JACKKNIFE AND WEIGHTED JACKKNIFE ESTIMATES OF THE VARIANCE OF M-ESTIMATORS IN LINEAR REGRESSION

by

D. Kushary
Rutgers University,
Camden

and

Arup Bose
Indian Statistical Institute
and
Purdue University

Technical Report #96-12

Department of Statistics
Purdue University
West Lafayette, IN USA

April 1996

*

JACKKNIFE AND WEIGHTED JACKKNIFE ESTIMATES OF THE VARIANCE OF M-ESTIMATORS IN LINEAR REGRESSION

by

D. Kushary and Arup Bose

Rutgers University, Camden

Indian Statistical Institute and Purdue University

Abstract

We consider the problem of estimating the mean squared error of M-estimates of the regression coefficient. We establish in probability expansions of jackknife and weighted jackknife estimators and use these expansions to compare the two estimators in terms of efficiency and robustness.

Key Words and Phrases: Regression, M-estimates, jackknife, weighted jackknife, efficiency, robustness

AMS 1991 Subject Classification Primary 62G09 Secondary 62J05, 62G20

1. INTRODUCTION.

Consider the linear regression model

$$y_j = \beta x_j + \epsilon_j, \quad j = 1, \dots, n \tag{1.1}$$

where the regressors $(x_j, j \geq 1)$ are nonrandom. The errors $(\epsilon_j, j \geq 1)$ are assumed to be independent. Let ψ be a real valued function. Then an M-estimate of β corresponding to ψ is defined as the value β_n of β which solves

$$\sum_{j=1}^{n} x_j \, \psi(y_j - \beta x_j) = 0 \; \dots \; \dots \tag{1.2}$$

Under appropriate conditions on ψ, $(\epsilon_j, j \geq 1)$ and $(x_j, j \geq 1)$, this estimate is asymptotic normal.

If $\psi(x) = x$, then β_n is the least squares (LS) estimator, given by

$$\beta_{n,LS} = \sum_{j=1}^{n} x_j \, y_j / \sum_{j=1}^{n} x_j^2 \; \dots \; \dots \tag{1.3}$$

Note that

$$V(\beta_{n,LS}) = \sum_{j=1}^{n} x_j^2 \, \sigma_j^2 / \left(\sum_{j=1}^{n} x_j^2 \right)^2 \; \dots \; \dots \tag{1.4}$$

where $\sigma_j^2 = V(\epsilon_j), j \geq 1$.

Several estimators of this variance were compared by Liu and Singh (1992). To explain their results, let

$$L_n = \sum_{j=1}^{n} x_j^2, \; v_n = \frac{\sum_{j=1}^{n} x_j^2 \, \sigma_j^2}{L_n^2}, \; v_n^* = \frac{\sum_{j=1}^{n} \sigma_j^2}{n L_n}$$

They classified their estimators as being either efficient or robust. Let $V_{n,E}$ and $V_{n,R}$ denote respectively any efficient and robust estimator. Then under certain conditions,

$$n(V_{n,E} - v_n^*) = L_n^{-1} \sum_{j=1}^{n} (\epsilon_j^2 - \sigma_j^2) + O_p(n^{-1}) \; \dots \; \dots \tag{1.5}$$

$$n(V_{n,R} - v_n) = n L_n^{-2} \sum_{j=1}^{n} x_j^2 (\epsilon_j^2 - \sigma_j^2) + O_p(n^{-1}) \dots \; \dots \; \dots \tag{1.6}$$

The bootstrap and the weighted jackknife estimators were shown to be of type $V_{n,E}$, satisfying (1.5). The jackknife, paired bootstrap, external bootstrap and the weighted bootstrap estimators were shown to be of type $V_{n,R}$ satisfying (1.6). Babu (1992) showed that the half sample estimator also belongs to the robust class and satisfies (1.6).

Note that $V(\beta_n) = v_n$. When $(\epsilon_j, j \geq 1)$ are heterogenous (that is, σ_j^2 are different) typically $v_n \neq v_n^*$. Thus from (1.5), the efficient estimators are inconsistent under heterogeneity. From (1.6), the robust estimators typically remain consistent under heterogeneity.

On the other hand, from (1.5) and (1.6), the ratio of the asymptotic variances of $V_{n,E}$ and $V_{n,R}$ is given by

$$1 + n^{-1} \sum_{j=1}^{n} \frac{(x_j^2 - n^{-1}L_n)^2}{(L_n/n)^2} \quad \dots \dots \qquad (1.6)$$

So typically, the robust estimators have larger variance than the efficient estimators. There appears to be a trade off and one cannot be both efficient and robust.

Liu and Singh (1992) and Babu (1992) provide a clear picture of the situation for the LS estimator. It does not, however, provide any clues to what might happen for general M-estimators. In this paper we address this issue but for simplicity restrict our attention to two estimators jackknife and weighted jackknife, one each from the two classes.

Under suitable conditions, we derive expansions similar to (1.5) and (1.6) for these two estimators. In general, these expansions involve extra terms, coming from the derivatives of ψ. Hence the issue of categorization by efficiency and robustness is not so straightforward. We demonstrate that in contrast to the least squares case, the jackknife may be more efficient than the weighted jackknife while remaining robust.

2. THE MAIN RESULTS. We assume $E\psi(\epsilon_j) = 0$ for all $j \geq 1$. Recall that the estimator β_n solves

$$\sum_{j=1}^{n} x_j \ \psi(y_j - \beta x_j) = 0 \quad \dots \dots \dots \qquad (2.1)$$

Using the delta method, it is seen that the asymptotic variance of β_n is given by

$$v_n = \frac{\sum_{j=1}^{n} x_j^2 \ E\ \psi^2(\epsilon_j)}{\left(\sum_{j=1}^{n} x_j^2 \ E\ \psi'(\epsilon_j)\right)^2} \quad \dots \dots \qquad (2.2)$$

The values $\beta_{(i)}$, $1 \le i \le n$ are obtained by solving

$$\sum_{\substack{j=1 \\ j \ne i}}^{n} x_j \ \psi(y_j - \beta x_j) = 0, \ 1 \le i \le n \ \ldots \ \ldots \tag{2.3}$$

The pseudovalues of β are then defined as

$$J_j = n\beta_n - (n-1)\beta_{(j)} \ \ldots \ \ldots \tag{2.4}$$

Exactly as in the i.i.d. case, the jackknife estimator of v_n is defined as

$$V_{n,J} = (n(n-1))^{-1} \sum_{j=1}^{n} (J_j - \beta_n)^2 \ \ldots \ \ldots \ \ldots \tag{2.5}$$

$$= \left(\frac{n-1}{n}\right) \sum_{j=1}^{n} (\beta_{(j)} - \beta_n)^2 \ \ldots \ \ldots \ \ldots \tag{2.6}$$

The weighted jackknife estimator of v_n is defined as

$$V_{n,WJ} = \frac{L_n}{n^2(n-1)} \sum_{j=1}^{n} x_j^{-2} \ (J_j - \beta_n)^2 \ \ldots \ \ldots \ \ldots \tag{2.7}$$

$$= \left(\frac{n-1}{n}\right) \frac{L_n}{n} \sum_{j=1}^{n} x_j^{-2} (\beta_{(j)} - \beta_n)^2 \ \ldots \ \ldots \ \ldots \tag{2.8}$$

To establish the probability expansions, we need to impose conditions on $(x_j, j \ge 1)$, $(\epsilon_j, j \ge 1)$ and ψ. For the sake of clarity we did not aim for the minimal set of conditions. Thus some of the conditions given below can surely be relaxed.

ASSUMPTION A.

(a) $\sup_j |x_j| \le c < \infty$,

(b) For all $n, 0 < c_1 < \dfrac{\sum_{j=1}^{n} x_j^2 \ E \ \psi'(\epsilon_j)}{n} < c_2 < \infty$

(c) ψ is twice differentiable and ψ'' is Lipschitz.

(d) $\inf_j \ E \ \psi'(\epsilon_j) > 0$

(e) $\sup_j E\,[|\psi'(\epsilon_j)+|\psi''(\epsilon_j)|+|\psi(\epsilon_j)|]^6<\infty$

Define

$$a_i = x_i\ \psi(\epsilon_i),\ \ a_i^* = a_i/x_i$$

$$b_i = x_i^2\ (\psi'(\epsilon_i) - E\psi'(\epsilon_i)),\ \ b_i^* = b_i/x_i$$

$$c_i = x_i^3\ E\ \psi''(\epsilon_i),$$

$$S_1 = \sum_{i=1}^n a_i,\ \ S_1^* = \sum_{i=1}^n a_i^*$$

$$S_2 = \sum_{i=1}^n (a_i^2 - Ea_i^2),\ \ S_2^* = \sum (a_i^{*2} - Ea_i^{*2})$$

$$S_3 = \sum_{i=1}^n b_i,\ \ S_3^* = \sum_{i=1}^n\ b_i^*$$

$$M_n = \sum_{i=1}^n x_i^2\ E\ \psi'(\epsilon_i),\ \ N_n = \sum_{i=1}^n\ c_i$$

$$P_n = \sum_{i=1}^n\ E(a_i b_i),\ \ P_n^* = \sum_{i=1}^n E(a_i^* b_i^*)$$

$$Q_n = \sum_{i=1}^n\ E(a_i^2),\ \ Q_n^* = \sum_{i=1}^n E(a_i^{*2})$$

$$v_n^* = \frac{L_n \sum_{i=1}^n E\ \psi^2(\epsilon_i)}{n\ M_n^2}$$

PROPOSITION. Under the Assumption (A), the following in probability expansions hold:

$$M_n(V_{n,J} - v_n) = \frac{S_2}{M_n} - 2\frac{S_1}{M_n}\left(\frac{P_n}{M_n} - \frac{Q_n N_n}{M_n^2}\right) - 2\frac{S_3}{M_n}\cdot\frac{Q_n}{M_n} + O_p(n^{-1})\ \ \ldots\ \ldots \tag{2.9}$$

$$\frac{nM_n}{L_n}(V_{n,WJ} - v_n^*) = \frac{S_2^*}{M_n} - 2\frac{S_1}{M_n}\left(\frac{P_n^*}{M_n} - \frac{Q_n^* N_n}{M_n^2}\right) - 2\frac{S_3}{M_n}\cdot\frac{Q_n^*}{M_n} + O_p(n^{-1}) \quad \ldots \ldots \ldots \tag{2.10}$$

PROOF OF PROPOSITION. In Lahiri (1992) a stochastic expansion is derived for β_n under the added restriction that $(\epsilon_j, j \geq 1)$ are i.i.d. This expansion is based on probability inequalities of Fuk and Nagaev (1971). It may be verified that with appropriate modifications, his arguments remain valid under our assumptions, thus yielding the following expansions for β_n and $\beta_{(i)}$. We will use "$'$" to denote expressions computed without involving the ith observation.

$$\beta_n - \beta = \frac{S_1}{M_n}\left(1 - \frac{S_3}{M_n}\right) + \frac{1}{2}\,\frac{S_1^2 N_n}{M_n^3} + R_n \quad \ldots \ldots \tag{2.11}$$

where $R_n = O_p(M_n^{-1}) = O_p(n^{-1})$.

$$\beta_{(i)} - \beta = \frac{S_1'}{M_n'}\left(1 - \frac{S_3'}{M_n'}\right) + \frac{1}{2}\,\frac{S_1'^2 N_n'}{M_n'^3} + R_n' \quad \ldots \ldots \tag{2.12}$$

$$\text{where } R_n' = O_p(M_n'^{-1}) = O_p(n^{-1}) \quad \ldots \ldots \tag{2.13}$$

Subtracting (2.12) from (2.11),

$$\beta_n - \beta_{(i)} = T_{1i} + T_{2i} - T_{3i} + T_{4i} + T_{5i} + \epsilon_{ni} \quad \ldots \ldots \tag{2.14}$$

where

$$
\begin{aligned}
T_{1i} &= \frac{a_i}{M_n}\left(1 - \frac{S_3}{M_n}\right) \\
T_{2i} &= \frac{b_i(a_i - S_1)}{M_n^2} \\
T_{3i} &= \frac{a_i(c_i - N_n)S_1}{M_n^3} \\
T_{4i} &= \frac{a_i^2(c_i - N_n)}{2M_n^3} \\
T_{Si} &= \frac{c_i S_1^2}{2M_n^3} \\
\epsilon_{ni} &= -\frac{S_1'}{M_n}\left(\frac{M_n}{M_n'} - 1\right) + \frac{S_1' S_3'}{M_n^2}\left(\frac{M_n^2}{M_n'^2} - 1\right) \\
&\quad - \frac{S_1'^2 N_n'}{2M_n^3}\left(\frac{M_n^3}{M_n'^3} - 1\right) + R_n - R_n'
\end{aligned}
\tag{2.15}
$$

It is easily seen that

$$\sum_{i=1}^{n} T_{1i}^2 = \left(1 - \frac{2S_3}{M_n}\right)\frac{\sum_{i=1}^{n} a_i^2}{M_n^2} + O_p(M_n^{-2}) \quad \dots \dots \tag{2.16}$$

$$\sum_{i=1}^{n} T_{1i}\, T_{3i} = \frac{-S_1 N_n}{M_n^4}\left(1 - \frac{S_3}{M_n}\right)\sum_{i=1}^{n} a_i^2 + O_p(M_n^{-2}) \quad \dots \dots \tag{2.17}$$

$$\sum_{i=1}^{n} T_{1i}\, T_{2i} = -\frac{S_1}{M_n^3}\left(1 - \frac{S_3}{M_n}\right)\sum_{i=1}^{n} a_i b_i + O_p(M_n^{-2}) \quad \dots \dots \tag{2.18}$$

If equation (2.14) is squared then it yields the above three terms and several other terms. Using the facts that $R_n = O_p(n^{-2})$ and $R_n' = O_p(n^{-2})$ and Assumption A which allows to obtain the variance of some of these terms, it is seen that all the other terms are $O_p(n^{-2})$. The details involve pages of tedious computation which we omit.

REMARK 1. If $\psi(x) = x$, then $S_1 = P_n = Q_n = N_n = 0, M_n = L_n$ and the expansions agree with (1.5) and (1.6) as they should.

REMARK 2. It is clear from the expansion (2.9), that $V_{n,J} - v_n \xrightarrow{P} 0$ even when the $(\epsilon_i, i \geq 1)$ are heteroscedastic. Thus $V_{n,J}$ is robust to departure from homogeneity of variance.

From (2.10), $V_{n,WJ} - v_n^* \xrightarrow{P} 0$. In general v_n and v_n^* will have different limits (see their definitions in (2.2) and before the Proposition) and hence $V_{n,WJ}$ is in general an inconsistent estimator of v_n^*. If $V(\psi(\epsilon_i)) = \sigma^2$ for all i, then $v_n^* = v_n$. Hence $V_{n,WJ}$ becomes consistent in this case. Thus it is interesting to compare the (asymptotic) variances of $V_{n,J}$ and $V_{n,WJ}$.

In the least squares case ($\psi(x) = x$), when $(\epsilon_i, i \geq 1)$ are i.i.d., Liu and Singh (1992) proved that $V_{n,WJ}$ always has a *smaller* asymptotic variance compared to $V_{n,J}$ (See equation (1.6) given in Section 1.). Compared to the least squares case, now the expansions involve two additional random terms, S_1 and S_3. This makes the comparison of the variance of $V_{n,J}$ and $V_{n,WJ}$ cumbersome. We will impose a few restrictions on ψ and the errors $(\epsilon_i, i \geq 1)$ to facilitate the comparison.

ASSUMPTION B.

(a) ψ is anti-symmetric. That is, $\psi(x) = -\psi(-x)$

(b) $(\epsilon_i, i \geq 1)$ are i.i.d. each with a distribution F which is symmetric around 0, $V(\psi(\epsilon_1)) = \sigma_0^2$, $E\psi'(\epsilon_1) = \mu$, $V(\psi^2(\epsilon_1)) = \sigma_2^2$. It may be noted that the restriction on ψ is quite reasonable. (See Hoaglin, Mosteller and Tukey (1983 page 365) for a discussion on desirable properties of ψ in general).

As a consequence of our assumptions, ψ' is symmetric and ψ'' is anti-symmetric. Hence $E\ \psi(\epsilon_1) = 0$, $E\ \psi(\epsilon_1)\psi'(\epsilon_1) = 0$ and $E\ \psi''(\epsilon_1) = 0$. This implies that $N_n = P_n = P_n^* = 0$, $M_n = \mu\ L_n, Q_n = \sigma_0^2\ L_n, Q_n^* = n\sigma_0^2$. Thus from (2.9) and (2.10) we obtain

$$M_n^2(V_J - v_n) = S_2 - \frac{2\sigma_0^2 S_3}{\mu} + O_p(1)$$

$$= J + O_p(1), \text{ say}$$

$$M_n^2(V_{WJ} - v_n^*) = \frac{L_n}{n}\ S_2^* - \frac{2\sigma_0^2 S_3}{\mu} + O_p(1)$$

$$= WJ + O_p(1), \text{ say}$$

Now

$$V(J)-V(WJ)=\sigma_2^2\sum_{i=1}^{n}x_i^4 \;-\; 4\frac{\sigma_0^2}{\mu}\sum_{i=1}^{n}x_i^4\ \text{Cov}\,(\psi^2(\epsilon_1),\psi'(\epsilon_1))$$

$$-\left[\sigma_2^2\,\frac{L_n^2}{n^2}\;-\;4\frac{\sigma_0^2}{\mu}\,\frac{L_n^2}{n}\ \text{Cov}\,(\psi^2(\epsilon_1),\psi'(\epsilon_1))\right] \tag{2.19}$$

$$=\left(\sum_{i=1}^{n}x_i^4-\frac{L_n^2}{n^2}\right)\left(\sigma_2^2-\frac{4\sigma_0^2}{\mu}\ \text{Cov}\,(\psi^2(\epsilon_1),\psi'(\epsilon_1))\right)\ \ldots\ \ldots$$

Unlike the least squares case, here the jackknife estimator will be *more* efficient than the weighted jackknife estimator if the second factor in the above expression is negative.

Example 1. (Huber's ψ-functions). Let

$$\psi(u)=\begin{cases}u & \text{if } |u|\le c\\ c & \text{if } |u|>c.\end{cases}$$

Then

$$\psi'(u)=\begin{cases}1 & \text{if } |u|<c\\ 0 & \text{if } |u|>c\end{cases}$$

Letting F denote the (continuous) cdf of ϵ_1, $p=P(-c<\epsilon_1<c)$,

$$\text{Cov}\,(\psi^2(\epsilon_1),\ \psi'(\epsilon_1))=\int_{-c}^{c}x^2\ dF(x)-\left(\int_{-c}^{c}x^2\ dF(x)+\int_{|x|>c}c^2dF(x)\right)\left(\int_{-c}^{c}1\cdot dF(x)\right)$$

$$=(1-p)\int_{-c}^{c}x^2\ dF(x)-(1-p)\ p\ c^2$$

$$=(1-p)\int_{-c}^{c}(x^2-c^2)\ dF(x)<0$$

Thus in this case $V_{n,J}$ is always *less* efficient than $V_{n,WJ}$ for all symmetric distributions.

Example 2. Let for $0<c<1$,

$$\psi(u)=\begin{cases}u, & |u|\le c\\ 0 & \text{otherwise}\end{cases}$$

Let $(\epsilon_i, i \geq 1)$ be i.i.d. $U(-1,1)$. It can be easily checked that

$$E\ \psi'(\epsilon_i) = c, \quad E\ \psi^2(\epsilon_1) = c^3/3,$$

$$E[\psi^2(\epsilon_1)\ \psi'(\epsilon_1)] = c^3/3, \quad E\ \psi^4(\epsilon_1) = c^5/5.$$

Upon simplifying (2.19), it follows that $V(J) < V(WJ)$ if

$$(E\psi'(\epsilon_1))[E\ \psi^4(\epsilon_1) + 3(E\ \psi^2(\epsilon_1))^2] < 4E[\psi^2(\epsilon_1)]E[\psi^2(\epsilon_1)\psi'(\epsilon_1)] \ \ldots\ \ldots \qquad (2.20)$$

Using the above expected values, it follows that $V(J) < V(WJ)$ if $c < 33/45$, $V(J) > V(WJ)$ if $c > 33/45$ and $V(J) = V(WJ)$ when $c = 33/45$.

Example 3. Let ψ be Tukeys' biweight function,

$$\psi(u) = \begin{cases} u(1-u^2)^2, & 0 < |u| < 1 \\ 0 & \text{otherwise} \end{cases}$$

Let $(\epsilon_i, i \geq 1)$ be i.i.d. with the density

$$f_\alpha(x) = \begin{cases} \left(\frac{1+\alpha}{2t}\right)\left(1 - \frac{|x|}{t}\right)^\alpha, & 0 < |x| < t \\ 0 & \text{otherwise} \end{cases}$$

where $0 < \alpha, t < \infty$.

In principle it is possible to write down all the expected values that are needed for the comparison. After a little trial and error, we chose $\alpha = 0.05$, $t = 11$. With this choice,

$$V(\psi^2(\epsilon_1)) = 1.986 \times 10^{-4}, \text{ Cov } (\psi^2(\epsilon_1), \psi'(\epsilon_1)) = 6.38 \times 10^{-7}$$

$$E\psi^2(\epsilon_1) = 34.5465 \times 10^{-5}, \ E\psi'(\epsilon_1) = 0.4436 \times 10^{-4}$$

Using these values, simple computation shows that $V(J) < V(WJ)$.

REMARK 3. It is expected that as in the LS case, the different bootstrap estimators will obey the appropriate expansions (2.9) or (2.10). The detailed technical proofs will be quite involved, long and tedious.

REFERENCES

Babu, G. J. (1992). Subsample and half-sample methods. *Ann. Inst. Stat. Math.* 44, 9, 703-720.

Fuk, D. K. and Nagaev, S. V. (1971). Probability inequalities for sums of independent random variables. *Theory Probab. Appl.* **16**, 643-660.

Hoaglin, D. C., Mosteller, F. and Tukey, J. K. (1983). *Understanding Robust and Exploratory Data Analysis.* John Wiley, New York.

Huber, J. (1980). *Robust Statistics*, Wiley, New York.

Lahiri, S. N. (1992). Bootstrapping M-estimators of a multiple linear regression parameter. *Ann. Statist.* 20, 3, 1548-1570.

Liu, R. Y and Singh, K. (1992). Efficiency and robustness in resampling *Ann. Statist.* 20, 1, 370-384.

Miller, R. G. (1974). The jackknife - a review. *Biometrika* 43, 353-360.

Shao J. and Wu, C. F. J. (1987). Heteroscedasticity-robustness of jackknife variance estimators in linear models. *Ann. Statist.*, 15, 1563-1579.

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