The Annals of Statistics
1987, Vol. 15, No. 4, 1563–1579

HETEROSCEDASTICITY-ROBUSTNESS OF JACKKNIFE VARIANCE ESTIMATORS IN LINEAR MODELS[1]

BY JUN SHAO AND C. F. J. WU[2]

University of Wisconsin-Madison

The asymptotic unbiasedness and consistency of three types of jackknife variance estimators in the presence of error variance heteroscedasticity in linear models are studied. The results are given in terms of the number of observations deleted and measures of imbalance of the model. The consistency of a class of Wu's weighted jackknife variance estimators for nonlinear parameters is also studied. A necessary and sufficient condition is given for the asymptotic unbiasedness and consistency of the unweighted delete-1 jackknife variance estimator and Hinkley's weighted delete-1 jackknife variance estimator. This condition is more stringent than those required for Wu's weighted jackknife. Comparison of the three delete-1 jackknife variance estimators in terms of their biases also favors the latter method.

1. Introduction. The jackknife method was proposed for bias reduction of a point estimator [Quenouille (1956)] and for variance estimation [Tukey (1958)]. Its theoretical justification has largely been the consistency of point and variance estimators and the asymptotic normality of the associated t-statistic. If alternative methods are available and possess the preceding properties, the jackknife does not have any apparent superiority in theoretical performance. When the standard (i.e., linearization) method for variance estimation is only available under restrictive conditions such as normality, resampling methods such as the jackknife and bootstrap are often available. They are not logically based on the same restrictive conditions. Their small sample performance is, therefore, less susceptible to violations of these assumptions. The distribution-robustness of the jackknife was recognized by Tukey and subsequent workers. Another robustness aspect of the jackknife was later pointed out by Hinkley (1977). In the context of regression models, he proposed a weighted delete-1 jackknife variance estimator v_H (1.3) and showed its desirable asymptotic performance even when the errors are heteroscedastic. Wu (1986) proposed a class of weighted jackknife variance estimators $v_{J(d)}$ (1.4), allowing the deletion of an arbitrary number of observations denoted by d. In the case of the delete-1 jackknife, Wu's weighting scheme is different from Hinkley's. For homoscedastic errors, $v_{J(d)}$ is unbiased, while v_H is not. Like v_H, the delete-1 version of $v_{J(d)}$, denoted by $v_{J(1)}$, was also shown to be heteroscedasticity-robust [Wu (1986)].

The main purposes of this paper are (i) to study rigorously this robustness aspect of $v_{J(d)}$ for general d, (ii) to find necessary and sufficient conditions for

Received December 1985; revised April 1987.
[1]Work supported by the joint NSF-AFOSR grant DMS-85-02303 and ISSA86-0068.
[2]Partially supported by the A. P. Sloan Foundation for Basic Research.
AMS 1980 *subject classifications*. Primary 62J05; secondary 62F35.
Key words and phrases. Weighted jackknife variance estimator, heteroscedasticity-robust, asymptotic unbiasedness, consistency, imbalance measure of a linear model.

the heteroscedasticity-robustness of v_H and the unweighted jackknife variance estimator v_J (1.2) and (iii) to compare the biases of v_J, v_H and $v_{J(1)}$.

We assume the linear model $y_i = x_i'\beta + e_i$, $i = 1, \dots, n$, where x_i is a $k \times 1$ deterministic vector, β is the $k \times 1$ vector of parameters and e_i are independent with mean zero and variances σ_i^2. We assume the σ_i^2 are uniformly bounded. Writing $y = (y_1, \dots, y_n)'$, $X = [x_1, \dots, x_n]'$ and $e = (e_1, \dots, e_n)'$, the model can be expressed as

$$y = X\beta + e, \qquad \operatorname{Var}(e) = \operatorname{diag}(\sigma_i^2). \tag{1.1}$$

When σ_i^2 are different (constant), the errors are called heteroscedastic (homoscedastic). The reader should be aware that e_i, σ_i^2, x_i, etc. may vary with n. A subscript n could be appended for clarity, but will usually be dropped for simplicity.

Let $M = X'X = \sum_1^n x_i x_i'$ be the moment matrix of (1.1). Assuming that M is invertible, the least squares estimator (LSE) of β for (1.1) and its variances are, respectively,

$$\hat{\beta} = M^{-1}X'y \quad \text{and} \quad \operatorname{Var}\hat{\beta} = M^{-1}\sum_1^n \sigma_i^2 x_i x_i' M^{-1}.$$

A straightforward extension of the i.i.d. jackknife to regression was studied by Miller (1974),

$$v_J = n^{-1}(n-1)\sum_1^n \big(\hat{\beta}_{(i)} - \hat{\beta}_{(\cdot)}\big)\big(\hat{\beta}_{(i)} - \hat{\beta}_{(\cdot)}\big)', \tag{1.2}$$

where $\hat{\beta}_{(i)}$ is the LSE of β after deleting the ith pair (x_i, y_i) and $\hat{\beta}_{(\cdot)}$ is the average of the $\hat{\beta}_{(i)}$. Hinkley (1977) pointed out several shortcomings of v_J. Based on the concept of weighted pseudovalues, he proposed a weighted jackknife variance estimator

$$v_H = n(n-k)^{-1}\sum_1^n (1 - w_i)^2\big(\hat{\beta}_{(i)} - \hat{\beta}\big)\big(\hat{\beta}_{(i)} - \hat{\beta}\big)', \tag{1.3}$$

where $w_i = x_i'M^{-1}x_i$. Wu (1986) argued that v_H could be further improved by choosing a different weighting scheme and/or allowing a larger number of observations to be deleted. His method is described as follows.

Let $\mathbb{S}_r$ be the collection of subsets of $\{1, \dots, n\}$ which have size r. For $s = \{i_1, \dots, i_r\} \in \mathbb{S}_r$ and A an $n \times p$ matrix, A_s is defined to be the submatrix of A consisting of the i_1th, $\dots$, i_rth rows of A. Denote $X_s'X_s$ by M_s. For simplicity, it is assumed that M_s is positive definite for all $s \in \mathbb{S}_r$. Let $\hat{\beta}_s = M_s^{-1}X_s'y_s$ be the LSE of β based on (x_i, y_i), $i \in s$. The weighted delete-d jackknife estimator of $\operatorname{Var}\hat{\beta}$ is defined to be

$$v_{J(d)} = \left(\binom{n-k}{d-1}|M|\right)^{-1}\sum_s |M_s|\big(\hat{\beta}_s - \hat{\beta}\big)\big(\hat{\beta}_s - \hat{\beta}\big)', \tag{1.4}$$

where $|M_s|$ is the determinant of M_s and $\sum_s$ is summation over all $s \in \mathbb{S}_r$. An important special case is the delete-1 ($d = 1$) estimator

$$v_{J(1)} = \sum_1^n (1 - w_i)\big(\hat{\beta}_{(i)} - \hat{\beta}\big)\big(\hat{\beta}_{(i)} - \hat{\beta}\big)'. \tag{1.5}$$

For large $\binom{n}{d}$, the computation of $v_{J(d)}$ is cumbersome. As in the bootstrap, Monte Carlo approximation by randomly selecting J distinct subsets $\left(J \ll \binom{n}{d}\right)$ can be used. Several theoretical results obtained here for $v_{J(d)}$ can be extended to this version [Shao (1987)].

The weighted delete-d jackknife can be used for other purposes such as confidence intervals and bias reduction [Wu (1986)]. For confidence intervals based on the histogram of $\hat{\beta}_s$, the choice of $d = 1$ and more generally bounded d is not favored since it is known that, in the case of one-sample mean and nonnormal errors, the histogram is asymptotically normal iff d and $n - d$ both diverge to ∞ [Wu (1987)]. On the other hand, for variance estimation $v_{J(1)}$ and v_H are more likely to be used in practice because they are simpler to compute.

These three types of jackknife variance estimators are designed primarily for homoscedastic errors since the original point estimator is the unweighted least squares $\hat{\beta}$. Stable performance under heteroscedastic errors of an estimator that does not make explicit use of such information is called *heteroscedasticity-robust*. Two such definitions are considered.

DEFINITION 1. Let v be an estimator of $\operatorname{Var}\hat{\beta}$ and $\operatorname{Var}(e) = \operatorname{diag}(\sigma_i^2)$.

(i) v is AU-robust if v is asymptotically unbiased, i.e., $nE(v - \operatorname{Var}\hat{\beta}) \to 0$.
(ii) v is C-robust if v is consistent, i.e., $n(v - \operatorname{Var}\hat{\beta}) \to 0$ in probability.

Note that the usual variance estimator in linear model theory,

$$\hat{v} = \hat{\sigma}^2 M^{-1}, \qquad \hat{\sigma}^2 = (n-k)^{-1}\sum_1^n r_i^2, \qquad r_i = y_i - x_i'\hat{\beta}, \tag{1.6}$$

is unbiased and consistent for homoscedastic errors, but is neither AU- nor C-robust. The bias of $\hat{v}$ may be large and is generally of the order n^{-1}, the same as that of $\operatorname{Var}\hat{\beta}$. A variance estimator based on "bootstrapping" the residuals r_i [Efron (1979)] is identical to $\hat{v}$ and is, therefore, not robust. Wu (1986) proposed a method for resampling residuals that gives robust variance estimators.

Our paper is organized as follows. Section 2 contains several useful technical lemmas. Theorem 1 gives an upper bound on the bias of $v_{J(d)}$. As a consequence, the AU-robustness of $v_{J(d)}$ is obtained under weak conditions. In the case of $d = 1$, it gives a more precise result than a previous one on $v_{J(1)}$ [Wu, (1986), Theorem 5]. If the variances get closer to each other asymptotically, the AU-robustness of $v_{J(d)}$ holds under no assumptions on d or X (Theorem 2). Section 4 contains results on the C-robustness of $v_{J(d)}$ and its extension to nonlinear parameters. Crucial to the proof in the nonlinear case is a useful result (Proposition 1), which gives a bound on the maximum mean deviation of $\hat{\beta}_s$ from $\hat{\beta}$ with s ranging over $\mathbb{S}_r$. Theorem 5 shows that $h_n \to 0$, h_n given in (2.2), is necessary and sufficient for the AU- and C-robustness of v_J and v_H. On the other hand, according to Theorem 2, $v_{J(d)}$ is asymptotically unbiased (for nearly homoscedastic errors) *without* any condition on h_n. Section 6 provides a more refined comparison of v_J, v_H and $v_{J(1)}$ in terms of their biases. The comparison is again more favorable for $v_{J(1)}$, i.e., up to a certain order, v_J is upward biased and v_H downward biased and the orders of their biases are no smaller than that of the bias of $v_{J(1)}$.

The asymptotic approach adopted here is different from the prevailing one for linear models. It is common to assume that $n^{-1}M$ converges to a positive definite matrix. This and the boundedness of $\{x_i\}$ imply that h_n is of the order n^{-1}, which may be too restrictive an assumption for unbalanced samples of small or moderate size. In the study of AU-robustness, we make weak or no assumptions on h_n, which can be interpreted as an imbalance measure of the model. The order of the bias depends monotonically on h_n. The smaller h_n is, the better the asymptotic approximation. Our comparison of the biases of v_J, v_H and $v_{J(1)}$ is made possible with the use of another imbalance measure g_n defined in (3.2). On the other hand, the common approach previously cited implies that g_n is of the order n^{-1} with the consequence that the biases of these three estimators cannot be differentiated.

2. Some technical lemmas. In this section, we state and prove several useful lemmas. Throughout the paper, we use c to denote a positive generic constant, i.e., c is a positive constant but may have different values in different places. The trace of a matrix A is denoted by $\operatorname{tr}(A)$. Denote a nonnegative (positive) definite matrix A by $A \geq 0$ ($A > 0$); $A \geq B$ means $A - B \geq 0$. Let $A_n = [a_{ij}^{(n)}]$ be a sequence of $k \times k$ matrices with fixed k and α_n a sequence of positive numbers. We say $A_n = O(\alpha_n)$ [$o(\alpha_n)$, $O_p(\alpha_n)$, $o_p(\alpha_n)$, respectively] iff $a_{ij}^{(n)} = O(\alpha_n)$ [$o(\alpha_n)$, $O_p(\alpha_n)$, $o_p(\alpha_n)$, respectively] for all $1 \leq i, j \leq k$. For simplicity, $O(\alpha_n)$ will be used for both numbers and matrices.

LEMMA 1. *Let $A_n \geq 0$ be a sequence of $k \times k$ matrices.*

(i) *The following conditions are equivalent*: (a) $A_n = O(\alpha_n)$. (b) $\operatorname{tr}(A_n) = O(\alpha_n)$. (c) *There is an N such that $A_n \leq c\alpha_n I_k$ for all $n > N$, where I_k is the $k \times k$ identity matrix.*

(ii) *The following conditions are equivalent*: (a) $A_n = o(\alpha_n)$. (b) $\operatorname{tr}(A_n) = o(\alpha_n)$. (c) *For any $\varepsilon > 0$, there is an N such that $A_n \leq \varepsilon\alpha_n I_k$ for all $n > N$.*

Similar results hold if the O or o in (i) *and* (ii) *are replaced by O_P or o_P, respectively.*

PROOF. We prove (i) only. The other proofs are similar.

(a) ⇒ (b) follows from the definition of $A_n = O(\alpha_n)$.

(b) ⇒ (c) holds because $A_n \geq 0$ implies $A_n \leq \operatorname{tr}(A_n) I_k$.

(c) ⇒ (a). Let $A_n = [a_{ij}^{(n)}]$. Since $a_{ii}^{(n)} \geq 0$, (c) implies $a_{ii}^{(n)} = O(\alpha_n)$. Then for all i, j,

$$|a_{ij}^{(n)}| \leq \left(a_{ii}^{(n)} a_{jj}^{(n)}\right)^{1/2} = O(\alpha_n). \qquad \square$$

The equivalence between (a) and (b) is a useful tool since it replaces a condition on matrices by a similar condition on scalars. As an immediate consequence of Lemma 1, we have

LEMMA 2. (i) *If* $0 \le A_n = O(\alpha_n)$, $0 \le B_n = O(\alpha_n)$, C_n *is symmetric and* $-B_n \le C_n \le A_n$, *then* $C_n = O(\alpha_n)$. *The same result holds if* $O(\alpha_n)$ *is replaced by* $o(\alpha_n)$, $O_P(\alpha_n)$, $o_P(\alpha_n)$.

(ii) *For random matrices* A_n *with* $A_n \ge 0$, $EA_n = O(\alpha_n)$ [*or* $o(\alpha_n)$] *implies* $A_n = O_P(\alpha_n)$ [*or* $o_P(\alpha_n)$].

LEMMA 3. *Let* x_i *and* M *be defined in* (1.1) *and*

$$M^{-1}x_i x_j' M^{-1} = \left[m_{pq}^{ij}\right]_{p,q} \tag{2.1}$$

be a $k \times k$ *matrix. Suppose that* $M^{-1} = O(n^{-1})$. *Then for any* p *and* q,

(i) $\sum_1^n m_{pq}^{ii} = O(n^{-1})$,
(ii) $\sum_1^n (m_{pq}^{ii})^2 = O(h_n n^{-2})$,

where

$$h_n = \max_{i \le n} w_i \quad \textit{and} \quad w_i = x_i' M^{-1} x_i \tag{2.2}$$

and
(iii) $\sum_{i<j}^n (m_{pq}^{ij})^2 = O(n^{-2})$.

PROOF. (i) Note that $\sum_1^n m_{pq}^{ii}$ is the (p,q)th element of the matrix $\sum_1^n M^{-1}x_i x_i' M^{-1} = M^{-1}MM^{-1} = M^{-1}$ and is, therefore, of the order n^{-1}.

(ii)
$$\sum_1^n \left(m_{pq}^{ii}\right)^2 \le \sum_1^n \left(m_{pp}^{ii}\right)\left(m_{qq}^{ii}\right) \le cn^{-1} \max_{i \le n}\left(m_{qq}^{ii}\right)$$
$$\le cn^{-1} \max_{i \le n}\left(x_i' M^{-2} x_i\right) \le ch_n n^{-2}.$$

(iii) This follows from (i) and $\sum_{i<j}^n (m_{pq}^{ij})^2 \le \sum_{i,j=1}^n (m_{pp}^{ii})(m_{qq}^{jj})$. □

LEMMA 4. *Let* h_n *and* w_i *be defined in* (2.2), $s \in \mathbb{S}_r$ *and* $d = n - r$. *Suppose that* $dh_n < 1$. *Then*

$$M_s^{-1} - M^{-1} \le (1 - dh_n)^{-1} M^{-1} X_{\bar{s}}' X_{\bar{s}} M^{-1} \le dh_n (1 - dh_n)^{-1} M^{-1}, \tag{2.3}$$

where $\bar{s}$ *is the complement of* s *and*

$$M_s^{-1} \le (1 - dh_n)^{-1} M^{-1} \quad \textit{and} \quad x_i' M_s^{-1} x_i \le (1 - dh_n)^{-1} w_i. \tag{2.4}$$

PROOF. Since (2.4) follows easily from (2.3), we only need to show (2.3). For any $s \in \mathbb{S}_r$,

$$\operatorname{tr}\left(X_{\bar{s}}' M^{-1} X_{\bar{s}}\right) \le dh_n \quad \text{and} \quad \left(I - X_{\bar{s}}' M^{-1} X_{\bar{s}}\right)^{-1} \le (1 - dh_n)^{-1} I.$$

Therefore,

$$M_s^{-1} - M^{-1} = M^{-1}X_{\bar{s}}'\left(I - X_{\bar{s}}'M^{-1}X_{\bar{s}}\right)^{-1}X_{\bar{s}}M^{-1}$$
$$\leq (1 - dh_n)^{-1}M^{-1}X_{\bar{s}}'X_{\bar{s}}M^{-1} \leq dh_n(1 - dh_n)^{-1}M^{-1}. \qquad \square$$

3. AU-robustness of $v_{J(d)}$. It was shown in Wu (1986) that $v_{J(d)}$ is an unbiased estimator of $\operatorname{Var}\hat{\beta}$ for homoscedastic errors and the usual variance estimator $\hat{v}$ (1.6) is biased under violations of homoscedasticity. The AU-robustness of the delete-1 jackknife $v_{J(1)}$ was proved in Wu (1986), but the technique does not handle the general case. In this and the next sections, we study the robustness issue of $v_{J(d)}$ for general d. The main result, Theorem 1, provides more information than AU-robustness, i.e., an upper bound on the order of magnitude of the bias of $v_{J(d)}$ is given in terms of n, d and h_n defined in (2.2).

As in Lemma 3, we assume for the rest of the paper

$$M^{-1} = O(n^{-1}). \tag{3.1}$$

This is much weaker than $n^{-1}M$ converging to a positive definite matrix, a condition assumed in Miller (1974). As remarked in Section 1, an important feature of our approach is the incorporation of an imbalance measure such as h_n in our results. Another imbalance measure closely related to h_n is

$$g_n = \sum_1^n w_i^2. \tag{3.2}$$

From $\sum_1^n w_i^2 \geq n^{-1}(\sum_1^n w_i)^2 = n^{-1}k^2$ and $h_n \geq n^{-1}\sum_1^n w_i$, we have

$$n^{-1}k \leq k^{-1}g_n \leq h_n. \tag{3.3}$$

That is, the orders of h_n and of g_n are at least n^{-1}. In Sections 5 and 6, the use of h_n and g_n plays a key role in distinguishing between three delete-1 jackknife variance estimators. For unbalanced models, h_n and g_n can be of larger order than n^{-1} (examples given in Section 6). Only in this situation, the biases of the three estimators are of different orders.

Throughout the paper, d_n, which may depend on n, is used to denote the number of observations deleted in the jackknife procedure.

THEOREM 1. *Assume that*

$$\sup_n d_n h_n < 1. \tag{3.4}$$

Then

$$Ev_{J(d)} = \operatorname{Var}\hat{\beta} + O\left(n^{-1}d_n h_n\right),$$

which implies the AU-robustness of $v_{J(d)}$ under $d_n h_n \to 0$.

REMARKS. (i) The condition $d_n h_n \to 0$ imposes a reciprocal relation between d_n and h_n. The more unbalanced the model is (i.e., slower rate of convergence of $h_n \to 0$), the smaller d_n has to be.

(ii) In view of (3.3), an important implication of condition (3.4) is

$$d_n \le h_n^{-1} \le k^{-1}n, \tag{3.5}$$

i.e., d_n is smaller than a fraction of n. If too many observations are deleted, the robustness of the estimator will be lost. For example, in the extreme case of $d_n = n - k$, the corresponding [delete-$(n-k)$] jackknife variance estimator is identical to $\hat{v}$ (1.6) [Wu (1986)] and is, therefore, not robust.

Proof of Theorem 1. For simplicity, we drop the subscript n in d_n. For any $s \in \mathbb{S}_r$, $r = n - d$,

$$\begin{aligned}\hat{\beta}_s - \hat{\beta} &= M_s^{-1}X_s'y_s - M^{-1}X'y \\ &= \left(M_s^{-1} - M^{-1}\right)X_s'y_s + M^{-1}(X_s'y_s - X'y).\end{aligned} \tag{3.6}$$

The two terms in (3.6) are uncorrelated since $X'y - X_s'y_s$ depends only on $\bar{s}$, the complement of s. Then

$$\operatorname{Var}\left(\hat{\beta}_s - \hat{\beta}\right) = \left(M_s^{-1} - M^{-1}\right)X_s'D_sX_s\left(M_s^{-1} - M^{-1}\right) + M^{-1}X_{\bar{s}}'D_{\bar{s}}X_{\bar{s}}M^{-1},$$

where $D_s = \operatorname{Var}(e_s)$. Hence, $Ev_{J(d)} = S_1 - S_2 + S_3$, where

$$S_1 = \binom{n-1}{d-1}\binom{n-k}{d-1}^{-1}\operatorname{Var}\hat{\beta},$$

$$S_2 = \binom{n-k}{d-1}^{-1}|M|^{-1}\sum_s(|M| - |M_s|)M^{-1}X_{\bar{s}}'D_{\bar{s}}X_{\bar{s}}M^{-1},$$

$$S_3 = \binom{n-k}{d-1}^{-1}|M|^{-1}\sum_s|M_s|\left(M_s^{-1} - M^{-1}\right)X_s'D_sX_s\left(M_s^{-1} - M^{-1}\right).$$

Note that

$$\begin{aligned}\binom{n-1}{d-1}\binom{n-k}{d-1}^{-1} &\le [n/(r-k+2)]^{k-1} \\ &= \left[1 - (d+k-2)/n\right]^{-k+1} \le 1 + cdn^{-1}.\end{aligned} \tag{3.7}$$

The last inequality follows from (3.5) since it implies that $(d+k-2)/n$ is bounded away from 1. Hence,

$$S_1 = \left(1 + O(n^{-1}d)\right)\operatorname{Var}\hat{\beta} = \operatorname{Var}\hat{\beta} + O(n^{-2}d)$$

by (3.1). From (3.1) and the identity $\binom{n-k}{d}|M| = \sum_s|M_s|$ [Wu (1986), Lemma 1],

$$\begin{aligned}\operatorname{tr}(S_2) &\le c\binom{n-k}{d-1}^{-1}|M|^{-1}\sum_s(|M| - |M_s|)\operatorname{tr}\left(M^{-1}X_{\bar{s}}'X_{\bar{s}}M^{-1}\right) \\ &\le cn^{-1}\binom{n-k}{d-1}^{-1}|M|^{-1}\sum_s(|M| - |M_s|)\sum_{i\in\bar{s}}w_i \\ &\le cdh_nn^{-1}\binom{n-k}{d-1}^{-1}\left[\binom{n}{d} - \binom{n-k}{d}\right] = O\left(n^{-1}dh_n\right),\end{aligned}$$

since

$$\binom{n-k}{d-1}^{-1}\left[\binom{n}{d}-\binom{n-k}{d}\right] \le d^{-1}(r-k+1)\left[n^{k}(r-k+1)^{-k}-1\right]$$
$$\le cn^{-1}(r-k+1) = O(1).$$

The last inequality follows from (3.5). From $X_s' D_s X_s \le cM_s \le cM$,

$$S_3 \le c\binom{n-k}{d-1}^{-1} |M|^{-1} \sum_s |M_s|\left(M_s^{-1} - M^{-1}\right)M\left(M_s^{-1} - M^{-1}\right).$$

Using (2.3), $(1 - dh_n)^{-1} = O(1)$ [under (3.4)], and $\operatorname{tr}(AB) \le \operatorname{tr}(AC)$ for $A \ge 0$ and $B \ge C$,

$$\operatorname{tr}(S_3) \le cdh_n\binom{n-k}{d-1}^{-1} |M|^{-1} \sum_s |M_s| \operatorname{tr}\left(M^{-1}X_{\bar{s}}'X_{\bar{s}}M^{-1}\right).$$

Then by (3.1) and (3.7),

$$\operatorname{tr}(S_3) \le cn^{-1}dh_n\binom{n-k}{d-1}^{-1} \sum_s\left(\sum_{i\in\bar{s}} w_i\right)$$
$$= cn^{-1}dh_n\binom{n-k}{d-1}^{-1}\binom{n-1}{d-1}\sum_1^n w_i = O\left(n^{-1}dh_n\right).$$

From Lemma 1, $S_2 = O(n^{-1}dh_n)$ and $S_3 = O(n^{-1}dh_n)$ and the result follows. □

Theorem 1 can be extended to a more general resampling procedure called the variable jackknife [Wu (1986)], i.e., the number of deleted observations may vary for fixed n. Details are in Shao and Wu (1985). Another extension of Theorem 1 to the estimation of variance of a nonlinear function of $\hat{\beta}$ can be found in Shao (1986).

We end this section with the following theorem, which provides another way of looking at the robustness issue. In Theorem 1, we impose restrictions on the matrix X in order to obtain asymptotic results. In the next theorem, we impose instead restrictions on the error structure so that the σ_{in}^2 are nearly equal for large n. Recall that the errors e_{in} in the model (1.1) may vary with n. No restriction on d or h_n is made.

THEOREM 2. *Assume that* $\max_{i\le n}\sigma_{in}^2 - \min_{i\le n}\sigma_{in}^2 = o(1)$. *Then*

$$nE\left(v_{J(d)} - \operatorname{Var}\hat{\beta}\right) \to 0.$$

PROOF. Recall that $\operatorname{Var}\hat{\beta} = M^{-1}\sum_1^n \sigma_{in}^2 x_i x_i' M^{-1}$. It is easy to see that

$$\left(\max_{i\le n}\sigma_{in}^2\right)nM^{-1} - \left(\min_{i\le n}\sigma_{in}^2\right)nM^{-1} \to 0$$

and

$$\left(\max_{i\le n}\sigma_{in}^2\right)nM^{-1} - n\operatorname{Var}\hat{\beta} \to 0.$$

Noting that $v_{J(\mathrm{d})}$ is unbiased when $\sigma_{in}^2 = \sigma^2$ for all i and n, we have

$$\Big(\min_{i \le n} \sigma_{in}^2\Big) nM^{-1} \le nEv_{J(d)} \le \Big(\max_{i \le n} \sigma_{in}^2\Big) nM^{-1}.$$

Hence,

$$nEv_{J(d)} - n\Big(\max_{i \le n} \sigma_{in}^2\Big) M^{-1} \to 0 \quad \text{and} \quad nE\Big(v_{J(d)} - \operatorname{Var}\hat{\beta}\Big) \to 0. \qquad \square$$

In the special case of constant variances ($\sigma_{in}^2 = \sigma^2$), this is simply the unbiasedness of $v_{J(d)}$.

4. C-robustness of $v_{J(d)}$. The C-robustness of $v_{J(d)}$ will be proved in Theorem 3 for any d_n with $d_n h_n \to 0$. It is useful for constructing approximate confidence intervals for β. The result is extended in Theorem 4 to the estimation of variance of a nonlinear function of $\hat{\beta}$.

THEOREM 3. *Assume that*

$$d_n h_n \to 0 \quad \text{as} \quad n \to \infty, \tag{4.1}$$

$$\max_{i \le n} Ee_i^4 \le c < \infty \quad \text{for all } n. \tag{4.2}$$

Then

$$v_{J(d)} - \operatorname{Var}\hat{\beta} = o_P(n^{-1}).$$

PROOF. Let

$$V_s = M^{-1}(X_s'e_s - X'e),$$
$$U_s = \big(M_s^{-1} - M^{-1}\big)X_s'e_s,$$
$$c_s = n\binom{n-k}{d-1}^{-1} |M|^{-1}|M_s|,$$
$$V = \big(c_s^{1/2}V_s'\big)_s, U = \big(c_s^{1/2}U_s'\big)_s,$$

where $(\ )_s$ denotes a matrix with rows indexed by $s \in \mathbb{S}_r$. Then

$$nv_{J(d)} = U'U + V'V + U'V + V'U.$$

Since $EU'U = nS_3 = o(1)$ by Theorem 1 and (4.1), $U'U = o_P(1)$ holds according to Lemma 2(ii). By the Cauchy–Schwarz inequality, the result follows if

$$V'V - n\operatorname{Var}\hat{\beta} = o_P(1). \tag{4.3}$$

Let $\bar{s}$ be the complement of s. Then $X'e - X_s'e_s = X_{\bar{s}}'e_{\bar{s}}$ and

$$V'V = n\binom{n-k}{d-1}^{-1} |M|^{-1} \sum_s |M_s| M^{-1}X_{\bar{s}}'e_{\bar{s}}e_{\bar{s}}'X_{\bar{s}}M^{-1}.$$

Define

$$F_n = n\binom{n-k}{d-1}^{-1} \sum_s M^{-1}(X_{\bar{s}}'e_{\bar{s}}e_{\bar{s}}'X_{\bar{s}})M^{-1}.$$

Since $|M_s| \le |M|$, $F_n - V'V \ge 0$ and its expectation is nS_2, where S_2 was defined in the proof of Theorem 1 and $\mathrm{tr}(S_2)$ was shown to be $O(d_n h_n n^{-1})$. Therefore, from (4.1) and Lemma 2(ii),

$$F_n - V'V = o_P(1). \tag{4.4}$$

Decompose

$$F_n = n\binom{n-k}{d-1}^{-1} \sum_s M^{-1} \sum_{i,j \in \bar{s}} (e_i e_j x_i x_j') M^{-1} = G_n + H_n,$$

where

$$G_n = n\binom{n-k}{d-1}^{-1} \sum_s M^{-1} \sum_{i \ne j \in \bar{s}} (e_i e_j x_i x_j') M^{-1} = q_n M^{-1} \sum_{i<j}^{n} e_i e_j x_i x_j' M^{-1}$$

with $q_n = 2n\binom{n-2}{d-2}\binom{n-k}{d-1}^{-1}$, $G_n = 0$ when $d_n = 1$ and

$$H_n = n\binom{n-k}{d-1}^{-1} \sum_s M^{-1} \sum_{i \in \bar{s}} (e_i^2 x_i x_i') M^{-1} = p_n M^{-1} \sum_1^n e_i^2 x_i x_i' M^{-1}$$

with $p_n = n\binom{n-1}{d-1}\binom{n-k}{d-1}^{-1}$. Then (4.3) follows from (4.4), $G_n = o_P(1)$ and $H_n - n \operatorname{Var}\hat{\beta} = o_P(1)$.

For $d_n \ge 2$, from (2.1), the (p,q)th element of G_n is $q_n(\sum_{i<j}^n e_i e_j m_{pq}^{ij})$. Since $\operatorname{Cov}(e_i e_j, e_l e_r) = 0$ if $i \ne l$ or $j \ne r$, its variance is $q_n^2[\sum_{i<j}^n \sigma_i^2 \sigma_j^2 (m_{pq}^{ij})^2] \to 0$ by Lemma 3 and $n^{-1}q_n \to 0$. Since $EG_n = 0$, this implies $G_n = o_P(1)$. Since $n^{-1}p_n \to 1$, $EH_n - n\operatorname{Var}\hat{\beta} = (p_n - n)\operatorname{Var}\hat{\beta} \to 0$. The (p,q)th element of H_n is $p_n(\sum_{i=1}^n e_i^2 m_{pq}^{ii})$. Its variance $p_n^2[\sum_{i=1}^n \operatorname{Var}(e_i^2)(m_{pq}^{ii})^2]$ converges to zero by (4.2) and Lemma 3. Thus, $H_n - n\operatorname{Var}\hat{\beta} = o_P(1)$ and (4.3) holds. □

Let $\theta = g(\beta)$ be a nonlinear function of β from $\mathbb{R}^k$ to $\mathbb{R}^m$ with first order derivatives. We now consider the problem of estimating the variance of the estimator $\hat{\theta} = g(\hat{\beta})$. The jackknife variance estimator $v_{J(d)}$ in (1.4) can be extended as follows [Wu (1986)]:

$$v_{J(d)}(\hat{\theta}) = \left(\binom{n-k}{d-1}|M|\right)^{-1} \sum_s |M_s| (\hat{\theta}_s - \hat{\theta})(\hat{\theta}_s - \hat{\theta})',$$

where $\hat{\theta}_s = g(\hat{\beta}_s)$. We want to establish the C-robustness of $v_{J(d)}(\hat{\theta})$, i.e.,

$$v_{J(d)}(\hat{\theta}) - \nabla g(\beta) \operatorname{Var}\hat{\beta} (\nabla g(\beta))' = o_P(n^{-1}), \tag{4.5}$$

where $\nabla g(\beta)$ is the $m \times k$ matrix whose jth row is the gradient of the jth component of g at β and $\nabla g(\beta)\operatorname{Var}\hat{\beta}(\nabla g(\beta))'$ is the asymptotic variance of $\hat{\theta}$.

To prove (4.5), we first prove a useful result that gives an upper bound on the mean of the maximum Euclidean distance between $\hat{\beta}_s$ and $\hat{\beta}$, with s ranging over $\mathbb{S}_r$. It shows that $\hat{\beta}_s - \hat{\beta}$ are close to zero uniformly over all $s \in \mathbb{S}_r$ if $d_n h_n \to 0$. A special case was given in Miller (1974) for $d_n = 1$.

PROPOSITION 1. *Let $r = n - d_n$. Suppose that* (3.4) *holds. Then*

$$E\left(\max_{s \in \mathbb{S}_r} \|\hat{\beta}_s - \hat{\beta}\|^2\right) = O(d_n h_n), \tag{4.6}$$

where $\|\ \|$ *is the Euclidean norm. In addition, if* $d_n h_n \to 0$, *then for any* $\delta > 0$,

$$\lim_{n\to\infty} \mathbb{P}\left(\|\hat\beta - \beta\| < \delta,\ \|\hat\beta_s - \hat\beta\| < \delta \text{ for all } s \in \mathbb{S}_r\right) = 1. \tag{4.7}$$

PROOF. For simplicity, we drop the subscript n in d_n. The main step is in decomposing $\hat\beta - \hat\beta_s$ into the sum of d successive differences and finding a tight upper bound for each difference. Let $\bar s$ be the complement of $s \in \mathbb{S}_r$, denoted by $\bar s = \{j_1, \dots, j_d\}$, and $s_i = \{j_1, \dots, j_i\} \cup s$, $i = 1, \dots, d$, $\hat\beta_{s_0} = \hat\beta_s$ 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

$$\hat\beta - \hat\beta_s = \sum_{i=1}^{d} \left(\hat\beta_{s_i} - \hat\beta_{s_{i-1}}\right) = \sum_{i=1}^{d} \left(1 - x_{j_i}' M_{s_i}^{-1} x_{j_i}\right)^{-1} M_{s_i}^{-1} r_{j_i} x_{j_i}, \tag{4.8}$$

where $r_{j_i} = y_{j_i} - x_{j_i}'\hat\beta_{s_i}$ is the j_ith residual from fitting the subset model $y_{s_i} = X_{s_i}\beta + e_{s_i}$. Let $\mu_i = d - i$ be the number of elements outside s_i. Using (2.4),

$$\left(1 - x_{j_i}' M_{s_i}^{-1} x_{j_i}\right)^{-1} \le \left[1 - (1 - \mu_i h_n)^{-1} h_n\right]^{-1} \le (1 - d h_n)^{-1}, \tag{4.9}$$

where the last inequality follows from $(d-1)/(d-i) \ge 1 > d h_n$. Then

$$\begin{aligned}
\|\hat\beta_s - \hat\beta\|^2 &\le d \sum_{i=1}^{d} \|\hat\beta_{s_{i-1}} - \hat\beta_{s_i}\|^2 \\
&\le d \sum_{i=1}^{d} (1 - d h_n)^{-2} r_{j_i}^2 x_{j_i}' M_{s_i}^{-2} x_{j_i} \quad [\text{using (4.8) and (4.9)}] \\
&\le c d n^{-1} \sum_{i=1}^{d} r_{j_i}^2 x_{j_i}' M_{s_i}^{-1} x_{j_i} \quad [\text{using (2.3), (3.1) and (3.4)}] \\
&\le c d n^{-1} h_n \sum_{i=1}^{d} r_{j_i}^2 \quad [\text{using (2.4) and (3.4)}].
\end{aligned}$$

Note that $r_{j_i} = e_{j_i} - \sum_{p \in s_i} x_{j_i}' M_{s_i}^{-1} x_p e_p$. Thus,

$$r_{j_i}^2 \le 2 e_{j_i}^2 + 2\left(\sum_{p \in s_i} x_{j_i}' M_{s_i}^{-1} x_p e_p\right)^2 \le 2 e_{j_i}^2 + 2n \sum_{p \in s_i} \left(x_{j_i}' M_{s_i}^{-1} x_p\right)^2 e_p^2.$$

Hence, for any $s \in \mathbb{S}_r$,

$$\begin{aligned}
\sum_{i=1}^{d} r_{j_i}^2 &\le c\left[\sum_{i=1}^{d} e_{j_i}^2 + n \sum_{i=1}^{d} \sum_{p \in s_i} \left(x_p' M_{s_i}^{-1} x_p\right)\left(x_{j_i}' M_{s_i}^{-1} x_{j_i}\right) e_p^2\right] \\
&\le c\left(\sum_1^n e_i^2 + n \sum_{i \in \bar s} \sum_{p=1}^{n} w_i w_p e_p^2\right).
\end{aligned}$$

The last inequality follows from (2.4) and (3.4). Thus,

$$\max_{s \in \mathbb{S}_r} \|\hat\beta_s - \hat\beta\|^2 \le c d h_n \left(n^{-1} \sum_1^n e_i^2 + d h_n \sum_1^n w_i e_i^2\right)$$

and (4.6) results since $E(n^{-1}\Sigma_1^n e_i^2)$ and $E(\Sigma_1^n w_i e_i^2)$ are bounded (which follows from the boundedness of σ_i^2).

(4.7) is an immediate consequence of (4.6), $d_n h_n \to 0$ and $\hat{\beta} - \beta \to 0$ in probability. □

THEOREM 4. *If* (4.1) *and* (4.2) *hold and* ∇g *is continuous in a neighborhood of* β, *then* (4.5) *holds.*

PROOF. Express $v_{J(d)}(\hat{\theta})$ as

$$\binom{n-k}{d-1}^{-1} |M|^{-1} \sum_s |M_s| \left[G(\zeta_s)(\hat{\beta}_s - \hat{\beta})\right]\left[G(\zeta_s)(\hat{\beta}_s - \hat{\beta})\right]',$$

where $G(\zeta_s)$ is an $m \times k$ matrix whose jth row is the gradient of the jth component of g at $\zeta_s^{(j)}$, which is a point on the line segment between $\hat{\beta}_s$ and $\hat{\beta}$. Let $V_s = \nabla g(\hat{\beta})(\hat{\beta}_s - \hat{\beta})$, $U_s = [G(\zeta_s) - \nabla g(\hat{\beta})](\hat{\beta}_s - \hat{\beta})$, $c_s = \left[n\binom{n-k}{d-1}^{-1}|M|^{-1}|M_s|\right]^{1/2}$, $V = (c_s V_s')_s$ and $U = (c_s U_s')_s$. Then

$$n v_{J(d)}(\hat{\theta}) = V'V + U'U + U'V + V'U.$$

Since $V'V = n\nabla g(\hat{\beta}) v_{J(d)} (\nabla g(\hat{\beta}))'$, $V'V - n\nabla g(\beta)\mathrm{Var}\,\hat{\beta}(\nabla g(\beta))' = o_P(1)$ follows from Theorem 3 and the continuity of ∇g at β. From the Cauchy–Schwarz inequality, it suffices to show

$$U'U = o_P(1). \tag{4.10}$$

For any $\varepsilon > 0$, there is a $\delta > 0$ such that $\|\nabla g(x) - \nabla g(y)\| < \varepsilon$ for $\|x - \beta\| + \|y - \beta\| \le \delta$. Let

$$A_n^\delta = \left\{\|\hat{\beta} - \beta\| < \delta/3,\ \|\hat{\beta}_s - \hat{\beta}\| < \delta/3 \text{ for all } s \in \mathbb{S}_r\right\}.$$

Then

$$\begin{aligned}\mathrm{tr}\left(I_{A_n^\delta} U'U\right) &= n\binom{n-k}{d-1}^{-1} |M|^{-1} \sum_s |M_s| \left\|\left[G(\zeta_s) - \nabla g(\hat{\beta})\right](\hat{\beta}_s - \hat{\beta})\right\|^2 I_{A_n^\delta} \\ &\le \varepsilon^2 n\, \mathrm{tr}\left(v_{J(d)}\right),\end{aligned}$$

where I_A is the indicator function of A. Hence, for any given ε_0,

$$\mathbb{P}\left(\mathrm{tr}(U'U) > \varepsilon_0\right) \le 1 - \mathbb{P}\left(A_n^\delta\right) + \mathbb{P}\left(n\, \mathrm{tr}\left(v_{J(d)}\right) > \varepsilon_0/\varepsilon^2\right).$$

From (4.7),

$$0 \le \limsup_{n\to\infty} \mathbb{P}\left(\mathrm{tr}(U'U) > \varepsilon_0\right) \le \limsup_{n\to\infty} \mathbb{P}\left(n\, \mathrm{tr}\left(v_{J(d)}\right) > \varepsilon_0/\varepsilon^2\right),$$

which can be made arbitrarily small from Theorem 3 and by choosing ε small. This proves (4.10). □

5. Robustness of other jackknife variance estimators. We now consider the robustness of the unweighted jackknife variance estimator v_J (1.2) and

Hinkley's weighted jackknife variance estimator v_H (1.3). They can be rewritten as

$$v_J = n^{-1}(n-1)M^{-1}\sum_1^n (1-w_i)^{-2} r_i^2 x_i x_i' M^{-1} - R, \tag{5.1}$$

where

$$R = n^{-2}(n-1)M^{-1}\left[\sum_1^n (1-w_i)^{-1} r_i x_i\right]\left[\sum_1^n (1-w_i)^{-1} r_i x_i\right]' M^{-1}$$

and

$$v_H = n(n-k)^{-1}M^{-1}\sum_1^n r_i^2 x_i x_i' M^{-1}.$$

The following theorem proves that $h_n \to 0$ is necessary and sufficient for v_J and v_H to be AU- and C-robust. The condition is rather weak since it is known to be necessary and sufficient for the asymptotic normality of the LSE $\hat{\beta}$ in the case of homoscedastic errors [Huber (1981)]. On the other hand, it is sufficient but not necessary for $v_{J(1)}$ to be robust in view of Theorem 2, which does not require any condition on h_n.

THEOREM 5. *Under*

$$\sup_n h_n < 1 \tag{5.2}$$

and

$$M = O(n), \tag{5.3}$$

the following statements are equivalent:

(a) $h_n \to 0$.
(b) v_J *is AU-robust.*
(c) v_J *is C-robust* [*under* (4.2)].
(d) v_H *is AU-robust.*
(e) v_H *is C-robust* [*under* (4.2)].

REMARK. The proof of the theorem shows that (a) is necessary for the weaker property that v_J and v_H are asymptotically unbiased and consistent under $\text{Var}(e_i) = \sigma^2$.

For the proof of Theorem 5, we need the following results.

LEMMA 5. *Let R be given in* (5.1). *If* (5.2) *holds, then* $ER = O(n^{-2})$.

A proof of this lemma is given in Shao and Wu (1985). The next result establishes relationships among v_J (1.2), v_H (1.3) and $v_{J(1)}$ (1.5).

PROPOSITION 2. *Under* (5.2), *we have* (i) $v_J - v_{J(1)} = W_1 + O_P(n^{-2})$ *and* $E(v_J - v_{J(1)}) = EW_1 + O(n^{-2})$ *with* $W_1 \geq 0$.

(ii) $v_{J(1)} - v_H = W_2 + O_p(n^{-2})$ *and* $E(v_{J(1)} - v_H) = EW_2 + O(n^{-2})$ *with* $W_2 \geq 0$.

(iii) $\operatorname{tr}(EW_i) \leq cn^{-1}g_n \leq ckn^{-1}h_n$, $i = 1, 2$, *where* h_n *and* g_n *are defined in* (2.2) *and* (3.2).

(iv) $\operatorname{tr}(EW_i) \geq cn^{-1}g_n$, $i = 1, 2$, *if in addition* (5.3) *holds and* $\operatorname{Var}(e) = \sigma^2 I$.

PROOF. (i) The delete-1 jackknife estimator (1.5) can be rewritten as

$$v_{J(1)} = M^{-1}\sum_1^n (1 - w_i)^{-1} r_i^2 x_i x_i' M^{-1}.$$

Write $v_J - v_{J(1)} = W_1 - G - R$, where R is given in (5.1),

$$\begin{aligned} W_1 &= M^{-1}\sum_1^n (1 - w_i)^{-2} r_i^2 x_i x_i' M^{-1} - v_{J(1)} \\ &= M^{-1}\sum_1^n w_i (1 - w_i)^{-2} r_i^2 x_i x_i' M^{-1} \geq 0, \qquad (5.4) \\ G &= n^{-1} M^{-1}\sum_1^n (1 - w_i)^{-2} r_i^2 x_i x_i' M^{-1} \geq 0. \end{aligned}$$

By Lemma 5, $ER = O(n^{-2})$ and, therefore, $R = O_p(n^{-2})$ since $R \geq 0$. We complete the proof by showing $EG = O(n^{-2})$, which also implies $G = O_p(n^{-2})$. From Lemma 2.1 of Shao (1986), $Er_i^2 \leq c$ for all i. Hence, $EG \leq cn^{-1}M^{-1} = O(n^{-2})$.

(ii) It is easy to show that $v_{J(1)} - v_H = W_2 - k(n-k)^{-1}v_{J(1)}$, where

$$W_2 = n(n-k)^{-1}M^{-1}\sum_1^n w_i(1 - w_i)^{-1} r_i^2 x_i x_i' M^{-1} \geq 0.$$

The result follows since $Ev_{J(1)} = O(n^{-1})$ and $v_{J(1)} \geq 0$.

(iii) From (3.3) and $Er_i^2 \leq c$,

$$\operatorname{tr}(EW_i) \leq c \operatorname{tr}\left(M^{-1}\sum_1^n w_i x_i x_i' M^{-1}\right) \leq cn^{-1}\sum_1^n w_i^2 = cn^{-1}g_n \leq ckn^{-1}h_n.$$

(iv) Under (5.3) and $\operatorname{Var}(e) = \sigma^2 I$,

$$\operatorname{tr}(EW_i) \geq \sigma^2 \sum_1^n w_i x_i' M^{-2} x_i \geq cn^{-1}\sum_1^n w_i^2. \qquad \square$$

PROOF OF THEOREM 5. We prove that (a), (b) and (c) are equivalent. The proof of the equivalence of (a), (d) and (e) is similar.

From Theorems 1 and 3 and Proposition 2, (a) implies (b) and (c). To show that (b) implies (a), it suffices to show that, in the special case $\operatorname{Var}(e) = \sigma^2 I$,

$$nE\left(v_J - \operatorname{Var}\hat{\beta}\right) \to 0 \qquad (5.5)$$

implies (a). By Proposition 2(i) and $Ev_{J(1)} = \operatorname{Var}\hat{\beta}$ under $\operatorname{Var}(e) = \sigma^2 I$, (5.5) holds iff $nEW_1 \to 0$, where W_1 is defined in (5.4). From Proposition 2(iv), $n\,\operatorname{tr}(EW_1) \geq cg_n \geq ch_n^2$. Therefore, $nEW_1 \to 0$ iff (a) holds. This proves (b) implies (a).

It remains to show that (c) implies (a). Consider again the special case $\operatorname{Var}(e) = \sigma^2 I$ and assume (4.2) holds. Let $U = M^{-1}\Sigma_1^n(1 - w_i)^{-2} r_i^2 x_i x_i' M^{-1}$ and u_{pq} and V_{pq} be the (p, q)th elements of U and $\operatorname{Var}\hat{\beta}$, respectively. From (5.1) and Lemma 5, we have $v_J = U + O_p(n^{-2})$. Hence, (c) implies

$$n\big(U - \operatorname{Var}\hat{\beta}\big) = o_p(1). \tag{5.6}$$

The assumption (5.2) implies that $\{(1 - w_i)^{-1}\}$ are uniformly bounded. Therefore, $0 \leq U \leq cv_{J(1)}$. Thus, $n^2E(u_{pq}^2)$ and $n^2V_{pq}^2$ are uniformly bounded from Theorem 3.2 of Shao (1986). Hence, $\{n|u_{pq} - V_{pq}|\colon n = 1, 2, \dots\}$ are uniformly integrable. From this and (5.6),

$$nEW_1 = n\big(EU - Ev_{J(1)}\big) = nE\big(U - \operatorname{Var}\hat{\beta}\big) = o(1),$$

which implies (5.5). Therefore, (c) implies (a). □

In view of Proposition 2, the three delete-1 jackknife variance estimators are indistinguishable up to the order $O(n^{-2})$ if g_n is of the order $O(n^{-1})$. However, the comparison is more favorable for $v_{J(1)}$ in general since Theorem 5 shows that when $\lim_{n\to\infty} h_n \neq 0$, v_J and v_H are not asymptotically unbiased or consistent (even for homoscedastic errors), whereas $v_{J(1)}$ is asymptotically unbiased for nearly homoscedastic errors (Theorem 2).

6. Comparison of the biases of v_J, v_H and $v_{J(1)}$. The AU-robustness property for an estimator of $\operatorname{Var}\hat{\beta}$ is desirable for the bias reduction of $g(\hat{\beta})$, where g is a smooth nonlinear function of β, since the latter is closely connected with the existence of an asymptotically unbiased variance estimator of $\operatorname{Var}\hat{\beta}$ [Wu (1986), Section 9 and Shao (1986), Sections 4 and 5]. Unlike $v_{J(1)}$, v_J and v_H are neither AU- nor C-robust if $\lim_{n\to\infty} h_n \neq 0$. Further analysis in this section shows that v_J and v_H are upward and downward biased, respectively, up to the order $n^{-1}g_n$. This supports the empirical results of Wu (1986). Also, the order of the bias of $v_{J(1)}$ is always no larger than that of the bias of v_J or v_H.

Let B_J, B_H and $B_{J(1)}$ be the biases of v_J, v_H and $v_{J(1)}$, respectively.

THEOREM 6. *Assume that* (5.2) *and* (5.3) *hold.*

(i) *Suppose that* $\min_{i\leq n}\sigma_{in}^2 \geq \alpha > 0$. *Then*

$$B_J = U_1 + O(n^{-2}) \quad \textit{with } U_1 > 0 \textit{ and } \operatorname{tr}(U_1) \geq c_1 n^{-1} g_n,$$

where c_1 *is a positive constant and* g_n *is defined in* (3.2).

(ii) *Suppose that* $2\min_{i\leq n}\sigma_{in}^2 - \max_{i\leq n}\sigma_{in}^2 \geq \delta > 0$ *for all* n. *Then*

$$B_H = -U_2 + O(n^{-2}) \quad \textit{with } U_2 > 0 \textit{ and } \operatorname{tr}(U_2) \geq c_2 n^{-1} g_n,$$

where c_2 *is a positive constant.*

(iii) *The elements of* $B_{J(1)}$ *are bounded in absolute value by* $c_3 n^{-1} g_n$ *for a positive constant* c_3.

PROOF. (i) Let $V = E[M^{-1}\Sigma_1^n (1-w_i)^{-2} r_i^2 x_i x_i' M^{-1}]$. Then from Lemma 5,

$$Ev_J = V + O(n^{-2}). \tag{6.1}$$

Since $Er_i^2 = (1-w_i)^2\sigma_{in}^2 + \Sigma_{j\neq i}^n w_{ij}^2\sigma_{jn}^2$, we have

$$\begin{aligned} V - \operatorname{Var}\hat{\beta} &= M^{-1}\sum_1^n \left[(1-w_i)^{-2} Er_i^2 - \sigma_{in}^2\right] x_i x_i' M^{-1} \\ &= M^{-1}\sum_{i=1}^n \sum_{j\neq i}^n w_{ij}^2 \sigma_{jn}^2 x_i x_i' M^{-1} \\ &\geq \alpha M^{-1}\sum_1^n w_i(1-w_i) x_i x_i' M^{-1} > 0. \end{aligned}$$

From (5.2) and (5.3),

$$\operatorname{tr}\left(V - \operatorname{Var}\hat{\beta}\right) \geq c\sum_1^n w_i x_i' M^{-2} x_i \geq c_1 n^{-1} g_n.$$

Hence, (i) follows from (6.1) with $U_1 = V - \operatorname{Var}\hat{\beta}$.

(ii) From $v_H = M^{-1}\Sigma_1^n r_i^2 x_i x_i' M^{-1} + k(n-k)^{-1} M^{-1}\Sigma_1^n r_i^2 x_i x_i' M^{-1}$,

$$Ev_H = W + O(n^{-2}) \tag{6.2}$$

by the proof of Proposition 2, where $W = M^{-1}\Sigma_1^n Er_i^2 x_i x_i' M^{-1}$. From $\Sigma_{j=1}^n w_{ij}^2 = w_i$,

$$\sigma_{in}^2 - Er_i^2 = 2w_i\sigma_{in}^2 - \Sigma_{j=1}^n w_{ij}^2\sigma_{jn}^2 \geq w_i\left(2\min_{i\leq n}\sigma_{in}^2 - \max_{i\leq n}\sigma_{in}^2\right) \geq \delta w_i$$

by the assumption on σ_{in}^2. Then

$$\operatorname{Var}\hat{\beta} - W = M^{-1}\sum_1^n \left(\sigma_{in}^2 - Er_i^2\right) x_i x_i' M^{-1} \geq \delta M^{-1}\sum_1^n w_i x_i x_i' M^{-1} > 0$$

and $\operatorname{tr}(\operatorname{Var}\hat{\beta} - W) \geq c_2 n^{-1} g_n$ by (5.3). Hence, (ii) follows from (6.2) with $U_2 = \operatorname{Var}\hat{\beta} - W$.

(iii) Since $B_{J(1)} = M^{-1}\Sigma_1^n (1-w_i)^{-1}\tau_i x_i x_i' M^{-1}$, where $\tau_i = \Sigma_{j=1}^n w_{ij}^2(\sigma_{jn}^2 - \sigma_{in}^2)$, we have $-\Lambda_n \leq B_{J(1)} \leq \Lambda_n$ with $\Lambda_n = M^{-1}\Sigma_1^n (1-w_i)^{-1}|\tau_i| x_i x_i' M^{-1} \geq 0$. Since σ_{in}^2 are uniformly bounded, $|\tau_i| \leq cw_i$ and

$$\operatorname{tr}(\Lambda_n) \leq c\operatorname{tr}\left(M^{-1}\sum_1^n w_i x_i x_i' M^{-1}\right) \leq c_3 n^{-1}\sum_1^n w_i^2 = c_3 n^{-1} g_n,$$

which together with Lemma 2(i) implies the result. □

The theorem shows that the bias of $v_{J(1)}$ always converges to zero at a rate no slower than that of v_J and v_H. Example 1 considers an extreme case in which n times the biases of v_J and v_H do not converge to zero whereas $v_{J(1)}$ is unbiased. The theorem also implies that $Ev_H < \operatorname{Var}\hat{\beta} < Ev_J$ up to the order $n^{-1}g_n$, which may be larger than n^{-2} if g_n is of a larger order than n^{-1}. An illustration is given in Example 2.

EXAMPLE 1. Let $k = 2$, n even, $x_{2j-1,n} = (n^{1/2}3^{-j/2}, 0)'$, $x_{2j,n} = (0, n^{1/2}3^{-j/2})'$, $\sigma^2_{2j-1,n} = \sigma^2_1$ and $\sigma^2_{2j,n} = \sigma^2_2$, $j = 1, \dots, n/2$. Then

$$M = \left(n \sum_{j=1}^{n/2} 3^{-j}\right) I = O(n),$$

$M^{-1} = O(n^{-1})$ and $h_n = 3^{-1}(\sum_{j=1}^{n/2} 3^{-j})^{-1} \to \frac{2}{3}$. Hence by Theorem 5, v_J and v_H are neither asymptotically unbiased nor consistent even if $\sigma^2_1 = \sigma^2_2$. On the other hand, since $x'_{2j-1,n} M^{-1} x_{2k,n} = 0$ for any j, k, $v_{J(1)}$ is unbiased according to Theorem 5 of Wu (1986).

Miller (1974) proved that $n^{-1}M$ converging to a positive definite matrix implies $h_n \to 0$. This is not applicable here since he assumes that x_i, $i = 1, 2, \dots,$ is a sequence, while $x_{i,n}$, $i = 1, \dots, n$, in Example 1 vary with n.

EXAMPLE 2. Let $k = 2$ and $x_{in} = (1, a_{in})'$, where a_{1n} equals $n^{5/12}$ for odd n and $1 + n^{5/12}$ for even n, and a_{in} equals 1 for odd n and -1 for even n, $i = 2, \dots, n$. A straightforward calculation shows that $n^{-1}M \to I$, $w_i = \zeta_{in}/\xi_n$, where $\xi_n = n[n^{5/6} + (n-1)] - n^{5/6}$ and $\zeta_{in} = n^{5/6} + (n-1) - 2a_{in}n^{5/12} + a^2_{in}n$ and $\max_{i \le n} \zeta_{in} = n^{5/6} + (n-1) - 2n^{5/6} + n^{11/5}$. Hence, $h_n = \max_{i \le n} w_i$ is of the order $n^{-1/6}$, which is substantially larger than n^{-1}. Also, $\sum_{i=1}^n \zeta^2_{in}$ is of the order $n^{11/3}$. Hence, $g_n = \sum_1^n w_i^2$ is of the order $n^{-1/3}$, which is larger than n^{-1}.

REFERENCES

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. (1981). *Robust Statistics*. Wiley, New York.

MILLER, R. G. (1974). An unbalanced jackknife. *Ann. Statist.* **2** 880–891.

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. (1987). Sampling and resampling: An application of sampling method to jackknife variance estimation. Technical Report 799, Dept. Statistics, Univ. Wisconsin-Madison.

SHAO, J. and WU, C. F. J. (1985). Heteroscedasticity-robustness of jackknife variance estimators in linear models. Technical Report 778, Dept. Statistics, Univ. Wisconsin-Madison.

TUKEY, J. (1958). Bias and confidence in not quite large samples (abstract). *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.

WU, C. F. J. (1987). On the asymptotic properties of the jackknife histograms. Technical Report, Dept. Statistics, Univ. Wisconsin-Madison.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN-MADISON
1210 WEST DAYTON STREET
MADISON, WISCONSIN 53706