Econometric Theory, **16**, 2000, 287–299. Printed in the United States of America.

PROBLEMS AND SOLUTIONS

PROBLEMS

00.2.1. *Degeneration of Feasible GLS to 2SLS in a Limited Information Simultaneous Equations Model*, proposed by Chuanming Gao and Kajal Lahiri. Consider a simple limited information simultaneous equations model,

$$y_1 = \gamma y_2 + u, \tag{1}$$

$$y_2 = X\beta + v, \tag{2}$$

where y_1, y_2 are $N \times 1$ vectors of observations on two endogenous variables. X is $N \times K$ matrix of predetermined variables of the system, and $K \geq 1$ such that (1) is identified. Each row of (u, v) is assumed to be i.i.d. $(0, \Sigma)$, and Σ is p.d.s.

Following Lahiri and Schmidt (1978), feasible GLS for (1) and (2) based on a consistent estimate of Σ yields a consistent estimate for γ. Pagan (1979) showed that an iterated Aitken estimator will generate LIML estimate of γ.

Denote $\hat{\gamma}_{2SLS} = (y_2' P y_2)^{-1} y_2' P y_1$, where $P = X(X'X)^{-1}X'$. The residuals $\hat{u} = y - \hat{\gamma}_{2SLS} y_2$ and $\hat{v} = My_2$, where $M = I_N - P$, may be used to generate a consistent estimate for Σ, e.g.,

$$\hat{\Sigma} = \frac{1}{N}\begin{bmatrix} \hat{u}'\hat{u} & \hat{u}'\hat{v} \\ \hat{v}'\hat{u} & \hat{v}'\hat{v} \end{bmatrix}.$$

Show that a feasible GLS estimate of γ using $\hat{\Sigma}$ (i.e., the first iterate of iterated Aitken) degenerates to $\hat{\gamma}_{2SLS}$.

REFERENCES

Lahiri, K. & P. Schmidt (1978) On the estimation of triangular structural systems. *Econometrica* 46, 1217–1222.

Pagan, A. (1979) Some consequences of viewing LIML as an iterated Aitken estimator. *Economics Letters* 3, 369–372.

00.2.2. *The Maximum Number of Omitted Variables*, proposed by Dmitri L. Danilov and Jan R. Magnus. Consider the standard partitioned regression model $y = X_1\beta_1 + X_2\beta_2 + u$, where $X \equiv (X_1 : X_2)$ is a nonstochastic $n \times k$ matrix with full column rank $k = k_1 + k_2$. We are interested in estimating β_1 and consider β_2 as a nuisance parameter. Let $r = \text{rank}(X_1' X_2)$. Show that we may assume, without loss of generality, that $k_2 = r$ and, hence, in particular that $k_2 \leq k_1$. Can we still make this simplifying assumption when drawing inferences about β_1?

In the special case where $r = 0$ and where consequently X_2 is orthogonal to X_1, we may delete X_2 altogether, a well-known result.

 0266-4666/00 $9.50

In another special case where $k_1 = 1$ (one "focus" parameter and the rest nuisance parameters), it is sufficient to consider just *one* nuisance parameter.

00.2.3. *Effects of Transforming the Duration Variable in Accelerated Failure Time (AFT) Models*, proposed by S.K. Sapra. Consider the following AFT model,

$$\ln t = \beta' x + \varepsilon, \tag{1}$$

where x is a $(m \times 1)$ vector of known constants, β is a $(m \times 1)$ vector of unknown parameters, $\varepsilon = \ln t_0 - E(\ln t_0)$, and t_0 is a random variable with a density function not involving x or β.

(a) Show that the following transformations of t lead to AFT models: (i) $y = kt, k > 0$, and (ii) $y = t^k$, where k is a constant.
(b) Show that the following transformations of t do not lead to AFT models: (i) $y = a + bt$, $a > 0, b > 0$, and (ii) $y = \exp(a + bt)$, where a and b are constants.
(c) Derive the hazard functions for the density functions of y in parts (a) and (b) by using the transformations of t defined therein.

00.2.4. *Conflict Among Criteria for Testing Hypotheses: Examples from Non-Normal Distributions*, proposed by Badi H. Baltagi. Berndt and Savin (1977) showed that $W \geq LR \geqq LM$ for the case of a multivariate regression model with normal disturbances. Ullah and Zinde-Walsh (1984) showed that this inequality is not robust to non-normality of the disturbances. In the spirit of the latter article, this problem considers simple examples from non-normal distributions and illustrates how this conflict among criteria is affected.

(a) Consider a random sample $x_1, x_2, \ldots, x_n$ from a Poisson distribution with parameter λ. Show that for testing $\lambda = 3$ versus $\lambda \neq 3$ yields $W \geq LM$ for $\bar{x} \leq 3$ and $W \leq LM$ for $\bar{x} \geq 3$.
(b) Consider a random sample $x_1, x_2, \ldots, x_n$ from an exponential distribution with parameter θ. Show that for testing $\theta = 3$ versus $\theta \neq 3$ yields $W \geq LM$ for $0 < \bar{x} \leq 3$ and $W \leq LM$ for $\bar{x} \geq 3$.
(c) Consider a random sample $x_1, x_2, \ldots, x_n$ from a Bernoulli distribution with parameter θ. Show that for testing $\theta = 0.5$ versus $\theta \neq 0.5$, we will always get $W \geq$ LM. Show also, that for testing $\theta = \frac{2}{3}$ versus $\theta \neq \frac{2}{3}$ we get $W \leq LM$ for $\frac{1}{3} \leq \bar{x} \leq \frac{2}{3}$ and $W \geq LM$ for $\frac{2}{3} \leq \bar{x} \leq 1$ or $0 < \bar{x} \leq \frac{1}{3}$.

REFERENCES

Berndt, E.R. & N.E. Savin (1977) Conflict among criteria for testing hypotheses in the multivariate linear regression model. *Econometrica* 45, 1263–1278.

Ullah, A. & V. Zinde-Walsh (1984) On the robustness of LM, LR and W tests in regression models. *Econometrica* 52, 1055–1065.

https://doi.org/10.1017/S0266466600002103 Published online by Cambridge University Press

SOLUTIONS

99.3.1 *The Eigenvalue Decomposition of a Symmetric Matrix*—Solution.[1] The following solutions have been proposed independently by Simo Puntanen, George P.H. Styan and Hans Joachim Werner, and by Geert Dhaene. These solutions are based on different types of interesting arguments.

Solution 1, proposed by Simo Puntanen, George P.H. Styan and Hans Joachim Werner. We give two different eigenvalue decompositions of the $2n \times 2n$ real symmetric matrix

$$\begin{pmatrix} 0 & -S \\ S & 0 \end{pmatrix} =: M,$$

say. Precisely, we prove that M (1) is unitarily similar and (2) also orthogonally similar (because M is real symmetric) to the $2n \times 2n$ diagonal matrix

$$\Lambda := \begin{pmatrix} -D & 0 & 0 & 0 & 0 & 0 \\ 0 & D & 0 & 0 & 0 & 0 \\ 0 & 0 & D & 0 & 0 & 0 \\ 0 & 0 & 0 & -D & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

of its eigenvalues. Corresponding eigenvectors may be chosen as the column vectors of (1) the complex unitary matrix

$$U := \frac{1}{2}\begin{pmatrix} X+iY & X+iY & X-iY & X-iY & N & N \\ Y-iX & -Y+iX & -Y-iX & Y+iX & -iN & iN \end{pmatrix}$$

and then

$$M = U\Lambda U^*,$$

where U^* is the conjugate transpose of U, or as the column vectors of (2) the real orthogonal matrix

$$P := \frac{1}{\sqrt{2}}\begin{pmatrix} X & X & Y & -Y & N & 0 \\ Y & -Y & X & X & 0 & N \end{pmatrix}.$$

and then

$$M = P\Lambda P'.$$

https://doi.org/10.1017/S0266466600002103 Published online by Cambridge University Press

In U and P the real $n\times(n-2k)$ matrix N is any matrix so that the columns of $N/\sqrt{2}$ constitute an orthonormal basis of the null space of the transpose of the $n\times 2k$ block-partitioned matrix $(X\quad Y)=Z$, say. Because $X'X=I_k$, $Y'Y=I_k$, and $X'Y=0$, we have $Z'Z=I_{2k}$ and hence $n\geq 2k$ and $\text{rank}(Z)=\text{rank}(X)+\text{rank}(Y)=2k$.

Proof 1. Clearly, by checking $MU=U\Lambda$, $U^*U=I_{2n}$, $MP=P\Lambda$, and $P'P=I_{2n}$, our two claims above are established.

Proof 2. This is a constructive proof based on the following two well-known results; see Chapter 21.11 in Harville (1997) or Theorem 6.19 in Zhang (1999) for Lemma 1 and Corollary 5 in Dhrymes (1978) or Theorem 2.8 in Zhang (1999) for Lemma 2.

LEMMA 1. *Let A be an $m\times m$ matrix and B a $p\times p$ matrix. Suppose that A has m (not necessarily distinct) eigenvalues, say $\lambda_1,\dots,\lambda_m$, and let $x_1,\dots,x_m$ represent a linearly independent set of eigenvectors with x_j corresponding to λ_j $(j=1,\dots,m)$. Moreover, let $\mu_1,\dots,\mu_p$ represent the (not necessarily distinct) eigenvalues of B, and let y_j $(j=1,\dots,p)$ be eigenvectors of B corresponding to μ_j $(j=1,\dots,p)$ such that the p eigenvectors $y_1,\dots,y_p$ are linearly independent. Then, the Kronecker product $A\otimes B$ has mp (not necessarily distinct) eigenvalues $\lambda_j\mu_k$ $(j=1,\dots,m;\ k=1,\dots,p)$, and $x_j\otimes y_k$ is an eigenvector of $A\otimes B$ corresponding to $\lambda_j\mu_k$.*

LEMMA 2. *Let A and B be, respectively, $m\times n$ and $n\times m$ matrices, where $m\leq n$. Then the eigenvalues of BA (an $n\times n$ matrix) consist of $n-m$ zeros and the m eigenvalues of AB (an $m\times m$ matrix).*

Clearly, $M=H\otimes S$, with

$$H:=\begin{pmatrix}0 & -1\\ 1 & 0\end{pmatrix}.$$

Because

$$S=(X\quad Y)\begin{pmatrix}0 & D\\ -D & 0\end{pmatrix}\begin{pmatrix}X'\\ Y'\end{pmatrix},$$

it follows from Lemma 2 that S has $n-2k$ eigenvalues zero and that the remaining eigenvalues of S are the $2k$ eigenvalues of the matrix

$$\begin{pmatrix}0 & D\\ -D & 0\end{pmatrix}\begin{pmatrix}X'\\ Y'\end{pmatrix}(X\quad Y)=\begin{pmatrix}0 & D\\ -D & 0\end{pmatrix}=H'\otimes D.$$

It is easy to check that i and $-i$, with $i=\sqrt{-1}$, are the two eigenvalues of H'; the vector $(1\quad i)'/\sqrt{2}$ is an eigenvector of H' corresponding to the eigenvalue i, and

https://doi.org/10.1017/S0266466600002103 Published online by Cambridge University Press

$(1 \quad -i)'/\sqrt{2}$ is an eigenvector corresponding to the eigenvalue $-i$. The positive definite diagonal matrix D has eigenvalues d_j $(j = 1,\ldots,k)$; the corresponding eigenvectors may be chosen as e_j (the jth column of I_k). According to Lemma 1, it thus follows that the eigenvalues of $H' \otimes D$ are id_j and $-id_j$ $(j = 1,\ldots,k)$; the corresponding eigenvectors may be chosen as

$$\frac{1}{\sqrt{2}}\begin{pmatrix}1\\ i\end{pmatrix} \otimes e_j, \qquad \frac{1}{\sqrt{2}}\begin{pmatrix}1\\ -i\end{pmatrix} \otimes e_j \qquad (j = 1,\ldots,k).$$

Although S and $H' \otimes D$ have the same nonzero eigenvalues, the corresponding eigenvectors are in general different. It is evident that a set of eigenvectors of S corresponding to these nonzero eigenvalues can be obtained from the corresponding eigenvectors of $H' \otimes D$ by premultiplication with the column-orthonormal matrix $(X \quad Y)$. We see, therefore, that the columns of the two matrices

$$\frac{1}{\sqrt{2}}(X + iY) \quad \text{and} \quad \frac{1}{\sqrt{2}}(X - iY)$$

are eigenvectors of S corresponding consecutively to its $2k$ nonzero eigenvalues $id_1,\ldots,id_k$ and $-id_1,\ldots,-id_k$. When choosing the (real) matrix $N/\sqrt{2}$ such that its columns constitute an orthonormal basis of the null space of $(X \quad Y)'$, the columns of this $n \times (n-2k)$ matrix $N/\sqrt{2}$ are trivially $n-2k$ linear independent eigenvectors of S corresponding to its $n-2k$ zero eigenvalues.

An eigenvalue decomposition of M can now be obtained by applying Lemma 1 once more, this time to the Kronecker product $M = H \otimes S$. Whereas H and H' have the same eigenvalues, the corresponding eigenvectors are different. The vector $(1 \quad -i)'/\sqrt{2}$ is an eigenvector of H corresponding to the eigenvalue i, and the vector $(1 \quad i)'/\sqrt{2}$ is an eigenvector of H corresponding to the eigenvalue $-i$. Combining our results and using Lemma 1, it follows that

$$MU = U\Lambda, \tag{*}$$

or, equivalently,

$$M = U\Lambda U^*,$$

which is, as claimed above, our first eigenvalue decomposition of M.

To see that M is also orthogonally similar to Λ, we recall that M is real and so (*) holds true not only for U but separately also for the real part of U and for the imaginary part of U. The real matrices

$$P_1 := \sqrt{2}\mathrm{Re}(U) = \frac{1}{\sqrt{2}}\begin{pmatrix} X & X & X & X & N & N\\ Y & -Y & -Y & Y & 0 & 0\end{pmatrix}$$

https://doi.org/10.1017/S0266466600002103 Published online by Cambridge University Press

and

$$P_2 := \sqrt{2}\mathrm{Im}(U) = \frac{1}{\sqrt{2}}\begin{pmatrix} Y & Y & -Y & -Y & 0 & 0 \\ -X & X & -X & X & -N & N \end{pmatrix}$$

of normalized eigenvectors of M are, however, not orthogonal because they are both rank deficient. We construct our orthogonal matrix P from P_1 and P_2 by selecting appropriately signed linearly independent block pairs (with $2n$ rows), and thus

$$M = P\Lambda P'$$

is, as claimed above, our second eigenvalue decomposition of M.

Solution 2, proposed by Geert Dhaene. The eigenvalue decomposition has the form

$$A = \begin{bmatrix} 0 & -S \\ S & 0 \end{bmatrix} = P\Lambda P',$$

with P orthogonal and Λ diagonal, given by

$$P = \frac{1}{\sqrt{2}}\begin{bmatrix} X & X & Y & Y & Z & Z \\ Y & -Y & X & -X & Z & -Z \end{bmatrix} \tag{1}$$

and

$$\Lambda = \begin{bmatrix} -D & 0 & 0 & 0 & 0 & 0 \\ 0 & D & 0 & 0 & 0 & 0 \\ 0 & 0 & D & 0 & 0 & 0 \\ 0 & 0 & 0 & -D & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \tag{2}$$

where Z is an $n \times (n-2k)$ matrix such that $[X \quad Y \quad Z]$ is orthogonal. (In the event that $n - 2k = 0$, assume Z and the zero rows and columns in Λ away.) Below, two *constructive* methods of proof of this decomposition are given. Use will be made of the matrix

$$B = \begin{bmatrix} X & 0 & Y & 0 & Z & 0 \\ 0 & X & 0 & Y & 0 & Z \end{bmatrix}.$$

Method a (Diagonalization of A). By successive transformations $A \to G'AG$, where G is orthogonal, and one can try to diagonalize A. Given the structure of A and noting that $SX = -YD$, $SY = XD$ and $SZ = 0$, an obvious candidate for G is B. This gives

https://doi.org/10.1017/S0266466600002103 Published online by Cambridge University Press

$$B'AB = \begin{bmatrix} 0 & 0 & 0 & -D & 0 & 0 \\ 0 & 0 & D & 0 & 0 & 0 \\ 0 & D & 0 & 0 & 0 & 0 \\ -D & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

It is not difficult now to find an orthogonal matrix that transforms $B'AB$ into a diagonal matrix. Such a matrix is given by

$$C = \frac{1}{\sqrt{2}} \begin{bmatrix} I & I & 0 & 0 & 0 & 0 \\ 0 & 0 & I & -I & 0 & 0 \\ 0 & 0 & I & I & 0 & 0 \\ I & -I & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & I & I \\ 0 & 0 & 0 & 0 & I & -I \end{bmatrix},$$

which gives $C'B'ABC = \Lambda$ and $BC = P$, as can be verified.

Method b (Decomposition of A^2) Taking the square of A yields

$$A^2 = \begin{bmatrix} -S^2 & 0 \\ 0 & -S^2 \end{bmatrix} = P\Lambda^2 P'. \tag{3}$$

The eigenvalue decomposition of $-S^2$ is easily found as

$$-S^2 = XD^2X' + YD^2Y' = [X \quad Y \quad Z] \begin{bmatrix} D^2 & 0 & 0 \\ 0 & D^2 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} X' \\ Y' \\ Z' \end{bmatrix}. \tag{4}$$

It follows from (3) and (4) that the eigenvalues of A^2 are the diagonal elements of D^2, each with multiplicity 4, and $2(n-2k)$ zeroes. The eigenvalues of A are the positive or negative square roots of those of A^2 and sum to zero because $\operatorname{tr}(A) = 0$. Hence, they are given by the diagonal elements of D and of $-D$, each with multiplicity 2 and $2(n-2k)$ zeroes. A solution for Λ is therefore given by (2). It also follows from (3) and (4) that the eigenvectors of A^2 are given by the columns of B and by (specific) linear combinations thereof. Due to the increased multiplicity of the nonzero eigenvalues in going from A to A^2, the eigenvectors of A are only a subset of those of A^2. In any case, they are also (specific) linear combinations of the columns of B. To find the correct linear combinations, observe that

$$AB = \begin{bmatrix} 0 & YD & 0 & -XD & 0 & 0 \\ -YD & 0 & XD & 0 & 0 & 0 \end{bmatrix}.$$

It is easy now to find a matrix Q such that $AQ = Q\Lambda$ and the columns of Q are mutually orthogonal. Such a matrix is given by

$$Q = \begin{bmatrix} X & X & Y & Y & Z & Z \\ Y & -Y & X & -X & Z & -Z \end{bmatrix}.$$

The solution for P given by (1) follows after normalizing Q.

Remark. It is easily checked that the above eigenvalue decomposition also holds under the weaker condition that D is a diagonal matrix.

NOTE

1. Excellent solutions have been proposed independently by P.R. Hansen, by S. Lawford, by M. van de Velden and H. Neudecker, by S. Zernov, and by R.W. Farebrother, the poser of the problem.

REFERENCES

Dhrymes, P.J. (1978) *Mathematics for Econometrics.* New York: Springer-Verlag.
Harville, D.A. (1997) *Matrix Algebra from a Statistician's Perspective.* New York: Springer-Verlag.
Zhang, F. (1999) *Matrix Theory: Basic Results and Techniques.* New York: Springer-Verlag.

99.3.2. *Maximum-Likelihood Estimation under Singularity*—Solution,[1] proposed by Steve Lawford. We may write the $m \times 1$ score vector as

$$\frac{\partial \ln L}{\partial a} = -\frac{n}{2}\frac{\partial \ln|A|}{\partial a} = -\frac{n}{2}\left(\frac{\partial \ln|A|}{\partial A'}\right)\left(\frac{\partial A}{\partial E'}\right)\left(\frac{\partial E}{\partial a}\right). \tag{1}$$

We note the three results

$$\frac{\partial \ln|A|}{\partial A'} = A^{-1}, \tag{2}$$

$$\frac{\partial A}{\partial E'} = \frac{2}{n}E' \tag{3}$$

and

$$\frac{\partial E}{\partial a} = -l_n, \tag{4}$$

where l_n is the $n \times 1$ summation vector $[1,1,\ldots,1]'$. Result (3) follows from $A = \frac{1}{n}E'E + \frac{1}{m}l_m l_m'$, given that $n \geq m+1$, because

$$\frac{\partial A}{\partial E'} = \frac{1}{n}\frac{\partial(E'E)}{\partial E'} + \frac{1}{m}\frac{\partial(l_m l_m')}{\partial E'} = \frac{1}{n}\frac{\partial(E'E)}{\partial E'} = \frac{2}{n}E'. \tag{5}$$

https://doi.org/10.1017/S0266466600002103 Published online by Cambridge University Press

Result (4) follows from $\varepsilon_i' = y_i' - a' - x_i'\Gamma'$, $(i = 1,2,\ldots,n)$, because

$$\frac{\partial \varepsilon_i'}{\partial a} = \frac{\partial y_i'}{\partial a} - \frac{\partial a'}{\partial a} - \frac{\partial x_i'}{\partial a}\Gamma' = -1. \tag{6}$$

Substituting (2), (3), and (4) into (1) gives

$$\begin{aligned}\frac{\partial \ln L}{\partial a} &= -\frac{n}{2}(A^{-1})\left(\frac{2}{n}E'\right)(-l_n) = A^{-1}E'l_n \\ &= A^{-1}\sum_{i=1}^{n}\varepsilon_i = A^{-1}\sum_{i=1}^{n}(y_i - a - \Gamma x_i) = 0\end{aligned} \tag{7}$$

as the first-order maximum conditions with respect to a. Thus,

$$\hat{a} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \Gamma x_i). \tag{8}$$

To check second-order conditions, we write the $m \times m$ Hessian matrix as

$$\frac{\partial^2 \ln L}{\partial a \partial a'} = \frac{\partial}{\partial a'}\left(\frac{\partial \ln L}{\partial a}\right) = \frac{\partial\left(A^{-1}\sum_{i=1}^{n}\varepsilon_i\right)}{\partial a'} = \frac{\partial}{\partial a'}\sum_{i=1}^{n}A^{-1}\varepsilon_i = \sum_{i=1}^{n}\frac{\partial}{\partial a'}A^{-1}\varepsilon_i. \tag{9}$$

Noting that

$$\frac{\partial}{\partial a'}A^{-1}\varepsilon_i = \left(\frac{\partial A^{-1}\varepsilon_i}{\partial \varepsilon_i'}\right)\left(\frac{\partial \varepsilon_i}{\partial a'}\right), \tag{10}$$

where

$$\frac{\partial A^{-1}\varepsilon_i}{\partial \varepsilon_i'} = A^{-1} \tag{11}$$

and

$$\frac{\partial \varepsilon_i}{\partial a'} = \frac{\partial y_i}{\partial a'} - \frac{\partial a}{\partial a'} - \Gamma\frac{\partial x_i}{\partial a'} = -\frac{\partial a}{\partial a'} = -l_m l_m', \tag{12}$$

we obtain

$$\left.\frac{\partial^2 \ln L}{\partial a \partial a'}\right|_{(a=\hat{a})} = -n\hat{A}^{-1}l_m l_m'. \tag{13}$$

From $|l_m l_m'| = 0$, it follows that the Hessian matrix is singular for all a (not just at the optimum):

$$|-nA^{-1}l_m l_m'| = (-n)^m |A|^{-1}|l_m l_m'| = 0. \tag{14}$$

Therefore, we cannot ascertain whether $\hat{a}$ is a maximum, a minimum, or a saddle point (Magnus and Neudecker, 1988, p. 123).

https://doi.org/10.1017/S0266466600002103 Published online by Cambridge University Press

NOTE

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

REFERENCE

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

99.3.3. *Non-Normality and Explanatory Data Analysis*—Solution, proposed by David M. Levy. The first step is to provide a workable definition of exploratory data analysis (EDA). I propose to define EDA pragmatically, that is, in terms of the beliefs and actions of the modeler. As the sample size increases, the posited model changes as variables move from the error term to the model. The second step is to ask whether such dynamic modeling is consistent with normality of the error term. In general, the answer is no.

Nonstandard analytical devices shall be employed to replace integration arguments with simple summation (Nelson, 1987). This substitution makes the proof entirely transparent for the linear case. Because "the model" under EDA is defined in terms of acting individuals who are allowed to learn from ever increasing samples and not in terms of a timeless platonic object in the realm of ideas, it is entirely appropriate that we employ nonstandard devices. For a platonist, consistency is the mark of existence. Both formalists (Robinson, 1974, p. 282) and pragmatists, as I take Tukey to be, would deny this link.

Let us write the traditional linear model that expresses a random variable Y_t as a function of a finite number (K) of independent variables $X_{1t} \dots X_{Kt}$ and an error term ε_t. Thus, for $t = 1 \dots T$,

$$Y_t = \sum_{k=1}^{K} \beta_k X_{tk} + \varepsilon_t.$$

Suppose that ε_t is normal. The Lindeberg–Feller (Feller, 1971) conditions allow us to decompose a normal distributed random variable into normal and non-normal components. Those components, which are normally distributed, we label η_t and proceed to other concerns. Now, consider the infinite number (H) of non-normal components of ε_t, which the Lindeberg–Feller conditions allow. If each of these random variables $X_{K+1t} \dots X_{Ht}$ is real valued, then there must be infinitesimals $\delta_{K+1} \dots \delta_H$ such that we can rewrite ε_t in terms of the product of reals and infinitesimals. Thus, we obtain

$$Y_t = \sum_{k=1}^{K} \beta_k X_{kt} + \sum_{k=K+1}^{H} \delta_k X_{kt} + \eta_t.$$

https://doi.org/10.1017/S0266466600002103 Published online by Cambridge University Press

Although this formulation is not sufficient to establish the normality of ε_t, it is necessary. If any of the δ_k were, contrary to supposition, noninfinitesimal and X_k were non-normal, then this contribution to ε_t would have to be a real-valued non-normal, thus violating the Lindeberg–Feller conditions.

The definition of EDA proposed here to moving a variable from the infinitesimal δ list to the real β list. At T_{N+1} but not at T_N, it may be possible to reject the hypothesis that $\beta_{K+1} = 0$. We consider two types of EDA: local and global. With local EDA, for some $T_2 > T_1$, the model expands from K noninfinitesimal contributions to $K+1$. With global EDA, in addition to local EDA at T_1, if local EDA at T_N, then local EDA at T_{N+1}.

THEOREM. *If local EDA and the discovered variable is non-normal, then there is non-normality at* T_1.

Proof. The Lindeberg–Feller conditions were violated at T_1 because not all of the excluded variables had the infinitesimal impact required for normality: the discovered non-normal variable, X_{K+1}, has real impact.

THEOREM. *If global EDA and the discovered variables are non-normal, then non-normality exists for all T.*

Proof. The induction step to establish the inconsistency of global EDA and normality is provided by the definition of global EDA. There is no T for which EDA is not possible.

It can be seen from this argument that normality is equivalent to the supposition that a regression model is correctly specified with respect to the non-normal variables. No non-normal variable with a noninfinitesimal impact is omitted from the model. Successful EDA violates this condition.

REFERENCES

Feller, W. (1971) *An Introduction to Probability Theory and Its Applications*, 2nd ed. New York: John Wiley & Sons.

Nelson, E. (1987) *Radically Elementary Probability Theory*. Princeton: Princeton University Press.

Robinson, A. (1974) *Non-Standard Analysis*, 2nd ed. Amsterdam: North-Holland.

99.3.4. *The Overdispersion Test in Count Data as Gauss–Newton Regression*—Solution,[1] proposed by M. Douglas Berg. For the negative binomial II regression model (e.g., Cameron and Trivedi, 1986), we have $E(Y_i|X_i) = e^{X_i'\beta}$ and $\text{var}(Y_i|X_i) = e^{X_i'\beta}(1 + \alpha e^{X_i'\beta})$. Thus, we have the following regression model

$$Y_i = E(Y_i|X_i) + \epsilon_i = e^{X_i'\beta} + \epsilon_i, \tag{1}$$

where ϵ_i has zero mean and $\text{var}(\epsilon_i|X_i) = \text{var}(Y_i|\epsilon_i) = e^{X_i'\beta}(1 + \alpha e^{X_i'\beta})$. Under H_0 of a Poisson model, we have $\text{var}(\epsilon_i|X_i) = e^{X_i'\beta}$ ($\alpha = 0$ under H_0). Below, we

https://doi.org/10.1017/S0266466600002103 Published online by Cambridge University Press

transform model (1) into a conditional homoskedastic error model under H_0. Multiplying (1) by $e^{-X_i'\beta/2}$ gives

$$e^{-X_i'\beta/2}Y_i = e^{-X_i'\beta/2}e^{X_i'\beta} + e^{-X_i'\beta}\epsilon_i$$
$$= e^{X_i'\beta/2} + \epsilon_i^*, \qquad \textbf{(2)}$$

where $\epsilon_i^* = e^{-X_i'\beta/2}\epsilon_i$ has zero mean and $\text{var}(\epsilon_i^*|X_i) = 1 + \alpha e^{X_i'\beta}$. Under H_0 of $\alpha = 0$, $\text{var}(\epsilon_i^*|X_i) = 1$. Therefore, testing overdispersion is equivalent to test conditional homoskedasticity of ϵ_i^*, i.e., testing $\alpha = 0$ in $\epsilon_i^{*2} = 1 + \alpha e^{X_i'\beta} +$ error. Or equivalently, testing $\alpha = 0$ in

$$\epsilon_i^{*2} - 1 = \alpha e^{X_i'\beta} + \text{error}. \qquad \textbf{(3)}$$

Using the Gauss–Newton regression (GNR) method suggested by Davidson and MacKinnon (1993), we regress $\epsilon_i^{*2} - 1$ on $\partial(\alpha e^{X_i'\beta})/\partial\alpha = e^{X_i'\beta}$ and $\partial(\alpha e^{X_i'\beta})/\partial\beta = \alpha X_i e^{X_i'\beta}$, i.e., we estimate the following artificial regression model:

$$\hat{\epsilon}_i^{*2} - 1 = e^{X_i'\beta}\gamma_0 + (\alpha X_i e^{X_i'\beta})\gamma_1 + \text{error}. \qquad \textbf{(4)}$$

Replacing β and ϵ_i^* by $\hat{\beta}$ and $\hat{\epsilon}_i^* = e^{-X_i'\hat{\beta}/2}\hat{\epsilon}_i$, where $\hat{\beta}$ is the maximum likelihood estimator of β under H_0 (a Poisson model) and $\hat{\epsilon}_i = Y_i - e^{X_i'\hat{\beta}}$. Also imposing H_0 of $\alpha = 0$ in (4) leads to

$$\hat{\epsilon}_i^{*2} - 1 = e^{X_i'\hat{\beta}}\gamma_0 + \text{error}. \qquad \textbf{(5)}$$

Davidson and MacKinnon's (1993) GNR test is to test $\gamma_0 = 0$ in (5). Let $\hat{\gamma}_0$ denote the least squares estimator of γ_0 from (5) and $\text{var}(\hat{\gamma}_0)$ be the estimated asymptotic variance of $\hat{\gamma}_0$. Then, $\hat{\gamma}_0/\sqrt{\text{var}(\hat{\gamma}_0)} \to N(0,1)$ in distribution under H_0 (large sample result).

For the second part of the question, note that the overdispersion test of Cameron and Trivedi (1990, p. 353) is to test $\gamma_0 = 0$ in the following regression equation:

$$(\sqrt{2}\hat{\mu}_i)^{-1}\{(Y_i - \hat{\mu}_i)^2 - Y_i\} = (\sqrt{2}\hat{\mu}_i)^{-1}(\hat{\mu}_i)^2 + \text{error}, \qquad \textbf{(6)}$$

where $\hat{\mu}_i = e^{X_i'\hat{\beta}}$. Multiplying (6) by $\sqrt{2}$ and using the identity $Y_i = e^{X_i'\hat{\beta}} + \hat{\epsilon}_i$, (6) becomes

$$[\hat{\epsilon}_i^{*2} - 1 - e^{-X_i'\hat{\beta}}\hat{\epsilon}_i] = e^{X_i'\hat{\beta}}\gamma_0 + \text{error}. \qquad \textbf{(7)}$$

Comparing (5) and (7), we see that Cameron and Trivedi's (1990) test (CT) differs from the GNR-based test of (5) in that the former test has an extra term, $-e^{-X_i'\hat{\beta}}$, on the dependent variable. Let $\hat{\gamma}_{0,\text{GNR}}$ and $\hat{\gamma}_{0,\text{CT}}$ denote the least squares

https://doi.org/10.1017/S0266466600002103 Published online by Cambridge University Press

estimator of γ_0 based on (5) and (7), respectively. Standard law of large numbers and central limit theorems arguments lead to

$$\sqrt{n}\hat{\gamma}_{0,\text{GNR}} = [E(e^{2X_i'\beta})]^{-1} n^{-1/2} \sum_{i=1}^{n} e^{X_i'\beta}(\epsilon_i^{*2} - 1) + o_p(1) \to N(0,\Omega), \qquad \textbf{(8)}$$

in distribution, where $\Omega = \{E[e^{2X_i'\beta}]\}^{-2}\text{var}(e^{X_i'\hat{\beta}}(\epsilon_i^{*2} - 1))$, and

$$\hat{\gamma}_{0,\text{CT}} = [E(e^{2X_i'\beta})]^{-1} n^{-1/2} \sum_{i=1}^{n} e^{X_i'\beta}[\epsilon_i^{*2} - 1 - e^{-X_i'\beta}\epsilon_i] + o_p(1) \to N(0,\Sigma), \qquad \textbf{(9)}$$

in distribution, where $\Sigma = \{E[e^{2X_i'\beta}]\}^{-2}\text{var}(e^{X_i'\hat{\beta}}[\epsilon_i^{*2} - 1 - e^{-X_i'\beta}\epsilon_i])$.

Consistent estimators of Ω and Σ can be easily obtained by using White's heteroskedastic robust method. Let $\hat{\Omega}$ and $\hat{\Sigma}$ denote the consistent estimators of Ω and Σ, respectively. Then we have, under H_0,

$$\sqrt{n}\hat{\gamma}_{0,\text{GNR}}\Big/\sqrt{\hat{\Omega}} \to N(0,1) \text{ in distribution,}$$

and

$$\sqrt{n}\hat{\gamma}_{0,\text{CT}}\Big/\sqrt{\hat{\Sigma}} \to N(0,1) \text{ in distribution.}$$

NOTE

1. An excellent solution has been proposed independently by Badi H. Baltagi and Dong Li, the posers of the problem.

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.

Davidson, R. & J. MacKinnon (1993) *Estimation and Inference in Econometrics*. New York: Oxford University Press.

https://doi.org/10.1017/S0266466600002103 Published online by Cambridge University Press