Econometric Theory, **20**, 2004, 223–229. Printed in the United States of America.
DOI: 10.1017/S0266466604201098

PROBLEMS AND SOLUTIONS

PROBLEMS

04.1.1. A Hausman Test Based on the Difference between Fixed Effects Two-Stage Least Squares and Error Components Two-Stage Least Squares

Badi H. Baltagi
Texas A&M University

Consider the following first structural equation of a simultaneous equation panel data model:

$$y_1 = Z_1 \delta_1 + u_1,$$

where $Z_1 = [Y_1, X_1]$ and $\delta_1' = (\gamma_1', \beta_1')$. As in the standard simultaneous equation literature, Y_1 is the set of g_1 right-hand-side endogenous variables, and X_1 is the set of k_1 included exogenous variables. Let $X = [X_1, X_2]$ be the set of all exogenous variables in the system. Let k_2, the number of excluded exogenous variables from the first equation (X_2), be larger than or equal to g_1. The disturbance u_1 follows the one-way error component model

$$u_1 = Z_\mu \mu_1 + \nu_1,$$

where $Z_\mu = (I_N \otimes \iota_T)$ with I_N being an identity matrix of dimension N and ι_T a vector of ones of dimension T. In this case, $\mu_1' = (\mu_{11}, \ldots, \mu_{N1})$ is a vector of random individual effects of dimension N and $\nu_1' = (\nu_{111}, \ldots, \nu_{NT1})$ is a vector of remainder disturbances of dimension NT. These vectors have zero means and covariance matrix

$$E\begin{pmatrix}\mu_1 \\ \nu_1\end{pmatrix}(\mu_1', \nu_1') = \begin{bmatrix}\sigma_{\mu_{11}}^2 I_N & 0 \\ 0 & \sigma_{\nu_{11}}^2 I_{NT}\end{bmatrix}.$$

Baltagi (1981) derives the error component two-stage least squares (EC2SLS) estimator of this model, which is the random effects (RE) counterpart of a classical error components panel data regression. Similarly, the fixed effects two-stage least squares (FE2SLS) estimator of this model is the fixed effects (FE) counterpart of a classical error components panel data regression. Both EC2SLS and FE2SLS allow for the endogeneity of Y_1.

 0266-4666/04 $12.00

Hausman (1978) suggests comparing the FE and RE estimators in the classical panel data regression. With endogenous right-hand-side regressors such as Y_1 this test can be generalized to test $H_0: E(u_1|Z_1) = 0$ based on $\hat{q}_1 = \tilde{\delta}_{1,FE2SLS} - \hat{\delta}_{1,EC2SLS}$.

(a) Show that under $H_0: E(u_1|Z_1) = 0$, $\text{plim}\,\hat{q}_1 = 0$ and the asymptotic $\text{cov}(\hat{q}_1, \hat{\delta}_{1,EC2SLS}) = 0$.
(b) Conclude that $\text{var}(\hat{q}_1) = \text{var}(\tilde{\delta}_{1,FE2SLS}) - \text{var}(\hat{\delta}_{1,EC2SLS})$, where var denotes the asymptotic variance. This is used in computing the Hausman test statistic given by

$$m_1 = \hat{q}_1'[\text{var}(\hat{q}_1)]^{-1}\hat{q}_1.$$

Under H_0, m_1 is asymptotically distributed as χ_r^2, where r denotes the dimension of the slope vector of the time varying variables in Z_1. This can be easily implemented using standard software packages, like STATA. The usual Hausman test can yield misleading inference in the presence of Y_1.

REFERENCES

Baltagi, B.H. (1981) Simultaneous equations with error components. *Journal of Econometrics* 17, 189–200.
Hausman, J. (1978) Specification tests in econometrics. *Econometrica* 46, 1251–1271.

04.1.2. Correcting for Heteroskedasticity of Unspecified Form

Tom Wansbeek
University of Groningen, The Netherlands

Consider the basic linear regression model with heteroskedasticity, $y_i = x_i'\beta + e_i$, with $i = 1, \ldots, n$, where e_i is a zero-mean random variable distributed independently from all x_i and all e_j with $i \neq j$. The variance of e_i is σ_i^2, where all $\sigma_i^2, i = 1, \ldots, n$ may be different. The "true" generalized least squares estimator (GLSE) $\hat{\beta}$ of β is obtained as the ordinary least squares estimator (OLSE) from the regression of the y_i/σ_i on the x_i/σ_i. When, as usual, the σ_i are unknown, we might instead be tempted to use the absolute value of the residual $\hat{\sigma}_i \equiv |y_i - x_i'b|$, with b the OLSE from the regression of the y_i on the x_i, and obtain a "feasible" GLSE $\tilde{\beta}$, say, by regressing the $y_i/\hat{\sigma}_i$ on the $x_i/\hat{\sigma}_i$. Intuitively, this is not a sensible approach, but one is hard put to find an argument in the literature. What are the (finite or asymptotic) distributional properties of $\tilde{\beta}$? (One might conjecture that $\tilde{\beta}$ is consistent and asymptotically normal.) Is there any sense in which $\tilde{\beta}$ constitutes an improvement over b?

https://doi.org/10.1017/S0266466604201098 Published online by Cambridge University Press

SOLUTIONS

03.1.1. Deriving the Observed Information Matrix in Ordered Probit and Logit Models Using the Complete-Data Likelihood Function—Solution

S.K. Sapra (the poser of the problem)
California State University, Los Angeles

The complete-data log likelihood function is

$$\ln L^*(y^*|x,\beta) = constant - \frac{1}{2}\sum_{i=1}^{n}(y_i^* - \beta' x_i)^2. \tag{2}$$

Define the categorical variables

$$\begin{aligned} Z_{ij} &= 1 \quad \text{if } y_i^* \text{ belongs to the } j\text{th category,} \\ &= 0 \quad \text{otherwise, } i = 1,2,\ldots,n,\ j = 1,2,\ldots,m. \end{aligned} \tag{3}$$

Therefore, the missing information matrix is given by

$$\begin{aligned} I_m &= \mathrm{Var}(\partial \ln L^*(y^*|x,\beta)/\partial\beta|Z_{ij}) \\ &= \mathrm{Var}\left(\sum_{i=1}^{n}(y_i^* - (\beta' x_i))x_i|Z_{ij} = z_{ij}\right) \\ &= \mathrm{Var}\left(\sum_{i=1}^{n}(u_i x_i)|Z_{ij} = z_{ij}\right) \\ &= \sum_{j=1}^{m}\sum_{i=1}^{n} z_{ij}[1 - ((\phi_{j-1,i} - \phi_{j,i})/(\Phi_{j-1,i} - \Phi_{j,i}))^2 \\ &\qquad + (w_{j-1,i}\phi_{j-1,i} - w_{j,i}\phi_{j,i})/(\Phi_{j-1,i} - \Phi_{j,i})]x_i x_i' \\ &= \sum_{i=1}^{n} x_i x_i' - \sum_{j=1}^{m}\sum_{i=1}^{n} z_{ij}/(\Phi_{j-1,i} - \Phi_{j,i})^2 \\ &\quad \times [(\phi_{j-1,i} - \phi_{j,i})^2 - (w_{j-1,i}\phi_{j-1,i} - w_{j,i}\phi_{j,i}) \\ &\qquad \times (\Phi_{j,i} - \Phi_{j-1,i})]x_i x_i', \end{aligned} \tag{4}$$

where $\phi_{j-1,i} = \phi(w_{j-1,i})$, $\Phi_{j-1,i} = \Phi(w_{j-1,i})$, $\phi_{j,i} = \phi(w_{j,i})$, $\Phi_{j-1,i} = \Phi(w_{j-1,i})$, $w_{j-1,i} = \alpha_{j-1} - \beta' x_i$, $w_{j,i} = \alpha_j - \beta' x_i$, and $\sum_{j=1}^{m} z_{ij} = 1$ because

$$\begin{aligned} \mathrm{Var}(u_i|Z_{ij} = z_{ij}) &= \sum_{j=1}^{m} z_{ij}\,\mathrm{Var}(u_i|\alpha_{j-1} < y_i^* < \alpha_j) \\ &= \sum_{j=1}^{m} z_{ij}[1 - ((\phi_{j-1,i} - \phi_{j,i})/(\Phi_{j-1,i} - \Phi_{j,i}))^2 \\ &\qquad + (w_{j-1,i}\phi_{j-1,i} - w_{j,i}\phi_{j,i})/(\Phi_{j-1,i} - \Phi_{j,i})], \end{aligned} \tag{5}$$

https://doi.org/10.1017/S0266466604201098 Published online by Cambridge University Press

where ϕ and Φ are, respectively, the probability density function and cumulative distribution functions of a $N(0,1)$ random variable using formulas for variances of doubly truncated distributions (Maddala, 1983, p. 366). Furthermore, the complete-data information matrix is given by

$$I_c = -E(\partial^2 \ln L^*(y^*|x,\beta)/\partial\beta\partial\beta') = \sum_{i=1}^{n} x_i x_i'. \quad \textbf{(6)}$$

From equations (4) and (6), it follows that the observed information matrix is

$$\begin{aligned} I_o &= I_c - I_m \\ &= \sum_{j=1}^{m}\sum_{i=1}^{n} z_{ij}/(\Phi_{j-1,i} - \Phi_{j,i})^2[(\phi_{j-1,i} - \phi_{j,i})^2 - (w_{j-1,i}\phi_{j-1,i} - w_{j,i}\phi_{j,i}) \\ &\qquad \times (\Phi_{j,i} - \Phi_{j-1,i})]x_i x_i', \end{aligned} \quad \textbf{(7)}$$

which is equal to the observed information matrix, $-\partial^2 \ln L/\partial\beta\partial\beta'$, obtained from the observed-data log likelihood function

$$\ln L(Z|x,\beta) = \sum_{j=1}^{m}\sum_{i=1}^{n} z_{ij} \ln[\Phi_{j,i} - \Phi_{j-1,i}]. \quad \textbf{(8)}$$

Alternatively, the observed information matrix may be computed by using the result that the observed score function is the conditional expectation of the latent score function given the observed variables (Louis, 1982, p. 227):

$$\begin{aligned} \partial \ln L(Zx,\beta)/\partial\beta &= E(\partial \ln L^*(Z|x,\beta)/\partial\beta|Z) \\ &= \sum_{i=1}^{n} E((y_i^* x_i - (\beta' x_i)x_i)|Z_{ij} = z_{ij}) \\ &= \sum_{j=1}^{m}\sum_{i=1}^{n} z_{ij}[(\phi_{j-1,i} - \phi_{j,i})/(\Phi_{j,i} - \Phi_{j-1,i})]x_i. \end{aligned} \quad \textbf{(9)}$$

By differentiation of the observed score function in equation (9) with respect to β', we obtain the observed information matrix in equation (7) as the negative of the hessian of the observed log likelihood function.

REFERENCES

Louis, T.A. (1982) Finding the observed information matrix when using the EM algorithm. *Journal of the Royal Statistical Society Series B* 44, 226–233.

Maddala, G.S. (1983) *Limited Dependent and Qualitative Variables in Econometrics*. Cambridge University Press.

https://doi.org/10.1017/S0266466604201098 Published online by Cambridge University Press

03.1.2. Redundancy of Lagged Regressors in a Conditionally Heteroskedastic Time Series Regression[1]—Solution

G. Dhaene
Katholieke Universiteit Leuven, Belgium

Let $g_{i,t} = (y_t - \beta x_t)x_{t-i}$ $(i = 0,1,\ldots)$. For any finite set I of nonnegative integers, let $\hat{\beta}_I$ be the optimal generalized method of moments estimator of β based on the moment conditions

$$Eg_{i,t} = 0, \qquad i \in I.$$

By the results of Breusch, Qian, Schmidt, and Wyhowski (1999), the ordinary least squares estimator, $\hat{\beta}_{\{0\}}$, is at least as efficient as any $\hat{\beta}_I$ if, and only if, for all $i \geq 1$, the condition $Eg_{i,t} = 0$ is redundant given $Eg_{0,t} = 0$; i.e.,

$$D_i = \Omega_{i0}\Omega_{00}^{-1}D_0, \qquad i \geq 1, \tag{1}$$

where $D_i = E(\partial g_{i,t}/\partial\beta)$ and $\Omega_{ij} = \sum_{l=-\infty}^{\infty} E(g_{i,t}g_{j,t-l})$. Let $\gamma_i = \text{Cov}(x_t, x_{t-i})$. Then, for $i \geq 0$,

$$D_i = -E(x_t x_{t-i}) = -\gamma_i,$$

$$\Omega_{i0} = E\sum_{l=-\infty}^{\infty} e_t x_{t-i} e_{t-l} x_{t-l} = E[(\omega + \lambda(x_t - \mu)^2)x_t x_{t-i}] = \frac{\gamma_i}{\gamma_0}\Omega_{00}, \tag{2}$$

so (1) is seen to hold. The last equality in (2) follows by writing

$$x_{t-i} = \frac{\gamma_i}{\gamma_0}x_t + u_t$$

with x_t and u_t independent. That this is possible follows from the bivariate normality of x_{t-i} and x_t.

NOTE

1. An excellent solution has been independently proposed by the S. Anatolyev, the poser of the problem.

REFERENCE

Breusch, T., H. Qian, P. Schmidt, & D. Wyhowski (1999) Redundancy of moment conditions. *Journal of Econometrics* 91, 89–111.

03.1.2 Redundancy of Lagged Regressors in a Conditionally Heteroskedastic Time Series Regression—Solution

Marine Carrasco
University of Rochester

A solution to the problem is to calculate the greatest lower bound for the asymptotic variance of GMM estimators based on moment function $h(i;\beta) = (y_t - \beta x_t)x_{t-i}$, $i = 0,1,2,\dots$. To do so, we apply Hansen (1985). In the sequel, we will denote $h(i)$ the moment functions $h(i;\beta_0)$. Let $L^2(h)$ be the linear space spanned by $\{h(i) : i = 0,1,\dots\}$; it is made of elements $G = \sum_{i=0}^{\infty} \alpha_i h(i)$ for any real α_i. Note that $\{h(i) : i = 0,1,\dots\}$ are martingale difference sequences. Moreover, they are linearly independent and in that sense are complete in $L^2(h)$.

We are looking for the solution of equation (4.8) of Hansen (1985). That is, we are looking for the element G in $L^2(h)$ (if it exists) such that

$$E\left[\frac{\partial h}{\partial \beta}(i)\right] = E[h(i)G] \quad \text{for all } i = 0,1,\dots. \tag{1}$$

Or, equivalently, we are looking for the constants α_j solutions of

$$-E(x_t x_{t-i}) = E\left[e_t x_{t-i}\left(\sum_{j=0}^{\infty} \alpha_j e_t x_{t-j}\right)\right]. \tag{2}$$

Note that the solution may not exist in $L^2(h)$, but there is always a solution in the closure of $L^2(h)$. If a solution in $L^2(h)$ exists, it is unique because $\{h(i) : i = 0,1,\dots\}$ is complete. Let $\alpha_j = 0$, for $j = 1,2,\dots$; equation (2) becomes

$$\begin{aligned}
-E(x_t x_{t-i}) &= \alpha_0 E[e_t^2 x_{t-i} x_t] \\
&= \alpha_0 E[(\omega + \lambda(x_t - \mu)^2) x_{t-i} x_t] \\
&= \alpha_0(\omega + \lambda\mu^2) E(x_t x_{t-i}) \\
&\quad + \alpha_0 \lambda E(x_t^3 x_{t-i}) \\
&\quad - \alpha_0 2\lambda\mu E(x_t^2 x_{t-i}).
\end{aligned}$$

Using the fact that η_t are i.i.d. standard normal, we have

$$E(x_t^2 x_{t-i}) = E\left[\left(\sum_l \varphi_l \eta_{t-l}\right)^2 \left(\sum_j \varphi_j \eta_{t-i-j}\right)\right] = 0,$$

$$\begin{aligned}
E(x_t^3 x_{t-i}) &= E\Bigg[\Bigg\{\sum_j \varphi_j^3 \eta_{t-j}^3 + 3\sum_{k\neq j} (\varphi_j^2 \eta_{t-j}^2)(\varphi_k \eta_{t-k}) \\
&\qquad + 6\sum_{k\neq j\neq l} (\varphi_j \eta_{t-j})(\varphi_k \eta_{t-k})(\varphi_l \eta_{t-l})\Bigg\}\left(\sum_j \varphi_j \eta_{t-i-j}\right)\Bigg] \\
&= 3\sum_j \varphi_j^3 \varphi_{j-i} + 3\sum_{k\neq j} \varphi_j^2 \varphi_k \varphi_{k-i} \\
&= 3\left(\sum_j \varphi_j^2\right)\left(\sum_k \varphi_k \varphi_{k-i}\right) \\
&= 3E(x_t^2)E(x_t x_{t-i}).
\end{aligned}$$

https://doi.org/10.1017/S0266466604201098 Published online by Cambridge University Press

Hence, Equation (2) becomes

$$-E(x_t x_{t-i}) = \alpha_0 E(x_t x_{t-i})(\omega + \lambda\mu^2 + 3\lambda E(x_t^2)),$$

which is satisfied for

$$\alpha_0 = -\frac{1}{\omega + \lambda\mu^2 + 3\lambda E(x_t^2)}. \qquad \textbf{(3)}$$

Therefore, we found a solution to (1), which is $G_0 = \alpha_0 x_t e_t$ with α_0 as in (3). By Lemma 4.3 of Hansen (1985), the greatest lower bound is given by

$$[E(G_0^2)]^{-1} = [\alpha_0^2 E(x_t^2 e_t^2)]^{-1}.$$

Note that α_0 can be rewritten as

$$\alpha_0 = -\frac{E(x_t^2)}{E(x_t^2 e_t^2)}.$$

Therefore, the GMM efficiency bound is

$$\frac{E(x_t^2 e_t^2)}{[E(x_t^2)]^2},$$

which corresponds to the asymptotic variance of the OLS estimator.

We have shown that the OLS estimator is not only as efficient as any GMM estimator that uses an arbitrary fixed number of instruments from $\{x_t, x_{t-1}, \ldots\}$ but also as efficient as any GMM estimator that uses an infinity of instruments from $\{x_t, x_{t-1}, \ldots\}$.

REFERENCES

Hansen, L. P. (1985) A method for calculating bounds on the asymptotic covariance matrices of generalized method of moments estimators. *Journal of Econometrics* 203–238.

https://doi.org/10.1017/S0266466604201098 Published online by Cambridge University Press