Econometric Theory, **17**, 2001, 853–858. Printed in the United States of America.

PROBLEMS AND SOLUTIONS

PROBLEMS

01.4.1. *An Alternative GLS-like Transformation in Regression Models with AR(1) Errors*, proposed by Oliver Linton and D. Kristensen. A common way of teaching GLS estimation is by making a transformation of the regression model to one that obeys the classical assumptions of i.i.d. error terms so that the OLS estimator in this transformed regression is known to be efficient. In practice, one has to use an estimated transformation, and the argument is that the resulting estimator is the same to the required order as the infeasible estimator and is thus efficient. The following example shows that not all transformations have this property. This transformation is itself of interest in nonparametric estimation, where it turns out to have advantageous properties (see Linton, Mammen, Lin, and Carroll, 2000).

Suppose that

$$Y_t = \beta X_t + u_t, \tag{1}$$

where

$$u_t = \rho u_{t-1} + \varepsilon_t \tag{2}$$

for $|\rho| < 1$. Here $\varepsilon = (\varepsilon_t)_t$ is a sequence of i.i.d. errors with mean zero and variance σ_ε^2, whereas $X = (X_t)_t$ is assumed to be a mean zero covariance stationary mixing sequence of random variables with variance σ_X^2 and one-step correlation $\rho_X = E(X_t X_{t-1})/\sigma_X^2$ with $|\rho_X| < 1$. We furthermore assume that

$$E\left[\begin{pmatrix} u_t \\ u_{t-1} \end{pmatrix} X_t\right] = 0. \tag{3}$$

Consider the following estimator:

$$\tilde{\beta} = \frac{\sum_{t=2}^{T} X_t (Y_t - \hat{\rho}(Y_{t-1} - \hat{\beta} X_{t-1}))}{\sum_{t=2}^{T} X_t^2}, \tag{4}$$

where

$$\hat{\beta} = \frac{\sum_{t=1}^{T} X_t Y_t}{\sum_{t=1}^{T} X_t^2}; \qquad \hat{\rho} = \frac{\sum_{t=2}^{T} \hat{u}_t \hat{u}_{t-1}}{\sum_{t=2}^{T} \hat{u}_{t-1}^2} \tag{5}$$

© 2001 Cambridge University Press 0266-4666/01 $9.50

https://doi.org/10.1017/S0266466601174086 Published online by Cambridge University Press

and where $\hat{u}_t = Y_t - \hat{\beta}X_t$ is the residual.

Derive the asymptotic properties of $\tilde{\beta}$ and show that it can improve on OLS but is not as good as GLS.

REFERENCE

Linton, O., E. Mammen, X. Lin, & R. Carroll (2000) Accounting for Correlation in Marginal Longitudinal Nonparametric Regression. Available at http://econ.lse.ac.uk/staff/olinton/.

01.4.2. *Nonorthogonal Hilbert Projections in Trend Regression*, proposed by P.C.B. Phillips and Y. Sun. An observed continuous time process $X(t)$ is generated by the linear system

$$X(t) = \beta' Z(t) + W(t), \qquad t \in [0,1], \tag{1}$$

where $W(t)$ is an unobservable standard Brownian motion, $Z(t) = (t, \ldots, t^p)'$ is a time polynomial vector, and β is a parameter vector to be estimated.

The following two estimators of β are proposed:

$$\hat{\beta} = \left(\int_0^1 Z(t)Z(t)'\, dt \right)^{-1} \left(\int_0^1 Z(t)X(t)\, dt \right)$$

and

$$\tilde{\beta} = \left(\int_0^1 Z^{(1)}(t)Z^{(1)}(t)'\, dt \right)^{-1} \left(\int_0^1 Z^{(1)}(t)\, dX(t) \right),$$

where $Z^{(1)}$ is the vector of the first derivatives of Z.

Part A

1. Show that both $\hat{\beta}'Z(t)$ and $\tilde{\beta}'Z(t)$ are Hilbert projections in $L_2[0,1]$. How do these projections differ?
2. Find the distributions of $\hat{\beta}$ and $\tilde{\beta}$ and compare them in the case where $p = 1$. What do you conclude?

Part B

Suppose the system generating $X(t)$ is

$$X(t) = \beta' Z(t) + J_c(t), \qquad t \in [0,1], \tag{2}$$

where $J_c(t) = \int_0^t e^{(t-s)c} dW(s)$ for some known constant $c < 0$ is a linear diffusion or Orstein–Uhlenbeck process.

1. Are $\hat{\beta}'Z(t)$ and $\tilde{\beta}'Z(t)$ still Hilbert projections?
2. Calculate the distributions of $\hat{\beta}$ and $\tilde{\beta}$ and compare them in the case where $p = 1$. What do you conclude?
3. Taking c to be known, can you suggest an unbiased linear estimator of β that has smaller variance than $\hat{\beta}$ and $\tilde{\beta}$? Does it correspond to another Hilbert projection?

https://doi.org/10.1017/S0266466601174086 Published online by Cambridge University Press

01.4.3. *The Limit Distribution of Cointegration Rank Tests of "Wald" Type*, proposed by P. Paruolo. Consider an n-dimensional VAR(k) process X_t in error-correction form

$$Z_{0t} = \alpha\beta' Z_{1t} + \mathrm{Y}Z_{2t} + \epsilon_t \qquad t = 1,\ldots,T, \tag{1}$$

where $Z_{0t} := \Delta X_t$, $Z_{1t} := X_{t-1}$, $Z_{2t} := (\Delta X'_{t-1} : \ldots : \Delta X'_{t-k+1})$, ϵ_t are i.i.d. with mean 0 and covariance matrix Ω. Here X_t, ϵ_t are $n \times 1$ vectors and A_i, Ω are $n \times n$ matrices; α and β are full rank $n \times r$ matrices, $r < n$. Assume that the system in (1) is $I(1)$, i.e., that it satisfies the assumptions of Theorem 4.2 in Johansen (1996).

Dhrymes (1998, Ch. 6.2.2) has proposed the following "Wald" test statistics for cointegration rank: $D_1(j) := T\sum_{i=j+1}^{n} \hat{\rho}_i$ or $D_2(j) := T\hat{\rho}_{j+1}$ where $\hat{\rho}_1 > \cdots > \hat{\rho}_n$ are the solutions of

$$|\rho S_{00.1} - S_{01}S_{11}^{-1}M_{11}S_{11}^{-1}S_{10}| = 0 \tag{2}$$

and where $M_{ij} := T^{-1}\sum_{t=1}^{T} Z_{it}Z'_{jt}$, $S_{ij} := M_{ij.2} := M_{ij} - M_{i2}M_{22}^{-1}M_{2j}$, $S_{ij.h} := S_{ij} - S_{ih}S_{hh}^{-1}S_{hj}$; this notation parallels the one in Johansen (1996).

(a) For the special case $k = 1$, show that $\rho = \lambda/(1-\lambda)$, where λ solves the canonical correlation problem $|\lambda S_{00} - S_{01}S_{11}^{-1}S_{10}| = 0$.
(b) For the case $k = 1$ the tests $D_1(r)$ and $D_2(r)$ have the same limit distribution of the LR trace and lambda max tests described in Johansen (1988).
(c) For the case $k > 1$, show that the results in part (b) continue to hold.

Observe that these distributions are different from the ones stated in Dhrymes (1998, Proposition 5, Ch. 6).

REFERENCES

Dhrymes, P. (1998) *Time Series, Unit Roots, and Cointegration.* London: Academic Press.
Johansen, S. (1988) Statistical analysis of cointegration vectors. *Journal of Economics Dynamics and Control* 12, 231–254.
Johansen, S. (1996) *Likelihood-Based Inference in Cointegrated Vector Auto-Regressive Models.* Oxford: Oxford University Press.

SOLUTIONS

00.4.1 *Simple Applications of the Cox–Tsiatis Result on Unidentifiability of Dependent Competing Risks Models with Regressors—Solution*, proposed by K. Sapra.

(i) For $T = \min(T_1, T_2)$, the overall survivor function for the dependent risks model (1) is

$$S_T(t) = P(T_1 > t, T_2 > t) = S(t,t) = \exp(-(\lambda_1 + \lambda_2)^{\alpha} t^{\alpha}). \tag{3}$$

https://doi.org/10.1017/S0266466601174086 Published online by Cambridge University Press

The crude hazard rate (in the presence of both risks) corresponding to the joint survivor function in eq. (1) is

$$h_i(t) = -\partial \ln S(t_1, t_2)/\partial t_i|_{t_1=t_2=t} = \lambda_i \alpha(\lambda_1 + \lambda_2)^{\alpha-1} t^{\alpha-1}, \qquad i = 1,2. \tag{4}$$

The corresponding survival functions are

$$S_i(t) = \exp\left(-\int_0^t h_i(u)\, du\right) = \exp(-\lambda_i(\lambda_1 + \lambda_2)^{\alpha-1} t^{\alpha}), \qquad i = 1,2. \tag{5}$$

If T_1^* and T_2^* are independent random variables with survivor functions

$$S_i^*(t_i) = \exp(-\lambda_i(\lambda_1 + \lambda_2)^{\alpha-1} t_i^{\alpha}), \qquad i = 1,2, \tag{6}$$

then the overall survivor function for the independent risks model is

$$\begin{aligned} S_{T^*}(t) &= P(T_1^* > t, T_2^* > t) \\ &= P(T_1^* > t)P(T_2^* > t) \\ &= S_1^*(t)S_2^*(t) = \exp(-(\lambda_1 + \lambda_2)^{\alpha} t^{\alpha}) = S_T(t), \end{aligned} \tag{7}$$

the overall survivor function for the dependent risks model in eq. (3). Furthermore,

$$\begin{aligned} Q_i(t) &= P(T > t \cap \delta = i) = \int_t^{\infty} h_i(u) S_T(u)\, du = Q_i^*(t) \\ &= \int_t^{\infty} h_i^*(u) S_{T^*}(u)\, du, \qquad i = 1,2 \end{aligned}$$

because $S_T(u) = S_{T^*}(u)$ from eq. (7) and $h_i(u) = h_i^*(u)$, where from eq. (6),

$$h_i^*(u) = -d \ln S_i^*(u)/du = \lambda_i \alpha(\lambda_1 + \lambda_2)^{\alpha-1} u^{\alpha}, \qquad i = 1,2. \tag{8}$$

Hence the two models are equivalent.

(ii) For $T = \min(T_1, T_2)$, the overall survivor function for the dependent risks model (2) is

$$S_T(t) = P(T_1 > t, T_2 > t) = [\beta/((\lambda_1 + \lambda_2)t + \beta)]^{\alpha+1}. \tag{9}$$

Following the steps in part (i), an equivalent independent competing risks model is constructed using crude hazard rates (in the presence of both risks)

$$h_i(t) = (\alpha + 1)\lambda_i/((\lambda_1 + \lambda_2)t + \beta), \qquad i = 1,2, \tag{10}$$

which leads to the survival functions

$$S_i^*(t_i) = [\beta/((\lambda_1 + \lambda_2)t_i + \beta)]^{(\alpha+1)\lambda_i/(\lambda_1+\lambda_2)}, \qquad i = 1,2 \tag{11}$$

for T_1^* and T_2^*.

https://doi.org/10.1017/S0266466601174086 Published online by Cambridge University Press

00.4.2. *A Necessary and Sufficient Condition for the Convergence of a Sum of Weighting Random Variables to a Normal Distribution*—Solution,[1] proposed by Norbert Christopeit and Hans Joachim Werner.

Proof of sufficiency. Without loss of generality, assume $\sigma^2 = 1$. Introduce $\xi_{ni} := a_{ni}X_i$, $\alpha_n := \max_{1\le i\le n} |a_{ni}|$. Then, for each $\epsilon > 0$,

$$\begin{aligned}\sum_{i-1}^{n} E[\xi_{ni}^2 \mathbf{1}_{\{|\xi_{ni}|\ge\epsilon\}}] &= \sum_{i=1;a_{ni}\neq 0}^{n} a_{ni}^2 E[X_i^2 \mathbf{1}_{\{|X_i|\ge\epsilon/|a_{ni}|\}}] \\ &\le E[X_i^2 \mathbf{1}_{\{|X_i|\ge\epsilon/\alpha_n\}}] \\ &\to 0\end{aligned}$$

if $\alpha_n \to 0$. Hence the Lindeberg condition is satisfied and $S_n := \sum_{i=1}^{n} a_{ni}X_i \xrightarrow{\mathcal{D}} N(0,1)$ by the classic central limit theorem (CLT) for triangular arrays.

Proof of necessity. If the condition $\alpha_n \to 0$ is violated, so is the classic condition of *negligibility in the limit*. Hence, for the necessary part, we must resort to *nonclassic CLTs*. Theorem 1 is from Rotar (1975).

THEOREM 1. *For each $n \ge 1$, let $\xi_{n1},\dots,\xi_{nk_n}$ be nondegenerate independent zero mean random variables with variances $\sigma_{ni} > 0$ satisfying $\sum_{i=1}^{k_n}\sigma_{ni}^2 = 1$. Let $S_n := \xi_{n1} + \cdots + \xi_{nk_n}$, F_{ni} be the distribution function of ξ_{ni} and $\Phi_{ni}(x) = \Phi(x/\sigma_{ni})$, where $\Phi(.)$ indicates the distribution function of a standardized normal random variable. Then a necessary and sufficient condition for $S_n \xrightarrow{\mathcal{D}} N(0,1)$ to hold is that*

$$\sum_{i=1}^{k_n}\int_{|x|>\epsilon} |x||F_{ni}(x) - \Phi_{ni}(x)|dx \xrightarrow{n\to\infty} 0 \qquad \textbf{(1)}$$

is satisfied for every $\epsilon > 0$.

Back to our situation, note first that by renumeration we can assume that $a_{ni} \neq 0$ for $i = 1,\dots,k_n$ where $k_n \le n$ is the number of all nonzero coefficients a_{ni}. Then $F_{ni}(x) = F(x/|a_{ni}|)$, where F is the common distribution function of X_i. Therefore, (1) becomes

$$R_n := \sum_{i=1}^{k_n} a_{ni}^2 \int_{|z|>\epsilon/|a_{ni}|} |z||F(z) - \Phi(z)|dz \xrightarrow{n\to\infty} 0. \qquad \textbf{(2)}$$

Supposed now that $\alpha_n \to 0$ does not hold. Then there is a $\delta > 0$ and a subsequence (n') together with numbers $1 \le i_{n'} \le k_{n'}$ such that

$$\alpha_{n'} = |a_{n'i_{n'}}| \ge \delta > 0$$

https://doi.org/10.1017/S0266466601174086 Published online by Cambridge University Press

for all n'. But then

$$R_{n'} \geq \delta^2 \int_{|z|>\epsilon/\delta} |z||F(z) - \Phi(z)|dz. \tag{3}$$

Unless $F = \Phi$, the right-hand side of (3) is > 0 for ϵ small enough, so that (2) cannot hold. This completes the proof. ■

Remarks. If $S_n \xrightarrow{\mathcal{D}} N(0,1)$ does not hold, then it must be true that

$$k_n \to \infty, \quad \text{as} \quad n \to \infty. \tag{4}$$

Otherwise, there would be a subsequence $(n') \subset (n)$ and a natural number N such that $k_{n'} \leq N < \infty$ for all n'. Then $S_n \xrightarrow{\mathcal{D}} N(0,1)$ would imply that the finite sums $S_{n'} = a_{n'1}X_1 + \cdots + a_{n'k_{n'}}X_{k_{n'}} \xrightarrow{\mathcal{D}} N(0,1)$. Write

$$S_{n'} = c_{n'1}X_1 + \cdots + c_{n'N}X_N,$$

with $c_{n'i} = a_{n'i}$ for $1 \leq i \leq k_{n'}$ and $c_{n'i} = 0$ for $k_{n'} + 1 \leq i \leq N$. Because the vectors $c_{n'} = (c_{n'1} \cdots c_{n'N})$ take values in the N-dimensional unit cube, we may extract a further subsequence $(n'') \subset (n')$ such that $c_{n''} \to c := (c_1 \cdots c_N)$; hence, because $S_{n''} \xrightarrow{\mathcal{D}} N(0,1)$,

$$S := c_1X_1 + \cdots + c_NX_N = \lim_{n''\to\infty} S_{n''} \sim N(0,1).$$

By the well-known characterization of the normal distribution by Lévy and Cramér, this would imply that the X_i, $1 \leq i \leq N$, for which c_i does not vanish are themselves normally distributed, which was excluded by assumption. Therefore, (4) must be true.

NOTE

1. Alternative solutions have been proposed independently by P.R. Hansen, C. Kleiber, J. Hill, X. Ming, and H. Liang, the poser of the problem.

REFERENCE

Rotar, V.I. (1975) An extension of the Lindeberg–Feller theorem. *Mathematics Notes* 18, 660–663.

https://doi.org/10.1017/S0266466601174086 Published online by Cambridge University Press