SOME EFFICIENCY COMPARISONS FOR ESTIMATORS FROM QUASI-LIKELIHOOD AND GENERALIZED ESTIMATING EQUATIONS

S. Liu
Australian National University

C.C. Heyde
Australian National University and *Columbia University*

This paper is concerned with general methods for making efficiency comparisons for estimators with the aid of matrix inequalities. It is shown that a unified approach is possible for several distinct cases in the general regression model, one involving quasi-likelihood estimation and the other two generalized estimating equations. The paper includes some specific comparisons between an ordinary least squares estimator and the best linear unbiased estimator for an error component model. We also present some numerical examples.

1. Introduction

It is always interesting and useful to make efficiency comparisons in estimation in statistics. For example, simple estimators which are suboptimal may suffer little loss in efficiency relative to likelihood based estimators which are difficult to use. For regression models and inferences, Heyde (1997) and Rao and Rao (1998) present various results. In the present paper, we focus on the following cases for which we give a unified treatment:

1. Heyde (1989) introduces composite quasi-likelihood estimators (QLE). The comparison ensures that composition is generally advantageous. Heyde and Lin (1992) and Heyde (1997) study quasi-likelihood estimators for the general linear model. Two alternatives are $\hat{\theta}_A$ and $\hat{\theta}_{QS(V)}$, the latter being prefered on efficiency grounds.

2. Balemi and Lee (1999) make an application to clustered binary regression in the context of generalized estimating equation (GEE) (see Liang and Zeger, 1986; McCullagh and Nelder, 1989, Section 9.4), and include an efficiency comparison involving a working correlation matrix R and the correct correlation matrix R_0.

3. Wang and Shao (1992) and Liu and Neudecker (1997) study Σ, the asymptotic variance matrix of an estimator $\hat{\beta}_I$ under the independence working assumption studied by Liang and Zeger (1986). Wang and Shao (1992) give Σ an upper bound in the Löwner order sense. Liu and Neudecker (1997) give both the determinant $|\Sigma|$ and the trace $tr\Sigma$ an upper bound.

Keywords and phrases: Regression models; asymptotic variances; correlation matrices; efficiency measures; upper bounds.

4. Mukhopadhyay and Schwabe (1998) present a comparison result between an ordinary least squares estimator (OLSE) and the best linear unbiased estimator (BLUE) for an error component model. This is an upper bound for the largest eigenvalue of the difference of the two variance matrices.

It is noted in the overview of Heyde (1997, Chapter 2) that GEE is closely linked with QLE. The QLE/GEE framework is very general. We then make a further study of comparisons for these cases. The structure of the rest of the paper is as follows: In Section 2 we give a brief technical introduction to the four cases. We introduce some comparison measures in Section 3, and then establish an upper bound for each of the comparison measures in Section 4. We present some numerical examples in Section 5. In the last section, we make some supplementary remarks.

2. Four cases

Case 1. Consider the general linear model (see Heyde and Lin, 1992; Heyde, 1997, Section 11.2)

$$y = X\beta + u, \tag{2.1}$$

where y is an $n \times 1$ vector of observations, $X = (X_1, \ldots, X_n)'$ is an $n \times k$ matrix of full column rank, β is a $k \times 1$ vector of unknown parameters and u is an $n \times 1$ vector of independent residuals with mean zero and variance matrix

$$\Omega = \operatorname{diag}\big(g_1(\theta), \ldots, g_n(\theta)\big),$$

where $g_i(\theta) = \sigma^2(X_i'\beta)^{2(1-\alpha)}$ is a scalar function of θ, $\theta = (\beta', \sigma^2, \alpha)'$ is a $p \times 1$ parameter vector ($p = k+2$), and $X_i = (X_{i1}, \ldots, X_{ik})'$ is a $k \times 1$ vector, $i = 1, \ldots, n$.

Using two different estimating functions we can obtain the two estimators of interest, namely $\hat\theta_A$ and $\hat\theta_{QS(V)}$, though Anh (1988) derives $\hat\theta_A$ via nonlinear least squares. For the efficient estimation of θ, it is shown that $\hat\theta_A$ is a strongly consistent and asymptotically normal estimator with asymptotic variance

$$V_1 = (A'A)^{-1}A'FA(A'A)^{-1}, \tag{2.2}$$

and $\hat\theta_{QS(V)}$ is a quasi-likelihood estimator with asymptotic variance

$$V_2 = (A'F^{-1}A)^{-1}, \tag{2.3}$$

where $F = F(\theta) > 0$ is a known function of θ and in this context is a diagonal and positive definite matrix, $A = (\partial g_i(\theta)/\partial\theta_j)$ is an $n \times p$ matrix of full column rank. We have in the Löwner order

$$V_1 \geq V_2. \tag{2.4}$$

This means that $V_1 - V_2 \geq 0$ is nonnegative definite, so that $\hat{\theta}_{QS(V)}$ is prefered to $\hat{\theta}_A$, in terms of asymptotic efficiency. This result can be derived from (and viewed as an analogous version of) the Gauss–Markov theorem; see Heyde (1989).

Case 2. Balemi and Lee (1999) study

$$\mu_{ki} = E(Y_{ki}), \tag{2.5}$$

where $Y_k = (Y_{k1}, \ldots, Y_{kn_k})'$ denotes the $n_k \times 1$ binary response vector for cluster k ($i = 1, \ldots, n_k$; $k = 1, \ldots, K$; $n_1 + \cdots + n_K = n$), Y_{ki} is the response from the ith unit in the kth cluster, μ_{ki} is a function of the covariates vector x_{ki} and a vector β of regression parameters, assuming a generalized linear model for the marginal responses.

They compare the efficiency of the estimator based on the exchangeable working correlation relative to the efficiency that could be achieved if the true correlation was known. The leading term in the variance of their estimator is

$$V_3 = (X'R^{-1}X)^{-1}X'R^{-1}R_0R^{-1}X(X'R^{-1}X)^{-1}, \tag{2.6}$$

where R is an $n \times n$ block diagonal working correlation matrix, $R_0 > 0$ is the $n \times n$ block diagonal correct correlation matrix and X is an $n \times p$ matrix. We define

$$V_4 = (X'R_0^{-1}X)^{-1},$$

and then obtain

$$V_3 \geq V_4. \tag{2.7}$$

This means that maximum efficiency is achieved by using R_0 as R. To study how much we can lose by getting the correlation wrong, we can calculate the relative efficiency $e(c)$ of $c'\hat{\beta}_{\mathrm{W}}$ and $c'\hat{\beta}_{\mathrm{T}}$ (for two estimators $\hat{\beta}_{\mathrm{W}}$ and $\hat{\beta}_{\mathrm{T}}$, the "W" and "T" standing for "working" and "true" respectively) with c, a $p \times 1$ vector:

$$e(c) = \frac{c'V_4c}{c'V_3c}. \tag{2.8}$$

However, for notation convenience we consider efficiency measure $\mathrm{E} = \mathrm{E}(c) = 1/e(c)$ instead. Following Hannan (1970, Section 7.2) or Scott and Holt (1982), Balemi and Lee (1999) give the inequalities (in our notation)

$$1 \leq \mathrm{E} \leq \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}, \tag{2.9}$$

where λ_1 and λ_n are the largest and smallest eigenvalues of $R^{-1}R_0$, respectively.

The first relation in (2.9) corresponds to (2.7) or the Gauss–Markov theorem, and the second to a counterpart to it. We see that the two estimators are equivalent if E reaches the lower bound (E $=1$), but not in general. To study E, we may need to examine its upper bound, i.e., examine how far it is from 1. If it is (relatively) close to 1, we accept that R functions (reasonably) well as R_0.

Case 3. Liang and Zeger (1986) discuss large-sample properties of the solution $\hat{\beta}_I$ of the equation $X'\Delta S = 0$, where $X = (x_1, \dots, x_n)'$, $\Delta = \operatorname{diag}\big(h(x_1'\beta), \dots, h(x_n'\beta)\big)$, $h(t) = d(g(\mu))^{-1}/dt$, $S = \big(y_1 - g^{-1}(x_1'\beta), \dots, y_n - g^{-1}(x_n'\beta)\big)'$, $E(y_i) = \mu(\theta_i)$, $D(y_i) = \phi\dot{\mu}(\theta_i)$, $g\big(\mu(\theta_i)\big) = x_i'\beta$, μ is a scalar function and $\dot{\mu}$ is the derivative, g is a known link function, x_i is a $p \times 1$ observable vector and β is a $p \times 1$ parameter vector, $i = 1, \dots, n$. See also Wang and Shao (1992), for the assumptions and notations. Under some regularity conditions, $\hat{\beta}_I$ is an asymptotically normal estimator with asymptotic variance

$$\Sigma = (X'\Delta\Lambda\Delta X)^{-1}X'\Delta D(y)\Delta X(X'\Delta\Lambda\Delta X)^{-1},$$

where $\Lambda = \operatorname{diag}[\dot{\mu}(\theta_1), \dots, \dot{\mu}(\theta_n)]$.

To study Σ, Wang and Shao (1992) consider the case in which the variance matrix $D(y) > 0$ of $y = (y_1, \dots, y_n)'$ is of the form of their (2.1), i.e., a block diagonal matrix with small block sizes. Let $\Gamma = [X'\Delta\Lambda D^{-1}(y)\Lambda\Delta X]^{-1}$, and observe that in the Löwner order

$$\Gamma \le \Sigma, \tag{2.10}$$

which again corresponds to (a version of) the Gauss–Markov theorem.

Wang and Shao (1992) give

$$\Sigma \le \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}\Gamma, \tag{2.11}$$

where $\lambda_1 \ge \cdots \ge \lambda_n$ are the (positive) eigenvalues of $\Lambda^{-1}D(y)$. Having (2.10) as a benchmark implying that Γ is the lower bound of Σ, we may just examine the number multiplied by Γ in the upper bound in (2.11); the number is larger than or equal to 1 and is a function only of the largest and smallest eigenvalues of $\Lambda^{-1}D(y)$. Essentially (2.11) is equivalent to the inequality established by Marshall and Olkin (1990).

In addition, for the same case Liu and Neudecker (1997) present

$$|\Sigma| \le \prod_{i=1}^{p} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}}|\Gamma|, \tag{2.12}$$

$$\operatorname{tr}\Sigma \le \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n}\operatorname{tr}\Gamma, \tag{2.13}$$

where Γ and $\lambda_1 \geq \cdots \geq \lambda_n$ are the same as given in (2.11) (from which (2.13) follows immediately). More than two eigenvalues λ_i $(i = 1, \ldots, p, n-p+1, \ldots, n; \ p > 1)$ are involved in the upper bound in (2.12), and then more information is used and improvements on (2.11) are expected.

Case 4. A sample survey model with OLSE has been investigated earlier by, e.g., Scott and Holt (1982) and Wang, Chow and Tse (1994). A random effects model in the context of panel or longitudinal data analysis is discussed by Liu and Neudecker (1997). An error component model is studied by Mukhopadhyay and Schwabe (1998), which covers the models in both Wang, Chow and Tse (1994) and Liu and Neudecker (1997). Most of these use OLSE and BLUE methods. We then make a further study in the same OLSE/BLUE context, which can be viewed as a special case for our general comparisons.

Consider the model

$$y = X\beta + \epsilon, \tag{2.14}$$

where y is an $n \times 1$ observation vector, $X = (x_1, \ldots, x_n)'$ is an $n \times p$ matrix, β is a $p \times 1$ unknown parameter vector and ϵ is an $n \times 1$ error vector with mean $E(\epsilon) = 0$ and variance $D(\epsilon) = \Omega$. We compare the estimators of β. Let $\hat{\beta}_0$ and $\hat{\beta}^*$ denote the OLSE and BLUE of β, respectively. Without loss of generality, we suppose $\Omega > 0$ is positive definite with eigenvalues $\lambda_1 \geq \cdots \geq \lambda_n$, X is of full column rank and $n \geq 2p$. We have

$$\hat{\beta}_0 = (X'X)^{-1}X'y, \tag{2.15}$$

$$V_5 = D(\hat{\beta}_0) = (X'X)^{-1}X'\Omega X(X'X)^{-1}, \tag{2.16}$$

$$\hat{\beta}^* = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y, \tag{2.17}$$

$$V_6 = D(\hat{\beta}^*) = (X'\Omega^{-1}X)^{-1}, \tag{2.18}$$

$$V_5 \geq V_6, \tag{2.19}$$

where (2.19) comes from a typical application of the Gauss-Markov theorem.

The error component model in Mukhopadhyay and Schwabe (1998) is a special case of (2.14), where

$$\begin{aligned} \Omega &= \sigma^2(1-\rho-\delta)I_{NT} + \sigma^2(\rho C + \delta B), \\ C &= I_N \otimes J_T, \\ B &= J_N \otimes I_T, \end{aligned}$$

$\sigma^2 > 0$, $\rho \geq 0$ and $\delta \geq 0$ are the only scalar parameters of the variance matrix, $1 - \rho - \delta > 0$, $n = NT$, I_m is an $m \times m$ identity matrix, J_m is an $m \times m$ matrix of ones and $\otimes$ indicates the Kronecker product of matrices.

For further details, see Mukhopadhyay and Schwabe (1998). In particular, when $T = 1$ this model reduces to the one in Wang et al. (1994).

To compare V_5 and V_6 in addition to (2.19), Mukhopadhyay and Schwabe's (1998) idea is to study d, defined as the largest eigenvalue of $V_5 - V_6$. They give an upper bound for d as follows:

$$d \leq \mu_p^{-1}(\lambda_1 - \lambda_n) \tag{2.20}$$

where $\lambda_1 \geq \cdots \geq \lambda_n$ are the eigenvalues of Ω, μ_p is the smallest eigenvalue of $X'X$ and the upper bound of d happens to be equal to the largest eigenvalue of $\rho C + \delta B$.

For the above cases, there are other comparison studies and results, in addition to (2.4), (2.9) and (2.10)–(2.13). For example, there are discussions on correlation structures and relative efficiency for GEE. However, most of them give calculations of efficencies themselves, based on simulation, see e.g. Sutradhar and Das (1999). In contrast, we will proceed in the following sections by studying alternative measures to E in (2.9) and establish further results. In Cases 1, 2 and 4, we will give the efficiency measures and their upper bounds. In Case 3, we will consider matrix determinants and traces involving Σ and give their upper bounds.

3. Comparison measures

For the first case, we see from (2.4) that $\hat{\theta}_A$ is not as good as $\hat{\theta}_{QS(V)}$. To further compare $\hat{\theta}_A$ with $\hat{\theta}_{QS(V)}$ from a different point of view, we can introduce and examine matrix determinants and traces as our comparison measures. The following measures are defined in a unified approach for all the four cases:

$$\begin{aligned}
d_1 &= |Z'WZZ'W^{-1}Z|^{1/p}, \\
d_2 &= \frac{1}{p}\operatorname{tr}(Z'WZZ'W^{-1}Z), \\
d_3 &= |Z'WZ - (Z'W^{-1}Z)^{-1}|^{1/p}, \\
d_4 &= \frac{1}{p}\operatorname{tr}[Z'WZ - (Z'W^{-1}Z)^{-1}], \\
d_5 &= |Z'WZZ'W^{-1}Z - I|^{1/p} = \frac{|Z'WZ - (Z'W^{-1}Z)^{-1}|^{1/p}}{|(Z'W^{-1}Z)^{-1}|^{1/p}},
\end{aligned}$$

where $W > 0$ is an $n \times n$ positive definite matrix and Z is an $n \times p$ matrix such that $Z'Z = I$. The measures d_1 and d_3 are geometric means of the eigenvalues of $Z'WZZ'W^{-1}Z$ and $Z'WZ - (Z'W^{-1}Z)^{-1}$, respectively, and d_2 and d_4 are arithmetic means. Note that d_5 is linked to d_1 and d_3 in a nice way. The idea of introducing d_1, d_2, d_3, d_4 and d_5 is to study the difference between $Z'WZ$ and $(Z'W^{-1}Z)^{-1}$, and is similar to, e.g., the one used to introduce (2.8) to study the difference between V_3 and V_4. However, they are each implemented with a specific function in a (slightly) different way. More information is then involved in and reflected by the determinant and the trace functions in the definitions, with which we expect to gain improvements in the comparisons. We choose $Z'WZZ'W^{-1}Z$ because it is a "relative" difference between $Z'WZ$ and $(Z'W^{-1}Z)^{-1}$, each of which can be a variance matrix of some estimator for the parameter vector. We use $Z'WZ - (Z'W^{-1}Z)^{-1}$ because it is an "absolute" difference. In the simplest form of the general linear model $\{y, Z\beta, W\}$, W can be the variance matrix of the error vector in the model, Z can be the design matrix with $Z'Z = I$, and then $Z'WZ$ is the variance matrix of an OLSE of β and $(Z'W^{-1}Z)^{-1}$ is the variance matrix of the BLUE. The result $Z'WZ \geq (Z'W^{-1}Z)^{-1}$ is the Gauss-Markov theorem, which implies those in which d_1, d_2, d_3, d_4 and d_5 are lower bounded. The inequalities in which d_1, d_2, d_3, d_4 and d_5 are upper bounded can be viewed as just countparts of the Gauss-Markov theorem. Also, d_1, d_2 and d_4 can be viewed as the modifications of the corresponding measures in Rao and Rao (1998, Section 14.8), by taking a geometric or arithmetic mean. It is important to take such a mean in the sense that the measures are "normalized." We wish to see how the differently "normalized" measures or their upper bounds behave or if a single one can be optimal (better than the others). Some commonly used measures whose motivation is to compare an OLSE and the BLUE can be found in Rao and Rao (1998). A collection of relevant matrix inequalities is given by Liu and Neudecker (1999). For further results and extensions, including those measures based on $Z'W^2Z - (Z'WZ)^2$, see Drury, Liu, Lu, Puntanen and Styan (2002) and references therein.

For the first case in which we compare $\hat{\theta}_A$ with $\hat{\theta}_{QS(V)}$, inserting $W = F$ and $Z = A(A'A)^{-1/2}$, we can present the comparison measures in the forms

$$f_1 = |V_1 V_2^{-1}|^{1/p}, \tag{3.1}$$

$$f_2 = \frac{1}{p}\operatorname{tr}(V_1 V_2^{-1}), \tag{3.2}$$

$$f_3 = |(A'A)^{1/2}(V_1 - V_2)(A'A)^{1/2}|^{1/p} = |A'A(V_1 - V_2)|^{1/p}, \tag{3.3}$$

$$f_4 = \frac{1}{p}\operatorname{tr}[A'A(V_1 - V_2)], \tag{3.4}$$

$$f_5 = |V_1 V_2^{-1} - I|^{1/p}, \tag{3.5}$$

where $(A'A)^{1/2}$ is a square root of $A'A$, V_1 is given in (2.2) and V_2 is in (2.3).

To compare $\hat{\beta}_W$ and $\hat{\beta}_T$ in the second case, inserting $W = R^{-1/2}R_0R^{-1/2}$ and $Z = R^{-1/2}X(X'R^{-1}X)^{-1/2}$ we have

$$g_1 = |V_3V_4^{-1}|^{1/p}, \tag{3.6}$$

$$g_2 = \frac{1}{p}\operatorname{tr}(V_3V_4^{-1}), \tag{3.7}$$

$$\begin{aligned} g_3 &= |(X'R^{-1}X)^{1/2}(V_3 - V_4)(X'R^{-1}X)^{1/2}|^{1/p} \\ &= |X'R^{-1}X(V_3 - V_4)|^{1/p}, \end{aligned} \tag{3.8}$$

$$g_4 = \frac{1}{p}\operatorname{tr}[X'R^{-1}X(V_3 - V_4)], \tag{3.9}$$

$$g_5 = |V_3V_4^{-1} - I|^{1/p}, \tag{3.10}$$

where $R^{1/2}$ is a square root of R, $(X'R^{-1}X)^{1/2}$ is a square root of $X'R^{-1}X$, V_3 is given in (2.6) and V_4 is in (2.7).

Notice that the reciprocals of f_1 and g_1 are genuine relative efficiency measures in the sense of (2.8). In addition to the differences of the (asymptotic) variance matrices, f_3, f_4, g_3 and g_4 have an extra term, namely $(A'A)^{1/2}$ or $(X'R^{-1}X)^{1/2}$, which plays a role like c in (2.8). Each of f_1, f_2, f_3, f_4, f_5, g_1, g_2, g_3, g_4 and g_5 has an upper bound. If we find such an upper bound not far from the lower bound 1 for f_1 and f_2, and not far from 0 for f_3, f_4 and f_5, we can accept $\hat{\theta}_A$ as an alternative to $\hat{\theta}_{QS(V)}$ in the first case. In the second case, if a chosen R leads to an upper bound not far from the lower bound 1 for g_1 and g_2, and not far from 0 for g_3, g_4 and g_5, it is then reasonable for us to use this R in practice.

For the third case in studying Σ, we can insert $W = \Lambda^{-1/2}D(y)\Lambda^{-1/2}$ and $Z = \Lambda^{1/2}\Delta X(X'\Delta\Lambda\Delta X)^{-1/2}$ to have a set of forms similar to g_1 through g_5. For the fourth, we can consider $W = \Omega$ and $Z = X(X'X)^{-1/2}$, and then rewrite d_1 through d_5 in terms of V_5, V_6 and $(X'X)^{1/2}$.

4. Upper bounds

We now present the upper bounds of the comparison measures. Without loss of generality, we assume $n \geq 2p$. We have the following inequalities:

$$1 \leq |Z'WZZ'W^{-1}Z| \leq \prod_{i=1}^{p} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}}, \tag{4.1}$$

$$p \leq tr(Z'WZZ'W^{-1}Z) \leq \sum_{i=1}^{p} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}}, \tag{4.2}$$

$$0 \leq |Z'WZ - (Z'W^{-1}Z)^{-1}| \leq \max_{(s,t)} \prod_{i=1}^{p} (\lambda_{s(i)}^{1/2} - \lambda_{t(i)}^{1/2})^2, \tag{4.3}$$

$$(4.4)\qquad 0 \le \operatorname{tr}[Z'WZ - (Z'W^{-1}Z)^{-1}] \le \sum_{i=1}^{p}(\lambda_i^{1/2} - \lambda_{n-i+1}^{1/2})^2,$$

$$(4.5)\qquad 0 \le |Z'WZZ'W^{-1}Z - I| \le \max_{(s,t)} \prod_{i=1}^{p} \frac{(\lambda_{s(i)} - \lambda_{t(i)})^2}{4\lambda_{s(i)}\lambda_{t(i)}},$$

where $W > 0$ is an $n \times n$ matrix, Z is an $n \times p$ matrix satisfying $Z'Z = I$, $\lambda_1 \ge \cdots \ge \lambda_n$ are the eigenvalues of W, and the maxima in (4.3) and (4.5) are each taken over all possible partial matchings (s,t) of $(1,2,\ldots,n)$ into p pairs with the ith pair being denoted $\big(s(i),t(i)\big)$.

The right-hand inequality in (4.1) is the well-known Bloomfield–Watson–Knott inequality established by Bloomfield and Watson (1975) and Knott (1975). For (4.2), see, e.g., Rao and Rao (1998). The upper bound in (4.3) was first established in Drury et al. (2002) improving an early version in a preprint of Liu and King (2002). The right-hand inequality in (4.4) is due to Rao (1985). The result (4.5) can be established in a similar way to the derivation in Liu and King (2002). By using (4.1) through (4.5), we obtain the following results for the "normalized" measures in the four cases each with a $W > 0$ which is given in Section 3:

$$(4.6)\qquad 1 \le d_1 \le \left[\prod_{i=1}^{p} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}}\right]^{1/p},$$

$$(4.7)\qquad 1 \le d_2 \le \frac{1}{p}\sum_{i=1}^{p} \frac{(\lambda_i + \lambda_{n-i+1})^2}{4\lambda_i\lambda_{n-i+1}},$$

$$(4.8)\qquad 0 \le d_3 \le \left[\max_{(s,t)} \prod_{i=1}^{p}(\lambda_{s(i)}^{1/2} - \lambda_{t(i)}^{1/2})^2\right]^{1/p},$$

$$(4.9)\qquad 0 \le d_4 \le \frac{1}{p}\sum_{i=1}^{p}(\lambda_i^{1/2} - \lambda_{n-i+1}^{1/2})^2,$$

$$(4.10)\qquad 0 \le d_5 \le \left[\max_{(s,t)} \prod_{i=1}^{p} \frac{(\lambda_{s(i)} - \lambda_{t(i)})^2}{4\lambda_{s(i)}\lambda_{t(i)}}\right]^{1/p}.$$

Clearly, more (distinct) eigenvalues in (4.6) though (4.10) than in, e.g., (2.9) are involved. To make efficiency comparisons in the first two cases, what we do is to examine the upper bounds of f_1, f_2, f_3, f_4 and f_5, or g_1, g_2, g_3, g_4 and g_5, in addition to the one of E. For Case 1 we define $\mathrm{E} = \mathrm{E}(c) = c'V_1c/c'V_2c$, with c being a $p \times 1$ vector, and have

$$1 \le \mathrm{E} \le \frac{(\lambda_1 + \lambda_n)^2}{4\lambda_1\lambda_n},$$

with $\lambda_1 \ge \cdots \ge \lambda_n$ being the eigenvalues of F.

Note that (4.6) for the third case is equivalent to (2.12) given by Liu and Neudecker (1997), though (4.7) is better than (2.13). The upper bound in (2.13) containing the same information as (2.9) depends on only two eigenvalues, but (4.7) does involve more information. In Case 3, (4.8) through (4.10) are useful and complementary.

In Case 4, we note that d in (2.20) is the spectral norm of the difference between the variance matrices of $\hat{\beta}_0$ and $\hat{\beta}^*$. Actually we can improve (2.20) to have

$$d \leq \mu_p^{-1}(\lambda_1^{1/2} - \lambda_n^{1/2})^2. \tag{4.11}$$

This is deduced as follows:

$$\begin{aligned} V_5 - V_6 &= (X'X)^{-1}X'\Omega X(X'X)^{-1} - (X'\Omega^{-1}X)^{-1} \\ &= (X'X)^{-1/2}[Z'\Omega Z - (Z'\Omega^{-1}Z)^{-1}](X'X)^{-1/2} \\ &\leq (X'X)^{-1/2}[(\lambda_1^{1/2} - \lambda_n^{1/2})^2 I_p](X'X)^{-1/2} \\ &= (\lambda_1^{1/2} - \lambda_n^{1/2})^2(X'X)^{-1} \\ &\leq \mu_p^{-1}(\lambda_1^{1/2} - \lambda_n^{1/2})^2 I_p, \end{aligned}$$

where $Z = X(X'X)^{-1/2}$, and we use in the third step a matrix Katorovich inequality in, e.g., Liu and King (2002). This upper bound is sharper than the one in (2.20), as $\lambda_1^{1/2} - \lambda_n^{1/2} < \lambda_1^{1/2} + \lambda_n^{1/2}$.

The five upper bounds can be presented in the same forms as in (4.6) through (4.10) with $\lambda_1 \geq \cdots \geq \lambda_n$, which are the eigenvalues of Ω. We see that the upper bounds of d_3 and d_4 involving more than two eigenvalues are better than those involving only two in the key terms $\lambda_1 - \lambda_n$ in (2.20) and $(\lambda_1^{1/2} - \lambda_n^{1/2})^2$ in (4.11).

5. Numerical examples

We study two examples and carry out numerical calculations to get an impression of the upper bounds, which are dependent on our generic W.

Example 1. Consider Case 2. As a candidate for $R = \mathrm{diag}(R_1, \ldots, R_K)$ in (2.6) consider

$$R_k = R_k(\rho) = (1-\rho)I_k + \rho J_k, \tag{5.1}$$

where I_k is an $n_k \times n_k$ identity matrix, J_k is a matrix of ones, ρ is an unknown parameter and n_k is the size of the kth cluster ($k = 1, \ldots, K$; $n_1 + \cdots + n_K = n$). This is the working correlation parameterized by ρ (which must be estimated); see (2) in Balemi and Lee (1999). If ρ becomes

ρ_0, then R_k becomes R_{k0} and R becomes R_0. We can find the n eigenvalues of $R^{-1}R_0$ by using (5.1) rewritten as

$$R_k = (1-\rho)M_k + (1+(n_k-1)\rho)J_k/n_k,$$

where $M_k = I_k - J_k/n_k$ and using $M_kM_k = M_k$, $J_kJ_k = n_kJ_k$ and $M_kJ_k = 0$. The n eigenvalues of $R^{-1}R_0$ consist of K groups with the k-th group being $n_k - 1$ eigenvalues equal to

$$\mu_m = \frac{1-\rho_0}{1-\rho} \tag{5.2}$$

and one eigenvalue equal to

$$\mu_s = \frac{1+(n_k-1)\rho_0}{1+(n_k-1)\rho}. \tag{5.3}$$

If $\rho = \rho_0$, then $\mu_m = \mu_s = 1$. If $\rho < \rho_0$, then $\mu_m < 1 < \mu_s$. If $\rho > \rho_0$, then $\mu_m > 1 > \mu_s$. Based on (5.2) and (5.3), we are able to establish the corresponding upper bounds, which are each a function of ρ.

To illustrate, we now give two figures, in which we draw plots for the upper bounds against ρ. We assume $n = 90$, $K = 30$, $n_1 = \cdots = n_{30} = 3$, $p = 2$ and run ρ from 0 to 0.95. In Figure 1, we set $\rho_0 = 0.1$, and then $W_1 = W_1(\rho) = R^{-1/2}(\rho)R_0(0.1)R^{-1/2}(\rho)$ which has the same eigenvalues as $R^{-1}(\rho)R_0(0.1)$. In Figure 2, we set $\rho_0 = 0.4$, and then $W_2 = W_2(\rho) = R^{-1/2}(\rho)R_0(0.4)R^{-1/2}(\rho)$. In each case, $R^{-1}R_0$ has 90 eigenvalues: 60 of μ_m and 30 of μ_s. We choose μ_m once and μ_s once to calculate the upper bound of E, and choose μ_m twice and μ_s twice to calculate the upper bounds of g_1, g_2, g_3, g_4 and g_5; the upper bounds of E, g_1 and g_2 are the same, so are the upper bounds of g_3 and g_4. We observe that all the plots reach a minimum at $\rho = \rho_0$ (which corresponds to the lower bounds of E, g_1, g_2, g_3, g_4 and g_5). The performances of the plots also vary. In both figures, g_3, g_4 and g_5 are much more sensitive than E, g_1 and g_2, and g_5 seems sharper and better than g_3 and g_4. The trends of E, g_1, g_2, g_3, g_4 and g_5 in Figure 2 seem similar to those in Figure 1, though the differences among E, g_1, g_2, g_3 and g_4 is more distinguishable in Figure 2 for $\rho_0 = 0.4$. The upper bounds stay (relatively) close to the lower bounds even when $|\rho - \rho_0|$ is close to 0.4, especially the one for g_5 in Figure 2. Thus we can be fairly certain of choosing a good R, close enough to R_0, with small loss of efficiency. However, the difference between using R and R_0 may still be small for a large range of ρ.

Example 2. Noting that F is diagonal in Case 1, we arbitrarily choose $W_3 = W_3(\rho)$ to be a diagonal matrix. We choose $W_4 = W_4(\rho)$ to be of a moving average MA(1) variance structure as studied by Fomby, Hill and Johnson

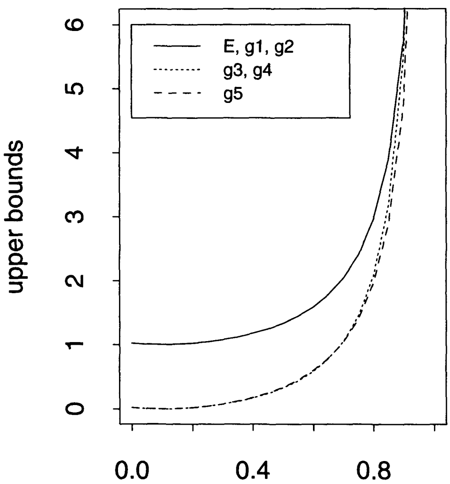

Figure 1. ρ for W_1

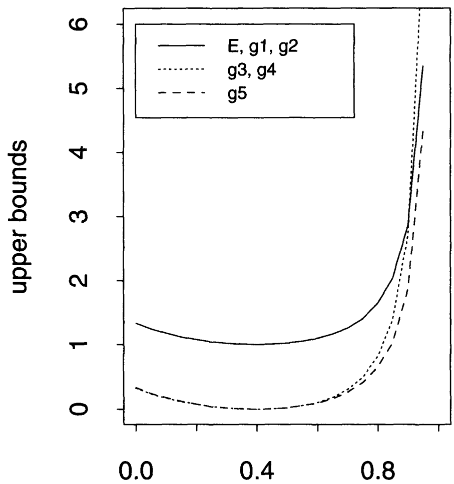

Figure 2. ρ for W_2

(1984, Section 10.7.2). Sometimes dependent variables are unavoidable. The matrices W_3 and W_4 are both $n \times n$ and defined as follows:

$$W_3 = \begin{pmatrix} 2n\rho+1 & 0 & 0 & \cdots & 0 \\ 0 & n\rho+1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix},$$

$$W_4 = \begin{pmatrix} 1+\rho^2 & \rho & 0 & \cdots & 0 \\ \rho & 1+\rho^2 & \rho & \cdots & 0 \\ 0 & \rho & 1+\rho^2 & \ddots & \vdots \\ \vdots & \vdots & \ddots & \ddots & \rho \\ 0 & 0 & \cdots & \rho & 1+\rho^2 \end{pmatrix},$$

where $n = 15$ and $0 \leq \rho \leq 0.95$. We find the eigenvalues of W_3 and W_4, and the corresponding upper bounds which are each a function of ρ. We draw plots for the upper bounds against ρ in Figures 3 and 4. The upper bounds of g_1 and g_2 are close and sharper than E, and g_1 seems the best. The upper bounds of g_3 and g_4 are close, and even indistinguishable in Figure 4, though g_3 is slightly sharper than g_4. In Figure 3 g_5 is sharper than g_3 and g_4, but in Figure 4 g_3 and g_4 are sharper. In both figures, g_5 has a similar trend to those of E, g_1 and g_2.

Among all the plots, the ones in Figure 3 seem most distinguishable, perhaps because the eigenvalues there are most distinct. However, in all the

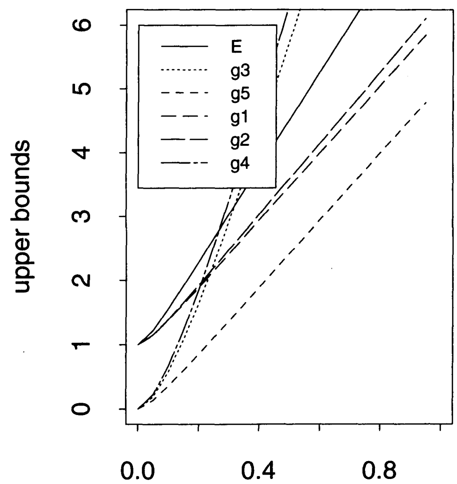

Figure 3. ρ for W_3

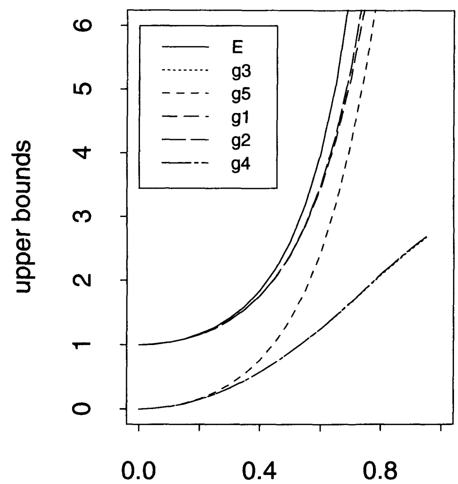

Figure 4. ρ for W_4

four figures each of (the function forms of) the upper bounds plays its own important role. Additional upper bounds may provide further insights. We could rescale all the measures by using E -1, $g_1 - 1$, $g_2 - 1$, g_3, g_4 and g_5 so that identity corresponds to a value of zero, but we have chosen no to do so in order to avoid obscuring the distinctions between the plots.

6. Remarks

The "normalized" measures d_1 through d_5 in Section 3 and their upper bounds in (4.6)–(4.10) are advocated. The measures based on the determinants and traces are useful, and their upper bounds involving more information are better than the one for E relying only on the largest and smallest eigenvalues of a positive definite matrix. The behaviour of the upper bounds of the measures depends on specific cases; no upper bound can always be optimal. This is clearly seen through the examples in Section 5, although even Example 1 is for a specific case in which (nondistinct) multiple eigenvalues are involved. Of course, the values of the measures themselves are not known. Further investigation in studying the measures themselves and then making efficiency comparisons is required.

REFERENCES

Anh, V.V. (1988) Nonlinear linear squares and maximum likelihood estimation of a heteroscedastic regression model. *Stochastic Process. Appl.* 29, 317–333.

Balemi, A. and Lee, A. (1999) Some properties of the Liang-Zeger method applied to clustered binary regression. *Aust. N. Z. J. Stat.* 41, 43–58.

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

Drury, S.W., Liu, S., Lu, C.-Y., Puntanen, S. and Styan, G.P.H. (2002) Some comments on several matrix inequalities with applications to canonical correlations: Historical background and recent developments. *Sankhyā Ser.* A 64, 453–507. Associated with "An International Conference in Honor of Professor C.R. Rao on the Occasion of his 80th Birthday, Statistics: Reflections on the Past and Visions for the Future, The University of Texas at san Antonio (March 2000)".

Fomby, T.B., Hill, R.C. and Johnson, S.R. (1984) *Advanced Econometric Methods.* Springer-Verlag, New York.

Hannan, E.J. (1970) *Multiple Time Series.* Wiley, New York.

Heyde, C.C. (1989) On efficiency for quasi-likelihood and composite quasi-likelihood methods. In *Statistical Data Analysis and Inference* (Y. Dodge, ed.). North-Holland, Amsterdam, pp. 209–213.

Heyde, C.C. (1997) *Quasi-likelihood and Its Application. A General Approach to Optimal Parameter Estimation.* Springer-Verlag, New York.

Heyde, C.C. and Lin, Y.-X. (1992) On quasi-likelihood methods and estimation for branching processes and heteroscedastic regression models. *Austral. J. Statist.* 34, 199–206.

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

Liang, K.-Y. and Zeger, S.L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika* 73, 13–22.

Liu, S. and King, M.L. (2002) Two Kantorovich-type inequalities and efficiency comparisons between the OLSE and BLUE. *J. Inequal. Appl.* 7, 169–177.

Liu, S. and Neudecker, H. (1997) A survey of Cauchy–Schwarz and Kantorovich-type matrix inequalities, *Statist. Papers* 40, 55–73.

Liu, S. and Neudecker, H. (1999) Kantorovich inequalities and efficiency comparisons for several classes of estimators in linear models. *Statist. Neerlandica* 51, 345–355.

Marshall, A.W. and Olkin, I. (1990) Matrix versions of Cauchy and Kantorovich inequalities, *Aequationes Math.* 40, 89-93.

McCullagh, P. and Nelder, J.A. (1989) *Generalized Linear Models*, 2nd ed. Chapman and Hall, London.

Mukhopadhyay, P. and Schwabe, R. (1998) On the performance of the ordinary least squares method under an error component model. *Metrika* 47, 215–226.

Rao, C.R. (1985) The inefficiency of least squares: extensions of the Kantorovich-type inequality. *Linear Algebra Appl.* 70, 249–255.

Rao, C.R. and Rao, M.B. (1998) *Matrix Algebra and Its Applications to Statistics and Econometrics.* World Scientific, Singapore.

Scott, A.J. and Holt, D. (1982) The effect of two-stage sampling on ordinary least squares methods. *J. Amer. Statist. Assoc.* 77, 848–854.

Sutradhar, B.C. and Das, K. (1999) On the efficiency of regression estimators in generalised linear models for longitudinal data. *Biometrika* 86, 459–465.

Wang, S.G., Chow, S.C. and Tse, S.K. (1994) On ordinary least-squares methods for sample surveys. *Statist. Probab. Lett.*, 20, 173–182.

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

S. Liu
Centre for Mathematics and its Applications
Mathematical Sciences Institute
Australian National University
Canberra, ACT 0200
Australia
lius@maths.anu.edu.au

C.C. Heyde
Centre for Mathematics and its Applications
Mathematical Sciences Institute
Australian National University
Canberra, ACT 0200
Australia
chris@maths.anu.edu.au