The Annals of Statistics
1988, Vol. 16, No. 3, 986–1008

ON RESAMPLING METHODS FOR VARIANCE AND BIAS ESTIMATION IN LINEAR MODELS[1]

BY JUN SHAO

University of Wisconsin-Madison

Let g be a nonlinear function of the regression parameters β in a heteroscedastic linear model and $\hat{\beta}$ be the least squares estimator of β. We consider the estimation of the variance and bias of $g(\hat{\beta})$ [as an estimator of $g(\beta)$] by using three resampling methods: the weighted jackknife, the unweighted jackknife and the bootstrap. The asymptotic orders of the mean squared errors and biases of the resampling variance and bias estimators are given in terms of an imbalance measure of the model. Consistency of the resampling estimators is also studied. The results indicate that the weighted jackknife variance and bias estimators are asymptotically unbiased and consistent and their mean squared errors are of order $o(n^{-2})$ if the imbalance measure converges to zero as the sample size $n \to \infty$. Furthermore, based on large sample properties, the weighted jackknife is better than the unweighted jackknife. The bootstrap method is shown to be asymptotically correct only under a homoscedastic error model. Bias reduction, a closely related problem, is also discussed.

1. Introduction. In statistical applications involving the point estimation of an unknown parameter θ, one needs to estimate the accuracy of $\hat{\theta}$ as an estimator of θ. Some important and commonly used measures of accuracy are the variance, the bias and the mean squared error (MSE) of $\hat{\theta}$. Having good estimators of accuracy not only provides some information about the performance of $\hat{\theta}$, but often suggests improvements to $\hat{\theta}$ and provides ways of making other statistical inferences (e.g., confidence regions). Resampling methods such as the jackknife [Quenouille (1956) and Tukey (1958)] and the bootstrap [Efron (1979)] provide convenient and widely applicable methods of estimating the accuracy of the chosen estimator in the independent and identically distributed (i.i.d.) setting. These methods are computer-based and can handle problems which are far too complicated for traditional statistical analysis. For certain types of estimators in the i.i.d. situation, the resampling methods were proved to be asymptotically correct [Miller (1964), Bickel and Freedman (1981), Parr (1985) and Shao and Wu (1986)].

The main objective of this paper is to study resampling variance and bias estimation in the context of linear models. Throughout the paper the following model is assumed:

$$y = X\beta + e, \tag{1.1}$$

Received September 1986; revised October 1987.
[1]Research supported by NSF Grant DMS-85-02303 and ISSA-860068.
AMS 1980 *subject classifications*. Primary 62J05; secondary 62F35.
Key words and phrases. Resampling variance and bias estimators, jackknife, weighted jackknife, bootstrap, bias reduction, homoscedastic and heteroscedastic linear models, asymptotic unbiasedness, consistency, mean squared error, imbalance measure of a linear model.

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics*.
www.jstor.org

where $X = (x_1, \ldots, x_n)'$, $x_i \in \mathbf{R}^k$ is known, $i = 1, \ldots, n$, $y = (y_1, \ldots, y_n)' \in \mathbf{R}^n$ are the observed data, $\beta \in \mathbf{R}^k$ is the unknown parameter, $e = (e_1, \ldots, e_n)' \in \mathbf{R}^n$ are the random errors and the e_i are independent with zero means and unknown variances σ_i^2. We assume that the σ_i^2 are bounded.

The model (1.1) is said to be *homoscedastic* if $\sigma_i^2 = \sigma^2$ for all i and *heteroscedastic* otherwise. We will assume that the model is heteroscedastic unless otherwise specified. Also, x_i, e_i and σ_i^2 may depend on n, but the subscript n will be suppressed for simplicity.

It is assumed that $M = X'X = \sum_{i=1}^n x_i x_i'$ is positive definite and

$$M^{-1} = O(n^{-1}). \tag{1.2}$$

Let g be a real-valued nonlinear function defined on $\mathbf{R}^k$. (The case of vector g can be treated similarly.) The parameter of interest and its point estimator are $\theta = g(\beta)$ and $\hat{\theta} = g(\hat{\beta})$, respectively, where $\hat{\beta} = M^{-1}X'y$ is the ordinary least squares estimator (LSE) of β. We focus on $\hat{\beta}$ instead of the weighted least squares estimator (WLSE) for the following reasons:

1. Choosing adequate weights in WLSE involves estimation of each individual σ_i^2. Unless there are many replicates at the design point x_i or σ_i^2 is a smooth function of x_i, a consistent estimator of σ_i^2 is not available.
2. If v_i^{-1} are used as the weights and v_i is an inconsistent estimator of σ_i^2, the asymptotic distribution of the WLSE is complicated and generally unknown. On the other hand, the asymptotic distribution of the LSE is well known and therefore statistical inferences can be made based on it, if we have a suitable estimate of the variance of the LSE.
3. As a point estimator of β, the WLSE may not be better than the LSE, especially when the σ_i^2 are not very different from each other [Jacquez, Mather and Crawford (1968)].

Denote the variance and bias of $\hat{\theta}$ by $\operatorname{Var}\hat{\theta}$ and $B(\hat{\theta})$, respectively. If x_i are observations of random vectors x_i^*, $\operatorname{Var}\hat{\theta}$ and $B(\hat{\theta})$ are defined to be the conditional variance and bias. We study the properties of the resampling estimators of $\operatorname{Var}\hat{\theta}$ and $B(\hat{\theta})$ (conditional on x_i, $i = 1, \ldots, n$, if x_i is the observed value of x_i^*). The problem of improving $\hat{\theta}$ is considered only in Section 5, where we discuss the use of the jackknife estimator for reducing bias.

Note that under model (1.1), the observations y_i are independent but not identically distributed. Because of this model unbalancedness, the straightforward extension of the jackknife method to the linear model, which will be called unweighted jackknife method henceforth, does not provide good estimators for $\operatorname{Var}\hat{\theta}$ and $B(\hat{\theta})$. Hinkley (1977) modified the delete-1 jackknife by putting weights on the pseudovalues. Wu (1986) proposed a weighted delete-d jackknife method for arbitrary d with a different weighting scheme. The variance estimators obtained from these weighted jackknives possess some desirable properties [Shao and Wu (1987)]. For the jackknife bias estimator, Hinkley (1977) conjectured that his weighted jackknife bias estimator (which coincides with Wu's weighted delete-1 jackknife bias estimator although their weighting schemes are different) estimates $B(\hat{\theta})$ unbiasedly up to the order $O(n^{-2})$ [hence the resulting jackknife estimator of θ eliminates the bias up to the order $O(n^{-2})$]. His justification was heuristic but is valid only in some special cases. In fact, determining the order of the second order term of $B(\hat{\theta})$ [i.e., the term of lower order than the leading term in the expansion of $B(\hat{\theta})$] is not a trivial matter due to the complexity of the model. As the results in Section 2 indicate, the first order term (i.e., the leading term) of $B(\hat{\theta})$ is generally of the order $O(n^{-1})$, but unlike the i.i.d. case, the order of the second order term of $B(\hat{\theta})$ is between $O(n^{-3/2})$ and $O(n^{-2})$. Hence it is reasonable to expect that the bias estimator estimates $B(\hat{\theta})$ unbiasedly up to the order of the second order term of $B(\hat{\theta})$ rather than $O(n^{-2})$, as it usually does for the i.i.d. case.

In Section 2, we obtain asymptotic expansions of $\operatorname{Var}\hat{\theta}$ and $B(\hat{\theta})$ which will be used in studying the properties of the resampling estimators. A technical lemma used in the proofs of the main results in Sections 3 and 4 is also given.

The asymptotic properties of several resampling variance estimators are studied in Section 3. In particular, we obtain asymptotic orders of the MSE of the resampling variance estimators. The bias of a variance estimator is also an important issue. It is found that the bias of the variance estimator based on bootstrapping residuals may become the dominating factor in its MSE under heteroscedastic models. As a consequence, the bootstrap variance estimator has larger MSE than the jackknife variance estimators and therefore is not preferred. A finite sample comparison of some variance estimators is made in an example.

Although $B(\hat{\theta})$ is shown to have a lower order than the standard deviation of $\hat{\theta}$ in Section 2, knowing the magnitude and the direction of the bias is still important in practice. Section 4 is devoted to studying the resampling bias estimators. The asymptotic properties and the orders of the MSE of resampling bias estimators are obtained after establishing a mathematical equivalence between the estimation of $B(\hat{\theta})$ and the estimation of $\operatorname{Var}\hat{\beta}$. The weighted jackknife bias estimator is shown to be asymptotically unbiased and consistent. The reason for the poor performance (inconsistency, large MSE) of the unweighted jackknife bias estimator is explored: The order of the unweighted jackknife bias estimator does not match that of $B(\hat{\theta})$. It is also shown that the bias estimator based on bootstrapping residuals has the same asymptotic properties as the weighted jackknife bias estimator if the model is homoscedastic but otherwise performs poorly.

The quantity

$$h_n = \max_{i \le n} w_i, \tag{1.3}$$
$$w_i = x_i' M^{-1} x_i = \text{the } i\text{th diagonal element of "hat" matrix } XM^{-1}X',$$

plays a crucial role in the asymptotic analysis throughout the paper. It was termed an imbalance measure of the model (1.1) by Shao and Wu (1987) and its importance was first stressed by Huber (1973). Most of the results obtained are in terms of h_n. In the extreme case where the model is asymptotically balanced in the sense that $h_n = O(n^{-1})$, all the jackknife estimators are indistinguishable. No weighting procedure is needed in this case since the model has nearly the same nature as the i.i.d. situation. However, for an unbalanced model [i.e., the order of h_n is higher than $O(n^{-1})$], the gain in using the weighted jackknife is substantial.

The jackknife method was originally proposed to reduce the bias of $\hat{\theta}$. Bias reduction is actually equivalent to asymptotically unbiased estimation of bias, i.e., the reduction of bias of $\hat{\theta}$ amounts to finding an asymptotically unbiased estimator of $B(\hat{\theta})$ up to the order of the second order term of $B(\hat{\theta})$. Hence the results of Section 4 show that the weighted jackknife reduces bias and the unweighted jackknife does not. A discussion of whether to use the jackknife for bias reduction is provided in Section 5.

2. Preliminaries. We first develop some more notation and terminology. For a matrix A, its trace and determinant are denoted by $\operatorname{tr} A$ and $|A|$, respectively. Let $\|A\| = [\operatorname{tr}(A'A)]^{1/2}$. Denote a nonnegative (positive) definite matrix A by $A \geq 0$ ($A > 0$), and $A \geq B$ means $A - B \geq 0$. c is used as a positive generic constant, i.e., c is positive and independent of n but may have different values in different places. We say that the order of a sequence $\{a_n\}$ is no higher than that of $\{b_n\}$ iff

$$|a_n| \leq c|b_n| \quad \text{for all } n \tag{2.1}$$

and is higher than that of $\{b_n\}$ if (2.1) does not hold for any constant c.

The exact form of $B(\hat{\theta})$ and $\operatorname{Var}\hat{\theta}$ is not easy to obtain under model (1.1). An approximate form is obtained via a Taylor expansion. Thus, similar to the i.i.d. case, some smoothness conditions of the function g are required. Assume that $z_1, \dots, z_n$ are i.i.d., $\bar{z} = n^{-1}\sum_{i=1}^{n} z_i$ and g has a third order Lipschitz-continuous derivative. Then, under certain moment conditions,

$$E(g(\bar{z})) = \mu + \frac{1}{2n} g''(\mu)\sigma^2 + O(n^{-2})$$

and

$$\operatorname{Var}(g(\bar{z})) = \frac{\sigma^2}{n}(g'(\mu))^2 + O(n^{-2}),$$

where $\mu = E\bar{z}$ and $\sigma^2 = \operatorname{Var}(z_1)$.

Under model (1.1), using a Taylor expansion, we can expand

$$B(\hat{\theta}) = 2^{-1}\operatorname{tr}\left[\nabla^2 g(\beta)\operatorname{Var}\hat{\beta}\right] + R_1 \tag{2.2}$$

and

$$\operatorname{Var}\hat{\theta} = \nabla g(\beta)\operatorname{Var}\hat{\beta}(\nabla g(\beta))' + R_2, \tag{2.3}$$

where $\nabla g(\beta)$ and $\nabla^2 g(\beta)$ are the gradient and Hessian matrix of g at β, respectively, and R_1 and R_2 are the remainder terms. From (1.2) and the uniform boundedness of σ_i^2, $\operatorname{tr}[\nabla^2 g(\beta)\operatorname{Var}\hat{\beta}]$ and $\nabla g(\beta)\operatorname{Var}\hat{\beta}(\nabla g(\beta))'$ are of the order $O(n^{-1})$. Due to the model imbalance, the orders of R_1 and R_2 are not necessarily $O(n^{-2})$ as in the i.i.d. case. The following results show that the orders of R_1 and R_2 depend on h_n (1.3), the imbalance measure of the model (1.1).

THEOREM 2.1. (i) *Suppose that*

$$\max_{i\le n} Ee_i^4 \le \rho < \infty \quad \textit{for all } n \tag{2.4}$$

and g has a second order derivative satisfying

$$\|\nabla^2 g(x) - \nabla^2 g(y)\| \le c \sum_{j=1}^{L} \|x - y\|^{\lambda_j}, \qquad \lambda_j \le 2 \textit{ with } \min_{j\le L} \lambda_j = \lambda > 0 \tag{2.5}$$

for an integer L and some constants λ_j. *Then the remainder term* R_1 *in* (2.2) *satisfies*

$$R_1 = O(n^{-1-\lambda/2}).$$

(ii) *If* (2.4) *holds and g has a third order Lipschitz-continuous derivative, then*

$$R_1 = O(n^{-3/2}h_n^{1/2}).$$

(iii) *Suppose that* (2.4) *holds and g and* g^2 *have third order Lipschitz-continuous derivatives. Then the remainder term* R_2 *in* (2.3) *satisfies*

$$R_2 = O(n^{-3/2}h_n^{1/2}).$$

(iv) *Suppose that*

$$\max_{i\le n} Ee_i^8 \le \phi < \infty \quad \textit{for all } n \tag{2.6}$$

and g has a third order Lipschitz-continuous derivative. Then

$$R_2 = O(n^{-3/2}h_n^{1/2}).$$

(v) *If g satisfies condition* (2.5) *and either* g^2 *satisfies condition* (2.5) *or* (2.6) *holds, then the order of* R_2 *is either* $O(n^{-1-\lambda/2})$ *or* $O(n^{-3/2}h_n^{1/2})$.

REMARK 2.1. The conditions in Theorem 2.1 are sufficient for

$$B(\hat\theta) = 2^{-1}\mathrm{tr}\left[\nabla^2 g(\beta)\mathrm{Var}\,\hat\beta\right] + o(n^{-1}) \tag{2.7}$$

and

$$\mathrm{Var}\,\hat\theta = \nabla g(\beta)\mathrm{Var}\,\hat\beta(\nabla g(\beta))' + o(n^{-1}). \tag{2.8}$$

That is, $2^{-1}\mathrm{tr}[\nabla^2 g(\beta)\mathrm{Var}\,\hat\beta]$ and $\nabla g(\beta)\mathrm{Var}\,\hat\beta(\nabla g(\beta))'$ are valid asymptotic approximations of $B(\hat\theta)$ and $\mathrm{Var}\,\hat\theta$, respectively. (2.7) and (2.8) can hold under very weak conditions for regular statistics. See also Parr [(1985), Remark (ii) of Theorem 2].

REMARK 2.2. From Theorem 2.1, we not only have (2.7) and (2.8), but also obtain asymptotic orders of R_1 and R_2. The restriction $\lambda_j \le 2$ in condition (2.5) can be relaxed if we assume higher moment conditions on the errors. In most applications, $0 < \lambda \le 1$. If λ is close to zero, the orders of R_1 and R_2 are much higher than $O(n^{-2})$. In parts (ii)–(iv) of Theorem 2.1, by assuming a stronger smoothness condition on g, more precise orders of R_1 and R_2 are obtained in terms of the imbalance measure h_n (1.3). The order of h_n is no lower than $O(n^{-1})$ since $h_n \geq n^{-1}\sum_{i=1}^{n} w_i = n^{-1}k$. Hence the orders of R_1 and R_2 are between $O(n^{-3/2})$ and $O(n^{-2})$ and are $O(n^{-2})$ if the model is asymptotically balanced in the sense that $h_n = O(n^{-1})$.

We give the proof of Theorem 2.1(ii) for illustration. Other proofs are in Shao (1986).

PROOF OF THEOREM 2.1(ii). Let l_{jp} be the pth component of $M^{-1}x_j$,

$$f_{pqm} = \frac{\partial^3 g(\beta)}{\partial\beta_p\, \partial\beta_q\, \partial\beta_m}.$$

Then by the Lipschitz-continuity of the third order derivative of g,

$$g(\hat{\beta}) = g(\beta) + \nabla g(\beta)(\hat{\beta} - \beta) + 2^{-1}(\hat{\beta} - \beta)'\nabla^2 g(\beta)(\hat{\beta} - \beta) + 6^{-1} \sum_{p,q,m=1}^{k} f_{pqm}\left(\sum_{j=1}^{n} l_{jp}e_j\right)\left(\sum_{j=1}^{n} l_{jq}e_j\right)\left(\sum_{j=1}^{n} l_{jm}e_j\right) + \Gamma,$$

with $|\Gamma| \leq c\|\hat{\beta} - \beta\|^4$. Since e_i are independent with mean zero,

$$B(\hat{\theta}) = 2^{-1}\mathrm{tr}\left[\nabla^2 g(\beta)\mathrm{Var}\,\hat{\beta}\right] + 6^{-1} \sum_{p,q,m=1}^{k} f_{pqm} \sum_{j=1}^{n} l_{jp}l_{jq}l_{jm}Ee_j^3 + E\Gamma. \tag{2.9}$$

From Lemma 2.1, $|E\Gamma| \leq cE\|\hat{\beta} - \beta\|^4 = O(n^{-2})$. Note that

$$|l_{jp}| \leq \left(x_j'M^{-2}x_j\right)^{1/2} \leq cn^{-1/2}w_j^{1/2}.$$

The result follows since the second term on the right-hand side of (2.9) is bounded in absolute value by

$$\sum_{p,q,m=1}^{k} |f_{pqm}| \sum_{j=1}^{n} |l_{jp}l_{jq}l_{jm}|\, |Ee_j^3| \leq cn^{-3/2} \sum_{j=1}^{n} w_j^{3/2} = O\left(n^{-3/2}h_n^{1/2}\right). \qquad \square$$

Lemma 2.1 was used in the preceding proof and will be used frequently in the sequel. Its proof can be found in Shao (1986).

LEMMA 2.1. *Suppose that*

$$\max_{i \leq n} E|e_i|^p \leq c_1 < \infty \quad \textit{for all } n, \tag{2.10}$$

where p is an even integer. Then for any q satisfying $0 < q \leq p$,

$$\max_{i \leq n} E|r_i|^q \leq c_2 < \infty \quad \textit{for all } n$$

and

$$E\|\hat{\beta} - \beta\|^q = O(n^{-q/2}),$$

where c_1 and c_2 are independent of n and $r_i = y_i - x_i'\hat{\beta}$ is the ith residual from fitting model (1.1).

3. Resampling variance estimators. In this section, we study the properties of the resampling variance estimators. Theorem 3.1(i) gives the order of the MSE of the weighted delete-d jackknife estimator of $\operatorname{Var}\hat{\beta}$. The result is extended in Theorem 3.2 to the case of estimating $\operatorname{Var}\hat{\theta}$, $\hat{\theta} = g(\hat{\beta})$. The orders of the MSE of other resampling estimators of $\operatorname{Var}\hat{\beta}$ are given in Theorem 3.1(ii). Some of these results will be used later for the estimation of the bias of $\hat{\theta}$.

We first define the weighted delete-d jackknife estimators. For any fixed integer $d \le n - k$, let $r = n - d$. Define $\mathbf{S}_r$ to be the collection of subsets of size r in $\{1, \dots, n\}$. Let $s = \{i_1, \dots, i_r\} \in \mathbf{S}_r$. For an $n \times m$ matrix A, let A_s denote the submatrix of A consisting of the i_1th, $\dots$, i_rth rows of A. Denote $X_s'X_s$ by M_s and assume that M_s is positive definite for all $s \in \mathbf{S}_r$. Let $\hat{\beta}_s = M_s^{-1}X_s'y_s$ be the least squares estimator of β for the model $y_s = X_s\beta + e_s$ and $\hat{\theta}_s = g(\hat{\beta}_s)$. Denote $|M|^{-1}|M_s|$ by ω_s. Then the weighted delete-d jackknife estimator of $\operatorname{Var}\hat{\theta}$ [Wu (1986)] is

$$v_{J(d)}(\hat{\theta}) = \binom{n-k}{d-1}^{-1} \sum_{s \in \mathbf{S}_r} \omega_s(\hat{\theta}_s - \hat{\theta})^2.$$

The preceding formula has the obvious extension when $\hat{\theta}$ is a vector. For the estimation of $\operatorname{Var}\hat{\beta}$, the weighted delete-$d$ jackknife estimator is

$$v_{J(d)} = \binom{n-k}{d-1}^{-1} \sum_{s \in \mathbf{S}_r} \omega_s(\hat{\beta}_s - \hat{\beta})(\hat{\beta}_s - \hat{\beta})'.$$

We consider the estimation of $\operatorname{Var}\hat{\beta}$ first. There are several other resampling variance estimators: the modified weighted delete-1 jackknife estimator $v_{J(1)}(c)$ [Wu (1986), rejoinder], which is very close to $v_{J(1)}$ and thus has the same asymptotic properties as $v_{J(1)}$, the unweighted jackknife estimator [Miller (1974)]

$$v_J = n^{-1}(n-1)M^{-1} \sum_{i=1}^{n} (1 - w_i)^{-2} r_i^2 x_i x_i' M^{-1} - (n-1)M^{-1}RR'M^{-1},$$

where $R = n^{-1}\sum_{i=1}^{n}(1 - w_i)^{-1} r_i x_i$ and $r_i = y_i - x_i'\hat{\beta}$, the weighted jackknife estimator [Hinkley (1977)]

$$v_H = n(n-k)^{-1} M^{-1} \sum_{i=1}^{n} r_i^2 x_i x_i' M^{-1}$$

and the bootstrap estimator [Efron (1979)],

$$v_b = \left[(n-k)^{-1} \sum_{i=1}^{n} r_i^2\right] M^{-1},$$

which is identical to the classical variance estimator in the homoscedastic linear model.

Theorem 3.1 shows that the MSE of $v_{J(d)}$, v_J and v_H are all of the order $O(n^{-2}h_n)$, which implies that these variance estimators are consistent in a stronger sense that n times the difference between the variance estimator and $\operatorname{Var}\hat{\beta}$ converges to zero in L_2 when $h_n \to 0$. On the other hand, the MSE of v_b is generally of a higher order than the other variance estimators. Under the heteroscedastic models, v_b is neither consistent nor asymptotically unbiased [Wu (1986)]. See also Remark 3.3.

A heuristic explanation of the poor performance of v_b in the heteroscedastic models is: The bootstrap method depends on the exchangeability of the distribution of the data from which the bootstrap sample is taken [see Efron (1979)]. v_b is obtained by bootstrapping the normalized residuals $r_i/(1-k/n)^{1/2}$, $i = 1,\dots,n$ (see Section 4), which are nearly i.i.d. under the homoscedastic model but are *not* under the heteroscedastic model. Beran (1986) described a heteroscedastic bootstrap method which yields the same variance estimator as the jackknife method. Wu [(1986), Sections 6 and 7] gave some other bootstrap methods which are robust against heteroscedasticity.

Denote the (p,q)th element of the MSE of a variance estimator v by $\mathrm{MSE}_{pq}(v)$.

THEOREM 3.1. (i) *Assume that* (2.4) *and*

$$\sup_n dh_n < 1 \tag{3.1}$$

hold. Then

$$\mathrm{MSE}_{pq}(v_{J(d)}) = O(n^{-2}h_n). \tag{3.2}$$

(ii) *Under* (2.4) *and* $\sup_n h_n < 1$, (3.2) *holds with* $v_{J(d)}$ *replaced by* v_J *or* v_H.

(iii) *Under* (2.4), *we have*

$$\mathrm{MSE}_{pq}(v_b) = O\left[\max\left(n^{-2}h_n, \alpha_{n,pq}^2\right)\right], \tag{3.3}$$

where $\alpha_{n,pq}$ *is the* (p,q)*th element of*

$$M^{-1}\sum_{i=1}^{n}\left(\bar{\sigma}^2-\sigma_i^2\right)x_ix_i'M^{-1}, \qquad \bar{\sigma}^2 = (n-k)^{-1}\sum_{i=1}^{n}(1-w_i)\sigma_i^2.$$

REMARK 3.1. If the model is asymptotically balanced in the sense that $h_n = O(n^{-1})$, then the MSE of $v_{J(d)}$, v_J and v_H are of the order $O(n^{-3})$, which is the same as in the i.i.d. situation.

REMARK 3.2. The MSE of unweighted jackknife variance estimators has the same order as that of weighted jackknife estimators. However, the performance of v_J is not as good as the weighted jackknife variance estimators, especially when the model is unbalanced. An example is given later. See also the discussion in Shao and Wu [(1987), Sections 5 and 6] and the simulation results in Wu [(1986), Section 10 and rejoinder] and Tibshirani (1986).

REMARK 3.3. From the proof of Theorem 3.1(iii), $\alpha_{n,pq}$ is the leading term of $\mathrm{bias}_{pq}(v_b)$, the (p,q)th element of the bias of v_b. Under the homoscedastic model, $\alpha_{n,pq} = 0$ since $(n-k)^{-1}\sum_{i=1}^n(1-w_i)\sigma^2 = \sigma^2$, $\mathrm{MSE}_{pq}(v_b)$ has the same order as the MSE of jackknife estimators. In general, if the σ_i^2 are not close to each other, $\alpha_{n,pq}^2$ is of the order $O(n^{-2})$ and therefore $n^2\mathrm{bias}_{pq}^2(v_b)$ and $n^2\mathrm{MSE}_{pq}(v_b)$ do not converge to zero. See Example 3.1.

Before proving Theorem 3.1, we state Lemmas 3.1–3.3, which are used in the proofs of the main results in this and the next sections. Their proofs are given in the Appendix.

LEMMA 3.1. *Assume* (2.10) *and* (3.1) *hold. Let* $0 < q \le p$. *Then*

$$E\|\hat{\beta}_s - \hat{\beta}\|^q \le cn^{-q/2} \sum_{i \in \bar{s}} w_i^{q/2},$$

where $\bar{s}$ *is the complement of* $s \in \mathbf{S}_r$.

LEMMA 3.2. *Assume* (2.4) *and* (3.1) *hold. Let* $\gamma_{pq}^{(s)}$ *be the* (p, q)*th element of* $(\hat{\beta}_s - \hat{\beta})(\hat{\beta}_s - \hat{\beta})'$ *for a given* $s \in \mathbf{S}_r$. *Let* s *and* t *be two subsets in* $\mathbf{S}_r$. *Then*

$$\left|\operatorname{Cov}\left(\gamma_{pq}^{(s)}, \gamma_{pq}^{(t)}\right)\right| \le c_1 n^{-2} h_n.$$

Furthermore, if $\bar{s}$ *and* $\bar{t}$ (*the complements of* s *and* t, *respectively*) *do not share any common element, then*

$$\left|\operatorname{Cov}\left(\gamma_{pq}^{(s)}, \gamma_{pq}^{(t)}\right)\right| \le c_2 n^{-2} h_n \sum_{i \in \bar{s}} w_i \sum_{i \in \bar{t}} w_i.$$

If $\bar{s}$ *and* $\bar{t}$ *have only one common element* l, *then*

$$\left|\operatorname{Cov}\left(\gamma_{pq}^{(s)}, \gamma_{pq}^{(t)}\right)\right| \le c_3 n^{-2} h_n \left(w_l + \sum_{i \in \bar{s}} w_i \sum_{i \in \bar{t}} w_i \right),$$

where c_i *are independent of* n, l, s *and* t, $i = 1, 2, 3$.

LEMMA 3.3. *Assume* (2.4) *and* (3.1) *hold. Let* v_{pq} *be the* (p, q)*th element of* $\nu_{J(d)}$. *Then*

$$|Ee_j v_{pq}| \le cn^{-1} w_j, \qquad j = 1, \dots, n,$$

where c *is independent of* j *and* n.

PROOF OF THEOREM 3.1. (i) Let V_{pq} be the variance of the (p, q)th element of $\nu_{J(d)}$. From Theorem 1 of Shao and Wu (1987), the bias of $\nu_{J(d)} = O(n^{-1}h_n)$. Hence it suffices to show that

$$V_{pq} = O\left(n^{-2} h_n\right).$$

Using the notation in Lemma 3.2 and $\omega_s \le 1$, we have

$$V_{pq} \le \binom{n-k}{d-1}^{-2} \sum_{v=0}^{2} \sum_{v} \left|\operatorname{Cov}\left(\gamma_{pq}^{(s)}, \gamma_{pq}^{(t)}\right)\right|,$$

where Σ_0 is over all the pairs of s, t such that $\bar{s}$ and $\bar{t}$ do not share any common element, Σ_1 is over all the pairs of s, t such that $\bar{s}$ and $\bar{t}$ have only one common element and Σ_2 is over the remainder of the pairs of s, t. By Lemma 3.2,

$$\begin{aligned}
&\binom{n-k}{d-1}^{-2} \sum_0 \left|\operatorname{Cov}\left(\gamma_{pq}^{(s)}, \gamma_{pq}^{(t)}\right)\right| \\
&\quad \le cn^{-2}h_n \binom{n-k}{d-1}^{-2} \sum_0 \sum_{i\in\bar{s}} w_i \sum_{i\in\bar{t}} w_i \\
&\quad \le ck^2 n^{-2}h_n \binom{n-k}{d-1}^{-2} \binom{n-1}{d-1}^2 \\
&\quad = O\left(n^{-2}h_n\right),
\end{aligned}$$

$$\begin{aligned}
&\binom{n-k}{d-1}^{-2} \sum_1 \left|\operatorname{Cov}\left(\gamma_{pq}^{(s)}, \gamma_{pq}^{(t)}\right)\right| \\
&\quad \le cn^{-2}h_n \binom{n-k}{d-1}^{-2} \left[\binom{n-d}{d-1} \sum_{s\in S_r} \sum_{i\in\bar{s}} w_i + \sum_1 \sum_{i\in\bar{s}} w_i \sum_{i\in\bar{t}} w_i\right] \\
&\quad \le O\left(n^{-2}h_n\right) + ckn^{-2}h_n \binom{n-k}{d-1}^{-2} \binom{n-d}{d-1}\binom{n-1}{d-1} \\
&\quad = O\left(n^{-2}h_n\right)
\end{aligned}$$

and

$$\begin{aligned}
&\binom{n-k}{d-1}^{-2} \sum_2 \left|\operatorname{Cov}\left(\gamma_{pq}^{(s)}, \gamma_{pq}^{(t)}\right)\right| \\
&\quad \le cn^{-2}h_n \binom{n-k}{d-1}^{-2} \binom{n}{d}\left[\binom{n}{d} - \binom{n-d}{d} - d\binom{n-d}{d-1}\right] \\
&\quad = O\left(n^{-2}h_n\right),
\end{aligned}$$

where the last equality follows from

$$\binom{n}{d}^{-1}\left[\binom{n}{d} - \binom{n-d}{d} - d\binom{n-d}{d-1}\right] = O(n^{-2}).$$

Hence (i) is proved.

(ii) The bias of v_J or v_H is of the order $O(n^{-1}h_n)$ [Shao and Wu (1987), Section 5]. The variance of the (p, q)th element of v_J or v_H is of the order $O(n^{-2}h_n)$ since from (2.4) and Lemma 2.1, $\operatorname{Var}(r_i^2) \le c$, and from (A4) in the Appendix, $\left|\operatorname{Cov}\left(r_i^2, r_j^2\right)\right| \le ch_n$ for $i \ne j$. Hence (ii) follows.

(iii) From the proof of (ii), the variance of the (p, q)th element of v_b has order $O(n^{-2}h_n)$. Let m_{pq} and $\operatorname{bias}_{pq}(v_b)$ be the (p, q)th elements of M^{-1} and the bias of v_b, respectively. Then

$$\operatorname{bias}_{pq}(v_b) = \alpha_{n,pq} + m_{pq}(n-k)^{-1} \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij}^2\left(\sigma_j^2 - \sigma_i^2\right) = \alpha_{n,pq} + O(n^{-2}),$$

where the last equality follows from (1.2) and $\sum_{i=1}^n \sum_{j=1}^n w_{ij}^2 = \sum_{i=1}^n w_i = k$. This proves (iii). □

The MSE of v_b may be dominated by the squared bias, which has the order $\alpha^2_{n,pq}$. This usually occurs when v_b is inconsistent. An example is the following:

EXAMPLE 3.1. We compare $v_{J(1)}$, v_J and v_b in the model

$$y_{ij} = \beta_j + e_{ij}, \qquad i = 1, \dots, n_j,\ j = 1, 2,$$

with independent e_{ij}, $Ee_{ij} = 0$ and $\mathrm{Var}(e_{ij}) = \sigma_j^2$, $j = 1, 2$. The variance of LSE is a diagonal matrix

$$\mathrm{diag}\big(n_1^{-1}\sigma_1^2, n_2^{-1}\sigma_2^2\big).$$

Let $n = n_1 + n_2$. Then

$$v_{J(1)} = \mathrm{diag}\big(n_1^{-1}(n_1 - 1)^{-1}SS_1, n_2^{-1}(n_2 - 1)^{-1}SS_2\big),$$
$$v_J = n^{-1}(n-1)\mathrm{diag}\big((n_1 - 1)^{-2}SS_1, (n_2 - 1)^{-2}SS_2\big),$$
$$v_b = (n-2)^{-1}(SS_1 + SS_2)\mathrm{diag}\big(n_1^{-1}, n_2^{-1}\big),$$

where $SS_j = \sum_{i=1}^{n_j}(y_{ij} - \bar{y}_j)^2$, $\bar{y}_j = n_j^{-1}\sum_{i=1}^{n_j} y_{ij}$, $j = 1, 2$. Assume that $h_n = \max(n_1^{-1}, n_2^{-1}) = O(n^{-1})$. It is easy to see that $v_{J(1)}$ and v_J are consistent, but v_b is inconsistent unless $\sigma_1^2 = \sigma_2^2$. For the biases of these estimators, $v_{J(1)}$ is unbiased and the biases of v_J and v_b are, respectively,

$$\mathrm{diag}\left(\frac{n_2\sigma_1^2}{nn_1(n_1 - 1)}, \frac{n_1\sigma_2^2}{nn_2(n_2 - 1)}\right)$$

and

$$\mathrm{diag}\left(\frac{(n_2 - 1)(\sigma_2^2 - \sigma_1^2)}{(n-2)n_1}, \frac{(n_1 - 1)(\sigma_1^2 - \sigma_2^2)}{(n-2)n_2}\right).$$

The bias of v_J is positive, but of the order $O(n^{-2})$. However, the bias of v_b is of the order $O(n^{-1})$ unless $\sigma_1^2 = \sigma_2^2$. From Theorem 3.1, the MSE of $v_{J(1)}$ and v_J are of the order $O(n^{-3})$. The MSE of v_b is dominated by the squared bias of v_b [note that the variance of v_b is of the order $O(n^{-3})$] and is of the order $O(n^{-2})$. Hence v_b is not as good as the jackknife variance estimators.

A comparison of $v_{J(1)}$ and v_J shows that $v_{J(1)}$ is preferred to v_J: $v_{J(1)}$ is unbiased and has smaller variance (hence smaller MSE) than v_J since $(n-1)/n(n_1 - 1) > 1/n_1$.

We now consider the estimation of $\mathrm{Var}\,\hat{\theta}$, $\hat{\theta} = g(\hat{\beta})$. We focus on the weighted jackknife only, since the preceding results show that it provides better variance estimators in the case of $\hat{\theta} = \hat{\beta}$. Theorem 3.2 is an extension of Theorem 3.1(i).

THEOREM 3.2. (i) *Assume that* (2.6) *and* (3.1) *hold. Suppose that g has a third order Lipschitz-continuous derivative. Let $B(v_{J(d)}) = Ev_{J(d)} - \mathrm{Var}\,\hat{\beta}$ be the bias of $v_{J(d)}$. Then*

$$Ev_{J(d)}(\hat{\theta}) = \nabla g(\beta)\mathrm{Var}\,\hat{\beta}(\nabla g(\beta))' + \nabla g(\beta)B(v_{J(d)})(\nabla g(\beta))' + O\big(n^{-3/2}h_n^{1/2}\big). \tag{3.4}$$

(ii) *Under the same conditions as in* (i), *we have*

$$\mathrm{MSE}\big(v_{J(d)}(\hat{\theta})\big) = O\big(n^{-2}h_n\big).$$

REMARK 3.4. Under the homoscedastic model,

$$Ev_{J(d)}(\hat{\theta}) = \nabla g(\beta)\mathrm{Var}\,\hat{\beta}(\nabla g(\beta))' + O\big(n^{-3/2}h_n^{1/2}\big)$$

since $B(v_{J(d)}) = 0$. In general, $B(v_{J(d)}) = O(n^{-1}h_n)$. Hence

$$Ev_{J(d)}(\hat{\theta}) = \nabla g(\beta)\mathrm{Var}\,\hat{\beta}(\nabla g(\beta))' + O\big(n^{-1}h_n\big).$$

From Theorem 2.1, the bias of $v_{J(d)}(\hat{\theta})$ in both cases is of an order no higher than $O(n^{-1}h_n)$.

REMARK 3.5. The condition on g in Theorem 3.2 can be relaxed so that g satisfies (2.5) with λ satisfying $h_n^{-1}n^{-(1+\lambda)/2} = O(1)$. The last term on the right-hand side of (3.4) is then $O(n^{-1}h_n)$.

PROOF OF THEOREM 3.2. (i) From a Taylor expansion,

$$\hat{\theta}_s - \hat{\theta} = \nabla g(\hat{\beta})\big(\hat{\beta}_s - \hat{\beta}\big) + 2^{-1}\big(\hat{\beta}_s - \hat{\beta}\big)'\nabla^2 g(\zeta_s)\big(\hat{\beta}_s - \hat{\beta}\big), \tag{3.5}$$

where ζ_s is a point on the line segment between $\hat{\beta}_s$ and $\hat{\beta}$. Then

$$v_{J(d)}(\hat{\theta}) = \nabla g(\hat{\beta})v_{J(d)}\big(\nabla g(\hat{\beta})\big)' + \Gamma,$$

where

$$\Gamma = \binom{n-k}{d-1}^{-1}\sum_{s\in \mathcal{S}_r}\omega_s\Big\{4^{-1}\big[\big(\hat{\beta}_s - \hat{\beta}\big)'\nabla^2 g(\zeta_s)\big(\hat{\beta}_s - \hat{\beta}\big)\big]^2 + \nabla g(\hat{\beta})\big(\hat{\beta}_s - \hat{\beta}\big)\big(\hat{\beta}_s - \hat{\beta}\big)'\nabla^2 g(\zeta_s)\big(\hat{\beta}_s - \hat{\beta}\big)\Big\}.$$

Under the conditions of Theorem 3.2, $E\Gamma = O(n^{-3})$. By Theorem 3.1(i), $\mathrm{Var}(\mathrm{tr}(v_{J(d)})) = O(n^{-2}h_n)$. Since $E(\mathrm{tr}(v_{J(d)})) = O(n^{-1})$, we have $E(\mathrm{tr}(v_{J(d)}))^2 = O(n^{-2})$. Then

$$\begin{aligned} &E\big(\nabla g(\hat{\beta}) - \nabla g(\beta)\big)v_{J(d)}\big(\nabla g(\hat{\beta}) - \nabla g(\beta)\big)' \\ &\quad\le \Big[E\big\|\nabla g(\hat{\beta}) - \nabla g(\beta)\big\|^4 E\big(\mathrm{tr}\, v_{J(d)}\big)^2\Big]^{1/2} = O(n^{-2}) \end{aligned}$$

follows from $E\|\nabla g(\hat{\beta}) - \nabla g(\beta)\|^4 = O(n^{-2})$ under the conditions of Theorem 3.2. From

$$\begin{aligned} &\nabla g(\hat{\beta})v_{J(d)}\big(\nabla g(\hat{\beta})\big)' \\ &\quad= \big(\nabla g(\hat{\beta}) - \nabla g(\beta)\big)v_{J(d)}\big(\nabla g(\hat{\beta}) - \nabla g(\beta)\big)' \\ &\qquad+ 2\big(\nabla g(\hat{\beta}) - \nabla g(\beta)\big)v_{J(d)}\big(\nabla g(\beta)\big)' + \nabla g(\beta)v_{J(d)}\big(\nabla g(\beta)\big)', \end{aligned}$$

what remains to be shown is

$$Ev_{J(d)}\big(\nabla g(\hat{\beta}) - \nabla g(\beta)\big)' = O\big(n^{-3/2}h_n^{1/2}\big). \tag{3.6}$$

Since $(\nabla g(\hat{\beta}) - \nabla g(\beta))' = \nabla^2 g(\beta)(\hat{\beta} - \beta) + S$ with $\|S\| \le c(\|\hat{\beta} - \beta\|^2 + \|\hat{\beta} - \beta\|^3)$ by the Lipschitz-continuity of the third order derivative of g, we have $E\nu_{J(d)}S = O(n^{-2})$. Then (3.6) is equivalent to

$$E\left[\nu_{J(d)}\nabla^2 g(\beta)(\hat{\beta} - \beta)\right] = O\left(n^{-3/2}h_n^{1/2}\right).$$

Denote the pth component of $\nabla^2 g(\beta)M^{-1}x_j$ by τ_{pj} and the (p, q)th element of $\nu_{J(d)}$ by v_{pq}. Note that $|\tau_{qj}| \le \|\nabla^2 g(\beta)\| \, \|M^{-1}x_j\| \le cn^{-1/2}w_j^{1/2}$ and $|Ev_{pq}e_j| \le cn^{-1}w_j$ for all j by Lemma 3.3. Then (3.6) follows from

$$\left|\sum_{q=1}^{k}\sum_{j=1}^{n}\tau_{qj}E(v_{pq}e_j)\right| \le c\sum_{q=1}^{k}\sum_{j=1}^{n} n^{-3/2}w_j^{3/2} \le cn^{-3/2}h_n^{1/2},$$

since the pth component of $E[\nu_{J(d)}\nabla^2 g(\beta)(\hat{\beta} - \beta)]$ is equal to $\sum_{q=1}^{k}\sum_{j=1}^{n}\tau_{qj}E(v_{pq}e_j)$.

(ii) We only sketch the proof for this part. From part (i) and Remark 3.4, it suffices to show that the variance of $\nu_{J(d)}(\hat{\theta})$ is of the order $O(n^{-2}h_n)$. By using a similar argument to that used in the proofs of part (i) and Theorem 3.1, it can be shown that the order of the variance of $\nabla g(\hat{\beta})\nu_{J(d)}(\nabla g(\hat{\beta}))'$ [the dominating term of $\nu_{J(d)}(\hat{\theta})$] is $O(n^{-2}h_n)$. □

4. Resampling bias estimator. In this section, we focus on estimating another measure of statistical accuracy: the bias of $\hat{\theta}$. Three types of resampling bias estimators are considered. That is, the weighted delete-d jackknife bias estimator

$$\hat{B}_{J(d)} = \binom{n-k}{d-1}^{-1}\sum_{s\in \mathbf{S}_r}\omega_s\left(\hat{\theta}_s - \hat{\theta}\right),$$

the unweighted jackknife bias estimator

$$\hat{B}_J = n^{-1}(n-1)\sum_{i=1}^{n}\left(\hat{\theta}_{(i)} - \hat{\theta}\right),$$

where $\hat{\theta}_{(i)} = g(\hat{\beta}_{(i)})$ and $\hat{\beta}_{(i)}$ is the LSE of β after deleting (x_i, y_i), and the bootstrap bias estimator

$$\hat{B}_b = E_*\theta^* - \hat{\theta},$$

where $\theta^* = g(\beta^*)$, $\beta^* = \hat{\beta} + M^{-1}X'e^*$, $e^* = (e_1^*, \dots, e_n^*)'$ and e_i^* are i.i.d. samples from the normalized residuals $\{r_i/(1 - k/n)^{1/2}, i = 1, \dots, n\}$ and E_* is the expectation under the bootstrap distribution. For convenience we assume that the first element of each x_i is 1 when we discuss the bootstrap estimator.

In the following, we study the properties of the resampling bias estimators (consistency, asymptotic unbiasedness and the order of MSE) by relating the bias estimation to the estimation of $\operatorname{Var}\hat{\beta}$.

4.1. *The weighted delete-d jackknife bias estimator.* Before examining the properties of $\hat{B}_{J(d)}$, we first consider the simple case of $g = a'\beta$, where a is a known vector. There is no need for bias estimation since $B(\hat{\theta}) = 0$ in this case. A natural property of a bias estimator $\hat{B}$ is

$$\hat{B} = 0 \quad \text{if } g = a'\beta. \tag{4.1}$$

From Wu [(1986), Theorem 2], (4.1) holds for $\hat{B}_{J(d)}$. We now prove the consistency of $\hat{B}_{J(d)}$.

Theorem 4.1. *Suppose that $h_n \to 0$ and g has a second order derivative which is continuous in a neighborhood of β. Then*

$$\hat{B}_{J(d)} - 2^{-1}\mathrm{tr}\left[\nabla^2 g(\beta)\mathrm{Var}\,\hat{\beta}\right] = o_p(n^{-1}).$$

If we assume (2.7), *then*

$$\hat{B}_{J(d)} - B(\hat{\theta}) = o_p(n^{-1}).$$

Proof. From the expansion (3.5) and Theorem 2 of Wu (1986), we have

$$\hat{B}_{J(d)} = 2^{-1}\mathrm{tr}\left[\nabla^2 g(\hat{\beta})v_{J(d)}\right] + \binom{n-k}{d-1}^{-1} \sum_{s \in \mathbb{S}_r} \omega_s \xi_s, \tag{4.2}$$

where $v_{J(d)}$ is the weighted delete-d jackknife estimator of $\mathrm{Var}\,\hat{\beta}$ defined in Section 3 and

$$\xi_s = 2^{-1}(\hat{\beta}_s - \hat{\beta})'\left(\nabla^2 g(\zeta_s) - \nabla^2 g(\hat{\beta})\right)(\hat{\beta}_s - \hat{\beta}).$$

From $v_{J(d)} - \mathrm{Var}\,\hat{\beta} = o_p(n^{-1})$ [Shao and Wu (1987), Theorem 3] and the continuity of $\nabla^2 g$ at β,

$$2^{-1}\mathrm{tr}\left[\nabla^2 g(\hat{\beta})v_{J(d)}\right] - 2^{-1}\mathrm{tr}\left[\nabla^2 g(\beta)\mathrm{Var}\,\hat{\beta}\right] = o_p(n^{-1}).$$

It remains to be shown that

$$\binom{n-k}{d-1}^{-1} \sum_{s \in \mathbb{S}_r} \omega_s \xi_s = o_p(n^{-1}). \tag{4.3}$$

By the continuity of $\nabla^2 g$ and $\mathrm{tr}(v_{J(d)}) = O_p(n^{-1})$, a similar argument used in the proof of Theorem 4 of Shao and Wu (1987) yields (4.3). □

Theorem 4.2 establishes a relation between finding an asymptotically unbiased estimator of $B(\hat{\theta})$ and the existence of an asymptotically unbiased estimator of $\mathrm{Var}\,\hat{\beta}$.

Theorem 4.2. *Suppose that* (2.4) *and* (3.1) *hold. Let* $B(v_{J(d)}) = Ev_{J(d)} - \mathrm{Var}\,\hat{\beta}$.

(i) *If g satisfies condition* (2.5) *for some $\lambda > 0$, then*

$$E\hat{B}_{J(d)} = B(\hat{\theta}) + 2^{-1}\mathrm{tr}\left[\nabla^2 g(\beta)B(v_{J(d)})\right] + O(n^{-1-\lambda/2}).$$

(ii) *If g has a third order Lipschitz-continuous derivative, then*

$$E\hat{B}_{J(d)} = B(\hat{\theta}) + 2^{-1}\mathrm{tr}\left[\nabla^2 g(\beta)B(v_{J(d)})\right] + O\left(n^{-3/2}h_n^{1/2}\right).$$

PROOF. We prove (ii) only. Since g has a third order Lipschitz-continuous derivative, (4.2) holds with ξ_s satisfying $|\xi_s| \le c(\|\hat{\beta}_s - \hat{\beta}\|^3 + \|\hat{\beta}_s - \hat{\beta}\|^4)$. Hence by Lemma 3.1,

$$E\binom{n-k}{d-1}^{-1} \sum_{s \in \mathcal{S}_r} \omega_s |\xi_s| \le c\binom{n-k}{d-1}^{-1} \sum_{s \in \mathcal{S}_r} \left(n^{-3/2} \sum_{i \in \bar{s}} w_i^{3/2} + n^{-2} \sum_{i \in \bar{s}} w_i^2\right)$$
$$\le c\binom{n-k}{d-1}^{-1}\binom{n-1}{d-1} n^{-3/2} h_n^{1/2} \sum_{i=1}^{n} w_i \le c n^{-3/2} h_n^{1/2}.$$

The proof is completed by showing

$$E\left[\operatorname{tr}\left(\left(\nabla^2 g(\hat{\beta}) - \nabla^2 g(\beta)\right) v_{J(d)}\right)\right] = O\left(n^{-3/2} h_n^{1/2}\right).$$

Denote the (p, q)th elements of $\nabla^2 g(\hat{\beta}) - \nabla^2 g(\beta)$, $\nabla^2 g$ and $v_{J(d)}$ by ζ_{pq}, f_{pq} and v_{pq}, respectively. Since g has a third order Lipschitz-continuous derivative, $\zeta_{pq} = \nabla f_{pq}(\beta)(\hat{\beta} - \beta) + s_{pq}$ with $|s_{pq}| \le c\|\hat{\beta} - \beta\|^2$. From Lemma 2.1 and Theorem 3.1,

$$|E v_{pq} s_{pq}| \le c\left[E\left(\operatorname{tr}(v_{J(d)})\right)^2 E\|\hat{\beta} - \beta\|^4\right]^{1/2} = O(n^{-2}).$$

Note that $|\nabla f_{pq}(\beta) M^{-1} x_j| \le c(x_j' M^{-2} x_j)^{1/2} \le c n^{-1/2} w_j^{1/2}$ and $|E e_j v_{pq}| \le c n^{-1} w_j$ by Lemma 3.3. Thus,

$$\left|E\left[\operatorname{tr}\left(\left(\nabla^2 g(\hat{\beta}) - \nabla^2 g(\beta)\right) v_{J(d)}\right)\right]\right|$$
$$= \left|\sum_{p,q=1}^{k} \left[E \nabla f_{pq}(\beta)(\hat{\beta} - \beta) v_{pq} + E v_{pq} s_{pq}\right]\right|$$
$$\le \sum_{p,q=1}^{k} \sum_{j=1}^{n} \left|\nabla f_{pq}(\beta) M^{-1} x_j\right| \left|E(e_j v_{pq})\right| + O(n^{-2})$$
$$\le c n^{-3/2} \sum_{j=1}^{n} w_j^{3/2} + O(n^{-2})$$
$$= O\left(n^{-3/2} h_n^{1/2}\right). \qquad \square$$

For the asymptotic unbiasedness of $\hat{B}_{J(d)}$, we have

THEOREM 4.3. (i) *Under the homoscedastic model, we have*

$$E\hat{B}_{J(d)} = B(\hat{\theta}) + R, \tag{4.4}$$

where $R = O(n^{-1-\lambda/2})$ *if* (2.5) *holds and* $R = O(n^{-3/2} h_n^{1/2})$ *if* g *has a third order Lipschitz-continuous derivative.*

(ii) *Under the heteroscedastic model,* (4.4) *holds with* $R = O(n^{-1} H_n)$ *if* (2.5) *holds, where* $H_n = \max\{h_n, n^{-\lambda/2}\}$, *and with* $R = O(n^{-1} h_n)$ *if* g *has a third order Lipschitz-continuous derivative.*

PROOF. The results follow directly from Theorem 4.2 and the asymptotic unbiasedness of $v_{J(d)}$. $\square$

The asymptotic order of the MSE of $\hat{B}_{J(d)}$ is given by Theorem 4.4.

THEOREM 4.4. *Assume* (2.6) *and g has a third order Lipschitz-continuous derivative. Then*

$$\mathrm{MSE}(\hat{B}_{J(d)}) = O(n^{-2}h_n).$$

PROOF. By Theorem 4.3, it suffices to show

$$\mathrm{Var}\,\hat{B}_{J(d)} = O(n^{-2}h_n).$$

From (2.6), Lemma 3.1 and the condition on g, the variance of the second term on the right-hand side of (4.2) is of order $O(n^{-2}h_n^2)$. Since $\mathrm{Var}(v_{J(d)}) = O(n^{-2}h_n)$, the result follows if

$$E\left[\mathrm{tr}\left(\nabla^2 g(\hat{\beta}) - \nabla^2 g(\beta)\right)v_{J(d)}\right]^2 = O(n^{-2}h_n).$$

But this follows from (2.6), the smoothness condition on g and Lemma 4.1. □

LEMMA 4.1. *If* (2.6) *and* (3.1) *hold, then*

$$E\left(\mathrm{tr}(v_{J(d)})\right)^4 = O(n^{-4}).$$

The proof is given in the Appendix.

4.2. *The unweighted jackknife bias estimator.* If the model (1.1) is unbalanced, i.e., h_n is not of the order $O(n^{-1})$, the unweighted jackknife bias estimator is not recommended since in general it is *inconsistent* and has larger bias and MSE than the weighted jackknife bias estimator. The reason for the poor performance of $\hat{B}_J$ is that the first term in the Taylor expansion of $\hat{B}_J$ does not vanish due to the unbalancedness of the model. As a consequence, the order of $\hat{B}_J$ does not match that of $B(\hat{\theta})$ unless $h_n = O(n^{-1/2})$.

For simplicity, we limit ourselves to the homoscedastic model in this section. One cannot expect $\hat{B}_J$ to perform better in the heteroscedastic case. Let

$$\Gamma_1 = n^{-1}(n-1)\nabla g(\hat{\beta}) \sum_{i=1}^{n} \left(\hat{\beta}_{(i)} - \hat{\beta}\right).$$

Then from a Taylor expansion,

$$\hat{B}_J = \Gamma_1 + \Gamma_2,$$

where $\Gamma_2 = O_p(n^{-1})$ under the weak condition that ∇g is Lipschitz-continuous in a neighborhood of β [Shao (1986)]. Note that for $g = a'\beta$, $\hat{B}_J$ is exactly equal to Γ_1. (4.1) is not satisfied since in general $\Gamma_1 \neq 0$. Theorem 4.5 shows that the order of Γ_1 is in general $O_p(n^{-1/2}h_n)$.

THEOREM 4.5. *Suppose that ∇g is Lipschitz-continuous in a neighborhood of β. Then*

$$\Gamma_1 = O_p(n^{-1/2}h_n).$$

PROOF. Since $\nabla g(\hat{\beta}) - \nabla g(\beta) = o_p(1)$, it suffices to show that

$$\sum_{i=1}^{n} \left(\hat{\beta}_{(i)} - \hat{\beta}\right) = \sum_{i=1}^{n} (1 - w_i)^{-1} w_i M^{-1} x_i r_i = O\left(n^{-1/2} h_n\right).$$

This is implied by

$$E \sum_{i=1}^{n} \sum_{j=1}^{n} (1 - w_i)^{-1} (1 - w_j)^{-1} w_i w_j w_{ij} r_i r_j = O\left(h_n^2\right) \tag{4.5}$$

since $M^{-1} = O(n^{-1})$, where $w_{ij} = x_i' M^{-1} x_j$. Since $E r_i r_j$ equals $(1 - w_i)\sigma^2$ for $i = j$ and $-w_{ij}\sigma^2$ for $i \neq j$, the left-hand side of (4.5) is equal to

$$\sigma^2 \left[\sum_{i=1}^{n} (1 - w_i)^{-1} w_i^3 - \sum_{i \neq j}^{n} (1 - w_i)^{-1} (1 - w_j)^{-1} w_i w_j w_{ij}^2 \right] \leq c h_n^2 \left(\sum_{i=1}^{n} w_i \right)^2$$
$$= c k^2 h_n^2. \qquad \square$$

From Theorem 4.5, if the order of h_n is higher than $O(n^{-1/2})$, then the order of Γ_1 is higher than $O_p(n^{-1})$ in general. Hence the order of $\hat{B}_J$ does not match that of $B(\hat{\theta})$ in view of Theorem 2.1. As a consequence, $n(\hat{B}_J - B(\hat{\theta}))$ does not converge to zero in probability, i.e., $\hat{B}_J$ is inconsistent. Theorem 4.5 does not show whether the order of Γ_1 can be lower than $O_p(n^{-1})$. But it is easy to find an example in which the order of Γ_1 is higher than $O_p(n^{-1})$.

EXAMPLE 4.1. Let $k = 1$, $x_{in} = 1$ for $i \neq n$ and $x_{nn} = a_n$, where $a_n \geq 1$. Let $\tau = \sum_{i=1}^{n} x_{in}^2 = n - 1 + a_n^2$. Then $M^{-1} = \tau^{-1}$, $w_i = \tau^{-1}$ for $i \neq n$ and $w_n = \tau^{-1} a_n^2$ and $h_n = \tau^{-1} a_n^2$. By a straightforward calculation, we have

$$\sum_{i=1}^{n} \left(\hat{\beta}_{(i)} - \hat{\beta}\right) = \frac{a_n^2 (1 - a_n^2)}{\tau(\tau - 1)(n - 1)} \sum_{i=1}^{n-1} e_i + \frac{a_n \left(a_n^2 - 1\right)}{\tau(\tau - 1)} e_n.$$

1. If $a_n = n^{1/2}$, then $h_n \to \frac{1}{2}$. Since $n^{1/2} a_n (a_n^2 - 1)/\tau(\tau - 1) \to 1$ as $n \to \infty$, the order of Γ_1 is exactly equal to $O_p(n^{-1/2} h_n) = O_p(n^{-1/2})$.
2. If $a_n = n^{5/12}$, then h_n is of order $n^{-1/6}$. Since $n^{3/4} a_n (a_n^2 - 1)/\tau(\tau - 1) \to 1$, the order of Γ_1 is $O_p(n^{-3/4})$, which is lower than $O_p(n^{-1/2} h_n) = O_p(n^{-2/3})$, but still higher than $O_p(n^{-1})$.

We now consider the bias of $\hat{B}_J$. It can be shown, assuming that g satisfies the condition in part (i) or (ii) of Theorem 2.1, that the dominating term of the bias of $\hat{B}_J$ is

$$L = 2^{-1} \sigma^2 \sum_{i=1}^{n} (1 - w_i)^{-1} w_i x_i' M^{-1} \nabla^2 g(\beta) M^{-1} x_i = O\left(n^{-1} h_n\right). \tag{4.6}$$

If the model is asymptotically balanced, i.e., $h_n = O(n^{-1})$, then $L = O(n^{-2})$. For an unbalanced model, in contrast to the result in Theorem 4.3(i), the order of the bias of $\hat{B}_J$ does not match that of the second order term of $B(\hat{\theta})$. In fact, Theorem 4.6 gives a lower bound for the order of L.

THEOREM 4.6. *Suppose that* $M = O(n)$ *and* $\nabla^2 g(\beta)$ *is either positive or negative definite at* β. *Then*

$$|L| \geq cn^{-1}g_n,$$

where $g_n = \sum_{i=1}^n w_i^2$.

REMARK 4.1. The condition that $\nabla^2 g(\beta)$ is either positive or negative definite is equivalent to $\nabla^2 g(\beta) \neq 0$ if β is a scalar.

REMARK 4.2. Since $h_n^2 \leq g_n$, Theorem 4.6 implies that $E\hat{B}_J - B(\hat{\theta}) = o(n^{-1})$ iff $h_n \to 0$. Thus, unlike $\hat{B}_{J(d)}$, $\hat{B}_J$ is not asymptotically unbiased if h_n does not converge to zero.

REMARK 4.3. Even if $h_n \to 0$, L may not be of the order $O(n^{-3/2}h_n^{1/2})$. For example, if g_n has the order $n^{-1/3}$ and h_n has the order $n^{-1/6}$ [an example given in Shao and Wu (1987)], then $|L| \geq cn^{-4/3}$ while $n^{-3/2}h_n^{1/2}$ has the order $n^{-19/12}$.

PROOF OF THEOREM 4.6. Suppose that $\nabla^2 g(\beta)$ is positive definite. Then $x'\nabla^2 g(\beta)x \geq \varepsilon x'x$ for any x and some positive ε. Note that $M = O(n)$ implies $x_i'M^{-2}x_i \geq cn^{-1}w_i$. Then from (4.6),

$$|L| \geq 2^{-1}\sigma^2\varepsilon \sum_{i=1}^n w_i x_i'M^{-2}x_i \geq cn^{-1}\sum_{i=1}^n w_i^2 = cn^{-1}g_n. \qquad \square$$

In the proof of Theorem 4.5 we have actually shown that

$$\operatorname{Var}\left(\sum_{i=1}^n w_i M^{-1}x_i r_i\right) = O\left(n^{-1}h_n^2\right). \tag{4.7}$$

Thus, we have

THEOREM 4.7. *If* (2.6) *holds and g satisfies the condition in part* (i) *or* (ii) *of Theorem* 2.1, *then*

$$\operatorname{MSE}\left(\hat{B}_J\right) = O\left(n^{-1}h_n^2\right).$$

PROOF. From (4.6) and (4.7), the result is true if $g = a'\beta$. For nonlinear g, a similar argument used in the proof of Theorem 4.4 yields the result. $\square$

Thus, $\hat{B}_J$ will usually have a much larger MSE than the weighted jackknife bias estimator. The unstable performance of $\hat{B}_J$ is again due to the fact that the $n^{-1/2}h_n$ order term in the Taylor expansion does not vanish for the unweighted jackknife.

4.3. *The bootstrap bias estimator.* Similar to the jackknife estimators, the behavior of bootstrap bias estimator $\hat{B}_b$ is closely related to that of bootstrap variance estimator v_b, as the following results indicate. The proofs of Theorems

4.8–4.10, which employ similar techniques to those of Theorems 4.1–4.4, are omitted here and can be found in Shao (1986).

THEOREM 4.8. *Suppose that* (2.4) *holds and g satisfies condition* (2.5) *with* $\lambda > 0$. *Then*:

(a) $\hat{B}_b = 2^{-1}\mathrm{tr}[\nabla^2 g(\hat{\beta})\nu_b] + O_p(n^{-1-\lambda/2})$.
(b) $\hat{B}_b = 2^{-1}\mathrm{tr}[\nabla^2 g(\hat{\beta})\mathrm{Var}\,\hat{\beta}] + o_p(n^{-1})$ *under the homoscedastic model.*
(c) $\hat{B}_b$ *is inconsistent under the heteroscedastic model.*

THEOREM 4.9. *Assume* (2.4) *and that g has a third order Lipschitz-continuous derivative. Then*:

(a) $E\hat{B}_b = B(\hat{\theta}) + 2^{-1}\mathrm{tr}[\nabla^2 g(\beta)B(\nu_b)] + O(n^{-3/2}h_n^{1/2})$, *where* $B(\nu_b) = E\nu_b - \mathrm{Var}\,\hat{\beta}$.
(b) $E\hat{B}_b = B(\hat{\theta}) + O(n^{-3/2}h_n^{1/2})$ *under the homoscedastic model.*
(c) *Under the heteroscedastic model,* $E\hat{B}_b = B(\hat{\theta}) + O(n^{-1})$ *in general. Therefore* $\hat{B}_b$ *is not asymptotically unbiased.*
(d) *If we only assume that g satisfies* (2.5), *the results still hold with* $O(n^{-3/2}h_n^{1/2})$ *in* (a) *and* (b) *replaced by* $O(n^{-1-\lambda/2})$.

THEOREM 4.10. *Assume* (2.6) *and that g has a third order Lipschitz-continuous derivative. Then*:

(a) $\mathrm{Var}\,\hat{B}_b = O(n^{-2}h_n)$.
(b) $\mathrm{MSE}(\hat{B}_b) = O(n^{-2}h_n)$ *under the homoscedastic model.*
(c) $\mathrm{MSE}(\hat{B}_b) = O(n^{-2})$ *under the heteroscedastic model.*

5. Comments on bias reduction. In the previous sections we have shown that the weighted jackknife is a handy and adequate tool for variance and bias (also MSE) estimation. We now briefly discuss a closely related problem: bias reduction. Since bias reduction is mathematically equivalent to asymptotically unbiased estimation of bias, Theorem 4.3 implies that under the homoscedastic model, the weighted jackknife estimator

$$\hat{\theta}_{J(d)} = \hat{\theta} - \hat{B}_{J(d)}$$

completely removes the leading term of $B(\hat{\theta})$ [i.e., the order of bias of $\hat{\theta}_{J(d)}$ matches that of the second order term of $B(\hat{\theta})$] no matter whether the model is balanced or not. Under the heteroscedastic model, the portion of bias removed by the weighted jackknife depends on the imbalance measure h_n. The leading term of $B(\hat{\theta})$ is completely removed iff the order of $n^{-1}h_n$ matches that of the second order term of $B(\hat{\theta})$.

On the other hand, the unweighted jackknife estimator

$$\hat{\theta}_J = \hat{\theta} - \hat{B}_J$$

cannot eliminate the leading term of $B(\hat{\theta})$ completely, as the results in Theorem 4.6 indicated.

One can also apply the bootstrap method to reduce bias. The bootstrap estimator of θ is

$$\hat{\theta}_b = \hat{\theta} - \hat{B}_b,$$

which eliminates the leading term of $B(\hat{\theta})$ completely under the homoscedastic model. Under the heteroscedastic model, $\hat{\theta}_b$ is not preferred due to the poor performance of $\hat{B}_b$ (see Section 4.3).

It is known that reducing the bias may increase the MSE of an estimator. For example, the weighted jackknife estimator $\hat{\theta}_{J(d)}$ of θ may have a larger MSE than that of the original estimator $\hat{\theta}$, although it reduces bias. The MSE of $\hat{\theta}$ is the sum of $\operatorname{Var}\hat{\theta}$ and $(B(\hat{\theta}))^2$ which have orders $O(n^{-1})$ and $O(n^{-2})$, respectively. The MSE of $\hat{\theta}_{J(d)}$ is equal to, up to the order of $O(n^{-2})$,

$$\operatorname{Var}\hat{\theta}_{J(d)} = \operatorname{Var}\hat{\theta} + \operatorname{Var}\hat{B}_{J(d)} - 2\operatorname{Cov}\left(\hat{\theta}, \hat{B}_{J(d)}\right). \tag{5.1}$$

From Theorem 4.4, the order of $\operatorname{Var}\hat{B}_{J(d)}$ is $O(n^{-2}h_n)$. By the Cauchy–Schwarz inequality, $\operatorname{Cov}(\hat{\theta}, \hat{B}_{J(d)}) = O(n^{-3/2}h_n^{1/2})$. Thus, in the worst case [the second and third terms on the right-hand side of (5.1) do not cancel out each other], the order of $\operatorname{Var}\hat{B}_{J(d)} - 2\operatorname{Cov}(\hat{\theta}, \hat{B}_{J(d)})$ is $O(n^{-3/2}h_n^{1/2})$, which is equal to $O(n^{-2})$ [the order of $(B(\hat{\theta}))^2$] if $h_n = O(n^{-1})$, but higher than $O(n^{-2})$ in general. Since the dominating term in the MSE of $\hat{\theta}$ and $\hat{\theta}_{J(d)}$ is of the order $O(n^{-1})$, the increase in MSE by using the jackknife is still asymptotically relatively negligible. If the model is very unbalanced and n is small, this increase may be large.

However, we should keep in mind that the jackknife estimator $\hat{\theta}_{J(d)}$ was originally designed to eliminate bias, i.e., the focus is on the bias of the estimator rather than other measures of statistical accuracy. Naturally, one may pay the price for an increased MSE. Thus, whether to use $\hat{\theta}_{J(d)}$ as an estimator of θ depends on how important bias is in practice. One needs to balance the advantage of unbiasedness against the drawback of a larger MSE.

APPENDIX

Proof of Lemma 3.1. For any $s \in \mathbf{S}_r$, let $\bar{s} = \{j_1, \dots, j_d\}$, $s_i = \{j_1, \dots, j_i\} \cup s$, $i = 1, \dots, d$, $\hat{\beta}_{s_0} = \hat{\beta}_s$ and $\hat{\beta}_{s_d} = \hat{\beta}$. Noting that $s_i = s_{i-1} \cup \{j_i\}$ and using an updating formula [Miller (1974)], we have

$$\hat{\beta} - \hat{\beta}_s = \sum_{i=1}^{d} \left(\hat{\beta}_{s_i} - \hat{\beta}_{s_{i-1}}\right) = \sum_{i=1}^{d} \delta_{j_i} M_{s_i}^{-1} r_{j_i} x_{j_i}, \tag{A1}$$

where $\delta_{j_i} = (1 - x_{j_i}' M_{s_i}^{-1} x_{j_i})^{-1}$ and $r_{j_i} = y_{j_i} - x_{j_i}' \hat{\beta}_{s_i}$ is the j_ith residual from fitting the subset model $y_{s_i} = X_{s_i}\beta + e_{s_i}$. Let $\mu_i = n - \#(s_i)$, $i = 1, \dots, d$, where $\#(s)$ is the number of elements in s. Then $\mu_i + 1 \le d$. By Lemma 4 of Shao and Wu (1987),

$$x_i' M_s^{-1} x_i \le (1 - dh_n)^{-1} w_i \tag{A2}$$

for any $s \in \mathbf{S}_r$, $d = n - r$. Then

$$\left(1 - x_{j_i}' M_{s_i}^{-1} x_{j_i}\right)^{-1} \le \left[1 - (1 - \mu_i h_n)^{-1} h_n\right]^{-1}$$
$$\le \left[1 - (\mu_i + 1) h_n\right]^{-1} \le (1 - d h_n)^{-1}.$$

Hence from (A1),

$$\text{(A3)}\quad \|\hat{\beta}_s - \hat{\beta}\|^q \le c \sum_{i=1}^{d} \|\hat{\beta}_{s_{i-1}} - \hat{\beta}_{s_i}\|^q \le c \sum_{i=1}^{d} \left[(1 - d h_n)^{-2} (r_{j_i})^2 x_{j_i}' M_{s_i}^{-2} x_{j_i}\right]^{q/2}.$$

Then by (3.1), (A2) and Lemma 2.1,

$$E\|\hat{\beta}_s - \hat{\beta}\|^q \le c n^{-q/2} \sum_{i \in \bar{s}} w_i^{q/2} E|r_{j_i}|^q \le c n^{-q/2} \sum_{i \in \bar{s}} w_i^{q/2}. \qquad \square$$

PROOF OF LEMMA 3.2. The first part of Lemma 3.2 follows directly from Lemma 3.1. Let $m_{j_i j_l}$ be the (p, q)th element of $M_{s_i}^{-1} x_{j_i} x_{j_l}' M_{s_l}^{-1}$, $\bar{t} = \{l_1, \ldots, l_d\}$ and $t_j = \{l_1, \ldots, l_j\} \cup t$, $j = 1, \ldots, d$. From (A1),

$$\operatorname{Cov}\left(\gamma_{pq}^{(s)}, \gamma_{pq}^{(t)}\right) = \sum_{i=1}^{d} \sum_{l=1}^{d} \sum_{j=1}^{d} \sum_{m=1}^{d} \delta_{j_i} \delta_{j_l} \delta_{l_j} \delta_{l_m} m_{j_i j_l} m_{l_j l_m} \operatorname{Cov}\left(r_{j_i} r_{j_l}, r_{l_j} r_{l_m}\right).$$

From (A2), $|m_{j_i j_l}| \le c n^{-1} (w_{j_i} w_{j_l})^{1/2}$. Since $\Sigma_{i=1}^d \Sigma_{l=1}^d (w_{j_i} w_{j_l})^{1/2} = \left(\Sigma_{i=1}^d w_{j_i}^{1/2}\right)^2 \le d \Sigma_{i=1}^d w_{j_i} = d \Sigma_{i \in \bar{s}} w_i$, the second and the third parts of Lemma 3.2 follow if

$$\text{(A4)}\qquad \left|\operatorname{Cov}\left(r_{j_i} r_{j_l}, r_{l_j} r_{l_m}\right)\right| \le c h_n$$

when $(j_i, j_l) \ne (l_j, l_m)$, where c is independent of j_i and l_j. Let $\tau = \min(\#(s_i), \#(t_j))$, $u_{j_i p}$ be $1 - x_{j_i}' M_{s_i}^{-1} x_{j_i}$ if $p = j_i$ and $-x_{j_i}' M_{s_i}^{-1} x_p$ if $p \ne j_i$, and $v_{l_j p}$ be $1 - x_{l_j}' M_{t_j}^{-1} x_{l_j}$ if $p = l_j$ and $-x_{l_j}' M_{t_j}^{-1} x_p$ if $p \ne l_j$. Assume that $j_i \le j_l$ and $l_j \le l_m$. Then

$$\operatorname{Cov}\left(r_{j_i} r_{j_l}, r_{l_j} r_{l_m}\right) = \sum_{p=1}^{\tau} u_{j_i p} u_{j_l p} v_{l_j p} v_{l_m p} \left(E e_p^4 - \sigma_p^4\right) + 2 \sum_{p \ne q}^{\tau} u_{j_i p} u_{j_l q} v_{l_j p} v_{l_m q} \sigma_p^2 \sigma_q^2.$$

Now (A4) follows since if $j_i \ne p$ and $l_j \ne p$, $|u_{j_i p}| \le c(w_{j_i} w_p)^{1/2} \le c h_n$ and $|v_{l_j p}| \le c(w_{l_j} w_p)^{1/2} \le c h_n$. $\square$

PROOF OF LEMMA 3.3. Let $\gamma_{pq}^{(s)}$ be the (p, q)th element of $(\hat{\beta}_s - \hat{\beta})(\hat{\beta}_s - \hat{\beta})'$, $\mathbf{S}_{r_1} = \{s \in \mathbf{S}_r \colon j \in s\}$ and $\mathbf{S}_{r_2} = \mathbf{S}_r - \mathbf{S}_{r_1}$. Then

$$\text{(A5)}\quad E e_j v_{pq} = \binom{n-k}{d-1}^{-1} \sum_{s \in \mathbf{S}_{r_1}} \omega_s E\left(e_j \gamma_{pq}^{(s)}\right) + \binom{n-k}{d-1}^{-1} \sum_{s \in \mathbf{S}_{r_2}} \omega_s E\left(e_j \gamma_{pq}^{(s)}\right).$$

For $s \in \mathbf{S}_{r_1}$, $j \in s$. Let $\bar{s} = \{j_1, \ldots, j_d\}$, $s_i = \{j_1, \ldots, j_i\} \cup s$, $i = 1, \ldots, d$, $\hat{\beta}_s = \hat{\beta}_{s_0}$, $\hat{\beta} = \hat{\beta}_{s_d}$. Denote the (p, q)th element of $(\hat{\beta}_{s_{i-1}} - \hat{\beta}_{s_i})(\hat{\beta}_{s_{m-1}} - \hat{\beta}_{s_m})'$ by b_{pq}^{im}. Then $E e_j \gamma_{pq}^{(s)} = \Sigma_{i=1}^d \Sigma_{m=1}^d E e_j b_{pq}^{im}$. Let r_{j_i} be defined as in the proof of Lemma 3.1. Then $E e_j b_{pq}^{im} = \delta_{j_i} \delta_{j_m} M_{s_i}^{-1} x_{j_i} x_{j_m}' M_{s_m}^{-1} E e_j r_{j_i} r_{j_m}$. Since $j \in s$, $j \ne j_i$,

$i = 1, \dots, d,$

$$|Ee_j r_{j_i} r_{j_m}| = \left|\left(x'_{j_i} M_{s_i}^{-1} x_j\right)\left(x'_{j_m} M_{s_m}^{-1} x_j\right) Ee_j^3\right| \le c w_{j_i}^{1/2} w_{j_m}^{1/2} w_j,$$

by (3.1) and (A2). Thus, $|Ee_j b_{pq}^{im}| \le cn^{-1} w_{j_i} w_{j_m} w_j$ and $|Ee_j \gamma_{pq}^{(s)}| \le cn^{-1} w_j (\sum_{i \in \bar{s}} w_i)$ since $\sum_{i \in \bar{s}} w_i \le dh_n < 1$. Then the first term on the right side of (A5) can be bounded in absolute value by

$$cn^{-1} w_j \binom{n-k}{d-1}^{-1} \sum_{s \in \mathbf{S}_{r_1}} \sum_{i \in \bar{s}} w_i \le cn^{-1} w_j \binom{n-k}{d-1}^{-1} \binom{n-2}{d-1} \le cn^{-1} w_j.$$

For $s \in \mathbf{S}_{r_2}$, $j \in \bar{s}$. Suppose that $\bar{s} = \{j_1, \dots, j_d\}$ and $j_d = j$. Denote $\hat{\beta}_{s_{d-1}}$ by $\hat{\beta}_{(j)}$. Using the same notation as before, since for $i < m = d$, $Ee_j r_j r_{j_i} = 0$ and for $i \le m < d$, $Ee_j b_{pq}^{im} = Ee_j Eb_{pq}^{im} = 0$, we have

$$Ee_j \gamma_{pq}^{(s)} = \sum_{i=1}^{d} Ee_j b_{pq}^{ii} + \sum_{i \ne m}^{d} Ee_j b_{pq}^{im} = Ee_j b_{pq}^{dd}.$$

Hence the second term on the right side of (A5) can be bounded in absolute value by

$$\begin{aligned}\binom{n-k}{d-1}^{-1} \sum_{s \in \mathbf{S}_{r_2}} E|e_j b_{pq}^{dd}| &\le \binom{n-k}{d-1}^{-1} \sum_{s \in \mathbf{S}_{r_2}} E|e_j| \, \|\hat{\beta}_{(j)} - \hat{\beta}\|^2 \\ &\le c\binom{n-k}{d-1}^{-1} \binom{n-1}{d-1} \left(E\|\hat{\beta}_{(j)} - \hat{\beta}\|^4\right)^{1/2} \le cn^{-1} w_j.\end{aligned}$$

The last inequality follows from Lemma 3.1. □

PROOF OF LEMMA 4.1. For $s_i \in \mathbf{S}_r$, $i = 1, 2, 3, 4$, by (A3), (3.1) and Lemma 2.1,

$$\begin{aligned}E\left(\operatorname{tr}(\nu_{J(d)})\right)^4 &\le cn^{-4} \binom{n-k}{d-1}^{-4} \prod_{i=1}^{4} \sum_{s_i \in \mathbf{S}_r} \sum_{l_i \in \bar{s}_i} w_{l_i} \\ &= ck^4 n^{-4} \binom{n-k}{d-1}^{-4} \binom{n-1}{d-1}^4 = O(n^{-4}). \qquad \square\end{aligned}$$

Acknowledgments. This paper is part of the author's Ph.D. thesis written at the University of Wisconsin-Madison. The author would like to thank Professor C. F. J. Wu for his guidance and valuable help during the preparation of the paper. Thanks are also due the Associate Editor and the referee for their comments and suggestions.

REFERENCES

BERAN, R. (1986). Discussion of "Jackknife, bootstrap and other resampling methods in regression analysis" by C. F. J. Wu. *Ann. Statist.* **14** 1295–1298.

BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *Ann. Statist.* **9** 1196–1217.

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

HINKLEY, D. V. (1977). Jackknifing in unbalanced situations. *Technometrics* **19** 285–292.
HUBER, P. J. (1973). Robust regression: Asymptotics, conjectures and Monte Carlo. *Ann. Statist.* **1** 799–821.
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.
MILLER, R. G. (1964). A trustworthy jackknife. *Ann. Math. Statist.* **35** 556–568.
MILLER, R. G. (1974). An unbalanced jackknife. *Ann. Statist.* **2** 880–891.
PARR, W. C. (1985). Jackknifing differentiable statistical functionals. *J. Roy. Statist. Soc. Ser. B* **47** 56–66.
QUENOUILLE, M. (1956). Notes on bias in estimation. *Biometrika* **43** 353–360.
SHAO, J. (1986). On resampling methods for variance and bias estimation in linear models. Technical Report 788, Dept. Statistics, Univ. Wisconsin-Madison.
SHAO, J. and WU, C. F. J. (1986). Some general theory for the jackknife. Technical Report 797, Dept. Statistics, Univ. Wisconsin-Madison.
SHAO, J. and WU, C. F. J. (1987). Heteroscedasticity-robustness of jackknife variance estimators in linear models. *Ann. Statist.* **15** 1563–1579.
TIBSHIRANI, R. (1986). Discussion of "Jackknife, bootstrap and other resampling methods in regression analysis" by C. F. J. Wu. *Ann. Statist.* **14** 1335–1339.
TUKEY, J. (1958). Bias and confidence in not quite large samples. *Ann. Math. Statist.* **29** 614.
WU, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *Ann. Statist.* **14** 1261–1350.

DEPARTMENT OF STATISTICS
MATHEMATICAL SCIENCES BUILDING
PURDUE UNIVERSITY
WEST LAFAYETTE, INDIANA 47907