Empirical likelihood for quantile regression

Taisuke Otsu[*†]
Department of Economics
University of Wisconsin-Madison

November 2003
Job Market Paper

Abstract

We propose new estimation and inference methods for quantile regression models based on the method of empirical likelihood and its extensions. We consider four concepts of nonparametric likelihood—conditional empirical likelihood (CEL), smoothed conditional empirical likelihood (SCEL), usual empirical likelihood (EL), and smoothed empirical likelihood (SEL)—and investigate the statistical properties of the derived estimators and test statistics. Our extensions to the empirical likelihood approach effectively deal with several problems of existing quantile regression estimation and inference methods, such as the efficiency of the estimators, variance estimation to construct confidence sets, and higher order refinements of confidence sets. In order to avoid practical and technical problems of non-smooth objective functions, we introduce kernel smoothing on quantile restrictions. As extensions, we consider multiple quantile regression models, tests for homoskedasticity and symmetry, confidence sets without parameter estimation, and consistent specification tests for quantile regression models.

JEL classification: C14; C21

Keywords: Quantile regression; Empirical likelihood

[*]E-mail: totsu@wisc.edu Website: http://www.ssc.wisc.edu/∼totsu/

[†]The author is deeply grateful to Bruce Hansen, Philip Haile, John Kennan, Yuichi Kitamura, and Gautam Tripathi for guidance and time. The author also thanks Meta Brown, Matthew Kim, and James Walker for helpful suggestions. Financial support from the Alice Gengler Wisconsin Distinguished Graduate Fellowship and Wisconsin Alumni Research Foundation Dissertation Fellowship is gratefully acknowledged.

1 Introduction

This paper studies new estimation and inference methods for quantile regression models based on the method of empirical likelihood and its extensions. Our extensions to the empirical likelihood approach effectively deal with several problems of existing quantile regression estimation and inference methods, such as the efficiency of the estimators, variance estimation to construct confidence sets, and higher order refinements of confidence sets. In order to avoid practical and technical problems of non-smooth objective functions, we introduce kernel smoothing on quantile restrictions. We consider four concepts of nonparametric likelihood—conditional empirical likelihood (CEL), smoothed conditional empirical likelihood (SCEL), usual empirical likelihood (EL), and smoothed empirical likelihood (SEL)—and investigate statistical properties of the derived estimators and test statistics. Each method has different advantages and disadvantages compared to conventional estimation and inference methods. Particularly, (i) the CEL and SCEL estimators are asymptotically efficient; (ii) all of the EL-based test statistics provide valid confidence sets without estimating the variances of estimators; (iii) SCEL- and SEL-based estimation and inference can be conducted by a Newton-type algorithm; (iv) SEL is Bartlett correctable and provides higher order refinement of the confidence sets; (v) however, CEL, SCEL, and SEL require some kernel smoothing, in which the choices of the kernel function and bandwidth may be arbitrary.

Since the seminal works of Koenker and Bassett (1978a,b), quantile regression has become a standard tool of empirical economic analysis, particularly in the fields of labor and public economics.[1] A familiar special case of quantile regression is the least absolute deviation (LAD) regression, in which the quantile of interest is the median. Since the distributional form of the error term is unspecified except for the conditional quantile restriction, quantile regression is regarded as a semiparametric model, which is robust to misspecification of the distributional form of the error term. There are five important features to consider when analyzing quantile regression models.

(i) Efficiency: Koenker and Bassett's (1978a) conventional quantile regression estimator is not efficient under the conditional quantile restriction, which is an asymptotically equivalent representation of a quantile regression model. Indeed, the conventional estimator is based on an unconditional moment restriction, which is an implication of the conditional quantile restriction. Based on the efficient score of quantile regression models, Newey and Powell (1990) proposed an efficient quantile regression estimator. However, implementation of Newey and Powell's (1990) efficient estimator requires a sample

[1]See Buchinsky (1998) for a review of quantile regression. For empirical applications of quantile regression, see the special issue of *Empirical Economics* (vol. 26:1).

splitting device for estimating the efficient score and estimation of the optimal weight, which is the conditional error density function evaluated at zero.

(ii) Variance estimation: Since the asymptotic variances of quantile regression estimators contain the conditional error density function evaluated at zero, variance estimation for constructing confidence sets is an important issue. The Wald-type test statistics and confidence sets (i.e., estimate $\pm$ 1.96 $\times$ standard error) require some variance estimator. Several variance estimation methods are proposed and compared, such as the order statistic, kernel smoothing, and bootstrap methods.[2]

(iii) Non-smooth moment restriction: Since the moment restrictions implied by quantile regression models contain indicator functions, we need to deal with non-smooth objective functions for estimation and inference. Thus, we cannot apply the usual argument based on Taylor expansions for discussing the asymptotic properties of estimators and test statistics, particularly for higher order properties.

(iv) Algorithm: Due to the non-smoothness of the implied moment restrictions, the choice of an algorithm to implement quantile regression estimation is a substantial practical issue. Koenker and Bassett's (1978a) conventional estimator employs a linear programming algorithm, which is stable and globally convergent by a finite number of iterations. We could use some generalized method of moments (GMM) type estimation to improve efficiency by including additional moment restrictions. However, in this case, we would not have such a linear programming representation for the optimization problem of the estimator. Thus, we would have to apply some non-derivative optimization algorithm, such as Nelder and Mead's (1965) simplex method or simulated annealing.

(v) Censoring: In contrast to the conditional mean restrictions of mean regression models, the conditional quantile restriction is useful for identifying censored regression (or Tobit) models. Since Powell (1984, 1986), several semiparametric methods for censored regression models are provided by using conditional quantile restrictions.

This paper considers uncensored quantile regression models and therefore deals with (i)-(iv). Extension to censored regression models is an important topic for future research.[3]

In this paper, we apply the method of EL by Owen (1988, 1990, 1991) and CEL by Kitamura, Tripathi, and Ahn (2003) and Zhang and Gijbels (2003) to quantile regression models, and propose new estimation

[2]See Buchinsky (1995b) and Koenker (1994) for simulation results and comparisons of these methods.
[3]The author is currently working on this extension, which introduces additional kernel smoothing for the indicator function by censoring.

and inference methods.[4] EL is a data-driven nonparametric method of estimation and inference for moment restriction models, which does not require weight matrix estimation like GMM and is invariant to linear transformations of moment restrictions.[5] Qin and Lawless (1994) showed that the EL estimator is asymptotically (first-order) equivalent to the optimally weighted GMM estimator by Hansen (1982), and that the EL ratio test statistics for parameter restrictions has the chi-square limiting distribution. Newey and Smith (2003) and Kitamura (2001) derived desirable properties of the EL approach from the viewpoints of higher-order bias and large deviation properties, respectively. CEL is an extension of EL to attain the efficiency bound for conditional moment restriction models, which imply infinitely many unconditional moment restrictions. LeBlanc and Crowley (1995) proposed a local likelihood approach to construct nonparametric likelihood for conditional moment restriction models. Although LeBlanc and Crowley (1995) showed that their approach is applicable to quantile regression by a numerical example, they did not provide any formal statistical theory. Kitamura, Tripathi, and Ahn (2003) assumed sufficiently smooth moment restrictions, and derived the consistency, asymptotic normality, and efficiency of the CEL estimator. Therefore, the CEL estimator is asymptotically equivalent to the optimal instrumental variable GMM estimator by Newey (1990, 1993), which attains the semiparametric efficiency bound derived by Chamberlain (1987). Zhang and Gijbels' (2003) setup allows for non-smooth moment restrictions and nonparametric regression models.[6] In contrast to Zhang and Gijbels (2003), Kitamura, Tripathi, and Ahn (2003) proposed the CEL ratio test statistic for parameter restrictions and derived its chi-square limiting distribution. These previous results show that EL and CEL have similar properties as usual parametric likelihood. Kitamura (2003) extended the CEL approach to possibly misspecified models, proposed the CEL-based measure of fit for conditional models, and discussed quantile regression models as an example.

In the quantile regression setup, we first show the consistency, asymptotic normality, and efficiency of the CEL estimator. An important advantage of the CEL estimator is that we do not need to estimate the optimal weight of Newey and Powell's (1990) efficient estimator; the optimal weight is automatically incorporated in CEL. In addition, we derive the chi-square limiting distribution of the CEL ratio test statistic for parameter restrictions. However, in contrast to the conventional inefficient quantile regression estimator by Koenker and Bassett (1978a), the optimization problem for CEL estimation does not have a linear programming

[4]A note on terminology. CEL is called "smoothed" and "sieve" empirical likelihood in Kitamura, Tripathi, and Ahn (2003) and Zhang and Gijbels (2003), respectively. Since we introduce additional smoothing on moment restrictions, we describe their method as "conditional" empirical likelihood in order to avoid confusion, which is adopted by Kitamura (2003).

[5]See Owen (2001) for a review of empirical likelihood.

[6]Otsu (2003a) extended the CEL approach to semiparametric models, i.e., conditional moment restriction models including unknown functions, and proposed the penalized empirical likelihood estimator (PELE).

representation. Since CEL is non-smooth with respect to unknown parameters, we must use some non-derivative optimization algorithm, which tends to have multiple local optima and converge slowly, for the implementation of CEL-based method. To solve this practical problem, we introduce kernel smoothing to the conditional quantile restriction. By replacing the conditional quantile restriction with a smoothed counterpart, we derive the SCEL estimator and SCEL ratio test statistic. Since the SCEL objective function is smooth with respect to unknown parameters, we can apply a popular Newton-type optimization algorithm. We show that the SCEL estimator and SCEL ratio test statistic are asymptotically equivalent to the CEL estimator and CEL ratio test statistic, respectively. Furthermore, by inverting the CEL or SCEL ratio test statistic, we can construct valid confidence sets for unknown parameters. In contrast to the conventional Wald-type confidence sets, both the CEL- and SCEL-based confidence sets avoid variance estimation for constructing confidence sets. While the Wald-type confidence set relies on a local quadratic approximation of likelihood, shapes of the CEL- and SCEL-based confidence sets are determined by observed data and are not necessarily symmetric around the estimates.

Higher order refinement of confidence sets is another reason for smoothing the moment restrictions. Due to technical difficulties for analyzing higher order properties in the conditional quantile restriction setup, we consider an unconditional quantile restriction, which is an implication of the conditional quantile restriction. In order to obtain higher order refinements for the EL ratio we must use Taylor series approximations, which require sufficiently smooth moment restrictions. For the LAD regression model, Horowitz (1998) considered a kernel-smoothed objective function, and derived higher order refinements of the t and Wald test statistics by bootstrapping. For distribution quantiles, Chen and Hall (1993) employed a smoothed moment restriction and obtained the Bartlett correction of their EL ratio test statistic. From the smoothed counterpart of the unconditional quantile restriction, we propose SEL, and derive the Bartlett correction of the SEL ratio test, which is an extension of Chen and Hall's (1993) result to the quantile regression setup. Using the Bartlett correction, the rejection probability error of the SEL ratio test becomes $O(n^{-2})$, which is better than the conventional rejection probability error, $O(n^{-1})$. Similarly to the cases of CEL and SCEL, the EL- and SEL-based confidence sets do not require variance estimation, and shapes of the confidence sets are automatically determined by data. After the completion of this draft, the author was informed that Whang (2003) independently derived similar results, i.e., the Bartlett correction for the SEL ratio test. While Whang's (2003) main purpose is to compare SEL to the bootstrap, this paper merely intends to provide a motivation for smoothing the quantile restriction and focuses mainly on the comparison to the conventional method (for detailed discussion, see Section 5).

As extensions, we consider multiple quantile regression models, tests for homoskedasticity and symmetry of error terms, confidence sets without parameter estimation, and consistent specification tests for quantile regression models. The CEL- or SCEL-based tests for homoskedasticity and symmetry are convenient tools for analyzing the distributional form of error terms. The CEL- or SCEL-based confidence set without parameter estimation, which is an extension of Tripathi and Kitamura's (2001) canonical version of the CEL ratio test statistic, is easy to implement if the number of unknown parameters is small. The CEL- or SCEL-based consistent specification test for quantile regression models, which is an extension of Tripathi and Kitamura's (2002) CEL ratio specification test statistic, is an important diagnostic statistic to check the validity of specification of the quantile regression models.

This paper is organized as follows. Section 2 describes the basic setup and background. In Section 3, we introduce CEL and derive the statistical properties of the CEL estimator and CEL ratio test statistic for quantile regression. In Section 4, we propose the SCEL estimator and SCEL ratio test statistic, and derive the statistical properties. Section 5 considers the unconditional quantile restriction, and investigates EL- and SEL-based inference methods; we also derive the Bartlett correction for SEL. Section 6 provides extensions of the proposed methods, multiple quantile regression, homoskedasticity and symmetry tests, confidence sets without estimation, and specification tests. Section 7 concludes. Appendices contain mathematical proofs and preliminary lemma.

The author is currently working on a Monte Carlo simulation and empirical application of the proposed methods. The simulation setup is based on that of Horowitz (1998). The empirical example is wage regression based on CPS data by Buchinsky (1994) and Bierens and Ginther (2001). Preliminary results are available from the author's website (http://www.ssc.wisc.edu/∼totsu/quantile.htm).

2 Setup and background

In this section, we describe the basic setup and background for quantile regression models. Let $\{y_i : i = 1, \ldots, n\}$ be a scalar random sample used as a regressand and $\{x_i : i = 1, \ldots, n\}$ be a $q \times 1$ vector of random samples used as regressors. Letting $F_{y|x}$ be the conditional distribution function of y given x, the τth conditional quantile function of y given x is defined as $Q_\tau(y|x) \equiv \inf\{y|F_{y|x}(y|x) \geq \tau\}$. The (linear) quantile regression model is written as

$$y_i = x_i'\beta_0 + \epsilon_i, \quad Q_\tau(y_i|x_i) = x_i'\beta_0, \tag{1}$$

for $i = 1, \ldots, n$, where $\tau \in (0,1)$ is a fixed and known quantile of interest, β_0 is a $q \times 1$ vector of unknown parameters, and ϵ_i is the error term.[7] As τ increases from 0 to 1, we can trace the entire conditional distribution of y given x. In general, β_0 and ϵ_i vary with the value of τ. If $\tau = 0.5$, (1) corresponds to the LAD regression model. The mth element of the regression coefficients $\beta_{0m} = \partial Q_\tau(y_i|x_i)/\partial x_{im}$ is interpreted as the marginal change in the τth conditional quantile $Q_\tau(y_i|x_i)$ by the marginal change in x_{im}. Let $z \equiv (y, x')'$. Apart from the conditional quantile restriction (1), the distribution form of z is unspecified; therefore, the quantile regression model is regarded as a semiparametric model. Furthermore, note that compared to the conditional mean restriction $E[y|x] = x'\beta_0$, the conditional quantile restriction is robust to outliers in y.

Assume that z is continuously distributed with the joint density f_z and conditional density $f_{y|x}$. Then the conditional quantile $Q_\tau(y|x)$ satisfies $\int_{-\infty}^{Q_\tau(y|x)} f_{y|x}(y|x)dy = \tau$, and the quantile regression model (1) is equivalent to the following conditional moment restriction,

$$E[g(z, \beta_0)|x] \equiv E[\tau - I(y - x'\beta_0 \leq 0)|x] = 0 \quad \text{(conditional quantile restriction)}, \tag{2}$$

where $I(\cdot)$ is the indicator function. CEL and SCEL discussed in the following sections are constructed from this conditional moment restriction.

The conventional quantile regression estimator by Koenker and Bassett (1978a) is defined as

$$\hat{\beta}_{KB} \equiv \arg\min_{\beta \in \mathcal{B}} \frac{1}{n} \sum_{i=1}^{n} \rho_\tau(y_i - x_i'\beta), \tag{3}$$

where $\rho_\tau(\epsilon) \equiv \epsilon(\tau - I(\epsilon \leq 0))$ is the so-called check function. Let $f_{\epsilon|x}$ be the conditional density of ϵ given x. Koenker and Bassett (1978a) showed that under certain regularity conditions,

$$n^{1/2}(\hat{\beta}_{KB} - \beta_0) \xrightarrow{d} N(0, V_{KB}),$$

where

$$V_{KB} \equiv \tau(1-\tau)E[f_{\epsilon|x}(0|x)xx']^{-1}E[xx']E[f_{\epsilon|x}(0|x)xx']^{-1}.$$

To construct the confidence set of β_0, we usually estimate the variance V_{KB}, which contains the conditional error density evaluated at zero (i.e., $f_{\epsilon|x}(0|x)$). While several variance estimation methods have been proposed, such as the order statistics, kernel smoothing estimation, and bootstrapping, our EL-based methods avoid the variance estimation for constructing confidence sets.

[7] Under certain additional regularity conditions, our methods can be easily extended to nonlinear parametric regression models or parametric transformation models, such as the Box-Cox transformation model by Buchinsky (1995a).

Since the first-order condition for the minimization problem in (3) is written as $n^{-1}\sum_{i=1}^n x_i(\tau - I(y_i - x_i'\hat{\beta}_{KB} \leq 0)) = o_p(n^{-1/2})$, $\hat{\beta}_{KB}$ can be interpreted as the GMM estimator of the following unconditional moment restriction,

$$E[xg(z,\beta_0)] = E[x(\tau - I(y - x'\beta_0 \leq 0))] = 0 \quad \text{(unconditional quantile restriction).} \tag{4}$$

Since the conditional quantile restriction (2) implies infinitely many unconditional moment restrictions in the form of $E[\psi(x)(\tau - I(y - x'\beta \leq 0))] = 0$ for any arbitrary function ψ, (4) is an implication of (2) (i.e., $\psi(x) = x$). Therefore, $\hat{\beta}_{KB}$ is not efficient under the conditional quantile restriction (2). This result is analogous to the inefficiency of the OLS estimator under the conditional mean restriction like $E[y|x] = x'\beta_0$ (see Chamberlain (1987)).

Since $\hat{\beta}_{KB}$ is based on the unconditional quantile restriction (4), $\hat{\beta}_{KB}$ and the GMM estimator for (4) (i.e., $\hat{\beta}_{GMM} \equiv \arg\min_{\beta\in\mathcal{B}} n^{-1}\sum_{i=1}^n g(z_i,\beta)^2 x_i'x_i$) are asymptotically equivalent.[8] We can gain the efficiency of the GMM estimator by adding moment restrictions in the form of $E[\psi(x)(\tau - I(y - x'\beta \leq 0))] = 0$. However, an important difference between $\hat{\beta}_{KB}$ and $\hat{\beta}_{GMM}$ in practice is the existence of a linear programming representation for the optimization problems. Since the minimization problem of $\hat{\beta}_{KB}$ in (3) has a linear programming representation, we can apply, for example, the simplex method by Barrodale and Roberts (1973), which is globally convergent in a finite number of iterations.[9] On the other hand, since the minimization problem of $\hat{\beta}_{GMM}$ does not have a linear programming representation, we must apply some non-derivative optimization algorithm, such as Nelder and Mead's (1965) simplex method or simulated annealing, which has typically multiple local optima and converges slowly. Therefore, although the GMM approach is useful for discussing the theoretical properties of quantile regression estimators, the conventional estimator $\hat{\beta}_{KB}$ is more appropriate in practice. Our SCEL and SEL methods do not require non-derivative optimization due to smoothing on the moment restrictions.

To attain the semiparametric efficiency bound for the conditional quantile restriction (2), Newey and Powell (1990) proposed the optimally weighted quantile regression estimator,[10] that is

$$\hat{\beta}_{NP} \equiv \arg\min_{\beta\in\mathcal{B}} \frac{1}{n}\sum_{i=1}^n f_{\epsilon|x}(0|x_i)\rho_\tau(y_i - x_i'\beta). \tag{5}$$

The asymptotic distribution of $\hat{\beta}_{NP}$ is

$$n^{1/2}(\hat{\beta}_{NP} - \beta_0) \xrightarrow{d} N(0, V_{NP}),$$

[8]Note that (4) is just identified and $g(z,\beta_0)$ is scalar.

[9]Note that the simplex method for solving linear programming problems is different from Nelder and Mead's (1965) simplex method for optimizing non-smooth objective functions.

[10]Newey and Powell (1990) allows censored regression models.

where

$$V_{NP} \equiv \tau(1-\tau)E[f_{\epsilon|x}(0|x)^2 xx']^{-1}.$$

The optimal weight is the conditional error density evaluated at zero ($f_{\epsilon|x}(0|x_i)$), which also appears in the variance V_{NP}. Note that if the conditional density evaluated at zero is independent of x (i.e., $f_{\epsilon|x}(0|x) = f_\epsilon(0)$), $\hat{\beta}_{NP} = \hat{\beta}_{KB}$ and then the variance is simplified to

$$V_{KB} = V_{NP} = \frac{\tau(1-\tau)}{f_\epsilon(0)^2}E[xx']^{-1}. \tag{6}$$

In Section 6.2, we propose CEL- and SCEL-based test statistics for testing $f_{\epsilon|x}(0|x) = f_\epsilon(0)$. Using nonparametric estimation of a component including $f_{\epsilon|x}(0|x_i)$ and a sample splitting device for estimating the efficient score, Newey and Powell (1990) proposed a two-step estimation procedure for $\hat{\beta}_{NP}$. The estimates depend both on the method of nonparametric estimation of $f_{\epsilon|x}(0|x_i)$ and on the way of splitting the sample. Since the CEL and SCEL methods, as discussed in the following sections, automatically incorporate the optimal weight, the CEL and SCEL estimators do not require any preliminary nonparametric estimation of $f_{\epsilon|x}(0|x_i)$ or the sample splitting device for the efficient estimation of β_0.

Based upon $\hat{\beta}_{KB}$ or $\hat{\beta}_{NP}$, the Wald-type confidence set for the mth component of β_0 is obtained as

$$(\hat{\beta}_{\bullet m} - z_{\alpha/2}\sqrt{(\hat{V}_\bullet)_{mm}},\ \hat{\beta}_{\bullet m} + z_{\alpha/2}\sqrt{(\hat{V}_\bullet)_{mm}}\ ),$$

where $z_{\alpha/2}$ is $(1-\alpha/2)$-th quantile of a standard normal variable, $\bullet$ is KB or NP, and $(\hat{V}_\bullet)_{mm}$ is (m,m)th component of a consistent estimator for the variance of $\hat{\beta}_{\bullet m}$. Note that the above confidence set requires estimation of the variance $\hat{V}_\bullet$ that contains $f_{\epsilon|x}(0|x_i)$; in addition, the shape of the confidence set is restricted to be symmetric around $\hat{\beta}_\bullet$. Our EL-based confidence sets do not require variance estimation. Furthermore, the shapes of confidence sets are automatically determined by observed data (i.e., confidence sets may be asymmetric around the estimators).

3 Conditional empirical likelihood

In this section, we introduce the notion of CEL and derive asymptotic properties of the CEL estimator and CEL ratio test statistic for quantile regression models. The idea of CEL was proposed by Kitamura, Tripathi, and Ahn (2003) and Zhang and Gijbels (2003). However, Kitamura, Tripathi, and Ahn (2003) ruled out non-smooth moment restrictions like the conditional quantile restriction. While Zhang and Gijbels (2003) discussed quantile regression as an example, we provide a formal argument for asymptotic properties of the CEL estimator and CEL ratio test. Consider a discrete distribution with support on $\{z_1, \dots, z_n\} \times \{x_1, \dots, x_n\}$. We do not make any notational distinction among a random variable, the value taken by it, and its discrete support. The distinction should be clear from the context. Let $p_{ji} \equiv Pr\{z = z_j | x = x_i\}$ be the conditional probability mass of z given x. For information about $Pr\{z | x = x_i\}$, only a single observation, z_i, is available. By borrowing sample information from nearby observations around x_i, we can construct the nonparametric likelihood for the conditional quantile restriction $E[g(z, \beta_0)|x] = 0$. Let w_{ji} be weight for the sample information from nearby data, which is defined as

$$w_{ji} \equiv \frac{K(\frac{x_j - x_i}{b_n})}{\sum_{j=1}^n K(\frac{x_j - x_i}{b_n})},$$

where $K : R^q \mapsto R$ is a kernel function and b_n is a bandwidth parameter. Using w_{ji}, the local empirical likelihood at x_i is defined as

$$\sum_{j=1}^n w_{ji} \log p_{ji},$$

which is interpreted as the nonparametric kernel smoothing estimator for $E[\log p_{\cdot i}|x_i]$. Let $\mathcal{B}$ be the parameter space of β. Based on this local likelihood, consider the following maximization problem for each $\beta \in \mathcal{B}$,

$$\max_{\{p_{ji}\}_{i,j=1}^n} \sum_{i=1}^n \sum_{j=1}^n w_{ji} \log p_{ji} \tag{7}$$

$$\text{s.t. } p_{ji} \geq 0, \quad \sum_{j=1}^n p_{ji} = 1, \quad \sum_{j=1}^n p_{ji} g(z_j, \beta) = 0, \quad i, j = 1, \dots, n.$$

Using the Lagrange multiplier method, the maximizer of (7) is written as

$$\hat{p}_{ji} = \frac{w_{ji}}{1 + \lambda_i(\beta) g(z_j, \beta)},$$

where $\lambda_i(\beta)$, the Lagrange multiplier for the restriction $\sum_{j=1}^n p_{ji} g(z_j, \beta) = 0$, satisfies[11]

$$\sum_{j=1}^n \frac{w_{ji} g(z_j, \beta)}{1 + \lambda_i(\beta) g(z_j, \beta)} = 0. \tag{8}$$

Without the restrictions $\sum_{j=1}^n p_{ji} g(z_j, \beta) = 0$ for $i = 1, \dots, n$, the maximizer of (7) is $\tilde{p}_{ji} = w_{ji}$. Using $\hat{p}_{ji}$ and $\tilde{p}_{ji}$, the conditional empirical log-likelihood ratio (CELR) is defined as

$$\begin{aligned} \text{CELR}(\beta) &\equiv \sum_{i=1}^n I_{in} \Big(\sum_{j=1}^n w_{ji} \log \hat{p}_{ji} - \sum_{j=1}^n w_{ji} \log \tilde{p}_{ji} \Big) \\ &= -\sum_{i=1}^n I_{in} \sum_{j=1}^n w_{ji} \log(1 + \lambda_i(\beta) g(z_j, \beta)), \end{aligned} \tag{9}$$

[11]Note that in our quantile regression setup, $\lambda_i(\beta)$ and $g(z, \beta)$ are scalar.

where $I_{in} \equiv I(x_i \in \mathcal{X}_n)$ is a trimming term to avoid the boundary bias of kernel estimators, and $\mathcal{X}_n$ is a subset of the support of x, $\mathcal{X}$ (see, Ai (1997) and Ai and Chen (1999)). Let X_L and X_U be known boundary points of $\mathcal{X}$, and ι be a $q \times 1$ vector of ones. $\mathcal{X}_n$ is defined as $[X_L + b_n^\mu \iota, X_U - b_n^\mu \iota]$ for some $0 < \mu < 1$. In general, the computation of the CELR requires a numerical solution for $\lambda_i(\beta)$ in (8). However, for the quantile restriction $g(z, \beta)$, there exists an analytical solution for $\lambda_i(\beta)$ (see LeBlanc and Crowley (1995, p.100) and Kitamura (2003)), i.e.,

$$\hat{\lambda}_i(\beta) = \frac{\tau - \sum_{j=1}^n w_{ji} I(y_j - x_j'\beta \leq 0)}{\tau(1-\tau)} \equiv \frac{\tau - W_i(\beta)}{\tau(1-\tau)}. \tag{10}$$

By plugging (10) into (9), CELR(β) can be written as

$$\mathrm{CELR}(\beta) = -\sum_{i=1}^n I_{in}\Big[(1 - W_i(\beta))\log\Big(\frac{1 - W_i(\beta)}{1-\tau}\Big) + W_i(\beta)\log\Big(\frac{W_i(\beta)}{\tau}\Big)\Big]. \tag{11}$$

The conditional empirical likelihood estimator (CELE) is defined as

$$\hat{\beta}_{CEL} \equiv \arg\max_{\beta \in \mathcal{B}} \mathrm{CELR}(\beta). \tag{12}$$

Since CELR(β) is non-smooth for β, we must use some non-derivative optimization algorithm, such as Nelder and Mead's (1965) simplex method or simulated annealing. However, in this case, we do not need to nest the computational routine for $\lambda_i(\beta)$.

Assumptions for the asymptotic properties of the CELE are as follows.

Assumption 1. *Assume that*

(i) $\{y_i, x_i' : i = 1, \ldots, n\}$ *are i.i.d.,*

(ii) *the support of x, $\mathcal{X}$, is compact,*

(iii) *let x_{1i} be the constant term, and $(y_i, x_{2i}, \ldots, x_{qi})$ is continuously distributed with the joint density function f_z and the conditional density function $f_{y|x}$ of y_i given $x_i = x$ for $i = 1, \ldots, n$,*

(iv) *the density function of x, f_x, is strictly positive and continuous on $\mathcal{X}$, and $\sup_{x\in\mathcal{X}} f_x(x) < \infty$.*

Assumption 2. *Assume that*

(i) $E[g(z, \beta_0)|x] \equiv E[\tau - I(y - x'\beta_0 \leq 0)|x] = 0$ *almost surely for almost every $x \in \mathcal{X}$,*

(ii) *the parameter space for β, $\mathcal{B}$, is compact,*

(iii) $\beta_0 \in int(\mathcal{B})$.

Assumption 3. *Assume that*

(i) $f_{\epsilon|x}(0|x) > 0$ *for every* $x \in \mathcal{X}$*, where* $f_{\epsilon|x}$ *is the conditional density for* ϵ_i *given* $x_i = x$*,*

(ii) $f_{\epsilon|x}(\epsilon|x)$ *is Lipschitz continuous (i.e.,* $|f_{\epsilon|x}(\epsilon_1|x) - f_{\epsilon|x}(\epsilon_2|x)| \leq f_1|\epsilon_1 - \epsilon_2|$ *for some constant* $0 < f_1 < \infty$ *and every* ϵ_1*,* ϵ_2*, and* $x \in \mathcal{X}$*),*

(iii) *there exists a constant* $0 < f_2 < \infty$ *such that* $f_{\epsilon|x}(\epsilon|x) < f_2$ *for every* ϵ *and* $x \in \mathcal{X}$*,*

(iv) $E[f_{\epsilon|x}(0|x)^2 xx']$ *is positive definite.*

Assumption 4. *Assume that*

(i) *for* $v = (v_1, \dots, v_q)'$*,* $K(v) = \prod_{i=1}^q \kappa(v_i)$*, where* $\kappa : R \to R$ *is a continuously differentiable density function with support* $[-1, 1]$*. Furthermore,* κ *is symmetric around the origin,*

(ii) $b_n^q \propto n^{-\eta}$*, where* $0 < \eta < 1/2$*.*

Assumption 5. *Assume that when we solve (8) with respect to* $\{\lambda_i(\beta) : i = 1, \dots, n\}$ *for each* $\beta \in \mathcal{B}$*, we search only on the set* $\{\lambda \in R : ||\lambda|| \leq \Lambda_n\}$ *with* $\Lambda_n = o(1)$*.*

Assumption 1 (i) excludes dependent data. If the moment restriction $\{g(z_i, \beta) : i = 1, \dots, n\}$ were a martingale difference sequence, we expect that similar results would hold under certain additional regularity conditions.[12] However, the extension to weakly dependent processes, in which we have to deal with the long-run variance matrix of moment restrictions, is a challenging task.[13] Assumption 1 (ii) implies that all moments of x exist, and excludes unbounded regressors x. The compactness assumption of $\mathcal{X}$ can be dropped by employing a trimming argument of Kitamura, Tripathi, and Ahn (2003). Assumption 1 (iii) is required to derive the conditional quantile restriction (2) from the quantile regression model (1). Assumption 1 (iii) and (iv) exclude discrete regressors like dummy variables. This assumption can be dropped by using a trimming argument to control for small density values of f_x (see Andrews (1995)). If we include discrete regressors, the weight w_{ji} for constructing CEL should be modified to $w_{ji} = \left\{K(\frac{x_j^c - x_i^c}{b_n}) I(x_j^d = x_i^d)\right\} / \left\{\sum_{j=1}^n K(\frac{x_j^c - x_i^c}{b_n}) I(x_j^d = x_i^d)\right\}$, where x^c and x^d are continuous and discrete regressors, respectively. If all regressors are discrete, we can use the minimum distance estimator by Chamberlain (1994).

Assumption 2 (i) is the conditional quantile restriction, which assumes that the quantile regression model (1) is correctly specified. This assumption combined with Assumptions 1 and 3 (i) guarantees the global

[12] Weiss (1991) derived asymptotic properties of the conventional LAD estimator under dependent data with a martingale structure.

[13] For unconditional moment restriction models, Kitamura (1997) extended the empirical likelihood method to weakly dependent data by employing a blocking procedure.

identification of $\beta_0 \in \mathcal{B}$ (see, e.g., the proof of Kim and White (2002, Lemma 1)). Instead of assuming the correct specification (2), Kim and White (2002) considered (4) as the model of interest, and allowed the quantile regression model (1) to be misspecified. In that case, the solution of (4) with respect to β_0 is regarded as the "pseudo-true" parameters. Kitamura (2003) generalized the misspecification analysis to conditional moment restriction models, and showed that the CELE also converges to some pseudo-true value. Assumption 2 (ii) and (iii) are used for obtaining the consistency and asymptotic normality of the CEL estimator, respectively. Assumption 3, which is based on Powell (1986) and Kim and White (2002), is a set of standard regularity conditions on the conditional density $f_{\epsilon|x}$.

Assumption 4 (i) constrains the shape of the kernel function K. This assumption implies that K belongs to the class of second order product kernels. In order for $\hat{p}_{ji}$ to take only positive values, we rule out kernels whose orders are higher than two. Assumption 4 (ii) is a condition on the bandwidth b_n. Due to the boundedness of $g(z,\beta) \equiv \tau - I(y - x'\beta_0 \leq 0)$, this simple condition on b_n is sufficient in our setup (see Zhang and Gijbels (2003, Theorem 3)). The optimal choices of K and b_n are open questions.[14] Assumption 5, which is employed by Kitamura, Tripathi, and Ahn (2003, Assumption 3.6), controls the order of the Lagrange multiplier $\lambda_i(\beta)$ and simplifies the proof of Theorem 3.2. Since $\hat{\lambda}_i(\beta)$ converges to zero under (2), this assumption is innocuous in practice.

Under these assumptions, we obtain the consistency and asymptotic normality of the CELE, $\hat{\beta}_{CEL}$.

Theorem 3.1. *Suppose that Assumptions 1-5 hold. Then*

(i) $\hat{\beta}_{CEL} - \beta_0 = o_p(1)$,

(ii) $n^{1/2}(\hat{\beta}_{CEL} - \beta_0) \xrightarrow{d} N(0, V)$, *where* $V \equiv \tau(1-\tau)E[f_{\epsilon|x}(0|x)^2 xx']^{-1}$.

Therefore, the CELE, $\hat{\beta}_{CEL}$, is consistent, asymptotically normal, and efficient, i.e., $\hat{\beta}_{CEL}$ is asymptotically equivalent to Newey and Powell's (1990) efficient estimator $\hat{\beta}_{NP}$ in (5). In contrast to Newey and Powell's (1990) efficient estimator, we do not need to estimate the optimal weight $f_{\epsilon|x}(0|x_i)$, which is automatically incorporated in the construction of CEL. Since the non-smooth optimization problem for $\hat{\beta}_{CEL}$ in (12) does not have any linear programming representation, we must use some non-derivative optimization algorithm to implement CEL estimation.

Now consider a test of nonlinear parameter restrictions on β_0, that is

$$H_0 : R(\beta_0) = 0,$$

[14] For the bandwidth b_n, Kitamura, Tripathi, and Ahn (2003) suggested to use the bandwidth obtained in the process of estimation of optimal instrumental variables by Newey (1993). For the kernel K, Kitamura, Tripathi, and Ahn (2003) employed the Gaussian kernel in the simulation.

where $R : \mathcal{B} \to R^r$ is an $r \times 1$ vector of functions with $r \leq q$. For testing H_0, we can use the Wald test statistic with a quadratic form of $R(\hat{\beta})'[\text{Var}(R(\hat{\beta}))]^{-1}R(\hat{\beta})$, where $\hat{\beta}$ is some $\sqrt{n}$-consistent estimator of β_0, such as $\hat{\beta}_{KB}$, $\hat{\beta}_{NP}$, or $\hat{\beta}_{CEL}$. However, the Wald test statistic requires estimation of $\text{Var}(R(\hat{\beta}))$ and is not invariant to how the parameter restrictions R are specified. The likelihood ratio test statistic avoids these problems. The constrained CELE under H_0 is

$$\hat{\beta}_{CEL}^R \equiv \arg\max_{\beta \in \mathcal{B}} \text{CELR}(\beta) \quad \text{s.t.} \quad R(\beta) = 0.$$

Following Kitamura, Tripathi, and Ahn (2003), the CELR test statistic under H_0 is defined as

$$\text{LR}_n \equiv 2\{\text{CELR}(\hat{\beta}_{CEL}) - \text{CELR}(\hat{\beta}_{CEL}^R)\}. \tag{13}$$

The derivation of the asymptotic distribution of LR_n requires the following assumption regarding R.

Assumption 6. *Assume that $R : \mathcal{B} \to R^r$ is twice continuously differentiable and rank$\left(\frac{\partial R(\beta_0)}{\partial \beta'}\right) = r$.*

This standard assumption is used to derive an alternative representation of the constrained CELE $\hat{\beta}_{CEL}^R$. The asymptotic distribution of LR_n is obtained as follows.

Theorem 3.2. *Suppose that Assumptions 1-6 hold. Then under H_0,*

$$LR_n \xrightarrow{d} \chi_r^2.$$

This result is analogous to that of the usual likelihood ratio test. Note that the CELR test statistic does not require any variance estimation, and is invariant to the specification of R. Based on the CELR test statistic, we can construct asymptotically valid confidence sets for β_0. The $(1-\alpha) \times 100\%$ confidence set for the mth component of β_0 is obtained as

$$\left\{\beta_m : \min_{\beta_1,\ldots,\beta_{m-1},\beta_{m+1},\ldots,\beta_q} 2\{\text{CELR}(\hat{\beta}_{CEL}) - \text{CELR}(\beta)\} \leq \chi_{1,\alpha}^2\right\}, \tag{14}$$

where $\chi_{1,\alpha}^2$ is the $(1-\alpha) \times 100\%$ critical value of the χ_1^2 distribution. Similar to usual empirical likelihood, the above confidence set automatically satisfies natural range restrictions, and the shape of the confidence set is determined by observed data (i.e., it is not necessarily symmetric around the estimator, $\hat{\beta}_{CEL}$).

In practice, (14) is computed as follows. First, set the support of β_m, $\mathcal{B}_m \in R$. If there is no prior information for $\mathcal{B}_m$, we can set the support to be a large interval around the conventional estimator $\hat{\beta}_{KB,m}$; e.g., set $\mathcal{B}_m = [\ \hat{\beta}_{KB,m} - 5\sqrt{(\hat{V}_{KB})_{mm}}, \hat{\beta}_{KB,m} + 5\sqrt{(\hat{V}_{KB})_{mm}}\ ]$. Next, search over the set $\mathcal{B}_m$, and find the root for $\min_{\beta_1,\ldots,\beta_{m-1},\beta_{m+1},\ldots,\beta_q} 2\{\text{CELR}(\hat{\beta}_{CEL}) - \text{CELR}(\beta)\} = \chi_{1,\alpha}^2$ with respect to β_m. To find the root, we can use some one-dimensional optimization algorithm, such as the Bracketing algorithm (see, e.g., Judd (1998, p.95)).

4 Smoothed conditional empirical likelihood

In this section, we introduce SCEL and derive asymptotic properties of the SCEL estimator and SCEL ratio test statistic. An important drawback of the CEL approach is the lack of practical algorithm for implementation. Due to the non-smoothness of the CEL objective function in (9) or (11), the implementation of the CEL estimation requires some non-derivative algorithm, which tends to have multiple local optima and be computationally expensive. Hence, we introduce kernel smoothing on the non-smooth moment restriction $g(z,\beta) \equiv \tau - I(y - x'\beta)$ in order to obtain a smooth objective function. Once the objective function is smooth, we can apply a popular Newton-type algorithm. The idea of the smoothed moment restriction is proposed by Horowitz (1998) in the unconditional median restriction model. We extend Horowitz's (1998) idea to the conditional quantile restriction setup.

By replacing the indicator function in $g(z,\beta)$ with a smooth integrated kernel function H (i.e., the first-order derivative $H^{(1)}$ corresponds to the kernel function), the smoothed counterpart of the conditional quantile restriction function $g(z,\beta)$ is defined as

$$\tilde{g}(z,\beta) \equiv \tau - H\Big(-\frac{y - x'\beta}{h_n}\Big), \tag{15}$$

where h_n is the bandwidth for H that converges to 0 as $n \to \infty$. Note that $E[\tilde{g}(z,\beta_0)|x] = 0$ does not hold exactly. Intuitively, if h_n converges to 0 with a sufficiently fast rate, estimators and test statistics based on $g(z,\beta)$ and $\tilde{g}(z,\beta)$ are asymptotically (first-order) equivalent.

Based on the CELR in (9), the smoothed conditional empirical likelihood ratio (SCELR) is defined by replacing $g(z,\beta)$ with $\tilde{g}(z,\beta)$, that is

$$\text{SCELR}(\beta) \equiv -\sum_{i=1}^{n} I_{in} \sum_{j=1}^{n} w_{ji} \log(1 + \tilde{\lambda}_i(\beta)\tilde{g}(z_j,\beta)), \tag{16}$$

where $\tilde{\lambda}_i(\beta)$ satisfies

$$\sum_{j=1}^{n} \frac{w_{ji}\tilde{g}(z_j,\beta)}{1 + \tilde{\lambda}_i(\beta)\tilde{g}(z_j,\beta)} = 0. \tag{17}$$

The smoothed conditional empirical likelihood estimator (SCELE) is

$$\hat{\beta}_{SCEL} \equiv \arg\max_{\beta \in \mathcal{B}} \text{SCELR}(\beta).$$

Since there is generally no analytical solution for $\tilde{\lambda}_i(\beta)$ as (10), the implementation of the SCELE requires numerical optimization for $\tilde{\lambda}_i(\beta)$. First, for each $\beta \in \mathcal{B}$, $\tilde{\lambda}_i(\beta)$ $(i = 1, \ldots, n)$ is obtained by solving (17) or equivalently by computing

$$\arg\max_{\tilde{\lambda}_i} \sum_{j=1}^{n} w_{ji} \log(1 + \tilde{\lambda}_i \tilde{g}(z_j, \beta)),$$

which is a globally concave optimization problem. Next, by nesting the optimization routine for $\tilde{\lambda}_i(\beta)$, the SCELE $\hat{\beta}_{SCEL}$ is obtained as the maximizer of SCELR(β) with respect to β. Since SCELR(β) is smooth for β, we can apply a Newton-type optimization algorithm for computing $\hat{\beta}_{SCEL}$, which is more practical and faster than a non-derivative algorithm.

Instead of smoothing the moment restriction $g(z, \beta)$, we can also consider a smoothed counterpart of the explicit formula of the CELR in (11), that is

$$\mathrm{SCELR}^*(\beta) = -\sum_{i=1}^{n} I_{in} \Big[(1 - \tilde{W}_i(\beta)) \log\Big(\frac{1 - \tilde{W}_i(\beta)}{1 - \tau} \Big) + \tilde{W}_i(\beta) \log\Big(\frac{\tilde{W}_i(\beta)}{\tau} \Big) \Big], \tag{18}$$

where $\tilde{W}_i(\beta) \equiv \sum_{j=1}^{n} w_{ji} H\Big(-\frac{y_j - x_j' \beta}{h_n} \Big)$. Although SCELR$^*(\beta)$ is not identical to SCELR(β) in (16), we can expect that similar asymptotic properties will hold. The advantage of using SCELR$^*(\beta)$ is that we can avoid the computation to obtain $\tilde{\lambda}_i(\beta)$ from (17).

Similarly to the CELR test statistic in (13), the SCELR test statistic for nonlinear parameter restrictions $H_0 : R(\beta_0) = 0$ is defined as

$$\tilde{\mathrm{LR}}_n \equiv 2\{\mathrm{SCELR}(\hat{\beta}_{SCEL}) - \mathrm{SCELR}(\hat{\beta}^R_{SCEL})\}, \tag{19}$$

where $\hat{\beta}^R_{SCEL}$ is the constrained SCELE under H_0, that is

$$\hat{\beta}^R_{SCEL} \equiv \arg\max_{\beta \in \mathcal{B}} \mathrm{SCELR}(\beta) \quad \text{s.t.} \quad R(\beta) = 0.$$

Additional assumptions to obtain the asymptotic properties of the SCELE and SCELR test statistic are as follows.

Assumption 7. *Assume that*

(i) *$H : R \to R$ is bounded, differentiable everywhere, $H(v) = 0$ if $v \le -1$, $H(v) = 1$ if $v \ge 1$, and $H^{(1)}(v)$ is symmetric about $v = 0$, bounded, and Lipschitz continuous,*

(ii) *$h_n \propto n^{-\xi}$, where $0 < \xi < 1/2$.*

Assumption 8. *Assume that when we solve (17) with respect to $\{\tilde{\lambda}_i(\beta) : i = 1, \ldots, n\}$ for each $\beta \in \mathcal{B}$, we search only on the set $\{\tilde{\lambda} \in R : ||\tilde{\lambda}|| \le \tilde{\Lambda}_n\}$ with $\tilde{\Lambda}_n = o(1)$.*

Assumption 7 (i) and (ii) are conditions on the integrated kernel function H and its bandwidth h_n, respectively. In order to derive higher order properties, we need stronger assumptions on H and h_n, such as higher order of the kernel function $H^{(1)}(v)$ and stronger order conditions on h_n (see Section 5). The optimal choices of H and h_n are also challenging questions. Horowitz (1998) suggested to use Müller's (1984) optimal higher order integrated kernel for H, and Hall and Horowitz's (1990) optimal bandwidth of LAD regression models for h_n. Assumption 8, which controls the order of $\tilde{\lambda}_i(\beta)$, is analogous to Assumption 5.

Under these additional assumptions, the asymptotic properties of the SCELE and CELR test statistic are obtained as follows.

Theorem 4.1. *Suppose that Assumptions 1-5, 7, and 8 hold. Then*

(i) $\hat{\beta}_{SCEL} - \beta_0 = o_p(1)$,

(ii) $n^{1/2}(\hat{\beta}_{SCEL} - \beta_0) \xrightarrow{d} N(0, V)$.

Theorem 4.2. *Suppose that Assumptions 1-8 hold. Then under H_0*

$$\tilde{LR}_n \xrightarrow{d} \chi^2_r.$$

Therefore, the SCELE and SCELR test statistic are asymptotically (first-order) equivalent to the CELE and CELR test statistic, respectively. Due to the smoothness of the objective function, the SCEL-based methods are easier to implement than the CEL-based methods. The asymptotically valid confidence set is constructed by the similar manner as (14), that is

$$\left\{\beta_m : \min_{\beta_1,\cdots,\beta_{m-1},\beta_{m+1},\cdots,\beta_q} 2\{\mathrm{SCELR}(\hat{\beta}_{SCEL}) - \mathrm{SCELR}(\beta)\} \leq \chi^2_{1,\alpha}\right\}. \tag{20}$$

Similar to the CEL-based confidence set (14), the above confidence set does not require variance estimation, automatically satisfies natural range restrictions, and the shape of the confidence set is determined by data. The implementation is similar as the CEL-base confidence set except for the fact that we can use a Newton-type algorithm.

5 Unconditional quantile restriction: higher order refinement

So far, we have focused on the conditional quantile restriction (2). In the previous two sections, we show that the CELE and SCELE are asymptotically first-order equivalent to Newey and Powell's (1990) efficient estimator. In the standard unconditional moment restriction setup, in which sufficiently smooth moment restrictions are assumed, Newey and Smith (2003) and Kitamura (2001) provided favorable results for the EL approach relative to the GMM approach, based on higher order bias and large deviation properties, respectively. However, due to technical difficulties, it is challenging to extend the above results to the conditional moment restriction setup even for sufficiently smooth moment restrictions.[15]

Hence, in this section, we consider an unconditional moment restriction, which is relatively easier to analyze higher order properties, and derive the Bartlett correction for the SEL-based confidence set. We consider the following unconditional moment restriction,

$$E[g_u(z,\beta_0)] \equiv E[xg(z,\beta_0)] = E[x(\tau - I(y - x'\beta_0 \leq 0))] = 0, \tag{21}$$

which is implied from the conditional quantile restriction (2). Koenker and Bassett's (1978a) conventional quantile regression estimator is asymptotically equivalent to the optimally weighted GMM estimator for (21). This result is analogous to the efficiency of the OLS estimator for the projection model, i.e., $E[x(y-x'\beta_0)] = 0$.

For unconditional moment restriction models, we can employ usual empirical likelihood by Owen (1988), that is

$$\max_{\{p_i\}_{i=1}^n} \sum_{i=1}^n \log p_i \tag{22}$$

$$\text{s.t. } p_i \geq 0, \quad \sum_{i=1}^n p_i = 1, \quad \sum_{i=1}^n p_i g_u(z_i,\beta) = 0, \quad i = 1,\ldots,n,$$

for each $\beta \in \mathcal{B}$, where $p_i \equiv Pr\{z = z_i\}$ is the unconditional probability mass at z_i. Using the Lagrange multiplier method, the maximizer of (22) is written as

$$\hat{p}_i = \frac{1}{n(1+\gamma(\beta)' g_u(z_i,\beta))},$$

where the Lagrange multiplier $\gamma(\beta)$ satisfies

$$\sum_{i=1}^n \frac{g_u(z_i,\beta)}{1+\gamma(\beta)' g_u(z_i,\beta)} = 0. \tag{23}$$

Without the restriction $\sum_{i=1}^n p_i g(z_i,\beta) = 0$, the maximizer of (22) is $\tilde{p}_i = n^{-1}$. Using $\hat{p}_i$ and $\tilde{p}_i$, the empirical likelihood ratio (ELR) is defined as

$$\begin{aligned} \mathrm{ELR}(\beta) &\equiv \sum_{i=1}^n \log \hat{p}_i - \sum_{i=1}^n \log \tilde{p}_i \\ &= -\sum_{i=1}^n \log(1+\gamma(\beta)' g_u(z_i,\beta)). \end{aligned} \tag{24}$$

[15]The difficulties are mainly due to kernel smoothing in CEL or SCEL. By using local polynomial smoothing with variable bandwidth, Linton (2002) derived a higher order asymptotic expansion of Newey's (1990, 1993) optimal instrumental variables GMM estimator.

To derive the asymptotic distribution of $\mathrm{ELR}(\beta_0)$, we impose the following assumptions.

Assumption 9. *Assume that*

(i) $E[g_u(z, \beta_0)] = 0$ *for* $\beta_0 \in \mathcal{B}$,

(ii) $E[g(z, \beta_0)^2 xx']$ *is finite and has full rank.*

The result for the asymptotic distribution of $\mathrm{ELR}(\beta_0)$ does not require the full rank assumption in Assumption 9 (ii). In that case, since β_0 satisfying $E[g_u(z, \beta_0)] = 0$ is not necessarily unique, the confidence set will not shrink to a single point as $n \to \infty$.[16] However, in order to derive higher order properties of the SEL ratio, we require this assumption. As a special case of Owen (2001, Theorem 3.4), we obtain the following corollary.

Corollary 5.1. *Suppose that Assumptions 1 and 9 hold. Then*

$$-2ELR(\beta_0) \xrightarrow{d} \chi^2_q.$$

Therefore, even if the unconditional moment restriction (21) is non-smooth with respect to β_0, the ELR follows the limiting chi-square distribution. The EL-based confidence set for β_0 is constructed as

$$\{\beta : -2\mathrm{ELR}(\beta) \leq \chi^2_{q,\alpha}\}. \tag{25}$$

However, Chen and Hall (1993, p.1169) showed that in the case of distribution quantiles (i.e., x is constant), we cannot improve the coverage accuracy of the confidence set (25) with order higher than $n^{-1/2}$ because of the non-smoothness of the unconditional moment restriction $g_u(z, \beta)$. Moreover, to establish an Edgeworth expansion for the ELR, we must use Taylor series approximations, which require sufficiently smooth moment restrictions. Therefore, similarly to (15), we consider the following smoothed unconditional quantile restriction, that is

$$\tilde{g}_u(z, \beta) \equiv x\tilde{g}(z, \beta) = x\Big(\tau - H\Big(-\frac{y - x'\beta}{h_n}\Big)\Big),$$

where H is the integrated kernel function and h_n is the bandwidth. By replacing $g_u(z, \beta)$ in (24) with $\tilde{g}_u(z, \beta)$, the smoothed empirical likelihood ratio (SELR) is defined as

$$\mathrm{SELR}(\beta) \equiv -\sum_{i=1}^{n} \log(1 + \tilde{\gamma}(\beta)'\tilde{g}_u(z_i, \beta)), \tag{26}$$

[16] In other words, we do not need any identification assumption to show the asymptotic distribution of $\mathrm{ELR}(\beta_0)$. Otsu (2003b) extended the empirical likelihood inference under no or weak identification assumption to nonlinear and time-series models.

where $\tilde{\gamma}(\beta)$ satisfies

$$\sum_{i=1}^{n} \frac{\tilde{g}_u(z_i, \beta)}{1 + \tilde{\gamma}(\beta)'\tilde{g}_u(z_i, \beta)} = 0. \tag{27}$$

The derivation of the asymptotic distribution of SELR(β_0) requires the following assumptions.

Assumption 10. *Assume that*

(i) *x and $\epsilon = y - x'\beta_0$ are independent (i.e., $f_{\epsilon|x} = f_\epsilon$),*

(ii) *$H^{(1)}(v) \equiv \frac{dH(v)}{dv}$ is a p-th order kernel function, that is*

$$\int_{-1}^{1} v^j H^{(1)}(v)dv = \begin{cases} 1 & \text{if } j = 0, \\ 0 & \text{if } 1 \leq j \leq p-1, \\ C_H & \text{if } j = p, \end{cases}$$

where C_H is a positive constant,

(iii) *$f_\epsilon^{(p-1)}$ exists in a neighborhood of 0 and is continuous at 0,*

(iv) *$nh_n^{2p} \to 0$.*

Although the above assumptions are too strong for the first-order asymptotic distribution of SELR(β_0), we need these assumptions to establish the Edgeworth expansion and Bartlett correction. Assumption 10 (i) implies that $f_{\epsilon|x}(0|x) = f_\epsilon(0)$ and therefore Koenker and Bassett's (1978a) conventional estimator is efficient.[17] Even though the conventional estimator is efficient, the Bartlett correction of the SELR provides more precise confidence sets for β_0. Assumption 10 (ii) requires $H^{(1)}$ to be a higher order kernel function, which controls the remainder term in the asymptotic expansion of the SELR. Assumption 10 (iii), employed by Chen and Hall (1993, p.1170), controls the order of $E[\tilde{g}(z, \beta_0)]$. Assumption 10 (iv) ensures that h_n converges with a sufficiently quick rate so that the difference between ELR(β_0) and SELR(β_0) is negligible. Currently, there is no statistical theory for the choice of h_n. We may apply a suggestion by Horowitz (1998, p.1338), which is based on the optimal bandwidth for the LAD t statistic.

Theorem 5.1. *Suppose that Assumptions 1, 3, 7, 9, and 10 hold. Then*

$$-2SELR(\beta_0) \xrightarrow{d} \chi_q^2.$$

[17] Horowitz (1998) dropped this assumption and derived higher order refinements for the t and Wald test statistics by bootstrapping.

Therefore, SELR(β_0) is asymptotically first-order equivalent to ELR(β_0). The valid confidence set for β_0 is constructed as

$$\{\beta : -2\text{SELR}(\beta) \leq \chi^2_{q,\alpha}\}. \tag{28}$$

In addition to Assumption 10, suppose that there exist sufficiently higher order moments for ϵ_i and x_i. We also assume that a multivariate analog of Cramér's condition in Chen and Hall (1993, pp.1178-1179) holds, i.e., for sufficiently small h_n, we impose a boundedness condition for the characteristic function of a stacked vector of sufficiently higher-order power functions of $\tilde{g}_u(z_i, \beta)$.[18] Then, similarly to Chen and Hall (1993, Theorem 3.2), we can establish an Edgeworth expansion and show that the order of the coverage error of (28) is $O(n^{-1})$, i.e.,

$$Pr\Big\{-2\text{SELR}(\beta_0) \leq t\Big\} = Pr\{\chi^2_q \leq t\} + O(n^{-1}).$$

In order to discuss higher order properties, we introduce additional notation. Let

$$\begin{aligned}
g_i &\equiv E[\tilde{g}(z_i, \beta_0)^2 x_i x_i']^{-1/2} x_i \tilde{g}(z_i, \beta_0), \\
g^{j_1 \cdots j_m} &\equiv n^{-1} \sum_{i=1}^n g_i^{j_1} \cdots g_i^{j_m}, \\
\alpha^{j_1 \cdots j_m} &\equiv n^{-1} \sum_{i=1}^n E[g_i^{j_1} \cdots g_i^{j_m}], \\
A^{j_1 \cdots j_m} &\equiv n^{-1} \sum_{i=1}^n (g_i^{j_1} \cdots g_i^{j_m} - \alpha^{j_1 \cdots j_m}),
\end{aligned}$$

where g_i^j is the jth component of g_i. Note that $g^{j_1 \cdots j_m} = A^{j_1 \cdots j_m} + \alpha^{j_1 \cdots j_m}$ by definition, and $A^{j_1 \cdots j_m} = O_p(n^{-1/2})$ by the central limit theorem. By a similar argument as DiCiccio, Hall, and Romano (1991), we obtain the signed root approximation for SELR(β_0) (see Appendix B for the derivation), that is

$$-2\text{SELR}(\beta_0) = nR'R + O_p((n^{-1/2} + h_n^p)^3), \tag{29}$$

where $R = R_1 + R_2 + R_3$, $R_1 = O_p(n^{-1/2} + h_n^p)$, $R_2 = O_p((n^{-1/2} + h_n^p)^2)$, and $R_3 = O_p((n^{-1/2} + h_n^p)^3)$. The jth components of R_1, R_2, and R_3 are written as

$$\begin{aligned}
R_1^j &\equiv g^j, \\
R_2^j &\equiv -\frac{1}{2} g^k A^{jk} + \frac{1}{3} g^k g^m \alpha^{jkm}, \\
R_3^j &\equiv \frac{3}{8} g^k A^{jm} A^{km} + \frac{1}{3} g^k g^l A^{jkm} - \frac{5}{12} g^k g^l \alpha^{jkm} A^{lm} - \frac{5}{12} g^k g^l \alpha^{klm} A^{jm} + \frac{4}{9} g^k g^l g^m \alpha^{jkn} \alpha^{lmn} - \frac{1}{4} g^k g^l g^m \alpha^{jklm},
\end{aligned}$$

[18] If x_i is constant (i.e., β is the distribution quantile), $nh_n \log n \to 0$ as $n \to \infty$ ensures Cramér's condition.

where repeated indices are summed over in the usual summation convention. In order to derive the Bartlett correction, suppose that $\sup_n n^3 h_n^{2p} < \infty$, which is used for deriving expansions of $E[nR_1^j R_3^j]$ and $E[nR_2^j R_2^j]$. Under the validity of the Edgeworth expansion for the distribution of $n^{1/2}R$, we can apply a similar argument as DiCiccio, Hall, and Romano (1991, p.1055); the higher order refinement for the SELR is obtained as

$$Pr\Big\{-2\mathrm{SELR}(\beta_0)\{E[n(R'R)/q]\}^{-1} \leq t\Big\} = Pr\{\chi_q^2 \leq t\} + O(n^{-2}). \tag{30}$$

Intuitively, the Bartlett correction is a multiplicative finite sample correction that ensures that the mean of the corrected statistic matches the mean of the limiting chi-square distribution (i.e., q). Since the asymptotic expansion for $E[-2\mathrm{SELR}(\beta_0)]$ does not exist in general, we use the higher order approximation for $E[nR'R/q]$ as a correction factor. The Bartlett correction term $\{E[n(R'R)/q]\}^{-1}$ is written as (see Appendix B for the derivation),

$$\{E[n(R'R)/q]\}^{-1} = 1 - an^{-1} + O(n^{-2}), \tag{31}$$

where

$$a \equiv q^{-1}\Big(\frac{1}{2}\alpha^{jjkk} - \frac{1}{3}\alpha^{jkl}\alpha^{jkl}\Big).$$

Therefore, from (30) and (31), the Bartlett correction for the SELR is obtained as

$$Pr\Big\{-2\mathrm{SELR}(\beta_0)(1 - an^{-1}) \leq t\Big\} = Pr\{\chi_q^2 \leq t\} + O(n^{-2}), \tag{32}$$

and the higher order refined confidence set is constructed as

$$\{\beta : -2\mathrm{SELR}(\beta)(1 - an^{-1}) \leq \chi_{q,\alpha}^2\}.$$

The Bartlett correction, i.e., the multiplication of $(1 - an^{-1})$ to $-2\mathrm{SELR}(\beta_0)$, reduces the rate of the error for the rejection probability from $O(n^{-1})$ to $O(n^{-2})$. The correction factor a can be consistently estimated by the sample analog. Baggerly (1998) showed that in the member of the Cressie and Read's (1984) family of discrepancy statistics, only empirical likelihood is Bartlett correctable. Thus, for example, exponential tilting likelihood (Kitamura and Stutzer (1997) and Imbens, Spady, and Johnson (1998)) and the continuous updating GMM objective function (Hansen, Heaton, and Yaron (1996)) are not Bartlett correctable. This result is due to the forms of the third- and fourth-order joint cumulants of the signed root of Cressie and Read's (1984) discrepancy statistics. We can expect that the same result will hold in our setup under some suitable regularity conditions for H and h_n.

As mentioned earlier, after the completion of this draft, the author was informed that Whang (2003) independently derived similar results, i.e., the Bartlett correction for the SELR. In contrast to Assumption 10 (i), Whang (2003) allows for some dependence between x and ϵ, establishes the Edgeworth expansion with rigorous proof, and extends the Bartlett correctability to censored regression models. However, Whang's (2003) main purpose is to compare the SELR to the bootstrap; this section is intended merely to provide a motivation for smoothing the quantile restriction. Based on Chen and Cui (2002, 2003), the author is currently working on extensions of the Bartlett correctability to (i) overidentified unconditional quantile restriction models (i.e., $E[\psi(x)g(z,\beta_0)] = 0$); and (ii) quantile restriction models with nuisance parameters.

6 Extensions

6.1 Multiple quantile regression

Instead of a single quantile regression model for a specific value of quantile τ, this subsection considers the following multiple quantile regression model at different values of quantile $\tau^M \equiv (\tau_1, \ldots, \tau_L)$, that is

$$y_i = x_i'\beta_0^\ell + \epsilon_i^\ell, \quad Q_{\tau_\ell}(y_i|x_i) = x_i'\beta_0^\ell, \quad \ell = 1, \ldots, L, \tag{33}$$

where $0 < \tau_1 < \cdots < \tau_L < 1$ without loss of generality. Multiple quantile regression is useful for testing parameter restrictions among different quantiles, such as the homoskedasticity and symmetry restrictions of the error term (see next subsection). In order to impose cross-restrictions for $\beta_0^M \equiv (\beta_0^{1\prime}, \ldots, \beta_0^{L\prime})'$, we need to estimate simultaneously the whole parameter vector β_0^M. To apply the empirical likelihood approach, we use the following conditional moment restrictions for (33),

$$E[(g_1(z,\beta_0^1), \ldots, g_L(z,\beta_0^L))'|x] = 0,$$

where $g_\ell(z,\beta_0^\ell) \equiv \tau_\ell - I(y - x'\beta_0^\ell \le 0)$. In this case, CEL and SCEL are defined by replacing $g(z,\beta_0)$ and $\tilde{g}(z,\beta_0)$ with $(g_1(z,\beta_0^1), \ldots, g_L(z,\beta_0^L))'$ and $(\tilde{g}_1(z,\beta_0^1), \ldots, \tilde{g}_L(z,\beta_0^L))'$ in (9) and (16), respectively. Since the statistical properties of the CEL and SCEL estimators and their test statistics in Sections 3 and 4 do not depend on the dimension of conditional moment restrictions, we obtain similar results as the single quantile case under some analogous regularity conditions to Assumptions 1-7. The asymptotic distribution of the CELE and SCELE for β_0^M (denote $\hat{\beta}_{CEL}^M$ and $\hat{\beta}_{SCEL}^M$, respectively) is obtained as[19]

$$n^{1/2}(\hat{\beta}_{CEL}^M - \beta_0^M) \overset{a}{=} n^{1/2}(\hat{\beta}_{SCEL}^M - \beta_0^M) \xrightarrow{d} N(0, V^M),$$

[19]Since the estimates for any pair of β_0^a and β_0^b are different in general, the estimated quantile regression lines (i.e., $x'\hat{\beta}^a$ and $x'\hat{\beta}^b$) cross each other at some point of x. Therefore, we need to check that those crosses do not appear within the relevant range of x.

where

$$(V^M)_{ab} \equiv (\min\{\tau_a, \tau_b\} - \tau_a\tau_b)E[f_{\epsilon_a|x}(0|x)f_{\epsilon_b|x}(0|x)xx']^{-1}.$$

Note that the asymptotic variance for $\hat{\beta}^{\ell}_{CEL}$ or $\hat{\beta}^{\ell}_{SCEL}$ is same as that of the single quantile case. The CEL- (or SCEL-) based parameter restriction test statistics in (13) (or (19)) and confidence sets in (14) (or (20)) are obtained by a similar manner. Similarly to the single quantile case, we require neither estimation of the optimal weights for efficient estimation nor variance estimation for obtaining the confidence set.

6.2 Tests for homoskedasticity and symmetry

In this subsection, we propose the CEL- and SCEL-based test statistics for homoskedasticity and symmetry of the error term by using the multiple quantile regression estimators $\hat{\beta}^M_{CEL}$ and $\hat{\beta}^M_{SCEL}$.[20] These test statistics are useful for investigating the shape of the conditional distribution $F_{y|x}$. Since the homoskedasticity and symmetry restrictions are written as cross parameter restrictions on $(\beta^1_0, \ldots, \beta^L_0)$, we can use the CELR or SCELR test statistic in (13) or (19), respectively.

First, consider a test for homoskedasticity. If the conditional error density evaluated at zero does not depend on x (i.e., $f_{\epsilon|x}(0|x) = f_\epsilon(0)$), then any pair of quantile regression parameters differs only in the intercepts and all the slope coefficients are identical (i.e., multiple quantile regression lines are parallel). Moreover, in this case, Koenker and Bassett's (1978a) conventional quantile regression estimator $\hat{\beta}_{KB}$ is efficient (see (6)). Letting β^{ℓ}_{0m} be the mth component of β^{ℓ}_0, the homoskedasticity restriction is written as

$$H^{homo}_0 : \ \beta^1_{0m} = \cdots = \beta^L_{0m} \ \text{ for } \ m = 2, \ldots, q. \tag{34}$$

Let $\hat{\beta}^{M,homo}_{CEL}$ and $\hat{\beta}^{M,homo}_{SCEL}$ be the CELE and SCELE under the restriction H^{homo}_0, respectively. By applying Theorems 3.2 and 4.2, the asymptotic distribution of the CELR and SCELR test statistics for H^{homo}_0 is obtained as

$$2\{\mathrm{CELR}(\hat{\beta}^M_{CEL}) - \mathrm{CELR}(\hat{\beta}^{M,homo}_{CEL})\} \overset{a}{=} 2\{\mathrm{SCELR}(\hat{\beta}^M_{SCEL}) - \mathrm{SCELR}(\hat{\beta}^{M,homo}_{SCEL})\} \xrightarrow{d} \chi^2_{Lq-(L+q-1)}. \tag{35}$$

Next, consider a test for symmetry. If the conditional density of the error term, $f_{\epsilon|x}$, is symmetric around zero, multiple quantile regression coefficients $(\beta^1_0, \ldots, \beta^L_0)$ also show a symmetric pattern. Suppose that L is an odd number, $\tau_{L-\ell} = 1 - \tau_{\ell+1}$ for $\ell = 0, \cdots, (L-1)/2 - 1$, and $\tau_{1+(L-1)/2} = 0.5$ (median). In words,

[20]While our test statistics are based on multiple quantile regression of discrete points of quantiles, Koenker and Xiao (2002) considered a continuous quantile regression process, and proposed test statistics for location shift and location-scale shift models by a similar manner as the Kolmogorov-Smirnov test.

the quantile points $(\tau_1, \cdots, \tau_L)$ are located symmetrically around the median. From Buchinsky (1998), the parameter restriction implied by the symmetric error density is written as

$$H_0^{sym}: \; \frac{1}{2}(\beta^{L-\ell} + \beta^{\ell+1}) = \beta^{1+(L-1)/2} \;\; \text{for} \;\; \ell = 0, \cdots, (L-1)/2 - 1. \tag{36}$$

Let $\hat{\beta}^{M,sym}_{CEL}$ and $\hat{\beta}^{M,sym}_{SCEL}$ be the CELE and SCELE under the restriction H_0^{sym}, respectively. Similar to (35), the asymptotic distribution of the CELR and SCELR test statistics for H_0^{sym} is obtained as

$$2\{\text{CELR}(\hat{\beta}^M_{CEL}) - \text{CELR}(\hat{\beta}^{M,sym}_{CEL})\} \overset{a}{=} 2\{\text{SCELR}(\hat{\beta}^M_{SCEL}) - \text{SCELR}(\hat{\beta}^{M,sym}_{SCEL})\} \overset{d}{\to} \chi^2_{(L-1)q/2}. \tag{37}$$

We can conduct a joint test for H_0^{homo} and H_0^{sym} by a similar manner. Note that compared to existing test statistics, such as the minimum distance or GMM test statistics, the CELR and SCELR test statistics do not need to estimate or choose some weight matrix.

6.3 Confidence set without estimation

In this subsection, we propose CEL- and SCEL-based confidence sets that do not require any estimation of parameters. The CEL- and SCEL-based confidence sets in (14) and (20), respectively, require parameter estimation for concentrating out nuisance parameters. However, if we are interested in the joint confidence set of β_0, we can employ Tripathi and Kitamura's (2001) canonical version of the CELR test statistic and obtain a valid confidence set without estimating parameters. This confidence set is particularly useful if the number of parameters is small.

While the CELR test statistic in (13) follows the chi-square limiting distribution, Tripathi and Kitamura (2001) showed that a normalized version of $\text{CELR}(\beta_0)$ (denote $\text{LR}^*(\beta_0)$) follows the normal limiting distribution under the null hypothesis $H_0: E[g(z, \beta_0)|x] = 0$. The normalized test statistic is defined as

$$\text{LR}^*(\beta_0) \equiv \{-2b_n^{q/2}\text{CELR}(\beta_0) - b_n^{q/2} T_n(\beta_0)\}/\sigma, \tag{38}$$

where

$$\begin{aligned} T_n(\beta_0) &\equiv \sum_{i=1}^n \sum_{j=1, j\neq i}^n I_{in} w_{ji}^2 \Big(\sum_{j=1}^n w_{ji} g(z_j, \beta_0)^2\Big)^{-1} g(z_j, \beta_0)^2, \\ \sigma^2 &\equiv 2\Big\{\int_{[-2,2]^q} \Big(\int_{[-1,1]^q} K(v)K(u-v)dv\Big)^2 du\Big\}. \end{aligned}$$

Note that σ is a constant and can be computed numerically. $T_n(\beta_0)$ is a correction term for the bias in $-2\text{CELR}(\beta_0)$. Since Assumptions 1-4 satisfy Tripathi and Kitamura's (2001) regularity conditions, we obtain the following corollary.

Corollary 6.1. *Suppose that Assumptions 1-4 hold. Furthermore, assume that* $b_n^q \propto n^{-\eta}$*, where* $0 < \eta < 1/3$*. Then*

$$LR^*(\beta_0) \xrightarrow{d} N(0,1).$$

If the conditional moment restriction is correctly specified, we can use $\mathrm{LR}^*(\beta_0)$ as a test statistic for the simple parameter hypothesis $H_0 : \beta_0 = \beta$. If researchers are interested in testing the validity of some specific values of β, $\mathrm{LR}^*(\beta)$ is a convenient test statistic. Since $\mathrm{LR}^*(\beta_0)$ does not contain any parameter estimator, we avoid the optimization problem for the non-smooth objective function $\mathrm{CELR}(\beta)$. In this case, the confidence set for β_0 is constructed as

$$\Big\{\beta \ : \ |\mathrm{LR}^*(\beta)| \le z_{\alpha/2}\Big\}, \tag{39}$$

where $z_{\alpha/2}$ is the $(1-\alpha/2) \times 100\%$ critical value for the standard normal distribution. If the dimension of β_0 is small (typically, fewer than three), (39) is a convenient way for obtaining joint confidence sets. The SCELR-based test statistic (i.e., $\{-2b_n^{q/2}\mathrm{SCELR}(\beta_0) - b_n^{q/2}T_n(\beta_0)\}/\sigma$) and confidence set can be derived by a similar manner.

6.4 Specification test for conditional quantile restriction

In this subsection, we propose CEL- and SCEL-based consistent specification tests for the quantile regression model. The specification test is useful to check the validity of specified functional forms and included regressors. We extend Tripathi and Kitamura's (2002) CEL-based specification test statistic to the quantile regression setup. Since Tripathi and Kitamura (2002) ruled out non-smooth moment restrictions, this extension is not trivial. The null hypothesis is

$$H_0 \ : \ Pr\{E[g(z,\beta)|x] = 0\} = 1 \ \text{ for some } \ \beta \in \mathcal{B},$$

and the alternative is that H_0 is false. Zheng (1998) and Bierens and Ginther (2001) developed consistent specification tests for H_0, which are based on a kernel smoothing method and integrated conditional moment test, respectively. Kim and White (2002) proposed a convenient specification test based on the information matrix equality, which is however inconsistent for H_0. Our CEL-based specification test statistic is consistent and is obtained as a by-product of the CEL estimation.

Let $\hat{\beta}$ be some $\sqrt{n}$-consistent estimator for β_0, such as Koenker and Bassett's (1978a) conventional quantile regression estimator, CELE, or SCELE. Following Tripathi and Kitamura (2002), the CEL-based specification test statistic for H_0 is obtained by replacing β_0 in (38) with $\hat{\beta}$, that is

$$\mathrm{LR}^*(\hat{\beta}) \equiv \{-2b_n^{q/2}\mathrm{CELR}(\hat{\beta}) - b_n^{q/2}T_n(\hat{\beta})\}/\sigma. \tag{40}$$

If $\hat{\beta}$ is the CELE, CELR$(\hat{\beta})$ is obtained as the maximized objective function of the CEL estimation. The following theorem is an extension of Tripathi and Kitamura's (2002) results to the non-smooth conditional quantile restriction setup.

Theorem 6.1. *Suppose that Assumptions 1-5 hold. Furthermore, assume that $b_n^q \propto n^{-\eta}$, where $0 < \eta < 1/3$. Then*

$$LR^*(\hat{\beta}) \xrightarrow{d} N(0,1).$$

Under additional regularity conditions, we can also show the asymptotic distribution of LR$^*(\hat{\beta})$ under local alternatives. Under sufficiently smooth moment restrictions, Tripathi and Kitamura (2002) showed the asymptotic optimality of the CELR-based specification test in terms of average local power. We can expect that the same result will hold in our setup. The SCELR-based test statistic (i.e., $\{-2b_n^{q/2}\mathrm{SCELR}(\hat{\beta}) - b_n^{q/2}T_n(\hat{\beta})\}/\sigma$) can be derived by a similar manner as the CEL case.

7 Conclusion

In this paper we propose new estimation and inference methods for quantile regression models. Our methods are based on empirical likelihood and its extensions, i.e., conditional empirical likelihood, smoothed conditional empirical likelihood, usual empirical likelihood, and smoothed empirical likelihood. The advantages of the proposed methods are that (i) the conditional empirical likelihood and smoothed conditional empirical likelihood estimators are asymptotically efficient for the quantile regression model; (ii) all of the empirical likelihood-based test statistics provide valid confidence sets without variance estimation of the estimators; (iii) the smoothed conditional empirical likelihood and smoothed empirical likelihood methods can be implemented by a Newton-type algorithm; and (iv) smoothed empirical likelihood is Bartlett correctable, and provides higher order refinements of the derived confidence sets. However, we introduce kernel smoothing devices, in which we need to choose the kernel function and bandwidth. Based on the proposed methods, we provide four extensions, i.e., multiple quantile regression models, tests for homoskedasticity and symmetry, confidence sets without parameter estimation, and consistent specification tests for quantile regression.

In order to investigate finite sample performance of the proposed methods, we are now conducting Monte Carlo simulations based on the setup of Horowitz (1998). In addition, we plan to include an empirical example of wage regression by Buchinsky (1994). Preliminary simulation results show better performance of our methods than conventional methods, both in terms of point estimates and confidence sets. Preliminary results are available from the author's website (http://www.ssc.wisc.edu/$\sim$totsu/quantile.htm).

Directions for future research include higher order comparisons of efficient estimators, choice rules for bandwidths, inclusion of endogenous regressors, and extensions to censored regression models, selection models, semiparametric models (e.g., partially linear quantile regression models), and weakly dependent data.

A Mathematical proofs

Notation

Let c denote a generic constant, which may vary from case to case, and define

$$\begin{aligned}
\delta &\equiv \beta-\beta_0, \quad \hat{f}_i \equiv \frac{1}{nb_n^q}\sum_{j=1}^n K\Big(\frac{x_j-x_i}{b_n}\Big), \\
\hat{g}(x_i,\beta) &\equiv \sum_{j=1}^n w_{ji}g(z_j,\beta), \quad \tilde{\hat{g}}(x_i,\beta) \equiv \sum_{j=1}^n w_{ji}\tilde{g}(z_j,\beta), \\
\hat{V}(x_i,\beta) &\equiv \sum_{j=1}^n w_{ji}g(z_j,\beta)^2, \quad \tilde{\hat{V}}(x_i,\beta) \equiv \sum_{j=1}^n w_{ji}\tilde{g}(z_j,\beta)^2, \quad V(x_i,\beta) \equiv E[g(z,\beta)^2|x_i].
\end{aligned}$$

Proof of Theorem 3.1

Check the assumptions of Zhang and Gijbels (2003, Theorems 2 and 3), i.e., conditions (X0), (K0), and (P1)-(P9) in Zhang and Gijbels (2003).

(X0) is satisfied by Assumption 1 (ii) and (iv). (K0) is satisfied by Assumption 4 (i) (i.e., the symmetry of κ implies $\int_{[-1,1]^q} vK(v)dv = 0$, and the continuity and compact support of κ imply $\int_{[-1,1]^q} K(v)^{(1+\delta_0)/\delta_0}dv < \infty$ for every $\delta_0 > 0$). (P1) is equivalent to Assumption 2 (i). Since $|g(z,\beta)| = |\tau - I(y-x'\beta \le 0)| \le \tau+1$, $E[\sup_{\beta\in\mathcal{B}}|g(z,\beta)|^{\alpha_0}] \le |\tau+1|^{\alpha_0} < \infty$ for every $\alpha_0 > 0$ (i.e., (P2) is satisfied), and $\sup_{x\in\mathcal{X},\beta\in\mathcal{B}} E[g(z,\beta)^4|x] \le (\tau+1)^4 < \infty$ (i.e., (P3) is satisfied).

Note that

$$E[g(z,\beta)^2|x] = \tau^2 + (-2\tau+1)E[I(\epsilon \le x'\delta)|x] = \tau^2 + (-2\tau+1)\int_{-\infty}^{x'\delta} f_{\epsilon|x}(\epsilon|x)d\epsilon.$$

From Assumption 3 (iii), $E[g(z,\beta)^2|x]$ is finite. From Assumption 1 (iii), $E[g(z,\beta)^2|x]$ is continuous with respect to $x \in \mathcal{X}$ and $\beta \in \mathcal{B}$. Furthermore, since $0 \le \int_{-\infty}^{x'\delta} f_{\epsilon|x}(\epsilon|x)d\epsilon \le 1$, $E[g(z,\beta)^2|x] \ge \min\{\tau^2,(\tau-1)^2\} > 0$ for every $0<\tau<1$, $x\in\mathcal{X}$, and $\beta\in\mathcal{B}$. Then (P4) is satisfied.

From Assumption 2 (i),

$$E[g(z,\beta)|x] = E[g(z,\beta)|x] - E[g(z,\beta_0)|x] = -E[I(\epsilon \le x'\delta)|x] + E[I(\epsilon\le 0)|x] = -\int_0^{x'\delta} f_{\epsilon|x}(\epsilon|x)d\epsilon.$$

By Leibniz's formula, Cauchy-Schwartz inequality, and Assumptions 1 (ii) and 3 (iv),

$$\begin{aligned}
\sup_{x\in\mathcal{X},\beta\in\mathcal{B}}\Big\|\frac{\partial E[g(z,\beta)|x]}{\partial\beta}\Big\| &= \sup_{x\in\mathcal{X},\beta\in\mathcal{B}}\Big\|-\frac{\partial\int_0^{x'\delta} f_{\epsilon|x}(\epsilon|x)d\epsilon}{\partial\delta}\Big\| = \sup_{x\in\mathcal{X},\beta\in\mathcal{B}}||f_{\epsilon|x}(x'\delta|x)x|| \\
&\le \sup_{x\in\mathcal{X},\beta\in\mathcal{B}}||f_{\epsilon|x}(x'\delta|x)||\sup_{x\in\mathcal{X}}||x|| \le f_2\sup_{x\in\mathcal{X}}||x|| < \infty.
\end{aligned}$$

Then (P5) is satisfied.

Let $\mathcal{F} \equiv \{g(z,\beta) : \beta \in \mathcal{B}\}$, and $\mathcal{N}(\nu, L_1, \mathcal{F})$ be the covering number for $\mathcal{F}$ by the L_1-norm, i.e., the smallest number of $\nu-$balls in L_1 metric required for covering $\mathcal{F}$. It is verified that $\mathcal{F}$ belongs to the type I class in Andrews (1994) (i.e., set $h(\cdot)$ in Andrews (1994, p. 2270) as $g(z,\beta)$). Furthermore, from Andrews (1994, p. 2284), the type I class in Andrews (1994) is a subset of the VC-subgraph class defined in, e.g., van der Vaart and Wellner (1996). From van der Vaart and Wellner (1996, Theorem 2.6.4), the upper bound of the covering number for the VC-subgraph class is written as

$$\mathcal{N}(\nu, L_1, \mathcal{F}) \leq c\nu^{1-V(\mathcal{F})},$$

where $V(\mathcal{F})$ is the VC-index defined in van der Vaart and Wellner (1996, p. 135). Since $V(\mathcal{F}) > 1$ by the definition of $V(\mathcal{F})$, (P6) is satisfied.

From Assumption 1 (iii),

$$\begin{aligned}
& \sup_{x\in\mathcal{X},\beta\in\mathcal{B}} |E[g(z,\beta)|x+x^*] - E[g(z,\beta)|x]| = \sup_{x\in\mathcal{X},\beta\in\mathcal{B}} |-E[I(\epsilon \leq (x+x^*)'\delta)|x+x^*] + E[I(\epsilon \leq x'\delta)|x]| \\
= & \sup_{x\in\mathcal{X},\beta\in\mathcal{B}} \left| -\int_{x'\delta}^{(x+x^*)'\delta} f_{\epsilon|x}(\epsilon|x)d\epsilon \right| \to 0,
\end{aligned}$$

and

$$\begin{aligned}
& \sup_{x\in\mathcal{X},\beta\in\mathcal{B}} |E[g(z,\beta)^2|x+x^*] - E[g(z,\beta)^2|x]| = \sup_{x\in\mathcal{X},\beta\in\mathcal{B}} |(-2\tau+1)\{E[I(\epsilon \leq (x+x^*)'\delta)|x+x^*] - E[I(\epsilon \leq x'\delta)|x]\}| \\
= & \sup_{x\in\mathcal{X},\beta\in\mathcal{B}} \left| (-2\tau+1) \int_{x'\delta}^{(x+x^*)'\delta} f_{\epsilon|x}(\epsilon|x)d\epsilon \right| \to 0,
\end{aligned}$$

as $x^* \to 0$. Then the first part of (P7) is satisfied. Assumptions 1 (iii) and 3 (i) yield that there exists f_0 such that $|\epsilon| < f_0$ implies $f_{\epsilon|x}(\epsilon|x) > f_0$ for every $x \in \mathcal{X}$. Thus,

$$\sup_{||\beta-\beta_0||\geq c_0} |E[g(z,\beta)]| = \sup_{||\delta||\geq c_0} \left| -E_x\left[\int_0^{x'\delta} f_{\epsilon|x}(\epsilon|x)d\epsilon\right] \right| \geq \sup_{||\delta||\geq c_0} \left| E_x\left[\int_0^{x'\delta} f_0 I(-f_0 \leq \epsilon \leq f_0)d\epsilon\right] \right| > 0$$

for every $c_0 > 0$, and,

$$\sup_{c_{2n}\geq||\beta-\beta_0||\geq c_{1n}} |E[g(z,\beta)]| \geq \sup_{c_{2n}\geq||\delta||\geq c_{1n}} \left| E_x\left[\int_0^{x'\delta} f_0 I(-f_0 \leq \epsilon \leq f_0)d\epsilon\right] \right| \geq \sup_{c_{2n}\geq||\delta||\geq c_{1n}} ||f_0 E_x[\min\{f_0, x'\delta\}]|| \geq cc_{1n}.$$

for every $0 < c_{1n} < c_{2n} \to 0$. Then the second part of (P7) is satisfied.

From Assumptions 1 (ii) and (iii) and 3 (iii) and a Taylor expansion,

$$\begin{aligned}
& E[|g(z,\beta) - g(z,\beta_0)|^2|x] = E[I(\epsilon \leq x'\delta)|x] + E[I(\epsilon \leq 0)|x] - 2E[I(\epsilon \leq x'\delta)I(\epsilon \leq 0)|x] \\
= \; & F_{\epsilon|x}(x'\delta|x) + F_{\epsilon|x}(0|x) - 2E[I(\epsilon \leq x'\delta)I(\epsilon \leq 0)|x] \leq |F_{\epsilon|x}(x'\delta|x) - F_{\epsilon|x}(0|x)| \\
= \; & |f_{\epsilon|x}(\epsilon^*|x)x'\delta| \leq f_2||x||||\delta|| \leq c||\delta||,
\end{aligned}$$

where ϵ^* is a point joining 0 and $x'\delta$. Then the first part of (P8) is satisfied. Moreover, from Assumption 3 (ii) and (iii) and Taylor expansions,

$$\begin{aligned} & |E[g(z,\beta)|x+x^*] - E[g(z,\beta)|x]| = |-F_{\epsilon|x}((x+x^*)'\delta) + F_{\epsilon|x}(x'\delta)| \\ \leq \ & |-f_{\epsilon|x}(\epsilon^{**}|x)(x+x^*)'\delta + f_{\epsilon|x}(\epsilon^{***}|x)x'\delta| \leq \Big| f_1|\epsilon^{**} - \epsilon^{***}|x'\delta\Big| + f_2\Big|x^{*\prime}\delta\Big| = O(||x^*||)||\delta|| = o(||\delta||), \end{aligned}$$

as $x^* \to 0$, where ϵ^{**} (ϵ^{***}, respectively) is point joining 0 and $x'\delta$ ($(x+x^*)'\delta$, respectively). Then the second part of (P8) is satisfied. Therefore, the regularity conditions of Zhang and Gijbels (2003, Theorem 2 and 3) are implied by Assumptions 1-5.

Proof of Theorem 3.2

From Lemma C.8 and $\hat{\beta}_{CEL} = O_p(n^{-1/2})$, a quadratic expansion of CEL is derived as

$$\mathrm{CELR}(\hat{\beta}_{CEL}) - \mathrm{CELR}(\beta_0) = -n(\hat{\beta}_{CEL} - \beta_0)'\bar{A} - \frac{1}{2}n(\hat{\beta}_{CEL} - \beta_0)'\bar{B}(\hat{\beta}_{CEL} - \beta_0) + o_p(1), \tag{41}$$

where $\bar{A} \equiv n^{-1}\sum_{i=1}^n A_i$, $\bar{B} \equiv n^{-1}\sum_{i=1}^n B_i$, and

$$A_i \equiv I_{in}\frac{\partial E[g(z,\beta_0)|x_i]}{\partial \beta}V(x_i,\beta_0)^{-1}g(z_i,\beta_0), \quad B_i \equiv I_{in}\frac{\partial E[g(z,\beta_0)|x_i]}{\partial \beta}V(x_i,\beta_0)^{-1}\frac{\partial E[g(z,\beta_0)|x_i]}{\partial \beta'}$$

The asymptotic linear form for $\hat{\beta}_{CEL}$ is

$$n^{1/2}(\hat{\beta}_{CEL} - \beta_0) = -\bar{B}^{-1}(n^{1/2}\bar{A}) + o_p(1). \tag{42}$$

Since $\mathrm{rank}\Big(\frac{\partial R(\beta_0)}{\partial \beta}\Big) = r$ (Assumption 6), it contains a nonsingular $r \times r$ submatrix. Without loss of generality, it can be assumed that $\Big[\frac{\partial R(\beta_0)}{\partial \beta^{(q-r+1)}}, \cdots, \frac{\partial R(\beta_0)}{\partial \beta^{(q)}}\Big]$ is such a submatrix. Let $\alpha \equiv (\beta^{(1)}, \ldots, \beta^{(p-r)})$. Using the implicit function theorem, there exist a neighborhood $\mathcal{B}_0$ of β_0, an open set $\mathcal{A}_0$ containing α_0, and twice differentiable function $r : \mathcal{A}_0 \to R^r$, such that every $\beta \in \mathcal{B}_0$ can be expressed as $\beta = \tilde{R}(\alpha) \equiv \begin{bmatrix} \alpha \\ r(\alpha) \end{bmatrix}$ for some $\alpha \in \mathcal{A}_0$. Note that $\tilde{R}$ is twice continuously differentiable and $\mathrm{rank}\Big(\frac{\partial \tilde{R}(\alpha_0)}{\partial \alpha}\Big) = q - r$. Therefore, the restricted CELE is expressed as $\hat{\beta}^R_{CEL} = \tilde{R}(\hat{\alpha}_{CEL})$, where $\hat{\alpha}_{CEL} \equiv \arg\max_{\alpha \in \mathcal{A},} \mathrm{CELR}(\tilde{R}(\alpha))$. Let $D \equiv \frac{\partial \tilde{R}(\alpha_0)}{\partial \alpha'}$. Similar to (41), the quadratic expansion of $\mathrm{CELR}(\tilde{R}(\hat{\alpha}_{CEL})) - \mathrm{CELR}(\beta_0)$ is obtained as

$$\mathrm{CELR}(\tilde{R}(\hat{\alpha}_{CEL})) - \mathrm{CELR}(\beta_0) = -n(\hat{\alpha}_{CEL} - \alpha_0)'D'\bar{A} - \frac{1}{2}n(\hat{\alpha}_{CEL} - \alpha_0)'D'\bar{B}D(\hat{\alpha}_{CEL} - \alpha_0) + o_p(1). \tag{43}$$

The asymptotic linear form for $\hat{\alpha}_{CEL}$ is

$$n^{1/2}(\hat{\alpha}_{CEL} - \alpha_0) = -(D'\bar{B}D)^{-1}D'(n^{1/2}\bar{A}) + o_p(1). \tag{44}$$

From (41)-(44), the CELR test statistic is written as

$$\begin{aligned}
\mathrm{LR}_n &= 2\{\mathrm{CELR}(\hat\beta_{CEL}) - \mathrm{CELR}(\tilde R(\hat\alpha_{CEL}))\} \\
&= -2n^{1/2}(\hat\beta_{CEL}-\beta_0)'(n^{1/2}\bar A) - n^{1/2}(\hat\beta_{CEL}-\beta_0)'\bar B n^{1/2}(\hat\beta_{CEL}-\beta_0) \\
&\quad +2n^{1/2}(\hat\alpha_{CEL}-\alpha_0)'D'(n^{1/2}\bar A) + n^{1/2}(\hat\alpha_{CEL}-\alpha_0)'D'\bar B D n^{1/2}(\hat\alpha_{CEL}-\alpha_0) + o_p(1) \\
&= (n^{1/2}\bar A)'\bar B^{-\frac12}[I_q - \bar B^{\frac12}D(D'\bar B D)^{-1}D'\bar B^{\frac12}]\bar B^{-\frac12}(n^{1/2}\bar A) + o_p(1).
\end{aligned}$$

By the central limit theorem of Pollard (1984, Theorem5, p.141), $\bar B^{-\frac12}(n^{1/2}\bar A) \xrightarrow{d} N(0, I_q)$. From rank$(D) = q-r$, the matrix $[I_q - \bar B^{\frac12}D(D'\bar B D)^{-1}D'\bar B^{\frac12}]$ is idempotent with rank r. Therefore, $\mathrm{LR}_n \xrightarrow{d} \chi^2(r)$.

Proof of Theorem 4.1

(17) implies that for each $\beta \in \mathcal{B}$ and $i = 1, \dots, n$,

$$\begin{aligned}
0 &= \sum_{j=1}^n w_{ji}\frac{\tilde g(z_j,\beta)}{1+\tilde\lambda_i(\beta)\tilde g(z_j,\beta)} \\
&= \sum_{j=1}^n w_{ji}\tilde g(z_j,\beta)\Big[1 - \tilde\lambda_i(\beta)\tilde g(z_j,\beta) + \frac{(\tilde\lambda_i(\beta)\tilde g(z_j,\beta))^2}{1+\tilde\lambda_i(\beta)\tilde g(z_j,\beta)}\Big]. \qquad (45)
\end{aligned}$$

From Lemma C.4 and $\inf_{x\in\mathcal{X}_n,\beta\in\mathcal{B}} V(x,\beta) > 0$, $\inf_{x\in\mathcal{X}_n,\beta\in\mathcal{B}} \hat{\tilde V}(x,\beta) > 0$ with probability 1 and $I_{in}\hat{\tilde V}(x_i,\beta)^{-1}$ is well-defined. Thus, from (45), $I_{in}\tilde\lambda_i(\beta)$ is written as

$$I_{in}\tilde\lambda_i(\beta) = I_{in}\hat{\tilde V}(x_i,\beta)^{-1}\hat{\tilde g}(x_i,\beta) + I_{in}\hat{\tilde V}(x_i,\beta)^{-1}\tilde r_i, \qquad (46)$$

where $\tilde r_i \equiv \sum_{j=1}^n w_{ji}\tilde g(z_j,\beta)\frac{(\tilde\lambda_i(\beta)\tilde g(z_j,\beta))^2}{1+\tilde\lambda_i(\beta)\tilde g(z_j,\beta)}$. From (17),

$$\sum_{j=1}^n w_{ji}\frac{(\tilde\lambda_i(\beta)\tilde g(z_j,\beta))^2}{1+\tilde\lambda_i(\beta)\tilde g(z_j,\beta)} = \sum_{j=1}^n w_{ji}\tilde\lambda_i(\beta)\tilde g(z_j,\beta), \qquad (47)$$

and therefore

$$\begin{aligned}
I_{in}|\tilde r_i| &\le I_{in}\max_{1\le j\le n}|\tilde g(z_j,\beta)|\sum_{j=1}^n w_{ji}\tilde\lambda_i(\beta)\tilde g(z_j,\beta) \le I_{in}c\sum_{j=1}^n w_{ji}\tilde\lambda_i(\beta)\tilde g(z_j,\beta) \\
&\le c\max_{1\le i,j\le n}\sup_{\beta\in\mathcal{B}} I_{in}|\tilde\lambda_i(\beta)\tilde g(z_j,\beta)| = o_p(1), \qquad (48)
\end{aligned}$$

uniformly for $i = 1, \dots, n$ and $\beta \in \mathcal{B}$. The second inequality follows from the boundedness of $\tilde g(z_j,\alpha)$ (Assumption 7 (i)), and the equality follows from Assumption 8, which ensures $Pr\{\max_{1\le i,j\le n}\sup_{\beta\in\mathcal{B}}|\tilde\lambda_i(\beta)\tilde g(z_j,\beta)| = o_p(1)\} = 1$ as $n\to\infty$. Using $\max_{1\le i\le n}\sup_{\beta\in\mathcal{B}} I_{in}\hat{\tilde V}(x_i,\beta) = O_p(1)$ (by Lemma C.4), (46) yields

$$I_{in}\tilde\lambda_i(\beta) = I_{in}\hat{\tilde V}(x_i,\beta)^{-1}\hat{\tilde g}(x_i,\beta) + o_p(1), \qquad (49)$$

uniformly for $i = 1, \ldots, n$ and $\beta \in \mathcal{B}$. By the Taylor expansion, for $i = 1, \ldots, n$ and each $\beta \in \mathcal{B}$,

$$\sum_{j=1}^{n} w_{ji} \log(1 + \tilde{\lambda}_i(\beta)\tilde{g}(z_j, \beta)) = \sum_{j=1}^{n} w_{ji} \Big[\tilde{\lambda}_i(\beta)\tilde{g}(z_j, \beta) - \frac{1}{2}(\tilde{\lambda}_i(\beta)\tilde{g}(z_j, \beta))^2 + \tilde{\zeta}_{ji}\Big],$$

where for some finite $c > 0$, $Pr\{|\tilde{\zeta}_{ji}| \leq c|\tilde{\lambda}_i(\beta)\tilde{g}(z_j, \beta)|^3,\ 1 \leq i, j \leq n\} \to 1$ as $n \to \infty$. Hence, using (49),

$$\begin{aligned} n^{-1}\text{SCELR}(\beta) &= -\frac{1}{n}\sum_{i=1}^{n} I_{in}\Big[\sum_{j=1}^{n} w_{ji}\tilde{\lambda}_i(\beta)\tilde{g}(z_j, \beta) - \frac{1}{2}\sum_{j=1}^{n} w_{ji}(\tilde{\lambda}_i(\beta)\tilde{g}(z_j, \beta))^2 + \sum_{j=1}^{n} w_{ji}\tilde{\zeta}_{ji}\Big] \\ &= -\frac{1}{n}\sum_{i=1}^{n} I_{in}\Big[\tilde{\lambda}_i(\beta)\tilde{\hat{g}}(x_i, \beta) - \frac{1}{2}\tilde{\lambda}_i(\beta)\tilde{\hat{V}}(x_i, \beta)^{-1}\tilde{\lambda}_i(\beta) + \sum_{j=1}^{n} w_{ji}\tilde{\zeta}_{ji}\Big] \\ &= -\frac{1}{2n}\sum_{i=1}^{n} I_{in}\Big[\tilde{\hat{g}}(x_i, \beta)\tilde{\hat{V}}(x_i, \beta)^{-1}\tilde{\hat{g}}(x_i, \beta) + 2\sum_{j=1}^{n} w_{ji}\tilde{\zeta}_{ji}\Big]. \end{aligned}$$

The second term is

$$\Big|\frac{1}{n}\sum_{i=1}^{n} I_{in}\sum_{j=1}^{n} w_{ji}\tilde{\zeta}_{ji}\Big| \leq \frac{c}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} I_{in}w_{ji}|\tilde{\lambda}_i(\beta)\tilde{g}(z_j, \beta)|^3 \leq \frac{c}{n}I_{in}\max_{1\leq i,j\leq n}\sup_{\beta\in\mathcal{B}}|\tilde{\lambda}_i(\beta)\tilde{g}(z_j, \beta)|^3\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ji} = o_p(1),$$

uniformly for $i = 1, \ldots, n$ and $\beta \in \mathcal{B}$. The equality follows from Assumption 8. Thus, the quadratic expansion of the SCELR is

$$n^{-1}\text{SCELR}(\beta) = -\frac{1}{2n}\sum_{i=1}^{n} I_{in}\tilde{\hat{g}}(x_i, \beta)\tilde{\hat{V}}(x_i, \beta)^{-1}\tilde{\hat{g}}(x_i, \beta) + o_p(1),$$

uniformly for $\beta \in \mathcal{B}$. From Lemma C.2 and C.4,

$$\begin{aligned} & \max_{1\leq i\leq n}\sup_{\beta\in\mathcal{B}} I_{in}|\tilde{\hat{g}}(x_i, \beta)\tilde{\hat{V}}(x_i, \beta)^{-1}\tilde{\hat{g}}(x_i, \beta) - \hat{g}(x_i, \beta)\hat{V}(x_i, \beta)^{-1}\hat{g}(x_i, \beta)| \\ \leq\ & \max_{1\leq i\leq n}\sup_{\beta\in\mathcal{B}} I_{in}\Big[|(\tilde{\hat{g}}(x_i, \beta) - \hat{g}(x_i, \beta))\tilde{\hat{V}}(x_i, \beta)^{-1}\tilde{\hat{g}}(x_i, \beta)| + |\hat{g}(x_i, \beta)\tilde{\hat{V}}(x_i, \beta)^{-1}(\tilde{\hat{g}}(x_i, \beta) - \hat{g}(x_i, \beta))| \\ & + |\hat{g}(x_i, \beta)(\tilde{\hat{V}}(x_i, \beta)^{-1} - \hat{V}(x_i, \beta)^{-1})\hat{g}(x_i, \beta)|\Big] \\ =\ & o_p(1). \end{aligned}$$

From Lemma C.7, $n^{-1}\text{SCELR}(\beta) = n^{-1}\text{CELR}(\beta) + o_p(1)$ uniformly for $\beta \in \mathcal{B}$. Therefore, the SCELE is asymptotically equivalent to the CELE. The conclusion is obtained.

Proof of Theorem 4.2

From Lemma C.9 and $\hat{\beta}_{SCEL} = O_p(n^{-1/2})$, a quadratic expansion of SCEL is derived as

$$\text{SCELR}(\hat{\beta}_{SCEL}) - \text{SCELR}(\beta_0) = -n(\hat{\beta}_{SCEL} - \beta_0)'\tilde{\bar{A}} - \frac{1}{2}n(\hat{\beta}_{SCEL} - \beta_0)'\bar{B}(\hat{\beta}_{SCEL} - \beta_0) + o_p(1), \quad (50)$$

where $\tilde{\bar{A}} \equiv n^{-1}\sum_{i=1}^{n}\tilde{A}_i$, $\bar{B} \equiv n^{-1}\sum_{i=1}^{n}B_i$, and

$$\tilde{A}_i \equiv I_{in}\frac{\partial E[g(z,\beta_0)|x_i]}{\partial\beta}V(x_i,\beta_0)^{-1}\tilde{g}(z_i,\beta_0), \quad B_i \equiv I_{in}\frac{\partial E[g(z,\beta_0)|x_i]}{\partial\beta}V(x_i,\beta_0)^{-1}\frac{\partial E[g(z,\beta_0)|x_i]}{\partial\beta'}$$

The asymptotic linear form for $\hat{\beta}_{SCEL}$ is

$$n^{1/2}(\hat{\beta}_{SCEL}-\beta_0) = -\bar{B}^{-1}(n^{1/2}\tilde{\bar{A}}) + o_p(1). \tag{51}$$

Similar to the proof of Theorem 3.2, the restricted SCELE is expressed as $\hat{\beta}^R_{SCEL} = \tilde{R}(\hat{\alpha}_{SCEL})$, where $\hat{\alpha}_{SCEL} \equiv \arg\max_{\alpha\in\mathcal{A},}\mathrm{SCELR}(\tilde{R}(\alpha))$. Therefore, by a similar argument as the proof of Theorem 3.2,

$$\begin{aligned}
\tilde{\mathrm{LR}}_n &= 2\{\mathrm{SCELR}(\hat{\beta}_{SCEL}) - \mathrm{SCELR}(\tilde{R}(\hat{\alpha}_{SCEL}))\} \\
&= -2n^{1/2}(\hat{\beta}_{SCEL}-\beta_0)'(n^{1/2}\tilde{\bar{A}}) - n^{1/2}(\hat{\beta}_{SCEL}-\beta_0)'\bar{B}n^{1/2}(\hat{\beta}_{SCEL}-\beta_0) \\
&\quad +2n^{1/2}(\hat{\alpha}_{SCEL}-\alpha_0)'D'(n^{1/2}\tilde{\bar{A}}) + n^{1/2}(\hat{\alpha}_{SCEL}-\alpha_0)'D'\bar{B}Dn^{1/2}(\hat{\alpha}_{SCEL}-\alpha_0) + o_p(1) \\
&= (n^{1/2}\tilde{\bar{A}})'\bar{B}^{-\frac{1}{2}}[I_q - \bar{B}^{\frac{1}{2}}D(D'\bar{B}D)^{-1}D'\bar{B}^{\frac{1}{2}}]\bar{B}^{-\frac{1}{2}}(n^{1/2}\tilde{\bar{A}}) + o_p(1).
\end{aligned}$$

By the central limit theorem of Pollard (1984, Theorem5, p.141), a similar argument as the proof of Lemma C.2, and $n^{1/2}h_n \to 0$ (Assumption 7 (ii)),

$$\begin{aligned}
\bar{B}^{-1/2}(n^{1/2}\tilde{\bar{A}}) &= \bar{B}^{-1/2}(n^{1/2}\bar{A}) + \bar{B}^{-1/2}(n^{1/2}\tilde{\bar{A}} - n^{1/2}\bar{A}) \\
&= \bar{B}^{-1/2}(n^{1/2}\bar{A}) + O(n^{1/2}h_n) \to N(0, I_q).
\end{aligned}$$

From $\mathrm{rank}(D) = q - r$, the matrix $[I_q - \bar{B}^{\frac{1}{2}}D(D'\bar{B}D)^{-1}D'\bar{B}^{\frac{1}{2}}]$ is idempotent with rank r. Therefore, $\tilde{\mathrm{LR}}_n \xrightarrow{d} \chi^2(r)$.

Proof of Theorem 5.1

See Appendix B.

Proof of Theorem 6.1

$\mathrm{LR}^*(\hat{\beta})$ is decomposed as

$$\begin{aligned}
\mathrm{LR}^*(\hat{\beta}) &= \{\mathrm{LR}^*(\hat{\beta}) - \mathrm{LR}^*(\beta_0)\} + \mathrm{LR}^*(\beta_0) \\
&= \sigma^{-1}\Big\{-2b_n^{\frac{q}{2}}(\mathrm{CELR}(\hat{\beta}) - \mathrm{CELR}(\beta_0)) - b_n^{\frac{q}{2}}(T_n(\hat{\beta}) - T_n(\beta_0))\Big\} + \mathrm{LR}^*(\beta_0) \\
&= O_p(b_n^{\frac{q}{2}}) - \sigma^{-1}nb_n^{\frac{q}{2}}(T_n(\hat{\beta}) - T_n(\beta_0)) + \mathrm{LR}^*(\beta_0),
\end{aligned}$$

The third equality follows from Theorem 3.2. Since $\mathrm{LR}^*(\beta_0) \xrightarrow{d} N(0,1)$ (Corollary 6.1), it is sufficient to show that $b_n^{\frac{q}{2}}(T_n(\hat{\beta}) - T_n(\beta_0)) = o_p(1)$. By the definition of $T_n(\beta)$,

$$b_n^{\frac{q}{2}}(T_n(\hat{\beta}) - T_n(\beta_0)) = b_n^{\frac{q}{2}} \sum_{i=1}^n \sum_{j=1, j\neq i}^n I_{in} w_{ji}^2 \Big\{ \hat{V}(x_i, \hat{\beta})^{-1}(g(z_j,\hat{\beta})^2 - g(z_j,\beta_0)^2) + (\hat{V}(x_i,\hat{\beta})^{-1} - \hat{V}(x_i,\beta_0)^{-1}) g(z_j,\beta_0)^2 \Big\}.$$

Let $K_{\max}$ be the maximum of $K(\cdot)$ (since $K(\cdot)$ is a continuous function with compact support, the maximum always exists), $\hat{\delta} \equiv \hat{\beta} - \beta_0$, and δ^* be a point joining 0 and $\hat{\delta}$. The first term is

$$\begin{aligned}
& b_n^{\frac{q}{2}} \sum_{i=1}^n I_{in} \hat{V}(x_i,\hat{\beta})^{-1} \sum_{j=1, j\neq i}^n w_{ji}^2 (g(z_j,\hat{\beta})^2 - g(z_j,\beta_0)^2) \\
\leq\ & b_n^{\frac{q}{2}} \sum_{i=1}^n I_{in} O_p(1) K_{\max} \hat{f}_i^{-1} \sum_{j=1, j\neq i}^n \frac{1}{n b_n^q} w_{ji} (g(z_j,\hat{\beta})^2 - g(z_j,\beta_0)^2) \\
=\ & O_p(n^{-1} b_n^{-\frac{q}{2}}) \sum_{i=1}^n \sum_{j=1, j\neq i}^n I_{in} w_{ji} (1-2\tau) \Big(I(y_j - x_j'\hat{\beta} \leq 0) - I(y_j - x_j'\beta_0 \leq 0) \Big) \\
=\ & O_p(n^{-1} b_n^{-\frac{q}{2}}) (1-2\tau) \sum_{i=1}^n I_{in} \Big(\tau - E[I(y - x'\hat{\beta} \leq 0) | x_i] + o_p(n^{-1/2}) \Big) \\
=\ & O_p(n^{-1} b_n^{-\frac{q}{2}}) (1-2\tau) \sum_{i=1}^n I_{in} \Big(F_{\epsilon|x}(0|x_i) - F_{\epsilon|x}(x_i'\hat{\delta}|x_i) + o_p(n^{-1/2}) \Big) \\
=\ & O_p(n^{-1} b_n^{-\frac{q}{2}}) (1-2\tau) \sum_{i=1}^n I_{in} \Big(-f_{\epsilon|x}(x_i'\delta^*) x_i'\hat{\delta} \Big) \\
=\ & O_p(n^{-1} b_n^{-\frac{q}{2}}) O_p(n^{1/2}) = O_p(n^{\frac{-1+q\eta}{2}}) = o_p(1),
\end{aligned}$$

where the first inequality follows from Lemma C.3, the first equality follows from the definition of $g(z,\beta)$ and the uniform convergence of the kernel density estimator $\hat{f}_i$, the second equality follows from Lemma C.5, the fourth equality follows from the Taylor expansion, and the fifth equality follows from $\hat{\beta} - \beta_0 = O_p(n^{-1/2})$ and Assumption 3 (iii). Similarly, the second term is

$$\begin{aligned}
& b_n^{\frac{q}{2}} \sum_{i=1}^n \sum_{j=1, j\neq i}^n I_{in} w_{ji}^2 (\hat{V}(x_i,\hat{\beta})^{-1} - \hat{V}(x_i,\beta_0)^{-1}) g(z_j,\beta_0)^2 \\
\leq\ & b_n^{\frac{q}{2}} \sum_{i=1}^n I_{in} (\hat{V}(x_i,\hat{\beta})^{-1} - \hat{V}(x_i,\beta_0)^{-1}) K_{\max} \hat{f}_i^{-1} \sum_{j=1, j\neq i}^n \frac{1}{n b_n^q} w_{ji} g(z_j,\beta_0)^2 \\
=\ & O_p(n^{-1} b_n^{-\frac{q}{2}}) \sum_{i=1}^n I_{in} (\hat{V}(x_i,\hat{\beta})^{-1} - \hat{V}(x_i,\beta_0)^{-1}).
\end{aligned}$$

Now, similar to the first term, $I_{in}(\hat{V}(x_i,\hat{\beta}) - \hat{V}(x_i,\beta_0)) = I_{in} \sum_{j=1}^n w_{ji}(1-2\tau)\Big(I(y_j - x_j'\hat{\beta} \leq 0) - I(y_j - x_j'\beta_0 \leq 0)\Big) = O_p(n^{-1/2})$. Thus, from Tripathi and Kitamura (2002, Lemma C.5), $I_{in}(\hat{V}^{-1}(x_i,\hat{\beta}) - \hat{V}^{-1}(x_i,\beta_0)) = O_p(n^{-1/2})$. Therefore, the second term is also $o_p(1)$, and the conclusion is obtained.

B Derivation of (29) and (31)

First, derive an expansion for $\gamma \equiv E[\tilde{g}(z_i,\beta_0)^2 x_i x_i']^{1/2}\tilde{\gamma}(\beta_0)$. By Lemma C.10, $\gamma = O_p(n^{-1/2}+h_n^p)$, and by the central limit theorem and Assumption 7, $g^j = A^j + \alpha^j = O_p(n^{-1/2}+h_n^p)$. A Taylor expansion of (27) yields

$$0 = g^j - \gamma^k g^{jk} + \gamma^k\gamma^l g^{jkl} - \gamma^k\gamma^l\gamma^m g^{jklm} + O_p\big((n^{-1/2}+h_n^p)^4\big).$$

Using $\alpha^{jk} = \delta^{jk}$, where δ^{jk} is the Kronecker delta,

$$\gamma^j = g^j - \gamma^k A^{jk} + \gamma^k\gamma^l A^{jkl} + \gamma^k\gamma^l\alpha^{jkl} - \gamma^k\gamma^l\gamma^m\alpha^{jklm} + O_p\big((n^{-1/2}+h_n^p)^4\big).$$

Expanding occurrences of γ^k, γ^l, and γ^m in $O_p\big((n^{-1/2}+h_n^p)^3\big)$ terms,

$$\begin{aligned}
\gamma^k A^{jk} &= g^k A^{jk} - g^l A^{jk} A^{kl} + g^l g^m \alpha^{klm} A^{jk} + O_p\big((n^{-1/2}+h_n^p)^4\big),\\
\gamma^k\gamma^l A^{jkl} &= g^k g^l A^{jkl} + O_p\big((n^{-1/2}+h_n^p)^4\big),\\
\gamma^k\gamma^l\alpha^{jkl} &= g^k g^l \alpha^{jkl} - g^l g^m \alpha^{jkl} A^{km} + g^l g^m g^n \alpha^{jkl}\alpha^{kmn} - g^k g^m \alpha^{jkl} A^{lm} + g^k g^m g^n \alpha^{jkl}\alpha^{lmn}\\
&\quad + O_p\big((n^{-1/2}+h_n^p)^4\big),\\
\gamma^k\gamma^l\gamma^m\alpha^{jklm} &= g^k g^l g^m \alpha^{jklm} + O_p\big((n^{-1/2}+h_n^p)^4\big).
\end{aligned}$$

Combining these terms, the expansion of γ is written as

$$\begin{aligned}
\gamma^j &= g^j - g^k A^{jk} + g^l A^{jk} A^{kl} - g^l g^m \alpha^{klm} A^{jk} + g^k g^l A^{jkl} + g^k g^l \alpha^{jkl}\\
&\quad -2 g^l g^m \alpha^{jkl} A^{km} + 2 g^l g^m g^n \alpha^{jkl}\alpha^{kmn} - g^k g^l g^m \alpha^{jklm} + O_p\big((n^{-1/2}+h_n^p)^4\big).
\end{aligned}$$

Next, derive an expansion for $\mathrm{SELR}(\beta_0)$.

$$\begin{aligned}
-2n^{-1}\mathrm{SELR}(\beta_0) &= \frac{2}{n}\sum_{i=1}^{n}\log(1+\gamma' g_i)\\
&= 2\Big(\gamma^j g^j - \frac{1}{2}\gamma^j\gamma^k g^{jk} + \frac{1}{3}\gamma^j\gamma^k\gamma^l g^{jkl} - \frac{1}{4}\gamma^j\gamma^k\gamma^l\gamma^m g^{jklm}\Big) + O_p\big((n^{-1/2}+h_n^p)^5\big)\\
&= 2\Big(\gamma^j g^j - \frac{1}{2}\gamma^j\gamma^j - \frac{1}{2}\gamma^j\gamma^k A^{jk} + \frac{1}{3}g^j g^k g^l A^{jkl} + \frac{1}{3}\gamma^j\gamma^k\gamma^l\alpha^{jkl} - \frac{1}{4}g^j g^k g^l g^m \alpha^{jklm}\Big)\\
&\quad + O_p\big((n^{-1/2}+h_n^p)^5\big).
\end{aligned}$$

Expanding occurrences of γ^j, γ^k, γ^l, and γ^m in $O_p\big((n^{-1/2}+h_n^p)^4\big)$ terms,

$$\begin{aligned}
\gamma^j g^j - \frac{1}{2}\gamma^j\gamma^j &= \frac{1}{2}g^j g^j - \frac{1}{2}g^k g^k A^{jk}A^{jk} - \frac{1}{2}g^k g^k g^l g^l \alpha^{jkl}\alpha^{jkl} + g^k g^k g^l \alpha^{jkl}A^{jk} + O_p\big((n^{-1/2}+h_n^p)^5\big),\\
\gamma^j\gamma^k A^{jk} &= g^j g^k A^{jk} - 2g^j g^l A^{jk}A^{kl} + 2g^j g^l g^m \alpha^{klm}A^{jk} + O_p\big((n^{-1/2}+h_n^p)^5\big),\\
\gamma^j\gamma^k\gamma^l\alpha^{jkl} &= g^j g^k g^l \alpha^{jkl} - 3g^k g^l g^m \alpha^{jkl}A^{jm} - 3g^k g^l g^m g^n \alpha^{jkl}\alpha^{jmn} + O_p\big((n^{-1/2}+h_n^p)^5\big).
\end{aligned}$$

Combining these terms, the expansion of $\mathrm{SELR}(\beta_0)$ is written as

$$\begin{aligned}
-2n^{-1}\mathrm{SELR}(\beta_0) &= g^j g^j - g^j g^k A^{jk} + \frac{2}{3}g^j g^k g^l \alpha^{jkl} + g^j g^k A^{jl}A^{kl} + \frac{2}{3}g^j g^k g^l A^{jkl} - 2g^j g^k g^l \alpha^{jkm}A^{lm}\\
&\quad + g^j g^k g^l g^m \alpha^{jkn}\alpha^{lmn} - \frac{1}{2}g^j g^k g^l g^m \alpha^{jklm} + O_p\big((n^{-1/2}+h_n^p)^5\big). \qquad (52)
\end{aligned}$$

The signed root expansion (29) is obtained by comparing the terms of (52) and $R'R$.

Using $nh_n^{2p}\to 0$ (Assumption 10 (iv)),

$$\begin{aligned}
-2\mathrm{SELR}(\beta_0) &= ng^j g^j + O_p\big(n(n^{-1/2}+h_n^p)^3\big)\\
&= (n^{1/2}A^j + n^{1/2}\alpha^j)(n^{1/2}A^j + n^{1/2}\alpha^j) + O_p\big(n(n^{-1/2}+h_n^p)^3\big)\\
&= (n^{1/2}A^j + O(n^{1/2}h_n^p))(n^{1/2}A^j + O(n^{1/2}h_n^p)) + O_p\big(n^{-1/2}\big)\\
&\xrightarrow{d} \chi_q^2.
\end{aligned}$$

Therefore, the conclusion of Theorem 5.1 is obtained.

Finally, derive the Bartlett correction term in (31). Note that

$$E[R^jR^j] = E[R_1^jR_1^j] + 2E[R_1^jR_2^j] + 2E[R_1^jR_3^j] + E[R_2^jR_2^j] + O_p(n^{-3}).$$

By some lengthy calculations, each term is evaluated as

$$\begin{aligned}
E[R_1^jR_1^j] &= n^{-1}q + O(h_n^{2p}),\\
E[R_1^jR_2^j] &= n^{-2}\Big(\frac{1}{3}\alpha^{jkl}\alpha^{jkl} - \frac{1}{2}\alpha^{jjkk} + \frac{1}{2}nq\Big) + O(h_n^{2p}),\\
E[R_1^jR_3^j] &= n^{-2}\Big(\frac{43}{72}\alpha^{jkl}\alpha^{jkl} - \frac{73}{72}\alpha^{jjk}\alpha^{kll} + \frac{5}{8}\alpha^{jjkk} - \frac{3}{8}nq\Big) + O_p(n^{-3}+h_n^{2p}),\\
E[R_2^jR_2^j] &= n^{-2}\Big(-\frac{7}{36}\alpha^{jkl}\alpha^{jkl} + \frac{1}{36}\alpha^{jjk}\alpha^{kll} + \frac{1}{4}\alpha^{jjkk} - \frac{1}{4}nq\Big) + O_p(n^{-3}+h_n^{2p}).
\end{aligned}$$

Since $\sup_n n^3h_n^{2p} < \infty$, $O(h_n^{2p}) = O(n^{-3})$ and $O_p(n^{-3}+h_n^{2p}) = O_p(n^{-3})$. Combining these terms yields

$$E[nR^jR^j] = q + n^{-1}\Big(\frac{1}{2}\alpha^{jjkk} - \frac{1}{3}\alpha^{jkl}\alpha^{jkl}\Big) + O_p(n^{-2}).$$

Therefore, (31) is derived.

C Auxiliary lemma

Lemma C.1. *Suppose that Assumptions 1-4 hold. Then*

$$\max_{1\leq i\leq n}\sup_{\beta\in\mathcal{B}} I_{in}|\hat{g}(x_i,\beta)-E[g(z,\beta)|x_i]| = o_p(1).$$

Proof. This lemma is a special case of Zhang and Gijbels (2003, Lemma 3). As shown in the proof of Theorem 3.1, Assumptions 1-4 imply the assumptions of Zhang and Gijbels (2003, Lemma 3). □

Lemma C.2. *Suppose that Assumptions 1-4, and 7 hold. Then*

$$\max_{1\leq i\leq n}\sup_{\beta\in\mathcal{B}} I_{in}|\tilde{\hat{g}}(x_i,\beta)-E[g(z,\beta)|x_i]| = o_p(1).$$

Proof. By the triangle inequality and Lemma C.1,

$$\begin{aligned}
&\max_{1\leq i\leq n}\sup_{\beta\in\mathcal{B}} I_{in}|\tilde{\hat{g}}(x_i,\beta)-E[g(z,\beta)|x_i]| \\
\leq\;& \max_{1\leq i\leq n}\sup_{\beta\in\mathcal{B}} I_{in}|\tilde{\hat{g}}(x_i,\beta)-\hat{g}(x_i,\beta)| + \max_{1\leq i\leq n}\sup_{\beta\in\mathcal{B}} I_{in}|\hat{g}(x_i,\beta)-E[g(z,\beta)|x_i]| \\
=\;& \max_{1\leq i\leq n}\sup_{\beta\in\mathcal{B}} I_{in}|\tilde{\hat{g}}(x_i,\beta)-\hat{g}(x_i,\beta)| + o_p(1).
\end{aligned}$$

Thus, it is sufficient to check the stochastic order of the first term.

$$\begin{aligned}
&|\tilde{\hat{g}}(x_i,\beta)-\hat{g}(x_i,\beta)| = \Big|\sum_{j=1}^n w_{ji}\Big(I(y_j-x_j'\beta\leq 0)-H\Big(-\frac{y_j-x_j'\beta}{h_n}\Big)\Big)\Big| \\
\leq\;& \sum_{j=1}^n w_{ji}\Big|I(y_j-x_j'\beta\leq 0)-H\Big(-\frac{y_j-x_j'\beta}{h_n}\Big)\Big| \\
=\;& \sum_{j=1}^n w_{ji}I(|y_j-x_j'\beta|\leq h_n)\Big|I(y_j-x_j'\beta\leq 0)-H\Big(-\frac{y_j-x_j'\beta}{h_n}\Big)\Big| \\
\leq\;& \sum_{j=1}^n w_{ji}I(|y_j-x_j'\beta|\leq h_n) \\
=\;& \sum_{j=1}^n w_{ji}(I(\epsilon_j\leq h_n+x_j'\delta)-I(\epsilon_j\leq -h_n+x_j'\delta)) \\
\leq\;& \sum_{j=1}^n w_{ji}\Big|I(\epsilon_j\leq h_n+x_j'\delta)-F_{\epsilon|x}(h_n+x_i'\delta)\Big| + \sum_{j=1}^n w_{ji}\Big|I(\epsilon_j\leq -h_n+x_j'\delta)-F_{\epsilon|x}(-h_n+x_i'\delta)\Big| \\
&+\Big|F_{\epsilon|x}(h_n+x_i'\delta)-F_{\epsilon|x}(-h_n+x_i'\delta)\Big|.
\end{aligned}$$

The second equality follows from the fact that the summand differs from zero only if $|y_j-x_j'\beta|\leq h_n$ by Assumption 7 (i). The second inequality follows from $\Big|I(y_j-x_j'\beta\leq 0)-H\Big(-\frac{y_j-x_j'\beta}{h_n}\Big)\Big|\leq 1$ by Assumption

7 (i). The first and third inequalities follow from the triangular inequality. By Lemma C.1, the first and second terms are

$$\max_{1\le i\le n}\sup_{\beta\in\mathcal{B}} I_{in}\Big|\sum_{j=1}^{n} w_{ji} I(\epsilon_j \le h_n + x_j'\delta) - F_{\epsilon|x}(h_n + x_i'\delta)\Big| = o_p(1),$$

and

$$\max_{1\le i\le n}\sup_{\beta\in\mathcal{B}} I_{in}\Big|\sum_{j=1}^{n} w_{ji} I(\epsilon_j \le -h_n + x_j'\delta) - F_{\epsilon|x}(-h_n + x_i'\delta)\Big| = o_p(1),$$

respectively. By Assumption 3 (ii) and a Taylor expansion, the third term is

$$\begin{aligned}\Big|F(h_n + x_i'\delta) - F(-h_n + x_i'\delta)\Big| &= h_n\Big|f_{\epsilon|x}(x_i'\delta + h^*) - f_{\epsilon|x}(x_i'\delta - h^{**})\Big| \\ &\le f_1 h_n |h^* + h^{**}| = O_p(h_n),\end{aligned}$$

where h^* and h^{**} are points on the line joining 0 and h_n. Therefore, $\max_{1\le i\le n}\sup_{\beta\in\mathcal{B}} I_{in}|\tilde{\hat{g}}(x_i,\beta) - \hat{g}(x_i,\beta)| = o_p(1)$, and then the conclusion is obtained. □

Lemma C.3. *Suppose that Assumptions 1-4 hold. Then*

$$\max_{1\le i\le n}\sup_{\beta\in\mathcal{B}} I_{in}|\hat{V}(x_i,\beta) - V(x_i,\beta)| = o_p(1).$$

Proof. This lemma is a special case of Zhang and Gijbels (2003, Lemma 4). As shown in the proof of Theorem 3.1, Assumptions 1-4 imply the assumptions of Zhang and Gijbels (2003, Lemma 4). □

Lemma C.4. *Suppose that Assumptions 1-4, and 7 hold. Then*

$$\max_{1\le i\le n}\sup_{\beta\in\mathcal{B}} I_{in}|\tilde{\hat{V}}(x_i,\beta) - V(x_i,\beta)| = o_p(1).$$

Proof. By the triangle inequality and Lemma C.3,

$$\begin{aligned}&\max_{1\le i\le n}\sup_{\beta\in\mathcal{B}} I_{in}|\tilde{\hat{V}}(x_i,\beta) - V(x_i,\beta)| \\ \le\ &\max_{1\le i\le n}\sup_{\beta\in\mathcal{B}} I_{in}|\tilde{\hat{V}}(x_i,\beta) - \hat{V}(x_i,\beta)| + \max_{1\le i\le n}\sup_{\beta\in\mathcal{B}} I_{in}|\hat{V}(x_i,\beta) - V(x_i,\beta)| \\ =\ &\max_{1\le i\le n}\sup_{\beta\in\mathcal{B}} I_{in}|\tilde{\hat{V}}(x_i,\beta) - \hat{V}(x_i,\beta)| + o_p(1).\end{aligned}$$

Thus, it is sufficient to check the stochastic order of the first term.

$$
\begin{aligned}
|\tilde{\hat{V}}(x_i,\beta) - \hat{V}(x_i,\beta)| &= \Big|\sum_{j=1}^n w_{ji}\Big(\tau - H\Big(-\frac{y_j - x_j'\beta}{h_n}\Big)\Big)^2 - \sum_{j=1}^n w_{ji}(\tau - I(y_j - x_j'\beta \le 0))^2\Big| \\
\le \quad & \sum_{j=1}^n w_{ji}\Big|2\tau\Big(I(y_j - x_j'\beta \le 0) - H\Big(-\frac{y_j - x_j'\beta}{h_n}\Big)\Big)\Big| + \sum_{j=1}^n w_{ji}\Big|H\Big(-\frac{y_j - x_j'\beta}{h_n}\Big)^2 - I(y_j - x_j'\beta \le 0)\Big| \\
= \quad & \sum_{j=1}^n w_{ji} I(|y_j - x_j'\beta| \le h_n)\Big\{\Big|2\tau\Big(I(y_j - x_j'\beta \le 0) - H\Big(-\frac{y_j - x_j'\beta}{h_n}\Big)\Big)\Big| + \Big|H\Big(-\frac{y_j - x_j'\beta}{h_n}\Big)^2 - I(y_j - x_j'\beta \le 0)\Big|\Big\} \\
\le \quad & \sum_{j=1}^n w_{ji} I(|y_j - x_j'\beta| \le h_n)(2\tau + 1).
\end{aligned}
$$

The second equality follows from the fact that the summand differs from zero only if $|y_j - x_j'\beta| \le h_n$ by Assumption 7 (i). The second inequality follows from the boundedness of H and $I(\cdot)$. By the same argument as in Lemma C.2, it is shown that $\max_{1\le i\le n}\sup_{\beta\in\mathcal{B}} I_{in}|\tilde{\hat{V}}(x_i,\beta) - \hat{V}(x_i,\beta)| = o_p(1)$. Therefore, the conclusion is obtained. $\square$

Lemma C.5. *Suppose that Assumptions 1-4 hold. Then*

$$\sum_{j=1}^n w_{ji}(g(z_j,\beta) - g(z_j,\beta_0)) - E[g(z,\beta)|x_i] = o_p(\max\{||\beta-\beta_0||, n^{-1/2}\}),$$

uniformly for $x \in \mathcal{X}_n$, $||\beta - \beta_0|| \le r_n \to 0$.

Proof. This lemma is a special case of Zhang and Gijbels (2003, Lemma 7). As shown in the proof of Theorem 3.1, Assumptions 1-4 imply the assumptions of Zhang and Gijbels (2003, Lemma 7). $\square$

Lemma C.6. *Suppose that Assumptions 1-4, and 7 hold. Then*

$$\sum_{j=1}^n w_{ji}(\tilde{g}(z_j,\beta) - \tilde{g}(z_j,\beta_0)) - E[g(z,\beta)|x_i] = o_p(\max\{||\beta-\beta_0||, n^{-1/2}\}),$$

uniformly for $x \in \mathcal{X}_n$, $||\beta - \beta_0|| \le r_n \to 0$.

Proof. From the triangular inequality and Lemma C.5,

$$
\begin{aligned}
& \sum_{j=1}^n w_{ji}(\tilde{g}(z_j,\beta) - \tilde{g}(z_j,\beta_0)) - E[g(z,\beta)|x_i] \\
\le \quad & \sum_{j=1}^n w_{ji}|\tilde{g}(z_j,\beta) - g(z_j,\beta)| + \sum_{j=1}^n w_{ji}|\tilde{g}(z_j,\beta_0) - g(z_j,\beta_0)| + \sum_{j=1}^n w_{ji}|(g(z_j,\beta) - g(z_j,\beta_0)) - E[g(z,\beta)|x_i]| \\
= \quad & \sum_{j=1}^n w_{ji}|\tilde{g}(z_j,\beta) - g(z_j,\beta)| + \sum_{j=1}^n w_{ji}|\tilde{g}(z_j,\beta_0) - g(z_j,\beta_0)| + o_p(\max\{||\beta-\beta_0||, n^{-1/2}\}),
\end{aligned}
$$

uniformly for $x \in \mathcal{X}_n$, $||\beta - \beta_0|| \leq r_n \to 0$. From the proof of Lemma C.2,

$$\sum_{j=1}^{n} w_{ji} |\tilde{g}(z_j, \beta) - g(z_j, \beta)| \leq O(h_n),$$

uniformly for $\beta \in \mathcal{B}$. Using $h_n = o(n^{-1/2})$ (Assumption 7 (ii)), the conclusion is obtained. $\square$

Lemma C.7. *Suppose that Assumptions 1-5 hold. Then*

$$n^{-1} CELR(\beta) = -\frac{1}{2n} \sum_{i=1}^{n} I_{in} \hat{g}(x_i, \beta) \hat{V}(x_i, \beta)^{-1} \hat{g}(x_i, \beta) + o_p(1),$$

uniformly for $\beta \in \mathcal{B}$.

Proof. (8) implies that for each $\beta \in \mathcal{B}$ and $i = 1, \ldots, n$,

$$\begin{aligned} 0 &= \sum_{j=1}^{n} w_{ji} \frac{g(z_j, \beta)}{1 + \lambda_i(\beta) g(z_j, \beta)} \\ &= \sum_{j=1}^{n} w_{ji} g(z_j, \beta) \Big[1 - \lambda_i(\beta) g(z_j, \beta) + \frac{(\lambda_i(\beta) g(z_j, \beta))^2}{1 + \lambda_i(\beta) g(z_j, \beta)} \Big]. \end{aligned} \tag{53}$$

From Lemma C.3 and $\inf_{x \in \mathcal{X}_n, \beta \in \mathcal{B}} V(x, \beta) > 0$, $\inf_{x \in \mathcal{X}_n, \beta \in \mathcal{B}} \hat{V}(x, \beta) > 0$ with probability 1 and $I_{in} \hat{V}(x_i, \beta)^{-1}$ is well-defined. Thus, solving for $\lambda_i(\beta)$ in (53) yields

$$I_{in} \lambda_i(\beta) = I_{in} \hat{V}(x_i, \beta)^{-1} \hat{g}(x_i, \beta) + I_{in} \hat{V}(x_i, \beta)^{-1} r_i, \tag{54}$$

where $r_i \equiv \sum_{j=1}^{n} w_{ji} g(z_j, \beta) \frac{(\lambda_i(\beta) g(z_j, \beta))^2}{1 + \lambda_i(\beta) g(z_j, \beta)}$. From (8),

$$\sum_{j=1}^{n} w_{ji} \frac{(\lambda_i(\beta) g(z_j, \beta))^2}{1 + \lambda_i(\beta) g(z_j, \beta)} = \sum_{j=1}^{n} w_{ji} \lambda_i(\beta) g(z_j, \beta), \tag{55}$$

and therefore

$$\begin{aligned} I_{in} |r_i| &\leq I_{in} \max_{1 \leq j \leq n} |g(z_j, \beta)| \sum_{j=1}^{n} w_{ji} \lambda_i(\beta) g(z_j, \beta) \leq I_{in} c \sum_{j=1}^{n} w_{ji} \lambda_i(\beta) g(z_j, \beta) \\ &\leq c \max_{1 \leq i, j \leq n} \sup_{\beta \in \mathcal{B}} I_{in} |\lambda_i(\beta) g(z_j, \beta)| = o_p(1), \end{aligned} \tag{56}$$

uniformly for $i = 1, \ldots, n$ and $\beta \in \mathcal{B}$. The second inequality follows from the boundedness of $g(z_j, \alpha)$, and the equality follows from Assumption 5, which ensures $Pr\{\max_{1 \leq i, j \leq n} \sup_{\beta \in \mathcal{B}} |\lambda_i(\beta) g(z_j, \beta)| = o_p(1)\} = 1$ as $n \to \infty$. Using $\max_{1 \leq i \leq n} \sup_{\beta \in \mathcal{B}} I_{in} \hat{V}(x_i, \beta) = O_p(1)$ (by Lemma C.3), (54) yields

$$I_{in} \lambda_i(\beta) = I_{in} \hat{V}(x_i, \beta)^{-1} \hat{g}(x_i, \beta) + o_p(1), \tag{57}$$

uniformly for $i = 1, \ldots, n$ and $\beta \in \mathcal{B}$. By the Taylor expansion, for $i = 1, \ldots, n$ and each $\beta \in \mathcal{B}$,

$$\sum_{j=1}^{n} w_{ji} \log(1 + \lambda_i(\beta)g(z_j, \beta)) = \sum_{j=1}^{n} w_{ji} \Big[\lambda_i(\beta)g(z_j, \beta) - \frac{1}{2}(\lambda_i(\beta)g(z_j, \beta))^2 + \zeta_{ji}\Big],$$

where for some finite $c > 0$, $Pr\{|\zeta_{ji}| \leq c|\lambda_i(\beta)g(z_j, \beta)|^3,\ 1 \leq i, j \leq n\} \to 1$ as $n \to \infty$. Hence, using (57),

$$\begin{aligned} n^{-1}\mathrm{CELR}(\beta) &= -\frac{1}{n}\sum_{i=1}^{n} I_{in}\Big[\sum_{j=1}^{n} w_{ji}\lambda_i(\beta)g(z_j, \beta) - \frac{1}{2}\sum_{j=1}^{n} w_{ji}(\lambda_i(\beta)g(z_j, \beta))^2 + \sum_{j=1}^{n} w_{ji}\zeta_{ji}\Big] \\ &= -\frac{1}{n}\sum_{i=1}^{n} I_{in}\Big[\lambda_i(\beta)\hat{g}(x_i, \beta) - \frac{1}{2}\lambda_i(\beta)\hat{V}(x_i, \beta)^{-1}\lambda_i(\beta) + \sum_{j=1}^{n} w_{ji}\zeta_{ji}\Big] \\ &= -\frac{1}{2n}\sum_{i=1}^{n} I_{in}\Big[\hat{g}(x_i, \beta)\hat{V}(x_i, \beta)^{-1}\hat{g}(x_i, \beta) + 2\sum_{j=1}^{n} w_{ji}\zeta_{ji}\Big]. \end{aligned}$$

The second term is

$$\Big|\frac{1}{n}\sum_{i=1}^{n} I_{in}\sum_{j=1}^{n} w_{ji}\zeta_{ji}\Big| \leq \frac{c}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ji}I_{in}|\lambda_i(\beta)g(z_j, \beta)|^3 \leq \frac{c}{n}\max_{1\leq i,j\leq n}\sup_{\beta\in\mathcal{B}} I_{in}|\lambda_i(\beta)g(z_j, \beta)|^3\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ji} = o_p(1),$$

uniformly for $i = 1, \ldots, n$ and $\beta \in \mathcal{B}$. The equality follows from Assumption 5. Thus, the quadratic expansion of the CELR is

$$n^{-1}\mathrm{CELR}(\beta) = -\frac{1}{2n}\sum_{i=1}^{n} I_{in}\hat{g}(x_i, \beta)\hat{V}(x_i, \beta)^{-1}\hat{g}(x_i, \beta) + o_p(1),$$

uniformly for $\beta \in \mathcal{B}$. The conclusion is obtained. □

Lemma C.8. *Suppose that Assumptions 1-5 hold. Then*

$$CELR(\beta) - CELR(\beta_0) = -(\beta - \beta_0)'\sum_{i=1}^{n} A_i - \frac{1}{2}(\beta - \beta_0)'\Big(\sum_{i=1}^{n} B_i\Big)(\beta - \beta_0) + o_p(\max\{n||\beta - \beta_0||^2, 1\}),$$

uniformly for $\beta \in \mathcal{B}$ *such that* $||\beta - \beta_0|| \leq r_n \to 0$*, where*

$$A_i \equiv I_{in}\frac{\partial E[g(z, \beta_0)|x_i]}{\partial \beta}V(x_i, \beta_0)^{-1}g(z_i, \beta_0), \quad B_i \equiv I_{in}\frac{\partial E[g(z, \beta_0)|x_i]}{\partial \beta}V(x_i, \beta_0)^{-1}\frac{\partial E[g(z, \beta_0)|x_i]}{\partial \beta'}.$$

Proof. From Lemma C.7,

$$\begin{aligned}
& n^{-1}(\mathrm{CELR}(\beta)-\mathrm{CELR}(\beta_0)) \\
= \; & -\frac{1}{2n}\sum_{i=1}^n I_{in}\hat{g}(x_i,\beta)\hat{V}(x_i,\beta)^{-1}\hat{g}(x_i,\beta)+\frac{1}{2n}\sum_{i=1}^n I_{in}\hat{g}(x_i,\beta_0)\hat{V}(x_i,\beta_0)^{-1}\hat{g}(x_i,\beta_0)+o_p(1) \\
= \; & -\frac{1}{2n}\sum_{i=1}^n I_{in}\Big[\hat{g}(x_i,\beta)\hat{V}(x_i,\beta_0)^{-1}\hat{g}(x_i,\beta)-\hat{g}(x_i,\beta_0)\hat{V}(x_i,\beta_0)^{-1}\hat{g}(x_i,\beta_0)\Big]+o_p(1) \\
= \; & -\frac{1}{2n}\sum_{i=1}^n I_{in}(\hat{g}(x_i,\beta)-\hat{g}(x_i,\beta_0))\hat{V}(x_i,\beta_0)^{-1}(\hat{g}(x_i,\beta)+\hat{g}(x_i,\beta_0))+o_p(1) \\
= \; & -\frac{1}{2n}\sum_{i=1}^n I_{in}H(x_i,\beta)\sum_{j=1}^n w_{ji}(g(z_j,\beta)+g(z_j,\beta_0))+o_p(1) \\
= \; & -\frac{1}{2n}\sum_{j=1}^n\Big(\sum_{i=1}^n I_{in}w_{ji}H(x_i,\beta)\Big)(g(z_j,\beta)+g(z_j,\beta_0))+o_p(1),
\end{aligned}$$

uniformly for $\beta\in\mathcal{B}$ such that $||\beta-\beta_0||\le r_n$, where the term $(1+o_p(1))$ is omitted, $H(x_i,\beta)\equiv(\hat{g}(x_i,\beta)-\hat{g}(x_i,\beta_0))\hat{V}(x_i,\beta_0)^{-1}$, the second equality follows from Lemma C.3, and the last equality follows from the exchange of the order of summations. From Zhang and Gijbels (2003, Lemma 7 and 8),

$$\begin{aligned}
\sum_{i=1}^n I_{in}w_{ji}H(x_i,\beta) &= \sum_{i=1}^n I_{in}w_{ji}\Big(E[g(z,\beta)|x_i]V(x_i,\beta_0)^{-1}(1+o_p(1))+o_p(\max\{||\beta-\beta_0||,n^{-1/2}\})\Big) \\
&= E[g(z,\beta)|x_j]V(x_j,\beta_0)^{-1}(1+o_p(1))+o_p(\max\{||\beta-\beta_0||,n^{-1/2}\}),
\end{aligned}$$

uniformly for $\beta\in\mathcal{B}$ such that $||\beta-\beta_0||\le r_n$ and $x\in\mathcal{X}_n$. Therefore, by omitting the term $(1+o_p(1))$

$$\begin{aligned}
& \mathrm{CELR}(\beta)-\mathrm{CELR}(\beta_0) \\
= \; & -\frac{1}{2n}\sum_{j=1}^n I_{jn}\Big(E[g(z,\beta)|x_j]V(x_j,\beta_0)^{-1}+o_p(\max\{||\beta-\beta_0||,n^{-1/2}\})\Big)(g(z_j,\beta)+g(z_j,\beta_0))+o_p(1) \\
= \; & -\frac{1}{2n}\sum_{j=1}^n I_{jn}E[g(z,\beta)|x_j]V(x_j,\beta_0)^{-1}(g(z_j,\beta)-g(z_j,\beta_0))-\frac{1}{2n}\sum_{j=1}^n I_{jn}E[g(z,\beta)|x_j]V(x_j,\beta_0)^{-1}2g(z_j,\beta_0) \\
& -o_p(\max\{||\beta-\beta_0||,n^{-1/2}\})\frac{1}{2n}\sum_{j=1}^n I_{jn}(g(z_j,\beta)-g(z_j,\beta_0)+2g(z_j,\beta_0)) \\
= \; & -\frac{1}{2n}\sum_{j=1}^n I_{jn}E[g(z,\beta)|x_j]V(x_j,\beta_0)^{-1}E[g(z,\beta)|x_j]-\frac{1}{n}\sum_{j=1}^n I_{jn}E[g(z,\beta)|x_j]V(x_j,\beta_0)^{-1}g(z_j,\beta_0) \\
& +o_p(\max\{||\beta-\beta_0||^2,n^{-1}\}) \\
= \; & -\frac{1}{2}(\beta-\beta_0)'\Big[\frac{1}{n}\sum_{j=1}^n I_{jn}\frac{\partial E[g(z,\beta_0)|x_j]}{\partial\beta}V(x_j,\beta_0)^{-1}\frac{\partial E[g(z,\beta_0)|x_j]}{\partial\beta'}\Big](\beta-\beta_0) \\
& -(\beta-\beta_0)'\frac{1}{n}\sum_{j=1}^n I_{jn}\frac{\partial E[g(z,\beta_0)|x_j]}{\partial\beta}V(x_j,\beta_0)^{-1}g(z_j,\beta_0)+o_p(\max\{||\beta-\beta_0||^2,n^{-1}\})
\end{aligned}$$

The third equality follows from Zhang and Gijbels (2003, Lemma 9 and 10). The fourth equality follows from the continuity of $E[g(z,\beta_0)|x]$ with respect to β and Taylor expansion. Therefore, the conclusion is obtained. $\square$

Lemma C.9. *Suppose that Assumptions 1-4, 7, and 8 hold. Then*

$$SCELR(\beta) - SCELR(\beta_0) = -(\beta-\beta_0)'\sum_{i=1}^n \tilde{A}_i - \frac{1}{2}(\beta-\beta_0)'\Big(\sum_{i=1}^n B_i\Big)(\beta-\beta_0) + o_p(\max\{n||\beta-\beta_0||^2, 1\}),$$

uniformly for $\beta \in \mathcal{B}$ such that $||\beta-\beta_0|| \le r_n \to 0$, where

$$\tilde{A}_i \equiv I_{in}\frac{\partial E[g(z,\beta_0)|x_i]}{\partial \beta}V(x_i,\beta_0)^{-1}\tilde{g}(z_i,\beta_0), \quad B_i \equiv I_{in}\frac{\partial E[g(z,\beta_0)|x_i]}{\partial \beta}V(x_i,\beta_0)^{-1}\frac{\partial E[g(z,\beta_0)|x_i]}{\partial \beta'}.$$

Proof. The proof is similar to Lemma C.8; instead of Lemma C.5, use Lemma C.6. $\square$

Lemma C.10. *Suppose that Assumptions 1, 3, 7, 9, and 10 hold. Then*

$$\tilde{\gamma}(\beta_0) = O_p(n^{-1/2} + h_n^p).$$

Proof. Let $\tilde{g}_{ui} \equiv \tilde{g}_u(z_i,\beta_0)$, $\tilde{\gamma} \equiv \tilde{\gamma}(\beta_0)$, and $\tilde{\gamma} = ||\tilde{\gamma}||\theta$, where θ is a unit vector. From (27),

$$\begin{aligned} 0 &= \frac{1}{n}\Big|\sum_{i=1}^n \frac{-\theta'\tilde{g}_{ui}}{1+\tilde{\gamma}'\tilde{g}_{ui}}\Big| = \frac{1}{n}\Big|\sum_{i=1}^n \frac{||\tilde{\gamma}||\theta'\tilde{g}_{ui}\tilde{g}_{ui}'\theta}{1+\tilde{\gamma}'\tilde{g}_{ui}} - \theta'\sum_{i=1}^n \tilde{g}_{ui}\Big| \\ &\ge \frac{1}{n}\Big|\sum_{i=1}^n \frac{||\tilde{\gamma}||\theta'\tilde{g}_{ui}\tilde{g}_{ui}'\theta}{1+\tilde{\gamma}'g_{ui}}\Big| - \frac{1}{n}\Big|\theta'\sum_{i=1}^n \tilde{g}_{ui}\Big| \\ &\ge \frac{||\tilde{\gamma}||}{\big|1+||\tilde{\gamma}||\max_{1\le i\le n}||\tilde{g}_{ui}||\big|}\frac{1}{n}\Big|\theta'\sum_{i=1}^n \tilde{g}_{ui}\tilde{g}_{ui}'\theta\Big| - \frac{1}{n}\Big|\theta'\sum_{i=1}^n \tilde{g}_{ui}\Big|. \end{aligned}$$

The second inequality follows from $\hat{p}_i = n^{-1}(1+\tilde{\gamma}'g_{ui})^{-1} \ge 0$. Letting $\max_{1\le i\le n}||x_i|| \equiv M < \infty$ (since the support of x is compact),

$$\max_{1\le i\le n}||\tilde{g}_{ui}|| \le M\max_{1\le i\le n}\Big|\tau - H\Big(-\frac{y_i - x_i'\beta}{h_n}\Big)\Big| \le 2M,$$

where the second inequality follows from Assumption 7 (i). Thus, letting $\bar{g}_{u1} \equiv n^{-1}\sum_{i=1}^n \tilde{g}_{ui}$ and $\bar{g}_{u2} \equiv n^{-1}\sum_{i=1}^n \tilde{g}_{ui}\tilde{g}_{ui}'$,

$$||\tilde{\gamma}||\theta'\bar{g}_{u2}\theta \le |\theta'\bar{g}_{u1}|\big|1+||\tilde{\gamma}||\max_{1\le i\le n}||\tilde{g}_{ui}||\big| \le |\theta'\bar{g}_{u1}| + 2||\tilde{\gamma}||\,|\theta'\bar{g}_{u1}|M$$

and then

$$||\tilde{\gamma}||\Big(\theta'\bar{g}_{u2}\theta - 2M|\theta'\bar{g}_{u1}|\Big) \le |\theta'\bar{g}_{u1}|.$$

By Assumption 9 and the weak law of large numbers, $\theta' \bar{g}_{u2} \theta = O_p(1)$. By the central limit theorem, $\bar{g}_{u1} - E[\bar{g}_{u1}] = O_p(n^{-1/2})$. By Assumptions 1, 2 (ii), and 7, $E[\bar{g}_{u1}] = O(h_n^p)$. Combining these results,

$$||\tilde{\gamma}||\{O_p(1) + O(1)O_p(n^{-1/2} + h_n^p)\} \leq O_p(n^{-1/2} + h_n^p).$$

From Assumption 10 (ii), $\tilde{\gamma}(\beta_0) = O_p(n^{-1/2} + h_n^p)$. The conclusion is obtained. □

References

[1] Ai, C. (1997) A semiparametric maximum likelihood estimator, *Econometrica*, 65, 933-963.

[2] Ai, C. and X. Chen (1999) Efficient estimation of models with conditional moment restrictions containing unknown functions, Working paper.

[3] Andrews, D.W.K. (1994) Empirical process methods in econometrics, in R.F. Engle and D.L. McFadden, eds., *Handbook of Econometrics*, vol. IV, 2247-2294, Elsevier, Amsterdam.

[4] Andrews, D.W.K. (1995) Nonparametric kernel estimation for semiparametric models, *Econometric Theory*, 11, 560-596.

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

[6] Barrodale, I. and F. Roberts (1973) An improved algorithm for discrete ℓ_1 linear approximation, *SIAM Journal of Numerical Analysis*, 10, 839-848.

[7] Bierens, H.J. and D.K. Ginther (2001) Integrated conditional moment testing of quadratic regression models, *Empirical Economics*, 26, 307-324.

[8] Buchinsky, M. (1994) Changes in the U.S. wage structure 1963-1987: application of quantile regression, *Econometrica*, 62, 405-458.

[9] Buchinsky, M. (1995a) Quantile regression Box-Cox transformation model, and the U.S. wage structure, 1963-1987, *Journal of Econometrics*, 65, 109-154.

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

[11] Buchinsky, M. (1998) Recent advances in quantile regression models: a practical guideline for empirical research, *Journal of Human Resources*, 33, 88-126.

[12] Chamberlain, G. (1987) Asymptotic efficiency in estimation with conditional moment restrictions, *Journal of Econometrics*, 34, 305-334.

[13] Chamberlain, G. (1994) Quantile regression, censoring, and the structure of wage, in C. Sims (ed.) *Advances in Econometrics*, New York: Cambridge University Press.

[14] Chen, S.X. and H. Cui (2002) On Bartlett correction of empirical likelihood in the presence of nuisance parameters, Working paper.

[15] Chen, S.X. and H. Cui (2003) On the second order properties of empirical likelihood for generalized estimation equations, Working paper.

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

[17] Cressie, N. and T. Read (1984) Multinomial goodness-of-fit tests, *Journal of the Royal Statistical Society*, B46, 440-464.

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

[19] Hansen, L.P. (1982) Large sample properties of generalized method of moments estimators, *Econometrica*, 50, 1029-1054.

[20] Hansen, L.P., J. Heaton and A. Yaron (1996) Finite-sample properties of some alternative GMM estimators, *Journal of Business and Economic Statistics*, 14, 262-280.

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

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

[23] Judd, K.L. (1998) *Numerical Methods in Economics*, MIT Press.

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

[25] Koenker, R. and G. Bassett (1978b) The asymptotic distribution of the least absolute error estimator, *Journal of the American Statistical Association*, 73, 618-622.

[26] Koenker, R. (1994) Confidence intervals for regression quantiles, in Mandl, P. and M. Huskova (eds.), *Proceedings of the 5th Prague Symposium on Asymptotic Statistics*, Heidelberg: Physica-Verlag.

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

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

[29] Kitamura, Y. (2003) A likelihood-based approach to the analysis of a class of nested and non-nested models, Working paper.

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

[31] Kitamura, Y., G. Tripathi and H. Ahn (2003) Empirical likelihood-based inference in conditional moment restriction models, Working paper.

[32] Kim, T.-H. and H. White (2002) Estimation, inference, and specification testing for possibly misspecified quantile regression, Working paper.

[33] Koenker, R. and Z. Xiao (2002) Inference on the quantile regression process, *Econometrica*, 70, 1583-1612.

[34] LeBlanc, M. and J. Crowley (1995) Semiparametric regression functionals, *Journal of the American Statistical Association*, 90, 95-105.

[35] Linton, O. (2002) Edgeworth approximations for semiparametric instrumental variable estimators and test statistics, *Journal of Econometrics*, 106, 325-368.

[36] Müller, H.-G. (1984) Smooth optimum kernel estimators of densities, regression curves and modes, *Annals of Statistics*, 12, 766-774.

[37] Nelder, J.A. and R. Mead (1965) A simplex algorithm for function minimization, *Computer Journal*, 7, 308-313.

[38] Newey, W.K. (1990) Efficient instrumental variables estimation of nonlinear models, *Econometrica* 58, 809-837.

[39] Newey, W.K. (1993) Efficient estimation of models with conditional moment restrictions, in: G.S. Maddala, C.R. Rao and H.D. Vinod, eds., *Handbook of Statistics*, Vol. 11, 419-454, North-Holland, Amsterdam.

[40] Newey, W.K. and J.L. Powell (1990) Efficient estimation of linear and type I censored regression models under conditional quantile restrictions, *Econometric Theory*, 6, 295-317.

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

[42] Otsu, T. (2003a) Penalized empirical likelihood estimation of conditional moment restriction models with unknown functions, Working paper.

[43] Otsu, T. (2003b) Generalized empirical likelihood inference under weak identification, Working paper.

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

[45] Owen, A. (1990) Empirical likelihood for confidence regions, *Annals of Statistics*, 18, 90-120.

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

[47] Owen, A. (2001) *Empirical Likelihood*, Chapman & Hall.

[48] Pollard, D. (1984) *Convergence of Stochastic Process*, Springer-Verlag, New York.

[49] Powell, J.L. (1984) Least absolute deviation estimator for the censored regression model, *Journal of Econometrics*, 25, 303-325.

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

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

[52] Tripathi, G. and Y. Kitamura (2001) On testing conditional moment restriction: the canonical case, Working Paper.

[53] Tripathi, G. and Y. Kitamura (2002) Testing conditional moment restrictions, forthcoming in *Annals of Statistics*.

[54] van der Vaart, A.W. and J.A. Wellner (1996) Weak convergence and empirical process, Springer, New York.

[55] Whang, Y.-J. (2003) Smoothed empirical likelihood methods for quantile regression models, Working paper.

[56] Weiss, A.A. (1991) Estimating nonlinear dynamic models using least absolute error estimation, *Econometric Theory*, 7, 46-68.

[57] Zhang, J. and I. Gijbels (2003) Sieve empirical likelihood and extensions of the generalized least squares, *Scandinavian Journal of Statistics*, 30, 1-24.

[58] Zheng, J.X. (1998) A consistent nonparametric test of parametric regression models under conditioning quantile restrictions, *Econometric Theory*, 14, 123-138.