Ann. Inst. Statist. Math.
Vol. 43, No. 4, 677–688 (1991)

ROBUST ESTIMATION OF COMMON REGRESSION COEFFICIENTS UNDER SPHERICAL SYMMETRY*

T. KUBOKAWA[1], C. ROBERT[2] AND A. K. Md. E. SALEH[3]

[1] *Department of Mathematical Engineering and Information Physics, Faculty of Engineering, University of Tokyo, Bunkyo-ku, Tokyo 113, Japan*
[2] *I.S.U.P.-L.S.T.A., Universite de Paris VI, Tour 45, 4, place Jussieu, 75252 Paris Cedex 05, France*
[3] *Department of Mathematics and Statistics, Carleton University, Ottawa, Canada, K1S 5B6*

(Received October 20, 1989; revised October 17, 1990)

Abstract. Consider the problem of estimating the common regression coefficients of two linear regression models where the two distributions of the errors may be different and unknown. Under the spherical symmetry assumption, the paper proves the superiority of a Graybill-Deal type combined estimator and the further improvement by the Stein effect which were exhibited by Shinozaki (1978, *Comm. Statist. Theory Methods*, **7**, 1421–1432) in the normal case. This shows the robustness of the dominations since the conditions for the dominations are independent of the errors distributions.

Key words and phrases: Elliptically contoured distribution, heteroscedastic linear model, Stein problem, common mean, Graybill-Deal estimator.

1. Introduction

1.1 *Motivations*

In the problem of estimating the coefficients of a linear regression model, the normality assumption on the errors distribution has often been criticized as being too restrictive. Cellier, Fourdrinier and Robert (1989)—later denoted by CFR in this paper—have shown that the usual Stein domination results hold under the much weaker assumption of *spherical symmetry.* More precisely, for the regression model

$$\underset{(n\times 1)}{y} = \underset{(n\times p)}{X}\;\underset{(p\times 1)}{\beta} + \underset{(n\times 1)}{\varepsilon}, \tag{1.1}$$

and for $p \geq 3$, the least squares estimator $\hat{\beta}$ is dominated by a shrinkage estimator if ε has a *spherically symmetric distribution* (i.e. the density of ε factorizes through

* Research supported by NSERC Grant No. A3088 and GR-5 Grant from Faculty of Graduate Studies, Carleton University, Ottawa, Canada.

$\|\varepsilon\|^2$), and *the domination conditions are independent of the distribution.* In this respect, we can speak of the *robustness of the Stein effect.* We give in Subsection 3.1 an application of CFR results to the linear regression model (1.1). Note that the results are also valid if ε has an *elliptically symmetric distribution* (i.e. when the symmetric matrix defining the norm is not the identity).

Now, it may happen that several linear regression models are available with the same unknown coefficients β:

$$\underset{(n_i\times 1)}{y_i} = \underset{(n_i\times p)}{X_i}\ \underset{(p\times 1)}{\beta} + \underset{(n_i\times 1)}{\varepsilon_i} \qquad (i=1,\dots,k). \tag{1.2}$$

For instance, one may want to consider the effect of several economic indicators (X_i) on the unemployment rate (y_i) for EEC (European Economic Community) countries ($k = 12$) and believe that the weights of each factor are the same for every country. The number of observations for each country (n_i) may differ, as well as the distribution of the errors (ε_i), due to different data collection methods for instance. These distributions may even be unknown but a spherical symmetry assumption is reasonable.

In this paper, we investigate the estimation of the parameter β in the special case where $k = 2$. We establish the domination of the naive least squares estimators, $\hat\beta_1$ and $\hat\beta_2$, by a compound estimator for every quadratic loss (Section 2). Furthermore, we prove that this combined estimator is itself dominated by a class of shrinkage estimators (Subsection 3.2). Both results are independent of the distributions of the errors and only depend on the spherical symmetry assumption.

1.2 *Previous results*

For the model (1.2), the least squares estimators are ($i = 1, 2$)

$$\hat\beta_i = H_i^{-1}X_i'y_i \tag{1.3}$$

where $H_i = X_i'X_i$, and the residual sum of squares are

$$s_i = y_i'[I_{n_i} - X_iH_i^{-1}X_i']y_i. \tag{1.4}$$

The model (1.2) is called a *heteroscedastic linear model* as the ε_i may have different distributions, and the estimation problem of the common coefficients β has been studied in several papers. For the references, see Kubokawa (1989). Of these, Shinozaki (1978), based on Graybill and Deal (1959), proposed a combined estimator

$$\hat\beta_{\mathrm{CM}} = W_1\hat\beta_1 + W_2\hat\beta_2, \tag{1.5}$$

where

$$\begin{aligned} W_1 &= (s_1H_2 + cs_2H_1)^{-1}cs_2H_1, \\ W_2 &= (s_1H_2 + cs_2H_1)^{-1}s_1H_2. \end{aligned} \tag{1.6}$$

Note that

$$\hat{\beta}_{\mathrm{CM}} = \left[\lambda\frac{H_1}{s_1} + (1-\lambda)\frac{H_2}{s_2}\right]^{-1}\left[\lambda\frac{H_1}{s_1}\hat{\beta}_1 + (1-\lambda)\frac{H_2}{s_2}\hat{\beta}_2\right],$$

where $\lambda = c/(1+c)$. Therefore, the form of $\hat{\beta}_{\mathrm{CM}}$ is quite natural, as a convex combination of the two original estimators, the matricial weights being inversely proportional to s_i. Note also that H_i^{-1} is proportional to the covariance matrix of $\hat{\beta}_i$ $(i = 1, 2)$ when it is defined.

Under the assumption that $\varepsilon_i \sim N_{n_i}(0, \sigma_i^2 I_{n_i})$, $i = 1, 2$, Shinozaki (1978) established that

$$\mathrm{Cov}(\hat{\beta}_{\mathrm{CM}}) \leq \mathrm{Cov}(\hat{\beta}_i), \qquad i = 1, 2, \tag{1.7}$$

holds uniformly in β, σ_1^2, σ_2^2, if and only if

$$\frac{n_1 - p + 2}{2(n_2 - p - 4)} \leq c \leq \frac{2(n_1 - p - 4)}{n_2 - p + 2}. \tag{1.8}$$

The inequality (1.7) means that the matrix $\mathrm{Cov}(\hat{\beta}_i) - \mathrm{Cov}(\hat{\beta}_{\mathrm{CM}})$ is nonnegative definite, and we call it the *covariance criterion.* As $\hat{\beta}_1$ and $\hat{\beta}_2$ are unbiased, $\hat{\beta}_{\mathrm{CM}}$ is also *unbiased* (as $W_1 + W_2 = I_p$ and these two matrices depend only on s_i, $i = 1, 2$). Therefore, (1.7) is equivalent to

$$E[(\hat{\beta}_{\mathrm{CM}} - \beta)'Q(\hat{\beta}_{\mathrm{CM}} - \beta)] \leq E[(\hat{\beta}_i - \beta)'Q(\hat{\beta}_i - \beta)] \tag{1.9}$$

for every nonnegative definite symmetric matrix Q; this is uniform domination over the class of quadratic losses. This result extends Graybill and Deal (1959) and Khatri and Shah (1974). As noted in Shinozaki (1978), we can choose c satisfying (1.8) if and only if $n_1 - p \geq 7$, $n_2 - p \geq 7$ and $(n_1 - p - 6)(n_2 - p - 6) \geq 16$.

In the non-normal case, similar results have been obtained by Swamy and Mehta (1979) (see also Cohen (1976), Bhattacharya (1981) and Akai (1982)). However, the domination conditions always assume a certain knowledge of the error distribution. In Section 2, we prove that (1.8) is necessary and sufficient under the spherical symmetry assumption.

Shinozaki (1978) also established a sufficient condition for the domination of $\hat{\beta}_{\mathrm{CM}}$ under normality assumption and a quadratic loss. We give in Subsection 3.2 the robust equivalent of this result.

The following lemma is essential for our purpose; it is derived from CFR and can be shown by integration by parts. A similar result is also to be found in Berger (1975).

LEMMA 1.1. *Let $0 \leq f \in L^1([0, +\infty))$ and h an absolutely continuous function. Define $F(x) = 2^{-1}\int_x^{+\infty} f(t)dt$ and assume that*

(a) $\int_{-\infty}^{+\infty} |h(x)||x - t| f((x-t)^2 + a^2)dx < \infty$, *for any real a, t,*

(b) $\lim_{t\to\pm\infty} h(t)F(t^2) = 0$.

Then

$$\int_{-\infty}^{+\infty} h(x)(x-t)f((x-t)^2 + a^2)dx = \int_{-\infty}^{+\infty} h'(x)F((x-t)^2 + a^2)dx. \tag{1.10}$$

2. Uniform domination by combined estimators

In the model (1.2) with $k = 2$, we suppose that ε_1 and ε_2 are independent and that for $i = 1, 2$, the density of ε_i is given by $f_i(\|\varepsilon_i\|^2)$ where $f_i \in L^1([0, +\infty))$. It is noted that we do not assume f_i to be known. Also note that the spherical symmetry assumption on ε_i allows to deal with models where the components may be dependent (though they are still uncorrelated). We now prove that Shinozaki's (1978) result still holds under the spherical symmetry assumption.

THEOREM 2.1. *Assume that $E[\hat\beta_i' H_i \hat\beta_i] < \infty$ and $n_i \geq p + 5$ for $i = 1, 2$. Then the combined estimator $\hat\beta_{\mathrm{CM}}$ given by (1.5) dominates both $\hat\beta_1$ and $\hat\beta_2$ under the covariance criterion if and only if*

$$\frac{n_1 - p + 2}{2(n_2 - p - 4)} \leq c \leq \frac{2(n_1 - p - 4)}{n_2 - p + 2}. \tag{2.1}$$

PROOF. We first write

$$\begin{aligned}
\mathrm{Cov}(\hat\beta_{\mathrm{CM}}) &= E[(\hat\beta_{\mathrm{CM}} - \beta)(\hat\beta_{\mathrm{CM}} - \beta)'] \\
&= E[W_1(\hat\beta_1 - \beta)(\hat\beta_1 - \beta)'W_1' + W_2(\hat\beta_2 - \beta)(\hat\beta_2 - \beta)'W_2' \\
&\qquad + W_1(\hat\beta_1 - \beta)(\hat\beta_2 - \beta)'W_2' + W_2(\hat\beta_2 - \beta)(\hat\beta_1 - \beta)'W_1'].
\end{aligned}$$

Note that there exists a nonsingular matrix P such that $H_1 = PP'$ and $H_2 = PD_\lambda P'$ for $D_\lambda = \mathrm{diag}(\lambda_1, \ldots, \lambda_p)$. Put $u_1 = (u_{11}, \ldots, u_{1p})' = P'\hat\beta_1$ and $\nu_1 = (\nu_{11}, \ldots, \nu_{1p})' = P'\beta$. It can be seen that there exists a $n_1 \times n_1$ orthogonal matrix P_1 such that $P_1 y_1 = (u_1', v_1')'$ and $y_1'(I_{n_1} - X_1 H_1^{-1} X_1')y_1 = \|v_1\|^2$ for $v_1 = (v_{11}, \ldots, v_{1,n_1-p})'$ (see Nickerson (1987), p. 98, for example). Then $\|\varepsilon_1\|^2 = \|y_1 - X_1\beta\|^2 = (\hat\beta_1 - \beta)'H_1(\hat\beta_1 - \beta) + y_1'(I_{n_1} - X_1 H_1^{-1} X_1')y_1 = \|u_1 - \nu_1\|^2 + \|v_1\|^2$. Similarly we can write $u_2 = D_\lambda^{1/2} P'\hat\beta_2$, $\nu_2 = D_\lambda^{1/2} P'\beta$ and $\|v_2\|^2 = s_2$ and get

$$\begin{aligned}
&P'\,\mathrm{Cov}(\hat\beta_{\mathrm{CM}})P \\
&= E\{(cs_2)^2[s_1 D_\lambda + cs_2 I]^{-1}(u_1 - \nu_1)(u_1 - \nu_1)'[s_1 D_\lambda + cs_2 I]^{-1} \\
&\qquad + s_1^2[s_1 D_\lambda + cs_2 I]^{-1} D_\lambda^{1/2}(u_2 - \nu_2)(u_2 - \nu_2)' D_\lambda^{1/2}[s_1 D_\lambda + cs_2 I]^{-1} \\
&\qquad + cs_1 s_2[s_1 D_\lambda + cs_2 I]^{-1}(u_1 - \nu_1)(u_2 - \nu_2)' D_\lambda^{1/2}[s_1 D_\lambda + cs_2 I]^{-1} \\
&\qquad + cs_1 s_2[s_1 D_\lambda + cs_2 I]^{-1} D_\lambda^{1/2}(u_2 - \nu_2)(u_1 - \nu_1)'[s_1 D_\lambda + cs_2 I]^{-1}\} \\
&= E\left[\mathrm{diag}\left\{\frac{(cs_2)^2}{(s_1\lambda_j + cs_2)^2}(u_{1j} - \nu_{1j})^2 \right.\right. \\
&\qquad \left.\left. + \frac{s_1^2 \lambda_j}{(s_1\lambda_j + cs_2)^2}(u_{2j} - \nu_{2j})^2\right\}_{j=1,\ldots,p}\right] \\
&= E\left[\mathrm{diag}\left\{\frac{(c\|v_2\|^2)^2}{(\|v_1\|^2\lambda_j + c\|v_2\|^2)^2}(u_{1j} - \nu_{1j})^2 \right.\right. \\
&\qquad \left.\left. + \frac{\|v_1\|^4 \lambda_j}{(\|v_1\|^2\lambda_j + c\|v_2\|^2)^2}(u_{2j} - \nu_{2j})^2\right\}_{j=1,\ldots,p}\right]
\end{aligned} \tag{2.2}$$

where the second equality in (2.2) follows from the symmetry of the distribution of $u_{ij} - \nu_{ij}$. Here using the identity (1.10) for $u_{1j} - \nu_{1j}$ conditionally on the other coordinates gives that

$$(2.3)\quad \iint \frac{(c\|v_2\|^2)^2}{(\|v_1\|^2\lambda_j + c\|v_2\|^2)^2}(u_{1j} - \nu_{1j})^2 f_1(\|u_1 - \nu_1\|^2 + \|v_1\|^2)du_1dv_1$$
$$= \iint \frac{(c\|v_2\|^2)^2}{(\|v_1\|^2\lambda_j + c\|v_2\|^2)^2} F_1(\|u_1 - \nu_1\|^2 + \|v_1\|^2)du_1dv_1,$$

where $du_i = \prod_{j=1}^{p} du_{ij}$ and $dv_i = \prod_{j=1}^{n_i-p} dv_{ij}$. Further, using the identity (1.10) for v_{2k} yields

$$(2.4)\quad \iint \frac{(c\|v_2\|^2)^2}{(\|v_1\|^2\lambda_j + c\|v_2\|^2)^2} f_2(\|u_2 - \nu_2\|^2 + \|v_2\|^2)du_2dv_2$$
$$= \sum_{k=1}^{n_2-p} \iint \frac{c^2\|v_2\|^2}{(\|v_1\|^2\lambda_j + c\|v_2\|^2)^2} v_{2k}v_{2k} f_2(\|u_2 - \nu_2\|^2 + \|v_2\|^2)du_2dv_2$$
$$= \iint \frac{c^2(n_2 - p + 2)(\|v_1\|^2\lambda_j + c\|v_2\|^2)\|v_2\|^2 - 4c^3\|v_2\|^4}{(\|v_1\|^2\lambda_j + c\|v_2\|^2)^3}$$
$$\cdot F_2(\|u_2 - \nu_2\|^2 + \|v_2\|^2)du_2dv_2.$$

Combining (2.3) and (2.4), we obtain that

$$(2.5)\quad E\left[\frac{(c\|v_2\|^2)^2}{(\|v_1\|^2\lambda_j + c\|v_2\|^2)^2}(u_{1j} - \nu_{1j})^2\right]$$
$$= \iiiint \frac{c^2(n_2 - p + 2)(\|v_1\|^2\lambda_j + c\|v_2\|^2)\|v_2\|^2 - 4c^3\|v_2\|^4}{(\|v_1\|^2\lambda_j + c\|v_2\|^2)^3}$$
$$\cdot F_1(\|u_1 - \nu_1\|^2 + \|v_1\|^2)F_2(\|u_2 - \nu_2\|^2 + \|v_2\|^2)du_1dv_1du_2dv_2.$$

Similarly,

$$(2.6)\quad \lambda_j^{-1}E\left[\frac{(\|v_1\|^2\lambda_j)^2}{(\|v_1\|^2\lambda_j + c\|v_2\|^2)^2}(u_{2j} - \nu_{2j})^2\right]$$
$$= \iiiint \frac{\lambda_j(n_1 - p + 2)(\|v_1\|^2\lambda_j + c\|v_2\|^2)\|v_1\|^2 - 4\lambda_j^2\|v_1\|^4}{(\|v_1\|^2\lambda_j + c\|v_2\|^2)^3}$$
$$\cdot F_1(\|u_1 - \nu_1\|^2 + \|v_1\|^2)$$
$$\cdot F_2(\|u_2 - \nu_2\|^2 + \|v_2\|^2)du_1dv_1du_2dv_2,$$

so that from (2.5) and (2.6),

$$(2.7)\quad E\left[\frac{(c\|v_2\|^2)^2}{(\|v_1\|^2\lambda_j + c\|v_2\|^2)^2}(u_{1j} - \nu_{1j})^2 + \frac{\|v_1\|^4\lambda_j}{(\|v_1\|^2\lambda_j + c\|v_2\|^2)^2}(u_{2j} - \nu_{2j})^2\right]$$
$$= \iiiint (\|v_1\|^2\lambda_j + c\|v_2\|^2)^{-3}$$
$$\cdot [\{c^2(n_2 - p + 2)\|v_2\|^2 + \lambda_j(n_1 - p + 2)\|v_1\|^2\}$$
$$\cdot (\|v_1\|^2\lambda_j + c\|v_2\|^2) - 4c^3\|v_2\|^4 - 4\lambda_j^2\|v_1\|^4]$$
$$\cdot F_1(\|u_1 - \nu_1\|^2 + \|v_1\|^2)$$
$$\cdot F_2(\|u_2 - \nu_2\|^2 + \|v_2\|^2)du_1dv_1du_2dv_2.$$

On the other hand, the similar representation of $P' \operatorname{Cov}(\hat{\beta}_1)P$ can be derived by taking $c = \infty$ in (2.7) as

$$(2.8) \qquad \begin{aligned} &P' \operatorname{Cov}(\hat{\beta}_1)P \\ &= \Big(\iiiint \frac{n_2 - p - 2}{\|v_2\|^2} F_1(\|u_1 - \nu_1\|^2 + \|v_1\|^2) \\ &\qquad \cdot F_2(\|u_2 - \nu_2\|^2 + \|v_2\|^2) du_1 dv_1 du_2 dv_2 \Big) I_p. \end{aligned}$$

Indeed, since $E[\|u_1\|^2] < \infty$, we can apply the dominated convergence theorem to get (2.8). Hence from (2.7) and (2.8), $\operatorname{Cov}(\hat{\beta}_{\mathrm{CM}}) \leq \operatorname{Cov}(\hat{\beta}_1)$ uniformly in β, σ_1^2, σ_2^2 if for $z = c\|v_2\|^2/(\lambda_j \|v_1\|^2)$,

$$\begin{aligned} &\frac{z}{c}(1+z)^{-3}[(n_2 - p + 2)cz(1+z) - 4cz^2 + (n_1 - p + 2)(1+z) - 4] \\ &\quad \leq n_2 - p - 2, \end{aligned}$$

which is always satisfied by the condition (2.1). Similarly, it can be verified that $\operatorname{Cov}(\hat{\beta}_{\mathrm{CM}}) \leq \operatorname{Cov}(\hat{\beta}_2)$ under the condition (2.1), and the sufficiency of Theorem 2.1 holds. If $\hat{\beta}_{\mathrm{CM}}$ dominates $\hat{\beta}_1$ and $\hat{\beta}_2$ under the covariance criterion for any f_1 and f_2, then the same domination holds for $\varepsilon_1 \sim N_{n_1}(0, \sigma_1^2 I_{n_1})$ and $\varepsilon_2 \sim N_{n_2}(0, \sigma_2^2 I_{n_2})$, so that the necessity follows from the result of Shinozaki (1978). Therefore the proof of Theorem 2.1 is complete.

Since $\hat{\beta}_{\mathrm{CM}}$ is unbiased, $\operatorname{Cov}(\hat{\beta}_{\mathrm{CM}}) = E[(\hat{\beta}_{\mathrm{CM}} - \beta)(\hat{\beta}_{\mathrm{CM}} - \beta)']$. This is the *matricial mean square error* (MMSE), and the domination under MMSE is equivalent to that under the set of quadratic losses as noted at (1.9).

3. Further domination by shrinkage estimators

In this section, we consider the quadratic loss function

$$(3.1) \qquad L(\hat{\beta}, \beta) = (\hat{\beta} - \beta)' Q (\hat{\beta} - \beta),$$

where Q is a positive definite and known matrix. As noted above, the domination result in Theorem 2.1 remains true under the loss (3.1). For $p \geq 3$, $\hat{\beta}_{\mathrm{CM}}$ is further dominated by a Stein type estimator. In fact, this was demonstrated by Shinozaki (1978) for the normal case. Under spherical symmetry assumption, we develop a shrinkage estimator dominating $\hat{\beta}_{\mathrm{CM}}$ relative to the loss (3.1). Since the essence of the derivation is given by CFR, we first introduce their result for one sample regression model (1.1). Notice that the proof of Theorem 3.1, as well as the model, makes things more readable than CFR original developments.

3.1 *Robust shrinkage estimators*

Let $y = X\beta + \varepsilon$ be a regression model given by (1.1). Assume that ε has a spherically symmetric and unknown density $f(\|\varepsilon\|^2)$ where $f \in L^1([0,\infty))$. Stein (1956) and James and Stein (1961) exhibited the inadmissibility of the least squares estimator $\hat\beta_0 = (X'X)^{-1}X'y$ for $p \geq 3$ and $Q = X'X$ in the normal case, and various extensions and investigations have been studied since. Recently, CFR showed a robustness of the Stein phenomena. Consider shrinkage estimators of the form

$$\hat\beta_0^S = (I - g^{-1}\phi(g,s)Q^{-1}H)\hat\beta_0,$$

where $g = \hat\beta_0'HQ^{-1}H\hat\beta_0/s$, $s = y'(I - XH^{-1}X')y$, $H = X'X$ and ϕ is a positive function.

THEOREM 3.1. (Cellier *et al.* (1989)) *Let $p \geq 3$. Assume that $E[\hat\beta_0'H\hat\beta_0] < \infty$ and $E[s^2/\hat\beta_0'H\hat\beta_0] < \infty$. Then $\hat\beta_0^S$ dominates $\hat\beta_0$ relative to the loss* (3.1) *if*

(a) *$\phi(x,y)$ is nondecreasing in x and nonincreasing in y,*

(b) $0 < \phi(x,y) \leq 2(p-2)/(n-p+2)$.

PROOF. We give a simplified proof in the case when ϕ is differentiable. The extension to more general functions is given in CFR. The risk difference is written as

$$\begin{aligned}\Delta &= E[(\hat\beta_0^S - \beta)'Q(\hat\beta_0^S - \beta)] - E[(\hat\beta_0 - \beta)'Q(\hat\beta_0 - \beta)]\\ &= E[g^{-2}\{\phi(g,s)\}^2\hat\beta_0'HQ^{-1}H\hat\beta_0 - 2g^{-1}\phi(g,s)\hat\beta_0'H(\hat\beta_0 - \beta)]\\ &= E\left[\frac{\|v\|^4}{u'Ru}\left\{\phi\left(\frac{u'Ru}{\|v\|^2}, \|v\|^2\right)\right\}^2 - 2\frac{\|v\|^2}{u'Ru}\phi\left(\frac{u'Ru}{\|v\|^2}, \|v\|^2\right)u'(u-\nu)\right],\end{aligned}$$

where $R = H^{1/2}Q^{-1}H^{1/2}$. By the identity (1.10), for $u = (u_1, \ldots, u_p)'$ and $\nu = (\nu_1, \ldots, \nu_p)'$,

$$\begin{aligned}&E\left[\frac{\|v\|^2}{u'Ru}\phi\left(\frac{u'Ru}{\|v\|^2}, \|v\|^2\right)u_j(u_j - \nu_j)\right]\\ &\quad= \iint\left[\frac{\|v\|^2}{u'Ru}\phi\left(\frac{u'Ru}{\|v\|^2}, \|v\|^2\right)\right.\\ &\qquad - 2\frac{\|v\|^2}{(u'Ru)^2}\phi\left(\frac{u'Ru}{\|v\|^2}, \|v\|^2\right)\left(\sum_{i=1}^p r_{ij}u_iu_j\right)\\ &\qquad \left.+ \frac{2}{u'Ru}\phi'_{(1)}\left(\frac{u'Ru}{\|v\|^2}, \|v\|^2\right)\left(\sum_{i=1}^p r_{ij}u_iu_j\right)\right]\\ &\qquad \cdot F(\|u-\nu\|^2 + \|v\|^2)dudv,\end{aligned} \tag{3.2}$$

where $R = (r_{ij})$ and $\phi'_{(i)}(x_1,x_2) = (\partial/\partial x_i)\phi(x_1,x_2)$. Also the identity (1.10) for v_j gives that for $v = (v_1, \ldots, v_{n-p})'$,

$$E\left[\frac{\|v\|^2}{u'Ru}\left\{\phi\left(\frac{u'Ru}{\|v\|^2}, \|v\|^2\right)\right\}^2 v_jv_j\right] \tag{3.3}$$

$$
\begin{aligned}
&= \iint \Bigg[\frac{\|v\|^2}{u'Ru} \left\{ \phi\left(\frac{u'Ru}{\|v\|^2}, \|v\|^2 \right) \right\}^2 + \frac{2}{u'Ru} \left\{ \phi\left(\frac{u'Ru}{\|v\|^2}, \|v\|^2 \right) \right\}^2 v_j^2 \\
&\qquad + 4 \frac{\|v\|^2}{u'Ru} \phi\left(\frac{u'Ru}{\|v\|^2}, \|v\|^2 \right) \\
&\qquad \cdot \left\{ \phi'_{(2)}\left(\frac{u'Ru}{\|v\|^2}, \|v\|^2 \right) - \frac{u'Ru}{\|v\|^4} \phi'_{(1)}\left(\frac{u'Ru}{\|v\|^2}, \|v\|^2 \right) \right\} v_j^2 \Bigg] \\
&\qquad \cdot F(\|u - \nu\|^2 + \|v\|^2) du dv.
\end{aligned}
$$

From (3.2), (3.3) and the condition of Theorem 3.1,

$$
\begin{aligned}
\Delta = \iint \Bigg[&(n - p + 2) \frac{\|v\|^2}{u'Ru} \left\{ \phi\left(\frac{u'Ru}{\|v\|^2}, \|v\|^2 \right) \right\}^2 - 4\phi'_{(1)}\left(\frac{u'Ru}{\|v\|^2}, \|v\|^2 \right) \\
&+ 4\phi\left(\frac{u'Ru}{\|v\|^2}, \|v\|^2 \right) \left\{ \frac{\|v\|^4}{u'Ru} \phi'_{(2)}\left(\frac{u'Ru}{\|v\|^2}, \|v\|^2 \right) - \phi'_{(1)}\left(\frac{u'Ru}{\|v\|^2}, \|v\|^2 \right) \right\} \\
&- 2(p-2) \frac{\|v\|^2}{u'Ru} \phi\left(\frac{u'Ru}{\|v\|^2}, \|v\|^2 \right) \Bigg] F(\|u - \nu\|^2 + \|v\|^2) du dv \\
\le 0, &
\end{aligned}
$$

which proves Theorem 3.1.

It should be noted that the conditions of Theorem 3.1 do not depend on the distribution of ε; the Stein effect is robust with respect to the distribution of the errors.

3.2 *Improving on the combined estimators*

Now we return to the common mean problem defined in Section 1. Recall that the combined estimator $\hat{\beta}_{\mathrm{CM}}$ given by (1.5) is better than both the uncombined ones $\hat{\beta}_1$ and $\hat{\beta}_2$ for the loss (3.1). Here we show that by the Stein effect, $\hat{\beta}_{\mathrm{CM}}$ is further dominated by

$$
\hat{\beta}^S_{\mathrm{CM}} = \sum_{i=1}^{2} W_i(\hat{\beta}_i - e_i \psi_i \hat{\beta}_i), \tag{3.4}
$$

where for $i = 1, 2$,

$$
\psi_i = a_i g_i^{-1} (W_i' Q W_i)^{-1} H_i, \qquad g_i = \hat{\beta}_i' H_i (W_i' Q W_i)^{-1} H_i \hat{\beta}_i / s_i,
$$

$H_i = X_i' X_i$ and a_1, a_2, e_1, e_2 are nonnegative constants satisfying $e_1 + e_2 = 1$.

THEOREM 3.2. *Let $p \ge 3$. Assume that $E[\hat{\beta}_i' H_i \hat{\beta}_i] < \infty$ and $E[s_i^2 / \hat{\beta}_i' H_i \hat{\beta}_i] < \infty$ for $i = 1, 2$. Also assume that for $i = 1, 2$,*

$$
0 < a_i \le \frac{2(p-2)}{n_i - p + 2} \cdot \frac{ch_1(Q^{-1} H_i)}{ch_p(Q^{-1} H_i)}, \tag{3.5}
$$

where $ch_1(A)$ *and* $ch_p(A)$ *designate the smallest and the largest eigenvalues of a* $p \times p$ *matrix* A. *Then* $\hat{\beta}^S_{\mathrm{CM}}$ *given by* (3.4) *dominates* $\hat{\beta}_{\mathrm{CM}}$ *under the loss* (3.1). *When* H_1, H_2 *and* Q *satisfy the inequality*

$$u'H_1Q^{-1}H_2u \geq 0 \quad \text{for any} \quad u \in R^p, \tag{3.6}$$

the above sufficient condition on a_i *is relaxed as*

$$0 < a_i \leq 2(p-2)/(n_i - p + 2). \tag{3.7}$$

PROOF. Since $\hat{\beta}^S_{\mathrm{CM}} = e_1\{W_1(I-\psi_1)\hat{\beta}_1 + W_2\hat{\beta}_2\} + e_2\{W_1\hat{\beta}_1 + W_2(I-\psi_2)\hat{\beta}_2\}$, from the convexity of the loss function, it suffices to show that

$$\Delta_1 = R(W_1(I-\psi_1)\hat{\beta}_1 + W_2\hat{\beta}_2) - R(\hat{\beta}_{\mathrm{CM}}) \leq 0, \tag{3.8}$$

$$\Delta_2 = R(W_1\hat{\beta}_1 + W_2(I-\psi_2)\hat{\beta}_2) - R(\hat{\beta}_{\mathrm{CM}}) \leq 0, \tag{3.9}$$

where $R(\hat{\beta}) = E[(\hat{\beta} - \beta)'Q(\hat{\beta} - \beta)]$. We first prove (3.8). Note that the risk difference Δ_1 is rewritten as

$$\begin{aligned} \Delta_1 &= E[a_1^2 s_1 g_1^{-1} - 2a_1 g_1^{-1}\hat{\beta}_1' H_1(\hat{\beta}_1 - \beta)] \\ &= E[a_1^2\|v_1\|^2 g_1^{-1} - 2a_1 g_1^{-1} u_1'(u_1 - \nu_1)], \end{aligned}$$

where

$$\begin{aligned} g_1 &= u_1'[s_1D_\lambda + cs_2I]R_1[s_1D_\lambda + cs_2I]u_1/\{s_1(cs_2)^2\} \\ &= \sum_{i,j}(\|v_1\|^2\lambda_i + cs_2)(\|v_1\|^2\lambda_j + cs_2)r_{ij}u_{1i}u_{1j}/\{\|v_1\|^2(cs_2)^2\}, \end{aligned}$$

and $R_1 = (r_{ij}) = P'Q^{-1}P$ for the matrix P defined in the proof of Theorem 2.1. Similar to (3.2) and (3.3),

$$\begin{aligned} &E[g_1^{-1}u_1'(u_1 - \nu_1) \mid y_2] \\ &\quad = \iint (p-2)g_1^{-1}F_1(\|u_1 - \nu_1\|^2 + \|v_1\|^2)du_1dv_1, \end{aligned} \tag{3.10}$$

$$\begin{aligned} &E[\|v_1\|^2 g_1^{-1} \mid y_2] \\ &\quad = \iint \Big\{ \frac{n_1 - p + 2}{g_1} - 4(cs_2g_1)^{-2} \\ &\qquad\qquad \cdot \Big[\sum_{i,j}(\|v_1\|^2\lambda_i + cs_2)\lambda_j r_{ij}u_{1i}u_{1j}\Big]\Big\} \\ &\qquad\quad \cdot F_1(\|u_1 - \nu_1\|^2 + \|v_1\|^2)du_1dv_1. \end{aligned} \tag{3.11}$$

If (3.6) holds, then $\sum_{i,j}(\|v_1\|^2\lambda_i + cs_2)\lambda_j r_{ij}u_{1i}u_{1j} \geq 0$, so that we get that

$$\Delta_1 \leq E\left[\iint \frac{a_1}{g_1}\{(n_1 - p + 2)a_1 - 2(p-2)\}F_1(\|u_1 - \nu_1\|^2 + \|v_1\|^2)du_1dv_1\right],$$

which is less than zero for the condition (3.7), and (3.8) is proved under the restrictive condition (3.6).

Next we treat the general case without assuming (3.6). Note that

$$(3.12) \qquad ch_1(P'Q^{-1}P)h_1 \leq g_1 \leq ch_p(P'Q^{-1}P)h_1,$$

where

$$h_1 = u_1'[s_1 D_\lambda + cs_2 I]^2 u_1/\{s_1(cs_2)^2\}.$$

For simplicity, put $d_i = ch_i(P'Q^{-1}P)$. From (3.10), (3.11) and (3.12), we observe that

$$\begin{aligned}
\Delta_1 &\leq E[a_1^2\|v_1\|^2(d_1h_1)^{-1} - 2a_1g_1^{-1}u_1'(u_1-\nu_1)] \\
&= E\left[\iint\left[\frac{a_1^2}{d_1}\left\{\frac{n_1-p+2}{h_1} - \frac{4}{(cs_2h_1)^2}\sum_{i=1}^{p}(\|v_1\|^2\lambda_i + cs_2)\lambda_i u_{1i}^2\right\}\right.\right. \\
&\qquad \left.\left. - 2(p-2)a_1g_1^{-1}\right]F_1(\|u_1-\nu_1\|^2 + \|v_1\|^2)du_1dv_1\right] \\
&\leq \iint [a_1(n_1-p+2)d_1^{-1} - 2(p-2)d_p^{-1}]\frac{a_1}{h_1}F_1(\|u_1-\nu_1\|^2 + \|v_1\|^2)du_1dv_1,
\end{aligned}$$

which is less than zero for the condition (3.5). Hence (3.8) is proved without the condition (3.6). (3.9) can be established in a similar way and the proof of Theorem 3.2 is complete.

The condition (3.7) is available if $Q = (1-\alpha)H_1 + \alpha H_2$ for $0 \leq \alpha \leq 1$. This type of loss is quite natural when we want to mix the two regression models according to their covariances.

4. Comments and generalizations

Theorem 3.2 shows that we can improve upon a simple composite estimator like $\hat{\beta}_{CM}$ by taking advantage of the Stein effect in a semi-nonparametric setting. In fact, we do not need to specify precisely the distribution of the errors for the two regression models; we have only to assume that they are spherically symmetric. Moreover, the two distributions may be completely different, apart from the removal of the much restrictive normality assumption. This point is interesting as it allows, in practice, to make use simultaneously of obviously different sets of observations. The best example is when the two regression equations deal with two places where the data collection methodology completely differs. In such cases, the error distributions are presumably different. Even under such unfriendly situations, we can yet improve upon estimators like $\hat{\beta}_{CM}$ uniformly.

A straightforward generalization of this result would be to consider k simultaneous regression equations with common coefficients as Shinozaki (1978) did for the normal case. As one can see in the proofs of this paper, the simultaneous diagonalization theorem we use cannot be extended to more than two equations. One would have then to use more complicated spectral algebra (see Fraisse *et* *al.* (1987)) or to consider other techniques, as those involved in ridge regression analysis (Casella (1980, 1985)). However, the problem is more technical than fundamental, in our eyes, as the (already) long history of shrinkage estimators indicates that such an extension is presumably true.

Another direction of work would be to examine the problem where the coefficients of the two regression models are presumably equal, but with some degree of uncertainty. In such a case, an empirical Bayes analysis could draw a bridge between the estimator $\hat{\beta}^S_{\rm CM}$ and some estimators like $(\hat{\beta}'_1, \hat{\beta}'_2)'$, i.e. estimators deduced from each model separately (see Ghosh *et al.* (1989) or Robert and Saleh (1989)).

A last step would be the removal of the spherical symmetry assumption, in the direction shown by Shinozaki (1984).

Finally it is noted that a corresponding domination of the joint least squares estimator is impossible. In fact, if the two equations have normal errors with the same variance, the common least square estimator is the best unbiased estimator (the Gauss-Markov theorem) and thus cannot be dominated by the weighted sum $\hat{\beta}_{\rm CM}$.

Acknowledgements

The authors are grateful to Professor J. N. K. Rao for his valuable comments concerning the common mean problem. They also thank the referees for their constructive remarks.

References

Akai, T. (1982). A combined estimator of a common parameter, *Keio Sci. Tech. Rep.*, **35**, 93–104.

Berger, J. (1975). Minimax estimation of location vectors for a wide class of densities, *Ann. Statist.*, **3**, 1318–1328.

Bhattacharya, C. G. (1981). Estimation of a common mean, *Comm. Statist. Theory Methods*, **10**, 955–961.

Casella, G. (1980). Minimax ridge regression estimation, *Ann. Statist.*, **8**, 1036–1056.

Casella, G. (1985). Condition number and minimax ridge regression estimators, *J. Amer. Statist. Assoc.*, **80**, 753–758.

Cellier, D., Fourdrinier, D. and Robert, C. (1989). Robust shrinkage estimators of the location parameter for elliptically symmetric distributions, *J. Multivariate Anal.*, **29**, 39–52.

Cohen, A. (1976). Combining estimates of location, *J. Amer. Statist. Assoc.*, **71**, 172–175.

Fraisse, A. M., Robert, C. and Roy, M. (1987). Estimateurs a retrecisseur matriciel differentiable, pour un covt quadratique general, *Ann. Économ. Statist.*, **8**, 161–175.

Ghosh, M., Saleh, A. K. Md. E. and Sen, P. K. (1989). Empirical Bayes subset estimation in regression models, *Statist. Decisions*, **7**, 15–35.

Graybill, F. A. and Deal, R. B. (1959). Combining unbiased estimators, *Biostatistics*, **15**, 543–550.

James, W. and Stein, C. (1961). Estimation with quadratic loss, *Proc. Fourth Berkeley Symp. on Math. Statist. Prob.*, Vol. 1, 361–379, University of California Press, Berkeley.

Khatri, C. G. and Shah, K. R. (1974). Estimation of location parameters from two linear models under normality, *Comm. Statist.*, **3**, 647–663.

Kubokawa, T. (1989). Estimation of the common mean and its applications, Technical Report Series of the Laboratory for Research in Statistics and Probability, No. 135, Carleton University, Ottawa, Canada.

Nickerson, D. M. (1987). Sequential shrinkage estimation of linear regression parameters, *Sequential Anal.*, **6**, 93–117.

Robert, C. and Saleh, A. K. Md. E. (1989). Point estimation and confidence set estimation in a parallelism model: an empirical Bayes approach, Technical Report No. 103, L.S.T.A., University of Paris VI.

Shinozaki, N. (1978). A note on estimating the common mean of k normal distributions and Stein problem, *Comm. Statist. Theory Methods*, **7**, 1421–1432.

Shinozaki, N. (1984). Simultaneous estimation of location parameters under quadratic loss, *Ann. Statist.*, **12**, 322–335.

Stein, C. (1956). Inadmissibility of the usual estimator for the mean of a multivariate normal distribution, *Proc. Third Berkeley Symp. on Math. Statist. Prob.*, Vol. 1, 197–206, University of California Press, Berkeley.

Swamy, P. A. V. B. and Mehta, J. S. (1979). Estimation of common coefficients in two regression equations, *J. Econometrics*, **10**, 1–14.