Econometric Theory, **22**, 2006, 173–205. Printed in the United States of America.
DOI: 10.1017/S0266466606060087

SMOOTHED EMPIRICAL LIKELIHOOD METHODS FOR QUANTILE REGRESSION MODELS

Yoon-Jae Whang
Seoul National University
and
Cowles Foundation for Research in Economics, Yale University

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 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 errors of order n^{-2}, where n denotes the sample size. Our result is an extension of the previous result of Chen and Hall (1993, *Annals of Statistics* 21, 1166–1181) to the regression context. 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, *Econometrica* 66, 1327–1351).

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 as a result of 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) and Koenker (2005) for a recent survey.

Koenker and Bassett (1978, 1982) give conditions under which their quantile regression (QR) estimator is $n^{1/2}$-consistent and asymptotically normal. This

I thank the co-editor Bruce Hansen and anonymous referees for valuable suggestions and comments. I also thank Song X. Chen, Joel Horowitz, Yuichi Kitamura, Oliver Linton, Whitney Newey, Peter Phillips, and Richard Smith for helpful comments. Parts of this paper were written while I was visiting the Cowles Foundation at Yale University, whose hospitality is gratefully acknowledged. Young-Hyun Cho has provided excellent research assistance. This work was supported by the Korea Research Foundation grant KRF 2003-041-B00072. Address correspondence to Yoon-Jae Whang, School of Economics, Seoul National University, Seoul 151-742; e-mail: whang@snu.ac.kr.

0266-4666/06 $12.00

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; Horowitz, 1997, 2001). However, the standard theory of the bootstrap cannot 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 (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 (EL) method (originally proposed by Owen, 1988, 1990) to estimate the parameters of the QR models and to construct confidence regions for the parameters, which are expected to be more accurate in finite samples than the confidence regions that can be constructed from the SLAD/bootstrap approach.[2] To discuss the higher order properties of the EL confidence regions, we consider smoothing the estimating equations of the standard QR estimator. We establish that the resulting smoothed EL (SEL) estimator is first-order asymptotically equivalent to the standard QR estimator and the confidence regions based on the SEL ratio statistic have coverage errors of order $O(n^{-1})$. Furthermore, we show that the SEL 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 of the concept of the "optimal" smoothing parameter is not necessary.

There are a number of papers in the literature that are related to this paper. Chen and Hall (1993) show that smoothed confidence intervals for quantiles *with no covariates* are Bartlett correctable, and our paper is an extension of the latter paper to the QR context, which perhaps should be of more interest to econometricians. On the other hand, contrary to De Angelis, Hall, and Young (1993), we do not assume that the error terms in the QRs are independent of regressors (X), and hence they may have an unknown form of conditional heteroskedasticity. Finally, we should note that Otsu (2003), independently of our work, has recently obtained a result similar to ours, assuming independence of the errors and regressors. However, the main focus of the latter paper is on the relative efficiency (over other competing estimators) of smoothed *conditional* EL estimators in the first-order approximations, whereas our main focus is on the higher order properties of smoothed *unconditional* EL confidence regions.

The remainder of this paper is organized as follows. Section 2 defines the SEL estimator and confidence region in QR models and discusses their asymptotic properties. Section 3 reports some Monte Carlo results. Section 4 makes concluding remarks and suggests some possible extensions. An Appendix contains proofs of the results.

2. SMOOTHED EMPIRICAL LIKELIHOOD FOR QUANTILE REGRESSIONS

2.1. Definition of the SEL Estimator and Confidence Regions

Consider the linear QR model given by

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

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 $0 < q < 1$. For simplicity, we assume that $\{(Y_i, X_i) : i = 1,\ldots,n\}$ are independent and identically distributed (i.i.d.).[4]

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.[5] 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 problems for our subsequent (higher order) asymptotic analysis because most of the theoretical development of EL has focused on the statistic that is a *smooth* function of random variables. 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 later. 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 $p = (p_1, \ldots, p_n)'$ be a vector of nonnegative numbers adding to unity. Then, the *smoothed empirical log likelihood ratio* is defined by

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

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

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

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

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

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 $\hat{\beta}_E$ of β_0 solves

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

where B is the parameter space. Using the just-identification assumption in (2), it can be shown that $\hat{\beta}_E$ satisfies

$$p_i(\hat{\beta}_E) = \frac{1}{n}, \quad t(\hat{\beta}_E) = 0, \quad \frac{1}{n} \sum_{i=1}^{n} Z_i(\hat{\beta}_E) = 0; \tag{10}$$

see Lemma 3 in the Appendix.

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

$$\min_{\beta \in B} H_n(\beta) = \frac{1}{n} \sum_{i=1}^{n} \rho_q(Y_i - X_i'\beta), \tag{11}$$

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

$$n^{-1}\sum_{i=1}^{n} g(Y_i, X_i, \beta) = n^{-1}\sum_{i=1}^{n}[1(Y_i \le X_i'\beta) - q]X_i = o_p(n^{-1/2}), \tag{12}$$

which approximately equals to the unsmoothed version (i.e., $h = 0$) of the estimating equations $\sum p_i Z_i(\beta) = 0$ for the smoothed EL (5). Under the regularity conditions given later, we shall show that the two estimators are (first-order) asymptotically equivalent in the sense that

$$\sqrt{n}(\hat{\beta}_E - \hat{\beta}_Q) = o_p(1) \quad \text{as } n \to \infty. \tag{13}$$

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

$$\sqrt{n}(\hat{\beta}_E - \beta_0) \xrightarrow{d} N(0, \Lambda_0), \quad \text{where} \tag{14}$$

$$\Lambda_0 = q(1-q)D_0^{-1}S_0D_0^{-1}, \tag{15}$$

$$S_0 = E[X_iX_i'], \qquad D_0 = E[f(0|X_i)X_iX_i'], \tag{16}$$

and $f(u|x)$ denote the conditional density of U given $X = x$. On the other hand, when $q = \frac{1}{2}$, the result (13) and Theorem 2.1 of Horowitz (1998) imply that $\hat{\beta}_E$ 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

$$I_{hc} = \{\beta : l_h(\beta) \le c\}, \tag{17}$$

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

$$\alpha_{hc} = P(\beta_0 \in I_{hc}) = P(l_h(\beta_0) \le c). \tag{18}$$

The coverage accuracy of I_{hc} depends on the asymptotic distribution of the $l_h(\beta_0)$ statistic. As we shall see subsequently, under suitable regularity conditions, $l_h(\beta_0)$ has an asymptotic χ^2_K distribution, and hence c might be chosen using this result.[6] As noted by Chen and Hall (1993), if G_h is a higher order kernel, then it is possible that I_{hc} 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, because 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 (36), (37), and (39) defined in Section 3.1), 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.[7] 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.

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

Assumption 1. $\{(Y_i, X_i) : i = 1,\dots,n\}$ are i.i.d. 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 $C_K \neq 0$, $K(\cdot)$ is an rth-order kernel, i.e., $\int u^j K(u)\,du = 1$ if $j = 0$; 0 if $1 \leq j \leq r-1$; C_K if $j = r$.
(c) Let $\tilde{G}(u) = ([G(u)],[G(u)]^2,\dots,[G(u)]^{L+1})'$ 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 $[-1,1]$, $-1 = a_0 < a_1 < \cdots < a_{L+1} = 1$ such that $\theta'\tilde{G}(u)$ is either strictly positive or strictly negative on (a_{l-1}, a_l) for $l = 1,\dots,L+1$.

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 $\sqrt{n}(\hat{\beta}_E - \beta_0)$ and to justify a Taylor expansion for the EL ratio statistic that 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 the Appendix. It can be removed at the cost of more complicated proofs. Also, although the distribution U_i given X_i is assumed to be continuous, the distribution of X_i can be discrete unless it is identically zero with probability 1, which is excluded by Assumption 3. Note that the discrete regressors are important for applications with dummy variables. Assumption 5(c) is used to verify a version of Cramér's condition (Lemma 4 in the Appendix) that 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 $r \geq 2$. 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, part (b) of Assumption 6 requires that h not be too small. It is needed to ensure a minimum level of smoothness of $l_h(\beta_0)$ that is necessary to derive Cramér's condition for the Edgeworth analysis. Intuitively this assumption makes sense, because Cramér's condition is mainly intended to ensure distributions of statistics to have an absolutely continuous component, but the latter might be hard to attain for $l_h(\beta_0)$ if h is chosen too small (for a general interpretation of Cramér's condition, see Hall, 1992, p. 57).

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), we have*

$$(a)\ \sqrt{n}(\hat{\beta}_E - \hat{\beta}_Q) = o_p(1); \qquad (b)\ \sqrt{n}(\hat{\beta}_E - \beta_0) \xrightarrow{d} N(0, \Lambda_0),$$

where Λ_0 is defined in (15).

The asymptotic covariance matrix Λ_0 can be estimated, e.g., by

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

$$\hat{D} = (nh)^{-1}\sum_{i=1}^{n} K((Y_i - X_i'\hat{\beta}_E)/h)X_iX_i', \qquad \hat{S} = n^{-1}\sum_{i=1}^{n} X_iX_i', \tag{20}$$

which is suggested by Powell (1984, 1986). Under the assumptions of Theorem 1, it is not difficult to show that the estimator is consistent for Λ_0. Another way to estimate Λ_0 is to use a bootstrap estimator as in Buchinsky (1995, 2000), e.g., (38) in Section 3.1. 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.[8]

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

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

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

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

where λ satisfies

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

We need to introduce additional notation. We let W_i^j denote the jth component of W_i and define

$$\alpha^{j_1 \dots j_k} = EW_i^{j_1} \dots W_i^{j_k}, \tag{24}$$

$$\bar{A}^{j_1 \dots j_k} = n^{-1} \sum_{i=1}^{n} W_i^{j_1} \dots W_i^{j_k}, \quad \text{and} \quad A^{j_1 \dots j_k} = \bar{A}^{j_1 \dots j_k} - \alpha^{j_1 \dots j_k}.$$

In particular, $\alpha^{jk} = \delta^{jk}$, where δ^{jk} is the Kronecker delta.[9]

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, as* $n \to \infty$,

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

Remarks.

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

2. From the expansion (A.6) and Lemma 1(a) in the Appendix, we can see that $n^{1/2} EZ_i \to 0$ if $nh^{2r} \to 0$ and, if $E[X_i f^{(r-1)}(0|X_i)] \neq 0$, $n^{1/2} EZ_i \to 0$ implies $nh^{2r} \to 0$. Therefore, if $E[X_i f^{(r-1)}(0|X_i)] \neq 0$, 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

$$P(\chi_K^2 \leq c_\alpha) = \alpha, \tag{25}$$

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 $n \to \infty$.

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, as $n \to \infty$,*

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

Remarks.

1. The expansion (A.23) in the 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 $E[X_i f^{(r-1)}(0|X_i)] \neq 0$. If $E[X_i f^{(r-1)}(0|X_i)] = 0$, 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.23) and the results (A.20)–(A.22) in the 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 practical of interest because of the availability of the Bartlett correction, which is discussed in the next section.

2.3. Bartlett Correction

In the previous section, the coverage error of the EL confidence region I_{hc} is 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 the 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 the 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.5), we can show that if $nh^r \to 0$

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

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

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*:

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

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

$$\hat{b} = K^{-1}(\hat{\alpha}^{iikk}/2 - \hat{\alpha}^{ikm}\hat{\alpha}^{ikm}/3), \quad \text{where} \tag{29}$$

$$\hat{\alpha}^{iikk} = n^{-1}\sum_{j=1}^{n}\hat{\varepsilon}_j^4(X_j'\hat{V}_n^{-1}X_j)^2, \qquad \hat{\alpha}^{ikm} = n^{-1}\sum_{j=1}^{n}\hat{\varepsilon}_j^3\hat{v}_{ni}^{-1/2}X_j\hat{v}_{nk}^{-1/2}X_j\hat{v}_{nm}^{-1/2}X_j,$$

$$\hat{V}_n = n^{-1}\sum_{j=1}^{n}\hat{\varepsilon}_j^2X_jX_j', \qquad \hat{\varepsilon}_j = G_h(X_j'\hat{\beta} - Y_j) - q, \tag{30}$$

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

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

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

$$I_{hc}^{\hat{b}} = \{\beta : l_h(\beta) \le c(1 + n^{-1}\hat{b})\}. \tag{32}$$

Theorem 4, which follows, 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 $\hat{b}$.

On the other hand, from (A.21) and (A.22) in the Appendix, we have

$$\alpha^{iikk} = q^{-1}(1-q)^{-1}(1 - 3q + 3q^2)E\{(X_j'S_0X_j)^2\} + O(h) \quad \text{and}$$

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

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

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

with the Bartlett factor given by

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

where $\bar{S} = n^{-1}\sum_{k=1}^{n}X_kX_k'$. Note that $\tilde{b}$ is computationally simpler than $\hat{b}$ because it does not depend on the bandwidth parameter h. However, if $\tilde{b}$ is used instead of b, we will not have the same asymptotic accuracy as b or $\hat{b}$ because of the 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 $O(n^{-2})$.[10]

The following theorem formally states the preceding 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, as* $n \to \infty$,

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

Remarks.

1. The result (A.30) in the Appendix implies that the condition $\sup_n n^3 h^{2r} < \infty$ is also necessary for the asymptotic coverage error of I^{b}_{hc} or $I^{\hat{b}}_{hc}$ to be of order $O(n^{-2})$ if $E[X_i f^{(r-1)}(0|X_i)] \neq 0$.

2. The Edgeworth expansion for the distribution of $l_h(\beta_0)$ (i.e., eqn. (A.18) in the Appendix) implies that we have, for any $c > 0$,

$$P(l_h(\beta_0) \leq c) = P(\chi^2_K \leq c) - n^{-1}\tilde{q}(c^{K/2})\exp(-c/2) + O(n^{-2}) + o(nh^{2r}), \tag{35}$$

where $\tilde{q}(u)$ is a polynomial of degree 1: i.e., $\tilde{q}(u) = C_0 \cdot u$, where $C_0 = [K2^{K/2}\Gamma(K/2)]^{-1}\mathrm{tr}(\Delta)$ is a constant. This property of $\tilde{q}(\cdot)$ is essential for our Bartlett correction result, because it enables us to remove the term of order n^{-1} from the right-hand side of (35) by a simple adjustment for the expected value of $l_h(\beta_0)$. On the other hand, the Edgeworth expansions of Horowitz (1998, Thm. 4.2) for the *t*- and Wald statistics and their bootstrap analogues have (odd) polynomials (i.e., $q(\nu_n, \cdot)$ on p. 1349 of the latter paper) with nonvanishing terms of degrees 3 and 5. The existence of the latter terms, in general, is the main reason why Bartlett correction is not available for bootstrap (for a general discussion of this phenomenon, see DiCiccio, Hall, and Romano, 1991, Sec. 2.3).

3. Bertail (2003) makes an ingenious use of convex duality and LAN (locally asymptotic normality) properties of the EL ratio to show that empirical likelihoods for Hadamard differentiable functionals are Bartlett correctable in a general setting.[11] However, it is not clear whether his result could be directly applicable to our context, because the result (Bertail, 2003, Thm. 2.1) has been derived under a set of high-level assumptions including Cramér's condition (i.e., eqn. (8) of the latter paper). In fact, inspection of our Lemma 3 in the Appendix shows that smoothness of the function $G : \mathbb{R} \to \mathbb{R}$ is crucial to verify Cramér's condition (or more specifically to verify the inequality (A.11)). Furthermore, contrary to our result, Bertail (2003) considers only an infeasible Bartlett corrected confidence region and does not suggest any explicit estimator of the Bartlett factor.

3. MONTE CARLO SIMULATIONS

3.1. Experimental Design

We consider a linear median regression model

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

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 EL using a second-order kernel (i.e., $r = 2$) $K(u) = (\frac{3}{4})(1 - u^2)1(|u| \leq 1)$, which is the so-called Bartlett or Epanečnikov kernel. The SEL confidence regions considered are I_{hc}, $I_{hc}^{\tilde{b}}$, and $I_{hc}^{\hat{b}}$, which are defined in (17), (33), and (32), respectively. In the simulation results given subsequently, we denote them by 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. To evaluate the effect of smoothing the estimating equations, we also consider unsmoothed EL confidence regions (i.e., the case $h = 0$), and we denote this by EL.

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

$$I_{LAD} = \{\beta : n(\hat{\beta}_Q - \beta)' \hat{\Lambda}^{-1} (\hat{\beta}_Q - \beta) \leq c_\alpha\}, \tag{36}$$

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

$$I_{BLAD} = \{\beta : n(\hat{\beta}_Q - \beta)' \Lambda^{*-1} (\hat{\beta}_Q - \beta) \leq c_\alpha\}, \tag{37}$$

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

$$\Lambda^* = \frac{n}{B} \sum_{b=1}^{B} (\hat{\beta}_{Qb}^* - \bar{\hat{\beta}}_Q^*)(\hat{\beta}_{Qb}^* - \bar{\hat{\beta}}_Q^*)', \tag{38}$$

where $\bar{\hat{\beta}}_Q^* = (1/B)\sum_{b=1}^{B} \hat{\beta}_{Qb}^*$ and $\{\hat{\beta}_{Qb}^* : b = 1, \ldots, B\}$ 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 $\{(Y_i, X_i): i = 1,\dots,n\}$. 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

$$I_{SLAD} = \{\beta : n(\tilde{\beta}_S - \beta)'\tilde{\Lambda}^{-1}(\tilde{\beta}_S - \beta) \le \tilde{c}_\alpha^*\}. \tag{39}$$

Here, $\tilde{\beta}_S$ is the SLAD estimator of β_0 that solves

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

and its variance is estimated by

$$\tilde{\Lambda} = D_n(\tilde{\beta}_S)^{-1}T_n(\tilde{\beta}_S)D_n(\tilde{\beta}_S)^{-1}, \quad \text{where}$$

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

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

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

$$\tilde{G}(u) = \begin{cases} 0 & \text{if } u < -1, \\ 0.5 + \dfrac{105}{64}\left[u - \dfrac{5}{3}u^3 + \dfrac{7}{5}u^5 - \dfrac{3}{7}u^7\right] & \text{if } |u| \le 1, \\ 1 & \text{otherwise,} \end{cases} \tag{40}$$

and $\tilde{G}^{(1)}(u) = d\tilde{G}(u)/du$. The constant $\tilde{c}_\alpha^*$ is computed from the following bootstrap procedure. (i) Generate a bootstrap sample $\{(Y_i^*, X_i^*): i = 1,\dots,n\}$ by sampling the original data $\{(Y_i, X_i): i = 1,\dots,n\}$ randomly with replacement. (ii) Using the bootstrap sample, compute the SLAD estimate $\tilde{\beta}_S^*$ and its variance estimate $\tilde{\Lambda}^*$ and get $S_n^* = n(\tilde{\beta}_S^* - \tilde{\beta}_S)'\tilde{\Lambda}^{*-1}(\tilde{\beta}_S^* - \tilde{\beta}_S)$. (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 $\tilde{c}_\alpha^*$ 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 $n^{-1/5}$, 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 (40) is of order $n^{-\kappa}$, where $\frac{2}{9} < \kappa < \frac{1}{3}$. 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 preceding rules are justified in an asymptotic sense, and hence they provide little practical guidance in how to choose h in finite samples. We consider a rule of thumb $h = c_0 n^{\gamma}$ in our simulations and take $\gamma \in [-1.0, -0.9, \ldots, -0.1]$. 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 the LAD, EL, and SEL confidence regions. This yields simulation standard errors of approximately 0.0015 and 0.0010 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 0.0094 and 0.0068 for nominal 90% and 95% levels, respectively. The number of bootstrap repetitions used is also restricted to $B = 100$ because of heavy computational cost. We consider eight different sample sizes $n \in [15, 20, \ldots, 50]$.

3.2. 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 $\hat{b}$) confidence regions for different values of γ values (which determine 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 the DGP1 and $n = 35$ case, the coverage probabilities of the nominal 95% LAD confidence region are 0.920 and 0.204 for $\gamma = -0.1$ and $\gamma = -0.9$,

Table 1. Estimated true coverage probabilities of α-level confidence regions (DGP1)

n	$-\gamma$	LAD	BLAD	SLAD	EL	SEL1	SEL2	SEL3
$\alpha = 0.90$								
20	0.1	0.828	0.932	0.947	0.877	0.860	0.861	0.878
	0.3	0.675	0.932	0.962	0.877	0.869	0.870	0.884
	0.5	0.499	0.932	0.970	0.877	0.873	0.874	0.889
	0.7	0.308	0.932	0.977	0.877	0.875	0.876	0.890
	0.9	0.158	0.932	0.978	0.877	0.876	0.876	0.890
35	0.1	0.866	0.919	0.921	0.889	0.885	0.885	0.896
	0.3	0.734	0.919	0.945	0.889	0.889	0.889	0.898
	0.5	0.562	0.919	0.952	0.889	0.890	0.890	0.898
	0.7	0.350	0.919	0.958	0.889	0.891	0.891	0.899
	0.9	0.167	0.919	0.968	0.889	0.890	0.890	0.899
50	0.1	0.879	0.910	0.926	0.895	0.892	0.892	0.899
	0.3	0.762	0.910	0.945	0.895	0.893	0.893	0.899
	0.5	0.597	0.910	0.948	0.895	0.895	0.895	0.900
	0.7	0.377	0.910	0.942	0.895	0.895	0.895	0.901
	0.9	0.172	0.910	0.956	0.895	0.895	0.895	0.900
$\alpha = 0.95$								
20	0.1	0.889	0.960	0.978	0.932	0.913	0.914	0.926
	0.3	0.750	0.960	0.987	0.932	0.921	0.922	0.933
	0.5	0.571	0.960	0.990	0.932	0.926	0.926	0.936
	0.7	0.367	0.960	0.994	0.932	0.929	0.929	0.939
	0.9	0.193	0.960	0.993	0.932	0.930	0.931	0.940
35	0.1	0.920	0.954	0.966	0.944	0.939	0.939	0.947
	0.3	0.805	0.954	0.977	0.944	0.942	0.943	0.949
	0.5	0.637	0.954	0.985	0.944	0.943	0.943	0.949
	0.7	0.412	0.954	0.988	0.944	0.944	0.944	0.949
	0.9	0.204	0.954	0.989	0.944	0.944	0.944	0.949
50	0.1	0.932	0.949	0.964	0.946	0.944	0.944	0.949
	0.3	0.830	0.949	0.976	0.946	0.946	0.946	0.950
	0.5	0.674	0.949	0.981	0.946	0.947	0.947	0.950
	0.7	0.444	0.949	0.980	0.946	0.945	0.945	0.949
	0.9	0.209	0.949	0.986	0.946	0.945	0.946	0.949

respectively. On the other hand, the coverage probabilities of the BLAD confidence regions are relatively very good and stable across different designs.

2. Both the SLAD and SEL confidence regions are robust to the choice of bandwidth. However, Figure 1 shows some evidence that the SEL3 con-

TABLE 2. Estimated true coverage probabilities of α-level confidence regions (DGP2)

n	$-\gamma$	LAD	BLAD	SLAD	EL	SEL1	SEL2	SEL3
$\alpha = 0.90$								
20	0.1	0.800	0.917	0.917	0.877	0.860	0.860	0.878
	0.3	0.648	0.917	0.926	0.877	0.868	0.869	0.885
	0.5	0.474	0.917	0.937	0.877	0.874	0.874	0.890
	0.7	0.292	0.917	0.946	0.877	0.875	0.875	0.891
	0.9	0.148	0.917	0.951	0.877	0.876	0.877	0.890
35	0.1	0.843	0.905	0.918	0.889	0.887	0.887	0.898
	0.3	0.713	0.905	0.930	0.889	0.890	0.891	0.900
	0.5	0.540	0.905	0.945	0.889	0.890	0.891	0.900
	0.7	0.337	0.905	0.946	0.889	0.890	0.890	0.899
	0.9	0.159	0.905	0.956	0.889	0.891	0.891	0.898
50	0.1	0.857	0.900	0.906	0.895	0.892	0.892	0.900
	0.3	0.742	0.900	0.928	0.895	0.893	0.893	0.899
	0.5	0.575	0.900	0.932	0.895	0.894	0.894	0.900
	0.7	0.358	0.900	0.934	0.895	0.895	0.895	0.901
	0.9	0.162	0.900	0.943	0.895	0.894	0.895	0.901
$\alpha = 0.95$								
20	0.1	0.863	0.951	0.964	0.932	0.915	0.915	0.927
	0.3	0.723	0.951	0.969	0.932	0.921	0.922	0.933
	0.5	0.544	0.951	0.974	0.932	0.926	0.926	0.936
	0.7	0.350	0.951	0.973	0.932	0.928	0.929	0.939
	0.9	0.183	0.951	0.976	0.932	0.930	0.931	0.940
35	0.1	0.901	0.945	0.963	0.944	0.938	0.938	0.945
	0.3	0.783	0.945	0.972	0.944	0.943	0.943	0.949
	0.5	0.615	0.945	0.978	0.944	0.943	0.943	0.949
	0.7	0.398	0.945	0.976	0.944	0.943	0.943	0.949
	0.9	0.194	0.945	0.981	0.944	0.944	0.944	0.949
50	0.1	0.913	0.940	0.955	0.946	0.943	0.943	0.948
	0.3	0.811	0.940	0.961	0.946	0.946	0.946	0.950
	0.5	0.649	0.940	0.960	0.946	0.946	0.946	0.950
	0.7	0.421	0.940	0.964	0.946	0.945	0.946	0.950
	0.9	0.198	0.940	0.967	0.946	0.945	0.945	0.949

fidence 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 $\tilde{b}$ (SEL2) perform similarly, though SEL2 is slightly better than SEL1 in almost all cases. This confirms the theory in Theo-

Table 3. Estimated true coverage probabilities of α-level confidence regions (DGP3)

n	$-\gamma$	LAD	BLAD	SLAD	EL	SEL1	SEL2	SEL3
$\alpha = 0.90$								
20	0.1	0.557	0.904	0.952	0.875	0.868	0.869	0.885
	0.3	0.387	0.904	0.965	0.875	0.871	0.872	0.887
	0.5	0.223	0.904	0.969	0.875	0.873	0.873	0.887
	0.7	0.104	0.904	0.973	0.875	0.874	0.874	0.887
	0.9	0.042	0.904	0.968	0.875	0.874	0.874	0.887
35	0.1	0.660	0.890	0.932	0.887	0.887	0.887	0.897
	0.3	0.497	0.890	0.950	0.887	0.889	0.889	0.897
	0.5	0.294	0.890	0.951	0.887	0.888	0.888	0.896
	0.7	0.132	0.890	0.960	0.887	0.887	0.887	0.896
	0.9	0.046	0.890	0.960	0.887	0.887	0.887	0.895
50	0.1	0.716	0.891	0.941	0.895	0.892	0.892	0.898
	0.3	0.563	0.891	0.950	0.895	0.893	0.893	0.899
	0.5	0.343	0.891	0.953	0.895	0.893	0.894	0.899
	0.7	0.151	0.891	0.955	0.895	0.894	0.894	0.899
	0.9	0.050	0.891	0.963	0.895	0.894	0.894	0.900
$\alpha = 0.95$								
20	0.1	0.632	0.940	0.977	0.931	0.923	0.924	0.935
	0.3	0.454	0.940	0.982	0.931	0.926	0.927	0.937
	0.5	0.269	0.940	0.989	0.931	0.928	0.928	0.938
	0.7	0.128	0.940	0.988	0.931	0.929	0.930	0.939
	0.9	0.053	0.940	0.988	0.931	0.930	0.930	0.940
35	0.1	0.732	0.931	0.966	0.943	0.940	0.941	0.947
	0.3	0.567	0.931	0.975	0.943	0.941	0.941	0.947
	0.5	0.351	0.931	0.978	0.943	0.942	0.942	0.946
	0.7	0.160	0.931	0.985	0.943	0.942	0.942	0.947
	0.9	0.058	0.931	0.983	0.943	0.942	0.942	0.947
50	0.1	0.786	0.932	0.975	0.946	0.944	0.944	0.948
	0.3	0.635	0.932	0.979	0.946	0.945	0.945	0.949
	0.5	0.407	0.932	0.984	0.946	0.945	0.945	0.949
	0.7	0.185	0.932	0.978	0.946	0.945	0.945	0.949
	0.9	0.062	0.932	0.975	0.946	0.946	0.946	0.949

rems 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 $\hat{b}$ (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.

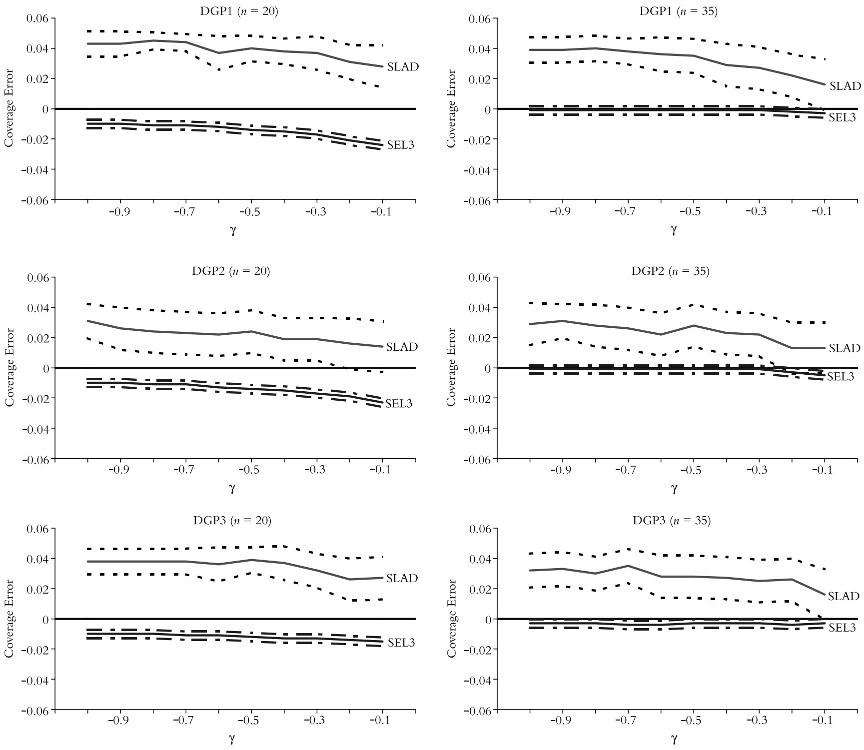

FIGURE 1. Sensitivity of coverage errors with respect to bandwidth parameters ($\alpha = 0.95$).

5. The overall performance of the unsmoothed EL confidence regions was similar to that of SEL1 and SEL2. This suggests that, in practice, smoothing of estimating equations is mainly useful to achieve higher order refinements via Bartlett correction.
6. The SLAD confidence regions perform fairly well, especially in small samples ($n \leq 20$) and in some cases, outperform SEL1 and SEL2.
7. The effect of increasing the sample size is to reduce coverage errors for almost all confidence regions.
8. 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 $O(n^{-2})$.
9. There is not much difference in relative performance of confidence regions under different DGPs.
10. The results for nominal 90% and 95% confidence regions are similar.

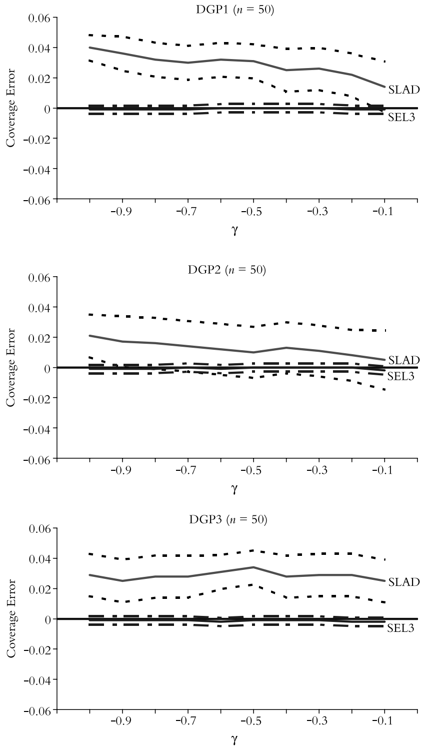

Figure 1. Continued.

11. 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 using the latter rule of thumb in practical applications, though the results seem to be very robust to the choice of γ.

4. CONCLUDING REMARKS AND EXTENSIONS

In this paper, we have used smoothed EL methods to obtain asymptotically valid point estimators and confidence regions about the parameters of QR models that allow for unknown form of heteroskedasticity. We further have shown that, if a simple correction is made, the smoothed EL confidence region can achieve higher order refinements, which are better than the refinements that might be obtained through the (smoothed) bootstrap approach.

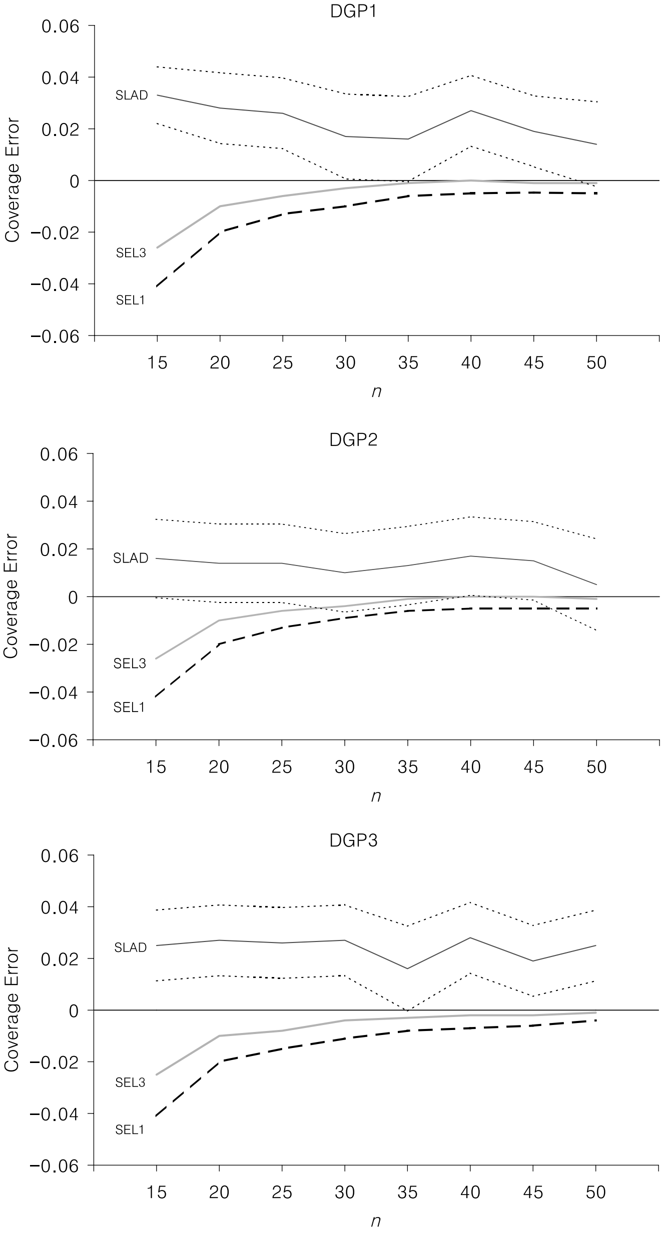

FIGURE 2. Coverage errors with varying sample sizes ($\alpha = 0.95$, $\gamma = -0.1$(SLAD), -0.9(SEL1, SEL3)).

The results of this paper can be extended in the following directions. First, this paper considers only the confidence region for the full parameter vector. If one is interested in a subvector $\beta_{10} \in \mathbb{R}^{K_1}$ of $\beta_0 = (\beta_{10}', \beta_{20}')'$, he or she may consider a confidence region after profiling out β_2, i.e.,

$$I_{hc,1} = \{\beta_1 : l_h(\beta_1, \tilde{\beta}_2(\beta_1)) \le c\},$$

where $\tilde{\beta}_2(\beta_1)$ minimizes $l_h(\beta_1, \beta_2)$ with respect to β_2 holding β_1 fixed. Under suitable assumptions and using an argument analogous to the proof of Theorem 2, it is not difficult to show that $l_h(\beta_{10}, \tilde{\beta}_2(\beta_{10})) \xrightarrow{d} \chi^2_{K_1}$, so that $I_{hc,1}$ has asymptotically correct coverage (in a first-order approximation) if c is chosen from the $\chi^2_{K_1}$ distribution. On the other hand, Bartlett correctability of $I_{hc,1}$ is a more challenging issue. Because the profiling requires an additional optimization step (which was not needed in the full parameter case), the Lagrange multiplier λ should now satisfy a first-order equation for this optimization problem, in addition to the first-order equation corresponding to (A.2). This implies that a new Edgeworth expansion for the distribution of $l_h(\beta_{10}, \tilde{\beta}_2(\beta_{10}))$ is needed to take into account the influence of the latter step. However, a systematic investigation of this expansion is beyond the scope of this paper and will be left to a future work.[12] In practice, one may consider the following bootstrap procedure to improve the finite-sample performance of $I_{hc,1}$.[13]

1. Using the original sample $\chi = \{(Y_i, X_i) : i = 1, \ldots, n\}$, compute $\hat{\beta} = (\hat{\beta}_1', \hat{\beta}_2')'$ by minimizing $l_h(\beta)$ with respect to β.
2. Draw bootstrap samples $\chi_b^* = \{(Y_{bi}^*, X_{bi}^*) : i = 1, \ldots, n\}$ for $1 \le b \le B$ randomly with replacement from the original sample χ.
3. 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 $l_{bh}^*(\hat{\beta}_1, \bar{\beta}_2)$, where $\bar{\beta}_2$ minimizes $l_{bh}^*(\hat{\beta}_1, \beta_2)$ with respect to β_2 holding $\hat{\beta}_1$ fixed.
4. Estimate the bootstrap Bartlett factor $\hat{b}_{1B}$ by solving the equation
$$B^{-1} \sum_{b=1}^{B} l_{bh}^*(\hat{\beta}_1, \bar{\beta}_2) = K_1(1 + n^{-1}\hat{b}_{1B}).$$
5. The Bartlett corrected confidence region is given by
$$I_{hc,1}^{\hat{b}_{1B}} = \{\beta_1 : l_h(\beta_1, \tilde{\beta}_2) \le c_{\alpha,1}(1 + n^{-1}\hat{b}_{1B})\}.$$

Finally, extensions to other econometric models with discontinuous estimating equations such as censored QR models and discrete choice models would be interesting future topics.

NOTES

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

2. Qin and Lawless (1994) link EL to general estimating equations for many interesting estimators. Recent econometric applications of the EL method include Bravo (2002, 2004), Donald, Imbens, and Newey (2003), Guggenberger and Smith (2003), Imbens, Spady, and Johnson (1998), Kitamura (1997, 2001), Kitamura and Stutzer (1997), Moon and Schorfheide (2003), Newey and Smith (2004), and Su and White (2003).

3. We do not claim here that EL is the only way of achieving such higher order refinements. An 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 EL is the only member of the Cressie–Read family that admits a Bartlett correction (see Jing and Wood, 1996; Baggerly, 1998). We expect that the same result will hold in our context under suitable assumptions.

4. The linearity assumption in (1) is not essential for our results. We expect that a similar result to ours will hold in nonlinear QR models with some additional work.

5. We note that the assumption of just-identification is important for our higher order results (especially Bartlett correction). The case of overidentification needs separate treatment and is beyond the scope of this paper; see Chen and Cui (2003) for an interesting result related to this issue. On the other hand, the first-order results in Theorem 1 can easily be shown to hold even under the overidentified case.

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

7. 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).

8. Another estimator of Λ_0 one might consider would be the negative of the second derivative of the log empirical likelihood, which should exist as a result of smoothing of the estimating equations.

9. This α-A notation was originally used by DiCiccio, Hall, and Romano (1991). For example, for $W = (W^1, \ldots, W^K)' \in \mathbb{R}^K$, we define $\alpha^j = EW^j, \alpha^{jj} = E(W^j)^2$, $\alpha^{jjkk} = E(W^j)^2(W^k)^2$, etc., for $1 \leq j,k \leq K$.

10. The Bartlett factor $\tilde{b}$ is derived in the same spirit (i.e., to remove the dependence on h) as β_0 in Chen and Hall (1993, Thm. 4.1). However, it is interesting to note that, in our case, $\tilde{b}$ depends not just on q but also on the covariates $\{X_i : i \geq 1\}$, in the regression framework.

11. See, e.g., van der Vaart and Wellner (1996, pp. 372–373) for the definition of Hadamard differentiability.

12. See Chen and Cui (2002, 2003) for some interesting results related to this issue in parametric models with no smoothing.

13. The idea of using a bootstrap procedure for Bartlett correction is originally due to Hall and LaScalar (1990). We extend their idea to account for the presence of nuisance parameters.

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 $O(n^{-2})$.

REFERENCES

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

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

Beran, R. (1988) Prepivoting test statistics: A bootstrap view of asymptotic refinements. *Journal of the American Statistical Association* 83, 687–697.

Bertail, P. (2003) Empirical Likelihood in Some Semiparametric Models. Manuscript, Laboratoire de Statistique, CREST.

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

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

Bravo, F. (2002) Testing linear restriction in linear models with empirical likelihood. *Econometrics Journal* 5, 104–130.

Bravo, F. (2004) Empirical likelihood based inference with applications to some econometric models. *Econometric Theory* 20, 231–264.

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

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

Chen, S.X. & H. Cui (2002) On Bartlett Correction of Empirical Likelihood in the Presence of Nuisance Parameters. Manuscript, Department of Statistics, Iowa State University.

Chen, S.X. & H. Cui (2003) On the Second Order Properties of Empirical Likelihood for Generalized Estimating Equations. Manuscript, Department of Statistics, Iowa State University.

Chen, S.X. & P. Hall (1993) Smoothed empirical likelihood confidence intervals for quantiles. *Annals of Statistics* 21, 1166–1181.

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

DiCiccio, T.J., P. Hall, & J.P. Romano (1991) Empirical likelihood is Bartlett-correctable. *Annals of Statistics* 19, 1053–1061.

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

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

Efron, B. (1982) *The Jackknife, the Bootstrap and Other Resampling Plans*. Society for Industrial Applied Mathematics.

Guggenberger, P. & R.J. Smith (2003) Generalized empirical likelihood estimators and tests under partial, weak and strong identification. Manuscript, Department of Economics, University of California at Los Angeles.

Hahn, J. (1995) Bootstrapping quantile regression estimators. *Econometric Theory* 11, 105–121.

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

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

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

Hall, P. & J.L. Horowitz (1990) Bandwidth selection in semiparametric estimation of censored linear regression models. *Econometric Theory* 6, 123–150.

Hall, P. & B. LaScalar (1990) Methodology and algorithms of empirical likelihood. *International Statistical Review* 58, 109–127.

Horowitz, J.L. (1997) Bootstrap methods in econometrics: Theory and numerical performance. In D. Kreps & K.W. Wallis (eds.), *Advances in Economics and Econometrics: Theory and Applications*, vol. 3, pp. 188–222. Cambridge University Press.

Horowitz, J.L. (1998) Bootstrap methods for median regression models. *Econometrica* 66, 1327–1351.

Horowitz, J.L. (2001) The bootstrap. In J.J. Heckman & E. Leamer (eds.), *Handbook of Econometrics*, vol. 5, pp. 3159–3228. North-Holland.

Imbens, G.W., R.H. Spady, & P. Johnson (1998) Information theoretic approaches to inference in moment condition models. *Econometrica* 66, 333–357.

Jing, B.-Y. & A.T.A. Wood (1996) Exponential empirical likelihood is not Bartlett correctable. *Annals of Statistics* 24, 365–369.

Kitamura, Y. (1997) Empirical likelihood methods with weakly dependent processes. *Annals of Statistics* 25, 2084–2102.

Kitamura, Y. (2001) Asymptotic optimality of empirical likelihood for testing moment restrictions. *Econometrica* 69, 1661–1672.

Kitamura, Y. & M. Stutzer (1997) An information-theoretic alternative to generalized method of moment estimation. *Econometrica* 65, 861–874.

Koenker, R. (2005) *Quantile Regression*. Cambridge University Press.

Koenker, R. & G. Bassett (1978) Regression quantiles. *Econometrica* 46, 33–50.

Koenker, R. & G. Bassett (1982) Robust tests for heteroskedasticity based on regression quantiles. *Econometrica* 50, 43–61.

Monti, A.C. (1997) Empirical likelihood confidence regions in time series models. *Biometrika* 84, 395–405.

Moon, H.R. & F. Schorfheide (2003) Empirical Likelihood Estimation Based on Inequality Constraints. Manuscript, Department of Economics, University of Southern California.

Newey, W.K. & R.J. Smith (2004) Higher order properties of GMM and generalized empirical likelihood estimators. *Econometrica* 72, 219–256.

Otsu, T. (2003) Empirical Likelihood for Quantile Regression. Manuscript, Department of Economics, University of Wisconsin.

Owen, A. (1988) Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* 75, 237–249.

Owen, A.B. (1990) Empirical likelihood ratio confidence regions. *Annals of Statistics* 18, 90–120.

Owen, A.B. (1991) Empirical likelihood for linear models. *Annals of Statistics* 19, 1725–1747.

Owen, A.B. (2001) *Empirical Likelihood*. Chapman and Hall.

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

Powell, J.L. (1986) Censored regression quantiles. *Journal of Econometrics* 32, 143–155.

Qin, J. & J. Lawless (1994) Empirical likelihood and general estimating equations. *Annals of Statistics* 22, 300–325.

Su, L. & H. White (2003) Testing Conditional Independence via Empirical Likelihood. Manuscript, Department of Economics, University of California at San Diego.

van der Vaart, A.W. & J.A. Wellner (1996) *Weak Convergence and Empirical Processes*. Springer-Verlag.

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

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

APPENDIX

Here we sketch the proof of the theorems in the main text. Some of the details of the proofs are available in our working paper version.

LEMMA 1. *Under Assumptions 1–5(b) and 6(a), we have, as $n \to \infty$,*

(a) $EZ_i(\beta_0) = (-h)^r (r!)^{-1} C_K E[X_i f^{(r-1)}(0|X_i)] + o(h^r)$,

(b) $EZ_i(\beta_0)Z_i(\beta_0)' = q(1-q)S_0 + o(1)$,

(c) $E\dfrac{\partial Z_i(\beta_0)}{\partial \beta'} = D_0 + o(1)$.

Proof of Lemma 1. The results of Lemma 1 hold using a standard technique to obtain biases and variances of kernel density estimators. ■

LEMMA 2. *Suppose Assumptions 1–5(b) and 6(a) hold. Then, as* $n \to \infty$, *we have (a)* $(1/n)\sum_{i=1}^{n} Z_i(\beta) = O(d_n)$ *a.s., (b)* $(1/n)\sum_{i=1}^{n} Z_i(\beta) Z_i(\beta)' = q(1-q)S_0 + o(1)$ *a.s., and (c)* $t(\beta) = O(d_n)$ *a.s. uniformly in* $\beta \in B_n \equiv \{\beta : \|\beta - \beta_0\| \le d_n\}$, *where* $t(\beta)$ *satisfies (7),* $d_n = n^{-1/3-\delta}$, *and* $0 < \delta < \frac{1}{6}$.

Proof of Lemma 2. The proof is similar to Owen (1990, proof of Thm. 1) and Qin and Lawless (1994, proof of Lem. 1). ■

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* $\hat{\beta}_E \in int(B)$ *such that* $l_h(\beta)$ *attains its minimum value at* $\hat{\beta}_E$ *and (b)* $\hat{\beta}_E$ *satisfies* $t(\hat{\beta}_E) = 0$ *and* $Q_n(\hat{\beta}_E) = 0$, *where* $Q_n(\beta) = n^{-1}\sum_{i=1}^{n} Z_i(\beta)$.

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. ■

Proof of Theorem 1. By Lemma 1 and the weak law of large numbers, we have $\partial Q_n(\beta_0)/\partial\beta' \xrightarrow{p} D_0$. Letting $G_{ni} \equiv [G(-U_i/h) - 1(U_i \le 0)]$ and rearranging terms, we rewrite $\sqrt{n}Q_n(\beta_0)$ as

$$\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[1(U_i \le 0) - q]X_i + \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[G_{ni}X_i - EG_{ni}X_i] + \sqrt{n}EG_{ni}X_i. \qquad \textbf{(A.1)}$$

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

$$P\left(\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}[G_{ni}X_i - EG_{ni}X_i]\right\| > \varepsilon\right) \le \varepsilon^{-2}E\left[G\left(\frac{-U_i}{h}\right)^2 1(|U_i| \le h)\right]\|X_i\|^2$$
$$= O(h),$$

which is $o(1)$, using Assumptions 3, 4(b), and 5(a). Also, the last term in (A.1) is also $o(1)$ using Assumption 6(a) because $\sqrt{n}EG_{ni}X_i = \sqrt{n}EZ_i(\beta_0) = O(n^{1/2}h^r) \to 0$. Now, by expanding $Q_n(\hat{\beta}_E)$ at β_0, we have the desired result. ■

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

$$\frac{1}{n}\sum\frac{W_i}{1+\lambda' W_i} = 0. \qquad \textbf{(A.2)}$$

Then, we have

$$\lambda = O_p(n^{-1/2} + h^r) \qquad \textbf{(A.3)}$$

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

Next we develop a Taylor expansion for λ and $l_h(\beta_0)$. By (A.2), we have

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

By Lemma 1(a), we have $\alpha^j = O(h^r)$. Also, observe that $\bar{A}^j = A^j + \alpha^j = O_p(n^{-1/2} + h^r)$, $A^{jk} = O_p(n^{-1/2})$, and $\bar{A}^{j_1\cdots j_k} = O_p(1)$ for $k \geq 3$. Solving for λ and then recursive substitutions in equation (A.4) give, for each $L \geq 1$,

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

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

Similarly, we have

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

Therefore, for any $k > 1$,

$$l_h(\beta_0) = n\bar{A}^j\bar{A}^j + O_p(n(n^{-1/2} + h^r)^k) \tag{A.6}$$

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

$$+ O_p(n(n^{-1/2} + h^r)^k).$$

Because $V_n - \mathrm{Var}(Z_i) \to 0$, $[\mathrm{Var}(Z_i)]^{-1/2}\cdot n^{-1/2}\sum(Z_i - EZ_i) \xrightarrow{d} N(0, I_K)$, 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. ■

Let

$$\bar{Q} = (A^1, \ldots, A^K, A^{11}, \ldots, A^{KK}, \ldots, A^{11\cdots 1}, \ldots, A^{KK\cdots K})' = \frac{1}{n}\sum_{i=1}^{n} Q_i \tag{A.7}$$

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

$$(G(-U_i/h) - q)^{|\nu|}X_i^{\nu_1}\ldots X_i^{\nu_k} \quad \text{for } 1 \le k \le L + 1, \tag{A.8}$$

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

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.7) and $i = (-1)^{1/2}$. Under Assumptions 1–6, we have, for each $\varepsilon > 0$, that there exists some $C(\varepsilon) > 0$ such that*

$$\sup_{\|t\|>\varepsilon} |I(t,h)| < 1 - C(\varepsilon)h \quad \textit{for all sufficiently small } h.$$

Proof of Lemma 4. The proof of Lemma 4 is analogous to those of Horowitz (1998, Lem. 9) and Hall (1992, Lem. 5.6). First, collect terms with the same polynomial order, so that we have

$$I(t,h) = E\{\exp[it'Q]\}$$

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

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

$$= I_1(t,h) + I_2(t,h), \quad \text{say,}$$

where $g_r(X)$ is a vector of the products of elements of X that correspond to the rth-order polynomial $[G(-U/h)]^r$ in the expansion of $t'Q$ and $\tau_r(t)$ denotes the corresponding subvector of $t \in \mathbb{R}^{\dim(Q)}$. For h sufficiently small, by a two-term Taylor expansion, we have

$$|I_1(t,h)| \leq 1 - Ef(0|X)h. \tag{A.9}$$

Furthermore, given $\varepsilon > 0$, there exists $\delta(\varepsilon) > 0$ such that

$$\int_{-\infty}^{\infty}\int_{-1}^{1} |f(hu|x) - f(0|x)|\, du\, dP(x) \leq 2\delta(\varepsilon)Ef(0|X) \quad \text{for } h \text{ sufficiently small.}$$

Take $\eta > 0$ and $\gamma_1 < 1$ such that $\int_{\|x\|\leq\eta} f(0|x)\, dP(x) = \gamma_1 Ef(0|X)$. Then, by a change of variables and triangle inequality, we have

$$|I_2(t,h)| \leq (2\delta(\varepsilon) + \gamma_1)hEf(0|X) + h\int_{\|x\|>\eta} \Psi(t,x)f(0|x)\, dP(x), \tag{A.10}$$

where $\Psi(t,x) = \int_{-1}^{1} \exp[i\sum_{r=0}^{L+1}[G(u)]^r\tau_r(t)'g_r(x)]\, du$. By partitioning of $[-1,1]$ into subsets $\{(a_{l-1},a_l): l = 1,\ldots,L+1\}$ as in Assumption 5(c) and using a change of variables argument similar to Horowitz (1998, pp. 1346–1347), we can show that

$$C_1(\varepsilon) \equiv \sup_{\|x\|>\eta}\ \sup_{\|t\|>\varepsilon} |\Psi(t,x)| < 1. \tag{A.11}$$

Now, by combining (A.9)–(A.11), we have, for each $\varepsilon > 0$,

$$\sup_{\|t\|>\varepsilon} |I(t,h)| \leq 1 - \{1 - 2\delta(\varepsilon) - [\gamma_1 + (1-\gamma_1)C_1(\varepsilon)]\}Ef(0|X)h = 1 - C(\varepsilon)h$$

for all $h > 0$ sufficiently small. Note that $C(\varepsilon) > 0\ \forall \varepsilon > 0$ because $[\gamma_1 + (1-\gamma_1) C_1(\varepsilon)] < 1$ and $\delta(\varepsilon)$ can be made arbitrarily small by choosing h sufficiently small. This completes the proof of Lemma 4. ■

Proof of Theorem 3. We first derive the signed root of $l_h(\beta_0)$ in (A.5), which is a K-dimensional vector $n^{1/2}S_{0L} = n^{1/2}(S_{0L}^1,\ldots,S_{0L}^K)'$ 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 $T_l = O_p((n^{-1/2} + h^r)^l)$ and $U_{1L} = O_p((n^{-1/2} + h^r)^{L+1})$. Some calculations yield

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

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

$$T_3^j = \frac{3}{8} A^{jm} A^{km} \bar{A}^k + \frac{4}{9} \bar{A}^{jkn} \bar{A}^{lmn} \bar{A}^m \bar{A}^k \bar{A}^l - \frac{1}{4} \bar{A}^{jklm} \bar{A}^m \bar{A}^k \bar{A}^l$$

$$- \frac{5}{12} \bar{A}^{jkm} A^{lm} \bar{A}^k \bar{A}^l - \frac{5}{12} \bar{A}^{klm} A^{jm} \bar{A}^k \bar{A}^l,$$

$$T_4^j = \frac{11}{16} A^{rk} A^{rj} A^{kl} \bar{A}^l$$

$$- \frac{53}{48} \bar{A}^{rkj} A^{rm} A^{kp} \bar{A}^m \bar{A}^p - \frac{53}{48} \bar{A}^{rkl} A^{rj} A^{kp} \bar{A}^p \bar{A}^l - \frac{7}{6} \bar{A}^{rkj} A^{rm} A^{mo} \bar{A}^o \bar{A}^k$$

$$- \frac{7}{6} \bar{A}^{rkl} A^{rm} A^{mj} \bar{A}^k \bar{A}^l$$

$$+ \frac{229}{108} \bar{A}^{rjl} \bar{A}^{lmo} A^{rp} \bar{A}^m \bar{A}^o \bar{A}^p + \frac{229}{108} \bar{A}^{rkl} \bar{A}^{ljo} A^{rp} \bar{A}^k \bar{A}^o \bar{A}^p + \frac{229}{108} \bar{A}^{rkl} \bar{A}^{lmo} A^{rj} \bar{A}^k \bar{A}^m \bar{A}^o$$

$$+ \frac{59}{36} \bar{A}^{rjl} \bar{A}^{mop} A^{rm} \bar{A}^o \bar{A}^p \bar{A}^l - \frac{25}{16} \bar{A}^{rjlm} A^{rp} \bar{A}^p \bar{A}^l \bar{A}^m - \frac{25}{16} \bar{A}^{rklm} A^{rj} \bar{A}^k \bar{A}^l \bar{A}^m$$

$$+ \frac{49}{24} \bar{A}^{rjlm} \bar{A}^{rpq} \bar{A}^l \bar{A}^m \bar{A}^p \bar{A}^q + \frac{49}{24} \bar{A}^{rklm} \bar{A}^{rjq} \bar{A}^k \bar{A}^l \bar{A}^m \bar{A}^q - \frac{56}{27} \bar{A}^{rjl} \bar{A}^{rpq} \bar{A}^{qmo} \bar{A}^p \bar{A}^m \bar{A}^o \bar{A}^l$$

$$- \frac{56}{27} \bar{A}^{rkl} \bar{A}^{rjq} \bar{A}^{qmo} \bar{A}^m \bar{A}^o \bar{A}^k \bar{A}^l - \frac{4}{5} \bar{A}^{jklmo} \bar{A}^k \bar{A}^l \bar{A}^m \bar{A}^o.$$

Also, by choosing L sufficiently large, we can ensure that

$$P(\|U_{1L}\| > n^{-5/2}) = O(n^{-2}).$$

Hence, for $c > 0$, we have

$$P(l_h(\beta_0) \le c) = P[n^{1/2}\|S_L + U_{1L}\| \le c^{1/2}],$$

and so

$$\max_{+,-} |P(l_h(\beta_0) \le c) - P(n^{1/2}\|S_L\| \le c^{1/2} \pm n^{-2})| = O(n^{-2}). \qquad \textbf{(A.12)}$$

We now develop an Edgeworth expansion for the distribution of $S_{nL} = n^{1/2}S_L$.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:

$$\kappa^{(j)} = n^{1/2}\alpha^j - n^{-1/2}\left(\frac{1}{6}\alpha^{jkk}\right) + O(n^{-1/2}h^r + n^{-3/2}), \qquad \textbf{(A.13)}$$

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

$$\kappa^{(i,j,k)} = O(n^{-1/2}h^{r}), \qquad \kappa^{(i,j,k,l)} = O(n^{-1}h^{2r}),$$

$$\kappa^{(j_1,j_2,\ldots,j_m)} = O(n^{-(m-2)/2}) \quad \text{for } m \geq 5.$$

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

$$\sup_{B\in\mathcal{B}} \int_{(\partial B)^{\varepsilon}} \phi_{0,I}(x)\,dx = O(\varepsilon) \quad \text{as } \varepsilon \downarrow 0, \tag{A.14}$$

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 $nh^{2r} \to 0$,

$$\sup_{B\in\mathcal{B}} \left| P(n^{1/2}S_L \in B) - \int_B p(x)\phi_{0,I}(x)\,dx \right| = O(n^{-2}) + o(nh^{2r}), \tag{A.15}$$

where

$$p(x) = 1 + p_1(x) + p_2(x),$$

$$p_1(x) = \frac{1}{2}\,n^{-1}\{x'\Delta x - \mathrm{tr}(\Delta)\},$$

$$p_2(x) = \text{odd polynomial in } x, \tag{A.16}$$

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

$$\Delta^{ij} = n^2\alpha^i\alpha^j + n\left\{\frac{1}{3}\alpha^{ijk}\alpha^k - \frac{1}{6}\alpha^{ikk}\alpha^j - \frac{1}{6}\alpha^{jkk}\alpha^i\right\}$$

$$+ \frac{1}{2}\alpha^{ijkk} - \frac{1}{3}\alpha^{ikm}\alpha^{jkm} - \frac{1}{36}\alpha^{ijm}\alpha^{mkk} + \frac{1}{36}\alpha^{ikk}\alpha^{jll}. \qquad \textbf{(A.17)}$$

Accepting that the Edgeworth expansion (A.15) is justified, we now develop an Edgeworth expansion for the distribution of $l_h(\beta_0)$. From (A.12), we have, for any $c > 0$,

$$\begin{aligned} P(l_h(\beta_0) \le c) &= \int_{\|x\|<c^{1/2}} p(x)\phi_{0,I}(x)\,dx + O(n^{-2}) + o(nh^{2r}) \\ &= P(\chi_K^2 \le c) \\ &\quad + \frac{1}{2}n^{-1}\int_{\|x\|<c^{1/2}}\left\{\sum_{i=1}^{K}\Delta^{ii}[(x^i)^2-1] - \sum_{i\ne j}\Delta^{ij}x^ix^j\right\}\phi_{0,I}(x)\,dx \\ &\quad + O(n^{-2}) + o(nh^{2r}) \\ &= P(\chi_K^2 \le c) - n^{-1}\mathrm{tr}(\Delta)K^{-1}cg_K(c) + O(n^{-2}) + o(nh^{2r}), \end{aligned} \qquad \textbf{(A.18)}$$

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

$$\mathrm{tr}(\Delta) = n^2\alpha^i\alpha^i + \frac{1}{2}\alpha^{iikk} - \frac{1}{3}\alpha^{ikm}\alpha^{ikm}. \qquad \textbf{(A.19)}$$

Let $\zeta \equiv E[Xf^{(r-1)}(0|X)]$. Then, using (A.12) and Lemma 1, we have

$$\begin{aligned} n^2\alpha^i\alpha^i &= n^2(EZ)'V_n^{-1}(EZ) \\ &= (nh^r)^2(r!)^{-2}C_K^2(\zeta'S^{-1}\zeta)q^{-1}(1-q)^{-1} + o((nh^r)^2). \end{aligned} \qquad \textbf{(A.20)}$$

Similarly, we have

$$\begin{aligned} \alpha^{iikk} &= E\{[G_h(-U)-q]^4(X'V_n^{-1}X)^2\} \\ &= q^{-1}(1-q)^{-1}(1-3q+3q^2)E\{(X'SX)^2\} + O(h) < \infty \end{aligned} \qquad \textbf{(A.21)}$$

and

$$\alpha^{ikm} = E\{[G_h(-U)-q]^3(v_{ni}^{-1/2}X)(v_{nk}^{-1/2}X)(v_{nm}^{-1/2}X)\} \qquad \textbf{(A.22)}$$

$$q^{-1/2}(1-q)^{-1/2}(1-2q)E\{(s_i^{-1/2}X)(s_k^{-1/2}X)(s_m^{-1/2}X)\} + O(h)$$

$$< \infty,$$

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

Therefore, (A.18)–(A.22) give

$$\begin{aligned} &P(l_h(\beta_0) \le c_\alpha) \\ &\quad = \alpha - n^{-1}\{(nh^r)^2 (r!)^{-2} C_K^2 (\zeta' S^{-1}\zeta) q^{-1}(1-q)^{-1} + O(1)\} K^{-1} c_\alpha g_K(c_\alpha) \\ &\qquad + o(n^{-1} + nh^{2r}). \end{aligned} \tag{A.23}$$

It now follows that, because $\sup_n nh^r < \infty$, we have $P(l_h(\beta_0) \le c_\alpha) = \alpha + O(n^{-1})$, as desired.

It remains to check that the formal expansion (A.15) is valid. For this purpose, we first need the following formal Edgeworth expansion for the distribution of $n^{1/2}\bar{Q}$: for each $m \ge 1$,

$$\sup_{A \in \mathcal{A}} \left| P(n^{1/2}\bar{Q} \in A) - \int_A \sum_{k=0}^{m} n^{-k/2} q_k(x) \phi_{0,\Sigma}(x)\, dx \right| = O(n^{-(m+1)/2}), \tag{A.24}$$

where $\mathcal{A}$ denote a class of Borel sets $A \subseteq \mathbb{R}^d$ for $d = \dim(Q)$ that satisfy $\sup_{A \in \mathcal{A}} \int_{(\partial\mathcal{A})^\varepsilon} \exp(-\frac{1}{2}\|x\|^2)\, dx = O(\varepsilon)$ as $\varepsilon \downarrow 0$, $\Sigma = \mathrm{Var}(n^{1/2}\bar{Q})$, and $q_k(x)\phi_{0,\Sigma}(x)$ is the density of the finite-signed measure whose Fourier–Stieltjes transform is $\exp(-t'\Sigma t/2)P_k(t)$. Here, the polynomial $P_k(t)$ is defined by the expansion

$$\exp\left[\sum_{l=0}^{\infty} \frac{(-1)^l}{u^2(l+1)} \left\{ \sum_{|r|=2}^{\infty} \frac{(it)^r}{r!} (EQ^r) u^r \right\}^{l+1} \right] = \exp\left(-\frac{1}{2} t'\Sigma t \right) \left\{ 1 + \sum_{k=1}^{\infty} P_k(t) u^k \right\},$$

where $r = (r_1, \dots, r_d)' \in \mathbb{R}^d$ denotes a vector of nonnegative integers, $Q^r = (Q^1)^{r_1} \dots (Q^d)^{r_d}$, $r! = r_1! \dots r_d!$, and $t = (t_1, \dots, t_d)' \in \mathbb{R}^d$. The result (A.24) holds by an argument very similar to the proof of Theorem 19.2 of Bhattacharya and Rao (1976), provided we establish an analogue of Cramér's condition (i.e., their eqn. (19.25) on p. 193). Because the latter is already established in Lemma 4, we have the result (A.24).

Next, write $n^{1/2}S_L = g_n(n^{1/2}\bar{Q})$, where $g_n : \mathbb{R}^d \to \mathbb{R}^K$ is a mapping. Then, because $g_n(\cdot)$ is continuous, we have

$$P(n^{1/2}S_L \in B) = \int_{\{g_n(x) \in B\}} \sum_{k=0}^{3} n^{-k/2} q_k(x) \phi_{0,\Sigma}(x)\, dx + O(n^{-2}), \tag{A.25}$$

uniformly over $B \in \mathcal{B}$, using (A.24) and the result of Bhattacharya and Ghosh (1978, eqns. (2.16)–(2.19), pp. 444–445). By Lemma 2.1 of the latter paper, there exists a polynomial $\tilde{p}_n(x)$ such that

$$\int_{\{g_n(x) \in B\}} \sum_{k=0}^{3} n^{-k/2} q_k(x) \phi_{0,\Sigma}(x)\, dx = \int_B \tilde{p}(x) \phi_{0,I}(x)\, dx + O(n^{-2}) \tag{A.26}$$

uniformly over $B \in \mathcal{B}$. Therefore, it remains to identify $\tilde{p}(x)$. Using the cumulants in (A.13), we can approximate (by a Taylor expansion) the characteristic function $\Psi_n(t)$ of $n^{1/2}S_L$ up to order $O(n^{-2})$ uniformly over $\{t : \|t\| \le 1\}$ and then take an inverse Fourier transformation of the approximation to get the desired result (A.15). This completes the proof of Theorem 3. ■

Proof of Theorem 4. By (A.18), we have for all $c > 0$

$$P(l_h(\beta_0) \leq c(1+n^{-1}b)) = P(\chi_K^2 \leq c(1+n^{-1}b)) - c\{n\alpha^i\alpha^i K^{-1} + n^{-1}b\}\{1+n^{-1}b\}g_K[c(1+n^{-1}b)] + O(n^{-2}) + o(nh^{2r}). \quad \textbf{(A.27)}$$

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

$$g_K[c(1+n^{-1}b)] = g_K(c) + O(n^{-1}) \quad \text{and} \quad \textbf{(A.28)}$$

$$P(\chi_K^2 \leq c(1+n^{-1}b)) = P(\chi_K^2 \leq c) + cn^{-1}bg_K(c) + O(n^{-2}). \quad \textbf{(A.29)}$$

By substituting (A.28) and (A.29) into (A.27), we have

$$\begin{aligned} P(l_h(\beta_0) &\leq c(1+n^{-1}b)) \\ &= P(\chi_K^2 \leq c) - cn\alpha^i\alpha^i K^{-1}g_K(c) + O(n^{-2}) + o(nh^{2r}) \\ &= P(\chi_K^2 \leq c) - nh^{2r}(r!)^{-2}C_K^2(\zeta'\Sigma^{-1}\zeta)q^{-1}(1-q)^{-1}\cdot cK^{-1}g_K(c) \\ &\quad + O(n^{-2}) + o(nh^{2r}), \end{aligned} \quad \textbf{(A.30)}$$

where the second equality follows from (A.20). Therefore, $\sup_n n^3h^{2r} < \infty$ implies that

$$P(l_h(\beta_0) \leq c(1+n^{-1}b)) = P(\chi_K^2 \leq c) + O(n^{-2}) \quad \textbf{(A.31)}$$

for all $c > 0$. The proof of Theorem 3 is complete by taking $c = c_\alpha$ in (A.31).

The case where b is replaced by $\hat{b}$ or $\tilde{b}$ may be treated in a similar way using the fact that $\hat{b} = b + O_p(n^{-1/2})$ and the parity properties of polynomials in Edgeworth expansions such as (A.16). ■