Econometric Theory, **14**, 1998, 151–159. Printed in the United States of America.

PROBLEMS AND SOLUTIONS

PROBLEMS

98.1.1. *Similarity and Distance Matrices*, proposed by Heinz Neudecker and Michel Van de Velden. Let $C = (c_{ij})$ be a positive semidefinite similarity matrix and $D = (d_{ij})$ be a distance matrix obtained from C by the definition $d_{ij} := (c_{ii} + c_{jj} - 2c_{ij})^{1/2}$. Show, by using the vector triangle inequality, that the distance d_{ij} satisfies the triangle inequality $d_{ij} \leq d_{ik} + d_{jk}$.

98.1.2. *Lower Eigenbound for* AR(1) *Disturbance Covariance Matrix*, proposed by Kyung-Taik Han and Eric Iksoon Im. Let $\Omega(T \times T)$ denote the disturbance covariance matrix for the standard linear model with AR(1) residuals. Then,

$$\Omega = \frac{1}{1-\rho^2}\begin{bmatrix} 1 & \rho & & \dots & \rho^{T-2} & \rho^{T-1} \\ \rho & 1 & \rho & \dots & \rho^{T-3} & \rho^{T-2} \\ & & & \ddots & & \\ \rho^{T-2} & \rho^{T-3} & & \dots \ \rho & 1 & \rho \\ \rho^{T-1} & \rho^{T-2} & & \dots & \rho & 1 \end{bmatrix}, \qquad (1)$$

where $|\rho| < 1$ and $2 \leq T < \infty$.

(a) Show that

$$\lambda_i(\Omega) > \frac{1}{(1+|\rho|)^2} \qquad (i = 1,2,\dots,T), \qquad (2)$$

where $\lambda_i(\cdot)$ denote the eigenvalues in descending order: $\lambda_1(\cdot) \geq \lambda_2(\cdot) \geq \cdots \geq \lambda_T(\cdot)$.

(b) Show that $\lim_{|\rho|\to 1} \lim_{T\to\infty} \lambda_i(\Omega) \geq \frac{1}{4}$.

98.1.3. *Equivalence of LR Test and Hausman Test*, proposed by Hailong Qian. Suppose that we have the following two equations:

$$y_{gt} = \alpha_g + \varepsilon_{gt}, \qquad g = 1,2; \qquad t = 1,2,\dots,T,$$

where $(\varepsilon_{1t}, \varepsilon_{2t})$ is independently and identically distributed normal with mean zero and variance $\Omega = \begin{bmatrix} \sigma_1^2 & \sigma_{12} \\ \sigma_{21} & \sigma_2^2 \end{bmatrix}$. Show that the likelihood ratio test of H_0: $\sigma_1^2 = \sigma_2^2$ is asymptotically equivalent to the Hausman test.

 0266-4666/98 $9.50

SOLUTIONS

97.1.1. *Standard Errors for the Long-Run Variance Matrix*—Solution, proposed by Paolo Paruolo. The solution follows by applying Cramér's theorem (i.e., the delta method) to the functions $g(C,\Omega) = C\Omega C'$ and $h(C,\Omega) = CA$. It is known that

$$T^{1/2}\,\mathrm{vec}(\hat{C} - C) \xrightarrow{w} N\,(0, \xi'\Sigma^{-1}\xi \otimes C\Omega C')$$

(see Johansen, 1995, Theorem 13.7; Paruolo, 1997, Theorem 7.1) and

$$T^{1/2}\,\mathrm{vec}(\hat{\Omega} - \Omega) \xrightarrow{w} N(0, 2P_D(\Omega \otimes \Omega)P_D)$$

(see, i.e., Lütkepohl, 1991, p. 85). Moreover, $\hat{C}$ and $\hat{\Omega}$ are asymptotically independent because $\hat{C}$ is a function of $\hat{\vartheta}$ where $\vartheta = (\alpha, \beta, \Gamma_i, i = 1, \ldots, k-1)$ and $\hat{\vartheta}$ and $\hat{\Omega}$ are asymptotically independent (see, e.g., Paruolo, 1997, Lemma 5.1). Hence, no covariance terms arise in (3) and (5).

(1) A first-order expansion of $g(\cdot,\cdot)$ gives

$$T^{1/2}\,\mathrm{vec}(\hat{C}\hat{\Omega}\hat{C}' - C\Omega C') \simeq T^{1/2}(I + K)(C\Omega \otimes I)\mathrm{vec}(\hat{C} - C) + T^{1/2}(C \otimes C)\mathrm{vec}(\hat{\Omega} - \Omega), \qquad \textbf{(6)}$$

where K is the commutation matrix of order p (see Magnus and Neudecker, 1988, Sec. 3.7) (henceforth MN). Note that $(I + K) = 2P_D$ (see MN, equation (3.8.7, p. 49). Because the two terms on the right-hand side (r.h.s.) of (6) are asymptotically independent, the asymptotic variance of the left-hand side is equal to the sum of the asymptotic variances of the terms on the r.h.s. The first term gives rise to the variance $4P_D(C\Omega\xi'\Sigma^{-1}\xi\Omega C' \otimes C\Omega C')P_D$ and the second to $2P_D(C\Omega C' \otimes C\Omega C')P_D$ because $(C \otimes C)P_D = P_D(C \otimes C)$ (cf. MN, equation (3.8.14), p. 50). Summing the preceding expressions, the asymptotic variance in (3) is obtained.

(2) Differentiating the Choleski decomposition one obtains

$$\frac{\partial\,\mathrm{vec}(A)}{\partial\,\mathrm{vec}(\Omega)'} = \frac{1}{2}D(D'(A \otimes I)D)^{-1}D'.$$

Thus, a first-order expansion of $h(\cdot,\cdot)$ gives

$$T^{1/2}\,\mathrm{vec}(\hat{C}^* - C^*) \simeq T^{1/2}(A' \otimes I)\mathrm{vec}(\hat{C} - C) + T^{1/2}B\,\mathrm{vec}(\hat{\Omega} - \Omega), \qquad \textbf{(7)}$$

where B is define after (5). Again, the two terms on the r.h.s. of (7) are asymptotically independent; the first term gives rise to $(A'\xi'\Sigma^{-1}\xi A \otimes C\Omega C')$ and the second to $\frac{1}{2}B(\Omega \otimes \Omega)B'$, thus proving (5).

(3) The asymptotic variance matrix in (3) is singular as a result of two factors: (a) the long-run variance is symmetric (which is reflected in the presence of the singular projection matrix P_D in (3)); (b) the impact matrix C is singular. The asymptotic variance matrix in (5) is singular because of factor (b) only.

To illustrate point (b), consider the linear combination $R_1'ZR_2$, where $Z = C\Omega C'$, i.e., $R'\,\mathrm{vec}(Z)$ with $R' = (R_2' \otimes R_1')$. This linear combination has asymptotic variance matrix

$$G = 2R'P_D(CFC' \otimes C\Omega C')P_D R,$$

where $F = \Omega + 2\Omega\xi'\Sigma^{-1}\xi\Omega$. Now

$$R'P_D = \tfrac{1}{2}(R_2' \otimes R_1')(I + K) = \tfrac{1}{2}((R_2' \otimes R_1') + K(R_1' \otimes R_2'))$$

(see MN, Theorem 3.9(a)), and it is simple to verify that $G = 0$ when either R_1 and/or R_2 belongs to span(β), such that $R_i'C = 0$.

Take now $Z = C^*$. The asymptotic variance matrix of $R'\,\text{vec}(Z)$ is in this case

$$G = (R_2'A'\xi'\Sigma^{-1}\xi AR_2' \otimes R_1'C\Omega C'R_1) + \tfrac{1}{2}R'B(\Omega \otimes \Omega)B'R,$$

where $R'B = (R_2' \otimes R_1'C)D(D'(A \otimes I)D)^{-1}D'$. It is easy to verify that $G = 0$ whenever $R_1 \in \text{span}(\beta)$, such that $R_1'C = 0$.

(4) If μ is added in (1), the representations (2) and (4) possibly present a linear trend. However nothing changes in the preceding results (3) and (5).

REFERENCES

Johansen, S. (1995) *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models.* Oxford: Oxford University Press.

Lütkepohl, H. (1991) *Introduction to Multiple Time Series Analysis.* Berlin: Springer-Verlag.

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

Paruolo, P. (1997) Asymptotic inference on the moving average impact matrix in cointegrated I(1) VAR systems. *Econometric Theory* 13, 79–118.

97.1.2. *Asymptotic Inefficiency of an Estimator Derived from a Kernel-Based Test Statistic*—Solution, proposed by Oliver Linton. Although the criterion function is a U-statistic of order two, its asymptotic properties follow from standard arguments. Write

$$\begin{aligned} Q_n(\beta) &= \frac{1}{n(n-1)h^d}\sum_i\sum_{j\neq i} u_i(\beta)u_j(\beta)K_{ij} \\ &= \frac{1}{n(n-1)h^d}\sum_i\sum_{j\neq i} u_i u_j K_{ij} \\ &\quad - \frac{1}{n(n-1)h^d}\sum_i\sum_{j\neq i} (\beta-\beta_0)^T\{x_i u_j + x_j u_i\}K_{ij} \\ &\quad + \frac{1}{n(n-1)h^d}\sum_i\sum_{j\neq i} (\beta-\beta_0)^T x_i^T x_j(\beta-\beta_0)K_{ij} \\ &= Q_{n1} + (\beta-\beta_0)^T Q_{n2} + (\beta-\beta_0)^T Q_{n3}(\beta-\beta_0), \end{aligned}$$

with Q_{nj} implicitly defined. In fact, $Q_{n1} = O_p(n^{-1}h^{-d/2})$, $Q_{n2} = O_p(n^{-1/2})$, and $Q_{n3} = O_p(1)$. Therefore,

$$Q_n(\beta) \to_p Q(\beta) \equiv (\beta-\beta_0)^T Q_3(\beta-\beta_0),$$

where $Q_{n3} \to_p Q_3 > 0$. The preceding convergences are uniform in β by inspection. Clearly, $Q(\beta)$ is uniquely minimized at β_0. In conclusion, $\tilde{\beta} \to_p \beta_0$.

By virtue of the quadratic shape of $Q_n(\beta)$, we have

$$0 = \frac{\partial Q_n(\tilde{\beta})}{\partial \beta} = Q_{n2} + 2Q_{n3}(\tilde{\beta} - \beta_0),$$

where

(1) $n^{1/2}Q_{n2}(\beta_0) \to N(0,4[E\{XX^T\sigma_u^2(X)f^2(X)\}])$ in distribution
(2) $Q_3 = E\{XX^T f(X)\} > 0$,

using the arguments of Fan and Li (1996). The result follows.

REFERENCE

Fan, J. & Q. Li (1996) Consistent model specification tests: Omitted variables and semiparametric functional forms. *Econometrica* 64, 865–890.

97.1.3. *A Joint Test for Functional Form and Random Individual Effects*—Solution,[1] proposed by Dong Li. We rewrite (3) as

$$f_{it}(y_{it},\phi) = \varepsilon_{it}, \qquad i = 1,\dots,N; \qquad t = 1,\dots,T, \quad \text{with } \varepsilon_{it} \sim NID(0,1), \tag{4}$$

where

$$f_{it}(y_{it},\phi) \equiv \frac{1}{\sigma_\nu}\left[B^*(y_{it},\lambda) - \sum_{k=1}^{K}\beta_k B^*(X_{itk},\lambda) - \sum_{s=1}^{S}\gamma_s Z^*_{its}\right] \tag{5}$$

and $\phi = (\beta,\gamma,\lambda,\theta,\sigma_\nu)$.

The contribution of the itth observation to the loglikelihood function $l(y,\theta)$ is

$$l_{it}(y_{it},\phi) = -\tfrac{1}{2}\log(2\pi) - \tfrac{1}{2}f_{it}^2(y_{it},\phi) + k_{it}(y_{it},\phi), \tag{6}$$

where

$$k_{it}(y_{it},\phi) \equiv \log\left|\frac{\partial f_{it}(y_{it},\phi)}{\partial y_{it}}\right|$$

is a Jacobian term. Define

$$F_{itj}(y_{it},\phi) = \frac{\partial f_{it}(y_{it},\phi)}{\partial \phi_j} \quad \text{and} \quad K_{itj}(y_{it},\phi) = \frac{\partial k_{it}(y_{it},\phi)}{\partial \phi_j}. \tag{7}$$

Then $F(y,\phi)$ and $K(y,\phi)$ are the $NT \times (K+S+3)$ matrices with typical elements $F_{itj}(y_{it},\phi)$ and $K_{itj}(y_{it},\phi)$, respectively.

Let $f(y,\phi)$ be the NT vector with typical elements $f_{it}(y_{it},\phi)$. Then the decreasing likelihood ratio (DLR) can be written as

$$\begin{bmatrix} f(y,\phi) \\ \iota_{NT} \end{bmatrix} = \begin{bmatrix} -F(y,\phi) \\ K(y,\phi) \end{bmatrix} b + \text{residuals}, \tag{8}$$

where ι_{NT} denotes a vector of ones of dimension NT. This artificial regression is double-length with $2NT$ observations.

https://doi.org/10.1017/S0266466698141075 Published online by Cambridge University Press

From (5), we can obtain

$$\frac{\partial f_{it}(y_{it},\phi)}{\partial\beta_k} = -\frac{1}{\sigma_\nu}B^*(X_{itk},\lambda), \qquad \frac{\partial f_{it}(y_{it},\phi)}{\partial\gamma_s} = -\frac{1}{\sigma_\nu}Z^*_{itk},$$

$$\frac{\partial f_{it}(y_{it},\phi)}{\partial\lambda} = \frac{1}{\sigma_\nu}\left[\left(C(y_{it},\lambda) - \theta\frac{\sum_{t=1}^{T}C(y_{it},\lambda)}{T}\right)\right.$$
$$\left. - \sum_{k=1}^{K}\beta_k\left(C(X_{itk},\lambda) - \theta\frac{\sum_{t=1}^{T}C(X_{itk},\lambda)}{T}\right)\right],$$

$$\frac{\partial f_{it}(y_{it},\phi)}{\partial\theta} = -\frac{1}{\sigma_\nu}\frac{\sum_{t=1}^{T}u_{it}}{T}, \qquad \frac{\partial f_{it}(y_{it},\phi)}{\partial\sigma_\nu} = -\frac{u^*_{it}}{\sigma_\nu^2},$$

where $C(x,\lambda) \equiv \lambda x^\lambda \log x - x^\lambda + 1/\lambda^2$. (Note that there is a constant term in Z.)

Also, from (5), the Jacobian term for the itth observation is given by

$$k_{it}(y_{it},\phi) \equiv \log\left|\frac{\partial f_{it}(y_{it},\phi)}{\partial y_{it}}\right| = \log\left(1 - \frac{\theta}{T}\right) + (\lambda - 1)\log y_{it} - \log(\sigma_\nu), \tag{9}$$

so that

$$\frac{\partial k_{it}(y_{it},\phi)}{\partial\beta_k} = \frac{\partial k_{it}(y_{it},\phi)}{\partial\gamma_s} = 0, \qquad \frac{\partial k_{it}(y_{it},\phi)}{\partial\lambda} = \log y_{it},$$

$$\frac{\partial k_{it}(y_{it},\phi)}{\partial\theta} = -\frac{1}{T-\theta}, \qquad \frac{\partial k_{it}(y_{it},\phi)}{\partial\sigma_\nu} = -\frac{1}{\sigma_\nu}.$$

For H_0^a: $\lambda = 1$ and $\theta = 0$, (3) becomes

$$y_{it} = \sum_{k=1}^{K}\beta_k X_{itk} + \sum_{s=1}^{S}\gamma_s Z_{its} + u_{it}. \tag{10}$$

The regressand of the DLR has itth element $\hat{u}_{it}/\hat{\sigma}_\nu$ and $(it + NT)$th element 1, where $\hat{u}_{it}$ and $\hat{\sigma}_\nu$ denote the restricted maximum likelihood (ML) estimates of the itth residual and σ_ν, respectively. The typical elements for the first NT and the second NT observations of the regressors are then

for β_k: $(1/\hat{\sigma}_\nu)(X_{itk} - 1)$ and 0;
for γ_s: $(1/\hat{\sigma}_\nu)Z_{its}$ and 0;
for λ: $-(1/\hat{\sigma}_\nu)[y_{it}\log(y_{it}) - y_{it} + 1 - \sum_{k=1}^{K}\hat{\beta}_k(X_{itk}\log(X_{itk}) - X_{itk} + 1)]$ and $\log(y_{it})$;
for θ: $(1/\hat{\sigma}_\nu)\sum_{t=1}^{T}\hat{u}_{it}/T$ and $-1/T - \hat{\theta}$;
for σ_ν: $(1/\hat{\sigma}_\nu^2)\hat{u}_{it}$ and $-1/\hat{\sigma}_\nu$.

The test statistic is $2NT$ − the residuals sum of squares. It is asymptotically distributed as χ_2^2 under the null hypothesis H_0^a.

https://doi.org/10.1017/S0266466698141075 Published online by Cambridge University Press

For H_0^b: $\lambda = 0$ and $\theta = 0$, (3) becomes

$$\log(y_{it}) = \sum_{k=1}^{K} \beta_k \log(X_{itk}) + \sum_{s=1}^{S} \gamma_s Z_{its} + u_{it}. \tag{11}$$

The regressand of the DLR has itth element $\tilde{u}_{it}/\tilde{\sigma}_\nu$ and $(it + NT)$th element 1, where $\tilde{u}_{it}$ and $\tilde{\sigma}_\nu$ denote the restricted ML estimates of the itth residual and σ_ν, respectively. The typical elements for the first NT and the second NT observations of the regressors are then

for β_k: $1/\tilde{\sigma}_\nu \log(X_{itk})$ and 0;
for γ_s: $(1/\tilde{\sigma}_\nu)Z_{its}$ and 0;
for λ: $-(1/\tilde{\sigma}_\nu)[(\log(y_{it}))^2/2 - \sum_{k=1}^{K}\hat{\beta}_k(\log(X_{itk}))^2/2]$ and $\log(y_{it})$ (by noting $\lim_{\lambda\to 0}(\lambda x^\lambda \log x - x^\lambda + 1/\lambda^2) = (\log x)^2/2)$;
for θ: $(1/\tilde{\sigma}_\nu)(\sum_{t=1}^{T}\tilde{u}_{it}/T)$ and $-1/T - \theta$;
for σ_ν: $(1/\tilde{\sigma}_\nu^2)\tilde{u}_{it}$ and $-1/\tilde{\sigma}_\nu$.

The test statistic is $2NT -$ the residuals sum of squares. It is asymptotically distributed as χ_2^2 under the null hypothesis H_0^b.

NOTE

1. An excellent solution has been proposed independently by Badi H. Baltagi, the poser of the problem.

97.1.4. *Least-Squares Approximation of Off-Diagonal Elements of a Variance Matrix in the Context of Factor Analysis*—Solution, proposed by Albert Satorra and Heinz Neudecker. Clearly, we need to maximize

$$\varphi = \operatorname{tr} E_u^2 = (\operatorname{vec} E_u)' \operatorname{vec} E_u = (\operatorname{vec} E)'(I - K_d)\operatorname{vec} E, \tag{1}$$

where $E := C - AA'$ and K is the commutation matrix. Note that generically $K_d \operatorname{vec} P = \operatorname{vec} P_d$. Clearly, $(I - K_d)^2 = I - K_d$.

The differential of φ is

$$\begin{aligned} d\varphi &= 2(\operatorname{vec} E)'(I - K_d)\operatorname{vec} dE \\ &= -2(\operatorname{vec} E)'(I - K_d)\operatorname{vec}[(dA)A' + AdA'] \\ &= -2(\operatorname{vec} E)'(I - K_d)(I + K)(A \otimes I)\operatorname{vec} dA \\ &= -4(\operatorname{vec} E)'(I - K_d)(A \otimes I)\operatorname{vec} dA \\ &= -4(\operatorname{vec} E_u)'(A \otimes I)\operatorname{vec} dA. \end{aligned}$$

We used the obvious equalities $\operatorname{vec}(dA)A' = (A \otimes I)\operatorname{vec} dA$, $\operatorname{vec} AdA' = (I \otimes A)\operatorname{vec} dA' = (I \otimes A)K \operatorname{vec} dA = K(A \otimes I)\operatorname{vec} dA$, and $(I - K_d)(I + K)\operatorname{vec} E = 2(I - K_d)\operatorname{vec} E$. Thus, the first-order condition for the solution A is

$$(A' \otimes I)\operatorname{vec} E_u = 0,$$

https://doi.org/10.1017/S0266466698141075 Published online by Cambridge University Press

which is equivalent to

$$E_u A = 0. \quad \textbf{(2)}$$

Because

$$E_u = (C - AA')_u = C_u + (AA')_d - AA',$$

we write (2) as

$$[C_u + (AA')_d]A = AA'A. \quad \textbf{(3)}$$

Using equation (3), we obtain

$$\begin{aligned} E_u^2 &= [C_u + (AA')_d][C_u + (AA')_d] + AA'AA' \\ &\quad - [C_u + (AA')d]AA' - AA'[C_u + (AA')_d] \\ &= [C_u + (AA')_d][C_u + (AA')_d] + AA'AA' - AA'AA' - AA'AA' \\ &= [C_u + (AA')_d][C_u + (AA')_d] - AA'AA'. \end{aligned}$$

Consequently, φ of (1) equals

$$\varphi = \mathrm{tr}[C_u + (AA')_d]^2 - \mathrm{tr}(A'A)^2. \quad \textbf{(4)}$$

Clearly the solution for A is invariant under orthogonal transformation. That is, if A is a solution, then $\tilde{A} = AT$ where T is orthogonal is also a solution. Without loss of generality, we take $A'A = \Lambda$ where $\Lambda = \Lambda_d$, a positive definite diagonal matrix. Under this normalization of A, equations (3) and (4) become

$$[C_u + (AA')_d]A = A\Lambda \quad \textbf{(5)}$$

and

$$\varphi = \mathrm{tr}[C_u + (AA')_d]^2 - \mathrm{tr}\,\Lambda^2. \quad \textbf{(6)}$$

To attain a minimum in (6), we need to search for a maximum of tr Λ^2.

Equations (5) and (6) suggest the following iterative procedure to find the matrix A:

(1) Choose an arbitrary diagonal matrix $U = U_d > 0$ and compute $C^* := C_u + U$.
(2) Compute the eigenvalues and associated eigenvectors of C^* and determine A and Λ so that the columns of A are appropriately scaled eigenvectors of C^* associated with the m largest eigenvalues. The diagonal elements of Λ will thus be the m largest eigenvalues of C^*.
(3) In step (1), replace U by $(AA')_d$.
(4) Iterate (1)–(3) until stability has been reached.

This solution is called the principal factor analysis solution. To obtain the diagonal matrix U_d in step (1), a simple procedure is to set $U_d = (WW')_d$ with W a $p \times m$ matrix whose columns are eigenvectors of C associated with the m largest eigenvalues and of squared length equal to the associated eigenvalue.

97.1.5. *Inconsistency of Minimum Variance Quadratic Unbiased Estimators under Non-Gaussian Compound Normal Distribution*—Solution,[1] proposed by Jean-Daniel Rolle.

1. Because $U \sim CN(0,\gamma^2 I,\phi_H)$, then $y = X\beta + U \sim CN(X\beta,\gamma^2 I,\phi_H)$. Let $y'Ay$ be a potential estimator for σ^2. Nonnegativity and unbiasedness ($E(y'Ay) = \sigma^2$) imply $AX = 0$, as one can show easily. Hence $A\mu = AX\beta = 0$, and we have that $E(y'Ay) = \sigma^2 \operatorname{tr} A$. From (3), we have $\operatorname{Var}(y'Ay) = \kappa_H \sigma^4 (\operatorname{tr} A)^2 + 2\sigma^4(\kappa_H + 1)\operatorname{tr} A^2$. Define the set of matrices $\mathcal{A} = \{A \in \mathbb{R}^{N\times N}; A \geqslant 0, AX = 0, \operatorname{tr} A = 1\}$. The quadratics $y'Ay$ with $A \in \mathcal{A}$ are unbiased for σ^2. We are seeking an optimal estimator $y'A^*y$ of σ^2, with $A^* \in \mathcal{A}$, such that $\operatorname{Var}(y'A^*y) \leq \operatorname{Var}(y'Ay)$ for all $A \in \mathcal{A}$. From a computational point of view, it is very useful to characterize the matrices $A \in \mathcal{A}$. To do so, let us define the open set of matrices $\mathcal{P} = \{P \in \mathbb{R}^{N\times N}; PM_X \neq 0\}$, where $M_X = I - XX^+$ and X^+ is the Moore–Penrose inverse of X. Let us prove the following lemma.

LEMMA 1. *The function* $h_1 : \mathcal{P} \to \mathcal{A}$ *given by* $h_1(P) = \|PM_X\|^{-2} M_X P'PM_X$ *is surjective* (*i.e.*, $h_1(\mathcal{P}) = \mathcal{A}$).

Proof. Let $A \geqslant 0$, with $AX = 0$. Here $A \geqslant 0$ implies the existence of an $N \times N$ matrix B such that $A = B'B$. Next, $0 = AX = B'BX$ implies $X'B'BX = 0$. Therefore $BX = 0$. The general solution of $BX = 0$ is $B = PM_X$ (see Magnus and Neudecker 1988, ex. 4 p. 38), where P is an arbitrary matrix of appropriate order. Hence $A = B'B = M_X P'PM_X$. The unbiasedness condition imposes that $\operatorname{tr} A = 1$, that is, $\operatorname{tr}(M_X P'P) = \|PM_X\|^2 = 1$. Hence, $A = \|PM_X\|^{-2} M_X P'PM_X$.

To continue the proof of Lemma 1, define the function $h_2 : \mathcal{A} \to \mathbb{R}$ by $h_2(A) = \operatorname{var}(y'Ay)$ and $h : \mathcal{P} \to \mathbb{R}$ by $h = h_2 \circ h_1$. We will find a $P^* \in \mathcal{P}$ such that $h(P^*) \leq h(P)$. Because the function h_1 is surjective, h_2 takes a minimum at $A^* = h_1(P^*)$, i.e., $h_2(A^*) \leq h_2(A)$ for all $A \in \mathcal{A}$. Using the representation $A = \|PM_X\|^{-2} M_X P'PM_X$ and (3), one has $h(P) = \operatorname{Var}(y'Ay) = \kappa_H \sigma^4 + 2(\kappa_H + 1)\sigma^4 [\operatorname{tr}(M_X P'P)^2][\operatorname{tr}(M_X P'P)]^{-2}$. Therefore we can equivalently minimize on the open set $\mathcal{P}$ the differentiable function $\mu(P) = \operatorname{tr}(M_X P'P)^2/(\operatorname{tr} M_X P'P)^2$. Using Theorem 2 in Rolle (1996, p. 265), a global minimizer P^* of $\mu(P)$ is given by $(N - rX)^{-1/2} J'$, where J is such that $JJ' = M_X$. Hence the optimal matrix $A^* \in \mathcal{A}$ at which h_2 takes a minimum is given by $A^* = h_1(P^*) = \|J'M_X\|^{-2} M_X JJ'M_X = M_X/(N - rX)$, noting that idempotence of M_X implies $rM_X = \operatorname{tr} M_X = N - rX$. The optimal estimator is then

$$\hat{\sigma}^2 = \frac{1}{N - rX}\, y'(I - XX^+)y. \tag{4}$$

Note that this estimator does not depend on any particular (mixing) distribution of τ and hence of any particular compound normal distribution (in that sense, it is a uniform minimum variance quadratic unbiased estimator). ∎

2. Noting that $M_X X = 0$, we have $\hat{\sigma}^2 = y'M_X y/(N - rX) = U'M_X U/(N - rX)$. Next, by assumption $U \sim CN(0,\gamma^2 I,\phi_H)$, for some H. It follows that $U =$

$\tau^{1/2}Z$, where Z is $N(0,\gamma^2 I)$ and τ, Z are independent. Hence $\hat{\sigma}^2 = (1/N - rX)U'M_X U = (1/N - rX)\tau Z'M_X Z = (\tau\gamma^2/N - rX)Z'(M_X/\gamma^2)Z = (\tau\gamma^2/N - rX)\chi^2_{N-rX}$. But $\chi^2_{N-rX}/(N - rX)$ converges in probability to 1, and $\hat{\sigma}^2$ converges in probability to $\gamma^2\tau$. That is, $\hat{\sigma}^2$ converges in probability to a random variable. Hence, $\hat{\sigma}^2$ is inconsistent (unless in the Gaussian case, where H is degenerated on $\tau = 1$ and where $\hat{\sigma}^2$ converges in probability to $\gamma^2 = \sigma^2$ because $-2\phi'(0) = 1$). Another way to see this is to note that the components of U are exchangeable (but not independent). The law of large numbers for such sequences allows the limit to be a random variable. This is what happens here.

NOTE

1. An excellent solution has been proposed independently by David Greene.

REFERENCES

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

Rolle, J.D. (1996) Optimization of functions of matrices with application in statistics and econometrics. *Linear Algebra and Its Applications* 234, 261–275.

ERRATUM

An excellent solution by G. Trenkler to Problem 96.2.3 was omitted from the reference to the published solution in vol. 13, no. 3, p. 466.

https://doi.org/10.1017/S0266466698141075 Published online by Cambridge University Press