Econometric Theory, **14**, 1998, 687–698. Printed in the United States of America.

PROBLEMS AND SOLUTIONS

PROBLEMS

98.5.1. *A Useful Result for Lipschitz Functions of Mixingales*, proposed by Halbert White. Provide a proof of the following proposition, useful for showing that Lipschitz functions of L_1-mixingales are themselves L_1-mixingales. How does this extend to the L_p case?

PROPOSITION. *Let $U \subset \mathbb{R}$ and let $g : U \to \mathbb{R}$ be Lipschitz. Let $(\Omega, \mathcal{F}, P)$ be a complete probability space and let $X : \Omega \to U$ be measurable such that $E(X) < \infty$ and $E(g(X)) < \infty$. Let $\mathcal{G}$ be a sub σ-field of $\mathcal{F}$. Then,*

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

with $0 \leq \Delta < \infty$ independent of $\mathcal{G}$.

98.5.2. *Within Two Way Is Equivalent to Two Withins One Way*, proposed by Badi H. Baltagi. Consider the fixed-effects two-way error component panel data model

$$y_{it} = \alpha + X_{it}'\beta + \mu_i + \lambda_t + \nu_{it} \qquad i = 1,2,\ldots,N; \qquad t = 1,2,\ldots,T,$$

where α is a scalar, X_{it}' is $1 \times K$, β is $K \times 1$, μ_i and λ_t denote fixed individual and time effects, respectively, and $\nu_{it} \sim$ i.i.d.$(0, \sigma_\nu^2)$.

(a) Show that the within estimator of β, which is best linear unbiased, can be obtained by applying two within (one-way) transformations on this model. The first is the within transformation ignoring the time effects followed by the within transformation ignoring the individual effects.
(b) Show that the order of these two within (one-way) transformations is unimportant. Give an intuitive explanation for this result.

98.5.3. *Contrast of the Asymptotic Properties of Least Squares Estimates in Symmetric Seasonal Processes*, proposed by Paulo M.M. Rodrigues. Consider the process

$$y_t = \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t, \tag{1.1}$$

where $y_{-1} = y_0 = 0$, $e_t \sim$ i.i.d.$(0, \sigma^2)$ and e_t satisfies Assumption 1 in Phillips (1986).

Show that if $\phi_1 = 0$ and $\phi_2 = -1$, the least squares estimates of ϕ_1 and ϕ_2 are asymptotically independent, whereas if $\phi_1 = 0$ and $\phi_2 = 1$, they are not. (The

© 1998 Cambridge University Press 0266-4666/98 $9.50

https://doi.org/10.1017/S0266466698145085 Published online by Cambridge University Press

former has a pair of complex conjugate roots on the unit circle, whereas the latter contains the two real unit roots ± 1. These properties are important in the context of seasonal unit roots.)

REFERENCE

Phillips, P.C.B. (1986) Understanding spurious regressions in econometrics. *Journal of Econometrics* 33, 311–340.

98.5.4. *Further Examples of Accelerated Time Regression Models That Are not Proportional Hazard Models*, proposed by S.K. Sapra. Consider the following duration data models with density functions given below.

1. Pareto Model

$$f(t) = ak^a/t^{a+1}, \qquad a > 0, \qquad t \geq k > 0. \tag{1}$$

2. Inverse Gaussian Model

$$f(t) = (2\pi t^3)^{-1/2} e^{-(t-\mu)^2/2\mu^2 t}, \qquad t > 0, \qquad \mu < 0. \tag{2}$$

Using suitable parameterizations, show that models (1) and (2) are accelerated failure time regression models but are not proportional hazard models (for definitions of these models, see Kalbfleisch and Prentice, 1980).

REFERENCE

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

98.5.5. *A Property of the Companion Matrix of a Reduced Rank VAR*, proposed by Farshid Vahid. Consider the following mean-subtracted k-dimensional VAR(p),

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-1} + \cdots + \Phi_p y_{t-p} + \varepsilon_t$$

in which rank of the $(k \times kp)$ parameter matrix $\Phi \equiv [\Phi_1 \vdots \Phi_2 \vdots \cdots \vdots \Phi_p]$ is $r < k$. A useful reformulation of a VAR is its kp-dimensional VAR(1) form:

$$z_t = Fz_{t-1} + \eta_t,$$

where F is the $(kp \times kp)$ companion matrix

$$F = \begin{pmatrix} & \Phi \\ I_{kp-k} & \vdots 0_{(kp-k)\times k} \end{pmatrix},$$

https://doi.org/10.1017/S0266466698145085 Published online by Cambridge University Press

and $z_t = (y_t' \vdots y_{t-1}' \vdots \cdots \vdots y_{t-p+1}')'$ and $\eta_t = (\varepsilon_t' \vdots 0_{1\times(kp-k)})'$. Show that F has at least $p(k-r)$ zero eigenvalues. (For introduction to reduced rank VARs, see Lütkepohl, 1991; Reinsel, 1993).

REFERENCES

Lütkepohl, H. (1991) *Introduction to Multiple Time Series Analysis.* New York: Springer-Verlag.
Reinsel, G.C. (1993) *Elements of Multivariate Time Series Analysis.* New York: Springer-Verlag.

SOLUTIONS

97.5.1. *Estimation of Time-Series Regressions with Autoregressive Disturbances and Missing Observations*—Solutions.[1] The following solutions have been proposed independently by José A.F. Machado and by J. Roderick McCrorie and Jeremy Penzer. Each contains an interesting derivation.

1. Solution, proposed by José A.F. Machado. Standard backward substitution reveals that the AR(1) model for the disturbances can be written at the observations times as

$$u_{t_j} = \rho^{t_j - t_{(j-1)}} u_{t_{(j-1)}} + \xi_{t_j}, \tag{1}$$

with

$$\xi_{t_j} = \epsilon_{t_j} + \rho\epsilon_{t_j-1} + \cdots + \rho^{(t_j - t_{(j-1)})-1}\epsilon_{t_{(j-1)}+1} \tag{2}$$

for $1 = t_1 < t_2 < \cdots < t_n = T$.

Because ξ_{t_j} and ξ_{t_l} for $j \neq l$ depend on different sets of ϵ's, it is clear that

$$\operatorname{cov}(\xi_{t_j}, \xi_{t_l}) = 0 \quad \text{for } j \neq l. \tag{3}$$

However, the ξ_{t_j} are heteroskedastic. Indeed,

$$\operatorname{var}(\xi_{t_j}) = \frac{1 - \rho^{2(t_j - t_{(j-1)})}}{1 - \rho^2}\,\sigma_\epsilon^2. \tag{4}$$

Quasi-differencing the regression model at consecutive observation times yields

$$(y_{t_j} - \rho^{t_j - t_{(j-1)}} y_{t_{(j-1)}}) = (x_{t_j} - \rho^{t_j - t_{(j-1)}} x_{t_{(j-1)}})'\beta + \xi_{t_j}, \qquad j = 2, \ldots, n. \tag{5}$$

The disturbances of (5) are uncorrelated but heteroskedastic. Correcting now for heteroskedasticity yields the transformed model

$$\tilde{y}_{t_j} = \tilde{x}_{t_j}'\beta + \tilde{\xi}_{t_j}, \qquad j = 1, \ldots, n, \tag{6}$$

https://doi.org/10.1017/S0266466698145085 Published online by Cambridge University Press

where

$$\tilde{y}_{t_1} = y_{t_1}$$

$$\tilde{x}_{t_1} = x_{t_1}$$

and, for $j = 2,\ldots,n$,

$$\tilde{\xi}_{t_1} = u_{t_1}$$

$$\tilde{y}_{t_j} = (y_{t_j} - \rho^{t_j - t_{(j-1)}} y_{t_{(j-1)}})/(1 - \rho^{2(t_j - t_{(j-1)})})^{1/2}$$

$$\tilde{x}_{t_j} = (x_{t_j} - \rho^{t_j - t_{(j-1)}} x_{t_{(j-1)}})/(1 - \rho^{2(t_j - t_{(j-1)})})^{1/2}$$

$$\tilde{\xi}_{t_j} = \xi_{t_j}/(1 - \rho^{2(t_j - t_{(j-1)})})^{1/2}$$

GLS on the original regression may be obtained as a simple OLS on (6).

2. Solution, proposed by J. Roderick McCrorie and Jeremy Penzer. Rewrite equations (1) and (2) in matrix notation as

$$y = X\beta + u \quad \text{and} \quad u = \rho u_{-1} + e.$$

Let $\Omega = [\omega_{ts}]_{T\times T}$, where

$$\omega_{ts} = \frac{\sigma_\varepsilon^2}{1-\rho^2}\,\rho^{|t-s|} \qquad (t,s = 1,\ldots,T).$$

The missing data case is then obtained by applying the transformation represented by the matrix J, the $T \times T$ identity matrix with row t deleted if observation t is missing. The matrix $J\Omega J'$ (which agrees with the expression stated by Shively, 1993, p. 235) is reduced to row-echelon form by a sequence of elementary row operations represented by the lower triangular matrix

$$D = \begin{bmatrix} 1 & & & & \\ -\rho^{(t_2-t_1)} & 1 & & & \\ 0 & -\rho^{(t_3-t_2)} & 1 & & \\ . & 0 & . & . & \\ 0 & . & 0 & -\rho^{(t_n-t_{n-1})} & 1 \end{bmatrix},$$

which is a special case of the general transformation for ARMA models stated by Penzer and Shea (1997, equation 6). Then,

$$\mathrm{var}(DJu) = \frac{\sigma_\varepsilon^2}{1-\rho^2}\,\mathrm{diag}(1, 1-\rho^{2(t_2-t_1)}, 1-\rho^{2(t_3-t_2)}, \ldots, 1-\rho^{2(t_n-t_{n-1})}).$$

To reduce weighted least squares to ordinary least squares, as is required for a Prais–Winsten–type transformation, we apply the transformation represented by the diagonal matrix

$$H = \mathrm{var}(DJu)^{-1/2},$$

https://doi.org/10.1017/S0266466698145085 Published online by Cambridge University Press

whose typical diagonal element is just the reciprocal of the square root of the corresponding element in var(DJu). Indeed, because H is diagonal and ρ is known, feasible GLS estimates can be obtained recursively from the transformed standard-case regression:

$$HDJy = HDJX\beta + HDJu,$$

or, changing notation in an obvious way,

$$y^* = X^*\beta + u^*.$$

By construction, this is ordinary least squares regression because

$$\hat{\beta} = [(X'J'D'H')I_n(HDJX)]^{-1}(X'J'D'H')I_n(HDJy) = (X^{*\prime}X^*)^{-1}X^{*\prime}y^*.$$

Applying the above transformations successively obviates the need to find the inverse of the square root of a product of different diagonal matrices, as would be required when using elementary row operations alone.

NOTE

1. An excellent solution has been proposed independently by Badi H. Baltagi and Ping X. Wu, the posers of the problem.

REFERENCES

Penzer, J. & B. Shea (1997) The exact likelihood of an autoregressive-moving average model with incomplete data. *Biometrika* 84, 919–928.

Shively, T.S. (1993) Testing for autoregressive disturbances in a time series regression with missing observations. *Journal of Econometrics* 57, 233–255.

97.5.2. *The Relationship Between Forward and Backward Representations of the Stationary VAR Models*—Solution, proposed by Jae H. Kim. The forward VAR(p) model (1) can be written in the Kp-dimensional mean-adjusted VAR(1) form as

$$W_t = \Pi W_{t-1} + U_t, \tag{3}$$

where W_t and U_t are $Kp \times 1$ vectors defined as

$$W_t = (w_t', \ldots, w_{t-p+1}')' \text{ with } w_t = y_t - E(y_t); \quad \text{and}$$

$$U_t = (u_t', 0, \ldots, 0)' \text{ with } E(U_tU_t') = \Sigma_U,$$

and Π is a $Kp \times Kp$ matrix of the form

$$\Pi = \left[\begin{array}{c|c} A_* & A_p \\ \hline I & \mathbf{0} \end{array}\right],$$

https://doi.org/10.1017/S0266466698145085 Published online by Cambridge University Press

where I is a $K(p-1)$ identity matrix, $\mathbf{0}$ is a $K(p-1)\times K$ null matrix, and $A_* = [A_1,\ldots,A_{p-1}]$. By postmultiplying W'_{t-1} on both sides of (3) and taking expectations, we obtain

$$\Pi = \Gamma_p(1)(1)\Gamma_p(0)^{-1}, \tag{4}$$

where

$$\Gamma_p(j) = E(W_t W'_{t-j}) = \begin{bmatrix} \Gamma(j) & \Gamma(j+1) & \ldots & \Gamma(j+p-1) \\ \Gamma(j-1) & \Gamma(j) & \ldots & \Gamma(j+p-2) \\ \vdots & \vdots & & \vdots \\ \Gamma(j-p+1) & \Gamma(j-p+2) & \ldots & \Gamma(j) \end{bmatrix}_{(Kp\times Kp)};$$

and

$$\Gamma(j) = E(w_t w'_{t-j}).$$

The backward VAR(p) model (2) can be expressed in the VAR(1) form as

$$W_t = \Omega W_{t+1} + V_t, \tag{5}$$

where $Kp\times 1$ vector V_t and $Kp\times Kp$ matrix Ω is defined as

$$V_t = (0',\ldots,0',\nu'_{t-p+1})' \text{ with } E(V_t V'_t) = \Sigma_V; \quad \text{and}$$

$$\Omega = \left[\begin{array}{c|c} \mathbf{0} & I \\ \hline H_p & H_* \end{array}\right] \text{ with } H_* = [H_{p-1},\ldots,H_1].$$

By postmultiplying W'_{t+1} on both sides of (5) and taking expectations, we obtain

$$\Omega = \Gamma_p(-1)\Gamma_p(0)^{-1}. \tag{6}$$

By combining (4) and (6), we have

$$\Omega' = \Gamma_p(0)^{-1}\Pi\Gamma_p(0); \quad \text{of} \quad \Pi = \Gamma_p(0)\Omega'\Gamma_p(0)^{-1}. \tag{7}$$

The relation given in (7) indicates that Π and Ω' are similar matrices, which share the identical characteristic polynomials (Horn and Johnson, 1985, p. 45), so that

$$\begin{aligned}\det(I_{Kp} - \Pi z) &= \det(I_K - A_1 z - \cdots - A_p z^p) = \det(I_{Kp} - \Omega' z) \\ &= \det(I_{Kp} - \Omega z) = \det(I_K - H_1 z - \cdots - H_p z^p).\end{aligned}$$

It can be seen that the variance-covariance structures of forward and backward innovations in the VAR(1) form are related as

$$\Sigma_U = \Gamma_p(0) - \Pi\Gamma_p(0)\Pi' = (I_{Kp} - \Pi\Omega)\Gamma_p(0); \quad \text{and} \tag{8}$$

$$\Sigma_V = \Gamma_p(0) - \Omega\Gamma_p(0)\Omega' = (I_{Kp} - \Omega\Pi)\Gamma_p(0). \tag{9}$$

https://doi.org/10.1017/S0266466698145085 Published online by Cambridge University Press

Consider

$$I_{Kp} - \Pi\Omega = \begin{bmatrix} T_1 & T_2 & \cdots & T_p \\ & 0_{(K-1)p\times Kp} & & \end{bmatrix} \quad \text{and}$$

$$(I_{Kp} - \Omega\Pi) = \begin{bmatrix} & 0_{(K-1)p\times Kp} & & \\ S_p & S_{p-1} & \cdots & S_1 \end{bmatrix},$$

where $T_i = I_K - (A_p H_{p-i+1} + A_{i-1})$ and $S_i = I_K - (H_p A_{p-i+1} + H_{i-1})$, with $A_0 = H_0 = O_{K\times K}$. From (8), it can be seen that

$$\begin{bmatrix} \Sigma_u & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} T_1 & T_2 & \cdots & T_p \\ & 0_{(K-1)p\times Kp} & & \end{bmatrix} \Gamma_p(0).$$

Define the matrix P to be a $Kp \times Kp$ matrix, with the first K rows of $\Gamma_p(0)$ being replaced by the first K rows of Σ_U. That is,

$$P = \begin{bmatrix} \Sigma_u & 0 & \cdots & 0 \\ & & P_1 & \end{bmatrix}; \qquad P_1 = \begin{bmatrix} \Gamma(-1) & \\ \vdots & \Gamma_{p-1}(0) \\ \Gamma(-p+1) & \end{bmatrix}_{K(p-1)\times Kp}.$$

Then, P can be written as

$$P = \begin{bmatrix} T_1 & T_2 & \cdots & T_p \\ 0_{K(p-1)\times K} & & I_{K(p-1)} & \end{bmatrix} \Gamma_p(0).$$

Similarly, using the relation in (9), define the matrix Q to be a $Kp \times Kp$ matrix, with the last K rows of $\Gamma_p(0)$ being replaced by the last K rows of Σ_V. Then, Q can be written as

$$Q = \begin{bmatrix} & I_{K(p-1)} & 0_{K(p-1)\times K} & \\ S_p & S_{p-1} & \cdots & S_1 \end{bmatrix} \Gamma_p(0).$$

By evaluating the determinants of P and Q, it can be seen that

$$\det(P) = \det(\Sigma_u)\det[\Gamma_{p-1}(0)] = \det(T_1)\det[\Gamma_p(0)]; \quad \text{and}$$

$$\det(Q) = \det(\Sigma_V)\det[\Gamma_{p-1}(0)] = \det(S_1)\det[\Gamma_p(0)].$$

The proof concludes as $\det(T_1) = \det(S_1)$, by rule (6) in A.10 of Lütkepohl (1991, p. 463).

REFERENCES

Horn, R.A. & C.R. Johnson (1985) *Matrix Analysis.* Cambridge: Cambridge University Press.
Lütkepohl, H. (1991) *Introduction to Multiple Time Series Analysis.* New York: Springer-Verlag.

97.5.3. *A Fundamental Matrix Result on Scaling in Multivariate Analysis*—Solutions.[1] The following solutions have been proposed independently by Jan Graffelman and by Simo Puntanen and George P.H. Styan. Each contains an interesting derivation.

https://doi.org/10.1017/S0266466698145085 Published online by Cambridge University Press

1. Solution, proposed by Jan Graffelman. Consider the singular value decomposition of T as $T = UDV'$, with $U'U = I$ and $V'V = I$. The singular values of matrix T are the square roots of the eigenvalues of matrix $T'T$ because $T'T = VDU'UDV' = VD^2V'$. Bounds on the eigenvalues therefore imply bounds on the singular values.

Because $T'T$ is a nonnegative definite matrix, all its eigenvalues are nonnegative, and thus the singular values of T are also nonnegative.

Bounds for the largest eigenvalue λ_F of a nonnegative matrix $A = T'T$ can be obtained by applying Gershgorin's theorem (see Barbolla and Sanz, 1998, pp. 334–340):

$$\min_i \left\{ \sum_{j=1}^{p} a_{ij} \right\} \leq \lambda_F \leq \max_i \left\{ \sum_{j=1}^{p} a_{ij} \right\}.$$

Note that $A = T'T = D_c^{-1/2} F' D_r^{-1} F D_c^{-1/2}$ has the same eigenvalues as $B = D_c^{-1} F' D_r^{-1} F$ because $D_c^{-1/2} F' D_r^{-1} F D_c^{-1/2} v = \lambda v$ implies $D_c^{-1} F' D_r^{-1} F D_c^{-1/2} v = \lambda D_c^{-1/2} v$ and thus $D_c^{-1} F' D_r^{-1} F u = \lambda u$, where $u = D_c^{-1/2} v$.

The rows of B sum to 1 because $D_c^{-1} F' D_r^{-1} F 1_p = D_c^{-1} F' D_r^{-1} D_r 1_n = D_c^{-1} F' 1_n = D_c^{-1} D_c 1_p = 1_p$. By applying Gershgorin's theorem, we obtain bounds for the eigenvalues of B:

$$\min_i \left\{ \sum_{j=1}^{p} b_{ij} \right\} = \max_i \left\{ \sum_{j=1}^{p} b_{ij} \right\} = 1.$$

Thus, $\lambda_F = 1$, which shows that the matrix B and thus A, always has an eigenvalue of 1 and that this is the largest one. Consequently, the singular values of T lie all in the closed interval $[0,1]$, and there will always be a singular value of 1.

This fundamental matrix result is illustrated by the fact that the principal inertias in a correspondence analysis are in the $[0,1]$ interval, which is evident because they are squared canonical correlations. The eigenvalue 1 corresponds to a trivial dimension in the correspondence analysis solution, when data have not been centered (for more details, see Greenacre, 1984, pp. 108–116).

2. Solution, proposed by Simo Puntanen and George P.H. Styan. We assume that the nonnegative data matrix F does not have any rows or columns that are completely zero, and thus both D_r and D_c are positive definite. We show that the singular values of the matrix $T = D_r^{-1/2} F D_c^{-1/2}$ lie in the closed interval $[0,1]$ with at least one singular value equal to $+1$.

The nonzero singular values of the matrix $T = D_r^{-1/2} F D_c^{-1/2}$ are defined as the positive square roots of the positive eigenvalues of the nonnegative definite matrix $TT' = D_r^{-1/2} F D_c^{-1} F' D_r^{-1/2}$. Because the eigenvalues of AB are identical to the eigenvalues of BA, whenever A and B are conformable square matrices, it follows at once that the eigenvalues of the matrix TT' are identical to the eigenvalues of $D_r^{-1} F D_c^{-1} F'$. Because, using (1),

$$D_r^{-1} F D_c^{-1} F' 1_n = D_r^{-1} F D_c^{-1} D_c 1_p = D_r^{-1} F 1_p = D_r^{-1} D_r 1_n = 1_n, \qquad \textbf{(2)}$$

https://doi.org/10.1017/S0266466698145085 Published online by Cambridge University Press

we see that all the row totals of $D_r^{-1}FD_c^{-1}F'$ are equal to 1; and because, in addition, all its elements are nonnegative, it follows that $D_r^{-1}FD_c^{-1}F'$ is row stochastic. It is well known (see Marcus and Minc, 1992, Sect. 5.13.3, p. 133) that the eigenvalues of a row-stochastic matrix all lie in the unit disc in the complex plane with at least one eigenvalue equal to $+1$. Here, however, all the eigenvalues are real and nonnegative and thus must lie in the closed interval $[0,1]$, with at least one eigenvalue equal to $+1$. It follows at once that the singular values of $T = D_r^{-1/2}FD_c^{-1/2}$ lie in the closed interval $[0,1]$, with at least one singular value equal to $+1$, and our proof is complete.

NOTE

1. An excellent solution has been proposed independently by Heinz Neudecker, Albert Satorra, and Michel van de Velden, the posers of the problem.

REFERENCES

Barbolla, R. & P. Sanz (1998) *Álgebra lineal y teoría de matrices*. Madrid: Prentice-Hall.

Greenacre, M. (1984). *Theory and Applications of Correspondence Analysis.* Orlando: Academic Press.

Marcus, M. & H. Minc (1992) *A Survey of Matrix Theory and Matrix Inequalities*, reprint. New York: Dover.

97.5.4. *Mahalanobis Distance for Multinomial Data*—Solution,[1] proposed by Simo Puntanen, George P.H. Styan, and Gerald E. Subak-Sharpe. We assume that all the elements of $\mathbf{a}$ are positive so that $\mathbf{A}$ is positive definite. We establish the more general result that

$$\mu = n_i n_j (n_i + n_j)^{-1} (\mathbf{x}_i - \mathbf{x}_j)' \boldsymbol{\Sigma}^{-} (\mathbf{x}_i - \mathbf{x}_j) \tag{3}$$

for *any* generalized inverse $\boldsymbol{\Sigma}^{-}$. Then (3) implies (2) because $\mathbf{A}^{-1}$ is a generalized inverse of $\boldsymbol{\Sigma}$,

$$\begin{aligned} \boldsymbol{\Sigma}\mathbf{A}^{-1}\boldsymbol{\Sigma} &= (\mathbf{A} - \mathbf{a}\mathbf{a}')\mathbf{A}^{-1}(\mathbf{A} - \mathbf{a}\mathbf{a}') = (\mathbf{A} - \mathbf{a}\mathbf{a}')(\mathbf{I} - \mathbf{A}^{-1}\mathbf{a}\mathbf{a}') \\ &= \mathbf{A} - \mathbf{a}\mathbf{a}' = \boldsymbol{\Sigma}, \end{aligned} \tag{4}$$

because $\mathbf{A}^{-1}\mathbf{a} = \mathbf{1}$ and $\mathbf{a}'\mathbf{1} = 1$ (here $\mathbf{1}$ generically denotes a unit-weight summation vector).

To prove that (3) is invariant for all choices of generalized inverse $\boldsymbol{\Sigma}^{-}$, we first show that $\boldsymbol{\Sigma}$ has rank equal to $p - 1$. We see that

$$\begin{aligned} \operatorname{rank}(\boldsymbol{\Sigma}) &= \operatorname{rank}(\mathbf{A} - \mathbf{a}\mathbf{a}') = \operatorname{rank}[\mathbf{A}(\mathbf{I} - \mathbf{A}^{-1}\mathbf{a}\mathbf{a}')] = \operatorname{rank}(\mathbf{I} - \mathbf{A}^{-1}\mathbf{a}\mathbf{a}') \\ &= \operatorname{rank}(\mathbf{I} - \mathbf{1}\mathbf{a}'), \end{aligned} \tag{5}$$

https://doi.org/10.1017/S0266466698145085 Published online by Cambridge University Press

as $\mathbf{A}^{-1}\mathbf{a} = \mathbf{1}$; because $\mathbf{a}'\mathbf{1} = 1$, it follows that $\mathbf{I} - \mathbf{1a}'$ is idempotent, and because the rank of an idempotent matrix is equal to its trace, we find that

$$\operatorname{rank}(\boldsymbol{\Sigma}) = \operatorname{rank}(\mathbf{I} - \mathbf{1a}') = \operatorname{tr}(\mathbf{I} - \mathbf{1a}') = p - \operatorname{tr}(\mathbf{1a}') = p - \mathbf{a}'\mathbf{1} = p - 1. \quad \textbf{(6)}$$

The dispersion matrix $\boldsymbol{\Sigma}$ satisfies

$$\boldsymbol{\Sigma}\mathbf{1} = (\mathbf{A} - \mathbf{aa}')\mathbf{1} = \mathbf{A1} - \mathbf{aa}'\mathbf{1} = \mathbf{a} - \mathbf{a} = \mathbf{0} \quad \textbf{(7)}$$

because $\mathbf{a}'\mathbf{1} = 1$, and thus all the row totals of $\boldsymbol{\Sigma}$ are 0; but $\boldsymbol{\Sigma}$ is symmetric, and hence all the column totals of $\boldsymbol{\Sigma}$ are 0 as well, and because $\boldsymbol{\Sigma}$ has maximal rank, it follows that every generalized inverse $\boldsymbol{\Sigma}^-$ must be of the form

$$\boldsymbol{\Sigma}^- = \boldsymbol{\Sigma}^+ + \mathbf{1b}' + \mathbf{c1}', \quad \textbf{(8)}$$

where $\mathbf{b}$ and $\mathbf{c}$ are arbitrary conformable vectors (see Sharpe and Styan, 1965; Styan and Subak-Sharpe, 1997). Hence,

$$\begin{aligned}(\mathbf{x}_i - \mathbf{x}_j)'\boldsymbol{\Sigma}^+(\mathbf{x}_i - \mathbf{x}_j) &= (\mathbf{x}_i - \mathbf{x}_j)'(\boldsymbol{\Sigma}^- - \mathbf{1b}' - \mathbf{c1}')(\mathbf{x}_i - \mathbf{x}_j) \\ &= (\mathbf{x}_i - \mathbf{x}_j)'\boldsymbol{\Sigma}^-(\mathbf{x}_i - \mathbf{x}_j), \quad \textbf{(9)}\end{aligned}$$

because $\mathbf{1}'(\mathbf{x}_i - \mathbf{x}_j) = \mathbf{1}'\mathbf{x}_i - \mathbf{1}'\mathbf{x}_j = 1 - 1 = 0$. Substituting (9) into (1) yields (3), and our proof is complete.

The result (4) that $\mathbf{A}^{-1}$ is a generalized inverse of $\boldsymbol{\Sigma}$ is also useful is establishing that Pearson's chi-squared statistic is asymptotically distributed as a chi-squared random variable (see Stuart and Ord, 1987, Example 15.3, pp. 487–488). We may define

$$\boldsymbol{\chi}^2 = \sum_{i=1}^{p} \frac{(O_i - E_i)^2}{E_i} = n(\mathbf{x} - \mathbf{a})'\mathbf{A}^{-1}(\mathbf{x} - \mathbf{a}), \quad \textbf{(10)}$$

where the random vector $\mathbf{x} = \{x_i\}$ follows a multinomial distribution with parameters n and $\mathbf{a} = \{a_i\}$, with the "observed" $O_i = nx_i$ and the "expected" $E_i = na_i$. From the central limit theorem, $n^{1/2}(\mathbf{x} - \mathbf{a})$ is asymptotically normally distributed $N(\mathbf{0}, \boldsymbol{\Sigma})$ and thus $\boldsymbol{\chi}^2$ is asymptotically distributed as a chi-squared random variable if and only if (see Styan, 1970, Theorem 2, p. 568),

$$\boldsymbol{\Sigma}\mathbf{A}^{-1}\boldsymbol{\Sigma}\mathbf{A}^{-1}\boldsymbol{\Sigma} = \boldsymbol{\Sigma}\mathbf{A}^{-1}\boldsymbol{\Sigma}, \quad \textbf{(11)}$$

which follows at once from (4): $\boldsymbol{\Sigma}\mathbf{A}^{-1}\boldsymbol{\Sigma} = \boldsymbol{\Sigma}$.

Matrices such as the dispersion matrix $\boldsymbol{\Sigma}$, with all rows and columns summing to 0, are said to be "double-centered" or "equicofactor"—all first cofactors are equal (see Sharpe and Styan, 1965; Styan and Subak-Sharpe, 1997). An easy computation shows that the common value of these first cofactors is equal to the product $a_1 a_2 \cdots a_p$ of the p components of the parameter vector $\mathbf{a}$. Because this product $a_1 a_2 \cdots a_p$ is positive, we have a further proof that $\boldsymbol{\Sigma}$ has maximal rank $p - 1$ (6).

https://doi.org/10.1017/S0266466698145085 Published online by Cambridge University Press

A third proof that $\boldsymbol{\Sigma}$ has maximal rank $p-1$ follows by noting that $\boldsymbol{\Sigma} = \mathbf{A} - \mathbf{a}\mathbf{a}'$ is the Schur complement of 1 in the partitioned matrix

$$\mathbf{F} = \begin{pmatrix} \mathbf{A} & \mathbf{a} \\ \mathbf{a}' & 1 \end{pmatrix}; \tag{12}$$

because rank is additive on the Schur complement (see Styan, 1985, p. 42), we see that

$$\begin{aligned} \operatorname{rank}(\boldsymbol{\Sigma}) &= \operatorname{rank}(\mathbf{A} - \mathbf{a}\mathbf{a}') = \operatorname{rank}(\mathbf{F}) - 1 \\ &= \operatorname{rank}(\mathbf{A}) + \operatorname{rank}(1 - \mathbf{a}'\mathbf{A}^{-1}\mathbf{a}) - 1 = p - 1, \end{aligned} \tag{13}$$

as $\mathbf{a}'\mathbf{A}^{-1}\mathbf{a} = 1$.

We also note that the quadratic form in (3),

$$(\mathbf{x}_i - \mathbf{x}_j)'\boldsymbol{\Sigma}^{-}(\mathbf{x}_i - \mathbf{x}_j), \tag{14}$$

is invariant with respect to the choice of generalized inverse $\boldsymbol{\Sigma}^{-1}$ if and only if

$$\mathbf{x}_i - \mathbf{x}_j \in \mathcal{C}(\boldsymbol{\Sigma}), \tag{15}$$

the column space (range) of $\boldsymbol{\Sigma}$, because by definition $\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-}\boldsymbol{\Sigma} = \boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}$ is symmetric. When all the components of the vector $\mathbf{a}$ are equal to (to $1/p$), then $\boldsymbol{\Sigma} = \mathbf{C}/p$, where the $p \times p$ "centering matrix"

$$\mathbf{C} := \mathbf{I} - \frac{1}{p}\,\mathbf{1}\mathbf{1}', \tag{16}$$

with rank $p-1$. Because $\mathbf{C}\boldsymbol{\Sigma} = \boldsymbol{\Sigma}$, we see that the column space $\mathcal{C}(\boldsymbol{\Sigma}) = \mathcal{C}(\mathbf{C}\boldsymbol{\Sigma}) \subseteq \mathcal{C}(\mathbf{C})$; because $\operatorname{rank}(\mathbf{C}) = \operatorname{rank}(\boldsymbol{\Sigma}) = p-1$, it follows that $\mathcal{C}(\boldsymbol{\Sigma}) = \mathcal{C}(\mathbf{C})$ and hence (15) follows because $\mathbf{x}_i - \mathbf{x}_j = \mathbf{C}(\mathbf{x}_i - \mathbf{x}_j)$.

A third proof of the invariance of (14) follows from the result that the Moore–Penrose inverse $\boldsymbol{\Sigma}^{+}$ may always be computed by "double-centering" any generalized inverse $\boldsymbol{\Sigma}^{-1}$, i.e.,

$$\boldsymbol{\Sigma}^{+} = \mathbf{C}\boldsymbol{\Sigma}^{-}\mathbf{C} \tag{17}$$

(see eq. (8)). Because $\mathbf{x}_i - \mathbf{x}_j = \mathbf{C}(\mathbf{x}_i - \mathbf{x}_j)$, we see that

$$\begin{aligned} (\mathbf{x}_i - \mathbf{x}_j)'\boldsymbol{\Sigma}^{-}(\mathbf{x}_i - \mathbf{x}_j) &= (\mathbf{x}_i - \mathbf{x}_j)'\mathbf{C}\boldsymbol{\Sigma}^{-}\mathbf{C}(\mathbf{x}_i - \mathbf{x}_j) \\ &= (\mathbf{x}_i - \mathbf{x}_j)'\boldsymbol{\Sigma}^{+}(\mathbf{x}_i - \mathbf{x}_j). \end{aligned} \tag{18}$$

NOTE

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

https://doi.org/10.1017/S0266466698145085 Published online by Cambridge University Press

REFERENCES

Sharpe, G.E. & G.P.H. Styan (1965) Circuit duality and the general network inverse. *IEEE Transactions on Circuit Theory* CT-12, 22–27.

Stuart, A. & J.K. Ord (1987) *Kendall's Advanced Theory of Statistics: Volume 1, Distribution Theory*, 5th ed. Oxford: Oxford University Press.

Styan, G.P.H. (1970) Notes on the distribution of quadratic forms in singular normal variables. *Biometrika* 57, 567–572.

Styan, G.P.H. (1985) Schur complements and linear statistical models. In T. Pukkila & S. Puntanen (eds.), *Proceedings of the First International Tampere Seminar on Linear Statistical Models and Their Applications*, pp. 37–75. Tampere, Finland: University of Tampere.

Styan, G.P.H. & G.E. Subak-Sharpe (1997) Inequalities and equalities associated with the Campbell–Youla generalized inverse of the indefinite admittance matrix of resistive networks. *Linear Algebra and Its Applications* 250, 349–370.

https://doi.org/10.1017/S0266466698145085 Published online by Cambridge University Press