of the bootstrap estimate of variance to the asymptotic variance. Nor does weak convergence to normality of the empirical distribution of the centered pseudovalues guarantee corresponding convergence of the jackknife estimate of variance. The situation begs for robustification—replacement of the variance functional by a scale equivariant functional that equals variance at normal distributions, but is weakly continuous there while retaining high asymptotic efficiency. One possibility is a standardized trimmed variance.

A similar argument exists for replacing the mean functional by a symmetrically trimmed mean (say) in bootstrap and jackknife estimates of bias.

REFERENCES

ABRAMOVITCH, L. and SINGH, K. (1985). Edgeworth corrected pivotal statistics and the bootstrap. *Ann. Statist.* **13** 116–132.

BERAN, R. (1984). Jackknife approximations to bootstrap estimates. *Ann. Statist.* **12** 101–118.

BERAN, R. and MILLAR, P. W. (1985). Asymptotic theory of confidence sets. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (L. M. Le Cam and R. A. Olshen, eds.). Wadsworth, Monterey, Calif.

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

DEPARTMENT OF STATISTICS
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720

RAYMOND J. CARROLL[1] AND DAVID RUPPERT[2]

University of North Carolina

Professor Wu is to be congratulated on a very interesting paper that advances our knowledge of jackknife methods and illustrates some problems of heteroscedastic data. Of course, Professor Wu's paper does not demonstrate a superiority of the jackknife over the bootstrap and is not intended as such. The bootstrap is a more general method. The bootstrap philosophy is to estimate the probability distribution of the data as accurately as possible and then find or approximate the sampling distribution of the relevant statistic under this estimated distribution. We agree with this philosophy. The present paper does a great service in underscoring the need for care about assumptions, both in this specific case and in statistics in general.

The robustness of the jackknife variance estimator to nonconstant variance is an interesting and potentially useful property, but what is its real importance for statistical practice? To answer this question we need to ask, "What types of heteroscedasticity can we expect in practice and what should be done about

[1]Research supported by the Air Force Office of Scientific Research under contract AFOSR-S-49620-85-C-0144.

[2]Research supported by the National Science Foundation under grant DMS-8400602.

them?" There are at least three basic forms of heteroscedasticity:

(1) σ_i is unrelated to x_i or is only slightly dependent on x_i.
(2) σ_i varies systematically and smoothly with x_i, perhaps as a function of the mean $x_i^t\beta$.
(3) σ_i is (nearly) constant for most observations but is large for one or at most a few cases, which will typically be response outliers.

A particularly simple case included in (1) is where the σ's are mutually independent, identically distributed, and independent of the x's. Unconditional on the unkown σ's the errors $u_i = \sigma_i \varepsilon_i$ are homoscedastic though heavier-tailed than the ε's. The least-squares estimator should probably be replaced by a robust estimate, but the usual (exchangeable errors) bootstrap can be used.

In case (2) careful data analysis will often allow us to model the relationship between σ and x. We examined several Monte Carlo samples from the "unequal variances" model in Section 10. For most samples the pattern of the residuals indicates that σ is an increasing function of x. In such a situation one very well might postulate the model

$$\sigma_i = \sigma x_i^{\theta}, \tag{1}$$

which is in fact the true model here. Other models, e.g., that σ is a linear function of x, would serve as acceptable approximations to the true model. Heteroscedastic regression models where the variance depends on unknown parameters are the subject of an extensive literature, e.g., Box and Hill (1974), Jobson and Fuller (1980), Carroll and Ruppert (1982, 1987). For larger data sets and possibly even when the sample size is as small as here ($n = 12$), σ can be estimated nonparametrically as a smooth function of x (Carroll, 1982).

We performed a Monte Carlo study of an estimator based on the true model (1) applied to Wu's unequal variances model. The parameters σ and θ were initially estimated by nonlinear least squares:

$$\sum_{i=1}^{n} \left\{ \left[r_i / \left(\hat{\sigma} x_i^{\theta} \right) \right]^2 - 1 \right\}^2 = \min!,$$

where $r_i = (y_i - x_i^t\hat{\beta})/(1 - w_i)$. These estimators of θ and σ are asymptotically equivalent to the MLE. In practice, one will want to estimate β by generalized least squares (least squares with inverse estimated variances as weights), and usually better estimates of θ can be obtained by GLS residuals. Most often in our study $\hat{\theta}$ was close to the true value $\theta = 0.5$, but occasionally $\hat{\theta}$ fell below 0 or was greater than 1. Typically the statistician has some prior knowledge of the possible types of heteroscedasticity, for example that σ neither decreases with x nor increases too rapidly with x, and this prior information can be particularly important for small data sets. We used the estimators $\tilde{\theta} = \min\{1, \max\{0, \hat{\theta}\}\}$ and

$$\tilde{\sigma}^2 = \sum_{i-1}^{n} \left[\left(y_i - x_i^t\hat{\beta} \right) / \left(x_i^{\tilde{\theta}} \right) \right]^2 / (N - p).$$

Then we estimated the covariance matrix of the unweighted LS estimate $\hat{\beta}$ by

$$(X^tX)^{-1}(X^t\tilde{\Sigma}X)(X^tX)^{-1},$$

where $\tilde{\Sigma} = \text{diag}[(\tilde{\sigma}x_i^{\tilde{\theta}})^2]$. The relative biases are

(0,0)	(0,1)	(0,2)	(1,1)	(1,2)	(2,2)
0.149	−0.120	0.118	0.095	−0.096	0.098.

Except for (0,0) and (0,1) the relative biases are close to those of the best jackknife estimators, and for all elements of the covariance matrix the squared biases were always less that 6% of the MSE. There are several advantages to modeling the heteroscedasticity rather than using least squares and a robust (to heteroscedasticity) variance estimate. The generalized least-squares estimator is asymptotically fully efficient. Prediction intervals for y given a "new x" can be extremely biased if a homoscedastic model is incorrectly assumed, even if the covariance matrix of $\hat{\beta}$ is properly estimated. The weighted residuals are (approximately) exchangeable and can be bootstrapped. The bootstrap would pick up the extra variability in the generalized least-squares estimator due to estimation of the weights. Even when the heteroscedasticity is modeled, there may be residual heterogeneity of variance due to modeling error, and a variance estimator robust to heteroscedasticity might sometimes be appropriate. For the same reason, a robust M-estimator would be preferable to least squares.

Heteroscedasticity of type (3) can occur when the regression model breaks down or the response suddenly becomes more variable at the extremes of the factor space. An interesting example of this is the Pamlico Sound salinity data in Carroll and Ruppert (1985). In such situations σ is large for high leverage points and the assumption of homoscedasticity can cause a severe downward bias in the estimated variance of $\hat{\beta}$. In this case Wu's jackknife estimates may be very useful, though it is quite likely that the bootstrap methodology can be extended to apply here as well. In any case a bounded influence estimate (Krasker and Welsch (1982)) would be preferable to least squares. It is interesting to note that Huber's (1983) objection to bounded-influence estimation, namely that the saddlepoint (where the Krasker–Welsch estimator is minimax) "corresponds to the (unrealistically pessimistic) assumption that Nature selectively places most of the contamination on points with the highest leverage," does *not* apply here. The pessimistic assumption *is* realistic in this situation, as well as in many others.

REFERENCES

Box, G. E. P. and Hill, W. J. (1974). Correcting inhomogeneity of variance with power transformation weighting. *Technometrics* **16** 385–389.

Carroll, R. J. (1982). Adapting for heteroscedasticity in linear models. *Ann. Statist.* **10** 1224–1233.

Carroll, R. J. and Ruppert, D. (1982). Robust estimation in heteroscedastic linear models. *Ann. Statist.* **10** 429–441.

Carroll, R. J. and Ruppert, D. (1985). Transformations in regression: a robust analysis. *Technometrics* **27** 1–12.

Carroll, R. J. and Ruppert, D. (1987). *Weighting and Transformations in Regression*. Book in preparation, Chapman and Hall, London.

HUBER, P. J. (1983). Minimax aspects of bounded-influence regression. *J. Amer. Statist. Assoc.* **78** 66–80.

JOBSON, J. D. and FULLER, W. A. (1980). Least squares estimation when the covariance matrix and parameter vector are functionally related. *J. Amer. Statist. Assoc.* **75** 176–181.

KRASKER, W. S. and WELSCH, R. E. (1982). Efficient bounded-influence regression estimation. *J. Amer. Statist. Assoc.* **77** 595–604.

DEPARTMENT OF STATISTICS
UNIVERSITY OF NORTH CAROLINA
CHAPEL HILL, NORTH CAROLINA 27514

B. EFRON

Stanford University

The remarks that follow are mainly critical, but that is not unusual when statisticians discuss difficult new areas of research. My criticism is not meant to obscure the paper's many positive achievements: the neat development of resampling methods for the linear regression problem, in particular Theorem 2; the extended class of weighted jackknives introduced in Section 4, and their justification in Theorem 3; and the intriguing suggestion in Section 8 for a more general weighted jackknife based on the Fisher information. The paper's main fault, in my opinion, is not the absence of interesting new ideas but rather an overinterpretation of results, which leads to bold distinctions not based on genuine differences.

(A) I reran part of the simulation experiment of Section 10, exactly as described except for the following change: Instead of taking the $e_i \sim N(0, x_i/2)$, I took them $N(0, |x_i - 5.5|)$. This gives nearly the same set of variances for the errors e_i, but with the large variances occurring at both ends of the x range, rather than just at the right end. Only the estimation of $\mathrm{Var}(\beta_0)$ (actually equal 3.64 in this situation) was considered, and only by the two estimators $v_{J(1)}$, definition (5.1), and $\hat{v}$, definition (2.9).

Here are summary statistics for 400 Monte Carlo trials:

	mean	st. dev.	rms
$v_{J(1)}$	3.47	3.14	3.14
$\hat{v}$	2.40	1.20	1.73

(rms indicates root mean square error). Now $\hat{v}$, the ordinary estimator (and also the "residual bootstrap" estimator v_b (2.9)), is biased sharply downward instead of upward as in Table 1; $v_{J(1)}$ is nearly unbiased, as it was designed to be.

However $v_{J(1)}$ *is much more variable than* $\hat{v}$, having nearly three times the standard deviation and twice the rms error for estimating $\mathrm{Var}(\beta_0)$. The percentiles of the two Monte Carlo distributions

	5%	10%	16%	50%	(true)	84%	90%	95%
$v_{J(1)}$	0.57	0.83	1.02	2.47	(3.64)	6.15	7.80	9.63
$\hat{v}$	0.88	1.12	1.27	2.14	(3.64)	3.65	4.06	4.56