Estimating Heteroscedastic Variances in Linear Models I: A Resampling Empirical Bayesian Approach

by

Jun Shao
Purdue University

Purdue University

Technical Report #87-42

Department of Statistics
Purdue University

October 1987

Estimating Heteroscedastic Variances in Linear Models I:

A Resampling Empirical Bayesian Approach

by
Jun Shao
Purdue University

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. The Resampling Empirical Bayes Estimators (REBE's) are shrinkage type estimators in general, and are always positive and analytically tractable. Their properties, including invariance, robustness, consistency, asymptotic unbiasedness and mean squared error (MSE) are discussed. In particular, a second order expansion of the MSE and an upper bound of the bias of the REBE are given in terms of the diagonal elements of the projection matrix. Using these results, we compare the REBE with other existing variance estimators. The MSE of the REBE is smaller than that of the customary estimator, i.e., the within-group sample variance, and the MINQUE when N is large. The problem of estimating a linear function of the error variances is also considered. Some simulation results are presented.

KEY WORDS: Data resampling; Empirical Bayes estimators; Shinkage estimators; Sample variance; MINQUE.

1. INTRODUCTION

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

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

Here y_{ij} is the response of the j-th replicate in the i-th group, x_i is a $k\times 1$ deterministic vector, β is a $k\times 1$ 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 quantity of interest is the parameter 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 inferences such as setting confidence regions for β. Also, one may utilize the estimates of σ_i^2 in improving the estimates of β.

The customary estimator of σ_i^2, the sample variance within the i-th group, is questionable when m_i is small (actually it is not defined when $m_i=1$). 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 n 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 although they are different.

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

(i) One assumes that σ_i^2 is a function of the design point x_i and possibly some other unknown quantities, i.e., $\sigma_i^2=H(x_i, \theta)$, and then obtains estimates of σ_i^2 by estimating H and θ. For more details of this approach, see Carroll (1982) and its references.

(ii) Instead of putting some restrictions on σ_i^2, one can impose some restrictions (such as unbiasedness and invariance) on the estimators, e.g., the MINQUE (C. R. Rao, 1970) and its modifications (J. N. K. Rao, 1973; Horn et al., 1975).

The MINQUE has the following well known deficiencies: (a) the MINQUE may not exist; (b) the MINQUE requires large computations; and more seriously, (c) the MINQUE can be negative. In fact, if $m_i=1$ for all i and $\sigma_i^2\neq\sigma_j^2$ for $i\neq j$, then the MINQUE of σ_i^2 is of the form $y'A_iy$, $y=(y_{11}\ y_{12}\ \cdots\ y_{nm_n})'$, with a symmetric matrix A_i which is *not* nonnegative definite (Shao, 1987). 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 in general but tends to underestimate σ_i^2. 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 propose a class of estimators by using the empirical Bayesian method incorporating data resampling techniques. The Bayes estimators are given in Section 2. Due to the heteroscedasticity of the model, the Bayes estimators are not analytically tractable. Also, the Bayes estimators are derived based on the assumption that the errors are normally distributed, and therefore their optimality may be lost under the violation of the normality assumption. However, the empirical Bayes estimators derived in Section 3 are analytically tractable and robust against non-normality. In Section 4, we discuss 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 of the bias of the empirical Bayes estimators are given in terms of diagonal elements of the projection (hat) matrix. Using these results, we compare the empirical Bayes estimators with other variance estimators. The MSE of the empirical Bayes estimators is 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 larger negative bias.

The empirical Bayes estimators are generally of the form

$$(1.2)\qquad B_i\hat{a}_i+(1-B_i)\bar{a}_i,\quad 0\le B_i\le 1,$$

which is a compromise between a local estimator $\hat{a}_i$ using the residuals within the i-th group (see (3.4)) and an ensemble estimator $\bar{a}_i$. A similar type of estimator was studied by Morris (1983) in a problem of estimating normal means, where B_i is called the shrinking coefficient. As Morris discussed in his paper, the shrinkage estimator is superior to the classical method in many statistical applications because the estimators of type (1.2) incorporate the auxiliary information provided by the data in other groups.

In Section 5, by using the estimators of σ_i^2 proposed in Section 3, we consider the estimation of linear functions of σ_i^2 (e.g., the variances and covariances of the least squares estimator (LSE) of β). Some extensions of the results are discussed in Section 6. The last section contains some simulation results, which indicate that the performances of the empirical Bayes estimators are generally better than those of the other variance estimators mentioned above.

2. BAYES ESTIMATORS

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

(2.1) $$\pi_i(\tau_i) \propto \tau_i^{p_i} exp(-\alpha_i\tau_i) \quad \tau_i>0,$$

and $p_i>0$, $\alpha_i>0$ are known constants. For the parameter β, we only assume that β is independent of τ_i, $i=1,\dots,n$, and has a known prior density $\pi(\beta)$ with respect to a measure μ on $\mathbf{R}^k$. Using standard techniques in Bayesian analysis, we obtain the following Bayes estimator of σ_i^2 (under squared error loss):

(2.2) $$v_i^B=\frac{2p_i}{2p_i+m_i}a_i+\frac{1}{2p_i+m_i}\sum_{j=1}^{m_i}(y_{ij}-\bar{y}_i)^2+\frac{m_i}{2p_i+m_i}\int(\bar{y}_i-x_i'\beta)^2 p(\beta|y)d\mu,$$

where $\bar{y}_i=m_i^{-1}\sum_{j=1}^{m_i}y_{ij}$, $a_i=E\sigma_i^2=E(2\tau_i)^{-1}=\alpha_i/(2p_i)$ is the prior mean of σ_i^2, and

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

is the posterior density of β.

In practice, if little is known about p_i, α_i and $\pi(\beta)$, the Bayesian analysis can be carried out via the following two commonly used techniques:

(i) The empirical Bayesian method, which we will discuss in the next section. The nonexchangeability of the prior (2.1) does not cause any difficulty in applying this method.

(ii) The hierarchical Bayesian method (Lindley, 1971). This method needs to assume that the prior of σ_i^2 (or the prior of p_i and α_i) is exchangeable in order to reduce the dimension of the parameter space.

Another form of v_i^B, which is quite easy to see from (2.2), is that

$$v_i^B=\frac{2p_i}{2p_i+m_i}a_i+\frac{1}{2p_i+m_i}\sum_{j=1}^{m_i}(y_{ij}-\bar{y}_i)^2+\frac{m_i}{2p_i+m_i}[(\bar{y}_i-x_i'\hat{\beta}_B)^2+x_i'\Sigma_y x_i], \tag{2.3}$$

where $\hat{\beta}_B$ and Σ_y are the posterior mean of β (Bayes estimator of β) and posterior variance of β, respectively. From (2.2), the Bayes estimator is a mixture of three components: the prior information, the within-group variation of y_{ij} (this component is zero when m_i=1) and a smooth average of the squared "residuals" $\bar{y}_i-x_i'\beta$, which captures the information from fitting model (1.1). Equation (2.3) further decomposes the third term on the right hand side of (2.2) into a squared residual obtained by estimating β by the Bayes estimator $\hat{\beta}_B$ and the variance of the posterior $p(\beta|y)$. (2.3) allows us to approximate the Bayes estimator by estimating the first two moments of the posterior density $p(\beta|y)$. See Section 3.

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 k, the dimension of β, is large. Also, v_i^B is obtained based on normality assumptions, and hence may not perform well under violation of these assumptions. Without normality assumptions, the Bayes estimators are much harder to evaluate and interpret. On the other hand, the empirical Bayes estimators derived in the next section are analytically tractable, easy to evaluate and quite robust against non-normality.

3. RESAMPLING EMPIRICAL BAYES ESTIMATORS

The Bayes estimator (2.2) depends on hyperparameters a_i, p_i and the posterior density $p(\beta|y)$. If some of these quantities are unknown, they can be estimated by the data. The resulting estimators are known as empirical Bayes estimators.

Two classes of empirical Bayes estimators of σ_i^2 are derived by using the usual moment estimators of the mean of the prior $\pi_i(\tau_i)$ and data resampling techniques to estimate the mean and variance of the posterior $p(\beta|y)$. The properties of the obtained estimators, which are called Resampling Empirical Bayes Estimators (REBE's) henceforth, are discussed in the next section.

To begin our procedure, suppose that the parameter β is estimated by $\hat{\beta}$. In this and the next two sections we will assume that $\hat{\beta}$ is obtained by using the ordinary least squares methods, i.e.,

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

where $y=(y_{11}\, y_{12}\, \cdots\, y_{nm_n})'$, $X=(X_1\, X_2\, \cdots\, X_n)$, X_i is a $k\times m_i$ matrix whose columns are x_i, $i=1,\ldots,n$, and $M=X'X$ is assumed to be nonsingular. Although all the results in this and next two sections are still valid if we use instead the weighted least squares estimator (see Section 6), we use the ordinary LSE for the following reasons:

(i) Since σ_i^2 are unknown, choosing adequate weights $\{w_i\}$ is hard. The weighted least squares method may not provide a better estimator of β even if one has estimates of σ_i^2. See the results in Section 7 and Jacquez et al. (1968) and Rao (1970).

(ii) When the weights are functions of data, the precision of the weighted least squares estimator is not tractable and needs to be estimated.

(iii) The inverse of the weight w_i can be thought of a prior guess of σ_i^2. When little is known about σ_i^2, the ordinary least squares method simply uses a noninformative prior guess of σ_i^2, i.e., $w_i\equiv$ constant.

3.1. Estimating the Prior Mean of σ_i^2

We start with the estimation of the prior mean a_i. Define the residuals by

$$r_{ij}=y_{ij}-x_i'\hat{\beta},\quad j=1,\ldots,m_i,\ i=1,\ldots,n.$$

The marginal mean of r_{ij}^2 is

$$\int r_{ij}^2 p(y)dy=\iiint r_{ij}^2 p(y\mid\tau,\beta)\pi(\tau)\pi(\beta)dyd\tau d\mu,$$

where $\tau=(\tau_1,\ldots,\tau_n)'$, $p(y\mid\tau,\beta)$ is the density of y given τ and β, $\pi(\tau)=\prod_{i=1}^n\pi_i(\tau_i)$, $\pi_i(\tau_i)$ and $\pi(\beta)$ are priors of τ_i and β, and $p(y)$ is the marginal density of y. Note that we do not assume that $p(y\mid\tau,\beta)$ is normal. Let $h_{il}=x_i'M^{-1}x_l$ and $h_i=h_{ii}$. Since

$$\int r_{ij}^2 p(y\,|\,\tau,\beta)dy=(1-h_i)\sigma_i^2+\sum_{l=1}^n h_{il}^2 m_l(\sigma_l^2-\sigma_i^2),\quad j=1,\ldots,m_i,\ i=1,\ldots,n,$$

which does not depend on β, we have

$$\int r_{ij}^2 p(y)dy=E_\tau E r_{ij}^2=(1-h_i)a_i+\sum_{l=1}^n h_{il}^2 m_l(a_l-a_i),\quad j=1,\ldots,m_i,\ i=1,\ldots,n, \tag{3.1}$$

where E_τ and E denote the expectations taken under the distributions $\pi(\tau)$ and $p(y\,|\,\tau,\beta)$, respectively.

If $a_i=a$, i.e., the prior means are all equal, we have

$$E_\tau E r_{ij}^2=(1-h_i)a,\quad j=1,\ldots,m_i.$$

Note that $\sum_{i=1}^n \sum_{j=1}^{m_i}(1-h_i)a=(N-k)a$. Hence the moment estimator of a is

$$s^2=(N-k)^{-1}\sum_{i=1}^n\sum_{j=1}^{m_i} r_{ij}^2. \tag{3.2}$$

However, if a_i are not all equal, we have

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

from (3.1). One could obtain the estimates of a_i by solving the linear system

$$[m_i(1-h_i)]^{-1}\sum_{j=1}^{m_i} r_{ij}^2=a_i+\sum_{l=1}^n(1-h_i)^{-1}h_{il}^2 m_l(a_l-a_i),\quad i=1,\ldots,n. \tag{3.3}$$

We will not do so for the following reasons: (i) The linear system (3.3) may not be solvable. (ii) Even if the solution of (3.3) exists, solving (3.3) may involve a large number of computations. (iii) The solution $\hat{a}_i$ may not be nonnegative.

We will instead use almost unbiased estimates

$$\hat{a}_i=[m_i(1-h_i)]^{-1}\sum_{j=1}^{m_i} r_{ij}^2, \tag{3.4}$$

which provide an approximate solution of (3.3).

The estimators $\hat{a}_i$ can also be used even if the a_i are equal. On the other hand, s^2 cannot be used if some of the a_i are not equal. We will employ $\hat{a}_i$ as an estimate of a_i in the sequel.

3.2. Estimating Posterior Moments of β

From (2.3), the Bayes estimator depends on the mean and variance of the posterior density $p(\beta|y)$. We now apply data resampling methods to approximate these moments. For given y, the resampling distributions described below have some similarities to the posterior distribution of β, e.g., they are both close to normal when n is large. In fact $p(\beta|y)$ is normal when σ_i^2 are equal and $\pi(\beta)$ is normal. Even though the resampling distributions may not be very close to the posterior of β, their first two moments may be close. When σ_i^2 are known, Lindley and Smith (1972) showed that if the prior of β is chosen to be noninformative, the posterior mean and variance of β is the same as the weighted least squares estimator and its variance, which are equal to (or very close to) the mean and variance of the resampling distributions.

There are many different data resampling techniques in the statistical literature. We describe and use two of them as follows.

(i) *Bootstrapping residuals* (Efron, 1979). For given y, let e^* be an N-vector whose components are i.i.d. samples from the normalized residuals $\{ (r_{ij}-\bar{r})/(1-k/N)^{1/2}$, $j=1,\ldots,m_i$, $i=1,\ldots,n$ $\}$, where $\bar{r}=N^{-1}\sum_{i=1}^{n}\sum_{j=1}^{m_i} r_{ij}$. Treat e^* as an error vector and $y^*=X\hat{\beta}+e^*$ as the observed data. The corresponding LSE is

$$\beta^*=M^{-1}X'y^*=\hat{\beta}+M^{-1}X'e^*.$$

Denote the expectation under the bootstrap distribution (given y) by E_y^*. We replace the posterior mean and variance of β in (2.3) by the mean and variance of the bootstrap distribution, which are respectively

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

where $s_b^2=s^2-N\bar{r}^2/(N-k)$, s^2 is defined in (3.2). $N\bar{r}^2/(N-k)$ is equal to zero when the first components of x_i, $i=1,\ldots,n$, are all equal to one, and generally has a lower order than s^2. For simplicity we ignore this lower order 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 hand side of (2.3) is approximated by

$$\frac{m_i}{2p_i+m_i}[(\bar{y}_i-x_i'\hat{\beta})^2+h_i s^2].$$

Assume that p_i in (2.1) are known and let

(3.5) $$\lambda_i=\frac{m_i}{2p_i+m_i}.$$

Then the resulting empirical Bayes estimator of σ_i^2 is equal to

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

which employs $\hat{a}_i$ (3.4) as an estimate of a_i.

Note that v_i^b is of the form (1.2) and is a convex combination of $\hat{a}_i$ and s^2, and s^2 is 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 weighted resampling distribution. By weighted resampling we mean that for given data y, select a subset model

(3.7) $$y_s=X_s\beta+e_s$$

with probability

(3.8) $$W_s \propto |X_s'X_s|,$$

where $s=\{ i_1,\ldots,i_r \}$ is a subset of $\{ 1,\ldots,N \}$, $r\leq N$, y_s, X_s and e_s are sub-vector and/or sub-matrix of y, X and e consisting of the i_1-th,...,i_r-th rows of y, X and e, respectively. For more details of this weighted resampling procedure, see Shao (1986).

Denote the LSE of β under subset model (3.7) by $\hat{\beta}_s$ and the expectation under the weighted resampling distribution (given y) by E_y^s. Then

(3.9) $$E_y^s\hat{\beta}_s=\hat{\beta} \quad \text{and} \quad E_y^s(\hat{\beta}_s-\hat{\beta})(\hat{\beta}_s-\hat{\beta})'=\sum_r W_s(\hat{\beta}_s-\hat{\beta})(\hat{\beta}_s-\hat{\beta})'=(N-r)(r-k+1)^{-1}V_J,$$

where $\sum_r$ is the summation over all distinct subsets s of size r and V_J is the weighted retain-r (or delete-d, $d=N-r$) jackknife estimator of the variance-covariance matrix of $\hat{\beta}$ (Wu, 1986). The unequal probability (3.8) in data resampling procedure takes account of the unbalanced nature of the regression data, since W_s is proportional to the determinant of the Fisher information matrix of the corresponding subset model (3.7) with i.i.d. errors. It is more appropriate to use a scaled variance of the resampling distribution, i.e., $(N-r)^{-1}(r-k+1)E_y^s(\hat{\beta}_s-\hat{\beta})(\hat{\beta}_s-\hat{\beta})'$, to estimate the variance-covariance matrix of $\hat{\beta}$ for the reason of moment matching (Shao, 1986). Thus, the resulting empirical Bayes estimator of σ_i^2 (again we assume p_i are known) is equal to

$$(3.10) \qquad v_i^w=(1-\lambda_i h_i)\hat{a}_i+\lambda_i h_i s_J^2,$$

where $s_J^2=h_i^{-1}x_i'V_J x_i$ and λ_i is defined in (3.5). This estimator is also of the form (1.2). If $r=N-1$ ($d=1$), $s_J^2=h_i^{-1}\sum_{l=1}^n h_{il}^2 m_l \hat{a}_l$ is another weighted average of $\hat{a}_i$'s.

3.3. The Hyperparameters λ_i

In the above procedure the hyperparameters λ_i (or p_i, see (3.5)) are assumed to be known. If little is known about λ_i, one may also use data to estimate λ_i. For example, since

$$\lambda_i=m_i/[2(SN_i+1)+m_i],$$

where $SN_i=(E\sigma_i^2)^2/Var(\sigma_i^2)$ is the signal-noise ratio of the prior distribution of σ_i^2, λ_i can be estimated by estimating the signal-noise ratio SN_i from the data. 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 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 prior mean.

However, the sampling properties of this kind of estimator are hardly known. Alternatively, we can let λ_i in (3.6) and (3.10) range over [0, 1] to obtain two classes of REBE's. Then choose an appropriate λ_i in terms of the sampling properties of the REBE under certain criteria. This will be discussed in the next section. Note that from (3.5), $0<\lambda_i<1$ since

$0<p_i<\infty$. But for the REBE, we can include the limiting cases $\lambda_i=0$ and $\lambda_i=1$.

4. PROPERTIES OF THE REBE AND COMPARISONS

Some properties of a class of REBE's: $v_i^b(\lambda_i)$ defined in (3.6) with $\lambda_i \in [0, 1]$, are studied in this section. In terms of MSE and biases, comparisons of the REBE's with different λ_i as well as comparisons of the REBE's and other variance estimators, such as the within-group sample variance, the MINQUE and the ARE, are given. The results are obtained for the situation where the m_i are small, but N is large. A discussion for the case of small N is given in Shao (1987). All the results stated in this section also hold for another class of REBE's: $v_i^w(\lambda_i)$ defined in (3.10).

The proofs of the results are omitted. The detailed proofs can be found in Shao (1987).

(i) *Invariance under the translation of* β. All the REBE's obtained in Section 3 are invariant under the translation of β since they depend on residuals (or residuals from fitting the subset model (3.7)). The Bayes estimators obtained in Section 2 are invariant iff the prior distribution of β is invariant. If $\pi(\beta)$ is a density with respect to the Lebesgue measure on $\mathbf{R}^k$, then the Bayes estimators are not invariant. This assertion follows since if $\pi(\beta)$ is invariant, then $\pi(\beta-\beta_o)=\pi(\beta)$ for all $\beta_o \in \mathbf{R}^k$, which implies that $\pi(\beta)$ is improper.

(ii) *Consistency when* $m_i \to \infty$. As mentioned in Section 1, in common situations m_i are rarely large. However, there are still some statistical applications which involve large m_i and small n (e.g., a two sample comparison problem). Also, consistency (as $m_i \to \infty$) is a basic property of an estimator. Any *inconsistent* estimator should not be used.

When $m_i \to \infty$ (hence $N \to \infty$), under some minor conditions such as

(4.1) $\qquad \sigma_i^2 \le \sigma_U^2$ for all i,

(4.2) $\qquad h_{\max} = \max_{i \le n} h_i \to 0$ as $N \to \infty$,

$v_i^b(\lambda_i)$ is consistent for any $\lambda_i \in [0,1]$.

(iii) *Asymptotic unbiasedness when* $N\to\infty$. From now on we consider the case that the m_i are fixed and $N\to\infty$. Under (4.1), the bias of $v_i^b(\lambda_i)$ is of the order $O(h_i)$ for any $\lambda_i\in[0, 1]$. Hence $v_i^b(\lambda_i)$ is asymptotically unbiased if $h_i\to 0$ as $N\to\infty$.

(iv) *Mean squared error.* The exact forms of the MSE of the REBE's are extremely complicated due to the non-identical distributions of the errors. Instead we obtain an asymptotic expansion of the MSE ($N\to\infty$). Assume that $\rho_i=Var(e_{ij}^2)$ exist and are uniformly bounded. Then the MSE of $v_i^b(\lambda_i)$ has the following expansion:

$$m_i^{-1}(1-h_i)^{-2}(1-\lambda_i h_i)^2\rho_i+m_i^{-1}(1-h_i)^{-2}(1-\lambda_i h_i)^2 O(h_i)+O(h_i h_{\max}). \tag{4.3}$$

(v) *Choice of* λ_i: *A comparison of* REBE's. We now compare $v_i^b(\lambda_i)$ with different λ_i. An immediate consequence of (4.3) is that for any $0\le s<t\le 1$,

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

if $h_{\max}\to 0$ as $N\to\infty$. This says that for $s<t$, the MSE of $v_i^b(t)$ is less than that of $v_i^b(s)$ when N is large. Hence in terms of lower MSE, $v_i^b(1)$ is a clear choice.

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 may not use an estimator of σ_i^2 which has a trend in its bias. When 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 bias of $v_i^b(\lambda_i)$ shows that $v_i^b(\lambda_i)$ with smaller λ_i usually has smaller bias. That is, for $0\le s<t\le 1$,

$$\liminf_{N\to\infty} |Bias(v_i^b(t))|/|Bias(v_i^b(s))| > 1 \tag{4.4}$$

under some conditions (see Theorem 5 of Shao, 1987).

The picture is clear now. If one uses MSE as the measure of accuracy, then $v_i^b(\lambda_i)$ with large λ_i is preferred. On the other hand, if one is concerned about the bias of the estimator, then $v_i^b(\lambda_i)$ with smaller λ_i is better. In general, one should balance the advantage of having a smaller bias against the drawback of a large MSE.

(vi) *Robustness against non-normality.* The REBE's are derived based on the Bayes estimator v_i^B (2.2), which is derived under the assumption that the errors are normally distributed. However, the above results (and also the comparisons below) are true *with or without* normality assumptions. Hence the REBE's are robust against non-normality.

(vii) *Comparisons between the* REBE *and other variance estimators.* We first focus on the MSE. For the within-group sample variance

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

we have for any $\lambda_i\in[0, 1]$,

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

if $h_{max}\to 0$ as $N\to\infty$.

The MSE of the MINQUE is not easy to obtain, even in approximate form. We only compare the REBE with the MINQUE in the following two special but quite broad situations.

(a) Consider a special case of model (1.1):

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

Let v_i^m be the MINQUE of σ_i^2. Then for any $\lambda_i\in[0, 1]$,

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

where $\rho_i=Var(e_{ij}^2)$.

(b) Consider the general linear model (1.1) but assume $m_i=m$ for all i. For the case $m=1$, we also assume that $h_i<0.5$ to ensure that v_i^m exists. Then for any $\lambda_i\in[0, 1]$,

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

Results (4.5)-(4.7) indicate that the REBE has a smaller MSE than s_i^2 and the MINQUE.

Finally, the MSE of the ARE

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

has the same expansion as that of $v_i^b(1)$.

We now consider the bias. The sample variance s_i^2 and the MINQUE are unbiased. The REBE's are biased but their biases vanish as $N\to\infty$. v_i^r is generally too small as an estimator of σ_i^2, i.e., it has a negative bias. Since

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

$\text{Bias}(v_i^r)<0$ if all the σ_i^2 are not far away from each other, say

$$sup_l\,|\sigma_l^2-\sigma_i^2|<\sigma_i^2.$$

This condition is clearly not necessary. The confidence regions of β constructed by using v_i^r as estimators of σ_i^2 are usually of low coverage probability. See Section 7.

In fact, both $v_i^b(1)$ and $v_i^w(1)$ can be thought of bias adjustments of v_i^r, since

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

and

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

5. ESTIMATING LINEAR FUNCTIONS OF σ_i^2

Very often one wants to estimate a linear function of σ_i^2, i.e., $\eta=\sum_{i=1}^n l_i\sigma_i^2$, where l_i are constants. For example, suppose that we want to estimate $Var\hat{\beta}$, the variance-covariance matrix of the LSE $\hat{\beta}$. Let η_{pq} be the (p,q)-th entry of $Var\hat{\beta}$. Then l_i= the (p,q)-th entry of $m_iM^{-1}x_ix_i'M^{-1}$.

A natural estimator of η is $\hat{\eta}=\sum_{i=1}^n l_iv_i$, where v_i is an estimator of σ_i^2, $i=1,\ldots,n$. Thus, we can apply the results in the previous sections.

Again we assume that all m_i are small but N is large. The estimators of σ_i^2 in previous sections are not consistent ($N\to\infty$) 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 $\sum_{i=1}^n l_iv_i$. Hence the resulting estimator $\hat{\eta}=\sum_{i=1}^n l_iv_i$ is consistent as $N\to\infty$. By smoothness of l_i we mean that l_i satisfy

(5.1) $$\sum_{i=1}^n |l_i|=O(N^{-1}) \quad\text{and}\quad \sum_{i=1}^n l_i^2=o(N^{-2}).$$

Note that if η is the (p,q)-th element of $Var\hat{\beta}$, then the corresponding coefficients l_i satisfy (5.1) under the weak conditions (4.2) and $M^{-1}=O(N^{-1})$. From Shao (1987), the estimators of η of the following form:

$$\hat{\eta}=\sum_{i=1}^{n} l_i[(1-B_i)\hat{a}_i+B_i\bar{a}_i], \tag{5.2}$$

are consistent in a stronger sense that the MSE of the estimator is of the order $o(N^{-2})$, where $\hat{a}_i$ is defined in (3.4), $\bar{a}_i=$ either s_2 or s_J^2, B_i satisfies $0\le B_i\le 1$, B_i possibly depends on data and $\max_{i\le n} sup_y [B_i(y)]\to 0$ as $N\to\infty$. Note that $\hat{\eta}_b(\lambda)=\sum_{i=1}^{n} l_i v_i^b(\lambda_i)$ and $\hat{\eta}_w(\lambda)=\sum_{i=1}^{n} l_i v_i^w(\lambda_i)$ ($\lambda=(\lambda_1, \ldots, \lambda_n)'$) are special cases of the estimators in (5.2).

As a consequence, one can find a class of consistent and asymptotically unbiased (as $N\to\infty$) estimators of $Var\hat{\beta}$.

It is interesting to note that $\hat{\eta}_b(0)=\hat{\eta}_w(0)$ is identical to the weighted delete-1 jackknife variance estimator proposed by Wu (1986). Its asymptotic properties are studied in Shao and Wu (1987). $\hat{\eta}_b(0)$ may be improved (in terms of MSE) by $\hat{\eta}_b(\lambda)$ or $\hat{\eta}_w(\lambda)$ with $\lambda\neq 0$, since the covariances of the variance estimators of different groups usually are of smaller orders than their variances. Also, $\hat{\eta}_b(\lambda)$ or $\hat{\eta}_w(\lambda)$ is better than $\hat{\eta}=\sum_{i=1}^{n} l_i v_i$, where v_i is the sample variance or v_i^m, the MINQUE of σ_i^2. Note that if $v_i=v_i^m$, then $\hat{\eta}$ is the MINQUE of η.

One may also use the estimator $\hat{\eta}_R=\sum_{i=1}^{n} l_i v_i^r$, where v_i^r is defined in (4.8). In the case of estimating $Var\hat{\beta}$, $\hat{\eta}_R$ leads to

$$\hat{V}_R=M^{-1}\sum_{i=1}^{n} x_i x_i'\sum_{j=1}^{m_i} r_{ij}^2 M^{-1},$$

which has a negative bias. Hinkley (1977) proposed a weighted jackknife estimator

$$\hat{V}_H=(N-k)^{-1}N\hat{V}_R$$

which improves $\hat{V}_R$ but still tends to underestimate if the model is not balanced (Shao and Wu, 1987). For example, consider the following model:

$$y_{ij}=\beta_i+e_{ij} \quad j=1,\dots,m_i,\ i=1,2,$$

where e_{ij} are independent and $Ee_{ij}=0$ and $Var(e_{ij})=\sigma_i^2$. The biases of $\hat{V}_R$ and $\hat{V}_H$ are respectively

$$diag\left(-\frac{\sigma_1^2}{m_1^2}, -\frac{\sigma_2^2}{m_2^2}\right) \quad \text{and} \quad diag\left(\frac{(m_1-m_2)\sigma_1^2}{(m_1+m_2-2)m_1^2}, \frac{(m_2-m_1)\sigma_2^2}{(m_1+m_2-2)m_2^2}\right).$$

The negative bias of $\hat{V}_R$ can be very large. $\hat{V}_H$ is unbiased if $m_1=m_2$ and has a small bias if m_1 and m_2 are nearly equal (balanced model). When m_1 and m_2 are quite different (unbalanced model), then the bias of $\hat{V}_H$ may be large.

6. SOME EXTENSIONS

In Sections 3 and 4, we use the LSE as the estimator of β. The same procedure and analysis can be carried out if we start with a weighted least squares estimator (WLSE) with weights $\{ w_i, i=1,\dots,n \}$. Let $W=diag(w_i)$ and $\tilde{X}=W^{1/2}X$, $\tilde{y}=W^{1/2}y$ and $\tilde{e}=W^{1/2}e$. Then we have a model

$$\tilde{y}=\tilde{X}\beta+\tilde{e}. \tag{6.1}$$

Note that the LSE $\tilde{\beta}$ under model (6.1) is exactly the same as the WLSE under model (1.1) with weights w_i. Hence we can obtain residuals $\tilde{r}_{ij}=\tilde{y}_{ij}-\tilde{x}_i'\tilde{\beta}$ and the REBE's $\tilde{v}_i^b(\lambda_i)/w_i$ and $\tilde{v}_i^w(\lambda_i)/w_i$. In the special case that $\lambda_i=0$,

$$\tilde{v}_i^b(0)/w_i=\tilde{v}_i^w(0)/w_i=m_i^{-1}(1-h_i)^{-1}\sum_{j=1}^{m_i}(y_{ij}-x_i'\tilde{\beta})^2.$$

When $m_i=m$ for all i, this is the AUE proposed by Horn et al. (1973), but their formula does not apply to general unequal m_i cases.

Improved estimators of β can be obtained by using the weighted least squares method with the reciprocal of the variance estimates as the weights. We can obtain estimates of β and σ_i^2 simultaneously through the following iterative procedure. That is, start with an initial guess of σ_i^2, say $v_i^{(0)}$, and let $w_i^{(0)}=(v_i^{(0)})^{-1}$. Obtain the WLSE $\tilde{\beta}^{(0)}$ with $w_i^{(0)}$ as weights. Then obtain the REBE's $v_i^{(1)}$ and WLSE $\tilde{\beta}^{(1)}$ with $(v_i^{(1)})^{-1}$ as weights. Treat $v_i^{(1)}$ as the new guess of σ_i^2 and repeat the above procedure. If little is known about σ_i^2 in the beginning, $w_i^{(0)}\equiv 1$ can be used in the initial step. The simulation results in Section 7 indicate that there is an improvement (i.e., $\tilde{\beta}^{(1)}$ is better than $\tilde{\beta}^{(0)}$) by using this procedure.

Another important extension is to the case of nonlinear regression, i.e.,

(6.2) $$y_{ij}=f(x_i,\beta)+e_{ij}, \quad j=1,\dots,m_i,\ i=1,\dots,n,$$

where $f(x_i,\beta)$ is a nonlinear function in β. All the estimation procedures can be extended to this case in a straightforward manner. All we need to do is to fit model (6.2) by least squares or weighted least squares, and use $r_{ij}=y_{ij}-f(x_i,\tilde{\beta})$ as the residuals.

7. A SIMULATION STUDY

In this section, we examine by simulation (a) the finite sample performances of the variance estimators considered in the previous sections; (b) the empirical coverage probabilities of the 95% approximate confidence intervals of β; (c) the performances of the WLSE with the reciprocal of the various variance estimators used as weights.

7.1. The Model and the Etsimators

(i) *Model.* In the simulation we considered the following quadratic regression model:

$$y_{ij}=\beta_o+\beta_1 x_i+\beta_2 x_i^2+e_{ij}, \quad j=1,2,\ i=1,\dots,20.$$

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. Also, e_{ij} are independently distributed as $N(0, \sigma_i^2)$. We studied two models with different variance patterns. For Model 7.1, the values of σ_i^2 are

(7.1) 0.20, 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, 1.0.

For Model 7.2, the values of σ_i^2 are proportional to x_i, i.e.,

(7.2) $\sigma_i^2 = x_i/4, \quad i=1,\ldots,20.$

For each case, we have 3000 replicates in the simulation.

(ii) *The variance estimators.* We investigate the following variance estimators: the within-group sample variance s_i^2, the MINQUE v_i^m, the ARE v_i^r, and the REBE's $v_i^b(0)=v_i^w(0)$, $v_i^b(0.5)$, $v_i^b(1)$, $v_i^w(0.5)$ and $v_i^w(1)$.

(iii) *The confidence intervals.* The 95% approximate normal confidence intervals of β_j are:

$$[\hat{\beta}_j-1.96\nu_j,\ \hat{\beta}_j+1.96\nu_j], \quad j=0,1,2,$$

where β_j and $\hat{\beta}_j$ are the jth components of β and $\hat{\beta}$, respectively, and ν_j is the estimated standard deviation of $\hat{\beta}_j$ obtained by using the variance estimators in (ii). Let $CI(v_i)$ be the confidence interval using v_i as estimates of σ_i^2, $i=1,\ldots,20$.

(iv) *The WLSE.* The WLSE of β is defined to be

$$\tilde{\beta}=(X'WX)^{-1}X'Wy,$$

where $W=diag(w_i)$. The following WLSE are considered: $\tilde{\beta}_s: w_i^{-1}=s_i^2$, $\tilde{\beta}_m: w_i^{-1}=v_i^m$, $\tilde{\beta}_r: w_i^{-1}=v_i^r$, $\tilde{\beta}_{b(0)}: w_i^{-1}=v_i^b(0)$, $\tilde{\beta}_{b(0.5)}: w_i^{-1}=v_i^b(0.5)$, $\tilde{\beta}_{b(1)}: w_i^{-1}=v_i^b(1)$, $\tilde{\beta}_{w(0.5)}: w_i^{-1}=v_i^w(0.5)$, and $\tilde{\beta}_{w(1)}: w_i^{-1}=v_i^w(1)$.

The true value of β in the simulation is $\beta=(\ 1 \ \ 4 \ \ -0.5\ )'$.

7.2. Summary of the Simulation Study

(i) *The performances of variance estimators.* Tables A1 and A2 (for Models 7.1 and 7.2, respectively) show the root mean squared errors (RMSE) and biases of the variance estimators. Some conclusions drawn from these tables are:

(a) In terms of the RMSE, the REBE's are better than the within-group sample variance and the MINQUE for all i and both variance patterns. The improvement can be as high as 43% under Model 7.1 and 40% under Model 7.2.

(b) The ARE has negative biases especially under Model 7.1, where the σ_i^2 are not related to the x_i.

(c) The REBE with larger parameter λ has smaller RMSE and larger bias (in absolute value), especially under Model 7.1.

(d) v_i^b is generally better than v_i^w if the σ_i^2 are not related to the x_i. There is no definite conclusion otherwise.

(ii) *The performances of confidence intervals.* The coverage probabilities and the average lengths of the confidence intervals of β are shown in Tables B1 and B2. The results show that $CI(v_i^b(\lambda_i))$ and $CI(v_i^w(\lambda_i))$ (the confidence intervals using the REBE's as estimates of σ_i^2) have higher coverage probabilities than $CI(s_i^2)$. The coverage probabilities of $CI(v_i^r)$ are even lower than those of $CI(s_i^2)$.

(iii) *The performances of WLSE.* The biases and RMSE of WLSE are given in Tables C1 and C2. The results indicate that

(a) The WLSE using the REBE's is better than the ordinary LSE $\hat{\beta}$ if the σ_i^2 are very different (Model 7.2). The improvement can be as high as 18%. When the σ_i^2 are not very different (Model 7.1), the performances of LSE and WLSE are almost the same.

(b) In terms of RMSE, the WLSE using the REBE's are better than the WLSE using ARE.

(c) The use of the REBE's with larger λ_i provides better WLSE of β.

(d) The WLSE using the within-group sample variances and the MINQUE have very large RMSE. We found that over 40% of the time the matrix $X'WX$ is nearly singular when

$w_i^{-1}=v_i^m$, due to the negative estimates of MINQUE.

(iv) *Overall conclusion.* Combining the results in (i)-(iii), we conclude that for the models under consideration, the resampling empirical Bayesian method with $\lambda_i=1$ generally provides better estimators of σ_i^2, confidence intervals and WLSE of β.

Table A1: RMSE and biases of variance estimators (Model 7.1).

(The biases are shown in the second row for each i.)

i	s_i^2	v_i^m	v_i^r	$v_i^b(0)$	$v_i^b(0.5)$	$v_i^w(0.5)$	$v_i^b(1)$	$v_i^w(1)$
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 A2: RMSE and biases of variance estimators (Model 7.2).

(The biases are shown in the second row for each i.)

i	s_i^2	v_i^m	v_i^r	$v_i^b(0)$	$v_i^b(0.5)$	$v_i^w(0.5)$	$v_i^b(1)$	$v_i^w(1)$
1	.1444	.1387	.1214	.1319	.1415	.1314	.1609	.1324
	.0002	−.0016	.0176	.0265	.0644	.0388	.1024	.0512
2	.1814	.1570	.1379	.1487	.1544	.1466	.1685	.1454
	.0005	−.0019	.0114	.0211	.0564	.0313	.0916	.0415
3	.2030	.1787	.1575	.1687	.1704	.1652	.1792	.1623
	−.0039	−.0028	.0053	.0157	.0485	.0240	.0812	.0323
4	.2614	.2046	.1816	.1938	.1941	.1895	.1999	.1856
	.0072	.0032	.0058	.0173	.0475	.0237	.0778	.0302
5	.2977	.2298	.2051	.2179	.2164	.2130	.2191	.2082
	.0064	.0060	.0042	.0164	.0444	.0213	.0725	.0262
6	.3562	.2652	.2396	.2522	.2477	.2465	.2460	.2408
	.0042	−.0025	−.0100	.0029	.0273	.0056	.0516	.0084
7	.5183	.3956	.3649	.3802	.3736	.3732	.3680	.3662
	.0005	.0053	−.0119	.0035	.0203	.0029	.0372	.0024
8	.7207	.5257	.4907	.5080	.5000	.4998	.4925	.4916
	−.0011	.0024	−.0164	.0009	.0131	.0004	.0253	−.0000
9	.9106	.6571	.6166	.6367	.6272	.6272	.6178	.6177
	−.0082	−.0029	−.0197	.0003	.0096	.0020	.0189	.0036
10	1.0895	.8103	.7599	.7851	.7733	.7735	.7617	.7619
	−.0032	.0020	−.0128	.0117	.0189	.0163	.0261	.0209
11	1.3276	.9588	.8942	.9266	.9117	.9119	.8968	.8973
	.0396	.0105	−.0040	.0271	.0324	.0348	.0376	.0424
12	1.4438	1.1247	1.0410	1.0843	1.0650	1.0660	1.0457	1.0480
	.0370	.0425	.0233	.0642	.0670	.0741	.0697	.0839
13	1.7199	1.3087	1.1866	1.2469	1.2178	1.2196	1.1887	1.1928
	−.0045	−.0215	−.0518	.0102	.0089	.0243	.0076	.0383
14	2.1934	1.6943	1.4998	1.5986	1.5521	1.5570	1.5058	1.5159
	.0554	.0432	−.0277	.0682	.0571	.0788	.0461	.0894
15	2.5213	1.9190	1.6668	1.7901	1.7293	1.7353	1.6688	1.6810
	.0089	−.0174	−.1264	−.0012	−.0206	.0054	−.0400	.0120
16	2.8948	2.1872	1.8722	2.0277	1.9516	1.9609	1.8762	1.8945
	.0172	.0260	−.1410	.0191	−.0131	.0159	−.0453	.0127
17	3.4850	2.7178	2.3080	2.5085	2.4091	2.4219	2.3111	2.3359
	−.0018	.0031	−.2485	−.0389	−.0923	−.0565	−.1457	−.0742
18	4.4101	3.3833	2.8535	3.1179	2.9898	3.0084	2.8639	2.8997
	.1895	.1552	−.1672	.0991	.0178	.0747	−.0634	.0503
19	5.0233	4.1520	3.2241	3.6649	3.4423	3.5116	3.2290	3.3679
	−.1233	−.0354	−.4946	−.0051	−.1710	−.0016	−.3368	.0019
20	6.1577	5.8889	3.6281	4.9287	4.1562	4.6980	3.5038	4.4724
	−.0793	−.0496	−1.7794	−.2651	−.8064	−.3223	−1.3478	−.3795

Table B1: Coverage probabilities of confidence intervals (Model 7.1).

The average lengths of confidence intervals are shown in brackets.

	$CI(s_i^2)$	$CI(v_i^m)$	$CI(v_i^r)$	$CI(v_i^b(0))$	$CI(v_i^b(0.5))$	$CI(v_i^w(0.5))$	$CI(v_i^b(1))$	$CI(v_i^w(1))$
β_o	.9207 (.9244)	.9353 (.9329)	.9223 (.8984)	.9347 (.9334)	.9343 (.9331)	.9347 (.9338)	.9347 (.9327)	.9353 (.9341)
β_1	.9260 (.3413)	.9330 (.3436)	.9163 (.3214)	.9307 (.3433)	.9310 (.3419)	.9317 (.3432)	.9307 (.3405)	.9320 (.3430)
β_2	.9037 (.0207)	.9120 (.0209)	.8887 (.0189)	.9160 (.0209)	.9203 (.0208)	.9173 (.0209)	.9223 (.0206)	.9183 (.0209)

Table B2: Coverage probabilities of confidence intervals (Model 7.2).

The average lengths of confidence intervals are shown in brackets.

	$CI(s_i^2)$	$CI(v_i^m)$	$CI(v_i^r)$	$CI(v_i^b(0))$	$CI(v_i^b(0.5))$	$CI(v_i^w(0.5))$	$CI(v_i^b(1))$	$CI(v_i^w(1))$
β_o	.9220 (.8625)	.9377 (.8683)	.9210 (.8120)	.9407 (.8731)	.9427 (.8758)	.9417 (.8757)	.9457 (.8782)	.9437 (.8783)
β_1	.9000 (.5227)	.9157 (.5275)	.8983 (.4798)	.9187 (.5264)	.9143 (.5149)	.9203 (.5260)	.9117 (.5028)	.9207 (.5256)
β_2	.8727 (.0353)	.8839 (.0356)	.8577 (.0313)	.8953 (.0357)	.8900 (.0345)	.8963 (.0356)	.8833 (.0332)	.8997 (.0356)

Table C1: RMSE and biases of WLSE (Model 7.1).

(The biases are shown in the second row for each j.)

j	$\hat{\beta}$	$\tilde{\beta}_s$	$\tilde{\beta}_m$	$\tilde{\beta}_r$	$\tilde{\beta}_{b(0)}$	$\tilde{\beta}_{b(0.5)}$	$\tilde{\beta}_{w(0.5)}$	$\tilde{\beta}_{b(1)}$	$\tilde{\beta}_{w(1)}$
0	.2410 .0021	2.4250 −.0017	2.0897 .0545	.3082 −.0021	.2524 −.0006	.2463 .0010	.2467 .0009	.2453 .0011	.2459 .0011
1	.0891 .0004	.5945 −.0413	.8066 .0055	.0977 .0006	.1710 −.0023	.0889 .0012	.0891 .0012	.0885 .0011	.0889 .0012
2	.0055 −.0001	.0564 .0052	.0580 −.0005	.0070 .0000	.0194 .0003	.0055 −.0001	.0055 −.0001	.0055 −.0001	.0055 −.0001

Table C2: RMSE and biases of WLSE (Model 7.2).

(The biases are shown in the second row for each j.)

j	$\hat{\beta}$	$\tilde{\beta}_s$	$\tilde{\beta}_m$	$\tilde{\beta}_r$	$\tilde{\beta}_{b(0)}$	$\tilde{\beta}_{b(0.5)}$	$\tilde{\beta}_{w(0.5)}$	$\tilde{\beta}_{b(1)}$	$\tilde{\beta}_{w(1)}$
0	.2219 .0038	1.2275 −.0308	1.2181 −.0934	.1951 .0010	.1954 .0018	.1831 .0006	.1856 .0007	.1810 .0003	.1819 .0004
1	.1391 .0004	.4666 −.0276	.6740 .0375	.1355 .0003	.1334 .0005	.1272 .0018	.1270 .0018	.1266 .0021	.1254 .0021
2	.0097 −.0001	.0491 .0042	.0454 −.0017	.0104 −.0000	.0102 −.0000	.0093 −.0002	.0093 −.0002	.0093 −.0002	.0092 −.0002

REFERENCES

Carroll, R. J. (1982), "Adapting for Heteroscedasticity in Linear Models," *The Annals of Statistics*, 10, 1224-1233.

Efron, B. (1979), "Bootstrap Methods: Another Look at the Jackknife," *The Annals of Statistics*, 7, 1-26.

Hinkley, D. V. (1977), "Jackknifing in Unbalanced Situations," *Technometrics*, 19, 285-292.

Horn, S. D., Horn, R. A. and Duncan, D. B. (1975), "Estimating Heteroscedastic Variances in Linear Models," *Journal of the American Statistical Association*, 70, 380-385.

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. (1971), "The Estimation of Many Parameters," in *Foundation of Statistical Inference*, eds. V. P. Godambe and D. A. Sprott, Toronto: Holt, Rinehart and Winston, 435-455.

Lindley, D. V. and Smith, A. F. M. (1972), "Bayes Estimates for the Linear Model," *Journal of the Royal Statistical Society*, Ser. B, 34, 1-41.

Morris, C. N. (1983), "Parametric Empirical Bayes Inference: Theory and Applications," *Journal of the American Statistical Association*, 78, 47-65.

Rao, C. R. (1970), "Estimation of Heteroscedastic Variances in Linear Models," *Journal of the American Statistical Association*, 65, 161-172.

Rao, J. N. K. (1973), "On the Estimation of Heteroscedastic Variances," *Biometrics*, 29, 11-24.

Shao, J. (1986), "Discussion of a Paper by Professor C. F. J. Wu," *The Annals of Statistics*, 14, 1322-1326.

Shao, J. (1987), "Estimating Heteroscedastic Variances in Linear Models II: Properties of the Resampling Empirical Bayes Estimators," Tech. Report, Department of Statistics, Purdue University.

Shao, J. and Wu, C. F. J. (1987), "Heteroscedasticity-Robustness of Jackknife Variance Estimators in Linear Models," *The Annals of Statistics*, 15, to appear.

Wu, C. F. J. (1986), "Jackknife, Bootstrap and Other Resampling Methods in Regression Analysis (with discussion)," *The Annals of Statistics*, 14, 1261-1350.