The Canadian Journal of Statistics
Vol. 28, No. 3, 2000, Pages 449–499
La revue canadienne de statistique

The estimating function bootstrap[1]

Feifang HU and John D. KALBFLEISCH

Key words and phrases: Bootstrap; common means problem; confidence intervals; estimating equation; estimating function; generalized score; linear regression; nonlinear regression; resampling; significance tests.

AMS 1991 subject classifications: Primary 62G09; secondary 62G05.

ABSTRACT

The authors propose a bootstrap procedure which estimates the distribution of an estimating function by resampling its terms using bootstrap techniques. Studentized versions of this so-called estimating function (EF) bootstrap yield methods which are invariant under reparametrizations. This approach often has substantial advantage, both in computation and accuracy, over more traditional bootstrap methods and it applies to a wide class of practical problems where the data are independent but not necessarily identically distributed. The methods allow for simultaneous estimation of vector parameters and their components. The authors use simulations to compare the EF bootstrap with competing methods in several examples including the common means problem and nonlinear regression. They also prove symptotic results showing that the studentized EF bootstrap yields higher order approximations for the whole vector parameter in a wide class of problems.

RÉSUMÉ

Les auteurs proposent une procédure bootstrap qui estime la loi d'une fonction d'estimation en rééchantillonnant ses termes au moyen de techniques d'auto-amorçage. Les versions studentisées de ce bootstrap dit de la fonction d'estimation (FE) conduisent à des méthodes invariantes par reparamétrisation. Cette approche, qui s'avère souvent plus rapide et plus précise que les méthodes bootstrap traditionnelles, s'applique à de très nombreuses situations concrètes où les observations sont indépendantes mais pas nécessairement de même loi. Elle permet l'estimation simultanée de plusieurs paramètres vectoriels et de leurs composantes. Les auteurs présentent des simulations permettant de comparer le bootstrap FE à ses compétiteurs dans différents contextes, notamment celui de la régression non linéaire et du problème des moyennes communes. Ils démontrent également des résultats asymptotiques prouvant que dans beaucoup de situations, le bootstrap FE studentisé fournit une meilleure approximation du vecteur des paramètres.

1. INTRODUCTION

Since the work of Efron (1979), the bootstrap has been among the most influential recent developments in statistics. Like many basic ideas, the principle behind the bootstrap, that of resampling from the empirical distribution function, is simple yet elegant and powerful. The methods are described, e.g., in Hall (1992), Efron & Tibshirani (1993), Davison & Hinkley (1997) and DiCiccio & Efron (1996). The greatest potential value of the bootstrap lies in complex situations, such as nonlinear regression, for example. In such situations, standard inferential methods encounter serious difficulty. However, in these more complex problems, resampling from the empirical distribution function is also not generally appropriate and extending the bootstrap to non iid situations requires special arguments.

---

[1]Presented at the 28th Annual Meeting of the Statistical Society of Canada in Ottawa (Ontario) on 4 June 2000. Discussion and rejoinder follow on pp. 482–499. The Editor would like to express his gratitude to the Centre de recherches mathématiques for their generous financial support for this discussion event.

The most studied problem in the bootstrap literature is that of determining reliable confidence limits; see, for examples Efron (1987), Efron & Tibshirani (1986), DiCiccio & Romano (1988) and DiCiccio & Efron (1996). In the iid situation, basic bootstrap methods are asymptotically correct to first order and various approaches have been suggested to obtain higher order accuracy. For example, the BC_a, the classical bootstrap percentile-t method and the ABC method are asymptotically second order accurate under fairly general conditions in the iid case. The percentile-t method requires a stable estimate of the variance which is sometimes very difficult to specify, and it has been found to give unreliable results in some instances. The BC_a and ABC methods are in many ways preferable as DiCiccio and Efron (1996) discuss and document in some detail. They are intuitively, however, less appealing in that they are less transparent than the percentile-t method and seem less connected to the basic bootstrap tenet of mimicking the population sampling through sampling the empirical distribution function.

In this paper, we investigate and extend a new bootstrap method, the Estimating Function (EF) bootstrap, a precursor of which was proposed and discussed in Hu & Kalbfleisch (1997). As the name suggests, the method concentrates attention, not on the estimator as is traditional, but rather on the estimating function and equation from which the estimator is obtained. We consider situations in which the estimating function is a sum of independent terms. The EF bootstrap then proceeds by resampling the estimates of these terms and so obtaining a bootstrap estimate of the distribution of the estimating function or a studentized version of the estimating function. By using the estimating function as the basis of a testing procedure, hypotheses about parameter values can be tested and corresponding confidence intervals obtained. The methods are extended to deal with nuisance parameter problems where the inference is based on the generalized score function and its bootstrap analog.

The EF bootstrap often has substantial advantages over more traditional approaches. First, it is often computationally much simpler than classical methods. The bootstrap calculations for approximating the distribution of the estimating function are essentially those for estimating a population mean or total and so are easily applied. Especially when the parameter estimate has no closed form and must be obtained iteratively, calculations are greatly simplified. Second, there are straightforward ways of defining studentized versions of the EF bootstrap that are functionally invariant under reparametrizations and require very little additional computation. Third, the method is widely applicable as it applies immediately to cases where the estimating function is the sum of independent mean zero terms. As a consequence, it handles quite complex situations that are not easily amenable to standard approaches. And fourth, the method applies to situations where the estimating function is not smooth. This is of particular importance in certain nonparametric or robust estimating functions.

As mentioned above, the EF bootstrap was introduced in Hu & Kalbfleisch (1997) who give some examples of its use. The idea is also closely related to, and derives from, that in Hu & Zidek (1995) who discuss bootstrap methods in the context of the linear model. Other related work is found in Parzen, Wei & Ying (1994) who consider parametric bootstraps in a related context.

In Section 2, the linear estimating function is introduced, and the EF bootstrap is defined. Although there is no usual or classical approach to bootstrap estimation in the class of problems being considered here, Section 2 also gives a natural generalization, termed the C bootstrap, of an approach that involves resampling the estimator. Section 3 considers a number of examples involving a single parameter including the classical problem of estimating a common mean (Neyman & Scott 1948).

In Section 4, the multiparameter case is considered explicitly. Methods based on generalized score statistics are developed to estimate subsets or functions of the parameters, or to test hypotheses specified in various ways. The methods are exemplified in an example on binary logistic regression. In Section 5, the EF bootstrap is related to calibration of the usual normal or chi squared asymptotic score tests, and a further calibration of the EF bootstrap itself is outlined.

Section 6 considers nonlinear regression models and compares the EF bootstrap, and an iid EF bootstrap suitable for iid errors, with other methods and procedures that have been proposed. Section 7 considers application to L_q estimation of a linear regression. Section 8 concludes the paper with a number of comments and suggestions for further investigation. The appendices summarize the asymptotic results.

2. ESTIMATING EQUATIONS AND THE BOOTSTRAP

Let $y_1, \ldots, y_n$ be a sequence of independent random vectors of dimension q and $\theta \in \Omega \subset \mathbf{R}^p$ be an unknown parameter vector. For specified functions $g_i : \mathbf{R}^q \to \mathbf{R}^p$, suppose that $\mathrm{E}\{g_i(y_i, \theta)\} = 0$ for all $i = 1, \ldots, n$ and $\theta \in \Omega$. We suppose that θ is to be estimated as the solution $\hat{\theta}$ of the *unbiased linear estimating equation*

$$S(y, \theta) = n^{-1/2} \sum g_i(y_i, \theta) = 0 \tag{1}$$

(cf., e.g., Godambe & Kale 1991). Note that the normalizing constant $n^{-1/2}$ is chosen for convenience of expressing asymptotic results. For simplicity, we suppose that $S(y, \theta)$ is a one-to-one function of θ and our main consideration will be the construction of confidence regions for the whole parameter vector θ, or for components or functions of θ that are of particular interest.

The estimating equation (1) typically arises through minimization (or maximization) of some objective function,

$$SS(y, \theta) = \sum_{i=1}^{n} G_i(y_i, \theta),$$

where $g_i(y_i, \theta) = \partial G_i(y_i, \theta)/\partial\theta$. This is the case, for example, when the estimating function is a score function from some likelihood or (possibly weighted) least squares. Note that a change of parameters to λ, where $\theta = \theta(\lambda)$ is a one-to-one differentiable function, results in an equivalent estimating function for λ

$$S^{\dagger}(y, \lambda) = \left(\frac{\partial\theta}{\partial\lambda'}\right)' S(y, \theta).$$

We generally prefer methods that are invariant or nearly invariant under such changes in parameterization.

If the random vector $S(y, \theta)$ is exactly pivotal, exact methods for obtaining confidence or fiducial distributions (e.g., Buehler 1983, Parzen, Wei & Ying 1994) may be available. More usually, however, $S(y, \theta)$ is only approximately pivotal and we rely on asymptotic normal and χ^2 approximations.

There are two standard asymptotic results. First from a central limit theorem, we obtain

$$S(y, \theta) \approx \mathrm{N}_p(0, \mathcal{V}(\theta)) \tag{2}$$

and second, from a Taylor expansion, we find

$$n^{1/2}(\hat{\theta} - \theta) \approx \mathrm{N}_p\left(0, \mathcal{W}(\theta)^{-1}\mathcal{V}(\theta)\mathcal{W}(\theta)^{-1}\right). \tag{3}$$

In these expressions,

$$\mathcal{V}(\theta) = \operatorname{var} S(y, \theta) = n^{-1} \sum \operatorname{var}\{g_i(y_i, \theta)\}$$

and

$$\mathcal{W}(\theta) = \mathrm{E}\left\{n^{-1} \sum \frac{\partial g_i(y_i, \theta)}{\partial\theta'}\right\}$$

are $p \times p$ matrices.

By estimating $\mathcal{V}(\theta)$ and $\mathcal{W}(\theta)$ as necessary, we can construct approximate confidence regions for θ. Two distinct estimates of $\mathcal{V}(\theta)$ are used here. In the first, $\mathcal{V}(\theta)$ is estimated, for given θ, with

$$V(y,\theta) = n^{-1} \sum \{g_i(y_i,\theta) - \bar{g}(y,\theta)\}\{g_i(y_i,\theta) - \bar{g}(y,\theta)\}' \tag{4}$$

where $\bar{g}(y,\theta) = n^{-1} \sum g_i(y_i,\theta)$. In the second, we use the estimate

$$\widehat{V} = V(y,\hat{\theta}). \tag{5}$$

Similarly, $\mathcal{W}(\theta)$ can be estimated as

$$W(y,\theta) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial g_i(y_i,\theta)}{\partial \theta'} \tag{6}$$

or $\widehat{W} = W(y,\hat{\theta})$.

In many applications such as nonlinear regression or generalized linear models, $\mathcal{W}(\theta)$ itself is easily computed and simpler in form than (6). In these situations, either $\mathcal{W}(\theta)$ or $\mathcal{W}(\hat{\theta})$ is often used. As we shall see, use of $\mathcal{W}(\theta)$ in the generalized score statistic also leads to procedures that are invariant under reparameterization.

The result (3) with $\mathcal{V}(\theta)$ and $\mathcal{W}(\theta)$ estimated using $\widehat{V}$ and $\widehat{W}$ or $\mathcal{W}(\hat{\theta})$ is most commonly used for inference. It has distinct advantage in the simplicity of the inference about components or functions of θ. It has, however, the serious drawback that it is not functionally invariant so that different results are obtained for different parameterizations. As a consequence, it can lead to very inaccurate inferences unless parametric representation are carefully selected.

2.1. The estimating function (EF) bootstrap.

The procedure we now define is allied with (2) since it approximates the distribution of the estimating function, and $S(y,\theta)$ itself forms the basic tool for inference.

The EF bootstrap *based on* $S(y,\theta)$ *in* (1):

Let $z_i = g_i(y_i,\hat{\theta})$, $1 \le i \le n$. *From* $(z_1,\ldots,z_n)$,

1. *Draw the bootstrap sample* $(z_1^*,\ldots,z_n^*)$;
2. *Compute* $S^* = n^{-1/2} \sum z_i^*$.

   The empirical or bootstrap distribution of S^* *approximates the distribution of* $S(y,\theta)$.

This approximation is usually accurate only to first order, $O_p(n^{-1/2})$. Studentization gives an approximation of asymptotic higher order.

The Studentized EF bootstrap:

We approximate the distribution of

$$S_t(y,\theta) = V(y,\theta)^{-1/2}\, S(y,\theta), \tag{7}$$

where $V(y,\theta)$ *is defined in* (4), *with the empirical distribution of*

$$S_t^* = V^{*-1/2}\, S^*,$$

where $V^* = n^{-1} \sum (z_i^* - \bar{z}^*)(z_i^* - \bar{z}^*)'$ *and* $\bar{z}^* = n^{-1} \sum z_i^*$.

Under fairly general conditions, the approximation to the distribution of $S_t(y,\theta)$ is asymptotically second order accurate and often leads to substantial improvement in the accuracy of confidence intervals. It should be emphasised that the approximate pivotal in the formula (7) and the Studentized *EF bootstrap* procedure are invariant under reparameterization.

In the next subsection, we consider the case of scalar θ. Vector and nuisance parameters are discussed in Section 4.

2.2. The case of a single parameter.

If the parameter θ is a scalar and $S(y,\theta)$ is monotone decreasing in θ, approximate confidence intervals for θ based on the EF bootstrap are particularly simple to obtain. For any specified α, we can find $S^*_{(\alpha)}$, the α quantile of the bootstrap distribution of S^*, which can be determined to any accuracy required. The interval $(-\infty, \theta^*_{(\alpha)})$ where $S(y,\theta^*_{(\alpha)}) = S^*_{(1-\alpha)}$ is the one-sided $100\alpha\%$ EF bootstrap confidence interval for θ. To obtain this interval, the equation $S(y,\theta) = S^*$ need only be solved at one point. The asymptotic accuracy of this interval is $O_p(n^{-1/2})$.

Higher order accuracy can be obtained by using the studentized version based on (7). Let $S^*_{t(\alpha)}$ be quantiles of the distribution of S^*_t. If $S_t(y,\theta)$ is monotone in θ, solving $S_t(y,\theta) = S^*_{t(\alpha)}$ yields the endpoints of the intervals. Figure 1 illustrates the relationship between the bootstrap approximation to the distribution of S_t and the confidence intervals for θ.

In the univariate case, when $S_t(y,\theta)$ is a monotone function of θ, the bootstrap distribution of S^*_t generates a bootstrap confidence distribution for θ. Specifically, even though it would not make sense to do so in practice, we can envisage solving the equation $S_t(y,\theta) = S^*_t$ to obtain θ^*_t for each bootstrap sample. It is easy to see that the bootstrap distribution of $\theta^*_t - \hat{\theta}$ approximates the sampling distribution of $\theta - \hat{\theta}$. Even using the studentized version, this approximation is typically $O_p(n^{-1/2})$ since it is based on the Taylor approximation used to obtain (3). Nonetheless, the order $O_p(n^{-1})$ approximation inherent in the studentized EF bootstrap translates here to a statement about the confidence distribution generated by θ^*_t; specifically the α quantile, $\theta^*_{t(\alpha)}$, as discussed above gives the upper bound of a $100\alpha\%$ confidence interval for θ. In this single parameter case, the sequence of bootstrap estimates θ^*_t generates an approximate confidence distribution for θ that is accurate to order $O_p(n^{-1})$. Symmetric intervals are accurate at least to order $O_p(n^{-3/2})$. The confidence distribution is also illustrated in Figure 1.

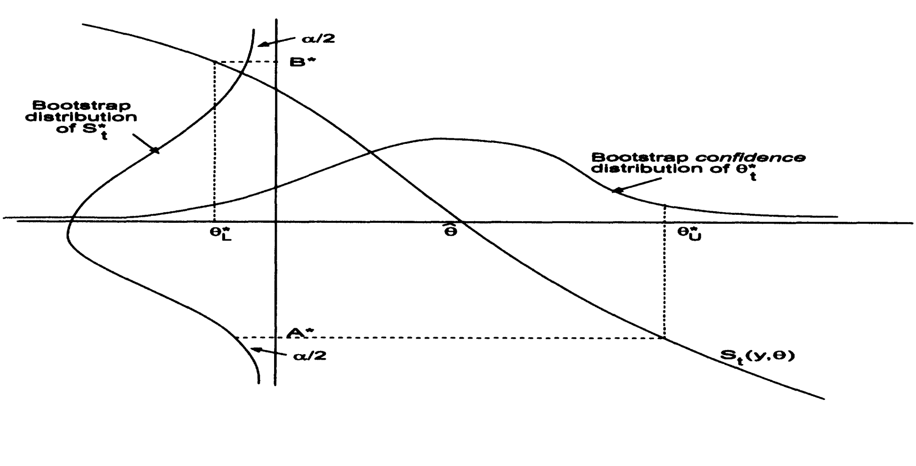


FIGURE 1: The EF-t statistic, $S_t(y,\theta)$, plotted as a function of θ. The EF bootstrap sample from S^* generates the upper and lower $\alpha/2$ percentiles A^* and B^* on the vertical axis. Confidence intervals for θ are obtained by solving $S_t(y,\theta)$ at two points to give (θ^*_L, θ^*_U).

Remark 1. This method of generating bootstrap confidence (or fiducial) distributions for θ, by solving $S_t(y,\theta) = S_t^*$ for θ_t^* would apparently extend to the multiparameter case. A conceptually simple, though computationally intensive, approach to constructing confidence intervals for components or functions of θ could be based on the corresponding marginal distributions. Such procedures would typically yield one side intervals that are accurate only to order $n^{-1/2}$. To obtain order n^{-1} accuracy, the procedure generating confidence intervals must be directly related to a test or pivotal based on $S_t(y,\theta)$. This relates to the fact that margins of a fiducial distribution are not in general confidence distributions.

Remark 2. The studentized statistic $S_t(y,\theta)$ in (7) and the related studentized EF bootstrap give results that are invariant under reparameterization. The unstudentized EF bootstrap is not invariant.

Remark 3. Hu & Kalbfleisch (1997) use a studentized statistic different from (7). Specifically, they suggest using S_t^* to approximate

$$S_{t1}(y,\theta) = \widehat{V}^{-1/2}\, S(y,\theta) \tag{8}$$

where $\widehat{V} = n^{-1}\sum_{i=1}^{n} z_i z_i'$; cf. (5). Second order results generally hold for (7) and not for (8) though in many situations, the standardization in (8) works well. It should be noted that (8) is not invariant under reparameterizations, and can break down rather badly in, e.g., highly nonlinear situations. In these instances, the variance of $S(y,\theta)$ changes rapidly as θ changes and the constant estimate $\widehat{V}$ is quite inadequate. We included (8) in almost all simulations but have not reported the results on it since methods based on (7) are generally preferable.

2.3. Bootstrap resampling of $\hat{\theta}$: the C bootstrap.

There has been no systematic discussion of bootstrap methods in the context of the linear estimating equation (1). In this section, we define an approach which focuses on the estimator $\hat{\theta}$ and seems in keeping with traditional ideas of the bootstrap.

The C bootstrap*:*

Let $w_i(\theta) = g_i(y_i,\theta), i = 1,\ldots,n$. From $\{w_1(\theta),\ldots,w_n(\theta)\}$

1. *Draw the bootstrap sample $\{w_1^*(\theta),\ldots,w_n^*(\theta)\}$. Note that the specific functions of θ are being resampled with replacement.*

2. *Define $\hat{\theta}_C^*$ as the solution to $\sum_{i=1}^{n} w_i^*(\theta) = 0$.*

The empirical distribution of $\hat{\theta}_C^* - \hat{\theta}$ approximates the distribution of $\hat{\theta} - \theta$.

This approach is akin to the asymptotic approximation in (3). Although it does not seem to have been proposed as a general technique, it does give the ordinary or classical bootstrap when the y_i are iid and $g_i = g$, $i = 1,\ldots,n$. More generally, it leads to well accepted practices in other problems such as, for example: bootstrap resampling of (y_i,σ_i), $i = 1,\ldots,n$ in the common means problem with known variances (cf. Example 2 below); or bootstrap resampling of (y_i,x_i) in the paired bootstrap procedure for linear or nonlinear regression (cf. Section 6). In these cases, C can be taken to stand for "classical." More generally, C stands for "comparison" in some of the examples below.

The C bootstrap is typically accurate to order $n^{-1/2}$ and, depending on the application, can be improved in various ways. We shall consider studentized versions based on the asymptotic approximation in (3). Specifically, we construct a *studentized C bootstrap* by considering an approximate pivotal of the form

$$\widehat{V}^{-1/2}\widehat{W}(\hat{\theta} - \theta).$$

The distribution of this can be approximated by the bootstrap distribution of

$$(V_C^*)^{-1/2} W_C^* (\hat{\theta}_C^* - \hat{\theta})$$

where $V_C^* = V(y^*, \hat{\theta}_C^*)$ and $W_C^* = W(y^*, \hat{\theta}_C^*)$; cf. Equations (4) and (6). This generally yields a better approximation although, in some instances, the variance estimates can be quite unstable. Alternative approaches might be based, e.g., on the BC_a or ABC approach as discussed, in DiCiccio & Efron (1996) although these may need some extensions to apply to the class of linear estimating equations.

Remark 4. The C bootstrap is computationally much more intensive than the EF Bootstrap whenever $\hat{\theta}$ must be calculated iteratively. With each bootstrap sample, a new estimating equation must be solved to obtain $\hat{\theta}_C^*$. In contrast, with a single parameter θ, the EF procedure only requires solving the estimating equation at the end points of the desired confidence interval.

Remark 5. Neither the C bootstrap nor the studentized C bootstrap is invariant under reparameterizations. It is important to select a good parameterization in order to get accurate intervals. Efron & Tibshirani (1993, pp. 162–166) give some discussion.

3. SOME EXAMPLES IN THE SINGLE PARAMETER CASE

Example 1 below relates to the estimation of a population mean based on an iid sample. In this classical, very simple problem, the EF bootstrap and the classical bootstrap yield identical numerical results. Nonetheless, the emphasis is different and the example serves to illustrate the different view. Examples 2 and 3 deal with the more complex and also classical problem of estimating a common mean.

Example 1 (Estimating the population mean). Observations $y_1, \ldots, y_n$ are made on independent and identically distributed random variables, each with an unspecified distribution function, F. Interest focuses on the mean, μ, which is estimated with $\hat{\mu} = \bar{y}$. In the usual (or C) bootstrap, we (i) draw the bootstrap sample $\{y_1^*, \ldots, y_n^*\}$ from $\{y_1, \ldots, y_n\}$ and (ii) calculate the bootstrap sample mean $\hat{\mu}_C^* = n^{-1} \sum y_i^*$. These steps are repeated and the empirical distribution of the $(\hat{\mu}_C^* - \hat{\mu})$ is the bootstrap approximation to the sampling distribution of $\hat{\mu} - \mu$.

In contrast, the EF bootstrap begins with the estimating equation $\sum(y_i - \mu) = 0$, whose solution is $\hat{\mu} = \bar{y}$. The component functions $y_i - \mu$ are estimated with $z_i = y_i - \bar{y}$, $i = 1, \ldots, n$. The method proceeds as follows: (i) draw a bootstrap sample $\{z_1^*, \ldots, z_n^*\}$ from $\{z_1, \ldots, z_n\}$; (ii) Calculate $S^* = n^{-1/2} \sum z_i^*$. The bootstrap distribution of S^* approximates the sampling distribution of $S(y, \mu) = \sqrt{n}\,(\hat{\mu} - \mu)$. Note that if μ^* is the solution to $S(y, \mu) = S^*$, the bootstrap distribution of $\mu^* - \hat{\mu}$ approximates the distribution of $\mu - \hat{\mu}$.

The difference between the methods is evident, even though they give, in the end, identical results. With the C bootstrap, $\hat{\mu}_C^* - \hat{\mu}$ approximates $\hat{\mu} - \mu$ whereas in the EF procedure, $\mu^* - \hat{\mu}$ approximates $\mu - \hat{\mu}$. As a consequence, μ^* is "bias corrected". The comparison between the studentized versions is similar.

Example 2 (Common mean with known variances). Suppose that $y_1, \ldots, y_n$ are from populations with $\mathrm{E}y_i = \mu$ and $\mathrm{var}(y_i) = \sigma_i^2$ with σ_i^2 known. The estimating equation

$$\sum (y_i - \mu)/\sigma_i^2 = 0$$

gives rise to the weighted least squares estimator,

$$\hat{\mu} = \left(\sum y_i/\sigma_i^2\right) / \left(\sum 1/\sigma_i^2\right).$$

The EF and C bootstraps can be applied to this problem in a straightforward way. (As noted above, the C bootstrap is equivalent to the classical procedure of resampling (y_i, σ_i), $i = 1, \ldots, n$.). If the σ_i's are not all equal, the EF and C procedures give different results here, as can be seen by directly applying the procedures above. Hu & Kalbfleisch (1997) compare the EF bootstrap with the classical bootstrap and the asymptotic normal approximation assuming normal and uniform errors. All methods do reasonably well, though the studentized versions of the EF and C bootstraps are better than the other methods with non normal errors.

Example 3 (Common means problem with unknown variances). Suppose there are k independent strata and, in the ith stratum, $y_{ij} \sim \mathsf{N}(\mu, \sigma_i^2)$, $j = 1, \ldots, n_i$, independently where $n_i \geq 3$ and $i = 1, \ldots, k$. The variances σ_i^2 are unknown and interest centers on the estimation of μ. This problem has received much attention in the literature; cf., e.g., Bartlett (1936), Neyman & Scott (1948), Kalbfleisch & Sprott (1970), Barndorff-Nielsen (1983) and Cox & Reid (1987). Neyman and Scott showed that the maximum likelihood estimator can be inefficient. They (and many others) proposed the estimating equation

$$\sum_{i=1}^{k} n_i(n_i - 2)(\bar{y}_i - \mu)/T_i(\mu) = 0 \tag{9}$$

where $T_i(\mu) = \sum_{j=1}^{n_i}(y_{ij} - \mu)^2$ and $\bar{y}_i = \sum_{j=1}^{n_i} y_{ij}/n_i$. More generally, we could relax the condition of normal errors and still use (9) for estimation of μ.

When the number k of strata is large and the individual n_i's are small, usual inferential techniques can cause substantial difficulty. This is the case considered here, though other situations are, also of interest and will be discussed elsewhere.

Let $y_i = (y_{i1}, \ldots, y_{in_i})$ and $g_i(y_i, \mu) = n_i(n_i - 2)(\bar{y}_i - \mu)/T_i(\mu)$. Thus, (9) can be rewritten

$$\sum_{i=1}^{k} g_i(y_i, \mu) = 0,$$

and EF and C bootstraps can now be applied in a straightforward manner.

We compare five methods:

1. *The Normal Approximation* (Norm1), $k^{1/2}(\hat{\mu} - \mu) \approx \mathsf{N}(0, \hat{\sigma})$, *where*

$$\hat{\sigma}^2 = \widehat{W}^{-1}\widehat{V}\widehat{W}^{-1} = k\sum_{i=1}^{k} g_i^2(y_i, \hat{\mu}) \Big/ \left\{\sum_{i=1}^{k} \frac{\partial}{\partial\mu} g_i(y_i, \hat{\mu})\right\}^2; \tag{10}$$

2. *The Normal Approximation* (Norm2), $S_t(y, \mu) \approx \mathsf{N}(0, 1)$;

3. *The C bootstrap* (C) *obtained by resampling* $w_i(\mu) = g_i(y_i, \mu)$;

4. *The Studentized C Bootstrap* (C-t) *using the variance estimator* (10);

5. *The Studentized EF Bootstrap* (EF-t).

We consider $k = 30$, $\mu = 0$, $n_1 = \cdots = n_{10} = 4$, $n_{11} = \cdots = n_{20} = 5$, $n_{21} = \cdots = n_{30} = 6$, $(\sigma_1, \ldots, \sigma_{10}) = (0.5, 0.6, \ldots, 1.4)$, $(\sigma_{11}, \ldots, \sigma_{20}) = (1.0, 1.2, \ldots, 2.8)$ and $(\sigma_{21}, \ldots, \sigma_{30}) = (0.5, 1.0, \ldots, 5.0)$. The errors were taken to be standard normal or Laplacian (p.d.f. $\exp(-|x - \mu|/2)/4$). Table 1 gives the estimated coverage probabilities and average confidence intervals based on 1000 simulations of 500 bootstrap samples.

TABLE 1: Coverage percentages and average confidence intervals (with standard deviations of the endpoints) for the common means problem (Example 3) with $k = 30$ strata. Entries are based on 1000 replications of 500 bootstrap samples.

Standard normal errors						
	80%		90%		95%	
Norm1	74	[-0.14,0.14]	81	[-0.18,0.18]	86	[-0.22,0.21]
(*s.d.*)		0.14, 0.14		0.16, 0.16		0.17,0.17
Norm2	77	[-0.15,0.15]	88	[-0.20,0.20]	95	[-0.25,0.25]
(*s.d.*)		0.13, 0.13		0.14,0.14		0.14,0.15
C	65	[-0.15,0.14]	75	[-0.19,0.18]	81	[-0.22,0.22]
(*s.d.*)		0.16, 0.16		0.17, 0.17		0.18, 0.18
C-t	72	[-0.23,0.23]	81	[-0.33,0.33]	86	[-0.44,0.44]
(*s.d.*)		1.13, 1.15		1.53, 1.50		1.86,1.89
EF-t	79	[-0.16,0.16]	89	[-0.21,0.21]	95	[-0.26,0.26]
(*s.d.*)		0.13, 0.13		0.15, 0.15		0.16, 0.16

Laplacian errors						
	80%		90%		95%	
Norm1	66	[-0.08,0.08]	76	[-0.10,0.10]	82	[-0.12,0.12]
Norm2	78	[-0.09,0.09]	89	[-0.11,0.12]	94	[-0.14,0.15]
C	66	[-0.08,0.08]	76	[-0.11,0.11]	83	[-0.12,0.13]
C-t	71	[-0.12,0.12]	80	[-0.17,0.17]	86	[-0.22,0.23]
EF-t	78	[-0.09,0.09]	90	[-0.11,0.12]	95	[-0.14,0.15]

The methods C, C-t and Norm1 all perform very poorly. In addition, the C bootstrap gives rise to substantial computational problems. Newton's method often does not converge for calculating μ_C^* and a slower bisection method is needed. The C-t bootstrap gives confidence intervals that are, on average, substantially wider than all other methods. This is because the variance estimator in (10) is not stable for small n_i's. For normal errors, the observed standard deviations of the terminals of the confidence intervals are reported in Table 1 and, for the C-t bootstrap, reflect the unstable variance estimate.

The EF-t bootstrap gives much more accurate results and, from the average lengths of the confidence intervals, appear to have good power properties. Calculation of a confidence interval require only two solutions to the estimating equations for each of the 1000 simulations. For the C bootstrap, 500 solutions are required for each simulation.

Norm2 performs very much better than Norm1. With $k = 30$, the EF bootstrap offers little gain over Norm2. To further compare EF-t with Norm2, we considered a smaller sample size with $k = 12$, $n_1 = \cdots = n_4 = 4$, $n_5 = \cdots = n_8 = 5$, $n_9 = \cdots = n_{12} = 6$ and variances similar in range to the larger sample size. The results are reported in Table 2. With the smaller sample size, EF-t does better, but the normal approximation still does reasonably well.

TABLE 2: Estimated coverage percentages for EF-t and the normal approximation (Norm2) for the common means problem of Example 3 with $k = 12$ strata. Entries are based on 1000 replications of 1000 bootstrap samples.

	80%		90%		95%	
Norm2	76.3	[-0.13,0.13]	86.7	[-0.18,0.18]	92.7	[-0.21,0.21]
EF-t	79.5	[-0.15,0.15]	89.7	[-0.19,0.20]	94.8	[-0.25,0.25]

The C and C-t bootstraps are included here for comparison purposes. This approach has not been advocated in the literature and there may be better ways to proceed, based on more traditional bootstrap methods. We do not, however, see simple alternatives. With small n_i, e.g., resampling within strata is not at all stable, since singular samples occur with relately high frequency. It is also not clear how to obtain a more stable variance estimator of $\hat{\theta}_C^*$.

4. THE MULTIPARAMETER CASE

In this section, we consider inference techniques for the joint estimation of the whole p dimensional parameter vector θ, or for estimation of components or functions of θ. We henceforth only consider studentized EF procedures since simulations we have done show substantial advantage to studentization.

4.1. Joint estimation of θ.

Consider the approximate pivotal

$$Q(y,\theta) = S(y,\theta)'V(y,\theta)^{-1}S(y,\theta) = S_t(y,\theta)'S_t(y,\theta) \tag{11}$$

which, under fairly general conditions, is asymptotically distributed as a χ^2 variate with p degrees of freedom. The distribution of $Q(y,\theta)$ can be estimated with the bootstrap distribution of

$$Q^* = S^{*\prime}V^{*-1}S^{*\prime} = S_t^{*\prime}\, S_t^*$$

using the calculations described in Section 2.

To construct a confidence region for θ, we find q_α^* to satisfy $P^*(Q^* > q_\alpha^*) = \alpha$. An approximate $100(1-\alpha)\%$ confidence region for θ is defined by

$$C_{1-\alpha}(y) = \{\theta : Q(y,\theta) \le q_\alpha^*\} \tag{12}$$

since

$$\mathrm{P}\{\theta \in C_{1-\alpha}(y)\} = \mathrm{P}\{Q(y,\theta) \le q_\alpha\} \approx P^*(Q^* \le q_\alpha) = 1-\alpha. \tag{13}$$

It is shown in the appendix that the approximation in (13) is asymptotically accurate to order $O_p(n^{-3/2})$. Computations here are again simplified by the fact that, for a given confidence coefficient $1-\alpha$, one need only solve (12) once for the relevant contour.

A test of the global hypothesis $\mathcal{H}_0 : \theta = \theta_0$ can be obtained using $Q(y,\theta_0)$ as a test statistic. The approximate significance level of the data with reference to $\mathcal{H}_0$ is

$$SL(\theta_0) = \mathrm{P}\{Q(y,\theta_0) \ge Q(y_{\mathrm{obs}},\theta_0)\} \approx P^*\{Q^* \ge Q(y_{\mathrm{obs}},\theta_0)\}$$

in an obvious notation. The confidence region $C_{1-\alpha}$ also has an interpretation as a significance interval. That is $C_{1-\alpha} = \{\theta : SL(\theta) \ge \alpha\}$.

4.2. Nuisance parameters.

Suppose that $\theta = (\theta_1', \theta_2')'$ where $\dim(\theta_1) = p_1$, $\dim(\theta_2) = p_2$ and $p_1+p_2 = p$. Suppose that θ_1 is of interest. The estimating function in (1) can be written as $S(y,\theta) = (S_1(y,\theta)', S_2(y,\theta)')'$, where S_1 is a vector of dimension p_1 and S_2 is of dimension p_2. Note that we are assuming that the estimating function arises from minimization of a certain objective function in Section 2 and so S_1 is associated with θ_1. Let $\tilde{\theta}_2(\theta_1)$ be the (assumed unique) solution in θ_2 to the equation $S_2(y,\theta_1,\theta_2) = 0$. Let $\tilde{\theta} = (\theta_1', \tilde{\theta}_2')'$. The matrix $V(y,\theta)$ is partitioned as

$$V(y,\theta) = \begin{pmatrix} V_{11}(y,\theta) & V_{12}(y,\theta) \\ V_{21}(y,\theta) & V_{22}(y,\theta) \end{pmatrix} = \begin{pmatrix} V^{11}(y,\theta) & V^{12}(y,\theta) \\ V^{21}(y,\theta) & V^{22}(y,\theta) \end{pmatrix}^{-1}$$

with similar expressions for $\widehat{V}$, V^*, $W(y,\theta)$, and $\mathcal{W}(\theta)$. Finally, we define

$$\widetilde{S}_1(y,\theta_1) = S_1(y,\tilde{\theta}).$$

As has been noted by several authors and is nicely summarized in Boos (1992), a standardized form of $\widetilde{S}_1(y,\theta_1)$ can be used as the basis of inference about θ_1. From a Taylor expansion of $S(y,\tilde{\theta})$ about θ_2, it can be seen that $\widetilde{S}_1(y,\theta_1) = S_P(y,\theta) + O_p(n^{-1/2})$, where

$$S_P(y,\theta) = S_1(y,\theta) - bS_2(y,\theta) + O_p(n^{-1/2}), \tag{14}$$

$b = \mathcal{W}_{12}(\theta)\mathcal{W}_{22}(\theta)^{-1}$, and the subscript "P" represents "projection." Thus, the asymptotic covariance matrix of $\widetilde{S}_1$ is the probability limit of

$$U_{11} = V_{11} - bV_{21} - V_{12}b' + b\,V_{22}\,b', \tag{15}$$

where $V_{ij} = V_{ij}(y,\theta)$. Now, (15) is not a useful estimate of the asymptotic variance of $\widetilde{S}_1$ since it involves θ_2. We therefore consider instead

$$\widetilde{U}_{11} = \widetilde{V}_{11} - \tilde{b}\widetilde{V}_{21} - \widetilde{V}_{12}\tilde{b}' + \tilde{b}\widetilde{V}_{22}\tilde{b}', \tag{16}$$

where V and $\mathcal{W}$ in (15) are replaced with $\widetilde{V} = V(y,\tilde{\theta})$ and $\widetilde{\mathcal{W}} = \mathcal{W}(\tilde{\theta})$. We consider the following approximate pivotals:

$$Q_{11}(y,\theta) = S_P(y,\theta)'U_{11}^{-1}S_P(y,\theta) \tag{17}$$

and

$$\widetilde{Q}_{11}(y,\theta_1) = \widetilde{S}_1(y,\theta_1)'\widetilde{U}_{11}^{-1}\widetilde{S}_1(y,\theta_1). \tag{18}$$

By standard theory, each of these has an asymptotic χ^2 distribution with p_1 degrees of freedom. The statistic (18) is sometimes referred to as the *generalized score statistic.*

We now consider a bootstrap approximation to (17) and (18). We can approximate $S_P(y,\theta)$ with the bootstrap quantity

$$S_P^* = S_1^* - b^*S_2^*,$$

where $b^* = \mathcal{W}_{12}^*\mathcal{W}_{22}^{*-1}$ and $\mathcal{W}_{12}^*$ and $\mathcal{W}_{22}^*$ are obtained as follows. Define the $p \times p$ matrices $a_i(\theta) = E\{\partial g_i(y_i,\theta)/\partial\theta'\}$. Then $\mathcal{W}^* = n^{-1}\sum_{i=1}^n a_i^*$ where $a_1^*,\ldots,a_n^*$ is the bootstrap sample of $a_1(\hat{\theta}),\ldots,a_n(\hat{\theta})$ corresponding to $z_1^*,\ldots,z_n^*$. Let $U_{11}^* = V_{11}^* - b^*V_{21}^* - V_{12}^*b^{*T} + b^*V_{22}\,b^{*T}$, and $Q_{11}^* = S_P^{*T}U_{11}^{*-1}S_P^*$ provides the required bootstrap approximations. As discussed in the appendix, Q_{11}^* is a direct approximation to Q_{11} in (17) to asymptotic order $O_p(n^{-3/2})$. In the linear model, Q_{11} depends only on θ_1 (in fact, in the linear model, $S_P(y,\theta) = \widetilde{S}_1$ and $Q_{11} = \widetilde{Q}_{11}$) and the $O_p(n^{-3/2})$ approximation applies. More generally, Q_{11} depends on θ_2 and is not directly useful for inference. But $\widetilde{Q}_{11}(y,\theta_1)$ approximates Q_{11} and its distribution is also estimated by Q_{11}^* but to asymptotic order $O_p(n^{-1})$ only.

Remark 6. The result (18), depends on the expected matrix of second derivatives, $\mathcal{W}(\theta)$, which could be replaced with $W(y,\theta)$. *We chose to use $\mathcal{W}(\theta)$ since the estimation procedures are then invariant under reparameterization.* This invariance does not in general hold with $W(y,\theta)$. The invariance of the pivotal (18) is also noted in Boos (1992). In some applications, $\mathcal{W}(\theta)$ is difficult to compute or, as in some examples with censoring, may not be obtainable at all. In those cases, $W(y,\theta)$ can be used. In other situations, such as those considered in Section 6, $\mathcal{W}$ also has the advantage of being simpler to compute.

4.3. Testing hypotheses specified by constraints.

We now consider situations in which we wish to test a hypothesis of the form $\mathcal{H}_0 : h(\theta) = 0$ where $h : \mathbf{R}^p \to \mathbf{R}^r$ is differentiable. Let

$$H(\theta) = \left(\frac{\partial h(\theta)}{\partial \theta'}\right)$$

and suppose that $H(\theta)$ is of full row rank $r < p$. This case is considered by Boos (1992), White (1982), Gallant (1987) and others who give approximate pivotals suitable for testing $\mathcal{H}_0$. Let $\tilde{\theta}$ be the constrained estimate of θ which satisfies $S(y,\theta) - n^{-1/2}H(\theta)'\lambda = 0$ and $h(\theta) = 0$, where $\lambda' = (\lambda_1, \dots, \lambda_r)$ is a vector of Lagrange multipliers.

Following arguments in White (1982), it can be shown that

$$\widetilde{S} = H'(H\mathcal{W}^{-1}H')^{-1}H\mathcal{W}^{-1}S + O_p(n^{-1/2}), \tag{19}$$

where $\widetilde{S} = S(y,\tilde{\theta})$, $H = H(\theta)$, $S = S(y,\theta)$ and $\mathcal{W} = \mathcal{W}(\theta)$. It follows that the asymptotic variance of $\widetilde{S}$ is the probability limit of

$$U = H'(H\mathcal{W}^{-1}H')^{-1}H\mathcal{W}^{-1}V\mathcal{W}^{-1}H'(H\mathcal{W}^{-1}H')^{-1}H, \tag{20}$$

where again the arguments y and θ are suppressed. It can be verified that

$$U^- = \mathcal{W}^{-1}H'(H\mathcal{W}^{-1}V\mathcal{W}^{-1}H')^{-1}H\mathcal{W}^{-1} \tag{21}$$

is the (Moore–Penrose) generalized inverse of U so that

$$Q_{h=0} = S(y,\theta)'U^-(y,\theta)S(y,\theta) \tag{22}$$

is asymptotically chi-squared with r degrees of freedom. Since Q generally involves all of θ, it is not useful for inference. However, the relationship (19) gives the statistic

$$\widetilde{Q}_{h=0} = S(y,\tilde{\theta})'U^-(y,\tilde{\theta})S(y,\tilde{\theta}) \tag{23}$$

where W, H and V in U^- have been replaced with $\widetilde{\mathcal{W}}$, $\widetilde{H}$, and $\widetilde{V}$. It is easy to see that $\widetilde{Q}_{h=0} = Q_{h=0} + O_p(n^{-1})$ is also asymptotically chi-squared and can be used to test $\mathcal{H}_0$. Note that equations (22) and (23) are analogous to (17) and (18) respectively.

A bootstrap approximation to (22) is given by

$$Q^*_{h=0} = S^{*T}U^{*-}S^*, \tag{24}$$

where U^{*-} is obtained by replacing $\mathcal{W}$ and V in U^- with their bootstrap analogs and replacing $H(\theta)$ with $H(\hat{\theta})$.

It can be verified that the results in the previous section arise as the special case $h(\theta) = \theta_1 - \theta_1^0$. It then follows that $H = (I_r \mid 0)$, where I_r is the $r \times r$ identity matrix. Substitution shows, for example, that (19) reduces exactly to (14), and $\widetilde{Q}_{h=0}$ reduces to $\widetilde{Q}_{11}$. Like $\widetilde{Q}_{11}$, $\widetilde{Q}_{h=0}$ is invariant under reparameterizations of θ.

The above results can also be extended to the estimation of a parameter $\gamma = h(\theta)$. In this case, we can consider testing the hypothesis $\gamma = \gamma_0$ or equivalently $\mathcal{H}_0 : h(\theta) - \gamma_0 = 0$. It follows that $H(\theta)$ and $H(\hat{\theta})$ are independent of γ_0. The bootstrap distribution of $Q^*_{h=0}$ is, therefore, unaffected by the value of γ_0 and confidence regions for γ can be obtained by solving the inequality

$$\widetilde{Q}_{h=\gamma} = \widetilde{S}(y,\gamma)U^-(y,\gamma)\widetilde{S}(y,\gamma) \le q^*_\alpha \tag{25}$$

in an obvious notation. Here, q^*_α is the upper α quantile of the bootstrap distribution of Q^* and $\tilde{\theta}(\gamma)$ has been replaced with γ.

Example 4 (Application to binary logistic regression). Consider a logistic regression model for a binary response Y, $\mathrm{P}(Y_i = 1 \mid x_i) = \pi_i$ where $\mathrm{logit}(\pi_i) = x_i'\theta$, $i = 1, \ldots, n$. In this, θ is a column vector of p regression parameters and x_i is a vector of constants with $x_{i1} = 1$. The corresponding estimating or maximum likelihood equation is

$$S(y, \theta) = \sum_{i=1}^{n} x_i(y_i - \pi_i).$$

We find that $V(y, \theta) = \sum x_i x_i'(y_i - \pi_i)^2$ and $\mathcal{W}(\theta) = \sum x_i x_i' \pi_i(1 - \pi_i)$. Hypotheses about the π_i's or about the θ's can now be evaluated using the results of this section.

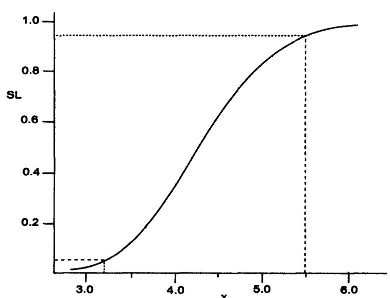
DiCiccio & Efron (1996, Table 4, p. 198) consider an example of logistic regression applied to a study on 1843 independent cell cultures. The response ($y = 1$ or 0) was the success or failure of the culture and there were two factors, each on five levels. The model was additive on the logit scale so that $\mathrm{logit}(\pi_{ij}) = \mu + \alpha_i + \beta_j$, $i, j = 1, \ldots, 5$, where, π_{ij} is the probability of a successful culture at levels i and j of the factors. For identifiability, we assume $\alpha_5 = \beta_5 = 0$. In the context of the preceding paragraph, we have $p = 9$ with $\theta = (\mu, \alpha_1, \ldots, \alpha_4, \beta_1, \ldots, \beta_4)$ and x_i, $i = 1, \ldots, n$ specifying the design matrix. DiCiccio and Efron consider the estimation of the parameter $\gamma = \pi_{15}/\pi_{51}$. The methods are invariant under reparameterizations, and it is simplest to work with $h(\theta) = \log(\pi_{15}/\pi_{51}) - \log \gamma_0$. Straightforward calculation gives the corresponding Jacobian matrix, $H(\theta)$. The signed square root of (25),

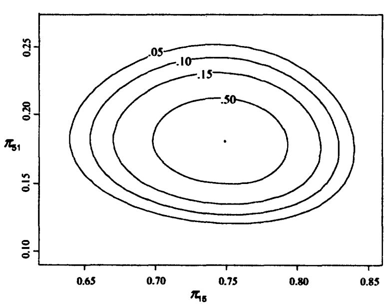
$$\widetilde{S}_{h=\gamma} = \mathrm{sign}\{\widetilde{S}(y, \gamma)\}\widetilde{Q}_{h=\gamma}^{1/2},$$

can be used for inference about γ.

A bootstrap sample of $S^*_{h=0}$, the signed square root of $Q^*_{h=0}$ from (24) based on 100,000 replications gave the critical values, $q^*_\alpha = (-1.908, -1.614, 1.681, 2.009)$ for $\alpha =$ .025, .05, .95, .975. (Note that the sign is that of $\widetilde{S}^*$ obtained from (19) using $\widehat{H}$, $\mathcal{W}^*$, and S^*.) The corresponding 90% confidence interval for γ is (3.20,5.48) in very good agreement with the ABC interval given by DiCiccio and Efron of (3.20,5.43). The standard Wald type interval is symmetric about the mle $\hat{\gamma} = 4.16$ and yields (3.06,5.26), so both the ABC and the EF bootstrap are making similar corrections. The interval based on the normal approximation to $\widetilde{S}_{h=\gamma}$ is (3.20,5.54). Corrections here are relatively small since the sample size (n=1843) is so large.

Figure 2 displays a plot of $SL(\gamma) = P^*(S^*_{h=0} \leq \widetilde{S}_h = \gamma)$. This can be interpreted as the confidence distribution function for γ and from it, all approximate confidence intervals are easily determined.





FIGURES 2 AND 3: On the left, the one-sided significance level from a test of γ in Example 4. This can be interpreted as a confidence distribution; the 90% confidence interval is indicated. On the right, approximate confidence regions from the EF bootstrap in Example 4. In this example, essentially the same contours arise from the chi-square approximation.

We took a very large bootstrap sample here in order to get a very accurate estimate of the bootstrap quantiles. One advantage of the EF approach is that large bootstrap samples can be easily taken. Even so, a bootstrap sample of 100,000 is much larger than would typically be needed for most practical purposes; 10,000 is more than adequate and can be quickly obtained.

We can also consider joint estimation of $\lambda = (\lambda_1, \lambda_2) = (\pi_{15}, \pi_{51})$. The Jacobian matrix $H = \partial\lambda/\partial\theta'$ is easily computed as is the test statistic (25). The corresponding bootstrap distribution is obtained from (24) and for $\alpha = (.50,\ .80,\ ,90,\ .95)$ yields critical values $q^*_\alpha = (1.384,\ 3.217,\ 4.609,\ 5.990)$ again from 100,000 bootstrap replications. The corresponding EF bootstrap confidence regions from (25) are displayed in Figure 3.

The bootstrap critical points here are essentially those of the $\chi^2_{(2)}$ distribution. A more interesting example can be obtained by considering a smaller sample size. For example, if the frequencies in this study are divided by 10, the EF bootstrap method makes a substantial adjustment from the asymptotic chi-squared approximation. The calculation of the contours for the joint estimation of (π_{15}, π_{51}) is, however, no different than that illustrated above for the larger sample size.

In this example, a strong case can be made for stratifying and resampling the terms in the estimating function holding constant the number n_{ij} of samples in each of the 25 treatment combinations. In this example, this constraint yields results that are essentially identical to those obtained without constraining the resampling as above. The question of constrained resampling needs further consideration. Some additional comments are made in Section 8.

4.4. Hypotheses that are parametrically specified.

In some problems, the hypothesis of interest is most naturally described in terms of a new set of parameters, η, say. Thus, the hypothesis is $\mathcal{H}_0 : \theta = k(\eta)$ where η is of dimension $p - r$ and $k : \mathbf{R}^{p-r} \to \mathbf{R}^p$. Goodness-of-fit tests, e.g., are typically formulated in this way where θ is the parameter of the multinomial model and the hypothesis corresponds to a parametric model with parameters η.

Let $K(\eta) = \partial k(\eta)/\partial\eta'$ and suppose that K is of full column rank $p - r$. In this case, the generalized score statistic for testing $\mathcal{H}_0$ is given by Boos (1992) as

$$\widetilde{Q}_{\theta=k(\eta)} = \widetilde{S}'\{\widetilde{V}^{-1} - \widetilde{V}^{-1}\widetilde{W}\widetilde{K}(\widetilde{K}'\widetilde{W}\widetilde{V}^{-1}\widetilde{W}\widetilde{K})^{-1}\widetilde{K}'\widetilde{W}\widetilde{V}^{-1}\}\widetilde{S}. \tag{26}$$

Note that the notation here follows the conventions of the last section with "tilde" indicating that the matrix is evaluated at the estimate subject to the hypothesis, i.e., at $\widetilde{\theta} = k(\hat{\eta})$.

Boos (1992) notes that (26) can be derived from the previous result (23) if there exists an alternative and equivalent representation of the hypothesis of the form $h(\theta) = 0$ where $h : \mathbf{R}^p \to \mathbf{R}^r$. From this it follows that $HK = 0$. Gallant (1987, p. 241) gives the identity

$$H'(HAH')^{-1}H = A^{-1} - A^{-1}K(K'A^{-1}K)^{-1}K'A^{-1} \tag{27}$$

which holds for arbitrary positive definite symmetric $A_{p\times p}$, $H_{r\times p}$ of full rank r, and $K_{p\times p-r}$ of rank $p - r$. Substitution of this into (19) gives the result

$$\widetilde{S} = \{I - WK(K'W\,K)^{-1}K'\}S + O_p(n^{-1/2})$$

which provides the needed link between $\widetilde{S}$ and S for a bootstrap procedure. The remaining results for parametrically specified hypotheses can be found by substituting (27) into (20) and (21).

Construction of the corresponding bootstrap procedure is straightforward. Applications of this to goodness-of-fit and other problems where parametric hypotheses naturally arise is the subject of further investigation.

5. CALIBRATION AND THE EF BOOTSTRAP

Calibration of approximate confidence procedures has been discussed by various authors. Hall (1986, 1987), Beran (1987) and Loh (1987, 1991) first discussed the procedures as applied to the bootstrap. See also Efron & Tibshirani (1993, Chapter 18) and DiCiccio & Efron (1996). For simplicity of presentation, we restrict discussion to the case of a scalar parameter.

The studentized EF bootstrap can be viewed as a calibration of the asymptotic normal approximation applied to the studentized statistic

$$S_t(y,\theta) = V(y,\theta)^{-1/2} S(y,\theta).$$

To make specific contact with the discussion in DiCiccio & Efron (1996), we note that the approximate (one-sided) $100\alpha\%$ confidence interval from the normal approximation is

$$\{\theta : S_t(y,\theta) > z_{(\alpha)}\} = \{\theta < \hat{\theta}_{(\alpha)}\}$$

where $z_{(\alpha)}$ is the standard normal quantile and it is assumed that $S_t(y,\theta)$ is monotone decreasing in θ for each y. The *calibration function*, $\beta(\alpha) = P(\theta < \hat{\theta}_{(\alpha)})$, can be estimated using the bootstrap approximation

$$\hat{\beta}(\alpha) = P^*(\hat{\theta} < \hat{\theta}^*_{(\alpha)}) = P^*(V^{*-1/2} S^* < z_{(\alpha)}).$$

This estimate can be used to obtain a corrected interval by finding λ such that $\alpha = \hat{\beta}(\lambda)$ for the given α, and then using the corresponding interval,

$$\{\theta : S_t(y,\theta) < z_{(\lambda)}\} = \{\theta : S_t(y,\theta) < q^*_\alpha\}.$$

This is the EF bootstrap interval since, by construction, $z_{(\lambda)} = q^*_\alpha$ is the $100\alpha\%$ percentile point of the bootstrap distribution of S^*_t.

In some instances, a further calibration may be useful Let $z^*_1, \ldots, z^*_n$ be a first level EF bootstrap sample as discussed in Section 2. Let $w_i(\theta) = g_i(y_i, \theta)$, $i = 1, \ldots, n$ and note that we can write $z^*_i = w^*_i(\hat{\theta})$. We proceed by finding $\hat{\theta}^*_C$ to satisfy

$$\sum w^*_i(\hat{\theta}^*_C) = 0.$$

Let $\hat{z}^*_i = w^*_i(\hat{\theta}^*_C)$. The second stage now proceeds exactly as the first: we sample $z^{**}_1, \ldots, z^{**}_n$ from $\hat{z}^*_1, \ldots, \hat{z}^*_n$ and, in an obvious notation, approximate the calibration curve $\beta^*(\alpha) = P(S_t < q^*_\alpha)$ with $\hat{\beta}^*(\alpha) = P^*(S^*_t < q^{**}_\alpha)$. Now $\hat{\beta}^*(\lambda)$ can be obtained at a grid of λ values from a suitably large number of bootstrap samples for each of the first stage samples.

This procedure is computationally intensive since it requires perhaps 1000 replications of each of 1000 first level bootstraps. At the second stage, we need to solve for the C bootstrap estimators $\hat{\theta}^*_C$. Thus we must solve the estimating equation 1001 times. This is, however, still within the bounds of reasonable computation in many problems and could be routinely done. By contrast, a 1000×1000 calibration of the C bootstrap would require the solution of about 1,000,000 equations. Simulation assessment of either calibrated procedure requires yet another level of replications.

6. NONLINEAR MODELS

In Section 6.1, we consider univariate nonlinear models and ordinary least squares, and the various methods are compared in simulations for scalar and vector parameter in Section 6.2. Section 6.3 considers extensions to multivariate nonlinear regression and weighted least squares.

6.1. Univariate nonlinear models.

Let y_i be a univariate response variable and consider the model

$$y_i = f(x_i, \beta) + \varepsilon_i, \quad i = 1, \ldots, n. \tag{28}$$

Here, β is a $p \times 1$ vector of unknown parameters, f is a known nonlinear function of β, x_i is a $k \times 1$ vector of constants, and the ε_i are independent errors with mean zero and variance σ_i^2. We suppose that the σ_i^2 are bounded away from 0 and ∞, and that

$$A(x_i, \beta) = \frac{\partial f(x_i, \beta)}{\partial \beta}$$

is continuous in β. The ordinary least squares estimating equation is

$$S(y, \beta) = n^{-1/2} \sum_{i=1}^{n} A(x_i, \beta)\{y_i - f(x_i, \beta)\} = 0,$$

and the corresponding estimator is $\hat{\beta}_{LS}$.

Following the outline above, we find

$$V(y, \beta) = n^{-1} \sum_{i=1}^{n} \left[\{y_i - f(x_i, \beta)\}^2 A(x_i, \beta) A(x_i, \beta)' - S(y, \beta) S(y, \beta)'\right].$$

The matrix of second partial derivatives is

$$W(y, \beta) = -n^{-1} \sum_{i=1}^{n} \left[A(x_i, \beta) A(x_i, \beta)' - \frac{\partial A(x_i, \beta)}{\partial \beta'} \{y_i - f(x_i, \beta)\} \right],$$

and $\mathcal{W}(\beta) = \mathrm{E}\{W(y, \beta)\}$ has the simpler form

$$\mathcal{W}(\beta) = -n^{-1} \sum_{i=1}^{n} A(x_i, \beta) A(x_i, \beta)'.$$

Let $\widehat{V} = V(y, \hat{\beta}_{LS})$ and $\widehat{\mathcal{W}} = \mathcal{W}(\hat{\beta}_{LS})$.

Consider inference about $c'\beta$, where c is a $p \times r$ matrix of known constants. All of the bootstrap procedures described in Sections 2 and 4 can be applied. In addition, there are more efficient approaches which can be applied when the errors are iid.

1. Classical bootstrap based on residuals iid C-t. Huet & Jolivet (1989) and Huet, Jolivet & Messéan (1990) suggest this approach for inference in the nonlinear model with iid errors; cf. also Efron (1979) for a parallel discussion of the linear case. Let $r_i = y_i - f(x_i, \hat{\beta}_{LS})$, $i = 1, \ldots, n$ and $\bar{r} = \sum r_i / n$. Let $e_1^*, \ldots, e_n^*$ be a bootstrap sample from $r_1 - \bar{r}, \ldots, r_n - \bar{r}$ and let

$$y_i^* = f(x_i, \hat{\beta}_{LS}) + \sqrt{\frac{n}{n-p}}\, e_i^*.$$

Note that $\sqrt{n/(n-p)}$ is an adjustment for the degrees of freedom. The bootstrap estimate $\hat{\beta}_{LS}^*$ is the solution to

$$\sum_{i=1}^{n} A(x_i, \beta)\{y_i^* - f(x_i, \beta)\} = 0.$$

The corresponding studentized bootstrap is obtained as follows: the asymptotic covariance matrix of $n^{1/2} c'(\hat{\beta} - \beta)$ is estimated by $c' \widehat{\mathcal{W}}^{-1} \widehat{V}_{\text{iid}} \widehat{\mathcal{W}}^{-1}\ c$ where $\widehat{V}_{\text{iid}} = \widehat{\mathcal{W}} \hat{\sigma}$ and $\hat{\sigma} = n^{-1} \sum (r_i - \bar{r})^2$. An appropriate studentized statistic is then

$$T = n^{1/2} \hat{\sigma}^{-1} (c' \widehat{\mathcal{W}}^{-1} c)^{-1/2} \{c'(\hat{\beta}_{LS} - \beta)\}$$

whose distribution can be approximated by that of its bootstrap analog, viz.

$$T^* = n^{1/2}\hat{\sigma}^{*-1}\{c'\mathcal{W}(\hat{\beta}^*_{LS})c\}^{-1/2}\{c'(\hat{\beta}^*_{LS} - \hat{\beta}_{LS})\},$$

where $\hat{\sigma}^{*2} = n^{-1}\sum(r^*_i - \bar{r}^*)^2$, $r^*_i = y^*_i - f(x_i, \hat{\beta}^*_{LS})$, and $\bar{r}^* = \sum r^*_i/n$.

2. Paired or C bootstrap (Paired-t). When the σ_i's are not equal, the above method is not consistent. We can, however, resample pairs (x^*_i, y^*_i), $i = 1, \ldots, n$, from (x_i, y_i), $i = 1, \ldots, n$. This is analogous to the paired bootstrap in the linear case discussed, e.g., in Freedman (1981) and Hinkley (1988), and is exactly equivalent to the C bootstrap of Section 2. The corresponding paired bootstrap estimator, $\hat{\beta}^*_P$, is a solution to

$$\sum_{i=1}^{n} A(x^*_i, \beta)\{y^*_i - f(x^*_i, \beta)\} = 0.$$

The studentized paired statistic is

$$T_P = n^{1/2}(c'\widehat{\mathcal{W}}^{-1}\hat{V}\widehat{\mathcal{W}}^{-1}c)^{-1/2}\{c'(\hat{\beta}_{LS} - \beta)\},$$

and the corresponding studentized bootstrap statistic is

$$T^*_P = n^{1/2}(c'\widehat{\mathcal{W}}^{*-1}_P\hat{V}^*_P\widehat{\mathcal{W}}^{*-1}_P c)^{-1/2}\{c'(\hat{\beta}^*_P - \hat{\beta}_{LS})\},$$

where $\widehat{\mathcal{W}}^*_P$ and $\hat{V}^*_P$ are exactly analogous to $\hat{V}$ and $\widehat{\mathcal{W}}$ except computed on the bootstrap sample, (y^*_i, x^*_i), $i = 1, \ldots, n$.

3. Studentized EF bootstrap (EF-t). Inference about $\gamma = c'\beta$ can be based on the results in Section 4.3. Specifically, consider a hypothesis of the form $h(\beta) = c'\beta = \gamma_0$ where γ_0 is a hypothesized value of γ. The relevant statistic from (23) is

$$\widetilde{Q}_{h=\gamma} = S(y, \tilde{\beta})'\widetilde{U}^-(y, \tilde{\beta})S(y, \tilde{\beta}),$$

where $\tilde{\beta} = \tilde{\beta}(\gamma_0)$ is the estimate of β constrained by the hypothesis $c'\beta = \gamma_0$ and $\widetilde{U}^-$ is given in (21) with $\widetilde{\mathcal{W}} = \mathcal{W}(\tilde{\beta})$, $H = c'$ and $\tilde{V} = V(y, \tilde{\beta})$. The analogous bootstrap statistic is obtainable directly from (24) and inference proceeds as described in (25). If $c'\beta = \beta_1$ where β_1 comprises the first r components of β, the simpler and equivalent expression (18) from Section 4.2 can be used. This *EF-t* procedure, like the paired bootstrap is valid for the heteroscedastic case.

4. EF bootstrap based on iid residuals (iid EF-t). When the ε_i's are iid, the asymptotic covariance matrix of $S(y, \tilde{\beta})$ is estimated more efficiently by replacing the variance estimate $V(y, \tilde{\beta})$ with

$$V_{\text{iid}}(y, \tilde{\beta}) = \tilde{\sigma}^2\mathcal{W}(\tilde{\beta}),$$

where $\tilde{\sigma}^2 = n^{-1}\sum(y_i - f_i - \bar{y} + \bar{f})^2$, $f_i = f(x_i, \tilde{\beta})$ and $\bar{f} = \sum f_i/n$. Thus, with reference to (23) and (21), we use the studentized statistic

$$\widetilde{Q}^{\text{iid}}_{h=0} = S(y, \tilde{\beta})'\widetilde{U}^-_{\text{iid}}S(y, \tilde{\beta}), \tag{29}$$

where

$$\widetilde{U}^-_{\text{iid}} = \tilde{\sigma}^{-2}\mathcal{W}(\tilde{\beta})^{-1}c\{c'\mathcal{W}(\tilde{\beta})^{-1}c\}^{-1}c'\mathcal{W}(\tilde{\beta})^{-1}.$$

As for the iid C bootstrap above, the iid EF bootstrap proceeds by resampling the residuals. Let $e^*_1, \ldots, e^*_n$ be a bootstrap sample from $(r_1 - \bar{r}), \ldots, (r_n - \bar{r})$ and let

$$S^*_{\text{iid}} = (n - p)^{-1/2}\sum_{i=1}^{n} A(x_i, \hat{\beta}_{LS})e^*_i.$$

The bootstrap statistic corresponding to (29) is

$$Q_{h=0}^{\text{iid}*} = S_{\text{iid}}^{*T} U_{\text{iid}}^{*-} S_{\text{iid}}^{*},$$

where

$$U_{\text{iid}}^{*-} = \sigma^{*-2}\widehat{\mathcal{W}}^{-1}c(c'\widehat{\mathcal{W}}^{-1}c)^{-1}c'\widehat{\mathcal{W}}^{-1}$$

and $\sigma^{*2} = (n-p)^{-1}\sum(e_i^* - \bar{e}^*)^2$.

If $c'\beta = \beta_1$, it can be verified that the statistic (29) reduces to

$$\widetilde{Q}_{11}^{\text{iid}} = \tilde{\sigma}^{-2}S_1(y,\tilde{\beta})'\widetilde{\mathcal{W}}^{11}S_1(y,\tilde{\beta}),$$

and the corresponding bootstrap statistic is

$$Q_{11}^{\text{iid}*} = \sigma^{*-2}\widetilde{S}_{\text{iid}1}^{*T}\widehat{\mathcal{W}}^{11}\widetilde{S}_{\text{iid}1}^{*}$$

where $\widetilde{S}_{\text{iid}1}^{*} = S_{\text{iid}1}^{*} - \hat{b}S_{\text{iid}2}^{*}$, and $\hat{b} = \widehat{\mathcal{W}}_{12}\widehat{\mathcal{W}}_{22}^{-1}$.

Remark 7. In the discussion of the iid EF bootstrap (Point 4. above) corrections of degrees of freedom are made in two places: in the definition of S_{iid}^{*} and in the definition of σ^{*2}. These two corrections are in opposite directions and cancel. Similar corrections could have been introduced in the studentized EF bootstrap and also would have cancelled. The corrections were included in the discussion of the iid EF bootstrap to keep a parallel with the iid C bootstrap as defined in Huet, Jolivet & Messéan (1990) or Efron (1979).

Remark 8. The paired and iid C bootstraps involve solving a system of nonlinear equations for every bootstrap sample and are computationally extremely intensive. The procedures based on EF and iid EF retain the substantial advantage of requiring solutions to the equation only at the final stage.

TABLE 3: Coverage percentages and average confidence intervals (with standard derivations of the endpoints) for Example 5. Entries are based on 1000 replications of 1000 bootstrap samples.

	Homoscedastic errors		Heteroscedastic errors	
	80%	95%	80%	95%
Norm1	**75.7**	**90.8**	**72.0**	**88.7**
Ave CI	[0.011,0.179]	[-0.033,0.223]	[-0.003,0.216]	[-0.06, 0.274]
(*s.d.*)	0.075, 0.068	0.080, 0.068	0.087,0.11	0.087,0.11
Norm2	**76.2**	**92.3**	**74.5**	**93.1**
Ave CI	[0.014,0.183]	[-0.035,0.220]	[-0.011,0.215]	[-0.048,0.288]
(*s.d.*)	0.074, 0.069	0.079, 0.071	0.081,0.103	0.083,0.111
iid C-t	**80.7**	**95.7**	**71.4**	**85.9**
Ave CI	[-0.001,0.185]	[-0.064,0.238]	[0.003,0.212]	[-0.061,0.271]
(*s.d.*)	0.075, 0.067	0.081,0.066	0.094,0.098	0.097,0.101
Paired-t	**80.5**	**95.2**	**77.0**	**91.5**
Ave CI	[-0.001,0.186]	[-0.067,0.243]	[-0.026,0.247]	[-0.13,0.339]
(*s.d.*)	0.076, 0.068	0.088,0.068	0.11,0.16	0.11,0.18
EF-t	**79.4**	**94.2**	**76.8**	**93.5**
Ave CI	[0.007,0.183]	[-0.050,0.235]	[-0.019,0.225]	[-0.068,0.293]
(*s.d.*)	0.076, 0.070	0.082,0.079	0.090,0.111	0.094,0.124
iid EF-t	**79.9**	**94.5**	**71.4**	**89.7**
Ave CI	[0.006,0.176]	[-0.044,0.221]	[-0.003,0.200]	[-0.061,0.245]
(*s.d.*)	0.073, 0.066	0.078, 0.068	0.084,0.097	0.082,0.102

Remark 9. Linear models ($f(x_i, \beta) = x_i'\beta$) are a special case of the above and the same four methods can be applied. In the linear model, the iid C bootstrap method above was proposed by Efron (1979) and the paired or C method is also often discussed. In the context of the linear model, Hu & Zidek (1995) considered the (unstudentized) EF and iid EF bootstraps and explored normality and robustness properties.

Remark 10. To summarize aspects of asymptotic properties, both *EF-t* and *iid EF-t* are generally accurate to first order for the estimation of components or linear functions of β and are second order accurate in the linear model. They are also second order accurate for estimation of the whole parameter vector in general. For the nonlinear regression model, Huet & Jolivet (1989) show that the iid C-t bootstrap is second order accurate for the homoscedastic case. Both it and the iid EF bootstrap are generally inconsistent if the errors are heteroscedastic.

6.2. Some examples in the nonlinear model.

Example 5 (A one parameter nonlinear model). We consider first the model

$$y_i = \exp(x_i\beta) + \varepsilon_i, \quad i = 1, \ldots, n,$$

for which computations and comparisons among all the methods can be relatively easily done. In Table 3, we report two simulations. For the first, the errors ε_i are iid $\mathsf{N}(0, 0.25)$. For the second, the errors are $\varepsilon_i = a_i u_i$ where $u_i \sim \mathsf{N}(0, 1)$ and $a_i = (0.05)i$, $i = 1, \ldots, 20$. For both, we chose $\beta = .1$, $n = 20$ and $x_i = -2, -1.9, \ldots, -1.1, 1.1, \ldots, 2$. The results in Table 3 are based on 1000 simulations of 1000 bootstrap samples. In addition to the four bootstrap procedures (iid C-t, Paired-t, EF-t, iid EF-t), we also include normal approximations, Norm1 based on (3), and Norm2 based on (7). Other simulations using non normal errors also give similar results.

For the homoscedastic case, all methods work reasonably well, though the Norm1 approximation is less accurate than the others. The iid EF-t procedure gives the best results, though the EF-t also performs well. In the homoscedastic case, the iid EF-t bootstrap is, as expected, somewhat more efficient.

In the heteroscedastic case, only the EF-t, Paired-t and Norm2 methods are consistent and the iid C and iid EF procedures perform less well. The paired, EF and Norm2 give the most accurate coverages, although the coverage probabilities are all somewhat low. A further calibration could be done here as in the next example.

The next example was also investigated in simulations by Huet, Jolivet & Messéan (1990). We compare the EF-t bootstrap, the asymptotic approximations applied to the estimating function (Norm2 or χ^2app2) and to the least squares estimate (Norm1 or χ^2app1, and the iid EF-t bootstrap. We have not included the Paired or iid C bootstraps. The iid C bootstrap was the subject of the investigation of Huet, Jolivet & Messéan (1990) who essentially conclude that the normal approximation (Norm1) works as well as iid C-t. We have included Norm1 in our simulations.

Example 6 (The exponential model). This model, an extension of that in Example 5, has been considered by several authors (cf. Seber & Wild 1988 for examples and references). Rasch & Schimke (1983) and Ratkowsky (1984), as well as Huet, Jolivet & Messéan (1990) have studied various bootstrap procedures and asymptotic approximations in simulations. The particular model examined has mean function

$$f(x, \beta) = \beta_1 + \beta_2 \exp(\beta_3\, x) \tag{30}$$

with (approximate) true values $\beta_1 = 1122$, $\beta_2 = -1309$, $\beta_3 = -0.087$. The independent variable x takes values $x_i = i$, $i = 1, \ldots, 15$. We consider one, two and three replicates of this for total sample sizes of 15, 30 or 45. Errors were taken to have mean zero and were either assumed to be homoscedastic with variance 40^2, or heteroscedastic with variance at $x_i = i$ of

$40^2(i/7.5)^2$. Error distributions considered were the normal, Laplacian and uniform. These parameter values were chosen to correspond to those used in Huet, Jolivet & Messéan (1990) for the homoscedastic case. They did not consider the heteroscedastic case.

TABLE 4: Estimated size and powers for tests ($\times 1000$) of a hypothesis about β_3 in the nonlinear exponential model of Example 6. The null (true) value is $-.0875$ and alternatives are considered at $\pm.005$. Sample sizes are $n = 15$ and $n = 30$ with homoscedastic double exponential errors ($\sigma = 40$) and normal heteroscedastic errors.

$n = 15$	size, $\beta_3 = -.087$				power, $\beta_3 = -.092$				power, $\beta_3 = -.082$			
homo.	025	050	950	975	025	050	950	075	025	050	950	975
Norm1	064	103	904	938	105	152	942	966	042	063	856	900
Norm2	041	073	938	967	065	134	958	975	021	054	913	925
L. R.	044	071	923	958	064	105	953	975	027	045	887	929
EF-t	022	054	943	974	041	088	968	990	014	042	913	955
iidEF-t	034	056	942	965	052	089	963	986	020	038	910	947

$n = 30$	size, $\beta_3 = -.087$				power, $\beta_3 = -.092$				power, $\beta_3 = -.082$			
homo.	025	050	950	975	025	050	950	075	025	050	950	975
Norm1	041	073	926	957	089	138	967	984	021	033	860	914
Norm2	034	063	929	966	064	120	973	988	016	031	872	922
L. R.	030	062	941	970	071	117	973	987	014	027	878	936
EF-t	035	063	933	969	067	118	973	988	015	032	878	929
iidEF-t	025	054	946	975	054	107	979	991	012	022	893	941

$n = 15$	size, $\beta_3 = -.087$				power, $\beta_3 = -.092$				power, $\beta_3 = -.082$			
hetero	025	050	950	975	025	050	950	075	025	050	950	975
Norm1	076	107	920	950	110	161	950	973	045	072	878	923
Norm2	051	087	922	955	080	133	949	972	035	055	887	929
L. R.	039	070	946	969	063	109	965	980	028	043	912	952
EF-t	025	049	956	981	039	079	969	990	018	034	932	967
iidEF-t	028	049	954	976	043	081	970	986	019	033	929	960

$n = 30$	size, $\beta_3 = -.087$				power, $\beta_3 = -.092$				power, $\beta_3 = -.082$			
hetero	025	050	950	975	025	050	950	075	025	050	950	975
Norm1	053	080	937	967	090	143	970	985	024	042	895	929
Norm2	033	071	935	966	072	125	968	984	017	031	896	931
L. R.	029	055	948	976	057	107	977	988	015	026	910	948
EF-t	028	058	949	972	057	110	975	988	016	028	911	947
iidEF-t	023	046	955	979	045	090	979	989	015	022	917	953

Table 4 reports the results of some simulations done on estimation of β_3 with sample sizes of $n = 15$ and $n = 30$, and with homoscedastic Laplacian errors and normal heteroscedastic errors. The entries in the table give 1000 times the estimated size or power of the corresponding one sided test procedure at specified nominal levels $\alpha = .025,\ .050,\ .950,\ .975$. Each of the four sections in the table is based on 2000 replications of 1000 bootstrap samples. Estimates of the true coverage probabilities of a symmetric 95% confidence interval, e.g., can be obtained as the difference of the "size" estimates in the first and fourth columns of the table at nominal values of .025 and .975.

The EF and the iid EF bootstraps do well in all simulations we have done on estimating a single parameter. Similar results, for example, are obtained for the estimation of β_2. Any gain in power or efficiency through use of the iid procedure with homogeneous errors is apparently relatively small. The iid procedure appears to be quite robust against heteroscedastic errors. The differences seen in Example 5 corresponded to a very high degree of heteroscedasticity. With the more moderate and practical differences here, the iid EF bootstrap performs well. The approximate Norm2 applied to the studentized EF statistic (or the generalized score statistic) does surprisingly well. In the nonlinear examples we considered, Norm2 substantially outperforms Norm1. The latter, often used as the basis for comparison with bootstrap methods (e.g., Huet, Jolivet & Messéan 1990), is itself very poor. The likelihood ratio test based on homoscedastic normal errors also generally does well. As expected, however, it does less well in the heteroscedastic case and its parametric dependence makes it somewhat less attractive than other methods considered.

Table 5 reports the results of simulations on the simultaneous estimation of $(\beta_1, \beta_2, \beta_3)$. We have only reported results for the size of a test based on the competing methods (or equivalently on the coverage probability of the corresponding confidence regions). The results in the table are for homoscedastic Laplacian error. Other error distributions and heteroscedastic errors gave similar results. The entries in each column of the table give the estimated probability of rejection of the (true) null hypothesis based on 2000 replications of 1000 bootstrap samples.

Only the iid EF bootstrap gives good results here and, in fact, it also works reasonably well when the errors are heteroscedastic, at least to the degree considered in this example. The χ^2 approximation (χ^2 app2) to the generalized score statistic (11) gives substantial under coverage. For example, with $n = 30$, the nominal 90% confidence region has estimated coverage probability of 85.0%. As noted in Section 5, the EF bootstrap can be viewed as a calibration of this procedure and gives an estimated coverage probability 93.7%, a substantial overcorrection. The χ^2 app1 based on the quadratic form from the least squares estimate performs very badly. The likelihood ratio is better but also has substantial under coverage.

TABLE 5: Estimated sizes of tests of a simple hypothesis about $(\beta_1, \beta_2, \beta_3)$ in the nonlinear model of Example 6. The iid-χ^2 is based on the χ^2 approximation to the iid quadratic form (29).

	$n = 15$			$n = 30$			$n = 45$		
test/size	.200	.100	.050	.200	.100	.050	.200	.100	.050
χ^2 app1	.617	.536	.485	.482	.393	.334	.414	.334	.279
χ^2 app2	.357	.233	.134	.270	.150	.074	.231	.118	.062
L. R.	.291	.168	.092	.235	.116	.067	.216	.105	.048
EF-t	.107	.024	.008	.168	.063	.025	.166	.072	.028
iid EF-t	.215	.105	.042	.200	.098	.046	.185	.094	.042
iid-χ^2	.265	.138	.068	.230	.111	.054	.205	.100	.043

The results in Table 5 suggest the need, in this example, for a further calibration of the EF bootstrap as described in Section 5. Table 6 gives the results of such a calibration for a sample size of $n = 30$, again with Laplacian errors. This further calibration results in estimated coverage probabilities much closer to the nominal values.

Some difficulties arise due to the singularity at $\beta_3 = 0$. With the sample size of 30, about one in one thousand first level replications results in one or more second level bootstraps for which the singularity causes difficulties. (In these instances, the maximized likelihood continues to increase as $\beta_3 \uparrow 0$.) This happens when the first level bootstrap gives a β_3 estimate that is unusually close to 0. For sample size 30, there was little lost through ignoring these cases. With smaller sample sizes, the problem occurs more frequently and some difficulties with calibration can result.

The Michaelis–Menten model,

$$f(x,\beta) = \beta_1\, x/\beta_2 + x$$

was also considered by Huet, Jolivet & Messéan (1990). We also investigated the procedures within that model and our simulations gave similar results to those in Example 6.

Remark 11. All of the calculations for the vector and nuisance parameter procedures were done on a PC with a Pentium II processor. The simulations took less than 20 minutes per 1000×1000 run. The calibration example ($2000 \times 600 \times 600$) was more extensive since it involved the construction of about 720 million samples, each of size 30. Each of the 2000 replications required solving the least squares equations 601 times. The calculation required about 50 hours. This corresponds to about 90 seconds for each replication, which is the time required to carry out the calibration for a particular sample. In an application, one might wish to have a larger calibration sample of say 1000×1000 and this is certainly feasible.

TABLE 6: Estimated size and power for tests of a simple hypothesis about $(\beta_1, \beta_2, \beta_3)$ in the model (30) for the uncalibrated and the calibrated EF bootstrap. Entries are based on 2000 replications of 600 first level and 600 second level bootstrap samples.

Uncalibrated EF-t bootstrap

size	(0,0)	(0,-1)	(0,+1)	(-1,0)	(-1,-1)	(-1,+1)	(+1,0)	(+1,-1)	(+1,+1)
.800	.825	.997	.995	.955	.956	1.000	.954	1.000	.969
.500	.508	.966	.975	.842	.827	1.000	.804	1.000	.859
.200	.174	.865	.896	.540	.527	.995	.481	.997	.560
.100	.068	.720	.787	.364	.323	.986	.292	.981	.388
.050	.027	.560	.656	.212	.203	.964	.175	.955	.238

Calibrated EF-t bootstrap

size	(0,0)	(0,-1)	(0,+1)	(-1,0)	(-1,-1)	(-1,+1)	(+1,0)	(+1,-1)	(+1,+1)
.800	.825	.997	.995	.955	.954	1.000	.953	1.000	.967
.500	.512	.964	.976	.848	.830	1.000	.810	1.000	.855
.200	.205	.871	.902	.563	.552	.995	.514	.995	.593
.100	.093	.766	.825	.411	.386	.988	.364	.985	.435
.050	.048	.633	.708	.280	.262	.972	.230	.966	.295

The column (i, j) designates $\beta_2 = -1309 + 25i$ and $\beta_3 = -.087 + .005j$.
The true values are $\beta_2 = -1309, \beta_3 = -.087; \beta_1 = 1122$ throughout.

6.3. Nonlinear regression with multivariate response.

Let $y_i' = (y_{i1}, \ldots, y_{in_i})$ be the response variable for the i-th individual and consider the model

$$y_i = f_i(x_i, \beta) + \varepsilon_i, \quad i = 1, \ldots, n \tag{31}$$

where, as before, x_i is a $k \times 1$ vector of known constants, $\beta' = (\beta_1, \ldots, \beta_p)$ is a vector of regression parameters, and f_i is a (nonlinear) function of β. The errors $\varepsilon_i' = (\varepsilon_{i1}, \ldots, \varepsilon_{in_i})$ in

(31) are independent with zero mean and unknown covariance matrix Σ_i. In order to estimate β, we consider the weighted least squares objective function

$$SS(y,\beta) = \sum_{i=1}^{n} \{y_i - f_i(x_i,\beta)\}' B_i^{-1} \{y_i - f_i(x_i,\beta)\} \tag{32}$$

where B_i is a $n_i \times n_i$ matrix of constants which can be interpreted as a "working" covariance matrix. The corresponding estimating equation is

$$S(y,\beta) = \sum_{i=1}^{n} A_i(x_i,\beta)' B_i^{-1} \{y_i - f_i(x_i,\beta)\} = 0, \tag{33}$$

where $A_i(x_i,\beta) = \partial f_i(x_i,\beta)/\partial\beta'$ is an $n_i \times p$ matrix. We suppose that $A_i(x_i,\beta)$ is continuous in β. The weighted least squares estimator $\hat{\beta}_{\text{wls}}$ is the (assumed unique) solution to (33). Note that the estimating equation (33) is exactly of the form (1) and methods developed above can be applied directly. We have not recorded specific results here, though they can be written in a straightforward manner.

Standard methods of analysis utilize the "sandwich" estimators of the variance of $\hat{\beta}_{\text{wls}}$. The methods based on the EF bootstrap provide a relatively simple alternative to this approach that has the potential to better reflect small sample properties and perhaps to increase accuracy. This area is the subject of further investigations.

In some applications, B_i is a function of β. When this is the case, it would be inappropriate to minimize (32), though the estimating function (33) can still be used. Typically it is solved through an iterative reweighting with the estimate of B_i being updated with each iteration. This is a further extension which would allow the applicability of these methods to the area of Generalized Estimating Equations and Generalized Linear Models (cf., e.g., Liang & Zeger 1986 and Zeger, Liang & Albert 1988).

7. L_q ESTIMATION

In this section, we consider a linear estimating equation in which $g_i(y_i,\theta)$ is not differentiable with respect to θ. Such situations arise in nonparametric and semiparametric models (cf., e.g., Koenker & Bassett 1978 and Hettmansperger 1984) and in robust regression (e.g., Huber 1981).

We consider the general regression model (28) and suppose that β is to be estimated by minimization of

$$\sum_{i=1}^{n} |y_i - f(x_i,\beta)|^a$$

for some a, $1 \le a \le 2$. When $a = 2$, this yields the least squares estimator, but for other values of a, a more robust procedure than least squares is obtained. For example, $a = 1$ yields the median regression; cf., e.g., Koenker & Bassett (1978). In general, we have the following estimating equation,

$$\sum_{i=1}^{n} \operatorname{sign}\{y_i - f(x_i,\beta)\} A(x_i,\beta) |y_i - f(x_i,\beta)|^{a-1} = 0. \tag{34}$$

We consider $a = 1.5$ for which both the multidimensional and nuisance parameter problems can be addressed. We confine attention to simple linear regression, $f(x_i,\beta) = x_i'\beta$, and the corresponding estimating equation

$$\sum_{i=1}^{n} \operatorname{sign}(y_i - x_i'\beta) x_i |y_i - x_i'\beta|^{1/2} = 0. \tag{35}$$

The EF bootstrap procedures for estimating the whole parameter β or components of β can be applied to this problem in a straightforward manner. The following simulations are based on 1000 replications of 1000 bootstrap samples.

For the first simulation, we consider two models for a problem with a single regressor variable:

- *homoscedastic errors*: The model is $Y_i = \beta x_i + e_i$, $i = 1, \ldots, n$, where $n = 20$, the errors are iid $\mathsf{N}(0, 0.25)$, $x_i = -2, -1.9, \ldots, -1.1, 1.1, \ldots, 2$ and $\beta = 1$.
- *heteroscedastic errors*: The model is $Y_i = \beta x_i + x_i^2 e_i$, $i = 1, \ldots, n$ with n, x_i, β, and errors defined as above.

TABLE 7: Coverage percentages and average confidence intervals (with standard deviations) for the L_q estimating function of Section 7.

	Homoscedastic errors		Heteroscedastic errors	
	80%	95%	80%	95%
Norm1	**75.4**	**89.7**	**75.3**	**90.4**
Ave CI	[0.907,1.091]	[0.858,1.139]	[0.75,1.25]	[0.62,1.38]
(*s.d.*)	0.082, 0.080	0.088, 0.087	0.21,0.22	0.23, 0.24
Norm2	**79.5**	**92.5**	**77.2**	**93.1**
Ave CI	[0.904,1.097]	[0.857,1.134]	[0.756,1.25]	[0.63,1.38]
(*s.d.*)	0.075, 0.074	0.077, 0.076	0.20,0.20	0.21, 0.21
iid C-t	**78.5**	**94.3**	**76.1**	**93.2**
Ave CI	[0.890,1.099]	[0.806,1.179]	[0.73,1.26]	[0.51,1.49]
(*s.d.*)	0.083, 0.080	0.104, 0.097	0.22,0.22	0.27, 0.27
Paired-t	**80.7**	**95.4**	**79.3**	**95.4**
Ave CI	[0.889,1.100]	[0.789,1.199]	[0.72,1.29]	[0.45,1.55]
(*s.d.*)	0.084, 0.081	0.114, 0.108	0.23,0.24	0.30, 0.31
EF-t	**81.0**	**95.0**	**79.0**	**94.5**
Ave CI	[0.900,1.092]	[0.845,1.148]	[0.74,1.27]	[0.59,1.42]
(*s.d.*)	0.075, 0.074	0.078, 0.077	0.21, 0.20	0.22, 0.21
iidEF-t	**80.7**	**95.2**	**76.4**	**93.6**
Ave CI	[0.900,1.093]	[0.843,1.150]	[0.76,1.25]	[0.61,1.39]
(*s.d.*)	0.075,0.074	0.078,0.077	0.21,0.20	0.22,0.21

Table 7 summarizes average confidence intervals for the various methods in this example and the results are similar to those reported in Table 3 and discussed in Section 6. The normal approximation (Norm2) based on the score or estimating function S_t performs considerably better than Norm1. The EF and iid EF procedures do well overall; they have good coverage properties and relatively smaller average interval size than the other bootstrap competitors.

For the second simulation, we consider the model

$$Y_i = \beta_1 + \beta_2 x_{1i} + \beta_3 x_{2i} + c_i e_i, \quad i = 1, \ldots, n$$

with x_{1i}, x_{2i} and e_i all generated independently from the standard normal distribution. (This is a random design matrix where, for each simulation, a new set of independent variables is generated). We take $n = 20$, and the true parameters β_i are set at 0. In the first simulation we took

$c_i = 1$ and in the second, we considered heteroscedastic errors with $c_i = 0.1i$, $i = 1, \ldots, 20$. Simulation results are reported in Table 8 for estimation of the vector parameter β and in Table 9 for estimation of β_3.

TABLE 8: Coverage probabilities of approximate confidence regions for the vector parameter in the L_q estimation problem of Section 7.

	homoscedastic errors				heteroscedastic errors			
test/nom level	0.80	0.90	0.95	0.975	0.80	0.90	0.95	0.975
χ^2 app2	0.67	0.792	0.879	0.925	0.718	0.834	0.885	0.933
EF-t	0.825	0.928	0.974	0.99	0.849	0.935	0.976	0.991
iid-χ^2	0.755	0.876	0.93	0.968	0.777	0.89	0.943	0.982
iid EF-t	0.793	0.897	0.948	0.974	0.82	0.91	0.96	0.983

The results in Tables 8 and 9 are similar to those in Tables 5 and 4 for the nonlinear model simulation of Example 6. The EF-t bootstrap does very well for both homoscedastic and heteroscedastic errors in the estimation of β_3 with accurate coverage probabilities. As before, there is some over coverage in the EF procedures for the vector case and an additional calibration would improve this.

TABLE 9: Coverage probabilities for the regression parameter β_3 in the linear regression model (L_q estimation) in Section 7.

	homoscedastic errors			heteroscedastic errors		
	0.80	0.90	0.95	0.80	0.90	0.95
χ^2 app2	0.785	0.878	0.93	0.772	0.875	0.923
EF-t	0.81	0.904	0.948	0.806	0.896	0.941

8. COMMENTS AND DISCUSSION

The estimating function bootstrap works well in many situations and provides a good alternative to other bootstrap methods or asymptotic approximations. The EF bootstrap methods are straightforward, intuitively appealing and simple to implement. They generally involve much less computation than many competing bootstrap methods.

It is mentioned in Remark 3 but bears repeating that our methods are based on the studentized variables with variance estimates based on $\widetilde{V}$ and $\widetilde{\mathcal{W}}$ as, for example, in (16). This gives a functionally invariant statistic so that issues of parameterization do not arise and the properties of inferential procedures based on this score appear to be very good. Standardizing or studentizing the estimating function or generalized score function through use of $\widehat{V}$ and $\widehat{\mathcal{W}}$ can give very poor results in models with a high degree of nonlinearity or with independent variables that exert a high degree of influence on the estimates.

In the results reported for estimating β_3 in Table 4 and in other simulations we have done, the EF bootstrap works well for the estimation of a single parameter in the nuisance parameter case. On the other hand, estimation of a vector parameter in Table 5 gave poorer results for the EF bootstrap. This difficulty with vector estimation seemed generally to be the case in examples where there are observations that have relatively high influence. The following points are worth mentioning:

i. It is encouraging that the EF bootstrap does so well in examples involving estimation in the presence of nuisance parameters. This is a more important applied problem than that of estimating a vector parameter.

ii. The vector EF procedure is shown in the appendix to generate confidence regions that are accurate to order $n^{-3/2}$ whereas, for nuisance parameters, the EF bootstrap confidence regions are generally accurate to asymptotic order n^{-1}. This is a striking illustration that the asymptotic results may not provide much guidance in a particular example; cf. Lee & Young (1996) for interesting comment on this point.

iii. The calibration of the EF bootstrap would apparently increase the asymptotic accuracy of the vector procedure to order n^{-2} (cf. Hall 1986 and Beran 1987). In examples we considered, this calibration does remedy the difficulty with the EF the vector results. Hall & Martin (1996) advocate the use of "the double bootstrap" on percentile methods as a general tool for constructing accurate confidence intervals. In following this recommendation, the EF bootstrap offers a substantial advantage since computation for the double bootstrap is well within reasonable bounds.

iv. There appears to be very little in the literature on simultaneous estimation of vector parameters. It appears however, that this is a difficult problem and one where many of the conventional bootstrap and other procedures perform very poorly. Yatrakos, in an unpublished report, presents some results which suggest that estimation of a vector parameter is inherently difficult for the bootstrap process.

There are many areas which require further investigation.

1. The interplay (and conflict) between bootstrap procedures and conditionality is a general area that requires careful consideration. There are potential advantages in estimating the distribution of the pivotal through reflecting more of the observed sample characteristics in the bootstrap sample. Respecting stratification in the study in the bootstrap sampling, or even introducing post hoc stratification may be a way to do this. On the other hand, the notion of conditioning on strata can also result in insufficient latitude to estimate the distribution with sufficient accuracy using bootstrap methods. The common means problem with small n_i and large k offers an example where one apparently cannot stratify in resampling. On the other hand, if n_i is large and k is small, then stratification is essential. What general strategies should be used?

2. It would be possible to develop an EF Jackknife procedure or a "wild" bootstrap (cf., e.g., Wu 1986) that we would expect to work well in some instances. In the context of a regression model with iid errors, the permutation (or randomization) distribution suggested by Fisher (1937) and further developed by Kempthorne (1954) and Hinkelman & Kempthorne (1994, pp. 41–44) could be applied. It would involve bootstrap resampling without replacement rather than with replacement, but otherwise would parallel the iid C or iid EF procedures.

3. In this paper, we have restricted attention to linear estimating functions with independent terms. The general idea, however, of estimating the distribution of the estimating function is useful more generally. Extending these ideas to include, e.g., simple time series models is an area of current investigation.

4. Asymptotic normal and χ^2 approximations to the generalized score statistics (the studentized EF statistics) do remarkably well. They give rise to procedures that are invariant under reparameterization and which have much better accuracy than methods based for example on the estimator directly. It may be possible to develop more accurate analytical approximations to the generalized score and this may be worth investigation.

APPENDIX: ASYMPTOTIC PROPERTIES

In this Appendix, we establish the asymptotic properties of the EF bootstraps. In subsections 1 and 2, we obtain the Edgeworth expansions of $S_t(y,\theta)$, and S_t^* as defined in Section 2 of the paper and so establish the second order approximations. We also show the validity of Edgeworth expansions for $Q(y,\theta)$ and Q^* as defined in Section 4. Subsection 3 establishes the asymptotic results for the EF bootstraps for nuisance parameters. In subsection 4, we consider the asymptotic results for the EF-iid bootstrap in both linear and non-linear regression applications with homogeneous errors discussed in Sections 6 and 7. The main aspect of these results which require new development relates to the replacement of θ with $\hat{\theta}$ in the resampling associated with S^* and S_t^*. The condition A5 is needed to address this and allows the proof of Lemma 3.

1. Edgeworth expansions and EF bootstraps.

From the central limit theorem, it follows that

$$\widehat{V}^{-1/2}S^* \xrightarrow{\mathcal{L}} N_p(0,I), \tag{36}$$

where $\widehat{V}$ is defined in (5). If $\widehat{V} = V_\theta + O_p(n^{-1/2})$, the first order approximation of the distribution of S^* to that of $S(y,\theta)$ follows directly from (2) and (36).

To examine whether a second order approximation holds, we compare the Edgeworth expansions of $S_t(y,\theta)$ and S_t^* for general $p = \dim(\theta)$. For this purpose, we introduce some notation that follows Hall & Horowitz (1996). Let $h_i(y_i,\theta)$ be the vector of dimension of k (say) that contains the unique components (eliminating repetitions) of $g_i(y_i,\theta)$ and $g_i(y_i,\theta)g_i(y_i,\theta)'$. [Note that, $k \le p + p(p+1)/2$.] Similarly, for the bootstrap, we define $\hat{h}_i = h_i(y_i,\hat{\theta})$ and let h_i^*, $i = 1,\ldots,n$ be the corresponding bootstrap sample from $\{\hat{h}_1,\ldots,\hat{h}_n\}$. Let $T_n = n^{-1}\sum_{i=1}^n h_i(y_i,\theta)$, $T_n^* = n^{-1}\sum_{i=1}^n h_i^*$, $\Psi_n = n^{1/2}\{T_n - \mathrm{E}(T_n)\}$ and $\Psi_n^* = n^{1/2}\{T_n^* - E^*(T_n^*)\}$.

We require the following assumptions:

A1. *The true value θ_0 is an interior point of the compact parameter space Ω. It is the unique solution in Ω to the equation $\mathrm{E}_{\theta_0}S(y,\theta) = 0$. Further $\hat{\theta} \to \theta_0$ (a.s.).*

A2. (i) *the smallest eigenvalue of $n^{-1}\sum_{i=1}^n \operatorname{cov}\{h_i(y_i,\theta_0)\}$ is bounded away from zero;*
(ii) *The average sixth moments $n^{-1}\sum_{i=1}^n \mathrm{E}\|h_i(y_i,\theta_0)\|^6$ are bounded away from infinity and for every positive ε,*

$$\lim_{n\to\infty} n^{-1}\sum_{i=1}^n \mathrm{E}[I\{\|h_i(y_i,\theta_0)\| > \varepsilon n^{1/2}\}\|h_i(y_i,\theta_0)\|^6] = 0.$$

A3. *Let m be the vector of all possible moments of the form*

$$\lim_{n\to\infty} n^{\omega(\ell)}\mathrm{E}\left(\prod_{k=1}^{\ell}\Psi_{nj_k}\right),$$

where $2 \le \ell \le 6$, $w(l) = 0$ if ℓ is even and $1/2$ if ℓ is odd, and Ψ_{nj_k} is the j_k component of the vector Ψ_n. Similarly, let m_n be the corresponding vector of moments of the form

$$n^{\omega(\ell)}E^*\left(\prod_{k=1}^{\ell}\Psi_{nj_k}^*\right),$$

where $2 \le \ell \le 6$ and $\Psi_{nj_k}^*$ is the j_k component of the vector Ψ_n^*. It is assumed that $m_n - m = O_p(n^{-1/2})$.

A4. *The characteristic function $\xi_n(t,\theta)$ of $h_n(y_n,\theta)$ satisfies*

$$\limsup_{n\to\infty}\sup_{\|t\|>b}|\xi_n(t,\theta_0)| < 1 \tag{37}$$

for every positive b.

A5. *Let*

$$f_n(t,\theta) = n^{-1} \sum_{j=1}^{n} \exp\{ih_j(y_j,\theta)t\}.$$

For every $b, \delta > 0$, the function

$$\limsup_{n\to\infty} \sup_{b\le\|t\|\le\exp(\delta n)} |f_n(t,\theta)|$$

is continuous at θ_0 a.s.

Assumption A1 ensures that there is an open ball inside the parameter space Ω, such that for all n large enough, the estimator is inside the open ball almost surely. Assumption A2(i) ensures that $S(y,\theta)$, $V(y,\theta)$, $S_t(y,\theta)$ and $Q(y,\theta)$ are smooth functions of T_n, so that the Taylor expansions about $\mathrm{E}(T_n)$ are valid. The main difference between the EF bootstrap and the classical bootstrap is that the EF bootstrap is based on resampling $\hat{h}_i$, which depends on the estimator $\hat{\theta}$. To ensure a valid Edgeworth expansion of the bootstrap Ψ_n^*, we need the condition (37), Assumption A5 and Assumption A1. Note that Assumption A4 implies that h_n is non lattice. The case h_n lattice is an important one that we do not consider here.

We now state the main results regarding the Edgeworth expansions. Let $U = \lim_{n\to\infty} \mathrm{var}(\Psi_n)$ and $U_n = \mathrm{var}^*(\Psi_n^*)$. From Assumption A3, $U_n = U + O_p(n^{-1/2})$. Since $S_t(y,\theta) = V(y,\theta)^{-1/2}S(y,\theta)$ and both $V(y,\theta)^{-1/2}$ and $S(y,\theta)$ are differentiable functions of T_n in a neighbourhood of $\nu = \lim_{n\to\infty} \mathrm{E}T_n$, we can write $S_t(y,\theta) = n^{-1/2}B(T_n)$. Let G be the $p\times k$ matrix, $G = \partial B(\nu)/\partial\nu'$ and note that $G'UG = I$, is the asymptotic variance of S_t. Let $\phi_q(x;\Sigma)$ be the q-variate normal density at x with mean 0 and covariance matrix Σ, and let $\Phi_q(x;\Sigma)$ be the corresponding distribution function. We will use $\phi_q(x) = \phi_q(x;I)$ and $\Phi_q(x) = \Phi_q(x;I)$ and the standard notation, $\phi = \phi_1$ and $\Phi = \Phi_1$ for the univariate case. Let $D = \partial/\partial u'$.

THEOREM 1. *Under Assumptions A1–A4, the following two results hold:*

$$\text{i)}\quad \sup_{x\in\mathbf{R}^p}\left|\mathrm{P}\{S_t(y,\theta)\le x\} - \Phi_p(x) - \sum_{r=1}^{2} n^{-r/2}\int_{u\le x}\psi_r(D;m)\phi_p(u)\,du\right| = O(n^{-3/2}), \quad (38)$$

where ψ_r is a polynomial function (in p variables $D_1,\dots,D_p$) whose coefficients are continuous functions of m.

$$\text{(ii)}\qquad \sup_z\left|\mathrm{P}\{Q(y,\theta) < z\} - \int_0^z \{1 + n^{-1}\pi(u,m)\}\,dP(\chi_p^2 < u)\right| = O(n^{-3/2}), \quad (39)$$

where $\pi(z,m)$ is an odd polynomial function of z whose coefficients are continuous functions of m and χ_p^2 is a χ^2 variable with p degrees of freedom.

THEOREM 2. *Under Assumptions A1–A5, the following expansions are valid.*

$$\text{(i)}\ \sup_{x\in\mathbf{R}^p}\left|P^*(S_t^*\le x) - \Phi_p(x) - \sum_{r=1}^{2} n^{-r/2}\int_{u\le x}\psi_r(D,m_n)\phi_p(u)\,du\right| = O_p(n^{-3/2}).$$

$$\text{(ii)}\ \sup_z\left|P^*(Q^* < z) - \int_0^z \{1 + n^{-1}\pi(u,m_n)\}\,dP(\chi_p^2 < u)\right| = O_p(n^{-3/2}).$$

THEOREM 3. *If Assumptions A1–A5 hold, then*

$$\sup_{x\in\mathbf{R}^p} |\mathrm{P}\{S_t(y,\theta) < x\} - P^*(S_t^*\le x)| = O_p(n^{-1})$$

and

$$\sup_z \|P(Q(y,\theta) < z) - P^*(Q^* < z)\| = O_p(n^{-3/2}).$$

To provide a better understanding of the above notation and theorems, we consider the scalar case ($p = 1$). In this case, $h_i(y_i, \theta) = (g_i(y_i, \theta), g_i^2(y_i, \theta))'$, $T_n = (\bar{g}, \bar{g}_2)$, where $\bar{g} = n^{-1}\sum g_i(y_i, \theta)$ and $\bar{g}_2 = n^{-1}\sum g_i^2(y_i, \theta)$, so that $S_t = S_t(y, \theta) = n^{1/2}H(T_n) = n^{1/2}\bar{g}/(\bar{g}_2 - \bar{g}^2)^{1/2}$. S_t^* has the corresponding bootstrap form. The Edgeworth expansions for $S_t(y, \theta)$ and S_t^* are, respectively,

$$P(S_t < x) = \Phi(x) + \frac{\mu_3}{6v^3 n^{1/2}}(2x^2 + 1)\phi(x) + n^{-1}\psi_2^\dagger(x, m)\phi(x) + O(n^{-3/2}) \tag{40}$$

and

$$P^*(S_t^* < x) = \Phi(x) + \frac{\hat{\mu}_3}{6\hat{v}^3 n^{1/2}}(2x^2 + 1)\phi(x) + n^{-1}\psi_2^\dagger(x, m_n)\phi(x) + O_p(n^{-3/2}) \tag{41}$$

where

$$v^2 = \lim_{n\to\infty} n^{-1}\sum E_\theta\{g_i(Y_i, \theta)\}^2, \quad \hat{v}^2 = n^{-1}\sum\{g_i(Y_i, \hat{\theta})\}^2$$
$$\hat{\mu}_3 = n^{-1}\sum\{g_i(Y_i, \hat{\theta})\}^3, \quad \mu_3 = \lim_{n\to\infty} n^{-1}\sum E_\theta\{g_i(Y_i, \theta)\}^3.$$

Here, $\psi_2^\dagger(x, m)$ is an odd polynomial function of x with coefficients depending on m. In the notation of (38), we find that $\psi_1(D, m) = \mu_3(-2D^3 + 3D)/(6v^3)$ and ψ_2 is related to $\psi_2^\dagger$ and is quite complicated in form. The leading terms in (40) and (41) are identical. The second terms depend on the second and third moments and, from Assumption A3, agree to order n^{-1} so that

$$\mathrm{P}\{S_t(y, \theta) < x\} - P^*(S_t^* < x) = O_p(n^{-1}).$$

Thus, the studentized EF bootstrap is accurate to second order.

Since $Q(y, \theta) = S_t^2$, it follows from (40) and (41) that

$$\begin{aligned}\mathrm{P}\{Q(y, \theta) < z\} &= \Phi(z^{1/2}) - \Phi(-z^{1/2}) + n^{-1}\psi_2^\dagger(z^{1/2}, m)\phi(z^{1/2}) \\ &\quad - n^{-1}\psi_2^\dagger(-z^{1/2}, m)\phi(-z^{1/2}) + O(n^{-3/2})\end{aligned}$$

and

$$\begin{aligned}P^*(Q^* < z) &= \Phi(z^{1/2}) - \Phi(-z^{1/2}) + n^{-1}\psi_2^\dagger(z^{1/2}, m_n)\phi(z^{1/2}) \\ &\quad - n^{-1}\psi_2^\dagger(-z^{1/2}, m_n)\phi(-z^{1/2}) + O_p(n^{-3/2}).\end{aligned}$$

The second term has been eliminated. Since $m_n - m = O_p(n^{-1/2})$, the terms involving ψ_2 agree to the higher order, $n^{-3/2}$. In other words,

$$\mathrm{P}\{Q(y, \theta) < z\} - P(Q^* < z) = O_p(n^{-3/2}).$$

This property of symmetric intervals is of course well known and, to some extent, can result from the cancellation of somewhat larger errors at both endpoints.

It follows in general from Theorem 3 that the confidence region (13) based on the quadratic form Q is third order accurate $O_p(n^{-3/2})$.

2. Proof of theorems.

LEMMA 1. *Suppose Assumptions A1–A4 hold. Then*

$$\sup_{x\in\mathbb{R}^k}\left|\mathrm{P}(\Psi_n \le x) - \Phi_k(x; U) - \int_{u\le x}\sum_{r=1}^{2} n^{-r/2}p_r(D, m)\phi_k(u; U)\right| = O(n^{-3/2}), \tag{42}$$

where $p_r(D, m)$ is a polynomial function in k variables whose coefficients are continuous functions of m and is defined in a manner similar to Bhattacharya & Rao (1976, p. 54, Equation 7.11).

Proof. The $\{h_i(y_i, \theta_0) : i \geq 1\}$ is a sequence of independent random vectors with value in $\mathbb{R}^k$. Assumptions A1–A4 ensure the conditions of Theorem 20.6 of Bhattacharya & Rao (1976, p. 216). The Edgeworth expansion (42) follows directly from that theorem and Corollary 20.4. □

LEMMA 2. *If Assumptions A1–A5 hold, then*

$$\sup_{x\in\mathbb{R}^k}\left|P^*(\Psi_n^* \leq x) - \Phi_k(x; U_n) - \int_{u\leq x}\sum_{r=1}^{2} n^{-r/2}p_r(D, m_n)\phi_k(u; U_n)\right| = O_p(n^{-3/2}).$$

For given n, $h_1^*, \ldots, h_n^*$ are iid random vectors with values in R^k, and Ψ_n^* has zero mean and a finite sixth moment. To apply Theorem 20.1 of Bhattacharya & Rao (1976, p. 208), it would be sufficient to show that Cramér's condition holds for h_1^*. For a given n, however, h_1^* is discrete so Cramér's condition is not satisfied. We therefore prove the following Lemma 3 first and then return to Lemma 2. We instead use a result Theorem 3.3 of Bhattacharya (1987), which in turn depends on a result of Babu & Singh (1984). The following Lemma 3 is analogous to Lemma 2 of Babu and Singh and essentially extends their result to independent sequences as we require.

LEMMA 3. *Suppose Assumptions A1–A5 hold. Then, for any $b > 0$, there exist positive constants ε and δ such that*

$$\limsup_{n\to\infty}\sup_{b\leq\|t\|\leq e^{\delta n}}|f_n(t,\hat{\theta})| \leq 1-\varepsilon \text{ a.s.}, \tag{43}$$

where $f_n(t,\hat{\theta}) = n^{-1}\sum_{j=1}^{n}\exp\{ih_j(y_j,\hat{\theta})t\}$ is the characteristic function of h_1^*.

Proof. First we note that $\mathrm{E}f_n(t,\theta_0) = n^{-1}\sum_{j=1}^{n}\xi_n(t,\theta_0)$ and from Assumption A4, it follows that

$$\limsup_{n\to\infty}\sup_{\|t\|>b}\mathrm{E}f_n(t,\theta_0) < 1$$

for any $b > 0$. This is an extension of Cramér's condition for independent but not identically distributed random variables related to conditions discussed, e.g., in Liu (1988). This, together with the condition on the moments in Assumption A2, is sufficient to extend the argument in Babu & Singh (1984) to apply to the independent sequence $\{h_n(y_n,\theta_0)\}$. Thus, we find that for any $b > 0$, there exist positive constants ε and δ such that

$$\limsup_{n\to\infty}\sup_{b\leq\|t\|\leq e^{\delta n}}|f_n(t,\theta_0)| < 1-\varepsilon, \text{ a.s.}$$

Now from Assumption A5, it follows that there exists a constant $r > 0$ such that

$$\sup_{\|\theta-\theta_0\|\leq r}\limsup_{n\to\infty}\sup_{b\leq\|t\|\leq e^{\delta n}}|f_n(t,\theta)| < 1-\varepsilon, \text{ a.s.}$$

From Assumption A1, $\|\hat{\theta}_n - \theta_0\| < r$ a.s. and hence

$$|f_n(t,\hat{\theta})| \leq \sup_{\|\theta-\theta_0\|\leq r}|f_n(t,\theta)|, \text{ a.s.}$$

This is sufficient to establish (43) since

$$\limsup_{n\to\infty}\sup_{b\leq\|t\|\leq e^{\delta n}}|f_n(t,\hat{\theta})| \leq \sup_{\|\theta-\theta_0\|\leq r}\limsup_{n\to\infty}\sup_{b\leq\|t\|\leq e^{\delta n}}|f_n(t,\theta)| < 1-\varepsilon \text{ a.s.}$$

Proof of Lemma 2. From Lemma 3, this lemma follows using a proof identical to that of Theorem 3.3 of Bhattacharya (1987). □

Proof of Theorem 1. Both $V(y,\theta)$ and $S(y,\theta)$ are smooth functions of T_n. From Lemma 1, we have the Edgeworth expansion of Ψ_n to order $O_p(n^{-3/2})$. From Theorem 2 and Remark 1.1 of Bhattacharya & Ghosh (1978), the Edgeworth expansion for $S_t(y,\theta) = H(T_n)$ can be obtained to order $O_p(n^{-3/2})$. Then let $s = 6$ and use Theorem 1 and Remark (2.2) of Chandra & Ghosh (1979) to show that $Q(y,\theta) = S(y,\theta)'V^{-1}(y,\theta)S(y,\theta)$ has Edgeworth expansion (39). □

The proof of Theorem 2 uses Lemma 2, but otherwise is identical to that of Theorem 1. Theorem 3 follows immediately from Theorems 1 and 2.

3. Nuisance parameters.

Following the notation of Section 4.2, we consider the Edgeworth expansions of $Q_{11}(y,\theta)$, $\widetilde{Q}_{11}(y,\theta_1)$ and Q_{11}^*. The only essential difference from Theorem 1 is that $Q_{11}(y,\theta)$ depends on the derivatives of $g_i(y_i,\theta)$. Thus we let $h_i(y_i,\theta)$ be a vector containing the unique components of $g_i(y_i,\theta)$, $g_i(y_i,\theta)g_i(y_i,\theta)'$ and $\partial g_i(y_i,\theta)/\partial\theta_2'$. Following Theorems 1–3, we have

THEOREM 4. *Suppose $S(y,\theta)$ is differentiable with respect to θ_2 and the Assumptions A1–A5 hold with $h_i(y_i,\theta)$ as redefined immediately above. Then*

$$\mathrm{P}\{Q_{11}(y,\theta) < z\} - P^*(Q_{11}^* < z) = O_p(n^{-3/2}).$$

For the case of linear regression, $Q_{11}(y,\theta)$ depends on θ_1 only, and one can construct a third order accurate confidence region of θ_1 based on Q_{11}^* and $Q_{11}(y,\theta)$. In the nonlinear models, however, $Q_{11}(y,\theta)$ depends on θ_2 and we use $\widetilde{Q}_{11}(y,\theta_1)$ instead. From Edgeworth expansions, it can be shown that

$$\mathrm{P}\{\widetilde{Q}_{11}(y,\theta_1) < z\} - P^*(Q_{11}^* < z) = O_p(n^{-1}),$$

which is only second order accurate.

4. Asymptotic properties of the EF-iid bootstraps.

Asymptotic properties of the EF-iid bootstrap and their studentized versions can also be derived. We require assumptions similar to A1–A5, to establish a result analogous to Theorem 3 for the iid EF bootstrap. Thus, we find

$$\sup_{x\in\mathbf{R}^p} \left|\mathrm{P}\{S_t^{\mathrm{iid}}(y,\theta) < x\} - P^*(S_t^{\mathrm{iid}*} \le x)\right| = O_p(n^{-1})$$

and

$$\sup_{z} \left|\mathrm{P}\{Q^{\mathrm{iid}}(y,\theta) < z\} - P^*(Q^{\mathrm{iid}*} < z)\right| = O_p(n^{-3/2}),$$

giving second and third order accurate results respectively. The results for the nuisance parameter case are identical to those discussed in Section A3 for the EF bootstraps.

ACKNOWLEDGEMENTS

This article formed the basis of the R. A. Fisher Lecture given by John D. Kalbfleisch to the Joint Statistical Meetings in Baltimore, Maryland, August 1999. We wish to thank Professors Vidyadhar P. Godambe, Alastair Scott, Mary E. Thompson and Christopher J. Wild for their comments on this work and for helpful discussions and suggestions. The comments of the referees and the Editor also were helpful in preparing this revision. This work was supported in part by a research grant from the Natural Sciences and Engineering Research Council of Canada to J. D. Kalbfleisch and from the National University of Singapore to F. Hu. Both authors acknowledge the generous research support of the National University of Singapore where much of this work was done.

REFERENCES

G. J. Babu & K. Singh (1984). On one term Edgeworth correction by Efron's bootstrap. *Sankhyā Series A*, 46, 219–232.

O. E. Barndorff-Nielsen (1983). On a formula for the distribution of a maximum likelihood estimator. *Biometrika*, 70, 343–365.

M. S. Bartlett (1936). The information available in small samples. *Proceedings of the Cambridge Philosophical Society*, 34, 33–40.

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

R. N. Bhattacharya (1987). Some aspects of Edgeworth expansions in statistics and probability. In *New Perspectives in Theoretical and Applied Statistics* (M. L. Puri, J. P. Vilaplana and W. Wertz, eds.), Wiley, New York, pp. 157–170.

R. N. Bhattacharya & J. K. Ghosh (1978). On the validity of the formal Edgeworth expansion. *The Annals of Statistics, 6, 434–451.*

R. N. Bhattacharya & R. R. Rao (1976). *Normal Approximation and Asymptotic Expansions*. Wiley, New York.

D. D. Boos (1992). On generalized score tests. *The American Statistician*, 46, 327–333.

R. J. Buehler (1983). Fiducial inference. In *Encyclopedia of Statistical Sciences vol. 3: Faà di Bruno's Formula to Hypothesis Testing* (S. Kotz and N. L. Johnson, eds.), Wiley, New York, pp. 76–79.

T. K. Chandra & J. K. Ghosh (1979). Valid asymptotic expansions for the likelihood ratio statistic and other perturbed chi-square variables. *Sankhyā Series A*, 41, 22–47.

D. R. Cox & N. M. Reid (1987). Parameter orthogonality and approximate conditional inference (with discussion). *Journal of the Royal Statistical Society Series B, 49, 1–39.*

A. Davison & D. V. Hinkley (1997). *Bootstrap Methods and Their Application*. Cambridge University Press, New York.

T. J. DiCiccio & B. Efron (1996). Bootstrap confidence intervals (with discussion). *Statistical Science*, 11, 189–228.

T. J. DiCiccio & J. P. Romano (1988). A review of bootstrap confidence intervals (with discussion). *Journal of the Royal Statistical Society Series B, 50, 338–354.*

B. Efron (1979). Bootstrap methods: another look at the jackknife. *The Annals of Statistics, 7, 1–26.*

B. Efron (1987). Better bootstrap confidence intervals. *Journal of the American Statistical Association, 82, 171–200.*

B. Efron & R. J. Tibshirani (1986). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. *Statistical Science*, 1, 54–77.

B. Efron & R. J. Tibshirani (1993). *An Introduction to Bootstrap*. Chapman & Hall, New York.

R. A. Fisher (1937). *Design of Experiments*, 2nd Edition. Oliver & Boyd, London.

D. A. Freedman (1981). Bootstrapping regression models. *The Annals of Statistics, 9, 1218–1228.*

A. R. Gallant (1987). *Nonlinear Statistical Models*. Wiley, New York.

V. P. Godambe & B. K. Kale (1992). Estimating functions: an overview. *Estimating Functions* (V. P. Godambe, ed.), Oxford University Press, Oxford, pp. 3–20.

P. Hall (1986) On the bootstrap and confidence intervals. *The Annals of Statistics, 14, 1431–1452.*

P. Hall (1987) On the bootstrap and likelihood based confidence intervals. *Biometrika*, 74, 481–493.

P. Hall (1992). *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, Berlin.

P. Hall & J. L. Horowitz (1996). Bootstrap critical values for tests based on generalized-method-of-moments estimators. *Econometrica*, 64, 891–916.

P. Hall & M. A. Martin (1996). Comment on a paper by DiCiccio and Efron. *Statistical Science*, 11, 189–228.

T. P. Hettmansperger (1984). *Statistical Inference Based on Ranks*. Wiley, New York.

K. Hinkelmann & O. Kempthorne (1994). *Design and Analysis of Experiments, Introduction to Experimental Design*. Wiley, New York.

D. V. Hinkley (1988). Bootstrap methods (with discussion). *Journal of the Royal Statistical Society Series B, 50, 321–337.*

F. Hu (1994). *Relevance Weighted Smoothing and a New Bootstrap*. Unpublished doctoral dissertation, The University of British Columbia, Vancouver, Canada.

F. Hu & J. D. Kalbfleisch (1997). Estimating equations and the bootstrap. *Selected Proceedings of the Symposium on Estimating Equations*. (I. V. Basawa, V. P. Godambe and R. L. Taylor, eds. IMS Lecture Note-Monograph Series, vol. 32, pp. 405–416.

F. Hu & J. V. Zidek (1995). A bootstrap based on the estimating equations of the linear model. *Biometrika*, 82, 263–275.

P. J. Huber (1981). *Robust Statistics*. Wiley, New York.

S. Huet & E. Jolivet (1989). Exactitude au second ordre des intervalles de confiance bootstrap pour les paramètres d'un modèle de régression non linéaire. *Comptes Rendus de l'Académie des Sciences de Paris, Séries I Mathématiques*, 308, 429–432.

S. Huet, E. Jolivet & A. Messéan (1990). Some simulation results about confidence intervals and bootstrap methods in nonlinear regression. *Statistics*, 21, 369–432.

O. Kempthorne (1954). *The Design and Analysis of Experiments*. Wiley, New York.

J. D. Kalbfleisch & D. A. Sprott (1970). Application of likelihood methods to models involving large numbers of nuisance parameters (with discussion). *Journal of the Royal Statistical Society Series B, 32, 175–208.*

R. Koenker & G. J. Bassett (1978). Regression quantiles. *Econometrica*, 84, 33–50.

S. M. S. Lee & G. A. Young (1996). Comment on a paper by DiCiccio and Efron. *Statistical Science*, 11, 189–228.

K.-Y. Liang & S. L. Zeger (1986). Longitudinal data analysis using generalised linear models. *Biometrika*, 73, 13–22.

R. Y. Liu (1988). Bootstrap procedures under some non-iid models. *The Annals of Statistics, 16, 1696–1708.*

W.-Y. Loh (1987). Calibrating confidence coefficients. *Journal of the American Statistical Association, 82, 155–162.*

W.-Y. Loh (1991). Bootstrap calibration for confidence interval construction and selection. *Statistica Sinica*, 1, 479–495.

J. Neyman & E. L. Scott (1948). Consistent estimates based on partially consistent observations. *Econometrica*, 16, 1–32.

M. I. Parzen, L. J. Wei & Z. Ying (1994). A resampling method based on pivotal estimating functions. *Biometrika*, 81, 341–50.

D. Rasch & E. Schimke (1983). Distribution of estimators on exponential regression. A simulation study. *Scandinavian Journal of Statistics, 10, 139–149.*

A. Ratkowsky (1983). *Nonlinear regression modeling*. Marcel Dekker, New York.

G. A. F. Seber and C. J. Wild (1989). *Nonlinear Regression*. Wiley, New York.

J. Shao & D. Tu (1995). *The Jackknife and Bootstrap*. Springer-Verlag, Berlin.

H. White (1982). Maximum likelihood estimation of misspecified models. *Econometrica*, 50, 1–26.

C. F. J. Wu (1986). Jackknife, bootstrap and other resampling methods in regression analysis (with discussion). *The Annals of Statistics, 14, 1261–1295.*

S. L. Zeger, K.-Y. Liang & P. A. Albert (1988). Models for longitudinal data: a generalized estimating equation approach. *Biometrics*, 44, 1049–1060.

Received 15 June 1999
Accepted 31 December 1999

Feifang HU: stahuff@nus.edu.sg
Department of Statistics and Applied Probability
National University of Singapore, Singapore 119260

John D. KALBFLEISCH: jdkalbfl@math.uwaterloo.ca
Department of Statistics and Actuarial Science
University of Waterloo, Waterloo, Ontario, Canada N2L 3G1