Econometric Theory, **16**, 2000, 791–796. Printed in the United States of America.

PROBLEMS AND SOLUTIONS

PROBLEMS

00.5.1. *An Exact Fitting Estimator for Linear Regression*, proposed by Oliver Linton. Suppose we have the linear regression model

$$y_i = \beta x_i + u_i,$$

where (u_i, x_i) are i.i.d. and mutually independent with finite variances and $E(u_i) = 0$. Suppose also that x_i is bounded strictly away from zero and infinity and that the density of u_i is positive at the origin. Now consider the following estimator.

(1) Compute the OLS estimator $\hat{\beta}$ and residuals $\hat{u}_i = y_i - \hat{\beta} x_i$.
(2) Sort the residuals so that $|\hat{u}_{(1)}| < \ldots < |\hat{u}_{(n)}|$ and denote by $(y_{(i)}, x_{(i)})$ the corresponding data values.
(3) Let

$$\tilde{\beta} = \frac{y_{(1)}}{x_{(1)}}.$$

Find the asymptotic distribution of $\tilde{\beta}$ and compare it with the distribution of the OLS estimator $\hat{\beta}$.

The natural extension of this procedure to k-variable regression is to use the k-best fitting observations and then solve the exact linear system that it implies. A generalization of this procedure is used by Fama and Bliss (1995) to estimate yield curves, the main difference being that their sorting procedure is based on "extraneous" information. The advantage is that this procedure exactly fits a particular subset of the observations.

REFERENCE

Fama, E.F. & R.R. Bliss (1995) The information in long-maturity forward rates. *American Economic Review* 77, 680–692.

00.5.2. *When Are Nested Reduced Rank Autoregressive Processes Integrated?* proposed by Paolo Paruolo. Consider the m-dimensional autoregressive (AR) process of order p

$$Y_t = (\Phi_1, \ldots, \Phi_p) \begin{pmatrix} Y_{t-1} \\ \cdots \\ Y_{t-p} \end{pmatrix} + \epsilon_t = FX_t + \epsilon_t, \quad \textbf{(1)}$$

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

https://doi.org/10.1017/S0266466600165089 Published online by Cambridge University Press

where Y_t and ϵ_t are $m \times 1$ vectors, $F := (\Phi_1, \ldots, \Phi_p)$, $X_t := (Y'_{t-1}, \ldots, Y'_{t-p})'$, and ϵ_t is a white-noise process. Reinsel and Velu (1998, Chap. 5) defined several processes that satisfy the restrictions

$$\Phi_j = \gamma \xi'_j \qquad j = 1, \ldots, p. \tag{2}$$

Derive conditions on γ and ξ_j under which Y_t is integrated of order 1, I(1), at frequency 0 (for a definition of integration, see Johansen, 1996, Chap. 3). Illustrate the results using the numerical example in Section 5.4 of Chapter 5 of Reinsel and Velu (1998, pp. 126–129) on U.S. hog data by applying the conditions on the estimated parameter matrices.

Hypothesis (2) is satisfied by nested reduced rank autoregressive processes, which are defined as the subclass of AR processes for which

$$span(\Phi_1) \supset span(\Phi_2) \supset \ldots \supset span(\Phi_p). \tag{3}$$

Defining γ to be a basis of $span(\Phi_1)$, restriction (3) can be represented in form (2), where $span(\xi'_1) \supset \ldots \supset span(\xi'_p)$. Reinsel and Velu (1998) also considered the class of processes satisfying the rank restriction $rank(F) =: q < m$, which is equivalent to $F = \gamma B$, where γ is $m \times q$ and B is $q \times n$, both of full rank q. This restriction can be written in form (2), where ξ'_j is of dimension $q \times m$ such that $B = (\xi'_1, \ldots, \xi'_p)$ is of full rank q.

REFERENCES

Johansen, S. (1996) *Likelihood Based Inference on Cointegrated Vector Autoregressive Models.* Oxford: Oxford University Press.

Reinsel, G.C. & P. Velu (1998) *Multivariate Reduced Rank Regression, Theory and Applications.* Lecture Notes in Statistics 136. New York: Springer-Verlag.

SOLUTIONS

99.5.1. *Use of the Moore–Penrose Inverse in Canonical Correspondence Analysis*—Solution,[1] proposed by Jan Graffelman. We can write the singular value decomposition of CCA as:

$$\mathbf{D}_r^{-1/2}\mathbf{P}\mathbf{Z}(\mathbf{Z}'\mathbf{D}_c\mathbf{Z})^{-1/2} = \hat{\mathbf{U}}\hat{\mathbf{D}}\hat{\mathbf{V}}' = \hat{\mathbf{U}}\hat{\mathbf{D}}\hat{\mathbf{\Gamma}}'\mathbf{D}_c\mathbf{Z}(\mathbf{Z}'\mathbf{D}_c\mathbf{Z})^{-1/2}.$$

Postmultiply by $(\mathbf{Z}'\mathbf{D}_c\mathbf{Z})^{-1/2}\mathbf{Z}'\mathbf{D}_c^{1/2}$ to obtain

$$\mathbf{D}_r^{-1/2}\mathbf{P}\mathbf{D}_c^{-1/2}\mathbf{D}_c^{1/2}\mathbf{Z}(\mathbf{Z}'\mathbf{D}_c\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{D}_c^{1/2} = \hat{\mathbf{U}}\hat{\mathbf{D}}\hat{\mathbf{\Gamma}}'\mathbf{D}_c\mathbf{Z}(\mathbf{Z}'\mathbf{D}_c\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{D}_c^{1/2}. \tag{1}$$

Consider now the case where $q > j$ and $\mathbf{Z}'\mathbf{D}_c\mathbf{Z}$ is singular. We define $\mathbf{X} = \mathbf{D}_c^{1/2}\mathbf{Z}$ and use the Moore–Penrose inverse of $\mathbf{Z}'\mathbf{D}_c\mathbf{Z}$, which we denote by $(\mathbf{X}'\mathbf{X})^+$. Then we replace $\mathbf{D}_c^{1/2}\mathbf{Z}(\mathbf{Z}'\mathbf{D}_c\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{D}_c^{1/2}$ by $\mathbf{X}(\mathbf{X}'\mathbf{X})^+\mathbf{X}' = \mathbf{X}\mathbf{X}^+ \times (\mathbf{X}')^+\mathbf{X}' = \mathbf{X}\mathbf{X}^+(\mathbf{X}^+)'\mathbf{X}' = \mathbf{X}\mathbf{X}^+(\mathbf{X}\mathbf{X}^+)'$, where proofs of the properties

https://doi.org/10.1017/S0266466600165089 Published online by Cambridge University Press

$(\mathbf{X}'\mathbf{X})^+ = \mathbf{X}^+(\mathbf{X}')^+$ and $(\mathbf{X}')^+ = (\mathbf{X}^+)'$ can be found in Graybill (1983, pp. 108–110).

We use two of the Moore–Penrose conditions, $\mathbf{X}^+ = \mathbf{X}^+\mathbf{X}\mathbf{X}^+$ and $\mathbf{X}^+\mathbf{X} = (\mathbf{X}^+\mathbf{X})'$, substituting the latter in the first: $\mathbf{X}^+ = (\mathbf{X}^+\mathbf{X})\mathbf{X}^+ = (\mathbf{X}^+\mathbf{X})'\mathbf{X}^+ = \mathbf{X}'(\mathbf{X}^+)'\mathbf{X}^+ = \mathbf{X}'(\mathbf{X}\mathbf{X}')^+$. Because $\mathbf{X}$ has full row rank j, $(\mathbf{X}\mathbf{X}')$ is nonsingular and thus $(\mathbf{X}\mathbf{X}')^+ = (\mathbf{X}\mathbf{X}')^{-1}$ so that $\mathbf{X}^+ = \mathbf{X}'(\mathbf{X}\mathbf{X}')^{-1}$. As a consequence, the term $\mathbf{X}\mathbf{X}^+ = \mathbf{X}\mathbf{X}'(\mathbf{X}\mathbf{X}')^{-1} = \mathbf{I}$. Consequently, $\mathbf{D}_c^{1/2}\mathbf{Z}(\mathbf{Z}'\mathbf{D}_c\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{D}_c^{1/2} = \mathbf{X}\mathbf{X}^+(\mathbf{X}\mathbf{X}^+)' = \mathbf{I}\mathbf{I}' = \mathbf{I}$.

Equation (1) thus reduces to $\mathbf{D}_r^{-1/2}\mathbf{P}\mathbf{D}_c^{-1/2}\mathbf{I} = \hat{\mathbf{U}}\hat{\mathbf{D}}\hat{\mathbf{\Gamma}}'\mathbf{D}_c^{1/2}\mathbf{I}$, from which it follows that $\mathbf{D}_r^{-1/2}\mathbf{P}\mathbf{D}_c^{-1/2} = \mathbf{U}\mathbf{D}\mathbf{V}' = \hat{\mathbf{U}}\hat{\mathbf{D}}\hat{\mathbf{\Gamma}}'\mathbf{D}_c^{1/2}$. Assuming no repetitions of singular values and ignoring the indeterminacy of the sign of the singular vectors, the singular value decomposition will be unique, and we have $\mathbf{U} = \hat{\mathbf{U}}$, $\mathbf{D} = \hat{\mathbf{D}}$ (thus implying $\mathbf{F} = \hat{\mathbf{F}}$) and $\mathbf{V} = \mathbf{D}_c^{1/2}\hat{\mathbf{\Gamma}}$. Because $\mathbf{V} = \mathbf{D}_c^{1/2}\mathbf{\Gamma} = \mathbf{D}_c^{1/2}\hat{\mathbf{\Gamma}}$, also $\mathbf{\Gamma} = \hat{\mathbf{\Gamma}}$.

NOTE

1. An alternative solution has been proposed independently by M. van de Velden.

REFERENCE

Graybill, F.A. (1983) *Matrices with Applications in Statistics.* London: Wadsworth and Brooks/Cole.

99.5.2. *Relationship Between Two Eigenmatrices of a (Real) Symmetric Matrix*—Solution,[1] proposed by Hong-Tu Zhu. Assume that

$$\Lambda = \begin{pmatrix} \lambda_1\mathbf{I}_{k_1} & & \\ & \ddots & \\ & & \lambda_p\mathbf{I}_{k_p} \end{pmatrix},$$

where k_i is the frequency of λ_i, $(i = 1,\ldots,p)$.

Partition $\mathbf{U}$ as $[U_1, U_2, \ldots, U_p]$ and $\mathbf{W}$ as $[W_1, \ldots, W_p]$, where U_i and W_i are $n \times k_i$ matrices, $i = 1,\ldots,p$. Because $\mathbf{A}\mathbf{U} = \mathbf{U}\mathbf{\Lambda}$ and $\mathbf{A}\mathbf{W} = \mathbf{W}\mathbf{\Lambda}$, it is easy to see that $\mathbf{A}U_i = \lambda_i U_i$ and $\mathbf{A}W_i = \lambda_i W_i$. Define the space

$$\mathcal{U}_\lambda = \{\mathbf{x} | A\mathbf{x} = \lambda\mathbf{x}\},$$

where $\mathbf{x}$ is an $n \times 1$ vector. Thus, $\mathcal{U}_{\lambda_i}$ is a vector space with $\dim(\mathcal{U}_{\lambda_i}) = k_i$ (Schott, 1997).

Because $\mathbf{U}'\mathbf{U} = I_k$, we can show that the columns of U_i is a basis of the space $\mathcal{U}_{\lambda_i}$. Similarly, the columns of W_i is also a basis of the space $\mathcal{U}_{\lambda_i}$. Therefore, there exists a nonsingular $k_i \times k_i$ matrix Φ_{ii} such that $W_i = U_i\Phi_{ii}$. Thus,

https://doi.org/10.1017/S0266466600165089 Published online by Cambridge University Press

$\mathbf{W} = \mathbf{U}\boldsymbol{\Phi}$, where $\boldsymbol{\Phi} = \text{diag}(\boldsymbol{\Phi}_{11},\ldots,\boldsymbol{\Phi}_{pp})$. Substituting $\mathbf{W} = \mathbf{U}\boldsymbol{\Phi}$ into $\mathbf{U}'\mathbf{U} = \mathbf{W}'\mathbf{W} = \mathbf{I}_k$, we have

$$\boldsymbol{\Phi}'\boldsymbol{\Phi} = \mathbf{I}_k.$$

This indicates that $\boldsymbol{\Phi}$ is orthogonal. This completes the proof.

NOTE

1. An excellent solution has been proposed independently by M. van de Velden and H. Neudecker, the posers of the problem.

REFERENCE

Schott, J.R. (1997) *Matrix Analysis for Statistics.* New York: John Wiley & Sons.

99.5.3. *Does the Link Matter?*—Solution, proposed by Joao M.C. Santos Silva. The log-likelihood function of the binary model can be written as

$$\sum_{i=1}^{n} \{y_i \ln[F(\beta_0 + \beta_1 x_i)] + (1 - y_i) \ln[1 - F(\beta_0 + \beta_1 x_i)]\},$$

where $P(Y_i = 1|x) = F(\beta_0 + \beta_1 x_i)$ and y_i denotes a realization of the random variable Y_i. The maximum likelihood estimates (MLE's) of the parameters are obtained by solving

$$\sum_{i=1}^{n} \frac{y_i - \hat{F}(x_i)}{\hat{F}(x_i)[1 - \hat{F}(x_i)]} \hat{f}(x_i) = 0, \tag{1}$$

$$\sum_{i=1}^{n} \frac{y_i - \hat{F}(x_i)}{\hat{F}(x_i)[1 - \hat{F}(x_i)]} \hat{f}(x_i)x_i = 0, \tag{2}$$

where $\hat{F}(x_i)$ denotes $F(\beta_0 + \beta_1 x_i)$ evaluated at the MLE and $\hat{f}(x_i)$ is the derivative of this function with respect to the index $\beta_0 + \beta_1 x_i$, also evaluated at the MLE.

1. Consider now the case in which x_i only assumes two different values. Without loss of generality, assume that $x_i \in \{0,1\}$. Then, (2) can be written as $\sum_{x_i=1}(y_i - \hat{F}(1))[\hat{f}(1)]/\{\hat{F}(1)[1 - \hat{F}(1)]\} = 0$, from where it results that $\hat{F}(1) = (\sum_{x_i=1} y_i)/n_1$, where n_1 is the number of observations for which $x_i = 1$. Writing (1) as $\sum_{x_i=0}(y_i - \hat{F}(0))[\hat{f}(0)]/\{\hat{F}(0)[1 - \hat{F}(0)]\} + \sum_{x_i=1}(y_i - \hat{F}(1))[\hat{f}(1)]/\{\hat{F}(1)[1 - \hat{F}(1)]\} = 0$, and given $\hat{F}(1)$, one has that $\hat{F}(0) = (\sum_{x_i=0} y_i)/(n - n_1)$. Therefore, in this case, the estimates of $F(\beta_0 + \beta_1 x_i)$ and the value of the like-

https://doi.org/10.1017/S0266466600165089 Published online by Cambridge University Press

lihood function evaluated at the MLE are the same, whatever the form of $F(\cdot)$. The obvious intuition behind this result is that $F(\beta_0 + \beta_1 x_i) = E[Y_i|x]$ and, because $x_i \in \{0,1\}$, any $F(\cdot)$ can be made to pass through the means of y_i, computed at the different values of x_i.

2. Under the conditions described above, it is easy to show that $\hat{\beta}_0 = F^{-1}([\sum_{x_i=0} y_i]/[n - n_1])$ and $\hat{\beta}_1 = F^{-1}([\sum_{x_i=1} y_i]/n_1) - F^{-1}([\sum_{x_i=0} y_i]/[n - n_1])$. If, as usual, $\lim_{x\to-\infty} F(x) = 0$ and $\lim_{x\to\infty} F(x) = 1$, identification of β_0 and β_1 requires $0 < \hat{F}(x) < 1$.

99.5.4. *Robustness of Tests for Functional Form to Serial Correlation*—Solution, proposed by Jeffrey M. Wooldridge.

(i) From Wooldridge (1991), the nR_u^2 statistic is identical to nR_u^2 from the regression $\hat{u}_t$ on $\hat{r}_t$, $t = 1,2,\ldots,n$, where $\hat{r}_t$ is the $1 \times q$ vector of OLS residuals from the regression $\hat{\mathbf{g}}_t$ on $\mathbf{x}_t$, $t = 1,2,\ldots,n$. So the statistic is

$$\left(n^{-1/2}\sum_{t=1}^{n}\hat{\mathbf{r}}_t'\hat{u}_t\right)'\left(\hat{\sigma}^2 n^{-1}\sum_{t=1}^{n}\hat{\mathbf{r}}_t'\hat{\mathbf{r}}_t\right)^{-1}\left(n^{-1/2}\sum_{t=1}^{n}\hat{\mathbf{r}}_t'\hat{u}_t\right),$$

where $\hat{\sigma}^2 = n^{-1}\sum_{t=1}^{n}\hat{u}_t^2$. Following Wooldridge (1991), under H_0, by the CLT and the law of large numbers, it can be shown that $n^{-1/2}\sum_{t=1}^{n}\hat{\mathbf{r}}_t'\hat{u}_t = n^{-1/2}\sum_{t=1}^{n}\mathbf{r}_t'u_t + o_p(1)$ and $\hat{\sigma}^2(n^{-1}\sum_{t=1}^{n}\hat{\mathbf{r}}_t'\hat{\mathbf{r}}_t) = \sigma^2 E(\mathbf{r}_t'\mathbf{r}_t) + o_p(1)$. Therefore, nR_u^2 is asymptotically equivalent to

$$\left(n^{-1/2}\sum_{t=1}^{n}\mathbf{r}_t'u_t\right)'[\sigma^2 E(\mathbf{r}_t'\mathbf{r}_t)]^{-1}\left(n^{-1/2}\sum_{t=1}^{n}\mathbf{r}_t'u_t\right),$$

where u_t is the error at time t and $\mathbf{r}_t \equiv \mathbf{g}(\mathbf{x}_t,\boldsymbol{\delta}) - \mathbf{x}_t\boldsymbol{\Pi}$, and $\boldsymbol{\Pi}$ is the $k \times q$ matrix of population regression coefficients. To prove the result, it suffices to show that (a) $\mathrm{Var}(\mathbf{r}_t'u_t) = \sigma^2 E(\mathbf{r}_t'\mathbf{r}_t)$ and (b) $E(u_t u_{t+j}\mathbf{r}_t'\mathbf{r}_{t+j}) = 0$, $j \geq 1$. But under (A1) and (A2), the law of iterated expectations implies that $\mathrm{Var}(\mathbf{r}_t'u_t) = E(u_t^2\mathbf{r}_t'\mathbf{r}_t) = E[E(u_t^2|\mathbf{x}_t)\mathbf{r}_t'\mathbf{r}_t] = \sigma^2 E(\mathbf{r}_t'\mathbf{r}_t)$. To show (b), we again use iterated expectations:

$$E(u_t u_{t+j}\mathbf{r}_t'\mathbf{r}_{t+j}) = E[E(u_t u_{t+j}|\mathbf{x}_t,\mathbf{x}_{t+j})\mathbf{r}_t'\mathbf{r}_{t-j}] = \eta_j E(\mathbf{r}_t'\mathbf{r}_{t+j})$$

by (A3), where $\eta_j \equiv E(u_t u_{t+j})$.

Finally, if $\{\mathbf{x}_t\}$ is serially independent, $\mathbf{r}_t$ and $\mathbf{r}_{t+j}$ are independent, and so $E(\mathbf{r}_t'\mathbf{r}_{t+j}) = E(\mathbf{r}_t')E(\mathbf{r}_{t+j}) = 0$, because $E(\mathbf{r}_t) = 0$ for all t. (Because $\mathbf{x}_t$ contains unity, the population residuals from regressing $\mathbf{g}(\mathbf{x}_t,\boldsymbol{\delta})$ on $\mathbf{x}_t$ have mean zero.)

(ii) Without (A4), $E(u_t u_{t+j}\mathbf{r}_t'\mathbf{r}_{t+j}) = \eta_j E(\mathbf{r}_t'\mathbf{r}_{t+j}) = \sigma^2\rho_j E(\mathbf{r}_t'\mathbf{r}_{t+j})$, where $\rho_j = \mathrm{Corr}(u_t,u_{t+j})$. Therefore, the asymptotic variance matrix of $n^{-1/2}\sum_{t=1}^{n}\mathbf{r}_t'u_t$ can be written as

https://doi.org/10.1017/S0266466600165089 Published online by Cambridge University Press

$$\sigma^2 E(\mathbf{r}_t'\mathbf{r}_t) + \sum_{j=1}^{\infty} \eta_j \{E(\mathbf{r}_t'\mathbf{r}_{t+j}) + E(\mathbf{r}_{t+j}'\mathbf{r}_t)\}$$

$$\equiv \sigma^2 \left(E(\mathbf{r}_t'\mathbf{r}_t) + \sum_{j=1}^{\infty} \rho_j (\Lambda_j + \Lambda_j') \right).$$

This differs from the standard formula by the infinite sum, which depends on the products $\rho_j \Lambda_j$.

ACKNOWLEDGMENT OF AUTHORSHIP

The author of the "Corrigendum to Solution to Problem 98.5.53" published in *ET* 16(3), pp. 459–461 was Seiji Nabeya, who pointed out the error both in the original problem and in the published solution given by Paulo M.M. Rodrigues.

The following typographical corrections should also be made:

page 459 line 3: "the process" becomes "in the process"

page 459 line 13: "U_3" becomes "$U_3 =$"

page 460 line 6 from below: "(2)" becomes "(4)"

https://doi.org/10.1017/S0266466600165089 Published online by Cambridge University Press