Ann. Inst. Statist. Math.
Vol. 52, No. 2, 332–342 (2000)

OPERATIONAL VARIANTS OF THE MINIMUM MEAN SQUARED ERROR ESTIMATOR IN LINEAR REGRESSION MODELS WITH NON-SPHERICAL DISTURBANCES

ALAN T. K. WAN[1] AND ANOOP CHATURVEDI[2]

[1] *Department of Management Sciences, City University of Hong Kong, Tat Chee Avenue, Kowloon, Hong Kong, China*
[2] *Department of Mathematics and Statistics, University of Allahabad, Allahabad-211002, India*

(Received July 13, 1998; revised October 26, 1998)

Abstract. There is a good deal of literature that investigates the properties of various operational variants of Theil's (1971, *Principles of Econometrics*, Wiley, New York) minimum mean squared error estimator. It is interesting that virtually all of the existing analysis to date is based on the premise that the model's disturbances are i.i.d., an assumption which is not satisfied in many practical situations. In this paper, we consider a model with non-spherical errors and derive the asymptotic distribution, bias and mean squared error of a general class of feasible minimum mean squared error estimators. A Monte-Carlo experiment is conducted to examine the performance of this class of estimators in finite samples.

Key words and phrases: Asymptotic expansion, quadratic loss, minimum mean squared error, risk, Stein-rule.

1. Introduction

Much of the literature of biased estimation in regression analysis is concerned with the search of estimators for improvements in mean squared error (MSE) terms over the unbiased least squares rule. See Mayer and Willke (1973), Draper and Van Nostrand (1979) and Vinod (1978) for surveys of the relevant literature. Within the context of the linear regression model, Theil (1971) exhibits an estimator which is shown to have the minimum MSE property among the class of all linear homogenous estimator for the regression coefficients. The problem with this estimator is that it is non-operational as it depends upon the unknown regression coefficients and disturbance variance. Farebrother (1975) consequently proposes a feasible minimum mean squared error estimator (FMMSEE) which is obtained through replacing the unknown parameters with their estimates from ordinary least squares (OLS). Several authors, including Vinod (1976) and Dwivedi and Srivastava (1978), have subsequently studied the FMMSEE's asymptotic and finite samples properties. Stahlecker and Trenkler (1985) suggest incorporating prior information and derive a minimum MSE heterogeneous estimator of the regression coefficients. Other authors, such as Liski *et al.* (1993) and Tracy and Srivastava (1994), have considered various other generalisations and extensions. Recently, there has been a resurgence of interest in the FMMSEE, as reflected in the work of Ohtani (1996*a*, 1996*b*). The former relates to the derivation of an expression of the exact finite sample risk of the FMMSEE under a quadratic loss function, whereas the latter considers an adjusted (for degrees of freedom) feasible minimum mean squared error estimator (AFMMSEE). A striking feature that emerges from these investigations is that the AFMMSEE can be a superior alternative to the Stein rule estimator (SRE) and Stein positive rule estimator

(SPRE) over a wide range of experimental settings. Other relevant work includes Ohtani (1997), where the exact MSE of the FMMSEE of individual coefficients is considered, and Ohtani (1999), which explores a pre-test strategy involving the AFMMSEE.

It is interesting to note that all the aforementioned studies invariably assume that the underlying model's disturbances are i.i.d., an assumption which is not satisfied in many practical situations. Some authors, such as Toutenburg *et al.* (1992) and Rao and Toutenburg (1995) discuss the (non-operational) minimum mean squared error linear estimator for models with non-spherical errors, but do not consider its operational variants. This gap is remedied in this article. In Section 2, we discuss a class of adaptive versions for the minimum mean squared error estimator for the case of non-spherical errors. Section 3 is devoted to the derivation of the asymptotic distributions of this class of estimators, along with the conditions of dominance of this class over the feasible generalized least squares estimator (FGLSE) using the risk under quadratic loss criterion. A result of Rothenberg (1984) on the asymptotic distribution of the FGLSE is nested as a special case in our results. By means of a Monte-Carlo experiment, Section 4 explores the performance of these estimators in finite samples. Some concluding remarks appear in the final section.

2. The model and estimators

To motivate discussion and establish notations, we consider the linear regression model, $y = X\beta + \varepsilon$, where y is a $n \times 1$ vector of observations on the dependent variable, X is a $n \times k$ matrix of observations on k independent variables, β is a $k \times 1$ vector of regression coefficients and ε is a $n \times 1$ vector of disturbances assumed to follow the Normal distribution $N(0, \sigma^2 W)$. We assume that the elements of the covariance matrix W are functions of a $q \times 1$ parameter vector θ belonging to an open subset of the q-dimensional Euclidean space.

It has been shown (see, for example, Rao and Toutenburg (1995)) that the estimator, $\beta_M = [\beta' X' W^{-1} y/(\sigma^2 + \beta' X' W^{-1} X\beta)]\beta$, is the minimum mean squared error linear estimator of β. It is obvious that β_M is not a true estimator as it depends upon the unknown parameters β, σ^2 and θ. In order to make β_M operational, we replace β and σ^2 by their estimates from feasible generalised least squares and θ by a consistent estimate. This gives rise to the following feasible minimum mean squared error estimator as an adaptive version of β_M:

$$\hat\beta_M = [\hat\beta' X' \hat W^{-1} X\hat\beta/(\tau/\nu + \hat\beta' X' \hat W^{-1} X\hat\beta)]\hat\beta, \tag{2.1}$$

where $\hat W$ is obtained by replacing θ by its consistent estimator, say, $\hat\theta$, in W, $\hat\beta = (X'\hat W^{-1}X)^{-1}X'\hat W^{-1}y$ is the FGLSE of β, $\tau = (y - X\hat\beta)'\hat W^{-1}(y - X\hat\beta)$ and $\nu = n - k$. A slight modification of the FMMSEE is to adjust $\hat\beta' X' \hat W^{-1} X\hat\beta$ in (2.1) by its degrees of freedom (Ohtani (1996*b*)), which leads to,

$$\hat\beta_{AM} = [(\hat\beta' X' \hat W^{-1} X\hat\beta/k)/(\tau/\nu + \hat\beta' X' \hat W^{-1} X\hat\beta/k)]\hat\beta,$$

the adjusted feasible minimum mean squared error estimator of β. Further, a general form of $\hat\beta_M$ and $\hat\beta_{AM}$ may be written as,

$$\hat\beta_C = [\hat\beta' X' \hat W^{-1} X\hat\beta/(\tau d/\nu + \hat\beta' X' \hat W^{-1} X\hat\beta)]\hat\beta. \tag{2.2}$$

For $d = 1$, $\hat\beta_C$ reduces to the FMMSEE and for $d = k$, it becomes $\hat\beta_{AM}$. Also, the FGLSE results when $d = 0$. Both $\hat\beta_M$ and $\hat\beta_{AM}$ are consistent estimators of β, but neither estimator has any minimum mean squared error property.

It is worth noting that for the case $E(\varepsilon\varepsilon') = I$, the FMMSEE satisfies Baranchik's (1970) condition and therefore dominates the OLS estimator in terms of risk under quadratic loss when $k \geq 3$. In the same context, Ohtani (1996a, 1996b) provide numerical evidence suggesting that even if $k < 3$, the FMMSEE is risk superior under a quadratic loss structure to the OLS estimator in a wide region of the non-centrality parameter $\lambda = \beta' X' X \beta / \sigma^2$; and that the AFMMSEE is risk preferred to both the FMMSEE and the Stein positive rule estimator over a wide range of λ values.

In the next section, we consider the properties of $\hat{\beta}_M$ and $\hat{\beta}_{AM}$ under the more general framework of non-spherical errors, which is heretofore unexamined.

3. Asymptotic distribution, bias and mean squared errors

In this section we first derive the asymptotic distributions of the proposed estimators. For convenience purposes, we write $W = \Omega^{-1}$ and $\hat{W} = \hat{\Omega}^{-1}$ so that (2.2) can be written as,

$$\hat{\beta}_M = [\hat{\beta}' X' \hat{\Omega} X \hat{\beta} / (\tau d/\nu + \hat{\beta}' X' \hat{\Omega} X \hat{\beta})] \hat{\beta}.$$

Now, let θ_j be the j-th element of θ,

$$\begin{aligned}
&\Omega_j = \partial\Omega/\partial\theta_j, \Omega_{jk} = \partial\Omega/\partial\theta_{jk}, \ldots \\
&A = X'\Omega X/n, A_j = X'\Omega_j X/n, A_{jk} = X'\Omega_{jk} X/n, \ldots \\
&\alpha = X'\Omega\varepsilon/\sqrt{n}, \alpha_j = X'\Omega_j\varepsilon/\sqrt{n}, \alpha_{jk} = X'\Omega_{jk}\varepsilon/\sqrt{n}, \ldots
\end{aligned}$$

and the set of all matrices having the same number of indices be denoted by that matrix subscripted in bracket by that number. For instance, $A_{(2)}$ denotes the set of matrices A_{jk}, $j, k = 1, 2, 3, \ldots, q$. Furthermore, we require the following regularity conditions for the validity of the asymptotic expansion for the distribution of $\hat{\beta}_C$ (see Rothenberg (1984) or Chaturvedi and Shukla (1990)):

C1. The matrix A converges to a finite matrix as $n \to \infty$;

C2. Each matrix in the set $A_{(1)}, \ldots, A_{(5)}$ and covariance matrix of each random vector $\alpha_{(1)}, \ldots, \alpha_{(5)}$ converges to a finite matrix as $n \to \infty$;

C3. For all matrices Γ in $\Omega_{(6)}$, $n^{-1} X' \Gamma^2 X$ is bounded as $n \to \infty$;

C4. The estimator $\hat{\theta}$ of θ has a stochastic expansion of the form,

$$\sqrt{n}(\hat{\theta} - \theta) = \delta = e + O_p(n^{-1})$$

where e follows a normal distribution with mean vector of order $O(n^{-1/2})$ and covariance matrix $\Lambda + O(n^{-1})$. Further, the third order cumulants of $(\alpha_{11}, \ldots, \alpha_{kk})$ are of order $O(n^{-1/2})$ and higher cumulants are of order $O(n^{-1})$.

Now, we denote,

$$\begin{aligned}
&P_j = (X'\Omega_j - A_j A^{-1} X'\Omega)/\sqrt{n}, \\
&P_{jk} = (X'\Omega_{jk} - 2A_j A^{-1} X'\Omega_k + 2A_j A^{-1} A_k A^{-1} X'\Omega - A_{jk} A^{-1} X'\Omega)/(2\sqrt{n}), \\
&\phi = \beta' A \beta, \quad \phi_j = \beta' A_j \beta, \quad \phi_{jk} = \beta' A_{jk} \beta, \\
&\varsigma = -d\sigma\beta/\phi\sqrt{n}, \quad \text{and} \\
&\Delta = A^{-1} + A^{-1} \left(\sum_{j,k}^{q} P_j \Omega^{-1} P_k' \lambda_{jk} \right) A^{-1}/n - 2d\sigma^2 (A^{-1} - 2\beta\beta'/\phi)/n\phi,
\end{aligned}$$

where λ_{jk} is the (j,k)-th element of Λ.

THEOREM 3.1. *Given the regularity conditions* C1–C4, *the asymptotic distribution of* $r = \sqrt{n}(\hat{\beta}_C - \beta)/\sigma$, *up to order*, $O_P(n^{-1})$, *is* $N(\varsigma, \Delta)$ *as* $n \to \infty$.

PROOF. Following Chaturvedi and Shukla (1990), up to order $O_P(n^{-1})$, we can write,

$$\sqrt{n}(\hat{\beta} - \beta)/\sigma = \eta_o + \eta_{-1/2} + \eta_{-1} + O_P(n^{-3/2})$$

where $\eta_o = A^{-1}\alpha/\sigma$,

$$\eta_{-1/2} = \sum_j^q A^{-1}P_j\varepsilon\delta_j/(\sigma\sqrt{n}), \qquad \eta_{-1} = \sum_{j,k}^q A^{-1}P_{jk}\varepsilon\delta_j\delta_k/(\sigma n)$$

and δ_i is the i-th element of δ. Further, we have, $\tau/\nu = \sigma^2 + O_P(n^{-1})$ and

$$\begin{aligned} 1/(d\tau/\nu + \hat{\beta}'X'\hat{\Omega}X\hat{\beta}) &= \left(d\sigma^2/n + \phi + 2\sigma\beta'A\eta_o/\sqrt{n} + \sum_j^q \phi_j\delta_j/\sqrt{n}\right)^{-1} \Big/ n \\ &= \left(1 - 2\sigma\beta'A\eta_o/\phi\sqrt{n} - \sum_j^q \phi_j\delta_j/\phi\sqrt{n}\right) \Big/ n\phi + O_P(n^{-2}). \end{aligned}$$

Therefore, up to order $O_P(n^{-1})$, we have,

$$\begin{aligned} (3.1) \qquad r &= \sqrt{n}[\hat{\beta} - \beta - (d\tau/\nu)/((d\tau/\nu) + \hat{\beta}'X'\hat{\Omega}X\hat{\beta})]\hat{\beta}/\sigma \\ &= \eta_o + \eta_{-1/2} + \eta_{-1} - d\sigma\left(1 - 2\sigma\beta'A\eta_o/\phi\sqrt{n} - \sum_j^q \phi_j\delta_j/\phi\sqrt{n}\right) \\ &\quad \cdot (\beta + \eta_o\sigma/\sqrt{n})/\phi\sqrt{n} \\ &= \eta_o + \eta_{-1/2} + \eta_{-1} \\ &\quad - d\sigma\left(\beta + (\mathrm{I} - 2\beta\beta'A/\phi)\eta_o\sigma/\sqrt{n} - \sum_j^q \phi_j\delta_j\beta/\phi\sqrt{n}\right) \Big/ \phi\sqrt{n}. \end{aligned}$$

Utilizing (3.1), we obtain the cumulant generating function of r, up to order $O(n^{-1})$, as,

$$\begin{aligned} K(h) = -id\sigma h'\beta/\phi\sqrt{n} + \ln E\Bigg\{ e^{ih'\eta_o}\Bigg[1 + ih'\eta_{-1/2} + ih'\eta_{-1} - (h'\eta_{-1/2})^2/2 \\ - id\sigma^2\left(h'(\mathrm{I} - 2\beta\beta'A/\phi)\eta_o - \sum_j^q \phi_j\delta_j\beta/\sigma\phi\right) \Big/ \phi n\Bigg]\Bigg\}. \end{aligned}$$

Now,

$$E(e^{ih'\eta_o}) = e^{-h'A^{-1}h/2}, \qquad E(h'\eta_{-1/2}e^{ih'\eta_o}) = E(h'\eta_{-1}e^{ih'\eta_o}) = 0,$$

$$E((h'\eta_{-1/2})^2e^{ih'\eta_o}) = e^{-h'Ah/2}\sum_{j,k}^q h'A^{-1}P_j\Omega^{-1}P_k'A^{-1}h\lambda_{jk}/n,$$

$$E(\eta_o e^{ih'\eta_o}) = e^{-h'Ah/2}(iA^{-1}h), \qquad \text{and} \qquad E\left(e^{ih'\eta_o}\sum_j^q \phi_j\delta_j\right) = 0.$$

Hence, up to order $O(n^{-1})$,

$$K(h) = -id\sigma h'\beta/\phi\sqrt{n} - h'A^{-1}h/2 + \ln\left[1 - \sum_{j,k}^{q} h'A^{-1}P_j\Omega^{-1}P_k'A^{-1}h\lambda_{jk}/(2n) + d\sigma^2 h'(A^{-1} - 2\beta\beta'/\phi)h/\phi n\right]$$

$$= -id\sigma h'\beta/\phi\sqrt{n} - h'\left[A^{-1} + A^{-1}\sum_{j}^{q}(P_j\Omega^{-1}P_k'\lambda_{jk})A^{-1}/n - 2d\sigma^2(A^{-1} - 2\beta\beta'/\phi)/(\phi n)\right]h/2$$

$$= ih'\varsigma - h'\Delta h/2$$

which is the cumulant generating function of a normal distribution $N(\varsigma, \Delta)$. Hence Theorem 3.1 follows.

Remark 3.1. The asymptotic distribution of $\hat{\beta}$ given in Rothenberg ((1984), 814–819) is a special case of Theorem 1 by substituting $d = 0$. The equivalence of ours and Rothenberg's expressions can be proven by observing that for any $k \times 1$ vector c, the asymptotic distribution of $\sqrt{n}c'(\hat{\beta} - \beta)/\sigma$, up to order $O(n^{-1})$, is $N(0, c'\Phi_1 c)$, where

$$\Phi_1 = A^{-1} + A^{-1}\left(\sum_{j,k=1}^{q} P_j\Omega^{-1}P_k'\lambda_{jk}\right)A^{-1}/n.$$

COROLLARY 3.1. *Utilizing Theorem* 3.1, *it is straightforward to show, up to order* $O(n^{-1})$, *that the bias vectors of* $\hat{\beta}_C$, $\hat{\beta}$, $\hat{\beta}_M$ *and* $\hat{\beta}_{AM}$ *are,*

$$E(\hat{\beta}_C - \beta) = -\sigma^2 d\beta/\phi n, \qquad E(\hat{\beta} - \beta) = 0,$$
$$E(\hat{\beta}_M - \beta) = -\sigma^2\beta/\phi n, \qquad \text{and} \qquad E(\hat{\beta}_{AM} - \beta) = -\sigma^2 k\beta/\phi n$$

respectively.

COROLLARY 3.2. *Using Theorem* 3.1 *again, the MSE matrices of* $\hat{\beta}_C$, $\hat{\beta}$, $\hat{\beta}_M$ *and* $\hat{\beta}_{AM}$, *up to order* $O(n^{-2})$, *are given by,*

$$E(\hat{\beta}_C - \beta)(\hat{\beta}_C - \beta)' = \sigma^2\left\{A^{-1} + A^{-1}\left(\sum_{j,k=1}^{q} P_j\Omega^{-1}P_k'\lambda_{jk}\right)A^{-1}/n - d\sigma^2(2A^{-1} - (4+d)\beta\beta'/\phi)/(n\phi)\right\}\bigg/ n,$$

$$E(\hat{\beta} - \beta)(\hat{\beta} - \beta)' = \sigma^2\left\{A^{-1} + A^{-1}\left(\sum_{j,k=1}^{q} P_j\Omega^{-1}P_k'\lambda_{jk}\right)A^{-1}/n\right\}\bigg/ n$$

$$E(\hat{\beta}_M - \beta)(\hat{\beta}_M - \beta)' = \sigma^2 \left\{ A^{-1} + A^{-1} \left(\sum_{j,k=1}^{q} P_j \Omega^{-1} P_k' \lambda_{jk} \right) A^{-1} \Big/ n \right.$$
$$\left. - \sigma^2 (2A^{-1} - 5\beta\beta'/\phi)/(n\phi) \right\} \Big/ n \quad \textit{and}$$

$$E(\hat{\beta}_{AM} - \beta)(\hat{\beta}_{AM} - \beta)' = \sigma^2 \left\{ A^{-1} + A^{-1} \left(\sum_{j,k=1}^{q} P_j \Omega^{-1} P_k' \lambda_{jk} \right) A^{-1}/n \right.$$
$$\left. - k\sigma^2 (2A^{-1} - (4+k)\beta\beta'/\phi)/(n\phi) \right\} \Big/ n$$

respectively.

THEOREM 3.2. *Up to the order of our approximation, the estimators $\hat{\beta}$, $\hat{\beta}_M$ and $\hat{\beta}_{AM}$ do not strictly dominate one another with respect to the criterion of MSE matrix.*

PROOF. First, consider the dominance of $\hat{\beta}_M$ over $\hat{\beta}$. From Corollary 3.2, this requires the matrix $(2A^{-1} - 5\beta\beta/\phi)$ to be positive definite. This cannot be true as can be seen by an application of Theorem A.57 of Rao and Toutenburg ((1995), p. 303), according to which a necessary and sufficient condition for $(2A^{-1} - 5\beta\beta/\phi)$ to be positive definite is $5\beta' A\beta/(2\phi) < 1$ or $5/2 < 1$, which can never hold. This means that $\hat{\beta}_M$ cannot dominate $\hat{\beta}$ with respect to the criterion of MSE matrix. Consider the converse of it, i.e., dominance of $\hat{\beta}$ over $\hat{\beta}_M$. This requires the matrix $(5\beta\beta'/\phi - 2A^{-1})$ to be positive definite. This also cannot be true by virtue of Theorem A.59 of Rao and Toutenburg ((1995), p. 304). So, in other words, $\hat{\beta}_M$ neither dominates $\hat{\beta}$ nor is dominated by $\hat{\beta}$ with respect to the MSE matrix criterion. It can be established in the same fashion that similar is the case between $\hat{\beta}_{AM}$ and $\hat{\beta}$, and between $\hat{\beta}_M$ and $\hat{\beta}_{AM}$.

It is thus seen that of the estimators considered, no one strictly dominates any of the others in terms of MSE matrix. Such is, however, not the case if we consider the following weighted quadratic loss function,

$$L(\hat{\vartheta}\vartheta) = (\hat{\vartheta} - \vartheta)' Q (\hat{\vartheta} - \vartheta) \tag{3.2}$$

as a (weaker) criterion to appraise estimators' performance, where $\hat{\vartheta}$ is any estimator of ϑ and Q is a $k \times k$, positive definite, symmetric weighted matrix of the loss function. In particular, if we choose $Q = A$, then we have the following *sufficient* conditions:

THEOREM 3.3. *Under the quadratic loss function* (3.2) *with $Q = A$, up to order $O(n^{-2})$, we have,*

i) *$\hat{\beta}_M$ dominates $\hat{\beta}$ whenever $k \geq 5/2$ or equivalently $k \geq 3$;*
ii) *$\hat{\beta}_{AM}$ dominates $\hat{\beta}$ whenever $k \geq 4$;*
iii) *$\hat{\beta}_{AM}$ dominates $\hat{\beta}_M$ whenever $k \geq 5$.*

PROOF. From the MSE expressions of $\hat{\beta}$ and $\hat{\beta}_M$, it is readily shown, up to order $O(n^{-2})$, that,

$$E(\hat{\beta} - \beta)' Q (\hat{\beta} - \beta) - E(\hat{\beta}_M - \beta)' Q (\hat{\beta} - \beta) = \sigma^4 [2\,\mathrm{tr}(A^{-1}Q) - 5\beta' Q\beta/\phi]/(n^2\phi).$$

Therefore, the FMMSEE dominates the FGLSE as long as, $w = [\mathrm{tr}(A^{-1}Q/\mu(A^{-1}Q)] \geq 5/2$, where μ denotes the maximum characteristic root of the matrix $A^{-1}Q$. Similarly, the difference between the risks of the estimators $\hat{\beta}$ and $\hat{\beta}_{AM}$, to the order of our approximation, is given by,

$$E(\hat{\beta} - \beta)'Q(\hat{\beta} - \beta) - E(\hat{\beta}_{AM} - \beta)'Q(\hat{\beta}_{AM} - \beta)$$
$$= \sigma^4 k[2\,\mathrm{tr}(A^{-1}Q) - (4 + k)\beta'Q\beta/\phi]/(n^2\phi)$$

which is non-negative whenever $w \geq 2 + k/2$. Further, comparing the risks of $\hat{\beta}_M$ and $\hat{\beta}_{AM}$, we observe, up to order $O(n^{-2})$, that,

$$E(\hat{\beta}_M - \beta)'Q(\hat{\beta}_M - \beta) - E(\hat{\beta}_{AM} - \beta)'Q(\hat{\beta}_{AM} - \beta)$$
$$= \sigma^4(k - 1)[2\,\mathrm{tr}(A^{-1}Q) - (5 + k)\beta'Q\beta/\phi]/(n^2\phi),$$

so that a sufficient condition for $\hat{\beta}_{AM}$ to dominate $\hat{\beta}_M$ is $w \geq (5 + k)/2$. If $Q = A$, then $w = k$ and Theorem 3.3 follows.

Notwithstanding the results being large sample approximations, Theorem 3.3 offers a simple prescription regarding the choice of estimators to be followed in practice, in a form that overcomes the unobservability of the model's parameters. The following simulation experiment sheds further light on the finite sample performance of $\hat{\beta}_M$ and $\hat{\beta}_{AM}$ in comparison with various other estimators including the SRE and PSRE.

4. Monte-Carlo results

Our experimental design is based on the following model with AR(1) error terms:

$$y = X\beta + \varepsilon, \qquad \varepsilon = \rho\varepsilon_{-1} + u,$$

where u is a random vector with elements $u_t \sim IN(0, \sigma_u^2)$, $t = 1, 2, \ldots, 20$. In addition, we consider the following parameter values:

$$\sigma_u^2 = 1.0, \qquad k = 2, 4, 6, \qquad \rho = 0.0, \pm 0.4, \pm 0.8, \qquad \lambda = \text{various values}.$$

Furthermore, the design matrix X is chosen such that $XX' = I$. Estimators' performance is compared on the basis of the loss function given in (3.2) with $Q = A$. The matrix $\hat{\Omega}$ is constructed using the Prais-Winsten (1954) transformation. Each part of the experiment is based on 5000 repetitions, and our computations are undertaken with the SHAZAM econometric package version 8. For the experiments with $k > 2$, we also include the Stein-rule estimator discussed in Chaturvedi and Shukla (1990) and its positive-rule counterpart, with $k - 2$ chosen as the (optimal) value of the Stein-rule's shrinkage parameter (see Chaturvedi and Shukla (1990) for details on the choice of value of this parameter). To the best of our knowledge, there are no previous results on the empirical performance of the SRE and SPRE in models with non-spherical disturbances. A selection of the results being representative of the general patterns are illustrated in Figs. 1 through 6, with the risk of the FGLSE scaled to 1 in each case for ease of comparisons.

The discussion that follows is based on a full set of results available upon request. Considering first the case of $k = 4$ (Figs. 1 and 2), it is found that the AFMMSEE is better than all the other estimators in almost all parts of the parameter space, except

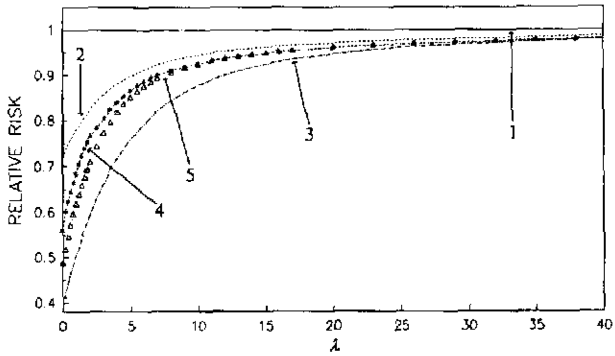

Fig. 1. Relative risks of estimators for $n = 20$, $k = 4$ and $\rho = -0.4$. Plotted for 1 = FGLSE, 2 = FMMSEE, 3 = AFMMSEE, 4 = SRE and 5 = SPRE.

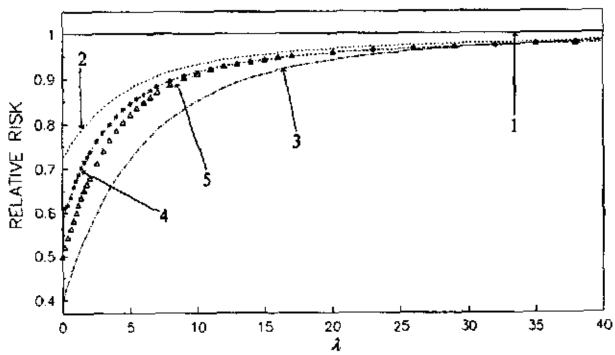

Fig. 2. Relative risks of estimators for $n = 20$, $k = 4$ and $\rho = 0.0$. Plotted for 1 = FGLSE, 2 = FMMSEE, 3 = AFMMSEE, 4 = SRE and 5 = SPRE.

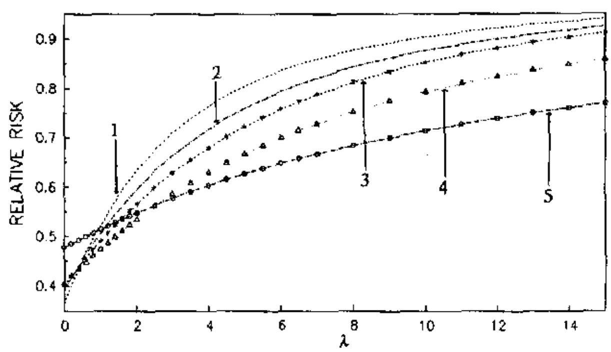

Fig. 3. Relative riskis of the AFMMSEE for $n = 20$ and $k = 4$. The curves labeled 1, 2, 3, 4 and 5 are for $\rho = -0.8$, -0.4, 0.0, 0.4 and 0.8 respectively.

when λ is relatively large (> 100), where the SRE and SPRE can have a slight advantage over the AFMMSEE. In general, the SPRE risk is never greater than the SRE risk, and the FGLSE usually has the largest risk followed by the FMMSEE. For all the cases that we have considered, the FMMSEE risk is always greater than the risks of the SRE, SPRE and AFMMSEE. At $\lambda = 0$, the risk reductions of the AFMMSEE and FMMSEE

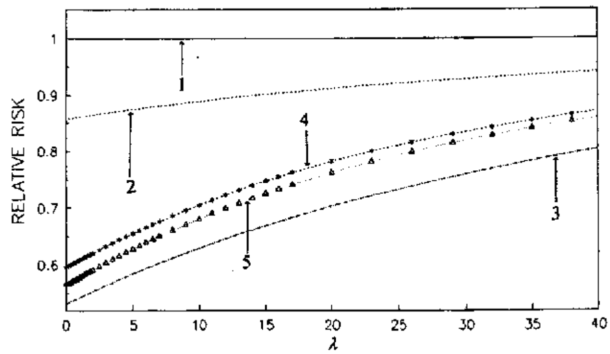

Fig. 4. Relative risks of estimators for $n = 20$, $k = 6$ and $\rho = 0.8$. Plotted for 1 = FGLSE, 2 = FMMSEE, 3 = AFMMSEE, 4 = SRE and 5 = SPRE.

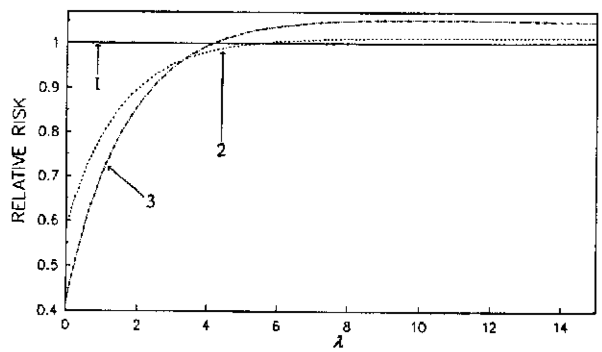

Fig. 5. Relative risks of estimators for $n = 20$, $k = 2$ and $\rho = -0.4$. Plotted for 1 = FGLSE, 2 = FMMSEE and 3 = AFMMSEE.

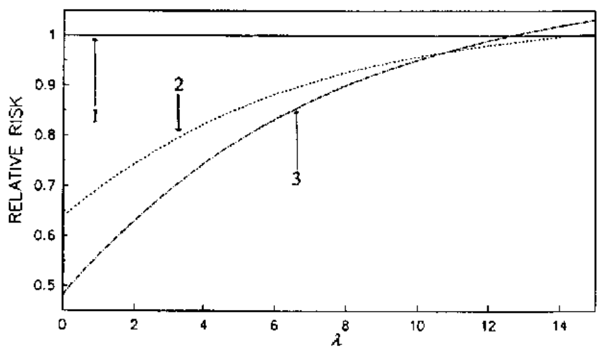

Fig. 6. Relative risks of estimators for $n = 20$, $k = 2$ and $\rho = 0.4$. Plotted for 1 = FGLSE, 2 = FMMSEE and 3 = AFMMSEE.

are quite striking relative to the FGLSE–depending on the values of ρ, the risks of $\hat{\beta}_{AM}$ and $\hat{\beta}_M$ are, at best, 36 and 70 percent, and at worst, 48 and 77 percent, respectively, of the risk of $\hat{\beta}$. Regardless of the ρ value, the relative risks of $\hat{\beta}_{AM}$ and $\hat{\beta}_M$ increase as λ increases, *ceteris paribus*. On the other hand, increasing ρ results in decreasing the risks of these estimators relative to the FGLSE risk over quite a large range of values for λ (Fig. 3). For small to moderate values of λ (say, < 10), the risk of $\hat{\beta}_{AM}$ averages

80 to 90 percent the risk of the SPRE. Generally speaking, the larger is λ, the smaller is the risk difference between the AFMMSEE and the SPRE, with the former estimator having the edge on the latter in most circumstances.

Much of these general patterns persist as k increases (Fig. 4). Exceptions occur at and near $\lambda = 0$, where the SPRE can be marginally superior to the AFMMSEE when k is large. Other things being equal, the risk difference between $\hat{\beta}_{AM}$ and the SPRE decreases as k increases. Over much of the parameter space, however, the former estimator continues to be the preferred estimator over the SPRE.

For $k = 2$, the SRE and SPRE reduce to the FGLSE. Comparing the risks of the FMMSEE and AFMMSEE, neither estimator dominates the other (Figs. 5 and 6), and both of these estimators can have greater risks than the FGLSE for larger λ values. At worst, the FMMSEE and AFMMSEE risks are, respectively, 5.6 and 12.9 percent larger than the FGLSE risk. This is more than compensated for by a substantial decline in risk of using the former estimators in place of the latter for small values of λ. Typically, the FMMSEE dominates the AFMMSEE in regions where both estimators are dominated by the FGLSE. On the other hand, over much of the rest of the parameter space, the AFMMSEE habitually enjoys a smaller risk than the FMMSEE. The value of λ at which the risk function of $\hat{\beta}_M$ crosses with that of $\hat{\beta}_{AM}$ seems to increase as ρ increase, *ceteris paribus.*

Overall, while none of the estimators performs best under all circumstances, the AFMMSEE is often found to provide the smallest risk in a variety of situations. For large values of λ, the choice of estimators appears to make very little difference. On the other hand, when λ is small, and especially for small number of regressors, moderate to large reductions in risk can often be made by replacing the FGLSE with the AFMMSEE.

5. Conclusions

In this paper, we have extended the existing results on a class of operational variants of the minimum mean squared error estimators, by considering models with non-spherical error structures. The motivation of considering this class of estimators is stressed in Sections 1 and 2. We examine approximations to the distribution of this class of estimators, and explore the finite sample performance of these estimators in a Monte Carlo study. In many cases, our results reinforce the conclusions of Ohtani (1996*b*) that the adjusted feasible minimum mean squared error estimator is often a superior and viable alternative to the Stein-rule estimators, and the gains of using this estimator generally outweigh the losses. Of course, as with all Monte-Carlo experiments, the simulation results presented here are of a limited nature. It remains for future research to consider other forms of autocorrelation processes, and the effects of the choice of design matrix have upon the properties of these estimators.

Acknowledgements

We are grateful to Kazuhiro Ohtani, Viren Srivastava, Götz Trenkler and the referees for helpful comments. The first author acknowledges financial support from the City University of Hong Kong.

References

Baranchik, A. J. (1970). A family of minimax estimators of the mean of a multivariate normal distribution, *Ann. Math. Statist.*, **41**, 642–645.

Chaturvedi, A. and Shukla G. (1990). Stein-rule estimation in linear model with non-scalar error covariance matrix, *Sankhyā Ser. B*, **52**, 293–304.

Draper, N. R. and Van Nostrand, R. C. (1979). Ridge regression and James Stein estimation: review and comments, *Technometrics*, **21**, 451–466.

Dwivedi, T. D. and Srivastava, V. K. (1978). On the minimum mean squared error estimator in a regression model, *Comm. Statist. Theory Methods*, **7**, 487–494.

Farebrother, R. W. (1975). The minimum mean square error linear estimator and ridge regression, *Technometrics*, **17**, 127–128.

Liski, E. P., Toutenburg, H. and Trenkler, G. (1993). Minimum mean square error estimation in linear regression, *J. Statist. Plann. Inference*, **37**, 203–214.

Mayer, L. S. and Willke, T. A. (1973). On biased estimation in linear models, *Technometrics*, **15**, 497–508.

Ohtani, K. (1996*a*). Exact small sample properties of an operational variant of the minimum mean squared error estimator, *Comm. Statist. Theory Methods*, **25**, 1223–1231.

Ohtani, K. (1996*b*). On an adjustment of degrees of freedom in the minimum mean squared error estimator, *Comm. Statist. Theory Methods*, **25**, 3049–3058.

Ohtani, K. (1997). Minimum mean squared error estimation of each individual coefficient in a linear regression model, *J. Statist. Plann. Inference*, **62**, 301–316.

Ohtani, K. (1999). MSE performance of a heterogeneous pre-test estimator, *Statist. Probab. Lett.*, **41**, 65–71.

Prais, S. J. and Winsten, C. B. (1954). Trend estimators and serial correlation, Cowles Commission Discussion, Paper No. 383, University of Chicago.

Rao, C. R. and Toutenburg, H. (1995). *Linear Models: Least Squares and Alternatives*, Springer, New York.

Rothenberg, T. J. (1984). Approximate normality of generalized least squares estimates, *Econometrica*, **52**, 811–825.

Stahlecker, P. and Trenkler, G. (1985). On heterogeneous versions of the best linear and the ridge estimator, *Proceedings of the First International Tampere Seminar on Linear Statistical Models and Their Applications* (eds. T. Pukkila and S. Puntanen), University of Tampere, Finland.

Theil, H. (1971). *Principles of Econometrics*, Wiley, New York.

Toutenburg, H., Trenkler, G. and Liski, E. (1992). Optimal estimation methods under weakened linear restrictions in regression, *Comput. Statist. Data Anal.*, **14**, 527–536.

Tracy, D. S. and Srivastava, A. K. (1994). Comparison of operational variants of best homogeneous and heterogeneous estimators in linear regression, *Comm. Statist. Theory Methods*, **23**, 2313–2322.

Vinod, H. D. (1976). Simulation and extension of a minimum mean squared error estimator in comparison with Stein's, *Technometrics*, **18**, 491–496.

Vinod, H. D. (1978). A survey of ridge regression and related techniques for improvements over ordinary least squares, *Review of Economics and Statistics*, **60**, 121–131.