Econometric Theory, **22**, 2006, 1191–1194. Printed in the United States of America.
DOI: 10.1017/S0266466606060567

AN ALTERNATIVE DERIVATION OF MUNDLAK'S FIXED EFFECTS RESULTS USING SYSTEM ESTIMATION

Badi H. Baltagi
Syracuse University

Mundlak (1978, *Econometrica* 46, 69–85) showed that the fixed effects estimator can be obtained as generalized least squares (GLS) for a panel regression model where the individual effects are random but are *all* hopelessly correlated with the regressors. This result was obtained by partitioned inversion after substituting the reduced form expression for the individual effects as a function of the means of *all* the regressors. This note shows that Mundlak's result can be obtained using *system estimation* without using partitioned inversion. System estimation has proved useful for deriving two-stage least squares (2SLS) and three-stage least squares (3SLS) counterparts for the random effects panel models by Baltagi (1981, *Journal of Econometrics* 17, 189–200). It also has been used for obtaining an alternative derivation of the Hausman tests that is robust to heteroskedasticity of unknown form (see Arellano, 1993, *Journal of Econometrics* 59, 87–97) and more recently, for obtaining generalized method of moments (GMM) estimators for dynamic panel models (see Arellano and Bover, 1995, *Journal of Econometrics* 68, 29–51; and Blundell and Bond, 1998, *Journal of Econometrics* 87, 115–143, to mention a few). We also show that a necessary and sufficient condition for ordinary least squares (OLS) to be equivalent to GLS is satisfied for this model.

1. MOTIVATION AND RESULTS

Mundlak (1978) considered a panel data regression model with error component disturbances

$$y_{it} = X_{it}'\beta + \mu_i + \nu_{it} \qquad i = 1,\ldots,N; \qquad t = 1,\ldots,T, \tag{1}$$

where the individual effects are a linear function of the averages of *all* the explanatory variables across time

$$\mu_i = \bar{X}_{i.}'\pi + \epsilon_i, \tag{2}$$

Address correspondence to Badi H. Baltagi, Syracuse University, Department of Economics and Center for Policy Research, 426 Eggers Hall, Syracuse, NY 13244-1020; e-mail: bbaltagi@maxwell.syr.edu.

 0266-4666/06 $12.00

where $\epsilon_i \sim \text{IIN}(0, \sigma_\epsilon^2)$, $\nu_{it} \sim \text{IIN}(0, \sigma_\nu^2)$, and $\bar{X}'_{i.}$ is $1 \times K$ vector of observations on the explanatory variables averaged over time. Mundlak showed that generalized least squares (GLS) on the resulting model,

$$y_{it} = X'_{it}\beta + \bar{X}'_{i.}\pi + \epsilon + \nu_{it} \qquad i = 1, \ldots, N; \qquad t = 1, \ldots, T, \tag{3}$$

yields

$$\hat{\beta}_{GLS} = \tilde{\beta}_{Within} = (X'QX)^{-1}X'Qy \tag{4}$$

and

$$\hat{\pi}_{GLS} = \hat{\beta}_{Between} - \tilde{\beta}_{Within} = (X'PX)^{-1}X'Py - (X'QX)^{-1}X'Qy \tag{5}$$

with

$$\begin{aligned} \text{var}(\hat{\pi}_{GLS}) &= \text{var}(\hat{\beta}_{Between}) + \text{var}(\tilde{\beta}_{Within}) \\ &= (T\sigma_\epsilon^2 + \sigma_\nu^2)(X'PX)^{-1} + \sigma_\nu^2(X'QX)^{-1}, \end{aligned} \tag{6}$$

where P is a matrix that averages the observation across time for each individual and $Q = I_{NT} - P$ is a matrix that obtains the deviations from individual means. This note gives an alternative derivation of this result using *system estimation*. Arellano (1993) applied system estimation to obtain an alternative derivation of the Hausman (1978) test. In fact, Arellano (1993) used the *forward orthogonal deviations* operator. Here, we use the usual *fixed effects* transformation. In particular, we write the panel model in vector form as

$$y = X\beta + PX\pi + \eta, \tag{7}$$

where $\eta = Z_\mu\epsilon + \nu$, $Z_\mu = I_N \otimes \iota_T$ with I_N denoting an identity matrix of dimension N and ι_T a vector of ones of dimension T. Here P is the projection matrix on Z_μ, i.e., $P = Z_\mu(Z'_\mu Z_\mu)^{-1}Z'_\mu = I_N \otimes \bar{J}_T$ where J_T is a matrix of ones of dimension T and $\bar{J}_T = J_T/T$. Premultiplying (7) by P one gets

$$Py = PX(\beta + \pi) + P\eta \tag{8}$$

because $P^2 = P$ and $PZ_\mu = Z_\mu$. Note that ordinary least squares (OLS) or GLS on (8) yields $(\widehat{\beta + \pi}) = (X'PX)^{-1}X'Py$, which is the usual between estimator of y on X. Similarly, premultiplying (7) by Q one gets

$$Qy = QX\beta + Q\eta \tag{9}$$

because $QP = 0$. OLS or GLS on (9) yields $\tilde{\beta}_{Within} = (X'QX)^{-1}X'Qy$, which is the usual within or fixed effects estimator of y on X. Stacking the system of equations (8) and (9), we get

$$\begin{pmatrix} Py \\ Qy \end{pmatrix} = \begin{pmatrix} PX \\ QX \end{pmatrix}\beta + \begin{pmatrix} PX \\ 0 \end{pmatrix}\pi + \begin{pmatrix} P\eta \\ Q\nu \end{pmatrix}, \tag{10}$$

https://doi.org/10.1017/S0266466606060567 Published online by Cambridge University Press

and the system error vector has mean 0 and variance-covariance matrix given by

$$\Sigma = \begin{pmatrix} \sigma_1^2 P & 0 \\ 0 & \sigma_\nu^2 Q \end{pmatrix}, \tag{11}$$

where $\sigma_1^2 = T\sigma_\epsilon^2 + \sigma_\nu^2$. This system estimation has been useful in deriving error components two-stage least squares (EC2SLS) and error components three-stage least squares (EC3SLS) (see Baltagi, 1981). It has also been used to derive GMM estimators for dynamic panel data models (see Arellano and Bover, 1995, and Blundell and Bond, 1998). For the Mundlak case, there is no need for partitioned inversion. In fact, the OLS normal equations on (10) yield

$$X'y = X'X\beta + X'PX\pi \tag{12}$$

and

$$X'Py = X'PX\beta + X'PX\pi \tag{13}$$

because $P + Q = I_{NT}$. Subtracting (13) from (12) one gets $X'Qy = (X'QX)\beta$, which yields $\tilde{\beta}_{Within} = (X'QX)^{-1}X'Qy$.

Solving (13) yields $(\widehat{\beta + \pi}) = (X'PX)^{-1}X'Py$. Similarly, the GLS normal equations on (10) yield

$$\left(\frac{X'Py}{\sigma_1^2} + \frac{X'Qy}{\sigma_\nu^2}\right) = \left(\frac{X'PX}{\sigma_1^2} + \frac{X'QX}{\sigma_\nu^2}\right)\beta + \left(\frac{X'PX}{\sigma_1^2}\right)\pi \tag{14}$$

and

$$\frac{X'Py}{\sigma_1^2} = \left(\frac{X'PX}{\sigma_1^2}\right)(\beta + \pi). \tag{15}$$

Equation (15) yields $(\widehat{\beta + \pi}) = (X'PX)^{-1}X'Py$. Subtracting (15) from (14) one gets $X'Qy = (X'QX)\beta$, which yields $\tilde{\beta}_{Within} = (X'QX)^{-1}X'Qy$. This proves that system OLS or GLS on (10) yields the same results that Mundlak found by applying GLS to (3).

In fact, one can prove that the Zyskind (1967) necessary and sufficient condition for OLS to be equivalent to GLS on the system of equations (10) is satisfied. This calls for $P_Z\Sigma = \Sigma P_Z$, where $Z = \begin{pmatrix} PX & PX \\ QX & 0 \end{pmatrix}$ is the matrix of regressors in (10) and Σ is the variance-covariance matrix of its disturbances. It is straightforward to show that

$$P_Z = Z(Z'Z)^{-1}Z' = \begin{pmatrix} PX(X'PX)^{-1}X'P & 0 \\ 0 & QX(X'QX)^{-1}X'Q \end{pmatrix},$$

https://doi.org/10.1017/S0266466606060567 Published online by Cambridge University Press

from which it follows that

$$P_Z \Sigma = \Sigma P_Z = \begin{pmatrix} \sigma_1^2 PX(X'PX)^{-1}X'P & 0 \\ 0 & \sigma_\nu^2 QX(X'QX)^{-1}X'Q \end{pmatrix}.$$

Note that the Hausman (1978) specification test based on the between minus within estimators is basically a test for $H_0, \pi = 0$ in (3), and this is based upon

$$\hat{\pi}'_{GLS}(\text{var}(\hat{\pi}_{GLS}))^{-1}\hat{\pi}_{GLS} \xrightarrow{H_0} \chi^2_K.$$

The $\text{var}(\hat{\pi}_{GLS})$ can be obtained from the GLS variance-covariance matrix of (10). This is given by the inverse of

$$(Z'\Sigma^{-1}Z) = \begin{pmatrix} \left(\dfrac{X'PX}{\sigma_1^2} + \dfrac{X'QX}{\sigma_\nu^2}\right) & \left(\dfrac{X'PX}{\sigma_1^2}\right) \\ \left(\dfrac{X'PX}{\sigma_1^2}\right) & \left(\dfrac{X'PX}{\sigma_1^2}\right) \end{pmatrix},$$

which can be easily shown by partitioned inversion to be

$$(Z'\Sigma^{-1}Z)^{-1} = \begin{pmatrix} \sigma_\nu^2(X'QX)^{-1} & -\sigma_\nu^2(X'QX)^{-1} \\ -\sigma_\nu^2(X'QX)^{-1} & \sigma_1^2(X'PX)^{-1} + \sigma_\nu^2(X'QX)^{-1} \end{pmatrix}.$$

Note that the second diagonal matrix is exactly the same as that given by (6), which completes the proof.

REFERENCES

Arellano, M. (1993) On the testing of correlated effects with panel data. *Journal of Econometrics* 59, 87–97.

Arellano, M. & O. Bover (1995) Another look at the instrumental variables estimation of error-component models. *Journal of Econometrics* 68, 29–51.

Baltagi, B.H. (1981) Simultaneous equations with error components. *Journal of Econometrics* 17, 189–200.

Blundell, R. & S. Bond (1998) Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* 87, 115–143.

Hausman, J.A. (1978) Specification tests in econometrics. *Econometrica* 46, 1251–1271.

Mundlak, Y. (1978) On the pooling of time series and cross-section data. *Econometrica* 46, 69–85.

Zyskind, G. (1967) On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *Annals of Mathematical Statistics* 36, 1092–1109.

https://doi.org/10.1017/S0266466606060567 Published online by Cambridge University Press