M. S. SRIVASTAVA

University of Toronto

This is technically an impressive article that is full of innovative ideas on resampling methods. Although the unbalanced nature of the regression model was first recognized by Hinkley (1977), it is this paper that brings to light various pitfalls in the estimation of the variance of regression estimates by different resampling schemes. Motivated by a representation for the least-squares estimator, first given by Subrahmanyam (1972), although not as rigorous as in this paper, Wu provides robust estimates of the variance. It also shows the failure of the bootstrap method.

The failure of the bootstrap method is not new. Many examples exist in the literature, even when the observations are not as unbalanced as in the regression case (see, for example, Singh (1981)). Another example is when the vectors (X_i, Y_i) are independently distributed with mean vector zero, $\mathrm{Var}(X_i) = \sigma_{11}$, $\mathrm{Cov}(X_i, Y_i) = \sigma_{12}$ and $\mathrm{Var}(Y_i) = a_i\sigma_{22}$, a_i's known, $i = 1, \dots, n$. An estimate of σ_{22} is given by

$$\hat{\sigma}_{22} = n(n-1)^{-1}(\Sigma a_i)^{-1}\Sigma(Y_i - \overline{Y})^2,$$

with a bootstrap estimate

$$\hat{\sigma}_{22}^* = n(n-1)^{-1}(\Sigma a_i)^{-1}\Sigma(Y_i^* - \overline{Y}^*)^2.$$

The bootstrap estimate of $\mathrm{Var}(\hat{\sigma}_{22})$ is not consistent, unless $a_i \to 1$.

It is thus clear that neither the bootstrap method nor the jackknife method can be applied indiscriminately. Care needs to be taken. For example, in the jackknife case the delete-one method does not yield a consistent estimate of the variance of the sample median. In the regression case, until this paper, the problem had remained unresolved. But for "inference purposes" both the nonparametric bootstrap method as well as jackknife method require the assumption that the error terms are independently and identically distributed with means 0 and constant variance. In addition, both require that at least $w_i = x_i^T(X^TX)^{-1}x_i \to 0$ as $n \to \infty$, a condition due to Srivastava (1971).

In Section 7, Wu provides a method for handling the unbalanced nature of the residuals for bootstrapping. However, it appears somewhat manipulative. Obviously, if the model is

$$Y_i^* = x_i^T\hat{\beta} + \frac{r_i}{(1-w_i)^{1/2}}t_i^*, \qquad i = 1\dots, n$$

(equation (7.1) of Wu), then the bootstrap estimator should be defined by

$$\hat{\beta}^*_{\text{new}} = (X^TD^{-1}X)^{-1}X^TD^{-1}Y^*, \qquad Y^* = (Y_1^*, \dots Y_n^*)^T,$$

where

$$D = \mathrm{diag}\left[(1-w_1)^{-1}r_1^2, \dots, (1-w_n)^{-1}r_n^2\right].$$