The jackknife method can also be used to obtain an approximately unbiased estimator $\tilde{\theta}_J$ of $\theta = g(\sigma_e^2, \sigma_v^2)$, i.e., $E(\tilde{\theta}_J) - \theta = o(t^{-1})$ for large t, without normality assumption. The estimator $\tilde{\theta}_J$ can be used in small area estimation to get approximately unbiased estimators of the weights in the best predictors. It may be noted that in the empirical Bayes literature (e.g., Morris (1983)), the weights are unbiasedly estimated under normality assumption in the balanced case, $n_i = m$.

Details of these results will be reported in a separate paper.

REFERENCES

ARVESEN, J. N. (1969). Jackknifing U-statistics. *Ann. Statist.* **40** 2076–2100.

MORRIS, C. N. (1983). Parametric empirical Bayes inference: theory and applications. *J. Amer. Statist. Assoc.* **78** 47–65.

RAO, J. N. K. (1973). On the estimation of heteroscedastic variances. *Biometrics* **29** 11–24.

DEPARTMENT OF MATHEMATICS AND STATISTICS
CARLETON UNIVERSITY
OTTAWA, ONTARIO K1S 5B6
CANADA

DEPARTMENT OF STATISTICS AND APPLIED PROBABILITY
UNIVERSITY OF ALBERTA
EDMONTON, ALBERTA T6G 2G1
CANADA

JUN SHAO

University of Wisconsin-Madison

Applications of the jackknife and other resampling methods to regression analysis have been thoroughly discussed in Professor Wu's paper. One interesting and stimulating aspect of his approach is the use of a weighting scheme that takes into account the unbalanced nature of regression data. He has provided a fundamental tool for handling very general non-i.i.d. problems for which the classical jackknife method may not work well. In Section 8, he considered extensions of his method to several non-i.i.d. situations. More research is needed and is being done in this area.

In this discussion, I would like to (A) propose another weighted resampling scheme that gives an interpretation of Wu's weighted jackknife and provides an alternative resampling estimation procedure, (B) discuss the use of Tukey's pseudovalue, and (C) obtain the stochastic order of the weighted jackknife bias estimator.

In the following, all notation will be the same as that of Wu.

(A) Another weighted resampling scheme. In the regression situation, the information contained in different subsets of data may be quite different. The idea of my proposed weighted resampling scheme is to take account of the unbalanced nature of the data in the resampling process. That is, the probability of selecting a subset of data is not a constant as is usually done, but is proportional to the determinant of the Fisher information matrix of the corresponding subset model with i.i.d. errors. We will see that the bias and variance estimators based on this resampling scheme are the same as those of Wu. However, the method to be described can be directly extended to other non-i.i.d. cases. Furthermore, this method can be applied to more general problems than bias and variance estimation, and the variable R defined in (2) can be random variables other than $\hat{\theta} - \theta$.

The weighted resampling procedure is described in detail as follows.

For any $r \leq n$, define a resampling vector $P^* = (P_1^*, \dots, P_n^*)^T$ by

$$\text{Prob}(P_i^* = 1, i \in s, P_i^* = 0, i \notin s) = W_s \quad \text{for any } s \in S_r, \tag{1}$$

where $W_s = |X_s^T X_s| / \Sigma_r |X_s^T X_s|$. For a fixed P^*, let $s^* = \{i | P_i^* = 1\}$. The selected subset model is then

$$Y_{s^*} = X_{s^*}\beta + e_{s^*}$$

and the corresponding LSE is

$$\beta^* = \left(X_{s^*}^T X_{s^*}\right)^{-1} X_{s^*}^T y_{s^*}.$$

Note that under (1), the probability of selecting a subset s is W_s. Denote the expectation and probability under the resampling scheme (1) by E_* and P_*, respectively. Let

$$R = \hat{\theta} - \theta, \tag{2}$$

where $\hat{\theta} = g(\hat{\beta})$, $\theta = g(\beta)$ for some function g of β. To estimate a functional $f(R)$ of R (e.g., mean or variance of $\hat{\theta} - \theta$), we calculate the resample analogue R^*. Then we use $f^*(R^*)$ to estimate $f(R)$, where f^* is an appropriate functional corresponding to f.

From

$$P_*(\beta^* = \hat{\beta}_s) = W_s,$$

we have

$$E_*\beta^* = \Sigma_r W_s \hat{\beta}_s = \Sigma_r |X_s^T X_s| \hat{\beta}_s / \Sigma_r |X_s^T X_s| = \hat{\beta},$$

where the last equality follows from Theorem 1. Since $R = \hat{\theta} - \theta = g(\hat{\beta}) - g(\beta)$ and $E\hat{\beta} = \beta$, the resample analogue of R is

$$R^* = g(\beta^*) - g(\hat{\beta}) = \theta^* - \hat{\theta},$$

where $\theta^* = g(\beta^*)$. Note that $ER = E\hat{\theta} - \theta$ is the bias of $\hat{\theta}$. The weighted resampling bias estimator is then

$$\begin{aligned} \hat{B}_{\text{WR}} &= \frac{r-k+1}{n-r} E_* R^* = \frac{r-k+1}{n-r} E_*(\theta^* - \hat{\theta}) \\ &= \frac{r-k+1}{n-r} \Sigma_r W_s(\hat{\theta}_s - \hat{\theta}), \end{aligned} \tag{3}$$

which coincides with Wu's weighted delete-d jackknife bias estimator (9.9). We take $(r-k+1)(n-r)^{-1}E_*$ as f^*. Multiplying E_* by the scalar factor $(r-k+1)/(n-r)$ is necessary since $(r-k+1)(n-r)^{-1}E_*R^*$ matches the order of ER (see the theorem in part (C) of this discussion and Section 4(iii) of

Wu's paper). The weighted resampling estimator of θ is simply

$$\hat{\theta}_{\mathrm{WR}} = \hat{\theta} - \hat{B}_{\mathrm{WR}}.$$

Similarly, we can estimate $E(RR^T)$ by $(r - k + 1)(n - r)^{-1}E_*(R^*R^{*T})$. Since $E(RR^T) = E(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T$ is close to $\operatorname{Var}\hat{\theta}$, $(r - k + 1)(n - r)^{-1} \times E_*(R^*R^{*T})$ can also be used as a variance estimator for $\hat{\theta}$. Note that

$$\frac{r-k+1}{n-r}E_*(R^*R^{*T}) = \frac{r-k+1}{n-r}\Sigma_r W_s(\hat{\theta}_s - \hat{\theta})(\hat{\theta}_s - \hat{\theta})^T,$$

which is identical to Wu's $v_{J,r}(\hat{\theta})$ (4.1). Again the factor $(r - k + 1)/(n - r)$ plays the role of order-matching (Theorem 1 of Shao and Wu (1985)).

As an example of another functional f, let us consider the estimation of the skewness $E(\hat{\theta} - \theta)^3$ of $\hat{\theta}$ in the case of scalar θ. By using a similar argument, $E(\hat{\theta} - \theta)^3$ can be estimated by $(r - k + 1)(n - r)^{-1}E_*R^{*3}$.

(B) Tukey's pseudovalue. As pointed out in Section 5, Tukey's pseudovalue works well for the i.i.d. case, but its extension to non-i.i.d. situations needs further theoretical investigation. Hinkley (1977) suggested the use of weighted pseudovalues for the delete-1 jackknife. His method gives the same jackknife point estimator as Wu's but does not provide a suitable variance estimator (see the discussions in Section 5 and in Shao and Wu (1985)). The following discussion shows that a better and more natural method is to use a weighted resampling scheme (1) instead of weighting the pseudovalues. All the results obtained coincide with those in part (A).

Define a generalized pseudovalue for any $s \in S_r$ by

$$p_s = \hat{\theta} - \frac{r-k+1}{n-r}(\hat{\theta}_s - \hat{\theta}).$$

If $k = 1$ and $r = n - 1$, the factor $(r - k + 1)/(n - r)$ is equal to $n - 1$ and p_s becomes Tukey's pseudovalue $\hat{\theta} - (n - 1)(\hat{\theta}_{(i)} - \hat{\theta})$.

Let p^* be a random variable satisfying

$$P_*(p^* = p_s) = W_s \quad \text{for any } s \in S_r.$$

Analogous to the usual jackknife, the weighted delete-d jackknife estimator of θ can be defined as

$$(5) \qquad \hat{\theta}_{J,r} = E_*p^* = \Sigma_r W_s p_s = \hat{\theta} - \frac{r-k+1}{n-r}\Sigma_r W_s(\hat{\theta}_s - \hat{\theta}) = \hat{\theta} - \hat{B}_{\mathrm{WR}},$$

where $\hat{B}_{\mathrm{WR}}$ is defined in (3) and $d = n - r$. Thus $\hat{\theta}_{J,r}$ is identical to $\hat{\theta}_{\mathrm{WR}}$ defined in (4). For the special case of $d = 1$, $\hat{\theta}_{J,n-1}$ also coincides with Hinkley's delete-1 jackknife estimator.

Similarly, we can obtain the weighted delete-d jackknife variance estimator by

$$(6) \qquad \frac{n-r}{r-k+1}E_*(p^* - \hat{\theta})(p^* - \hat{\theta})^T = \frac{r-k+1}{n-r}\Sigma_r W_s(\hat{\theta}_s - \hat{\theta})(\hat{\theta}_s - \hat{\theta})^T = v_{J,r}(\hat{\theta}).$$

The factor $(n - r)/(r - k + 1)$ on the left side of (6) is necessary for order matching. If $k = 1$ and $r = n - 1$, $(n - r)/(r - k + 1) = 1/(n - 1)$ is the same as in Tukey's jackknife variance estimator.

(C) The stochastic order of the bias estimator. In this part we study the stochastic order of the weighted delete-d bias estimator $\hat{B}_{\mathrm{WR}}$ (3). Under certain conditions (see Shao (1986)), the bias of $\hat{\theta}$ has order n^{-1}. Hence a natural requirement for $\hat{B}_{\mathrm{WR}}$ is that the stochastic order of $\hat{B}_{\mathrm{WR}}$ is $O_p(n^{-1})$. Another requirement for $\hat{B}_{\mathrm{WR}}$, which ensures that the weighted jackknife estimator $\hat{\theta}_{J,r}$ reduces bias, is that

$$E\hat{B}_{\mathrm{WR}} = E\hat{\theta} - \theta + L,$$

where L is of a smaller order than n^{-1}. This has been studied by Shao (1986) under some smoothness conditions on the function g.

The stochastic order of $\hat{B}_{\mathrm{WR}}$ is given in the following theorem. One consequence of the theorem is that

$$\sqrt{n}\,\hat{B}_{WR} \to 0 \quad \text{in probability.}$$

Hence $\sqrt{n}\,(\hat{\theta}_{J,r} - \theta)$ converges to a normal limit distribution if $\sqrt{n}\,(\hat{\theta} - \theta)$ does. A special case of this result for the delete-1 jackknife and $\sigma_i^2 = \sigma^2$ for all i was proved in Weber and Welsh (1983) under the stronger condition that g has a bounded second-order derivative in a neighborhood of β. A similar result for the unweighted jackknife can be found in Miller (1974). We state the following lemma first, whose proof can be found in Shao and Wu (1985).

LEMMA. *Suppose that σ_i^2 are uniformly bounded, $(X^TX)^{-1} = O(n^{-1})$ and $\lim_{n\to\infty}(n - r)h_n = 0$ where $h_n = \max_{1\le i\le n} x_i^T(X^TX)^{-1}x_i$. Then for any $\delta > 0$,*

$$\lim_{n\to\infty} \mathrm{Prob}(\|\hat{\beta} - \beta\| < \delta, \|\hat{\beta}_s - \hat{\beta}\| < \delta \textit{ for all } s \in S_r) = 1.$$

THEOREM. *Suppose that g is a function from R^k to R^m with Lipschitz-continuous first-order derivatives in a neighborhood of β. Then, under the same conditions as in the lemma,*

$$\hat{B}_{\mathrm{WR}} = O_p(n^{-1}).$$

PROOF. From Theorem 1 of Wu's paper and the mean-value theorem, we have

$$\hat{B}_{\mathrm{WR}} = \frac{r - k + 1}{n - r}\Sigma_r W_s[G(\zeta_s) - G(\hat{\beta})](\hat{\beta}_s - \hat{\beta}),$$

where $G(\zeta_s)$ and $G(\hat{\beta})$ are $m \times k$ matrices whose jth rows are the gradient of the jth component of g at ζ_{sj} and $\hat{\beta}$, respectively, and ζ_{sj} is on the line segment connecting $\hat{\beta}_s$ and $\hat{\beta}$. Let

$$A_n^\delta = \{\|\hat{\beta} - \beta\| < \delta, \|\hat{\beta}_s - \hat{\beta}\| < \delta \text{ for all } s \in S_r\}.$$

Then by the Lipschitz-continuity of g in a neighborhood of β, there is a $\delta > 0$ such that on A_n^δ,

$$\|G(\zeta_s) - G(\hat{\beta})\| \leq c\|\hat{\beta}_s - \hat{\beta}\|,$$

where c is a positive constant. Let $I_{A_n^\delta}$ be the indicator function of A_n^δ. Then

$$\begin{aligned} E|\hat{B}_{\mathrm{WR}} I_{A_n^\delta}| &\leq \frac{r-k+1}{n-r} \Sigma_r W_s \left[E\left(\|G(\zeta_s) - G(\hat{\beta})\|^2 I_{A_n^\delta} \right) E\|\hat{\beta}_s - \hat{\beta}\|^2 \right]^{1/2} \\ &\leq c \frac{r-k+1}{n-r} \Sigma_r W_s E\|\hat{\beta}_s - \hat{\beta}\|^2 \\ &= c\,\mathrm{Tr}[E v_{J,r}(\hat{\beta})] \\ &= O(n^{-1}), \end{aligned}$$

where the last equality follows from Theorem 1 of Shao and Wu (1985). Hence

$$\hat{B}_{\mathrm{WR}} I_{A_n^\delta} = O_p(n^{-1}).$$

From the lemma, $\mathrm{Prob}(A_n^\delta) \to 1$ as $n \to \infty$. Thus

$$\hat{B}_{\mathrm{WR}} = O_p(n^{-1}). \qquad \square$$

REFERENCES

HINKLEY, D. V. (1977). Jackknifing in unbalanced situations. *Technometrics* **19** 285–292.

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

SHAO, J. (1986). On resampling methods for variance and bias estimation in linear models. Technical report, Department of Statistics, Univ. of Wisconsin-Madison.

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

WEBER, N. C. and WELSH, A. H. (1983). Jackknifing the general linear model. *Austral. J. Statist.* **25** 425–436.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN 53706

JEFFREY S. SIMONOFF AND CHIH-LING TSAI

New York University

We would like to congratulate the author on a very interesting paper, and discuss some issues arising from jackknifing nonlinear models (Section 8). Much of what is presented here is based on Simonoff and Tsai (1986); V is the $n \times p$ matrix of first partial derivatives of $f(\cdot)$ with respect to $\boldsymbol{\theta}$, while W is the $n \times p \times p$ array of second partial derivatives.

1. Alternative weighting schemes. The weighted jackknife originally suggested by Hinkley (1977) was applied to nonlinear models by Fox et al. (1980),