(2), $\hat{\theta}_{ADJ1(\psi)}(\alpha)$ is *third-order* equivalent to both $\hat{\theta}_{STUD}(\alpha)$ and $\hat{\theta}_{ABC}(\alpha)$ in the sense that their expansions match up to and including the term of order $n^{-3/2}$.

REFERENCES

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

HALL, P. (1983). Inverting an Edgeworth expansion. *Ann. Statist.* **11** 569–576.

HALL, P. (1986). On the bootstrap and confidence intervals. *Ann. Statist.* **14** 1431–1452.

LOH, W.-Y. (1984). Estimating an endpoint of a distribution with resampling methods. *Ann. Statist.* **12** 1543–1550.

LOH, W.-Y. (1987). Calibrating confidence coefficients. *J. Amer. Statist. Assoc.* **82** 155–162.

DEPARTMENT OF STATISTICS
UNIVERSITY OF WISCONSIN
1210 WEST DAYTON STREET
MADISON, WISCONSIN 53706

J. ROBINSON

University of Sydney

My major comment concerns the relative importance of the approximation of the critical points and of coverage error. It appears to me that much greater emphasis should be placed on the accuracy of the approximation of the bootstrap critical points to the theoretical points. The theoretical critical points based on θ should have been chosen as the best ones, in the nontechnical sense that they are thought to be better than any others available and the interval based on these has exact coverage α. Then, because we are not, in fact, able to find these critical points, we need an approximation; this can be provided by the bootstrap. Then we need to examine first the closeness of the approximating confidence interval to the theoretical one which we would use if we could. Finally, the coverage error for the approximation can be examined.

The point is more strongly made with reference to bootstrap simulations, which are not directly referred to here, the assumption being in this work that the number of simulations B, say, is very large. There is a discussion of them in Hall (1986), where it is shown that even for small B, say 19 for a 95% one-side interval, the coverage error is of the same order as if we simulated an infinite number of times. However, the accuracy of the approximation to the simulated critical point in the Studentized case is of order $n^{-1/2}B^{-1/2}$ compared to an accuracy for the infinitely resampled bootstrap of order $n^{-3/2}$. So B must be at least of size n^2 to make these approximations comparable. The reason for the accuracy of the coverage is that an averaging over all possible bootstrap simulations of size B has taken place in its calculation. Thus the particular approximation based on B simulations, compared to the real bootstrap approximation, may be gravely in error, although an average of these errors, taken over an inappropriate set, is small. I believe that in this case it is apparent that the accuracy of the coverage is much less important than the inaccuracy of the approximation to the critical points.

A rather extreme and somewhat trivial example of hypothesis testing might illustrate the preceding comments. In order to save funds, in each experiment analysed we toss a coin and if it is heads we analyse the data, whereas if it is tails we decide significance on a uniform random variable drawn independently of the data, all this being performed in a "black box" so that we do not know which was used. The test still has size α, analogous to coverage, but its power is reduced. Here and with few bootstrap simulations we have, for the sake of economy, introduced extraneous randomness not present in the experiment, whereas the "infinite" bootstrap approximation is fixed conditional on the observations. We would be unhappy explaining this test technique to the particular experimenter. Should we be any happier explaining doing only a few bootstrap simulations?

In the case of bootstrapping residuals in simple regression, a similar anomaly can be noted. Robinson (1987) shows that the error in the critical point equivalent to $\hat{y}_\alpha$ is of order n^{-1} [although I want to thank Peter Hall for pointing out that the expression for $u_U^* - u_U(\underset{\sim}{F})$ appearing there lacks a term with the same properties as the one given]; however, the coverage error is shown to be of smaller order than n^{-1}. Again considering the coverage error gives a false impression of the accuracy. It is worthwhile pointing out in passing that, in the regression case, the proofs of results based on Edgeworth and Cornish–Fisher expansions are simplified due to the possibility of approximating the conditional probabilities given the order statistics of the estimated residuals, so that it is impossible to avoid expansions of higher-order moments and the more complex results of Bhattacharya and Ghosh (1978).

None of the preceding points is a criticism of this paper, which does not attempt to discuss the merits of these two approximations; however, this lack of discussion could be interpreted as giving equal weight to the two. In fact I think the author is to be congratulated on his clarification of the several bootstrap confidence intervals and in particular I wish to thank him for the amusing, but penetrating remark that "using the percentile method amounts to looking up the wrong tables backwards."

REFERENCES

BHATTACHARYA, R. N. and GHOSH, J. K. (1978). On the validity of the formal Edgeworth expansion. *Ann. Statist.* **6** 435–451.

HALL, P. (1986). On the number of bootstrap simulations required to construct a confidence interval. *Ann. Statist.* 14 1453–1462.

ROBINSON, J. (1987). Nonparametric confidence intervals in regression: The bootstrap and randomization methods. In *New Perspectives in Theoretical and Applied Statistics* (M. L. Puri, J. P. Vilaplana and W. Wertz, eds.) 243–255. Wiley, New York.

DEPARTMENT OF MATHEMATICAL STATISTICS
UNIVERSITY OF SYDNEY
SYDNEY NSW 2006
AUSTRALIA