University of Michigan School of Public Health

The University of Michigan Department of Biostatistics Working Paper Series

Year 2004 *Paper* 33

Resampling methods for estimating functions with U-statistic structure

Wenyu Jiang* Jack Kalbfleisch†

*University of Waterloo, wjiang@uwaterloo.ca

†University of Michigan, jdkalbfl@umich.edu

Resampling methods for estimating functions with U-statistic structure

Wenyu Jiang and Jack Kalbfleisch

Abstract

Suppose that inference about parameters of interest is to be based on an unbiased estimating function that is U-statistic of degree 1 or 2. We define suitable studentized versions of such estimating functions and consider asymptotic approximations as well as an estimating function bootstrap (EFB) method based on resampling the estimated terms in the estimating functions. These methods are justified asymptotically and lead to confidence intervals produced directly from the studentized estimating functions. Particular examples in this class of estimating functions arise in La estimation as well as Wilcoxon rank regression and other related estimation problems. The proposed methods are evaluated in examples and simulations and compared with a recent suggestion for inference in such problems which relies on resampling an underlying objective functions with U-statistic structure.

Resampling methods for estimating functions with U-statistic structure

By WENYU JIANG

Department of Statistics and Actuarial Science,

University of Waterloo,

200 University Avenue West,

Waterloo, Ontario, Canada N2L 3G1

wjiang@uwaterloo.ca

AND JOHN D. KALBFLEISCH

Department of Biostatistics, School of Public Health,

University of Michigan,

1420 Washington Heights,

Ann Arbor, Michigan, USA 48109

jdkalbfl@umich.edu

April 15, 2004

Summary

Suppose that inference about parameters of interest is to be based on an unbiased estimating function that is U-statistic of degree 1 or 2. We define suitable studentized versions of such estimating functions and consider asymptotic approximations as well as an estimating function bootstrap (EFB) method based on resampling the estimated terms in the estimating functions. These methods are justified asymptotically and lead to confidence intervals produced directly from the studentized estimating functions. Particular examples in this class of estimating functions arise in L_a estimation as well as Wilcoxon rank regression and other related estimation problems. The proposed methods are evaluated in examples and simulations and compared with a recent suggestion for inference in such problems which relies on resampling an underlying objective functions with U-statistic structure.

Key words: Bootstrap; Estimating functions; L_a estimation; Resampling methods; U-statistics; Studentization;

1 Introduction

Let $Z_1, \ldots, Z_n$ be independent and identically distributed random vectors. Let $\theta \in \mathcal{R}^p$ be a vector of parameters and suppose that estimation of θ is based on an unbiased estimating function

$$S_n(\theta) = \binom{n}{K}^{-1} \sum_{1 \le i_1 < i_2 < \ldots < i_K \le n} h(Z_{i_1}, \ldots, Z_{i_K}; \theta) \tag{1}$$

which has the structure of a U-statistic of degree K. Thus, $E\{h(Z_1, \ldots, Z_k; \theta)\} = \mathit{0}$ where $\mathit{0}$ is a vector of zeros in $\mathcal{R}^p$ and h is symmetric in its arguments $Z_1, \ldots, Z_k$. Often such an estimating function arises from optimizing an underlying objective function,

$$U_n(\theta) = \binom{n}{K}^{-1} \sum_{1 \le i_1 < i_2 < \ldots < i_K \le n} H(Z_{i_1}, \ldots, Z_{i_K}; \theta), \tag{2}$$

which also has U-statistic structure. In this, $h = \partial H / \partial \theta$. The methods in this paper make inference about θ by resampling the estimating function (1) and are closely related and compared to a recent proposal of Jin, Ying and Wei (2001) which focuses on resampling the objective function (2).

We propose an estimating function bootstrap (EFB) method for resampling $S_n(\theta)$ with the purpose of constructing confidence regions for θ or components of θ. This is an extension of the work of Hu and Kalbfleisch (2000) who propose an estimating function based bootstrap method (EF-t) for inference based on linear estimating functions which are, in effect, U-statistics of degree $K = 1$. Their EF-t method provides an approximation to the distribution of a studentized version of the estimating function by bootstrap resampling of its estimated terms; the proposed EFB method is a natural generalization of the EF-t method to the estimating function (1). We consider primarily the case $K \le 2$ since this class subsumes the majority of applications of U-statistics in the literature. The case of general K is briefly discussed. In the EFB method, estimated terms in the studentized estimating function of $S_n(\theta)$ are weighted using a symmetric product of random variables generated from some selected distribution. When applied to linear estimating equations, this provides a useful generalization of the standard EF-t method as discussed in Hu and Kalbfleisch (2000). Though the resampling scheme of the EFB method for estimating function (1) treats the terms as the original sample, it links naturally to the classical bootstrap in which the resampling focuses on the original data. Hu and Kalbfleisch (2000) point out that their EF-t method is invariant under reparameterization and comment on the advantages that follow. This invariance property carries over to the EFB in U-statistic context as well.

Studentized estimating functions from $S_n(\theta)$ can be defined by utilizing known results for variance estimation of U-statistics. Hoeffding (1948) establishes the foundations of U-statistics, derives the theoretical form of the variance, and proves the asymptotic normality under quite mild conditions on h. Sen (1960) gives an estimator of the variance by utilizing a decomposition of S_n into identically distributed and asymptotically uncorrelated terms, and Arvesen (1969) derives a variance estimator by using jackknife techniques. These two variance estimations turn out to be essentially equivalent for U-statistics of any degree K. The asymptotic normality and order of the normal approximation are established by Callaert and Veraverbeke (1981) for the studentized U-statistics of degree two.

The classical bootstrap in the U-statistic framework has received much attention since bootstrap method was introduced in Efron (1979). The work of Bickel and Freedman (1981) indicates that a bootstrapped U-statistic has the same asymptotic distribution as the original U-statistic. Athreya et al. (1984) reveals the consistency of the bootstrapped variance estimator of a U-statistic and the asymptotic normality of the bootstrapped version of the studentized U-statistic. Helmers (1991) investigates the asymptotic improvements in accuracy obtained by approximating studentized U-statistics of degree two with Edgeworth expansions or by Efron's bootstrap techniques.

In most discussions, a U-statistic is scalar-valued and arises from estimating an unknown quantity such as a moment, a quantile, a correlation, or a regression coefficient. In this paper, however, we consider multidimensional U-statistics arising from estimating functions (1) where $S_n : \mathcal{R}^n \to \mathcal{R}^p$. Therefore, to studentize the estimating function $S_n(\theta)$ we develop multivariate variance estimators by extension of the work of Sen (1960) and Arvesen (1969).

In section 2, we generalize some results for U-statistics to p-dimensional estimating functions and propose the estimating function bootstrap (EFB) method. The procedure to define confidence intervals through studentized estimating functions is recalled and an iterative reweighted algorithm is suggested to address some computational issues. Section 3 presents a number of examples in which simulation studies suggest that the EFB method is very accurate and has substantial advantage over the resampling method of Jin et al. (2001). We conclude with some discussion in section 4.

2 U-statistics from Estimating Functions and the Bootstrap

Let $Z_i \in \mathcal{R}^q$, $i = 1, \ldots, n$ be independent random vectors from a distribution F. Let $\theta \in \mathcal{R}^p$ be a vector of unknown parameters. Suppose that inferences on θ are to be based on an estimating function of the form

$$S_n(\theta) = \binom{n}{2}^{-1} \sum_{1 \le i_1 < i_2 \le n} h(Z_{i_1}, Z_{i_2}; \theta), \qquad (3)$$

where the kernel $h(Z_1, Z_2; \theta)$ takes value on $\mathcal{R}^p$, is symmetric in Z_1, Z_2 and has expectation *0*. Our goal is to construct confidence regions for the parameter θ using the U-statistic feature of the underlying estimating function.

One example of an estimating function in the class (3) arises from minimizing the following objective function with respect to the regression parameter β, $\beta \in \mathcal{R}^p$,

$$U_n(\beta) = \binom{n}{2}^{-1} \sum_{1 \le i_1 < i_2 \le n} |Y_{i_1} - Y_{i_2} - (X_{i_1} - X_{i_2})^{\mathrm{T}}\beta|^a$$

where $1 \le a \le 2$, $Y_i = \gamma + X_i^{\mathrm{T}}\beta + e_i$ for a constant γ and independent and identically distributed errors $e_i, i = 1, \ldots, n$, with $E(e_i) = 0$. The regression estimator $\hat{\beta}$ is the solution to the estimating equation $S_n(\beta) = 0$ where

$$S_n(\beta) = \binom{n}{2}^{-1} \sum_{1 \le i_1 < i_2 \le n} \mathrm{sign}\{Y_{i_1} - Y_{i_2} - (X_{i_1} - X_{i_2})^{\mathrm{T}}\beta\}(X_{i_1} - X_{i_2})|Y_{i_1} - Y_{i_2} - (X_{i_1} - X_{i_2})^{\mathrm{T}}\beta|^{a-1}. \tag{4}$$

For the case with $a = 1$, the estimating function (4) gives rise to the Wilcoxon rank regression estimator (Hettmansperger, 1984). For $a = 2$, the above estimating function leads to an approach of least squares. In general, the derivatives of $S_n(\beta)$ are not well behaved for $1 \le a < 2$. For example, if $p = 1$ and $1 \le a < 2$, the derivative of $S_n(\beta)$ is undefined whenever β coincides with a value of $(Y_{i_1} - Y_{i_2})/(X_{i_1} - X_{i_2})$ for $1 \le i_1 < i_2 \le n$. Estimating functions with badly behaved derivatives also appear in ordinary L_a regression when $1 \le a < 2$. It then becomes difficult to apply traditional inference procedures involving the sandwich variance estimator and Wald-type statistic based on $\hat{\beta}$. To circumvent this problem, we base inferences on the studentized estimating function.

2.1 Some Results for U-statistics

The variance of S_n of degree $K = 2$ can be found following derivations of Hoeffding (1948) or Serfling(1980) for scalar valued U-statistics. Let $\{a_1, a_2\}$ and $\{b_1, b_2\}$ be sets of distinct integers from $\{1, \ldots, n\}$ with exactly c integers in common, and define

$$\zeta_c = E\{h(Z_{a_1}, Z_{a_2}; \theta) h(Z_{b_1}, Z_{b_2}; \theta)^{\mathrm{T}}\}$$

for $c = 1, 2$. It can then be seen that

$$\begin{aligned} \mathrm{var}\{S_n\} &= \binom{n}{2}^{-1} \{2(n-2)\zeta_1 + \zeta_2\} \\ &= \frac{4\zeta_1}{n} + \mathrm{O}_p(n^{-2}). \end{aligned}$$

In general, the variance of S_n of degree K has the form $\mathrm{var}\{S_n\} = n^{-1}K^2\zeta_1 + \mathrm{O}_p(n^{-2})$ where $\zeta_1 = E\{h(Z_{a_1}, \ldots, Z_{a_K}; \theta) h(Z_{b_1}, \ldots, Z_{b_K}; \theta)\}$ and $\{a_1, \ldots, a_K\}$, $\{b_1, \ldots, b_K\}$ are sets of distinct integers from $\{1, \ldots, n\}$ with exactly one elements in common. An alternative expression of the variance follows from writing the U-statistic as a sum of its projection and orthogonal complement (Serfling (1980), Callaert and Veraverbeke (1981)).

The asymptotic normality for S_n of degree $K = 2$ is a direct consequence of Theorem 7.1 of Hoeffding (1948). In fact, the asymptotic normality for an $\mathcal{R}^p$-valued U-statistic of any degree K can be established in the same way.

Theorem 2.1. Let S_n be an unbiased estimating function of the form (1) for $\theta \in \mathcal{R}^p$. If $E\{h_i(Z_1, \ldots, Z_K; \theta)\}^2 < \infty$ for all components $i = 1, \ldots, p$ of the vector h and the determinant $|\zeta_1| > 0$, then as $n \to \infty$,

$$n^{1/2} S_n \to \mathrm{N}_p(0, 4\zeta_1) \text{ in distribution.}$$

Since the variance of S_n is unknown in general, inference is based upon a studentized version of S_n. We consider two ways of studentizing S_n, both of which adapt established variance estimators in the literature to S_n.

The first variance estimator of S_n follows from Sen (1960). For a U-statistic of degree $K = 2$, define

$$q_i(\theta) = \frac{1}{n-1} \sum_{j:j \neq i} h(Z_i, Z_j; \theta)$$

for $i = 1, \ldots, n$. The q_i's are identically distributed and $S_n(\theta) = n^{-1} \sum_{i=1}^{n} q_i(\theta)$. Let s_q^2 be the sample covariance matrix of $q_1, \ldots, q_n$. For scalar-valued U-statistics, Sen (1960) shows that the q_i's are asymptotically uncorrelated and $s_q^2 \to \zeta_1$ in probability as $n \to \infty$. In the vector case, the corresponding variance estimator for S_n is

$$\begin{aligned} V_S(\theta) &= 4n^{-1}s_q^2 \\ &= \frac{4}{n(n-1)} \sum_{i=1}^{n} \{q_i(\theta) - S_n\}^{\otimes 2} \\ &= \frac{4}{n^2(n-1)} \sum_{1 \leq i < j \leq n} \{q_i(\theta) - q_j(\theta)\}^{\otimes 2}. \end{aligned} \tag{5}$$

where $a^{\otimes 2} = aa^{\mathrm{T}}$ for a column vector a.

In the same way, Sen (1960) defines, for U-statistic of degree K,

$$q_i(\theta) = \binom{n-1}{K-1}^{-1} \sum_{C_i} h(Z_i, Z_{l_1}, \ldots, Z_{l_{K-1}}; \theta)$$

where $C_i = \{(l_1, \ldots, l_{K-1}) : 1 \leq l_1 < \ldots < l_{K-1} \leq n, \text{ and } l_1, \ldots, l_{K-1} \neq i\}$. As above $S_n(\theta) = n^{-1} \sum_{i=1}^{n} q_i(\theta)$, and the resulting variance estimator of $S_n(\theta)$ is

$$\begin{aligned} V_S(\theta) &= K^2 n^{-1} s_q^2 \\ &= \frac{K^2}{n^2(n-1)} \sum_{1 \leq i < j \leq n} \{q_i(\theta) - q_j(\theta)\}^{\otimes 2}. \end{aligned} \tag{6}$$

For any $K = 1, 2, \ldots$, the corresponding studentized U-statistic is

$$S_{t_S}(\theta) = \{V_S(\theta)\}^{-1/2} S_n(\theta). \tag{7}$$

The second variance estimator of S_n of degree K follows from Arvesen (1969), and arises from an argument utilizing the jackknife. The resulting studentized U-statistic has received much theoretical consideration in Callaert and Veraverbeke (1981), Athreya et al. (1984) and Helmers (1991). Let

$$\hat{\mu}_i = nS_n - (n-1)S^i_{n-1},$$

where S^i_{n-1} is the U-statistic based upon $Z_1, \ldots, Z_{i-1}, Z_{i+1}, \ldots, Z_n$. Let $s^2_{\hat{\mu}}$ be the sample covariance matrix of $\hat{\mu}_1, \ldots, \hat{\mu}_n$. Then $V_J(\theta) = s^2_{\hat{\mu}}/n$ is a jackknife estimator of the variance of S_n and it is easy to see that

$$V_J(\theta) = \frac{K^2(n-1)}{n^2(n-K)^2} \sum_{1 \le i < j \le n} \{q_i(\theta) - q_j(\theta)\}^{\otimes 2}.$$

The corresponding studentized U-statistic is

$$S_{t_J}(\theta) = \{V_J(\theta)\}^{-1/2} S_n(\theta).$$

The two variance estimators are asymptotically equivalent and their ratio is exactly $(n - K)^2(n-1)^{-2}$. Consequently, the statistics S_{t_S} and S_{t_J} are also asymptotically equivalent.

When a scalar-valued U-statistic is studentized by these variance estimators, Sen (1960), Arvesen (1969), Callaert and Veraverbeke (1981) and Helmers (1991) have shown that the resulting studentized U-statistic is asymptotically normally distributed.

From these results and through considering asymptotic results for linear combinations of the components of S_n, the following multivariate central limit theorem can be established. It is shown in the Appendix that the conclusion holds for U-statistics of any degree K.

Theorem 2.2. Let $V_S(\theta)$ and $S_{t_S}(\theta)$ be defined as in (6) and (7). Under the same conditions as Theorem 2.1,

$$\frac{n}{K^2} V_S(\theta) \to \zeta_1 \text{ in probability,} \tag{8}$$

and

$$S_{t_S}(\theta) \to \mathrm{N}_p(\mathit{0}, \mathbf{1}) \text{ in distribution} \tag{9}$$

as $n \to \infty$, where $\mathbf{1}$ is a $p \times p$ identity matrix.

Asymptotic approximations can be used to obtain confidence regions for θ and one can estimate component-wise confidence intervals using the procedures described in section 2.3. Alternatively, we can use resampling methods to approximate the distribution of S_t.

2.2 Resampling the Estimating Function

To resample the estimating function (3), we propose the following generalization to the EF-t method of Hu and Kalbfleisch (2000). We proceed by replacing θ with $\hat{\theta}$ in the terms of the estimating function (3) and estimating the distribution of S_n or its studentized version S_{t_S} by resampling the estimated terms.

Specifically, we define **the EFB method for S_{t_S} and $K = 2$** as follows:

i. Generate $(V_1, \ldots, V_n)$ from Multinomial$(n, \frac{1}{n}, \ldots, \frac{1}{n})$.

ii. Let $S_n^* = n^{-1} \sum_{i=1}^n V_i \tilde{q}_i^*$ where

$$\tilde{q}_i^* = \frac{1}{n-1} \sum_{l:l \neq i} V_l h(Z_i, Z_l; \hat{\theta}).$$

Following (5), the variance estimator of S_n^* is

$$V_S^* = \frac{4}{n^2(n-1)} \sum_{1 \le i < j \le n} V_i V_j \left(\tilde{q}_i^* - \tilde{q}_j^* \right)^{\otimes 2} \tag{10}$$

iii. Finally we set

$$S^*_{t_S} = V_S^{*-1/2} S^*_n. \tag{11}$$

We proceed by repeating the above steps a large number B times and the empirical distribution of $S^*_{t_S}$ provides an approximation to that of S_{t_S}.

The next theorem states that the resampled statistic $S^*_{t_S}$ has the same asymptotic distribution as S_{t_S}, and its proof is given in the Appendix.

Theorem 2.3. Let S_n be an estimating function of the form (3). Suppose that Assumptions 1–5 in the Appendix hold. If the kernel function of S_n has the property that $h(z, z; \theta) = 0$ for any z, then under the EFB procedure,

$$\frac{n}{4} V_S^* \to \zeta_1 \text{ in probability,} \tag{12}$$

and

$$S^*_{t_S} \to \mathrm{N}_p(0, \mathbf{1}) \text{ in distribution} \tag{13}$$

as $n \to \infty$, where V_S^* and $S^*_{t_S}$ are given by (10) and (11).

Some remarks:

1. Since the variance estimators are proportional, the EFB method based on S_{t_J} gives identical results to that based on S_{t_S}.

2. The multinomial weights $(V_1, \ldots, V_n)$ give rise to a nonparametric bootstrap sample $(Z_1^*, \ldots, Z_n^*)$ from the observations $(Z_1, \ldots, Z_n)$. If the kernel function h satisfies $h(z, z; \theta) = 0$ for any z, then $S^*_{t_S}$ is identical to the bootstrap version of S_{t_S} obtained by replacing $Z_1, \ldots, Z_n$ and θ by $(Z_1^*, \ldots, Z_n^*)$ and $\hat{\theta}$.

3. Other choices of the weights are possible. For example, one might choose $V_1, \ldots, V_n$ to be independent and identically distributed Poisson(1) variates, or indeed as independent and identically distributed variates from any distribution with unit mean and unit variance.

The bootstrap approach that we have described above for the U-statistic of degree $K = 2$ generalizes easily to the case where $S_n(\theta)$ is a U-statistic of degree K for any $K \geq 1$. We can restate **the EFB method for S_{t_S} and** $K \geq 1$ as follows:

i. Generate $(V_1, \ldots, V_n)$ from Multinomial$(n, \frac{1}{n}, \ldots, \frac{1}{n})$.

ii. Let $S_n^* = n^{-1} \sum_{i=1}^n V_i \tilde{q}_i^*$ where

$$\tilde{q}_i^* = \binom{n-1}{K-1}^{-1} \sum_{C_i} V_{l_1} V_{l_2} \cdots V_{l_{K-1}} h(Z_i, Z_{l_1}, \ldots, Z_{l_{K-1}}; \hat{\theta}),$$

and let

$$V_S^* = \frac{K^2}{n^2(n-1)} \sum_{1 \leq i < j \leq n} V_i V_j \left(\tilde{q}_i^* - \tilde{q}_j^*\right)^{\otimes 2}$$

iii. Set $S_{t_S}^* = V_S^{*-1/2} S_n^*$.

When $K = 1$, the above procedure coincides with the EF-t method for linear estimating functions proposed by Hu and Kalbfleisch (2000). For K=2, it reduces to the method described earlier.

2.3 Confidence Regions via Studentized Estimating Functions

Denote the studentized estimating function of $S_n(\theta)$ by $S_t(\theta)$ where $S_t = S_{t_S}$ or $S_t = S_{t_J}$. If θ is a scalar, we define a one-sided $100(1-\alpha)\%$ confidence interval for θ by the collection

$\{\theta : S_t(\theta) > \hat{S}_{t_\alpha}\}$, where $\hat{S}_{t_\alpha}$ stands for an estimate of the αth quantile of S_t. We can either take $\hat{S}_{t_\alpha} = Z_\alpha$ according to the normal approximation to S_t, or take $\hat{S}_{t_\alpha} = S^*_{t_\alpha}$, where $S^*_{t_\alpha}$ is the empirical αth quantile from the replications of the EFB procedure. If $S_t(\theta)$ is monotone non-increasing in θ, an aproximate one-sided $100(1-\alpha)\%$ confidence interval is given by $(-\infty, \hat{\theta}_{1-\alpha}]$ where $\hat{\theta}_{1-\alpha}$ is simply the solution to $S_t(\theta) = \hat{S}_{t_\alpha}$. In this case, the EFB procedure that approximates the distribution of $S_t(\theta)$ results in significant reduction in computation of confidence intervals. It requires solving the $S_t(\theta)$ only at the terminals of the interval, and not for each resampling run as would be required for the classical bootstrap.

If θ is a vector, an approximate one-sided $100(1-\alpha)\%$ component-wise confidence interval for θ is $(-\infty, \theta^*_{i,1-\alpha}]$ for $i = 1, \ldots, p$. The confidence limit $\theta^*_{i,1-\alpha}$ is defined as the empirical $(1-\alpha)$ quantile of the set $\{\theta_i^{*(b)} : S_t(\theta^{*(b)}) = u^{(b)}, b = 1, \ldots, B\}$ for a large replication number B. When the normal approximation to S_t is used, we can implement this by repeatedly generating the p components of u as independent standard normal variates a large number B times. In the EFB procedure, we take $u = S^*_t$ where S^*_t is the resampled value. Such confidence procedures involve solving equations $S_t(\theta) = u^{(b)}$ repeatedly. Note that the left side of the equation is always the same. Such problems are often easier to handle and are numerically more stable than methods based on the classical bootstrap where resampling would result in a new estimating function each time.

In this derivation of component-wise intervals, $S_t(\theta)$ is being used as a multivariate pivotal and, in essence, $\theta^{*(b)}$ is defining a distribution for θ that is reminiscent of Fisher's fiducial distribution. Jin et al. (2001) carry out similar calculations in their approach of minimizing the objective function and use the term "fiducial intervals". It should be noted that even if $S_t(\theta)$ is exactly pivotal with known distribution, the resulting intervals from the marginalization procedure defined above or in Jin et al. (2001) would not be exact confidence intervals. Nonetheless, the intervals for θ_i are asymptotic confidence intervals as $n \to \infty$.

Solving $S_t(\theta) = u$ can be somewhat more complicated than solving $S_n(\theta) = u'$. We have found that numerical problems are typically easily handled through the following iteratively reweighted method. Begin with an initial estimate $\theta^{(0)}$. At the kth step, solve $\theta^{(k)}$ from

$$S_n(\theta^{(k)}) = V(\theta^{(k-1)})^{1/2}u, \quad k = 1, 2, \dots \; . \tag{14}$$

A convenient choice of initial value is often $\theta^{(0)} = \hat{\theta}$.

3 Examples

In this section, we implement the EFB method in estimating functions which are U-statistics of degree $K = 1$ and $K = 2$ with examples arising from L_1 regression and those taking the form (4) with $a = 1$ and $a = 1.5$. For comparison, the resampling method of Jin et al. (2001) and the normal approximations to $S_t(\theta)$ are also examined.

3.1 Description of Methods and Simulation Study

Since the derivatives of these estimating functions are poorly behaved, traditional approaches involving Wald-type statistics using sandwich variance estimators are infeasible and/or behave badly. Inference procedures that focus on the estimating function or on the objective function are more appealing.

Jin et al. (2001), henceforth JYW, consider inference methods based on an objective function of the form (2) in order to obtain an estimate of the standard error of $\hat{\theta}$. Their resampling method, the JYW method, involves perturbing or weighting the terms of the objective function $U_n(\theta)$ with symmetric sums of independent and identically distributed weights. For each replication of the resampled objective function, the minimizer θ^* is found. The purpose of their approach is to assess the variation of $\hat{\theta} - \theta$ from the resampling distribution of $\theta^* - \tilde{\theta}$, where $\tilde{\theta}$ indicates the observed value of $\hat{\theta}$. A direct application of this idea yields $\text{pr}\{(\hat{\theta} - \theta)_i \geq (\theta^* - \tilde{\theta})_{i,\alpha}\} \approx 1 - \alpha$ where $(\theta^* - \tilde{\theta})_{i,\alpha}$ stands for the αth quantile of the ith component of $(\theta^* - \tilde{\theta})$, $i = 1, \ldots, p$.

For the JYW method, we use $(-\infty, \hat{\theta} - (\theta^* - \tilde{\theta})_{i,\alpha}]$ as a $100(1-\alpha)\%$ confidence interval of θ_i. We assign the resampling weights as indpendent $Gamma(1,1)$ in Example 1 and $Gamma(.25, .5)$ in Examples 2 and 3 which are in conformity with recommendations of JYW. In addition to the JYW and EFB methods, we also consider NORM-S and NORM-J which refer to normal approximations to S_{t_S} and S_{t_J} respectively.

Example 1. L_1 regression.

In a regression problem, suppose that β is to be estimated by minimizing the objective function $U_n(\beta) = n^{-1} \sum_{i=1}^n |Y_i - X_i^{\mathrm{T}}\beta|$ and the estimating function is

$$S_n(\beta) = \frac{1}{n} \sum_{i=1}^n \text{sign}(Y_i - X_i^{\mathrm{T}}\beta) X_i.$$

We suppose that the data arises from the following two models:

1. *homoscedastic errors:* $Y_i = \beta x_i + e_i$, $i = 1, \ldots, n$, where $n = 20$, the e_i's are independent and identically distributed from $\mathrm{N}(0, 0.25)$, $x_i = -2, -1.9, \ldots, -1.1, 1.1, \ldots, 2$ and $\beta = 1$.

2. *heteroscedastic errors:* $Y_i = \beta x_i + x_i^2 e_i$, $i = 1, \ldots, n$ with n, x_i, β and e_i defined as above.

An example for L_a regression with $a = 1.5$ is investigated by Hu and Kalbfleisch (2000) with the same distributional inputs.

Table 1 (homoscedastic errors) and Table 2 (heteroscedastic errors) present coverage percentages, average end points and standard errors of the end points for one-sided confidence intervals at six nominal levels. In this case, the EF-t approach of Hu and Kalbfleisch (2000) coincides with our EFB method and the NORM method refers to the normal approximation to $S_{t_S}(\theta) = S_{t_J}(\theta) = S_t(\theta)$. Each simulation study consists of $N = 10,000$ replications of $B = 1,000$ resampling runs.

Insert Table 1 and Table 2 about here

Example 2. Wilcoxon rank regression.

Let $Y_i = \gamma + x_i^{\mathrm{T}}\beta + e_i$ where e_i are independent and identically distributed with $E(e_i) = 0$, $i = 1, \dots, n$. The slope β is estimated from the Wilcoxon rank regression:

$$S_n(\beta) = \binom{n}{2}^{-1} \sum_{1 \le i_1 < i_2 \le n} \mathrm{sign}\{Y_{i_1} - Y_{i_2} - (X_{i_1} - X_{i_2})^{\mathrm{T}}\beta\}(X_{i_1} - X_{i_2}).$$

We consider homoscedastic errors $e_i \sim \mathrm{N}(0, .25)$ for $i = 1, \dots, 20$. Table 3 presents the simulation results based on 10,000 replications of 1,000 resampling runs for a single parameter example where $\gamma = 0$, $\beta = 1$ and $x_1, \dots, x_{10} = 1$, $x_{11}, \dots, x_{20} = 0$. Table 4 presents the results of a two-dimensional parameter example where $\gamma = 0$, $\beta_1 = 1, \beta_2 = 0.5$, $(x_{1,1}, \dots, x_{20,1}) = (-2, -1.9, \dots, -1.1, 1.1, \dots, 2)$, $x_{1,2}, \dots, x_{5,2} = 0, x_{6,2}, \dots, x_{10,2} = 1, x_{11,2}, \dots, x_{15,2} = 2, x_{16,2}, \dots, x_{20,2} = 3$ and each simulation consists of $5,000$ replications of $1,000$ resampling runs.

Insert Table 3 and Table 4 about here

Example 3. L_a $(1 < a < 2)$ regression on paired differences.

Let $Y_i = \gamma + x_i^{\mathrm{T}}\beta + e_i$ where e_i are independent and identically distributed with $E(e_i) = 0$, $i = 1, \ldots, n$. The estimating function of β is given by:

$$S_n(\beta) = \binom{n}{2}^{-1} \sum_{1 \le i_1 < i_2 \le n} \mathrm{sign}\{Y_{i_1} - Y_{i_2} - (X_{i_1} - X_{i_2})^{\mathrm{T}}\beta\}(X_{i_1} - X_{i_2})|Y_{i_1} - Y_{i_2} - (X_{i_1} - X_{i_2})^{\mathrm{T}}\beta|^{a-1}.$$

In Table 5, we report a simulation with independent and identically distributed errors $e_i \sim \mathrm{N}(0, .25)$ where $i = 1, \ldots, 20$, $a = 1.5$, $\gamma = 0$, $\beta = 1$, $p = 1$ and $x_1, \ldots, x_8 = 1$, $x_9, \ldots, x_{20} = 0$. Each method is investigated in 10,000 simulation replications of 1,000 resampling runs.

Insert Table 5 about here

The JYW method appears less accurate in terms of coverage and gives less stable confidence limits than the NORM and EFB methods. Generally speaking, the normal approximation to the estimating function yields confidence intervals with good coverage properties, although the EFB method provides some further improvement. In the simulations, we found that all of these methods work well even when the errors are heteroscedastic; for simplicty of presentation, results for the heteroscedastic case are reported in the first example only (Table 2).

Jin et al. (2001) examine the JYW method in a simulation study of L_1 regression with sample size $n = 41$, and found it to perform reasonably well. With larger sample sizes like this, we found the normal approximations to be very accurate and again offered an improvement on the JYW method. With the larger sample size, however, there was relatively little additional benefit accrued through the EFB approach. Jin et al. (2001) have not investigated their method in problems involving U-statistics of degree two such as the Wilcoxon regression.

3.2 Computational Issues

In examples 1 and 2, the estimating functions are derived through minimizing objective functions in L_1 distance. Typically, such estimating functions are not continuous in the parameters and cannot be solved exactly. Instead we define $\hat{\beta} = \arg\ \min_\beta ||S_n(\beta)||$ where $||a|| = \max_{1\le i\le p} |a_i|$ for $a \in \mathcal{R}^p$. When the iteratively reweighted algorithm is applied as in (14), the kth step solution $\beta^{(k)}$ has been redefined accordingly as $\beta^{(k)} = \arg\ \min_\beta ||S_n(\beta) - V(\beta^{(k-1)})^{1/2}u||$. In the EFB approach, we take $S_t^* = V_S^{*-\frac{1}{2}}\{S^* - S(\hat{\theta})\}$ to adjust for the discontinuity in $S_n(\theta)$. When n is small, such adjustment results in slight corrections in coverage property of the confidence intervals. Example 3 deals with a continuous estimating function involving L_a criterion with $1 < a < 2$ and no adjustment is needed in the numerical computation.

We use the algorithm of Barrodale and Roberts (1973, 1974) to solve $S_n(\beta) = 0$ in Example 1 for the L_1 regression estimator $\hat{\beta}$. The alogorithm is also described in Bloomfield and Steiger (1983). Example 2 can be fitted through the same algorithm as an extended L_1 regression problem with $n(n-1)/2$ data points, each consists of the differences in response $Y_i - Y_j$ and covariates $X_i - X_j$ between one pair of distinct individuals. When we solve the equations of the form $S_n(\beta) = u'$ in Example 1, we can transform it into solving an L_1 regression of $n+1$ data points,$\sum_{i=1}^{n+1} \mathrm{sign}(Y_i - X_i^{\mathrm{T}}\beta)X_i=0$, where Y_{n+1} takes an extremely large number and $X_{n+1} = -nu'$ so that $Y_{n+1} - X_{n+1}^{\mathrm{T}}\beta > 0$ always holds. Similarly, solving $S_n(\beta) = u'$ in Example 2 is equivalent to solving an L_1 regression of $n(n-1)/2+1$ data where the extra data point takes an extremely large number as response and $-n(n-1)u'/2$ as convariates. Parzen et al. (1994) and Jin et al. (2001) handles similar computational issues in a similar way.

We can view the computation involved in Examples 1 and 2 as minimizing objective functions of the form (2) with $K = 1$ or $K = 2$ and its extended form with one extra data point in each problem as described above. An alternative algorithm is also used in simulation to minimize such objective functions on account of their continuity and convexity properties with respect to $\beta \in \mathcal{R}^p$. Denote the objective function by $f(\beta)$. There always exists a non empty set B such that $f(\beta)$ is minimized when $\beta \in B$. Let $\beta^{(0)}$ be an initial value of the algorithm, $\beta^{(0)} \in \mathcal{R}^p$. Let $s > 0$ be the current choice of step length. Denote the value of the kth iteration by $\beta^{(k)}$, we define

$$\beta^{(k+1)} = \beta^{(k)} - s \times \frac{f'(\beta^{(k)})}{1 + ||f'(\beta^{(k)})||}$$

where $f'(\cdot)$ is the derivative with respect to β. Since $\beta^{(k+1)}$ moves from $\beta^{(k)}$ along the negative direction of $f'(\beta^{(k)})$, we can always adjust step length s such that $f(\beta^{(k+1)})$ is smaller than $f(\beta^{(k)})$. This iteration can be repeated until $|f(\beta^{(k+1)}) - f(\beta^{(k)})|$ falls below some given tolerance level. This algorithm turns out to be a useful alternative to the algorithm of Barrodale and Roberts (1973, 1974) and appears quite stable even in Example 2 with two parameters.

In Example 3, we use a bisection method to solve the estimating functions for the single slope parameter involving L_a $(1 < a < 2)$ distance. If a multiple parameter example is considered in L_a $(1 < a < 2)$ regression, a more advanced optimization procedure such as convex programming is required.

In the one parameter examples, the EFB and the normal approximation to $S_t(\beta)$ only solve $S_t(\beta)$ at the estimated percentiles and are much faster than the resampling method of Jin et al. (2001). In the two parameter example (Table 4), all methods are comparable in computing time due to the requirement of solving for estimates for each resampling run, the method of Jin et al. (2001) is often faster than the rest. When each normal approximation or the EFB method is applied in Table 4, the function $S_t(\beta)$ can not be solved at about .01% of the simulated standard normal variates or .02% of the bootstrapped quantities S_t^*, while all runs can be solved in the simulation for the method of Jin et al. (2001).

4 Discussion

The EFB method adds a powerful tool for estimation when the estimating or objective function has U-statistic structure. These methods are particularly useful when it is difficult to apply conventional approaches that focus on studentized Wald-tlype statistics based on asymptotic distributions for $\hat{\theta}$. Such examples arise, for instance, when the objective functions are related to L_a norm with $1 \leq a < 2$. Even in more regular type problems, however, methods based on the score tend to have better coverage properties than those based on the Wald type statistics. As noted by Hu and Kalbfleisch (2000), the EFB procedures are invariant under reparametrization which is an advantage over the Wald procedures where the choice of the parametrization can be very important.

As noted above, Jin et al. (2001) offer an alternative approach to inference based on perturbations or resampling of the objective function. We have found that the methods based on the estimating function improve on coverage probabilities and on the stability of confidence intervals in all the examples considered. In the EFB methods, we have used resampling weights that correspond to those in the usual Efron bootstrap. One could also study other weights closer to those of Jin et al. (2001) but the weights we have chosen seem more simply interpreted in the context of classical and familiar approaches to resampling.

In the one parameter case, the EFB approach has considerable computational advantage over other methods since one needs only solve the estimating function at the appropriate quantiles of the bootstrap distribution. In considering multiparameter problems, we have considered a method whereby the estimates are obtained following each bootstrap simulation and quantile methods are used to determine intervals for parameters. It is of some considerable advantage that the left side of the studentized estimating equation is the same for each simulation; an iteratively reweighted algorithm simplifies the computation.

For S_n of degree $K \leq 2$, we have shown that the studentized estimating function $S_t(\theta)$ is first order accurate. This is similar to the results of Jin et al. (2001) who established the first order accuracy of their resampling method as well. For S_n of degree one, Hu and Kalbfleisch (2000) show that the EFB leads to higher order approximation to the studentized estimating function than the Normal approximation for simultaneous estimation of all parameters in the model. Higher order approximations to studentized U-statistics of degree two have been studied by Callaert and Veraverbeke (1981) and Helmers (1991). On account of these studies, it is possible that the EFB method may yield higher order accuracy than the normal approximation when the estimating function is a U-statistic of degree $K = 2$ and perhaps for general K. The estimating function S_n of degree $K > 2$, however, seems rarely to appear in practice, and it is not surprising that studentized U-statistics of degree higher than two have received little attention in the literature.

When $S_n(\theta)$ is a U-statistic of degree $K \geq 1$, simultaneous confidence regions for elements of the parameter vector can be obtained by solving $S_t(\theta)^{\mathrm{T}} S_t(\theta) = q_\alpha^*$ at estimated quantiles q_α^* obtained from either the chi-squared or the bootstrap approximation. When $S_n(\theta)$ is differentiable in θ, the approach suggested by Hu and Kalbfleisch (2000) for the case $K = 1$

could be generalized to obtain confidence regions for subsets of the parameters. Essentially, they propose use of generalized score statistics (see Boos, 1992) and develop methods for resampling these statistics. This approach avoids the need to resolve the estimating equation with each bootstrap simulation and so yields very simple methods which are computationally very easy to implement.

Another approach aimed at avoiding repeated solutions of the full estimating equations has been suggested by He and Hu (2002). In the context of M estimators in linear regression, they propose a bootstrap procedure that solves sequentially for one component of the parameter vector at a time. By this approach, they produce a Markov chain to approximate the distribution of the estimator. Further investigation is needed to see whether those methods are applicable in the presence estimating functions with dependent terms, such as those with U-statistic structure of degree $K > 1$.

5 Appendix

For a matrix M, let $||M|| = \max_{i,j}\{|m_{ij}|\}$. It is easy to see that

Lemma A.1 A sequence of square random matrices W_n converges to W in probability if and only if $a^{\mathrm{T}} W_n a \to a^{\mathrm{T}} W a$ in probability for all $a \in \mathcal{R}^p$.

Proof of Theorem 2.2: We prove the results for U-statistic S_n in (1) of any degree $K \geq 1$. For any $a \in \mathcal{R}^p$, let $S_n^\dagger(\theta) = a^{\mathrm{T}} S_n(\theta)$. The Sen (1960) variance estimator of $S_n^\dagger(\theta)$ is $V_S^\dagger(\theta) = a^{\mathrm{T}} V_S(\theta) a$ where V_S is defined in (6). From Sen (1960), it can be seen that $nK^{-2} V_S^\dagger(\theta) \to a^{\mathrm{T}} \zeta_1 a$ in probability. From Lemma A.1, we obtain (8) and (9) follows from Slutsky's theorem and Theorem 2.1.

In preparation for Theorem 2.3, let

$$
\begin{aligned}
A_n(\theta) &= \frac{1}{n^3}\sum_{i=1}^n\sum_{j=1}^n\sum_{k:k\neq j} h(Z_i,Z_j;\theta)h^{\mathrm{T}}(Z_i,Z_k;\theta),\\
B_n(\theta) &= \frac{1}{n^3}\sum_{i=1}^n\sum_{j=1}^n\sum_{k:k\neq j} h(Z_i,Z_j;\theta)h^{\mathrm{T}}(Z_i,Z_k;\theta_0),\\
\tilde{\zeta}_2(\theta) &= \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n h^{\otimes 2}(Z_i,Z_j;\theta), \qquad \text{(A1)}\\
C_n(\theta) &= \frac{1}{n^2}\sum_{i=1}^n\sum_{j=1}^n h(Z_i,Z_j;\theta)h^{\mathrm{T}}(Z_i,Z_j;\theta_0).
\end{aligned}
$$

The following assumptions are used to establish the results of Theorem 2.3.

Assumptions:

1. The true value of θ is θ_0, an interior point of the parameter space Ω. Further, $E_{\theta_0}\{S_n(\theta)\}=0$ has the unique solution θ_0 and $\hat{\theta}\rightarrow\theta_0$ in probability.

2. $||\zeta_2||<\infty$ and the determinant $|\zeta_1|>0$.

3. $A_n(\hat{\theta})$, $A_n(\theta_0)$, $B_n(\hat{\theta})$, and $B_n(\theta_0)$ converge in probability to ζ_1.
   $\tilde{\zeta}_2(\hat{\theta})$, $\tilde{\zeta}_2(\theta_0)$, $C_n(\hat{\theta})$, and $C_n(\theta_0)$ converge in probability to ζ_2.

4. There exists a function $f(Z_1,Z_2)$ with $E\{f(Z_1,Z_2)\}<\infty$ and a neighbourhood $\mathcal{B}(\theta_0)$ of θ_0 such that $\sup_{\theta\in\mathcal{B}(\theta_0)}||\tilde{\zeta}_2(\theta)||\leq f(Z_1,Z_2)$.

5. For almost every sequence of random vectors $Z_1,Z_2,\ldots,Z_n,\ldots$ in $\mathcal{R}^p$ and any $\epsilon>0$,

$$\sum_{i=1}^n ||T_{n,i}||^2 I\{||T_{n,i}||>\epsilon\}\rightarrow 0,$$

   where $T_{n,i}=\tilde{\zeta}_1(\hat{\theta})^{-1/2}\sum_{j=1}^n h(Z_i,Z_j;\hat{\theta})/n$ and $\tilde{\zeta}_1(\hat{\theta})=n^{-2}(n-1)^3V_S(\hat{\theta})/4$.

The Assumption 3 is needed to be able to substitute $\hat{\theta}$ into the estimating function before bootstrapping the terms. It is of some interest to note that these results would follow from conditions of uniform convergence. For example, if $A_n(\theta)$ converges to its probability limit $\zeta_1(\theta,\theta_0) = E_{\theta_0}\{h(Z_1, Z_2;\theta)h^{\mathrm{T}}(Z_1, Z_3;\theta)\}$ uniformly in a neighbourhood of θ_0 and $\zeta_1(\theta,\theta_0)$ is continuous at θ_0, then $A_n(\theta_0) \to \zeta_1$ and $A_n(\hat{\theta}) \to \zeta_1$.

Proof of Theorem 2.3: Some approaches in the following proof derive from Theorems 1 and 4 of Athreya et al. (1984) and Theorem 7.1 of Hoeffding (1948).

Let F_n be the empirical distribution function of $Z_1, \ldots, Z_n$, and let $Z_1^*, \ldots, Z_n^*$ be a random sample from F_n. The expressions (13) and (14) can be written as

$$V_S^*(\hat{\theta}) = \frac{4}{n(n-1)} \sum_{i=1}^{n} \left\{ \frac{1}{n-1} \sum_{j:j\neq i} h(Z_i^*, Z_j^*;\hat{\theta}) - S_n^*(\hat{\theta}) \right\}^{\otimes 2}$$

and $S_{t_S}^*(\hat{\theta}) = V_S^*(\hat{\theta})^{-1/2} S_n^*(\hat{\theta})$ where $S_n^*(\hat{\theta}) = 2n^{-1}(n-1)^{-1}\sum_{1\leq i_1<i_2\leq n} h(Z_{i_1}^*, Z_{i_2}^*;\hat{\theta})$.

Let $E^*(\cdot) = E(\cdot|F_n)$ and define $h_1^*(z_1^*;\theta) = E^*\{h(z_1^*, Z_2^*;\theta)\} = n^{-1}\sum_{i=1}^{n} h(z_1^*, Z_i;\theta)$. Given F_n, $h_1^*(Z_1^*;\hat{\theta}), \ldots, h_1^*(Z_n^*;\hat{\theta})$ are independent and identically distributed with mean 0 and conditional covariance matrix $E^*\{h_1^{*\otimes 2}(Z_1^*;\hat{\theta})\} = \tilde{\zeta}_1(\hat{\theta})$ where we define

$$\tilde{\zeta}_1(\theta) = \frac{1}{n^3} \sum_{i_1=1}^{n} \left\{ \sum_{i_2=1}^{n} h(Z_{i_1}, Z_{i_2};\theta) \right\}^{\otimes 2}.$$

Simple calculation indicates that $\tilde{\zeta}_1(\hat{\theta}) = n^{-2}(n-1)^3 V_S(\hat{\theta})/4$. By the definition of $\tilde{\zeta}_2(\theta)$ in (A1), we have the repeatedly used fact that $\tilde{\zeta}_1(\hat{\theta})_{(l,l)} \leq \tilde{\zeta}_2(\hat{\theta})_{(l,l)}$ for any given F_n and $l = 1, \ldots, p$.

To show (12), consider first the case $p = 1$ for which

$$nV_S^*(\hat{\theta}) = \frac{4}{n-1} \sum_{i=1}^{n} b_i^2(\hat{\theta}) - \frac{4n}{n-1} S_n^{*^2}(\hat{\theta}) \qquad \text{(A2)}$$

where $b_i(\hat{\theta}) = (n-1)^{-1} \sum_{j:j\neq i} h(Z_i^*, Z_j^*; \hat{\theta})$.

Given F_n, $S_n^*(\hat{\theta})$ is a U-statistic and from (3.17) of Athreya et al. (1984) we obtain,

$$E\{S_n^*(\hat{\theta})\}^2 = E[E^*\{S_n^*(\hat{\theta})\}^2] = \binom{n}{2}^{-1} [2(n-2)E\{\tilde{\zeta}_1(\hat{\theta})\} + E\{\tilde{\zeta}_2(\hat{\theta})\}] \leq \frac{4}{n} E\{\tilde{\zeta}_2(\hat{\theta})\}. \quad \text{(A3)}$$

By Assumptions 3, 4 and the dominated convergence theorem (Loève, 1977), $E\{S_n^*(\hat{\theta})\}^2 \to 0$ so that $S_n^*(\hat{\theta}) \to 0$ in probability as $n \to \infty$. In view of (A2), it remains to show that

$$\frac{1}{n} \sum_{i=1}^{n} \{b_i(\hat{\theta})\}^2 \to \zeta_1 \text{ in probablity.} \quad \text{(A4)}$$

Similar to (3.21) of Athreya et al. (1984), we have

$$n^{-1} \sum_{i=1}^{n} E\{b_i(\hat{\theta}) - h_1^*(Z_i^*; \hat{\theta})\}^2 = E\{b_i(\hat{\theta}) - h_1^*(Z_i^*; \hat{\theta})\}^2 \leq (n-1)^{-1} E\tilde{\zeta}_2(\hat{\theta}).$$

Thus,

$$\frac{1}{n} \sum_{i=1}^{n} \{b_i(\hat{\theta}) - h_1^*(Z_i^*; \hat{\theta})\}^2 \to 0 \quad \text{(A5)}$$

in probability. According to Lemma 1 of Sen (1960) and (A5), the result (A4) follows if $n^{-1} \sum_{i=1}^{n} h_1^{*^2}(Z_i^*; \hat{\theta}) \to \zeta_1$ in probability. Following arguments for (3.24) and (3.25) in Athreya et al. (1984), it follows that $n^{-1} \sum_{i=1}^{n} h_1^{*^2}(Z_i^*; \theta_0) \to \zeta_1$ in probability, and it suffices to show that

$$\frac{1}{n} \sum_{i=1}^{n} \{h_1^*(Z_i^*; \hat{\theta}) - h_1^*(Z_i^*; \theta_0)\}^2 \to 0 \text{ in probability.} \quad \text{(A6)}$$

Given F_n, the left side of (A6) is an average of conditionally independent and identically distributed terms and has characteristic function $\phi_n(t) = E\{\phi_n^*(t/n)\}^n$ where $\phi_n^*(t/n) = E^*[\exp\{it(h_1^*(Z_i^*; \hat{\theta}) - h_1^*(Z_i^*; \theta_0))^2/n\}]$. To show (A6), we show that $\lim_{n\to\infty} \phi_n(t) \to 1$ and, by the dominated convergence theorem, this follows if $\{\phi_n^*(t/n)\}^n \to 1$ in probability. Let $r_n(t/n) = \int_{-\infty}^{\infty} [\exp\{it(h_1^*(u; \hat{\theta}) - h_1^*(u; \theta_0))^2/n\} - 1] dF_n(u)$ so that $\phi_n^*(t/n) = 1 + r_n(t/n)$. It now remains to show $n|r_n(t/n)| \to 0$ in probability. From the inequality for characteristic functions,

$$\begin{aligned} n|r_n(t/n)| &\leq |t| \int_{-\infty}^{\infty} \{h_1^*(u;\hat{\theta}) - h_1^*(u;\theta_0)\}^2 dF_n(u) \\ &= \frac{|t|}{n} \sum_{i=1}^{n} \left\{ \frac{1}{n} \sum_{j=1}^{n} h(Z_i, Z_j; \hat{\theta}) - \frac{1}{n} \sum_{j=1}^{n} h(Z_i, Z_j; \theta_0) \right\}^2 \qquad \text{(A7)} \\ &= \frac{|t|}{n} \{\tilde{\zeta}_2(\hat{\theta}) + \tilde{\zeta}_2(\theta_0) - 2C_n(\hat{\theta})\} + |t|\{A_n(\hat{\theta}) + A_n(\theta_0) - 2B_n(\hat{\theta})\}. \end{aligned}$$

By Assumption 3, $n|r_n(t/n)| \to 0$ in probability.

This establishes (12) in the single parameter case ($p = 1$). To show (12) in the multiple parameter case, we replicate the above steps for the scalar valued U-statistic $a^{\mathrm{T}} S_n(\theta)$ and use Lemma A.1.

To show (13), we first show that

$$V_S(\hat{\theta})^{-1/2} S_n^*(\hat{\theta}) \to N(\mathit{0}, \mathbf{1}) \text{ in distribution.} \qquad \text{(A8)}$$

Let $\mathcal{X}_{n,i} = \tilde{\zeta}_1(\hat{\theta})^{-1/2} h_1^*(Z_i^*; \hat{\theta})$. For $n = 1, 2, \ldots$ and $i = 1, \ldots, n$, $\{\mathcal{X}_{n,i}\}$ is a triangular array whose elements in the nth row are independent and identically distributed conditional on F_n. The corresponding Lindeberg condition is that

$$\sum_{i=1}^{n} E^* ||\mathcal{X}_{n,i}||^2 I\{||\mathcal{X}_{n,i}|| > \epsilon\} \to 0$$

for all $\epsilon > 0$. This has the same form as Assumption 5 so that $\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathcal{X}_{n,i} \to N(\mathit{0}, \mathbf{1})$ in distribution.

Let $\mathcal{Y}_n = (n-1)^{3/2} n^{-1} S_n^*(\hat{\theta})$ and $\mathcal{Z}_n = 2n^{-1/2} \sum_{i=1}^{n} h_1^*(Z_i^*; \hat{\theta})$. Now,

$$V_S(\hat{\theta})^{-\frac{1}{2}} S_n^*(\hat{\theta}) - \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \mathcal{X}_{n,i} = \{4\tilde{\zeta}_1(\hat{\theta})\}^{-\frac{1}{2}} (\mathcal{Y}_n - \mathcal{Z}_n).$$

To show (A8), we show that $\tilde{\zeta}_1(\hat{\theta}) = (n-1)^3 n^{-2} V_S(\hat{\theta})/4 \to \zeta_1$ and $\mathcal{Y}_n - \mathcal{Z}_n \to 0$ in probability. By Theorem 2.2, $na^{\mathrm{T}} V_S(\theta_0) a/4 \to a^{\mathrm{T}} \zeta_1 a$ in probability. Since $S_n(\theta_0) \to 0$ in probability and

$$na^{\mathrm{T}} V_S(\theta) a = \frac{4}{n-1} \sum_{i=1}^{n} \{ \frac{1}{n-1} \sum_{j:j\neq i} a^{\mathrm{T}} h(Z_i, Z_j; \theta) \}^2 - \frac{4n}{n-1} \{ a^{\mathrm{T}} S_n(\theta) \}^2,$$

Lemma 1 of Sen (1960) and Lemma A.1 imply that $\tilde{\zeta}_1(\hat{\theta}) \to \zeta_1$ if

$$n^{-1} \sum_{i=1}^{n} \left\{ \frac{1}{n} \sum_{j=1}^{n} a^{\mathrm{T}} h(Z_i, Z_j; \hat{\theta}) - \frac{1}{n} \sum_{j=1}^{n} a^{\mathrm{T}} h(Z_i, Z_j; \theta_0) \right\}^2 \to 0$$

in probability. This follows from a similar argument used to verify (A7).

Finally, $\mathcal{Y}_n - \mathcal{Z}_n \to \mathit{0}$ in probability if $E(\mathcal{Y}_n^{(l)} - \mathcal{Z}_n^{(l)})^2 \to 0$ for each $l = 1, \ldots, p$, where $a^{(l)}$ stands for the lth component of a vector a. Given F_n, $S_n^*(\hat{\theta})$ is a U-statistic with mean $\mathit{0}$ so that

$$E^*(\mathcal{Y}_n^{(l)})^2 = \frac{4(n-1)^2(n-2)}{n^3} \tilde{\zeta}_1(\hat{\theta})_{(l,l)} + \frac{2(n-1)^2}{n^3} \tilde{\zeta}_2(\hat{\theta})_{(l,l)}.$$

Since $\mathcal{Z}_n$ is a sum of conditional independent vectors, $E^*(\mathcal{Z}_n^{(l)})^2 = 4\tilde{\zeta}_1(\hat{\theta})_{(l,l)}$. Finally,

$$\begin{aligned} E^*(\mathcal{Y}_n^{(l)} \mathcal{Z}_n^{(l)}) &= \frac{4}{n(n-1)} \left(\frac{n-1}{n} \right)^{\frac{3}{2}} \sum_{i=1}^{n} \sum_{1 \le i_1 < i_2 \le n} E^* \{ h_1^*(Z_i^*; \hat{\theta})^{(l)} h(Z_{i_1}^*, Z_{i_2}^*; \hat{\theta})^{(l)} \} \\ &= 4 \left(\frac{n-1}{n} \right)^{\frac{3}{2}} \tilde{\zeta}_1(\hat{\theta})_{(l,l)} \end{aligned}$$

where the sum contains $n(n-1)$ nonzero terms $E^* \{ h_1^*(Z_i^*; \hat{\theta})^{(l)} h(Z_{i_1}^*, Z_{i_2}^*; \hat{\theta})^{(l)} \} = \tilde{\zeta}_1(\hat{\theta})_{(l,l)}$ when $i_1 = i$ or $i_2 = i$. It follows that

$$E(\mathcal{Y}_n^{(l)} - \mathcal{Z}_n^{(l)})^2 = E\{ E^*(\mathcal{Y}_n^{(l)} - \mathcal{Z}_n^{(l)})^2 \} \le \frac{10}{n} E\{ \tilde{\zeta}_2(\hat{\theta})_{(l,l)} \},$$

and by an argument similar to that leading to (A3), $E(\mathcal{Y}_n^{(l)} - \mathcal{Z}_n^{(l)})^2 \to 0$ which establishes (A8).

The result (13) follows from Slutsky's theorem.

Reference

ARVESEN, J. N. (1969). Jackknifing U-statistics. *Annals of Mathematical Statistics* **40**, 2076-2100.

ATHREYA, K. B., GHOSH, M., LOW, L. Y. and SEN, P. K. (1984). Laws of large numbers for bootstrapped U-statistics. *Journal of Statistical Planning and Inference* **9**, 185-194.

BARRODALE, I. and ROBERTS, F. D. K. (1973). An improved algorithm for discrete L_1 linear approximations. *SIAM Journal of Numerical Analysis* **10**, 839-848.

BARRODALE, I. and ROBERTS, F. D. K. (1974). Solution of an overdetermined system of equations in the L_1 norm. *Communications of the ACM* **17**, 319-320.

BICKEL, P. J. and FREEDMAN, D. A. (1981). Some asymptotic theory for the bootstrap. *The Annals of Statistics* **9**, 1196-1217.

BLOOMFIELD, P. and STEIGER, W. L. (1983). *Least absolute deviations: theory, applications and algorithms.* Birkhauser, Boston, Mass.

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

CALLAERT, H. and VERAVERBEKE, N. (1981). The order of the normal approximation for a studentized U-statistic. *The Annals of Statistics* **9**, 194-200.

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

HE, X. and HU, F. (2002). Markov chain marginal bootstrap. *Journal of the American Statistical Association* **97**, 783-795.

HELMERS, R. (1991). On the Edgeworth expansion and the bootstrap approximation for a studentized U-statistic. *The Annals of Statistics* **19**, 470-484.

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

HOEFFDING, W. (1948). A class of statistics with asymptotically normal distribution. *Annals of Mathematical Statistics* **19**, 293-325.

HU, F. and KALBFLEISCH, J. D. (2000). The estimating function bootstrap (with Discussion). *Canadian Journal of Statistics* **28**, 449-499.

JIN, Z., YING, Z., and WEI, L. J. (2001). A simple resampling method by perturbing the minimand. *Biometrika* **88**, 381-390.

LOÈVE, M. (1977). *Probability Theory I.* New York: Springer-Verlag.

PARZEN, M. I., WEI, L. J., and YING, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81**, 341-350.

SEN, P. K. (1960). On some convergence properties of *U*-statistics. *Calcutta Statistical Association Bulletin* **10**, 1-18.

SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* New York: John Wiley.

Table 1. Upper Confidence Intervals for Example 1 with Homoscedastic Errors. N=10,000, B=1,000, n=20.

Nominal	2.5%	5.0%	10.0%	90.0%	95.0%	97.5%
JYW:CP(%)	9.36	12.89	18.24	81.03	86.38	90.30
Avg.CI	.82	.85	.89	1.11	1.14	1.17
SE.CI	.14	.13	.13	.12	.13	.14
Norm:CP(%)	3.60	6.26	11.34	89.04	94.09	96.53
Avg.CI	.83	.86	.89	1.11	1.14	1.17
SE.CI	.09	.09	.09	.09	.09	.09
EF-t:CP(%)	2.91	5.23	10.30	89.78	94.85	97.34
Avg.CI	.82	.85	.88	1.12	1.15	1.18
SE.CI	.10	.09	.09	.09	.09	.10

Table 2. Upper Confidence Intervals for Example 1 with Heteroscedastic Errors. N=10,000, B=1,000, n=20.

Nominal	2.5%	5.0%	10.0%	90.0%	95.0%	97.5%
JYW:CP(%)	9.23	12.75	17.80	81.73	87.12	90.68
Avg.CI	.56	.63	.72	1.28	1.36	1.44
SE.CI	.34	.33	.31	.31	.33	.34
Norm:CP(%)	3.55	6.27	11.29	89.07	94.11	96.56
Avg.CI	.58	.65	.72	1.28	1.35	1.42
SE.CI	.24	.23	.23	.23	.23	.24
EF-t:CP(%)	2.94	5.25	10.42	89.54	94.74	97.12
Avg.CI	.55	.62	.71	1.29	1.38	1.45
SE.CI	.25	.24	.23	.23	.24	.25

Table 3.Upper Confidence Intervals for Example 2. p=1, N=10,000, B=1,000, n=20.

Nominal	2.5%	5.0%	10.0%	90.0%	95.0%	97.5%
JYW:CP(%)	5.42	8.88	14.45	85.50	91.15	94.42
Avg.CI	.57	.65	.73	1.27	1.35	1.43
SE.CI	.28	.27	.26	.26	.27	.28
Norm-S:CP(%)	4.00	6.19	10.74	89.23	93.91	96.15
Avg.CI	.58	.64	.71	1.29	1.36	1.42
SE.CI	.24	.24	.24	.24	.24	.24
Norm-J:CP(%)	3.34	5.43	9.76	90.28	94.66	96.73
Avg.CI	.56	.62	.70	1.31	1.38	1.44
SE.CI	.24	.24	.24	.24	.24	.24
EFB:CP(%)	2.75	5.03	10.24	89.99	94.64	97.25
Avg.CI	.53	.61	.70	1.30	1.40	1.48
SE.CI	.24	.24	.24	.24	.24	.25

Table 4.Upper Confidence Intervals for Example 2. p=2, N=5,000, B=1,000, n=20.

Nominal	2.5%	5.0%	10.0%	90.0%	95.0%	97.5%
JYW:β_1-CP(%)	10.92	15.84	23.10	83.78	89.28	92.62
Avg.CI	.63	.70	.77	1.27	1.35	1.43
SE.CI	.31	.30	.29	.27	.28	.29
β_2-CP(%)	7.40	10.74	16.46	76.80	83.84	89.22
Avg.CI	-.10	.00	.12	.83	.93	1.02
SE.CI	.40	.39	.38	.41	.42	.43
Norm-S:β_1-CP(%)	4.44	6.98	11.24	89.38	93.74	96.32
Avg.CI	.58	.65	.73	1.28	1.36	1.43
SE.CI	.25	.24	.23	.23	.24	.25
β_2-CP(%)	3.30	5.78	10.76	88.66	93.78	95.88
Avg.CI	-.10	-.01	.10	.90	1.01	1.10
SE.CI	.35	.34	.33	.33	.34	.35
Norm-J:β_1-CP(%)	3.56	6.22	10.44	89.30	94.04	96.86
Avg.CI	.55	.62	.71	1.29	1.38	1.45
SE.CI	.25	.25	.24	.24	.25	.26
β_2-CP(%)	2.98	5.40	10.34	90.04	94.34	96.48
Avg.CI	-.13	-.03	.08	.93	1.04	1.14
SE.CI	.35	.35	.34	.33	.34	.35
EFB:β_1-CP(%)	2.68	5.08	9.70	90.38	95.06	97.54
Avg.CI	.51	.60	.69	1.31	1.41	1.50
SE.CI	.26	.25	.24	.24	.25	.26
β_2-CP(%)	2.58	4.28	8.46	91.44	96.02	97.88
Avg.CI	-.20	-.08	.05	.95	1.09	1.20
SE.CI	.35	.34	.33	.34	.35	.36

Table 5. Upper Confidence Intervals for Example 3. p=1, N=10,000, B=1,000, n=20.

Nominal	2.5%	5.0%	10.0%	90.0%	95.0%	97.5%
JYW:CP(%)	5.45	8.91	14.45	85.39	91.03	94.59
Avg.CI	.60	.67	.74	1.25	1.33	1.40
SE.CI	.25	.24	.24	.24	.24	.25
Norm-S:CP(%)	3.89	6.52	11.52	88.83	93.59	96.10
Avg.CI	.58	.64	.72	1.28	1.36	1.42
SE.CI	.24	.24	.23	.23	.24	.24
Norm-J:CP(%)	3.14	5.32	9.90	89.88	94.48	96.85
Avg.CI	.55	.62	.70	1.30	1.38	1.44
SE.CI	.24	.24	.23	.23	.24	.24
EFB:CP(%)	2.43	5.04	9.82	89.93	94.89	97.29
Avg.CI	.52	.60	.69	1.30	1.39	1.48
SE.CI	.26	.24	.25	.24	.24	.25