2}\sum_{i,j=1}^{m}\gamma_i^{-1}\gamma_j^{-1} \times\left(\operatorname{tr} M\Sigma_i^{-1}M\Sigma M\Sigma_j^{-1}M\Sigma\right)\Omega X'\Sigma_i^{-1} \times[\Sigma - X\Omega X']\Sigma_j^{-1}X\Omega + \mathrm{o}(T^{-2}), \tag{3.11}$$

where $\gamma_i = \lim_{T\to\infty}(1/T)\operatorname{tr}\Sigma(\partial\Sigma_i^{-1}/\partial\tau_i)$, Σ_i^{-1} is the derivative $\partial\Sigma_\infty^{-1}/\partial\tau_i$ and is as given in appendix 2, $\Omega = (X'\Sigma^{-1}X)^{-1}$ is $\mathrm{O}(T^{-1})$, and the second term in (3.11) is $\mathrm{O}(T^{-2})$.

Proof. The proof involves several steps: first we show that $\beta^{\mathrm{ML}} = \beta^{\mathrm{MOM}} + \mathrm{O}_P(T^{-3/2})$, next we derive the general form of the terms in the expansion of τ, then we expand β to include $\beta_{-3/2}$, and finally we demonstrate that some of the terms in the expansion of the variance of β vanish and that $\operatorname{var}\beta^{\mathrm{ML}} = \operatorname{var}\beta^{\mathrm{MOM}} + \mathrm{o}(T^{-2})$, and derive the expression for the term of $\mathrm{O}(T^{-2})$. Details are given in appendix 3.

Note that the expansions of the estimators do not require normality of the errors.

The expression for the second term in the variance can be simplified further. Recall that $\Sigma_i^{-1} = \partial\Sigma_\infty^{-1}/\partial\tau_i$, where τ_i is a parameter from either the autoregressive or the moving-average part. Denote the infinite covariance matrix for the autoregressive part by Σ_{A} and for the moving-average part by Σ_{M}; then $\Sigma_\infty = \Sigma_{\mathrm{A}}\Sigma_{\mathrm{M}} = \Sigma_{\mathrm{M}}\Sigma_{\mathrm{A}}$ where the product is taken as the product of infinite matrices. It is shown in the course of the proof of Theorem 1 (see appendix 3) that up to O(1),

$$\operatorname{tr}\left(M\Sigma_i^{-1}M\Sigma M\Sigma_j^{-1}M\Sigma\right)$$

$$= \operatorname{tr}\frac{\partial\Sigma_{\mathrm{A}}^{-1}}{\partial\tau_i}\Sigma_{\mathrm{A}}\cdot\frac{\partial\Sigma_{\mathrm{A}}^{-1}}{\partial\tau_j}\Sigma_{\mathrm{A}},$$

if both τ_i and τ_j are parameters of the AR part;

$$= \operatorname{tr}\frac{\partial\Sigma_{\mathrm{A}}^{-1}}{\partial\tau_i}\Sigma_{\mathrm{A}}\cdot\frac{\partial\Sigma_{\mathrm{A}}^{-1}}{\partial\tau_j}\Sigma_{\mathrm{M}}^{-1},$$

if τ_i is a parameter of the AR part and τ_j of the MA part;

$$= \operatorname{tr} \frac{\partial \Sigma_A^{-1}}{\partial \tau_i} \Sigma_M^{-1} \cdot \frac{\partial \Sigma_A^{-1}}{\partial \tau_j} \Sigma_M^{-1},$$

if both τ_i and τ_j are parameters of the MA part. (3.12)

These simpler expressions can be used in calculations without affecting the order of magnitude of the result.

Corollary 1. For an AR(1) error process,

$$\mathrm{E}(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = \sigma^2 \Omega + \frac{\sigma^2}{T}(1 - \rho^2)\Omega X' \Sigma_1^{-1}[\Sigma - X\Omega X']\Sigma_1^{-1} X\Omega + \mathrm{o}(T^{-2}),$$

where

$$\begin{aligned} \{\Sigma_1^{-1}\}_{ij} &= 2\rho \quad \textit{for} \quad i = j, \\ &= -1 \quad \textit{for} \quad |i - j| = 1, \\ &= 0 \quad\ \ \textit{otherwise}. \end{aligned}$$

Proof. This representation can be obtained from Corollary 2, noting that Σ_θ^{-1} is eliminated, and cancellation performed in the expression for Σ_ρ^{-1}, before θ is set to zero. It also follows, after some small notational changes, from Magee (1985).

Corollary 2. For an ARMA(1, 1) process in the errors,

$$\mathrm{E}(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = \sigma^2 \Omega + \frac{\sigma^2}{T}\{A\} + \mathrm{o}(T^{-2}),$$

where

$$\begin{aligned} A = \Omega X' \Big\{ &(1 - \theta^2) \cdot \Sigma_\theta^{-1}[\Sigma - X\Omega X']\Sigma_\theta^{-1} \\ &+ (1 - \rho^2) \cdot \Sigma_\rho^{-1}[\Sigma - X\Omega X']\Sigma_\rho^{-1} \\ &+ 2 \cdot \frac{(1 - \rho^2)(1 - \theta^2)}{(1 - \theta\rho)} \cdot \Sigma_\theta^{-1}[\Sigma - X\Omega X']\Sigma_\rho^{-1} \Big\} \cdot X\Omega, \end{aligned}$$

with

$$\Sigma_\theta^{-1} = -\rho\theta^{-2}I + \left[\frac{2\theta(1+\rho^2) - \rho(\theta^2 - \theta^{-2} + 4)}{(1-\theta^2)^2}\right] \cdot W_1 + \left[\frac{1+\rho^2 - \rho(\theta + \theta^{-1})}{(1-\theta^2)}\right] \cdot W_2,$$

$$\Sigma_\rho^{-1} = \left[\frac{2\rho - (\theta + \theta^{-1})}{(1-\theta^2)}\right] \cdot W_1 + \theta^{-1} \cdot I,$$

and finally,

$$\{W_1\}_{s,t} = \theta^{|s-t|} \quad \text{and} \quad \{W_2\}_{s,t} = |s-t| \cdot \theta^{|s-t|-1}.$$

Proof. Follows from (3.11) and (3.12); the details are available on request from the authors.

4. A simplified two-stage estimator

We can now summarise the algorithm leading to the two-stage (method-of-moments) estimator that we have proposed. The data-generating process is (2.1)–(2.2), and the model is presumed to be correctly specified (except insofar as there may be an extra variable in the model for certain values of the parameters below). The algorithm comprises the following stages:

(1) Calculate an initial OLS estimate of the parameter vector β in (2.1), and use this to form the vector of OLS-estimated residuals, $u_{\text{OLS}} = y - X\beta^{\text{OLS}}$.

(2) Calculate estimates of the elements of the vector τ of parameters of the ARMA process by solving (3.10) for τ^{MOM}. An explicit procedure for solving an equivalent equation is given by Godolphin (1984). The procedure uses sample serial correlations of the stochastic process (the OLS residuals, here) and involves, for an ARMA(p, q), the solution of a $p \times p$ system of linear equations (Yule–Walker) to obtain the autoregressive parameters, followed by an iterative procedure to obtain the moving-average parameters, each step of which solves a $q \times q$ system of linear equations. Convergence is typically achieved after a small number of iterations.

(3) Use the estimated parameter vector τ^{MOM} to form an estimate of $\Sigma_T^{-1} \equiv (\pi\Sigma_\infty\pi)^{-1}$. The true GLS estimate of the vector β is then

$$\beta^{\text{GLS}} = \left(X'\Sigma_T^{-1}X\right)^{-1} X'\Sigma_T^{-1}y;$$

the feasible GLS estimator β^{MOM} replaces Σ_T^{-1} with $\hat{\Sigma}_T^{-1}$, which can be calculated using the estimates τ^{MOM} from (2). In doing so we use the known structure of the covariance matrix of an ARMA process [see, e.g., Box and Jenkins (1970)], whereby elements of this matrix are expressed as functions of the individual τ_i. With this structure it is easy to form Σ_∞^{-1} (or $\pi\Sigma_\infty^{-1}\pi$), the *finite part* of the covariance matrix of the infinite inverse process (see footnote 1). To form the inverse of the covariance matrix of the *finite* process, $(\pi\Sigma_\infty\pi)^{-1}$, which is the quantity which exactly corresponds to Σ_T^{-1}, we must apply the low-rank correction referred to in (2.7). The elements of the matrices used to compute the correction V are given explicitly in Zinde-Walsh (1988).[3]

Because this correction is of rank independent of T, it will become irrelevant asymptotically. It is therefore natural to explore its importance in small samples; this is one object of the simulation study reported below.

Such simulation studies as exist concerning the properties of alternative estimators for the linear regression model with autocorrelated errors deal primarily with the case in which those errors follow an AR(1) process; see Rao and Griliches (1969), Spitzer (1979), and Krämer (1980). These authors generally find an efficiency improvement attributable to feasible GLS estimators such as that proposed here, *if* the errors in the regression model are sufficiently strongly autocorrelated [Krämer (1980) notes the importance of the presence or absence of a constant in the regression model for the results concerning relative efficiency, finding little improvement over OLS if the model contains a constant].

Our results are generally consistent with the suggestions of this literature concerning two-stage estimation procedures applicable in the AR(1) case. We compare the small-sample properties of three estimators of the vector β, corresponding to $\Sigma_T^{-1} \equiv I_T$ (the OLS estimator), $\Sigma_T^{-1} \equiv \pi\Sigma_\infty^{-1}\pi$ (the feasible GLS estimator without the low-rank correction), and finally $\Sigma_T^{-1} \equiv (\pi\Sigma_\infty\pi)^{-1}$ (feasible GLS *with* the correction V). In the AR(1) case, the low-rank correction corresponds to the difference between the Cochrane–Orcutt and Prais–Winsten estimators (the latter incorporating V).

The Monte Carlo results use the following process to generate the data, a special case of (2.1)–(2.2):

$$y_t = \beta_1 + \beta_2 x_t + u_t,$$

with

$$(1+\alpha B)u_t = (1+\theta B)\varepsilon_t, \quad \varepsilon_t \sim \mathrm{IN}(0,1).$$

[3]See also Errata, 1990.

Finally we need a generating process for the exogenous variable x_t, which we take to be AR(1) in order to allow us to select a stationary or nonstationary process with a single parameter. Hence

$$(1-\lambda B)x_t = \nu_t, \qquad \nu_t \sim \mathrm{IN}(0,1), \qquad \mathrm{E}(\nu_t \varepsilon_s) = 0, \qquad \forall\, t, s.$$

The parameter space is $\beta_1 = \{1\}$, $\beta_2 = \{1\}$; $\lambda = \{0, 0.5, 1.0\}$ and $(\alpha, \theta) = \{(-0.9, -0.5), (-0.9, 0.5), (0.4, -0.5), (-0.4, 0.5), (-0.5, 0), (0, 0.5), (0, 0)\}$; these values of α and θ are chosen to give a representative set of examples of possible autocorrelation structures in the errors of the regression model. We examine sample sizes $T = \{10, 20, 40\}$. For a given sample size, one set of random numbers per replication is used for all sets of parameter values.[4] The number of replications (samples) taken is typically very small in this literature because of the large execution times required for each feasible GLS estimation; by concentrating on a selection of representative cases we are however able to provide a somewhat larger number of samples.

Note finally that we examine the case in which the model (and data-generating process) contain a constant term, shown by Krämer (1980) to be the case most favourable to OLS relative to the two-stage procedures that he considered. Unlike Krämer, who treated the Cochrance–Orcutt and Prais–Winsten estimators for the AR(1) case, we do find for the ARMA(1, 1) a substantial efficiency improvement attributable to GLS *vis-à-vis* OLS. Krämer's measure of efficiency differs somewhat from our own and that of Rao–Griliches and Spitzer, however. We present several cases in tables 1 and 2.

Case 7 involves a white-noise error process so that OLS is the true GLS estimator; the sacrifice in efficiency implied by the use of the feasible GLS estimator is generally quite small. In cases 5 and 6, although one of the parameters has a true value of zero and the other has an absolute value of 0.5, the feasible GLS estimator is generally superior. Cases 2, 3, and 4 represent fairly substantial autocorrelation, and the superiority of the feasible GLS estimates is sometimes quite marked, particularly in estimating the coefficient on the exogenous variable. Case 1 is one in which there is positive autocorrelation in the AR part and negative in the MA, so that while each parameter is substantial, there is some offset between them.

In summary, a number of features of the results in tables 1 and 2 appear noteworthy. First, our results appear broadly consistent with those already cited concerning feasible estimators applicable to the AR(1) case, in that the

[4] In order to compare all points in the parameter space on the same random numbers, it is necessary to repeat a replication in which the estimation algorithm for the MA parameter fails for *any one* of the sets of parameters. Hence, and also to keep execution times manageable, we choose a coarse convergence criterion. If anything, this practice will reduce the estimated efficiencies of our estimators relative to OLS, as a finer criterion will tend to lead to more accurate parameter estimates when convergence is achieved.

Table 1

Efficiency of OLS relative to GLS (MOM) estimator *with* low-rank term.

Model: $Y = \beta_1 + \beta_2 X + u$.

$\mathrm{MSE}_{\mathrm{GLS+}}(\hat\beta_1)/\mathrm{MSE}_{\mathrm{OLS}}(\hat\beta_1)$; $\mathrm{MSE}_{\mathrm{GLS+}}(\hat\beta_2)/\mathrm{MSE}_{\mathrm{OLS}}(\hat\beta_2)$.

		λ					
		0.0		0.5		1.0	
Case	(α,θ)	β_1	β_2	β_1	β_2	β_1	β_2
		$T = 10$; 500 samples					
1	(−0.9, −0.5)	1.00	0.92	0.99	0.95	1.01	1.04
2	(−0.9, 0.5)	0.94	0.40	0.94	0.53	0.89	0.75
3	(0.4, −0.5)	0.59	0.50	0.51	0.46	0.67	0.61
4	(−0.4, 0.5)	0.94	0.63	0.94	0.71	0.87	0.86
5	(−0.5, 0.0)	0.98	0.88	0.99	0.92	0.98	1.00
6	(0.0, 0.5)	1.01	0.91	1.01	0.94	1.00	1.02
7	(0.0, 0.0)	1.02	1.07	1.02	1.03	1.07	1.08
		$T = 20$; 500 samples					
1	(−0.9, −0.5)	1.02	0.68	1.02	0.67	0.90	0.86
2	(−0.9, 0.5)	0.91	0.12	0.91	0.18	0.75	0.42
3	(0.4, −0.5)	0.64	0.38	0.57	0.37	0.55	0.38
4	(−0.4, 0.5)	0.96	0.44	0.92	0.54	0.79	0.79
5	(−0.5, 0.0)	0.98	0.79	0.96	0.84	0.90	0.97
6	(0.0, 0.5)	0.99	0.85	0.97	0.92	0.95	1.04
7	(0.0, 0.0)	1.00	1.10	1.00	1.06	1.00	1.00
		$T = 40$; 250 samples					
1	(−0.9, −0.5)	1.05	0.52	1.04	0.48	0.83	0.71
2	(−0.9, 0.5)	0.86	0.05	0.86	0.07	0.64	0.19
3	(0.4, −0.5)	0.76	0.31	0.71	0.35	0.72	0.59
4	(−0.4, 0.5)	0.97	0.30	0.94	0.39	0.64	0.65
5	(−0.5, 0.0)	0.98	0.63	0.97	0.69	0.79	0.85
6	(0.0, 0.5)	1.00	0.71	0.98	0.82	0.86	0.95
7	(0.0, 0.0)	1.00	1.03	1.00	1.03	1.00	1.03

feasible estimator is more efficient than OLS where there is substantial autocorrelation and in that the sacrifice made where there is none is reasonably small. Moreover, we find that the relative efficiency of the OLS estimator *in estimating the coefficient on the exogenous variable* x_t generally rises with the autocorrelation in that exogenous series. In estimating the constant, however, the relative efficiency of GLS appears to decline with λ. Over the range of 10 to 40 observations, the relative efficiency of OLS declines as sample size increases.

Finally, table 2 indicates that the magnitude of the improvement made by incorporating the low-rank correction [that is, using $\Sigma_T^{-1} \equiv (\pi\Sigma_\infty\pi)^{-1}$ rather than the approximation $\Sigma_T^{-1} \equiv \pi\Sigma_\infty^{-1}\pi$] makes a generally small improvement which is at its greatest when autocorrelation in the errors is greatest.

Table 2

Efficiency of GLS (MOM) estimator *without* low-rank term relative to GLS (MOM) estimator *with* low-rank term.

Model: $Y=\beta_1+\beta_2 X+u$.

$\mathrm{MSE}_{\mathrm{GLS}+}(\hat{\beta}_1)/\mathrm{MSE}_{\mathrm{GLS}-}(\hat{\beta}_1)$; $\mathrm{MSE}_{\mathrm{GLS}+}(\hat{\beta}_2)/\mathrm{MSE}_{\mathrm{GLS}-}(\hat{\beta}_2)$.

		λ					
		0.0		0.5		1.0	
Case	(α,θ)	β_1	β_2	β_1	β_2	β_1	β_2
		$T=10$; 500 samples					
1	$(-0.9,-0.5)$	0.99	0.99	1.00	0.98	1.02	0.98
2	$(-0.9,\ 0.5)$	1.04	0.90	1.03	0.93	1.06	0.94
3	$(\ 0.4,-0.5)$	0.91	0.94	0.91	0.94	0.98	0.95
4	$(-0.4,\ 0.5)$	0.96	0.98	0.97	0.98	0.99	0.97
5	$(-0.5,\ 0.0)$	0.98	0.99	0.98	0.99	1.00	0.99
6	$(\ 0.0,\ 0.5)$	0.95	1.00	0.96	1.00	0.99	0.99
7	$(\ 0.0,\ 0.0)$	0.98	1.00	0.99	1.00	1.00	0.99
		$T=20$; 500 samples					
1	$(-0.9,-0.5)$	0.97	0.97	0.96	0.95	1.03	1.00
2	$(-0.9,\ 0.5)$	1.06	0.77	1.06	0.79	1.17	0.86
3	$(0.4,-0.5)$	0.94	0.99	0.96	0.99	0.88	0.89
4	$(-0.4,\ 0.5)$	0.93	0.97	0.93	0.97	0.97	0.95
5	$(-0.5,\ 0.0)$	0.97	0.99	0.98	1.00	0.99	0.99
6	$(0.0,\ 0.5)$	0.96	0.99	0.98	1.00	0.98	0.98
7	$(0.0,\ 0.0)$	1.00	1.00	1.00	1.00	1.00	1.00
		$T=40$; 250 samples					
1	$(-0.9,-0.5)$	0.89	0.97	0.90	0.94	1.03	0.99
2	$(-0.9,\ 0.5)$	1.04	0.82	1.04	0.79	1.28	0.78
3	$(0.4,-0.5)$	0.97	0.96	0.98	0.96	0.90	0.91
4	$(-0.4,\ 0.5)$	0.88	1.00	0.89	0.98	0.93	0.96
5	$(-0.5,\ 0.0)$	0.98	1.00	0.98	1.00	1.00	0.99
6	$(0.0,\ 0.5)$	0.97	1.00	0.97	1.00	0.99	0.98
7	$(0.0,\ 0.0)$	1.00	1.00	1.00	1.00	1.00	1.00

We note an analogy to results concerning the approximation of exact likelihoods, in the context of estimation of ARMA processes, by the omission of low-rank terms; see again Godolphin (1984), who finds that the effect on estimates appears small even for small sample sizes. As there is a considerable saving in computation to be made here in ignoring the low-rank term, the results suggest that it may be sensible in many circumstances to do so.

It is important to emphasise in closing that rejection of the null of white-noise residuals in a regression model of course does not imply the correctness of any alternative ARMA process, nor that we should necessarily perform some procedure such as the proposed one which in effect allows us to model the errors and incorporate the information which they contain into the regression model. It may be that re-specification of the regression model will allow the investigator to eliminate the problem detected in a test for autocorrelation, obviating the need for modelling the error process. Nonetheless, in circumstances where this is unsuccessful, or in which autocorrelated errors are a plausible hypothesis, this method appears to be straightforward, applicable quite generally, and of potentially substantial benefit.

Appendix 1

The relation between the residuals from feasible GLS and OLS

Proposition. The following relationships hold among residuals:

$$u_{\mathrm{FGLS}} = u_{\mathrm{GLS}} + \tilde{u}_1,$$

$$u_{\mathrm{ML}} = u_{\mathrm{GLS}} + \tilde{u}_2,$$

$$u_{\mathrm{GLS}} = u_{\mathrm{OLS}} + P(I - \Sigma^{-1})(M + P\Sigma^{-1})^{-1} u_{\mathrm{OLS}},$$

with $u_{\mathrm{OLS}} = Mu$ *and with* $\tilde{u}_i = P(I - \Sigma^{-1})(M + P\Sigma^{-1})^{-1}\Sigma^{-1}_{-1/2}(M + P\Sigma^{-1})^{-1}Mu - P\Sigma^{-1}_{-1/2}(M + P\Sigma^{-1})^{-1}Mu + \tilde{\tilde{u}}$; $\Sigma^{-1}_{-1/2}$ *involves terms of* $\mathrm{O}_p(T^{-1/2})$ *in the expansions of the estimators of* τ, *which for* $i = 1$ *are the first-stage estimators and for* $i = 2$ *are ML estimators, and the elements of* $\tilde{\tilde{u}}$ *are of* $\mathrm{O}_p(T^{-1})$.

Proof. Denote $I - \Sigma^{-1}$ by D. From the restrictions on the roots of the AR, MA polynomials for the operator norm of D, the inequality $\|D\| < 1$ holds. We have

$$\begin{aligned}(X'\Sigma^{-1}X)^{-1} &= (X'X - X'DX)^{-1} \\ &= (X'X)^{-1}\left[I - X'DX(X'X)^{-1}\right]^{-1} \\ &= (X'X)^{-1}\Big[I + X'DX(X'X)^{-1} + \left[X'DX(X'X)^{-1}\right]^2 \\ &\quad + \cdots + \left[X'DX(X'X)^{-1}\right]^k + \cdots\Big].\end{aligned}$$

This representation of the inverse matrix is easily checked by multiplication.

Recall that $P = X(X'X)^{-1}X'$, $M = I - P$. Then

$$\begin{aligned} u_{\mathrm{GLS}} &= Mu + PD\left[I + PD + (PD)^2 + \cdots\right]Mu \\ &= Mu + P[I - \Sigma^{-1}][I - P + P\Sigma^{-1}]^{-1}Mu \\ &= Mu + P(I - \Sigma^{-1})(M + P\Sigma^{-1})^{-1}Mu \\ &= Mu + LMu \quad \text{where} \quad L = P(I - \Sigma^{-1})(M + P\Sigma^{-1})^{-1}. \end{aligned}$$

Expanding $\hat{\Sigma}$ around the true parameter values τ, $\tilde{u}_i$ can now be given as

$$\tilde{u}_i = L \cdot \Sigma_{-1/2}^{-1}(M + P\Sigma^{-1})^{-1}Mu - P\Sigma_{-1/2}^{-1}(M + P\Sigma^{-1})^{-1}Mu + \tilde{\tilde{u}},$$

where $\Sigma_{-1/2}^{-1}$ involves terms of order $O_p(T^{-1/2})$ in the expansions of the estimators of τ, and the elements of $\tilde{\tilde{u}}$ involve terms of $O_p(T^{-1})$, from Σ_{-1}^{-1}. ■

Appendix 2

The derivatives of the ARMA covariance matrix

The following formulae provide derivatives of the inverse process as well, by interchanging AR and MA parameters; hence the following theorem gives the elements of Σ_i^{-1} of Theorem 1. Recall first the formulae for elements of the ARMA(p, q) covariance matrix under the assumption of no multiple roots, given by Zinde-Walsh (1988):

$$\{\Sigma\}_{st} = \sum_{j=1}^{p} \xi_j \mu_j^k + \sum_{m=k+1}^{q} \alpha_m \sum_{j=1}^{p} \zeta_j\left(\mu_j^{m-k} - \mu_j^{k-m}\right), \tag{A2.1}$$

where

$$\zeta_j = \mu_j^{p-1} \Big/ \left[\left(1 - \mu_j^2\right)\prod_{i \neq j}(\mu_i - \mu_j)(1 - \mu_i\mu_j)\right], \qquad k = |s - t|,$$

$$\xi_j = \zeta_j \sum_{l=0}^{q} \alpha_l\left(\mu_j^l + \mu_j^{-l}\right), \qquad i \le j \le p, \tag{A2.2}$$

and

$$\alpha_0 = \tfrac{1}{2}\left(1 + \theta_1^2 + \cdots + \theta_q^2\right), \quad \alpha_m = -\theta_m + \sum_{h=1}^{q-m} \theta_h\theta_{h+m}, \qquad 1 \le m \le q.$$

Theorem A2. For an ARMA(p,q) process with no multiple roots in the autoregressive parts, the derivatives of the covariance matrix with respect to the MA parameters are given by

$$\left\{\frac{\partial\Sigma}{\partial\theta_i}\right\}_{st} = A_{i,k} + B_{i,k} + C_{i,k}, \qquad k = |s-t|,$$

where

$$A_{i,k} = \sum_{j=1}^{p} \mu_j^k \zeta_j \left[-\left(\mu_j^i + \mu_j^{-i}\right) + \sum_{l=1}^{i-1} \theta_{i-l}\left(\mu_j^l + \mu_j^{-l}\right) + \sum_{l=0}^{q-i} \theta_{i+l}\left(\mu_j^l + \mu_j^{-l}\right) \right],$$

$$\begin{aligned} B_{i,k} &= \sum_{m=k+1}^{q-i} \theta_{i+m} \sum_{j=1}^{p} \zeta_j\left(\mu_j^{m-k} - \mu_j^{k-m}\right) && \text{if } i < q-k, \\ &= 0 && \text{otherwise}, \end{aligned}$$

$$\begin{aligned} C_{i,k} &= \sum_{m=k+1}^{i-1} \theta_{i-m} \sum_{j=1}^{p} \zeta_j\left(\mu_j^{m-k} - \mu_j^{k-m}\right) \\ &\quad - \sum_{j=1}^{p} \zeta_j\left(\mu_j^{i-k} - \mu_j^{k-i}\right) && \text{if } i > k+1, \\ &= -\sum_{j=1}^{p} \zeta_j\left(\mu_j - \mu_j^{-1}\right) && \text{if } i = k+1, \\ &= 0 && \text{otherwise}. \end{aligned}$$

The derivatives of the covariance matrix with respect to the roots of the AR part are given by $\partial\Sigma/\partial\mu_i$ with elements

$$\begin{aligned} \left\{\frac{\partial\Sigma}{\partial\mu_i}\right\}_{st} &= k\xi_i\mu_i^{k-1} \\ &\quad + \mu_i^k \left[\sum_{j=1}^{p} \frac{\partial\zeta_j}{\partial\mu_i} \sum_{l=0}^{q} \alpha_l\left(\mu_j^l + \mu_j^{-l}\right) + \zeta_j \sum_{l=1}^{q} l\alpha_l\left(\mu_i^{l-1} - \mu_i^{-l-1}\right) \right] \\ &\quad + \sum_{l=k+1}^{q} \alpha_l \left\{ \sum_{j=1}^{p} \left(\mu_j^{l-k} - \mu_j^{k-l}\right) \frac{\partial\zeta_j}{\partial\mu_i} \right. \\ &\quad \left. + \zeta_i(l-k)\left(\mu_i^{l-k-1} + \mu_i^{k-l-1}\right) \right\}, \end{aligned}$$

where

$$\frac{\partial \zeta_j}{\partial \mu_i} = \frac{\mu_j^{p-1}\left[1 - 2\mu_i\mu_j + \mu_j^2\right]}{(\mu_i - \mu_j)(1 - \mu_i\mu_j)\left(1 - \mu_j^2\right) \prod_{m \neq i} (\mu_i - \mu_m)(1 - \mu_i\mu_m)} \qquad \text{for} \quad i \neq j,$$

$$= \frac{\mu_i^{p-2}\left[p - 1 + 2\mu_i^2/\left(1 - \mu_i^2\right)\right] - \mu_i^{p-1} \sum_{m \neq i} \left(1 - 2\mu_i\mu_m + \mu_i^2\right)/(1 - \mu_i\mu_m)(\mu_i - \mu_m)}{\left(1 - \mu_i^2\right) \prod_{m \neq i} (\mu_i - \mu_m)(1 - \mu_i\mu_m)} \qquad \text{for} \quad i = j.$$

The derivatives with respect to the AR parameters have as elements $\{\partial\Sigma/\partial\rho_i\}_{st}$, each of which can be obtained as the ith coordinate of the row vector $\{\partial\Sigma/\partial\rho\}_{st} = (\{\partial\Sigma_1/\partial\rho_1\}_{st}, \ldots, \{\partial\Sigma/\partial\rho_p\}_{st})$, with

$$\left\{\frac{\partial\Sigma}{\partial\rho}\right\}_{st} = \left\{\frac{\partial\Sigma}{\partial\mu}\right\}_{st} W_{p\times p}^{-1},$$

where

$$\{W_{p\times p}\}_{ij} = \frac{\partial\rho_i}{\partial\mu_j} = (-1)^{i-1} \sum_{l \neq j} {}^{*}\mu_{l_1}, \ldots, \mu_{l_{i-1}},$$

with the asterisk (*) indicating that $l_1, l_2, \ldots, l_i$ are different.

Proof. The results follow from straightforward differentiation with respect to the roots μ of the AR polynomial and parameters θ of the MA part; for parameters of the AR part, the relationship between these parameters and roots μ is then used.

Appendix 3

Proofs of Lemmas 1 and 2 and Theorem 1

Proof of Lemma 1. Denote by S_T the matrix of an orthogonal transformation of H_∞ such that $S_T'\Sigma^{1/2}A_T\Sigma^{1/2}S_T = \Lambda_T$, with Λ_T a diagonal matrix; clearly Λ_T has r nonzero uniformly bounded elements on the main diagonal. Denote by $\bar{\varepsilon}_i$ the vector $S_T\varepsilon_i$, the set $\{\bar{\varepsilon}_i\}$ being an orthogonal transformation of the orthonormal basis in H_T, the T-dimensional subspace of the Hilbert space H_∞, and therefore a set of identically distributed orthogonal random variables with finite variances. We have $u'A_T u = \bar{\varepsilon}'\Lambda_T\bar{\varepsilon} = \sum_{i=1}^{r}\lambda_i\bar{\varepsilon}_i^2$, with each λ_i representing a nonzero bounded eigenvalue of $\Sigma^{1/2}A_T\Sigma^{1/2}$, and therefore bounded in probability. It follows that $T^{-1}u'A_T u$ is $O_p(T^{-1})$. ■

Proof of Lemma 2. From the formula for the determinant of an ARMA covariance matrix [Zinde-Walsh (1988)] we can see that D is a rational function of the parameter vector τ and that $\det[\Sigma_T]$ is nonzero at τ, so that D is bounded in a neighbourhood of τ.

To find the orders of the quadratic forms, represent the maximum-likelihood residuals as $u_{\mathrm{ML}} = u + \eta$, where $u = \Sigma^{1/2}\varepsilon$ and η can be found from the formula in appendix 1. All matrices in the quadratic forms in (3.8) are symmetric and can be viewed as matrices of linear operators in H_∞. Recall that the operator norm for a linear operator A in a Hilbert space is defined as $\sup_{x\in H_\infty}\|Ax\|/\|x\|$. The operators corresponding to the matrices in (3.8) are rational symmetric functions of projection operators and the backshift operator, and the parameter space defined in section 2 above is such that these functions are analytic functions of these operators. The operator norm of a projection operator in H_∞, or of the backshift operator, is 1. The operator norm for each of the matrices in the quadratic forms is bounded and provides an upper bound for the eigenvalues of each matrix.

We need only show that there exists a uniform bound independent of T: from the matrix expressions we can see that it is sufficient that V and $\partial V/\partial\tau$ remain bounded in the norm as T increases. Consider the expression for Σ_T^{-1} in (3.21)–(3.22) of Zinde-Walsh (1988); these formulae indicate that the difference $\Sigma_{T_1}^{-1} - \Sigma_{T_2}^{-1}$ declines as $\min(T_1, T_2)$ increases (exponentially, elementwise). It is straightforward to show that a similar decline obtains for the difference in the derivatives, $\partial\Sigma_{T_1}^{-1}/\partial\tau - \partial\Sigma_{T_2}^{-1}/\partial\tau$. This is sufficient to imply uniform boundedness in the norm. Thus for all matrices in (3.8) with finite rank, Lemma 1 immediately applies.

To obtain the results for quadratic forms in u_{ML} note that $\eta = Pu + LMu + \tilde{u}$, where $\tilde{u}$ is $O_p(T^{-1/2})$. If Lemma 1 applies to $u'Bu$, where B represents any of the three matrices considered here, then by an analogous argument it also applies to $u'PBPu$ and $u'ML'BLMu$. Moreover, $\tilde{u}'B\tilde{u}$ is of order $O_p(T^{-1})$. Recalling that $u'B\eta \le (u'Bu)^{1/2}(\eta'B\eta)^{1/2}$, we have that the order of the quadratic form in u_{ML} is $O_p(T^{-1})$ for each of the matrices considered.

Finally, for E the diagonalised matrix Λ_T converges to a diagonal matrix with elements equal to the eigenvalues of $\Sigma_\infty^{1/2}(\partial\Sigma_\infty^{-1}/\partial\tau)\Sigma_\infty^{1/2}$ so that $T^{-1}E$ is a weighted sum of $\bar{\varepsilon}_i^2$, the eigenvalues are bounded, and $T^{-1}E$ is bounded in mean and therefore also in probability. ■

Proof of Theorem 1. The feasible GLS estimator can be represented, after expanding the denominator, as follows:

$$\beta^{\mathrm{GLS}} = \beta + \beta_{-1/2} + \beta_{-1} + O_P(T^{-3/2}), \tag{A3.2}$$

where

$$\beta_{-1/2} = \Omega X'\Sigma^{-1}u,$$

$$\beta_{-1} = \Omega X'\overline{\frac{\partial\Sigma^{-1}}{\partial\tau}}\bar{\tau}_{-1/2}u - \Omega X'\overline{\frac{\partial\Sigma^{-1}}{\partial\tau}}\bar{\tau}_{-1/2}X\Omega X'\Sigma^{-1}u. \tag{A3.3}$$

The expression $\overline{(\partial\Sigma^{-1}/\partial\tau)}\bar{\tau}_{-1/2}$ denotes the product of the vectors $\overline{\partial\Sigma^{-1}/\partial\tau} = (\Sigma_1^{-1},\dots,\Sigma_m^{-1})'$, with matrix coordinates $\Sigma_i^{-1} = \partial\Sigma^{-1}/\partial\tau_i$, and $\bar{\tau}_{-1/2}$ with $\bar{\tau}_{-1/2} = (\tau_{-1/2}^1 I_T,\dots,\tau_{-1/2}^m I_T)$, where $\tau_{-1/2}^i$ comes from the expansion of τ_i and I_T is the $T\times T$ identity matrix.

Thus we see that the expansion in (A3.2) is the same regardless of the methods of estimation, i.e., regardless of whether maximum-likelihood or method-of-moments estimators based on (3.10) are used for τ.

For a comparison of the variances of the estimators a further term is required. Assume that we have the expansion

$$\tau^{\mathrm{MOM}} = \tau + \tau_{-1/2}^{\mathrm{MOM}} + \tau_{-1}^{\mathrm{MOM}} + \mathrm{O}_P(T^{-3/2});$$

then

$$\tau^{\mathrm{ML}} = \tau^{\mathrm{MOM}} + \tilde{\tau}_{-1} + \mathrm{O}_{\mathrm{P}}(\mathrm{T}^{-3/2}). \tag{A3.4}$$

We next determine the general form of the terms $\tau_{-1/2}^{\mathrm{MOM}}$, τ_{-1}^{MOM}, and $\tilde{\tau}_{-1}$. The following representations hold for *each of the coordinates of the parameter vector*, τ:

$$\tau_{-1/2}^{\mathrm{MOM}} = -\frac{1}{T}\zeta^{-1}u'MU_1Mu, \tag{A3.5}$$

$$\tau_{-1}^{\mathrm{MOM}} = \frac{1}{T}\zeta^{-2}u'MU_1Mu\left[\frac{1}{T}u'MU_2Mu - \zeta\right]$$
$$- T^{-2}\zeta^{-2}(u'MU_1Mu)^2u'MU_3Mu, \tag{A3.6}$$

$$\tilde{\tau}_{-1} = \frac{2}{T}\left[R(\tau) + u'MU_4Mu\right]\left[\frac{1}{T}u'MU_2Mu - \zeta\right], \tag{A3.7}$$

where U_1, U_2, U_3, U_4 are the symmetric matrices

$$U_1 = \frac{\partial\Sigma_\infty^{-1}}{\partial\tau},\qquad U_2 = \frac{\partial U_1}{\partial\tau},\qquad U_3 = \frac{\partial U_2}{\partial\tau},\qquad U_4 = \frac{\partial U_3}{\partial\tau},$$

$R(\tau)$ is a rational function defined below, and $\zeta = \lim_{T\to\infty}(1/T)u'MU_2Mu$. Note that ζ can be expressed as $\lim(1/T)\mathrm{tr}\,\Sigma[\partial^2\Sigma^{-1}/\partial\tau_i^2]$ and is O(1).

To prove (A3.5)–(A3.7) consider the eq. (3.10); the following must hold for each element of the vector τ (for notational convenience we continue to omit subscripts i):

$$\frac{\partial \hat{\Sigma}_\infty^{-1}}{\partial \tau} = U_1 + U_2(\tau_{-1/2} + \tau_{-1}) + U_3 \tau_{-1/2}^2 + \mathrm{O}_P(T^{-3/2}). \tag{A3.8}$$

Then $u'MU_1Mu + \tau_{-1/2}u'MU_2Mu = \mathrm{O}_P(T^{-1})$, from (3.10).

Consider $((1/T)u'MU_2Mu)^{-1}$. It is equal to

$$\zeta^{-1}\left[1 + \zeta^{-1} \cdot \left[\frac{1}{T}u'MU_2Mu - \zeta\right]\right]^{-1}$$

$$= \zeta^{-1}\left[1 - \zeta^{-1}\left[\frac{1}{T}u'MU_2Mu - \zeta\right] + \mathrm{O}(T^{-2})\right]. \tag{A3.9}$$

We obtain $\tau_{-1/2}$ by substitution and ignoring terms to $\mathrm{O}(T^{-1})$; to get τ_{-1}, the next term and the expression for $\tau_{-1/2}^2$ can be substituted and (A3.6) results.

Denote by $R(\tau)$ the rational function $[\partial \det \Sigma/\partial\tau][\det \Sigma]^{-1}$. It is easily proven that $\tilde{\tau}_{-1}$ solves

$$R(\tau) + u'_{\mathrm{OLS}}U_2u_{\mathrm{OLS}}\tilde{\tau}_{-1} + u'_{\mathrm{OLS}}\frac{\partial V}{\partial \tau}u_{\mathrm{OLS}} + u'_{-1/2}\frac{\partial \Sigma_\infty^{-1}}{\partial \tau_i}u_{\mathrm{OLS}}$$

$$+ u'_{\mathrm{OLS}}\frac{\partial \Sigma_\infty^{-1}}{\partial \tau}u_{-1/2} + u'_{-1/2}\frac{\partial \Sigma_\infty^{-1}}{\partial \tau}u_{-1/2} = \mathrm{O}(T^{-1}).$$

Here $u_{-1/2} = P(I - \Sigma^{-1})(M + P\Sigma^{-1})^{-1}Mu$ (see appendix 1).

The conclusion of the proof of (A3.5)–(A3.7) requires solving for $\tilde{\tau}_{-1}$ and expanding the denominator, which is (A3.9).

Next we derive the expansion for the variance of the slope estimator. Note that $\mathrm{E}(\hat{\beta} - \beta) = 0$ because $(\hat{\beta} - \beta)$ is an odd function of u. In fact, all terms in the expansion of τ are functions of quadratic forms in u; this holds true both for the two-stage and the maximum-likelihood estimator. In order to obtain the variances up to $\mathrm{O}(T^{-2})$ we need to expand the estimators further, to $\mathrm{O}_P(T^{-3/2})$. If we know that $\hat{\beta} - \beta = \beta_{-1/2} + \beta_{-1} + \beta_{-3/2} + \mathrm{O}_P(T^{-2})$, we obtain

$$\begin{aligned}\mathrm{E}(\hat{\beta} - \beta)(\hat{\beta} - \beta)' &= \mathrm{E}(\beta_{-1/2}\beta'_{-1/2}) + \mathrm{E}(\beta_{-1/2}\beta'_{-1} + \beta_{-1}\beta'_{-1/2}) \\ &\quad + \mathrm{E}(\beta_{-1}\beta'_{-1} + \beta_{-1/2}\beta'_{-3/2} + \beta_{-3/2}\beta'_{-1/2}) \\ &\quad + \mathrm{o}(T^{-2}).\end{aligned}$$

In addition to the terms in the expansion of β given in (A3.3), which are the same for the maximum-likelihood and for the two-stage estimator, we need to consider $\beta_{-3/2}$ here which depends upon τ_{-1} and thus differs for the two estimators. It is not difficult to obtain by further expansion

$$\begin{aligned}\beta_{-3/2} = {} & \Omega X' \overline{\frac{\partial \Sigma^{-1}}{\partial \tau}} \bar{\tau}_{-1} u - \Omega X' \overline{\frac{\partial \Sigma^{-1}}{\partial \tau}} \bar{\tau}_{-1} X \Omega X' \Sigma^{-1} u \\ & + \Omega X' \bar{\tau}'_{-1/2} \overline{\frac{\partial^2 \Sigma^{-1}}{\partial \tau^2}} \bar{\tau}_{-1/2} u \\ & - \Omega X' \bar{\tau}'_{-1/2} \overline{\frac{\partial^2 \Sigma^{-1}}{\partial \tau^2}} \bar{\tau}_{-1/2} X \Omega X' \Sigma^{-1} u \\ & - \Omega X' \overline{\frac{\partial \Sigma^{-1}}{\partial \tau}} \bar{\tau}_{-1/2} X \Omega X' \overline{\frac{\partial \Sigma^{-1}}{\partial \tau}} \bar{\tau}_{-1/2} u \\ & + \Omega X' \overline{\frac{\partial \Sigma^{-1}}{\partial \tau}} \bar{\tau}_{-1/2} X \Omega X' \overline{\frac{\partial \Sigma^{-1}}{\partial \tau}} \bar{\tau}_{-1/2} X \Omega X' \Sigma^{-1} u, \end{aligned} \qquad \text{(A3.10)}$$

where the vector $\overline{\partial \Sigma^{-1}/\partial \tau}$ and the matrix $\overline{\partial^2 \Sigma^{-1}/\partial \tau^2}$ are composed of first derivatives and second and cross-derivatives of Σ^{-1} with respect to the τ_i, respectively.

From (A3.5)–(A3.7) we see that each of $\tau^2_{-1/2}$, τ^{MOM}_{-1}, or τ^{ML}_{-1} can be represented as sums of products of at most two quadratic forms in u.

For $\mathrm{E}(u'MAMu)(u'MBMu)uu'$ with positive definite A and B, we obtain from Srivastava and Tiwari (1976) the expression

$$\phi \Sigma + \Sigma M \Phi M \Sigma,$$

where ϕ is a constant and Φ a symmetric matrix, each of which depends on the given matrices. Substituting, we get owing to cancellations of terms that $\mathrm{E}(\beta_{-1/2}\beta'_{-3/2}) = 0$ and $\mathrm{E}(\beta_{-1/2}\beta'_{-1}) = 0$. Thus $\mathrm{E}(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = \sigma^2 \Omega + \mathrm{E}(\beta_{-1}\beta'_{-1}) + \mathrm{o}(T^{-2})$, which is independent of whether $\hat{\beta}$ was obtained by maximum likelihood or by a two-stage procedure using (3.10).

Next, we obtain $\mathrm{E}(\beta_{-1}\beta'_{-1})$. Substituting β_{-1} from (A3.3) we have

$$\begin{aligned}\mathrm{E}(\beta_{-1}\beta'_{-1}) = \sum_{i,j=1}^{m} \Big\{ & \Omega X' \Sigma_i^{-1} \mathrm{E}\big(\tau^i_{-1/2} \tau^j_{-1/2} u u'\big) \Sigma_j^{-1} X \Omega \\ & - \Omega X' \Sigma_i^{-1} X \Omega X' \Sigma^{-1} \mathrm{E}\big(\tau^i_{-1/2} \tau^j_{-1/2} u u'\big) \\ & \times \Sigma^{-1} X \Omega X' \Sigma_j^{-1} X \Omega \Big\}. \end{aligned}$$

Using the formulae for the expectations from the right-hand side in each term, we are left after some cancellations and ignoring terms of $o(T^{-2})$ with terms involving $\operatorname{tr}(M\Sigma_i^{-1}M\Sigma)$ and $\operatorname{tr}(M\Sigma_i^{-1}M\Sigma M\Sigma_j^{-1}M\Sigma)$ as coefficients. We demonstrate next that $\operatorname{tr}(M\Sigma_i^{-1}M\Sigma)=O(1)$ and that $\operatorname{tr}(M\Sigma_i^{-1}M\Sigma M\Sigma_j^{-1}M\Sigma)$ $=O(T)$.

Note that the quantity $M=I-P$ appearing in expressions for traces can be replaced by I without affecting the order of the expression. If the matrix M is diagonalised, the trace of the product of M with other matrices has two components, one in which M is replaced by I and another in which the trace of the product of the matrix P with other matrices appears. Since P is of constant rank, the trace of such a product is O(1). Hence the order of the overall expression is unchanged.

Recall that $\Sigma_i^{-1}=\partial\Sigma_\infty^{-1}/\partial\tau_i$. To evaluate the order of the first trace we need only consider $\operatorname{tr}((\partial\Sigma_\infty^{-1}/\partial\tau_i)\Sigma_\infty)$. Note that $(\partial\Sigma_\infty^{-1}/\partial\tau)\Sigma_\infty=\Sigma_\infty^{-1}(\partial\Sigma_\infty/\partial\tau)$, so that without loss of generality we can assume that τ is a parameter of the autoregressive part. Then, ignoring low-rank effects, we have (subscripting the moving-average part by M, the autoregressive part by A):

$$\begin{gathered}\Sigma_\infty\cong\Sigma_{\mathrm{M}}\Sigma_{\mathrm{A}},\qquad \Sigma_\infty^{-1}\cong\Sigma_{\mathrm{A}}^{-1}\Sigma_{\mathrm{M}}^{-1},\\ \frac{\partial\Sigma_\infty^{-1}}{\partial\tau}=\frac{\partial\Sigma_{\mathrm{A}}^{-1}}{\partial\tau}\Sigma_{\mathrm{M}}^{-1},\quad\text{and so}\quad \frac{\partial\Sigma_\infty^{-1}}{\partial\tau}\Sigma_\infty\cong\frac{\partial\Sigma_{\mathrm{A}}^{-1}}{\partial\tau}\Sigma_{\mathrm{A}}.\end{gathered}\tag{A3.11}$$

Recalling the well-known Yule–Walker equations it is easy to see that all but a finite number of terms on the diagonal of $(\partial\Sigma_{\mathrm{A}}^{-1}/\partial\tau)\Sigma_A$ are zeros. For the second trace we need to examine the diagonal of $(\partial\Sigma_{\mathrm{A}'}^{-1}/\partial\tau_i)\Sigma_{\mathrm{A}'}\times(\partial\Sigma_{\mathrm{A}''}^{-1}/\partial\tau_j)\Sigma_{\mathrm{A}''}$; here each of A′ and A″ could be either A or M^{-1}, the autoregressive process which represents the inverse of our moving-average part. Note that each term on the diagonal of this matrix is a sum of at most $\max(q^2,p^2)$ terms; hence the trace is of $O(T)$.

Substituting the term of the appropriate order into the expression for variance concludes the proof of Theorem 1. ∎

References

Amemiya, T., 1985, Advanced econometrics (Harvard University Press, Cambridge, MA).

Box, G.E.P. and G.M. Jenkins, 1970, Time series analysis: Forecasting and control (Holden-Day, San Francisco, CA).

Godolphin, E.J., 1978, Modified maximum likelihood estimation of Gaussian moving averages using a pseudoquadratic convergence criterion, Biometrika 65, 203–206.

Godolphin, E.J., 1984, A direct representation for the large-sample maximum likelihood estimator of a Gaussian autoregressive – moving average process, Biometrika 71, 281–289.

Godolphin, E.J. and J.G. de Gooijer, 1982, On the maximum likelihood estimation of the parameters of a Gaussian moving average process, Biometrika 69, 443–451.

Harvey, A.C. and G.D.A. Phillips, 1979, Maximum likelihood estimation of regression models with autoregressive–moving average disturbances, Biometrika 66, 49–58.

Judge, G.G., R.C. Hill, W.E. Griffiths, H. Lütkepohl, and T.C. Lee, 1985, The theory and practice of econometrics, 2nd ed. (Wiley, New York, NY).

Krämer, W., 1980, Finite-sample efficiency of ordinary least squares in the linear regression model with autocorrelated errors, Journal of the American Statistical Association 75, 1005–1009.

Magee, L., 1985, Efficiency of iterative estimators in the regression model with AR(1) disturbances, Journal of Econometrics 29, 275–287.

Pagan, A., 1986, Two stage and related estimators and their applications, Review of Economic Studies 53, 517–538.

Pierce, D.A., 1971, Least-squares estimation in the regression model with autoregressive-moving average errors, Biometrika 58, 299–312.

Prais, S.J. and C.B. Winsten, 1954, Trend estimators and serial correlation, Discussion paper (Cowles Foundation, Chicago, IL).

Rao, P. and Z. Griliches, 1969, Small-sample properties of several two-stage regression methods in the context of auto-correlated errors, Journal of the American Statistical Association 64, 253–272.

Robinson, P, 1988, The stochastic difference between econometric statistics, Econometrica 56, 531–548.

Rothenberg, T.J., 1984, Approximate normality of generalized least squares estimates, Econometrica 52, 811–826.

Spitzer, J.J., 1979, Small-sample properties of nonlinear least squares and maximum likelihood estimators in the context of autocorrelated errors, Journal of the American Statistical Association 74, 41–47.

Srivastava, V.K. and R. Tiwari, 1976, Evaluation of expectations of products of stochastic matrices, Scandinavian Journal of Statistics 3, 135–138.

Ullah, A., V.K. Srivastava, L. Magee, and A. Srivastava, 1983, Estimation of linear regression model with autocorrelated disturbances, Journal of Time Series Analysis 4, 127–135.

Wise, J., 1955, The autocorrelation function and the spectral density function, Biometrika 42, 151–160.

White, H., 1984, Asymptotic theory for econometricians (Academic Press, New York, NY).

Zinde-Walsh, V., 1988, Some exact formulae for autoregressive-moving average processes, Econometric Theory 4, 384–402. Also: Errata, 1990, forthcoming.