Econometric Theory, **17**, 2001, 483–491. Printed in the United States of America.

PROBLEMS AND SOLUTIONS

PROBLEMS

01.2.1. *A Determinantal Inequality*, proposed by Heinz Neudecker. Consider a positive definite matrix A and its diagonal submatrix A_d. Show then that $|A| \leq |A_d|$ without using induction.

REFERENCE

Anderson, T.W. (1958) *An Introduction to Multivariate Statistical Analysis.* New York: John Wiley & Sons.

01.2.2. *The R/S Statistics as a Unit Root Test*, proposed by Giuseppe Cavaliere. Given a time series $\{X_t\}_{t=0,\ldots,T}$, consider the *rescaled range statistics* (see Hurst, 1951; and the recent generalization by Lo, 1991) computed on the differenced process ΔX_t:

$$R/S = \frac{1}{\hat{\lambda}_T\sqrt{T}}\left(\max_{t=1,\ldots,T}\sum_{i=1}^{t}(\Delta X_i - \hat{\mu}) - \min_{t=1,\ldots,T}\sum_{i=1}^{t}(\Delta X_i - \hat{\mu})\right)$$

$$\Delta X_t = X_t - X_{t-1}, \hat{\mu} = T^{-1}\sum_{t=1}^{T}\Delta X_t = T^{-1}(X_T - X_0),$$

where $\hat{\lambda}_T^2$ is a kernel HAC estimator of the long-run variance of ΔX_t. Show that R/S can be used to test the null hypothesis

$$H_0 : X \sim I(1), X_t := X_0 + \mu t + S_t, S_t := \sum_{i=1}^{t} u_i, |X_0| < \infty \text{ a.s.}$$

(a) by deriving the asymptotic distribution of R/S under H_0 and by showing that a right tail test based on R/S is consistent against
(b) $I(2)$ *alternatives*, i.e., $X_t = X_0 + \mu t + \sum_{i=1}^{t}\sum_{j=1}^{i} u_j$
(c) $I(1)$ *with trend breaks*, i.e., $X_t = X_0 + \mu t + (\mu_0 - \mu)\mathbb{I}(t > [\alpha T]) + \sum_{i=1}^{t} u_i$, $\alpha \in (0,1)$ under the assumption that $\{u_t\}$ is strong mixing (see Hansen, 1992, Condition V1) and that $\hat{\lambda}_T^2$ has kernel function $k(\ )$ satisfying Assumption 1 of de Jong (2000) and truncation lag $q_T = cT^{\gamma}$, $\gamma < \frac{1}{2} - 1/r$.

REFERENCES

de Jong, R.M. (2000) A strong consistency proof for heteroskedasticity and autocorrelation consistent covariance matrix estimators. *Econometric Theory* 16, 262–268.
Hansen, B.E. (1992) Consistent covariance matrix estimation for dependent heterogeneous processes. *Econometrica* 60, 967–972.
Hurst, H. (1951) Long-term storage capacity of reservoirs. *Transactions of the American Society of Civil Engineers* 116, 770–799.
Lo, A.W. (1991) Long-term memory in stock market prices. *Econometrica* 59, 1279–1313.

 0266-4666/01 $9.50

https://doi.org/10.1017/S0266466601172105 Published online by Cambridge University Press

SOLUTIONS

00.2.1. *Degeneration of Feasible GLS to 2SLS in a Limited-Information Simultaneous Equation Model*—Solution, proposed by Chuanming Gao and Kajal Lahiri. The limited-information simultaneous equations model may be written as a SUR model:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} y_2 & 0 \\ 0 & X \end{bmatrix} \begin{bmatrix} \gamma \\ \beta \end{bmatrix} + \begin{bmatrix} u \\ v \end{bmatrix}.$$

Based on a consistent

$$\hat{\Sigma} = \begin{bmatrix} \hat{\sigma}_{11} & \hat{\sigma}_{12} \\ \hat{\sigma}_{12} & \hat{\sigma}_{22} \end{bmatrix},$$

feasible GLS is carried out as

$$\begin{bmatrix} \hat{\gamma}_{\text{FGLS}} \\ \hat{\beta}_{\text{FGLS}} \end{bmatrix} = \left[\begin{pmatrix} y_2 & 0 \\ 0 & X \end{pmatrix}' \hat{\Sigma}^{-1} \begin{pmatrix} y_2 & 0 \\ 0 & X \end{pmatrix} \right]^{-1} \left[\begin{pmatrix} y_2 & 0 \\ 0 & X \end{pmatrix}' \hat{\Sigma}^{-1} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right].$$

Substituting in

$$\hat{\Sigma}^{-1} = \frac{1}{\hat{\sigma}_{11}\hat{\sigma}_{22} - \hat{\sigma}_{12}^2} \begin{bmatrix} \hat{\sigma}_{22} & -\hat{\sigma}_{12} \\ -\hat{\sigma}_{12} & \hat{\sigma}_{11} \end{bmatrix}$$

and using partitioned inversion, we have

$$\begin{aligned} \hat{\gamma}_{\text{FGLS}} &= Q[\hat{\sigma}_{22} y_2' y_1 - \hat{\sigma}_{12} y_2' y_2 - \hat{\sigma}_{12}^2 \hat{\sigma}_{11}^{-1} y_2' P y_1 + \hat{\sigma}_{12} y_2' P y_2] \\ &= Q[\hat{\sigma}_{22} y_2' y_1 - \hat{\sigma}_{12} y_2' M y_2 - \hat{\sigma}_{12}^2 \hat{\sigma}_{11}^{-1} y_2' P y_1], \end{aligned}$$

where $Q^{-1} = \hat{\sigma}_{22} y_2' y_2 - \hat{\sigma}_{12}^2 \hat{\sigma}_{11}^{-1} y_2' P y_2$.

Finally, noting that $\hat{u} = y_1 - \hat{\gamma}_{2\text{SLS}} y_2$, $\hat{v} = M y_2$, we have $\hat{\sigma}_{22} = (1/N)\hat{v}'\hat{v} = (1/N) y_2' M y_2$, $y_2' y_1 = y_2' y_2 \hat{\gamma}_{2\text{SLS}} + N\hat{\sigma}_{12}$, where $\hat{\sigma}_{12} = (1/N)\hat{v}'\hat{u}$. Therefore,

$$\begin{aligned} \hat{\gamma}_{\text{FGLS}} &= Q[\hat{\sigma}_{22} y_2' y_2 \hat{\gamma}_{2\text{SLS}} + \hat{\sigma}_{22} N \hat{\sigma}_{12} - \hat{\sigma}_{12} N \hat{\sigma}_{22} - \hat{\sigma}_{12}^2 \hat{\sigma}_{11}^{-1} y_2' P y_1] \\ &= Q[\hat{\sigma}_{22} y_2' y_2 - \hat{\sigma}_{12}^2 \hat{\sigma}_{11}^{-1} y_2' P y_2] \hat{\gamma}_{2\text{SLS}} \\ &= \hat{\gamma}_{2\text{SLS}}. \end{aligned}$$

The conclusion holds for a general model with multiple included endogenous variables and predetermined variables in the structural equation; see Gao and Lahiri (2000).

REFERENCE

Gao, C. & K. Lahiri (2000) Further consequences of viewing LIML as an iterated Aitken estimator. *Journal of Econometrics* 98, 187–202.

00.2.2. *The Maximum Number of Omitted Variables*—Solution, proposed by Dmitri L. Danilov and Jan R. Magnus. If $r = 0$, the result is simple and well known. We assume that $r \geq 1$. Let $(S : T)$ be an orthogonal $k_2 \times k_2$ matrix such that

$$X_2' X_1 X_1' X_2 S = S\Lambda, \qquad X_1' X_2 T = 0,$$

where Λ denotes an $r \times r$ diagonal matrix with positive diagonal elements. Notice that the dimensions of S and T are $k_2 \times r$ and $k_2 \times (k_2 - r)$, respectively. Because $X_2' X_2$ has full rank k_2, we obtain

$$r(X_2 T) = r(X_2' X_2 T) = r(T) = k_2 - r,$$

so that the $n \times (k_2 - r)$ matrix $W_2 \equiv X_2 T$ has full column rank. Hence, we may define the idempotent matrix $M_2 = I_n - W_2(W_2' W_2)^{-1} W_2'$.

Now, let $W_1 \equiv M_2 X_2 S$, an $n \times r$ matrix. Because

$$W_1 = M_2 X_2 S = X_2 S - W_2 (W_2' W_2)^{-1} W_2' X_2 S,$$

we obtain $X_1' W_1 = X_1' X_2 S$ and, hence,

$$X_2' X_1 X_1' W_1 = X_2' X_1 X_1' X_2 S = S\Lambda,$$

so that $r = r(X_2' X_1 X_1' W_1) \leq r(W_1) \leq r$ and, hence, $r(W_1) = r$.

Next, let $W \equiv (W_1 : W_2)$. We already know that $r(W_1) = r$ and $r(W_2) = k_2 - r$. Because $M_2 W_2 = 0$, it follows that $W_1' W_2 = 0$ and, hence, that $r(W) = r(W_1) + r(W_2) = r + k_2 - r = k_2$.

Finally, we observe that

$$M_2 X_2 = X_2 - X_2 T (W_2' W_2)^{-1} W_2' X_2 = X_2 P$$

for some matrix P and, hence,

$$W = (W_1 : W_2) = (M_2 X_2 S : X_2 T) = (X_2 PS : X_2 T) = X_2 Q$$

for some $k_2 \times k_2$ matrix Q. Because $r(W) = k_2$, Q is non-singular.

It is now easy to see that W_2 is orthogonal to both W_1 and X_1. Also, the space spanned by the k_2 columns of W is identical to the space spanned by the k_2 columns of X_2, so that $X_2 \beta_2 = W\delta$ for some choice of δ (namely $\delta = Q^{-1}\beta_2$). Hence, the estimator $\hat{\beta}_1$ obtained from a regression of y on X_1 and X_2 will be identical to the estimator obtained from a regression of y on X_1 and W_1, and W_1 only has r columns.

When drawing inferences about β_1, we assume that $u \sim N(0, \sigma^2 I_n)$. The estimator of σ^2 will be biased upward if we delete W_2 from our regression, even though W_2 is orthogonal to both X_1 and W_1, just as in the standard textbook case.

00.2.3. *Effects of Transforming the Duration Variable in Accelerated Failure Time (AFT) Models—Solution*,[1] proposed by S.K. Sapra.

(a) (i) $y = kt$: The moment generating function of $\ln y$ is

$$M_{\ln y}(s) = M_{\ln(kt)}(s) = k^s E(t^s) = k^s M_{\ln t}(s). \tag{2}$$

Therefore, the cumulant generating function of $\ln y$ is

$$K_{\ln y}(s) = s \ln k + K_{\ln t}(s). \tag{3}$$

Equation (3) yields

$$K'_{\ln y}(s) = \ln k + K'_{\ln t}(s), \tag{4}$$

and

$$K''_{\ln y}(s) = K''_{\ln t}(s), \tag{5}$$

where the primes denote the derivatives. Therefore,

$$E(\ln y) = K'_{\ln y}(0) = \ln k + K'_{\ln t}(0) = \ln k + E(\ln t) = \ln k + \beta' x, \tag{6}$$

$$\text{Var}(\ln r) = K''_{\ln y}(0) = K''_{\ln t}(0) = \text{Var}(\ln t) = \sigma^2 \text{ (constant)}. \tag{7}$$

Hence, we may write

$$\ln y = \ln k + \beta' x + \varepsilon, \tag{8}$$

which is an AFT model because $\ln y$ and ε are homoskedastic, as seen in equation (7).

(ii) $y = t^k$: The derivation in part (i) can be easily modified for this case to show that

$$E(\ln y) = K'_{\ln y}(0) = k(K'_{\ln t}(0)) = kE(\ln t) = k(\beta' x), \tag{9}$$

$$\text{Var}(\ln y) = K''_{\ln y}(0) = k^2(K''_{\ln t}(0)) = k^2 \text{ Var}(\ln t) = k^2\sigma^2 \text{ (constant)}. \tag{10}$$

Hence, we may write $\ln y = k(\beta' x) + k\varepsilon$, which is an AFT model because $\ln y$ and $k\varepsilon$ are homoskedastic.

(b) (i) $y = a + bt$: From equation (1) in the problem, we have

$$y = a + bt = a + b \exp(\beta' x + \varepsilon). \tag{12}$$

Therefore, we may write

$$\ln y = \ln(a + b \exp(\beta' x + \varepsilon))$$
$$= \ln(b \exp(\beta' x + \varepsilon)\{1 + a/b \exp(-(\beta' x + 3))\}), \tag{13}$$

or

$$\ln y = \ln b + \beta' x + \varepsilon^*, \tag{14}$$

https://doi.org/10.1017/S0266466601172105 Published online by Cambridge University Press

where $\varepsilon^* = \varepsilon + \ln(1 + a/b\exp(-(\beta'x + \varepsilon)))$ is heteroskedastic because $\mathrm{Var}(\varepsilon^*)$ depends on $\beta'x$.

Hence, it follows that (14) is not an AFT model.

(ii) $y = \exp(a + bt)$: From (1) of the problem, we have

$$\ln y = a + \varepsilon^*, \tag{15}$$

where $\varepsilon^* = b\exp(\beta'x + \varepsilon)$ is heteroskedastic because $\mathrm{Var}(\varepsilon^*)$ depends on $\beta'x$. Hence, it follows that (15) is not an AFT model.

(c) Without loss of generality, assume that $E(\ln t_0) = 0$ (for $E(\ln t_0) = \alpha \neq 0$, simply replace $\exp(-\beta'x)$ with $\exp(\alpha - \beta'x)$ below). Then the hazard function of the density function of t is

$$\lambda_x(t) = \exp(-\beta'x)\lambda_0(t\exp(-\beta'x)), \tag{16}$$

where λ_0 is the baseline hazard (hazard function of the density function of t_0).

For the density function of $y = f(t)$, where f is a differentiable function of t, the hazard function of the density function of y is given by

$$\lambda_x(y) = (dy/dt)^{-1}\exp(-\beta'x)\lambda_0(\exp(-\beta'x)f^{-1}(y)). \tag{17}$$

Therefore, substituting for dy/dt and $f^{-1}(y)$ for each definition of y into equation (17), we have

1. $y = kt$: $\lambda_x(y) = k^{-1}\exp(-\beta'x)\lambda_0(yk^{-1}\exp(-\beta'x))$, **(18)**

2. $y = t^k$: $\lambda_x(y) = k^{-1}y^{(1-k)/k}\exp(-\beta'x)\lambda_0(y^{1/k}\exp(-\beta'x))$, **(19)**

3. $y = a + bt$: $\lambda_x(y) = b^{-1}\exp(-\beta'x)\lambda_0(b^{-1}(y - a)\exp(-\beta'x))$, **(20)**

4. $y = \exp(a + bt)$:

$$\lambda_x(y) = (by)^{-1}\exp(-\beta'x)\lambda_0(b^{-1}(\ln y - a)\exp(-\beta'x)). \tag{21}$$

NOTE

1. A solution has been proposed independently by Walter Distaso and Steve Lawford.

REFERENCE

Kalbfleisch, J.D. & R.L. Prentice (1980) *The Statistical Analysis of Failure Time Data.* New York: John Wiley & Sons.

00.2.4. *Conflict among Criteria for Testing Hypotheses: Examples from Non-Normal Distributions*—Solution,[1] proposed by N.K. Dastoor. By considering a scalar parameter case of the univariate exponential family of distributions, this solution provides a theorem and corollary that not only contain as special cases the results for the distributions specified in the problem but also provide other simple examples. Let X be a continuous (or discrete) random variable with support $S \subseteq \mathbb{R}$ and probability density function (or probability function) $f(x,\theta)$,

https://doi.org/10.1017/S0266466601172105 Published online by Cambridge University Press

where $\theta \in \Omega \subseteq \mathbb{R}$ and the parameter space Ω is an open interval. For $\theta \in \Omega$, let $f(x,\theta)$ denote the exponential family

$$f(x,\theta) = \exp\{a(\theta) + b(x) + c(\theta)d(x)\} \quad \text{for } x \in S, \tag{1}$$

where $a(\theta)$, $b(x)$, $c(\theta)$, and $d(x)$ are scalar functions, $a(\theta)$ and $c(\theta)$ are (nonconstant) twice continuously differentiable on Ω, the first derivative of $c(\theta)$, $c'(\theta)$, is finite and positive for all $\theta \in \Omega$, and $\theta \equiv E[d(X)]$. This definition of θ is convenient for showing the results stated in the problem and is the mean-value parametrization of the exponential family (see Lehmann and Casella, 1998, equation (5.17), p. 116; LC). If $d(x) = x$, then (1) reduces to the scalar parameter case of the linear exponential family as defined by Gourieroux, Monfort, and Trognon (1984, Definition 1, p. 683; GMT). The usual regularity conditions are assumed to hold so that, in particular, S does not depend on θ. By noting that $\theta = E[d(X)]$, if X is a continuous (or discrete) random variable, then differentiating (with respect to θ) the identity $\int_{x\in S} f(x,\theta)\,dx = 1$ (or $\sum_{x\in S} f(x,\theta) = 1$) yields

$$a'(\theta) = -\theta c'(\theta); \tag{2}$$

see GMT (Property 1, p. 683) and LC (Problem 5.6(a), p. 66).

Given a random sample $x_1, x_2, \ldots, x_n$, the log-likelihood function is

$$L(\theta) = n\{a(\theta) + \bar{b} + \bar{d}c(\theta)\},$$

where $\bar{b} = (1/n)\sum_{t=1}^{n} b(x_t)$ and $\bar{d} = (1/n)\sum_{t=1}^{n} d(x_t)$. Using (2), the score function $s(\theta) \equiv L'(\theta) = n\{a'(\theta) + \bar{d}c'(\theta)\}$ can be written as

$$s(\theta) = n(\bar{d} - \theta)c'(\theta). \tag{3}$$

Because $s'(\theta) = n\{(\bar{d} - \theta)c''(\theta) - c'(\theta)\}$ and $E[d(X)] = \theta$, the information matrix $I(\theta) \equiv -E[s'(\theta)]$ simplifies to

$$I(\theta) = nc'(\theta). \tag{4}$$

For a given sample, it is now assumed that $\bar{d} \in \Omega$, which ensures the existence of the unrestricted maximum likelihood estimate (MLE) of θ; i.e., because Ω is an open interval and $c'(\theta)$ is finite and positive for all $\theta \in \Omega$, (3) shows that, if $\bar{d} \in \Omega$, then $s(\theta) \gtreqless 0$ for $\theta \lesseqgtr \bar{d}$, so the MLE of θ is $\bar{d}$ as $L(\theta)$ has a global maximum at $\theta = \bar{d}$.

To test the null hypothesis H_0: $\theta = \theta_0$ against the alternative H_1: $\theta \neq \theta_0$, where $\theta_0 \in \Omega$, the Wald and Lagrange multiplier statistics are given by Buse (1982, equations (3) and (6), pp. 154–155) as $W = (\hat{\theta} - \theta_0)^2 I(\hat{\theta})$ and $LM = s(\theta_0)^2/I(\theta_0)$, respectively, where $\hat{\theta}$ is the unrestricted MLE of θ. Using (3), (4), and $\hat{\theta} = \bar{d}$, these test statistics can then be written as

$$W = n(\bar{d} - \theta_0)^2 c'(\bar{d}) \quad \text{and} \quad LM = n(\bar{d} - \theta_0)^2 c'(\theta_0), \tag{5}$$

which show that the relationship between them depends on the behavior of $c'(\theta)$. Now let $g(\theta) = 1/c'(\theta)$; then, by differentiating $\theta = E[d(X)]$ with respect to θ and using (2), it is easily seen that $g(\theta) = V[d(X)]$; see GMT (Property 3, p. 684) and LC (equation (5.18), p. 116).

THEOREM. *Given the framework above,*

$$W \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} LM \Leftrightarrow g(\bar{d}) \begin{Bmatrix} > \\ = \\ < \end{Bmatrix} g(\theta_0).$$

Proof. The test statistics in (5) with $c'(\theta) = 1/g(\theta)$ yield

$$W - LM = \frac{n(\bar{d} - \theta_0)^2 [g(\theta_0) - g(\bar{d})]}{g(\bar{d}) g(\theta_0)}$$

from which the result follows as the denominator is finite and positive. ■

COROLLARY. *Given the framework above, if d(x) = x and* $g(\theta)$ *is a strictly increasing function on* Ω, *then*

$$W \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} LM \Leftrightarrow \bar{x} \begin{Bmatrix} > \\ = \\ < \end{Bmatrix} \theta_0,$$

where $\bar{x} = (1/n)\sum_{t=1}^{n} x_t$.

Proof. If $d(x) = x$, then $\bar{d} = \bar{x}$ and the result follows from the theorem and the definition of a strictly increasing function. ■

Some members of (1) are considered below, where $\mathbb{R}^+ \equiv \{y | y > 0\}$; parts (a)–(c) contain the distributions specified in the problem. For each distribution in parts (a)–(d) below, the forms of Ω, $f(x,\theta)$, and $c(\theta)$ are conveniently provided in Table I of GMT (p. 685).

(a) Let X be a Poisson variate with $\Omega = \mathbb{R}^+$, $S = \{0\} \cup \mathbb{N}$, and $f(x,\theta) = (1/x!)e^{-\theta}\theta^x$. Here, $d(x) = x$, $c(\theta) = \ln\theta$ and $g(\theta) = \theta$. Because $g(\theta)$ is a strictly increasing function, the corollary provides the relationship between W and LM. In particular, the corollary with $\theta_0 = 3$ shows that $W \geq LM$ iff $0 < \bar{x} \leq 3$ and $W \leq LM$ iff $\bar{x} \geq 3$, the result required in part (a) of the problem. In this example, if every observation in a sample is zero, then $L(\theta) = -n\theta$; therefore, (as $0 \notin \Omega$) the unrestricted MLE of θ does not exist. Hence, it is customary to assume that at least one observation is positive, which is equivalent to the assumption that $\bar{d} \in \Omega$.

(b) Let X be a gamma variate with mean θ, known $\alpha > 0$, $\Omega = S = \mathbb{R}^+$, and $f(x,\theta) = (1/(\Gamma(\alpha)))(\alpha/\theta)^\alpha x^{\alpha-1} e^{-\alpha x/\theta}$. Here, $d(x) = x$, $c(\theta) = -(\alpha/\theta)$ and $g(\theta) = \theta^2/\alpha$. For any known $\alpha > 0$, $g(\theta)$ is a strictly increasing function on Ω, so the relationship between W and LM is given by the corollary. In this example, if $\alpha = 1$, then

X is an exponential variate with mean θ, so the corollary with $\theta_0 = 3$ shows that $W \geq LM$ iff $0 < \bar{x} \leq 3$ and $W \leq LM$ iff $\bar{x} \geq 3$, the result required in part (b) of the problem.

(c) Let X be a binomial variate with mean θ, known $N \in \mathbb{N}$, $\Omega = (0, N)$, $S = \{0,1,2,\dots,N\}$, and

$$f(x,\theta) = \binom{N}{x}\left(\frac{\theta}{N}\right)^x \left(1 - \frac{\theta}{N}\right)^{N-x}.$$

Here, $d(x) = x$, $c(\theta) = \ln(\theta/(N-\theta))$ and $g(\theta) = (\theta(N-\theta))/N$. For any known $N \in \mathbb{N}$, $g(\theta)$ is an even function around $\theta = N/2$ (with a global maximum of $N/4$ at $\theta = N/2$), so now let $\theta_0 \in [N/2, N)$. Then it is easily seen that $g(\bar{x}) > g(\theta_0)$ iff $\bar{x} \in (N - \theta_0, \theta_0)$, $g(\bar{x}) = g(\theta_0)$ iff $\bar{x} = N - \theta_0$, θ_0, and $g(\bar{x}) < g(\theta_0)$ iff $\bar{x} \in (0, N-\theta_0) \cup (\theta_0, N)$; thus, the theorem with $\bar{d} = \bar{x}$ shows that, for $\theta_0 \in [N/2, N)$,

$$W \begin{Bmatrix} < \\ = \\ > \end{Bmatrix} LM \Leftrightarrow \bar{x} \in \begin{Bmatrix} (N-\theta_0, \theta_0) \\ \{N-\theta_0, \theta_0\} \\ (0, N-\theta_0) \cup (\theta_0, N) \end{Bmatrix}. \tag{6}$$

In (6), if $\theta_0 = N/2$, then the first partition of Ω, $(N-\theta_0, \theta_0)$, is the empty set, thus

$$\theta_0 = \frac{N}{2} \Rightarrow W \geq LM \quad \text{for all } \bar{x} \in \Omega = (0, N), \tag{7}$$

with the equality holding when $\bar{x} = N/2$. In this example, if $N = 1$, then X is a Bernoulli variate with mean θ and $\Omega = (0,1)$. Setting $N = 1$ in (7) shows that, if $\theta_0 = \frac{1}{2}$, then $W \geq LM$, the first result required in part (c) of the problem. Setting $N = 1$ and $\theta_0 = \frac{2}{3}$ in (6) gives $W \leq LM$ iff $\frac{1}{3} \leq \bar{x} \leq \frac{2}{3}$ and $W \geq LM$ iff $0 < \bar{x} \leq \frac{1}{3}$ or $\frac{2}{3} \leq \bar{x} < 1$, the second result required in part (c) of the problem. In this case, where X is a Bernoulli variate, if the parameter space is extended to the closed interval $[0,1]$, then $\bar{x}$ is still the unrestricted MLE of θ; i.e., if $\bar{x} = 0$ ($\bar{x} = 1$), then the likelihood function is strictly decreasing (increasing) on the closed interval, so $\hat{\theta} = 0$ ($\hat{\theta} = 1$). However, as $W = (n(\bar{x} - \theta_0)^2)/(\bar{x}(1-\bar{x}))$ here, the Wald statistic is not finite if $\bar{x} = 0,1$. Therefore, the solution presented here excludes $\bar{x} = 0,1$, whereas the problem (as stated) only excludes $\bar{x} = 0$.

(d) Let X be a negative binomial variate with mean θ, known $r > 0$, $\Omega = \mathbb{R}^+$, $S = \{0\} \cup \mathbb{N}$, and $f(x,\theta) = ((\Gamma(r+x))/(\Gamma(r)\Gamma(x+1)))(r/(r+\theta))^r(\theta/(r+\theta))^x$. Here, $d(x) = x$, $c(\theta) = \ln(\theta/(r+\theta))$ and $g(\theta) = (\theta(r+\theta))/r$. For any known $r > 0$, $g(\theta)$ is a strictly increasing function on Ω, so the relationship between W and LM is given by the corollary. Note that, by setting $r = 1$ in this example, the corollary provides the relationship between W and LM when X has a geometric distribution with mean θ.

(e) If X is a normal variate with mean θ and known variance $\sigma^2 > 0$, then $W = LM$, as shown by Buse (1982, pp. 155–156). Now let X be a normal variate with variance θ, known mean $\mu \in \mathbb{R}$, $\Omega = \mathbb{R}^+$, $S = \mathbb{R}$, and $f(x,\theta) = (1/\sqrt{2\pi\theta})\exp\{-(1/2\theta)(x-\mu)^2\}$. In this case, $d(x) = (x-\mu)^2$, $c(\theta) = -(1/2\theta)$, and $g(\theta) = 2\theta^2$. Using the fact that $g(\theta)$ is a strictly increasing function on Ω, the theorem shows that $W \lesseqqgtr LM$ iff $\bar{d} \lesseqqgtr \theta_0$. This example shows that a non-normal distribution is not necessary for the well-known inequality ($W \geq LM$) to be non-robust.

The definition of θ employed above was motivated by the need to show the results stated in the problem. However, it is important to note that the Wald statistic depends on the definition of θ and on the specification of the null hypothesis (see Dagenais and Dufour, 1991, p. 1607). For example, the probability density function of an exponential variate is often written in the form $f(x,\theta) = \theta e^{-\theta x}$ with $\Omega = S = \mathbb{R}^+$, in which case, $E[X] = 1/\theta$. By using this version of the probability density function to test the null hypothesis specified as H_0: $\theta = \theta_0$, it can be shown that $W = LM = n(1 - \bar{x}\theta_0)^2$, a result different than that stated in part (b) of the problem.

NOTE

1. Five excellent solutions have been proposed independently by (in alphabetical order): Badi H. Baltagi (the poser of the problem), Walter Distaso and Steve Lawford, Francisco J. Goerlich, Yulia Kotlyarova, and Diego Lubian. Francisco J. Goerlich noted that example (c) was solved by Godfrey (1988, Sect. 2.6, pp. 59–60).

REFERENCES

Buse, A. (1982) The likelihood ratio, Wald, and Lagrange multiplier tests: An expository note. *American Statistician* 36, 153–157.

Dagenais, M.G. & J.-M. Dufour (1991) Invariance, nonlinear models, and asymptotic tests. *Econometrica* 59, 1601–1615.

Godfrey, L.G. (1988) *Misspecification Tests in Econometrics.* Cambridge: Cambridge University Press.

Gourieroux, C., A. Monfort, & A. Trognon (1984) Pseudo maximum likelihood methods: Theory. *Econometrica* 52, 681–700.

Lehmann, E.L. & G. Casella (1998) *Theory of Point Estimation*, 2nd ed. New York: Springer-Verlag.

https://doi.org/10.1017/S0266466601172105 Published online by Cambridge University Press