Journal of Econometrics 42 (1989) 267–273. North-Holland

AN EFFICIENT GLS ESTIMATOR OF TRIANGULAR MODELS WITH COVARIANCE RESTRICTIONS*

Manuel ARELLANO

Institute of Economics and Statistics, Oxford OX1 3UL, England

Received May 1987, final version received November 1987

A GLS estimator of a triangular model with covariance restrictions is proposed that is asymptotically equivalent to the optimal minimum distance estimator regardless of nonnormality. The estimator used for the error covariance matrix Σ is a matrix-weighted average of the efficient restricted and unrestricted estimators of Σ. The relative weights are determined by a measure of the departure from normality.

1. Introduction

The purpose of this article is to propose an asymptotically efficient generalized least squares estimator of the structural parameters in a simultaneous equations system where the matrix of coefficients of the endogenous variables has a triangular structure and the error covariance matrix is constrained but independent of the slope parameters. Normality is not imposed, and asymptotic efficiency is defined to be relative to the (joint) optimal minimum distance (MD) estimator of slope and covariance parameters.

Models that can be represented as a triangular system with covariance restrictions have appeared in a variety of contexts. These include recursive models in which the error covariance matrix is required to be diagonal, but also certain models for panel data. For example, consider a dynamic random effects single-equation model observed over a fixed number of periods:

$$y_{it} = \alpha y_{i(t-1)} + \beta x_{it} + u_{it},$$

$$u_{it} = \eta_i + v_{it}, \qquad i = 1,\dots N, \quad t = 2,\dots,T,$$

$$\mathrm{E}(y_{i1}|x_{i1},\dots,x_{iT}) = \lambda_1 x_{i1} + \cdots + \lambda_T x_{iT}.$$

This model can be represented as a triangular system of T equations with linear cross-equation restrictions and an error covariance matrix that exhibits

*This article is based on material included in a paper I presented at the European Meeting of the Econometric Society, Budapest, 1986, under the title 'Estimating Simultaneous Equations with Covariance Restrictions'. I wish to thank J.D. Sargan for his help, advice, and encouragement. Responsibility for any errors is mine alone.

the typical error components structure. Models of this type have been analysed by Bhargava and Sargan (1983), Chamberlain (1984), and Arellano (1985).

The basic discussion on the properties of GLS estimators in triangular systems without covariance restrictions is due to Lahiri and Schmidt (1978). They showed that GLS estimators of triangular systems are only consistent if they use consistent estimators of the error covariance matrix, and that they are only asymptotically equivalent to full information simultaneous equations estimators (e.g., QML, 3SLS) if they are based on efficient estimators of the error covariance matrix. Thus, as Lahiri and Schmidt point out, if no prior information on variances and covariances is available, GLS estimators of triangular systems are not too interesting, except perhaps as an algorithm for the computation of the QML estimator.

However, in the presence of covariance restrictions, 3SLS and covariance restricted QML estimators are both inefficient and cannot be ordered [unless normality holds – see Arellano (1989)], and the optimal MD estimator is difficult to compute. Thus it is of interest to investigate whether there exists a simpler GLS estimator that attains the same efficiency as the optimal MD estimator.

2. Preliminaries

We consider the model

$$BY' + \Gamma Z' = AX' = U', \tag{1}$$

where Y' is the $n \times N$ matrix of N observations on n endogenous variables, Z' is the $k \times N$ matrix of exogenous variables, and U' is the $n \times N$ matrix of structural errors whose ith column is given by u_i. The elements of the $n \times (n+k)$ coefficient matrix $A = (B \vdots \Gamma)$ are linear functions of a $p \times 1$ vector of parameters θ:

$$\text{vec}(A) = S\theta - s,$$

where S and s are, respectively, a $n(n+k) \times p$ matrix and a $n(n+k) \times 1$ vector of known constants, B is lower triangular and nonsingular, and the errors are assumed to be independent and identically distributed with zero mean and finite moments up to the fourth order, such that, if we let $m = n(n+1)/2$,

$$\text{E}(u_i u_i') = \Sigma,$$

$$\text{E}\left[u_i(u_i' \otimes u_i')\right] L' = \Delta_3,$$

$$L\, \text{E}(u_i u_i' \otimes u_i u_i') L' = \Delta_4,$$

where L is a $m \times n^2$ selection matrix that eliminates from Δ_3 and Δ_4 some of the repeated cross-moments.[1] Σ is assumed to be nonsingular and its elements are linear homogeneous functions of a $q \times 1$ vector of parameters γ:

$$\nu(\Sigma) = \sigma = G\gamma,$$

where G is an $m \times q$ matrix of known constants. We assume that Z is independent of U and that $\operatorname{plim}_{N\to\infty} N^{-1}(Z'Z) = M$ exists and is nonsingular. We also assume that $\Delta_3 = 0$ or, alternatively, $\operatorname{plim}_{N\to\infty} Z'\iota/N = 0$, where ι is a $N \times 1$ vector of ones. This is a simplifying assumption which ensures that the formulae below will not depend on third-order moments [see Arellano (1989)].

Let $\hat{\theta}$ be a Σ-unrestricted efficient estimator of θ (e.g., 3SLS) and let $\hat{\Sigma}$ be

$$\hat{\Sigma} = \frac{1}{N} A(\hat{\theta})(X'X)A'(\hat{\theta}),$$

with $\hat{\sigma} = \nu(\hat{\Sigma})$. Under the assumptions of our model, $\sqrt{N}(\hat{\theta} - \theta)$ and $\sqrt{N}(\hat{\sigma} - \sigma)$ have a joint limiting normal distribution with mean zero and covariance matrix with blocks given by

$$V_{\theta\theta} = \left[S'(\Sigma^{-1} \otimes \Pi M \Pi')S\right]^{-1}, \tag{2}$$

$$V_{\sigma\theta} = 2L(I \otimes M_{ux})SV_{\theta\theta}, \tag{3}$$

$$V_{\sigma\sigma} = (\Delta_4 - \sigma\sigma') + V_{\sigma\theta}V_{\theta\theta}^{-1}V_{\theta\sigma}, \tag{4}$$

where

$$\Pi' = \left(-\Gamma'B'^{-1} \vdots I_k\right) \quad \text{and} \quad M_{ux} = \operatorname{plim}(U'X/N) = \left(\Sigma B'^{-1} \vdots 0\right).$$

Moreover, let $\tilde{\theta}_{\text{OMD}}$ and $\tilde{\gamma}_{\text{OMD}}$ be the joint optimal MD estimators of θ and γ based on the statistics $Y'Z(Z'Z)^{-1}$ and $(Y'Y - Y'Z(Z'Z)^{-1}Z'Y)/N$. The avm's of $\tilde{\theta}_{\text{OMD}}$ and $\tilde{\gamma}_{\text{OMD}}$ are, respectively, given by

$$C_{\theta\theta} = V_{\theta\theta} - V_{\theta\sigma}V_{\sigma\sigma}^{-1}(I - P_v)V_{\sigma\sigma}(I - P_v)'V_{\sigma\sigma}^{-1}V_{\sigma\theta}, \tag{5}$$

$$C_{\gamma\gamma} = \left(G'V_{\sigma\sigma}^{-1}G\right)^{-1}, \tag{6}$$

[1] The following conventions are adopted: For any matrix B, vec(B) is obtained by stacking the rows of B. For an $n \times n$ symmetric matrix A, $\nu(A)$ is the $n(n+1)/2$ column vector obtained stacking by rows the lower triangle of A. $\nu(A)$ 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., $D\nu(A) = \text{vec}(A)$. Furthermore, since $(D'D)$ is nonsingular we also have $\nu(A) = L\,\text{vec}(A)$ with $L = (D'D)^{-1}D'$. The properties of D and L are extensively studied in Magnus and Neudecker (1980).

where $P_v = G(G'V_{\sigma\sigma}^{-1}G)^{-1}G'V_{\sigma\sigma}^{-1}$. Expression (2) is the well-known avm of standard simultaneous equations estimators. Expressions (3) to (6) are adaptions of formulae given in Arellano (1985, 1989) where proofs can also be found. The following expression will also be needed below:

$$A_{\sigma\sigma} = 2L(\Sigma \otimes \Sigma)L' + V_{\sigma\theta}V_{\theta\theta}^{-1}V_{\theta\sigma}. \tag{7}$$

Under normality, $V_{\sigma\sigma} = A_{\sigma\sigma}$ and fourth-order moments are no longer required.

Finally, let $\tilde{\gamma}$ be a separate MD estimator of γ based on $\hat{\sigma}$ of the form

$$\tilde{\gamma} = \left(G'\hat{V}_{\sigma\sigma}^{-1}G\right)^{-1}G'\hat{V}_{\sigma\sigma}^{-1}\hat{\sigma}, \tag{8}$$

where $\hat{V}_{\sigma\sigma}$ is a consistent estimator of $V_{\sigma\sigma}$. Clearly, the avm of $\tilde{\gamma}$ is also given by $C_{\gamma\gamma}$ and $\tilde{\gamma}$ is therefore asymptotically efficient.

3. Generalized least squares estimators

GLS estimators of θ are of the form

$$\tilde{\theta}_{\text{GLS}} = \left[S'(\Sigma^{*-1} \otimes X'X)S\right]^{-1}S'(\Sigma^{*-1} \otimes X'X)s. \tag{9}$$

As showed by Lahiri and Schmidt (1978), the consistency of Σ^* is necessary for the consistency of $\tilde{\theta}_{\text{GLS}}$. Here we assume that $\sigma^* = \nu(\Sigma^*)$ is some consistent estimator of σ that can be written as a transformation of $\hat{\sigma}$:

$$\sigma^* = P\hat{\sigma},$$

where P is an $m \times m$ matrix such that $(I - P)G = 0$. If $P = I$, the resulting GLS estimator has avm given by $V_{\theta\theta}$. More generally choices of P of the form

$$I - P = K\left(I - G(G'QG)^{-1}G'Q\right),$$

where K and Q are $m \times m$ nonsingular matrices, will all produce consistent estimates of σ. For example, if $K = I$ and Q is a consistent estimator of $A_{\sigma\sigma}^{-1}$, then we obtain an estimator of σ which is asymptotically equivalent to the Σ-restricted QMLE, and since (9) is the form of the first-order conditions for the QMLE, the associated GLS estimator of θ will be asymptotically equivalent to the Σ-restricted QMLE.

The asymptotic distribution of $\tilde{\theta}_{\text{GLS}}$ for arbitrary P can be easily obtained by relating this estimator to $\hat{\theta}$ and $\hat{\sigma}$ whose joint asymptotic distribution is known. It can be shown that (a detailed derivation is given in the appendix)

$$\sqrt{N}\left(\tilde{\theta}_{\mathrm{GLS}}-\theta\right)=\sqrt{N}\left(\hat{\theta}-\theta\right)-V_{\theta\sigma}A_{\sigma\sigma}^{-1}(I-P)\sqrt{N}\left(\hat{\sigma}-\sigma\right)+\mathrm{o}_p(1), \tag{10}$$

so that

$$\operatorname{avm}\left(\tilde{\theta}_{\mathrm{GLS}}\right)=V_{\theta\theta}-V_{\theta\sigma}A_{\sigma\sigma}^{-1}\left[(I-P)A_{\sigma\sigma}+A_{\sigma\sigma}(I-P)' - (I-P)V_{\sigma\sigma}(I-P)'\right]A_{\sigma\sigma}^{-1}V_{\sigma\theta}. \tag{11}$$

Direct comparison of (10) and (11) with (5) immediately reveals that there exists an optimal choice for $(I-P)$ which will replace $A_{\sigma\sigma}$ by the optimal norm $V_{\sigma\sigma}$ and will use P_v as the projector, hence setting

$$(I-P)=A_{\sigma\sigma}V_{\sigma\sigma}^{-1}(I-P_v), \tag{12}$$

in which case $\operatorname{avm}(\tilde{\theta}_{\mathrm{GLS}})=C_{\theta\theta}$.

It is interesting to notice that the GLS estimator that uses the efficient estimator of σ, $\tilde{\sigma}=G\tilde{\gamma}$, as the choice for σ^* is inefficient.

The transformation defined by (12) corresponds to the following choice of σ^*:

$$\bar{\bar{\sigma}}=\left(I-\hat{A}_{\sigma\sigma}\hat{V}_{\sigma\sigma}^{-1}\right)\hat{\sigma}+\left(\hat{A}_{\sigma\sigma}\hat{V}_{\sigma\sigma}^{-1}\right)\tilde{\sigma},$$

where $\hat{A}_{\sigma\sigma}$ and $\hat{V}_{\sigma\sigma}$ are consistent estimators of $A_{\sigma\sigma}$ and $V_{\sigma\sigma}$, respectively. Thus $\bar{\bar{\sigma}}$ is defined as a matrix-weighted average of $\hat{\sigma}$ and $\tilde{\sigma}$. Under normality $V_{\sigma\sigma}=A_{\sigma\sigma}$ and then $\bar{\bar{\sigma}}=\tilde{\sigma}$, but as $V_{\sigma\sigma}$ becomes large relative to $A_{\sigma\sigma}$, an increasing weight is being put on $\hat{\sigma}$ relative to $\tilde{\sigma}$. This also suggests that the efficiency gain that is obtained by enforcing the covariance restrictions over Σ-unrestricted estimators, using MD methods, is smaller when the standardized fourth-order moments are large.

Finally, note that the previous results generalise to nonlinear restrictions both in A and Σ.

4. Conclusion

We have proposed an efficient GLS estimator of a triangular model with covariance restrictions that uses as the estimator for the error covariance matrix Σ a linear combination of the efficient restricted and unrestricted estimators of Σ. The relative weights are determined by a measure of the departure from normality. Under normality, the unrestricted estimator of Σ is given a zero weight, and the GLS estimator is asymptotically equivalent to the Σ-restricted QML estimator.

Appendix

We may rewrite (9) as

$$\left[S'\left(\Sigma^{*-1}\otimes\frac{X'X}{N}\right)S\right]\sqrt{N}\left(\tilde{\theta}_{\mathrm{GLS}}-\theta\right)=-\frac{1}{\sqrt{N}}S'\operatorname{vec}(\Sigma^{*-1}U'X). \tag{A.1}$$

Using a first-order expansion of the RHS of (A.1) about σ, we obtain

$$\begin{aligned}&\left[S'(\Sigma^{-1}\otimes M_{xx})S\right]\sqrt{N}\left(\tilde{\theta}_{\mathrm{GLS}}-\theta\right)\\&=-\frac{1}{\sqrt{N}}S'\operatorname{vec}(\Sigma^{-1}U'X)\\&\quad+S'(I\otimes M_{xu})(\Sigma^{-1}\otimes\Sigma^{-1})D\sqrt{N}\left(\sigma^{*}-\sigma\right)+\mathrm{o}_p(1),\end{aligned} \tag{A.2}$$

where

$$M_{xx}=\operatorname{plim}(X'X/N)=\Pi M\Pi'+M_{xu}\Sigma^{-1}M_{ux}.$$

Now letting $\hat{\theta}$ and $\hat{\sigma}$ be the Σ-unrestricted QML estimators of θ and σ, since (9) is the form of the first-order conditions for these estimators, an expression similar to (A.2) can be written where $\tilde{\theta}_{\mathrm{GLS}}$ and σ^{*} are replaced by $\hat{\theta}$ and $\hat{\sigma}$. Subtracting this expression from (A.2) we have

$$\begin{aligned}&\left[S'(\Sigma^{-1}\otimes M_{xx})S\right]\sqrt{N}\left(\tilde{\theta}_{\mathrm{GLS}}-\hat{\theta}\right)\\&=S'(I\otimes M_{xu})(\Sigma^{-1}\otimes\Sigma^{-1})D\sqrt{N}\left(\sigma^{*}-\hat{\sigma}\right)+\mathrm{o}_p(1).\end{aligned}$$

Let

$$H=2L(I\otimes M_{ux})S\quad\text{and}\quad K=2L(\Sigma\otimes\Sigma)L',$$

whose inverse is given by [cf. Richard (1975)]

$$K^{-1}=\tfrac{1}{2}D'(\Sigma^{-1}\otimes\Sigma^{-1})D.$$

Using properties of D and L given by Magnus and Neudecker (1980), and the triangularity of B, it can be verified that

$$S'(\Sigma^{-1} \otimes M_{xx})S = V_{\theta\theta}^{-1} + H'K^{-1}H$$

and

$$S'(I \otimes M_{xu})(\Sigma^{-1} \otimes \Sigma^{-1})D = H'K^{-1}.$$

Therefore,

$$\sqrt{N}\left(\tilde{\theta}_{\mathrm{GLS}} - \hat{\theta}\right) = \left(V_{\theta\theta}^{-1} + H'K^{-1}H\right)^{-1} H'K^{-1}\sqrt{N}(\sigma^* - \hat{\sigma}) + \mathrm{o}_p(1),$$

and using the matrix inversion lemma,

$$\sqrt{N}\left(\tilde{\theta}_{\mathrm{GLS}} - \hat{\theta}\right) = V_{\theta\theta}H'(K + HV_{\theta\theta}H')^{-1}\sqrt{N}(\sigma^* - \hat{\sigma}) + \mathrm{o}_p(1).$$

Finally, using $\sigma^* = P\hat{\sigma}$, $(I - P)\sigma = 0$, $V_{\theta\sigma} = V_{\theta\theta}H'$, and $A_{\sigma\sigma} = K + HV_{\theta\theta}H'$, we can write

$$\sqrt{N}\left(\tilde{\theta}_{\mathrm{GLS}} - \theta\right) = \sqrt{N}(\hat{\theta} - \theta) - V_{\theta\sigma}A_{\sigma\sigma}^{-1}(I - P)\sqrt{N}(\hat{\sigma} - \sigma) + \mathrm{o}_p(1).$$

References

Arellano, M., 1985, Estimation and testing of dynamic econometric models from panel data, Unpublished Ph.D. thesis (London School of Economics, London).

Arellano, M., 1989, On the efficient estimation of simultaneous equations with covariance restrictions, this issue.

Bhargava, A. and J.D. Sargan, 1983, Estimating dynamic random effects models from panel data covering short time periods, Econometrica 51, 1635–1659.

Chamberlain, G., 1984, Panel data, Ch. 22 in: Z. Griliches and M.D. Intriligator, eds., Handbook of econometrics, Vol. II (North-Holland, Amsterdam).

Lahiri, K. and P. Schmidt, 1978, On the estimation of triangular structural systems, Econometrica 46, 1217–1221.

Magnus, J.R. and H. Neudecker, 1980, The elimination matrix: Some lemmas and applications, SIAM Journal of Algebraic and Discrete Methods 1, 422–449.

Richard, J.-F., 1975, A note on the information matrix of the multivariate normal distribution, Journal of Econometrics 3, 57–60.