CESifo *Working Paper Series*

REEXAMINING THE ESTIMATION OF SIMULTANEOUS EQUATIONS SYSTEMS

Yoshio Kimura

Working Paper No. 187

May 1999

CESifo
Poschingerstr. 5
81679 Munich
Germany
Phone: +49 (89) 9224-1344
Fax: +49 (89) 9224-2344
http://www.lrz-muenchen.de/~ces/c02.htm

CESifo *Working Paper No. 187*
May 1999

REEXAMINING THE ESTIMATION OF SIMULTANEOUS EQUATIONS SYSTEMS

Abstract

This paper generalises a classical theorem on the minimisation of the ratio of two quadratic forms so as to permit the denominator to be nonnegative definite, provides a modified formula for the minimum variance ratio estimation including the limited information maximum likelihood, and collaterally shows that the use of the principal components of some predetermined variables in the first stage of two stage least squares is afraid of leading to biased estimators.

Yoshio Kimura
Chukyo University
Institute of Economics
101-2 Yagoto-Honmachi
Showa-Ku, Nagoya
Japan

Reexamining the Estimation of Simultaneous Equations Systems

Yoshio Kimura
Chukyo University
Faculty of Economics

The purpose of this paper lies primarily in strengthening limited information maximum likelihood(LIML) estimation by extending the existing theorem for minimising the ratio of two quadratic forms and secondarily in reexamining two stage least squares(TSLS) proposed by Kloek and Mennes(1960) under the assumption that the matrix of the observations on the predetermined variables has the rank not necessarily greater than the number of all predetermined variables.

As will be shown by the numerical examples presented later, the existing theorem aforementioned(henceforth, PD-theorem, for short) is often unapplicable to the first step of LIML which requires to minimise a ratio of two variances expressed as quadratic forms, since the positive definiteness of the denominator, an essential assumption of PD-theorem, is apt to collapse even when the sample size is enough large to exceed the number of all predetermined variables. Moreover, the first step of the LIML modified by Fuller(1977), Fujikoshi and others(1982) and Morimune(1983) remains intact. Therefore, the extension of PD-theorem is inevitable for securing the wider applicability of both LIML and modified LIML.

Turning to TSLS, Kloek and Mennes(1960) proposed to use the principal components of predetermined variables in the first stage of TSLS, the least squares estimation of the reduced form, to overcome the apparent difficulty occurred when the sample size is smaller than the number of all predetermined variables. However, it would be natural to reexamine their proposals since the normal equations of ordinary least squares(OLS) are known always solvable. To achieve these ends, we proceed as follows.

Section 1 is concerned with the revised LIML based on the generalised PD-theorem.

In Section 2, we reexamine the proposeals of Kloek and Mennes(1960) from the view point of the generalised inverses of a matrix and refer to a characterisation of the so called Ridge estimation, a variant of OLS.

A brief summary of the arguments of Sections 1 and 2 as well as the problems to be investigated further will be stated in Section 3.

Appendix 1 is devoted to the generalisation of PD-theorem, admitting the singularity of the denominator.

Finally, in Appendix 2, we state and prove the properties of generalised inverses of matrices so as to make the paper self-contained.

1. On the Revision of Limited Information Maximum Likelihood Estimation

Throughout the paper, we consider a standard linear statistical model consisting of G structural relations with K predetermined variables. In matrix form, the model is written as

$$\mathbf{YB}' + \mathbf{X}\,\Gamma = \mathbf{U}\ , \tag{1}$$

where n signifies the sample size, $\mathbf{Y}$ = an $n \times G$ matrix of observations on the endogenous variables, $\mathbf{B}$ = a $G \times G$ matrix of coefficients of current endogenous variables, $\mathbf{X}$ = an $n \times K$ matrix of observations on all predetermined variables, Γ = a $G \times K$ matrix of coefficients of all predetermined variables, $\mathbf{U}$ = an $n \times G$ matrix of all the sample disturbances, and $\mathbf{B}'(\Gamma')$ is of course the transposition of $\mathbf{B}(\Gamma)$.

As is well known, it is customary to assume that n is enough large to ensure that the rank of $\mathbf{X}$($r(\mathbf{X})$, in symbol) equals K. However, we dare to leave $r(\mathbf{X})$ unspecified, because in handling multiregional econometric models and/or econometric models of large scale it often occurs that we are obliged to be contended with the sample size insufficient to secure the assumption $r(\mathbf{X}) = K$ and mainly from this reason we substitute the generalised inverse(g-inverse) of a matrix for the usual inverse of a matrix.

Since LIML and TSLS estimate each equation contained in the model one by one, it is convenient to express a representative equation to be estimated as

$$\mathbf{y} = \mathbf{Y}_1 \beta + \mathbf{X}_1 \gamma + \mathbf{u}\ , \tag{2}$$

where y is an $n \times 1$ vector of observations on the dependent variable of equation (2), $\mathbf{X}_1(\mathbf{Y}_1)$ is the $n \times K_1$($n \times (g\text{-}1)$) matrix of observations on explanatory predetermined(endogenous) variables appearing in this equation and $\mathbf{u}$ is an $n \times 1$ vector of sample disturbances of $\mathbf{y}$.

Moreover, by $\mathbf{Y}_\Delta$ and β'_Δ denote $(n \times g)$ matrix $(\mathbf{y} \vdots \mathbf{Y}_1)$ and $(1 \times g)$ vector $(1, -\beta')$. Then (2) is equivalent to

$$\mathbf{Y}_\Delta \beta_\Delta - \mathbf{X}_1 \gamma = \mathbf{u}\ . \tag{3}$$

On the other hand, for the observations of the remaining variables, we specify $\mathbf{Y}_2(\mathbf{X}_2)$ as the submatrix obtained from $\mathbf{Y}(\mathbf{X})$ by deleting $\mathbf{Y}_\Delta(\mathbf{X}_1)$.

The main argument of this section begins with summarising the usual computational scheme of LIML. Following Johnston(1983, pp.483-486), the scheme consists of two procedures.

<u>Procedure (i)</u> To find the minimum value $\underline{\ell}$ of the ratio

$$\ell = (\beta'_\Delta \mathbf{W}_1 \beta_\Delta) / (\beta'_\Delta \mathbf{W} \beta_\Delta)\ , \tag{4}$$

where $\mathbf{W}_1 = \mathbf{Y}'_\Delta(\mathbf{I} - \mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1)^- \mathbf{X}'_1)\mathbf{Y}_\Delta$, $\mathbf{W} = \mathbf{Y}'_\Delta(\mathbf{I} - \mathbf{X}(\mathbf{X}'\mathbf{X})^- \mathbf{X}')\mathbf{Y}_\Delta$, $(\mathbf{X}'\mathbf{X})^-$ is the generalised inverse(g-inverse) of $\mathbf{X}'\mathbf{X}$(similarly for $(\mathbf{X}'_1\mathbf{X}_1)^-$), and $\mathbf{I}$ of course denotes an identity matrix whose order is to be understood from the context.

For later convenience, we further introduce $\mathbf{H}_1 = \mathbf{X}_1(\mathbf{X}'_1\mathbf{X}_1)^- \mathbf{X}'_1$ and

$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$. Conventionally, assuming implicitly that the denominator of (4)(or equivalently $\mathbf{W}$) is positive definite, $\underline{\ell}$ is obtained as the minimum root of the following determinantal equation

$$\left| \mathbf{W}_1 - \ell \mathbf{W} \right| = 0 \ . \tag{5}$$

Once $\underline{\ell}$ is obtained, $\hat{\beta}_\Delta$ can be found as an eigen vector satisfying

$$(\mathbf{W}_1 - \underline{\ell}\mathbf{W})\hat{\beta}_\Delta = 0 \, . \tag{6}$$

<u>Procedure (ii)</u> Using $\hat{\beta}_\Delta$ obtained above, we have the estimate $\hat{\gamma}$ of γ as

$$\hat{\gamma} = (\mathbf{X}_1'\mathbf{X}_1)^{-}\mathbf{X}_1'\mathbf{Y}_\Delta\hat{\beta}_\Delta \, . \tag{7}$$

We now examine the positive definiteness of $\mathbf{W}$. By assertions (ii-1), (ii-2.2) and (ii-2.3) of Theorem AP.2, $(\mathbf{I} - \mathbf{H})$ is idempotent as well as symmetric and is of rank n-r($\mathbf{X}$). Hence, $\mathbf{W} = ((\mathbf{I} - \mathbf{H})\mathbf{Y}_\Delta)'((\mathbf{I} - \mathbf{H})\mathbf{Y}_\Delta)$, which is clearly nonnegative definite, is positive definite if and only if $r((\mathbf{I} - \mathbf{H})\mathbf{Y}_\Delta) = g$. In addition, it does not harm the generality to assume that $r(\mathbf{Y}_\Delta) = g$, for otherwise one column of $\mathbf{Y}_\Delta$ is expressed as the linear combination of the others, which in turn makes the estimation unnecessary. Therefore

$$r((\mathbf{I} - \mathbf{H})\mathbf{Y}_\Delta) \leqq \min\{\, r(\mathbf{I} - \mathbf{H}), r(\mathbf{Y}_\Delta)\,\} = \min\{\, n\text{-}r(\mathbf{X}), g\,\} \, .$$

Accordingly, if n-r($\mathbf{X}$) < g, $\mathbf{W}$ is by no means positive definite. Furthermore, the numerical examples are to satisfy the order condition for identifiability which is prerequisite to LIML estimation. Bearing these in mind we enumerate two numerical examples which validate the necessity for the modification of Procedure (i) of LIML by illustrating that the conventionally assumed positive definiteness of $\mathbf{W}$ is apt to collapse.

<u>Example 3</u> (standard case, ie, r($\mathbf{X}$) = K)

Consider the following example ;

n=20, K=17, g=4(three explanatory endogenous variables), K_1=2(two explanatory predetermined variables; hence five explanatory variables in all). Then $\mathbf{W}$ is seen not to be positive definite because

$$n\text{-}r(\mathbf{X}) = 20\text{-}17 = 3 < 4 = g \, ,$$

while the order condition for identifiability is met since

$$K_2 = K\text{-}K_1 = 15 > 4 = g \, .$$

<u>Example 4</u> (n<K, as was in Kloek and Mennes(1960))

Since n is less than K, r($\mathbf{X}$) does not exceed n. If, in addition, n = r($\mathbf{X}$) then the matrix $(\mathbf{I} - \mathbf{H})$ vanishes for $\mathbf{X}$ is of now full rank. Consequently, r($\mathbf{X}$) should be less than is in order for the LIML to be applicable. Thus, we take up the case where n=10, K=12, r($\mathbf{X}$)=9, g=2, and K_1=2. Therefore, we have

$$n\text{-}r(\mathbf{X}) = 10\text{-}9 = 1 < 4 = g \quad (\mathbf{W} \text{ is not positive definite})$$

and

$$K_2 = K\text{-}K_1 = 12\text{-}2 = 10 > 4 = g \, .$$

(the order condition for identifiability holds)

Preliminary considerations being over, we now state and prove our main theorem which modifies Procedure (i) of LIML.

Theorem 1 (i) $N((\mathbf{I}-\mathbf{H}_1)\mathbf{Y}_\Delta) \subseteq N((\mathbf{I}-\mathbf{H})\mathbf{Y}_\Delta)$, and
(ii) the smallest root $\underline{\ell}$ of the determinantal equation

$$|\mathbf{P}'(\mathbf{W}_1-(\mathbf{W}_1\mathbf{Q})(\mathbf{Q}'\mathbf{W}_1\mathbf{Q})^-(\mathbf{Q}'\mathbf{W}_1))\mathbf{P}-\ell(\mathbf{P}'\mathbf{W}\mathbf{P})|=0$$

gives the nontrivial minimum of the variance ratio (4), by employing $(\underline{\mathbf{u}}, \mathbf{v}_{\underline{\mathbf{u}}})$ which satisfies the following two equations

$$\begin{cases} \mathbf{P}'(\mathbf{W}_1-(\mathbf{W}_1\mathbf{Q})(\mathbf{Q}'\mathbf{W}_1\mathbf{Q})^-(\mathbf{Q}'\mathbf{W}_1))\mathbf{P}\underline{\mathbf{u}}=\underline{\ell}(\mathbf{P}'\mathbf{W}\mathbf{P})\underline{\mathbf{u}} \\ (\mathbf{Q}'\mathbf{W}_1\mathbf{Q})\mathbf{v}_{\underline{\mathbf{u}}}=-(\mathbf{Q}'\mathbf{W}_1\mathbf{P}\underline{\mathbf{u}}) , \end{cases}$$

the estimate $\hat{\beta}_\Delta$ of β_Δ is expressed as

$$\hat{\beta}_\Delta=\mathbf{P}\underline{\mathbf{u}}+\mathbf{Q}\mathbf{v}_{\underline{\mathbf{u}}} ,$$

and finally, $\hat{\beta}_\Delta$ is independent of the choice of $\mathbf{P}$ and $\mathbf{Q}$, where $\mathbf{P}$ represents a basis of $R(((\mathbf{I}-\mathbf{H})\mathbf{Y}_\Delta)')$ and $\mathbf{Q}$ a basis of $N((\mathbf{I}-\mathbf{H})\mathbf{Y}_\Delta)$.

Proof (i) For any vector $\mathbf{q}$ of $N((\mathbf{I}-\mathbf{H}_1)\mathbf{Y}_\Delta)$, define $\mathbf{e}_1=(\mathbf{I}-\mathbf{H}_1)\mathbf{Y}_\Delta\mathbf{q}$ and $\mathbf{e}=(\mathbf{I}-\mathbf{H})\mathbf{Y}_\Delta\mathbf{q}$. Then, by definition, $\mathbf{e}_1=\mathbf{0}$ and $\mathbf{e}_1(\mathbf{e})$ is the vector of residuals in the regression of $\mathbf{Y}_\Delta\mathbf{q}$ on $\mathbf{X}_1(\mathbf{X})$. Moreover, it is well known that the sum of squared residuals of ordinary least squares decreases as the number of explanatory variables increases. This, in conjunction with the fact that $\mathbf{e}_1=\mathbf{0}$, yields $0 \leqq \mathbf{e}'\mathbf{e} \leqq \mathbf{e}_1'\mathbf{e}_1=0$. Hence, $\mathbf{0}=\mathbf{e}=(\mathbf{I}-\mathbf{H})\mathbf{Y}_\Delta\mathbf{q}$.

(ii) Applying assertions (ii-1) and (ii-2.3) of Theorem AP.2, both $(\mathbf{I}-\mathbf{H}_1)$ and $(\mathbf{I}-\mathbf{H})$ are idempotent as well as symmetric. Hence, we at once see that

$$R(\mathbf{W})=R(((\mathbf{I}-\mathbf{H})\mathbf{Y}_\Delta)'), \tag{8.1}$$

$$N(\mathbf{W})=N((\mathbf{I}-\mathbf{H})\mathbf{Y}_\Delta), \tag{8.2}$$

and

$$N(\mathbf{W}_1)=N((\mathbf{I}-\mathbf{H}_1)\mathbf{Y}_\Delta). \tag{8.3}$$

(8.1) and (8.2) ensure that the choice of $\mathbf{P}$ and $\mathbf{Q}$ in the theorem is conformable with that of bases in Theorem AP.1, and (8.2) and (8.3) show that the condition of assertion (ii) of Theorem AP.1 is met. Therefore the assertions to be verified are seen to be the direct consequence of assertions (ii) and (iii) of Theorem AP.1.

(Q.E.D.)

Turning to the estimation procedure, there are two paths to follow. For the usual LIML estimation it suffices to exert Procedure (ii) with $\hat{\beta}_\Delta$ given in Theorem 1. On the other hand, to conduct the revised LIML estimation due to Fuller(1977), we are to compute $\tilde{\ell}=\underline{\ell}-\frac{c}{n}$ (Fuller(1977, p.442)) and

$\tilde{\gamma} = (X_1'X_1)^- X_1'Y_\Delta\hat{\beta}_\Delta$, where c is a constant not less than unity, $\hat{\beta}_\Delta = P\tilde{u} + Q\tilde{v}$, and the pair $(\tilde{u}, \tilde{v})$ is determined so that $(P'(W_1 - (W_1Q)(Q'W_1Q)^-(Q'W_1)P)\tilde{u} = \tilde{\ell}(P'WP)\tilde{u}$ and $(Q'W_1Q)\tilde{v} = -(Q'W_1P\tilde{u})$. This completes the modified estimation scheme.

To simplify terminology, let's abbreviate the family of estimates obtained by applying the modified procedures stated above as MMVE(modified minimum variance estimator). Then, the MMVE derived from the data set fulfilling the condition that r(X) = K is known to be the LIML estimator, while it remains indistinct whether the MMVE derived from the data set such that r(X) < K coincides with LIML estimator. Nevertheless, there is found a value of n, say $\underline{n}$, such that r(X) = K for all $n \geqq \underline{n}$ in virtual process where the sample size tends to infinity. Consequently, MMVE becomes LIML estimator after $\underline{n}$ is reached. Resorting again to the consistency of LIML estimator, it is now clear that MMVE is consistent, irrespective of the sample size actually observed. In addition, it is needless to say that MMVE obtained by applying Fuller's formula exhibits the asymptotic properties shown by Fuller(1977) and Morimune(1983).

2. Further Consideration on TSLS

In this section, we reconsider the arguments of Kloek and Mennes(1960) from the view point of applying the generalised inverse to the first stage of TSLS and, as a byproduct, characterise the Ridge estimator in terms of the Moore-Penrose generalised inverse.

Kloek and Mennes(1960;p.49), assuming the singularity of $X'X$, proposed to replace $X = (X_1 \vdots X_2)$ by $Z = (X_1 \vdots F)$, so as to ensure the nonsingularity of $Z'Z$, where X_2 is a submatrix obtained by deleting X_1 from X and F is a suitable principal components of X_2. Let $\Upsilon = (Z'Z)^{-1}Z'Y_1$. Then, Υ is not necessarily uncorrelated to u, since $(Z'Z)^{-1}Z'Y_1$ is not a submatrix of the least squares estimate of the reduced form. Therefore, the estimators obtained by the estimation method Kloek and Mennes(1960) proposed are obliged to be biased and hence they are not always consistent. On the other hand, the least squares estimate $\hat{Y}_1$ of Y_1 regressed on X does exist irrespective of the rank of X and assertions (ii-1),(ii-2.1) and (iii) of Theorem AP.2 assure that $\hat{Y}_1 = X(X'X)^- X'Y_1$ is invariant for any choice of $(X'X)^-$. Clearly $\hat{Y}_1$ is uncorrelated to u because, in view of (1), $V = Y - X(X'X)^- X'Y$ gives the estimate of $U(B')^{-1}$ and because $V'\hat{Y} = [\,\theta\,]$ by the very nature of least squares. Thus the resultant TSLS estimators are undoubtedly consistent even if $X'X$ is singular. Hence, we can say that the proposals of Kloek and Mennes(1960) are of little use and that it suffices to apply g-inverse to each stage of TSLS, paying attention to the use of g-inverse obtainable with the least computational complexity.[1)] Though the use of principal components of X_2 is rather denied, it seems interesting to relate $\hat{Y}_1$ (hereafter, the generalised inverse estimate of Y_1, for short) to the principal components of X.

Let r be the rank of $\mathbf{X}$ and recall that $(\mathbf{X}'\mathbf{X})$ is nonnegative definite as well as symmetric. Then there exists an orthogonal matrix $\mathbf{S}$ such that

$$\mathbf{S}'(\mathbf{X}'\mathbf{X})\mathbf{S} = \begin{pmatrix} \overbrace{\Lambda_r}^{r} & \vdots & \overbrace{\mathbf{0}}^{K-r} \\ \cdots & & \cdots \\ \mathbf{0} & \vdots & \mathbf{0} \end{pmatrix}, \tag{9}$$

where Λ_r is all diagonal matrix with r positive eigen values of $(\mathbf{X}'\mathbf{X})$ on the main diagonal.

By $\mathbf{S}_1(\mathbf{S}_2)$ denote the submatrix consisting of the first r(the remaining K-r) columns of S. Then (9) further implies that

$$(\mathbf{XS}_1)'(\mathbf{XS}_1) = \Lambda_r , \tag{9'}$$

and that

$$\mathbf{XS}_2 = \mathbf{0} . \tag{9''}$$

Moreover the Moore-Penrose generalised inverse of $(\mathbf{X}'\mathbf{X})$ is expressed as

$$(\mathbf{X}'\mathbf{X})^{+} = \mathbf{S}\begin{pmatrix} (\Lambda_r)^{-1} & \vdots & \mathbf{0} \\ \cdots & & \cdots \\ \mathbf{0} & \vdots & \mathbf{0} \end{pmatrix}\mathbf{S}' = \mathbf{S}_1(\Lambda_r)^{-1}\mathbf{S}_1' . \tag{10}$$

In view of the invariance of $\hat{\mathbf{Y}}_1$ for the choice of $(\mathbf{X}'\mathbf{X})^{-}$ and of (9'), we have,

$$\begin{aligned} \hat{\mathbf{Y}}_1 &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}_1 \\ &= \mathbf{X}(\mathbf{X}'\mathbf{X})^{+}\mathbf{X}'\mathbf{Y}_1 \\ &= (\mathbf{XS}_1)(\Lambda_r)^{-1}(\mathbf{XS}_1)'\mathbf{Y}_1 && \Leftarrow\!\!= \text{by (10)} \\ &= (\mathbf{XS}_1)((\mathbf{XS}_1)'(\mathbf{XS}_1))^{-1}(\mathbf{XS}_1)'\mathbf{Y}_1 . && \Leftarrow\!\!= \text{by (9')} \end{aligned}$$

Consequently, the generalised inverse estimate of Y_1 is shown to be the least squares estimate of Y_1 regressed on the non-zero principal components $\mathbf{XS}_1$ of $\mathbf{X}$.

We turn now to the Ridge estimation, another contrivance to cope with the singularity or near singularity of $X'X$. The Ridge estimate $\tilde{\mathbf{Y}}_1$ of $\mathbf{Y}_1$ is given by

$$\tilde{\mathbf{Y}}_1 = \mathbf{X}(k\mathbf{I} + \mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_1 , \tag{11}$$

where k is an arbitrarily chosen positive number. Utilising (9), (9") and the fact that $\mathbf{I} = \mathbf{SS}'$, (11) further reduces to

$$\tilde{\mathbf{Y}}_1 = (\mathbf{XS}_1)(k\mathbf{I}_r + \Lambda_r)^{-1}(\mathbf{S}_1'\mathbf{X}')\mathbf{Y}_1 ,$$

where $\mathbf{I}_r$ of course signifies an identity matrix of order r. Therefore it is rather immediate that

$$\lim_{k \to 0} \tilde{\mathbf{Y}}_1 = (\mathbf{XS}_1)(\Lambda_r)^{-1}(\mathbf{S}_1'\mathbf{X}')\mathbf{Y}_1 = \mathbf{X}(\mathbf{X}'\mathbf{X})^{+}\mathbf{X}'\mathbf{Y}_1 = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}_1 .$$

Thus $\tilde{\mathbf{Y}}_1$ ultimately coincides with the generalised inverse estimate of $\mathbf{Y}_1$ and as a direct implication of the above argument on OLS we easily claim that the Ridge estimator $(k\mathbf{I} + \mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}_1$ converges to $(\mathbf{X}'\mathbf{X})^{+}\mathbf{X}'\mathbf{Y}_1$, the least squares estimator of unknown parameters in terms of $(\mathbf{X}'\mathbf{X})^{+}$ as k tends to 0.[2)]

3. Concluding Remark

The arguments so far established a generalised estimation procedure of MMVE and/or LIML by extending a classical theorem concerning the extremisation of a ratio of two quadratic forms(PD-theorem), and showed that recognising the normal equations of least squares are always solvable, the application of g-inverse to each stage of TSLS is rather desirablem since the artificial contrivance of Kloek and Mennes(1960) results in biased estimators. However, it remains to be studied further to clarify the statistical properties of MMVE and to elaborate computational schemes so as to promote the accuracy of numerical calculations needed in the estimation processes.

Appendix 1. On the Extrema of the Ratio of Quadratic Forms

In this appendix, we are concerned with the extrema of a real valued function denoted by $F(\mathbf{x}) = (\mathbf{x}'\mathbf{A}\mathbf{x})/(\mathbf{x}'\mathbf{B}\mathbf{x})$, where $\mathbf{A}$ and $\mathbf{B}$ are real symmetric matrices of order n and $\mathbf{x}$ is an n-dimensional real vector. Moreover, unless otherwise specified, $\mathbf{B}$ is supposed to be nonnegative definite. Obviously, $F(\mathbf{x})$ can be defined only for $\mathbf{x}$ for which $\mathbf{x}'\mathbf{B}\mathbf{x} \neq 0$ and the assumed nonnegative definiteness of $\mathbf{B}$ implies that the denominator vanishes if and only if $\mathbf{x} \in N(\mathbf{B}) \equiv \{x \in R^n \mid \mathbf{B}\mathbf{x} = 0\}$, where R^n is of course the n-dimensional real vector space. Thus it is taken for granted that F is a function from $\{R^n \setminus N(\mathbf{B})\} \equiv \{x \in R^n \mid \mathbf{x} \notin N(\mathbf{B})\}$ into R, the set of all real numbers. If $\mathbf{B}$ is positive definite the minimum(maximum) of $F(\mathbf{x})$ is known to be the smallest(largest) root of the determinantal equation $|\mathbf{A} - \lambda\mathbf{B}| = 0$ (Henceforth, PD-Theorem for short). However, if $\mathbf{B}$ is singular PD-Theorem becomes unapplicable. To see this, let $\mathbf{L}$ be the intersection of $N(\mathbf{B})$ and the n-dimensional unit circle centered at the origin $\mathbf{0}$. If $\mathbf{B}$ is singular $\mathbf{L}$ can not be compact, for L is then by no means closed. On the other hand, the nonsingularity of $\mathbf{B}$ implies that $\mathbf{L}$ coincides with the unit circle which is compact. Hence, the famous Weierstrass Theorem does guarantee that $F(\mathbf{x})$ attains its extrema on its substantial domain $\mathbf{L}$. On the contrary, the singularity of $\mathbf{B}$ makes Weierstrass Theorem unapplicable by preventing $\mathbf{L}$ from being compact.

For our main task to generalise PD-Theorem, the following two numerical examples prove useful because they illustrate that PD-Theorem can no longer remain valid under the nonsingularity of $\mathbf{B}$ and suggest how to generalise PD-Theorem.

<u>Example 1</u> $\mathbf{A} = \begin{pmatrix} a_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -a_3 \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} b_1 & 0 & 0 \\ 0 & b_2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ with $a_1>0$, $a_3>0$ and $b_i>0 (i = 1, 2)$.

Evidently, $N(\mathbf{B}) = \left\{\lambda \begin{pmatrix} 0 \\ 0 \\ 1 \end{pmatrix} : \lambda \in R \right\}$. Hence, $\mathbf{x} \notin N(\mathbf{B})$ if and only if x_1 or x_2 is nonzero with x_3 arbitrary. Therefore

$$F(\mathbf{x}) = ((a_1x_1^2 - a_3x_3^2)/(b_1x_1^2 + b_2x_2^2)) \quad \text{for any } \mathbf{x} \in \{R^3 \setminus N(\mathbf{B})\} .$$

Choose an $\tilde{\mathbf{x}} \in \{R^3 \setminus N(\mathbf{B})\}$. Then $\lim_{x_3 \to \infty} F(\mathbf{x}) = -\infty$ and $\lim_{x_3 \to -\infty} F(\mathbf{x}) = \infty$, provided that $\tilde{x}_1$ and $\tilde{x}_2$ are kept intact. Thus $F(x)$ attains neither minimum nor maximum on $\{R^3 \setminus N(\mathbf{B})\}$ and any root of $|\mathbf{A} - \lambda\mathbf{B}| = 0$ has no relevance to the extrema of F(x).

<u>Example 2</u> $\mathbf{A} = \begin{pmatrix} a_1 & 0 & 0 \\ 0 & a_2 & 0 \\ 0 & 0 & 0 \end{pmatrix}$, $\mathbf{B} = \begin{pmatrix} b_1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}$ with $a_i>0 (i = 1, 2)$ and $b_1>0$. A direct calculation shows that $N(\mathbf{B}) = \left\{ x_2\begin{pmatrix}0\\1\\0\end{pmatrix} + x_3\begin{pmatrix}0\\0\\1\end{pmatrix} : \begin{pmatrix}x_2\\x_3\end{pmatrix} \in R^2 \right\}$. Accordingly, $\{R^3 \setminus N(\mathbf{B})\} = \{x \in R^3$ with the property that $x_1 \neq 0\}$ and $F(\mathbf{x}) = (a_1/b_1) + (a_2/b_1)(x_2/x_1)^2$. Evidently, F(x) attains its minimum value = (a_1/b_1) at $\mathbf{x} = (x_1, 0, x_3)'$ where $x_1 \neq 0$ and x_3 is arbitrary.

From these examples we can infer that if **B** is singular(the matrices **B** of the above examples are nonnegative definite but not positive definite) then (i) PD-Theorem collapses, (ii) F(x) does not always attain the extreme and (iii) the nonnegative definiteness of **A** may serve to ensure the existence of $\min_{\mathbf{x} \notin N(\mathbf{B})} F(\mathbf{x})$.

Bearing in mind what are implied by the above numerical examples, we can now turn to the generalisation of PD-Theorem. To this end, let's recall the well known fact that to N(**B**) there corresponds a linear subspace **W** with the following properties;

$$\left.\begin{aligned} &\dim \mathbf{W} = \text{the rank of } \mathbf{B}, \quad &\text{(AP.1.1)} \\ &R^n = \mathbf{W} \oplus N(\mathbf{B}) \quad &\text{(AP.1.2)} \end{aligned}\right\} \text{(AP.1)}$$

where dim **W** denotes the dimension of **W** , $\mathbf{W} \oplus N(\mathbf{B})$ is a linear subspace $\{\mathbf{x} + \mathbf{y} \mid \mathbf{x} \in \mathbf{W}, \mathbf{y} \in N(\mathbf{B})\}$ obeying the side condition $\mathbf{W} \cap N(\mathbf{B}) = \{\theta\}$, and $\{\theta\}$ signifies a set consisting solely of the origin θ of R^n. Then, we can state and prove;

<u>Lemma AP.1</u> Let **W** and r be as in (AP.1) and by **P**(**Q**) denote the matrix representing any given basis of **W** (N(**B**)). Then,

$$\mathbf{V} \equiv \{R^n \setminus N(\mathbf{B})\} = \{\mathbf{Pu} + \mathbf{Qv} \mid \mathbf{u} \in R_0^r \equiv \{R^r \setminus \{0\}\},\ \mathbf{v} \in R^{n-r}\} \equiv \mathbf{U},$$

where by the phrase "a matrix representing a basis of a linear subspace", we mean that the matrix obtained by arranging all vectors of the basis as its columns.

<u>Proof</u> Let $\mathbf{x} \in \mathbf{V}$. Then, (AP.1.2) asserts that there exist $\mathbf{u}_x \in R_0^r$ and $\mathbf{v}_x \in R^{n-r}$ for which $\mathbf{x} = \mathbf{Pu}_x + \mathbf{Qv}_x$, because by definition $\mathbf{V} \subseteq R^n$. Obviously, $\mathbf{u}_x \neq \mathbf{0}$, for otherwise $\mathbf{x} = \mathbf{Qv}_x \in N(\mathbf{B})$, violating the assumption that $\mathbf{x} \notin N(\mathbf{B})$.

Conversely, if $(\mathbf{Pu}+\mathbf{Qv})$ is in $\mathbf{U}$ then $\mathbf{B}(\mathbf{Pu}+\mathbf{Qv}) = \mathbf{B}(\mathbf{Pu})$, for by definition $\mathbf{BQ}$ vanishes. Since the definition of $\mathbf{P}$ implies that $\mathbf{Pu}$ is a nonzero vector in $\mathbf{W}$ and since $\mathbf{W} \cap N(\mathbf{B}) = \{0\}$, $\mathbf{Pu}$ is not in $N(\mathbf{B})$, or equivalently $\mathbf{B}(\mathbf{Pu}) \neq \mathbf{0}$. In summary, $(\mathbf{Pu}+\mathbf{Qv}) \notin N(\mathbf{B})$ for any $(\mathbf{u}, \mathbf{v}) \in (R_0^r \times R^{n-r})$. (Q.E.D.)

The lemma just represented enables us to reckon $F(\mathbf{x})$ as a real valued function defined on the Cartesian product $(R_0^r \times R^{n-r})$ which we hereafter abbreviate by $\mathbf{Z}$. Let r, $\mathbf{P}$ and $\mathbf{Q}$ be as in Lemma AP.1 throughout the remaining part of this appendix. Then we can establish the following mathematical results.

<u>Lemma AP.2</u> Let f and G be the functions from $\mathbf{Z}$ into R specified respectively by

$$f(\mathbf{u}, \mathbf{v}) = (\mathbf{Qv})'\mathbf{A}(\mathbf{Qv}) + 2(\mathbf{Qv})'\mathbf{APu} + (\mathbf{Pu})'\mathbf{A}(\mathbf{Pu}) \qquad \text{(AP.2)}$$

and

$$G(\mathbf{u}, \mathbf{v}) = f(\mathbf{u}, \mathbf{v}) \,/\, (\mathbf{u}'(\mathbf{P}'\mathbf{BP})\mathbf{u}) \,. \qquad \text{(AP.2')}$$

Then,

$$F(\mathbf{V}) = G(\mathbf{Z}) \,.$$

<u>Proof</u> Rearranging the terms of the right hand side of (AP.2), we see that

$$f(\mathbf{u}, \mathbf{v}) = (\mathbf{Pu}+\mathbf{Qv})'\mathbf{A}(\mathbf{Pu}+\mathbf{Qv}) \quad \text{for all } (\mathbf{u}, \mathbf{v}) \in \mathbf{Z} \,. \qquad \text{(AP.3)}$$

By the definition of $\mathbf{Q}$, $\mathbf{BQ}$ vanishes. This, coupled with the assumed symmetry of $\mathbf{B}$, implies that $\mathbf{Q}'\mathbf{B}$ vanishes too. Therefore,

$$\mathbf{u}'(\mathbf{P}'\mathbf{BP})\mathbf{u} = (\mathbf{Pu}+\mathbf{Qv})'\mathbf{B}(\mathbf{Pu}+\mathbf{Qv}) \quad \text{for all } (\mathbf{u}, \mathbf{v}) \in \mathbf{Z} \,. \qquad \text{(AP.4)}$$

From (AP.3) and (AP.4), it follows that

$$\begin{aligned} G(\mathbf{u}, \mathbf{v}) &= (\mathbf{Pu}+\mathbf{Qv})'\mathbf{A}(\mathbf{Pu}+\mathbf{Qv})/(\mathbf{Pu}+\mathbf{Qv})'\mathbf{B}(\mathbf{Pu}+\mathbf{Qv}) \\ &= F(\mathbf{Pu}+\mathbf{Qv}) \qquad \text{for all } (\mathbf{u}, \mathbf{v}) \in \mathbf{Z} \,. \end{aligned} \qquad \text{(AP.5)}$$

Let ℓ be in $G(\mathbf{Z})$. Then, by definition, to ℓ there corresponds $(\mathbf{u}, \mathbf{v}) \in \mathbf{Z}$ such that $\ell = G(\mathbf{u}, \mathbf{v})$. Applying (AP.5) and noting Lemma AP.1 assures that $(\mathbf{Pu}+\mathbf{Qv}) \in \mathbf{V}$, it is clear that $G(\mathbf{Z}) \subseteq F(\mathbf{V})$.

Conversely, if ℓ is in $F(\mathbf{V})$, $\ell = F(\mathbf{x})$ for some $\mathbf{x} \in \mathbf{V}$ and by Lemma AP.1 again there exists $(\mathbf{u}, \mathbf{v}) \in \mathbf{Z}$ such that $\mathbf{x} = \mathbf{Pu}+\mathbf{Qv}$. Hence, by (AP.5), $\ell = F(\mathbf{x}) = F(\mathbf{Pu}+\mathbf{Qv}) = G(\mathbf{u}, \mathbf{v})$. Thus $F(\mathbf{V})$ is surely involved in $G(\mathbf{Z})$. (Q.E.D.)

<u>Lemma AP.3</u> Suppose that the matrix $\mathbf{A}$ is nonnegative definite and let $\mathbf{u}$ be arbitrarily specified vector of R_0^r. Then the equation

$$(\mathbf{Q}'\mathbf{AQ})\mathbf{v} = -(\mathbf{Q}'\mathbf{APu}) \qquad \text{(AP.6)}$$

is consistent, and

$$\min_{v \in R^{n-r}} f(u, v) = f(u, v_u) = u'(P'(A-(AQ)(Q'AQ)^-(Q'A))P)u \ , \qquad \text{(AP.7)}$$

where v_u is any solution of (AP.6).

Proof Since **A** is supposed to be nonnegative definite, there exists an $n \times n$ matrix α such that $\alpha'\alpha = A$. Hence $(Q'AQ) = (\alpha Q)'(\alpha Q)$ and $(Q'A) = (\alpha Q)'\alpha$. Consequently, $R(Q'AQ) = R((\alpha Q)'(\alpha Q)) = R((\alpha Q)')$ in which $-(Q'APu) = (\alpha Q')(-\alpha Pu)$ is contained. Therefore, (AP.6) is consistent.

Noticing that (AP.7) is the necessary and sufficient condition for minimising f(u, v) with respect to v, it is immediate that

$$\begin{aligned}\min_{v \in R^{n-r}} f(u, v) &= f(u, v_u) \\ &= u'(P'AP)u + (v_u)'(Q'APu) \\ &= (Pu)'A(Pu) - (v_u)'(Q'AQ)v_u \ , \qquad \text{(AP.8)}\end{aligned}$$

provided that (AP.6) holds for v_u.

Since v_u is known to be expressed as the sum of $(Q'AQ)^-(-Q'APu)$ and the linear combinations of vectors in $N(Q'AQ)$ and since the assumed nonnegative definiteness of **A** again allows that if $(Q'AQ)y = 0$ then $y'QA = 0$, it is easy to see that

$$\begin{aligned}f(u, v_u) &= f(u, (Q'AQ)^-(-Q'APu)) \\ &= u'(P'(A-(AQ)(Q'AQ)^-(Q'A))P)u \ . \qquad \text{(Q.E.D.)}\end{aligned}$$

With the aid of Lemmas AP.1 through AP.3, we can now state and prove ;

Theorem AP.1 Assume that **A** and **B** are nonnegative definite. Then, the following assertions hold true.

(i) If N(**A**) is not contained in N(**B**), F(**x**) attains the minimum which is 0 at any $x \in \{N(A) \setminus N(B)\}$, and

(ii) in the other case ($N(A) \subseteq N(B)$),

(ii-1) F(**x**) attains the minimum equals the smallest root $\underline{\ell}$ of the determinantal equation

$$| P'(A-(AQ)(Q'AQ)^-(Q'A))P - \ell(P'BP) | = 0 \qquad \text{(AP.9)}$$

at $\underline{x} \in V$ expressed as $P\underline{u} + Qv_{\underline{u}}$ by $(\underline{u}, v_{\underline{u}}) \in Z$ satisfying

$$\begin{cases} (P'(A-(AQ)(Q'AQ)^-(Q'A))P)\underline{u} = \underline{\ell}(P'BP)\underline{u} & \text{(AP.10)} \\ (Q'AQ)v_{\underline{u}} = -Q'AP\underline{u} & \text{(AP.10')} \end{cases}$$

and finally

(ii-2) $P\underline{u} + Qv_{\underline{u}}$ is independent of the choice of bases of **W** and N(**B**).

Proof (i) Since **A** and **B** are nonnegative definite, $F(x) \geqq 0$ for all $x \in V$.

By assumption, $\{N(\mathbf{A})\setminus N(\mathbf{B})\}$ is nonempty and for any $\mathbf{x}\in\{N(\mathbf{A})\setminus N(\mathbf{B})\}$, $F(\mathbf{x}) = 0$.

(ii-1) Let $\mathbf{u}$ be any vector in R_0^r. Then, Lemma AP.3 asserts that there exists $\mathbf{v}_\mathbf{u} \in R^{n-r}$ with the property that

$$G(\mathbf{u}, \mathbf{v}_\mathbf{u}) \leqq G(\mathbf{u}, \mathbf{v}) \quad \text{for all } \mathbf{v} \in R^{n-r}$$

and Lemma AP.1 implies that $\mathbf{Pu} = (\mathbf{Pu} + \mathbf{Q0}) \in \mathbf{V}$ for all $\mathbf{u} \in R_0^r$. Hence

$$\mathbf{u}'(\mathbf{P}'\mathbf{B}\mathbf{P})\mathbf{u} > 0 \quad \text{for all } \mathbf{u} \in R_0^r .$$

Thus PD-Theorem applied to $G(\mathbf{u}, \mathbf{v}_\mathbf{u})$, in conjunction with (AP.7), guarantees that $\min_{\mathbf{u}\in R_0^r} G(\mathbf{u}, \mathbf{v}_\mathbf{u}) = \underline{\ell}$, the minimum root of (AP.9), is attained at $\underline{\mathbf{u}}$ satisfying (AP.10).

Recalling (AP.8), $\underline{\ell}$ is explicitly expressed as the function of $(\underline{\mathbf{u}}, \mathbf{v}_{\underline{\mathbf{u}}}) \in \mathbf{Z}$ so far as $(\underline{\mathbf{u}}, \mathbf{v}_{\underline{\mathbf{u}}})$ is determined by (AP.10) and (AP.10'). Then, by definition, we have

$$G(\mathbf{u}, \mathbf{v}) \geqq G(\mathbf{u}, \mathbf{v}_\mathbf{u}) \geqq G(\underline{\mathbf{u}}, \mathbf{v}_{\underline{\mathbf{u}}}) \text{ for all} (\mathbf{u}, \mathbf{v}) \in \mathbf{Z} .$$

Thus

$$\underline{\ell} = G(\underline{\mathbf{u}}, \mathbf{v}_{\underline{\mathbf{u}}}) = \min_{(\mathbf{u},\mathbf{v})\in \mathbf{Z}} G(\mathbf{u}, \mathbf{v}) . \tag{AP.11}$$

Recalling that $(\mathbf{P}\underline{\mathbf{u}} + \mathbf{Q}\mathbf{v}_{\underline{\mathbf{u}}}) \equiv \underline{\mathbf{x}}$ is in $\mathbf{V}$ by virtue of Lemma AP.1, Lemma AP.2 and (AP.11) surely guarantee that

$$\min_{\mathbf{x}\in \mathbf{V}} F(\mathbf{x}) = F(\underline{\mathbf{x}}) = \underline{\ell} .$$

(ii-2) Let $\mathbf{\Pi}(\mathbf{\Omega})$ be the matrix representing the basis of $R(\mathbf{B})(N(\mathbf{B}))$ different from the basis represented by $\mathbf{P}(\mathbf{Q})$. Then, there exist nonsingular matrices $\mathbf{S}$ and $\mathbf{T}$ such that $\mathbf{P} = \mathbf{\Pi S}$ and $\mathbf{Q} = \mathbf{\Omega T}$ and it is easily confirmed that

$$(\mathbf{Q}'\mathbf{A}\mathbf{Q})^- = \mathbf{T}^{-1}(\mathbf{\Omega}'\mathbf{A}\mathbf{\Omega})^-(\mathbf{T}')^{-1} \qquad \text{for any } (\mathbf{\Omega}'\mathbf{A}\mathbf{\Omega})^- .$$

Hence a direct computation yields

$$\begin{aligned} &\mathbf{P}'(\mathbf{A} - (\mathbf{AQ})(\mathbf{Q}'\mathbf{AQ})^-(\mathbf{Q}'\mathbf{A}))\mathbf{P} - \ell(\mathbf{P}'\mathbf{B}\mathbf{P}) \\ &= \mathbf{S}'(\mathbf{\Pi}'(\mathbf{A} - (\mathbf{A\Omega})(\mathbf{\Omega}'\mathbf{A\Omega})^-(\mathbf{\Omega}'\mathbf{A}))\mathbf{\Pi} - \ell(\mathbf{\Pi}'\mathbf{B}\mathbf{\Pi}))\mathbf{S} . \end{aligned} \tag{AP.12}$$

Obviously the nonsingularity of $\mathbf{S}$, together with (AP.12), further implies that $\underline{\ell}$ is the smallest root of

$$|\ \mathbf{\Pi}'(\mathbf{A} - (\mathbf{A\Omega})(\mathbf{\Omega}'\mathbf{A\Omega})^-(\mathbf{\Omega}'\mathbf{A}))\mathbf{\Pi} - \ell(\mathbf{\Pi}'\mathbf{B}\mathbf{\Pi})\ | = 0 ,$$

and that (AP.10) is true if and only if

$$(\mathbf{\Pi}'(\mathbf{A} - (\mathbf{A\Omega})(\mathbf{\Omega}'\mathbf{A\Omega})^-(\mathbf{\Omega}'\mathbf{A}))\mathbf{\Pi})(\mathbf{S}\underline{\mathbf{u}}) = \underline{\ell}(\mathbf{\Pi}'\mathbf{B}\mathbf{\Pi})(\mathbf{S}\underline{\mathbf{u}}) .$$

Moreover, (AP.10') is equivalent to

$$(\mathbf{\Omega}'\mathbf{A\Omega})(\mathbf{T}\mathbf{v}_{\underline{\mathbf{u}}}) = -(\mathbf{\Omega}'\mathbf{A\Pi})(\mathbf{S}\underline{\mathbf{u}}) .$$

Combining the above three observations with the fact that the bases represented by $\mathbf{P}$ and $\mathbf{Q}$ are chosen arbitrarily, $F(\mathbf{x})$ surely attains the minimum(=

$\underline{\ell}$) at

$$\Pi(S\underline{u}) + \Omega(Tv_{\underline{u}}) = P\underline{u} + Qv_{\underline{u}}. \qquad \text{(Q.E.D.)}$$

Theorem AP.1 confines itself to the case where both **A** and **B** are nonnegative definite. However, in view of the facts that a matrix **H** is nonpositive definite if and only if −**H** is nonnegative definite and that N(**H**) = N(-**H**), it is clear that Theorem AP.1 applied to all possible patterns of definiteness of **A** and **B** yields the following results which we state as a corollary.

Corollary AP.1. (i) If A and B are nonpositive definite, Theorem AP.1 remains intact, and

(ii) in the remaining case(A and B obey different patterns of definiteness), it suffices to replace the " minimum " and " smallest-root " in the theorem by " maximum " and " greatest-root " respectively.

Remarks concerning Theorem AP.1 (i) The condition that $N(\mathbf{A}) \subseteq N(\mathbf{B})$ assures that $F(\mathbf{x}) > 0$ for all $\mathbf{x} \in \mathbf{V}$. Hence $\underline{\ell}$ is positive.

(ii) As is well known, the linear subspaces possessing the property (AP.1) are not unique. However, in view of the symmetry of **B**, the most natural candidate of **W** is $R(\mathbf{B}) = \{\mathbf{Bx} \mid \mathbf{x} \in R^n\}$, the image space of **B**.

Appendix 2. On the Generalised Inverse of a Matrix

Here, we state and prove some fundamental properties of the generalised inverse of a given matrix(abbreviated as g-inverse, for short) and refer to the implication of g-inverse to the least squares method, so long as the implication is relevant to the limited-information maximum likelihood estimation.

<u>Lemma AP.4</u> Let $\mathbf{A}$ be an $m \times n$ complex matrix. Then the following two statements are mutually equivalent.

(i) There exists an $n \times m$ matrix $\mathbf{A}^-$ such that

$$\mathbf{A}(\mathbf{A}^- \mathbf{y}) = \mathbf{y} \qquad \text{for all } \mathbf{y} \in R(\mathbf{A}), \text{ and}$$

(ii) there exists an $n \times m$ matrix $\mathbf{A}^-$ for which

$$\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}.$$

<u>Proof</u> <u>(i)⇒(ii)</u> By $\mathbf{a}^j$ denote the j-th column of $\mathbf{A}$. Since $\mathbf{a}^j (j = 1, \cdots, n)$ are in $R(\mathbf{A})$, we have

$$\mathbf{A}(\mathbf{A}^- \mathbf{a}^j) = \mathbf{a}^j \qquad j = 1, \cdots, n,$$

or equivalently,

$$\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}.$$

<u>(ii)⇒(i)</u> With any $\mathbf{y} \in R(\mathbf{A})$, there is associated an $\mathbf{x}$ satisfying that $\mathbf{A}\mathbf{x} = \mathbf{y}$. Hence,

$$\mathbf{A}\mathbf{A}^-\mathbf{y} = \mathbf{A}\mathbf{A}^-\mathbf{A}\mathbf{x} = \mathbf{A}\mathbf{x} = \mathbf{y}. \qquad \text{(Q.E.D.)}$$

The lemma just shown implies that it makes no difference whether we choose (i) or (ii) to define the g-inverse. Since we prefer (ii) to (i), we arrive at the following definition.

<u>Definition AP.1</u> For any $m \times n$ matrix $\mathbf{A}$, an $n \times m$ matrix $\mathbf{A}^-$ such that $\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A}$ is said to be the g-inverse of $\mathbf{A}$.

<u>Remark concerning Definition AP.1</u> The usual elimination method guarantees that for any $\mathbf{A}$, there exists $\mathbf{A}^-$ such that

$$\mathbf{A}\mathbf{A}^-\mathbf{A} = \mathbf{A} \quad \text{and} \quad \mathbf{A}^-\mathbf{A}\mathbf{A}^- = \mathbf{A}^-.$$

For the detail, see Rao(1973, pp.26-27).

<u>Lemma AP.5</u> Let $L_i (i = 1, 2)$ be two linear subspaces with the properties that $L_2 \subseteq L_1$ and that $\dim L_1 = \dim L_2 = d$. Then $L_1 = L_2$, where by $\dim L_i$ we mean the number of vectors forming a basis of L_i.

<u>Proof</u> Suppose the contrary. Then there exists an $\mathbf{x}$ of L_1 which does not

belong to L_2. Let $\{y^1, y^2, \cdots, y^d\}$ be a basis of L_2. Since $\mathbf{x}$ is not in L_2, the vectors $\mathbf{x}$, y^1, y^2, $\cdots$, y^d are linearly independent, for otherwise there exists a nonzero vector $(c_0, c_1, c_2, \cdots, c_d)$ such that

$$c_0\mathbf{x} + c_1 y^1 + \cdots\cdots + c_d y^d = \mathbf{0}$$

If $c_0 = 0$, $c_i = 0 (i = 1, \cdots, d)$, for $y^1, y^2, \cdots, y^d$ are linearly independent by definition. Thus, $c_0 \neq 0$, whence $\mathbf{x}$ is expressed as a linear combination of $y^j (j = 1, \cdots, d)$. Consequently, $\mathbf{x}$ should be in L_2, which violates the hypothesis that $\mathbf{x}$ is not in L_2. Hence, $\mathbf{x}$, y^1, y^2, $\cdots$, y^d are linearly independent and in L_1 because $y^j \in L_2 \subseteq L_1 (j = 1, \cdots, d)$. Therefore, $d = \dim L_1 \geqq d+1$, a self-contradiction. (Q.E.D.)

Theorem AP.2 Let $\mathbf{A}$ be an $m \times n$ complex matrix. Then $\mathbf{A}^-$ possesses the properties stated below(assertions (i) through (iii) are due to Rao and Mitra(1971; Lemma 2.2.6)).

(i) $(\mathbf{A}^-\mathbf{A}) = (\mathbf{A}^-\mathbf{A})^2$, that is, $\mathbf{A}^-\mathbf{A}$ is idempotent,

(ii) if $\mathbf{V}$ is a matrix such that $r(\mathbf{A}^*\mathbf{V}\mathbf{A}) = r(\mathbf{A})$ then

(ii-1) $\mathbf{A}(\mathbf{A}^*\mathbf{V}\mathbf{A})^-(\mathbf{A}^*\mathbf{V}\mathbf{A}) = \mathbf{A}$ and $(\mathbf{A}^*\mathbf{V}\mathbf{A})(\mathbf{A}^*\mathbf{V}\mathbf{A})^-\mathbf{A}^* = \mathbf{A}^*$,

(ii-2) $\mathbf{A}(\mathbf{A}^*\mathbf{V}\mathbf{A})^-\mathbf{A}^*$ has the properties;

(ii-2.1) $\mathbf{A}(\mathbf{A}^*\mathbf{V}\mathbf{A})^-\mathbf{A}^*$ is invariant for any choice of $(\mathbf{A}^*\mathbf{V}\mathbf{A})^-$,

(ii-2.2) $r(\mathbf{A}(\mathbf{A}^*\mathbf{V}\mathbf{A})^-\mathbf{A}^*) = r(\mathbf{A})$, and

(ii-2.3) if $\mathbf{A}^*\mathbf{V}\mathbf{A}$ is hermitian, so is $\mathbf{A}(\mathbf{A}^*\mathbf{V}\mathbf{A})^-\mathbf{A}^*$,

and finally

(iii) for any $\mathbf{y} \in R(\mathbf{A})$, the general solution of $\mathbf{A}\mathbf{x} = \mathbf{y}$ is given by

$$\mathbf{A}^-\mathbf{y} + (\mathbf{I} - \mathbf{A}^-\mathbf{A})\mathbf{v},$$

where $\mathbf{v}$ is an arbitrary vector.

<u>Proof</u> <u>(i)</u> $(\mathbf{A}^-\mathbf{A})(\mathbf{A}^-\mathbf{A}) = \mathbf{A}^-(\mathbf{A}\mathbf{A}^-\mathbf{A}) = \mathbf{A}^-\mathbf{A}$.

<u>(ii-1)</u> The definition of g-inverse yields

$$(\mathbf{A}^*\mathbf{V}\mathbf{A})(\mathbf{A}^*\mathbf{V}\mathbf{A})^-(\mathbf{A}^*\mathbf{V}\mathbf{A}) = \mathbf{A}^*\mathbf{V}\mathbf{A}. \quad \text{(AP.13)}$$

Rearranging the terms of (AP.13), we have

$$(\mathbf{A}^*\mathbf{V}\mathbf{A})((\mathbf{A}^*\mathbf{V}\mathbf{A})^-(\mathbf{A}^*\mathbf{V}\mathbf{A}) - \mathbf{I}) = [0].$$

In other words, every column of $((\mathbf{A}^*\mathbf{V}\mathbf{A})^-(\mathbf{A}^*\mathbf{V}\mathbf{A}) - \mathbf{I})$ is in $N(\mathbf{A}^*\mathbf{V}\mathbf{A})$ which equals $N(\mathbf{A})$ by Lemma AP.5. This verifies the first half of assertion (ii-1).

Let $\mathbf{y}$ be a representative row of $((\mathbf{A}^*\mathbf{V}\mathbf{A})(\mathbf{A}^*\mathbf{V}\mathbf{A})^- - \mathbf{I})$. Then, (AP.13) implies that $\mathbf{y}^* \in N((\mathbf{A}^*\mathbf{V}\mathbf{A})^*) = N(\mathbf{A}^*\mathbf{V}^*\mathbf{A})$. Since $\dim N(\mathbf{A}^*\mathbf{V}^*\mathbf{A}) = n - r((\mathbf{A}^*\mathbf{V}\mathbf{A})^*) = n - r(\mathbf{A}^*\mathbf{V}^*\mathbf{A}) = n - r(\mathbf{A})$ and since $N(\mathbf{A}^*\mathbf{V}^*\mathbf{A}) \supseteq N(\mathbf{A})$, Lemma AP.5 ensures that $N(\mathbf{A}^*\mathbf{V}^*\mathbf{A}) = N(\mathbf{A})$. Hence $\mathbf{y}\mathbf{A}^* = \mathbf{0}$.

<u>(ii-2.1)</u> By $\mathbf{D}_1$ and $\mathbf{D}_2$ denote any two g-inverses of $(\mathbf{A}^*\mathbf{V}\mathbf{A})$. Applying the

second half of (ii-1), we obtain

$$((A^*VA)D_i - I)A^* = [0] \qquad i = 1, 2.$$

This further reduces to

$$(A^*VA)(D_1 - D_2)A^* = [0]\,. \qquad \text{(AP.14)}$$

Employing Lemma AP.5 as in the proof of the first half of assertion (ii-1), [AP.14] implies that $A(D_1A^* - D_2A^*) = [0]$, from which it immediately follows that

$$AD_1A^* = AD_2A^* \qquad \text{for any g-inverse of } (A^*VA).$$

(ii-2.2) $r(A) \geqq r(A(A^*VA)^-A^*) \geqq r(A(A^*VA)^-A^*VA) = r(A)$

(ii-2.3) Let H be an hermitian matrix. Then, as is well known, there exists a unitary matrix Q such that $Q^*HQ = \begin{pmatrix} \Psi & \vdots & 0 \\ \cdots & & \cdots \\ 0 & \vdots & 0 \end{pmatrix}$, where Ψ is a real diagonal matrix, the diagonal elements of which are real nonzero eigen values of H. Furthermore, a direct calculation yields

$$H\left(Q\begin{pmatrix} \Psi^{-1} & \vdots & 0 \\ \cdots & & \cdots \\ 0 & \vdots & 0 \end{pmatrix}Q^*\right)H = Q\begin{pmatrix} \Psi & \vdots & 0 \\ \cdots & & \cdots \\ 0 & \vdots & 0 \end{pmatrix}Q^* = H\,.$$

And the matrix $Q\begin{pmatrix} \Psi^{-1} & \vdots & 0 \\ \cdots & & \cdots \\ 0 & \vdots & 0 \end{pmatrix}Q^*$ is clearly hermitian. In other words, an hermitian matrix has an hermitian g-inverse. Let D be an hermitian g-inverse of A^*VA. Then $A^*DA = (A^*DA)^*$. Thus, the assertion follows directly from assertion (ii-2.1).

(iii) A direct computation, with Lemma AP.4 and Definition AP.1 in mind, assures that

$$A(A^-y + (I - A^-A)v) = AA^-y = y\,. \qquad \text{(Q.E.D.)}$$

Consider now to estimate β of the model

$$y = X\beta + u\,,$$

where $X(y)$ is an $n \times K$ matrix($n \times 1$ vector) of observations, u the vector of disturbances, and n of course the sample size.

Then, Theorem AP.2(iii) assures that the least squares estimates $\hat{\beta}$ of β is expressed as ;

$$\hat{\beta} = (X'X)^-X'y + (I - (X'X)^-(X'X))v\,, \qquad \text{(AP.15)}$$

where v can be any vector of K-dimensional real vector space. Define the estimate $\hat{y}$ of y by $X\hat{\beta}$ and the residual e by $y - \hat{y}$. Then in view of

Theorem AP.2(ii-1: the first half), we have

$$e = (I - X(X'X)^{-}X')y \ . \qquad \text{(AP.16)}$$

Therefore, Theorem AP.2(ii-2.1) asserts that e is independent of the g-inverse used but dependent sorely on the data observed. And (AP.16) affords a basis for interpreting the denominator(numerator) of (4) in the text as the sum of squared residuals obtained from regressing $z = Y_{\Delta}\beta_{\Delta}$ on $X(X_1)$.

References

A. Albert, *Regression and the Moore-Penrose Pseudoinverse*, Academic Press, New York, 1972.

Y. Fujikoshi, K. Morimune, N. Kunitomo and M. Taniguchi, "Asymptotic Expansions of the Distributions of the Estimates of Coefficients in a Simultaneous Equation System", *Journal of Econometrics*, 1982, *18*, 191-205.

W.A. Fuller, "Some Properties of a Modification of the Limited Information Estimator", *Econometrica*, 1977, *45*, 939-953.

J. Johnston, *Econometric Methods* (Third edition), McGraw-Hill, New York, 1984.

T. Kloekand L.B.M. Mennes, "Simultaneous Equations Estimation Based on Principal Components of Predetermined Variables", *Econometrica*, 1960, 28, 45-61.

K. Morimune, "Approximate Distributions of the *k*-class Estimators When the Degree of Overidentifiability is Large Compared to the Sample Size", *Econometrica*, 1983, *51*, 821-841.

C.R. Rao and S.K. Mitra, *Generalized Inverse of Matrices and Its Applications*, John Wiley & Sons, New York, 1971.

C.R. Rao, *Linear Statistical Inference and Its Applications* (Second edition), John Wiley & Sons, New York, 1973.

Footnotes

*/ Acknowledgement;

This paper is a revised version of one presented at the CES Research Seminar on September 7, 1998.

The author is grateful to Professor Dr. Hans Werner Sinn, the Director of CES, for having provided him with a comfortable research environment and thanks Professor Dr. Hans Schneeweiß who kindly chaired the Seminar and gave valuable comments on the earlier version. He owes a great deal to Professor Dr. Martin J. Beckmann, a Councillor of CES, who recommended him for the Visiting Research Fellow of CES and is indebted to Dr. Alfons J. Weichenrieder of CES for his ingenious organisation of the Research Seminar as well as the friendly encouragement during the author's stay in CES. Thanks also go to the financial aid from CES.

1) One candidate of $(X'X)^-$ would be a g-inverse of $X'X$ to be obtained by the elimination stated in Rao(1973, pp.26 – 27).

2) An alternative proof of this assertion is found in Albert(1972, pp.19 – 20) of which Professor Schneeweiß notified the author.

CESifo *Working Paper Series*

---

125 John Douglas Wilson and Xiwen Fan, Tax Evasion and the Optimal Tax Treatment of Foreign-Source Income, January 1997

126 Terje Lensberg, Investment Behaviour under Knightian Uncertainty - An Evolutionary Approach, January 1997

127 David F. Bradford, On the Uses of Benefit-Cost Reasoning in Choosing Policy Toward Global Climate Change, January 1997

128 David F. Bradford and Kyle D. Logue, The Influence of Income Tax Rules on Insurance Reserves, January 1997

129 Hans-Werner Sinn and Alfons J. Weichenrieder, Foreign Direct Investment, Political Resentment and the Privatization Process in Eastern Europe, February 1997

130 Jay Pil Choi and Marcel Thum, Market Structure and the Timing of Technology Adoption with Network Externalities, February 1997

131 Helge Berger and Jakob de Haan, A State within the State? An Event Study on the Bundesbank, February 1997

132 Hans-Werner Sinn, Deutschland im Steuerwettbewerb (Germany Faces Tax Competition), March 1997

133 Francesca Cornelli and David D. Li, Large Shareholders, Private Benefits of Control, and Optimal Schemes of Privatization, May 1997

134 Hans-Werner Sinn and Holger Feist, Eurowinners and Eurolosers: The Distribution of Seigniorage Wealth in EMU, May 1997

135 Peter J. Stemp, What Happens when Inflation Targets Change?, May 1997

136 Torsten Persson, Gerard Roland and Guido Tabellini, Separation of Powers and Political Accountability, June 1997

137 Avinash Dixit and John Londregan, Ideology, Tactics, and Efficiency in Redistributive Politics, June 1997

138 Hans Haller, Inefficient Household Decisions and Efficient Markets, June 1997

139 Avinash Dixit and Mancur Olson, Does Voluntary Participation Undermine the Coase Theorem?, September 1997

140 Frank R. Lichtenberg, The Allocation of Publicly-Funded Biomedical Research, September 1997

141 Hans-Werner Sinn, The Value of Children and Immigrants in a Pay-as-you-go Pension System: A Proposal for a Partial Transition to a Funded System, September 1997

142 Agnar Sandmo, Redistribution and the Marginal Cost of Public Funds, September 1997

143 Erkki Koskela and Ronnie Schöb, Payroll Taxes vs. Wage Taxes: Non-Equivalence Results, September 1997

144 Luis H. R. Alvarez and Vesa Kanniainen, Valuation of Irreversible Entry Options under Uncertainty and Taxation, October 1997

145 Patricia F. Apps and Ray Rees, The Optimal Taxation of Couples, October 1997

146 Patricia F. Apps and Ray Rees, On the Taxation of Trade Within and Between Households, October 1997

147 Patricia F. Apps and Ray Rees, Individual vs. Joint Taxation in Models with Household Production, October 1997

148 David F. Bradford, Transition to and Tax Rate Flexibility in a Cash-Flow Type Tax, November 1997

149 Bas van Aarle and Florence Huart, Monetary and Fiscal Unification in the EU: A Stylized Analysis, November 1997

150 Timur Kuran, Ethnic Norms and their Transformation through Reputational Cascades, December 1997

151 Erkki Koskela, Ronnie Schöb and Hans-Werner Sinn, Pollution, Factor Taxation and Unemployment, January 1998

152 Guttorm Schjelderup and Alfons J. Weichenrieder, Trade, Multinationals, and Transfer Pricing Regulations, January 1998

153 Hans-Werner Sinn, European Integration and the Future of the Welfare State, January 1998

154 Hans-Werner Sinn, The Pay-as-you-go Pension System as a Fertility Insurance and Enforcement Device, January 1998

155 M. Ayhan Kose and Raymond Riezman, External Shocks and Economic Dynamics: The Case of African Countries, February 1998

156 Geir B. Asheim, Green National Accounting: Why and How?, February 1998

157 Robert M. Kunst, Unit Roots, Change, and Decision Bounds, March 1998

158 Bengt-Arne Wickström, Can Bilingualism be Dynamically Stable? A Simple Model of Language Choice, March 1998

159 Manfred J. Holler and Bengt-Arne Wickström, The Scandal Matrix: The Use of Scandals in the Progress of Society, March 1998

160 Georg Nöldeke and Klaus M. Schmidt, Sequential Investments and Options to Own, March 1998

161 Robert Cooter, Expressive Law and Economics, April 1998

162 Peter J. Lambert, Horizontal Inequity: Some New Perspectives, May 1998

163 Syed M. Ahsan and Peter Tsigaris, The Design of a Consumption Tax under Capital Risk, May 1998

164 Jaleel Ahmad and Jing Yang, Trade Liberalization in Eastern European Countries and the Prospects of their Integration into the World Trading System, May 1998

165 Michael Hoy and Mattias Polborn, The Value of Genetic Information in the Life Insurance Market, June 1998

166 Constantin Colonescu and Nicolas Schmitt, Market Segmentation, Market Integration and Tacit Collusion, June 1998

167 Ronnie Schöb and David E. Wildasin, Economic Integration and Labor Market Institutions; Worker Mobility, Earnings Risk, and Contract Structure, June 1998

168 Christian Gollier and Jean-Charles Rochet, Discounting an Uncertain Future, June 1998

169 John G. Treble, Point Schemes, August 1998

170 Jay Pil Choi, Tying and Innovation: A Dynamic Analysis of Tying Arrangements, August 1998

171 Michael Reiter and Alfons J. Weichenrieder, Public Goods, Club Goods and the Measurement of Crowding, September 1998

172 Jay Pil Choi and Marcel Thum, The Economics of Repeated Extortion, October 1998

173 Lars Calmfors, Unemployment, Labour-Market Reform and Monetary Union, November 1998

174 Lars Calmfors, Monetary Union and Precautionary Labour-Market Reform, November 1998

175 Erkki Koskela and Ronnie Schöb, Why Governments Should Tax Mobile Capital in the Presence of Unemployment, November 1998

176 Erkki Koskela, Ronnie Schöb and Hans-Werner Sinn, Green Tax Reform and Competitiveness, December 1998

177 Sheilagh Ogilvie and Jeremy Edwards, Women and the "Second Serfdom": Evidence From Bohemia, December 1998

178 James A. Robinson, When is a State Predatory?, January 1999

179 Hans-Werner Sinn, Inflation and Welfare, Comment on Robert Lucas, Februar 1999

180 Jeremy Edwards and Marcus Nibler, Corporate Governance in Germany: The Influence of Banks and Large Equity-Holders, February 1999

181 Alessandro Balestrino, Alessandro Cigno and Anna Pettini, Direct and Indirect Taxation when Households Differ in Market and Non-market Abilities, March 1999

182 Hans-Werner Sinn, EU Enlargement, Migration, and Lessons from German Unification, April 1999

183 Ernst Fehr and Simon Gächter, Cooperation and Punishment in Public Goods Experiments, April 1999

184 Ernst Fehr and Jean-Robert Tyran, Does Money Illusion Matter? An Experimental Approach, April 1999

185 Holger Feist and Ronnie Schöb, Workfare in Germany and the Problem of Vertical Fiscal Externalities, May 1999

186 Alfons J. Weichenrieder, How Efficient is a Contestable Natural Monopoly?, May 1999

187 Yoshio Kimura, Reexamining the Estimation of Simultaneous Equations Systems, May 1999