Biometrika (2001), **88**, 2, *pp.* 381–390

Printed in Great Britain

A simple resampling method by perturbing the minimand

BY ZHEZHEN JIN

Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, Massachusetts 02115, U.S.A.

zjin@hsph.harvard.edu

ZHILIANG YING

Department of Statistics, Columbia University, 2990 Broadway, New York, New York 10027, U.S.A.

zying@stat.columbia.edu

AND L. J. WEI

Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, Massachusetts 02115, U.S.A.

wei@hsph.harvard.edu

SUMMARY

Suppose that under a semiparametric setting an estimator of a vector of parameters of interest is obtained by optimising an objective function which has a U-process structure. The covariance matrix of the estimator is generally a function of the underlying density function, which may be difficult to estimate well by conventional methods. In this paper, we present a simple resampling method by perturbing the objective function repeatedly. Inferences of the parameters can then be made based on a large collection of the resulting optimisers. We illustrate our proposal by three examples with a heteroscedastic regression model.

Some key words: Bootstrap; Heteroscedastic regression; L_p norm; Resampling method; Truncated regression; U-process.

1. INTRODUCTION

Let θ be an $r \times 1$ vector of parameters of interest for the distribution of a random vector Z. With n independent copies of Z, a consistent estimator $\hat{\theta}$ for θ_0, the true value of θ, can often be obtained by minimising an objective function $U_n(\theta)$. If the minimand is smooth enough in θ, generally a large-sample approximation to the distribution of $(\hat{\theta} - \theta_0)$ can be obtained easily. If $U_n(\theta)$ is not smooth, however, it may be difficult to estimate the covariance of $\hat{\theta}$ well by conventional methods, especially under the semiparametric setting with unknown underlying density functions. Resampling methods, such as bootstrap, jackknife and so on, are particularly useful in dealing with this situation (Efron & Tibshirani, 1993; Shao & Tu, 1995; Davison & Hinkley, 1997). Generally a resampling method generates 'artificial' samples from the observed Z's and, for each generated sample, we obtain an estimate of θ_0 by minimising the corresponding minimand. Inferences about θ_0 can

then be made based on a large collection of these estimates. Theoretical justification of the existing resampling methods is generally nontrivial and has to be made on a case by case basis.

In this paper, we consider a large class of objective functions and propose a simple resampling method by perturbing the minimand directly. As with the bootstrap method, the new procedure does not involve any complicated and subjective nonparametric functional estimator. Our proposal is illustrated with three examples and, for each example, we show how the new method can be implemented easily with existing statistical software. Recently, Parzen et al. (1994) and Hu & Kalbfleisch (2000) proposed resampling methods by perturbing the 'score function' of the minimand directly. Often, however, the estimating function may not be continuous and solving the corresponding equations numerically can be rather challenging, especially when the dimension of the parameter vector is large.

2. A GENERAL RESAMPLING METHOD

Suppose that Z_i $(i = 1, \ldots, n)$ are independent and identically distributed random vectors and that θ belongs to a compact space $\Theta \subset \Re^r$. Let $\hat{\theta}$ be a minimiser of a U-process of degree K,

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

where $h(.)$ is symmetric in the Z's, and $\sum$ denotes the summation over subsets of K integers $(i_1, \ldots, i_K)$ from $\{1, \ldots, n\}$. For fixed θ, $U_n(\theta)$ is simply a standard U-statistic with kernel $h(.)$. Most of the minimands for the estimation procedures in the literature are U-processes. In Propositions 1 and 2 of the Appendix we show that, if, for large n, $U_n(\theta)$ has a 'good' quadratic approximation around θ_0, $\hat{\theta}$ is strongly consistent and asymptotically normal. When $h(.; \theta)$ is not twice differentiable with respect to θ, it is difficult to estimate the covariance matrix of $\hat{\theta}$, which generally involves the unknown underlying density functions.

Let $\{z_i\}$ be the observed value of $\{Z_i\}$ and let $\{V_i\}$ be n independent, identical copies of a nonnegative, completely known random variable V with mean μ and variance $K^2\mu^2$. Consider a stochastic perturbation $\tilde{U}_n(\theta)$ of the observed $U_n(\theta)$, where

$$\tilde{U}_n(\theta) = \binom{n}{K}^{-1} \sum_{1 \leqslant i_1 < i_2 < \ldots < i_K \leqslant n} (V_{i_1} + \ldots + V_{i_K}) h(z_{i_1}, \ldots, z_{i_K}; \theta).$$

Let θ^* be the minimiser of $\tilde{U}_n(\theta)$. Note that the only random quantities in $\tilde{U}_n(\theta)$ are $\{V_i\}$. In Proposition 3 of the Appendix, we show that, if $\tilde{U}_n(\theta)$ also has a 'good' quadratic expansion around θ_0, the distribution of $n^{\frac{1}{2}}(\theta^* - \tilde{\theta})$ can be used to approximate the distribution of $n^{\frac{1}{2}}(\hat{\theta} - \theta_0)$, where $\tilde{\theta}$ is the observed value of $\hat{\theta}$.

In practice, the distribution of θ^* can be estimated by generating a large number, M, say, of random samples $\{V_i, i = 1, \ldots, n\}$. For the jth realised sample from $\{V_i\}$, we obtain θ_j^* by minimising $\tilde{U}_n(\theta)$ $(j = 1, \ldots, M)$. The theoretical distribution of θ^* or $\hat{\theta}$ can then be approximated by the usual empirical distribution function based on $\{\theta_j^*, j = 1, \ldots, M\}$. The covariance matrix of $\hat{\theta}$ can then be estimated by the sample covariance matrix constructed from $\{\theta_j^*, j = 1, \ldots, M\}$.

To make our resampling method practically useful, one needs reliable and efficient optimisation algorithms for minimising $U_n(\theta)$ and $\tilde{U}_n(\theta)$. In § 3, we consider three examples to illustrate our proposal. For each case, easily accessible software exists for solving the above optimisation problem.

3. Examples

3·1. *Model and dataset*

Consider a heteroscedastic linear regression model

$$Y_i = \alpha + X_i^{\mathrm{T}}\beta + \varepsilon_i, \tag{3·1}$$

where $\theta^{\mathrm{T}} = (\alpha, \beta^{\mathrm{T}})$, $E\varepsilon_i = 0$, $Z_i = (X_i, Y_i)$, X_i is bounded, and the components of X_i are linearly independent in a nonnull set, $i = 1, \ldots, n$. For $X = x$, let $f(.|x)$ be the density function of ε, which may depend on x. Under this model, we consider three cases to illustrate our resampling method using a well-known dataset from 41 prepubescent boys (Hettmansperger & McKean, 1998, p. 204). For each individual subject in the dataset, the response is the level of free fatty acid while the independent variables are age, weight and skin-fold thickness.

3·2. *Estimation based on L_p norm*

The estimator $\hat{\theta}$ using the L_p norm for model (3·1) is a minimiser of the U-process

$$U_n(\theta) = n^{-1} \sum_{i=1}^{n} |Y_i - \alpha - X_i^{\mathrm{T}}\beta|^p.$$

Here, the degree K of the U-process is 1. The corresponding $\tilde{U}_n(\theta)$ is

$$\tilde{U}_n(\theta) = n^{-1} \sum_{i=1}^{n} V_i |y_i - \alpha - x_i^{\mathrm{T}}\beta|^p,$$

where (y, x) is the observed value of (Y, X). When $p \geqslant 1$, $U_n(\theta)$ is a convex function of θ. The program RLLP in the IMSL statistical package (1991) can be used to estimate the regression coefficients of (3·1) with the general L_p norm. For the case with $p = 1$, minimisation of $U_n(\theta)$ can be efficiently handled by linear programming techniques, and an efficient algorithm developed by Koenker & D'Orey (1987) is available in S-Plus to obtain $\hat{\theta}$. Furthermore, when $p \geqslant 2$, $U_n(\theta)$ is twice differentiable and $\hat{\theta}$ can be obtained trivially. Note that when $p < 2$ all the existing estimators for the covariance matrix of $\hat{\theta}$ are derived under the assumption that the distribution of the error term ε is free of the covariate X. Otherwise, further parametric structures on the error term or high-dimensional nonparametric function estimators are needed for the variance estimation.

For the resampling method proposed in § 2, when $p \neq 2$, assume that, with respect to t, $f(t|x)$ satisfies the usual Lipschitz condition and is symmetric about 0. Also, assume that $f(0|x) > 0$ and $E|\varepsilon|^{(p-2)} < \infty$. Under these mild assumptions, all the conditions in Propositions A1–A3 of the Appendix are satisfied. The details of the proof are given in an unpublished report by Z. Jin, Z. Ying and L. J. Wei. It follows that, for the general L_p norm, the distribution of $\hat{\theta}$ can be approximated by that of θ^*. Furthermore, the aforementioned software can also be used to obtain θ^* numerically. It is important to note that our procedure is valid without imposing any parametric structure on $f(.|x)$.

For the free fatty acid example, estimates for the regression coefficients of model (3·1) were obtained with various L_p norms. For each case, 1000 θ^*'s were generated to estimate the covariance matrix of $\hat{\theta}$ with several different types of random variable V in $\tilde{U}$. In Table 1(a), we report the results with $p = 1$, 1·5, 2 and 2·5 and with V being the gamma distribution Ga(1, 1) and beta distribution Be($\sqrt{2} - 1$, 1). For comparison, we also analysed the data with the subroutine RLLP from IMSL (1991) and with the heteroscedastic bootstrap method (Efron, 1982, p. 36). The standard error estimates under the heading

Table 1. *Estimates of regression coefficients for the free fatty acid example, with estimated standard errors in parentheses*

(*a*) L_p *norm*

		Gamma	Beta	Bootstrap	IMSL		Gamma	Beta	Bootstrap	IMSL
			$p=1$					$p=1{\cdot}5$		
Intercept	1·520	(0·402)	(0·427)	(0·384)	(0·534)	1·576	(0·286)	(0·313)	(0·312)	(0·310)
Age	−0·003	(0·005)	(0·005)	(0·005)	(0·005)	−0·002	(0·003)	(0·003)	(0·003)	(0·003)
Weight	−0·011	(0·007)	(0·008)	(0·008)	(0·008)	−0·014	(0·005)	(0·005)	(0·005)	(0·005)
SFT	0·177	(0·304)	(0·305)	(0·302)	(0·272)	0·195	(0·176)	(0·192)	(0·203)	(0·158)
			$p=2$					$p=2{\cdot}5$		
Intercept	1·702	(0·276)	(0·320)	(0·321)	(0·327)	1·818	(0·299)	(0·337)	(0·345)	(0·300)
Age	−0·002	(0·003)	(0·003)	(0·003)	(0·003)	−0·003	(0·002)	(0·003)	(0·003)	(0·003)
Weight	−0·015	(0·004)	(0·004)	(0·004)	(0·005)	−0·016	(0·003)	(0·004)	(0·004)	(0·004)
SFT	0·205	(0·137)	(0·160)	(0·155)	(0·167)	0·223	(0·120)	(0·151)	(0·155)	(0·153)

(*b*) *Wilcoxon rank estimation*

		Gamma	Beta	Minitab	Bootstrap
Intercept	—	—	—	—	—
Age	−0·001	(0·002)	(0·002)	(0·003)	(0·003)
Weight	−0·015	(0·003)	(0·003)	(0·004)	(0·004)
SFT	0·275	(0·121)	(0·130)	(0·137)	(0·185)

(*c*) *Truncated regression*

		Gamma	Beta	Bootstrap
Intercept	1·507	(0·448)	(0·488)	(0·432)
Age	−0·003	(0·004)	(0·005)	(0·004)
Weight	−0·011	(0·006)	(0·007)	(0·006)
SFT	0·163	(0·273)	(0·312)	(0·255)

IMSL, method based on program RLLP in the IMSL package; SFT, skin fold thickness.

'Bootstrap' in Table 1(a) are based on 1000 bootstrap samples. The results from both methods are fairly similar to those obtained from our resampling procedure. The variance estimates of the regression coefficient estimates used in IMSL were obtained under a strong assumption, namely that the errors are independent of $\{X_i\}$.

3·3. *Estimation based on Wilcoxon statistic*

The rank estimator based on the Wilcoxon test statistic for linear regression model (3·1) is a minimiser of

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

a U-process with degree 2, which is a convex function of β. Note that the intercept term in (3·1) is not estimable with this type of rank estimation. The corresponding $\tilde{U}_n(\beta)$ is

$$\binom{n}{2}^{-1} \sum_{1 \leqslant i_1 < i_2 \leqslant n} (V_{i_1} + V_{i_2}) |y_{i_1} - y_{i_2} - (x_{i_1} - x_{i_2})^{\mathrm{T}} \beta|.$$

Minimisation of $U_n(\beta)$ or $\tilde{U}_n(\beta)$ is a linear programming problem. If $f(.|x)$ and its derivative $f'(.|x)$ are bounded, one can show that all the conditions in Propositions A1–A3 of the Appendix are satisfied. It is important to note that, for our resampling method, there is no need to assume that the distribution of ε is free of X. On the other hand, this assumption is crucial for existing methods, which involve nonparametric density function estimation, for estimating the variance of $\hat{\beta}$.

For the free fatty acid example, we generated 1000 β*'s with various random variables V. In Table 1(b), we report the results with V being Ga(0·25, 0·5) and Be(0·125, 1·125). For comparison, we also report the standard error estimates of $\hat{\beta}$ using nonparametric density function estimates with the software Minitab (Hettmansperger & McKean, 1998, p. 181), and those using the heteroscedastic bootstrap method.

3·4. *Truncated median regression*

Suppose that ε in (3·1) has a unique median at 0 for any given x. Let $c < d$ be two known constants. It is not unusual that, because of the limitation of instruments, the response variable Y may not be accurately measured when it is above d or below c. This results in a truncated variable T from Y such that

$$T_i = \begin{cases} c, & \text{if } Y_i \leqslant c, \\ Y_i, & \text{if } c < Y_i < d, \\ d, & \text{if } Y_i \geqslant d. \end{cases}$$

Recently, Fitzenberger (1997) proposed an estimation procedure defined by minimising

$$U_n(\theta) = n^{-1} \sum_{i=1}^{n} |T_i - \min\{d, \max(c, \alpha + X_i^{\mathrm{T}} \beta)\}|,$$

and showed that the resulting estimator $\hat{\theta}$ is consistent and asymptotically normal under some regularity conditions. However, it is quite difficult to estimate well the covariance matrix of $\hat{\theta}$. For our resampling method, $\tilde{U}_n(\theta)$ is

$$\tilde{U}_n(\theta) = n^{-1} \sum_{i=1}^{n} V_i |t_i - \min\{d, \max(c, \alpha + x_i^{\mathrm{T}} \beta)\}|,$$

where t is the observed value of T. Minimisation of $U_n(\theta)$ or $\tilde{U}_n(\theta)$ can be implemented by the algorithm developed by Koenker & Park (1996), or by an adaptation of the Barrodale–Roberts algorithm for the one-sided censored quantile regression (Fitzenberger, 1997). In the econometrics literature, the one-sided truncated regression model, with either the $c = -\infty$ or $d = \infty$, has been extensively studied, for example by Powell (1984, 1986) and Pollard (1990).

If

$$\text{pr}\{X^{\mathrm{T}}\theta_0 \in (c, d)\} > 0, \quad \text{pr}(X^{\mathrm{T}}\theta_0 = c) = 0, \quad \text{pr}(X^{\mathrm{T}}\theta_0 = d) = 0,$$

one can show that the conditions in Propositions 1–3 of the Appendix are satisfied. For the free fatty acid example, we artificially set $c = 0{\cdot}325$ and $d = 1{\cdot}016$ to create a dataset with 20% truncation for the response variable. The results, using the algorithm by Koenker & Park (1996), are reported in Table 1(c) with V being Ga(1, 1) and Be($\sqrt{2}-1$, 1). For comparison, the standard error estimates based on 1000 bootstrap samples are also reported in Table 1(c).

4. REMARKS

In this paper, we present a rather simple resampling method by perturbing the minimand directly. In generating an approximation to the distribution of $\hat{\theta}$, the problem of choosing an appropriate number M of samples $\{V_i\}$ is similar to that of choosing the number of bootstrap samples. One may start with 500 θ^*'s, say, for constructing a confidence interval of the parameter of interest and then repeat the same kind of analysis with 500 additional θ^*'s. If the discrepancy between the two intervals is practically insignificant, no further resampling is needed.

Several extensive simulation studies were conducted to evaluate the adequacy of the new proposals. The results indicate that the new resampling method behaves well even for small sample sizes. For example, in one of the numerical studies, we mimicked the set-up of the free fatty acid example to examine if the new interval estimation procedure based on L_1 norm for the regression coefficient of model (3·1) has correct coverage probabilities. To this end, we generated 1000 samples $\{(y_i, z_i), i = 1, \ldots, 41\}$ from the model

$$Y = 1{\cdot}702 - 0{\cdot}002 \times \text{Age} - 0{\cdot}015 \times \text{Weight} + 0{\cdot}205 \times \text{SFT} + \varepsilon,$$

where SFT is the skin-fold thickness, the regression coefficients are the least squares estimates from the free fatty acid data and ε is a normal error with various variance structures. For each of these 1000 samples, the covariates were fixed and taken from the free fatty acid data, and 1000 θ^*'s were generated for estimating the covariance matrix of $\hat{\theta}$ with several distinct random variables V. The empirical coverage probabilities and estimated average lengths for the resulting intervals are summarised in Table 2 with V being Ga(1, 1) and Be($\sqrt{2}-1$, 1). For comparison, we also report the results based on the unconditional bootstrap method and the procedure in RLLP from IMSL. We find that, when the variance of the error term in the model is constant, the intervals obtained from the bootstrap method are similar to ours in terms of the empirical coverage probabilities and estimated average lengths. The procedure used in the commercial software IMSL is slightly better than these two resampling methods. This superiority, however, vanishes when the sample size is moderate or large, $n \geqslant 100$, say. If the variance of the error term is not constant, for example if the variance is proportional to the square of the standardised weight, the empirical coverage probabilities of the existing procedure obtained from the IMSL can be

extremely low. On the other hand, the new method performs well; see Table 2. Through our simulation studies, we also find that the choice of V for the resampling method is rather robust.

Table 2. *Empirical coverage probabilities and estimated mean lengths of confidence intervals for various procedures based on L_1 norm with $n = 41$*

(*a*) *Gaussian error with mean* 0 *and constant variance* 0·0464

	Confidence level	Gamma		Beta		Bootstrap		IMSL	
		ECP	EML	ECP	EML	ECP	EML	ECP	EML
Intercept	0·95	0·97	1·92	0·98	2·03	0·98	2·00	0·95	1·76
Age		0·98	0·02	0·99	0·02	0·98	0·02	0·94	0·02
Weight		0·98	0·03	0·99	0·03	0·98	0·03	0·95	0·03
SFT		0·98	1·01	0·99	1·10	0·99	1·08	0·94	0·90
Intercept	0·90	0·94	1·61	0·95	1·71	0·95	1·68	0·91	1·48
Age		0·95	0·02	0·96	0·02	0·96	0·02	0·90	0·02
Weight		0·96	0·02	0·97	0·02	0·97	0·02	0·91	0·02
SFT		0·95	0·84	0·96	0·92	0·96	0·90	0·91	0·75
Intercept	0·85	0·91	1·41	0·93	1·49	0·92	1·47	0·86	1·30
Age		0·89	0·01	0·93	0·02	0·92	0·01	0·86	0·01
Weight		0·91	0·02	0·94	0·02	0·93	0·02	0·87	0·02
SFT		0·92	0·74	0·94	0·80	0·94	0·79	0·87	0·66

(*b*) *Gaussian error with mean* 0 *and heteroscedastic variance*

	Confidence level	Gamma		Beta		Bootstrap		IMSL	
		ECP	EML	ECP	EML	ECP	EML	ECP	EML
Intercept	0·95	0·95	1·86	0·97	1·95	0·96	1·92	0·71	0·91
Age		0·99	0·01	0·99	0·01	0·99	0·01	0·94	0·01
Weight		0·97	0·02	0·98	0·03	0·97	0·03	0·76	0·01
SFT		0·99	0·75	1·00	0·81	1·00	0·79	0·90	0·46
Intercept	0·90	0·91	1·56	0·92	1·64	0·92	1·61	0·61	0·77
Age		0·96	0·01	0·97	0·01	0·97	0·01	0·89	0·01
Weight		0·93	0·02	0·94	0·02	0·94	0·02	0·68	0·01
SFT		0·97	0·63	0·98	0·68	0·98	0·66	0·83	0·39
Intercept	0·85	0·87	1·36	0·89	1·44	0·88	1·41	0·55	0·67
Age		0·94	0·01	0·95	0·01	0·95	0·01	0·85	0·01
Weight		0·89	0·02	0·91	0·02	0·91	0·02	0·62	0·01
SFT		0·93	0·55	0·95	0·60	0·95	0·58	0·79	0·34

ECP, empirical coverage probabilities; EML, estimated mean lengths.
IMSL, method based on program RLLP in the IMSL package; SFT, skin fold thickness.

The estimating function bootstrap method proposed by Hu & Kalbfleisch (2000) is valid only for the case with a linear estimating equation with independent terms. Using similar techniques presented in this paper, one may be able to generalise their method to the case in which the estimating function has a U-process structure.

Acknowledgement

We are very grateful to the editor and referees for helpful comments on the paper. This research was partially supported by grants from the US National Institutes of Health.

APPENDIX

Large-sample properties of $\hat{\theta}$ and θ^*

PROPOSITION A1. *Assume that $\{h(z_1, \ldots, z_K; \theta): \theta \in \Theta\}$ is Euclidean (Nolan & Pollard, 1987, Definition 8, p. 789), and there exists a function $H(Z_1, \ldots, Z_K)$ such that $EH(Z_1, \ldots, Z_K) < \infty$ and $|h(Z_1, \ldots, Z_K; \theta)| \leqslant H(Z_1, \ldots, Z_K)$ almost surely for all $\theta \in \Theta$. Furthermore, assume that $Eh(Z_1, \ldots, Z_K; \theta)$ is continuous and has a unique minimum at θ_0. Then $\hat{\theta} \to \theta_0$ almost surely as $n \to \infty$.*

Proof. Since $\{h(z_1, \ldots, z_K; \theta): \theta \in \Theta)$ is Euclidean, by Theorem 3.1 of Arcones & Giné (1993), $\sup_{\theta \in \Theta} |U_n(\theta) - EU_n(\theta)| \to 0$ almost surely. This, coupled with the fact that $Eh(Z_1, \ldots, Z_K; \theta)$ has a unique minimum at θ_0, implies that $\hat{\theta} \to \theta_0$ almost surely. □

Note that '$h(.)$ is Euclidean' is not a strong assumption (Nolan & Pollard, 1987, Lemma 22, p. 797). For the weak convergence of $\hat{\theta}$, assume that there exists an r-dimensional function $q(z_1, \ldots, z_K; \theta)$, the gradient of $h(z_1, \ldots, z_K; \theta)$ with respect to θ if it exists, such that

$$Eq(Z_1, Z_2, \ldots, Z_K; \theta) = \partial\{Eh(Z_1, \ldots, Z_K; \theta)\}/\partial\theta,$$

and $E\|q(Z_1, \ldots, Z_K; \theta)\|^2 < \infty$. Furthermore, assume that $Eq(Z_1, \ldots, Z_K; \theta)$ is continuously differentiable and

$$D = \frac{\partial}{\partial\theta} E\{q(Z_1, \ldots, Z_K; \theta)\}\bigg|_{\theta=\theta_0}$$

is nonsingular. Let

$$W_n(\theta) = n^{\frac{1}{2}} \binom{n}{K}^{-1} \sum_{1 \leqslant i_1 < \ldots < i_K \leqslant n} q(Z_{i_1}, \ldots, Z_{i_K}; \theta).$$

By a central limit theorem for U-statistics (Serfling, 1980, p. 192), $W_n(\theta_0)$ converges in distribution to normal with mean 0 and covariance matrix Γ.

PROPOSITION A2. *Assume that all the conditions listed in Proposition* A1 *are satisfied. In addition, assume that, almost surely,*

$$U_n(\theta_1) - U_n(\theta_2) = n^{-\frac{1}{2}} W_n^{\mathrm{T}}(\theta_2)(\theta_1 - \theta_2) + (\theta_1 - \theta_2)^{\mathrm{T}} D(\theta_1 - \theta_2)/2 + o(\|\theta_1 - \theta_2\|^2 + n^{-1}) \quad \text{(A·1)}$$

holds uniformly in $\|\theta_1 - \theta_0\| \leqslant d_n$ and $\|\theta_2 - \theta_0\| \leqslant d_n$, where $\{d_n\}$ is any sequence of positive random variables, converging to 0 almost surely. Then

$$n^{\frac{1}{2}}(\hat{\theta} - \theta_0) = -D^{-1} W_n(\theta_0) + o(1 + \|W_n(\theta_0)\| + n^{\frac{1}{2}}\|\hat{\theta} - \theta_0\|) \quad \text{(A·2)}$$

almost surely.

Proof. Since $\hat{\theta}$ is strongly consistent, by (A·1),

$$U_n(\hat{\theta}) - U_n(\theta_0) = n^{-\frac{1}{2}} W_n^{\mathrm{T}}(\theta_0)(\hat{\theta} - \theta_0) + (\hat{\theta} - \theta_0)^{\mathrm{T}} D(\hat{\theta} - \theta_0)/2 + o(\|\hat{\theta} - \theta_0\|^2 + n^{-1}). \quad \text{(A·3)}$$

Also, since $n^{-\frac{1}{2}} W_n(\theta_0)$ converges to 0, again by (A·1),

$$U_n\{\theta_0 - n^{-\frac{1}{2}} D^{-1} W_n(\theta_0)\} - U_n(\theta_0) = -\tfrac{1}{2} n^{-1} W_n(\theta_0)^{\mathrm{T}} D^{-1} W_n(\theta_0) + o\left(\frac{\|W_n(\theta_0)\|^2}{n} + n^{-1}\right). \quad \text{(A·4)}$$

Furthermore, since $\hat{\theta}$ is a minimiser of $U_n(\theta)$, $U_n(\hat{\theta}) \leqslant U_n\{\theta_0 - n^{-\frac{1}{2}} D^{-1} W_n(\theta_0)\}$. It follows from (A·3) and (A·4) that

$$\{\hat{\theta} - \theta_0 + n^{-\frac{1}{2}} D^{-1} W_n(\theta_0)\}^{\mathrm{T}} D\{\hat{\theta} - \theta_0 + n^{-\frac{1}{2}} D^{-1} W_n(\theta_0)\}/2 + o(\|\hat{\theta} - \theta_0\|^2 + n^{-1}\|W_n(\theta_0)\|^2 + n^{-1}) \leqslant 0,$$

almost surely. This implies (A·2). □

For the resampling part, we consider the unconditional perturbed objective function

$$\hat{U}_n(\theta) = \binom{n}{K}^{-1} \sum_{1 \leqslant i_1 < \ldots < i_K \leqslant n} (V_{i_1} + \ldots + V_{i_K}) h(Z_{i_1}, \ldots, Z_{i_K}; \theta).$$

Now let $\hat{\theta}^*$ be the minimiser of $\hat{U}_n(\theta)$. Without loss of generality, we assume that the mean and variance of V are $1/K$ and 1, respectively. Since $\hat{U}_n$ is also a U-process of degree K, exactly the same arguments as for Propositions A1 and A2 are applicable to obtain the following proposition, but with $W_n(\theta)$ replaced by

$$\hat{W}_n(\theta) = n^{\frac{1}{2}} \binom{n}{K}^{-1} \sum_{1 \leqslant i_1 < \ldots < i_K \leqslant n} (V_{i_1} + \ldots + V_{i_K}) q(Z_{i_1}, \ldots, Z_{i_K}; \theta).$$

PROPOSITION A3. *Assume that the conditions in Proposition* A1 *are satisfied. Then* $\hat{\theta}^*$ *is strongly consistent. Assume furthermore that, almost surely,*

$$\hat{U}_n(\theta_1) - \hat{U}_n(\theta_2) = n^{-\frac{1}{2}} \hat{W}_n^{\mathrm{T}}(\theta_2)(\theta_1 - \theta_2) + (\theta_1 - \theta_2)^{\mathrm{T}} D(\theta_1 - \theta_2)/2 + o(\|\theta_1 - \theta_2\|^2 + n^{-1}), \quad \text{(A·5)}$$

uniformly in $\|\theta_1 - \theta_0\| \leqslant d_n$, $\|\theta_2 - \theta_0\| \leqslant d_n$. *Then, in probability space generated by* $\{V, Z\}$,

$$n^{\frac{1}{2}}(\hat{\theta}^* - \hat{\theta}) = -D^{-1}\hat{W}_n(\hat{\theta}) + o(1 + \|\hat{W}_n(\hat{\theta})\| + n^{\frac{1}{2}}\|\hat{\theta}^* - \hat{\theta}\|) \quad \text{(A·6)}$$

almost surely.

Note that, since $Eh(Z_1, \ldots, Z_K; \theta) = E\{(V_1 + \ldots + V_K)h(Z_1, \ldots, Z_K; \theta)\}$, the matrix D in (A·5) is the same as that in (A·1). It follows from Proposition A2 that the asymptotic distribution of $n^{\frac{1}{2}}(\hat{\theta} - \theta_0)$ is the same as that of $D^{-1}W_n(\theta_0)$. Thus, in view of (A·6), to show that for every realisation of $\{Z\}$ the conditional distribution of $n^{\frac{1}{2}}(\hat{\theta}^* - \hat{\theta})$ converges to the same limiting distribution as that of $n^{\frac{1}{2}}(\hat{\theta} - \theta_0)$, it suffices to show that for every realisation of $\{Z\}$ the conditional distribution of $\hat{W}_n(\hat{\theta})$ converges to normal with mean 0 and covariance matrix Γ, the limiting distribution of $W_n(\theta_0)$.

Let $\theta_1 = \hat{\theta} - n^{-\frac{1}{2}}D^{-1}W_n(\hat{\theta})$ and $\theta_2 = \hat{\theta}$ in (A·1). Since $\hat{\theta}$ is a minimiser, it follows that

$$0 \leqslant -n^{-1}W_n^{\mathrm{T}}(\hat{\theta})D^{-1}W_n(\hat{\theta})/2 + o(n^{-1}\|W_n(\hat{\theta})\|^2 + n^{-1})$$

almost surely, which implies $\|W_n(\hat{\theta})\| = o(1)$ almost surely in the space of $\{Z\}$. Thus, up to an almost surely negligible term,

$$\begin{aligned}\hat{W}_n(\hat{\theta}) &= n^{\frac{1}{2}} \binom{n}{K}^{-1} \sum_{1 \leqslant i_1 < \ldots < i_K \leqslant n} \left\{\left(V_{i_1} - \frac{1}{K}\right) + \ldots + \left(V_{i_K} - \frac{1}{K}\right)\right\} q(Z_{i_1}, \ldots, Z_{i_K}; \hat{\theta}) \\ &= \frac{K}{n^{\frac{1}{2}}} \sum_{i=1}^{n} \left(V_i - \frac{1}{K}\right) \frac{1}{(n-1)\ldots(n-K)} \sum q(Z_i, Z_{i_2}, \ldots, Z_{i_K}; \hat{\theta}), \end{aligned} \quad \text{(A·7)}$$

where the last summation is over $i_2, \ldots, i_K$ such that no pair of indices among $i, i_2, \ldots, i_k$ are equal. By the strong law of large number for U-processes (Arcones & Giné, 1993, Theorem 3.1) and Proposition A1, the covariance matrix of (A·7) converges almost surely to Γ. Moreover, by the usual multivariate central limit theorem (Serfling, 1980, p. 30), for every realisation of $\{Z\}$ the conditional distribution of (A·7) converges to normal with mean 0 and covariance matrix Γ. Hence, $n^{\frac{1}{2}}(\hat{\theta} - \theta_0)$ has the same asymptotic distribution as that of $n^{\frac{1}{2}}(\theta^* - \tilde{\theta})$.

REFERENCES

ARCONES, M. A. & GINÉ, E. (1993). Limit theorems for U-processes. *Ann. Prob.* **21**, 1494–542.

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

EFRON, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans.* Philadelphia, PA: SIAM.

EFRON, B. & TIBSHIRANI, R. J. (1993). *An Introduction to the Bootstrap.* New York: Chapman and Hall.

FITZENBERGER, B. (1997). A guide to censored quantile regressions. In *Handbook of Statistics*, **15**, Ed. G. S. Maddala and C. R. Rao, pp. 405–37. Amsterdam: Elsevier Science.
HETTMANSPERGER, T. P. & MCKEAN, J. W. (1998). *Robust Nonparametric Statistical Methods.* New York: John Wiley.
HU, F. & KALBFLEISCH, J. D. (2000). The estimating function bootstrap (with Discussion). *Can. J. Statist.* **28**, 449–99.
IMSL (1991). *STAT/LIBRARY User's Manual, Version 2.0.* Houston, TX: IMSL.
KOENKER, R. & D'OREY, V. (1987). Computing regression quantiles. *Appl. Statist.* **36**, 383–93.
KOENKER, R. & PARK, B. J. (1996). An interior point algorithm for nonlinear quantile regression. *J. Economet.* **71**, 265–83.
NOLAN, D. & POLLARD, D. (1987). *U*-processes: rates of convergence. *Ann. Statist.* **15**, 780–99.
PARZEN, M. I., WEI, L. J. & YING, Z. (1994). A resampling method based on pivotal estimating functions. *Biometrika* **81**, 341–50.
POLLARD, D. (1990). *Empirical Processes: Theory and Applications,* Regional Conference Series Probability and Statistics 2. Hayward, CA: Institute of Mathematical Statistics.
POWELL, J. L. (1984). Least absolute deviations for the censored regression models. *J. Economet.* **25**, 303–25.
POWELL, J. L. (1986). Censored regression quantiles. *J. Economet.* **32**, 143–55.
SERFLING, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* New York: John Wiley.
SHAO, J. & TU, D. S. (1995). *The Jackknife and Bootstrap.* New York: Springer-Verlag.

[*Received February* 2000. *Revised August* 2000]