A finite-sample version of this disaster is even easier to arrange. For example, if $1 \le i \le n$, let

$$\delta_i = 0 \quad \text{with probability } 1 - \lambda/n,$$

$$\delta_i = \sqrt{n}\, W_i \quad \text{with probability } \lambda/n.$$

The moral seems clear: Second-order moment conditions or no, with skew long-tailed errors that change from observation to observation, the jackknife cannot be relied upon. On the other hand, preliminary calculations suggest that in our special case, with independence, in the presence of Lindeberg's condition both the jackknife and the bootstrap will perform adequately. Wu admits dependent errors, and this introduces further complications.

REFERENCES

DAGGETT, R. and FREEDMAN, D. (1985). Econometrics and the law: a case study in the proof of antitrust damages. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (L. M. Le Cam and R. A. Olshen, eds.) **1** 123–172. Wadsworth, Monterey, Calif.

FREEDMAN, D. (1981). Bootstrapping regression models. *Ann. Statist.* **9** 1218–1228.

FREEDMAN, D. (1984). On bootstrapping two-stage least-squares estimates in stationary linear models. *Ann. Statist.* **12** 827–842.

FREEDMAN, D. and PETERS, S. (1984a). Bootstrapping a regression equation: some empirical results. *J. Amer. Statist. Assoc.* **79** 97–106.

FREEDMAN, D. and PETERS, S. (1984b). Bootstrapping an econometric model: some empirical results. *J. Bus. Econ. Statist.* **2** 150–158.

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720

MALAY GHOSH[1]

University of Florida

Professor Wu is to be complimented for bringing out several important issues on jackknife, bootstrap and other resampling methods in regression analysis. Using a representation of the full-data least-squares estimator as a weighted average of corresponding least-squares estimators for appropriately chosen subsets, he has been able to motivate very successfully general-weighted jackknife in regression. I agree with the author that a jackknife that allows for the deletion of an arbitrary number of observations at a time is more flexible than the delete-one jackknife. However, I will be surprised if, for estimating nonsmooth functions such as the median, a delete-d jackknife estimator will necessarily rectify the deficiency of a delete-one jackknife estimator.

Although $v_{J(1)}$ enjoys the same robustness property of $v_{H(1)}$ when the errors are independent, but not identically distributed, and the design matrix satisfies

[1] Research supported by National Science Foundation grant DMS-8218091.

conditions (C1)–(C3) given in Section 5 of this paper, I will be hesitant to recommend use of either $v_{J(1)}$ or $v_{H(1)}$ when there is dependence among the errors. To see this, consider the following two sample problem.

Let $y_{ij} = \beta_j + e_{ij}$ $(i = 1, \dots, n_j;\ j = 1, 2)$, where $E(e_{ij}) = 0$, $V(e_{ij}) = \sigma^2$, $\mathrm{Cov}(e_{ij}, e_{i'j}) = \rho_j \sigma^2 (0 < \rho_j < 1)$ when $i \neq i'$ and $\mathrm{Cov}(e_{ij}, e_{i'j'}) = 0$ when $j \neq j'$, $1 \leq i \leq n_j$, $j = 1, 2$. This situation arises quite often in mixed models and is a generalization of the model considered by Professor Wu in Section 6 of his paper when $\rho_1 = \rho_2 = 0$. In our setup, the design matrix X can be written as

$$X^T = (x_{11}, \dots, x_{n_1 1}, x_{12}, \dots, x_{n_2 2}), \tag{1}$$

where $x_{i1}^T = (1 \quad 0)$ and $x_{i2}^T = (0 \quad 1)$. Also, writing I_u for the identity matrix of order u and 1_u for the u-component column vector with all its elements equal to 1, one can express the variance–covariance matrix Σ as

$$\Sigma = \begin{bmatrix} (1-\rho_1)I_{n_1} + \rho_1 1_{n_1} 1_{n_1}^T & 0 \\ 0 & (1-\rho_2)I_{n_2} + \rho_2 1_{n_2} 1_{n_2}^T \end{bmatrix} \tag{2}$$

Since, in this case $\Sigma X = XQ$ with $Q = \mathrm{diag}(1 + (n_1 - 1)\rho_1, 1 + (n_2 - 1)\rho_2)$, the BLUE $\hat{\beta}$ of β is the same as its LSE (see Zyskind (1967)) and is given by

$$\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} = \begin{pmatrix} \overline{Y}_1 \\ \overline{Y}_2 \end{pmatrix}.$$

Thus,

$$V(\hat{\beta}) = \sigma^2 \mathrm{diag}\big(n_1^{-1}(1 + (n_1 - 1)\rho_1), n_2^{-1}(1 + (n_2 - 1)\rho_2)\big). \tag{3}$$

We shall now see that $Ev_{J(1)} \nrightarrow V(\hat{\beta})$ when $\min(n_1, n_2) \to \infty$.

To see this, define

$$r_{ij} = y_{ij} - x_{ij}^T \hat{\beta} = y_{ij} - \bar{y}_j, \qquad i = 1, \dots, n_j;\ j = 1, 2.$$

Easy calculations give

$$E(r_{ij}) = 0, \qquad E\big(r_{ij}^2\big) = \sigma^2\big(1 - n_j^{-1}\big)(1 - \rho_j). \tag{4}$$

Also, write $w_{ij} = x_{ij}^T (X^T X)^{-1} x_{ij}$ so that

$$w_{ij} = n_j^{-1} \quad \text{for every } i = 1, \dots, n_j \text{ and } j = 1, 2. \tag{5}$$

Note that

$$v_{J(1)} = (X^T X)^{-1} \left\{ \sum_{j=1}^{2} \sum_{i=1}^{n_j} r_{ij}^2 (1 - w_{ij})^{-1} x_{ij} x_{ij}^T \right\} (X^T X)^{-1}, \tag{6}$$

so that after some algebra, one gets

$$E[v_{J(1)}] = \sigma^2 \mathrm{diag}\big(n_1^{-1}(1 - \rho_1), n_2^{-1}(1 - \rho_2)\big). \tag{7}$$

Thus, when $\min(n_1, n_2) \to \infty$, $E[v_{J(1)}] \to 0$, while $V(\hat{\beta}) \to \sigma^2 \mathrm{diag}(\rho_1, \rho_2)$. A similar phenomenon holds for $v_{H(1)}$. Our model is suitable when there is exchangeability among observations receiving the same treatment, and is as realistic as the assumption of their independence. Routine use of $v_{J(1)}$ or $v_{H(1)}$ without any modification seems to be dangerous in this situation as it may lead to inconsistent estimators of $V(\hat{\beta})$.

As a second example, consider the simple linear regression model $y_i = \alpha + \beta(x_i - \bar{x}) + e_i$, where $E(e_i) = 0$, $V(e_i) = \sigma^2$ and $\text{Cov}(e_i, e_j) = \rho\sigma^2 (0 < \rho < 1)$, $1 \le i \ne j \le n$. This example, though similar to the previous one, brings out some extra features not found earlier. In particular, we shall see that while the appropriate component of $v_{J(1)}$ unbiasedly estimates $V(\hat{\beta})$, the other diagonal element of $v_{J(1)}$ gives an inconsistent estimator of $V(\hat{\alpha})$.

To see this, write

$$(8)\quad X^T = \begin{pmatrix} 1 & \cdots & 1 \\ x_1 - \bar{x} & \cdots & x_n - \bar{x} \end{pmatrix} \quad \text{and} \quad \Sigma = V(e) = \sigma^2\left[(1-\rho)I_n + \rho 1_n 1_n^T\right].$$

Hence, $\Sigma X = XQ$ with $Q = \text{diag}(1 + (n-1)\rho, 1 - \rho)$. Thus, once again, the BLUE of $\binom{\alpha}{\beta}$ is the same as its LSE, and is given by $\binom{\hat{\alpha}}{\hat{\beta}}$, where $\hat{\alpha} = \bar{y}$ and

$$\hat{\beta} = \sum_{i=1}^{n} y_i(x_i - \bar{x}) \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2.$$

Hence,

$$V(\hat{\alpha}) = \sigma^2 n^{-1}(1 + (n-1)\rho), \qquad V(\hat{\beta}) = \sigma^2(1-\rho) \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2,$$

and $\text{Cov}(\hat{\alpha}, \hat{\beta}) = 0$. Also, using Professor Wu's notation, in this case

$$w_i = n^{-1} + (x_i - \bar{x})^2 \Big/ \sum_{i=1}^{n} (x_i - \bar{x})^2,$$

$E(r_i) = 0$ and $E(r_i^2) = \sigma^2(1-\rho)(1-w_i)$. Hence,

$$E[v_{J(1)}] = \sigma^2(1-\rho)(X^T X)^{-1} = \sigma^2(1-\rho)\,\text{diag}\left(n^{-1}, \left(\sum_{i=1}^{n}(x_i - \bar{x})^2\right)^{-1}\right).$$

Thus as $n \to \infty$, $V(\hat{\alpha}) \to \rho\sigma^2$, but the corresponding component of $E[v_{J(1)}]$ converges to zero. What I would like to see is a general theory for jackknife with some correlated error structure as in the preceding text.

It is possibly difficult to make analytic comparison of the MSE's of the different estimators of $V(\hat{\beta})$, but it would have been instructive to see their Monte Carlo performances at least in some of the specific examples considered by the author.

REFERENCE

ZYSKIND, G. (1967). On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *Ann. Math. Statist.* **38** 1092–1109.

DEPARTMENT OF STATISTICS
UNIVERSITY OF FLORIDA
GAINESVILLE, FLORIDA 32611