Econometric Theory, **15**, 1999, 427–432. Printed in the United States of America.

PROBLEMS AND SOLUTIONS

PROBLEMS

99.3.1. *The Eigenvalue Decomposition of a Symmetric Matrix*, proposed by R.W. Farebrother. Let X and Y be $n \times k$ matrices satisfying $X'X = I_k$, $Y'Y = I_k$, and $X'Y = 0$ and let D be a $k \times k$ diagonal matrix with positive diagonal elements; then $S = XDY' - YDX'$ is an $n \times n$ skew symmetric matrix. Determine the eigenvalue decomposition of the associated $2n \times 2n$ symmetric matrix

$$\begin{bmatrix} 0 & -S \\ S & 0 \end{bmatrix}.$$

99.3.2. *Maximum-Likelihood Estimation under Singularity*, proposed by Heinz Neudecker. The regression equation $y = a + \Gamma x + \varepsilon$ with $E(\varepsilon) = 0$ and $D(\varepsilon) = \Omega$ and further $l_m'a = l_m'\varepsilon = 0$ under normality yields the Ω-concentrated loglikelihood of the sample $Y' := (y_1 \ldots y_n)$

$$\log L_c = -\tfrac{1}{2}n \log|A| + \text{constant},$$

where

$$A := \frac{1}{n} E'E + \frac{1}{m} l_m l_m', \qquad E' := (\varepsilon_1 \ldots \varepsilon_n).$$

The matrix A is nonsingular.

Find first-order maximum conditions with respect to a.

99.3.3. *Non-Normality and Exploratory Data Analysis*, proposed by David M. Levy. Two of John Tukey's memorable contributions to statistical practice have been to teach statisticians the importance of non-normal error distributions (Tukey, 1960; Andrews et al., 1972) and to legitimize "exploratory data analysis" (EDA; Tukey, 1977). EDA remains an avowedly vague concept; indeed Mosteller and Tukey (1977) offer a subtle defense of vagueness. The question addressed is whether there is an analytical connection between these two concerns of Tukey. The statistical context is that of linear regression.

The problem has two parts. First, provide a formal definition of EDA in the linear regression context. Second, prove that in general when EDA is possible, the errors will not have a normal distribution.

REFERENCES

Andrews, D.F., P.J. Bickel, F.R. Hampel, P.J. Huber, W.H. Rogers, & J.W. Tukey. (1972) *Robust Estimates of Location.* Princeton: Princeton University Press.

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

https://doi.org/10.1017/S0266466699153088 Published online by Cambridge University Press

Mosteller, F. & J.W. Tukey (1977) *Data Analysis and Regression.* Reading, Massachusetts: Addison-Wesley.
Tukey, J.W. (1960) A survey of sampling from contaminated distributions. In I. Olkin, S. Ghwiye, W. Hoeffding, G. Madow, & H. Mann (eds.), *Contributions to Probabilities and Statistics.* Stanford: Stanford University Press.
Tukey, J.W. (1977) *Exploratory Data Analysis.* Reading, Massachusetts: Addison-Wesley.

99.3.4. *The Overdispersion Test in Count Data as a Gauss–Newton Regression*, proposed by Badi H. Baltagi and Dong Li. Consider the Poisson regression model

$$\Pr(y_i|\mu_i) = \frac{e^{-\mu_i}\mu_i^{y_i}}{y_i!}, \qquad y_i = 0,1,2,\ldots$$

where the parameter of the Poisson μ_i is related to the regressors X_i as follows:

$$\mu_i = \exp(X_i\beta)$$

where β is $k\times 1$. It is well known that the Poisson distribution has the property that $E(y_i|X_i) = \mathrm{Var}(y_i|X_i) = \mu_i$. Criticisms of the Poisson regression model lead to the negative binomial regression model (see Cameron and Trivedi, 1986), where the mean and variance are not necessarily equal to each other. A popular specification is the Negbin II model, which satisfies $\mathrm{Var}(y_i|X_i) = E(y_i|X_i)(1+\alpha E(y_i|X_i))$. This implies that the variance-mean ratio is linear in the mean. For this model, a test for overdispersion is a test for $H_0: \alpha = 0$. Show that a test for overdispersion can be derived as a Gauss–Newton regression for heteroskedasticity. How does it differ from the optimal regression based test in Cameron and Trivedi (1990)?

REFERENCES

Cameron, A.C. & P.K. Trivedi (1986) Econometric models based on count data: Comparisons and applications of some estimates and tests. *Journal of Applied Econometrics* 1, 29–53.
Cameron, A.C. & P.K. Trivedi (1990) Regression-based tests for overdispersion in the Poisson model. *Journal of Econometrics* 46, 347–364.

SOLUTIONS

98.3.1. *Model Identification Based on Conditional Moments*—Solution, proposed by Xing Ming. Solving (2) and (3) in the problem for the reduced form and expressing them in the form of conditional moments, we obtain

$$\mathsf{E}(y_1|x) = \mathsf{E}[\exp(e_1)]\cdot\exp(-a_0)\cdot\exp(-a_3x),$$

$$\mathsf{E}(y_2|x) = -b_0 - b_1\mathsf{E}[\exp(e_1)]\cdot\exp(-a_0)\cdot\exp(-a_3x) - b_3x.$$

https://doi.org/10.1017/S0266466699153088 Published online by Cambridge University Press

Assume there exists another set of parameters $\tilde{a}_0$, $\tilde{a}_3$, $\tilde{b}_0$, $\tilde{b}_1$, and $\tilde{b}_3$ that gives the same conditional moment. Then

$$\mathsf{E}[\exp(e_1)]\cdot\exp(-a_0 - a_3 x) = \mathsf{E}[\exp(e_1)]\cdot\exp(-\tilde{a}_0 - \tilde{a}_3 x) \quad \textbf{(4)}$$

$$b_0 + b_1\mathsf{E}[\exp(e_1 - a_0)]\cdot\exp(-a_3 x) + b_3 x = \tilde{b}_0 + \tilde{b}_1\mathsf{E}[\exp(e_1 - \tilde{a}_0)]\cdot\exp(-\tilde{a}_3 x) + \tilde{b}_3 x \quad \textbf{(5)}$$

1. From (4),

   $$\tilde{a}_0 - a_0 = (a_3 - \tilde{a}_3)x.$$

   If $a_3 \neq \tilde{a}_3$, the left hand side is a constant and the right hand side is a random variable, which is impossible. Thus we must have

   $$\tilde{a}_0 = a_0 \quad \text{and} \quad \tilde{a}_3 = a_3,$$

   i.e., equation (2) is identified.
2. When $a_3 \neq 0$, as a result of the fact that b_0, $\exp(-a_3 x)$, and x are linearly independent, from equation (5) and $\tilde{a}_0 = a_0$, we have

   $$b_0 = \tilde{b}_0, \qquad b_1 = \tilde{b}_1, \qquad b_3 = \tilde{b}_3,$$

   i.e., equation (3) is identified.
3. When $a_3 = 0$, equation (5) can only give

   $$b_0 + b_1\mathsf{E}[\exp(e_1)]\exp(a_0) + b_3 x = \tilde{b}_0 + \tilde{b}_1\mathsf{E}[\exp(e_1)]\exp(a_0) + \tilde{b}_3 x.$$

   From the preceding equation, we can obtain $b_3 = \tilde{b}_3$, i.e., b_3 is identified. But b_0, b_1 and $\tilde{b}_0, \tilde{b}_1$ are interchangeable, i.e., b_0, b_1 are not identified.

98.3.2. *One–One Transformation of Multidimensional General Distribution to Multidimensional Uniform Distribution*—Solution, proposed by Xing Ming. The following transformation is a one–one transformation and $(Y_1, Y_2, \ldots, Y_k)$ is uniformly distributed on $[0,1]^k$.

$$Y_1 = \int_{-\infty}^{X_1} f(x_1)\, dx_1 \quad \textbf{(1)}$$

$$Y_2 = \int_{-\infty}^{X_2} f(x_2|X_1)\, dx_2 \quad \textbf{(2)}$$

$$Y_3 = \int_{-\infty}^{X_3} f(x_3|X_1, X_2)\, dx_3 \quad \textbf{(3)}$$

$$Y_k = \int_{-\infty}^{X_k} f(x_k|X_1, X_2, \ldots, X_{k-1})\, dx_k \quad \textbf{(4)}$$

https://doi.org/10.1017/S0266466699153088 Published online by Cambridge University Press

where $f(x_1)$ is the marginal density function of X_1; $f(x_2|X_1)$ is the density function of X_2 conditional on X_1; $f(x_3|X_1,X_2)$ is the density function X_3 conditional on X_1 and X_2; $f(x_k|X_1,X_2,\ldots,X_{k-1})$ is the density function X_k conditional on $X_1,X_2,\ldots,X_{k-1}$.

Then the joint density function of $(Y_1,Y_2,\ldots,Y_k)$ is

$$
\begin{aligned}
&p(y_1,y_2,\ldots,y_k)\\
&\quad= \frac{f(x_1,x_2,\ldots,x_k)}{\left|\dfrac{\partial(y_1,y_2,\ldots,y_k)}{\partial(x_1,x_2,\ldots,x_k)}\right|}\\
&\quad= \frac{f(x_1,x_2,\ldots,x_k)}{\begin{vmatrix} f(x_1) & 0 & 0 & 0 & 0\\ * & f(x_2|X_1) & 0 & 0 & 0\\ * & * & f(x_3|X_1,X_2) & 0 & 0\\ \cdots & \cdots & \cdots & \ddots & \cdots\\ * & * & * & * & f(x_k|X_1,X_2,\ldots,X_{k-1}) \end{vmatrix}}\\
&\quad= f(x_1,x_2,\ldots,x_k)/f(x_1,x_2,\ldots,x_k)=1.
\end{aligned}
$$

Obviously $(Y_1,Y_2,\ldots,Y_k)$ is valued on $[0,1]^k$ and has a joint density function $p(y_1,y_2,\ldots,y_k)=1$. Thus $(Y_1,Y_2,\ldots,Y_k)$ is a uniformly distributed random vector on $[0,1]^k$.

One can see from transformations (1)–(4) that Y_1 is a strictly monotonic increasing function of X_1, Y_2 a strictly monotonic increasing function of X_2 given X_1, and Y_k a strictly monotone increasing function of X_k given $X_1,X_2,\ldots,X_{k-1}$. Hence one can first solve for X_1 in terms of Y_1 from (1). Then solve for X_2 in terms of Y_1 and Y_2 from (2). Finally solve for Y_k in terms of $Y_1,Y_2,\ldots,Y_k$ from (4). It is easily seen that the inverse mapping is also unique.

98.3.3. *Prediction in the Equicorrelated Regression Model*—Solution,[1] proposed by Seuck Heun Song. In matrix form the variance-covariance of $u=(u_1,\ldots,u_T)'$ is written as

$$E(uu')=\Omega=\sigma^2[(1-\rho)I_T+\rho J_J], \tag{1}$$

where I_T is the $T\times T$ identity matrix, $J_T=i_T i_T'$, and $i_T=(1,\ldots,1)'$ is the $T\times 1$ vector of ones. This can be rearranged as

$$\Omega=\sigma^2[(T\rho+1-\rho)\bar{J}_T+(1-\rho)E_T], \tag{2}$$

where $\bar{J}_T=J_T/T$ and $E_T=I_T-\bar{J}_T$. This is also the spectral decomposition of Ω. This means that

$$\Omega^{-1}=\frac{1}{\sigma^2}\left[\frac{1}{T\rho+1-\rho}\,\bar{J}_T+\frac{1}{1-\rho}\,E_T\right]. \tag{3}$$

https://doi.org/10.1017/S0266466699153088 Published online by Cambridge University Press

Solution a. It is well known (Kruskal, 1968, p. 71) that ordinary least squares (OLS) is equivalent to generalized least squares (GLS) if and only if $\Omega X \in C(X)$, where $C(X)$ is the column space of X.

Let x^j be the jth column of X. In the equicorrelated regression model, because x^j and $i_T \in C(X)$,

$$\Omega x^j = \sigma^2 \left[(1-\rho)x^j + \rho \sum_{i=1}^{T} x_{ij} i_T \right] \in C(X), \quad \text{for all } j = 1, \ldots, k. \tag{4}$$

Therefore, the result follows from Kruskal.

Solution b. Let $\hat{u}$ be the vector of the OLS residuals. Then $\hat{u} = Mu$, where $M = I - X(X'X)^{-1}X'$. We have

$$\begin{aligned} E(\hat{u}\hat{u}') &= E(uMu') = \operatorname{tr}\Omega - \operatorname{tr}[X(X'X)^{-1}X'\Omega] \\ &= \sigma^2 T - \sigma^2 \operatorname{tr}\{X(X'X)^{-1}X'[(1-\rho)I_T + \rho J_T]\} \\ &= \sigma^2 T - \sigma^2[(1-\rho)\operatorname{tr}(X(X'X)^{-1}X') + \rho\, tr(X(X'X)^{-1}X'J_T)] \\ &= \sigma^2[(T-k)(1-\rho)], \quad \text{because } X(X'X)^{-1}X'i_T = i_T. \end{aligned}$$

Therefore,

$$E(s^2) = E\left(\frac{\hat{u}\hat{u}'}{T-k}\right) = \sigma^2(1-\rho). \tag{5}$$

From equation (5), if $0 \le \rho \le 1$, we have $0 \le E(s^2) \le \sigma^2$. Also, if $-1/(T-1) \le \rho \le 0$, we have $0 \le E(s^2) \le T/(T-1)$. Finally, we obtain that

$$0 \le E(s^2) \le T/(T-1), \quad \text{for any value of } \sigma^2. \tag{6}$$

Solution c. Goldberger (1962) showed that, for any Ω, the best linear unbiased predictor (BLUP) for y_{T+s} is given by

$$\hat{y}_{T+s} = x'_{T+s}\hat{\beta}_{GLS} + w'\Omega^{-1}\hat{u}_{GLS}, \tag{7}$$

where $w' = E(u_{T+s}u')$ and $\hat{u}_{GLS} = y - X\hat{\beta}_{GLS}$. We need to show that the last term of (7) is zero. For this model, the last term of (7) is given by

$$\begin{aligned} w'\Omega^{-1}\hat{u}_{GLS} &= E[u_{T+s}u']\Omega^{-1}\hat{u}_{GLS} \\ &= [\sigma^2\rho i'_T]\frac{1}{\sigma^2}\left(\frac{1}{T\rho + 1 - \rho}\bar{J}_T + \frac{1}{1-\rho}E_T\right)\hat{u}_{GLS} \\ &= \frac{\rho}{T\rho + 1 - \rho} i'_T\hat{u}_{GLS} \\ &= \frac{\rho}{T\rho + 1 - \rho}\sum_{i=1}^{T}\hat{u}_i = 0, \end{aligned}$$

https://doi.org/10.1017/S0266466699153088 Published online by Cambridge University Press

because $i_T' E_T = 0$ and $i_T' \bar{J}_T = i_T'$. Now by (a), $\hat{\beta}_{GLS} = \hat{\beta}_{OLS}$ and it is verified that $x_{T+s}' \hat{\beta}_{OLS}$ is the BLUP for y_{T+s}.

NOTE

1. Excellent solutions have been proposed independently by J. Li, by S. Puntanen, G.P.H. Styan, and H.J. Werner, by G. Trenkler, and by B.H. Baltagi, the poser of the problem.

REFERENCES

Goldberger, A.S. (1962) Best linear unbiased prediction in the generalized linear regression model. *Journal of the American Statistical Association* 57, 369–375.

Kruskal, W. (1968) When are Gauss-Markov and least squares estimator identical? A coordinate-free approach. *Annals of Mathematical Statistics* 39, 70–75.

https://doi.org/10.1017/S0266466699153088 Published online by Cambridge University Press