Econometric Theory, **15**, 1999, 257–264. Printed in the United States of America.

PROBLEMS AND SOLUTIONS

PROBLEMS

99.2.1. *An Empirical Likelihood Ratio Test for a Unit Root*, proposed by Jonathan Wright. Consider the univariate AR(1) model

$$y_t = \alpha y_{t-1} + u_t, \qquad t = 1, \ldots, T,$$

where y_0 is fixed and known and u_t is an independent and identically distributed (i.i.d.) random variable with mean zero, variance σ^2, and six finite moments. Define weights $p_1, \ldots, p_T$ to solve the maximization problem

$$\underset{p_1,\ldots,p_T}{\text{Max}} \sum_{t=1}^{T} \log(p_t)$$

subject to the restrictions

(i) $\sum_{t=1}^{T} p_t \psi_t = 0$,
(ii) $\sum_{t=1}^{T} p_t = 1$,

where $\psi_t = y_{t-1}(y_t - y_{t-1})$. Clearly $E(\psi_t) = 0$ iff $\alpha = 1$. This maximization problem is identical to that which arises in using empirical likelihood to test a moment restriction with i.i.d. data (see, e.g., Owen, 1988, 1990; Qin and Lawless, 1994) except that the weights $p_1, \ldots, p_T$ do not have the same interpretation as in the i.i.d. case. Correspondingly define the empirical likelihood ratio statistic for testing the unit root hypothesis ($H_0: \alpha = 1$) as

$$ELR = -2 \sum_{t=1}^{T} \log(Tp_t).$$

In i.i.d. applications, the empirical likelihood ratio statistic is asymptotically χ^2 (Owen, 1988, 1990). Prove that, under the null, as $T \to \infty$, $ELR \Rightarrow (2\int_0^1 W(r)^2\, dr)^{-1}(W(1)^2 - 1)^2$, the square of the simple Dickey–Fuller distribution.

REFERENCES

Owen, A. (1988) Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* 75, 237–249.

Owen, A. (1990) Empirical likelihood ratio confidence regions. *Annals of Statistics* 18, 90–120.

Qin, J., & J. Lawless (1994) Generalized estimating equations. *Annals of Statistics* 22, 300–325.

© 1999 Cambridge University Press 0266-4666/99 $9.50

https://doi.org/10.1017/S0266466699152071 Published online by Cambridge University Press

99.2.2. *Asymptotic Normality of the Nonparametric Part in Partially Linear Heteroskedastic Models*, proposed by Hua Liang, Wolfgang Härdle, and Axel Werwatz. Consider partially linear models

$$Y_i = X_i^T\beta + g(T_i) + \varepsilon_i, \qquad i = 1,\ldots,n,$$

where X_i and T_i are p-dimensional and one-dimensional random regressors, respectively, β is a vector of unknown parameters, $g(\cdot)$ is an unknown smooth function, and ε_i are independent (possibly) nonidentically distributed with mean zero conditional on X and T. Define an estimator of $g(\cdot)$ as follows:

$$g_n(t) = \sum_{i=1}^{n} \omega_{ni}(t)(Y_i - X_i^T\beta_{LS}),$$

where $\omega_{ni}(t) - \omega_{ni}(t;T_1,T_2,\ldots,T_n)$ are weight functions based on kernel function $K(\cdot)$ and bandwidth h and β_{LS} is the least squares estimator of β (see Speckman, 1988)

$$\beta_{LS} = (\tilde{\mathbf{X}}^T\tilde{\mathbf{X}})^{-1}\tilde{\mathbf{X}}^T\tilde{\mathbf{Y}},$$

where $\tilde{\mathbf{X}} = (\tilde{X}_1,\ldots,\tilde{X}_n)^T$ and $\tilde{\mathbf{Y}} = (\tilde{Y}_1,\ldots,\tilde{Y}_n)^T$ are the presmoothed design and response variables.

Denote $h_j(T_i) = E(x_{ij}|T_i)$ and $u_{ij} = x_{ij} - E(x_{ij}|T_i)$. Let us take $h_n = Cn^{-1/3}$ for some $C > 0$. Make the following assumptions.

Assumption 1. $E(u_1u_1^T)$ is a positive definite matrix, where $u_1 = (u_{11},\ldots,u_{1p})^T$.

Assumption 2. $\operatorname{supp}(K) = [-1,1]$, $\sup|K(x)| \le C < \infty$, $\int K(u)\,du = 1$, $K(u) = K(-u)$ and $\int_{|u|\ge ah_n^{-1}} K(u)\,du = O(n^{-1/3}\log n)$ for any given $a > 0$.

Assumption 3. $n^{-2/3}\sum_{i=1}^{n}\omega_{ni}^2(t)E\varepsilon_i^2 = \sigma_0^2 + o(1)$ for some $\sigma_0^2 > 0$.

Then, prove the following theorem.

THEOREM. *Assume that $\varepsilon_1,\ldots,\varepsilon_n$ are independent random variables with $E\varepsilon_i = 0$ and $\inf_i E\varepsilon_i^2(=\sigma_i^2) > c_\sigma > 0$ for some c_σ and that there exists a function $G(u)$ satisfying $\int_0^\infty uG(u)\,du < \infty$ such that*

$$P(|\varepsilon_i| > u) \le G(u), \quad \text{for } i = 1,\ldots,n \quad \text{and} \quad \text{large enough } u. \tag{1}$$

Then

$$\frac{g_n(t) - Eg_n(t)}{\sqrt{\operatorname{Var}\{g_n(t)\}}} \xrightarrow{\mathcal{L}} N(0,1) \quad \text{as} \quad n \to \infty.$$

REFERENCE

Speckman, P. (1988) Kernel smoothing in partial linear models. *Journal of the Royal Statistical Society, Series B* 50, 413–436.

https://doi.org/10.1017/S0266466699152071 Published online by Cambridge University Press

99.2.3. *Asymptotic Efficiency of OLS Estimator in Models with Deterministic Regressors*, proposed by Jinyoung Hahn. Consider a linear regression model

$$y_i = x_i'\beta + \varepsilon_i, \qquad i = 1, \ldots, n,$$

where x_i is deterministic and $\varepsilon_i \sim N(0, \sigma^2)$ is independent and identically distributed (i.i.d.). Suppose that

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} x_i x_i' = \Upsilon, \qquad \lim_{n\to\infty} \frac{\max_{1\le i\le n} x_i' x_i}{n} = 0,$$

where Υ is nonsingular. Show that, for any regular estimator b of β, the limit distribution of $\sqrt{n}(b - \beta)$ can be expressed as the distribution of a sum of $N(0, \sigma^2\Upsilon^{-1})$ and some other random variable independent of the normal random variable. Conclude that OLS is asymptotically efficient. (Observe that y's are non-i.i.d. Therefore, maximum likelihood estimation, which is asymptotically efficient for i.i.d. sample, cannot justify the asymptotic efficiency.)

99.2.4. *Prediction in the Spatially Autocorrelated Error Component Model*, proposed by Badi H. Baltagi and Dong Li. Testing for spatial autocorrelation is important in regional science and urban and transportation economics (see Anselin, 1988). This problem considers the panel regression model

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

with error components that are spatially autocorrelated (see Anselin, 1988, p. 152). The disturbance vector for time t is given by

$$\varepsilon_t = \mu + \phi_t \tag{2}$$

where $\varepsilon_t = (\varepsilon_{1t}, \ldots, \varepsilon_{Nt})'$, $\mu = (\mu_1, \ldots, \mu_N)'$ denote the random region effects that are i.i.d.$(0, \sigma_\mu^2)$, and $\phi_t = (\phi_{1t}, \ldots, \phi_{Nt})'$ are the remainder disturbances that are independent of μ. The ϕ_t's follow the spatial error dependence model

$$\phi_t = \lambda W \phi_t + \nu_t \tag{3}$$

where W is the matrix of known spatial weights and λ is the spatial autoregressive coefficient; $\nu_t = (\nu_{1t}, \ldots, \nu_{Nt})'$ is i.i.d.$(0, \sigma_\nu^2)$ and is independent of ϕ_t and μ. Derive the best linear unbiased predictor of $y_{i,T+S}$ for the ith region at period $T + S$.

REFERENCE

Anselin, L. (1988) *Spatial Econometrics: Methods and Models.* Dordrecht: Kluwer.

https://doi.org/10.1017/S0266466699152071 Published online by Cambridge University Press

SOLUTIONS

98.2.1. *Consistency of OLS in the Presence of Lagged Dependent Variable and Serially Correlated Errors*—Solution, proposed by Jeffrey M. Wooldridge.

(i) $E(u_t) = E(y_t) - \beta E(y_{t-1}) = 0$. Further, $E(y_{t-1}u_t) = \sigma_1 - (\sigma_1/\sigma_y^2)\sigma_y^2 = 0$, where $\sigma_y^2 = \text{Var}(y_t)$ and $\sigma_1 = \text{Cov}(y_{t-1}, y_t)$.

(ii) This follows from part (i) by basic algebra and the law of large numbers:

$$p\lim \hat{\beta} = \beta + p\lim\left(n^{-1}\sum_{t=1}^{n} y_{t-1}^2\right)^{-1}\left(n^{-1}\sum_{t=1}^{n} y_{t-1}u_t\right)$$
$$= \beta + (1/\sigma_y^2)E(y_{t-1}u_t) = \beta + (1/\sigma_y^2)\cdot 0 = \beta.$$

(iii) In general, $E(u_t u_{t-1}) \neq 0$. In fact, we can write

$$E(u_t u_{t-1}) = E[u_t(y_{t-1} - \beta y_{t-2})] = -\beta E(y_{t-2}u_t) = -\beta(\sigma_2 - \beta\sigma_1),$$

where $\sigma_j \equiv E(y_t y_{t-j})$, $j \geq 1$. We have assumed nothing about σ_2. If $\beta \neq 0$ and $\sigma_2 \neq \beta\sigma_1$, then u_t and u_{t-1} are correlated. (Incidentally, if $\{y_t\}$ follows an AR(1) process then $\sigma_2 = \beta\sigma_1$, and so $E(u_t u_{t-1}) = 0$.)

(iv) Parts (ii) and (iii) show that the statement given is generally false. The resolution is that the statement is correct in a very specific case. If we write

$$y_t = \beta y_{t-1} + u_t$$

and then *assume* that $\{u_t\}$ is a stable AR(1) process—with, say, parameter ρ—then u_t and y_{t-1} are necessarily correlated. This means that OLS of y_t on y_{t-1} does not consistently estimate β—but it will consistently estimate the linear projection of y_t on y_{t-1}, as part (ii) shows. In some cases we might be interested in $E(y_t|y_{t-1})$, or especially $E(y_t|\mathbf{z}_t, y_{t-1})$, where $\mathbf{z}_t$ is a vector of explanatory variables of interest. Then, we do not care if the errors are serially correlated, except that our inference would have to be adjusted.

It is useful to show directly that, if $E(y_{t-1}u_t) = 0$, then $\{u_t\}$ cannot follow an AR(1) process. If $\{u_t\}$ is an AR(1) process with parameter ρ, $|\rho| < 1$, then the autocovariances must satisfy

$$E(u_t u_{t-j}) = \rho E(u_t u_{t-j+1}), \quad \text{all } j \geq 1. \tag{1}$$

We can arrange (1) for finite j by choosing the autocovariances of $\{y_t\}$ appropriately. But (1) cannot hold for all j if $\beta \neq 0$. To see this, note that, for $j \geq 3$, (1) requires

$$\alpha_j - \beta\alpha_{j+1} = \rho(\alpha_{j-1} - \beta\alpha_j), \tag{2}$$

where $\alpha_j \equiv E(y_{t-j}u_t)$. Because the process is asymptotically uncorrelated, we must be able to choose ρ so that $\alpha_j \to 0$ as $j \to \infty$. We can easily see that this is impossible by rewriting (2):

$$\alpha_{j+1} = (1/\beta + \rho)\alpha_j - (\rho/\beta)\alpha_{j-1}.$$

https://doi.org/10.1017/S0266466699152071 Published online by Cambridge University Press

This second-order difference equation is stable if and only if

$$1 - (1/\beta + \rho) + \rho/\beta > 0,$$
$$1 + (1/\beta + \rho) + \rho/\beta > 0,$$
$$1 - \rho/\beta > 0.$$

The first two conditions can be written as

$$(1 - 1/\beta)(1 - \rho) > 0, \tag{3}$$
$$(1 + 1/\beta)(1 + \rho) > 0. \tag{4}$$

If $|\rho| < 1$ then $1 - \rho > 0$ and $1 + \rho > 0$. But with $|\beta| < 1$, either $1 - 1/\beta < 0$ (if $\beta > 0$) or $1 + 1/\beta < 0$ (if $\beta < 0$). Therefore, (3) and (4) cannot both be true with $|\beta| < 1$ and $|\rho| < 1$. This completes the proof.

98.2.2. *Upper Bounds for Eigenvalues of Nonnegative Definite Matrices*—Solutions.[1] The following solutions have been proposed independently by George P.H. Styan and Hans Joachim Werner, and by Egon Zakrajšek. These solutions are based on different types of interesting arguments.

1.—Solution, proposed by George P.H. Styan and Hans Joachim Werner.

Let $H = A \odot B$ denote the Hadamard (Schur or elementwise) product of the $T \times T$ nonnegative definite matrices A and B. From the "Augmented Schur Theorem" on p. 142 of Im (1997) we have that

$$\lambda_i(H) \le \alpha_i(A)\lambda_1(B) \qquad (i = 1,2,\ldots,T). \tag{2}$$

Putting $B = ee'$ in (2), where e denotes the $T \times 1$ column vector of ones, yields $\lambda_i(A) \le T\alpha_i(A)$ at once, because now $H = A \odot ee' = A$ and the largest eigenvalue of $B = ee'$ is T (all other eigenvalues are 0).

2.—Solution, proposed by Egon Zakrajšek.

According to Courant–Fischer minimax theorem (Demmel, 1997, Theorem 5.2) we have

$$\lambda_i(A) = \min_{\dim(S)=T-i+1} \max_{0\neq s\in S} \frac{s'As}{s's},$$

so we have

$$\lambda_i(A) \le \max_{0\neq s\in S} \frac{s'As}{s's},$$

where S is some particular $T - i + 1$ dimensional subspace, say the subspace spanned by the last $T - i + 1$ columns of the identity matrix. If we partition matrix A as

https://doi.org/10.1017/S0266466699152071 Published online by Cambridge University Press

$$A = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix},$$

we see that $\lambda_i(A) \leq \lambda_{\max}(A_{22})$, where submatrix A_{22} is $(T - i + 1) \times (T - i + 1)$. In our case the matrix A is nonnegative definite; so is the matrix A_{22}, which means $\lambda_i(A_{22}) \geq 0$, and as no eigenvalue can exceed the sum of the eigenvalues, which equals the trace of matrix A_{22}, we have

$$\lambda_i(A) \leq \sum_{k=i}^{T} \alpha_k(A).$$

We can always permute the rows and columns of the matrix without changing its eigenvalues, so we can assume that matrix A has the diagonal elements already sorted in descending order and we get

$$\lambda_i(A) \leq \sum_{k=i}^{T} \alpha_k(A) \leq \sum_{k=i}^{T} \alpha_i(A) = (T - i + 1)\alpha_i(A).$$

As this is a somewhat sharper bound the proposition is proved.

NOTE

1. An excellent solution has been proposed independently by H. Neudecker and M. van de Velden and by Kyung-Taik Han and Eric I. Im (the posers of the problem).

REFERENCES

Demmel J.W. (1997) *Applied Numerical Linear Algebra.* Philadelphia: SIAM.

Im, Eric Iksoon (1997) Narrower eigenbounds for Hadamard products. *Linear Algebra and Its Applications* 264, 141–144.

98.2.3. *Partial Observability in a Bivariate Logit Model*—Solution, proposed by S.K. Sapra. The log-likelihood for model P is

$$L(\beta_1, \beta_2) = \sum_{i=1}^{n} w_i \ln F_i + (1 - w_i) \ln (1 - F_i), \tag{1}$$

where $F_i = F(\beta_1' x_i, \beta_2' x_i) = 1/(1 + e^{-\beta_1' x_i} + e^{-\beta_2' x_i})$.

The information matrix based on n observations is

$$I_{2k \times 2k} = \begin{pmatrix} I_{11} & I_{12} \\ I_{21} & I_{22} \end{pmatrix}, \tag{2}$$

https://doi.org/10.1017/S0266466699152071 Published online by Cambridge University Press

where $I_{11} = \sum_{i=1}^{n}[\{(e^{-2\beta_1' x_i})F_i^2\}x_i x_i']/C_i$, $I_{12} = \sum_{i=1}^{n}[\{(e^{-(\beta_1'+\beta_2')x_i}F_i^2\}x_i x_i']/C_i$, $I_{21} = I_{12}$, and $I_{22} = \sum_{i=1}^{n}[\{(e^{-2\beta_2' x_i})F_i^2\}x_i x_i']/C_i$, where $C_i = e^{-\beta_1' x_i} + e^{-\beta_2' x_i}$.

To show that model P is not locally identified under each of the configurations in the problem, we show that the information matrix I is singular in each case (Rothenberg, 1971).

Under configuration (i), $\beta_1 = \beta_2$, and therefore,

$$I_{11} = I_{12} = I_{21} = I_{22} = J, \text{ say.} \tag{3}$$

Then $I = J \otimes E$, where

$$E = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix}. \tag{4}$$

Therefore, $\det(I) = (\det(J))^k(\det(E))^2 = 0$ because $\det(E) = 0$. Hence, I is a singular matrix.

Under configuration (ii), $\beta_1 = (\beta_{11},0,0,\ldots,0)'$, $\beta_2 = (\beta_{21},0,0,\ldots,0)'$, and $\beta_1' x_i = \beta_{11}$, $\beta_2' x_i = \beta_{21}$. Substituting these expressions into (2) yields

$$I = A \otimes \sum_{i=1}^{n} x_i x_i', \tag{5}$$

where

$$A = 1/D_1 \begin{pmatrix} e^{-2\beta_{11}} & e^{-(\beta_{11}+\beta_{21})} \\ e^{-(\beta_{11}+\beta_{21})} & e^{-2\beta_{21}} \end{pmatrix},$$

$$D_1 = (e^{-\beta_{11}+e^{-\beta_{21}}})(1 + e^{-\beta_{11}+e^{-\beta_{21}}})^2. \tag{6}$$

Therefore, from (5) and (6),

$$\det(I) = (\det(A))^2 \left(\det\left(\sum_{i=1}^{n} x_i x_i'\right)\right)^k = 0 \tag{7}$$

because $\det(A) = 0$. Hence I is a singular matrix.

Under configuration (iii), $x_{1i} = 1$ and x_{2i} is a dummy variable taking the values 1 or 0. Following Poirier (1980, p. 213), order the observations so that $x_{2i} = 0$ $(i = 1,2,\ldots,r)$ and $x_{2i} = 1$ $(i = r + 1, r + 2,\ldots,n)$. Then the information matrix I reduces to

$$I = A_1 \otimes B_1 + A_2 \otimes B_2, \tag{8}$$

https://doi.org/10.1017/S0266466699152071 Published online by Cambridge University Press

where

$$A_1 = rA, \qquad B_1 = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix},$$

$$A_2 = (n-r)/D_2 \begin{pmatrix} e^{-2(\beta_{11}+\beta_{12})} & e^{-(\beta_{11}+\beta_{12}+\beta_{21}+\beta_{22})} \\ e^{-(\beta_{11}+\beta_{12}+\beta_{21}+\beta_{22})} & e^{-2(\beta_{21}+\beta_{22})} \end{pmatrix},$$

$$B_2 = \begin{pmatrix} 1 & 1 \\ 1 & 1 \end{pmatrix} = E,$$

$$D_2 = (e^{-(\beta_{11}+\beta_{12})} + e^{-(\beta_{21}+\beta_{22})})(1 + e^{-(\beta_{11}+\beta_{12})} + e^{-(\beta_{21}+\beta_{22})}).^2$$

(The term A is defined in equation (6), and E is defined in equation (4).)

It is easily verified that $\det(I) = 0$. In fact, $\text{rank}(I) \leq 2$ because $\text{rank}(I) \leq \text{rank}(A_1 \otimes B_1) + \text{rank}(A_2 \otimes B_2)$ and each rank on the right-hand side of the inequality is 1. Hence I is a singular matrix.

REFERENCES

Poirier, D. (1980) Partial observability in bivariate probit model. *Journal of Econometrics* 12, 209–217.

Rothenberg, T. (1971) Identification in parametric models. *Econometrica* 39, 577–591.

https://doi.org/10.1017/S0266466699152071 Published online by Cambridge University Press