Review of Economic Studies (1990) 57, 127-134 0034-6527/90/00070127$02.00

Testing for Autocorrelation in Dynamic Random Effects Models

MANUEL ARELLANO
London School of Economics

First version received April 1987; *final version accepted April* 1989 (*Eds.*)

This article develops tests of covariance restrictions after estimating by three-stage least squares a dynamic random effects model from panel data. The asymptotic distribution of covariance matrix estimates under non-normality is obtained. It is shown how minimum chi-square tests for interesting covariance restrictions can be calculated from a generalised linear regression involving the sample autocovariances and dummy variables. Asymptotic efficiency exploiting covariance restrictions can also be attained using a GLS estimator.

1. INTRODUCTION AND PRELIMINARIES

A popular method of estimation for linear models from short panels including lags of the dependent variable and individual effects is three-stage least squares (3SLS) applied to the system of T equations corresponding to each time period available in the sample. A desirable feature of this technique is that the resulting estimates are robust to residual autocorrelation of arbitrary form, since the $T\times T$ autocovariance matrix of the errors is left unrestricted. However, in a model with lagged dependent variables and sufficient strictly exogenous variables, if the autocovariances satisfy some set of restrictions, more efficient estimates of the regression parameters can be obtained. This is so because of lack of orthogonality between slope and covariance parameters in the autoregressive model. In addition, in a typical situation the number of constraints under consideration will be large. For example, an error components structure with a first-order moving average scheme is characterised by 3 parameters which, for $T=10$, imposes 52 restrictions on the autocovariance matrix. On the other hand, it can be argued that often the economic model suggests restrictions in the autocorrelation properties of the errors (e.g. white noise shocks) and that finding the converse could be interpreted as an indication of mis-specification. Therefore, it is of interest to be able to perform tests of covariance restrictions after estimation by 3SLS or asymptotically equivalent techniques, and also to obtain more efficient estimates of the parameters if a set of restrictions is accepted.

The model we consider is

$$y_{it}=\gamma_t+\alpha y_{i(t-1)}+\beta z_{it}+u_{it} \qquad (i=1,\ldots,N;\ t=2,\ldots,T)$$

$$u_{it}=\eta_i+v_{it} \tag{1}$$

We assume that $z_i=(1 \quad z_{i1} \quad \cdots \quad z_{iT})'$ is strictly exogenous conditional on the individual effect η_i

$$E(v_{it}\,|\,z_i,\eta_i)=0$$

and that u_{it} is homoskedastic across individuals. To complete the model the distributions of y_{i1} and η_i conditional on z_i are assumed to satisfy

$$E(y_{i1}|z_i)=\mu' z_i$$

$$E(\eta_i|z_i)=\lambda' z_i$$

The coefficient γ_t represents a time effect and α and β are both scalar for simplicity of presentation.

The basic discussion on this model is due to Chamberlain (1984), where the following facts are shown.[1] The model can be written as a multivariate regression of T equations subject to restrictions and, provided plim $N^{-1}\sum' z_i z_i'$ is nonsingular and $T\geqq 3$, α, β, μ and λ are identified. In the absence of covariance restrictions an efficient minimum distance estimator is readily available. In addition, Chamberlain shows that an asymptotically equivalent estimator of α and β can be obtained by 3SLS in the system of $(T-2)$ equations (1) in first-difference form. This alternative is more convenient because in this form the cross-equation restrictions are linear and μ and λ need not be explicitly estimated. If $\lambda=0$, η_i and z_{it} are uncorrelated in which case 3SLS in the system of the $(T-1)$ equations (1) in levels is efficient. Since λ is a vector of regression coefficients, conventional tests of $\lambda=0$ can easily be implemented.

Bhargava and Sargan (1983) have developed maximum likelihood estimators for a similar model with normal errors. They propose likelihood ratio (LR) tests of the error components structure. One problem with this approach is that LR tests of covariance restrictions are not robust to non-normality (see MaCurdy (1981) and Arellano (1985)) and thus are bound to reject too often when the distribution of the errors has long tails. Another problem is that enforcing covariance restrictions by quasi-ML methods does not necessarily represent an efficiency improvement over unrestricted covariance QML when the errors are non-normal (see Arellano (1989*a*)). MaCurdy (1982) has proposed to use time-series methods for selecting autoregressive moving-average schemes for v_{it} in static models. While his approach could be extended to dynamic models, the possibility of obtaining consistent estimates of the covariances in the absence of restrictions suggests basing a formal specification search on a sequence of tests of particular schemes against the unrestricted autocovariance matrix.

Section 2 states the limiting distribution of the elements of the estimated covariance matrix under non-normality and presents robust Wald and minimum chi-square (MCS) tests of covariance restrictions. Section 3 discusses how to calculate MCS tests for particular useful cases in the form of a generalised regression involving the sample autocovariances and dummy variables. Section 4 provides a summary and some concluding remarks on how the GLS estimator of Arellano (1989*b*) can be applied to obtain efficient estimates of α, β and γ_t imposing covariance restrictions. Proofs are collected in an Appendix.

2. THE LIMITING DISTRIBUTION OF UNRESTRICTED COVARIANCE MATRIX ESTIMATES

It is convenient to write model (1) in the form

$$y_{i1}=\mu' z_i+u_{i1}$$

$$y_i=X_i\delta+u_i \qquad (i=1,\ldots,N) \tag{2}$$

1. Chamberlain's model is more general since he does not assume homoskedasticity and replaces conditional expectations by linear predictors.

With uncorrelated individual effects ($\lambda=0$), (2) is the notation for the levels model so that $\delta=(\alpha\,\beta\,\gamma_2\cdots\gamma_T)'$, $u_i=(u_{i2}\cdots u_{iT})'$, $y_i=(y_{i2}\cdots y_{iT})'$ and X_i is a matrix containing the time series of $y_{i(t-1)}$, z_{it} and time dummies for the ith unit. While with correlated effects, (2) represents the model in first differences (y_{i1}, y_i, X_i and u_i—but not z_i—represent first-differences of the original variables and the time series are one period shorter). In either case let Ω^* be the covariance matrix of the complete system given by

$$\Omega^*=\mathrm{Var}\begin{pmatrix}u_{i1}\\ u_i\end{pmatrix}=\begin{pmatrix}\omega_{11} & \omega_{12}\\ \omega_{21} & \Omega\end{pmatrix}$$

In the levels model, Ω will typically contain an error-component structure while in the first-differences model Ω may be expected to contain a moving average unit root.

Theorem. *Assume* (i) $(u_{i1}\cdots u_{iT})$ *is an iid random vector with* $E(u_{it})=0$ $(t=1,\ldots,T)$ *and finite moments up to the fourth order*; (ii) plim $N^{-1}\sum_{i=1}^{N}z_iz_i'$ *exists and is non-singular. Let* $\hat{\omega}_{ts}$ *be an estimate of the* (t,s) *covariance based on the* 3SLS *residuals* $\hat{u}_{it}$ *and* $\hat{u}_{is}$:

$$\hat{\omega}_{ts}=\frac{1}{N}\sum_{i=1}^{N}\hat{u}_{it}\hat{u}_{is}\qquad (t.\ s=1,\ldots,T) \tag{3}$$

and let $\hat{\omega}$ *be a* $T(T+1)/2$ *vector of the distinct elements* $\hat{\omega}_{ts}$. *Then* $\sqrt{N}(\hat{\omega}-\omega)$ *has a limiting multivariate normal distribution with zero mean and covariance matrix with elements*

$$\mathrm{acov}\,(\hat{\omega}_{ts},\hat{\omega}_{t's'})=\mathrm{avar}\,(\hat{\alpha})a_{ts}a_{t's'}+\mu_{tst's'}-\omega_{ts}\omega_{t's'} \tag{4}$$

where $\mu_{tst's'}=E(u_{it}u_{is}u_{it'}u_{is'})$, $\hat{\alpha}$ *is the* 3SLS *estimator of* α *and*

$$\begin{aligned} a_{ts}&=a_{st}=\sum_{k=1}^{t-1}\alpha^{(k-1)}\omega_{(t-k)s}+\sum_{l=1}^{s-1}\alpha^{(l-1)}\omega_{(s-l)t}, && (t,s=2,\ldots T),\\ a_{t1}&=a_{1t}=\sum_{k=1}^{t-1}\alpha^{(k-1)}\omega_{(t-k)1}, && (t=2,\ldots,T),\\ a_{11}&=0. \end{aligned} \tag{5}$$

Proof. See Appendix. ||

Note that third-order moments of u_{it}, though unrestricted, are absent from (4). However, this result does not hold when the intercepts γ_t are constrained unless the third-order moments vanish.

Natural estimates of the elements given in (4) can be obtained replacing α, ω_{ts} and $\mu_{tst's'}$ by $\hat{\alpha}$, $\hat{\omega}_{ts}$ and $\hat{\mu}_{tst's'}=N^{-1}\sum_{i=1}^{N}\hat{u}_{it}\hat{u}_{is}\hat{u}_{it'}\hat{u}_{is'}$, and avar$(\hat{\alpha})$ by the square of the standard error of $\hat{\alpha}$. Having obtained such estimates, we can construct Wald and MCS tests of covariance restrictions. Suppose $f(\omega)=0$ is a vector of r restrictions on the elements of ω, which can alternatively be parameterised as $\omega=\omega(\psi)$ where ψ is a $T(T+1)/2-r$ vector of constraint parameters. A Wald test is

$$W=Nf(\hat{\omega})'(F\hat{V}F')^{-1}f(\hat{\omega})\xrightarrow{d}\chi_r^2 \tag{6}$$

where $F=\partial f(\hat{\omega})/\partial\omega'$ and $\hat{V}$ is a consistent estimate of the covariance matrix of $\hat{\omega}$. Let $\hat{\psi}$ be the minimizer of

$$s(\psi)=[\hat{\omega}-\omega(\psi)]'H'(\psi)(\tilde{H}\hat{V}\tilde{H}')^{-1}H(\psi)[\hat{\omega}-\omega(\psi)] \tag{7}$$

where $H(\psi)$ is a nonsingular transformation matrix and $\tilde{H}$ is $H(\psi)$ evaluated at some preliminary consistent estimate $\tilde{\psi}$ (cf. Chamberlain (1984)). An MCS test is

$$MCS = N \cdot s(\hat{\psi}) \xrightarrow{d} \chi^2_r. \tag{8}$$

Under normality the form for the asymptotic covariance matrix of $\hat{\omega}$ simplifies since fourth-order moments are no longer required. Indeed we can write $\mu_{tst's'}$ as

$$\mu_{tst's'} = \omega_{ts}\omega_{t's'} + \omega_{tt'}\omega_{ss'} + \omega_{ts'}\omega_{st'} \tag{9}$$

so that

$$\text{acov}(\hat{\omega}_{ts}, \hat{\omega}_{t's'}) = \text{avar}(\hat{\alpha})a_{ts}a_{t's'} + \omega_{tt'}\omega_{ss'} + \omega_{ts'}\omega_{st'} \tag{10}$$

However, Wald and MCS tests based on (10) will not be distributed as a χ^2 under the null unless (9) holds.

3. THE MCS TEST FOR SOME USEFUL CASES

In some cases of interest, like error-components or first-differences structures with a moving average scheme, the restrictions implied on the elements ω_{ts} are linear and the calculation of the Wald test is straightforward.[2] However for the majority of useful cases, the MCS statistic can be calculated as the residual sum of squares from an auxiliary generalised linear regression in which the elements $\hat{\omega}_{ts}$ and dummy variables are the observations. This makes the MCS test particularly attractive in this context. Below we give the details for particular cases.

(i) *Error components with a white noise or moving average scheme.*

Let $\text{Cov}(v_{it}, v_{i(t-j)}) = g_j$, $\text{Var}(\eta_i) = c$ and $h_j = c + g_j$. The white noise specification is

$$\omega_{st} = \begin{cases} c + g_0 & \text{if } s = t \\ c & \text{if } s < t \end{cases} \quad (t = 2, \ldots, T;\ s = 2, \ldots t)$$

and the MCS test (8) with H equal to an identity matrix can be obtained as the generalised sum of squares from the regression

$$\hat{\omega}_{st} = h_0 d(s = t) + cd(s < t) + \varepsilon_{st} \qquad (t = 2, \ldots, T;\ s = 2, \ldots t)$$

using $\hat{V}$ with elements $\hat{v}_{sts't'} = \text{ac\hat{o}v}(\hat{\omega}_{st}, \hat{\omega}_{s't'})$ as the weighting matrix. The variable $d(A)$ takes the value 1 if A is true and 0 otherwise. Similarly, for the first-order moving-average specification, the associated regression is

$$\hat{\omega}_{st} = h_0 d(s = t) + h_1 d(s = t-1) + cd(s < t-1) + \varepsilon_{st}.$$

(ii) *First differences with white noise or moving average errors in levels.*

Let $\text{Cov}(\Delta_{it}, \Delta v_{i(t-j)}] = \bar{g}_j$. The white noise case is equivalent to a simple first-order moving-average scheme with the restriction $\bar{g}_0 + 2\bar{g}_1 = 0$. The relevant regression is

$$\hat{\omega}_{st} = \bar{g}_1[d(s = t-1) - 2d(s = t)] + \varepsilon_{st}$$

2. Wald tests of moving average errors have been independently studied by Arellano (1985) and Bhargava (1987), including applications to empirical earnings functions estimated from the Michigan PSID data.

with a similar weighting matrix as for the cases in (i). On the same lines, for a first-order moving-average scheme for the error in levels we have

$$\hat{\omega}_{st} = \bar{g}_1[d(s=t-1)-2d(s=t)]+\bar{g}_2[d(s=t-2)-2d(s=t)]+\varepsilon_{st}.$$

(iii) *Error components with an autoregressive scheme.*

The structure is

$$\omega_{st} - c = \begin{cases} g_0 & \text{if } s=t \\ \rho(\omega_{s(t-1)} - c) & \text{if } s<t. \end{cases}$$

This specification suggests the regression

$$\hat{\omega}_{st} = h_0 d(s=t) + \bar{c} d(s<t) + \rho\hat{\omega}_{s(t-1)} d(s<t) + \varepsilon_{st}$$

with $\bar{c} = (1-\rho)c$. However, this regression corresponds to the criterion function of the form (7) that uses $H = I - \rho D$, where I is an identity matrix and D is a 0–1 matrix that maps the vector ω_1, say, of elements $\omega_{s(t-1)} d(s<t)$ into the vector ω of elements ω_{ts}: $\omega = D\omega_1$. Hence the appropriate weighting matrix in this case is

$$\hat{V}^* = (I - \tilde{\rho}D)\hat{V}(I-\tilde{\rho}D)'$$

with elements

$$\hat{v}^*_{sts't'} = \hat{v}_{sts't'} - \tilde{\rho}\hat{v}_{s(t-1)s't'} d(s<t) - \tilde{\rho}\hat{v}_{sts'(t'-1)} d(s'<t') + \tilde{\rho}^2\hat{v}_{s(t-1)s'(t'-1)} d(s<t)d(s'<t')$$

where $\tilde{\rho}$ is the OLS estimate of ρ in the regression above.

(iv) *First differences with autoregressive errors in levels.*

This is the ARMA (1, 1) scheme

$$\omega_{st} = \begin{cases} \bar{g}_0 & \text{if } s=t \\ \bar{g}_1 & \text{if } s=t-1 \\ \rho\omega_{s(t-1)} & \text{if } s<t-1 \end{cases}$$

with the restriction $(1-\rho)\bar{g}_0 + 2\bar{g}_1 = 0$, which is nonlinear. However the distance function can be suitably transformed to produce linearity. We can write:

$$\omega_{st} = \begin{cases} \bar{g}_0 & \text{if } s=t \\ (\rho\omega_{s(t-1)} - \bar{g}_0)/2 & \text{if } s=t-1 \\ \rho\omega_{s(t-1)} & \text{if } s<t-1 \end{cases}$$

which gives rise to the linear regression

$$\hat{\omega}_{st} = \bar{g}_0[d(s=t) - \tfrac{1}{2}d(s=t-1)] + \rho\hat{\omega}_{s(t-1)}[d(s<t-1)+\tfrac{1}{2}d(s=t-1)] + \varepsilon_{st}.$$

In this case we use

$$\text{acôv}\,(\hat{\omega}_{st} - \tilde{\rho}\hat{\omega}_{s(t-1)} d^*_{st},\ \hat{\omega}_{s't'} - \tilde{\rho}\hat{\omega}_{s'(t'-1)} d^*_{s't'})$$

with $d^*_{st} = d(s<t-1) + d(s=t-1)/2$, as the elements of the weighting matrix.

The previous discussion has assumed that the elements in the first row of Ω^*, ω_{11} and ω_{12} have been left unrestricted and are not used in performing the calculation of

MCS or Wald statistics. However a T degrees of freedom test of $\omega_{12}=0$ is a test of the exogeneity of y_{i1}:

$$W_1 = N\hat{\omega}_{12}[\text{av\^ar}(\hat{\omega}_{21})]^{-1}\hat{\omega}_{21}.$$

This seems an odd hypothesis to test in the presence of individual effects, since the model postulates correlation between y_{i1} and η_i (W_1 is effectively measuring whether small T biases are statistically significant). It would be more natural to test the exogeneity of y_{i1} by testing the absence of individual effects and the lack of serial correlation in v_{it}.

4. SUMMARY AND CONCLUDING REMARKS CONCERNING EFFICIENT ESTIMATION

This paper has developed tests for specific schemes of autocorrelation after estimating by three-stage least squares a dynamic random effects model. We have derived the asymptotic distribution of unrestricted autocovariance matrix estimates without imposing the assumption of normal errors. In particular we have shown how MCS tests for various error schemes commonly found in practical applications can be calculated from simple generalised linear regressions involving the sample autocovariances and dummy variables.

If Ω^* is found to satisfy a set of restrictions $\omega=\omega(\psi)$, 3SLS estimates are inefficient. However a GLS estimator can be used to achieve asymptotic efficiency. The form of a GLS estimator of μ and δ is

$$\begin{pmatrix}\tilde{\mu}\\ \tilde{\delta}\end{pmatrix} = \left[\sum_{i=1}^{N}\begin{bmatrix}\tilde{\omega}^{11}z_iz_i' & z_i\tilde{\omega}^{12}X_i\\ X_i'\tilde{\omega}^{21}z_i' & X_i'\tilde{\Omega}^{22}X_i\end{bmatrix}\right]^{-1}\sum_{i=1}^{N}\begin{bmatrix}\tilde{\omega}^{11}z_iy_{i1}+z_i\tilde{\omega}^{12}y_i\\ X_i'\tilde{\omega}^{21}y_{i1}+X_i'\tilde{\Omega}^{22}y_i\end{bmatrix} \tag{11}$$

where

$$\tilde{\Omega}^* = \begin{bmatrix}\tilde{\omega}^{11} & \tilde{\omega}^{12}\\ \tilde{\omega}^{21} & \tilde{\Omega}^{22}\end{bmatrix}^{-1}$$

is some estimate of Ω^*. As before, let $\tilde{\omega}$ be a vector containing the elements in the upper triangle of $\tilde{\Omega}^*$. GLS estimates of triangular systems like (2) are only consistent if $\tilde{\omega}$ is consistent and they are asymptotically equivalent to 3SLS if $\tilde{\omega}$ is an efficient estimator in the absence of restrictions (cf. Lahiri and Schmidt (1978)). In Arellano (1989*b*) it is proved that under normality the GLS estimator that uses $\tilde{\omega}=\omega(\hat{\psi})$ is asymptotically efficient, and that more generally an optimal choice of $\tilde{\omega}$ is a matrix weighted average of $\hat{\omega}$ and $\omega(\hat{\psi})$ given by

$$\tilde{\omega} = (I-\hat{V}_0\hat{V}^{-1})\hat{\omega} + \hat{V}_0\hat{V}^{-1}\omega(\hat{\psi}) \tag{12}$$

where $\hat{V}$ is as defined earlier and $\hat{V}_0$ is an estimate of avar$(\hat{\omega})$ under normality, that is, its elements are sample counterparts of those in (10). The GLS estimator based on (12) is asymptotically efficient in the sense of attaining the same limiting distribution as the optimal joint minimum distance estimator of slope and covariance parameters.

APPENDIX

The following conventions are adopted: for any $m\times n$ matrix B, vec (B) is obtained by stacking the rows of B. The $mn\times mn$ commutation matrix K performs the transformation K vec $(B)=$ vec (B'). For a square $n\times n$ matrix A, $\nu(A)$ is the $n(n+1)/2$ column vector obtained stacking by rows the upper triangle of A. If A is symmetric $\nu(A)$ and vec(A) can be connected by mean of a $n^2\times n(n+1)/2$ duplication matrix D that maps $\nu(A)$ into vec (A), i.e. $D\nu(A)=$ vec (A). Furthermore, since $(D'D)$ is non-singular we also have $\nu(A)=L$ vec (A) with $L=(D'D)^{-1}D'$. For any $n\times n$ matrix A we have L vec $(A)=\nu(A+A')/2$ (cf. Magnus and Neudecker (1980)).

Lemma. *Let u_i $(i=1,\ldots,N)$ be an iid random vector sequence with $E(u_i)=0$, $E(u_iu_i')=\Omega$, $E(u_iu_i'\otimes u_i')L'=\Delta_3$ and $LE(u_iu_i'\otimes u_iu_i')L'=\Delta_4$, and let z_i be a random vector independent of u_i such that $plim(N^{-1}\sum_{i=1}^N z_i)=\bar{z}$ and $plim(N^{-1}\sum_{i=1}^N z_iz_i')=M$. Letting $\omega=\nu(\Omega)$, $\xi_{1i}=u_i\otimes z_i$, $\xi_{2i}=L(u_i\otimes u_i)-\omega$ and $\xi_i=(\xi_{1i}'\ \xi_{2i}')'$ we have*

$$N^{-1/2}\sum_{i=1}^N \xi_i \xrightarrow{d} N\left[0,\begin{pmatrix}\Omega\otimes M & (I\otimes\bar{z})\Delta_3\\ \Delta_3'(I\otimes\bar{z}') & \Delta_4-\omega\omega'\end{pmatrix}\right].$$

Proof. Since ξ_i is independently distributed, the Lemma follows from the Liapunov central limit theorem. ||

We first obtain the distribution of structural covariance estimates in a linear simultaneous equations system with unrestricted constant terms and then specialise the result to our model. Let

$$B(\theta)y_i+\Gamma(\theta)z_i+\gamma=A(\phi)x_i=u_i \tag{A.1}$$

where $\phi=(\theta'\gamma')'$and $x_i=(y_i'z_i'1)'$. We are concerned with statistics of the form

$$\hat{\Omega}=A(\hat{\phi})\left(\frac{1}{N}\sum_i x_ix_i'\right)A'(\hat{\phi}). \tag{A.2}$$

A first-order expansion of vec $(\hat{\Omega})$ about the true value ϕ gives

$$\begin{aligned}\text{vec}\,(\hat{\Omega})&=\frac{1}{N}\sum_i u_i\otimes u_i+\frac{\partial\,\text{vec}\,\Omega}{\partial\phi'}\cdot(\hat{\phi}-\phi)+o_p(1)\\&=\frac{1}{N}\sum_i u_i\otimes u_i+(I+K)(I\otimes\Omega B'^{-1})\frac{\partial\,\text{vec}\,B}{\partial\theta'}(\hat{\theta}-\theta)+o_p(1).\end{aligned}$$

This is so in view of

$$\frac{\partial\,\text{vec}\,(\Omega)}{\partial\phi'}=(I+K)\left(I\otimes\frac{1}{N}\sum_i u_ix_i'\right)\frac{\partial\,\text{vec}\,A}{\partial\phi'}\xrightarrow{p}(I+K)[I\otimes(\Omega B'^{-1}:0)]\frac{\partial\,\text{vec}\,A}{\partial\phi'}$$

and the fact that the partial derivatives of vec (B) with respect to γ vanish. Finally

$$\sqrt{N}(\hat{\omega}-\omega)=N^{-1/2}\sum_i[L(u_i\otimes u_i)-\omega]+2L(I\otimes\Omega B'^{-1})\frac{\partial\,\text{vec}\,B}{\partial\theta'}\sqrt{N}(\hat{\theta}-\theta)+o_p(1) \tag{A.3}$$

where we have used the fact that $L(I+K)=2L$

If $\hat{\theta}$ is the 3SLS estimator, it minimises

$$s(\theta)=\frac{1}{N}\left(\sum_i u_i^*\otimes z_i^*\right)'\left(\tilde{\Omega}\otimes\sum_i z_i^*z_i^{*\prime}\right)^{-1}\left(\sum_i u_i^*\otimes z_i^*\right)$$

where u_i^* and z_i^* are deviations from sample means of u_i and z_i respectively, and $\tilde{\Omega}$ is the 2SLS estimate of Ω. Note that $s(\theta)$ is the 3SLS criterion function concentrated with respect to γ. Expanding $\partial s(\hat{\theta})/\partial\theta$ about the true value θ we obtain

$$\sqrt{N}(\hat{\theta}-\theta)=\text{avar}\,(\hat{\theta})(\partial\zeta/\partial\theta')\left(\Omega\otimes\sum_i z_i^*z_i^{*\prime}N\right)^{-1}N^{-1/2}\sum_i(u_i\otimes z_i^*)+o_p(1) \tag{A.4}$$

where $\zeta=N^{-1}\sum_i u_i^*\otimes z_i^*$ and we have made use of

$$N^{-1/2}\sum_i(u_i^*\otimes z_i^*)=N^{-1/2}\sum_i(u_i\otimes z_i^*)+o_p(1)$$

Combining (A.3) and (A.4), a limiting normal distribution for $\sqrt{N}(\hat{\omega}-\omega)$ can be obtained using the Lemma. Moreover, since plim $N^{-1}\sum_i z_i^*=0$, the two terms on the R.H.S. of (A.3) are asymptotically independent so that

$$\text{avar}\,(\hat{\omega})=H\,\text{avar}\,(\hat{\theta})H'\Delta_4-\omega\omega' \tag{A.5}$$

with

$$H=2L(I\otimes\Omega B'^{-1})(\partial\,\text{vec}\,B/\partial\theta')$$

Note that a similar result can be obtained if the constant terms are restricted but $\Delta_3=0$. The elements of avar $(\hat{\omega})$ in (A.5) have the form

$$\text{acov}\,(\hat{\omega}_{ts},\hat{\omega}_{t's'})=\sum_{jk}\text{acov}\,(\hat{\theta}_j,\hat{\theta}_k)h_{jts}h_{kt's'}+\mu_{tst's'}-\omega_{ts}\omega_{t's'} \tag{A.6}$$

where h_{jts} is the (t, s) element of the matrix

$$H_j = (\partial B/\partial \theta_j)B^{-1}\Omega + \Omega B'^{-1}(\partial B/\partial \theta_j)'.$$

Turning now to the random effects model (2), in this case the elements of the matrix B are given by

$$b_{ts} = \begin{cases} 1 & \text{if } t = s \\ -\alpha & \text{if } t = s+1 \\ 0 & \text{otherwise} \end{cases}$$

and the elements of B^{-1} are

$$b^{ts} = \begin{cases} 1 & \text{if } t = s \\ \alpha^{(t-s)} & \text{if } t > s \\ 0 & \text{otherwise} \end{cases}$$

Hence, only derivatives of B with respect to α are retained in (A.6), thus obtaining

$$\text{acov}\,(\hat{\omega}_{ts}, \hat{\omega}_{t's'}) = \text{avar}\,(\hat{\alpha})a_{ts}a_{t's'} + \mu_{tst's} - \omega_{ts}\omega_{t's'}$$

where a_{ts} is the (t, s) element of the matrix:

$$H_\alpha = -(\partial B/\partial \alpha)B^{-1}\Omega - \Omega B'^{-1}(\partial B/\partial \alpha)'$$

which completes the proof of the Theorem. ||

Acknowledgement. I would like to thank Denis Sargan for his constant help and advice in the preparation of Arellano (1985) on which this paper is partly based. Special thanks are due to Olympia Bover for innumerable discussions on this material and encouragement. I am also indebted to Steve Nickell for various useful discussions, and to an anonymous referee and the participants of the econometrics workshops at LSE and Oxford for helpful comments. Responsibility for any errors is mine alone. Financial support from an LSE Suntory-Toyota scholarship and from Caja de Ahorros de Barcelona is gratefully acknowledged.

REFERENCES

ARELLANO, M. (1985), "Estimation and Testing of Dynamic Econometric Models from Panel Data" (unpublished Ph.D. Thesis, London School of Economics).

ARELLANO, M. (1989*a*), "On the Efficient Estimation of Simultaneous Equations with Covariance Restrictions", *Journal of Econometrics*, **41**, 247-265.

ARELLANO, M. (1989*b*), "An Efficient GLS Estimator of Triangular Models with Covariance Restrictions", *Journal of Econometrics*, **41**, 267-273.

BHARGAVA, A. and SARGAN, J. D. (1983), "Estimating Dynamic Random Effects Models from Panel Data Covering Short Time Periods", *Econometrica*, **51**, 1635-1659.

BHARGAVA, A. (1987), "Wald Tests and Systems of Stochastic Equations", *International Economic Review*, **28**, 789-808.

CHAMBERLAIN, G. (1984), "Panel Data", in Griliches, Z. and Intriligator, M. D. (eds.), *Handbook of Econometrics, Volume II* (Amsterdam: Elsevier Science Publishers).

LAHIRI, K. and SCHMIDT, P. (1978) "On the Estimation of Triangular Structural Systems", *Econometrica*, **46**, 1217-1221.

MACURDY, T. E. (1981), "Asymptotic Properties of Quasi-Maximum Likelihood Estimators and Test Statistics" (National Bureau of Economic Research Technical Paper No. 14).

MACURDY, T. E. (1982), "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis", *Journal of Econometrics*, **18**, 83-114.

MAGNUS, J. R. and NEUDECKER, H. (1980), "The Elimination Matrix: Some Lemmas and Applications", *SIAM Journal of Algebraic and Discrete Methods*, **1**, 422-449.