Statistica Sinica 2(1992), 495-518

EMPIRICAL BAYES ESTIMATION OF HETEROSCEDASTIC VARIANCES

Jun Shao

University of Ottawa

Abstract: Based on data resampling techniques, two classes of empirical Bayes estimators are proposed for estimating the error variances in a heteroscedastic linear model. We concentrate primarily on the situation in which only a few replicates are available at each design point but the total number of observations n is relatively large. Properties of the empirical Bayes estimators, including invariance, robustness, consistency, asymptotic unbiasedness and mean squared error (MSE), are studied. In particular, a second order expansion of the MSE and an upper bound on the bias of the empirical Bayes estimator are given in terms of the diagonal elements of the projection matrix. Using these results, we compare the empirical Bayes estimator with other existing variance estimators. The MSE of the empirical Bayes estimator is smaller than that of the within-group sample variance and the MINQUE when n is large. Applications in inferences are also discussed. Some simulation results are presented.

Key words and phrases: Data resampling, empirical Bayes estimators, sample variance, MINQUE, consistency, bias, mean squared error, weighted least squares.

1. Introduction

In statistical applications, the following linear model is widely used:

$$y_{ij} = x_i^t \beta + e_{ij}, \quad j = 1, \ldots, m_i, \quad i = 1, \ldots, k, \quad \sum_i m_i = n, \qquad (1.1)$$

where y_{ij} is the response of the jth replicate in the ith group, x_i is a $p \times 1$ deterministic design vector and x_i^t is its transpose, β is a $p \times 1$ vector of parameters, and e_{ij} are mutually independent with means zero and variances σ_i^2, $j = 1, \ldots, m_i$. The σ_i^2 are unknown and different (heteroscedastic).

Although in most situations the parameter of primary interest is β, the statistical accuracy of any estimator of β depends on σ_i^2. Having good estimates of σ_i^2 is necessary for judging the performances of the estimators of β and other statistical inference such as setting confidence regions for β. Furthermore, one may utilize the estimates of σ_i^2 in improving the estimates of β. In some situations estimation of individual σ_i^2 is also required.

When the number of replicates m_i is small, the use of the sample variance within the ith group as an estimator of σ_i^2 is not adequate. The case of small m_i is important since it is often impractical to obtain more than 4 or 5 replicates at a design point in the regression problem (Jacquez et al. (1968)). Usually, the number of groups k is large. Improving the within-group sample variance is possible by using data in other groups, since very often σ_i^2 have some features in common.

A considerable amount of literature on this subject can be found. Common approaches have traditionally fallen into one of the areas described below.

(i) In many practical situations σ_i^2 is a function of the design ($\sigma_i^2 = H(x_i)$) or the mean ($\sigma_i^2 = H(x_i^t\beta)$). If the function H is known up to an unknown parameter θ, one can estimate σ_i^2 through estimating θ. This is called the parametric approach. If the function H is completely unknown but smooth, one may use a nonparametric method (e.g., kernel estimation) to obtain estimates of σ_i^2. For more details of these approaches, see Carroll (1982), Muller and Stadtmuller (1987), and Davidian and Carroll (1987), which also provide many references.

(ii) In some situations the variances do not vary with the design or there is no functional relation between σ_i^2 and x_i. For example, σ_i^2 are random (see Carroll and Ruppert (1986)) and independent of x_i or have distributions depending on x_i. Heteroscedasticity may also be caused by some factor other than the design (e.g., day-to-day, person-to-person and batch-to-batch variations). In these situations nonparametric estimators of σ_i^2 such as the MINQUE (C. R. Rao (1970)) and its modifications (J. N. K. Rao (1973), Horn et al. (1975)) were proposed.

(iii) When an appropriate loss function is available, a decision theory approach can be used (Das Gupta (1986)) .

The MINQUE has the following well known deficiencies: (a) it may not exist; (b) it requires large computations; and more seriously, (c) it can be negative. Because of (c), the MINQUE is not admissible.

J. N. K. Rao (1973) proposed a modified MINQUE: the within-group average of squared residuals (ARE). The ARE has a smaller mean squared error (MSE) than the MINQUE but tends to underestimate σ_i^2 (see Sections 5 and 7). Horn et al. (1975) proposed an estimator which is called AUE by the authors. The AUE was proved to have smaller MSE than the MINQUE but under a rather unrealistic condition, i.e., one can choose correct weights (in the weighted least squares fitting model (1.1)) *before* having estimates of σ_i^2.

In this paper, we focus on the situation where the variances do not vary with the design and propose a class of estimators by using the empirical Bayesian method incorporating data resampling techniques (Section 3). The empirical

Bayes estimator is generally a compromise between a local estimate using the residuals within the ith group and an ensemble estimate. This type of estimator is often superior to the within-group sample variance because it incorporates the auxiliary information provided by the other groups.

In Section 4, we study properties of these empirical Bayes estimators, such as invariance, consistency, asymptotic unbiasedness and MSE. In particular, a second order expansion of the MSE and an upper bound on the bias of the empirical Bayes estimator are given in terms of diagonal elements of the projection (hat) matrix. Using these results, we compare the empirical Bayes estimator with other variance estimators in Section 5. The MSE of the empirical Bayes estimator is shown to be smaller than that of the within-group sample variance and the MINQUE when n is large. We also show that the ARE has the same second order MSE expansion as the empirical Bayes estimators but generally has a large negative bias. The performances of various variance estimators in the case of small n are discussed through an example.

In Section 6, applications in statistical inference by using the proposed empirical Bayes estimators of σ_i^2 are studied. Using the estimates of σ_i^2 to obtain an improved estimate of β is also discussed. Section 7 contains some simulation studies including comparisons between the empirical Bayes estimators and other variance estimators. An example is given in Section 8.

2. Bayes Estimators

In this section, we assume that the errors e_{ij} in (1.1) have a normal distribution $N(0, \sigma_i^2)$, $j = 1, \ldots, m_i$. Let $\tau_i = (2\sigma_i^2)^{-1}$. Suppose that τ_i are independently distributed as

$$\pi_i(\tau_i) \propto \tau_i^{p_i} \exp(-\alpha_i \tau_i) \quad \tau_i > 0, \quad i = 1, \ldots, k, \tag{2.1}$$

where $p_i > 0$, $\alpha_i > 0$ are known constants. For the parameter β, we only assume that β is independent of $\tau = (\tau_1 \ldots \tau_k)$ and has a known prior density $\pi(\beta)$ with respect to a measure μ on $\mathbf{R}^p$. Let s_i^2 be the sample variance within the ith group, i.e.,

$$s_i^2 = (m_i - 1)^{-1} \sum_{j=1}^{m_i} (y_{ij} - \bar{y}_i)^2, \quad \bar{y}_i = m_i^{-1} \sum_{j=1}^{m_i} y_{ij}. \tag{2.2}$$

Under squared error loss, the Bayes estimator of σ_i^2 is

$$v_i^B = (1 - \lambda_i)a_i + \lambda_i(1 - m_i^{-1})s_i^2 + \lambda_i \int (\bar{y}_i - x_i^t \beta)^2 p(\beta | y) \, d\mu, \tag{2.3}$$

where $a_i = E\sigma_i^2 = E(2\tau_i)^{-1} = \alpha_i/(2p_i)$ is the prior mean of σ_i^2, $\lambda_i = m_i/(2p_i + m_i)$ and

$$p(\beta|y) \propto \pi(\beta) \prod_{i=1}^{k} \Big[\alpha_i + \sum_{j=1}^{m_i}(y_{ij} - x_i^t\beta)^2\Big]^{-(p_i+1+m_i/2)}$$

is the posterior density of β. When $m_i = 1$, the second term on the right side of (2.3) is defined to be zero. Another form of v_i^B, which is easy to see from (2.3), is that

$$v_i^B = (1-\lambda_i)a_i + \lambda_i(1-m_i^{-1})s_i^2 + \lambda_i[(\bar{y}_i - x_i^t\hat{\beta}_B)^2 + x_i^tV_yx_i], \qquad (2.4)$$

where $\hat{\beta}_B$ and V_y are the posterior mean and variance of β, respectively. From (2.3), the Bayes eatimator is a mixture of three components: the prior information, the within-group variation of y_{ij} and a smooth average of the squared "residuals" $\bar{y}_i - x_i^t\beta$ which captures the information from fitting model (1.1). Equation (2.4) further decomposes the third term on the right side of (2.3) into a squared residual obtained by estimating β by the Bayes estimator $\hat{\beta}_B$ and the variance of the posterior $p(\beta|y)$. Using (2.4), the Bayes estimator may be approximated by estimating the first two moments of the posterior density $p(\beta|y)$.

There is no explicit form for the Bayes estimator v_i^B. Numerical integration or Monte Carlo integration is necessary in order to evaluate v_i^B. This may cause problems when p, the dimension of β, is large. The empirical Bayes estimators derived in the next section are analytically tractable and easy to evaluate.

3. Resampling Empirical Bayes Estimators

The Bayes estimator (2.3) depends on hyperparameters a_i, λ_i and the posterior density $p(\beta|y)$. If some of these quantities are unknown, they can be estimated from the data. The resulting estimators are known as empirical Bayes estimators. Two classes of Resampling Empirical Bayes Estimators (REBEs) of σ_i^2 are derived by (1) estimating the mean of the prior $\pi_i(\tau_i)$ using the method of moments and (2) estimating the mean and variance of the posterior $p(\beta|y)$ by data resampling.

3.1. Estimating the prior mean of σ_i^2

We begin with the ordinary least squares estimator (OLSE)

$$\hat{\beta} = M^{-1}X^ty,$$

where $y = (y_{11}\cdots y_{1m_1}\cdots\cdots y_{k1}\cdots y_{km_k})^t_{n\times 1}$, $X = (x_1\dots x_1\cdots\cdots x_k\dots x_k)^t_{n\times p}$, and $M = X^tX > 0$ is assumed. We use the OLSE of β instead of a weighted least squares estimator (WLSE) for the following reasons:

(i) Since σ_i^2 are unknown, choosing the appropriate weights is very difficult.

(ii) The weighted least squares method may not provide a better estimator of β even if one has estimates of σ_i^2 (Jacquez et al. (1968) and C. R. Rao (1970)).

(iii) The inverse of the weight can be thought of as a prior guess of σ_i^2. The OLSE of β is essentially an estimate based on a noninformative prior guess of σ_i^2, i.e., the weights $\equiv$ a constant.

Define the residuals by

$$r_{ij} = y_{ij} - x_i^t\hat{\beta}, \quad j = 1, \ldots, m_i, \; i = 1, \ldots, k.$$

Let $h_{il} = x_i^t M^{-1} x_l$, the (i, l)th element of the hat matrix, and $h_i = h_{ii}$. The marginal mean of r_{ij}^2 is (normality assumption is not required):

$$E_\tau E r_{ij}^2 = (1 - h_i)a_i + \sum_{l=1}^{k} h_{il}^2 m_l (a_l - a_i), \quad j = 1, \ldots, m_i, \; i = 1, \ldots, k,$$

where E_τ and E denote the expectations taken under $\pi(\tau)$ and $p(y|\tau, \beta)$, respectively. If $a_i = a$, i.e., the prior means are all equal, we have $E_\tau E r_{ij}^2 = (1 - h_i)a$, $j = 1, \ldots, m_i$. Since $\sum_i m_i(1 - h_i)a = (n - p)a$, the moment estimator of a is

$$s^2 = (n - p)^{-1} \sum_{i=1}^{k} \sum_{j=1}^{m_i} r_{ij}^2. \tag{3.1}$$

If a_i are not all equal, $E_\tau E[m_i^{-1}(1-h_i)^{-1} \sum_{j \le m_i} r_{ij}^2] = a_i + \sum_{l \le k}(1-h_i)^{-1} h_{il}^2 m_l (a_l - a_i)$, $i = 1, \ldots, k$. One could obtain unbiased estimates of a_i by solving the linear system

$$[m_i(1 - h_i)]^{-1} \sum_{j=1}^{m_i} r_{ij}^2 = a_i + \sum_{l=1}^{k} (1 - h_i)^{-1} h_{il}^2 m_l (a_l - a_i), \quad i = 1, \ldots, k. \tag{3.2}$$

However, it may not be possible to solve (3.2). Even if the solution of (3.2) exists, solving it may involve large computations. In addition, the solution $\hat{a}_i$ may be negative. Since the off-diagonal elements of the hat matrix are small relative to the h_i, we ignore the second term on the right side of (3.2) and use almost unbiased estimates

$$\hat{a}_i = [m_i(1 - h_i)]^{-1} \sum_{j=1}^{m_i} r_{ij}^2. \tag{3.3}$$

These estimators can be used even if the a_i are equal. On the other hand, the use of s^2 requires all a_i to be equal.

3.2. Estimating posterior moments of β

From (2.4), the Bayes estimator depends on the mean and variance of the posterior density $p(\beta|y)$. We now approximate these moments by the corresponding moments of a resampling distribution. This approximation is valid in some situations. For example, Lindley and Smith (1972) showed that if σ_i^2 are known and the prior of β is chosen to be noninformative, the posterior mean (variance) of β is the same as the mean (variance) of a WLSE, which is equal to or very close to the mean (variance) of a resampling distribution. In general, the approximation may have a non-negligible error. Nevertheless, we will use this approximation to obtain estimators of σ_i^2 and study their properties in the next section.

There are many different data resampling techniques in the statistical literature. We describe and use two of them. Again, the OLSE is used for the reasons given in the previous section.

(i) *Bootstrapping residuals* (Efron (1979)). For given y, let e^* be an n-vector whose components are i.i.d. samples from $\{(r_{ij} - \bar{r})/(1-p/n)^{1/2},\ j = 1,\dots,m_i,\ i = 1,\dots,k\}$, where $\bar{r} = n^{-1}\sum_{i\le k}\sum_{j\le m_i} r_{ij}$. Treat e^* as an error vector and $y^* = X\hat{\beta} + e^*$ as the observed data. The corresponding OLSE is $\beta^* = M^{-1}X^t y^*$. Denote the expectation under the bootstrap distribution (given y) by E_y^*. We replace the posterior mean and variance of β in (2.4) by the mean and variance of the bootstrap distribution, which are

$$E_y^*\beta^* = \hat{\beta} \text{ and } E_y^*(\beta^* - \hat{\beta})(\beta^* - \hat{\beta})^t = s_b^2 M^{-1},$$

respectively, where $s_b^2 = s^2 - n\bar{r}^2/(n-p)$ and s^2 is defined in (3.1). The term $n\bar{r}^2/(n-p)$ is of lower order than s^2 and is equal to zero when the first components of x_i, $i = 1,\dots,k$, are all equal to one. For simplicity we ignore this term (or simply assume there is a constant term in model (1.1)) so that $s_b^2 = s^2$. Thus, the third term on the right side of (2.4) is approximated by $\lambda_i[(\bar{y}_i - x_i^t\hat{\beta})^2 + h_i s^2]$. Assume that p_i in (2.1) are known. Hence λ_i are known and the resulting REBE of σ_i^2 is

$$v_i^b(\lambda_i) = (1-\lambda_i h_i)\hat{a}_i + \lambda_i h_i s^2, \tag{3.4}$$

which employs $\hat{a}_i$ defined in (3.3) as an estimate of a_i. Note that $v_i^b(\lambda_i)$ is a convex combination of $\hat{a}_i$ and s^2, a weighted average of $\hat{a}_i$'s.

(ii) *Weighted resampling* (Shao (1986)). We can also approximate the posterior mean and variance of β by the mean and variance of a weighted resampling distribution. For given data y, select a subset model $y_s = X_s\beta + e_s$ with probability $W_s \propto |X_s^t X_s|$, where $s = \{i_1,\dots,i_r\}$ is a subset of $\{1,\dots,n\}$, $r \le n$, y_s, X_s and e_s are sub-vector and/or sub-matrix of y, X and e consisting of the i_1th, ..., i_rth rows of y, X and e, respectively. The unequal probability W_s in the resampling procedure takes account of the unbalanced nature of the regression data. More details of this resampling procedure can be found in Shao (1986). Denote the OLSE of β under the subset model $y_s = X_s\beta + e_s$ by $\hat{\beta}_s$ and the expectation under the weighted resampling distribution (given y) by E_y^s. Then the posterior mean is approximated by $E_y^s\hat{\beta}_s = \hat{\beta}$. In order to match moments, we use a scaled variance of the resampling distribution, i.e., $(n-r)^{-1}(r-p+1)E_y^s(\hat{\beta}_s-\hat{\beta})(\hat{\beta}_s-\hat{\beta})^t$, to approximate the posterior variance of β. We focus on the computationally simplest case of $r = n-1$. The resulting REBE of σ_i^2 (again we assume λ_i are known) is

$$v_i^w(\lambda_i) = (1-\lambda_i h_i)\hat{a}_i + \lambda_i h_i s_1^2, \quad s_1^2 = h_i^{-1}\sum_{l=1}^{k} h_{il}^2 m_l \hat{a}_l. \tag{3.5}$$

Note that $v_i^w(\lambda_i)$ is also a convex combination of $\hat{a}_i$ and a weighted average of $\hat{a}_i$'s.

3.3. The hyperparameters λ_i

In the above, the hyperparameters λ_i are assumed to be known. Note that $\lambda_i = m_i/[2(SN_i+1)+m_i]$, where $SN_i = (E\sigma_i^2)^2/\text{Var}(\sigma_i^2)$ is the signal-noise ratio of the prior distribution of σ_i^2. Thus, an estimate of the signal-noise ratio, SN_i, will provide an estimate of λ_i. Note that λ_i is a decreasing function of SN_i. When SN_i (or equivalently, p_i) is large, the prior is highly concentrated on its mean a_i. Then λ_i is small and the REBE puts more weight on the estimate of the prior mean. On the other hand if SN_i (or p_i) is small, the prior is vague. Hence λ_i is large and the REBE puts less weight on the estimate of the prior mean.

Alternatively, we can let λ_i in (3.4)–(3.5) range over $[0,1]$ to obtain two classes of REBEs. Then choose a λ_i in terms of the sampling properties of the REBE under certain criteria. This is discussed in Section 4.4. Note that $0 < \lambda_i < 1$ since $0 < p_i < \infty$. But for the REBE, we can include the limiting cases $\lambda_i = 0$ and $\lambda_i = 1$.

3.4. Extensions

When some of the variances are equal, say $\sigma_i^2 = \sigma_{i_u}^2$, $i = i_u+1,\ldots,i_{u+1}$, where $0 = i_0 \le i_1 \le \cdots \le i_g = k$, the REBEs can be used with a simple modification by grouping the residuals. That is, to estimate $\sigma_{i_u}^2$, we use (3.4)–(3.5) with $\hat{a}_i$, λ_i and h_i replaced by $\sum_{i_u<i\le i_{u+1}}\sum_{j\le m_i} r_{ij}^2/(\overline{m}_u - \overline{q}_u)$, λ_{i_u} and $\overline{q}_u/\overline{m}_u$, respectively, and s_1^2 in (3.5) by $\sum_{i_u<i\le i_{u+1}} m_i \sum_{l\le k} h_{il}^2 m_l \hat{a}_l/\overline{q}_u$, where $\overline{m}_u = \sum_{i_u<i\le i_{u+1}} m_i$ and $\overline{q}_u = \sum_{i_u<i\le i_{u+1}} m_i h_i$.

The REBEs can also be used in the nonlinear regression model

$$y_{ij} = f(x_i, \beta) + e_{ij}, \quad j = 1,\ldots,m_i, \; i = 1,\ldots,k$$

with $r_{ij} = y_{ij} - f(x_i, \hat{\beta})$ and h_{ij} = the (i, j)th element of $Z(Z^tZ)^{-1}Z^t$, where $\hat{\beta}$ is the OLSE of β under the nonlinear model and $Z = X$ with x_i replaced by $\partial[f(x_i, \hat{\beta})] / \partial\beta$.

4. Properties of the REBEs

Some properties of the REBEs $v_i^b(\lambda_i)$ and $v_i^w(\lambda_i)$ with $\lambda_i \in [0, 1]$ are studied in this section. The results will be used to compare the different variance estimators.

The REBEs are based on the Bayes estimator (2.3), which is derived under assumption (2.1) and the assumption that the errors are normal. However, the results in the rest of this paper are true *with or without* these assumptions. For example, the MSE expansion given in (4.1) only depends on the design, the variances and kurtosis of the error distributions. Hence, the use of the REBEs is still justified by looking at their sampling properties when assumption (2.1) and/or the normality assumption on the errors are dropped.

We assume that $m_i \le m_\infty$ for a fixed m_∞, and n is large.

4.1. Invariance and consistency

All the REBEs obtained in Section 3 are invariant under the translation of β since they depend on the residuals r_{ij}. For each fixed i, the REBEs are not consistent (as $n \to \infty$) unless $m_i \to \infty$. In fact, when m_i is small and σ_i^2 is not smoothly related to the design, no consistent estimator of σ_i^2 is available. However it is shown in Section 6 that for estimating linear combinations of σ_i^2 (such as the variance of the OLSE), the REBEs are consistent.

4.2. The bias

For a variance estimator v_i, let $\mathrm{Bias}(v_i) = Ev_i - \sigma_i^2$. The following result gives an upper bound on the order of the magnitude of the bias of the REBE. The result implies that $v_i^b(\lambda_i)$ and $v_i^w(\lambda_i)$ are asymptotically unbiased if $h_i \to 0$ as $n \to \infty$. The quantities h_i measure the balance of the design of model (1.1). The condition $h_i \to 0$ is weak and is necessary for the asymptotic normality of the OLSE (Huber (1981)).

Theorem 1. *Let* $\lambda_i \in [0, 1]$. *Assume* $\limsup_{n\to\infty} h_{\max} < 1$, *where* $h_{\max} = \max_{i\le k} h_i$, *and* $\sup_i \sigma_i^2 < \infty$. *Then there is a constant* $c > 0$ *(independent of* i *and* n*) such that*

$$|\,\mathrm{Bias}(v_i^b(\lambda_i))\,| \le ch_i \quad \textit{and} \quad |\,\mathrm{Bias}(v_i^w(\lambda_i))\,| \le ch_i.$$

Proof. Note that $Ev_i^b(\lambda_i) = (1 - \lambda_i h_i)E\hat{a}_i + \lambda_i h_i Es^2$. From (3.3),

$$E\hat{a}_i = m_i^{-1}(1 - h_i)^{-1}\sum_{j=1}^{m_i} Er_{ij}^2 = \sigma_i^2 + (1 - h_i)^{-1}\sum_{j=1}^{k} h_{ij}^2 m_j(\sigma_j^2 - \sigma_i^2).$$

Under the given conditions, there is a constant $c > 0$ such that for sufficiently large n,

$$\left|\sum_{j=1}^{k}(1-h_i)^{-1}h_{ij}^2 m_j(\sigma_j^2-\sigma_i^2)\right| \le c\sum_{j=1}^{k} h_{ij}^2 m_j = ch_i.$$

This, together with the boundedness of σ_i^2, implies that Es^2 is bounded and $E\hat{a}_i = \sigma_i^2 + O(h_i)$. Thus, the first assertion of the theorem follows. The proof for $v_i^w(\lambda_i)$ is similar.

4.3. The MSE

The exact form of the MSE of the REBE is complicated due to the nonidentical distributions of the errors. The following theorem gives asymptotic ($n \to \infty$) expansions of the MSE of the REBEs.

Theorem 2. *Assume that* $\limsup_{n\to\infty} h_{\max} < 1$ *and* $\rho_i = \mathrm{Var}(e_{ij}^2)$, $j = 1,\ldots,m_i$, $i = 1,\ldots,k$, *are bounded. Then for any* $\lambda_i \in [0,1]$,

$$\mathrm{MSE}(v_i^b(\lambda_i)) = m_i^{-1}(1-h_i)^{-2}(1-\lambda_i h_i)^2[\rho_i + O(h_i)] + O(h_i h_{\max}). \tag{4.1}$$

The same result holds if $v_i^b(\lambda_i)$ is replaced by $v_i^w(\lambda_i)$.

Remark. From the proof of Theorem 2, the expansion in (4.1) holds uniformly in i, i.e., there is an absolute constant $c > 0$ (independent of i and n) such that $O(h_i)$ and $O(h_i h_{\max})$ in (4.1) are bounded in absolute value by ch_i and $ch_i h_{\max}$, respectively.

Proof. We first show that there is an absolute constant $c > 0$ such that if $(i, j) \ne (t, r)$,

$$|\mathrm{Cov}(r_{ij}^2, r_{tr}^2)| \le ch_i h_t \quad \text{and} \quad |\mathrm{Var}(r_{ij}^2) - \rho_i| \le ch_i. \tag{4.2}$$

Let $u_{ijls} = 1 - h_i$ if $l = i$ and $j = s$, and $u_{ijls} = -h_{il}$ otherwise. Then

$$\mathrm{Cov}(r_{ij}^2, r_{tr}^2) = \sum_{l=1}^{k}\sum_{s=1}^{m_l} u_{ijls}^2 u_{trls}^2 \rho_l + 2\sum_{(l,s)\ne(m,v)} u_{ijls}u_{ijmv}u_{trls}u_{trmv}\sigma_l^2\sigma_m^2. \tag{4.3}$$

When $(i, j) \ne (t, r)$, the first and second terms on the right side of (4.3) are bounded by $(2+p)\rho_\infty h_i h_t$ and $(2+p^2)\sigma_\infty^4 h_i h_t$, respectively, where $\rho_\infty = \sup_l \rho_l$ and $\sigma_\infty = \sup_l \sigma_l$. Hence the first assertion in (4.2) holds. Also, from (4.3),

$$\mathrm{Var}(r_{ij}^2) = [(1-h_i)^4 - h_i^4]\rho_i + \sum_{l=1}^{k} m_l h_{il}^4 \rho_l + 2\sum_{(l,s)\ne(m,v)} u_{ijls}^2 u_{ijmv}^2 \sigma_l^2 \sigma_m^2.$$

Note that

$$\sum_{l=1}^{k} m_l h_{il}^4 \rho_l \le \rho_\infty \sum_{l=1}^{k} m_l h_{il}^2 = \rho_\infty h_i$$

and

$$\sum_{(l,s)\neq(m,v)} u_{ijls}^2 u_{ijmv}^2 \sigma_l^2 \sigma_m^2 \le 2\sigma_\infty^4 \sum_{l=1}^{k} m_l h_{il}^2 + \sigma_\infty^4 \Big(\sum_{l=1}^{k} m_l h_{il}^2\Big)^2 \le 3\sigma_\infty^4 h_i.$$

This proved (4.2). From Theorem 1, the bias of $v_i^b(\lambda_i)$ is of order $O(h_i)$. Hence

$$\begin{aligned}\mathrm{MSE}(v_i^b(\lambda_i)) &= (1-\lambda_i h_i)^2 \,\mathrm{Var}(\hat{a}_i) + 2\lambda_i h_i (1-\lambda_i h_i)\,\mathrm{Cov}(\hat{a}_i, s^2) \\ &\quad + \lambda_i^2 h_i^2 \,\mathrm{Var}(s^2) + O(h_i^2).\end{aligned}$$

By (4.2), $\mathrm{Var}(s^2)$ is bounded. Therefore $\lambda_i^2 h_i^2 \,\mathrm{Var}(s^2) = O(h_i^2)$. Also, from (4.2),

$$\begin{aligned}\mathrm{Var}(\hat{a}_i) &= m_i^{-2}(1-h_i)^{-2}\Big[\sum_{j=1}^{m_i} \mathrm{Var}(r_{ij}^2) + 2\sum_{1\le j<l\le m_i} \mathrm{Cov}(r_{ij}^2, r_{il}^2)\Big] \\ &= m_i^{-1}(1-h_i)^{-2}[\rho_i + O(h_i)]\end{aligned}$$

and

$$\mathrm{Cov}(\hat{a}_i,\, s^2) = [(n-p)m_i(1-h_i)]^{-1} \sum_{j=1}^{m_i}\sum_{l=1}^{k}\sum_{s=1}^{m_l} \mathrm{Cov}(r_{ij}^2,\, r_{ls}^2) = O(h_{\max}).$$

This proves (4.1). The proof for $v_i^w(\lambda_i)$ is similar.

4.4. The choice of λ_i

A consequence of Theorem 2 is that for any $0 \le s < t \le 1$,

$$h_i^{-1}\Big[\mathrm{MSE}(v_i^b(s)) - \mathrm{MSE}(v_i^b(t))\Big] \to 2m_i^{-1}(t-s)\rho_i > 0$$

if $h_{\max} \to 0$ as $n \to \infty$. The same result holds if v_i^b is replaced by v_i^w. Thus, if $0 \le s < t \le 1$, the MSE of $v_i^b(t)$ (or $v_i^w(t)$) is less than that of $v_i^b(s)$ (or $v_i^w(s)$) when n is large enough.

However, in practice the MSE is not the only measure of the precision of an estimator. The bias of the estimator, for example, is also important in some situations. When the purpose of estimating σ_i^2 is to set a confidence interval for β, one ought not use an estimator of σ_i^2 which has a trend in its bias. If the bias of the variance estimator is always negative, the resulting confidence interval will have a too low coverage probability. See the simulation results in Section 7.

A refined analysis of the biases of the REBEs shows that the REBE with smaller λ_i usually have smaller bias.

Theorem 3. *Let* $A_n = h_i^{-1}\sum_{l\le k} h_{il}^2 m_l(\sigma_l^2 - \sigma_i^2)$, $B_n = (n-p)^{-1}\sum_{l\le k} m_l(1-h_l)(\sigma_l^2-\sigma_i^2)$, *and* $0 \le s < t \le 1$. *Assume the conditions in Theorem 2 and* $h_i \to 0$ *as* $n \to \infty$.

(i) *If* $\liminf_{n\to\infty} |A_n| > 0$, *then*

$$\liminf_{n\to\infty}\left[\frac{|\operatorname{Bias}(v_i^w(t))|}{|\operatorname{Bias}(v_i^w(s))|}\right] > 1. \tag{4.4}$$

(ii) *If* $\liminf_{n\to\infty} |B_n| > 0$ *and* $A_nB_n \ge 0$, *then* (4.4) *holds with* v_i^w *replaced by* v_i^b.

Remarks. (1) $\liminf_{n\to\infty} |A_n| > 0$ (or $\liminf_{n\to\infty} |B_n| > 0$) ensures that the biases of $v_i^w(s)$ and $v_i^w(t)$ (or $v_i^b(s)$ and $v_i^b(t)$) are comparable in terms of their first order terms.
(2) The condition $A_nB_n \ge 0$ is satisfied for some balanced models. For example, any model satisfying condition (5.4) of Wu (1986).

Proof. We prove (ii) only. The proof of (i) is similar. For $0 \le t \le 1$,

$$\operatorname{Bias}(v_i^b(t)) = t(n-p)^{-1}h_i\sum_{l=1}^{k} m_l(1-h_l)(\sigma_l^2-\sigma_i^2) + \sum_{l=1}^{k} h_{il}^2 m_l(\sigma_l^2-\sigma_i^2) + O(h_i^2).$$

Let $\xi_n = h_i^{-1}\operatorname{Bias}(v_i^b(t))$, $\gamma_n = h_i^{-1}\operatorname{Bias}(v_i^b(s))$ and $c = \liminf_{n\to\infty}|\xi_n/\gamma_n|$. If $c = \infty$, the result follows. Suppose that $c < \infty$. Since γ_n are bounded, there is a subsequence $\{n(j)\}$ such that $\lim_{j\to\infty}|\xi_{n(j)}/\gamma_{n(j)}| = c$ and the limits $\lim_{j\to\infty}\xi_{n(j)}$ and $\lim_{j\to\infty}\gamma_{n(j)}$ exist. Since $\xi_{n(j)} = A_{n(j)} + tB_{n(j)} + o(1)$ and $\gamma_{n(j)} = A_{n(j)} + sB_{n(j)} + o(1)$, the limits $A = \lim_{j\to\infty} A_{n(j)}$ and $B = \lim_{j\to\infty} B_{n(j)}$ exist. Under the conditions in (ii), $B \ne 0$ and $A/B \ge 0$. Then $c = |A+tB|/|A+sB| = |1+(t-s)(A/B+s)^{-1}| > 1$.

Hence, if one uses MSE as the measure of accuracy, then the REBE with a large λ_i is preferred. On the other hand, if one is concerned about the bias of the estimator, then the REBE with a small λ_i is better. In general, one should balance the advantage of having a smaller bias against the drawback of a larger MSE.

One may also choose a λ_i by considering the performances of the statistical procedures using the REBEs as estimates of σ_i^2 (e.g., confidence intervals and the WLSE). See Section 6. Simulation results in Section 7 favor the REBE $v_i^b(1)$ (or $v_i^w(1)$). Note that the choice of $\lambda_i = 1$ corresponds to the use of a vague prior of σ_i^2 (Section 3.3).

5. Comparisons between the REBE and Other Variance Estimators

We compare the REBE with the other variance estimators, such as the within-group sample variance, the MINQUE and the ARE, in the case where m_i are small but n is large (Section 5.1). The case where n is also small is discussed in Section 5.2.

5.1. The case of large n

(a) *The REBE and the within-group sample variance.* For s_i^2 defined in (2.2),

$$\text{MSE}(s_i^2) = m_i^{-1}\rho_i + 2m_i^{-1}(m_i-1)^{-1}\sigma_i^4.$$

From Theorem 2, for any REBE defined in (3.4) or (3.5),

$$\text{MSE}(s_i^2) - \text{MSE(REBE)} \to 2m_i^{-1}(m_i-1)^{-1}\sigma_i^4 > 0$$

if $h_{\max} \to 0$ as $n \to \infty$. Hence the MSE of s_i^2 is larger than that of the REBEs for large n.

(b) *The REBE and the MINQUE.* The MINQUE is exactly unbiased, which often leads to a negative estimate of σ_i^2. As usual, a slightly biased estimator (the bias vanishes as the sample size tends to infinity) such as the REBE may perform better. The following results show that the MSE of the REBE is smaller than that of the MINQUE in broad situations. Let us first consider a special case where

$$y_{ij} = \mu + e_{ij}, \quad j = 1, \ldots, m_i, \quad i = 1, \ldots, k. \tag{5.1}$$

In this case, the MINQUE of σ_i^2 is $v_i^m = m_i^{-1}(n-2)^{-1}n\sum_{j\le m_i}(y_{ij}-\bar{y})^2 - (n-2)^{-1}s^2$, where $\bar{y} = n^{-1}\sum_{i\le k}\sum_{j\le m_i} y_{ij}$ and s^2 is given by (3.2). A direct calculation shows that

$$\text{MSE}(v_i^m) = m_i^{-1}(n-2)^{-2}n^2[\rho_i + O(n^{-1})] + O(n^{-2}).$$

Under the conditions in Theorem 2, for any $\lambda_i \in [0, 1]$,

$$n[\text{MSE}(v_i^m) - \text{MSE}(v_i^b(\lambda_i))] \to 2m_i^{-1}(1+\lambda_i)\rho_i > 0$$

as $n \to \infty$. The same result holds if $v_i^b(\lambda_i)$ is replaced by $v_i^w(\lambda_i)$.

Under the general model (1.1), the MSE of the MINQUE is not easy to obtain. We consider the special but important case where $m_i = m$ for all i.

Theorem 4. *Let v_i^m be the* MINQUE *of σ_i^2. Assume the conditions of Theorem 2 and that $m_i \equiv m$. If $m = 1$, we also assume $h_i < 0.5$, $i = 1, \ldots, k$, to ensure the existence of v_i^m. If $\lim_{n\to\infty}(h_{\max}^2/h_i) = 0$, then for any $\lambda_i \in [0,1]$,*

$$\liminf_{n\to\infty} h_i^{-1}[\text{MSE}(v_i^m) - \text{MSE}(v_i^b(\lambda_i))] \ge 2m^{-1}(1+\lambda_i)\rho_i > 0.$$

The same result holds if $v_i^b(\lambda_i)$ is replaced by $v_i^w(\lambda_i)$.

Proof. Let $G = (g_{ij})_{k\times k}$, where $g_{ij} = 1 - 2h_i + mh_i^2$ if $j = i$, and $g_{ij} = mh_{ij}^2$ if $j \neq i$. Since $m \geq 2$ (or $m = 1$ and $h_{\max} \leq 0.5$), $G^{-1} = (g^{ij})_{k\times k}$ exists and $\max_i \sum_{p\leq k} | g^{ip} | < \infty$. Then

$$\max_{1\leq i\leq k} \sum_{1\leq p<q\leq k} | g^{ip}g^{iq} | \leq \max_{1\leq i\leq k} \left(\sum_{p=1}^{k} | g^{ip} | \right)^2 < \infty. \tag{5.2}$$

From Lemma 4.5 of C. R. Rao (1970), $v_i^m =$ the ith component of $G^{-1}R$, where R is a k-vector whose ith component is $m^{-1}\sum_{j\leq m_i} r_{ij}^2$. Hence,

$$\mathrm{MSE}(v_i^m) = \sum_{j=1}^{k}(g^{ij})^2(1-h_j)^2\,\mathrm{Var}(\hat{a}_j) + 2\sum_{1\leq p<q\leq k} g^{ip}g^{iq}z_pz_q\mathrm{Cov}(\hat{a}_p, \hat{a}_q), \tag{5.3}$$

where $\hat{a}_i$ is defined in (3.3) and $z_i = 1 - h_i$. From (4.2) and (5.2), the second term of the right side of (5.3) is $O(h_{\max}^2)$. Since $g^{ii} \geq g_{ii}^{-1} = (1 - 2h_i + mh_i^2)^{-1}$, the first term of the right side of (5.3) is not smaller than $m^{-1}g_{ii}^{-2}[\rho_i + O(h_i)]$. Thus, $\mathrm{MSE}(v_i^m) \geq m^{-1}g_{ii}^{-2}[\rho_i + O(h_i)]$ up to the order $O(h_{\max}^2)$, which and Theorem 2 imply that

$$\begin{aligned} &\mathrm{MSE}(v_i^m) - \mathrm{MSE}(v_i^b(\lambda_i)) \\ \geq\ & m^{-1}[g_{ii}^{-2} - (1-h_i)^{-2}(1-\lambda_ih_i)^2][\rho_i + O(h_i)] + O(h_{\max}^2) \\ \geq\ & 2m^{-1}(1+\lambda_i)h_i[\rho_i + O(h_i)] + O(h_{\max}^2). \end{aligned}$$

The result follows. The proof for $v_i^w(\lambda_i)$ is the same.

(c) *The REBE and the ARE.* J. N. K. Rao (1973) proved that the ARE

$$v_i^r = m_i^{-1}\sum_{j=1}^{m_i} r_{ij}^2$$

has smaller MSE than the MINQUE in some situations. Under the same conditions as in Theorem 2, v_i^r has the same MSE as $v_i^b(1)$ (or $v_i^w(1)$) up to the order $O(h_ih_{\max})$. The bias of v_i^r, however, tends to be negative. Since

$$\mathrm{Bias}(v_i^r) = -h_i\sigma_i^2 + \sum_{l=1}^{k} h_{il}^2m_l(\sigma_l^2 - \sigma_i^2),$$

$\mathrm{Bias}(v_i^r) < 0$ if $\sup_l | \sigma_l^2 - \sigma_i^2 | < \sigma_i^2$. The condition $\sup_l | \sigma_l^2 - \sigma_i^2 | < \sigma_i^2$ is clearly not necessary for $\mathrm{Bias}(v_i^r) < 0$ (see Section 5.2). The confidence regions for β constructed by using v_i^r as the estimators of σ_i^2 usually have low coverage probabilities. See the simulation results in Section 7. In fact, $v_i^b(1)$ and $v_i^w(1)$ are bias adjustments of v_i^r, since

$$v_i^b(1) = v_i^r + h_i s^2 \text{ and } v_i^w(1) = v_i^r + h_i s_1^2.$$

5.2. The case of small n: an example

When n is small (consequently, m_i and k are small), it is hard to compare variance estimators analytically. The improvements obtained using empirical Bayesian methods become "small", since there is little auxiliary information. We compare the REBE with other variance estimators through the following example. Consider the model

$$y_{ij} = \beta_i + e_{ij}, \quad j = 1, \ldots, m_i, \quad i = 1, 2, \quad n = m_1 + m_2.$$

This can also be viewed as a two sample problem. If m_i are large, the estimators under comparison perform equally well. The MINQUE and the REBE $v_i^w(\lambda_i)$ $(0 \le \lambda_i \le 1)$ in this case are the same as s_i^2 and therefore the use of them does not achieve any improvement on s_i^2. The ARE v_i^r equals $(1 - m_i^{-1})s_i^2$. All these estimators do not use the data from the other group. The REBE $v_i^b(\lambda_i)$ equals

$$v_i^b(\lambda_i) = (1 - c_i)s_i^2 + c_i s_p^2, \quad c_i = \lambda_i \,/\, m_i,$$

where $s_p^2 = [(m_1-1)s_1^2+(m_2-1)s_2^2] \,/\, (n-2)$ is the pooled variance estimator when σ_i^2 are assumed to be equal or nearly equal. $v_i^b(\lambda_i)$ is a compromise between the within-group sample variance and the pooled estimator s_p^2. When $\lambda_i = 0$, $v_i^b(\lambda_i)$ equals s_i^2.

To compare these estimators, let us first look at their biases. s_i^2 is unbiased. The bias of v_i^r is $-m_i^{-1}\sigma_i^2$, which is always negative and can be large. The bias of $v_i^b(\lambda_i)$ is

$$\lambda_i(m_i - 1)(\sigma_j^2 - \sigma_i^2)/m_i(n - 2) = \lambda_i(\sigma_j^2 - \sigma_i^2)/2m \ \ (\text{if } m_1 = m_2 = m), \ \ j \neq i.$$

Hence $v_i^b(\lambda_i)$ does correct the negative bias of v_i^r, i.e., its bias does not have any deterministic trend and is smaller than that of v_i^r. The bias of $v_i^b(\lambda_i)$ is small if σ_i^2 are close. Also, $v_i^b(\lambda_i)$ with a smaller λ_i has smaller bias.

Next, we consider the MSE of these estimators. For simplicity we assume that $m_1 = m_2 = m$, and e_{ij} are distributed as $N(0, \sigma_i^2)$. The MSE of s_i^2 and v_i^r are $2\sigma_i^4 \,/\, (m_i - 1)$ and $(2m_i - 1)\sigma_i^4 \,/\, m_i^2$, respectively. Let $\theta_1 = \sigma_2^2/\sigma_1^2$ and $\theta_2 = \sigma_1^2/\sigma_2^2$. For $t \in [0, 1]$,

$$\text{MSE}(v_i^b(t)) = m^{-2}(m-1)^{-1}\sigma_i^4\Big[2(m-t/2)^2 + 2(t\theta_i/2)^2 + (t/2)^2(m-1)(\theta_i - 1)^2\Big],$$

which is decreasing in t when $\theta_i \le (3m-1)/(m+1)$. Hence if $\max(\theta_1, \theta_2) \le (3m-1)/(m+1)$,

$$\text{MSE}(v_i^b(t)) < \text{MSE}(v_i^b(s)), \quad i = 1, 2,$$

for $s < t$. In particular, the MSE of $v_i^b(\lambda_i)$ is less than that of the MINQUE or s_i^2.

It is not difficult to see that the MSE of v_i^r is generally less than that of $v_i^b(\lambda_i)$ and is therefore less than that of the MINQUE or s_i^2. v_i^r is further improved (in terms of MSE) by $v_i^c = (m+1)^{-1}(m_i-1)s_i^2$. But v_i^r and v_i^c are rarely used when m is small, since they underestimate σ_i^2 seriously. For example, when $m = 2$,

$$s_i^2 = (y_{i1} - y_{i2})^2/2, \quad v_i^r = (y_{i1} - y_{i2})^2/4, \quad v_i^c = (y_{i1} - y_{i2})^2/6$$

and

$$v_i^b(\lambda_i) = (1 - 0.25\lambda_i)(y_{i1} - y_{i2})^2/2 + 0.25\lambda_i(y_{j1} - y_{j2})^2/2, \quad j \ne i.$$

Clearly, v_i^r and v_i^c are too small. In fact, in this case the silly estimator $v_i \equiv 0$ has MSE half that of s_i^2 ! As we commented earlier, the MSE should not be the only criterion for choosing an estimator.

6. Applications to Inference

The use of the REBEs in assessing statistical accuracy, constructing confidence intervals for β and improving the OLSE are briefly discussed in this section.

6.1. Estimating linear functions of σ_i^2

Some statistical accuracy measures, such as the variance or covariance of the OLSE, are of the form $\gamma = \sum_i l_{in}\sigma_i^2$, where l_{in} are constants. A natural estimate of γ is obtained by replacing σ_i^2 by its estimate. Consider the following general class of REBEs:

$$\hat{\gamma} = \sum_{i=1}^{k} l_{in}[(1 - c_i)\hat{a}_i + c_i\bar{a}_i], \tag{6.1}$$

where $\hat{a}_i$ is defined in (3.3), $\bar{a}_i$ =either s^2 or s_1^2, c_i possibly depend on data and satisfy

$$0 \le c_i \le 1 \text{ and } \max_{i \le k} \sup_y c_i(y) \to 0 \text{ as } n \to \infty. \tag{6.2}$$

Note that $\sum_i l_{in}v_i^b(\lambda_i)$ and $\sum_i l_{in}v_i^w(\lambda_i)$ are special cases of (6.1)–(6.2).

Theorem 5. *Assume the conditions of Theorem 2 and that $h_{\max} \to 0$ as $n \to \infty$. If $\sum_i |l_{in}| = O(n^{-1})$ and $\sum_i l_{in}^2 = o(n^{-2})$, then $\hat{\gamma}$ defined in (6.1)–(6.2) satisfies*

$$\text{MSE}(\hat{\gamma}) = o(n^{-2}). \tag{6.3}$$

Remarks. (1) (6.3) means that $\hat\gamma$ is consistent in a stronger sense that $n^2E(\hat\gamma-\gamma)^2 \to 0$. The asymptotic unbiasedness and consistency of $\hat\gamma$ follow from (6.3).
(2) The REBEs of σ_i^2 are not consistent if m_i are small. However, for asymptotically unbiased estimators v_i of σ_i^2, "smooth" coefficients l_{in} (e.g., l_{in} satisfies the conditions in the theorem) will reduce the variance of $\sum_i l_{in}v_i$ and therefore $\hat\gamma$ is consistent as $n \to \infty$.
(3) For the asymptotic unbiasedness of $\hat\gamma$, the condition $\sum_i l_{in}^2 = o(n^{-2})$ is not required.

Proof. From (6.2), Theorem 1 and $\sum_i |\, l_{in}\,| = O(n^{-1})$,

$$\begin{aligned} n|\,E\hat\gamma-\gamma\,| &= n\Big|\sum_{i=1}^k l_{in}E(c_i\hat a_i) - \sum_{i=1}^k l_{in}E(c_i\overline{a}_i)\Big| + o(1) \\ &\le c_1(\max_{i\le k}\sup_y c_i)\Big[\sum_{i=1}^k |\, l_{in}\,|(E\hat a_i + E\overline{a}_i)\Big] + o(1) \to 0 \end{aligned}$$

since $E\hat a_i$ and $E\overline{a}_i$ are bounded, where c_1 is a positive constant. Also, from (4.2),

$$\max_{i\le k} \mathrm{Var}(\hat a_i) = O(1)$$

and

$$\max_{i\ne j} |\mathrm{Cov}(\hat a_i,\, \hat a_j)| \le \Big[m_i m_j(1-h_i)(1-h_j)\Big]^{-1}\sum_{p=1}^{m_i}\sum_{q=1}^{m_i} |\,\mathrm{Cov}(r_{ip}^2,\, r_{jq}^2)\,| = O(h_{\max}).$$

Then from (6.2) and $\sum_i l_{in}^2 = o(n^{-2})$,

$$n^2\mathrm{Var}(\hat\gamma) \le c_2 n^2\sum_{i=1}^k l_{in}^2 + c_3 n^2 h_{\max}\Big(\sum_{i=1}^k |\, l_{in}\,|\Big)^2 \to 0,$$

where c_2 and c_3 are positive constants. Thus the result follows.

The conditions in Theorem 5 are quite weak. As an example, we consider the estimation of $\mathrm{Var}\hat\beta$, the variance-covariance matrix of $\hat\beta$. Let l_{in}^{st} be the $(s,\, t)$th element of $m_iM^{-1}x_ix_i^tM^{-1}$. Then the $(s,\, t)$th element of $\mathrm{Var}\hat\beta$ is $\gamma_{st} = \sum_{i\le k} l_{in}^{st}\sigma_i^2$. Assume $M^{-1} = O(n^{-1})$ and $m_i \le m_\infty$. Then there is a constant $c > 0$ such that

$$\sum_{i=1}^k |\, l_{in}^{st}\,| \le \sum_{i=1}^k m_i x_i^t M^{-2}x_i \le cn^{-1}\sum_{i=1}^k m_i h_i = pcn^{-1}$$

and

$$\sum_{i=1}^k (l_{in}^{st})^2 \le \sum_{i=1}^k m_i^2(x_i^tM^{-2}x_i)^2 \le c^2\sum_{i=1}^k m_i^2h_i^2 \le c^2 m_\infty p h_{\max}.$$

Hence (7.3) holds if $h_{\max} \to 0$.

Monte Carlo comparisons of the REBEs with the estimators of γ_{st} based on the with-group sample variance, the MINQUE and the ARE are given in Section 7.

6.2. Confidence intervals

Let β_j and $\hat{\beta}_j$ be the jth components of β and $\hat{\beta}$, respectively, and s_{jj} be an estimator of the standard deviation of $\hat{\beta}_j$. If

$$n^{1/2}(s_{jj} - \text{the standard deviation of } \hat{\beta}_j) \xrightarrow{p} 0, \tag{6.4}$$

then the interval

$$[\hat{\beta}_j - z(1-\alpha/2)s_{jj},\ \hat{\beta}_j + z(1-\alpha/2)s_{jj}] \tag{6.5}$$

is an approximate $100(1-\alpha)\%$ confidence interval for β_j, where $z(\alpha)$ is the αth percentile of the standard normal distribution.

From Section 6.1, (6.4) is satisfied with s_{jj} = the square root of the REBE $\hat{\gamma}_{jj}$. Hence the confidence interval (6.5) based on the REBEs is asymptotically correct.

One can also use other estimators of σ_i^2 in constructing confidence interval (6.5). Monte Carlo comparisons of the coverage probabilities of confidence intervals based on various estimators of σ_i^2 are presented in Section 7.

6.3. The estimation of β

The Bayes estimator $\hat{\beta}_B$ in Section 2 does not have an explicit form and is hard to compute. Lindley and Smith (1972) showed that if σ_i^2 are known and $\pi(\beta) \equiv 1$, the Bayes estimator of β is

$$\beta_\tau = (X^tDX)^{-1}X^tDy,\ \ D = block\ diag.(\sigma_1^{-2}I_{m_1\times m_1} \cdots \sigma_k^{-2}I_{m_k\times m_k}). \tag{6.6}$$

Hence if $p(\tau\,|\,y)$ is the posterior of the parameter τ (see Section 2), then

$$\hat{\beta}_B = E_{p(\tau\,|\,y)}\beta_\tau = \int [(X^tDX)^{-1}X^tDy]p(\tau\,|\,y)\,d\tau.$$

If v_i is the empirical Bayes estimator of σ_i^2 (e.g., $v_i = v_i^b(\lambda_i)$ or $v_i^w(\lambda_i)$), then an approximate empirical Bayes estimator of β is

$$\tilde{\beta} = (X^tWX)^{-1}X^tWy,\ \ W = block\ diag.(w_1I_{m_1\times m_1} \cdots w_kI_{m_k\times m_k}), \tag{6.7}$$

where $w_i = v_i^{-1}$. Note that $\tilde{\beta}$ is a WLSE with the reciprocals of the REBEs as weights. The simulation results in Section 7 indicate that $\tilde{\beta}$ improves the

OLSE if $m_i \geq 3$ and is better than the WLSEs based on other estimators of σ_i^2. Asymptotic properties of $\tilde{\beta}$ are studied in Shao (1989).

This estimation procedure can be used iteratively. Note that $\tilde{\beta}$ is the OLSE of β under the model $w_i^{1/2} y_{ij} = w_i^{1/2} x_i^t \beta + w_i^{1/2} e_{ij}$ by treating w_i as constants. Thus, we can obtain estimates of β and σ_i^2 simultaneously through an iterative procedure:

obtain an estimate of $\beta \rightarrow$ obtain estimates of $\sigma_i^2 \rightarrow$ feedback and repeat.

With a fixed number of iterations, however, the estimator of β may not be asymptotically efficient, especially when $m_i \leq 2$ (Carroll and Cline (1988)).

7. Simulation Results

In this section, we examine by simulation (a) the finite sample performances of the estimators of σ_i^2 considered in the previous sections; (b) the performances of the estimators of Var$\hat{\beta}$ based on various estimators of σ_i^2; (c) the empirical coverage probabilities of the approximate 95% confidence intervals of β using various estimators of σ_i^2; (d) the performances of the WLSEs with the reciprocals of the estimators of σ_i^2 used as weights. We consider the following quadratic regression model:

$$y_{ij} = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_{ij}, \quad j = 1, 2, \quad i = 1, \ldots, 20. \tag{7.1}$$

The values of x_i are: 0.4, 0.5, 0.6, 0.7, 0.8, 1, 1.5, 2, 2.5, 3, 3.5, 4, 5, 6, 7, 8, 10, 12, 15, and 18. The random errors e_{ij} are independently distributed as $N(0, \sigma_i^2)$, where the values of σ_i^2 are: 0.2, 0.8, 0.5, 0.9, 0.8, 0.5, 0.91, 0.65, 0.77, 0.81, 0.21, 0.81, 0.12, 0.52, 0.9, 0.94, 0.67, 0.53, 0.88, and 1.0.

We study the following estimators for estimating σ_i^2 and Var$\hat{\beta}_j$: the within-group sample variance s_i^2, the MINQUE v_i^m, the ARE v_i^r, and the REBEs $v_i^b(0) = v_i^w(0), v_i^b(1/2), v_i^b(1), v_i^w(1/2)$ and $v_i^w(1)$.

The approximate 95% confidence intervals of β_j (see (6.5)) using v_i as an estimate of σ_i^2, $i = 1, \ldots, k$, are denoted by $C(v_i)$. We study the coverage probability of $C(v_i)$ with $v_i =$ one of estimators of σ_i^2 given above. The OLSE and the WLSE (6.7) with $w_i = v_i^{-1}$ are also studied. The "optimal" WLSE (6.6) with the true σ_i^2 as weights is included for comparison. The value of β in these studies is $(1\ 4\ -0.5)^\tau$.

The following is a summary for the simulation study. All the results are based on 3,000 simulations on a VAX 11/780 at the Purdue University.

(i) *The performance of the estimators of* σ_i^2. Table 1 shows the root mean squared errors (RMSE) and biases of the estimators of σ_i^2. Some conclusions drawn from this table are:

(a) In terms of the RMSE, the REBEs are better than the within-group sample variance and the MINQUE for all i. The improvement can be as high as 43%.
(b) The ARE has smaller RMSE then the REBEs but has a large negative bias.
(c) The REBEs with larger λ have smaller RMSE and larger bias (in absolute value).

(ii) *The performance of the estimators of* $\mathrm{Var}\hat{\beta}_j$. The RMSE and biases of the estimators of $\mathrm{Var}\hat{\beta}_j$ are reported in Table 2. The REBEs are better than the with-group sample variance and the MINQUE. The best REBE is $\sum_{i\leq k} l_{in}^{jj} v_i^b(1)$ (see Section 6.1), which improves the with-group sample variance about 30-41% and improves the MINQUE about 20-36%. The RMSE of the ARE are similar to those of the REBEs. But the ARE has large negative bias; its relative bias can be as high as 20%.

(iii) *The performance of the confidence intervals.* The coverage probabilities and the average lengths of the confidence intervals of β_j are shown in Table 3. The results show that $C(v_i^b(\lambda_i))$ and $C(v_i^w(\lambda_i))$ have higher coverage probabilities than $C(s_i^2)$. The coverage probabilities of $C(v_i^r)$ may be even lower than those of $C(s_i^2)$.

(iv) *The performance of the OLSE and WLSEs.* Table 4 contains the biases and RMSE of the OLSE and WLSEs. For comparison, we also include the biases and RMSE of the OLSE and WLSEs under model (7.1) but with three (or four) replicates in each group (the WLSE using s_i^2 or the MINQUE is only studied for two replicates case). The results indicate that
(a) The performances of the WLSEs are not as good or almost the same as that of the OLSE when there are two replicates in each group. However, the WLSEs using the REBEs are better than the OLSE when there are three or four replicates in each group, although the improvement is not large.
(b) The WLSEs using the REBEs with large λ_i are better than the WLSEs using the ARE. The WLSE using the within-group sample variances or the MINQUE has very large RMSE and is not recommended.

8. An Example

We consider the following example taken from the pharmaceutical industry. To determine the relationship between y, a characteristic of a pharmaceutical compound, and a covariate x, the concentration of the compound, some assay results were obtained using several standard concentrations. Usually a simple linear regression models is appropriate, i.e., $y = \beta_0 + \beta_1 x$. The experiment is typically run on several days with the same x's on every day, but different batches of compound on different days. Within each day, it is reasonable to assume that the error variances are the same. The error variances for different days, however, may be different due to batch-to-batch variation. Table 5 displays standard concentrations (x) and peak responses (y) collected in 15 different days. It is convenient to use two indices d (day) and j (within day replication) instead of one index i. Thus,

$$y_{dj} = \beta_0 + \beta_1 x_{dj} + e_{dj}, \quad d = 1, \ldots, 15, \quad j = 1, \ldots, 4,$$

where $Ee_{dj} = 0$ and $\mathrm{Var}(e_{dj}) = \sigma_d^2$.

In addition to the estimation of β_0 and β_1, we also need to estimate individual variances σ_d^2 for the purpose of quality control. That is, the error variance has to be kept under a specific value; otherwise we need to modify the manufacturing process to improve the quality.

Since, in this example, the variances do not vary with the x's, we use the REBEs $v_d^b(1)$ computed using the formula in Section 3.4. The OLSE $\hat{\beta}_0 = 0.0504$ and $\hat{\beta}_1 = 0.1026$ are used as initial estimates. The estimates of σ_d^2 (shown in Table 5) are very different for different days. Since there are four observations (with the same variance) for each day, the WLSE using appropriate variance estimates is more efficient than the OLSE. The WLSE based on the variance estimates in Table 5 are $\tilde{\beta}_0 = 0.3799$ and $\tilde{\beta}_1 = 0.1036$. Note that $\hat{\beta}_0$ and $\tilde{\beta}_0$ (estimates of intercept) are quite different, although estimates of slope are almost the same.

Acknowledgements

This research was partially supported by an operating grant from the Natural Science and Engineering Research Council of Canada. I would like to thank the referees for their helpful comments. Thanks also go to Professor Robert Gentleman for his valuable suggestions.

Table 1. RMSE and biases of estimators of σ_i^2
(The biases are shown in the second row for each i.)

i	s_i^2	v_i^m	v_i^r	$v_i^b(0)$	$v_i^b(1/2)$	$v_i^w(1/2)$	$v_i^b(1)$	$v_i^w(1)$
1	.2889	.2525	.2198	.2379	.2329	.2337	.2292	.2309
	.0004	−.0025	.0159	.0323	.0476	.0472	.0630	.0622
2	1.1611	.8564	.7528	.8035	.7784	.7806	.7534	.7578
	.0031	−.0015	−.0640	−.0117	−.0156	−.0160	−.0195	−.0204
3	.6768	.5373	.4749	.5058	.4908	.4922	.4759	.4787
	−.0131	−.0090	−.0295	.0021	.0074	.0070	.0127	.0118
4	1.3443	.9599	.8533	.9060	.8805	.8823	.8551	.8588
	.0368	.0220	−.0490	.0050	−.0020	−.0025	−.0090	−.0100
5	1.1908	.8624	.7705	.8161	.7941	.7956	.7722	.7752
	.0256	.0238	−.0325	.0135	.0094	.0089	.0054	.0044
6	.7124	.5308	.4796	.5050	.4928	.4935	.4807	.4822
	.0084	−.0032	−.0207	.0051	.0093	.0088	.0135	.0125
7	1.2577	.9566	.8828	.9145	.9019	.9025	.8844	.8856
	.0013	.0116	−.0372	−.0001	−.0050	−.0055	−.0099	−.0182
8	.9369	.6862	.6405	.6631	.6522	.6524	.6414	.6416
	−.0014	.0019	.0209	.0017	.0020	.0016	.0024	.0016
9	1.1219	.8022	.7531	.7773	.7655	.7657	.7538	.7541
	−.0101	−.0040	−.0322	−.0079	−.0093	−.0096	−.0108	−.0114
10	1.1766	.8705	.8167	.8434	.8306	.8307	.8178	.8181
	−.0035	.0031	−.0281	−.0021	−.0043	−.0045	−.0065	−.0069
11	.3186	.2428	.2267	.2353	.2323	.2323	.2296	.2297
	.0095	.0035	.0117	.0197	.0273	.0272	.0349	.0347
12	1.1694	.8871	.8210	.8543	.8388	.8392	.8233	.8241
	.0300	.0346	−.0046	.0277	.0245	.0245	.0213	.0213
13	.1651	.1549	.1411	.1490	.1489	.1492	.1497	.1505
	−.0004	.0026	.0192	.0265	.0394	.0395	.0523	.0525
14	.7604	.5891	.5216	.5560	.5402	.5417	.5245	.5275
	.0192	.0154	−.0086	.0247	.0285	.0287	.0324	.0328
15	1.2967	.9631	.8391	.8992	.8689	.8720	.8386	.8449
	.0046	−.0083	−.0876	−.0250	−.0323	−.0321	−.0396	−.0393
16	1.3606	1.0102	.8666	.9367	.9018	.9059	.8671	.8753
	.0081	.0119	−.0857	−.0121	−.0223	−.0222	−.0325	−.0323
17	.9340	.7395	.6265	.6821	.6546	.6583	.6273	.6347
	−.0005	−.0001	−.0563	.0008	.0008	.0012	.0008	.0017
18	.7791	.6251	.5261	.5774	.5546	.5580	.5320	.5390
	.0335	.0298	−.0036	.0459	.0499	.0524	.0540	.0590
19	1.1788	.9680	.7532	.8555	.8028	.8197	.7506	.7860
	−.0289	−.0080	−.1193	−.0049	−.0183	−.0048	−.0316	−.0048
20	1.3684	1.3109	.8026	1.0975	.9137	1.0461	.7369	.9958
	−.0176	−.0088	−.3845	−.0419	−.0933	−.0506	−.1447	−.0593

Table 2. RMSE and biases (shown in the second row) of estimators of $\text{Var}\hat{\beta}_j$

j	$\text{Var}\hat{\beta}_j$	Variance estimators s_i^2	v_i^m	v_i^r	$v_i^b(0)$	$v_i^b(1/2)$	$v_i^w(1/2)$	$v_i^b(1)$	$v_i^w(1)$
0	5.7762	2.6545	1.9232	1.8162	1.9018	1.8771	1.8944	1.8541	1.8872
		.0611	.0409	-.3836	.0434	.0356	.0459	.0278	.0485
1	.7899	.3648	.3020	.2481	.2772	.2593	.2713	.2441	.2656
		.0062	.0047	-.1000	-.0005	-.0089	-.0018	-.0173	-.0031
2	.0030	.0020	.0018	.0013	.0016	.0014	.0015	.0012	.0015
		-.0000	.0000	-.0006	-.0000	-.0001	-.0001	-.0002	-.0001

All the numbers in this table have been multiplied by 100.

Table 3. Coverage probabilities and lengths (shown in brackets) of $C(v_i)$

j	$C(s_i^2)$	$C(v_i^m)$	$C(v_i^r)$	$C(v_i^b(0))$	$C(v_i^b(1/2))$	$C(v_i^w(1/2))$	$C(v_i^b(1))$	$C(v_i^w(1))$
0	.9207	.9353	.9223	.9347	.9343	.9347	.9347	.9353
	(.9244)	(.9329)	(.8984)	(.9334)	(.9331)	(.9338)	(.9327)	(.9341)
1	.9260	.9330	.9163	.9307	.9310	.9317	.9307	.9320
	(.3413)	(.3436)	(.3214)	(.3433)	(.3419)	(.3432)	(.3405)	(.3430)
2	.9037	.9120	.8887	.9160	.9203	.9173	.9223	.9183
	(.0207)	(.0209)	(.0189)	(.0209)	(.0208)	(.0209)	(.0206)	(.0209)

Table 4. RMSE and biases (shown in the second row) of the OLSE and WLSE

j	OLSE	Variance estimators used in WLSE s_i^2	v_i^m	v_i^r	$v_i^b(0)$	$v_i^b(1/2)$	$v_i^w(1/2)$	$v_i^b(1)$	$v_i^w(1)$	σ_i^2
Two replicates in each group										
0	.2410	2.4250	2.0897	.3082	.2524	.2463	.2467	.2453	.2459	.2121
	.0021	-.0017	.0545	-.0021	-.0006	.0010	.0009	.0011	.0011	.0018
1	.0891	.5945	.8066	.0977	.1710	.0889	.0891	.0885	.0889	.0765
	.0004	-.0413	.0055	.0006	-.0023	.0012	.0012	.0011	.0012	.0003
2	.0055	.0564	.0580	.0070	.0194	.0055	.0055	.0055	.0055	.0048
	-.0001	.0052	-.0005	.0000	.0003	-.0001	-.0001	-.0001	-.0001	-.0000
Three replicates in each group										
0	.1999	–	–	.2022	.2021	.2008	.2009	.1999	.2006	.1732
	-.0006	–	–	-.0019	-.0020	-.0018	-.0019	-.0017	-.0018	-.0007
1	.0754	–	–	.0745	.0743	.0739	.0739	.0736	.0737	.0624
	.0006	–	–	.0014	.0014	.0014	.0014	.0013	.0014	.0006
2	.0047	–	–	.0046	.0046	.0046	.0046	.0045	.0045	.0039
	-.0001	–	–	-.0001	-.0001	-.0001	-.0001	-.0001	-.0001	-.0001
Four replicates in each group										
0	.1700	–	–	.1699	.1700	.1693	.1693	.1688	.1689	.1500
	-.0011	–	–	-.0019	-.0019	-.0018	-.0018	-.0018	-.0017	-.0012
1	.0642	–	–	.0628	.0627	.0625	.0625	.0622	.0623	.0541
	.0008	–	–	.0009	.0009	.0009	.0009	.0009	.0008	.0007
2	.0040	–	–	.0039	.0039	.0038	.0039	.0038	.0038	.0034
	-.0001	–	–	-.0001	-.0001	-.0001	-.0001	-.0001	-.0001	-.0001

Table 5. Example

x= standard concentration (ng/ml)

Day	Responses $x=0$	$x=1$	$x=10$	$x=30$	Variance estimates
1	0.186	0.240	1.400	3.230	0.0358
2	0.110	0.220	1.360	3.270	0.0277
3	0.111	0.200	1.290	3.160	0.0137
4	0.000	0.130	1.100	3.250	0.0051
5	0.039	0.112	1.060	3.140	0.0011
6	0.001	0.120	1.000	3.231	0.0055
7	0.032	0.109	0.920	2.820	0.0309
8	0.030	0.100	0.919	3.190	0.0084
9	0.100	0.202	1.270	3.080	0.0117
10	0.032	0.100	0.920	2.809	0.0328
11	0.030	0.110	0.920	3.090	0.0075
12	0.003	0.100	0.870	3.380	0.0283
13	0.049	0.150	1.030	3.020	0.0040
14	0.051	0.130	1.210	3.210	0.0068
15	0.038	0.140	0.980	3.010	0.0064

References

Carroll, R. J. (1982). Adapting for heteroscedasticity in linear models. *Ann. Statist.* **10**, 1224–1233.

Carroll, R. J. and Ruppert, D. (1986). Discussion of "Jackknife, bootstrap and other resampling methods in regression analysis" by C. F. J. Wu. *Ann. Statist.* **14**, 1298–1301.

Carroll, R. J. and Cline, D. B. H. (1988). An asymptotic theory for weighted least-squares with weights estimated by replication. *Biometrika* **75**, 35–43.

Das Gupta, A. (1986). Simultaneous estimation in the multiparameter gamma distribution under weighted quadratic losses. *Ann. Statist.* **14**, 206–219.

Davidian, M. and Carroll, R. J. (1987). Variance function estimation. *J. Amer. Statist. Assoc.* **82**, 1079–1091.

Efron, B. (1979). Bootstrap methods: another look at the jackknife. *Ann. Statist.* **7**, 1–26.

Horn, S. D., Horn, R. A. and Duncan, D. B. (1975). Estimating heteroscedastic variances in linear models. *J. Amer. Statist. Assoc.* **70**, 380–385.

Huber, P. J. (1981). *Robust Statistics.* John Wiley, New York.

Jacquez, J. A., Mather, F. J. and Crawford, C. R. (1968). Linear regression with nonconstant, unknown error variances: sampling experiments with least squares, weighted least squares and maximum likelihood estimators. *Biometrics* **24**, 607–626.

Lindley, D. V. and Smith, A. F. M. (1972). Bayes estimates for the linear model. *J. Roy. Statist. Soc. Ser.B* **34**, 1–41.

Muller, H.-G. and Stadtmüller, U. (1987). Estimation of heteroscedasticity in regression analysis. *Ann. Statist.* **15**, 610–625.

Rao, C. R. (1970). Estimation of heteroscedastic variances in linear models. *J. Amer. Statist. Assoc.* **65**, 161–172.

Rao, J. N. K. (1973). On the estimation of heteroscedastic variances. *Biometrics* **29**, 11–24.

Shao, J. (1986). Discussion of "Jackknife, bootstrap and other resampling methods in regression analysis" by C. F. J. Wu. *Ann. Statist.* **14**, 1322–1326.

Shao, J. (1989). Asymptotic distribution of the weighted least squares estimator. *Ann. Inst. Statist. Math.* **41**, 365–382.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Ann. Statist.* **14**, 1261–1350.

Department of Mathematics, University of Ottawa, Ottawa, Ontario K1N 6N5, Canada.

(Received August 1989; accepted January 1992)