Econometrica, Vol. 52, No. 1 (January, 1984)

NOTES AND COMMENTS

ON THE ASYMPTOTIC EFFICIENCY OF FEASIBLE AITKEN ESTIMATORS FOR SEEMINGLY UNRELATED REGRESSION MODELS WITH ERROR COMPONENTS

BY INGMAR R. PRUCHA

1. INTRODUCTION[1]

IN RECENT ARTICLES Avery [**4**] and Baltagi [**6**] considered a seemingly unrelated regression model with error components (ECSUR). Avery proposed a feasible generalized least squares (GLS) estimator based on ordinary least squares (OLS) residuals. Alternatively Baltagi suggested a feasible GLS estimator based on least squares dummy variable (LSDV) residuals and showed, among other things, that his feasible GLS estimator is asymptotically equivalent to the true GLS estimator. Baltagi [**6**, p. 1547 and 1550] also asserted that Avery's feasible GLS estimator is asymptotically inefficient (relative to true GLS). It is shown in the following that this assertion is incorrect; specifically, although Avery's estimator of the covariance components is asymptotically less efficient than Baltagi's, Avery's feasible GLS estimator is asymptotically efficient. Indeed, we give a general theorem which demonstrates, for the ECSUR model considered, the asymptotic equivalence of a wide class of feasible GLS estimators with the true GLS estimator and therefore with the seemingly unrelated dummy variable (SURDV) estimator.[2] Practically any feasible GLS estimator for the ECSUR model that is of interest will fall into this class, including Avery's and Baltagi's feasible GLS estimators.

2. THE MODEL

Consider the model of Baltagi [**6**]; let there be N cross sectional units observed over T periods and let there exist a set of M linear stochastic relationships of the form $(j = 1, \dots, M)$

$$(1) \qquad y_j = e_{NT}\alpha_j + X_j\beta_j + u_j, \qquad u_j = (I_N \otimes e_T)\mu_j + (e_N \otimes I_T)\lambda_j + v_j,$$

where y_j is the $NT \times 1$ vector of observations on the dependent variable and X_j is the $NT \times K_j$ matrix of observations on the K_j (exogeneous) explanatory variables of the jth equation and where e_{NT}, e_N, and e_T are $NT \times 1$, $N \times 1$, and $T \times 1$ vectors of ones, respectively. The parameter vector β_j is of the order $K_j \times 1$ and the scalar α_j represents the explicit intercept.[3] The ith element of the vector $\mu_j = [\mu_{1j}, \dots, \mu_{Nj}]'$ and the tth element of the vector $\lambda_j = [\lambda_{1j}, \dots, \lambda_{Tj}]'$ represent the error components specific to the ith unit and tth period (in the jth equation) respectively; the $NT \times 1$ vector v_j contains the error components (in the jth equation) specific to both. It proves convenient to rewrite model

[1] I would like to thank Roger Betancourt, Charlie Brown, Harry Kelejian, Eric Stud, and *Econometrica*'s editor and referees for their helpful comments. All errors remain my responsibility.

[2] The SURDV estimator is a generalization of the single equation LSDV estimator to the seemingly unrelated regression case. The asymptotic equivalence of the SURDV estimator and the true GLS estimator has been demonstrated in Baltagi [**5**].

[3] Note that Baltagi [**6**] measures his regressors as deviations from sample means and supresses the intercept. As expected, this formulation results in the same estimator for the slope parameters.

(1) in stacked notation as

$$y = (I_M \otimes e_{NT})\alpha + X\beta + u \tag{2}$$

where $y = [y_1', \ldots, y_M']'$, $u = [u_1', \ldots, u_M']'$, $X = \text{diag}_M(X_i)$, $\alpha = [\alpha_1, \ldots, \alpha_M]'$, and $\beta = [\beta_1', \ldots, \beta_M']'$. Define the error component vectors $\mu = [\mu_1', \ldots, \mu_M']'$, $\lambda = [\lambda_1', \ldots, \lambda_M']'$, and $v = [v_1', \ldots, v_M']'$; then μ, λ and v are assumed to be stochastically independent from each other with zero mean and finite fourth moments. Further $E(\mu_j\mu_l') = \sigma_{\mu jl}I_N$, $E(\lambda_j\lambda_l') = \sigma_{\lambda jl}I_T$, and $E(v_jv_l') = \sigma_{vjl}I_{NT}$ for $j, l = 1, \ldots, M$ with $\Omega_\mu = [\sigma_{\mu jl}]$, $\Omega_\lambda = [\sigma_{\lambda jl}]$, and $\Omega_v = [\sigma_{vjl}]$ positive definite. It is not assumed that the disturbances are normally distributed; however, zero covariance between disturbances is assumed to also indicate their independence. In particular, let $\Omega_v = PP'$; then it is assumed that $v = (P \otimes I_{NT})w$ where w is an $MNT \times 1$ vector of i.i.d. random variables with zero mean and unit variance. The above assumptions imply that the covariance matrix of the disturbance vectors of the jth and lth equation is given by $E(u_ju_l') = \sigma_{\mu jl}A + \sigma_{\lambda jl}B + \sigma_{vjl}I_{NT}$ with $A = I_N \otimes e_Te_T'$ and $B = e_Ne_N' \otimes I_T$. It is then readily seen [6, p. 1547] that the variance-covariance matrix of the stacked disturbance vector u can be written as

$$\Omega = E(uu') = \Omega_1 \otimes \left(\frac{A}{T} - \frac{J}{NT}\right) + \Omega_2 \otimes \left(\frac{B}{N} - \frac{J}{NT}\right) + \Omega_3 \otimes \frac{J}{NT} + \Omega_v \otimes Q \tag{3}$$

with $J = e_{NT}e_{NT}'$ and $Q = I_{NT} - A/T - B/N + J/NT$; further $\Omega_1 = \Omega_v + T\Omega_\mu$, $\Omega_2 = \Omega_v + N\Omega_\lambda$, and $\Omega_3 = \Omega_v + T\Omega_\mu + N\Omega_\lambda$. Finally, the matrices $X_j'X_l/NT$ and $X_j'QX_l/NT$ are assumed to tend to finite positive definite matrices as both N and T tend to infinity in every possible way ($j, l = 1, \ldots, M$).

4. A CLASS OF ASYMPTOTICALLY EFFICIENT FEASIBLE GLS ESTIMATORS

Baltagi [6, p. 1548] demonstrated that the GLS estimator of the vector of slope parameters β is given by $\hat{\beta}_{\text{GLS}} = (X'\Phi^{-1}X)^{-1}X'\Phi^{-1}y$ where $\Phi^{-1} = \Omega_1^{-1} \otimes (A/T - J/NT) + \Omega_2^{-1} \otimes (B/N - J/NT) + \Omega_v^{-1} \otimes Q$. Consider the covariance component estimators $\hat{\Omega}_\mu$, $\hat{\Omega}_\lambda$, and $\hat{\Omega}_v$. Then the corresponding feasible GLS estimator is defined as $\hat{\beta}_{\text{FGLS}} = (X'\hat{\Phi}^{-1}X)^{-1}X'\hat{\Phi}^{-1}y$ where $\hat{\Phi}^{-1} = \hat{\Omega}_1^{-1} \otimes (A/T - J/NT) + \hat{\Omega}_2^{-1} \otimes (B/N - J/NT) + \hat{\Omega}_v^{-1} \otimes Q$ with $\hat{\Omega}_1 = \hat{\Omega}_v + T\hat{\Omega}_\mu$ and $\hat{\Omega}_2 = \hat{\Omega}_v + N\hat{\Omega}_\lambda$. The proof of the following theorem is given in the Appendix.

THEOREM 1: *Suppose the assumptions of Section 2 are satisfied. Consider the class of feasible* GLS *estimators*, $\hat{\beta}_{\text{FGLS}}$, *with*

$$\text{plim}\,\hat{\Omega}_v = \Omega_v, \qquad \text{plim}\,\hat{\Omega}_\mu = \Omega_\mu^*, \qquad \text{plim}\,\hat{\Omega}_\lambda = \Omega_\lambda^*, \tag{4}$$

where Ω_μ^* *and* Ω_λ^* *are finite positive definite matrices.*[4] *Then any members of this class is asymptotically equivalent to the true* GLS *estimator in the sense that* $\text{plim}\sqrt{NT}(\hat{\beta}_{\text{FGLS}} - \hat{\beta}_{\text{GLS}}) = 0$.

REMARK 1: Condition (4) is rather weak and should be satisfied by any reasonable covariance component estimator. Consequently the above theorem makes case by case studies of the asymptotic properties of the feasible GLS estimators $\hat{\beta}_{\text{FGLS}} = (X'\hat{\Phi}^{-1}X)^{-1}X'\hat{\Phi}^{-1}y$ for the above defined ECSUR model essentially unnecessary.[5]

[4] Here and in the following (probability) limits are always understood to be taken as both N and T tend to infinity. Note that the consistency of the estimators for Ω_μ and Ω_λ is not assumed.

[5] For a wide class of asymptotically equivalent estimators in the single equation case, see Swamy and Arora [8].

Baltagi [5] showed the asymptotic equivalence of the seemingly unrelated dummy variable estimator, $\hat{\beta}_{\text{SURDV}} = [X'(\Omega_v^{-1} \otimes Q)X]^{-1}X'(\Omega_v^{-1} \otimes Q)y$, with the true GLS estimator. Hence the class of feasible GLS estimators defined in Theorem 1 is also asymptotically equivalent with the SURDV estimator; the limiting distribution of $\sqrt{NT}(\hat{\beta}_{\text{FGLS}} - \beta)$ is $N(0, M^{-1})$ with $M = \lim X'(\Omega_v^{-1} \otimes Q)X/NT$.

The analysis of variance estimators of the covariance components Ω_μ, Ω_λ and Ω_v are given by

$$\bar{\Omega}_\mu = \frac{1}{T(N-1)(T-1)} U'\left[\frac{T-1}{T}A - \frac{T-1}{NT}J - Q\right]U,$$

$$\bar{\Omega}_\lambda = \frac{1}{N(N-1)(T-1)} U'\left[\frac{N-1}{N}B - \frac{N-1}{NT}J - Q\right]U,$$

$$\bar{\Omega}_v = \frac{1}{(N-1)(T-1)} U'QU$$

with $U = [u_1, \ldots, u_M]$. It is not difficult to show that these estimators are unbiased and (given the fourth moments of the error components are finite) consistent. Let us now define two sets of feasible covariance component estimators. The first set of estimators, denoted as $\tilde{\Omega}_\mu$, $\tilde{\Omega}_\lambda$, and $\tilde{\Omega}_v$, is obtained by replacing in the above formulae the matrix U by the matrix of OLS residuals $\tilde{U} = [\tilde{u}_1, \ldots, \tilde{u}_M]$ with $\tilde{u}_j = [I_{NT} - J/NT - X_{*j}(X_{*j}'X_{*j})^{-1}X_{*j}']u_j$; the matrices $X_{*j} = (I_{NT} - J/NT)X_j$ are centered around sample means. The corresponding feasible GLS estimator, proposed by Avery [4], is denoted by $\tilde{\beta}_{\text{FGLS}}$. The second set of covariance component estimators, denoted as $\check{\Omega}_\mu$, $\check{\Omega}_\lambda$, and $\check{\Omega}_v$, is obtained by replacing the matrix U by $\check{U} = [\check{u}_1, \ldots, \check{u}_M]$ with $\check{u}_j = [I_{NT} - J/NT - X_{*j}(X_{*j}'QX_{*j})^{-1}X_{*j}'Q]u_j$ where $\check{u}_j$ is the vector of estimated residuals obtained by performing least squares with dummy variables (LSDV) on the jth equation. The corresponding feasible GLS estimator, proposed by Baltagi [6], is denoted as $\check{\beta}_{\text{FGLS}}$.[6] The proof of the following lemma is not difficult and hence omitted here.[7]

LEMMA 1: *Suppose the assumptions of Section 2 are satisfied. Then both, the covariance component estimators* $\tilde{\Omega}_v$, $\tilde{\Omega}_\mu$, *and* $\tilde{\Omega}_\lambda$ *based on* OLS *residuals and the covariance component estimators* $\check{\Omega}_v$, $\check{\Omega}_\mu$, *and* $\check{\Omega}_\lambda$ *based on* LSDV *residuals are consistent for, respectively,* Ω_v, Ω_μ, *and* Ω_λ *(independently of the relative speed with which* N *and* T *tend to infinity).*

Theorem 1 and Lemma 1 together imply the following corollary.

COROLLARY 1: *The feasible* GLS *estimator based on* OLS *residuals,* $\tilde{\beta}_{\text{FGLS}}$, *and that based on* LSDV *residuals,* $\check{\beta}_{\text{FGLS}}$, *are asymptotically equivalent to each other as well as to the true* GLS *estimator,* $\hat{\beta}_{\text{GLS}}$, *and the* SURDV *estimator,* $\hat{\beta}_{\text{SURDV}}$.

REMARK 2: Corollary 1 also proves that Avery's feasible GLS estimator, $\tilde{\beta}_{\text{FGLS}}$, based on OLS residuals is asymptotically efficient (relative to the true GLS estimator).

[6] The matrix J in the formulae for $\tilde{u}_j$ and $\check{u}_j$ stems from the intercept.

[7] For an explicit proof, see Prucha [7]. Note that the following lemma is not in contradiction with Amemiya's [1, p. 6] finding that the asymptotic distribution of the variance component estimators based on OLS residuals depends on the relative speed of increase of N and T.

While Theorem 1 demonstrates the asymptotic equivalence of a wide class of feasible GLS estimators for the ECSUR model it also suggests that, in order to be able to discriminate among different estimators, further research is needed on the small sample properties of those estimators.[8]

University of Maryland

Manuscript received December, 1981; final revision received March, 1983.

APPENDIX

PROOF OF THEOREM 1: In the following the assumptions of Theorem 1 are maintained. According to Theorem 8.4 of Theil [**9**] the following conditions are sufficient for the asymptotic equivalence of the feasible and true GLS estimators: $\operatorname{plim} X'(\hat{\Phi}^{-1} - \Phi^{-1})X/NT = 0$ and $\operatorname{plim} X'(\hat{\Phi}^{-1} - \Phi^{-1})u/\sqrt{NT} = 0$. It is readily seen that those conditions can be written more explicitly as

$$\text{(A.1)} \qquad \operatorname{plim}\left\{ (\hat{\sigma}_1^{jl} - \sigma_1^{jl}) \frac{X_j'(A/T - J/NT)X_l}{NT} + (\hat{\sigma}_2^{jl} - \sigma_2^{jl}) \frac{X_j'(B/N - J/NT)X_l}{NT} + (\hat{\sigma}_v^{jl} - \sigma_v^{jl}) \frac{X_j'QX_l}{NT} \right\} = 0 \qquad (j, l = 1, \ldots, M),$$

$$\text{(A.2)} \qquad \operatorname{plim} \sum_{l=1}^{M} \left\{ T^{3/4}(\hat{\sigma}_1^{jl} - \sigma_1^{jl}) \frac{X_j'(A/T - J/NT)u_l}{\sqrt{NT^{5/2}}} + N^{3/4}(\hat{\sigma}_2^{jl} - \sigma_2^{jl}) \frac{X_j'(B/N - J/NT)u_l}{\sqrt{N^{5/2}T}} + (\hat{\sigma}_v^{jl} - \sigma_v^{jl}) \frac{X_j Qu_l}{\sqrt{NT}} \right\} = 0 \qquad (j = 1, \ldots, M),$$

where $\hat{\sigma}_\alpha^{jl}$ and σ_α^{jl} denote the (j, l)th element of the inverse matrix of $\hat{\Omega}_\alpha$ and Ω_α, respectively ($\alpha = 1, 2, v$). By assumption the matrices $X_j'X_l/NT$ and $X_j'QX_l/NT$ and hence

$$X_j'((A/T) - (J/NT))X_l/NT$$

and

$$X_j'((B/T) - (J/NT))X_l/NT$$

all tend to finite limiting matrices as both N and T tend to infinity. It is furthermore readily seen that

$$\operatorname{plim} X_j'\left(\frac{A}{T} - \frac{J}{NT}\right)u_l/\sqrt{NT^{5/2}} = \operatorname{plim} X_j'\left(\frac{B}{N} - \frac{J}{NT}\right)u_l/\sqrt{N^{5/2}T} = 0.$$

It follows further from a standard central limit theorem that $X_j'Qu_l/\sqrt{NT} = X_j'Qv_l/\sqrt{NT}$ converges in distribution to a random variable.[9] Consequently, in order to prove that Conditions (A.1) and

[8]Another avenue of future research would be to generalize the present model to include lagged dependent variables among the regressors. See, for example, Anderson and Hsiao [**2, 3**] for recent work along those lines in the single equation case.

[9]Compare Theil [**9**, pp. 380–381]. Observe that the elements of v_l are i.i.d. and $\lim X_j'QX_j/NT$ is finite positive definite.

(A.2) are satisfied it is sufficient to show that

(A.3) $$\operatorname{plim}(\hat{\sigma}_v^{jl} - \sigma_v^{jl}) = \operatorname{plim} T^{3/4}\hat{\sigma}_1^{jl} = \operatorname{plim} T^{3/4}\sigma_1^{jl} = \operatorname{plim} N^{3/4}\hat{\sigma}_2^{jl} = \operatorname{plim} N^{3/4}\sigma_2^{jl} = 0.$$ [10]

Recall that $\Omega_v, \Omega_\mu, \Omega_\lambda, \Omega_\mu^*, \Omega_\lambda^*$ are finite positive definite matrices. The first probability limit in (A.3) then follows immediately from the consistency of $\hat{\Omega}_v$. That the second probability limit is zero can be shown as follows:

(A.4) $$\operatorname{plim} T^{3/4}\hat{\Omega}_1^{-1} = \operatorname{plim}(T^{-1/4})\operatorname{plim}\left[\frac{1}{T}\hat{\Omega}_v + \hat{\Omega}_\mu\right]^{-1} = \operatorname{plim}(T^{-1/4}) \cdot \Omega_\mu^{*-1} = 0.$$

The remaining probability limits of (A.3) follow in a completely analogous manner. Consequently Conditions (A.1) and (A.2), implying the asymptotic equivalence of $\hat{\beta}_{\mathrm{FGLS}}$ and $\hat{\beta}_{\mathrm{GLS}}$, are satisfied.
Q.E.D.

[10] Clearly this implies that $\operatorname{plim} T^{3/4}(\hat{\sigma}_1^{jl} - \sigma_1^{jl}) = \operatorname{plim}(\hat{\sigma}_1^{jl} - \sigma_1^{jl}) = \operatorname{plim} N^{3/4}(\hat{\sigma}_2^{jl} - \sigma_2^{jl}) = \operatorname{plim}(\hat{\sigma}_2^{jl} - \sigma_2^{jl}) = 0$. In [5] Baltagi gives a set of formal arguments in behalf of the assertion that Avery's feasible GLS estimator is asymptotically inefficient. His analysis is also based on a check of Theil's [9] Theorem 8.4. However, rather than considering probability limits of the kind $\operatorname{plim}(\hat{\sigma}_\alpha^{jl} - \sigma_\alpha^{jl})$ Baltagi [5, pp. 13–14 and 29–31] essentially checks the probability limits $\operatorname{plim}(\hat{\sigma}_{\alpha jl} - \sigma_{\alpha jl})$, $\alpha = v, 1, 2$. It is found that in the case in which the covariance component estimators are based on OLS residuals the latter probability limits depend for $\alpha = 1, 2$ on the relative speed with which T and N tend to infinity. Note, however, that since $\hat{\sigma}_{\alpha jl}$ and $\sigma_{\alpha jl}$, $\alpha = 1, 2$ tend to infinity as both T and N go to infinity this does not imply that also the (relevant) probability limits $\operatorname{plim}(\hat{\sigma}_\alpha^{jl} - \sigma_\alpha^{jl})$, $\alpha = 1, 2$, depend on the relative speed of increase of T and N. Intuitively speaking, from the fact that $\infty - \infty$ is indeterminate we cannot conclude that $1/\infty - 1/\infty$ is also indeterminate; in fact the latter difference is zero.

REFERENCES

[1] AMEMIYA, T.: "The Estimation of the Variances in a Variance-Components Model," *International Economic Review*, 12(1971), 1–13.
[2] ANDERSON, T. W., AND C. HSIAO: "Estimation of Dynamic Models with Error Components," *Journal of the American Statistical Association*, 76(1981), 598–606.
[3] ———: "Formulation and Estimation of Dynamic Models Using Panel Data," *Journal of Econometrics*, 18(1982), 47–82.
[4] AVERY, R. B.: "Error Components and Seemingly Unrelated Regressions," *Econometrica*, 45(1977), 199–209.
[5] BALTAGI, B. M.: "On Seemingly Unrelated Regressions with Error Components," unpublished manuscript, University of Houston, 1979.
[6] ———: "On Seemingly Unrelated Regressions with Error Components," *Econometrica*, 48(1980), 1547–1551.
[7] PRUCHA, I. R.; "On the Asymptotic Efficiency of Feasible Aitken Estimators for Seemingly Unrelated Regression Models with Error Components," University of Maryland, Department of Economics, Working Paper No. 15, 1981.
[8] SWAMY, P. A. V. B., AND S. S. ARORA: "The Exact Finite Sample Properties of the Estimators of Coefficients in the Error Components Regression Models," *Econometrica*, 40(1972), 261–275.
[9] THEIL, H.: *Principles of Econometrics*. New York: John Wiley and Sons, 1971.