Econometric Theory, **14**, 1998, 525–537. Printed in the United States of America.

PROBLEMS AND SOLUTIONS

PROBLEMS

98.4.1. *A Determinantal Identity*, proposed by Heinz Neudecker and Michel van de Velden. Show that

$$|\mathbf{X}'\mathbf{A}\mathbf{X}||\mathbf{X}'\mathbf{A}^{-1}\mathbf{X}| = |\mathbf{A}+\mathbf{B}|/|\mathbf{A}|,$$

for an $n \times k$ matrix $\mathbf{X}$ satisfying $\mathbf{X}'\mathbf{X} = \mathbf{I}_k$, a positive definite matrix $\mathbf{A}$ and $\mathbf{B} = \mathbf{X}\mathbf{X}'\mathbf{A} - \mathbf{A}\mathbf{X}\mathbf{X}'$.

REFERENCE

Bloomfield, P. & G.S. Watson (1975) The inefficiency of least squares. *Biometrika* 62, 121–128.

98.4.2. *An Alternative Way to Calculate the SUR Estimator*, proposed by Paolo Paroulo. Consider the following multivariate regression model under standard assumptions,

$$\underset{p\times 1}{y_t} = \underset{p\times s}{\varphi}\ \underset{s\times 1}{x_t} + \underset{p\times 1}{\varepsilon_t}\ \underset{p\times m}{(\gamma}\ \underset{p\times u}{\eta)}\begin{pmatrix} z_t \\ {\scriptstyle m\times 1} \\ v_t \\ {\scriptstyle u\times 1}\end{pmatrix} + \varepsilon_t, \qquad t = 1,\ldots,T, \tag{1}$$

with exclusion restrictions on γ, which do not involve complete columns of γ. The regressors in v_t are thus common to all the equations, whereas z_t contains the regressors that appear in some but not all the equations. Let ε_t be i.i.d. $N(0,\Omega)$ with Ω-positive definite and define $\alpha = (\alpha_1',\alpha_2')' = \mathrm{vec}(\varphi) = (\mathrm{vec}(\gamma)', \mathrm{vec}(\eta)')'$, where vec is the column stacking operator. The exclusion restrictions can be stated as $\alpha_1 = \mathrm{vec}(\gamma) = H_{11}\xi_1$, where H_{11} is a known matrix or, in terms of α,

$$\alpha = H\xi = (H_1,H_2)\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix} = \begin{pmatrix} H_{11} & 0 \\ 0 & I_{pu}\end{pmatrix}\begin{pmatrix}\xi_1\\ \xi_2\end{pmatrix}, \tag{2}$$

where $\alpha_2 = \xi_2 = \mathrm{vec}(\eta)$. It is well known that the SUR estimator of ξ can be written as follows:

$$\hat{\xi} = \begin{pmatrix}\hat{\xi}_1\\ \hat{\xi}_2\end{pmatrix} = (H'(M_{xx}\otimes\hat{\Omega}^{-1})H)^{-1}H'\,\mathrm{vec}(\hat{\Omega}^{-1}M_{yx}), \tag{3}$$

where $M_{ab} = T^{-1}\sum_{t=1}^{T} a_t b_t'$, are sample moment matrices, $a_t, b_t = x_t, y_t$, $\hat{\Omega}$ is a suitable estimator of Ω, and $\otimes$ is the Kronecker product. Finally, assume that the usual conditions under which the SUR estimator collapses to the OLS estimator

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

https://doi.org/10.1017/S0266466698144079 Published online by Cambridge University Press

do not hold. Show that the SUR estimator $\hat{\xi}$ can be calculated by using the alternative following steps.

1. Regress y_t and z_t on v_t, obtaining residuals $y_{t.v}$ and $z_{t.v}$; compute the resulting moment matrices $M_{zz.v}$ and $M_{zy.v}$, where $M_{ab.c} = M_{ab} - M_{ac}M_{cc}^{-1}M_{cb}$.
2. Calculate the SUR estimator of ξ_1 based on the equation $y_{t.v} = \gamma z_{t.v} + \text{error}$, i.e.,

$$\hat{\xi}_1 = (H_{11}'(M_{zz.v} \otimes \hat{\Omega}^{-1})H_{11})^{-1}H_{11}' \operatorname{vec}(\hat{\Omega}^{-1}M_{yz.v}); \tag{4}$$

   this defines $\hat{\alpha}_1 = \operatorname{vec}(\hat{\gamma}) = H_{11}\hat{\xi}_1$ and hence $\hat{\gamma}$.
3. Calculate $\hat{\xi}_2$ by regressing $y_t - \hat{\gamma}z_t$ on v_t, i.e.,

$$\hat{\xi}_2 = \operatorname{vec}((M_{yv} - \hat{\gamma}M_{zv})M_{vv}^{-1}), \tag{5}$$

   or $\hat{\eta} = (M_{yv} - \hat{\gamma}M_{zv})M_{vv}^{-1}$.

Note that $(M_{xx} \otimes \hat{\Omega}^{-1})$ in (3) is of dimension $ps \times ps$, whereas $(M_{zz.v} \otimes \hat{\Omega}^{-1})$ in (4) is of dimension $pm \times pm$, which can be substantially smaller $(s = m + u)$; this entails smaller memory requirements and hence faster calculations.

98.4.3. *Regression Specification Error Test as a Gauss–Newton Regression*, proposed by Badi H. Baltagi. Davidson and MacKinnon (1993) showed that Ramsey's (1969) regression error specification test (RESET) can be derived as a Gauss–Newton regression. This problem is a simple extension of their results. Suppose that the linear regression model under test is given by:

$$y_t = X_t'\beta + u_t \qquad t = 1,2,\ldots,T, \tag{1}$$

where β is a $k \times 1$ vector of unknown parameters. Suppose that the alternative is the nonlinear regression model between y_t and X_t:

$$y_t = X_t'\beta[1 + \theta(X_t'\beta) + \gamma(X_t'\beta)^2 + \lambda(X_t'\beta)^3] + u_t, \tag{2}$$

where θ, γ, and λ are unknown scalar parameters. It is well known that Ramsey's (1969) RESET is obtained by regressing y_t on $X_t, \hat{y}_t^2, \hat{y}_t^3$, and $\hat{y}_t^4$ and by testing that the coefficients of all powers of $\hat{y}_t$ are jointly zero. Show that this RESET can be derived from a Gauss–Newton regression on (2), which tests $\theta = \gamma = \lambda = 0$. For further readings on this subject, see MacKinnon and Magee (1990).

REFERENCES

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

MacKinnon, J.G. & L. Magee (1990) Transforming the dependent variable in regression models. *International Economic Review* 31, 315–339.

Ramsey, J.B. (1969) Tests for specification errors in classical linear least-squares regression analysis. *Journal of the Royal Statistical Society Series B* 31, 535–552.

https://doi.org/10.1017/S0266466698144079 Published online by Cambridge University Press

98.4.4. *Regression with Missing Observations on a Explanatory Variable*, proposed by Enrique Sentana. Consider the linear regression of a variable, y_i, on a constant and two explanatory variables, x_i and z_i say, and suppose that there are available two random samples of sizes N_1 and N_2 (with $N_1 \geq 2$, $N_2 \geq 3$), the first of which contains no observations on z_i.

Under the assumption of joint normality, show that OLS estimates based on the second sample are numerically identical to maximum likelihood estimates of the regression parameters computed from the combined sample if the sample means, variances, and covariance of y_i, x_i are the same in both samples.

SOLUTIONS

97.4.1. *Hausman's Specification Test as a Gauss–Newton Regression*—Solution, proposed by Badi H. Baltagi. (a) The GNR on (3) runs the restricted residuals under H_0 on the derivatives of the regression function with respect to b and c evaluated at the restricted estimates. Knowing θ, the restricted estimates of (3) under H_0 are $c = 0$ and $b = \hat{\beta}_{\mathrm{GLS}}$. The latter estimate follows from the fact that y_{it}^* regressed on X_{it}^* yields the GLS estimator of β from model (1) (see Hausman, 1978). The restricted residuals from (3) under H_0 will therefore be the GLS residuals $(y_{it}^* - X_{it}^{*\prime}\hat{\beta}_{\mathrm{GLS}})$. These residuals are regressed on X_{it}^* and $\widetilde{X}_{it}$, the derivatives of the regression function in (3) with respect to b and c, respectively, evaluated at $\hat{\beta}_{\mathrm{GLS}}$ and $c = 0$. Because we are dealing with linear regressions, the latter evaluation is not applicable.

(b) From the definitions of X_{it}^* and $\widetilde{X}_{it}$, these regressors span the same space as $\widetilde{X}_{it}$ and $\overline{X}_i$. Note that X_{it} can be written as the sum of $\widetilde{X}_{it}$ and $\overline{X}_i$. Hence, the GNR in (a) yields the same regression sums of squares and therefore the same test statistic as that proposed by Ahn and Lo (1996).

REFERENCES

Ahn, S.C. & S. Lo (1996) A reformulation of the Hausman test for regression models with pooled cross-section time-series data. *Journal of Econometrics* 71, 309–319.

Hausman, J.A. (1978) Specification tests in econometrics. *Econometrica* 46, 1251–1271.

97.4.2. *Asymptotic Properties of the LSE of the Variance in a Linear Model*—Solutions. The following solutions have been proposed independently by Benedikt M. Pötscher and Xing Ming. They both noticed a small mistake in the original draft of the Problem as proposed by Risto Heijmans.

1. Solution, proposed by Benedikt M. Pötscher.

We first note that the claim in the original formulation of the problem is not correct. The problem as posed maintains the condition $\lim_{a\to\infty} \max_{1\leq i\leq n} \times \int_{|x|>a} x^4 f_i(x)\, dx = 0$ (presumably for every n, but the formulation is not specific

https://doi.org/10.1017/S0266466698144079 Published online by Cambridge University Press

about a quantifier for n). This condition is superfluous given the other assumptions because it already follows from the assumption of a finite fourth moment (note that the max is over a finite set). Now, under the remaining assumptions the claimed result cannot hold, and counterexamples can indeed be constructed (which are available on request). The reason essentially is that, in view of Steps 1 and 2, the claim will hold if and only if ε_t^2 satisfy the appropriate central limit theorem. Because the ε_t^2 are not necessarily identically distributed, a condition like the Lindeberg condition needs to be verified, which is not possible from the given assumptions.

The solution presented below will therefore be based on the following set of assumptions, which is—apart from containing an appropriate modification of the above mentioned condition that allows verification of the Lindeberg condition—weaker than the set of assumptions maintained in the original formulation of the problem. Let X be a (possibly) random $n \times k$ matrix that is independent of the random vector $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)'$ for every n. (The elements of X may even depend on n, i.e., $X = (x_{ti,n})$.) The errors ε_t, $t \geq 1$, are independent with $E\varepsilon_t = 0$, $E\varepsilon_t = \sigma^2$, and with a common finite fourth moment μ_4. Assume furthermore that $\limsup_n \max_{1 \leq t \leq n} E\varepsilon_t^4 \mathbf{1}(|\varepsilon_t| > a) \to 0$ as $a \to \infty$. The claimed central limit theorem for $S^2 = \varepsilon'\varepsilon/(n-k) - \varepsilon' P_X \varepsilon/(n-k)$, where P_X is the projection matrix onto the column space of X, is now a direct consequence of the following two steps.

Step 1: Central limit theorem for $n^{1/2}[\varepsilon'\varepsilon/(n-k) - \sigma^2]$. Observe that $\operatorname{var}(n^{1/2}[\varepsilon'\varepsilon/n - \sigma^2]) = \mu_4 - \sigma^4$. If $\mu_4 - \sigma^4 = 0$, then $n^{1/2}[\varepsilon'\varepsilon/n - \sigma^2] \xrightarrow{d} N(0, \mu_4 - \sigma^4)$ in a trivial sense. If $\mu_4 - \sigma^4 > 0$, we only have to check the Lindeberg condition, all other assumptions in the classical central limit theorem for independent random variables being clearly satisfied (cf. Shiryaev, 1995, Ch. III.4). The Lindeberg condition amounts to showing convergence to zero of $n^{-1} \sum_{t=1}^n E[(\varepsilon_t^2 - \sigma^2)^2 \mathbf{1}(|\varepsilon_t^2 - \sigma^2| > \alpha n^{1/2})]$ as $n \to \infty$ for every $\alpha > 0$. Now, for every $b > 0$, the limsup of this expression is bounded above by $\limsup_n n^{-1} \sum_{t=1}^n E[(\varepsilon_t^2 - \sigma^2)^2 \mathbf{1}(|\varepsilon_t^2 - \sigma^2| > b^2)]$. For $b^2 > \sigma^2$, this last expression reduces to $\limsup_n n^{-1} \sum_{t=1}^n E[(\varepsilon_t^2 - \sigma^2)^2 \mathbf{1}(\varepsilon_t^2 > b^2 + \sigma^2)]$, which is bounded above by $\limsup_n n^{-1} \sum_{t=1}^n E[\varepsilon_t^4 \mathbf{1}(\varepsilon_t^2 > a^2)]$, where $a^2 = b^2 + \sigma^2$. This bound is certainly not smaller than $\limsup_n \max_{1 \leq t \leq n} E\varepsilon_t^4 \mathbf{1}(|\varepsilon_t| > a)$, which converges to zero for $a \to \infty$ by assumption. This establishes $n^{1/2}[\varepsilon'\varepsilon/n - \sigma^2] \xrightarrow{d} N(0, \mu_4 - \sigma^4)$, also in case $\mu_4 - \sigma^4 > 0$. It follows immediately that also $n^{1/2}[\varepsilon'\varepsilon/(n-k) - \sigma^2] \xrightarrow{d} N(0, \mu_4 - \sigma^4)$ because the difference is obviously $o_p(1)$.

Step 2: Negligibility of $n^{1/2}\varepsilon' P_X \varepsilon/(n-k)$. We show that $n^{1/2}\varepsilon' P_X \varepsilon/(n-k) \to 0$ in probability. From Markov's inequality, $\operatorname{Prob}(n^{1/2}\varepsilon' P_X \varepsilon/(n-k) > \alpha) \leq \alpha^{-1} n^{1/2} E(\varepsilon' P_X \varepsilon)/(n-k)$, and hence it suffices to show that $E(\varepsilon' P_X \varepsilon)$ is bounded. Using standard properties of the trace operator, X-measurability of P_X, and independence between X and ε, we obtain $E(\varepsilon' P_X \varepsilon | X) = E(\operatorname{tr}(\varepsilon' P_X \varepsilon)|X) = \operatorname{tr}[P_X E(\varepsilon\varepsilon'|X)] = \operatorname{tr}[P_X E(\varepsilon\varepsilon')] = \sigma^2 \operatorname{tr}(P_X) \leq \sigma^2 k$, and hence $E(\varepsilon' P_X \varepsilon) \leq \sigma^2 k$. Note that the (conditional) expectations involved are well defined because ε has

https://doi.org/10.1017/S0266466698144079 Published online by Cambridge University Press

finite second moments and because the elements p_{ij} of the random matrix P_X satisfy $|p_{ij}| \leq k^{1/2}$ in view of $\sum_{i,j} p_{ij}^2 = \mathrm{tr}(P_X' P_X) = \mathrm{tr}(P_X) \leq k$.

Comments.

(i) It is noteworthy that, apart from the independence between X and ε, *no* other conditions on X have been imposed for the central limit theorem for S^2 to hold. In particular, we do not assume X to have full column rank, and the assumptions made above do not rule out dependencies between regressors in the sample period $1 \leq t \leq n$ and "future" errors ε_s, $s > n$.

(ii) Clearly, the condition $\lim_{a\to\infty} \limsup_n \max_{1\leq t\leq n} E\varepsilon_t^4 \mathbf{1}(|\varepsilon_t| > a) = 0$ can be replaced by the weaker, but more complex, condition (Lindeberg condition)

$$\lim_{n\to\infty} n^{-1} \sum_{t=1}^{n} E[\varepsilon_t^2 - \sigma^2)^2 \mathbf{1}(|\varepsilon_t^2 - \sigma^2| > \alpha n^{1/2})] = 0 \quad \text{for every } \alpha > 0,$$

without affecting any of the results.

(iii) From the proof it is evident that a central limit theorem will hold for S^2 under any alternative set of assumptions that implies a central limit theorem for $\varepsilon'\varepsilon/(n-k)$ as well as the relation $E(\varepsilon' P_X \varepsilon) = o(n^{1/2})$. As the proof has shown, the latter property holds (even with $o(n^{1/2})$ replaced by $O(1)$), e.g., if only $E(\varepsilon\varepsilon') = \sigma^2 I_n$, $\sigma^2 < \infty$, and X and ε are independent (or, more generally, if $E(\varepsilon\varepsilon'|X) = E(\varepsilon\varepsilon')$). In particular, the temporal independence of the errors and the higher moment assumptions have not been used in the proof of Step 2. For further sufficient conditions for negligibility of the $\varepsilon' P_X \varepsilon$ term, see the theorem below.

(iv) If we also allow the errors and/or the number of regressors k_n to depend on n, then S^2 will satisfy a central limit theorem provided $\varepsilon'\varepsilon/(n-k_n)$ satisfies a central limit theorem and $E(\varepsilon' P_X \varepsilon) = o((n-k_n)/n^{1/2})$. For example, if k_n depends on n, but otherwise the set of assumptions used above holds, S^2 will continue to satisfy $n^{1/2}(S^2 - \sigma^2) \xrightarrow{d} N(0, \mu_4 - \sigma^4)$ whenever $k_n = o(n^{1/2})$.

(v) Obviously, similar results are true if we use n^{-1} rather than $(n-k)^{-1}$ as the normalization factor in the definition of S^2.

(vi) Related results: Many textbooks prove asymptotic properties of S^2 under additional but unnecessary assumptions on the regressors, like convergence of $X'X/n$. Gleser (1966) proved strong consistency of S^2 only under the assumption that the errors are i.i.d. with zero mean and finite variance and that X is nonstochastic and of full column rank. Schmidt (1976) noted that the full column rank assumption can be dropped and extended the result by Gleser to SUR models. Apparently unaware of Gleser (1966), Kloek (1972) proved weak consistency under the same assumptions as did Gleser (1966), and Drygas (1975) noted that the full column rank assumption in Kloek's paper can be dropped. For more related results, see the comments in the next section.

Extension to correlated errors and to residual autocovariance estimation. The solution given above proceeded by reducing asymptotic normality of S^2 (in Step 2) to asymptotic normality of its unobservable counterpart $\varepsilon'\varepsilon/(n-k)$. Asymptotic normality of this latter quantity can be obtained directly from central limit theorems available in the literature, and thus generalizations of Step 1 are

https://doi.org/10.1017/S0266466698144079 Published online by Cambridge University Press

quite straightforward. Thus, we shall concentrate on obtaining more general conditions under which the reduction step, i.e., Step 2, still carries through. The generalizations concern more general error structures, in particular correlated errors, and address estimation not only of the residual variance but also of residual autocovariances. The results and methods used in this section draw on results and methods used in Findley, Pötscher and Wei (1997; see the comments below for more discussion).

Let $X = (x_{ti,n})$ be a (possibly) random $n \times k_n$ matrix, and let $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)'$ be a random $n \times 1$ vector with finite second moments and $E(\varepsilon) = 0$. (The mean zero assumption is only made so that we can talk about variances and autocovariances rather than about noncentral second moments. It is not necessary for the mathematical development below.) Let Σ_n denote $E(\varepsilon\varepsilon')$, and put $\hat{\varepsilon} = (I - P_X)\varepsilon$, the ordinary least squares residual vector. Consider for $j \geq 0$ the estimator $\hat{\gamma}_{n,j} = n^{-1}\sum_{t=1}^{n-j} \hat{\varepsilon}_t \hat{\varepsilon}_{t+j}$ and its unobservable analogue $\tilde{\gamma}_{n,j} = n^{-1}\sum_{t=1}^{n-j} \varepsilon_t \varepsilon_{t+j}$, with the convention that $\hat{\gamma}_{n,j} = \tilde{\gamma}_{n,j} = 0$ if $j \geq n$ (e.g., if ε_t is stationary, then $\hat{\gamma}_{n,j}$ will be an estimator for the autocovariance $E(\varepsilon_t \varepsilon_{t+j})$). Similarly as in Step 1, convergence or asymptotic distributional results for the unobservable $\tilde{\gamma}_{n,j}$ often follow easily from laws of large numbers or central limit theorems, and hence analogous results will hold for the corresponding estimator $\hat{\gamma}_{n,j}$ if the difference $\tilde{\gamma}_{n,j} - \hat{\gamma}_{n,j}$ can be shown to converge to zero (at an appropriate rate). Note that in case of residual variance estimation, i.e., $j = 0$, the difference $\tilde{\gamma}_{n,0} - \hat{\gamma}_{n,0}$ precisely equals $n^{-1}\varepsilon' P_X \varepsilon$, the quantity appearing in Step 1. In what follows, δ_n always denotes a sequence of positive real numbers that is bounded away from zero.

We first show how the differences $\tilde{\gamma}_{n,j} - \hat{\gamma}_{n,j}$ can be bounded by $\tilde{\gamma}_{n,0} - \hat{\gamma}_{n,0}$, thus allowing one to concentrate on the latter difference. In the following lemma, we use the convention that an empty sum is assigned the value zero.

LEMMA. *Let $u_{t,n}$ and $v_{t,n}$, $1 \leq t \leq n$, $n \geq 1$, be triangular arrays of random variables. Then*

$$\max_{0 \leq j < \infty} n^{-1} \left| \sum_{t=1}^{n-j} u_{t,n} u_{t+j,n} - \sum_{t=1}^{n-j} v_{t,n} v_{t+j,n} \right|$$
$$\leq n^{-1} \sum_{t=1}^{n} (u_{t,n} - v_{t,n})^2 + 2 \left\{ n^{-1} \sum_{t=1}^{n} (u_{t,n} - v_{t,n})^2 \right\}^{1/2} \left\{ n^{-1} \sum_{t=1}^{n} u_{t,n}^2 \right\}^{1/2}.$$

Proof. We have

$$\sum_{t=1}^{n-j} v_{t,n} v_{t+j,n} - \sum_{t=1}^{n-j} u_{t,n} u_{t+j,n}$$
$$= \sum_{t=1}^{n-j} (v_{t,n} - u_{t,n})(v_{t+j,n} - u_{t+j,n})$$
$$+ \sum_{t=1}^{n-j} (v_{t,n} - u_{t,n}) u_{t+j,n} + \sum_{t=1}^{n-j} u_{t,n} (v_{t+j,n} - u_{t+j,n}).$$

https://doi.org/10.1017/S0266466698144079 Published online by Cambridge University Press

Applying the triangle and the Cauchy–Schwarz inequality, we obtain

$$\begin{aligned}
&\left|\sum_{t=1}^{n-j} u_{t,n}u_{t+j,n} - \sum_{t=1}^{n-j} v_{t,n}v_{t+j,n}\right| \\
&\quad\le \left\{\sum_{t=1}^{n-j}(v_{t,n}-u_{t,n})^2\right\}^{1/2}\left\{\sum_{t=1}^{n-j}(v_{t+j,n}-u_{t+j,n})^2\right\}^{1/2} \\
&\qquad + \left\{\sum_{t=1}^{n-j}(v_{t,n}-u_{t,n})^2\right\}^{1/2}\left\{\sum_{t=1}^{n-j}u_{t+j,n}^2\right\}^{1/2} \\
&\qquad + \left\{\sum_{t=1}^{n-j}u_{t,n}^2\right\}^{1/2}\left\{\sum_{t=1}^{n-j}(v_{t+j,n}-u_{t+j,n})^2\right\}^{1/2} \\
&\quad\le \sum_{t=1}^{n}(u_{t,n}-v_{t,n})^2 + 2\left\{\sum_{t=1}^{n}(u_{t,n}-v_{t,n})^2\right\}^{1/2}\left\{\sum_{t=1}^{n}u_{t,n}^2\right\}^{1/2}.
\end{aligned}$$

■

The lemma shows that the distance between the empirical second cross-moments of the arrays $u_{t,n}$ and $v_{t,n}$ can be bounded by (a function of) the size of $n^{-1}\sum_{t=1}^{n}(u_{t,n}-v_{t,n})^2$, which itself can be viewed as a measure of closeness between the arrays. For example, it follows that, given $n^{-1}\sum_{t=1}^{n}u_{t,n}^2 = O_p(1)$ [$O(1)$ a.s.], the difference between the empirical second cross-moments of $u_{t,n}$ and $v_{t,n}$ converges to zero i.p. [a.s.] if $n^{-1}\sum_{t=1}^{n}(u_{t,n}-v_{t,n})^2$ has the same property.

In the regression context, i.e., when $u_{t,n}=\varepsilon_t$, $v_{t,n}=\hat{\varepsilon}_t$, $1\le t\le n$, the measure of closeness becomes $n^{-1}\sum_{t=1}^{n}(\varepsilon_t-\hat{\varepsilon}_t)^2$. Because $\hat{\varepsilon}$ is obtained through projection, this expression coincides with $\tilde{\gamma}_{n,0}-\hat{\gamma}_{n,0}$. Hence, the lemma translates into

$$\max_{0\le j<\infty}|\tilde{\gamma}_{n,j}-\hat{\gamma}_{n,j}| \le (\tilde{\gamma}_{n,0}-\hat{\gamma}_{n,0}) + 2(\tilde{\gamma}_{n,0}-\hat{\gamma}_{n,0})^{1/2}(n^{-1}\varepsilon'\varepsilon)^{1/2}, \tag{1}$$

which provides the required upper bound in terms of $\tilde{\gamma}_{n,0}-\hat{\gamma}_{n,0}$. Of course, $\tilde{\gamma}_{n,0}-\hat{\gamma}_{n,0}$ is also a lower bound for $\max_{0\le j<\infty}|\tilde{\gamma}_{n,j}-\hat{\gamma}_{n,j}|$. Consequently, we obtain that convergence of $\tilde{\gamma}_{n,0}-\hat{\gamma}_{n,0}$ to zero i.p. [a.s.] is necessary and sufficient for convergence to zero of $|\tilde{\gamma}_{n,j}-\hat{\gamma}_{n,j}|$ i.p. [a.s.] for all $j\ge 0$ (even uniformly in j), provided $n^{-1}\varepsilon'\varepsilon$ is $O_p(1)$ [$O(1)$ a.s.]. Furthermore, $\delta_n(\tilde{\gamma}_{n,0}-\hat{\gamma}_{n,0})\to 0$ i.p. [a.s.] implies $\delta_n^{1/2}|\tilde{\gamma}_{n,j}-\hat{\gamma}_{n,j}| = o_p(1)$ [$o(1)$ a.s.] for all $j\ge 0$ (even uniformly in j), provided $n^{-1}\varepsilon'\varepsilon$ is $O_p(1)$ [$O(1)$ a.s.].

As already noted, observe that $\tilde{\gamma}_{n,0}-\hat{\gamma}_{n,0} = n^{-1}\varepsilon' P_X\varepsilon$. Hence, for the more general problem of residual autocovariance estimation, the crucial step is to establish (a bound for) the rate of convergence to zero of $n^{-1}\varepsilon' P_X\varepsilon$. This will then imply (a bound for) the rate of convergence of $\tilde{\gamma}_{n,j}-\hat{\gamma}_{n,j}$ in view of (1). This leads to the following result which is a generalization of the result in Step 1.

THEOREM. *Suppose X is independent of ε for every n, and let λ_n denote the largest eigenvalue of Σ_n. Then the following holds:*

https://doi.org/10.1017/S0266466698144079 Published online by Cambridge University Press

(a) $n^{-1}E\varepsilon' P_X \varepsilon \leq \lambda_n k_n/n$.
(b) *If* $\delta_n \lambda_n k_n = o(n)$, *then* $\delta_n |\tilde{\gamma}_{n,0} - \hat{\gamma}_{n,0}| \to 0$ *in* L^1, *and hence in probability, as* $n \to \infty$.
(c) *If* $\delta_n \lambda_n k_n = o(n)$ *and* $n^{-1}\operatorname{tr}(\Sigma_n) = O(1)$, *then* $\max_{0 \leq j < \infty} \delta_n^{1/2} |\tilde{\gamma}_{n,j} - \hat{\gamma}_{n,j}| \to 0$ *in* L^1, *and hence in probability, as* $n \to \infty$.
(d) *If* $\delta_n k_n = o(n)$ *and if the largest eigenvalue of* Σ_n *is bounded, i.e.,* $\lambda_n \leq M < \infty$ *for all* n, *then* $\delta_n |\tilde{\gamma}_{n,0} - \hat{\gamma}_{n,0}| \to 0$ *and* $\max_{0 \leq j < \infty} \delta_n^{1/2} |\tilde{\gamma}_{n,j} - \hat{\gamma}_{n,j}| \to 0$ *in* L^1, *and hence in probability, as* $n \to \infty$.

Proof. Part (a) follows from $E(E(\varepsilon' P_X \varepsilon | X)) = E(E(\operatorname{tr}(\varepsilon' P_X \varepsilon) | X)) = E(\operatorname{tr}(P_X E(\varepsilon\varepsilon' | X))) = E(\operatorname{tr}(P_X \Sigma_n)) \leq E(\lambda_n \operatorname{tr}(P_X)) \leq \lambda_n k_n$. Part (b) follows immediately from part (a), noting that $\tilde{\gamma}_{n,0} - \hat{\gamma}_{n,0}$ equals $n^{-1}\varepsilon' P_X \varepsilon$ and is nonnegative. Part (c) follows from part (b), the Cauchy–Schwarz inequality applied to (1), and $n^{-1}\operatorname{tr}(\Sigma_n) = O(1)$. Part (d) follows from (b) and (c), observing that $n^{-1}\operatorname{tr}(\Sigma_n) \leq \lambda_n$. ■

Comments.

(i) If the condition $n^{-1}\operatorname{tr}(\Sigma_n) = O(1)$ is replaced by the weaker condition $n^{-1}\varepsilon'\varepsilon = O_p(1)$, then part (c) still holds with convergence in probability. Both conditions are clearly satisfied if the variances of ε_t are bounded (as is, e.g., the case under weak stationarity).
(ii) The condition $\lambda_n \leq M < \infty$ for all n is certainly satisfied if the errors are uncorrelated and their variances are bounded. Another sufficient condition for boundedness of λ_n is provided by the case in which the errors ε_t form a mean zero weakly stationary process with an essentially bounded spectral density function f. To see that this is a sufficient condition, observe that $\lambda_n = \sup_{w'w=1} w'\Sigma_n w = \int_{-\pi}^{\pi} |w(\lambda)|^2 f(\lambda)\, d\lambda$, where $w(\lambda) = \sum_{j=1}^{n} w_j \exp(-ij\lambda)$. Hence, if C is an essential bound for f, we obtain $\lambda_n \leq C \int_{-\pi}^{\pi} |w(\lambda)|^2\, d\lambda \leq 2\pi C$.
(iii) The results presented above continue to hold without any changes if ε_t is allowed to depend on n. Also, the independence between X and ε can be replaced by the weaker assumption that $E(\varepsilon\varepsilon' | X) = E(\varepsilon\varepsilon')$.
(iv) In the important case in which $k_n - k$ and $\lambda_n \leq M < \infty$ for all n, the condition on δ_n in the theorem is implied by $\delta_n/n \to 0$. The theorem then implies, in particular, that consistency results for $\hat{\gamma}_{n,j}$ can immediately be deduced from corresponding results for $\tilde{\gamma}_{n,j}$, which typically follow from laws of large numbers. Concerning distributional convergence results, note that part (d) then implies $n^{1/2}(\tilde{\gamma}_{n,0} - \hat{\gamma}_{n,0}) = o_p(1)$, and hence $\hat{\gamma}_{n,0}$ will satisfy a central limit theorem if $\tilde{\gamma}_{n,0}$ does so. In contrast, we cannot deduce $n^{1/2}(\tilde{\gamma}_{n,j} - \hat{\gamma}_{n,j}) = o_p(1)$ for $j > 0$ from part (d) of the theorem. In fact, there is no hope to obtain such a result from (1) because this would require establishing $\varepsilon' P_X \varepsilon = o_p(1)$ (note that $n^{-1}\varepsilon'\varepsilon$ will typically be bounded away from zero). This is, however, impossible in general because, under typical assumptions on X and ε, the quantity $\varepsilon' P_X \varepsilon$ has a nondegenerate limit distribution (frequently the distribution of a multiple of a chi-square). (Of course, this does not mean that $n^{1/2}(\tilde{\gamma}_{n,j} - \hat{\gamma}_{n,j}) = o_p(1)$ is impossible or that a central limit theorem for $\tilde{\gamma}_{n,j}$, $\hat{\gamma}_{n,j}$ cannot hold if $j > 0$.)
(v) Related results: The results above are closely related to results in Findley et al. (1997). For example, the result given in the lemma represents one of the steps in

https://doi.org/10.1017/S0266466698144079 Published online by Cambridge University Press

the proof of Proposition 1.1 in Findley et al. (1997; cf. with (A.11) in that paper). In Section 6.1 of that paper, a.s. convergence of $\hat{\gamma}_{n,j}$ to the same limit as $\tilde{\gamma}_{n,j}$ is established for a wide class of dependent processes, without imposing any conditions on the (nonstochastic) regressors. This result substantially extends Gleser's (1966) result for i.i.d. errors. Convergence in probability results, similar to the ones presented above, and a related consistency result for $\hat{\gamma}_{n,j}$ in case of Gaussian errors due to Štulajter (1991) are discussed in Section 6.2 of Findley et al. (1997).

2. Solution, proposed by Xing Ming.

Notice that S^2 is an unbiased estimator of σ^2 and that the following asymptotic results hold: $\varepsilon^T X/n \xrightarrow{p} 0$, $n/(n-k) \to 1$, $k\sqrt{n}\sigma^2/(n-k) \to 0$, $X^T X/n \xrightarrow{p} \Sigma_x$, $X^T\varepsilon/\sqrt{n} \xrightarrow{d} N(0, \sigma^2\Sigma_x)$. But

$$\begin{aligned}\sqrt{n}(S^2-\sigma^2) &= \sqrt{n}\left(\frac{\varepsilon^T\varepsilon - \varepsilon^T X(X^TX)^{-1}X^T\varepsilon}{n-k} - \sigma^2\right)\\ &= \frac{\varepsilon^T\varepsilon - n\sigma^2}{\sqrt{n}}\cdot\frac{n}{n-k} + \frac{k\cdot\sqrt{n}}{n-k}\sigma^2\\ &\quad - \left(\frac{\varepsilon^T X}{n}\cdot\frac{n}{n-k}\right)\left(\frac{X^TX}{n}\right)^{-1}\cdot\frac{X^T\varepsilon}{\sqrt{n}}.\end{aligned}$$

Hence, we only need to prove that

$$\frac{\varepsilon^T\varepsilon - n\sigma^2}{\sqrt{n}} \xrightarrow{d} N(0, \mu_4 - \sigma^4), \tag{2}$$

where $\mu_4 - \sigma^4$ is the variance of ε_i^2. We prove (2) by using the Lindeberg–Feller central limit theorem by checking the Lindeberg–Feller condition (Spanos, 1995, p. 174). Because ε_i^2, $i = 1,2,3,\ldots$, are positive, have constant mean σ^2, and constant variance $\mu_4 - \sigma^4$, the Lindeberg–Feller condition for any $\delta > 0$ becomes,

$$\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}\int_{x^2>\sigma^2+\sqrt{n}\delta}(x^2-\sigma^2)^2 f_i(x)\,dx = 0. \tag{3}$$

The LHS of (3) is equal to

$$\begin{aligned}\lim_{n\to\infty}\Bigg[&\frac{1}{n}\sum_{i=1}^{n}\int_{x^2>c_n} x^4 f_i(x)\,dx\\ &- \frac{2\sigma^2}{n}\sum_{i=1}^{n}\int_{x^2>c_n} x^2 f_i(x)\,dx + \frac{\sigma^4}{n}\sum_{i=1}^{n}\int_{x^2>c_n} f_i(x)\,dx\Bigg],\end{aligned}$$

where $c_n = \sigma^2 + \sqrt{n}\delta$. As long as we can claim that the first term of the above expression converges to 0, the second and the third terms will converge to 0 automatically. For

https://doi.org/10.1017/S0266466698144079 Published online by Cambridge University Press

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \int_{x^2>c_n} x^4 f_i(x)\, dx = 0$$

to hold, we need to make the following assumption.

$$\lim_{a\to\infty} \max_{1\le i<\infty} \int_{|x|>a} x^4 f_i(x)\, dx = 0, \tag{4}$$

which is stronger than the assumption suggested by the author

$$\lim_{a\to\infty} \max_{1\le i<n} \int_{|x|>a} x^4 f_i(x)\, dx = 0 \text{ for any } n. \tag{5}$$

I would like to point out that assumption (5) is equivalent to the assumption that $E\varepsilon_i^4 = \mu_4$ for all i, i.e., it is not an additional condition. Thus, questions we should ask are: (a) do we really need a stronger condition (4) for the asymptotic result to hold and (b) does there exist such a sequence that satisfies (5) but not satisfies (4)? Because condition (4) is equivalent to the Lindeberg–Feller condition (3), the answer to (a) is yes. The example that follows satisfies condition (5) but not condition (4). Thus, the answer to (b) is also yes.

Example. Define ε_1 to be any random variable with $E\varepsilon_1 = 0$, $\text{var}(\varepsilon_i) = \sigma^2 = 1/8$ and $E\varepsilon_1^4 = \mu_4 = 1/4$. Define ε_i, for $i \ge 2$, as follows,

$$\varepsilon_i = \begin{cases} 1 & \text{with prob.} \quad p_1 = (i-2)/16(i-1) \\ -1 & \text{with prob.} \quad p_2 = (i-2)/16(i-1) \\ \sqrt{i} & \text{with prob.} \quad p_3 = 1/16i(i-1) \\ -\sqrt{i} & \text{with prob.} \quad p_4 = 1/16i(i-1) \\ 0 & \text{otherwise,} \end{cases}$$

from which we obtain

$$\varepsilon_i^2 = \begin{cases} 1 & \text{with prob.} \quad (i-2)/8(i-1) \\ i & \text{with prob.} \quad 1/8i(i-1) \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Then we have

$$E\varepsilon_i = 0,$$

$$E\varepsilon_i^2 = p_1 + p_2 + ip_3 + ip_4 = 1/8 = \sigma^2,$$

$$E\varepsilon_i^4 = p_1 + p_2 + i^2 p_2 + i^2 p_4 = 1/4 = \mu_4.$$

Hence, sequence ε_i^2, $i = 1,2,3,\ldots$, satisfies the requirements proposed by the author, but it does not satisfy the Lindeberg–Feller condition, which is equivalent

https://doi.org/10.1017/S0266466698144079 Published online by Cambridge University Press

to condition (4). From (6), we see that for any δ, the LHS of (3) will be no less than

$$\lim_{n\to\infty} \frac{1}{n} \sum_{i=2}^{n} \frac{(i-1/8)^2}{8i(i-1)} = \frac{1}{8}.$$

Hence, the central limit theorem does not hold. One can also check directly that for this special sequence condition (4) does not hold.

REFERENCES

Drygas, H. (1975) A note on a paper by T. Kloek concerning the consistency of variance estimation in the linear model. *Econometrica* 43, 175.

Gleser, L.J. (1966) Correction to "On the asymptotic theory of fixed-size sequential confidence bounds for linear regression parameters." *Annals of Mathematical Statistics* 37, 1053–1054.

Kloek, T. (1972) Note on consistent estimation of the variance of the disturbances in the linear model. *Econometrica* 40, 911–912.

Findley, D.F., B.M. Pötscher, & C.Z. Wei (1997) Convergence Results for Short-Memory Modeling of Time Series Arrays by Multistep Prediction or Likelihood Methods, with an Application to Model Selection. Unpublished manuscript.

Schmidt, W.H. (1976) Strong consistency of variance estimation and asymptotic theory for tests of the linear hypothesis in multivariate linear models. *Mathematische Operationsforschung und Statistik* 7, 701–705.

Spanos, A. (1995) *Statistical Foundations of Econometric Modelling*. Cambridge: Cambridge University Press.

Štulajter, F. (1991) Consistency of linear and quadratic least squares estimators in regression models with covariance stationary errors. *Applications of Mathematics* 36, 149–155.

97.4.3. *Multivariate Regression Subject to Orthogonality Conditions*—Solution,[1] proposed by Michel van de Velden and Heinz Neudecker. Consider the problem

$$\min_{\mathbf{B}} \operatorname{tr}(\mathbf{Y} - \mathbf{XB})'(\mathbf{Y} - \mathbf{XB}), \tag{1}$$

where $\mathbf{Y}$ is a given $n \times p$ matrix, $\mathbf{X}$ is a given $n \times q$ matrix of full column rank, and $\mathbf{B}$ is an unknown $q \times p$ matrix.

Without restrictions on $\mathbf{B}$, we have the solution

$$\mathbf{B} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}. \tag{2}$$

We will solve (1) when $\mathbf{B}$ is subjected to one of the following normalizations:

1. $\mathbf{B}'\mathbf{B} = \mathbf{I}_p$,
2. $\mathbf{BB}' = \mathrm{I}_q$.

Note that in both cases we have

$$\mathbf{B}^+ = \mathbf{B}', \tag{3}$$

where $\mathbf{B}^+$ denotes the Moore–Penrose inverse.

https://doi.org/10.1017/S0266466698144079 Published online by Cambridge University Press

It follows from (2) and (3) that

$$[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}]^{+} = \mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}. \tag{4}$$

For the Moore–Penrose inverse of a matrix $\mathbf{A}$, we have

$$\mathbf{A}^{+} = (\mathbf{A}'\mathbf{A})^{+}\mathbf{A}' = \mathbf{A}'(\mathbf{A}\mathbf{A})')^{+}.$$

Use these equalities to see that (4) yields

$$\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = [\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-2}\mathbf{X}'\mathbf{Y}]^{+}\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1},$$

and

$$\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}]^{+}.$$

Now, consider the two cases:

1. rank of $\mathbf{X}'\mathbf{Y}$ is p or equivalently $\mathbf{Y}'\mathbf{X}\mathbf{X}'\mathbf{Y} > 0, \quad (p \le q \le n)$
2. rank of $\mathbf{Y}'\mathbf{X}$ is q or equivalently $\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X} > 0, \quad (q \le p \le n)$

Then, for case 1 we get

$$\begin{aligned}\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} &= [\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-2}\mathbf{X}'\mathbf{Y}]^{-1}\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} \\ &\rightarrow \mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-2}\mathbf{X}'\mathbf{Y} = \mathbf{I}_p, \end{aligned} \tag{5}$$

and for case 2 we get

$$\begin{aligned}\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} &= \mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}[(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}]^{-1} \\ &= \mathbf{Y}'\mathbf{X}(\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X} \rightarrow (\mathbf{X}'\mathbf{X})^{-1} = (\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X})^{-1/2}. \end{aligned} \tag{6}$$

We see that, when $\mathbf{X}'\mathbf{Y}$ has full row rank (case 2), then

$$\mathbf{B} = (\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X})^{-1/2}\mathbf{X}'\mathbf{Y}.$$

This matrix satisfies the constraint $\mathbf{B}\mathbf{B}' = \mathbf{I}_q$.

It is clear that $\mathbf{B}'\mathbf{B} = \mathbf{I}_p$ when $\mathbf{X}'\mathbf{Y}$ has full column rank (case 1). This follows immediately from (5). A unique expression for $\mathbf{B}$ is not obtainable.

For the special case where $p = q$ and $\mathbf{X}'\mathbf{Y}$ is nonsingular, we have

$$\mathbf{B} = (\mathbf{X}'\mathbf{Y}\mathbf{Y}'\mathbf{X})^{-1/2}\mathbf{X}'\mathbf{Y} = \mathbf{X}'\mathbf{Y}(\mathbf{Y}'\mathbf{X}\mathbf{X}'\mathbf{Y})^{-1/2}. \tag{7}$$

The proof is immediate, e.g., by the singular value decomposition theorem or by the Schur decomposition theorem.

The problem can also be solved by using a standard optimization technique. For example, rewrite the problem in (1) as

$$\max_{\mathbf{B}} 2 \operatorname{tr} \mathbf{Y}'\mathbf{X}\mathbf{B} - \operatorname{tr} \mathbf{B}'\mathbf{X}'\mathbf{X}\mathbf{B},$$

$$\text{s.t. } \mathbf{B}'\mathbf{B} = \mathbf{I}_p.$$

https://doi.org/10.1017/S0266466698144079 Published online by Cambridge University Press

Define the Lagrangean function

$$\phi = 2 \operatorname{tr} \mathbf{Y}'\mathbf{X}\mathbf{B} - \operatorname{tr} \mathbf{B}'\mathbf{X}'\mathbf{X}\mathbf{B} - \operatorname{tr} \mathbf{L}(\mathbf{B}'\mathbf{B} - \mathbf{I}_p).$$

We consider variations in $\mathbf{B}$, hence

$$d\phi = 2 \operatorname{tr} \mathbf{Y}'\mathbf{X}d\mathbf{B} - 2 \operatorname{tr} \mathbf{B}'\mathbf{X}'\mathbf{X}d\mathbf{B} - 2 \operatorname{tr} \mathbf{L}\mathbf{B}'d\mathbf{B},$$

where we have used the symmetry of the matrix $\mathbf{L}$ of Lagrangean multipliers. Equating to zero yields a first-order condition together with the constraint

$$\mathbf{Y}'\mathbf{X} = \mathbf{B}'\mathbf{X}'\mathbf{X} + \mathbf{L}\mathbf{B}' \tag{8}$$

$$\mathbf{B}'\mathbf{B} = \mathbf{I}_p.$$

Postmultiply (8) by $\mathbf{B}$ to get

$$\mathbf{Y}'\mathbf{X}\mathbf{B} = \mathbf{B}'\mathbf{X}'\mathbf{X}\mathbf{B} + \mathbf{L}.$$

Due to the symmetry of $\mathbf{L}$, we have

$$(\mathbf{B}'\mathbf{X}'\mathbf{X}\mathbf{B} + \mathbf{L})^2 = \mathbf{Y}'\mathbf{X}\mathbf{B}\mathbf{B}'\mathbf{X}'\mathbf{Y}.$$

Hence

$$(\mathbf{Y}'\mathbf{X}\mathbf{B}\mathbf{B}'\mathbf{X}'\mathbf{Y})^{1/2} = \mathbf{Y}'\mathbf{X}\mathbf{B} = \mathbf{B}'\mathbf{X}'\mathbf{Y}. \tag{9}$$

For the case $p = q$, we have $\mathbf{B}\mathbf{B}' = \mathbf{I}_p$, so that

$$(\mathbf{Y}'\mathbf{X}\mathbf{X}'\mathbf{Y})^{1/2} = \mathbf{B}'\mathbf{X}'\mathbf{Y}.$$

Now, if we assume the rank of $\mathbf{Y}'\mathbf{X}$ to be p, we obtain expression (7), i.e.,

$$\mathbf{B} = \mathbf{X}'\mathbf{Y}(\mathbf{Y}'\mathbf{X}\mathbf{X}'\mathbf{Y})^{-1/2}.$$

NOTE

1. Excellent solutions have been proposed independently by S. Lawford and by R.W. Farebrother, the poser of the problem.

https://doi.org/10.1017/S0266466698144079 Published online by Cambridge University Press