I thank the Editor for the opportunity to contribute to the discussion of this valuable paper.

REFERENCES

EFRON, B. (1987). Better bootstrap confidence intervals (rejoinder). *J. Amer. Statist. Assoc.* **82** 198–200.

FREEDMAN, D. A. and NAVIDI, W. C. (1986). Regression models for adjusting the 1980 census. *Statist. Sci.* **1** 3–11.

VEALL, M. R. (1987a). Bootstrapping the probability distribution of peak electricity demand. *Int. Econom. Rev.* **28** 203–212.

VEALL, M. R. (1987b). Bootstrapping forecast uncertainty: A Monte Carlo analysis. In *Time Series in Econometric Modelling* (I. B. MacNeil and G. J. Umphrey, eds.) 203–212. Reidel, Dordrecht.

DEPARTMENT OF ECONOMICS
MCMASTER UNIVERSITY
HAMILTON, ONTARIO L8S 4M4
CANADA

REJOINDER

PETER HALL

Australian National University

All the discussants make informed and penetrating comments. I am most grateful for the time they have devoted to this project. It is very interesting to see the lively debate among discussants—some of them strongly favor non-pivotal methods, others definitely like a pivotal approach. If I had to make predictions, I would say that in many years' time, when most of the dust has settled, pivotal methods (e.g., percentile-t) will tend to be favored for simple problems such as estimation of a mean, particularly when computational resources are limited, and often after appropriate transformations to stablize variance or to put the parameter space into a more useful form. Bootstrap iteration and coverage correction (e.g., the double bootstrap) may find favor as a robust, utilitarian tool, suitable for complex problems provided adequate computational resources are available. See my reply to *Beran*'s comments. The non-pivotal methods which are presently most favored by practitioners, will be largely confined to exploratory studies, highly complex problems, and certain parametric problems. I wonder how kindly time will judge these predictions!

I appreciate *Bai and Olshen*'s point that my results cannot be expected to go over automatically to random parameter models. I am fascinated by their comments following their equation (6), and look forward to seeing their forthcoming note with Bickel. Concerning their remarks about regularity conditions in their second paragraph, I must admit that things like moment assumptions did not weigh heavily on my mind while preparing my paper. I feel sure that a

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

smoothness condition and moments of order $j + 2$ are sufficient for an expansion of coverage error to order $n^{-j/2}$ in the case of the mean, say, but readily concede that our present technology is insufficient to allow verification of this suggestion!

Beran argues that both accelerated bias correction and percentile-t "do not generalize readily to confidence sets for a multidimensional parameter." While this is fine for accelerated bias correction, I believe it is false in the case of percentile-t. In fact, percentile-t generalizes immediately and straightforwardly to a higher (finite) dimensional setting. If θ and $\hat{\theta}$ are p-dimensional, $\hat{\theta}$ with asymptotic variance matrix $n^{-1}\Sigma$ (Σ nonsingular), construct a $\sqrt{n}$ consistent estimator $\hat{\Sigma}$ of Σ, and base inference on the asymptotically normal $N(0, I)$ statistic $n^{1/2}\hat{\Sigma}^{-1/2}(\hat{\theta} - \theta)$. Estimate the distribution of this quantity using $n^{1/2}\hat{\Sigma}^{*-1/2}(\hat{\theta}^* - \hat{\theta})$, where $\hat{\theta}^*$ and $\hat{\Sigma}^*$ are versions of $\hat{\theta}$ and $\hat{\Sigma}$ computed for a resample rather than the sample. Results for this multivariate problem, including second-order correctness of confidence region boundaries, generalize *easily* from the univariate case discussed in my paper; see for example Hall (1987).

In small-sample high-dimensional problems there may be difficulties with percentile-t due to the determinant of $\hat{\Sigma}$ or $\hat{\Sigma}^*$ being close to zero. However, there are rarely problems in the case of estimating a two-dimensional mean, for example, and any difficulties which do arise are not due to failure of the percentile-t *argument* to generalize to higher dimensions.

Beran argues that certain bootstrap methods may be used in infinite dimensional (nonparametric) circumstances. That may be the case, but careful attention should be paid to bias in such problems. It is not clear to me that the bootstrap adequately estimates bias in nonparametric confidence interval problems.

I agree with *Beran* that the double bootstrap is a useful tool, provided sufficient computational resources are available. However, I now view the double bootstrap somewhat differently from the way it was portrayed in the articles where it was first discussed [Hall (1986a), Beran (1987)]. In both these places, emphasis was on double bootstrapping to correct even further a confidence interval (or testing) procedure which was already rather accurate. It may be better to start with a procedure which has somewhat inferior coverage accuracy at the expense of greater stability—examples include "backwards" and "hybrid" methods discussed in the present paper. These methods produce meaningful confidence regions even in awkward problems, such as intervals for mean ratios or correlation coefficients. Use the double bootstrap to coverage-correct the basic regions, thereby producing new regions which have at once good coverage accuracy and high stability. This approach is both robust and utilitarian, and could conceivably form part of a software package for high-powered pc's.

As a final comment on *Beran*'s discussion, I should point out that there *is* a pivotal, percentile-t approach to the confidence cone problem discussed by Ducharme, Jhun, Romano and Truong (1985); see Fisher and Hall (1988).

I find *Bickel*'s argument appealing and attractive. It would be good to have some idea of when this sort of argument fails. I suspect that the answer is, when key quantities such as variances and coefficients of $n^{-1/2}$ and n^{-1} terms in Edgeworth expansions depend on quantities which cannot be estimated $\sqrt{n}$

consistently. Thus, there might be difficulties when working with quantiles, where variances involve densities evaluated at isolated points.

Of course, one of the important results to flow the *Bickel*'s discussion is that the acceleration constant quite generally admits the formula "one-sixth skewness," which I proved only under rather restrictive conditions.

The objections which *Buckland, Garthwaite and Lovell* have to the percentile-t and accelerated bias correction methods are often easily overcome by transformation. Since a smooth transformation of the smooth function model produces another smooth function model, and since most statisticians are extremely familiar with the use of transformations in the context of constructing range-preserving confidence intervals, I did not say much about transformations on the paper. However, the correlation coefficient affords an excellent example. Do not apply percentile-t directly to the sample correlation coefficient, but to Fisher's z-transformed coefficient. Having constructed the confidence interval, transform it back by untransforming its endpoints. Not only does this procedure construct intervals which do not violate the range of the parameter space, but the variance-stablizing effect of Fisher's transformation enhances the performance of percentile-t, even in a nonparametric context.

The Robbins–Monro procedure proposed by *Buckland, Garthwaite and Lovell* holds out promise for the special case where it is designed to work: parametric problems with a single unknown parameter and no nuisance parameters.

I hope my paper does not convey the impression that I claim to have "a complete theory of confidence intervals," as suggested by *DiCiccio and Romano*, for I do not. One has to draw the line somewhere and I chose to direct attention at the smooth function model, which does include many very important examples—means, variances, F-ratios, correlation coefficients, etc. I think the paper constructs the outline of an *asymptotic* theory *for the smooth function model*, but it does not claim to discuss cases outside the model which do not share key properties of the model. I had thought this was clear. Such cases include a bootstrap–Studentized quantile, in which circumstance the polynomial in the $n^{-1/2}$ term of an Edgeworth expansion is not necessarily even.

It is easy to see that in terms of estimator variance, smoothing can yield no first-order asymptotic improvement under the smooth function model, and so I believe I am justified in neglecting the issue of smoothing in my article. However, smoothing can be beneficial in *other* examples—variance estimation for the sample quantile is a case in point. It is impossible to treat all these cases in a single paper!

The TILT method discussed by *DiCiccio and Romano* promises to offer the performance of percentile-t and accelerated bias correction, without the need for algebraic calculation. This is certainly a bonus.

Efron's comments on position of critical point seem to be more a criticism of equal-tailed intervals than of percentile-t or of "looking up tables the right way." He is correct in suggesting that critical points of equal-tailed intervals are often placed in a manner which is suboptimal if minimization of interval length is important. I would argue that the right way of solving this problem is to employ a bootstrap method designed to minimize interval length; the "shortest interval" method is but one example. A technique such as accelerated bias-correction is surely susceptible to the problems which Efron ascribes to methods such as percentile-t, for like these methods, ABC is second-order correct relative to a theoretical *equal-tailed* interval.

I am particularly grateful for Wei-Yin *Loh*'s contribution, which points out that coverage accuracy of *general* confidence intervals may be improved by application of the bootstrap. If the initial interval were a bootstrap interval, than the result of Wei-Yin *Loh*'s argument is a double bootstrap interval of the type discussed by Hall (1986a) and Beran (1987); but of course the nice thing about the "calibration method" is that it is applicable quite generally, starting from an arbitrary interval.

I'm afraid I disagree with *Robinson* that "much greater emphasis should be placed on the accuracy of the approximation of the bootstrap critical points to the theoretical points." The present paper devotes considerable attention to second-order correctness, which (it points out) is equivalent to $O(n^{-1})$ coverage accuracy of one-sided intervals when endpoints of those intervals admit the usual Cornish–Fisher expansion. However, it would be very misleading to delve more deeply into the subject of critical point accuracy. In particular, third-order correctness is usually an unattainable goal and so there is little reason for discussing it.

To appreciate why second-order correctness usually cannot be bettered, observe that coefficients in the term of order $n^{-1/2}$ in an Edgeworth or Cornish–Fisher expansion are usually unknowns, such as parameters describing skewness. We know from Cramér–Rao theory that those quantities cannot be estimated with any more than $n^{-1/2}$ accuracy. Hence there is an unavoidable error term, of precise size $n^{-1/2}n^{-1/2} = n^{-1}$, in our estimate of the true quantile used to construct a confidence interval. Since that quantile is multiplied by $n^{-1/2}$ when constructing the critical point, the unavoidable error is of size $n^{-1/2}n^{-1} = n^{-3/2}$. This clearly prevents third-order correctness from being attained. Only in very special cases, such as parametric problems with known skewness, will third-order correctness be possible.

I do agree with *Robinson* that, when discussing the influence of a finite number of simulations, the selection of B can have a significant effect on accuracy of critical point (as distinct from coverage). In an earlier study [Hall (1986b)] of the effect of simulation order on coverage accuracy, I pointed out that my theoretical conclusions applied only to coverage. I stressed that those results "do not amount to a suggestion that B can be taken relatively small without penalty."

I wish I could respond to *Singh and Liu*'s request for an intuitive explanation of why "short" confidence intervals simultaneously reduce interval length and increase coverage probability. It is not true in all circumstances, and why it should hold in the important case of the mean I do not really know.

I do agree with the tenor of *Singh and Liu*'s comments concerning percentile-t and accelerated bias correction. However, it is worth reiterating comments made in my paper on the extent to which practical conclusions can be drawn from a theoretical comparison: Theoretical arguments "comprise only part of the information needed for complete evaluation of bootstrap methods," and indeed "in some situations there are practical reasons for using 'suboptimal' procedures."

Veall and some other contributors point to difficulties with percentile-t when either it is not feasible to estimate the variance or when the only variance estimate available fluctuates erratically (e.g., has large variance itself). A solution to both problems is bootstrap iteration, provided one can afford the numerical cost. That is, construct a "bad" bootstrap confidence interval, such as backwards or hybrid, use the bootstrap to estimate the true coverage of this interval for a variety of nominal coverages and then coverage-correct, selecting a nominal coverage level which makes the estimated true coverage close to the desired true coverage. Simulations with the correlation coefficient example, a notoriously bad performer using percentile-t without a transformation, show that this technique works well. It is none other than the bootstrap iteration idea discussed by Hall (1986a) and Beran (1987).

A solution to the second problem noted by *Veall* is to apply a variance stabilizing transformation, then use percentile-t and then untransform. See the previous reply to *Buckland, Garthwaite and Lovell.*

My thanks again to all the discussants.

REFERENCES

BERAN, R. (1987). Prepivoting to reduce level error of confidence sets. *Biometrika* **74** 457–468.

DUCHARME, G. R., JHUN, M., ROMANO, J. P. and TRUONG, K. N. (1985). Bootstrap confidence cones for directional data. *Biometrika* **72** 637–645.

FISHER, N. I. and HALL, P. (1988). Bootstrap confidence regions for directional data. Unpublished.

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

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

HALL, P. (1987). On the bootstrap and likelihood-based confidence regions. *Biometrika* **74** 481–493.

DEPARTMENT OF STATISTICS
AUSTRALIAN NATIONAL UNIVERSITY
GPO BOX 4
CANBERRA, ACT 2601
AUSTRALIA