Econometric Theory, **16**, 2000, 1043–1050. Printed in the United States of America.

PROBLEMS AND SOLUTIONS

PROBLEMS

00.6.1. *First and Second Moments of Newey and West's HAC Covariance Matrix Estimator under Normality*, proposed by José Matos Passos. Consider the simple case of a linear regression model $y = X\beta + \varepsilon$, where ε can be autocorrelated and heteroskedastic. Let $\varepsilon \sim N(0, \Omega)$, Ω positive definite, and let n be the number of observations. In this case, the variance matrix of the least squares estimator of $\beta, \hat{\beta}$, conditional on X, is given by $(X'X)^{-1}X'\Omega X(X'X)^{-1}$. Newey and West (1987) suggested estimating Ω by $\hat{\Omega} = [\hat{\omega}_{ij}]_{i,j=1,\ldots,n}$, $\hat{\omega}_{ij} = \kappa_{i-j}\hat{\varepsilon}_i\hat{\varepsilon}_{i-j}$, where $\hat{\varepsilon}_i$, $i = 1,2,\ldots,n$, are the OLS residuals, and

$$\kappa_h = \left(1 - \frac{|h|}{m+1}\right)\mathbf{1}(|h| \leq m) \tag{1}$$

are the weights associated with the autocovariances. $\mathbf{1}(\cdot)$ is the indicator function, and m is the maximum lag, with $n > m$. Derive the first and second moments of Newey and West's HAC covariance matrix estimator $\hat{\Omega}$.

REFERENCE

Newey, W. & K. West (1987) A simple, positive semi-definite, heteroskedasticity and autocorrelation consistent covariance matrix. *Econometrica* 55, 703–708.

00.6.2. *A Distributional Equality*, proposed by Paolo Paruolo. Consider the random walk $y_t = y_{t-1} + \epsilon_t$, $y_0 = 0$, ϵ_t i.i.d. $N(0,1)$, $t = 1,\ldots,T$, and the statistics $a_T := T^{-1}\sum_{t=1}^{T}(y_t - \bar{y})^2$ and $b_T := T^{-1}\sum_{t=1}^{T}(y_t - T^{-1}ty_T)^2$. Show that $\mathcal{L}(a_T) = \mathcal{L}(b_T)$, and that this implies the known result

$$\mathcal{L}\left(\int_0^1 (B_u - \bar{B})^2 du\right) = \mathcal{L}\left(\int_0^1 (B_u - uB_1)^2 du\right), \tag{1}$$

where $\mathcal{L}(\cdot)$ indicates the distribution of the argument, B_u is a standard Brownian motion, and $\bar{B} := \int_0^1 B_u du$. Equality (1) is documented, e.g., in Tanaka (1996), equations (1.18) and (1.26).

REFERENCE

Tanaka, K. (1996) *Time Series Analysis—Nonstationary and Noninvertible Distribution Theory.* New York: Wiley.

© 2000 Cambridge University Press 0266-4666/00 $9.50

https://doi.org/10.1017/S0266466600166095 Published online by Cambridge University Press

SOLUTIONS

99.6.1. *Central Limit Theorem for Squared MA Infinity Processes*—Solution,[1] proposed by Christian Kleiber. Under the stated assumptions, the limit of appropriately scaled partial sums of $\{y_t^2\}$ may or may not be Gaussian; hence, $\{y_t^2\}$ does not, in general, follow the central limit theorem (CLT). Non-Gaussian limits arise in the strongly dependent case where the autocorrelations of $\{y_t\}$ are not square-summable.

Consider, for example, fractionally integrated white noise

$$y_t = (1 - B)^{-d}\varepsilon_t, \tag{1}$$

where B is the backshift operator, $\{\varepsilon_t\}$ is a Gaussian (for simplicity) white-noise sequence, and $0 < d < \frac{1}{2}$. It is well known that $\{y_t\}$ is a stationary but strongly dependent Gaussian process. (Functional) Limit theorems for $\{f(y_t)\}$, f some sufficiently regular function, have been derived by Taqqu (1975). In particular, the transformation $f(y) = y^2$ has Hermite rank two, and, for $\frac{1}{4} < d < \frac{1}{2}$, the partial sum process of $\{y_t^2\}$ converges weakly to the Rosenblatt process, with non-Gaussian marginals (Taqqu, 1975, Theorem 6.1). This can be generalized to non-Gaussian inputs $\{\varepsilon_t\}$ using the concept of Appell rank (Avram and Taqqu, 1987); Giraitis and Taqqu (1999) provide a survey of such limit theorems.

NOTE

1. The statement of the problem is correct under the stronger assumption of absolute summability of $\{\alpha_j\}$ (see e.g., Anderson, 1971, Theorem 8.4.2; Brockwell and Davies, 1991, Theorem 7.3.2). Phillips and Solo (1992, Theorem 3.8) showed that the CLT applies under the alternative assumption $\sum_j j\alpha_j^2 < \infty$. An excellent solution implicitly assuming this alternative assumption was proposed by Antonis Demos, the poser of the problem.

REFERENCES

Anderson, T.W. (1971) *The Statistical Analysis of Time Series.* New York: John Wiley & Sons.

Avram, F. & M.S. Taqqu (1987) Noncentral limit theorems and Appell polynomials. *Annals of Probability* 15, 767–775.

Brockwell, P.J. & R.A. Davies (1991) *Time Series: Theory and Methods*, 2nd ed. New York: Springer-Verlag.

Giraitis, L. & M.S. Taqqu (1999) Convergence of normalized quadratic forms. *Journal of Statistical Planning and Inference* 80, 15–35.

Phillips, P.C.B. & V. Solo (1992) Asymptotics for linear processes. *Annals of Statistics* 20, 971–1001.

Taqqu, M.S. (1975) Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete* 31, 287–302.

99.6.2. *Common Principle for the Decomposition of the Total Variability in Principal Component Analysis and in Linear Regression*—Solution, proposed by Jean-Daniel Rolle. Consider $z = (x', y)' \in \mathbb{R}^{p+1}$, with $x \in \mathbb{R}^p$. We denote I_p

https://doi.org/10.1017/S0266466600166095 Published online by Cambridge University Press

the identity matrix of order p, e_{p+1} the last column of the identity matrix of order $p+1$, and 0_p, the $p \times 1$ vector of zeros. The notation $[A|B]$ is used for a matrix formed by horizontal blocks A and B. Then the seminorm to use is defined by $\|z\|^* = |y|$ and the projection matrix by $p_v = v[I_p|0_p][v|e_{p+1}]^{-1}$. In other words, the couple $(\|\cdot\|^*, p_v)$ must replace the couple $(\|\cdot\|, P_v)$ in (1) and (2). We must show that p_v is the projection matrix of the least squares, i.e., that p_v projects onto v parallel to e_{p+1}. Let us first verify that p_v is idempotent. To see that, define $V = [v|e_{p+1}]$ and note that $p_v = VHV^{-1}$, where

$$H = \begin{bmatrix} I_p & 0_p \\ 0_p' & 0 \end{bmatrix}.$$

It immediately follows that $p_v p_v = p_v$. We still have to show that $p_v v = v$ and that $p_v e_{p+1} = 0_{p+1}$. But $[p_v v|p_v e_{p+1}] = p_v[v|e_{p+1}] = p_v V = VH = [v|e_{p+1}]H = [v|0_{p+1}]$. With p_v and $\|\cdot\|^*$ instead of P_v and $\|\cdot\|$, the inertia in (2) becomes proportional to $\sum(y_i - \tilde{y}_i)^2$, where $\tilde{y}_i$ is the last coordinate of the projected point $p_v z_i$. It is well known that the optimal hyperplane Δ^* minimizing $\sum(y_i - \tilde{y}_i)^2$ is the mean squares hyperplane, in which case $\tilde{y}_i = \hat{y}_i$, the fitted value of linear regression. When Δ^* is the hyperplane of the mean squares, (3) becomes $SST = SSR + SSE$ and Q_Z in (4) becomes R^2.

99.6.3. *Variances of Truncated versus Full Distributions for Continuous Random Variables*—Solution, proposed by S.K. Sapra. In what follows, the expressions for $\mathrm{Var}(y^*)$ can be found in Johnson and Kotz (1972) and the expressions for $\mathrm{Var}(y)$ are easily derived.

(1) Pareto. We have

$$\mathrm{Var}(y^*) = a(a-2)^{-1}(a-1)^{-2}, \qquad a > 2, \quad \text{and} \tag{1}$$

$$\mathrm{Var}(y) = ak^2(a-2)^{-1}(a-1)^{-2}, \qquad a > 2, k > 1. \tag{2}$$

From equations (1) and (2), it follows that $\mathrm{Var}(y) > \mathrm{Var}(y^*)$ for $a > 2, k > 1$.

(2) Uniform. We have

$$\mathrm{Var}(y^*) = (d-c)^2/12, \quad \text{and} \tag{3}$$

$$\mathrm{Var}(y) = (d-k)^2/12. \tag{4}$$

Therefore, it follows from (3) and (4) that $\mathrm{Var}(y) < \mathrm{Var}(y^*)$ for $c < k < d$.

(3) Standard exponential. We have

$$\mathrm{Var}(y^*) = 1, \quad \text{and} \tag{5}$$

$$\mathrm{Var}(y) = \mathrm{Var}(y^*|y^* > k) = 1 \tag{6}$$

because $E(y^*|y^* > k) = k+1$, and $E(y^{*2}|y^* > k) = k^2 + 2k + 2$.
Hence, $\mathrm{Var}(y) = \mathrm{Var}(y^*)$.

https://doi.org/10.1017/S0266466600166095 Published online by Cambridge University Press

(4) Gamma (with scale parameter = 1, shape parameter = 2). We have

$$\mathrm{Var}(y^*) = 2, \quad \text{and} \tag{7}$$

$$\mathrm{Var}(y) = \mathrm{Var}(y^* | y^* > k) = 2 + [(2k^3 + k^2 + 4k)/(k+1)^2] \tag{8}$$

because $E(y^*|y^* > k) = (k^2 + 2k + 2)/(k+1)$, and $E(y^{*2}|y^* > k) = (k^3 + 3k^2 + 6k + 6)/(k+1)$.

Hence, it follows from equations (7) and (8) that $\mathrm{Var}(y) > \mathrm{Var}(y^*)$.

(5) Rayleigh. We have

$$\mathrm{Var}(y^*) = 2 - \pi/2, \quad \text{and}$$

$$\mathrm{Var}(y) = (\mathrm{Var}(y^*|y^* > k) = 2 + k^2 - (\pi/2)e^{k^2}[1 - F(k^2/2, 3/2)] = S, \quad \text{say,}$$

where $F(k^2/2,3/2) = 2/\sqrt{\pi}\int_0^{k^2/2} w^{1/2}e^{-w}dw$ is the incomplete gamma function at $k^2/2$.

After repeated integration by parts and dropping terms of order smaller than k^{-3}, we have

$$S = k^{-2} + 2k^{-4} - k^{-6}.$$

Therefore, $dS/dk = -2k^{-3} - 8k^{-5} + 6k^{-7} < 0$ for large k.

Hence, $\mathrm{Var}(y) < \mathrm{Var}(y^*)$ for large k.

For additional discussion on the variance of continuous nonnegative truncated distributions, see Mullooly (1988).

REFERENCES

Johnson, N.L. & S. Kotz (1972) *Distributions in Statistics—Continuous Univariate Distributions—1*. New York: Wiley.

Mullooly, J.P. (1988) The variance of left-truncated continuous nonnegative distributions. *American Statistician* 42, 208–210.

99.6.4. *Variances of Censored versus Full Distributions for Continuous Random Variables*—Solution, proposed by S.K. Sapra. In what follows, we derive conclusions about the relationship between $\mathrm{Var}(y)$ and $\mathrm{Var}(y^*)$ by examining the sign of dS/dk for each p.d.f. in the problem. The expressions for $\mathrm{Var}(y)$ are easily derived.

(1) Pareto. We have

$$\mathrm{Var}(y) = [2k^{2-a}(a-1) - k^{2-2a}(a-2)](a-1)^{-2}(a-2)^{-1} = S, \quad \text{say,}$$

$$\text{for} \quad a > 2, k > 1.$$

$$\text{Therefore, } dS/dk = 2k^{1-2a}(1-k^a)(a-1)^{-1} < 0 \quad \text{for} \quad a > 2, k > 1,$$

and S monotonically increases as k decreases to 1 (degree of censoring approaches 0).

Hence, $\mathrm{Var}(y) < \mathrm{Var}(y^*)$ for $a > 2, k > 1$.

https://doi.org/10.1017/S0266466600166095 Published online by Cambridge University Press

(2) Uniform. We have

$$\mathrm{Var}(y) = \tfrac{1}{3}[(d-c)^{-1}(2k^3 + d^3 - 3ck^2)]$$

$$- \tfrac{1}{4}[(d-c)^{-2}(k^2 + d^2 - 2cd)^2] = S, \quad \text{say, for} \quad c < k < d.$$

Therefore, for $c = 0, d = 1$, for example,

$$dS/dk = 2k^2 - k - k^3 = (k^2 - k^3) - (k - k^2) < 0$$

for $0 < k < 1$ because $(k^2 - k^3) < (k - k^2)$ for $0 < k < 1$.

Hence, $\mathrm{Var}(y) < \mathrm{Var}(y^*)$ for $0 < y^*, k < 1$.

It is not clear if a general relationship between $\mathrm{Var}(y)$ and $\mathrm{Var}(y^*)$ can be derived for arbitrary values of c and d.

(3) Standard exponential (with shape parameter = 1). We have

$$\mathrm{Var}(y) = e^{-k}(2 - e^{-k}) = S, \quad \text{say.}$$

Therefore, $dS/dk = 2(e^{-2k} - e^{-k}) < 0 \quad \text{for} \quad k > 0.$

Hence, $\mathrm{Var}(y) < \mathrm{Var}(y^*)$ for $k > 0$.

(4) Gamma (with scale parameter = 1, shape parameter = 2). We have

$$\mathrm{Var}(y) = 2ke^{-k} + 6e^{-k} - (k^2 + 4k + 4)e^{-2k} = S, \quad \text{say.}$$

Therefore, $dS/dk = 2(k+2)e^{-2k}(1 + k - e^k) < 0 \quad \text{for} \quad k > 0.$

Hence, $\mathrm{Var}(y) < \mathrm{Var}(y^*)$ for $k > 0$.

ADDENDUM

Two solutions to problem 99.3.1, the eigenvalues decomposition of a symmetric matrix, were published in *ET* 16(2); three other excellent solutions had been proposed, one of which was by Michel van de Velden and Heinz Neudecker, the posers of the problem. Heinz Neudecker has subsequently pointed out that their solution was shorter than the published solutions and asked to reconsider its publication. Although *ET* does not normally publish multiple solutions, in the present case the original solution is substantially shorter and is reported below.

99.3.1. *The Eigenvalues Decomposition of a Symmetric Matrix*—Solution, proposed by Michel van de Velden and Heinz Neudecker. The $n \times n$ matrix $\mathbf{S}$ is defined as

$$\mathbf{S} = \mathbf{XDY}' - \mathbf{YDX}',$$

https://doi.org/10.1017/S0266466600166095 Published online by Cambridge University Press

where $\mathbf{X}'\mathbf{X} = \mathbf{Y}'\mathbf{Y} = \mathbf{I}_k$, $\mathbf{X}'\mathbf{Y} = \mathbf{0}$, and $\mathbf{D}$ is a diagonal matrix with positive elements. Define $\mathbf{U} \equiv (\mathbf{X}, -\mathbf{Y})$, and $\mathbf{V} \equiv (\mathbf{Y}, \mathbf{X})$. Then,

$$\mathbf{S} = \mathbf{U}(\mathbf{I}_2 \otimes \mathbf{D})\mathbf{V}',$$

and $\mathbf{U}'\mathbf{U} = \mathbf{V}'\mathbf{V} = \mathbf{I}_{2k}$. Then, as $\text{rank}(\mathbf{S}) = 2k$, it follows that $n \geq 2k$ and $\text{rank}(\tilde{\mathbf{S}}) = 4k$, where

$$\tilde{\mathbf{S}} \equiv \begin{pmatrix} \mathbf{0} & -\mathbf{S} \\ \mathbf{S} & \mathbf{0} \end{pmatrix}.$$

We need to find matrices $\mathbf{A}$ and $\mathbf{B}$ (both of order $n \times 4k$), and a $4k \times 4k$ diagonal matrix Λ such that

$$\tilde{\mathbf{S}} \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} = \begin{pmatrix} \mathbf{A} \\ \mathbf{B} \end{pmatrix} \Lambda,$$

or, equivalently,

$$\begin{pmatrix} -\mathbf{SB} \\ \mathbf{SA} \end{pmatrix} = \begin{pmatrix} \mathbf{A}\Lambda \\ \mathbf{B}\Lambda \end{pmatrix},$$

and $\mathbf{A}'\mathbf{A} + \mathbf{B}'\mathbf{B} = \mathbf{I}_{4k}$.

As $\mathbf{SX} = -\mathbf{YD}$, and $\mathbf{SY} = \mathbf{XD}$, we can write $\tilde{\mathbf{S}} = \mathbf{T}\Lambda\mathbf{T}'$ where

$$\mathbf{T} = \frac{1}{\sqrt{2}} \begin{pmatrix} \mathbf{X} & \mathbf{Y} & \mathbf{X} & \mathbf{Y} \\ -\mathbf{Y} & \mathbf{X} & \mathbf{Y} & -\mathbf{X} \end{pmatrix}, \quad \Lambda = \begin{pmatrix} \mathbf{D} & & & \\ & \mathbf{D} & & \\ & & -\mathbf{D} & \\ & & & -\mathbf{D} \end{pmatrix},$$

and $\mathbf{T}'\mathbf{T} = \mathbf{I}_{4k}$.

CORRIGENDUM

Heinz Neudecker has pointed out that the solution to problem 99.3.2, maximum-likelihood estimation under singularity, proposed by Steve Lawford, contained questionable matrix differential calculus, and that the expression of the Hessian, equation (13), was incorrect and nonsymmetric. Heinz Neudecker's correct solution to the problem 99.3.2 is reported below.[1]

99.3.2. *Maximum-Likelihood Estimation under Singularity*—Solution, proposed by Heinz Neudecker. In the statement of the problem, subscript i is missing from y, x, ϵ in the regression equation $y_i = a + \Gamma x_i + \epsilon_i$ and in the expressions $E(\epsilon_i) = 0$, $D(\epsilon_i) = \Omega$. The proof that maximizing $\log L_c$ corresponds to maxi-

https://doi.org/10.1017/S0266466600166095 Published online by Cambridge University Press

mizing the likelihood function of $(I_{m-1},0)Y'$ concentrated with respect to Ω can be found in Barten (1969). We note that Magnus and Neudecker (1988, p. 329) gave a partially wrong derivation concerning the parameter a.

We first derive first-order conditions with respect to a and then calculate second-order conditions.

FIRST-ORDER CONDITION, SOLUTION 1

Implicit Treatment of the Structure of *a*

Start with the differential of $\log L_c$ due to perturbing a, namely

$$d\log L_c = -\frac{n}{2}\, trA^{-1}dA = -trA^{-1}E'dE = trA^{-1}E'1_n(da)' = (da)'A^{-1}E'1_n.$$

Note that $E' = Y' - a1_n' - \Gamma X'$, X and Y being defined similarly to E. Then write

$$a = \begin{pmatrix} \tilde{a} \\ a_m \end{pmatrix} = \begin{pmatrix} \tilde{a} \\ -1_{m-1}'\tilde{a} \end{pmatrix} = \begin{pmatrix} I_{m-1} \\ -1_{m-1}' \end{pmatrix}\tilde{a} = J_m\tilde{a}$$

using $1_m'a = 0$. Hence, $da = J_m d\tilde{a}$. This yields the first-order maximum condition

$$0 = \frac{\partial \log L_c}{\partial \tilde{a}} = J_m'A^{-1}E'1_n.$$

By using $1_m'E' = 0$, we get

$$\begin{pmatrix} J_m' \\ 1_m'A \end{pmatrix}A^{-1}E'1_n = 0.$$

Clearly

$$1_m'A = 1_m'(n^{-1}E'E + m^{-1}1_m1_m') = 1_m' \qquad \textbf{(1)}$$

implies

$$\begin{pmatrix} J_m' \\ 1_m' \end{pmatrix}A^{-1}E'1_n = 0.$$

The matrix $(J_m, 1_m)'$ is nonsingular, its determinant being equal to

$$|1 + 1_{m-1}'I_{m-1}1_{m-1}| = m \neq 0.$$

Hence, we get the first-order condition $E'1_n = 0$.

https://doi.org/10.1017/S0266466600166095 Published online by Cambridge University Press

FIRST-ORDER CONDITION, SOLUTION 2

Explicit Treatment of the Structure of *a*

Define the Lagrangean function $\psi(a) = \log L_c - \lambda 1'_m a$. Then $d\psi = d\log L_c - \lambda 1'_m da = (1'_n EA^{-1} - \lambda 1'_m)da$, from which issue the two equations

$$0 = \frac{\partial \psi}{\partial a} = A^{-1}E'1_n - \lambda 1_m \quad \text{and} \quad 0 = 1'_m a. \tag{2}$$

It is then straightforward to obtain

$$E'1_n = \lambda A 1_m = \lambda 1_m$$

by (1). Premultiplying by $1'_m$, we find $0 = 1'_m E'1_n = \lambda 1'_m 1_m = m\lambda$, and thus $\lambda = 0$. Substitution in (2) yields the first-order condition $E'1_n = 0$.

SECOND-ORDER CONDITIONS

Continuing from Solution 1, we derive the Hessian $\partial^2 \log L_c / \partial\tilde{a}\partial\tilde{a}'$ in the alleged maximum. We get

$$\begin{aligned} d^2 \log L_c &= d(1'_n EA^{-1}da) = 1'_n(dE)A^{-1}da + 1'_n E(dA^{-1})da \\ &= -1'_n 1_n (da)'A^{-1}da = -n(d\tilde{a})'J'_m A^{-1}J_m d\tilde{a}. \end{aligned}$$

Hence, the Hessian in the alleged maximum equals $-nJ'_m A^{-1}J_m$, which is negative definite. This means that a strict local maximum has been found.

When we treat the structure of a explicitly (Solution 2), we get as second-order condition $d^2\psi = d^2 \log L_c = -n(da)'A^{-1}da < 0$ for all $da \neq 0$ satisfying $1'_m da = 0$.

NOTE

1. The derivations can be slightly simplified by considering the likelihood function of $J'_m Y' = (I_{m-1}, -1_{m-1})Y'$ instead of the one of $(I_{m-1}, 0)Y'$, as done in Barten (1969) and in the solution by Heinz Neudecker. In the latter case, the first $m - 1$ equations of the system are retained, whereas the former case considers the difference of the first $m - 1$–dependent variables w.r.t. the last dependent variable.

REFERENCES

Barten, A.P. (1969) Maximum likelihood estimation of a complete system of demand equations. *European Economic Review* 1, 7–73.

Magnus, J.R. & H. Neudecker (1988) *Matrix Differential Calculus with Applications in Statistics and Econometrics.* Chichester: John Wiley & Sons.

https://doi.org/10.1017/S0266466600166095 Published online by Cambridge University Press