Yale University

EliScholar – A Digital Platform for Scholarly Publishing at Yale

Cowles Foundation Discussion Papers

Cowles Foundation

3-1-2004

Smoothed Empirical Likelihood Methods for Quantile Regression Models

Yoon-Jae Whang

SMOOTHED EMPIRICAL LIKELIHOOD METHODS
FOR QUANTILE REGRESSION MODELS

By

Yoon-Jae Whang

March 2004

COWLES FOUNDATION DISCUSSION PAPER NO. 1453

Smoothed Empirical Likelihood Methods for Quantile Regression Models

Yoon-Jae Whang[1]

Department of Economics

Korea University

and

Cowles Foundation For Research in Economics

Yale University

November 10, 2003

Revised: Feb. 26, 2004

Abstract

This paper considers an empirical likelihood method to estimate the parameters of the quantile regression (QR) models and to construct confidence regions that are accurate in finite samples. To achieve the higher-order refinements, we smooth the estimating equations for the empirical likelihood. We show that the smoothed empirical likelihood (SEL) estimator is first-order asymptotically equivalent to the standard QR estimator and establish that confidence regions based on the smoothed empirical likelihood ratio have coverage errors of order n^{-1} and may be Bartlett-corrected to produce regions with an error of order n^{-2}, where n denotes the sample size. We further extend these results to censored quantile regression models. Our results are extensions of the previous results of Chen and Hall (1993) to the regression contexts. Monte Carlo experiments suggest that the smoothed empirical likelihood confidence regions may be more accurate in small samples than the confidence regions that can be constructed from the smoothed bootstrap method recently suggested by Horowitz (1998).

1 Introduction

The quantile regression models, originally introduced by Koenker and Bassett (1978, 1982), have recently been very popular in both theoretical and applied econometrics literature, particularly due to their usefulness in characterizing the entire conditional distribution of a dependent variable given regressors and the robustness property of the quantile regression estimators to outlier observations. See Buchinsky (2000) for a recent survey.

Koenker and Bassett (1978, 1982) give conditions under which their quantile regression (hereafter QR) estimator is $n^{1/2}$-consistent and asymptotically normal. This result enables one to construct a standard asymptotic confidence region on the true parameters. However, the first-order approximation might be inaccurate with samples of the sizes encountered in many applications and hence it might yield a substantial gap between the true and the nominal coverage probabilities in practice. On the other hand, it is well known that bootstrap generally provides asymptotic refinements to the coverage probabilities of confidence regions under regularity conditions, see Beran (1988), Hall (1986, 1992), and Horowitz (1997, 2001). However, the standard theory of the bootstrap can not be directly applied to the confidence regions based on the QR estimator because the statistic of interest is not a smooth function of sample moments that has an Edgeworth expansion.[1] In his important recent contribution, Horowitz (1998) considers a median regression model and shows that one can overcome this difficulty by smoothing the least absolute deviation (LAD) objective function to make it differentiable. He shows that the resulting smoothed LAD (hereafter SLAD) estimator is asymptotically equivalent to the standard LAD estimator and bootstrap provides asymptotic refinements in the sense that, with bootstrap critical value, the rejection probabilities of symmetrical t and χ^2 tests (of linear restrictions) based on the SLAD estimator are correct up to order $O(n^{-a})$ under the null hypothesis, where $a < 1$ and n denotes the sample size. He suggests that his results also apply to coverage probabilities of confidence regions.

This paper considers an empirical likelihood method to estimate the parameters of the QR models and to construct confidence regions for the parameters. The empirical likelihood method was originally introduced by Owen (1988, 1990, 1991) and has received a lot of attention in recent econometrics literature. Examples include Bravo (2002, 2004), Donald, Imbens and Newey (2003), Guggenberger and Smith (2003), Imbens, Spady and Johnson (1998), Kitamura (1997, 2001), Kitamura and Stuzer (1997), Moon and Schorfheide (2003), Newey and Smith (2003) and Su and White (2003), to mention only a few.[2] Qin and Lawless (1994) link empirical likelihood to general estimating equations for many interesting estimators. One of the advantages

---

[1] For a first-order consistency result of bootstrap estimators in (non-smooth) QR models, see Hahn (1995).

[2] Visit also the empirical likelihood homepage of Owen (http://www-stat.stanford.edu/~owen/ empirical) for a recent update of the literature.

of empirical likelihood confidence regions is that they do not require estimation of the asymptotic covariance matrix of point estimators and allow the shapes of confidence regions to be determined automatically by the data. In contrast, classical (including bootstrap) confidence regions that depend on estimates of the asymptotic covariance matrix might sensitively depend on the quality of the estimates and typically require some subjective judgement on the shapes and orientations of the confidence regions. Also, in certain regular cases, empirical likelihood confidence regions are Bartlett correctable so that their asymptotic coverage accuracy can be improved, see e.g. DiCiccio, Hall and Romano (1991) and Hall and La Scala (1990). However, to get the asymptotic refinements, most of the existing empirical likelihood theory requires the statistic of interest to be a smooth function of sample moments. This implies that one can not directly apply the empirical likelihood method to QR models since the estimating equations for the standard QR estimator are not smooth.

In this paper, we avoid these problems by appropriately smoothing the estimating equations. We establish that the resulting smoothed empirical likelihood (SEL) estimator is first-order asymptotically equivalent to the standard QR estimator and the confidence regions based on the smoothed empirical likelihood ratio statistic have coverage errors of order $O(n^{-1})$. Furthermore, we show that the smoothed empirical likelihood (for the full parameter vector) is Bartlett correctable under suitable conditions, so that the coverage errors of confidence regions can be further reduced from order $O(n^{-1})$ to order $O(n^{-2})$.[3] We demonstrate that this improvement is possible for a wide range of smoothing parameter values and hence discussion on the concept of the "optimal" smoothing parameter is not necessary. We also provide a (heuristic) discussion on Bartlett correctability of SEL confidence regions for a sub-vector of the true parameters. We further extend our results to the *censored* quantile regression (CQR) models of Powell (1984, 1986).

There are a number of papers in the literature that are related to this paper. Previous research by Chen and Hall (1993) has shown that the smoothed confidence intervals for quantiles *with no covariates* have coverage error of order $O(n^{-1})$ and may be Bartlett-corrected to produce intervals with an error of order only $O(n^{-2})$. Our paper extends the results of Chen and Hall (1993) to the quantile regression contexts which perhaps should be more of interest to econometricians. The extension is not trivial, at least to us, because the necessary multivariate Edgeworth expansions of the smoothed models have terms that depend on bandwidth parameters, which complicates the asymptotic analysis substantially, and the proofs of the validity of

---

[3]We do not claim here that empirical likelihood is the only way of achieving such higher-order refinements. Alternative method such as double bootstrap (initially suggested by Hall (1986) and Beran (1987)) is known to enable further refinements over the standard bootstrap and hence might also yield results analogous to those obtained in this paper. However, the latter procedure can be computationally very expensive. On the other hand, in certain regular cases, it is known that empirical likelihood is the only member of the Cressie-Read family which admits a Bartlett correction, see Jing and Wood (1996) and Baggerly (1998). We expect that the same result will hold in our context under suitable assumptions.

Bartlett correction in the standard parametric and nonparametric (i.e., empirical likelihood) contexts are substantially different. For example, the standard results of Di Ciccio, Hall and Romano (1991) can not be directly applied to our contexts. On the other hand, contrary to De Angelis et. al (1993) and some of the other papers in the literature, we do not assume that the error terms in the uncensored (and censored) quantile regressions are independent of regressors (X) and hence can have unknown form of conditional heteroskedasticity. Finally, independently of our work, Otsu (2003) has recently proposed that similar results to ours hold in the uncensored quantile regression model, but he assumes independence of the error and regressors and does not provide a rigorous proof. However, the main focus of the latter paper is on the relative efficiency of smoothed conditional empirical likelihood estimators over other competing estimators and hence is different from ours.

The remainder of this paper is organized as follows: Section 2 defines the SEL estimator and confidence region in quantile regression models and discusses their asymptotic properties. Section 3 extends the previous results to censored quantile regression models. Section 4 reports some Monte Carlo results. Section 5 is a conclusion. An appendix contains proofs of the results.

2   Smoothed Empirical Likelihood for Quantile Regressions

2.1   Definition of the SEL estimator and confidence regions

In this section, we define the smoothed empirical likelihood estimator and confidence regions for the quantile regression models.

Consider the linear quantile regression model given by:

$$Y_i = X_i'\beta_0 + U_i \text{ for } i = 1, ..., n, \tag{1}$$

where $Y_i \in \mathbb{R}$ is an observed dependent variable, X_i is an observed $K \times 1$ vector of regressors, β_0 is a $K \times 1$ vector of constant parameters, and U_i is an unobserved error that satisfies $P[U_i \leq 0|X_i] = q$ a.s. $\forall i \geq 1$ for $0 \leq q \leq 1$. For simplicity, we assume that $\{(Y_i, X_i) : i = 1, ..., n\}$ are i.i.d.

To motivate our estimator, consider the following estimating equations:

$$Eg(Y_i, X_i, \beta_0) = 0, \tag{2}$$

where

$$g(Y_i, X_i, \beta) = [1(Y_i \leq X_i'\beta) - q] X_i \tag{3}$$

and $1(\cdot)$ denotes the indicator function. Note that the function $g(Y_i, X_i, \beta)$ is not differentiable at points β such that $Y_i = X_i'\beta$ for some i. This causes some problem to our subsequent (higher-order) asymptotic analysis because most of theoretical

development of empirical likelihood has focused on the statistic which is a *smooth function of sample moments*. In this paper, we circumvent this problem by smoothing the function g, i.e., by replacing the indicator function in g with a smooth function.

For this purpose, let $K(\cdot)$ denote a kernel function that is bounded, compactly supported on $[-1, 1]$ and integrated to one. Additional assumptions on $K(\cdot)$ are given below. Define $G(x) = \int_{u<x} K(u)du$ and $G_h(x) = G(x/h)$. Then, a smoothed version of g in (3) may be given by

$$Z_i(\beta) = (G_h(X_i'\beta - Y_i) - q) X_i. \tag{4}$$

Let $p = (p_1, ..., p_n)'$ be a vector of nonnegative numbers adding to unity. Then, the *smoothed empirical log likelihood ratio* is defined by

$$l_h(\beta) = -2 \min_{p:\, \sum p_i Z_i(\beta)=0} \sum_{i=1}^{n} \log(np_i). \tag{5}$$

For given β, using the standard Lagrange multiplier arguments, the optimal value for p_i solving (5) can be shown to be

$$p_i(\beta) = n^{-1} \left(1 + t(\beta)' Z_i(\beta)\right)^{-1}, \tag{6}$$

where $t(\beta)$ is a $K \times 1$ vector of Lagrange multipliers satisfying

$$n^{-1} \sum_{i=1}^{n} Z_i(\beta) / \left(1 + t(\beta)' Z_i(\beta)\right) = 0. \tag{7}$$

This gives the (profile) smoothed empirical log likelihood ratio statistic:

$$l_h(\beta) = 2 \sum_{i=1}^{n} \log(1 + t(\beta)' Z_i(\beta)), \tag{8}$$

where $t(\beta)$ satisfies (7). By definition, the SEL estimator $\widehat{\beta}_E$ of β_0 solves

$$\min_{\beta \in B} l_h(\beta) \tag{9}$$

where B is the parameter space.[4]

---

[4]In practice, since $l_h(\beta)$ is a smooth function of β, $\widehat{\beta}_E$ can be computed by using a nested algorithm as in Owen (1990) in which the inner stage solves for $t(\beta)$ that satisfies (7) for fixed values of β and the outer stage minimizes $l_h(\beta)$ in (8) over $\beta \in B$. Alternatively, as in Hall and La Scala (1990), one can use a multivariate Newton's algorithm that jointly solves the nonlinear system of $2K$ first-order conditions given in Lemma 3 in Appendix. See also Owen (2001, Ch. 12) for more examples of alternative algorithms.

We now compare the SEL estimator with the standard QR estimator. The standard QR estimator $\widehat{\beta}_Q$ of β_0 solves

$$\min_{\beta \in B} H_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \rho_q(Y_i - X_i'\beta), \tag{10}$$

where $\rho_q(x) = [q - 1(x \leq 0)] x$ is the check function. When $q = 1/2$, the estimator is the standard LAD estimator. Koenker and Bassett (1978, 1982) show that $\widehat{\beta}_Q$ is $n^{1/2}$-consistent and asymptotically normal. Intuitively, it is reasonable to expect that $\widehat{\beta}_Q$ and $\widehat{\beta}_E$ are asymptotically equivalent if h goes to zero sufficiently fast as $n \to \infty$. This is because, under regularity conditions, $\widehat{\beta}_Q$ satisfies the first-order condition (FOC)

$$n^{-1} \sum_{i=1}^{n} g(Y_i, X_i, \beta) = n^{-1} \sum_{i=1}^{n} [1(Y_i \leq X_i'\beta) - q] X_i = 0 \tag{11}$$

with probability that goes to one as $n \to \infty$, which is also an unsmoothed version (i.e., $h = 0$) of the estimating equations $\sum p_i Z_i(\beta) = 0$ for the smoothed empirical likelihood (5). Under the regularity conditions given below, we shall show that the two estimators are (first-order) asymptotically equivalent in the sense that

$$\sqrt{n} \left(\widehat{\beta}_E - \widehat{\beta}_Q \right) = o_p(1) \text{ as } n \to \infty. \tag{12}$$

This result implies that the asymptotic distribution of the SEL estimator is given by that of the usual QR estimator, i.e.,

$$\sqrt{n} \left(\widehat{\beta}_E - \beta_0 \right) \xrightarrow{d} N(0, \Lambda_0), \tag{13}$$

where

$$\Lambda_0 = q(1-q) D_0^{-1} S_0 D_0^{-1}, \tag{14}$$

$$S_0 = E[X_i X_i'], \ D_0 = E[f(0|X_i) X_i X_i'], \tag{15}$$

and $f(u|x)$ denote the conditional density of U given $X = x$. On the other hand, when $q = 1/2$, the result (12) and Horowitz (1998)'s theorem 2.1 imply that $\widehat{\beta}_E$ is also asymptotically equivalent to the SLAD estimator of Horowitz (1998) in the first-order approximation.

Now, we define the SEL confidence regions. Consider the smoothed empirical log likelihood ratio statistic given in (8). The *SEL confidence region* for $\beta_0 \in \mathbb{R}^K$ is defined by

$$I_{hc} = \{\beta : l_h(\beta) \leq c\}, \tag{16}$$

where $c > 0$ is a constant which determines the coverage probability α_{hc} of I_{hc} :

$$\alpha_{hc} = P(\beta_0 \in I_{hc}) = P(l_h(\beta_0) \leq c). \tag{17}$$

The coverage accuracy of I_{hc} depends on the asymptotic distribution of $l_h(\beta_0)$ statistic. As we shall see below, under suitable regularity conditions, $l_h(\beta_0)$ has an asymptotic χ_K^2 distribution and hence c might be chosen using this result. On the other hand, the SEL confidence region for a subvector $\beta_{10} \in \mathbb{R}^{K_1}$ of the parameter vector $\beta_0 = (\beta_{10}', \beta_{20}')'$ is defined by

$$I_{hc,1} = \left\{\beta_1 : l_h(\beta_1, \widetilde{\beta}_2) \leq c\right\},$$

where $\widetilde{\beta}_2$ minimizes $l_h(\beta_1, \beta_2)$ with respect to β_2 holding β_1 fixed. We shall also establish that $l_h(\beta_{10}, \widetilde{\beta}_2)$ converges in distribution to $\chi_{K_1}^2$ distribution under suitable conditions, and so c might be chosen using the latter distribution.[5] As noted by Chen and Hall (1993), if G_h is a higher-order kernel, then it is possible that I_{hc} or $I_{hc,1}$ might be a union of disjoint convex sets for small n and unusual values of h.

Now we comment on the main features of the SEL confidence regions. First, since they are based on the likelihood function, they do not depend on any explicit estimate of Λ_0. This is an advantage over the confidence regions that are based on Wald-type statistics (such as (45) - (48) below), which depend on explicit estimates of Λ_0 and might subsequently create problems regarding the quality of the estimates. Second, the shapes of the SEL confidence regions are not restricted *a priori* to be elliptical or rectangular and are allowed to be determined by the likelihood or, equivalently, by the data.[6] See also Wu (1986). Furthermore, as in the standard parametric contexts, we shall show that the SEL confidence regions are Bartlett-correctable provided the smoothing parameter is suitably chosen and other regularity conditions hold, improving higher-order accuracy of inferences.

2.2   Asymptotic Equivalence and Coverage Accuracy

In this section, we derive the asymptotic distribution of the SEL estimator and establish asymptotic equivalence of the SEL and QR estimators. We also discuss asymptotic coverage accuracy of the SEL confidence regions.

---

[5] In practice, the contours of I_{hc} or $I_{hc,1}$ can be computed using a multivariate Newton's algorithm as in Hall and La Scala (1990). In our simulation experiments below, we used the modified Newton algorithm written in gauss codes by Bruce Hansen (available at http://www.ssc.wisc.edu/~bhansen/progs/elike.prc).

[6] This feature is also shared by bootstrap confidence regions constructed by multivariate kernel density estimation applied to the resampled data (viz. Hall (1987)) or by constructing polygons to the resampled data (viz. Owen (1990)), but these methods do not seem to be very satisfactory, see Owen (2001, Ch.1).

Let $r \geq 2$ be an integer. We denote $F(\cdot|x)$ to be the CDF of U_i conditional on $X_i = x$ and define $f(\cdot|x)$ to be the conditional density of U_i with respect to Lebesgue measure whenever it exists. We need the following assumptions for our main results.

Assumption 1: $\{(Y_i, X_i) : i = 1, ..., n\}$ are independent and identically distributed random vectors.

Assumption 2: The parameter vector β_0 is an interior point of the compact parameter space B in $\mathbb{R}^K$.

Assumption 3: X_i has bounded support and S_0 and D_0 are nonsingular.

Assumption 4: (a) $F(0|x) = q$ for almost every x. (b) For all u in a neighborhood of 0 and almost every x, $f(u|x)$ exists, is bounded away from zero, and is r times continuously differentiable with respect to u.

Assumption 5: (a) $K(\cdot)$ is bounded and compactly supported on $[-1, 1]$. (b) For some constant $C_K \neq 0$, $K(\cdot)$ satisfies

$$\int u^j K(u) du = \begin{cases} 1, & \text{if } j = 0, \\ 0, & \text{if } 1 \leq j \leq r - 1, \\ C_K, & \text{if } j = r. \end{cases} \tag{18}$$

(c) Let $\widetilde{G}(u) = \left([G(u)], [G(u)]^2, ..., [G(u)]^{L+1}\right)'$ for some $L \geq 1$, where $G(u) = \int_{v < u} K(v) dv$. For any $\theta \in \mathbb{R}^{L+1}$ satisfying $\|\theta\| = 1$, there is a partition of $[-1, 1]$, $-1 = a_0 < a_1 < \cdots < a_M = 1$ such that $\theta' \widetilde{G}(u)$ is either strictly positive or strictly negative on (a_{m-1}, a_m) for $l = 1, ..., L + 1$.

Assumption 6: h satisfies (a) $nh^{2r} \to 0$ and (b) $nh/\log n \to \infty$ as $n \to \infty$.

Assumptions 1-5 are similar to Assumptions 1-5 of Horowitz (1998, p.1333), which were used to establish asymptotic refinement of the SLAD estimator-based t and χ^2 tests through bootstrap. Assumptions 1-5(b) are used to establish the asymptotic normality of $\sqrt{n}(\widehat{\beta}_E - \beta_0)$ and to justify a Taylor expansion for the empirical likelihood ratio statistic which in turn is used to calculate the coverage probabilities of our SEL confidence regions. The boundedness assumption for X_i (Assumption 3) is made to simplify the proofs in Appendix. It can be removed at the cost of more complicated proofs. Assumption 5(c) is used to verify a version of the Cramér's condition (Lemma 4 of Appendix) which is necessary to justify a formal Edgeworth expansion for the distribution of $l_h(\beta_0)$.

Assumption 6 requires that h goes to zero as $n \to \infty$ at a suitable rate. It is satisfied if $h \propto n^{-\kappa}$ for $1/(2r) < \kappa < 1$, where $r \geq 2$. The part (a) of Assumption 6 ensures that the smoothing has an asymptotically negligible effect on the distribution of $l_h(\beta_0)$. On the other hand, the part (b) of Assumption 6 requires that h should not be too small. It is needed to ensure a minimum level of smoothness of $l_h(\beta_0)$ which is necessary to derive the Cramér's condition for the Edgeworth analysis. Intuitively this assumption makes sense, because the Cramér's condition is usually intended to ensure distributions of statistics to have an absolutely continuous component but the

7

latter might be hard to attain for $l_h(\beta_0)$ if h is chosen too small, see Hall (1992, p.57) for a general interpretation of the Cramér's condition.

We now derive the asymptotic distribution of the SEL estimator and establish asymptotic equivalence of the SEL and QR estimators.

Theorem 1 *Under Assumptions 1-5(b) and 6(a) of Section 2.2, we have*

$$(a) \ n \left(\widehat{\beta}_E - \widehat{\beta}_Q \right) = o_p(1) \ and$$

$$(b) \ \sqrt{n} \left(\widehat{\beta}_E - \beta_0 \right) \overset{d}{\rightarrow} N(0, \Lambda_0),$$

where Λ_0 is defined in (14).

The asymptotic covariance matrix Λ_0 can be estimated, for example, by

$$\widehat{\Lambda} = q(1-q)\widehat{D}^{-1}\widehat{S}\widehat{D}^{-1} \ , \ \text{where} \tag{19}$$

$$\widehat{D} = h^{-1}\sum_{i=1}^{n} n^{-1} K \left(\left(Y_i - X_i'\widehat{\beta}_E \right)/h \right) X_i X_i' \ \text{and} \ \widehat{S} = \sum_{i=1}^{n} n^{-1} X_i X_i'. \tag{20}$$

This estimator is analogous to the covariance matrix estimator of Powell (1984, 1986) in the standard QR model. Alternatively, one may estimate Λ_0 using $\widehat{\Lambda}$ with n^{-1} in (20) replaced by $p_i(\widehat{\beta}_E)$, where $p_i(\cdot)$ is as defined in (6). As shown by Qin and Lawless (1994), the latter estimator should be more efficient than the former in finite samples because it fully exploits the restriction $\sum_{i=1}^{n} p_i(\widehat{\beta}_E) Z_i(\widehat{\beta}_E) = 0$. Under the assumptions of Theorem 1, it is not difficult to show that both estimators are consistent for Λ_0. Another way to estimate Λ_0 is to use a bootstrap estimator as in Buchinsky (1995, 2000), see Section 4 below for an example. The bootstrap estimator has an advantage in the sense that it does not require a choice of h, but its computation can be more demanding than the kernel-based estimators.

We now discuss coverage properties of the SEL confidence regions. To this end, it is convenient to write the empirical log likelihood-ratio statistic $l_h(\beta)$ (given by (8) and (7)) at $\beta = \beta_0$ in terms of standardized variables. That is, we let

$$\lambda = V_n^{1/2}t \ \text{and} \ W_i = V_n^{-1/2}Z_i \tag{21}$$

for $i = 1, ..., n$, where $t = t(\beta_0)$, $Z_i = Z_i(\beta_0)$ and $V_n = EZ_iZ_i'$. Then, in terms of the standardized variables λ and W_i, $l_h(\beta_0)$ can be re-written as

$$l_h(\beta_0) = 2 \sum_{i=1}^{n} \log(1 + \lambda'W_i), \tag{22}$$

8

where λ satisfies

$$n^{-1} \sum_{i=1}^{n} W_i / (1 + \lambda' W_i) = 0. \qquad (23)$$

We need to introduce a few more notation. We let W_i^j denote the j-th component of W_i and define

$$
\begin{aligned}
\alpha^{j_1 \cdots j_k} &= E W_i^{j_1} \cdots W_i^{j_k}, \qquad (24)\\
\overline{A}^{j_1 \cdots j_k} &= n^{-1} \sum_{i=1}^{n} W_i^{j_1} \cdots W_i^{j_k}, \text{ and } A^{j_1 \cdots j_k} = \overline{A}^{j_1 \cdots j_k} - \alpha^{j_1 \cdots j_k}.
\end{aligned}
$$

In particular, $\alpha^{jk} = \delta^{jk}$, where δ^{jk} is the Kronecker delta.[7]

Under the regularity conditions, we first establish that $l_h(\beta_0)$ has an asymptotic χ_K^2 distribution.

Theorem 2 *Suppose Assumptions 1-5(b) and 6(a) hold. Then, we have*

$$l_h(\beta_0) \xrightarrow{d} \chi_K^2$$

as $n \to \infty$.

Remarks: 1. Theorem 2 is a nonparametric version of Wilks (1938)' theorem, which has first been proved by Owen (1991) in the standard linear regression models. Chen and Hall (1993, Theorem 3.1) have also established a similar result for the quantiles (with no covariates).

2. From the expansion (A.12) and Lemma 1(a) in Appendix, we can see that $n^{1/2} E Z_i \to 0$ if $nh^{2r} \to 0$ and, if $E\left[X_i f^{(r-1)}(0|X_i)\right] \neq 0$, $n^{1/2} E Z_i \to 0$ implies $nh^{2r} \to 0$. Therefore, if $E\left[X_i f^{(r-1)}(0|X_i)\right] \neq 0$, the bandwidth condition 6(a) is in fact a necessary and sufficient condition for $l_h(\beta_0)$ to have an asymptotic χ_K^2 distribution.

If $c = c_\alpha$ is chosen such that

$$P(\chi_K^2 \leq c_\alpha) = \alpha, \qquad (25)$$

then Theorem 2 implies that the asymptotic coverage of the SEL confidence region I_{hc} will be α, i.e.,

$$P(\beta_0 \in I_{hc}) = P(l_h(\beta_0) \leq c_\alpha) = \alpha + o(1)$$

as n goes to infinity.

Similarly, when one is interested in constructing a confidence region for a subvector $\beta_{10} \in \mathbb{R}^{K_1}$ of the parameter vector $\beta_0 = (\beta_{10}', \beta_{20}')' \in \mathbb{R}^K$, one can use the following result.

---

[7]This α-A notation was originally used by Di Ciccio, Hall and Romano (1991).

Corollary 1 *Suppose Assumptions 1-5(b) and 6(a) hold. Then, we have*

$$l_h(\beta_{10}, \widetilde{\beta}_2) \xrightarrow{d} \chi^2_{K_1}$$

as $n \to \infty$, where $\widetilde{\beta}_2$ minimizes $l_h(\beta_{10}, \beta_2)$ with respect to β_2.

Corollary 1 implies that the SEL confidence region $I_{hc,1} = \left\{ \beta_1 : l_h(\beta_1, \widetilde{\beta}_2) \leq c_{\alpha,1} \right\}$ with $c_{\alpha,1}$ satisfying $P(\chi^2_{K_1} \leq c_{\alpha,1}) = \alpha$ has asymptotic coverage error α, as desired.

We now discuss the higher-order properties of the SEL confidence regions. Using an Edgeworth expansion of the distribution of $l_h(\beta_0)$, we can show that the asymptotic coverage accuracy of I_{hc} is in fact of order $O(n^{-1})$:

Theorem 3 *Define $c = c_\alpha$ by (25). Suppose Assumptions 1- 6 hold. If we further assume that $\sup_n nh^r < \infty$, then we have*

$$P(\beta_0 \in I_{hc}) = \alpha + O(n^{-1}) \tag{26}$$

as $n \to \infty$.

Remarks: 1. The expansion (A.31) in Appendix implies that the bandwidth condition $\sup_n nh^r < \infty$ is not only sufficient but also necessary for the asymptotic coverage error to be of order $O(n^{-1})$ if $E\left[X_i f^{(r-1)}(0|X_i)\right] \neq 0$. If $E\left[X_i f^{(r-1)}(0|X_i)\right] = 0$, then the result of Theorem 3 still holds even if nh^r diverges as long as $nh^{2r} \to 0$ and $nh/\log n \to \infty$, i.e. Assumption 6 holds.

2. When $nh^r \to C \ (< \infty)$, the expansion (A.31) and the results (A.28) - (A.30) in Appendix may be used to derive the "optimal" value of C that minimizes the $O(n^{-1})$ term on the right hand side of (26). However, this possibility is not practically of interest because of the availability of Bartlett correction, which is discussed in the next section.

2.3 Bartlett Correction

In the previous section, the coverage error of the empirical likelihood confidence region is I_{hc} of order $O(n^{-1})$. This error might be partly explained by the fact that the mean of the distribution of $l_h(\beta_0)$ does not agree with that of χ^2_K distribution, i.e., $E[l_h(\beta_0)] \neq K$. Therefore, one might suspect that this discrepancy might be diminished by rescaling $l_h(\beta_0)$ so that it has correct mean. This idea is known as the *Bartlett correction* in the literature. In this section we show that, provided h is chosen suitably, the Bartlett correction reduces the coverage error to $O(n^{-2})$.

From expansion (A.11), we can show that if $nh^r \to 0$

$$E[l_h(\beta_0)] = K\left(1 + n^{-1}b\right) + o(n^{-1}),$$

10

where
$$b = K^{-1} \left(\alpha^{iikk}/2 - \alpha^{ikm} \alpha^{ikm}/3 \right). \tag{27}$$

Here and throughout this paper, we use the convention that terms with repeated superscripts are to be summed over. The result (27) suggests that we might consider a confidence region corrected with the *Bartlett factor* b:

$$I_{hc}^b = \left\{ \beta : l_h(\beta) \leq c(1 + n^{-1}b) \right\}. \tag{28}$$

In practice, b is not observed and has to be estimated. Let $\widehat{\beta}$ denote any $n^{1/2}$- consistent estimator of β_0 such as the SEL estimator $\widehat{\beta}_E$ or the usual quantile regression estimator $\widehat{\beta}_Q$. Define the estimated Bartlett factor to be

$$\widehat{b} = K^{-1} \left(\widehat{\alpha}^{iikk}/2 - \widehat{\alpha}^{ikm} \widehat{\alpha}^{ikm}/3 \right), \tag{29}$$

where

$$\widehat{\alpha}^{iikk} = n^{-1} \sum_{j=1}^{n} \widehat{\varepsilon}_j^4 \left(X_j' \widehat{V}_n^{-1} X_j \right)^2, \tag{30}$$

$$\widehat{\alpha}^{ikm} = n^{-1} \sum_{j=1}^{n} \widehat{\varepsilon}_j^3 \widehat{v}_{ni}^{-1/2} X_j \widehat{v}_{nk}^{-1/2} X_j \widehat{v}_{nm}^{-1/2} X_j,$$

$$\widehat{V}_n = n^{-1} \sum_{j=1}^{n} \widehat{\varepsilon}_j^2 X_j X_j', \quad \widehat{\varepsilon}_j = G_h \left(X_j' \widehat{\beta} - Y_j \right) - q,$$

and $\widehat{v}_{ni}^{-1/2}$ is the i-th row of $\widehat{V}_n^{-1/2}$. With some calculation, one can show that

$$\widehat{\alpha}^{ikm} \widehat{\alpha}^{ikm} = n^{-2} \sum_{j=1}^{n} \sum_{l=1}^{n} \widehat{\varepsilon}_j^3 \widehat{\varepsilon}_l^3 \left(X_j' \widehat{V}_n^{-1} X_l \right)^3. \tag{31}$$

The confidence region corrected with $\widehat{b}$ is now defined to be

$$I_{hc}^{\widehat{b}} = \left\{ \beta : l_h(\beta) \leq c(1 + n^{-1} \widehat{b}) \right\}. \tag{32}$$

Theorem 4 below shows that the coverage error of the SEL confidence region is of order $O(n^{-2})$ if it is Bartlett corrected by either b or $\widehat{b}$.

On the other hand, from (A.29) and (A.30) in Appendix, we have

$$\alpha^{iikk} = q^{-1}(1 - q)^{-1}(1 - 3q + 3q^2)E \left\{ \left(X_j' S_0 X_j \right)^2 \right\} + O(h)$$

and

$$\alpha^{ikm} = q^{-1/2}(1 - q)^{-1/2}(1 - 2q)E \left\{ \left(s_i^{-1/2} X_j \right) \left(s_k^{-1/2} X_j \right) \left(s_m^{-1/2} X_j \right) \right\} + O(h),$$

11

where $s_i^{-1/2}$ denotes the i-th row of $S_0^{-1/2}$. This suggests that one might also consider a confidence region

$$\tilde{I}_{hc}^{b} = \left\{ \beta : l_h(\beta) \le c(1 + n^{-1}\tilde{b}) \right\}. \tag{33}$$

with Bartlett factor given by

$$
\begin{aligned}
\tilde{b} &= K^{-1}\left[2^{-1}(1 - 3q + 3q^2)q^{-1}(1-q)^{-1}\left\{ n^{-1}\sum_{j=1}^{n}\left(X_j'\overline{S}^{-1}X_j \right)^2 \right\} \right. \\
&\quad \left. -3^{-1}(1-2q)^2 q^{-1}(1-q)^{-1}\left\{ n^{-2}\sum_{j=1}^{n}\sum_{l=1}^{n}\left(X_j'\overline{S}^{-1}X_l \right)^3 \right\} \right], \tag{34}
\end{aligned}
$$

where $\overline{S} = n^{-1}\sum_{k=1}^{n} X_k X_k'$. However, if $\tilde{b}$ is used instead of b, we will not have the same asymptotic accuracy as b or $\hat{b}$ due to relatively large estimation error of $\tilde{b}$. This is because we have $\tilde{b} = b + O(n^{-1/2}) + O(h)$ and hence, with Bartlett factor $\tilde{b}$, the coverage error is of order $O(n^{-1}h)$ instead of $O(n^{-2})$.

The following theorem formally states the above results:

Theorem 4 *Define $c = c_\alpha$ by (25). Suppose Assumptions 1- 6 hold. If we further assume that $\sup_n n^3 h^{2r} < \infty$, then we have*

(a) $P(\beta_0 \in I_{hc}^{b}) = \alpha + O(n^{-2})$; (b) $P(\beta_0 \in I_{hc}^{\hat{b}}) = \alpha + O(n^{-2})$; (c) $P(\beta_0 \in \tilde{I}_{hc}^{b}) = \alpha + O(n^{-1}h)$

as $n \to \infty$.

Remark: The result (A.35) in Appendix implies that the condition $\sup_n n^3 h^{2r} < \infty$ is also necessary for the asymptotic coverage error of I_{hc}^{b} or $I_{hc}^{\hat{b}}$ to be of order $O(n^{-2})$ if $E\left[X_i f^{(r-1)}(0|X_i) \right] \ne 0$.

We now discuss Bartlett correctability of $I_{hc,1}$. Lazar and Mykland (1999) show that the empirical likelihood defined by two estimating equations in the presence of one nuisance parameter is not Bartlett correctable. This casts a serious doubt on the Bartlett correctability of $I_{hc,1}$. Recently, however, Chen and Cui (2002, 2003) show that, if the nuisance parameter is profiled out given the parameter of interest, the empirical likelihood is still Bartlett correctable. They propose that "the real cause of not being Bartlett correctable found in Lazar and Mykland (1999) is due to plugging-in a global maximum likelihood estimate for the nuisance parameter rather than any fundamental differences between estimating equations and the smooth function of means." Therefore, it would be interesting to see if one could extend the Chen and Cui (2002, 2003)'s result to $I_{hc,1}$. However, a formal investigation of such result is beyond the scope of this paper, because profiling out β_2 given $\beta_1 = \beta_{10}$ requires an additional Edgeworth expansion of the profile empirical likelihood which is substantially more complicated than the one given in Appendix as well as that of Chen and Cui (2002,

2003).[8] However, a practical solution in this situation would be to use the following bootstrap procedure[9] :

(i) Using the original sample $\chi = \{(Y_i, X_i) : i = 1, ..., n\}$, compute $\widehat{\beta} = (\widehat{\beta}'_1, \widehat{\beta}'_2)'$ by minimizing $l_h(\beta)$ with respect to β.

(ii) Draw bootstrap samples $\chi^*_b = \{(Y^*_{bi}, X^*_{bi}) : i = 1, ..., n\}$ for $1 \leq b \leq B$ randomly with replacement from the original sample χ.

(iii) Letting $l^*_{bh}(\beta_1, \beta_2)$ be the value of $l_h(\beta_1, \beta_2)$ computed from χ^*_b instead of χ, compute $l^*_{bh}(\widehat{\beta}_1, \overline{\beta}_2)$, where $\overline{\beta}_2$ minimizes $l^*_{bh}(\widehat{\beta}_1, \beta_2)$ with respect to β_2 holding $\widehat{\beta}_1$ fixed.

(iii) Estimate the bootstrap Bartlett factor $\widehat{b}_{1B}$ by solving the equation [10]

$$B^{-1} \sum_{b=1}^{B} l^*_{bh}(\widehat{\beta}_1, \overline{\beta}_2) = K_1 \left(1 + n^{-1} \widehat{b}_{1B}\right).$$

(iv) The Bartlett corrected confidence region is given by

$$I^{\widehat{b}_{1B}}_{hc,1} = \left\{\beta_1 : l_h(\beta_1, \widetilde{\beta}_2) \leq c_{\alpha,1}(1 + n^{-1}\widehat{b}_{1B})\right\}.$$

Although we do not prove here that $I^{\widehat{b}_{1B}}_{hc,1}$ has an asymptotic coverage error of order $O(n^{-2})$ as in the full parameter vector case, we expect that this correction may still be expected to improve upon the approximation of the distribution of the (smoothed) empirical likelihood.[11]

3 Extension to Censored Quantile Regressions

In this section, we extend the previous results to the censored quantile regression model of Powell (1984, 1986). The model is given by

---

[8]For Bartlett correction of $I_{hc,1}$, we need a higher-order Taylor expansion of the $2K$ first-order equations (A.13)-(A.14) around $(\beta_{20}, 0)$. But, the expansion of (A.14) introduces many additional terms, which makes computation of higher-order cumulants of the signed root of $l_h(\beta_{10}, \widetilde{\beta}_2)$ complicated. They are even more complicated than in Chen and Cui (2002, 2003) because the terms in general depend on bandwidth parameter h which interacts with the sample size n.

[9]The idea of using a bootstrap procedure for Bartlett correction is orginally due to Hall and La Scala (1990). We extend their idea to account for the presence of nuisance parameters.

[10]A bootstrap procedure similar to this can also be used to estimate Bartlett factor b for the confidence region I^b_{hc} for the full parameter vector. In this case, the bootstrap estimator $\widehat{b}_B$ solves $B^{-1} \sum_{b=1}^{B} l^*_{bh}(\widehat{\beta}) = K_1(1 + n^{-1}\widehat{b}_B)$.

[11]In a context different from ours, Monti (1997) shows that a Bartlett correction via bootstrapping might still yield asymptotic refinements in finite samples, even if it does not reduce the coverage error to $O(n^{-2})$.

$$Y_i = \max\{0, \ X_i'\beta_0 + U_i\} \ \text{for } i = 1, ..., n, \tag{35}$$

where Y_i, X_i, and U_i are as defined in (1).

The censored quantile regression (CQR) estimator $\widehat{\beta}^*$ of β_0 solves

$$\min_{\beta \in B} H_n^*(\beta) = n^{-1} \sum_{i=1}^{n} \rho_q \left(Y_i - \max\{0, \ X_i'\beta\}\right), \tag{36}$$

where B is the parameter space and $\rho_q(x)$ is the check function as in (10). Under regularity conditions, $\widehat{\beta}^*$ satisfies the first-order condition (FOC)

$$n^{-1} \sum_{i=1}^{n} \left[1(Y_i \leq X_i'\widehat{\beta}^*) - q\right] 1(X_i'\widehat{\beta}^* > 0)X_i = 0$$

with probability that goes to one as $n \to \infty$. This motivates us to consider the estimating function

$$g^*(Y_i, X_i, \beta) \equiv \left[1(Y_i \leq X_i'\beta) - q\right] 1(X_i'\beta > 0)X_i$$

for our empirical likelihood. However, like the function g in (3), g^* is not smooth. Therefore, we replace the indicator functions in g^* with smooth functions and consider

$$Z_i^*(\beta) = \left(G_h(X_i'\beta - Y_i) - q\right) G_h(X_i'\beta)X_i, \tag{37}$$

as our estimating functions, where G_h is as in (4). Given this, the smoothed empirical log likelihood ratio statistic for the CQR model is now defined by

$$l_h^*(\beta) = 2 \sum_{i=1}^{n} \log(1 + t^*(\beta)'Z_i^*(\beta)), \tag{38}$$

where $t^*(\beta)$ satisfies $n^{-1} \sum_{i=1}^{n} Z_i^*(\beta) / \left(1 + t^*(\beta)'Z_i^*(\beta)\right) = 0$. By definition, the SEL estimator $\widehat{\beta}_E^*$ of β_0 solves $\min_{\beta \in B} l_h^*(\beta)$, where B is the parameter space. Under assumptions given below, we may show that the CQR and SEL estimators are asymptotically equivalent in the sense that $\sqrt{n}\left(\widehat{\beta}_E^* - \widehat{\beta}^*\right) = o_p(1)$ as $n \to \infty$. Therefore, this result and asymptotic normality of $\sqrt{n}\left(\widehat{\beta}^* - \beta_0\right)$ (see Powell (1984, 1986)) imply that the SEL estimator satisfies

$$\sqrt{n}\left(\widehat{\beta}_E^* - \beta_0\right) \xrightarrow{d} N(0, \Lambda_0^*),$$

where

$$\begin{aligned}
\Lambda_0^* &= q(1-q)D_0^{*-1}S_0^*D_0^{*-1}, \ S_0^* = E[1(X_i'\beta_0 > 0)X_iX_i'], \\
D_0^* &= E\left[f(0|X_i)1(X_i'\beta_0 > 0)X_iX_i'\right].
\end{aligned}$$

14

For a discussion on consistent estimators of Λ_0^*, see Powell (1984, 1986) or Buchinsky (1995, 2000).

We now discuss the confidence region for β_0. The confidence region for $\beta_0 \in \mathbb{R}^K$ based on the smoothed empirical log likelihood ratio is defined by

$$I_{hc}^* = \{\beta : l_h^*(\beta) \le c\}, \tag{39}$$

where $c > 0$ is a constant. Under conditions given below, $l_h^*(\beta_0) \xrightarrow{d} \chi_K^2$ and hence, if c is chosen from χ_K^2 distribution, the SEL confidence region I_{hc}^* has asymptotically correct coverage. Similarly, the confidence region for a sub-vector $\beta_{10} \in R^{K_1}$ is defined by

$$I_{hc,1}^* = \left\{\beta_1 : l_h^*(\beta_1, \widetilde{\beta}_2) \le c\right\}, \tag{40}$$

where $\widetilde{\beta}_2$ minimizes $l_h(\beta_1, \beta_2)$ with respect to β_2 holding β_1 fixed, has asymptotically correct coverage if c is chosen from $\chi_{K_1}^2$ distribution.

If the bandwidth h is chosen suitably, we may ensure that the coverage accuracy of I_{hc}^* is of order $O(n^{-1})$. The coverage error can be further reduced to order $O(n^{-2})$ if we apply a Bartlett correction to the confidence region I_{hc}^* and h is chosen suitably. To define the Bartlett factor, let $\alpha^{j_1 \cdots j_k}$ be defined as in (24), but with W_i replaced by $W_i^* = V_n^{*-1/2} Z_i^*$, where $V_n^* = E Z_i^* Z_i^*$ and $Z_i^* = Z_i^*(\beta_0)$. After this change, the Bartlett factor b^* is defined to be the same as b in (27). The estimated Bartlett factor $\widehat{b}^*$ is defined to be

$$\widehat{b}^* = K^{-1} \left((2n)^{-1} \sum_{j=1}^n \widehat{\varepsilon}_j^{*4} \left(X_j' \widehat{V}_n^{*-1} X_j \right)^2 - (3n^2)^{-1} \sum_{j=1}^n \sum_{l=1}^n \widehat{\varepsilon}_j^{*3} \widehat{\varepsilon}_l^{*3} \left(X_j' \widehat{V}_n^{*-1} X_l \right)^3 \right) \tag{41}$$

where

$$\widehat{V}_n^* = n^{-1} \sum_{j=1}^n \widehat{\varepsilon}_j^{*2} X_j X_j', \; \widehat{\varepsilon}_j^* = \left(G_h \left(X_j' \widehat{\beta} - Y_j \right) - q \right) G_h(X_j' \widehat{\beta}), \tag{42}$$

and $\widehat{\beta}$ is a $n^{1/2}$ - consistent estimator such as $\widehat{\beta}_E^*$ or $\widehat{\beta}^*$. On the other hand, by the same reasoning as in (34), we might also consider the Bartlett factor

$$\begin{aligned} \widetilde{b}^* &= K^{-1} \left[2^{-1}(1 - 3q + 3q^2) q^{-1} (1-q)^{-1} \left\{ n^{-1} \sum_{j=1}^n \left(X_j^{*'} \overline{S}^{*-1} X_j^* \right)^2 \right\} \right. \\ &\quad \left. - 3^{-1}(1-2q)^2 q^{-1} (1-q)^{-1} \left\{ n^{-2} \sum_{j=1}^n \sum_{l=1}^n \left(X_j^{*'} \overline{S}^{*-1} X_l^* \right)^3 \right\} \right], \end{aligned} \tag{43}$$

where $\overline{S}^* = n^{-1} \sum_{m=1}^n X_m^* X_m^{*'}$, $X_m^* = 1(X_m' \widehat{\beta} > 0) X_m$ for $m = 1, ..., n$ and $\widehat{\beta}$ is as in (42). We define the SEL confidence region corrected with Bartlett factor b given by $b^*, \widehat{b}^*$ or $\widetilde{b}^*$ to be

$$I_{hc}^{*b} = \left\{\beta : l_h^*(\beta) \le c(1 + n^{-1}b)\right\}. \tag{44}$$

15

To establish the above claims, we need to modify Assumption 3 as follows:

Assumption 3*: X_i has bounded support, $P(X_i\beta_0 = 0) = 0$, and $E\left[1\left(X_i'b > \varepsilon\right)X_iX_i'\right]$ is nonsingular for some $\varepsilon > 0$ and all b in a neighborhood of β_0.

The following theorem shows that the SEL estimator and CQR estimator are asymptotically equivalent and the SEL confidence region has asymptotically correct coverage and we may achieve an asymptotic higher-order improvement through Bartlett correction of I_{hc}^*.

Theorem 5 *Suppose that Assumptions 1, 2, 3*, 4, 5(b) and 6(a) hold. Define $c = c_\alpha$ by (25). Then, as $n \to \infty$, we have*

$$(a) \sqrt{n}\left(\widehat{\beta}_E^* - \widehat{\beta}^*\right) = o_p(1),$$

$$(b)\ l_h^*(\beta_0) \xrightarrow{d} \chi_K^2,$$

$$(c)\ l_h^*(\beta_{10}, \widetilde{\beta}_2) \xrightarrow{d} \chi_{K_1}^2,$$

where $\widetilde{\beta}_2$ minimizes $l_h^*(\beta_{10}, \beta_2)$ with respect to β_2. If Assumptions 1-6 hold and $\sup_n nh^r < \infty$, then

$$(d)\ P(\beta_0 \in I_{hc}^*) = \alpha + O(n^{-1}).$$

If Assumptions 1–6 hold and $\sup_n n^3h^{2r} < \infty$, then

$$(e)\ P(\beta_0 \in I_{hc}^{b*}) = \alpha + O(n^{-2});\ (f)\ P(\beta_0 \in I_{hc}^{\widehat{b}*}) = \alpha + O(n^{-2});\ (g)\ P(\beta_0 \in I_{hc}^{\widetilde{b}*}) = \alpha + O(n^{-1}h).$$

4   Monte Carlo Simulations

In this section, we describe some Monte Carlo simulation results that are designed to investigate coverage probability accuracy of the SEL confidence regions.

4.1   Experimental Design

We consider a linear median regression model

$$Y_i = X_i'\beta_0 + U_i \text{ for } i = 1, ..., n,$$

where $X_i = (1, X_{2i})'$, $\beta_0 = (\beta_{01}, \beta_{02})'$ is a 2×1 parameter vector whose true value is $\beta_0 = (1, 1)'$, the regressor X_{2i} is generated from a uniform distribution $U[1, 5]$, and error satisfies $P[U_i \leq 0 | X_{2i}] = 0.5$. We consider three different distributions for the error U_i : (i) Student t distribution with 3 degrees of freedom rescaled to have variance 2 (DGP1), (ii) $U_i = 0.25(1 + X_{2i})V_i$, where $V_i \sim N(0, 1)$ (DGP2), and (iii) chi-square distribution with 3 degrees of freedom recentered to have median zero (DGP3). In DGP2, U_i is heteroskedastic and, in DGP3, the distribution is skewed.

DGP1 and DGP2 are the same as the simulation designs of Horowitz (1998) and DGP3 is considered by Chen and Hall (1993).

We consider confidence regions for the parameter vector β_0. We smooth the empirical likelihood using a second-order kernel (i.e., $r = 2$) $K(u) = (3/4)(1-u^2)1\,(|u| \leq 1)$, which is the so-called Bartlett or Epanečnikov kernel. The SEL confidence regions considered are I_{hc}, $\tilde{I}_{hc}^b$, and $\hat{I}_{hc}^{\hat{b}}$ which are defined in (16), (32), and (33) respectively. In simulation results given below, we denote them SEL1, SEL2, and SEL3 respectively. The confidence region corrected with the true Bartlett factor b, i.e. I_{hc}^b defined in (28) is not considered, because it is not of practical interest.

As benchmarks of our simulation experiments, we considered the confidence regions based on the unsmoothed LAD and the SLAD estimators. The former is defined to be

$$I_{LAD} = \left\{ \beta : n \left(\widehat{\beta}_Q - \beta \right)' \widehat{\Lambda}^{-1} \left(\widehat{\beta}_Q - \beta \right) \leq c_\alpha \right\} \qquad (45)$$

where $\widehat{\beta}_Q$ is the LAD estimator of β_0, c_α is the α- quantile of χ_2^2 distribution, and $\widehat{\Lambda}$ is as in (19) with the kernel function given by the second-order kernel $K_1(u) = (15/16)\,(1 - u^2)^2 1\,(|u| \leq 1)$, which was also used by Horowitz (1998). We also considered a confidence region

$$I_{BLAD} = \left\{ \beta : n \left(\widehat{\beta}_Q - \beta \right)' \Lambda^{*-1} \left(\widehat{\beta}_Q - \beta \right) \leq c_\alpha \right\}, \qquad (46)$$

where Λ^* is a bootstrap estimator of Λ_0. The latter is computed by

$$\Lambda^* = \frac{n}{B} \sum_{b=1}^{B} \left(\widehat{\beta}_{Qb}^* - \overline{\widehat{\beta}}_Q^* \right) \left(\widehat{\beta}_{Qb}^* - \overline{\widehat{\beta}}_Q^* \right)', \qquad (47)$$

where $\overline{\widehat{\beta}}_Q^* = (1/B) \sum_{b=1}^{B} \widehat{\beta}_{Qb}^*$ and $\{\widehat{\beta}_{Qb}^* : b = 1, ..., B\}$ are the B bootstrap estimates for β_0, for the B samples (each of size n) drawn from the empirical joint distribution of original data $\{(Y_i, X_i) : i = 1, ..., n\}$. The estimate Λ^* is based on the original idea of Efron (1979, 1982) and is also used by Buchinsky (1995) in the QR models.

On the other hand, the confidence region based on the SLAD estimator is given by

$$I_{SLAD} = \left\{ \beta : n \left(\widetilde{\beta}_S - \beta \right)' \widetilde{\Lambda}^{-1} \left(\widetilde{\beta}_S - \beta \right) \leq \widetilde{c}_\alpha^* \right\}. \qquad (48)$$

Here, $\widetilde{\beta}_S$ is the SLAD estimator of β_0 which solves

$$\min_{b \in B} \widetilde{H}_n(b) = \frac{1}{n} \sum_{i=1}^{n} (Y_i - X_i'b) \left[2\widetilde{G} \left(\frac{Y_i - X_i'b}{h} \right) - 1 \right],$$

17

and its variance is estimated by

$$\widetilde{\Lambda} = D_n(\widetilde{\beta}_S)^{-1} T_n(\widetilde{\beta}_S) D_n(\widetilde{\beta}_S)^{-1},$$

where

$$D_n(b) = 2(nh)^{-1} \sum_{i=1}^{n} X_i X_i' \widetilde{G}^{(1)} \left(\frac{Y_i - X_i'b}{h} \right),$$

$$T_n(b) = n^{-1} \sum_{i=1}^{n} X_i X_i' \left\{ \left[2\widetilde{G} \left(\frac{Y_i - X_i'b}{h} \right) - 1 \right] \right.$$
$$\left. + 2 \left(\frac{Y_i - X_i'b}{h} \right) \widetilde{G}^{(1)} \left(\frac{Y_i - X_i'b}{h} \right) \right\}^2,$$

$\widetilde{G}(\cdot)$ is the integral of a fourth-order kernel given by

$$\widetilde{G}(u) = \begin{cases} 0 & \text{if } u < -1 \\ 0.5 + \frac{105}{64} \left[u - \frac{5}{3}u^3 + \frac{7}{5}u^5 - \frac{3}{7}u^7 \right] & \text{if } |u| \leq 1, \\ 1 & \text{otherwise,} \end{cases} \tag{49}$$

and $\widetilde{G}^{(1)}(u) = d\widetilde{G}(u)/du$. The constant $\widetilde{c}_\alpha^*$ is computed from the following bootstrap procedure: (i) Generate a bootstrap sample $\{(Y_i^*, X_i^*) : i = 1, ..., n\}$ by sampling the original data $\{(Y_i, X_i) : i = 1, ..., n\}$ randomly with replacement. (ii) Using the bootstrap sample, compute the SLAD estimate $\widetilde{\beta}_S^*$ and its variance estimate $\widetilde{\Lambda}^*$ and get $S_n^* = n \left(\widetilde{\beta}_S^* - \widetilde{\beta}_S \right)' \widetilde{\Lambda}^{*-1} \left(\widetilde{\beta}_S^* - \widetilde{\beta}_S \right)$. (iii) Estimate the bootstrap distribution of S_n^* by the empirical distribution that is obtained by repeating steps (i) and (ii) many (B) times. (iv) Take $\widetilde{c}_\alpha^*$ to be the α-quantile of this empirical distribution.

Computing the LAD, SLAD, and SEL confidence regions requires choosing a bandwidth h for each. Existing theories suggest the following rules for choosing h: Hall and Horowitz (1990) show that the bandwidth that minimizes the asymptotic mean squared error of the LAD standard error is of order $n^{-1/2}$, so this rule might be useful for the LAD confidence regions. Also, using the duality of confidence region and hypothesis testing and Assumption 6 of Horowitz (1998), the bandwidth that is compatible with the SLAD confidence region based on the fourth order kernel (49) is of order $n^{-\kappa}$, where $2/9 < \kappa < 1/3$. On the other hand, our Theorems 3 and 4 show that, when the kernel order $r = 2$, the uncorrected and Bartlett corrected SEL confidence regions have coverage errors of order $O(n^{-1})$ and $O(n^{-2})$ if h is of order smaller than $n^{-1/2}$ and $n^{-3/4}$, respectively. However, all of the above rules are justified in an asymptotic sense and hence they provide little practical guidance how to choose h in finite samples. We consider a rule of thumb $h = c_0 n^\gamma$ in our simulations and take $\gamma \in [-1.0, -0.9, ..., -0.1]$. We take $c_0 = 1.0$ in our experiments but, as will be seen, the coverage probabilities of the SEL confidence regions vary little over a wide range of c_0 and γ values.

The number of simulation repetitions used is 40,000 for LAD and SEL confidence regions. This yields simulation standard errors of approximately .0015 and .0010 for the simulated coverage probabilities of nominal 90% and 95% confidence regions respectively. For the BLAD and SLAD confidence regions, however, the number of repetitions is merely 1,000 because of the very long computing times required for simulations with bootstrapping. In this case, the simulation standard errors are approximately .0094 and .0068 for nominal 90% and 95% levels respectively. The number of bootstrap repetitions used is also restricted to $B = 100$ due to heavy computational cost. We consider eight different sample sizes $n \in [15, 20, ..., 50]$.

4.2   Simulation Results

Tables 1-3 summarize results for simulated coverage probabilities of confidence regions. Figure 1 shows coverage errors of SLAD and SEL3 (i.e., Bartlett corrected with $\widehat{b}$) confidence regions for different values of γ values (which determines bandwidth h). The dotted lines surrounding the solid lines are Bonferroni uniform 95% confidence bands for the coverage errors, which were computed by connecting $(1 - 0.05/m)$ pointwise confidence intervals where m $(= 10)$ is the number of points at which the coverage error was estimated. Figure 2 shows coverage errors of SLAD and SEL1 (i.e., no Bartlett correction), and SEL3 confidence regions for varying sample sizes n. Here, we draw the Bonferroni uniform confidence band only for the SLAD case to make the picture less complicated. (The simulation standard errors for SEL1 and SEL3 are virtually negligible because of the large number of repetitions, i.e., 40,000.)

Our simulation results can be summarized as follows:

1. The coverage probabilities of the LAD confidence regions are relatively poor and very sensitive to the choice of bandwidth. For example, in DGP1 and $n = 35$ case, the coverage probabilities of the nominal 95% LAD confidence region are .920 and .204 for $\gamma = -0.1$ and $\gamma = -0.9$ respectively. On the other hand, the coverage probabilities of the BLAD confidence regions are relatively very good and stable across different designs.

2. Both SLAD and SEL confidence regions are robust to the choice of bandwidth. However, Figure 1 shows some evidence that the SEL3 confidence region is less sensitive to bandwidth than the SLAD confidence region especially for DGP1 and DGP2 and for $n \geq 35$.

3. The SEL confidence regions with no Bartlett correction (SEL1) or Bartlett corrected with b^* (SEL2) perform similarly, though SEL2 is slightly better than SEL1 in almost all cases. This confirms the theory in Theorem 3 and 4(c), which shows that the coverage errors are $O(n^{-1})$ and $O(n^{-1}h)$ for SEL1 and SEL2 respectively.

4. The SEL confidence regions Bartlett corrected with $\widehat{b}$ (i.e., SEL3) dominate the other confidence regions in most cases. For example, for $n = 50$, the SEL3 coverage error is virtually zero (up to simulation errors) in almost all cases.

5. The SLAD confidence regions perform fairly well especially in small samples ($n \leq 20$) and, in some case, out-perform SEL1 and SEL2.

6. The effect of increasing the sample size is to reduce coverage errors for almost all confidence regions.

7. Figure 2 shows that, as the sample size increases, SEL3 coverage errors decrease to zero at a faster speed than the SLAD coverage errors. This confirms our theory because the SLAD confidence region has coverage errors of order $O(n^{-a})$ for $a < 1$, whereas the SEL3 confidence region has coverage errors of order $O(n^{-2})$.

8. There is not much difference in relative performance of confidence regions under different DGP's.

9. The results for nominal 90% and 95% confidence regions are similar.

10. The bandwidth that gives the best overall performance for the SEL3 confidence regions is $h = n^\gamma$ for $\gamma = -0.8$. This result is compatible with Theorem 4 which requires $-1 < \gamma < -0.75$ for Bartlett correction. Therefore, we recommend to use the latter rule of thumb in practical applications, though the results seem to be very robust to the choice of γ.

5   Conclusion

In this paper, we have used smoothed empirical likelihood methods to obtain asymptotically valid point estimators and confidence regions about the parameters of uncensored and censored quantile regression models that allow for unknown form of heteroskedasticity. We further have shown that, if simple corrections are made, the smoothed empirical likelihood confidence regions can achieve higher order refinements, which are better than the refinements that might be obtained through the (smoothed) bootstrap approach. Extensions to other econometric models with discontinuous estimating equations and a rigorous investigation of higher-order properties of the confidence regions in the presence of nuisance parameters in smoothed models would be an interesting future topics.

6 Appendix

Lemma 1 *Under Assumptions 1-5(b) and 6(a) , the following relations hold as $n \to \infty$:*

$$
\begin{aligned}
(a)\ EZ_i(\beta_0) &= (-h)^r (r!)^{-1} C_K E\left[X_i f^{(r-1)}(0|X_i)\right] + o(h^r),\\
(b)\ EZ_i(\beta_0)Z_i(\beta_0)' &= q(1-q)S_0 + o(1),\\
(c)\ E\frac{\partial Z_i(\beta_0)}{\partial \beta'} &= D_0 + o(1),
\end{aligned}
$$

where $S_0 = E[X_iX_i']$ and $D_0 = E\left[f(0|X_i)X_iX_i'\right].$

Proof of Lemma 1: By a change of variables, we have

$$
EZ_i(\beta_0) = E\left\{X_i \int [F(-hu|X_i) - F(0|X_i)]\,K(u)du\right\}.
$$

Then, apply a Taylor expansion to establish part (a). Similarly, parts (b) and (c) hold by noting that

$$
\begin{aligned}
EZ_i(\beta_0)Z_i(\beta_0) &= q(1-q)E\left[X_iX_i'\right]\\
&\quad +2E\left\{X_iX_i'\int [F(-hu|X_i) - F(0|X_i)][G(u) - q]\,K(u)du\right\}
\end{aligned}
$$

and

$$
E\frac{\partial Z_i(\beta_0)}{\partial \beta'} = E[f(0|X_i)X_iX_i'] + E\left\{X_iX_i'\int [f(-hu|X_i) - f(0|X_i)]\,K(u)du\right\}.
$$

$\blacksquare$

Lemma 2 *Suppose Assumptions 1-5(b) and 6(a) hold. Then, with probability 1 as $n \to \infty$,*

$$
\begin{aligned}
(a)\frac{1}{n}\sum_{i=1}^{n} Z_i(\beta) &= O(d_n),\\
(b)\frac{1}{n}\sum_{i=1}^{n} Z_i(\beta)Z_i(\beta)' &= q(1-q)S_0 + o(1),\\
(c)\,t(\beta) &= O(d_n)
\end{aligned}
$$

uniformly in $\beta \in B_n \equiv \{\beta : \|\beta - \beta_0\| \leq d_n\}$, where $t(\beta)$ satisfies (7), $d_n = n^{-1/3-\delta}$ and $0 < \delta < 1/6$.

Proof of Lemma 2: By a Taylor expansion,

$$\frac{1}{n}\sum_{i=1}^{n} Z_i(\beta) = \frac{1}{n}\sum_{i=1}^{n} \{Z_i(\beta_0) - EZ_i(\beta_0)\} + EZ_i(\beta_0) + R_n(\beta), \qquad \text{(A.1)}$$

where

$$R_n(\beta) = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial Z(\beta^*)}{\partial \beta'}(\beta - \beta_0)$$

and β^* lies between β and β_0. Using Cauchy-Schwartz inequality, triangle inequality and an argument similar to the proof of Lemma 1, we have

$$\sup_{\beta \in B_n} \|R_n(\beta)\|$$

$$\leq d_n \cdot \sup_{\beta \in B_n} \left\{ \left\| \frac{1}{n}\sum_{i=1}^{n} \left(\frac{\partial Z(\beta^*)}{\partial \beta'} - E\frac{\partial Z(\beta^*)}{\partial \beta'} \right) \right\| + \left\| E\frac{\partial Z(\beta^*)}{\partial \beta'} \right\| \right\}$$

$$= O(d_n) \text{ a.s.} \qquad \text{(A.2)}$$

Therefore, using (A.1), (A.2), law of iterated logarithm, Lemma 1(a), and Assumption 6, we have

$$\sup_{\beta \in B_n} \left\| \frac{1}{n}\sum_{i=1}^{n} Z_i(\beta) \right\| = O\left(n^{-1/2} (\log\log n)^{1/2} \right) + O(h^r) + O(d_n)$$

$$= O(d_n) \text{ a.s.}$$

as desired. The proof of part (b) is similar to part (a).

To prove part (c), fix β such that $\|\beta - \beta_0\| \leq d_n$. Write $t \equiv t(\beta) = \rho\alpha$, where $\rho \geq 0$ and $\|\alpha\| = 1$. We have

$$0 = \left\| \frac{1}{n}\sum \frac{Z_i(\beta)}{1 + t'Z_i(\beta)} \right\| = \left\| \frac{1}{n}\sum \frac{Z_i(\beta)}{1 + \rho\alpha'Z_i(\beta)} \right\|$$

$$\geq \left| \frac{1}{n}\sum \frac{\alpha'Z_i(\beta)}{1 + \rho\alpha'Z_i(\beta)} \right|$$

$$= \left| \frac{1}{n}\sum \left\{ \rho\frac{\alpha'Z_i(\beta)Z_i(\beta)\alpha}{1 + \rho\alpha'Z_i(\beta)} - \alpha'Z_i(\beta) \right\} \right|$$

$$\geq \left| \frac{1}{n}\sum \rho\frac{\alpha'Z_i(\beta)Z_i(\beta)\alpha}{1 + \rho\alpha'Z_i(\beta)} \right| - \left| \frac{1}{n}\sum \alpha'Z_i(\beta) \right|$$

$$\geq \frac{\rho}{1 + \rho\max_i \|Z_i(\beta)\|} \cdot \alpha'\left(\frac{1}{n}\sum Z_i(\beta)Z_i(\beta) \right)\alpha - \left\| \frac{1}{n}\sum Z_i(\beta) \right\|,$$

where the last inequality follows from the positivity of $1 + \rho\alpha'Z_i(\beta)$ (which holds from $p_i = n^{-1}(1 + t'Z_i(\beta))^{-1} \geq 0$). Rearranging terms, we have

$$\frac{\rho}{1 + \rho\max_i \|Z_i(\beta)\|} \cdot \alpha'\left(\frac{1}{n}\sum Z_i(\beta)Z_i(\beta) \right)\alpha \leq \left\| \frac{1}{n}\sum Z_i(\beta) \right\|. \qquad \text{(A.3)}$$

Observe that $\max_i \|Z_i(\beta)\| = O(1)$ uniformly in $\beta \in B_n$. Therefore, (A.3) and the results of parts (a) and (b) yield

$$\rho \leq \left\{ q^{-1}(1-q)^{-1}\lambda_{\min}^{-1}(S) + o(1) \right\} \cdot O(d_n) \text{ a.s.} \tag{A.4}$$

uniformly in $\beta \in B_n$, where $\lambda_{\min}(\cdot)$ denotes the minimum eigenvalue of $\cdot$. (A.4) now establishes part (c) since $\|\lambda\| = \rho$. $\blacksquare$

Lemma 3 *Suppose Assumptions 1-5(b) and 6(a) hold. Then, with probability 1 as $n \to \infty$, (a) there exists a $K \times 1$ vector $\widehat{\beta}_E \in int(B)$ such that $l_h(\beta)$ attains its minimum value at $\widehat{\beta}_E$ and (b) $\widehat{\beta}_E$ and $\widehat{t} = t(\widehat{\beta}_E)$ satisfy*

$$Q_n(\widehat{\beta}_E, \widehat{t}) = 0$$

where

$$
\begin{aligned}
Q_n(\beta, t) &= (Q_{1n}(\beta, t)', \ Q_{1n}(\beta, t)')' \\
Q_{1n}(\beta, t) &= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{1 + t'Z_i(\beta)} Z_i(\beta) = 0, \\
Q_{2n}(\beta, t) &= \frac{1}{n}\sum_{i=1}^{n} \frac{1}{1 + t'Z_i(\beta)} \frac{\partial Z_i(\beta)}{\partial \beta'} t = 0.
\end{aligned}
$$

Proof of Lemma 3: This lemma is a slight modification of Lemma 1 of Qin and Lawless (1994) and can be proved using a similar argument to theirs and Lemma 2 above. $\blacksquare$

Proof of Theorem 1: By Lemma 1 and WLLN, we have

$$
\begin{aligned}
\frac{\partial Q_{1n}(\beta_0, 0)}{\partial \beta'} &= \frac{\partial Q_{2n}(\beta_0, 0)}{\partial t'} = \frac{1}{n}\sum_{i=1}^{n} \frac{\partial Z_i(\beta_0)}{\partial \beta'} \xrightarrow{p} D_0, \\
\frac{\partial Q_{1n}(\beta_0, 0)}{\partial t'} &= -\frac{1}{n}\sum_{i=1}^{n} Z_i(\beta_0)Z_i(\beta_0) \xrightarrow{p} -q(1-q)S_0, \text{ and} \tag{A.5} \\
\frac{\partial Q_{2n}(\beta_0, 0)}{\partial \beta'} &= 0.
\end{aligned}
$$

Below, we establish that

$$
\begin{aligned}
\sqrt{n}Q_{1n}(\beta_0, 0) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n} \left[1\left(U_i \leq 0\right) - q\right] X_i + o_p(1) \tag{A.6} \\
&= O_p(1).
\end{aligned}
$$

23

Expanding $Q_n(\widehat{\beta}_E, \widehat{t})$ at $(\beta_0, 0)$, by the conditions of Lemma 3 and using (A.5) and (A.6), we have

$$\sqrt{n}\left(\widehat{\beta}_E - \beta_0\right) = D_0^{-1}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[1\left(U_i \leq 0\right) - q\right]X_i + o_p(1),$$

which, in turn, is the Bahadur representation of the quantile regression estimator.

We now establish (A.6). Letting $G_{ni} \equiv \left[G\left(-U_i/h\right) - 1\left(U_i \leq 0\right)\right]$ and rearranging terms, we have

$$
\begin{aligned}
\sqrt{n}Q_{1n}(\beta_0, 0) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[G\left(\frac{-U_i}{h}\right) - q\right]X_i \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[1\left(U_i \leq 0\right) - q\right]X_i + \quad\quad\quad\quad\text{(A.7)} \\
&\quad \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[G_{ni}X_i - EG_{ni}X_i\right] + \sqrt{n}EG_{ni}X_i.
\end{aligned}
$$

The second term on the right hand side of (A.7) is $O_p(h^{1/2})$ and hence $o_p(1)$ since, for each $\varepsilon > 0$

$$
\begin{aligned}
P\left(\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left[G_{ni}X_i - EG_{ni}X_i\right]\right\| > \varepsilon\right) &\leq \varepsilon^{-2}E\left[G\left(\frac{-U_i}{h}\right) - 1\left(U_i \leq 0\right)\right]^2 \|X_i\|^2 \\
&\leq C \cdot P\left(-h \leq U \leq h\right) = O(h) \to 0.
\end{aligned}
$$

Also, the last term in (A.7) is $o(1)$ using Assumption 6(a) since

$$\sqrt{n}EG_{ni}X_i = \sqrt{n}EZ_i(\beta_0) = O(n^{1/2}h^r) \to 0$$

as desired. ∎

Proof of Theorem 2: Let $\lambda \equiv \lambda(\beta_0)$ denote the solution of the equation

$$\frac{1}{n}\sum\frac{W_i}{1 + \lambda'W_i} = 0. \quad\quad\quad\quad\text{(A.8)}$$

Then, we have

$$\lambda = O_p(n^{-1/2} + h^r). \quad\quad\quad\quad\text{(A.9)}$$

using the same arguments as in the proof of Lemma 2(c) after noting that we now have $n^{-1}\sum W_iW_i' \overset{p}{\to} EW_iW_i' = I_K$ by a WLLN, $n^{-1}\sum W_i = O_p(n^{-1/2} + h^r)$, and $\max_i \|W_i\| = O_p(1)$ by Assumption 3.

24

Next we develop a Taylor expansion for λ and $l_h(\beta_0)$. By (A.8) , we have

$$
\begin{aligned}
0 &= \frac{1}{n}\sum \frac{W_i}{1+\lambda'W_i} \\
&= \frac{1}{n}\sum W_i\left\{1-(\lambda'W_i)+(\lambda'W_i)^2-(\lambda'W_i)^3+(\lambda'W_i)^4-\cdots\right\} \\
&= \frac{1}{n}\sum W_i - \left(\frac{1}{n}\sum W_iW_i'\right)\lambda + \frac{1}{n}\sum(\lambda'W_i)^2W_i - \frac{1}{n}\sum(\lambda'W_i)^3W_i + \frac{1}{n}\sum(\lambda'W_i)^4W_i - \cdots.
\end{aligned}
\tag{A.10}
$$

By Lemma 1(a), we have $\alpha^j = O(h^r)$. Also, observe that $\overline{A}^j = A^j + \alpha^j = O_p(n^{-1/2} + h^r)$, $A^{jk} = O_p(n^{-1/2})$, and $\overline{A}^{j_1\cdots j_k} = O_p(1)$ for $k \geq 3$. Solving for λ and then recursive substitutions in equation (A.10) give, for each $L \geq 1$,

$$
\begin{aligned}
\lambda^j &= \overline{A}^j - A^{jk}\overline{A}^k + \overline{A}^{jkl}\overline{A}^k\overline{A}^l + A^{jk}A^{kl}\overline{A}^l - 2\overline{A}^{jkl}A^{kp}\overline{A}^p\overline{A}^l - \overline{A}^{klm}A^{jk}\overline{A}^l\overline{A}^m \\
&\quad + 2\overline{A}^{jkl}\overline{A}^{lmo}\overline{A}^k\overline{A}^m\overline{A}^o - \overline{A}^{jklm}\overline{A}^k\overline{A}^l\overline{A}^m + \sum_{l=4}^{L} R_{1l} + O_p((n^{-1/2}+h^r)^{L+1}),
\end{aligned}
$$

where R_{il} denotes a sum of the products of terms of the form $\overline{A}^j$, A^{jk}, and $\overline{A}^{j_1\cdots j_m}$ for $m \in \{3,...,l+1\}$ so that $R_{il} = O_p((n^{-1/2}+h^r)^l)$ for $i=1,2$.

Similarly, we have

$$
\begin{aligned}
\frac{1}{n}l_h(\beta_0) &= \frac{2}{n}\sum_{i=1}^{n}\log(1+\lambda'W_i) \\
&= 2\sum_{k=2}^{L+1}(-1)^k\frac{k-1}{k}\left\{\sum_{i=1}^{n}\frac{(\lambda'W_i)^k}{n}\right\} + O_p\left((n^{-1/2}+h^r)^{L+2}\right) \\
&= \overline{A}^j\overline{A}^j + \frac{2}{3}\overline{A}^{jkl}\overline{A}^j\overline{A}^k\overline{A}^l - A^{jk}\overline{A}^j\overline{A}^k \\
&\quad + \overline{A}^{jkl}\overline{A}^{jmo}\overline{A}^k\overline{A}^l\overline{A}^m\overline{A}^o - \frac{1}{2}\overline{A}^{jklm}\overline{A}^j\overline{A}^k\overline{A}^l\overline{A}^m + A^{jk}A^{jl}\overline{A}^k\overline{A}^l - 2\overline{A}^{jkl}A^{kp}\overline{A}^j\overline{A}^p\overline{A}^l \\
&\quad + 8\overline{A}^{jklm}\overline{A}^{jpq}\overline{A}^k\overline{A}^l\overline{A}^m\overline{A}^p\overline{A}^q - 8\overline{A}^{jkl}\overline{A}^{jpq}\overline{A}^{qmo}\overline{A}^p\overline{A}^m\overline{A}^o\overline{A}^k\overline{A}^l - \frac{8}{5}\overline{A}^{jklmo}\overline{A}^j\overline{A}^k\overline{A}^l\overline{A}^m\overline{A}^o \\
&\quad + 12\overline{A}^{jkl}\overline{A}^{lmo}A^{jp}\overline{A}^k\overline{A}^m\overline{A}^o\overline{A}^p + 3\overline{A}^{jkl}\overline{A}^{mop}A^{jm}\overline{A}^o\overline{A}^p\overline{A}^k\overline{A}^l + A^{jk}A^{jm}A^{kl}\overline{A}^m\overline{A}^l \\
&\quad - 4\overline{A}^{jkl}A^{jm}A^{kp}\overline{A}^m\overline{A}^p\overline{A}^l - 4\overline{A}^{jkl}A^{jm}A^{mo}\overline{A}^o\overline{A}^k\overline{A}^l - 6\overline{A}^{jklm}A^{jp}\overline{A}^p\overline{A}^k\overline{A}^l\overline{A}^m \\
&\quad + \sum_{l=6}^{L+1} R_{2l} + O_p\left((n^{-1/2}+h^r)^{L+2}\right).
\end{aligned}
\tag{A.11}
$$

Therefore, for any $k > 1$,

$$
l_h(\beta_0) = n\overline{A}^j\overline{A}^j + O_p\left(n\left(n^{-1/2}+h^r\right)^k\right)
\tag{A.12}
$$

$$= \left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_i - EZ_i) + n^{1/2}EZ_i\right]' V_n^{-1} \left[\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(Z_i - EZ_i) + n^{1/2}EZ_i\right]$$
$$+ O_p\left(n\left(n^{-1/2} + h^r\right)^k\right).$$

Since $V_n - Var(Z_i) \to 0$, $[Var(Z_i)]^{-1/2} \cdot n^{-1/2}\sum(Z_i - EZ_i) \xrightarrow{d} N(0, I_K)$, and $nh^{2r} \to 0$ as $n \to \infty$, $l_h(\beta_0)$ has an asymptotic central chi-square distribution with K degrees of freedom if $n^{1/2}EZ_i \to 0$. The latter holds by Lemma 1(a) and Assumption 6. ∎

Proof of Corollary 1: Let $W_i(\beta_{10}, \widetilde{\beta}_2) = W_i(\widetilde{\beta}_2)$ and $\lambda(\beta_{10}, \widetilde{\beta}_2) = \widetilde{\lambda}$. Lemma 3 implies that $\widetilde{\beta}_2$ and $\widetilde{\lambda}$ satisfy $H_n(\widetilde{\beta}_2, \widetilde{\lambda}) = 0$, where $H_n(\beta_2, \lambda) = (H_{1n}(\beta_2, \lambda)', \; H_{2n}(\beta_2, \lambda)')'$,

$$H_{1n}(\beta_2, \lambda) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{1 + \lambda' W_i(\beta_2)}W_i(\beta_2) = 0, \qquad (A.13)$$

$$H_{2n}(\beta_2, \lambda) = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{1 + \lambda' W_i(\beta_2)}\frac{\partial W_i(\beta_2)}{\partial \beta_2'}\lambda = 0. \qquad (A.14)$$

A Taylor expansion of $H_n(\widetilde{\beta}_2, \widetilde{\lambda})$ around $H_n(\beta_{20}, 0)$ yields

$$\widetilde{\lambda} = \left\{I_K - D_{20}(D_{20}'D_{20})^{-1}D_{20}'\right\}H_{1n}(\beta_{20}, 0) + o_p(n^{-1} + h^r), \quad (A.15)$$
$$(\widetilde{\beta}_2 - \beta_{20}) = -(D_{20}'D_{20})^{-1}D_{20}'H_{1n}(\beta_{20}, 0) + o_p(n^{-1} + h^r), \qquad (A.16)$$

where $D_{20} = V_n^{-1/2}E\left[f(0|X_i)X_iX_{i2}'\right]$ is a $K \times K_2$ matrix. Note that $\sqrt{n}H_{1n}(\beta_{20}, 0) \xrightarrow{d} N(0, I_K)$. Therefore, we have

$$\frac{1}{n}l_h(\beta_{10}, \widetilde{\beta}_2) = \frac{2}{n}\sum\log\left[1 + \widetilde{\lambda}'W_i(\widetilde{\beta}_2)\right]$$
$$= \widetilde{\lambda}'\widetilde{\lambda} + o_p(n^{-1} + h^r) \xrightarrow{d} \chi^2_{K_1},$$

as desired, where the second equality holds by a two-term Taylor expansion. ∎

Let

$$\overline{Q} = \left(A^1, ..., A^K, A^{11}, ..., A^{KK}, ..., A^{11\cdots1}, ..., A^{KK\cdots K}\right)' \equiv \frac{1}{n}\sum_{i=1}^{n}Q_i \qquad (A.17)$$

denote a vector of all distinct first $L + 1$ order multivariate centered moments of $W_i = V_n^{-1/2}Z_i$. Note that Q_i consists of elements of the form

$$(G(-U_i/h) - q)^{|\nu|}W_i^{\nu_1}\cdots W_i^{\nu_k} \text{ for } 1 \le k \le L + 1, \qquad (A.18)$$

where $|\nu| = \nu_1 + \cdots + \nu_k$. We first establish the following modified version of the Cramér's condition for the Edgeworth expansion, which will be needed later:

26

Lemma 4 *Let $t \in \mathbb{R}^{\dim(Q)}$ be a vector and $I(t, h) = E\{\exp[it'Q]\}$, where $Q \,(= Q_i)$ is given by (A.17) and $i = (-1)^{1/2}$. Under Assumptions 1-6, we have: for each $\varepsilon > 0$, there exists some $C > 0$ such that*

$$\sup_{\|t\| > \varepsilon} |I(t, h)| < 1 - Ch$$

for all sufficiently small h.

Proof of Lemma 4: The proof of Lemma 4 is analogous to those of Horowitz (1998, lemma 9) and Hall (1992, lemma 5.6). We just briefly sketch the main idea.

Note that the terms such as (A.18) can be expanded to be polynomials in $[G(-U_i/h)]^r$ for $0 \le r \le L+1$ with coefficients given by (not necessarily distinct) elements of X_i. Therefore, by collecting terms with the same polynomial order, we may write

$$
\begin{aligned}
I(t, h) &= E\{\exp[it'Q]\} \\
&= E\left\{\exp\left[i\sum_{r=0}^{L+1}[G(-U/h)]^r \tau_r(t)'g_r(X_i)\right]\right\},
\end{aligned}
$$

where $g_r(X)$ is a vector of the products of elements of X that correspond to the r-th order polynomial $[G(-U/h)]^r$ in the expansion of $t'Q$ and $\tau_r(t)$ denotes the corresponding sub-vector of $t \in \mathbb{R}^{\dim(Q)}$.

Since G satisfies $G(v) = 1$ if $v \ge 1$ and $G(v) = 0$ if $v \le -1$, we can write

$$
\begin{aligned}
I(t, h) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp\left[i\sum_{r=0}^{L+1}[G(-u/h)]^r \tau_r(t)'g_r(x)\right] f(u|x)dudP(x) \\
&= I_1(t, h) + I_2(t, h),
\end{aligned}
$$

where

$$I_1(t, h) = E\left\{[1 - F(h|X)] + F(-h|X)\exp\left[i\sum_{r=0}^{L+1}\tau_r(t)'g_r(x)\right]\right\}$$

and

$$I_2(t, h) = \int_{-\infty}^{\infty}\int_{-h}^{h}\exp\left[i\sum_{r=0}^{L+1}[G(-u/h)]^r \tau_r(t)'g_r(x)\right] f(u|x)dudP(x).$$

First, for h sufficiently small, we have

$$
\begin{aligned}
|I_1(t, h)| &\le E\{1 - F(h|X) + F(-h|X)\} \\
&\le 1 - Ef(0|X)h \qquad\qquad\qquad\qquad (A.19)
\end{aligned}
$$

by a two-term Taylor expansion using Assumption 4.

27

Next, given $\varepsilon > 0$, choose h so small that $\int_{-\infty}^{\infty} \int_{-1}^{1} |f(hu|x) - f(0|x)| \, du \, dP(x) \leq 2\varepsilon E f(0|X)$. Take $\eta > 0$ and $\gamma_1 < 1$ such that $\int_{\|x\| \leq \eta} f(0|x) dP(x) = \gamma_1 E f(0|X)$. Then, by a change of variables and triangle inequality, we have

$$|I_2(t, h)| \leq (2\varepsilon + \gamma_1) h E f(0|X) + \tag{A.20}$$
$$+ h \int_{\|x\| > \eta} \Psi(t, x) f(0|x) dP(x),$$

where

$$\Psi(t, x) = \int_{-1}^{1} \exp \left[i \sum_{r=0}^{L+1} [G(u)]^r \tau_r(t)' g_r(x) \right] du$$

Let $\xi = \|t\|$ and fix $t/\|t\|$ (and hence $\tau(t)/\|t\|$ trivially). Define

$$f(u, x) = \sum_{r=1}^{L+1} [G(u)]^r \tau_r(t)' g_r(x) / \|t\|.$$

Let $\{(a_{m-1}, a_m) : m = 1, ..., L+1\}$ be the partition of $[-1, 1]$ that satisfies Assumption 5(c). Then,

$$
\begin{aligned}
C_1 &\equiv \sup_{\|x\| > \eta} \sup_{\|t\| > \varepsilon} |\Psi(t, x)| \\
&\leq \sup_{\|x\| > \eta} \sup_{\xi > \varepsilon} \left| \int_{-1}^{1} \exp \left[i\, \xi f(u, x) \right] du \right|, \\
&= \sup_{\|x\| > \eta} \sup_{\xi > \varepsilon} \left| \sum_{m=1}^{M} \int_{a_{m-1}}^{a_m} \exp \left[i\, \xi f(u, x) \right] du \right| < 1 \tag{A.21}
\end{aligned}
$$

where the first inequality uses $|e^{itz}| \leq 1$ and the last inequality holds by an argument similar to Horowitz (1998, pp.1346-1347). Now, by combining (A.19), (A.20), and (A.21), we have

$$
\begin{aligned}
\sup_{\|t\| > \varepsilon} |I(t, h)| &\leq 1 - \{1 - 2\varepsilon - [\gamma_1 + (1 - \gamma_1) C_1]\} E f(0|X) h \\
&\equiv 1 - Ch
\end{aligned}
$$

for all $h > 0$ sufficiently small and $\varepsilon > 0$. This completes the proof of Lemma 4. $\blacksquare$

Define $\Sigma = Var\left(n^{1/2} \overline{Q}\right)$ and $d = \dim(Q)$. Let $r = (r_1, ..., r_d) \in \mathbb{R}^d$ denote a vector of nonnegative integers and $|r| = r_1 + \cdots + r_d$. Let $Z^r \equiv (Z^1)^{r_1} \cdots (Z^d)^{r_d}$ for $Z \in \mathbb{R}^d$ and $r! = r_1! \cdots r_d!$. Put $t = (t_1, ..., t_d)' \in \mathbb{R}^d$ and define the polynomial $P_k(t)$ by the following formal expansion:

$$\exp \left[u^{-2} \sum_{l=0}^{\infty} (-1)^l (l+1)^{-1} \left\{ \sum_{|r|=2}^{\infty} (r!)^{-1} (it)^r (EQ^r) u^r \right\}^{l+1} \right]$$

$$= \exp\left(-\frac{1}{2}t'\Sigma t\right)\left\{1 + \sum_{k=1}^{\infty} P_k(t)u^k\right\},$$

where u is a real number. Let $q_k(x)\phi_{0,\Sigma}(x)$ be the density of the finite signed measure whose Fourier-Stieltjes transform is $\exp(-t'\Sigma t/2)P_k(t)$, i.e.

$$\int \exp(it'x)q_k(x)\phi_{0,\Sigma}(x)dx = \exp\left(-\frac{1}{2}t'\Sigma t\right)P_k(t).$$

Let $\partial\mathcal{A}$ denote a boundary of a set $\mathcal{A}$ and $(\partial\mathcal{A})^\varepsilon$ for the set of all points distant at most ε from $\partial\mathcal{A}$. The formal Edgeworth expansion for the distribution of $n^{1/2}\overline{Q}$ is given by the following lemma:

Lemma 5 *Suppose Assumptions 1-6 hold. Let A denote a class of Borel sets $\mathcal{A} \subseteq \mathbb{R}^d$ that satisfy*

$$\sup_{A\in\mathcal{A}} \int_{(\partial\mathcal{A})^\varepsilon} \exp\left(-\frac{1}{2}\|x\|^2\right)dx = O(\varepsilon)$$

as $\varepsilon \downarrow 0$. Then, for each integer $m \geq 1$, we have

$$\sup_{A\in\mathcal{A}}\left|P\left(n^{1/2}\overline{Q} \in A\right) - \int_A \sum_{k=0}^{m} n^{-k/2}q_k(x)\phi_{0,\Sigma}(x)dx\right| = O(n^{-(m+1)/2}).$$

Proof of Lemma 5: Lemma 2 can be proved using an argument very similar to the proof of Theorem 5.8 of Hall (1992), which in turn relies on Hall (1992)'s Lemmas 5.6 and 5.7. We just note that Hall's lemma 5.6 corresponds to our Lemma 1 above and the result analogous to Hall's Lemma 5.7 can be proved using a technique which is similar to (but substantially simpler than) the Hall's method after replacing the norming constant $(nh)^{1/2}$ by $n^{1/2}$. ∎

Proof of Theorem 3: We first derive the signed root of $l_h(\beta_0)$ in (A.11), which is a K-dimensional vector $n^{1/2}S_{0L} = n^{1/2}(S_{0L}^1, ..., S_{0L}^K)'$ such that $l_h(\beta_0) = (n^{1/2}S_{0L})'(n^{1/2}S_{0L})$. Consider the expansion

$$S_{0L} = \sum_{l=1}^{L} T_l + U_{1L} \equiv S_L + U_{1L},$$

where $T_l = O_p\left((n^{-1/2} + h^r)^l\right)$ and $U_{1L} = O_p\left((n^{-1/2} + h^r)^{L+1}\right)$. Some calculations yield that we have

$$T_1^j = \overline{A}^j,$$

$$T_2^j = \frac{1}{3}\overline{A}^{jkl}\overline{A}^k\overline{A}^l - \frac{1}{2}A^{jk}\overline{A}^k,$$

29

$$T_3^j = \frac{3}{8}A^{jm}A^{km}\overline{A}^k + \frac{4}{9}\overline{A}^{jkn}\overline{A}^{lmn}\overline{A}^m\overline{A}^k\overline{A}^l - \frac{1}{4}\overline{A}^{jklm}\overline{A}^m\overline{A}^k\overline{A}^l$$
$$-\frac{5}{12}\overline{A}^{jkm}A^{lm}\overline{A}^k\overline{A}^l - \frac{5}{12}\overline{A}^{klm}A^{jm}\overline{A}^k\overline{A}^l,$$

$$T_4^j = \frac{11}{16}A^{rk}A^{rj}A^{kl}\overline{A}^l$$
$$-\frac{53}{48}\overline{A}^{rkj}A^{rm}A^{kp}\overline{A}^m\overline{A}^p - \frac{53}{48}\overline{A}^{rkl}A^{rj}A^{kp}\overline{A}^p\overline{A}^l - \frac{7}{6}\overline{A}^{rkj}A^{rm}A^{mo}\overline{A}^o\overline{A}^k - \frac{7}{6}\overline{A}^{rkl}A^{rm}A^{mj}\overline{A}^k\overline{A}^l$$
$$+\frac{229}{108}\overline{A}^{rjl}\overline{A}^{lmo}A^{rp}\overline{A}^m\overline{A}^o\overline{A}^p + \frac{229}{108}\overline{A}^{rkl}\overline{A}^{ljo}A^{rp}\overline{A}^k\overline{A}^o\overline{A}^p + \frac{229}{108}\overline{A}^{rkl}\overline{A}^{lmo}A^{rj}\overline{A}^k\overline{A}^m\overline{A}^o$$
$$+\frac{59}{36}\overline{A}^{rjl}\overline{A}^{mop}A^{rm}\overline{A}^o\overline{A}^p\overline{A}^l - \frac{25}{16}\overline{A}^{rjlm}A^{rp}\overline{A}^p\overline{A}^l\overline{A}^m - \frac{25}{16}\overline{A}^{rklm}A^{rj}\overline{A}^k\overline{A}^l\overline{A}^m$$
$$+\frac{49}{24}\overline{A}^{rjlm}\overline{A}^{rpq}\overline{A}^l\overline{A}^m\overline{A}^p\overline{A}^q + \frac{49}{24}\overline{A}^{rklm}\overline{A}^{rjq}\overline{A}^k\overline{A}^l\overline{A}^m\overline{A}^q - \frac{56}{27}\overline{A}^{rjl}\overline{A}^{rpq}\overline{A}^{qmo}\overline{A}^p\overline{A}^m\overline{A}^o\overline{A}^l$$
$$-\frac{56}{27}\overline{A}^{rkl}\overline{A}^{rjq}\overline{A}^{qmo}\overline{A}^m\overline{A}^o\overline{A}^k\overline{A}^l - \frac{4}{5}\overline{A}^{jklmo}\overline{A}^k\overline{A}^l\overline{A}^m\overline{A}^o.$$

Also, by choosing L sufficiently large, we can ensure that
$$P\left(\|U_{1L}\| > n^{-5/2}\right) = O(n^{-2}).$$
Hence, for $c > 0$, we have
$$P\left(l_h(\beta_0) \leq c\right) = P\left[n^{1/2}\|S_L + U_{1L}\| \leq c^{1/2}\right]$$
and so
$$\max_{+,-}\left|P\left(l_h(\beta_0) \leq c\right) - P\left(n^{1/2}\|S_L\| \leq c^{1/2} \pm n^{-2}\right)\right| = O(n^{-2}). \qquad \text{(A.22)}$$

We now develop an Edgeworth expansion for the distribution of $S_{nL} \equiv n^{1/2}S_L$. We first derive the (multivariate) cumulants of S_{nL}. By very tedious and lengthy calculations, we may show that the cumulants satisfy the following results:
$$cum(S_{nL}^j) = n^{1/2}\alpha^j - \frac{1}{n^{1/2}}\left(\frac{1}{6}\alpha^{jkk}\right) + O(n^{-1/2}h^r + n^{-3/2}),$$

$$cum(S_{nL}^i, S_{nL}^j) = \delta^{ij} + \frac{1}{3}\alpha^{ijk}\alpha^k$$
$$+\alpha^i\alpha^j - \frac{9}{24}\alpha^{jkmm}\alpha^i\alpha^k - \frac{9}{24}\alpha^{ikmm}\alpha^j\alpha^k - \frac{7}{12}\alpha^{ijkm}\alpha^k\alpha^m$$
$$-\frac{1}{18}\alpha^{ikl}\alpha^{jmm}\alpha^k\alpha^l - \frac{1}{18}\alpha^{ikk}\alpha^{jml}\alpha^m\alpha^l + \frac{13}{18}\alpha^{ikl}\alpha^{jkl}\alpha^k\alpha^m$$
$$+\frac{1}{36}\alpha^{jkl}\alpha^{mml}\alpha^i\alpha^k + \frac{1}{36}\alpha^{ikl}\alpha^{mml}\alpha^j\alpha^k + \frac{1}{18}\alpha^{jkl}\alpha^{kml}\alpha^i\alpha^m$$
$$+\frac{1}{18}\alpha^{ikl}\alpha^{kml}\alpha^j\alpha^m + \frac{1}{18}\alpha^{ijk}\alpha^{klm}\alpha^l\alpha^m$$
$$+\frac{1}{n}\left(\frac{1}{2}\alpha^{ijkk} - \frac{1}{3}\alpha^{ikm}\alpha^{jkm} - \frac{1}{36}\alpha^{ijm}\alpha^{mkk}\right) + O(n^{-1}h^r + n^{-2}),$$

$$\begin{aligned}
cum(S_{nL}^i, S_{nL}^j, S_{nL}^k) &= O(n^{-1/2}h^r), \\
cum(S_{nL}^i, S_{nL}^j, S_{nL}^k, S_{nL}^l) &= O(n^{-1}h^{2r}), \\
cum(S_{nL}^{j_1}, ..., S_{nL}^{j_m}) &= O(n^{-(m-2)/2}) \text{ for } m \geq 5.
\end{aligned}$$

Let $\mathcal{B}$ be a class of Borel sets satisfying

$$\sup_{B \in \mathcal{B}} \int_{(\partial B)^\varepsilon} \phi_{0,I}(x) dx = O(\varepsilon) \text{ as } \varepsilon \downarrow 0, \tag{A.23}$$

where $(\partial B)^\varepsilon$ denotes the set of all points distant at most ε from the boundary of B and $\phi_{0,I}$ is the density function of the standard K - dimensional normal distribution. A formal Edgeworth expansion for the distribution of $n^{1/2}S_L$ is given as follows: assuming $nh^{2r} \to 0$,

$$\sup_{B \in \mathcal{B}} \left| P\left(n^{1/2}S_L \in B\right) - \int_B p(x)\phi_{0,I}(x) dx \right| = O(n^{-2}) + o(nh^{2r}), \tag{A.24}$$

where

$$\begin{aligned}
p(x) &= 1 + p_1(x) + p_2(x), \\
p_1(x) &= \frac{1}{2}n^{-1}\left\{x'\Delta x - tr(\Delta)\right\}, \\
p_2(x) &= \text{odd polynomial in } x
\end{aligned} \tag{A.25}$$

and $\Delta = (\Delta^{ij})$ is a $K \times K$ matrix with

$$\begin{aligned}
\Delta^{ij} &= n^2\alpha^i\alpha^j + n\left\{\frac{1}{3}\alpha^{ijk}\alpha^k - \frac{1}{6}\alpha^{ikk}\alpha^j - \frac{1}{6}\alpha^{jkk}\alpha^i\right\} \\
&+ \frac{1}{2}\alpha^{ijkk} - \frac{1}{3}\alpha^{ikm}\alpha^{jkm} - \frac{1}{36}\alpha^{ijm}\alpha^{mkk} + \frac{1}{36}\alpha^{ikk}\alpha^{jll}.
\end{aligned}$$

Accepting that the Edgeworth expansion (A.24) is justified, we now develop an Edgeworth expansion for the distribution of $l_h(\beta_0)$. From (A.22), we have: for any $c > 0$

$$\begin{aligned}
P\left(l_h(\beta_0) \leq c\right) &= \int_{\|x\|<c^{1/2}} p(x)\phi_{0,I}(x)dx + O(n^{-2}) + o(nh^{2r}) \\
&= P(\chi_K^2 \leq c) \\
&\quad + \frac{1}{2}n^{-1}\int_{\|x\|<c^{1/2}} \left\{\sum_{i=1}^{K}\Delta^{ii}\left[(x^i)^2 - 1\right] - \sum_{i \neq j}\Delta^{ij}x^ix^j\right\}\phi_{0,I}(x)dx \\
&\quad + O(n^{-2}) + o(nh^{2r}) \\
&= P(\chi_K^2 \leq c) - n^{-1}tr(\Delta)K^{-1}cg_K(c) + O(n^{-2}) + o(nh^{2r}), \quad \text{(A.26)}
\end{aligned}$$

31

where the second inequality holds by the symmetry of $\phi_{0,I}(\cdot)$ and oddness of the polynomial $p_2(x)$ and the third inequality holds by the symmetry of $\phi_{0,I}(\cdot)$ and $g_K(\cdot)$ denotes the density of χ^2_K distribution. It is straightforward to see that

$$tr(\Delta) = n^2 \alpha^i \alpha^i + \frac{1}{2} \alpha^{iikk} - \frac{1}{3} \alpha^{ikm} \alpha^{ikm}. \tag{A.27}$$

Let

$$\zeta \equiv E\left[X f^{(r-1)}(0|X)\right].$$

Then, using (A.22) and Lemma 1, we have

$$
\begin{aligned}
n^2 \alpha^i \alpha^i &= n^2 (EZ)' V_n^{-1} (EZ) \\
&= (nh^r)^2 (r!)^{-2} C_K^2 \left(\zeta' S^{-1} \zeta\right) q^{-1}(1-q)^{-1} + o\left((nh^r)^2\right) \tag{A.28}
\end{aligned}
$$

Similarly, we have

$$
\begin{aligned}
\alpha^{iikk} &= E\left\{[G_h(-U) - q]^4 \left(X' V_n^{-1} X\right)^2\right\} \\
&= q^{-1}(1-q)^{-1}(1 - 3q + 3q^2) E\left\{(X'SX)^2\right\} + O(h) \\
&< \infty \tag{A.29}
\end{aligned}
$$

and

$$
\begin{aligned}
\alpha^{ikm} &= E\left\{[G_h(-U) - q]^3 \left(v_{ni}^{-1/2} X\right) \left(v_{nk}^{-1/2} X\right) \left(v_{nm}^{-1/2} X\right)\right\} \\
&= q^{-1/2}(1-q)^{-1/2}(1 - 2q) E\left\{\left(s_i^{-1/2} X\right) \left(s_k^{-1/2} X\right) \left(s_m^{-1/2} X\right)\right\} + O(h) \\
&< \infty, \tag{A.30}
\end{aligned}
$$

where $v_{ni}^{-1/2}$ and $s_i^{-1/2}$ denote the i-th row of $V_n^{-1/2}$ and $S^{-1/2}$ respectively..

Therefore, (A.26), (A.27), (A.28), (A.29) and (A.30) give

$$
\begin{aligned}
&P\left(l_h(\beta_0) \le c_\alpha\right) \\
&= \alpha - n^{-1}\left\{(nh^r)^2 (r!)^{-2} C_K^2 \left(\zeta' S^{-1} \zeta\right) q^{-1}(1-q)^{-1} + O(1)\right\} K^{-1} c_\alpha \xi_p(c_\alpha) \\
&\quad + o(n^{-1} + nh^{2r}). \tag{A.31}
\end{aligned}
$$

It now follows that, since $\sup_n nh^r < \infty$, we have

$$P\left(l_h(\beta_0) \le c_\alpha\right) = \alpha + O(n^{-1}),$$

as desired.

It remains to check that the formal expansion (A.24) is valid. Since $\overline{A}^{j_1 \cdots j_k} = A^{j_1 \cdots j_k} + \alpha^{j_1 \cdots j_k}$ for each $k \ge 1$, we can see that $n^{1/2} S_L$ is a "smooth function of the means of independent and identically distributed random variables Q_i", where Q_i is defined in (A.17). Note that the validity of Edgeworth expansion for the distribution

of $n^{1/2}\overline{Q}$ has been established in Lemma 5 above. Therefore, from Lemma 2.1 and Theorem 2 of Bhattacharya and Ghosh (1978), the Edgeworth expansion in Lemma 5 can be transformed to yield a valid Edgeworth expansion (A.24) under Assumptions 1-6. This proves Theorem 2. $\blacksquare$

Proof of Theorem 3: By (A.26), we have for all $c > 0$

$$
\begin{aligned}
& P\left(l_h(\beta_0) \leq c\left(1 + n^{-1}b\right)\right) \\
= \; & P(\chi_K^2 \leq c\left(1 + n^{-1}b\right)) \\
& - c\left\{n\alpha^i\alpha^i K^{-1} + n^{-1}b\right\}\left\{1 + n^{-1}b\right\} g_K\left[c\left(1 + n^{-1}b\right)\right] + O(n^{-2}) + o(nh^{2r}).
\end{aligned}
\tag{A.32}
$$

Note that since g_K is the density of χ_K^2 distribution,

$$
g_K\left[c\left(1 + n^{-1}b\right)\right] = g_K(c) + O(n^{-1})
\tag{A.33}
$$

and

$$
P(\chi_K^2 \leq c\left(1 + n^{-1}b\right)) = P(\chi_K^2 \leq c) + cn^{-1}bg_K(c) + O(n^{-2}).
\tag{A.34}
$$

By substituting (A.33) and (A.34) into (A.32), we have

$$
\begin{aligned}
& P\left(l_h(\beta_0) \leq c\left(1 + n^{-1}b\right)\right) \\
= \; & P(\chi_K^2 \leq c) \\
& - cn\alpha^i\alpha^i K^{-1} g_K(c) + O(n^{-2}) + o(nh^{2r}) \\
= \; & P(\chi_K^2 \leq c) \\
& - nh^{2r}(r!)^{-2}C_K^2\left(\zeta'\Sigma^{-1}\zeta\right)q^{-1}(1-q)^{-1} \cdot cK^{-1}g_K(c) \\
& + O(n^{-2}) + o(nh^{2r}),
\end{aligned}
\tag{A.35}
$$

where the second equality follows from (A.28). Therefore, $\sup_n n^3h^{2r} < \infty$ implies that

$$
P\left(l_h(\beta_0) \leq c\left(1 + n^{-1}b\right)\right) = P(\chi_K^2 \leq c) + O(n^{-2})
\tag{A.36}
$$

for all $c > 0$. The proof of Theorem 3 is complete by taking $c = c_\alpha$ in (A.36).

The case where b is replaced by $\widehat{b}$ or $\widetilde{b}$ may be treated in a similar way using the fact $\widehat{b} = b + O_p(n^{-1/2})$ and the parity properties of polynomials in Edgeworth expansions such as (A.25). $\blacksquare$

Proof of Theorem 5: Theorem 5 can be verified by repeating the proofs of Lemmas 1-5 and Theorems 1-4 with $Z_i(\beta) = [G_h(X_i'\beta - Y_i) - q] G_h(X_i'\beta)X_i$ and with Assumption 3* in place of Assumption 3. $\blacksquare$

References

[1] Baggerly, K. A. (1998), "Empirical likelihood as a goodness-of-fit measure," *Biometrika* 85, 535-547.

[2] Barndorff, O. E. and P. Hall (1988), "On the level-error after Bartlett adjustment of the likelihood ratio statistic," *Biometrika* 75, 374-378.

[3] Bhattacharya, R. N. and J. K. Ghosh (1978), "On the validity of the formal Edgeworth expansion," *Annals of Statistics* 6, 434-451.

[4] Beran, R. (1987), "Prepivoting to reduce level error of confidence sets," *Biometrika* 74, 457-468.

[5] Beran, R. (1988), "Prepivoting test statistics: A bootstrap view of asymptotic refinements," *Journal of the American Statistical Association* 83, 697-697.

[6] Bravo, F. (2002), "Testing linear restriction in linear models with empirical likelihood," *The Econometrics Journal* 5, 104 - 130.

[7] Bravo, F. (2004), "Empirical likelihood based inference with applications to some econometric models," forthcoming in *Econometric Theory*.

[8] Buchinsky, M. (1995), "Estimating the asymptotic covariance matrix for quantile regression models: A Monte Carlo study," *Journal of Econometrics* 68, 303-338.

[9] Buchinsky, M. (2000), "Recent advances in quantile regression models: A practical guideline for empirical research," *Journal of Human Resources* 33, 88-126.

[10] Chen, S. X. (1993), "On the accuracy of empirical likelihood confidence regions for linear regression models," *Annals of Institute of Statistical Mathematics* 45, 621-637.

[11] Chen, S. X. and H. Cui (2002), "On Bartlett correction of empirical likelihood in the presence of nuisance parameters," unpublished manuscript.

[12] Chen, S. X. and H. Cui (2003), "On the second order properties of empirical likelihood for generalized estimating equations," unpublished manuscript.

[13] Chen, S. X. and P. Hall (1993), "Smoothed empirical Likelihood confidence intervals for quantiles," *Annals of Statistics* 21, 1166-1181.

[14] De Angelis, D. P. Hall, and G. A. Young (1993), "Analytical and bootstrap approximations to estimator distributions in L^1 regressions," *Journal of the American Statistical Association* 88, 1310-1316.

[15] Di Ciccio, T. J., P. Hall, and J. P. Romano (1991), "Empirical likelihood is Bartlett-correctable," *Annals of Statistics* 19, 1053-1061.

[16] Donald, S. G., G. W. Imbens and W. K. Newey (2003), "Empirical likelihood estimation and consistent tests with conditional moment restrictions," *Journal of Econometrics* 117, 55-93.

[17] Efron, B. (1979), "Bootstrap methods: Another look at the jackknife," *Annals of Statistics* 7, 1-26.

[18] Efron, B. (1982), *The jackknife, the bootstrap and other resampling plans*, Society for Industrial Applied Mathematics, Philadelphia, PA.

[19] Guggenberger, P. and R. J. Smith (2003), "Generalized empirical likelihood estimators and tests under partial, weak and strong identification," unpublished manuscript.

[20] Hahn, J. (1995), "Bootstrapping quantile regression estimators," *Econometric Theory* 11, 105-121.

[21] Hall, P. (1986), "On the bootstrap and confidence intervals," *Annals of Statistics* 14, 1431-1452.

[22] Hall, P. (1987), "On the bootstrap and likelihoood-based confidence regions," *Biometrika* 74, 481-493.

[23] Hall, P. (1992), *The Bootstrap and Edgeworth Expansions*, New York: Springer-Verlag.

[24] Hall, P. and La Scala (1990), "Methodology and algorithms of empirical likelihood," *International Statistical Review* 58, 109-127.

[25] Hall, P. and J. L. Horowitz (1990), "Bandwidth selection in semiparametric estimation of censored linear regression models, " *Econometric Theory* 6, 123-150.

[26] Horowitz, J. L. (1997), "Bootstrap methods in econometrics: Theory and numerical performance," in *Advances in Economics and Econometrics: Theory and Applications*, Vol. 3, ed. by D. Kreps and K. W. Wallis, Cambridge: Cambridge University Press, pp. 188-222.

[27] Horowitz, J. L. (1998), "Bootstrap methods for median regression models," *Econometrica* 66, 1327-1351.

[28] Horowitz, J. L. (2001), "The Bootstrap," in *Handbook of Econometrics*, Vol. 5, ed. by J. J. Heckman and E. Leamer, New York: North-Holland, pp.3159-3228.

[29] Imbens, G. W., R. H. Spady and P. Johnson (1998), "Information theoretic approaches to inference in moment condition models," *Econometrica* 66, 333-357.

[30] Jing, B. -Y. and A. T. A. Wood (1996), "Exponential empirical likelihood is not Bartlett correctable," *Annals of Statistics* 24, 365-369.

[31] Kitamura, Y. (1997), "Empirical likelihood methods with weakly dependent processes," *Annals of Statistics* 25, 2084-2102.

[32] Kitamura, Y. (2001), "Asymptotic optimality of empirical likelihood for testing moment restrictions," *Econometrica* 69, 1661-1672.

[33] Kitamura, Y. and M. Stutzer (1997), "An information-theoretic alternative to generalized method of moment estimation," *Econometrica* 65, 861-874.

[34] Koenker, R. and G. Bassett (1978), "Regression quantiles," *Econometrica* 46, 33-50.

[35] Koenker, R. and G. Bassett (1982), "Robust tests for heteroskedasticity based on regression quantiles," *Econometrica* 50, 43-61.

[36] Lazar, N. A. and P. A. Mykland (1999), "Empirical likelihood in the presence of nuisance parameters," *Biometrika* 86, 203-211.

[37] Monti, A. C. (1997), "Empirical likelihood confidence regions in time series models," *Biometrika* 84, 395-405.

[38] Moon, H. R. and F. Schorfheide (2003), "Empirical likelihood estimation based on inequality constraints," unpublished manuscript.

[39] Newey, W. K. and R. J. Smith (2003), "Higher order properties of GMM and generalized empirical likelihood estimators," forthcoming in *Econometrica*.

[40] Otsu, T. (2003), "Empirical likelihood for quantile regression," unpublished manuscript.

[41] Owen, A. (1988), "Empirical likelihood ratio confidence intervals for a single functional," *Biometrika* 75, 237-249.

[42] Owen, A. B. (1990), "Empirical likelihood ratio confidence regions," *Annals of Statistics* 18, 90-120.

[43] Owen, A. B. (1991), "Empirical likelihood for linear models," *Annals of Statistics* 19, 1725-1747.

[44] Owen, A. B. (2001), *Empirical Likelihood*, New York: Chapman and Hall.

[45] Powell, J. L. (1984), "Least absolute deviations estimation for the censored regression model," *Journal of Econometrics* 25, 303-325.

[46] Powell, J. L. (1986), "Censored regression quantiles," *Journal of Econometrics* 32, 143-155.

[47] Qin, J. and J. Lawless (1994), "Empirical likelihood and general estimating equations," *Annals of Statistics* 22, 300-325.

[48] Su, L. and H. White (2003), "Testing conditional independence via empirical likelihood, " unpublished manuscript.

[49] Wilks, S. S. (1938), "The large-sample distribution of the likelihood ratio for testing composite hypothesis," *Annals of Mathematical Statistics* 9, 60-62.

[50] Wu, C. F. J. (1986), "Jackknife, bootstrap and other resampling methods in regression analysis (with discussion)," *Annals of Statistics* 14, 1261-1350.

Table 1. Estimated True Coverage Probabilities of α-Level Confidence Regions
(DGP1)

n	$-\gamma$	$\alpha = .90$					
		LAD	BLAD	SLAD	SEL1	SEL2	SEL3
	.1	.828	.932	.947	.860	.861	.878
	.3	.675	.932	.962	.869	.870	.884
20	.5	.499	.932	.970	.873	.874	.889
	.7	.308	.932	.977	.875	.876	.890
	.9	.158	.932	.978	.876	.876	.890
	.1	.866	.919	.921	.885	.885	.896
	.3	.734	.919	.945	.889	.889	.898
35	.5	.562	.919	.952	.890	.890	.898
	.7	.350	.919	.958	.891	.891	.899
	.9	.167	.919	.968	.890	.890	.899
	.1	.879	.910	.926	.892	.892	.899
	.3	.762	.910	.945	.893	.893	.899
50	.5	.597	.910	.948	.895	.895	.900
	.7	.377	.910	.942	.895	.895	.901
	.9	.172	.910	.956	.895	.895	.900

n	$-\gamma$	$\alpha = .95$					
		LAD	BLAD	SLAD	SEL1	SEL2	SEL3
	.1	.889	.960	.978	.913	.914	.926
	.3	.750	.960	.987	.921	.922	.933
20	.5	.571	.960	.990	.926	.926	.936
	.7	.367	.960	.994	.929	.929	.939
	.9	.193	.960	.993	.930	.931	.940
	.1	.920	.954	.966	.939	.939	.947
	.3	.805	.954	.977	.942	.943	.949
35	.5	.637	.954	.985	.943	.943	.949
	.7	.412	.954	.988	.944	.944	.949
	.9	.204	.954	.989	.944	.944	.949
	.1	.932	.949	.964	.944	.944	.949
	.3	.830	.949	.976	.946	.946	.950
50	.5	.674	.949	.981	.947	.947	.950
	.7	.444	.949	.980	.945	.945	.949
	.9	.209	.949	.986	.945	.946	.949

Table 2. Estimated True Coverage Probabilities of α-Level Confidence Regions
(DGP2)

		$\alpha = .90$					
n	$-\gamma$	LAD	BLAD	SLAD	SEL1	SEL2	SEL3
	.1	.800	.917	.917	.860	.860	.878
	.3	.648	.917	.926	.868	.869	.885
20	.5	.474	.917	.937	.874	.874	.890
	.7	.292	.917	.946	.875	.875	.891
	.9	.148	.917	.951	.876	.877	.890
	.1	.843	.905	.918	.887	.887	.898
	.3	.713	.905	.930	.890	.891	.900
35	.5	.540	.905	.945	.890	.891	.900
	.7	.337	.905	.946	.890	.890	.899
	.9	.159	.905	.956	.891	.891	.898
	.1	.857	.900	.906	.892	.892	.900
	.3	.742	.900	.928	.893	.893	.899
50	.5	.575	.900	.932	.894	.894	.900
	.7	.358	.900	.934	.895	.895	.901
	.9	.162	.900	.943	.894	.895	.901

		$\alpha = .95$					
n	$-\gamma$	LAD	BLAD	SLAD	SEL1	SEL2	SEL3
	.1	.863	.951	.964	.915	.915	.927
	.3	.723	.951	.969	.921	.922	.933
20	.5	.544	.951	.974	.926	.926	.936
	.7	.350	.951	.973	.928	.929	.939
	.9	.183	.951	.976	.930	.931	.940
	.1	.901	.945	.963	.938	.938	.945
	.3	.783	.945	.972	.943	.943	.949
35	.5	.615	.945	.978	.943	.943	.949
	.7	.398	.945	.976	.943	.943	.949
	.9	.194	.945	.981	.944	.944	.949
	.1	.913	.940	.955	.943	.943	.948
	.3	.811	.940	.961	.946	.946	.950
50	.5	.649	.940	.960	.946	.946	.950
	.7	.421	.940	.964	.945	.946	.950
	.9	.198	.940	.967	.945	.945	.949

Table 3. Estimated True Coverage Probabilities of α-Level Confidence Regions
(DGP3)

n	$-\gamma$	$\alpha = .90$					
		LAD	BLAD	SLAD	SEL1	SEL2	SEL3
	.1	.557	.904	.952	.868	.869	.885
	.3	.387	.904	.965	.871	.872	.887
20	.5	.223	.904	.969	.873	.873	.887
	.7	.104	.904	.973	.874	.874	.887
	.9	.042	.904	.968	.874	.874	.887
	.1	.660	.890	.932	.887	.887	.897
	.3	.497	.890	.950	.889	.889	.897
35	.5	.294	.890	.951	.888	.888	.896
	.7	.132	.890	.960	.887	.887	.896
	.9	.046	.890	.960	.887	.887	.895
	.1	.716	.891	.941	.892	.892	.898
	.3	.563	.891	.950	.893	.893	.899
50	.5	.343	.891	.953	.893	.894	.899
	.7	.151	.891	.955	.894	.894	.899
	.9	.050	.891	.963	.894	.894	.900

n	$-\gamma$	$\alpha = .95$					
		LAD	BLAD	SLAD	SEL1	SEL2	SEL3
	.1	.632	.940	.977	.923	.924	.935
	.3	.454	.940	.982	.926	.927	.937
20	.5	.269	.940	.989	.928	.928	.938
	.7	.128	.940	.988	.929	.930	.939
	.9	.053	.940	.988	.930	.930	.940
	.1	.732	.931	.966	.940	.941	.947
	.3	.567	.931	.975	.941	.941	.947
.35	.5	.351	.931	.978	.942	.942	.946
	.7	.160	.931	.985	.942	.942	.947
	.9	.058	.931	.983	.942	.942	.947
	.1	.786	.932	.975	.944	.944	.948
	.3	.635	.932	.979	.945	.945	.949
.50	.5	.407	.932	.984	.945	.945	.949
	.7	.185	.932	.978	.945	.945	.949
	.9	.062	.932	.975	.946	.946	.949

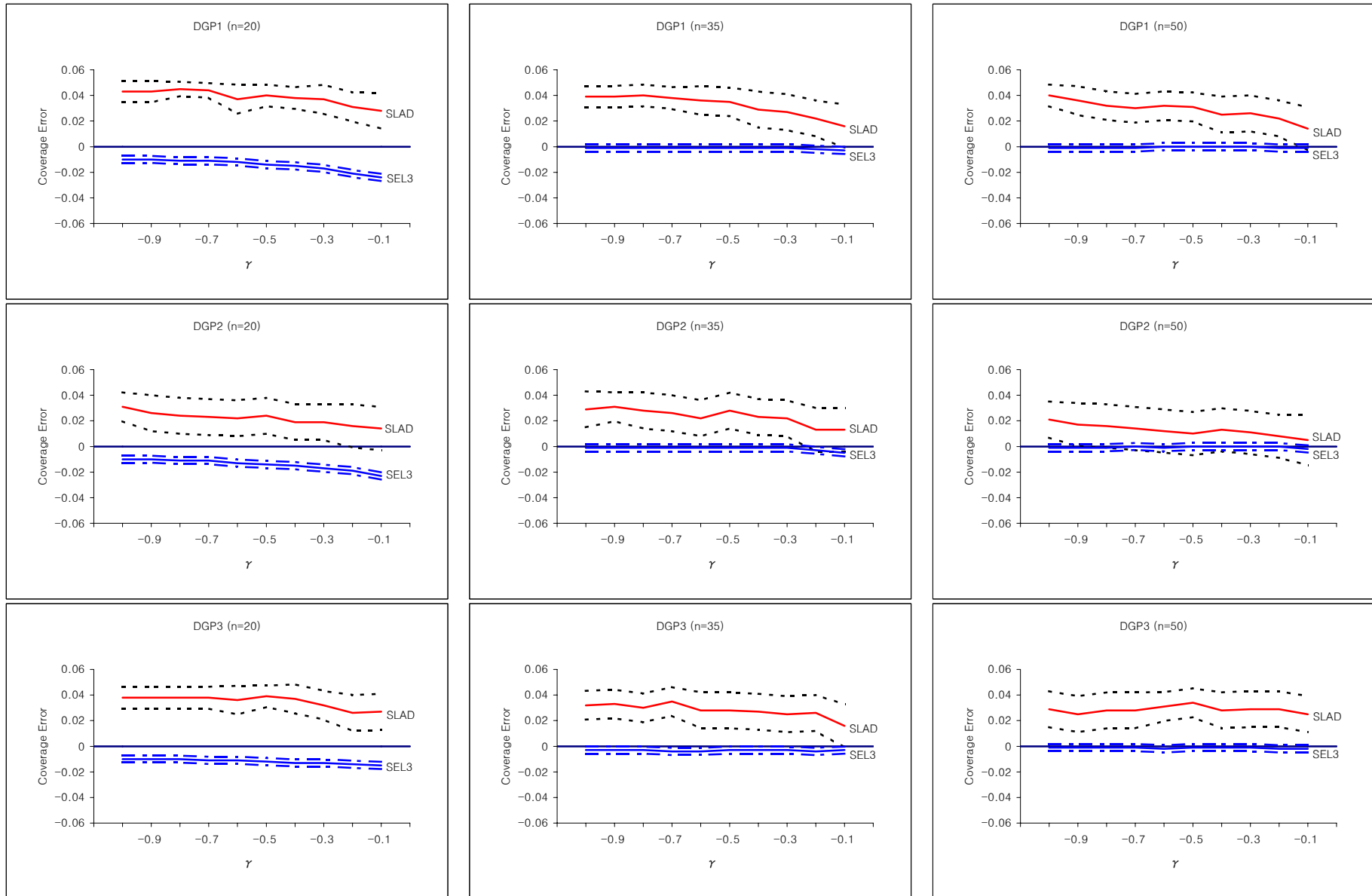Fig. 1 Sensitivity of Coverage Errors with respect to Bandwidth Parameters
[α=0.95]

Fig. 2 Coverage Errors with Varying Sample Sizes
[α=0.95, γ= -0.1(SLAD), -0.9(SEL1, SEL3)]