Econometric Theory, **15**, 1999, 901–909. Printed in the United States of America.

PROBLEMS AND SOLUTIONS

PROBLEMS

99.6.1. *Central Limit Theorem for Squared MA Infinity Processes*, proposed by Antonis Demos. Assume that the following stochastic process y_t follows the usual $MA(\infty)$ process, i.e., $y_t = \varepsilon_t + \sum_{i=1}^{\infty} \alpha_i \varepsilon_{t-i}$, where the α_i's satisfy the usual conditions of stationarity, i.e., $\sum_{i=0}^{\infty} \alpha_i^2 < \infty$. Furthermore, assume that the ε_t's are identically and independently distributed random variables with finite fourth moment, i.e., $E(\varepsilon_t) = 0$ (with no loss of generality), $E(\varepsilon_t^2) = \sigma^2$, $E(\varepsilon_t^3) = 0$ (with no loss of generality), and $E(\varepsilon_t^4) = \kappa^4$ for all t's. Prove that under these assumptions, y_t^2 follows the central limit theorem.

99.6.2. *Common Principle for the Decomposition of the Total Variability in Principal Components Analysis and in Linear Regression*, proposed by Jean-Daniel Rolle. Denote $\mathcal{N}_Z$ a cloud of n points $z_i = (x_i', y_i)'$, where the x_i are $p \times 1$ vectors of observations of regressor variables $X_1, \ldots, X_p$ on n units and the y_i are observations of a response variable Y on the same units. Without restriction of the generality, we assume that the variables are centered (the center of gravity of $\mathcal{N}_Z$ is therefore the origin). The p first principal components form the "best" fitting hyperplane onto which the points z_i of $\mathcal{N}_Z$ are projected orthogonally. More precisely, let $\|\cdot\|$ be the norm inherited from the standard scalar product of $\mathbb{R}^{p+1}$ and, for a $(p+1) \times p$ matrix v of rank p, let $P_v = v(v'v)^{-1}v'$ be the matrix of the orthogonal projection onto $\Delta = \mathcal{M}(v)$, the column space of v. The inertia (spread) of $\mathcal{N}_Z$ with regard to the origin and to Δ is defined, respectively, by

$$I(\mathcal{N}_Z, 0) = \frac{1}{n-1} \sum_{i=1}^{n} \|z_i\|^2 \qquad \text{and} \tag{1}$$

$$I(\mathcal{N}_Z, \Delta) = \frac{1}{n-1} \sum_{i=1}^{n} \|z_i - P_v z_i\|^2. \tag{2}$$

We note that inertia depends on a norm and a projection. Denote $\Delta^* = \mathcal{M}(v^*)$ the (optimal) hyperplane minimizing (2). It is well known from principal components analysis that the vectors $v_1^*, \ldots, v_p^*$ forming the columns of v are orthonormal eigenvectors associated with the p largest eigenvalues $\lambda_1 > \ldots > \lambda_p$ of the covariance matrix of the variables $X_1, \ldots, X_p, Y$. Moreover, one has the following property of decomposition of inertia:

$$I(\mathcal{N}_Z, 0) = I(\mathcal{N}_Z^*, 0) + I(\mathcal{N}_Z, \Delta^*), \tag{3}$$

where $\mathcal{N}_Z^*$ is the cloud $\mathcal{N}_Z$ projected (by P_{v^*}) on Δ^*. The total inertia is thus decomposed into the inertia explained by the hyperplane Δ^* (first term on the

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

https://doi.org/10.1017/S0266466699156068 Published online by Cambridge University Press

right) plus the inertia of $\mathcal{N}_Z$ with regard to Δ^*. The global quality of the hyperplane Δ^* is measured by the ratio of explained inertia

$$Q_Z = I(\mathcal{N}_Z^*,0)/I(\mathcal{N}_Z,0) = \sum_{j=1}^{p} \lambda_j \Big/ \sum_{j=1}^{p+1} \lambda_j. \tag{4}$$

Question: Replace in (1) and (2) the norm $\|\cdot\|$ by a seminorm and P_v by the matrix p_v of a (nonorthogonal) projection in such a way that Δ^* becomes the optimal hyperplane in the sense of the mean squares. In this case, (3) becomes $SST = SSR + SSE$, the well-known decomposition formula of the linear regression, and Q_Z of (4) becomes the coefficient of determination R^2. The matrix p_v, a function of v, must be explicitly given. The conclusion that one may draw from this exercise is that the formulae of decomposition of total variability in linear regression and in principal components analysis stem from a common principle. In the first case, inertia is defined from a seminorm and a nonsymmetric idempotent matrix (p_v); in the second case, it is defined from a norm based on the standard scalar product of $\mathbb{R}^{p+1}$ and a symmetric idempotent matrix (P_v).

99.6.3. *Variances of Truncated Versus Full Distributions for Continuous Random Variables*, proposed by S.K. Sapra. Let y^* be a latent continuous random variable and let $y = y^*$ if $y^* > k$, where k is a known constant and y is the observed counterpart of y^*. Consider the following probability density functions (p.d.f.'s) for y^*.

1. Pareto: $f(y^*) = ay^{*-(a+1)}, \quad a > 0, \quad y^* > 1.$
2. Uniform: $f(y^*) = 1/(d-c), \quad c < y^* < d, \quad -\infty < c < d < \infty.$
3. Standard exponential: $f(y^*)\, e^{-y^*}, \quad y^* > 0.$
4. Gamma (with scale parameter = 1, shape parameter = 2): $f(y^*) = y^* e^{-y^*}, \quad y^* > 0.$
5. Rayleigh: $f(y^*) = y^*\, e^{-y^{*2}/2}, \quad y^* > 0.$

Derive the relationship between $Var(y^*)$ and $Var(y)$ for each of the p.d.f.'s.

99.6.4. *Variances of Censored Versus Full Distributions for Continuous Random Variables*, proposed by S.K. Sapra. Let y^* be a continuous random variable and let $y = \max(y^*, k)$, where k is a known constant. Consider the following probability density functions (p.d.f.'s) for y^*.

1. Pareto: $f(y^*) = ay^{*-(a+1)}, \quad a > 0, \quad y^* > 1.$
2. Uniform: $f(y^*) = 1/(d-c), \quad c < y^* < d, \quad -\infty < c < d < \infty.$
3. Standard exponential: $f(y^*) = e^{-y^*}, \quad y^* > 0.$
4. Gamma (with scale parameter = 1, shape parameter = 2): $f(y^*) = y^* e^{-y^*}, \quad y^* > 0.$

Derive the relationship between $Var(y^*)$ and $Var(y)$ for each of the p.d.f.'s above.

https://doi.org/10.1017/S0266466699156068 Published online by Cambridge University Press

SOLUTIONS

99.1.1. *Minimum Distance Estimator with L_1 Norm*—Solutions.[1] The following solutions have been proposed independently by Geert Dhaene and Yulia Kotlyarova. Each contains an interesting derivation.

1. Solution, proposed by Geert Dhaene. Let $\hat{\theta}_{(1)}, \hat{\theta}_{(2)}, \hat{\theta}_{(3)}$ be the order statistics corresponding to $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3$. Because $Q_n(\delta) = \sum_{j=1}^{3} |\hat{\theta}_j - \delta|$ is strictly decreasing in δ when $\delta < \hat{\theta}_{(2)}$ and strictly increasing in δ when $\delta > \hat{\theta}_{(2)}$, Q_n attains its minimum at $\hat{\delta} = \hat{\theta}_{(2)}$, the median of $\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3$. Noting that $\sqrt{n}(\hat{\theta}_j - \delta_0), j = 1,2,3$, are asymptotically i.i.d. with c.d.f. Φ, we can straightforwardly apply standard results about order statistics of i.i.d. samples (e.g., Mood, Graybill, and Boes, 1987, pp. 252–253). This gives the asymptotic density function of $\sqrt{n}(\hat{\delta} - \delta_0)$ as $6\Phi(1 - \Phi)\phi$, wherefrom

$$\Pr[\sqrt{n}(\hat{\delta} - \delta_0) \le x] \to 6 \int_{-\infty}^{x} \Phi(t)[1 - \Phi(t)]\phi(t)\, dt = [\Phi(x)]^2[3 - 2\Phi(x)].$$

2. Solution, proposed by Yulia Kotlyarova. In the proposed example,

$$\hat{\delta} = \arg\min_{\delta} \sum_{j=1}^{3} |\hat{\theta}_j - \delta| = \text{median}\{\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3\}; \tag{1}$$

$3! = 6$ disjoint events of the form $\hat{\theta}_i < \hat{\theta}_j < \hat{\theta}_k, i \ne j \ne k, i, j, k \in \{1,2,3\}$ may occur; the probability of any strict equality is zero because the distribution of $\hat{\theta}$ is continuous; thus, $\sum_{l=1}^{6} P_1 = 1$. The $\hat{\theta}_i$ are i.i.d. $N(\delta_0, n^{-1})$; therefore, the probabilities of all six events are equal, and we can write

$$\begin{aligned}\Pr[\sqrt{n}(\hat{\delta} - \delta_0) \le x] &= \Pr[\sqrt{n}(\text{median}\{\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3\} - \delta_0) \le x] \\ &= 6\Pr[\sqrt{n}(\hat{\theta}_2 - \delta_0) \le x \text{ and } \hat{\theta}_1 \le \hat{\theta}_2 \text{ and } \hat{\theta}_2 \le \hat{\theta}_3].\end{aligned}$$

Define $z_i \equiv \sqrt{n}(\hat{\theta}_i - \delta_0)$, $i = 1,2,3$. Then vector $z \Rightarrow N(0, I_3)$, and the joint density function is $f(z_1, z_2, z_3) = \phi(z_1) \cdot \phi(z_2) \cdot \phi(z_3)$, where $\phi(\,)$ is a standard normal density function. Hence,

$$\begin{aligned}\Pr[\sqrt{n}(\hat{\delta} - \delta_0) \le x] &= 6\Pr[z_2 \le x, z_1 \le z_2, z_3 \ge z_2] \\ &\Rightarrow 6 \int_{-\infty}^{x} \left[\phi(z_2) \int_{-\infty}^{z_2} \phi(z_1)\, dz_1 \cdot \int_{z_2}^{\infty} \phi(z_3)\, dz_3 \right] dz_2 \\ &= 6 \int_{-\infty}^{x} \phi(z) \cdot \Phi(z) \cdot (1 - \Phi(z))\, dz \\ &= 6 \int_{-\infty}^{x} \Phi(z) \cdot (1 - \Phi(z))\, d\Phi(z) = 3\Phi^2(x) - 2\Phi^3(x).\end{aligned}$$

https://doi.org/10.1017/S0266466699156068 Published online by Cambridge University Press

This result does not coincide with the cumulative distribution function Φ of the standard normal. The choice of the norm function $f(\cdot)$ was considered to have little qualitative influence on the asymptotic properties of the estimates of δ, and it is true as long as we can apply the continuous mapping theory and the delta method:

$$\text{Let } \hat{\delta} = \arg\min\|\hat{\theta} - g(\delta)\| = \arg\min f(\hat{\theta} - g(\delta)) = h(\hat{\theta}).$$

$$\text{If } \sqrt{n}(\hat{\theta} - \theta_0) \Rightarrow N(0,\Sigma), \text{ then } \sqrt{n}(\hat{\delta} - \delta_0) \Rightarrow N\left(0, \frac{\partial h(\theta_0)}{\partial \hat{\theta}'} \Sigma \frac{\partial h'(\theta_0)}{\partial \hat{\theta}}\right),$$

i.e., the estimates of δ will be asymptotically normal. To insure it, $h(\,)$ should be continuously differentiable at θ_0, and the variance–covariance matrix of the δ estimates should be of full rank.

From the first-order condition for (1),

$$F = \frac{\partial f(\hat{\theta} - g(\delta))}{\partial \delta} = \frac{\partial g'}{\partial \delta} \cdot \frac{\partial f}{\partial \hat{\theta}} = 0.$$

By the implicit function theorem,

$$\frac{\partial h(\theta_0)}{\partial \hat{\theta}'} = \frac{\partial \hat{\delta}}{\partial \hat{\theta}'}\bigg|_{\hat{\delta}=\delta_0} = -\left(\frac{\partial F'}{\partial \delta}\right)^{-1}\left(\frac{\partial F}{\partial \hat{\theta}'}\right) = -\left(\frac{\partial F'}{\partial \delta}\right)^{-1} \cdot \frac{\partial g'}{\partial \delta} \cdot \frac{\partial^2 f}{\partial \hat{\theta} \partial \hat{\theta}'}\bigg|_{\hat{\theta}=\theta_0}.$$

Thus, the sufficient conditions for asymptotic normality of the estimates of δ include the existence of the non-zero full-rank second-order derivative of the norm function around the true value of θ. Therefore, any nondifferentiable function (e.g., an absolute value) or any differentiable function of a type $\sum x^{2k}$, $k > 1$, may not yield an expected standard distribution.

NOTE

1. An excellent solution has been proposed independently by O. Linton, the poser of the problem.

REFERENCE

Mood, A.M., F.A. Graybill, & D.C. Boes (1987) *Introduction to the Theory of Statistics*, 3rd ed. Singapore: McGraw-Hill.

99.1.2. *An Interesting Property of the Mahalanobis Distance*—Solution,[1] proposed by Tom Steerneman and Tom Wansbeek. Let $A = I - (1/(p+1))11'$ denote the usual centering matrix of order $p + 1$. Define

$$D = (d_{ij}) = pAX(X'AX)^{-1}X'A,$$

then

$$d_{ij} = (x_i - \bar{x})'S^{-1}(x_j - \bar{x}).$$

https://doi.org/10.1017/S0266466699156068 Published online by Cambridge University Press

Because $(1/p)D$ is idempotent of order $p+1$, has rank p, and $D1 = 0$, it follows that $D = pA$. Hence,

$$d_i^2 = d_{ii} = p^2/(p+1),$$

and for $i \neq j$

$$d_{ij} = -p/(p+1).$$

NOTE

1. Excellent solutions have been proposed independently by J. Graffelman, by M. van de Velden and H. Neudecker, and by A.S. Hadi, the poser of the problem.

99.1.3. *An Application of Beinstein's Inequality*—Solution, proposed by Hua Liang. Before proving this theorem, we state Beinstein's inequality for easy reference.

BEINSTEIN'S INEQUALITY. *Let $V_1,\ldots,V_n$ be independent random variables with zero means and bounded ranges $|V_i| \le M$. Then, for each $\eta > 0$,*

$$P\left(\left|\sum_{i=1}^n V_i\right| > \eta\right) \le 2\exp\left[-\eta^2 \Big/ \left\{2\left(\sum_{i=1}^n \operatorname{var} V_i + M\eta\right)\right\}\right].$$

Proof of Theorem. Denote $V_j' = V_j I(|V_j| \le n^{1/r})$ and $V_j'' = V_j - V_j'$ for $j = 1,\ldots,n$. Let $M = Cn^{-p_1}n^{1/r}$ and $\eta = n^{-s}\log n$. By Beinstein's inequality,

$$\begin{aligned} P&\left\{\max_{1\le i\le n}\left|\sum_{j=1}^n a_{ji}(V_j' - EV_j')\right| > C_1\eta\right\} \\ &\le \sum_{i=1}^n P\left\{\left|\sum_{j=1}^n a_{ji}(V_j' - EV_j')\right| > C_1\eta\right\} \\ &\le 2n\exp\left(-\frac{C_1 n^{-2s}\log^2 n}{2\sum_{j=1}^n a_{ji}^2 EV_j^2 + 2n^{-p_1+1/r-s}\log n}\right) \\ &\le 2n\exp(-C_1^2 C\log n) \le Cn^{-3/2} \end{aligned}$$

for some large $C_1 > 0$. The last second inequality from

$$\sum_{j=1}^n a_{ji}^2 EV_j^2 \le \sup|a_{ji}|\sum_{j=1}^n a_{ji}EV_j^2 = n^{-p_1+p_2} \quad \text{and} \quad n^{-p_1+1/r-s}\log n \le n^{-p_1+p_2}.$$

By the Borel–Cantelli lemma,

$$\max_{1\le i\le n}\left|\sum_{j=1}^n W_{nj}(T_i)(V_j' - EV_j')\right| = O(n^{-s}\log n) \quad \text{a.s.} \tag{1}$$

https://doi.org/10.1017/S0266466699156068 Published online by Cambridge University Press

Let $1 \le p < 2$, $1/p + 1/q = 1$ such that $1/q < (p_1 + p_2)/2 - 1/r$. By Hölder's inequality,

$$\max_{1\le i\le n}\left|\sum_{j=1}^{n} a_{ji}(V_j'' - EV_j'')\right| \le \max_{1\le i\le n}\left(\sum_{j=1}^{n}|a_{ji}|^q\right)^{1/q}\left(\sum_{j=1}^{n}|V_j'' - EV_j''|^p\right)^{1/p}$$
$$\le Cn^{-(p_1q-1)/q}\left(\sum_{j=1}^{n}|V_j'' - EV_j''|^p\right)^{1/p}. \tag{2}$$

Observe that

$$\frac{1}{n}\sum_{j=1}^{n}(|V_j'' - EV_j''|^p - E|V_j'' - EV_j''|^p) \to 0 \quad \text{a.s.} \tag{3}$$

and $E|V_j''|^p \le E|V_j|^r n^{-1+p/r}$, and then

$$\sum_{j=1}^{n} E|V_j'' - EV_j''|^p \le Cn^{p/r} \quad \text{a.s.} \tag{4}$$

Combining (2) and (3) with (4), we obtain

$$\max_{1\le i\le n}\left|\sum_{k=1}^{n} a_{ki}(V_k'' - EV_k'')\right| \le Cn^{-p_1+1/q+1/r} = o(n^{-s}) \quad \text{a.s.} \tag{5}$$

The theorem follows directly from (1) and (5).

Remark. In nonparametric regression, one generally has to deal with the following term:

$$\max_{1\le i\le n}\left|\sum_{k=1}^{n}\omega_{nk}(T_i)\varepsilon_k\right| \tag{6}$$

for some design points T_i. Where ε_k are i.i.d. random variables with $E\varepsilon = 0$ and $E|\varepsilon|^3 < \infty$, and $\omega_{nk}(\cdot)$ are probability weight functions, which are supposed to be satisfied such that:

(i) $\max_{1\le i\le n}\sum_{j=1}^{n}\omega_{ni}(T_j) = O(1)$,

(ii) $\max_{1\le i,j\le n}\omega_{ni}(T_j) = O(b_n)$,

(iii) $\max_{1\le i\le n}\sum_{j=1}^{n}\omega_{nj}(T_i)I(|T_j - T_i| > c_n) = O(c_n)$,

with $b_n = n^{-2/3}$ and $c_n = n^{-1/3}\log n$.

Taking $r = 3$, $p_1 = \frac{2}{3}$, and $p_2 = 0$ in the theorem, it follows that the order of (6) is $O(n^{-1/3}\log n)$.

https://doi.org/10.1017/S0266466699156068 Published online by Cambridge University Press

99.1.4. *Equality of Two IV Estimators*—Solution, proposed by Hailong Qian. To simplify notation, let $P_1 = P_{[Z_1]}$ and $M_1 = M_{[Z_1]}$. Define $\hat{\gamma} = (X_2'P_1X_2)^{-1}X_2'P_1X_1$, $\gamma = \text{p lim}\,\hat{\gamma} = \text{p lim}(N^{-1}X_2'P_1X_2)^{-1}N^{-1}X_2'P_1X_1$, and $e_1 = X_1 - X_2\hat{\gamma}$. Thus, e_1 is the residual in the IV regression of X_1 on X_2, using Z_1 as instruments. With this notation, the condition $Z_2'M_1[X_1 - X_2(X_2'P_1X_2)^{-1}X_2'P_1X_1] = 0$ becomes simply $Z_2'M_1e_1 = 0$.

The model to be estimated is $y = X_1\beta_1 + X_2\beta_2 + \varepsilon$. We transform it as follows:

$$y = e_1\beta_1 + X_2\pi + \eta, \tag{2}$$

where $\pi = \beta_2 + \gamma\beta_1$ and $\eta = \varepsilon + X_2(\hat{\gamma} - \gamma)\beta_1$. We note that $\text{p lim}\,N^{-1}Z'\eta = 0$ and $\hat{X}_2'e_1 = X_2'P_1e_1 = 0$.

Let $\hat{\beta}_1$, $\hat{\pi}$ be the IV estimates from (2) when the instruments are Z_1. Because $X_2'P_1e_1 = 0$, we have simply

$$\hat{\beta}_1 = (e_1'P_1e_1)^{-1}e_1'P_1y. \tag{3}$$

However, noticing that $P_1e_1 = P_1(X_1 - X_2\hat{\gamma}) = \hat{X}_1 - \hat{X}_2\hat{\gamma} = \hat{X}_1 - \hat{X}_2(\hat{X}_2'\hat{X}_2)^{-1}\hat{X}_2'\hat{X}_1 = M_{[\hat{X}_2]}\hat{X}_1$, where the third equality is from $\hat{\gamma} = (X_2'P_1X_2)^{-1}X_2'P_1X_1 = (\hat{X}_2'\hat{X}_2)^{-1}\hat{X}_2'\hat{X}_1$, we obtain

$$\hat{\beta}_1 = (\hat{X}_1'M_{[\hat{X}_2]}\hat{X}_1)^{-1}\hat{X}_1'M_{[\hat{X}_2]}y. \tag{4}$$

This is indeed the same as the corresponding subvector of $\hat{\beta} = (\hat{X}'\hat{X})^{-1}\hat{X}'y$, the IV estimator of the untransformed equation using Z_1 as instruments.

Now let $\tilde{\beta}_1$, $\tilde{\pi}$ be the IV estimates from (2) when the instruments are Z. Thus,

$$\begin{bmatrix}\tilde{\beta}_1\\ \tilde{\pi}\end{bmatrix} = \begin{bmatrix}e_1'P_Ze_1 & e_1'P_ZX_2\\ X_2'P_Ze_1 & X_2'P_ZX_2\end{bmatrix}^{-1}\begin{bmatrix}e_1'P_Zy\\ X_2'P_Zy\end{bmatrix}. \tag{5}$$

Thus, using the partitioned-matrix inverse rule, we have

$$\tilde{\beta}_1 = (\tilde{e}_1'M_{[\tilde{X}_2]}\tilde{e}_1)^{-1}\tilde{e}_1'M_{[\tilde{X}_2]}y, \tag{6}$$

where $\tilde{e}_1 = P_Ze_1$, but $M_{[\tilde{X}_2]}\tilde{e}_1 = M_{[\tilde{X}_2]}P_Ze_1 = M_{[\tilde{X}_2]}P_Z(X_1 - X_2\hat{\gamma}) = M_{[\tilde{X}_2]}(\tilde{X}_1 - \tilde{X}_2\hat{\gamma}) = M_{[\tilde{X}_2]}\tilde{X}_1$. With this substitution, we obtain

$$\tilde{\beta}_1 = (\tilde{X}_1'M_{[\tilde{X}_2]}\tilde{X}_1)^{-1}\tilde{X}_1'M_{[\tilde{X}_2]}y, \tag{7}$$

which is the same as the corresponding subvector of $\tilde{\beta} = (\tilde{X}'\tilde{X})^{-1}\tilde{X}'y$, the IV estimator of the untransformed equation using Z as instruments.

We now show that $\hat{\beta}_1 = \tilde{\beta}_1$ if $Z_2'M_1e_1 = 0$. We note that $P_Ze_1 = (P_1 + P_{[M_1Z_2]})e_1 = P_1e_1$ when $Z_2'M_1e_1 = 0$. Thus, $X_2'P_Ze_1 = X_2'P_1e_1 = 0$. Thus, substituting it into (5), we obtain $\tilde{\beta}_1 = (e_1'P_Ze_1)^{-1}e_1'P_Zy = (e_1'P_1e_1)^{-1}e_1'P_1y = \hat{\beta}_1$, where the second equality is from $P_Ze_1 = P_1e_1$, and the last equality is from (3). ■

https://doi.org/10.1017/S0266466699156068 Published online by Cambridge University Press

Comments.

1. The expression $\hat{\beta}_1 = (e_1'P_1e_1)^{-1}e_1'P_1y$ of (3) can be interpreted as an extension of the well-known partial-least-squares regression formula $\tilde{\beta}_1 = (X_1'M_{[X_2]}X_1)^{-1}X_1'M_{[X_2]}y$ to the case of linear IV estimation.
2. Breusch et al. (1997) showed that $AV(\hat{\beta}_1) = AV(\tilde{\beta}_1)$ if and only if $\operatorname{plim} N^{-1}Z_2'M_1[X_1 - X_2(X_2'P_1X_2)^{-1}X_2'P_1X_1] = 0$. Thus, this note provides an intuitive explanation for their condition.
3. It is easy to show that $\hat{\beta} = \tilde{\beta}$ if $Z_2'M_1X = 0$. Thus, the condition $Z_2'M_1[X_1 - X_2(X_2'P_1X_2)^{-1}X_2'P_1X_1] = 0$ for $\hat{\beta}_1 = \tilde{\beta}_1$ can be interpreted as an extension of the condition $Z_2'M_1X = 0$ for $\hat{\beta} = \tilde{\beta}$.

REFERENCE

Breusch, T., H. Qian, P. Schmidt, & D. Wyhowski (1997) Redundancy of moment conditions. *Journal of Econometrics*, forthcoming.

99.1.5. *The Justification of Multidimensional Scaling under Euclidean Conditions*—Solution,[1] proposed by Jan Graffelman. The proposed problem is in fact a reexpression of the solution of a principal coordinates analysis. Let $\mathbf{X}$ be the $n \times p$ data matrix of rank k. The matrix of squared Euclidean distances $\mathbf{D}$ can then be obtained from the data as $\mathbf{D} = \mathbf{s1}' + \mathbf{1s}' - \mathbf{2S}$, where $\mathbf{S} = \mathbf{XX}'$ and $\mathbf{s} = \operatorname{diag}(\mathbf{S})$. Matrix $\mathbf{A}$ then becomes

$$\mathbf{A} = -\tfrac{1}{2}\mathbf{D} = -\tfrac{1}{2}\mathbf{s1}' - \tfrac{1}{2}\mathbf{1s}' + \mathbf{S}, \tag{1}$$

and by double centering this matrix, we get $\mathbf{B} = \mathbf{MAM} = \mathbf{MSM} = \mathbf{MXX}'\mathbf{M}$, because the first two terms on the right-hand side of (1) become zero because of the centering operation $\mathbf{s1}'\mathbf{M} = \mathbf{s1}'(\mathbf{I} - (1/n)\mathbf{11}') = \mathbf{s}(\mathbf{1}' - \mathbf{1}') = \mathbf{0}$ and, analogously, $\mathbf{M1s}' = \mathbf{0}$. We will normally have more cases than variables, and so we find for the rank of $\mathbf{B}$, $r(\mathbf{B}) = r(\mathbf{MXX}'\mathbf{M}) \leq \min(r(\mathbf{M}), r(\mathbf{XX}')) = r(\mathbf{XX}') = k$. Thus, $r(\mathbf{B}) \leq k$, and $\mathbf{B}$ will normally have $n - k$ zero eigenvalues. Because $\mathbf{B}$ is symmetric, we can do the spectral decomposition $\mathbf{B} = \mathbf{T\Lambda T}'$, which we can also write as

$$\mathbf{BT} = \mathbf{T\Lambda}, \tag{2}$$

because $\mathbf{T}'\mathbf{T} = \mathbf{I}$. Indeed, $\mathbf{B}$ can still perfectly be recovered from $\mathbf{\Lambda}$ and $\mathbf{T}$ by omitting the $n - k$ zero eigenvalues and their corresponding columns in $\mathbf{T}$, and so from now on we take $\mathbf{\Lambda}$ to be $k \times k$ and nonsingular, and $\mathbf{T}$ to be $n \times k$ and semiorthogonal. We can then obtain solution $\mathbf{Z}$ by postmultiplying (2) by $\mathbf{\Lambda}^{-1/2}$:

$$\mathbf{Z} = \mathbf{T\Lambda}^{1/2} = \mathbf{BT\Lambda}^{-1/2} = \mathbf{MXX}'\mathbf{MT\Lambda}^{-1/2} = \mathbf{MXL}, \tag{3}$$

where $\mathbf{L}$ is $\mathbf{X}'\mathbf{MT\Lambda}^{-1/2}$. The $p \times k$ matrix $\mathbf{L}$ is semiorthogonal because $\mathbf{L}'\mathbf{L} = \mathbf{\Lambda}^{-1/2}\mathbf{T}'\mathbf{M}'\mathbf{XX}'\mathbf{MT\Lambda}^{-1/2} = \mathbf{\Lambda}^{-1/2}\mathbf{T}'\mathbf{BT\Lambda}^{-1/2} = \mathbf{\Lambda}^{-1/2}\mathbf{T}'\mathbf{T\Lambda T}'\mathbf{T\Lambda}^{-1/2} = \mathbf{\Lambda}^{-1/2}\mathbf{\Lambda\Lambda}^{-1/2} = \mathbf{I}_k$.

https://doi.org/10.1017/S0266466699156068 Published online by Cambridge University Press

The equivalence between principal components analysis and principal coordinates analysis is well known (Gower, 1966). The solution of a principal components analysis can be obtained by postmultiplying the centered data by the eigenvectors of the covariance matrix. Equation (3) precisely expresses this, where $\mathbf{Z}$ is the matrix of principal components (centered but not standardized), and where matrix $\mathbf{L}$ contains the eigenvectors of the covariance matrix $c\mathbf{X}'\mathbf{M}\mathbf{X}$:

$$c(\mathbf{X}'\mathbf{M}\mathbf{X})\mathbf{L} = c\mathbf{X}'\mathbf{M}\mathbf{X}(\mathbf{X}'\mathbf{M}\mathbf{T}\boldsymbol{\Lambda}^{-1/2}) = c\mathbf{X}'\mathbf{M}(\mathbf{M}\mathbf{X}\mathbf{X}'\mathbf{M})\mathbf{T}\boldsymbol{\Lambda}^{-1/2}$$
$$= c\mathbf{X}'\mathbf{M}\mathbf{T}\boldsymbol{\Lambda}\mathbf{T}'\mathbf{T}\boldsymbol{\Lambda}^{-1/2} = c\mathbf{X}'\mathbf{M}\mathbf{T}\boldsymbol{\Lambda}^{1/2} = c\mathbf{L}\boldsymbol{\Lambda},$$

where we used the idempotency of $\mathbf{M}$ and where c is the factor $1/(n-1)$.

NOTE

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

REFERENCE

Gower, J. (1966) Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika* 53(3); 325–338.

https://doi.org/10.1017/S0266466699156068 Published online by Cambridge University Press