This analysis shows that plants with more starting capital (larger seeds) or an earlier start on life (early germination date) attain larger final size, in accord with data from other plant species (Harper (1977)). Without a procedure such as Wu's resampling approach, formal analysis would have been far more difficult.

Estimation of natural selection gradients is a straightforward problem in multiple regression, but significance testing is complicated by constraints on transformations necessary for preservation of genetic interpretation and the relationship to population genetic theory. Resampling techniques provide a useful solution to this problem that is accessible to many biological practitioners.

Acknowledgment. Thanks are extended to E. Nordheim for helpful comments on the manuscript.

REFERENCES

ARNOLD, S. J. and WADE, M. J. (1984). On the measurement of natural and sexual selection: theory. *Evolution* **38** 709–719.

ELLSTRAND, N. C. and ANTONOVICS, J. (1985). Experimental studies of the evolutionary significance of sexual reproduction II. A test of the density-dependent selection hypothesis. *Evolution* **39** 657–666.

GRANT, B. R. (1985). Selection on bill characters in a population of Darwin's finches: *Geospiza conirostris* on Isla Genovesa, Galapagos. *Evolution* **39** 523–532.

HARPER, J. L. (1977). *Population Biology of Plants*. Academic, New York.

LANDE, R. and ARNOLD, S. J. (1983). The measurement of selection on correlated characters. *Evolution* **37** 1210–1227.

MITCHELL-OLDS, T. and WALLER, D. M. (1985). Relative performance of selfed and outcrossed progeny in *Impatiens capensis*. *Evolution* **39** 533–544.

SCHLUTER, D. and SMITH, J. N. M. (1986). Natural selection on beak and body size in the song sparrow. *Evolution* **40** 221–231.

SEBER, G. A. F. (1977). *Linear Regression Analysis*. Wiley, New York.

DEPARTMENT OF GENETICS AND BOTANY
UNIVERSITY OF WISCONSIN
MADISON, WISCONSIN 53706

RICHARD A. OLSHEN

University of California, San Diego

Professor Wu's paper raises many interesting points, only a few of which are touched upon in these comments.

If the bootstrap mentality is that the bootstrap sample bears approximately the same relationship to the empirical distribution of the data that the data bear to the distribution from which they were drawn, and if, in addition, the bootstrap process is to sample (x, y) pairs of residuals as if they were iid, then the deterministic predictor regression model studied here is not one for which bootstrap ideas ought to work well. Since the x_i's of (2.1) may be stratified, perhaps any resampling plan should reflect this stratification. In particular,

Institute of Mathematical Statistics is collaborating with JSTOR to digitize, preserve, and extend access to *The Annals of Statistics.*
www.jstor.org

then, neither assumption (A) nor assumption (B1) seems appropriate for Wu's model. Issues of stratification of resampling in a cross-validatory scenario different from his are touched upon in Breiman et al. (1984) and Olshen et al. (1985).

If the goal of techniques studied here is to protect against bias, then I believe one might take the following view, that I learned in conversations with graduate student Chongen Bai and which may bear further study. Any candidate bias correction has first to complete with the candidate correction 0. Thus, a preliminary test of the null hypothesis that the bias is 0 might precede the bias adjustment. So whenever a jackknife or other bias adjustment has a standard error that can conveniently be estimated, even approximately, then the value of a t-like statistic might be examined before the original parametric or resubstitution analysis is discarded. This tentative proposal is motivated in part by Stein's (1964) preliminary t-test in his argument for the quadratic loss inadmissibility of the best affine equivariant estimator of a normal variance.

That heterogeneity of scale can degrade the bootstrap has arisen with the study of "$o(1/\sqrt{n})$" convergence of bootstrap distributions of certain statistics. This phenomenon is commented on by Hartigan (1986) in his discussion of the paper by Efron and Tibshirani (1986). I first learned a similar argument from Athreya. In the context I have in mind, one studies statistics that are functions of a vector of sample moments. The Edgeworth-like expansions that arise in these studies require finite third moments of the data, and since sample moments figure in the vector, finite moments of the raw data three times the highest moment in the vector are required. Theoretical bootstrap distributions are necessarily discrete, but their asymptotic expansions are facilitated if at least they are strongly nonlattice. This is accomplished almost surely if the data have a continuous underlying distribution. Finally, a multivariate Edgeworth-like expansion needs conversion to a univariate one as in Lemma 5 of Babu and Singh (1983) or Lemma 2.1 of Bhattacharya and Ghosh (1978), and in the conversion, the scales of the bootstrapped statistic and the statistic computed from the original data must be synchronous in an obvious way. The conversion from multivariate to univariate expansion requires that the statistic under question is a sufficiently smooth function of the vector of sample moments. If any of the preceding requirements fails, then hyperefficiency seems lost. As Hartigan notes, asynchronous scales preclude cancellation of the lead terms when two expansions are subtracted; but still, if all else holds then at least $O(1/\sqrt{n})$ convergence of theoretical bootstrap distribution to true sampling distribution can hold. (Actually, the multivariate–univariate conversion is not required for $O(1/\sqrt{n})$ convergence, though some smoothness of the statistic is.) With sufficiently heavy tails to the raw data, even consistency of the theoretical bootstrap distribution can fail. Interested readers should consult the paper of Ghosh et al. (1984) and the forthcoming paper of Athreya (1987). The first paper regards bootstrapping a median; the latter two are concerned with the mean. In each instance a way out of the problem that arises from heavy-tailed data is offered. Wu emphasizes that heterogeneity of scale in his context also destroys the bootstrap.

In closing I note that the bootstrap can be an inconsistent procedure when it is employed to correct potentially large biases of some nonparametric techniques. See Section 11.7 of Breiman et al. (1984).

REFERENCES

ATHREYA, K. B. (1987). Bootstrap of the mean in the infinite variance case. To appear in *Ann. Statist.*

BABU, G. J. and SINGH, K (1983). Inference on means using the bootstrap. *Ann. Statist.* **11** 999–1003.

BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 434–451.

BREIMAN, L., FRIEDMAN, J. H., OLSHEN, R. A. and STONE, C. J. (1984). *Classification and Regression Trees.* Wadsworth, Belmont, Calif.

EFRON, B. and TIBSHIRANI, R. (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statist. Sci.* **1** 54–75, 77.

GHOSH, M., PARR, W. C., SINGH, K. and BABU, G. J. (1984). A note on bootstrapping the sample median. *Ann. Statist.* **12** 1130–1135.

HARTIGAN, J. A. (1986). Comment on "Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy" by Efron and Tibshirani. *Statist. Sci.* **1** 75–77.

OLSHEN, R. A., GILPIN, E. A., HENNING, H., LEWINTER, M. L., COLLINS, D. and ROSS, J., JR. (1985). Twelve month prognosis following myocardial infarction: classification trees, logistic regression, and stepwise linear discrimination. In *Proceedings of the Berkeley Conference in Honor of Jerzy Neyman and Jack Kiefer* (L. M. Le Cam and R. A. Olshen, eds.) **1** 245-267. Wadsworth, Monterey, Calif.

STEIN, C. (1964). Inadmissibility of the usual estimator for the variance of a normal distribution with unknown mean. *Ann. Inst. Statist. Math.* **16** 155–160.

DEPARTMENT OF MATHEMATICS
UNIVERSITY OF CALIFORNIA
LA JOLLA, CALIFORNIA 92093

J. N. K. RAO AND N. G. N. PRASAD

Carleton University and University of Alberta

1. General comments. This important paper is a major contribution to jackknife methodology. A major strength of the proposed weighted jackknife method is its ready extensibility to nonlinear situations, including the important generalized linear models with uncorrelated errors briefly discussed in Section 8.

In the case of a linear regression model with uncorrelated errors, the delete-1 jackknife variance estimator, $v_{J(1)}$, is shown to be exactly unbiased for $\mathrm{Var}(\hat{\beta})$ under $\mathrm{Var}(e) = \sigma^2 I$, and approximately unbiased (as $n \to \infty$) under $\mathrm{Var}(e) = \mathrm{diag}(\sigma_i^2)$. However, $v_{J(1)}$ seems to have no special advantage over the MINQUE (minimum norm quadratic unbiased estimator) of $\mathrm{Var}(\hat{\beta})$ (Rao (1973)) under the criterion of bias robustness since the latter estimator is exactly unbiased under $\mathrm{Var}(e) = \mathrm{diag}(\sigma_i^2)$ unlike $v_{J(1)}$. It may also be noted that the MINQUE of $\mathrm{Var}(\hat{\beta})$ seldom becomes negative definite even though the MINQUE of individual σ_i^2 may assume negative values. If $\theta = g(\beta)$, then the linearization technique can be used to get a MINQUE-based estimator of the variance of $\hat{\theta} = g(\hat{\beta})$. This variance estimator should be satisfactory since Wu's simulation study shows that