NORTH-HOLLAND

# Kantorovich and Cauchy-Schwarz Inequalities Involving Positive Semidefinite Matrices, and Efficiency Comparisons for a Singular Linear Model

Shuangzhe Liu
*Institut für Statistik und Oekonometrie*
*der Universität Basel*
*Holbeinstrasse 12, CH-4051 Basel, Switzerland*

and

Heinz Neudecker
*Institute of Actuarial Science and Econometrics*
*University of Amsterdam*
*Roetersstraat 11*
*1018 WB Amsterdam, The Netherlands*

Submitted by Simo Puntanen

ABSTRACT

Matrix Kantorovich inequalities involving two positive semidefinite matrices are presented. Corresponding Cauchy-Schwarz inequalities are discussed. Some of these are used to compare several efficient and inefficient estimators for a singular linear model. © Elsevier Science Inc., 1997

## 1. INTRODUCTION

The last decade has witnessed considerable progress in the study of the Kantorovich inequality (KI) and extensions, and applications in statistics. For recent developments we mention Wang and Chow (1994) and references therein. Marshall and Olkin (1990) and Baksalary and Puntanen (1991) presented matrix versions of the KI involving one positive definite or semidefinite matrix. Wang and Shao (1992) presented a matrix version of the KI involving two positive definite matrices and obtained an upper bound for the

*LINEAR ALGEBRA AND ITS APPLICATIONS* 259:209–221 (1997)


655 Avenue of the Americas, New York, NY 10010

0024-3795/97/$17.00
PII S0024-3795(96)00284-4

asymptotic covariance matrix of a weighted least-squares estimator. Mond and Pečarić (1993), Liu and Neudecker (1996), and Pečarić, Puntanen, and Styan (1996) derived some matrix Kantorovich-type inequalities (KTIs) involving one positive definite or semidefinite matrix. Mond and Pečarić (1994) and Liu (1995) gave additional results.

In this paper we obtain new matrix versions of the KI involving one or two positive semidefinite matrices and give statistical applications for some of them. Note that the KIs link nicely with Cauchy-Schwarz inequalities (CSIs), which are applied widely. As corresponding complements of the KIs, several Cauchy-Schwarz or CS-type inequalities (CSTIs) are briefly studied.

In Section 2 we report some basic results. In Section 3 we introduce Theorem 1, which is a matrix version of the KI for the case of one positive semidefinite matrix. We then generalize Theorem 1 to derive Theorems 2 through 4 for two positive semidefinite matrices. As appropriate complements, several CSIs (or CSTIs) are included. Mainly the first two theorems will be applied. In Section 4, we examine the relative efficiency of an extended weighted least-squares estimator (EWLSE) in comparison with the optimal extended weighted least-squares estimator (OEWLSE). We finish the paper with a concluding remark in the last section.

## 2. BASICS OF THE KI AND EFFICIENCY COMPARISON

All matrices, vectors and scalars discussed in this paper will be real. For symmetric matrices $G$ and $H$, $G \geqslant H$ will mean that $G - H$ is positive semidefinite. Let $\lambda_1 \geqslant \cdots \geqslant \lambda_r$ indicate the nonzero eigenvalues of an $n \times n$ positive semidefinite matrix of rank $r$. Let $(\cdot)^{1/2}$ be the positive semidefinite square root of a positive semidefinite matrix. Let $(\cdot)^-$ indicate a generalized matrix inverse, $(\cdot)^+$ indicate the Moore-Penrose inverse, $\mathfrak{R}(\cdot)$ indicate the column space of a matrix, and $r(\cdot)$ indicate its rank. For basic matrix algebra and its statistical applications in the context of linear models, see e.g. Rao (1973), Magnus and Neudecker (1991), Toutenburg (1992), and Wang and Chow (1994). Here we mention briefly (a vector version of) the KI and the relative efficiency of least-squares estimators (LSEs) in comparison with generalized least-squares estimators (GLSEs).

The KI asserts that

$$(x'x)^{-1}x'Ax(x'x)^{-1} \leqslant \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}(x'A^{-1}x)^{-1}, \tag{2.1}$$

where $A > 0$ is an $n \times n$ positive definite matrix with eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_n$, and $x$ is an arbitrary nonzero $n \times 1$ vector.

Consider then the following nonsingular linear model:

$$y = X\beta + \epsilon, \tag{2.2}$$

where $y$ is an $n \times 1$ observation vector, $X$ is an $n \times k$ model matrix with possibly deficient rank, $\beta$ is a $k \times 1$ parameter vector, and $\epsilon$ is an $n \times 1$ error vector with mean $E(\epsilon) = 0$ and covariance matrix $D(\epsilon) = \Omega > 0$.

For any estimable function $c'\beta$, the LSE and its variance are

$$c'b_L = c'(X'X)^+ X'y, \tag{2.3}$$

and

$$D(c'b_L) = c'(X'X)^+ X'\Omega X(X'X)^+ c; \tag{2.4}$$

the GLSE and its variance are

$$c'b_G = c'(X'\Omega^{-1}X)^+ X'\Omega^{-1}y, \tag{2.5}$$

and

$$D(c'b_G) = c'(X'\Omega^{-1}X)^+ c, \tag{2.6}$$

where $c \in \Re(X')$ and $(\cdot)^+$ can be replaced by $(\cdot)^-$ in (2.3)–(2.6).

In the model (2.2) the GLSE is the best linear unbiased estimator (BLUE). The relative efficiency of the LSE in comparison with the GLSE in this case is customarily defined as

$$\mathrm{eff}(c'b_L) := \frac{D(c'b_G)}{D(c'b_L)}. \tag{2.7}$$

Obviously $\mathrm{eff}(c'b_L) \leqslant 1$. Using (2.1) yields [for a nice proof in which the CSI was used, see Wang and Chow (1994, pp. 211–212); see also Magness and McGuire (1962)]

$$\mathrm{eff}(c'b_L) \geqslant \frac{4\lambda_1\lambda_n}{(\lambda_1 + \lambda_n)^2},$$

which can be rewritten in the following way:

$$D(c'b_L) \leqslant \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n} D(c'b_G), \tag{2.8}$$

where $\lambda_1 \geqslant \cdots \geqslant \lambda_n$ are the eigenvalues of $\Omega > 0$.

In Section 3, (2.8) will be used to prove part of Theorem 1. In Section 4, further results in the style of (2.8) will be presented.

## 3. MATRIX VERSIONS OF THE KI AND CSI

We now introduce the following result.

THEOREM 1.

$$Y^+ U Y^{+\prime} \leqslant \frac{(\lambda_1 + \lambda_r)^2}{4\lambda_1 \lambda_r} (Y' U^+ Y)^+, \tag{3.1}$$

*where the $n \times n$ positive semidefinite matrix $U \geqslant 0$ of rank $r$ has nonzero eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_r$, $Y$ is of order $n \times k$, and $\mathfrak{R}(Y) \subset \mathfrak{R}(U)$.*

*Proof.* Let us derive a result concerning the earlier-mentioned matrices $\Omega$ and $X$. Noting that $c \in \mathfrak{R}(X')$, we write $c = X'a$ with $a$ being an arbitrary nonzero $k \times 1$ vector. From (2.4), (2.6), and (2.8) we obtain

$$X^+ \Omega X^{+\prime} \leqslant \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1 \lambda_n} (X' \Omega^{-1} X)^+, \tag{3.2}$$

for $\Omega > 0$ and $X$ of order $n \times k$.

We then consider the case $U \geqslant 0$ and $\mathfrak{R}(Y) \subset \mathfrak{R}(U)$. Write $U = T\Lambda T'$, $Y = TZ$, and $T'T = I_r$, where the $r \times r$ diagonal matrix $\Lambda > 0$, $T$ is of order $n \times r$, $Z$ is of order $r \times k$, and $I_r$ is the $r \times r$ identity matrix. Therefore $U^+ = T\Lambda^{-1}T'$, $Y'U^+Y = Z'\Lambda^{-1}Z$, $Y^+ = Z^+T'$, and $Y^+UY^{+\prime} = Z^+\Lambda Z^{+\prime}$. As $\Lambda$ contains the nonzero eigenvalues of $U$, using (3.2) we get (3.1) immediately. Note that $TT' \neq I_n$, as $\Lambda > 0$. ∎

The way of proving (3.2) is similar to that of proving Theorem 2 in Wang and Shao (1992). In (3.1), $Y'U^+Y = Y'U^-Y$ for $\mathfrak{R}(Y) \subset \mathfrak{R}(U)$. Noting this point we only use the Moore-Penrose inverse for such expressions in the following.

Before proceeding further, we present a useful lemma.

LEMMA. *Under either of the two conditions*

(i) $BC = CB$, $B, C \geqslant 0$, *or*
(ii) $\mathfrak{R}(B) = \mathfrak{R}(C)$, $B, C \geqslant 0$,

*we have*

$$(B^{1/2}C^{+}B^{1/2})^{+} = B^{+1/2}CB^{+1/2}. \tag{3.3}$$

*Proof.* (i): As $BC = CB$, $B, C \geqslant 0$, we obviously have $B = T\Lambda T'$, $C = T\Gamma T'$, where $\Lambda, \Gamma \geqslant 0$ are $n \times n$ diagonal matrices, $T'T = TT' = I_n$. Therefore $B^{1/2} = T\Lambda^{1/2}T'$, $B^{+1/2} = T\Lambda^{+1/2}T'$, $C^{+} = T\Gamma^{+}T'$, and $(B^{1/2}C^{+}B^{1/2})^{+} = (T\Lambda^{1/2}\Gamma^{+}\Lambda^{1/2}T')^{+} = T\Lambda^{+1/2}\Gamma\Lambda^{+1/2}T' = B^{+1/2}CB^{+1/2}$.

(ii): Using $\mathfrak{R}(C) \subset \mathfrak{R}(B)$ and $r(C) = r(B) = r$, we write $B = S\Delta S'$, where the diagonal matrix $\Delta > 0$ is of order $r \times r$, $S$ is of order $n \times r$, $S'S = I_r$, and $C^{1/2} = SL = L'S'$ for some $L$. Then $C = C^{1/2}C^{1/2} = S\Pi S'$ with $\Pi = LL' > 0$, where $r(\Pi) = r(L) = r(C) = r(B) = r(\Delta) = r(S) = r$. Hence we obtain $(B^{1/2}C^{+}B^{1/2})^{+} = (S\Delta^{1/2}\Pi^{-1}\Delta^{1/2}S')^{+} = S\Delta^{-1/2}\Pi\Delta^{-1/2}S'$ $= B^{+1/2}CB^{+1/2}$. Note that $SS' \neq I_n$. ■

Now we are in a position to present several matrix versions of the KI involving two positive semidefinite matrices. This covers the case of two positive definite matrices as considered by Wang and Shao (1992).

THEOREM 2. *Let*

(i) $BC = CB$, $B, C \geqslant 0$, $\mathfrak{R}(X) \subset \mathfrak{R}(B)$, *and* $\mathfrak{R}(X) \subset \mathfrak{R}(C)$, *or*
(ii) $\mathfrak{R}(B) = \mathfrak{R}(C)$, $B, C \geqslant 0$, *and any compatible* $X$.

*Also let* $\lambda_1 \geqslant \cdots \geqslant \lambda_r$ *be the nonzero eigenvalues of* $BC^{+}$. *Then*

$$(X'BX)^{+}X'BC^{+}BX(X'BX)^{+} \leqslant \frac{(\lambda_1 + \lambda_r)^2}{4\lambda_1\lambda_r}(X'CX)^{+}. \tag{3.4}$$

*Proof.* Let $U = B^{1/2}C^{+}B^{1/2} \geqslant 0$. The lemma then implies $U^{+} = B^{+1/2}CB^{+1/2}$. Let $Y = B^{1/2}X$; then $X'BX = Y'Y$, $X'BC^{+}BX = Y'UY$, and $(X'BX)^{+}X'BC^{+}BX(X'BX)^{+} = Y^{+}UY^{+\prime}$. As (i) $\mathfrak{R}(X) \subset \mathfrak{R}(B)$, i.e. $X = B^{+1/2}B^{1/2}X$, or (ii) $\mathfrak{R}(B) = \mathfrak{R}(C)$, and hence $B^{1/2}B^{+1/2}C = C$, we always have $X'CX = X'B^{1/2}B^{+1/2}CB^{+1/2}B^{1/2}X = Y'U^{+}Y$. In order to apply Theorem 1, we have to show that $\mathfrak{R}(Y) \subset \mathfrak{R}(U)$. In fact, we have in the two cases:

(i): If $BC = CB$, then $B^{1/2}C^{+} = B^{1/2}C^{+}B^{1/2}B^{+1/2} = UB^{+1/2}$. As $\mathfrak{R}(X) \subset \mathfrak{R}(C)$, we have $X = C^{+}L$ for some $L$. Hence $Y = B^{1/2}X = B^{1/2}C^{+}L = UB^{+1/2}L$, i.e. $\mathfrak{R}(Y) \subset \mathfrak{R}(U)$.

(ii): If $\mathfrak{R}(B) = \mathfrak{R}(C)$, then $B^{1/2} = B^{1/2}C^{+}C$ and $C^{+} = C^{+}B^{1/2}B^{+1/2}$. Hence $Y = B^{1/2}X = B^{1/2}C^{+}B^{1/2}B^{+1/2}CX = UB^{+1/2}CX$, i.e. $\mathfrak{R}(Y) \subset \mathfrak{R}(U)$.

Then (3.4) follows. ■

We can obtain two additional theorems, which are similar to Theorem 2.

THEOREM 3. *Let $G$ and $H$ be $n \times n$ symmetric matrices and $GH = HG \geqslant 0$ (then $GH^{+} \geqslant 0$). Let $\lambda_1 \geqslant \cdots \geqslant \lambda_r$ be the nonzero eigenvalues of $GH^{+}$, $\mathfrak{R}(X) \subset \mathfrak{R}(G)$, and $\mathfrak{R}(X) \subset \mathfrak{R}(H)$. We have*

$$(X'GHX)^{+} X'G^2X(X'GHX)^{+} \leqslant \frac{(\lambda_1 + \lambda_r)^2}{4\lambda_1\lambda_r}(X'H^2X)^{+}. \quad (3.5)$$

*Proof.* This result follows by inserting $U = GH^{+}$, $Y = (GH)^{1/2}X$ [clearly $U \geqslant 0$ and $\mathfrak{R}(Y) \subset \mathfrak{R}(U)$ in this case], and therefore $Y'Y = X'GHX$, $Y'UY = X'G^2HH^{+}X = X'G^2X$, and $Y'U^{+}Y = X'H^2GG^{+}X = X'H^2X$ in (3.1). ■

For related results on matrix determinants and traces, see Khatri and Rao (1981, 1982).

THEOREM 4. *Let $\lambda_1 \geqslant \cdots \geqslant \lambda_r$ generically denote the nonzero eigenvalues of $n \times n$ matrix $U^{k-h}$, where $U \geqslant 0$, and $k$ and $h$ are arbitrary scalars. We then have for any compatible matrix $X$*

$$(X'U^kX)^{+} X'U^{2k-h}X(X'U^kX)^{+} \leqslant \frac{(\lambda_1 + \lambda_r)^2}{4\lambda_1\lambda_r}(X'U^hX)^{+}, \quad (3.6)$$

*Proof.* By replacing $U$ and $Y$ in (3.1) by $U^{k-h}$ and $U^{k/2}X$ respectively, we get this result immediately. ■

Keep in mind that $U^0 = UU^+$ and $U^{-\alpha} = (U^+)^\alpha$ for $\alpha$ a positive scalar. In fact, Theorem 4 is a special case of Theorem 2(ii) with $B = U^k$ and $C = U^h$.

REMARK.

(a) It is also worthwhile to present several complementary CSIs (or CSTIs). A more general inequality is

$$Z'F'X(X'CX)^+X'FZ \leqslant Z'F'C^+FZ, \tag{3.7}$$

where $Z$, $F$, and $X$ are compatible matrices, $C \geqslant 0$, and $\mathfrak{R}(X) \subset \mathfrak{R}(C)$ or $\mathfrak{R}(F) \subset \mathfrak{R}(C)$. This is proved as follows. As $\mathfrak{R}(X) \subset \mathfrak{R}(C)$ or $\mathfrak{R}(F) \subset \mathfrak{R}(C)$, then $X'F = X'C^{1/2}C^{+1/2}F$. Pre- and postmultiplying $C^{1/2}X(X'CX)^+X'C^{1/2} \leqslant I$ by $Z'F'C^{+1/2}$ and $C^{+1/2}FZ$ respectively, we get (3.7). We can also prove (3.7) by using (due to a referee)

$$\begin{aligned} &F'C^+F - F'X(X'CX)^+X'F \\ &\quad = F'C^+CC^+F - F'X(X'CX)^+X'CX(X'CX)^+X'F \\ &\quad = \left[F'C^+ - F'X(X'CX)^+X'\right]C\left[C^+F - X(X'CX)^+X'F\right] \geqslant 0. \end{aligned}$$

Consequently for $B, C \geqslant 0$, and (i) $\mathfrak{R}(X) \subset \mathfrak{R}(C)$ or (ii) $\mathfrak{R}(B) \subset \mathfrak{R}(C)$, we have

$$X'BX(X'CX)^+X'BX \leqslant X'BC^+BX. \tag{3.8}$$

The counterpart of (3.4) is the following CSI:

$$(X'CX)^+ \leqslant (X'BX)^+X'BC^+BX(X'BX)^+, \tag{3.9}$$

for $B, C \geqslant 0$, and

(i) $r(X'BX) = r(X)$ and $\mathfrak{R}(X) \subset \mathfrak{R}(C)$ or
(ii) $r(X'BX) = r(X)$ and $\mathfrak{R}(B) \subset \mathfrak{R}(C)$ or
(iii) $\mathfrak{R}(B) = \mathfrak{R}(C)$ and any $X$.

From (3.8) we can easily derive (3.9) under (i) and (ii) by using $X(X'BX)^+X'BX = X$, which is equivalent to $r(X'BX) = r(X)$ provided above [we see this equivalence by noting that $\mathfrak{R}(X'BX) \subset \mathfrak{R}(X')$ with $r(X'BX) = r(X')$ yields $\mathfrak{R}(X'BX) = \mathfrak{R}(X')$; for related matters, see Rao and Mitra (1971, pp. 22–23), Baksalary and Puntanen (1989), and Searle (1994)], or under (iii) by using $\mathfrak{R}(B) = \mathfrak{R}(C)$, $CB^+B = C$, and hence $CX(X'BX)^+X'BX = CX$ for any $X$. For special cases of (3.9), see e.g. Gaffke and Krafft (1977), Pukelsheim and Styan (1983), and Toutenburg (1992, pp. 104–105).

(b) Under the conditions of Theorem 3 we have

$$X'G^2X \leqslant \frac{(\lambda_1 + \lambda_r)^2}{4\lambda_1\lambda_r} X'GHX(X'H^2X)^+ X'GHX. \tag{3.10}$$

We use $X'GHX(X'GHX)^+X'G^2 = X'G^2$ and $X'GHX(X'GHX)^+X'H^2 = X'H^2$, and hence prove that (3.10) is equivalent to (3.5). From $HX(X'H^2X)^+X'H \leqslant I$, a CSI counterpart follows:

$$X'GHX(X'H^2X)^+ X'GHX \leqslant X'G^2X, \tag{3.11}$$

for $H$ symmetric, $G$ such that $HG = GH$, and $X$ a compatible matrix.

(c) The following result is equivalent to (3.6):

$$X'U^{2k-h}X \leqslant \frac{(\lambda_1 + \lambda_r)^2}{4\lambda_1\lambda_r} X'U^kX(X'U^hX)^+ X'U^kX. \tag{3.12}$$

We easily prove this equivalence by noting that $U^{2k-h} = U^{2k-h}U^0$, $U^h = U^hU^0$, and $U^0X(X'U^kX)^+X'U^kX = U^0X$. A CSTI counterpart of (3.12) is

$$Z'U^kX(X'U^hX)^+ X'U^kZ \leqslant Z'U^{2k-h}Z, \tag{3.13}$$

where $U \geqslant 0$, $X$ is of order $n \times p$, and $Z$ is any compatible matrix. Insertion of $C = U^h$ and $F' = F = U^k$ in (3.7) yields (3.13) immediately. For (3.13) with $h$ and $2k - h$ being any integers, and various CSTIs, see Pečarić, Puntanen, and Styan (1996).

We shall now look into statistical applications of some of these results.

## 4. A SINGULAR MODEL AND A COMPARISON BETWEEN AN EWLSE AND THE OEWLSE

Consider the singular model

$$y = X\beta + \epsilon, \qquad E(\epsilon) = 0, \qquad D(\epsilon) = \Psi \geqslant 0, \tag{4.1}$$

where $\Psi$ has nonzero eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_r$.

It is well known that $X\beta$ is an estimable function. To estimate $X\beta$, we use the following extended weighted least-squares estimator (EWLSE) with $W \geqslant 0$ a given positive semidefinite weight matrix, viz.

$$Xb_W = X(X'WX)^+ X'Wy. \tag{4.2}$$

The covariance matrix of this estimator is

$$D(Xb_W) = X(X'WX)^+ X'W\Psi WX(X'WX)^+ X'. \tag{4.3}$$

We first make the following alternative assumptions:

$$r(X'WX) = r(X) \quad \text{and} \quad \mathfrak{R}(X) \subset \mathfrak{R}(\Psi) \tag{4.4}$$

or

$$r(X'WX) = r(X) \quad \text{and} \quad \mathfrak{R}(W) \subset \mathfrak{R}(\Psi) \tag{4.5}$$

or

$$\mathfrak{R}(W) = \mathfrak{R}(\Psi). \tag{4.6}$$

Of these three assumptions the first is very common. The condition $r(X'WX) = r(X)$, i.e. $X(X'WX)^+X'WX = X$, makes of (4.2) an unbiased estimator. For using $(X'WX)^-$ instead of $(X'WX)^+$, see e.g. Baksalary and Puntanen (1989) and Searle (1994).

There are several choices for $W$ in (4.2). An important one is $W_O = \Psi^+$. It yields the estimator

$$Xb_O = X(X'\Psi^+X)^+ X'\Psi^+ y. \tag{4.7}$$

Its covariance matrix is

$$D(Xb_O) = X(X'\Psi^+X)^+ X'\Psi^+X(X'\Psi^+X)^+ X' = X(X'\Psi^+X)^+ X'. \tag{4.8}$$

Under any of the three groups of assumptions stated above, we have by using the CSI (3.9)

$$D(Xb_O) \leqslant D(Xb_W). \tag{4.9}$$

In this sense, (4.7) is an optimal extended weighted least squares estimator (OEWLSE).

We further focus attention on the following alternative assumptions:

$$\Re(X) \subset \Re(W), \qquad \Re(X) \subset \Re(\Psi), \quad \text{and} \quad W\Psi = \Psi W \tag{4.10}$$

or

$$r(X'WX) = r(X) \quad \text{and} \quad \Re(W) = \Re(\Psi) \tag{4.11}$$

or

$$\Re(W) = \Re(\Psi). \tag{4.12}$$

The two assumptions (4.10) and (4.11) are stronger than (4.4) and (4.5) respectively, and the assumption (4.12) is the same as (4.6). Under any of these three groups of assumptions (4.10), (4.11), and (4.12), we obtain an efficiency comparison by applying Theorem 2, viz.

$$D(Xb_W) \leqslant \frac{(\lambda_1 + \lambda_r)^2}{4\lambda_1\lambda_r} D(Xb_O), \tag{4.13}$$

where $\lambda_1 \geqslant \cdots \geqslant \lambda_r$ are the nonzero eigenvalues of $W\Psi$.

From (4.9) we see that in the class of estimators concerned, the OEWLSE is more efficient than the EWLSE [in fact, if $\Re(X) \subset \Re(\Psi)$, then $W = \Psi^-$ can be used to give the unique BLUE estimator—see Theorem 1 in Baksalary and Puntanen (1989) and Theorem 2 in Searle (1994); for an interesting related study in the context of minimum-distance estimators, see Satorra and Neudecker (1994)]. Hence (4.9) is an extended Gauss-Markov-

Aitken theorem. From a different point of view, (4.13) also compares the two covariance matrices and presents the relative (in)efficiencies of EWLSE vis-a-vis OEWLSE. Note that (4.13) is in the style of (2.8). If $(\lambda_1 + \lambda_r)^2/4\lambda_1\lambda_r$ in (4.13) is not too far from one, we can say that $D(Xb_W)$ is close to $D(Xb_O)$, and therefore can use $Xb_W$ instead of $Xb_O$ in practical situations.

We shall now study in detail the special case of $W = I$. It is easy to see that two conditions $\mathfrak{R}(X) \subset \mathfrak{R}(W)$ and $W\Psi = \Psi W$ of the assumption (4.10) are satisfied [the condition $\mathfrak{R}(X) \subset \mathfrak{R}(\Psi)$ has still to be assumed]. In this case the EWLSE, i.e. (4.2), becomes the LSE, which has covariance matrix $XX^+\Psi X^{+\prime}X'$, and (4.7) remains the GLSE. As $\mathfrak{R}(X) \subset \mathfrak{R}(\Psi)$, the GLSE is actually the BLUE now; see e.g. Corollary 2 in Baksalary and Puntanen (1989) and Theorem 2 in Searle (1994). Compare also (2.3) through (2.6) in Section 2 for the LSE, the GLSE, and their covariance matrices.

By applying Theorem 1 we get the following comparison between the LSE and the GLSE:

$$D(Xb_L) \leqslant \frac{(\lambda_1 + \lambda_r)^2}{4\lambda_1\lambda_r} D(Xb_G), \tag{4.14}$$

where $\Psi \geqslant 0$ has nonzero eigenvalues $\lambda_1 \geqslant \cdots \geqslant \lambda_r$, and $\mathfrak{R}(X) \subset \mathfrak{R}(\Psi)$.

Also, under the assumptions (4.11) and (4.12), $W = I$ implies $\Psi > 0$. Then (4.14) holds automatically. For related discussions for the case $\Psi > 0$, see Wang and Chow (1994) and Pečarić, Puntanen, and Styan (1996). See also (2.8). For corresponding measures of efficiency in terms of matrix determinants and traces, and more related results, see e.g. Bloomfield and Watson (1975), Knott (1975), Khatri and Rao (1981, 1982), Styan (1983), Rao (1985), and Liski, Puntanen, and Wang (1992).

## 5. CONCLUDING REMARK

We have shown in this paper that KIs can be used to study the relative efficiencies of several LS-type estimators. KIs can cover not only the case of two positive definite matrices [see e.g. Wang and Shao (1992)], but also the case of two positive semidefinite matrices as discussed already. The algebraic treatment in Section 4 is very general and can be applied widely, e.g. to compare minimum-distance estimators whose asymptotic covariance matrices also involve two positive semidefinite matrices [see e.g. Satorra and Neudecker (1994)].

*We thank the referees for very helpful comments. An earlier version of this paper was presented at the 4th International Workshop on Matrix Methods for Statistics, Montréal, Canada. Financial supports for Liu from the Netherlands Organization for Scientific Research (NWO) and Shell Nederland are much appreciated.*

## REFERENCES

Baksalary, J. K. and Puntanen, S. 1989. Weighted-least-squares estimation in the general Gauss-Markov model, in *Statistical Data Analysis and Inference* (Y. Dodge, Ed.), Elsevier Science (North-Holland), Amsterdam, pp. 355–368.

Baksalary, J. K. and Puntanen, S. 1991. Generalized matrix versions of the Cauchy-Schwarz and Kantorovich inequalities, *Aequationes Math.* 41:103–110.

Bloomfield, P. and Watson, G. S. 1975. The inefficiency of least squares, *Biometrika* 62:121–128.

Gaffke, N. and Krafft, O. 1977. Optimum properties of Latin square designs and a matrix inequality, *Statistics* 8:345–350.

Khatri, C. G. and Rao, C. R. 1981. Some extensions of the Kantorovich inequality and statistical applications, *J. Multivariate Anal.* 11:498–505.

Khatri, C. G. and Rao, C. R. 1982. Some generalizations of the Kantorovich inequality, *Sankhyā Ser. A* 44:91–102.

Knott, M. 1975. On the minimum efficiency of least squares, *Biometrika* 62:129–132.

Liski, E. P., Puntanen, S. and Wang, S. G. 1992. Bounds for the trace of the difference of the covariance matrices of the OLSE and BLUE, *Linear Algebra Appl.* 176:121–130.

Liu, S. 1995. An inequality involving submatrices, *Econometric Theory* 11:191.

Liu, S. and Neudecker, H. 1996. Several matrix Kantorovich-type inequalities, *J. Math. Anal. Appl.* 197:23–26.

Magness, T. A. and McGuire, J. B. 1962. Comparison of least squares and minimum variance estimates of regression parameters, *Ann. Math. Statist.* 33:462–470.

Magnus, J. R. and Neudecker, H. 1991. *Matrix Differential Calculus with Applications in Statistics and Econometrics*, rev. ed., Wiley, Chichester.

Marshall, A. W. and Olkin, I. 1990. Matrix versions of the Cauchy and Kantorovich inequalities, *Aequationes Math.* 40:89–93.

Mond, B. and Pečarić, J. E. 1993. Matrix versions of some means inequalities, *Austral. Math. Soc. Gaz.* 20:117–120.

Mond, B. and Pečarić, J. E. 1994. A matrix version of the Ky Fan generalization of the Kantorovich inequality, *Linear and Multilinear Algebra* 36:217–221.

Pečarić, J. E., Puntanen, S., and Styan, G. P. H. 1996. Some further matrix extensions of the Cauchy-Schwarz and Kantorovich inequalities, with some statistical applications, *Linear Algebra Appl.* 237/238:455–476.

Pukelsheim, F. and Styan, G. P. H. 1983. Convexity and monotonicity of dispersion matrices of estimators in linear models, *Scand. J. Statist.* 10:145–149.

Rao, C. R. 1973. *Linear Statistical Inference and Its Applications*, 2nd ed., Wiley, New York.

Rao, C. R. 1985. The inefficiency of least squares: Extensions of the Kantorovich inequality, *Linear Algebra Appl.* 70:249–255.

Rao, C. R. and Mitra, S. K. 1971. *Generalized Inverse of Matrices and its Applications*, Wiley, New York.

Satorra, A. and Neudecker, H. 1994. On the asymptotic optimality of alternative minimum-distance estimators in linear latent-variable models, *Econometric Theory* 10:867–883.

Searle, S. R. 1994. Extending some results and proofs for the singular linear model, *Linear Algebra Appl.* 210:139–151.

Styan, G. P. H. 1983. On some inequalities associated with ordinary least squares and the Kantorovich inequality, in Festschrift for Eino Haikala on his Seventieth Birthday, *Acta Univ. Tamper. Ser. A* 153:158–166.

Toutenburg, H. 1992. *Lineare Modelle*, Physica-Verlag, Heidelberg.

Wang, S. G. and Chow, S. C. 1994. *Advanced Linear Models*. Dekker, New York.

Wang, S. G. and Shao, J. 1992. Constrained Kantorovich inequalities and relative efficiency of least squares, *J. Multivariate Anal.* 42:284–298.

*Received 23 September 1995; final manuscript accepted 6 May 1996*