Econometric Theory, **15**, 1999, 777–788. Printed in the United States of America.

PROBLEMS AND SOLUTIONS

PROBLEMS

99.5.1. *Use of the Moore–Penrose Inverse in Canonical Correspondence Analysis*, proposed by Jan Graffelman. Consider a non-negative data matrix $\mathbf{N}$ (e.g., a contingency table or count data). The correspondence analysis (CA) of this matrix can be performed by the singular value decomposition

$$\mathbf{D}_r^{-1/2}\mathbf{P}\mathbf{D}_c^{-1/2} = \mathbf{U}\mathbf{D}\mathbf{V}',$$

with $\mathbf{U}'\mathbf{U} = \mathbf{I}$ and $\mathbf{V}'\mathbf{V} = \mathbf{I}$ and where $\mathbf{P}$ is the so-called $i \times j$ correspondence matrix, formed by dividing the data matrix by its grand total: $\mathbf{P} = \mathbf{N}/\mathbf{1}'\mathbf{N}\mathbf{1}$ (see Gifi, 1990, Sect. 8.3; or Greenacre, 1993). Here $\mathbf{D}_r$ and $\mathbf{D}_c$ are diagonal matrices of the row and column sums of $\mathbf{P}$, respectively. Using a particular scaling, the coordinates plotted in a CA biplot (principal row coordinates $\mathbf{F}$ and standard column coordinates $\mathbf{\Gamma}$, representing row and column categories of the data matrix, respectively) are given by

$$\mathbf{F} = \mathbf{D}_r^{-1/2}\mathbf{U}\mathbf{D}, \qquad \mathbf{\Gamma} = \mathbf{D}_c^{-1/2}\mathbf{V}. \tag{1}$$

Canonical correspondence analysis (CCA) is a restricted form of ordinary correspondence analysis and can be performed by the singular value decomposition

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

with $\hat{\mathbf{U}}'\hat{\mathbf{U}} = \mathbf{I}$, $\hat{\mathbf{V}}'\hat{\mathbf{V}} = \mathbf{I}$ and where $\mathbf{Z}$ is a $j \times q$ matrix of external, usually quantitative, variables. In CCA the standard coordinates of the column categories of the data matrix $\mathbf{N}$, say, $\hat{\mathbf{\Gamma}}$, are restricted to be linear combinations of these external variables (cf. Ter Braak, 1986, App.). The CCA solution for the column and row categories of $\mathbf{N}$ is given by

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

If $\mathbf{Z}$ contains more external variables than there are column categories in $\mathbf{N}$ ($q > j$), matrix $\mathbf{Z}'\mathbf{D}_c\mathbf{Z}$ is singular, and we run into computational problems. These can be avoided if we use the Moore–Penrose inverse of $\mathbf{Z}'\mathbf{D}_c\mathbf{Z}$. Show that, if $q > j$, by using the Moore–Penrose inverse of $\mathbf{Z}'\mathbf{D}_c\mathbf{Z}$, the solutions of CA (equation (1) and CCA (equation (2)) will be identical, that is, $\mathbf{F} = \hat{\mathbf{F}}$, $\mathbf{D} = \hat{\mathbf{D}}$, and $\mathbf{\Gamma} = \hat{\mathbf{\Gamma}}$.

REFERENCES

Gifi, A. (1990) *Nonlinear Multivariate Analysis.* New York: Wiley.

Greenacre, M.J. (1993) Biplots in correspondence analysis. *Journal of Applied Statistics* 20, 251–269.

Ter Braak, C.J.F. (1986) Canonical correspondence analysis: A new eigenvector technique for multivariate direct gradient analysis. *Ecology* 67, 1167–1179.

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

https://doi.org/10.1017/S0266466699155075 Published online by Cambridge University Press

99.5.2. *Relationship between Two Eigenmatrices of a (Real) Symmetric Matrix*, proposed by Heinz Neudecker. Consider the following two eigenequations for a (real) symmetric matrix $\mathbf{A}$ of order n, rank k, and p distinct (nonzero) eigenvalues λ_i:

$$\mathbf{AU} = \mathbf{U\Lambda},$$

$$\mathbf{AW} = \mathbf{W\Lambda},$$

where

$$\mathbf{U}'\mathbf{U} = \mathbf{W}'\mathbf{W} = \mathbf{I}_k$$

and $\mathbf{\Lambda}$ is diagonal and nonsingular. Show then that

$$\mathbf{W} = \mathbf{U\Phi},$$

where

$$\mathbf{\Phi} = \begin{pmatrix} \mathbf{\Phi}_{11} & & \\ & \ddots & \\ & & \mathbf{\Phi}_{pp} \end{pmatrix}$$

is orthogonal and the order of $\mathbf{\Phi}_{ii}$ equals the frequency of λ_i, $(i = 1 \ldots p)$.

99.5.3. *Does the Link Matter?* proposed by Joao M.C. Santos Silva. Consider a binary random variable Y_i such that

$$P(Y_i = 1|x) = F(\beta_0 + \beta_1 x_i), \qquad i = 1, \ldots, n,$$

where the link $F(\cdot)$ is a continuous distribution function.

(1) Show that under certain conditions on the sample of x_i, the value of the log-likelihood function evaluated at the maximum likelihood estimates of β_0 and β_1 is the same, independent of the form of the link function.
(2) What are the maximum likelihood estimates of β_0 and β_1 under the conditions described previously?

99.5.4. *Robustness of Tests for Functional Form to Serial Correlation*, proposed by Jeffrey M. Wooldridge. This question is motivated by the many simulation studies that have appeared in the last 20 years to evaluate the robustness of functional form misspecification tests to serial correlation in the errors. Ramsey's (1969) RESET has perhaps been the most studied (see, e.g., Thursby, 1979, 1981, 1989; Porter and Kashyap, 1984), but it is no more difficult to treat the general case.

Let $\{(\mathbf{x}_t, y_t) : t = 1, 2, \ldots\}$ be a time series process with $\mathbf{x}_t$ a $1 \times k$ vector and y_t a scalar. In what follows, assume that this process is strictly stationary and weakly

https://doi.org/10.1017/S0266466699155075 Published online by Cambridge University Press

dependent and has enough finite moments to apply the weak law of large numbers and the central limit theorem.

A general class of tests for functional form in linear models can be constructed as follows. The null model is

$$y_t = \mathbf{x}_t\boldsymbol{\beta} + u_t, \qquad E(u_t|\mathbf{x}_t) = 0, \tag{1}$$

where $\mathbf{x}_t$ contains unity to allow for a nonzero intercept. To test the null hypothesis that $E(y_t|\mathbf{x}_t) = \mathbf{x}_t\boldsymbol{\beta}$, let $\mathbf{g}(\mathbf{x}_t,\boldsymbol{\delta})$ be a $1 \times q$ vector of nonlinear functions of $\mathbf{x}_t$ and some nuisance parameters, $\boldsymbol{\delta}$. Under H_0, $E[\mathbf{g}(\mathbf{x}_t,\boldsymbol{\delta})'u_t] = \mathbf{0}$. Let $\hat{\boldsymbol{\delta}}$ be a $\sqrt{n}$-consistent estimator of $\boldsymbol{\delta}$ and define $\hat{\mathbf{g}}_t \equiv \mathbf{g}(\mathbf{x}_t,\hat{\boldsymbol{\delta}})$. For many tests, $\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\beta}}$. A leading example is RESET, where, for example, $\hat{\mathbf{g}}_t = (\hat{y}_t^2, \hat{y}_t^3, \hat{y}_t^4)$ and $\hat{y}_t = \mathbf{x}_t\hat{\boldsymbol{\beta}}$ are the ordinary least squares (OLS) fitted values. For the Davidson and MacKinnon (1981) test against nonnested alternatives, the $\mathbf{g}(\mathbf{x}_t,\hat{\boldsymbol{\delta}})$ are scalar fitted values for the alternative, nonlinear conditional mean function (for a summary, see Wooldridge, 1991).

The standard score-type statistic is obtained as nR_u^2 from the regression

$$\hat{u}_t \quad \text{on} \quad \mathbf{x}_t, \hat{\mathbf{g}}_t, \qquad t = 1,2,\ldots,n, \tag{2}$$

where $\hat{u}_t$ are the OLS residuals and R_u^2 is the usual R-squared. An F-test is also available, but we will focus on the n-R-squared form.

(i) Show that, under any needed regularity conditions, the following assumptions are sufficient for nR_u^2 to be asymptotically χ_q^2.
(A1) $E(u_t|\mathbf{x}_t) = 0$
(A2) $E(u_t^2|\mathbf{x}_t) = \sigma^2$
(A3) $E(u_t u_{t+j}|\mathbf{x}_t,\mathbf{x}_{t+j}) = E(u_t u_{t+j}), j \geq 1$
(A4) $\{\mathbf{x}_t: t = 1,2,\ldots,\}$ is serially independent.

Assumption A1 states the null hypothesis, whereas A2 is the appropriate homoskedasticity assumption (under H_0). Assumption A3 is a kind of strict exogeneity assumption, which holds when $\{\mathbf{x}_t\}$ is independent of $\{u_t\}$, something that is virtually always the case in simulation studies. Importantly, A3 does *not* restrict the serial dependence in $\{u_t\}$. Assumption A4 is the key to understanding a variety of simulation outcomes: if the $\mathbf{x}_t$ are generated to be independent over time, the amount of serial correlation in $\{u_t\}$ is irrelevant (subject to stationarity and weak dependence being met, of course).

(ii) Let $\mathbf{r}_t \equiv \mathbf{g}(\mathbf{x}_t,\boldsymbol{\delta}) - \mathbf{x}_t\boldsymbol{\Pi}$, where $\boldsymbol{\Pi} \equiv [E(\mathbf{x}_t'\mathbf{x}_t)]^{-1}E(\mathbf{x}_t'\mathbf{r}_t)$ is the $k \times q$ matrix of population regression coefficients from $\mathbf{g}(\mathbf{x}_t,\boldsymbol{\delta})$ on $\mathbf{x}_t$. Show that without A4, the size distortions due to serial dependence in $\{u_t\}$ depend on $\rho_j\boldsymbol{\Lambda}_j, j = 1,2,\ldots,$ where the ρ_j are the autocorrelations of $\{u_t\}$ and $\boldsymbol{\Lambda}_j \equiv E(\mathbf{r}_t'\mathbf{r}_{t+j})$.

REFERENCES

Davidson, R. & J. MacKinnon (1981) Several tests for model specification in the presence of alternative hypotheses. *Econometrica* 49, 781–793.

Porter, R. & A. Kashyap (1984) Autocorrelation and the sensitivity of RESET. *Economics Letters* 14, 229–233.

https://doi.org/10.1017/S0266466699155075 Published online by Cambridge University Press

Ramsey, J. (1969) Tests for specification errors in the classical linear least squares regression analysis. *Journal of the Royal Statistical Society, Series B* 31, 350–371.
Thursby, J. (1979) Alternative specification error tests: A comparative study. *Journal of the American Statistical Association* 74, 222–225.
Thursby, J. (1981) A test strategy for discriminating between autocorrelation and misspecification in regression analysis. *Review of Economics and Statistics* 63, 117–123.
Thursby, J. (1989) A comparison of several specification error tests for a general alternative. *International Economic Review* 30, 217–230.
Wooldridge, J. (1991) On the application of robust, regression-based diagnostics to models of conditional means and conditional variances. *Journal of Econometrics* 47, 5–46.

SOLUTIONS

98.5.1. *A Useful Result for Lipschitz Functions of Martingales*—Solution,[1] proposed by Atsushi Inoue. Let p be a natural number and $\|\cdot\|_p$ denote the L^p norm, i.e., $\|X\|_p = (E|X|^p)^{1/p}$. We replace $E(X) < \infty$ and $E(g(X)) < \infty$ by $E(X^p) < \infty$ and $E(g(X)^p) < \infty$, respectively, and we shall prove the L^p case,

$$\|g(X) - E(g(X)|\mathcal{G})\|_p \leq \Delta \|X - E(X|\mathcal{G})\|_p,$$

which includes the proposition as a special case.

By the Minkowski inequality,

$$\begin{aligned}\|g(X) - E(g(X)|\mathcal{G})\|_p &= \|g(X) - g(E(X|\mathcal{G})) + g(E(X|\mathcal{G})) - E(g(X)|\mathcal{G})\|_p \\ &\leq \|g(X) - g(E(X|\mathcal{G}))\|_p + \|g(E(X|\mathcal{G})) - E(g(X)|\mathcal{G}\|_p. \end{aligned} \quad \textbf{(1)}$$

By definition of the Lipschitz condition, there is a constant $M \in (0,\infty)$ such that

$$|g(x') - g(x)| \leq M|x' - x|$$

for all $x, x' \in U$. Thus we have

$$\begin{aligned}\|g(X) - g(E(X|\mathcal{G}))\|_p &= \left(\int_U |g(x) - g(E(X|\mathcal{G}))|^p \, dF(x) \right)^{1/p} \\ &\leq \left(\int_U M^p |x - E(X|\mathcal{G})|^p \, dF(x) \right)^{1/p} \\ &= M\|X - E(X|\mathcal{G})\|_p, \end{aligned} \quad \textbf{(2)}$$

where $F(\cdot)$ is the cumulative distribution function (c.d.f.) of X. By the conditional Jensen inequality (Williams, 1991, p. 88),

$$\begin{aligned}\|g(E(X|\mathcal{G})) - E(g(X)|\mathcal{G})\|_p &= \|E[g(E(X|\mathcal{G})) - g(X)|\mathcal{G}]\|_p \\ &\leq \|g(E(X|\mathcal{G})) - g(X)\|_p. \end{aligned} \quad \textbf{(3)}$$

https://doi.org/10.1017/S0266466699155075 Published online by Cambridge University Press

It follows from (1)–(3) that

$$\|g(X) - E(g(X)|\mathcal{G})\|_p \leq \Delta \|X - E(X|\mathcal{G})\|_p,$$

where $\Delta = 2M$.

NOTE

1. Excellent solutions have been proposed independently by H. White, the poser of the problem.

REFERENCE

Williams, D. (1991) *Probability with Martingales.* New York: Cambridge University Press.

98.5.2. *Within Two Way Is Equivalent to Two Withins One Way*—Solution,[1] proposed by Dong Li.

(a) Consider the fixed-effects two-way error component panel data model

$$y_{it} = \alpha + X_{it}'\beta + \mu_i + \lambda_t + \nu_{it}. \quad \textbf{(1)}$$

Stack the data in the way that index t grows faster than index i, i.e., $y = (y_{11}, y_{12}, \ldots, y_{1T}, y_{21}, \ldots, y_{NT})'$.

We will show a formal proof first and then give some intuitive explanations.

One can obtain the within two-way estimator of β by applying the following within two-way transformation given by Wallace and Hussain (1969),

$$Q = E_N \otimes E_T = I_N \otimes I_T - I_N \otimes \bar{J}_T - \bar{J}_N \otimes I_T + \bar{J}_N \otimes \bar{J}_T, \quad \textbf{(2)}$$

where $E_N = I_N - \bar{J}_N$, $E_T = I_T - \bar{J}_T$, $\bar{J}_N = J_N/N$, $\bar{J}_T = J_T/T$, J_N is an $N \times N$ matrix of ones, J_T is a $T \times T$ matrix of ones, I_N and I_T are $N \times N$ and $T \times T$ identity matrices, respectively. The within two-way transformation "sweeps out" both the individual and time effects and the constant.

The within estimator of β is therefore given by $\hat{\beta}_W = (X'QX)^{-1}X'Qy$, by noting $Q' = Q$ and $Q^2 = Q$ (for details, see Baltagi 1995).

Now we will show that $\hat{\beta}_W$ can be obtained by applying two within one-way transformations.

The within one-way transformation ignoring the time effect is given by

$$Q_1 = I_N \otimes E_T = I_N \otimes I_T - I_N \otimes \bar{J}_T. \quad \textbf{(3)}$$

This transformation removes the individual means (and the constant α) from observations.

The within one-way transformation ignoring the individual effect is given by

$$Q_2 = E_N \otimes I_T = I_N \otimes I_T - \bar{J}_N \otimes I_T. \quad \textbf{(4)}$$

This transformation removes the time means from observations.

https://doi.org/10.1017/S0266466699155075 Published online by Cambridge University Press

The resulting regression, after the preceding two within one-way transformations, is given by the following vector form:

$$Q_2 Q_1 y = Q_2 Q_1 X\beta + Q_2 Q_1 \nu. \tag{5}$$

The ordinary least squares (OLS) estimator of β in the preceding regression, $\tilde{\beta}$, is given by $\tilde{\beta} = (X'Q_1'Q_2'Q_2Q_1X)^{-1}X'Q_1'Q_2'Q_2Q_1y$. By noting that $Q_1' = Q_1$, $Q_2' = Q_2$, and

$$Q_1 Q_2 = (I_N \otimes E_T)(E_N \otimes I_T) = E_N \otimes E_T = Q, \tag{6}$$

we have $\tilde{\beta} = (X'QX)^{-1}X'Qy = \hat{\beta}_W$; i.e., the within two-way transformation is equivalent to two within one-way transformations.

Now we will give some intuitive explanations.

The typical element in the resulting regression after the within two-way transformation given in (2) is

$$(y_{it} - \bar{y}_{i\cdot} - \bar{y}_{\cdot t} + \bar{y}_{\cdot\cdot}) = (x_{it} - \bar{x}_{i\cdot} - \bar{x}_{\cdot t} + \bar{x}_{\cdot\cdot})\beta + (\nu_{it} - \bar{\nu}_{i\cdot} - \bar{\nu}_{\cdot t} + \bar{\nu}_{\cdot\cdot}), \tag{7}$$

where $\bar{y}_{i\cdot}$, $\bar{y}_{\cdot t}$, and $\bar{y}_{\cdot\cdot}$ are average over time, average over individual, and overall average, respectively, by noting that the transformation $I_N \otimes \bar{J}_T$ yields the average over time for each individual and that the transformation $\bar{J}_N \otimes I_T$ yields the average over individual for each time period and that the transformation $\bar{J}_N \otimes \bar{J}_T$ yields the overall average.

Now consider the two within one-way transformations. For the first one-way transformations given in (3), the typical element in the resulting regression is

$$(y_{it} - \bar{y}_{i\cdot}) = (x_{it} - \bar{x}_{i\cdot})\beta + \lambda_t - \bar{\lambda} + (\nu_{it} - \bar{\nu}_{i\cdot}). \tag{8}$$

After applying the second within one-way transformation to the preceding regression, the typical element is

$$(y_{it} - \bar{y}_{i\cdot} - \bar{y}_{\cdot t} + \bar{y}_{\cdot\cdot}) = (x_{it} - \bar{x}_{i\cdot} - \bar{x}_{\cdot t} + \bar{x}_{\cdot\cdot})\beta + (\nu_{it} - \bar{\nu}_{i\cdot} - \bar{\nu}_{\cdot t} + \bar{\nu}_{\cdot\cdot}), \tag{9}$$

which is the same as the typical element in (7).

(b) It is easy to show that the order of these two within one-way transformations is unimportant. Note that

$$Q_1 Q_2 = (I_N \otimes E_T)(E_N \otimes I_T) = E_N \otimes E_T = Q_2 Q_1. \tag{10}$$

Thus the order of these two transformations is unimportant.

For the regression, these two within one-way transformations remove the constant, time effects, and individual effects. Also, for each variable in X, the transformations remove the average over time and the average over individual and add the overall average. The order of these operations is not important.

NOTE

1. Excellent solutions have been proposed independently by B.C. Jung and S.H. Song, and by Badi H. Baltagi, the poser of the problem.

https://doi.org/10.1017/S0266466699155075 Published online by Cambridge University Press

REFERENCES

Baltagi, B.H. (1995) *Econometric Analysis of Panel Data*. New York: Wiley.
Wallace, T.D. & A. Hussain (1969) The use of error components models in combining cross section with time series data. *Econometrica* 37, 55–72.

98.5.3. *Contrast of the Asymptotic Properties of Least Squares Estimates in Symmetric Seasonal Processes*—Solution, proposed by Paulo M.M. Rodrigues. If $\{y_t\}_1^T$ is generated from (1.1) with $y_{-1} = y_0 = 0$, $\phi_1 = 0$, and $\phi_2 = -1$, then

$$y_t = \begin{cases} S_{1,n} & t \bmod 2 = 1 \\ S_{0,n} & t \bmod 2 = 0, \end{cases} \qquad y_{t-1} = \begin{cases} S_{0,n-1} & t \bmod 2 = 1 \\ S_{1,n} & t \bmod 2 = 0 \end{cases}$$

and

$$y_{t-2} = \begin{cases} S_{1,n-1} & t \bmod 2 = 1 \\ S_{0,n-1} & t \bmod 2 = 0, \end{cases} \tag{2.1}$$

where $n = [(t-1)/2] + 1$, $[\cdot]$ represents the greatest integer equal or less than $(t-1)/2$ and S_{in} is a partial sum process defined by

$$S_{in} = (-1)\left(\sum_{i=1}^{[(t-1)/2]} (-1)^{i+1} e_{t-2i}\right) + e_t \tag{2.2}$$

or more concisely as

$$S_{in} = (-1)S_{in-1} + e_t. \tag{2.3}$$

From the functional central limit theorem,

$$\frac{1}{\sigma\sqrt{N}} S_{in} \Rightarrow W_i(r), \tag{2.4}$$

where $W_i(r)$ is a standard Wiener process and $N = T/2$. Thus, from the continuous mapping theorem, (2.1) and (2.3) imply the following convergence results:

$$\frac{1}{T^{1/2}} \sum_{t=1}^{T} y_{t-1} \Rightarrow \frac{\sigma}{\sqrt{2}} \int_0^1 \underset{\sim}{\mathbf{1}}' \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix} \mathbf{W}(\mathbf{r})\, dr, \tag{2.5}$$

$$\frac{1}{T^{1/2}} \sum_{t=1}^{T} y_{t-2} \Rightarrow \frac{\sigma}{\sqrt{2}} \int_0^1 \underset{\sim}{\mathbf{1}}' \mathbf{W}(\mathbf{r})\, dr, \tag{2.6}$$

$$\frac{1}{T^2} \sum_{t=1}^{T} y_{t-i}^2 \Rightarrow \frac{\sigma^2}{4} \int_0^1 \mathbf{W}(\mathbf{r})' \mathbf{W}(\mathbf{r})\, dr, \qquad i = 1,2, \tag{2.7}$$

$$\frac{1}{T^2} \sum_{t=1}^{T} y_{t-1} y_{t-2} \to 0 \quad \text{a.s.}, \tag{2.8}$$

$$\frac{1}{T} \sum_{t=1}^{T} y_{t-1} e_t \Rightarrow \frac{\sigma^2}{2} \int_0^1 \mathbf{W}(\mathbf{r})' \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix} d\mathbf{W}(\mathbf{r}), \tag{2.9}$$

$$\frac{1}{T} \sum_{t=1}^{T} y_{t-2} e_t \Rightarrow \frac{\sigma^2}{2} \int_0^1 \mathbf{W}(\mathbf{r})'\, d\mathbf{W}(\mathbf{r}), \tag{2.10}$$

https://doi.org/10.1017/S0266466699155075 Published online by Cambridge University Press

where $\Rightarrow$ denotes weak convergence, $\rightarrow$ indicates convergence in probability, $\underset{\sim}{\mathbf{1}} = (1,1)'$, and $\mathbf{W}(\mathbf{r})$ is the vector of standard Wiener processes $\mathbf{W}(\mathbf{r}) = (W_1(r), W_0(r))'$. Consequently, the normalized least squares biases

$$T\begin{bmatrix}\hat{\phi}_1 - \phi_1 \\ \hat{\phi}_2 - \phi_2\end{bmatrix} = \begin{bmatrix}\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}^2 & \frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}y_{t-2} \\ \frac{1}{T^2}\sum_{t=1}^{T} y_{t-2}y_{t-1} & \frac{1}{T^2}\sum_{t=1}^{T} y_{t-2}^2\end{bmatrix}^{-1}\begin{bmatrix}\frac{1}{T}\sum_{t=1}^{T} y_{t-1}e_t \\ \frac{1}{T}\sum_{t=1}^{T} y_{t-2}e_t\end{bmatrix} \tag{2.11}$$

converge to

$$T\begin{bmatrix}\hat{\phi}_1 - \phi_1 \\ \hat{\phi}_2 - \phi_2\end{bmatrix} \Rightarrow \begin{bmatrix}\frac{\sigma^2}{4}\int_0^1 \mathbf{W}(\mathbf{r})'\mathbf{W}(\mathbf{r})\,dr & 0 \\ 0 & \frac{\sigma^2}{4}\int_0^1 \mathbf{W}(\mathbf{r})'\mathbf{W}(\mathbf{r})\,dr\end{bmatrix}^{-1} \times \begin{bmatrix}\frac{\sigma^2}{2}\int_0^1 \mathbf{W}(\mathbf{r})'\begin{bmatrix}0 & -1 \\ 1 & 0\end{bmatrix} d\mathbf{W}(\mathbf{r}) \\ \frac{\sigma^2}{2}\int_0^1 \mathbf{W}(\mathbf{r})'\,d\mathbf{W}(\mathbf{r})\end{bmatrix}.$$

Thus,

$$T(\hat{\phi}_1 - \phi_1) \Rightarrow \frac{2\int_0^1 \mathbf{W}(\mathbf{r})'\begin{bmatrix}0 & -1 \\ 1 & 0\end{bmatrix} d\mathbf{W}(\mathbf{r})}{\int_0^1 \mathbf{W}(\mathbf{r})'\mathbf{W}(\mathbf{r})\,dr}, \tag{2.12}$$

$$T(\hat{\phi}_2 - \phi_2) \Rightarrow \frac{2\int_0^1 \mathbf{W}(\mathbf{r})'\,d\mathbf{W}(\mathbf{r})}{\int_0^1 \mathbf{W}(\mathbf{r})'\mathbf{W}(\mathbf{r})\,dr}. \tag{2.13}$$

On the other hand, if $\phi_1 = 0$ and $\phi_2 = 1$, (2.1) continues to apply except that now,

$$S_{in} = \sum_{i=1}^{[(t-1)/2]} e_{t-2i} + e_t \tag{2.14}$$

and

$$S_{in} = S_{in-1} + e_t. \tag{2.15}$$

https://doi.org/10.1017/S0266466699155075 Published online by Cambridge University Press

Consequently, this will imply some changes to the results presented in (2.5)–(2.10). In particular, (2.5), (2.8), and (2.9) will become

$$\frac{1}{T^{1/2}}\sum_{t=1}^{T} y_{t-1} \Rightarrow \frac{\sigma}{\sqrt{2}}\int_0^1 \underset{\sim}{\mathbf{1}}'\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\mathbf{W}(\mathbf{r})\,dr, \tag{2.16}$$

$$\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}y_{t-2} \Rightarrow \frac{\sigma^2}{4}\int_0^1 \mathbf{W}(\mathbf{r})'\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\mathbf{W}(\mathbf{r})\,dr, \tag{2.17}$$

$$\frac{1}{T}\sum_{t=1}^{T} y_{t-1}e_t \Rightarrow \frac{\sigma^2}{2}\int_0^1 \mathbf{W}(\mathbf{r})'\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}d\mathbf{W}(\mathbf{r}). \tag{2.18}$$

It is the result presented in (2.17) that determines that the covariance matrix will not be diagonal and therefore the regressors will not be orthogonal when $\phi_1 = 0$ and $\phi_2 = 1$.

Proof of (2.8) and (2.17). Under the assumption that y_t is generated from (1.1) with $y_{-1} = y_0 = 0$, $\phi_1 = 0$, and $\phi_2 = -1$, we can see from (2.1) and (2.3) that

$$\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}y_{t-2} = \frac{1}{4N^2}\left\{\sum_{n=1}^{N}\begin{bmatrix}S_{1,n-1}\\ S_{0,n-1}\end{bmatrix}'\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}\begin{bmatrix}S_{1,n-1}\\ S_{0,n-1}\end{bmatrix} + \sum_{n=1}^{N}\begin{bmatrix}e_{1,n}\\ 0\end{bmatrix}'\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}S_{1,n-1}\\ S_{0,n-1}\end{bmatrix}\right\}. \tag{2.19}$$

Notice that

$$e_t = \begin{cases} e_{1,n} & t \bmod 2 = 1\\ e_{0,n} & t \bmod 2 = 0\end{cases},$$

because

$$\frac{1}{N^2}\sum_{n=1}^{N}\begin{bmatrix}e_{1,n}\\ 0\end{bmatrix}'\begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}\begin{bmatrix}S_{1,n-1}\\ S_{0,n-1}\end{bmatrix} = o_p(1). \tag{2.20}$$

Thus as $N \to \infty$,

$$\frac{1}{T^2}\sum_{t=1}^{T} y_{t-1}y_{t-2} \to 0 \quad \text{a.s.}$$

The proof of (2.17) follows similar lines except that the matrix

$$\begin{bmatrix}0 & -1\\ 1 & 0\end{bmatrix}$$

https://doi.org/10.1017/S0266466699155075 Published online by Cambridge University Press

considered in (2.19) has to be modified to

$$\begin{bmatrix} 0 & 1 \\ 1 & 0 \end{bmatrix},$$

which then delivers the result given in (2.17). ■

98.5.4. *Further Examples of Accelerated Time Regression Models That Are Not Proportional Hazard Models*—Solution, proposed by S.K. Sapra.

Model 1 (Pareto model). Assume $s < a$. The moment generating function (m.g.f.) of $\log t$ is

$$M_{\log t}(s) = E(e^{s \log t}) = E(t^s) = ak^s/(a - s). \tag{3}$$

Therefore, the cumulant generating function (c.g.f.) of $\log t$ is

$$K_{\log t}(s) = \log a + s \log k - \log(a - s). \tag{4}$$

Equation (4) yields

$$K'(s) = \log k + (a - s)^{-1} \tag{5}$$

and

$$K''(s) = (a - s)^{-2}, \tag{6}$$

where the primes denote the derivatives.

Hence,

$$E(\log t) = K'(0) = \log k + a^{-1}, \tag{7}$$

$$\text{Var}(\log t) = K''(0) = a^{-2}. \tag{8}$$

Using the parametrization $k = e^{\beta' x}$, equation (7) yields the regression-like equation

$$\log t = \beta' x + u, \tag{9}$$

where u has mean a^{-1} and variance a^{-2} and is therefore homoskedastic. Hence, model (1) is an accelerated failure time regression model.

To show that model (1) is not a proportional hazard model, note that the hazard function for model (1) is

$$h(t) = a/t = h_0(t) \neq h_0(t)e^{\beta' x} \quad \text{for } x \neq 0,$$

$h_0(t)$ being the baseline hazard for which $x = 0$.

Model 2 (inverse Gaussian model). Using formulas in Johnson and Kotz (1970, p. 139), the m.g.f. and c.g.f. of $\log t$ are easily derived as

$$M_{\log t}(s) = m(s)e^{1/\mu}\mu^{2s}, \tag{10}$$

where $m(s) = (-1)^s 1.3.5.7 \ldots (2s - 3)$,

$$K_{\log t}(s) = \log m(s) + 1/\mu + 2s \log \mu. \tag{11}$$

https://doi.org/10.1017/S0266466699155075 Published online by Cambridge University Press

Therefore,

$$E(\log t) = K'_{\log t}(0) = m'(0)/m(0) + 2\log\mu, \tag{12}$$

$$\mathrm{Var}(\log t) = K''_{\log t}(0) = [m''(0)m(0) - (m'(0))^2]/(m(0))^2. \tag{13}$$

Thus, for $\mu = e^{\beta' x}$, equation (12) leads to a regression like equation (9) with a homoskedastic error U with variance in equation (13). Hence, model (2) is an accelerated failure time model.

To show that model (2) is not a proportional hazard model, we note that the survivor function is

$$S(t|x) = 1 - \Phi(t - \mu/\mu\sqrt{t}) - e^{2/\mu}(1 - \Phi(t + \mu/\mu\sqrt{t})), \qquad \mu = e^{\beta' x}, \tag{14}$$

$\Phi(\cdot)$ being the cumulative distribution function of a standard normal random variable. It follows from (14) that

$$S(t|x) \neq [S_0(t|x)]^{e^{\beta' x}}, \tag{15}$$

where $S_0(t|x) = S(t|x)$ for $x = 0$. Hence, model (2) is not a proportional hazard model for which equality holds in (15).

REFERENCE

Johnson, N.L. & S. Kotz (1970) *Continuous Univariate Distributions—1*. New York: Wiley.

98.5.5. *A Property of the Companion Matrix of a Reduced VAR*—Solution, proposed by Farshid Vahid. We show that F has at least $p(k-r)$ zero eigenvalues by showing that rank of F^p is at most pr. Consider F^2:

$$F^2 = \begin{pmatrix} \Phi \\ I_{pk-k} \quad \vdots 0_{(pk-k)\times k} \end{pmatrix} F = \begin{pmatrix} \Phi F \\ \begin{pmatrix} \text{matrix } F \text{ without} \\ \text{its last } k \text{ rows} \end{pmatrix} \end{pmatrix}$$

$$= \begin{pmatrix} \Phi F \\ \Phi \\ I_{pk-2k} \quad \vdots 0_{(pk-2k)\times 2k} \end{pmatrix}.$$

Similarly,

$$F^3 = \begin{pmatrix} \Phi \\ I_{pk-k} \quad \vdots 0_{(pk-k)\times k} \end{pmatrix} F^2 = \begin{pmatrix} \Phi F^2 \\ \begin{pmatrix} \text{matrix } F^2 \text{ without} \\ \text{its last } k \text{ rows} \end{pmatrix} \end{pmatrix}$$

$$= \begin{pmatrix} \Phi F^2 \\ \Phi F \\ \Phi \\ I_{pk-3k} \quad \vdots 0_{(pk-3k)\times 3k} \end{pmatrix}.$$

https://doi.org/10.1017/S0266466699155075 Published online by Cambridge University Press

Hence it becomes apparent that F^p will be

$$F^p = \begin{pmatrix} \Phi F^{p-1} \\ \Phi F^{p-2} \\ \vdots \\ \Phi F \\ \Phi \end{pmatrix}.$$

Because Φ has rank r, ΦF^j has at most rank r for $j = 1, \ldots, p-1$, and therefore there are at most only r linearly independent rows in the jth block of k rows of F^p for all $j = 1, \ldots, p$, which implies that the rank of F^p is at most pr. Therefore, F^p has at least $p(k-r)$ zero eigenvalues. Because the eigenvalues of F^p are the pth power of the eigenvalues of F, we have thus shown that F must have at least $p(k-r)$ zero eigenvalues.

https://doi.org/10.1017/S0266466699155075 Published online by Cambridge University Press