Estimating Heteroscedastic Variances in Linear Models II: Properties of the Resampling Empirical Bayes Estimators

by

Jun Shao
Purdue University

Purdue University

Technical Report #87-43

Department of Statistics
Purdue University

October 1987

Estimating Heteroscedastic Variances in Linear Models II:

Properties of the Resampling Empirical Bayes Estimators

by
Jun Shao
Purdue University

Abstract

Properties of the resampling empirical Bayes estimators (REBE's) of the error variances in a heteroscedastic linear model (Shao, 1987) are studied. We concentrate on (i) the consistency, bias and mean squared error (MSE) of the REBE and (ii) the comparisons between the REBE and other variance estimators such as the within-group sample variance, MINQUE and the within-group average of squared residuals. In particular, we obtain an upper bound for the bias and a second order expansion of the MSE of REBE, and show that the REBE has smaller MSE than the within-group sample variance and MINQUE if the total number of observations is large. The consistency of a class of estimators of a linear function of the error variances is also studied.

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

1. Introduction.

We consider the estimation of the variances σ_i^2 in the following heteroscedastic linear model:

$$y_{ij} = x_i'\beta + e_{ij}, \quad j=1,...,m_i,\ i=1,...,n,\ \sum_{i=1}^n m_i=N, \tag{1.1}$$

where $\beta\in\mathbf{R}^k$ is the unknown parameter, $x_i\in\mathbf{R}^k$ are deterministic, and e_{ij} are mutually independent with $Ee_{ij}=0$ and $Ee_{ij}^2=\sigma_i^2$, $j=1,...,m_i$, $i=1,...,n$. The σ_i^2 are assumed uniformly bounded (i.e., $\sigma_i^2\le\sigma_U^2$ for all i) but otherwise unknown.

Since in most practical situations m_i are small although n and N may be large, it is difficult to obtain good estimators of σ_i^2 without putting any restrictions on σ_i^2 or their estimators. A great deal of research work has been done in this area by assuming $\sigma_i^2=H(x_i)$, where H is unknown or is known up to several unknown parameters, and estimating H from data. See Carroll (1982) and its references for further details. On the other hand, C. R. Rao (1970) developed the MINQUE (minimum norm quadratic unbiased estimator(s)) by imposing some restrictions on the estimators of σ_i^2 (see Section 6.2).

By incorporating data resampling techniques, Shao (1987) proposed two classes of empirical Bayes estimators (1.3)-(1.5). The purposes of this paper are: (i) to study the properties of the empirical Bayes estimators (1.3)-(1.5), and (ii) to compare them with other variance estimators. Formally we define the Resampling Empirical Bayes Estimator (REBE) as follows. For their derivations, we refer to Shao (1987).

Let

$$y = (y_{11}\ y_{12}\ \cdots\ y_{nm_n})'_{N\times 1}$$

and

$$X = (X_1\ X_2\ \cdots\ X_n)'_{N\times k},\quad X_i = (x_i\ x_i\ \cdots\ x_i)_{k\times m_i},\quad i=1,...,n.$$

Assume $M=X'X$ is nonsingular. Let

$$r_{ij} = y_{ij} - x_i'\hat{\beta},$$

where $\hat{\beta}=M^{-1}X'y$ is the least squares estimator of β, and

$$\hat{a}_i = m_i^{-1}(1-h_i)^{-1}\sum_{j=1}^{m_i} r_{ij}^2, \tag{1.2}$$

where $h_i = x_i' M^{-1} x_i$, $i=1,\dots,n$, are the diagonal elements of the "hat" matrix $XM^{-1}X'$. A class of REBE's of σ_i^2 obtained by using the bootstrap method is

$$(1.3) \qquad v_i^b(\lambda_i) = (1-\lambda_i h_i)\hat{a}_i + \lambda_i h_i s^2,$$

where $\lambda_i \in [0,1]$ and $s^2 = (N-k)^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i} r_{ij}^2$ is the usual variance estimator when $\sigma_i^2 = \sigma^2$ for all i. A class of REBE's obtained by using the weighted resampling method is

$$(1.4) \qquad v_i^w(\lambda_i, r) = (1-\lambda_i h_i)\hat{a}_i + \lambda_i h_i s_{J,r}^2,$$

where $\lambda_i \in [0,1]$, r is an integer satisfying $k \le r \le N$ and $r/N \to 1$, $s_{J,r}^2 = h_i^{-1} x_i' V_{J,r} x_i$, and $V_{J,r}$ is the weighted retain-r jackknife estimator of the variance-covariance matrix of $\hat{\beta}$ (Wu, 1986). In particular, if $r = N-1$, (1.4) reduces to

$$(1.5) \qquad v_i^w(\lambda_i) = (1-\lambda_i h_i)\hat{a}_i + \lambda_i h_i s_J^2, \quad s_J^2 = h_i^{-1}\sum_{l=1}^{n} h_{il}^2 m_l \hat{a}_l,$$

where $h_{ij} = x_i' M^{-1} x_j$.

The estimators (1.3)-(1.5) are shrinkage estimators. Similar to the MINQUE, they estimate σ_i^2 by using not only the data in the ith group, but also the data in the other groups. These estimators are usually superior to the customary estimator, the within-group sample variance, especially when σ_i^2 have some features in common.

We study the properties of REBE's in Sections 2-5. Section 2 contains a result for the consistency of the REBE. An upper bound of the bias of the REBE and a second order expansion of the mean squared error (MSE) of the REBE are given in Sections 3 and 4, respectively, in terms of the diagonal elements of the "hat" matrix. The problem of choosing a "best" estimator within class (1.3) or (1.4) is discussed in Section 5. Except in Section 2, we concentrate on the situation where m_i are small but N is large.

Comparisons between the REBE and other variance estimators are given in Sections 6 and 7. In addition to such properties such as invariance, asymptotic unbiasedness and robustness against non-normality (Shao, 1987), the REBE has smaller MSE than the within-group sample variance and the MINQUE when N is large (Sections 6.1 and 6.2). In Section 6.3, we compare the REBE with the within-group average of squared residuals (ARE), which is proposed by J. N. K. Rao (1973) as a modification of the MINQUE. It turns out that the ARE has the same second order MSE expansion as the REBE $v_i^b(1)$ and $v_i^w(1)$ but has a larger negative bias, and $v_i^b(1)$ and $v_i^w(1)$ are actually bias adjustments of the ARE. The performances of these variance estimators in the case of small N is discussed through an example in Section 7. Shao (1987) contains some simulation results which show that the REBE's generally perform better than the other variance estimators under consideration.

A brief discussion of estimating linear functions of σ_i^2, $i=1,\ldots,n$, by using the REBE of individual σ_i^2 is given in the last section.

2. Consistency of the REBE when m_i is large.

We consider the consistency of the REBE (when $m_i \to \infty$) for the following reasons:

(1) Although in common situations m_i are small, there are some statistical applications considering large m_i and small n.

(2) Consistency is a *basic* requirement for any estimator.

Theorem 1. Suppose that

(2.1) $$h_{\max} = \max_{i \le n} h_i \to 0 \quad \text{as } N \to \infty.$$

Then for any $\lambda_i \in [0,1]$, as $m_i \to \infty$,

(i) $v_i^b(\lambda_i) \to \sigma_i^2$ in probability;

(ii) $v_i^w(\lambda_i, r) \to \sigma_i^2$ in probability.

Proof. Under (2.1), it is easy to see that as $m_i \to \infty$,

$$\hat{a}_i \to \sigma_i^2 \quad \text{in probability.}$$

From (1.3) and (1.4), it remains to be shown that Es^2 and $Es_{J,r}^2$ are bounded. From Lemma 2.1 of Shao (1986), $\max_{i,j} Er_{ij}^2$ are bounded if the σ_i^2 are. Hence Es^2 are bounded. That $Es_{J,r}^2$ are bounded is proved in the following lemma. □

Lemma 1. Suppose that

(2.2) $$limsup_{N\to\infty}(N-r)h_{\max} < 1.$$

Then

(2.3) $$Ex_i'V_{J,r}x_i = O(h_i).$$

Proof. Let $s=\{i_1, \ldots, i_p\}\subset\{1, \ldots, N\}$ be a subset of integers and X_s be the submatrix of X containing the i_1th,..., i_pth rows of X. Let $M_s=X_s'X_s$ and $\bar{s}$ be the complement of s. From the proof of Theorem 1 of Shao and Wu (1987),

$$EV_{J,r} = Var\hat{\beta} + S_1 - S_2 + S_3,$$

where

$$S_1 = O[N^{-1}(N-r)]Var\hat{\beta},$$
$$S_2 \le \sigma_U^2 \binom{N-k}{r-k+1}^{-1} \sum_s |M|^{-1}(|M|-|M_s|)M^{-1}X_{\bar{s}}'X_{\bar{s}}M^{-1},$$
$$S_3 \le \sigma_U^2 \binom{N-k}{r-k+1}^{-1} \sum_s (M_s^{-1}-M^{-1})M(M_s^{-1}-M^{-1}),$$

and $\sum_s$ is the summation over all distinct subsets s of size r. Now, there is a constant c_1>0 such that

$$x_i'S_1x_i \le c_1N^{-1}(N-r)x_i'(Var\hat{\beta})x_i \le c_1\sigma_U^2N^{-1}(N-r)h_i$$

$$\le k^{-1}c_1\sigma_U^2(N-r)h_ih_{\max} = O[(N-r)h_ih_{\max}],$$

where the last inequality follows from $h_{\max}\ge N^{-1}\sum_{i=1}^n m_ih_i=kN^{-1}$. Also,

$$x_i'S_2x_i \le \sigma_U^2 \binom{N-k}{r-k+1}^{-1} \sum_s |M|^{-1}(|M|-|M_s|)\, x_i'M^{-1}X_{\bar{s}}'X_{\bar{s}}M^{-1}x_i$$
$$\le \sigma_U^2 \binom{N-k}{r-k+1}^{-1} \sum_s |M|^{-1}(|M|-|M_s|)\sum_{j\in\bar{s}}h_ih_j$$
$$\le \sigma_U^2(N-r)h_ih_{\max}\binom{N-k}{r-k+1}^{-1} \sum_s |M|^{-1}(|M|-|M_s|)$$
$$= \sigma_U^2(N-r)h_ih_{\max}\binom{N-k}{r-k+1}^{-1}[\binom{N}{r}-\binom{N-k}{r-k}] = O[(N-r)h_ih_{\max}],$$

and there is a constant c_2>0 such that

$$x_i'S_3x_i \le \sigma_U^2 \binom{N-k}{r-k+1}^{-1} \sum_s x_i'(M_s^{-1}-M^{-1})M(M_s^{-1}-M^{-1})x_i$$

$$= \sigma_U^2 \binom{N-k}{r-k+1}^{-1} \sum_s \sum_{l=1}^n m_l [x_i'(M_s^{-1} - M^{-1})x_l]^2$$
$$\le c_2 (N-r) h_i h_{\max} \binom{N-k}{r-k+1}^{-1} \sum_s \sum_{l=1}^n m_l x_l' M^{-1} X_{\bar{s}}' X_{\bar{s}} M^{-1} x_l$$
$$\le c_2 (N-r) h_i h_{\max} \binom{N-k}{r-k+1}^{-1} \sum_s \sum_{l=1}^n m_l \sum_{j \in \bar{s}} h_l h_j$$
$$= k^2 c_2 (N-r) h_i h_{\max} \binom{N-k}{r-k+1}^{-1} \binom{N-1}{r} = O[(N-r) h_i h_{\max}],$$

where the second inequality follows from

$$[x_i'(M_s^{-1} - M^{-1})x_l]^2 \le [x_i'(M_s^{-1} - M^{-1})x_i][x_l'(M_s^{-1} - M^{-1})x_l]$$

and the fact that

$$M_s^{-1} - M^{-1} \le [1-(N-r)h_{\max}]^{-1} M^{-1} X_{\bar{s}}' X_{\bar{s}} M^{-1} \le (N-r) h_{\max} [1-(N-r)h_{\max}]^{-1} M^{-1}$$

(Lemma 4, Shao and Wu, 1987). Thus (2.3) follows. □

3. The bias of the REBE.

From now on we consider the case that m_i are small but N is large. Asymptotic unbiasedness of $v_i^b(\lambda_i)$ and $v_i^w(\lambda_i, r)$ follows from the following result which gives an upper bound on the order of the magnitude of the bias of the REBE. For a variance estimator v_i, let $Bias(v_i) = Ev_i - \sigma_i^2$.

Theorem 2. Let $\lambda_i \in [0,1]$.

(i) If

(3.1) $\qquad limsup_{N \to \infty} h_{\max} < 1,$

then there is a constant $c>0$ (independent of i and N) such that

(3.2) $\qquad |Bias(v_i^b(\lambda_i))| \le ch_i.$

(ii) Under (2.2), (3.2) holds with $v_i^b(\lambda_i)$ replaced by $v_i^w(\lambda_i, r)$.

Proof. From the proof of Theorem 1, Es^2 and $Es_{J,r}^2$ are bounded under the given conditions. Note that $(1-\lambda_i h_i)\sigma_i^2 = \sigma_i^2 + O(h_i)$. The results follow if

(3.3) $$E\hat{a}_i = \sigma_i^2 + O(h_i).$$

From (1.2),

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

Now (3.3) follows from

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

for sufficiently large N, where $c=\sigma_U^2(1-limsup_N h_{\max})^{-1}$. □

Hence if $h_i \to 0$ as $N\to\infty$ (which is implied by (2.1)), then $v_i^b(\lambda_i)$ and $v_i^w(\lambda_i, r)$ are asymptotically unbiased. Condition (2.1) is quite weak since it is known to be necessary and sufficient for the asymptotic normality of $\hat{\beta}$ in the case of homoscedastic errors (Huber, 1981).

4. The MSE of the REBE.

The exact form of the MSE of the REBE is extremely complicated due to the nonidentical distributions of the errors. The following theorem gives asymptotic ($N\to\infty$) expansions of the MSE of REBE's. Assume that the fourth moments of the error distributions exist and $\tau_i = Var(e_{ij}^2)$, $j=1,...,m_i$, $i=1,...,n$, are uniformly bounded.

Theorem 3. For any $\lambda_i \in [0,1]$,

(i) $MSE(v_i^b(\lambda_i)) = m_i^{-1}(1-h_i)^{-2}(1-\lambda_i h_i)^2[\tau_i + O(h_i)] + O(h_i h_{\max})$ if (3.1) holds.

(ii) $MSE(v_i^w(\lambda_i, r)) = m_i^{-1}(1-h_i)^{-2}(1-\lambda_i h_i)^2[\tau_i + O(h_i)] + O(h_i h_{\max})$ if (2.2) holds and $N-r$ is fixed (independent of N).

Remark. From the proof of Theorem 3, the above expansions hold 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 the above expansions are bounded in absolute value by ch_i and $ch_i h_{\max}$, respectively.

We need the following results for the proof of Theorem 3.

Lemma 2. Let ε_i be independent with $E\varepsilon_i=0$, $Var(\varepsilon_i^2)<\infty$, $i=1,\dots,n$, and c_{pq} be some constants, $1\le p, q\le n$. Then

$$Var(\sum_{p=1}^{n}\sum_{q=1}^{n} c_{pq}\varepsilon_p\varepsilon_q) = \sum_{p=1}^{n} c_{pp}^2 Var(\varepsilon_p^2) + 4\sum_{1\le p<q\le n} c_{pq}^2 E\varepsilon_p^2\varepsilon_q^2.$$

The proof of this lemma is straightforward and is omitted.

Lemma 3. There is an absolute constant $c>0$ such that if $(i, j)\ne(t, r)$,

(4.1) $$|Cov(r_{ij}^2, r_{tr}^2)| \le ch_i h_t,$$

and

(4.2) $$Var(r_{ij}^2) = \tau_i + \zeta_i \quad \text{with } |\zeta_i|\le ch_i.$$

Proof. Let

$$u_{ijlp} = \begin{cases} 1-h_i & \text{if } l=i \text{ and } j=p \\ -h_{il} & \textit{otherwise.} \end{cases}$$

Then $r_{ij} = \sum_{l=1}^{n}\sum_{p=1}^{m_l} u_{ijlp} e_{lp}$ and

(4.3) $$Cov(r_{ij}^2, r_{tr}^2) = \sum_{l=1}^{n}\sum_{p=1}^{m_l} u_{ijlp}^2 u_{trlp}^2 \tau_l + 2\sum_{(l,p)\ne(m,q)} u_{ijlp}u_{ijmq}u_{trlp}u_{trmq}\sigma_l^2\sigma_m^2.$$

If $(i, j)\ne(t, r)$, then

$$|\sum_{l=1}^{n}\sum_{p=1}^{m_l} u_{ijlp}^2 u_{trlp}^2 \tau_l| \le 2\tau h_{it}^2 + \tau|\sum_{l=1}^{n}\sum_{p=1}^{m_l} h_{il}^2 h_{tl}^2| \le (2+k)\tau h_i h_t,$$

where $\tau = sup_l \tau_l$, and

$$|\sum_{(l,p)\ne(m,q)} u_{ijlp}u_{ijmq}u_{trlp}u_{trmq}\sigma_l^2\sigma_m^2| \le 2\sigma_U^4 h_{it}^2 + \sigma_U^4 h_i h_t(\sum_{l=1}^{n}\sum_{p=1}^{m_l} h_l)^2$$
$$\le (2+k^2)\sigma_U^4 h_i h_t.$$

Also, from (4.3),

$$Var(r_{ij}^2) = (1-h_i)^4\tau_i + \sum_{l\neq i}^{n} m_l h_{il}^4 \tau_l + \sum_{p\neq j}^{m_i} h_i^4 \tau_i + 2\sum_{(l,\,p)\neq(m,\,q)} u_{ijlp}^2 u_{ijmq}^2 \sigma_l^2 \sigma_m^2.$$

Note that

$$\sum_{l\neq i}^{n} m_l h_{il}^4 \tau_l + \sum_{p\neq j}^{m_i} h_i^4 \tau_i \leq \tau(1+k)h_i$$

and

$$\sum_{(l,\,p)\neq(m,\,q)} u_{ijlp}^2 u_{ijmq}^2 \sigma_l^2 \sigma_m^2 \leq 2\sigma_U^4 \sum_{l=1}^{n} m_l h_{il}^2 + \sigma_U^4 (\sum_{l=1}^{n} m_l h_{il}^2)^2 \leq 3\sigma_U^4 h_i.$$

Hence the result follows. □

Proof of Theorem 3. (i) From Theorem 2(i), the bias of $v_i^b(\lambda_i)$ is of order $O(h_i)$. Hence

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

$$+ \lambda_i^2 h_i^2 Var(s^2) + O(h_i^2).$$

By Lemma 3, $Var(s^2)$ is bounded. Therefore,

$$\lambda_i^2 h_i^2 Var(s^2) = O(h_i^2).$$

Also, from Lemma 3,

(4.4) $$Var(\hat{a}_i) = m_i^{-2}(1-h_i)^{-2}[\sum_{j=1}^{m_i} Var(r_{ij}^2) + 2\sum_{1\leq j<l\leq m_i} Cov(r_{ij}^2, r_{il}^2)]$$

$$= m_i^{-1}(1-h_i)^{-2}[\tau_i + O(h_i)],$$

and

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

The result follows.

(ii) We only give a proof for the case of $r=N-1$ for illustration. From (1.5), (4.4) and Theorem 2(ii), it suffices to show that

(4.5) $$Cov(\hat{a}_i, s_J^2) = O(h_{\max})$$

and

(4.6) $$Var(s_J^2) = O(1).$$

From Lemma 3, $\max_i Var(\hat{a}_i)=O(1)$. Since $s_J^2=h_i^{-1}\sum_{l=1}^n h_{il}^2 m_l \hat{a}_l$, (4.5) and (4.6) are implied by (4.4) and

$$\max_{i\neq j} |Cov(\hat{a}_i, \hat{a}_j)| = O(h_{\max}),$$

which follows directly from Lemma 3. □

5. The choice of λ_i.

A consequence of Theorem 3 is the following.

Theorem 4. Under the same conditions as in Theorem 3, for any $0\leq s<t\leq 1$, we have

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

if $h_{\max}\to 0$ as $N\to\infty$. The same result holds if v_i^b is replaced by v_i^w.

Proof. From Theorem 3, the difference of the MSE between $v_i^b(s)$ and $v_i^b(t)$ is

$$m_i^{-1}(1-h_i)^{-1}(t-s)[2-(t+s)h_i]h_i[\tau_i+O(h_i)] + O(h_i h_{\max}).$$

The result follows. The proof for v_i^w is the same. □

Thus, if $0\leq s<t\leq 1$, 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. If one wants to choose a variance estimator in terms of lower MSE, then a clear choice is $v_i^b(1)$ (or $v_i^w(1)$).

However, the MSE is not the only measure of the accuracy of an estimator. In practice, other measures of accuracy, such as the bias of the estimator, are also important. If we want to construct a confidence region for β by estimating σ_i^2 by v_i, the coverage probability of the confidence region will be too low if v_i always has a negative bias. See the discussion in Section 7 and the simulation results in Shao (1987).

A refined analysis of the biases of $v_i^b(\lambda_i)$ and $v_i^w(\lambda_i)$ gives the following theorem. The result indicates that $v_i^b(\lambda_i)$ (or $v_i^w(\lambda_i)$) with a smaller λ_i will usually have a smaller bias (in absolute value).

Theorem 5. Let $A_N = h_i^{-1}\sum_{l=1}^n h_{il}^2 m_l(\sigma_l^2 - \sigma_i^2)$, $B_N = (N-k)^{-1}\sum_{l=1}^n m_l(1-h_l)(\sigma_l^2-\sigma_i^2)$, and $0 \le s < t \le 1$. Assume that $h_i \to 0$ as $N \to \infty$.

(i) If $liminf_{N\to\infty} |A_N| > 0$, then

(4.7) $$liminf_{N\to\infty}[\,|Bias(v_i^w(t))| / |Bias(v_i^w(s))|\,] > 1.$$

(ii) If $liminf_{N\to\infty} |B_N| > 0$ and $A_N B_N \ge 0$, then (4.7) holds with v_i^w replaced by v_i^b.

Remarks. (1) The condition $liminf_{N\to\infty} A_N > 0$ (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 $A_N B_N \ge 0$ is satisfied for some balanced models. An example is model (5.9) of Wu (1986) or any model satisfying condition (5.4) of Wu (1986).

Proof. (i) Note that for any $0 \le t \le 1$,

$$\begin{aligned} Bias(v_i^w(t)) &= -t\, h_i\sigma_i^2 + \sum_{l=1}^n h_{il}^2 m_l(\sigma_l^2 - \sigma_i^2) + t\sum_{l=1}^n h_{il}^2 m_l \sigma_l^2 + o(h_i) \\ &= (1+t)\sum_{l=1}^n h_{il}^2 m_l(\sigma_l^2-\sigma_i^2) + o(h_i) = (1+t)h_i A_N + o(h_i). \end{aligned}$$

Hence

$$|Bias(v_i^w(t))| / |Bias(v_i^w(s))| = |(1+t)A_N + o(1)| / |(1+s)A_N + o(1)|.$$

Since $liminf_{N\to\infty} |A_N| > 0$,

$$liminf_{N\to\infty}[\,|Bias(v_i^w(t))| / |Bias(v_i^w(s))|\,] = (1+t)/(1+s) > 1.$$

(ii) For $0 \le t \le 1$,

$$\begin{aligned} Bias(v_i^b(t)) &= t(N-k)^{-1} h_i \sum_{l=1}^n m_l(1-h_l)(\sigma_l^2-\sigma_i^2) + \sum_{l=1}^n h_{il}^2 m_l(\sigma_l^2-\sigma_i^2) + O(h_i^2) \\ &= h_i[A_N + tB_N + O(h_i)]. \end{aligned}$$

Let $\xi_N = h_i^{-1}Bias(v_i^b(t))$, $\eta_N = h_i^{-1}Bias(v_i^b(s))$ and $\zeta_N = \xi_N/\eta_N$. If $liminf_{N\to\infty}|\zeta_N| = \infty$, the result follows. Suppose that $liminf_{N\to\infty}|\zeta_N| = a < \infty$. Then there is a subsequence $\{\zeta_{N(l)}\}$ such that either $\lim_{l\to\infty}\zeta_{N(l)}=a$ or $\lim_{l\to\infty}\zeta_{N(l)}=-a$. Since η_N are bounded, there is a subsequence $\{N(j)\}\subset\{N(l)\}$ such that $\lim_{j\to\infty}\eta_{N(j)}=\eta$. Then $\lim_{j\to\infty}\xi_{N(j)}=\lim_{j\to\infty}\zeta_{N(j)}\eta_{N(j)}$ equals either $a\eta$ or $-a\eta$. Note that

$$\xi_{N(j)}=A_{N(j)}+tB_{N(j)}+o(1),$$

$$\eta_{N(j)}=A_{N(j)}+sB_{N(j)}+o(1).$$

Hence the limits of $A_{N(j)}$ and $B_{N(j)}$ exist. Let $A=\lim_{j\to\infty}A_{N(j)}$ and $B=\lim_{j\to\infty}B_{N(j)}$. Under the conditions of the theorem, $B\neq 0$ and $A/B\geq 0$. Then

$$a=\lim_{j\to\infty}|\zeta_{N(j)}| = |A+tB|/|A+sB| = |1+(t-s)(A/B+s)^{-1}| > 1. \square$$

Hence for the choice of λ_i, we need to balance the advantage of having a smaller MSE against the drawback of a larger bias. λ_i may also be determined by other theoretical or practical considerations (Shao, 1987).

6. Comparisons between the REBE and other variance estimators when N is large.

We compare the REBE with the other variance estimators, such as the within-group sample variance, the MINQUE and the ARE, in the case that m_i are small but N is large. The case that N is also small is discussed in the next section.

6.1. The REBE and the within-group sample variance.

The customary variance estimator is the within-group sample variance

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

It can be shown by using Lemma 2 that

$$MSE(s_i^2)=m_i^{-1}\tau_i+2m_i^{-1}(m_i-1)^{-1}\sigma_i^4,$$

where $\tau_i=Var(e_{ij}^2)$, $j=1,\ldots,m_i$, $i=1,\ldots,n$. Thus from Theorem 3, we have

Theorem 6. Assume the conditions of Theorem 3. If (2.1) holds, then for any $\lambda_i \in [0,1]$,

$$MSE(s_i^2) - MSE(v_i^b(\lambda_i)) \to 2m_i^{-1}(m_i-1)^{-1}\sigma_i^4 > 0$$

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

6.2. The REBE and the MINQUE.

The MINQUE of σ_i^2 is generally of the form

(6.1) $\quad y'A_i y,$

where $A_i = (a_{pq}^{(i)})_{N \times N}$ is an symmetric matrix satisfying

(6.2) $\quad A_i X = 0.$

Note that all the REBE and s_i^2 are of the form (6.1)-(6.2). The MINQUE requires an additional unbiasedness requirement

(6.3) $\quad E(y'A_i y) = \sigma_i^2,$

which often leads to a negative estimate of σ_i^2. For example, if $m_i=1$ for all i, then (6.3) implies A_i is not non-negative definite. For if $A_i \geq 0$, then from (6.3), $a_{pq}^{(i)}=0$ unless $p=q=i$, and $a_{ii}^{(i)}=1$. Then (6.2) holds unless $x_i=0$.

Exact unbiasedness may not always provide a good estmator. As usual, a slightly biased estimator (the bias vanishes as the sample size tends to infinity) such as the REBE may perform better. Since the MSE of the MINQUE is not easy to obtain in general, we only compare the REBE with the MINQUE in the following two special but quite broad situations.

(a) A special case of model (1.1) is

$$y_{ij} = \mu + e_{ij}, \quad j=1,\ldots,m_i,\ i=1,\ldots,n,\ N=\sum_{i=1}^n m_i.$$

The MINQUE of σ_i^2 is

$$v_i^m = m_i^{-1}(N-2)^{-1}N\sum_{j=1}^{m_i}(y_{ij}-\bar{y})^2 - (N-2)^{-1}s^2,$$

where $\bar{y}=N^{-1}\sum_{i=1}^n\sum_{j=1}^{m_i}y_{ij}$, $s^2=(N-1)^{-1}\sum_{i=1}^n\sum_{j=1}^{m_i}(y_{ij}-\bar{y})^2$.

Theorem 7. Assume (2.1) and the conditions in Theorem 3. Then for any $\lambda_i \in [0,1]$,

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

as $N \to \infty$, where $\tau_i = Var(e_{ij}^2)$. The same result holds if $v_i^b(\lambda_i)$ is replaced by $v_i^w(\lambda_i)$.

Proof. Since v_i^m is unbiased,

$$MSE(v_i^m) = Var(v_i^m) = (N-2)^{-2}N^2 Var[m_i^{-1}\textstyle\sum_{j=1}^{m_i}(y_{ij}-\bar{y})^2] + O(N^{-2})$$

$$= m_i^{-1}(N-2)^{-2}N^2[\tau_i + O(N^{-1})] + O(N^{-2}).$$

Note that $h_i = N^{-1}$, $i=1,...,n$. From Theorem 3,

$$MSE(v_i^b(\lambda_i)) = m_i^{-1}(N-1)^{-2}(N-\lambda_i)^2[\tau_i + O(N^{-1})] + O(N^{-2}).$$

Since $(N-2)^{-2}N^2 - (N-1)^{-2}(N-\lambda_i)^2 = 2(1+\lambda_i)(N-1)^{-2}(N-2)^{-2}N^3 + O(N^{-2})$,

$$MSE(v_i^m) - MSE(v_i^b(\lambda_i)) = 2m_i^{-1}(1+\lambda_i)\tau_i(N-1)^{-2}(N-2)^{-2}N^3 + O(N^{-2}).$$

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

(b) Consider the general model (1.1) with $m_i = m$ for all i. For the case $m=1$, we also assmue that $h_i < 0.5$, $i=1,...,n$, to ensure the existence of v_i^m. A similar result to Theorem 7 can be obtained.

Theorem 8. Let v_i^m be the MINQUE of σ_i^2. Assume (2.1) and the conditions of Theorem 3. If $\lim_{N\to\infty}(h_{\max}^2/h_i) = 0$, then for any $\lambda_i \in [0,1]$,

$$\liminf_{N\to\infty} h_i^{-1}[MSE(v_i^m) - MSE(v_i^b(\lambda_i))] \geq 2m^{-1}(1+\lambda_i)\tau_i > 0.$$

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

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

(6.4) $$\max_i \sum_{1\leq p<q\leq n} |a^{ip}a^{iq}| \leq \max_i (\textstyle\sum_{p=1}^n |a^{ip}|)^2 < \infty.$$

From Lemma 4.5 of C. R. Rao (1970), $v_i^m =$ the ith component of $A^{-1}R$, where R is an n-vector whose ith component is $m^{-1}\sum_{j=1}^{m_i} r_{ij}^2$. Hence, (6.5) $$MSE(v_i^m) = \sum_{j=1}^n (a^{ij})^2(1-h_j)^2 Var(\hat{a}_j)$$

$$+ 2\sum_{1\le p<q\le n} a^{ip}a^{iq}(1-h_p)(1-h_q)Cov(\hat{a}_p, \hat{a}_q),$$

where $\hat{a}_i$ is defined in (1.2). From (4.1) and (6.4), the second term of the right hand side of (6.5) is $O(h_{\max}^2)$. Since $a^{ii} \ge a_{ii}^{-1} = (1-2h_i+mh_i^2)^{-1}$, the first term of the right hand side of (6.5) is not smaller than $m^{-1}a_{ii}^{-2}[\tau_i+O(h_i)]$. Thus,

$$MSE(v_i^m) \ge m^{-1}a_{ii}^{-2}[\tau_i+O(h_i)] + O(h_{\max}^2),$$

which and Theorem 3 imply that

$$MSE(v_i^m) - MSE(v_i^b(\lambda_i)) \ge m^{-1}[a_{ii}^{-2} - (1-h_i)^{-2}(1-\lambda_i h_i)^2][\tau_i+O(h_i)] + O(h_{\max}^2).$$

Note that

$$\begin{aligned} a_{ii}^{-2} - (1-h_i)^{-2}(1-\lambda_i h_i)^2 &= \{[1-h_i-(1-\lambda_i h_i)a_{ii}][1-h_i+(1-\lambda_i h_i)a_{ii}]\}/a_{ii}^2(1-h_i)^2 \\ &\ge [1-h_i-(1-\lambda_i h_i)(1-2h_i+mh_i^2)][2-O(h_i)] \\ &= 2(1+\lambda_i)h_i + O(h_i^2). \end{aligned}$$

Hence

$$\begin{aligned} MSE(v_i^m) - MSE(v_i^b(\lambda_i)) &\ge 2m^{-1}(1+\lambda_i)h_i[\tau_i+O(h_i)] + O(h_{\max}^2) \\ &= 2m^{-1}(1+\lambda_i)h_i\tau_i + O(h_{\max}^2). \end{aligned}$$

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

6.3. 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. The following result indicates that v_i^r has the same MSE as $v_i^b(1)$ (or $v_i^w(1)$) up to the order $O(h_i h_{\max})$, and generally has negative bias.

Theorem 9. The MSE of the ARE has the following expansion:

(6.6) $$MSE(v_i^r) = m_i^{-1}[\tau_i + O(h_i)] + O(h_i h_{max}).$$

If

(6.7) $$sup_{p \neq i} |\sigma_p^2 - \sigma_i^2| < \sigma_i^2,$$

then

(6.8) $$Bias(v_i^r) < 0.$$

Proof. Since $v_i^r = (1-h_i)\hat{a}_i$, the proof of (6.6) is the same as that of Theorem 3. (6.8) follows from (6.7) and

$$Bias(v_i^r) = -h_i \sigma_i^2 + \sum_{l=1}^n h_{il}^2 m_l (\sigma_l^2 - \sigma_i^2). \square$$

Condition (6.7) is clearly not necessary for (6.8) (see Section 7). Because of (6.8), the confidence regions for β obtained by using v_i^r as the estimators of σ_i^2, $i=1,...,n$, usually have low coverage probabilities. See the simulation results in Shao (1987).

Note that

(6.9) $$v_i^b(1) = v_i^r + h_i s^2$$

and

(6.10) $$v_i^w(1) = v_i^r + h_i s_J^2.$$

Hence $v_i^b(1)$ and $v_i^w(1)$ are bias adjustments of v_i^r. The second terms of the right hand side of (6.9) and (6.10) are positive but are of low orders so that they have small effects on the MSE of $v_i^b(1)$ and $v_i^w(1)$.

7. The case of small N: an example.

When N is small (consequently, m_i and n are small), it is hard to compare variance estimators analytically, and there is no definite conclusion in general. The improvements by using empirical Bayesian methods become "small", since there is little auxiliary information to be used.

We compare the REBE with other variance estimators through the following example. Consider the model

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

Let

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

This model can also be viewed as a two sample problem. If m_i are large, the estimators under comparison perform equally well. The MINQUE, s_i^2 and the REBE $v_i^w(\lambda_i)$ ($0\le\lambda_i\le1$) in this case are the same and equal to

$$(m_i-1)^{-1}SS_i, \quad i=1,2.$$

Hence the use of the MINQUE and $v_i^w(\lambda_i)$ does not achieve any improvement on s_i^2. v_i^r equals

$$m_i^{-1}SS_i.$$

All the above estimators do not use the data from the other group. The REBE $v_i^b(\lambda_i)$ equals

$$v_i^b(\lambda_i) = (1-c_i)SS_i/(m_i-1) + c_i(SS_1+SS_2)/(N-2), \quad c_i=\lambda_i/m_i.$$

Note that $s_p^2=(SS_1+SS_2)/(N-2)$ is the pooled variance estimator when σ_i^2 are assumed to be equal or nearly equal. The REBE $v_i^b(\lambda_i)$ is a compromise between the within-group sample variance s_i^2 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. The MINQUE and s_i^2 are unbiased. The bias of v_i^r is $-m_i^{-1}\sigma_i^2$, which is always negative and can be very large. The bias of $v_i^b(\lambda_i)$ is

$$\begin{aligned}&\lambda_i(m_i-1)(\sigma_j^2-\sigma_i^2)/m_i(N-2), \quad j\neq i\\ &= \lambda_i(\sigma_j^2-\sigma_i^2)/2m, \quad \text{if } m_1=m_2=m.\end{aligned}$$

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. $v_i^b(\lambda_i)$ will perform well if σ_i^2 are close (since we use the data from two groups). The bias of $v_i^b(\lambda_i)$ may be small even if m_i is not large.

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 respectively

$$2\sigma_i^4/(m_i-1) \quad \text{and} \quad (2m_i-1)\sigma_i^4/m_i^2.$$

For $t \in [0,1]$,

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

where $\theta_1=\sigma_2^2/\sigma_1^2$ and $\theta_2=\sigma_1^2/\sigma_2^2$. This is a decreasing function of t when $0<\theta_i\le(3m-1)/(m+1)$. Hence if $\max(\theta_1, \theta_2)\le(3m-1)/(m+1)$, then for $s<t$,

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

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 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

$$v_i^c=(m+1)^{-1}SS_i.$$

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 $v_i\equiv0$ has MSE half that of s_i^2! As we commented earlier, the MSE should not be the only criterion for choosing an estimator.

8. Estimating linear functions of σ_i^2.

We consider in this section the estimation of $\eta=\sum_{i=1}^n l_i\sigma_i^2$, where l_i are known constants.

For example, η is the (p, q)th element of $Var\hat{\beta}$, the variance-covariance matrix of $\hat{\beta}$, if $l_i=$ the (p, q)th element of $m_iM^{-1}x_ix_i'M^{-1}$. Again we assume that m_i are small but N is large. Consider the following general class of REBE's:

$$\hat{\eta} = \sum_{i=1}^n l_i[(1-B_i)\hat{a}_i+B_i\bar{a}_i], \tag{8.1}$$

where $\hat{a}_i$ is defined in (1.2), $\bar{a}_i$= either s^2 or $s^2_{J,r}$, B_i possibly depend on data and satisfy $0 \le B_i \le 1$ and

(8.2) $$\max_{i \le n} sup_y B_i(y) \to 0 \quad \text{as } N \to \infty.$$

Note that $\sum_{i=1}^n l_i v_i^b(\lambda_i)$ and $\sum_{i=1}^n l_i v_i^w(\lambda_i)$ are special cases of (8.1).

Theorem 10. Assume (2.1) and the conditions of Theorem 3. If

(8.3) $$\sum_{i=1}^n |l_i| = O(N^{-1})$$

and

(8.4) $$\sum_{i=1}^n l_i^2 = o(N^{-2}),$$

then $\hat{\eta}$ defined in (8.1)-(8.2) satisfies

(8.5) $$MSE(\hat{\eta}) = o(N^{-2}).$$

Remarks. (1) (8.5) means that $\hat{\eta}$ is consistent in a stronger sense that

$$N^2 E(\hat{\eta} - \eta)^2 \to 0.$$

The asymptotic unbiasedness and consistency of $\hat{\eta}$ follow from (8.5).

(2) The REBE's 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_i will stablize the variance of $\sum_{i=1}^n l_i v_i$ and therefore $\hat{\eta}$ is consistent as $N \to \infty$. Conditions (8.3) and (8.4) are the smoothness conditions for l_i. These conditions are quite weak. See the corollary below.

(3) For the asymptotic unbiasedness of $\hat{\eta}$, (8.3) is sufficient.

Proof. From (8.2), (8.3) and Theorem 2,

$$N\,|E\hat{\eta}-\eta| = N\,|\sum_{i=1}^{n} l_i E(B_i\hat{a}_i)-\sum_{i=1}^{n} l_i E(B_i\bar{a}_i)| + o(1)$$

$$\leq c_1(\max_{i\leq n} sup_y B_i)[\sum_{i=1}^{n} |l_i|(E\hat{a}_i+E\bar{a}_i)] + o(1) \to 0$$

since $E\hat{a}_i$ and $E\bar{a}_i$ are bounded, where c_1 is a positive constant. Also, from Lemma 3,

$$\max_{i\leq n} Var(\hat{a}_i) = O(1)$$

and

$$\max_{i\neq j} |Cov(\hat{a}_i, \hat{a}_j)| \leq [m_i m_j(1-h_i)(1-h_j)]^{-1}\sum_{p=1}^{m_i}\sum_{q=1}^{m_j} |Cov(r_{ip}^2, r_{jq}^2)| = O(h_{\max}).$$

Then from (8.2) and (8.4),

$$N^2 Var(\hat{\eta}) \leq c_2 N^2 \sum_{i=1}^{n} l_i^2 + c_3 N^2 h_{\max}(\sum_{i=1}^{n} |l_i|)^2 \to 0,$$

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

The following result provides a class of asymptotically unbiased and consistent estimators of $Var\hat{\beta}$.

Corollary. Let l_i^{pq} be the (p,q)th element of $m_i M^{-1}x_i x_i' M^{-1}$, $Var\hat{\beta}=(\eta_{pq})_{k\times k}$, and $\hat{\eta}_{pq}=\sum_{i=1}^{n} l_i^{pq} v_i$, where $v_i=v_i^b(\lambda_i)$ or $v_i^w(\lambda_i)$, $\lambda_i\in[0,1]$. Assume (2.1), $M^{-1}=O(N^{-1})$ and $m_i\leq m_o$ for all i. Then

$$E(\hat{\eta}_{pq}-\eta_{pq})^2 = o(N^{-2}).$$

Proof. We only need to check (8.3) and (8.4). From $M^{-1}=O(N^{-1})$, there is a constant $c>0$ such that $M^{-1} \le cN^{-1}I_{k\times k}$. Then

$$\sum_{i=1}^{n} |l_i^{pq}| \le \sum_{i=1}^{n} m_i x_i' M^{-2} x_i \le cN^{-1} \sum_{i=1}^{n} m_i h_i = kcN^{-1}.$$

Similarly,

$$\sum_{i=1}^{n} (l_i^{pq})^2 \le \sum_{i=1}^{n} m_i^2 (x_i' M^{-2} x_i)^2 \le c^2 \sum_{i=1}^{n} m_i^2 h_i^2 \le c^2 m_o k h_{\max}.$$

This completes the proof. □

REFERENCES

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

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

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). On resampling methods for variance and bias estimation in linear models. Tech. Report 788, Dept. of Statistics, University of Wisconsin-Madison.

Shao, J. (1987). Estimating heteroscedastic variances in linear models I: a resampling empirical Bayesian approach. Tech. Report, Dept. of Statistics, Purdue University.

Shao, J. and Wu, C. F. J. (1987). Heteroscedasticity-robustness of jackknife variance estimators in linear models. *Ann. Statist.*, 15, to appear.

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

Estimating Heteroscedastic Variances in Linear Models II: Properties of the Resampling Empirical Bayes Estimators

by

Jun Shao
Purdue University

Purdue University

Technical Report #87-43

Department of Statistics
Purdue University

October 1987