Metrika (2002) 56: 73–81

Metrika

Combining information for prediction in linear regression

K. Krishnamoorthy[1], **Brett C. Moore**[2]

[1] Department of Mathematics, University of Louisiana at Lafayette, Lafayette, Louisiana 70504, U.S.A. (e-mail: krishna@louisiana.edu)
[2] Service Sector Statistics Division, Census Bureau, Mailstop 6500, Room 2754-3, Washington, DC 20233

Abstract. This article deals with the prediction problem in linear regression where the measurements are obtained using k different devices or collected from k different independent sources. For the case of $k = 2$, a Graybill-Deal type combined estimtor for the regression parameters is shown to dominate the individual least squares estimators under the covariance criterion. Two predictors $\hat{y}_c$ and $\hat{y}_p$ are proposed. $\hat{y}_c$ is based on a combined estimator of the regression coefficient vector, and $\hat{y}_p$ is obtained by combining the individual predictors from different models. Prediction mean square errors of both predictors are derived. It is shown that the predictor $\hat{y}_p$ is better than the individual predictors for $k \geq 2$ and the predictor $\hat{y}_c$ is better than the individual predictors for $k = 2$. Numerical comparison between $\hat{y}_c$ and $\hat{y}_p$ shows that the former is superior to the latter for the case $k = 2$.

Key words: Covariance Criterion, Graybill-Deal Estimator, MINQUE, Prediction Mean Square Error

1. Introduction

The problem of combining the results of independent investigations for the purpose of a common objective has been considered in numerous papers in various contexts. This problem arises, for example, when different methods or laboratories have been used to measure the amount of carbon monoxide in the upper atmosphere and it is desired to pool the results since measurements are difficult or expensive to obtain. This type of problem is known in the biological and social sciences as meta-analysis. A practical example is given in Skinner (1991) where different laboratory methods are used to estimate the density of nitrogen in clinical trials. Another example is from Eberhardt et al. (1989) where the results from several experiments are combined for an estimate of the selenium level in non-fat milk.

In this article, we are interested in the use of combined data for prediction in a linear regression problem. Suppose that we have from independent sources the k experimental models

$$Y_i = X_i\beta + \varepsilon_i \quad i = 1, \ldots, k, \tag{1.1}$$

and the prediction model

$$y_0 = x_0\beta + \varepsilon, \tag{1.2}$$

where the Y_i's are $n_i \times 1$ vectors, the X_i's are $n_i \times p$ fixed matrices, β is a $p \times 1$ unknown common parameter vector, the random errors ε_i's are independent with $\varepsilon_i \sim N_{n_i}(0, \sigma_i^2 I)$, x_0 is a $1 \times p$ known vector observation of a future unit, y_0 is the unknown dependent variable of the future unit, and $\varepsilon \sim N(0, \sigma^2)$ independently of ε_i, $i = 1, \ldots, k$. We note that the error variance σ^2 of model (1.2) need not be equal to σ_i^2 for any i, and all the experimental and calibration models have the same vector of regression parameters β.

Let $\hat{\beta}_i$ denote the least squares estimators of β, and s_i^2 denote the unbiased estimator (based on the residual sum of squares with $n_i - p$ degrees of freedom) of σ_i^2 based on the i th model, $i = 1, \ldots, k$. Many authors have considered the problem of estimating the common parameter β by combining the $\hat{\beta}_i$'s. Rao and Subrahmaniam (1971) considered simple linear regression models and proposed the weighted least squares estimator with weights depending on the MINQUE (Minimum Norm Quadratic Unbiased Estimator) of σ_i^2 instead of the usual unbiased estimator s_i^2. Their estimator is more efficient than the usual weighted least squares estimator when all n_i are equal and small (less than 8). Shinozaki (1978) considered the problem of estimating a linear parametric function $a'\beta$ under the models in (1.1). He showed that a combined estimator $\widehat{(a\beta)}$ (which is obtained by combining $a'\hat{\beta}_1, \ldots, a'\hat{\beta}_k$) dominates each $a'\hat{\beta}_i$. He also mentioned that an estimator of the form $a'\hat{\beta}_*$, where $\hat{\beta}_*$ is a combined estimator of β, would perform better than $\widehat{(a\beta)}$. Kubokawa (1990) and Kubokawa et al. (1991) considered this problem in a decision theoretic setup, and studied the minimaxity and admissibility of some combined estimators under different loss functions.

Although the problem of estimation of β has been considered by many authors, the important problem of prediction in this context has not been addressed in the literature. Johnson and Krishnamoorthy (1995), and Mathew and Sharma (1999) considered the problem of combining information in the context of calibration where the interest is to estimate x_0 based on known y_0. A combined estimator of β may be better than another estimator of β under a criterion, but the prediction of y values based on the former need not be better than the prediction based on the latter in some sense. Suppose that for a future observation the x value is known to be x_0 and the y value is unknown, the problem is to predict the unknown y value y_0 based on the fitted models and x_0. A suitable measure of predictive variability is predictive mean square error (PMSE), which for a predictor $\hat{y}_0$ of y_0 is $E[(\hat{y}_0 - y_0)^2]$ where E denotes the expectation. The problem of interest here is to combine data so that the PMSE will be minimum.

In Section 2, we first present a combined estimator $\hat{\beta}_c$ which is obtained by combining $\hat{\beta}_i$'s with weights "proportional" to the inverse of $\widehat{\mathrm{Cov}(\hat{\beta}_i)}$ (that is $s_i^{-2}X_i'X_i$). For the case of $k = 2$, we show that the combined estimator dominates each $\hat{\beta}_i$ under the covariance criterion. That is, $\mathrm{Cov}(\hat{\beta}_i) - \mathrm{Cov}(\hat{\beta}_c)$ is positive definite for $i = 1, 2$. We then, in Section 3, propose two combined predictors $\hat{y}_p$ and $\hat{y}_c$ for an unknown y_0. The predictor $\hat{y}_p$ is a linear combination of the predictors based on each sample with weights proportional to the inverse of each predictor's sample variance. The predictor $\hat{y}_c$ is based on the combined estimator $\hat{\beta}_c$. Expressions for the PMSE of $\hat{y}_p$ and the PMSE for $\hat{y}_c$ are derived. Using these expressions for the PMSEs, we show that $\hat{y}_p$ is better than each of the individual predictor $\hat{y}_{0i}$ for $k \geq 2$, and $\hat{y}_c$ is better than $\hat{y}_{0i}$ for $k = 2$. In Section 4, variances of $\hat{y}_c$ and $\hat{y}_p$ are derived for the case $k = 2$. Furthermore, numerical comparison studies based on the variance expressions show that $\hat{y}_c$ is better than $\hat{y}_p$ for the case of $k = 2$. Some concluding remarks are given in Section 5.

2. A combined estimator of β

Let $V_i = (X_i'X_i)^{-1}$, and $m_i = n_i - p$, $i = 1, \ldots, k$. A combined unbiased estimator of β is given by

$$\hat{\beta}_c = \left(\sum_{j=1}^{k} V_j^{-1} s_j^{-2} \right)^{-1} \sum_{i=1}^{k} (V_i^{-1} s_i^{-2} \hat{\beta}_i). \tag{2.1}$$

We recall some useful results in the following lemma.

Lemma 2.1. *For $i = 1, \ldots, k$,*

(i) $\hat{\beta}_i \sim N_p(\beta, \sigma_i^2 (X_i'X_i)^{-1})$,
(ii) $(\hat{\beta}_i - \beta)'(X_i'X_i)(\hat{\beta}_i - \beta) \sim \sigma_i^2 \chi_p^2$,
(iii) $m_i s_i^2 / \sigma_i^2 \sim \chi_{m_i}^2$, *and*
(iv) $\hat{\beta}_1, \ldots, \hat{\beta}_k$, $s_1^2, \ldots, s_k^2$ *are all statistically independent.*

From the above lemma we have s_i^2's and $\hat{\beta}_i$'s are all independent and $E(\hat{\beta}_i) = \beta$ for $i = 1, \ldots, k$. Therefore, we have $E(\hat{\beta}_c \mid s_1^2, \ldots, s_k^2) = \beta$. That is, $\hat{\beta}_c$ is an unbiased estimator of β. Furthermore, for the case $k = 2$, we show in the following theorem that $\hat{\beta}_c$ dominates each $\hat{\beta}_i$ under the covariance criterion.

Theorem 2.1. *For the case $k = 2$, the combined estimator*

$$\hat{\beta}_c = (s_1^{-2} V_1^{-1} + s_2^{-2} V_2^{-1})^{-1} (V_1^{-1} s_1^{-2} \hat{\beta}_1 + V_2^{-1} s_2^{-2} \hat{\beta}_2)(s_1^{-2} V_1^{-1} + s_2^{-2} V_2^{-1})^{-1}$$

dominates both $\hat{\beta}_1$ and $\hat{\beta}_2$ under the covariance criterion provided $m_1 \geq 10$ and $m_2 \geq 10$ or $m_1 = 9$ and $m_2 \geq 18$, where $m_i = n_i - p$, $i = 1, 2$.

Proof. The covariance of $\hat{\beta}_c$ is given by

$$\begin{aligned} \mathrm{Cov}(\hat{\beta}_c) &= E[\mathrm{Cov}(\hat{\beta}_c \mid s_1^2, s_2^2)] \\ &= E[(s_1^{-2} V_1^{-1} + s_2^{-2} V_2^{-1})^{-1} (\sigma_1^2 s_1^{-4} V_1^{-1} + \sigma_2^2 s_2^{-4} V_2^{-1}) \\ &\quad \times (s_1^{-2} V_1^{-1} + s_2^{-2} V_2^{-1})^{-1}] \end{aligned} \tag{2.2}$$

Since V_1^{-1} and V_2^{-1} are positive definite matrices, there exists a nonsingular matrix C such that $V_1^{-1} = CC'$ and $V_2^{-1} = CD_\lambda C'$, where $D_\lambda = \mathrm{diag}(\lambda_1, \dots, \lambda_p)$ with $\lambda_1 \geq \cdots \geq \lambda_p$, and λ_i's are the eigenvalues of $V_1 V_2^{-1}$ (see Anderson 1958, p. 341). Using these results, (2.2) can be written as

$$\begin{aligned}\mathrm{Cov}(\hat{\beta}_c) &= C'^{-1} E[(s_1^{-2} I_p + s_2^{-2} D_\lambda)^{-1}(\sigma_1^2 s_1^{-4} I_p + \sigma_2^2 s_2^{-4} D_\lambda) \\ &\qquad \times (s_1^{-2} I_p + s_2^{-2} D_\lambda)^{-1}] C^{-1} \\ &= \sigma_1^2 C'^{-1} \,\mathrm{diag}(\delta_1, \dots, \delta_p) C^{-1}, \end{aligned} \tag{2.3}$$

where I_p denotes the identity matrix of order p, and

$$\delta_i = E\left[\frac{(\sigma_1^2 s_1^{-4} + \lambda_i \sigma_2^2 s_2^{-4})}{(s_1^{-2} + \lambda_i s_2^{-2})^2}\right], \quad i = 1, \dots, p. \tag{2.4}$$

Since $\mathrm{Cov}(\hat{\beta}_1) = \sigma_1^2 V_1 = \sigma_1^2 C'^{-1} C^{-1}$, it follows from (2.3) and (2.4) that $\mathrm{Cov}(\hat{\beta}_1) - \mathrm{Cov}(\hat{\beta}_c)$ is positive definite provided $\delta_i/\sigma_1^2 - 1 < 0$ for $i = 1, \dots, p$. If $\lambda_i = 1$ for a fixed i, then the results of Khatri and Shah (1974) can be used to show that $\delta_i/\sigma_1^2 - 1 < 0$. Since λ_i's are arbitrary positive numbers, we will take a slightly different approach. For a fixed i, let $\theta = \lambda_i \sigma_1^2/\sigma_2^2$ and $F = (s_1^2/\sigma_1^2)/(s_2^2/\sigma_2^2)$. In terms of θ and F, we see that

$$\delta_i = \sigma_1^2 E\left[\frac{1 + \theta F^2}{(1 + \theta F)^2}\right]. \tag{2.5}$$

For $\theta > 0$, it can be easily verified that

$$\begin{aligned}\frac{1 + \theta F^2}{(1 + \theta F)^2} &< \frac{1}{1 + \theta} + \frac{\theta}{1 + \theta}(F - 1)^2 \\ &= 1 + \frac{\theta}{1 + \theta}(F^2 - 2F). \end{aligned} \tag{2.6}$$

In view of (2.5) and (2.6), we see that $\delta_i/\sigma_1^2 - 1 < 0$ if $E(F^2) \leq 2E(F)$. Using the fact that F follows an F distribution with the numerator df $= m_1$ and denominator df $= m_2$, it can be readily verified that $E(F^2) \leq 2E(F)$ under the conditions on m_1 and m_2 stated in the statement of the theorem. Thus we complete the proof.

Since $\mathrm{Cov}(\hat{\beta}_i) - \mathrm{Cov}(\hat{\beta}_c)$ is positive definite, we have

$$\mathrm{trace}(\mathrm{Cov}(\hat{\beta}_c)) = \sum_{j=1}^{p} \mathrm{Var}(\hat{\beta}_{cj}) < \sum_{j=1}^{p} \mathrm{Var}(\hat{\beta}_{ij}) = \mathrm{trace}(\mathrm{Cov}(\hat{\beta}_i)), \quad i = 1, 2,$$

and

$$|\mathrm{Cov}(\hat{\beta}_c)| < |\mathrm{Cov}(\hat{\beta}_i)|, \quad i = 1, 2,$$

where $|.|$ denotes the determinant of a matrix (see Harville 1997, p. 418). That is, the combined estimator $\hat{\beta}_c$ is also better than the individual estimators $\hat{\beta}_1$ and $\hat{\beta}_2$ under the so called trace criterion and the determinant criterion provided the degrees of freedoms m_i's satisfy the conditions given in Theorem 2.1. Furthermore, Krishnamoorthy and Sarkar (1993) showed that if an estimator is better than another estimator under the covariance criterion, then the same relation holds between these two estimators under the quadratic loss function $(\hat{\theta}-\theta)'Q(\hat{\theta}-\theta)$, where Q is a known positive definite matrix. Therefore, $\hat{\beta}_c$ is also better than $\hat{\beta}_1$ and $\hat{\beta}_2$ under this quadratic loss function if m_i's satisfy the conditions given in Theorem 2.1.

Shinozaki (1978) proposed a combined estimator $\hat{\beta}_s$ which is similar to $\hat{\beta}_c$ in (2.1). He also claimed (without a proof) that $\hat{\beta}_s$ may dominate each $\hat{\beta}_i$ under the covariance criterion provided all the V_i's are simultaneously diagonalizable. Under such condition, our estimator $\hat{\beta}_c$ can also be shown to dominate each $\hat{\beta}_i$, $i=1,\dots,k$. This can be proved along the lines of Theorem 2.1, and using the inequalities of Bhattacharya (1984).

3. Two combined predictors

We consider two predictors for y_0. The first one, denoted by $\hat{y}_c$, is based on the combined estimator $\hat{\beta}_c$ in (2.1) and the second one, denoted by $\hat{y}_p$, is a linear combination of the individual predictors $\hat{y}_{0i}=x_0'\hat{\beta}_i$, $i=1,\dots,k$. Specifically, we consider

$$\hat{y}_c = x_0'\hat{\beta}_c, \tag{3.1}$$

where $\hat{\beta}_c$ is given in (2.1).

It follows from Lemma 2.1(i) and (iv) that $\hat{y}_{0i}\sim N(x_0'\beta,(x_0'(X_i'X_i)^{-1}x_0)\sigma_i^2)$ and $\hat{y}_{0i}$'s are independent. Furthermore, we have unbiased estimator s_i^2 of σ_i^2 for $i=1,\dots,k$. Then our Graybill-Deal type combined predictor of y_0 is defined by

$$\hat{y}_p = \sum_{i=1}^{k} w_i\hat{y}_{oi}, \tag{3.2}$$

where $w_i = c_i^{-1}/\sum_{j=1}^{k} c_j^{-1}$ and $c_i = x_0'(X_i'X_i)^{-1}x_0 s_i^2$, $i=1,\dots,k$.

We shall now show that PMSE of $\hat{y}_c$ is smaller than the PMSE of $\hat{y}_{0i}$ for $i=1,2$. Noting that $\hat{\beta}_c$ is an unbiased estimator of β, it can be easily checked that the PMSE of $\hat{y}_c = x_0'\hat{\beta}_c$ is

$$\mathrm{PMSE}(\hat{y}_c) = \sigma^2 + x_0'\,\mathrm{Cov}(\hat{\beta}_c)x_0 = \sigma^2 + \mathrm{Var}(\hat{y}_c). \tag{3.3}$$

The PMSE of $\hat{y}_{oi} = x_0'\hat{\beta}_i$ is

$$\mathrm{PMSE}(\hat{y}_{oi}) = \sigma^2 + x_0'\,\mathrm{Cov}(\hat{\beta}_i)x_0 = \sigma^2 + \mathrm{Var}(\hat{y}_{0i}) \tag{3.4}$$

for $i=1,\dots,k$. Since $\mathrm{Cov}(\hat{\beta}_i)-\mathrm{Cov}(\hat{\beta}_c)$ is positive definite for $i=1,2$, we have $x_0'[\mathrm{Cov}(\hat{\beta}_c)-\mathrm{Cov}(\hat{\beta}_i)]x_0 = x_0'\,\mathrm{Cov}(\hat{\beta}_c)x_0 - x_0'\,\mathrm{Cov}(\hat{\beta}_i)x_0 > 0$, and hence it follows from (3.3) and (3.4) that $\mathrm{PMSE}(\hat{y}_c)\le\mathrm{PMSE}(\hat{y}_{oi})$ for $i=1,2$ provided $m_1\ge 10$ and $m_2\ge 10$ or $m_1=9$ and $m_2\ge 18$.

We now show that $\hat{y}_p$ has smaller PMSE than $\hat{y}_{0i}$ for $i = 1, \ldots, k$. Since $E(\hat{y}_p) = E[\sum_{i=1}^k w_i x_0'\beta] = x_0'\beta = E(y_0)$, the PMSE of $\hat{y}_p$ is

$$\begin{aligned}\mathrm{PMSE}(\hat{y}_p) &= E(\hat{y}_p - x_0'\beta - \varepsilon_0)^2 \\ &= \sigma^2 + E(\hat{y}_p - x_0'\beta)^2 = \sigma^2 + \mathrm{Var}(\hat{y}_p).\end{aligned} \tag{3.5}$$

It is clear from (3.4) and (3.5) that the $\mathrm{PMSE}(\hat{y}_p) \le \mathrm{PMSE}(\hat{y}_{0i})$ if and only if $\mathrm{Var}(\hat{y}_p) \le \mathrm{Var}(\hat{y}_{0i})$, $i = 1, \ldots, k$. Since $\hat{y}_{01}, \ldots, \hat{y}_{0k}$ are independent normal random variables with common mean $x_0'\beta$, and $\hat{y}_p$ is a Graybill-Deal type combined predictor, $\hat{y}_p$ is better than each $\hat{y}_{0i}$ provided $m_i \ge 10$ for $i = 1, \ldots, k$ (see Norwood and Hinkelman 1977). When $k = 2$ this result also holds if $m_1 = 9$ and $m_2 \ge 18$ (see Khatri and Shah 1974).

It is to be noted that $\hat{y}_c$ is better than each of the $\hat{y}_{0i}$'s only for $k = 2$ whereas $\hat{y}_p$ is better than each of the $\hat{y}_{0i}$'s for any $k \ge 2$.

4. Comparison between $\hat{y}_c$ and $\hat{y}_p$

Although these two predictors $\hat{y}_c$ in (3.1) and $\hat{y}_p$ in (3.2) are better than $\hat{y}_{01}$ and $\hat{y}_{02}$ under some conditions on the m_i's, comparison between them has not been done. We will see that $\hat{y}_c$ does offer improved performance over $\hat{y}_p$ for the case $k = 2$ and will look at the nature of the improvement.

From (3.3) and (3.5) we have $\mathrm{PMSE}(\hat{y}_c) - \mathrm{PMSE}(\hat{y}_p) = \mathrm{Var}(\hat{y}_c) - \mathrm{Var}(\hat{y}_p)$. So, to compare $\mathrm{PMSE}(\hat{y}_p)$ and $\mathrm{PMSE}(\hat{y}_c)$ it is sufficient to compare $\mathrm{Var}(\hat{y}_p)$ and $\mathrm{Var}(\hat{y}_c)$. Exact expressions for $\mathrm{Var}(\hat{y}_p)$ and $\mathrm{Var}(\hat{y}_c)$ for $k = 2$ may be obtained using the approach of Nair (1980) as follows.

Variances of $\hat{y}_p$ and $\hat{y}_c$

For $k = 2$, let $V_i = (X_i'X_i)^{-1}$, $a_i = x_0'V_i x_0$, and $m_i = n_i - p$ for $i = 1, 2$. From (3.2), we have

$$\hat{y}_p = \frac{a_2 s_2^2 \hat{y}_{o1} + a_1 s_1^2 \hat{y}_{o2}}{a_1 s_1^2 + a_2 s_2^2}. \tag{4.1}$$

Using the fact that $\hat{y}_{0i}$'s are independent among themselves and are independent of s_i^2's, we have

$$\begin{aligned}\mathrm{Var}(\hat{y}_p) &= E\left[\frac{a_2^2 a_1 s_2^4 \sigma_1^2 + a_1^2 a_2 s_1^4 \sigma_2^2}{(a_1 s_1^2 + a_2 s_2^2)^2}\right] \\ &= a_1 \sigma_1^2 E\left[\frac{1 + \delta m_2 v_1^2/(m_1 v_2^2)}{(1 + \delta v_1/v_2)^2}\right],\end{aligned} \tag{4.2}$$

where $\delta = \sigma_1^2 m_2 (x_0' V_1 x_0)/[\sigma_2^2 m_1 (x_0' V_2 x_0)]$, $v_i = m_i s_i^2/\sigma_i^2 \sim \chi^2_{m_i}$, $i = 1, 2$. Since v_1 and v_2 are independent, we can write $v_1/v_2 = w/(1 - w)$, where w is a $\mathrm{Beta}(m_1/2, m_2/2)$ random variable. In these notations, we have

$$\mathrm{Var}(\hat{y}_p) = a_1\sigma_1^2 E\left[\frac{(1-w)^2 + \delta m_2 w^2/m_1}{(1-(1-\delta)w)^2}\right]. \tag{4.3}$$

If $\delta < 1$, then using the Taylor series expansion and then taking term by term expectations, we get

$$\mathrm{Var}(\hat{y}_p) = (x_0' V_1 x_0)\sigma_1^2 \sum_{i=0}^{\infty} \frac{(i+1)(1-\delta)^i}{B(m_1/2, m_2/2)} \times \left[B\left(\frac{m_1}{2}+i, \frac{m_2}{2}+2\right) + \frac{m_2}{m_1}\delta B\left(\frac{m_1}{2}+i+2, \frac{m_2}{2}\right)\right]. \tag{4.4}$$

This series converges provided $\delta < 1$ for all $x_0 \neq 0$. Let λ_1 denote the largest eigenvalue of $V_1V_2^{-1}$. Since $(x_0'V_1x_0)/(x_0'V_2x_0) < \lambda_1$ for every $x_0 \neq 0$, from the definition of δ it follows that the series (4.4) converges for every $x_0 \neq 0$ if $\eta_1 = \lambda_1\sigma_1^2 m_2/(\sigma_2^2 m_1) < 1$. We now derive an expression for $\mathrm{Var}(\hat{y}_c)$ for the case of $k = 2$.

Noticing that $\mathrm{Var}(\hat{y}_c) = x_0' \mathrm{Cov}(\hat{\beta})x_0$ and using the expression for $\mathrm{Cov}(\hat{\beta}_c)$ given in (2.2), we get

$$\begin{aligned}\mathrm{Var}(\hat{y}_c) &= x_0'C'^{-1}E[(s_1^{-2}I_p + s_2^{-2}D_\lambda)^{-1}(\sigma_1^2 s_1^{-4}I_p + \sigma_2^2 s_2^{-4}D_\lambda) \\ &\qquad \times (s_1^{-2}I_p + s_2^{-2}D_\lambda)^{-1}]C^{-1}x_0 \\ &= \sigma_1^2 x_0' C'^{-1}\,\mathrm{diag}(\delta_1, \ldots, \delta_p)C^{-1}x_0, \end{aligned} \tag{4.5}$$

where I_p denotes the identity matrix of order p and δ_j's are given in (2.4). Another expression for δ_j can be derived in a manner similar to the derivation of (4.3) and is given by

$$\delta_j = E\left[\frac{(1-w)^2 + \eta_j m_2 w^2/m_1)}{[1-(1-\eta_j)w]^2}\right], \tag{4.6}$$

where $\eta_j = \lambda_j\sigma_1^2 m_2/(\sigma_2^2 m_1)$ for $j = 1, \ldots, p$. Further, if $\eta_j < 1$ for all j, then

$$\delta_j = \sum_{i=0}^{\infty} \frac{(i+1)(1-\eta_j)^i}{B(m_1/2, m_2/2)}\left[B\left(\frac{m_1}{2}+i, \frac{m_2}{2}+2\right) + \frac{m_2}{m_1}\eta_j B\left(\frac{m_1}{2}+i+2, \frac{m_2}{2}\right)\right]. \tag{4.7}$$

The series (4.7) converges if $\eta_1 < 1$ which is the same condition for the series expression of $\mathrm{Var}(\hat{y}_p)$ in (4.4) to converge.

The series expressions for $\mathrm{Var}(\hat{y}_p)$ and $\mathrm{Var}(\hat{y}_c)$ are useful to compute the PMSEs of the predictors $\hat{y}_p$ and $\hat{y}_c$; however, they are of limited use to compare $\hat{y}_p$ and $\hat{y}_c$ because they converge only under some restrictions on the parameter space and the matrices V_1 and V_2. Therefore, we use the expressions (4.3) and (4.6) for comparison because they can be evaluated numerically for all parameters, V_1 and V_2.

When all the eigenvalues of $V_1V_2^{-1}$ are the same or equivalently $V_1 = cV_2$ for a positive constant c, then the predictors $\hat{y}_p$ and $\hat{y}_c$ are identical and hence no comparison is needed. When this condition is not met, for comparison purpose we can assume that the σ_i^2's are equal and the m_i's are equal. Let $Z = (z_1, \ldots, z_p)' = C^{-1}x_0$ and $q_i = z_i^2/(Z'Z)$, $i = 1, \ldots, p$. Then

$$\mathrm{Var}(\hat{y}_p) = \sigma_1^2 (Z'Z) E\left[\frac{(1-w)^2 + w^2/(\sum_{i=1}^p q_i/\lambda_i)}{(1 - w + w/(\sum_{i=1}^p q_i/\lambda_i))^2}\right], \tag{4.8}$$

and

$$\mathrm{Var}(\hat{y}_c) = \sigma_1^2 \sum_{i=1}^p z_i^2 E\left[\frac{(1-w)^2 + \lambda_j w^2)}{(1 - w + \lambda_j w)^2}\right]. \tag{4.9}$$

Thus,

$$\frac{\mathrm{Var}(\hat{y}_c)}{\mathrm{Var}(\hat{y}_p)} = \frac{\sum_{i=1}^p q_i E\left[\frac{(1-w)^2 + \lambda_j w^2)}{(1 - w + \lambda_j w)^2}\right]}{E\left[\frac{(1-w)^2 + w^2/(\sum_{i=1}^p q_i/\lambda_i)}{(1 - w + w/(\sum_{i=1}^p q_i/\lambda_i))^2}\right]}. \tag{4.10}$$

Noting that w is a Beta$(m_1/2, m_2/2)$ random variable, the expectations in (4.10) can be evaluated numerically. We computed the variance ratio using the Simpson's rule when $p = 2$ and 3, and for various values of λ's, q_1 and q_2. They are given in Table 1. The values in Table 1 indicate that $\hat{y}_c$ has smaller variance than $\hat{y}_p$ when the eigenvalues λ's are different or equivalently when the design matrices are different. Recall that $\hat{y}_c$ is known to be better than $\hat{y}_{0i}$'s only for $k = 2$ whereas $\hat{y}_p$ is better than $\hat{y}_{0i}$'s for $k \geq 2$; however, when $k = 2$, $\hat{y}_c$ is preferable to $\hat{y}_p$ as a predictor of y_0.

Table 1. $\mathrm{Var}(\hat{y}_c)/\mathrm{Var}(\hat{y}_p)$

q_1	$p = 2$, (λ_1, λ_2)				$(1, 100)$
	$(1, 2)$	$(1, 5)$	$(1, 10)$	$(1, 20)$	
.1	.99	.91	.82	.74	.61
.3	.97	.87	.79	.73	.67
.5	.97	.89	.83	.79	.76
.7	.98	.92	.89	.87	.85
.9	.99	.97	.96	.96	.95
q_1	$p = 3$, $(\lambda_1, \lambda_2, \lambda_3)$				$(1, 8, 16)$
	q_2	$(1, 2, 3)$	$(1, 2, 7)$	$(1, 2, 15)$	
.1	.1	.96	.88	.79	.77
.1	.3	.97	.89	.83	.80
.2	.4	.96	.90	.86	.78
.3	.3	.95	.89	.85	.77
.5	.3	.96	.93	.91	.83

5. Concluding remarks

We showed in this article that the Graybill-Deal combined estimator $\hat{\beta}_c$ dominates the individual regression estimators $\hat{\beta}_i$ under various criteria for the case of $k = 2$. In addition, for the case of $k = 2$, our numerical study showed that the predictor $\hat{y}_c$ based on $\hat{\beta}_c$ has lower PMSE than the combined predictor $\hat{y}_p$. For $k \geq 3$, it is difficult to evaluate an expression for the variance of $\hat{y}_c$. If the design matrices $X_1'X_1, \ldots, X_k'X_k$ are simultaneously diagonalizable, then an expression for the variance of $\hat{y}_c$ can be obtained, and it can be shown that $\hat{y}_c$ is superior to the individual predictors $\hat{y}_{0i}$'s. This simultaneous diagonalizability requirement appears to be impractical. For arbitrary design matrices, it is difficult to prove theoretically that $\hat{y}_c$ dominates $\hat{y}_i$, $i = 1, \ldots, k$. However, our preliminary Monte Carlo simulation studies (for $k = 3$ and 4; not reported here) indicated that $\hat{y}_c$ is not only better than $\hat{y}_{0i}$ but also superior to $\hat{y}_p$ for many parameter configurations. Therefore, we conjecture that the predictor $\hat{y}_c$ is superior to the predictor $\hat{y}_p$ in general.

Acknowledgement. The authors are thankful to two reviewers for their valuable comments and suggestions.

References

Anderson TW (1958) An introduction to multivariate statistical analysis. Wiley, New York

Bhattacharya CG (1984) Two inequalities with an application. Annals of the Institute of Statistical Mathematics 36:129–134

Eberhardt KR, Reeve CP, Spiegelman CH (1989) A minimax approach to combine means, with practical examples. Chemometrics and Intelligent Laboratory Systems 5:129–148

Graybill FA, Deal RB (1959) Combining unbiased estimators. Biometrics 15:543–550

Harville DA (1997) Matrix algebra from a statistician's perspective. Springer, New York

Johnson D, Krishnamoorthy K (1996) Combining independent studies in a calibration problem. Journal of the American Statistical Association 91:1707–1715

Khatri CG, Shah KR (1974) Estimation of location parameters from two linear models under normality. Communications in Statistics 3:647–663

Krishnamoorthy K, Sarkar S (1993) Simultaneous estimation of independent normal mean vectors with unknown covariance matrices. Journal of Multivariate Analysis 47:329–338

Kubokawa T (1990) Minimax estimation of the common coefficients of several regression models under quadratic loss. Journal of Statistical Planning and Inference 24:337–345

Kubokawa T, Robert C, Saleh AK Md. E (1991) Robust estimation of common regression coefficients under spherical symmetry. Annals of the Institute of Statistical Mathematics 43:677–688

Mathew T, Sharma MK (1999) On the construction of multiple use confidence regions based on combined information in univariate calibration. To appear in Journal of Statistical Planning and Inference

Nair AK (1980) Variance and distribution of the Graybill-Deal estimator of the common mean of two normal populations. The Annals of Statistics 8:212–216

Norwood TE, Hinkelmann K (1977) Estimating the common mean of several normal populations. The Annals of Statistics 5:1047–1050

Rao JNK, Subrahmaniam K (1971) Combining independent estimators and estimation in linear regression with unequal variances. Biometrics 27:971–990

Shinozaki N (1978) A note on estimating the common mean of k normal distributions and the stein problem. Communication in Statistics—Theory and Methods A 7:1421–1432

Skinner JB (1991) On combining studies. Drug Information Journal 25:395–403