A Direct Approach to Inference in Nonparametric and Semiparametric Quantile Regression Models

Yanqin Fan and Ruixuan Liu*
Department of Economics
Vanderbilt University
VU Station B #351819
2301 Vanderbilt Place
Nashville, TN 37235-1819

This version: December 2012

Abstract

This paper makes two main contributions. First, we construct "density-free" confidence intervals and confidence bands for conditional quantiles in nonparametric and semiparametric quantile regression models. They are based on pairs of symmetrized k-NN quantile estimators at two appropriately chosen quantile levels. In contrast to Wald-type confidence intervals or bands based on the asymptotic distributions of estimators of the conditional quantiles, our confidence intervals and bands circumvent the need to estimate the conditional quantile density function, do not require the covariate to have a density function, and are very easy to compute. The advantages of our new confidence intervals are borne out in a simulation study. Second, we present a generic confidence interval for conditional quantiles using the rearranged quantile curves that is asymptotically valid for any quantile regression (parametric, nonparametric, or semiparametric), any method of estimation, and any data structure, provided that the conditional quantile function satisfies some mild smoothness assumptions and the original quantile estimator is such that its associated quantile process converges weakly to a Gaussian process with a covariance kernel proportional to the conditional quantile density function.

Keywords: Confidence Interval; Confidence Band; Partially Linear Quantile Regression; Single-Index Quantile Regression; Rearranged Quantile Curve

JEL Codes: C12; C14; C21

*We thank Matias Cattaneo, Xuming He, Sang Soo Park, and seminar participants at University of Michigan for helpful comments and discussions. We also thank Song Song for sharing his code with us and Jon Wellner for finding the reference, Thompson (1936), for us.

1 Introduction

In their seminal paper, Koenker and Bassett (1978) propose to use linear quantile regression to examine effects of an observable covariate on the distribution of a dependent variable other than the mean. Since then, linear quantile regression has become a dominant approach in empirical work in economics, see e.g., Buchinsky (1994) and Koenker (2005). Following Koenker and Bassett (1978), this approach has been extended to censored data in Powell (1986), Buchinsky and Hahn (1998), Khan and Powell (2001), Chernozhukov and Hong (2002), Honore, Khan, and Powell (2002), and to unit root quantile regression models in Koenker and Xiao (2004), further broadening its scope of applications.

Linearity adopted in Koenker and Bassett (1978) has been relaxed to accommodate possibly nonlinear effects of the covariates on the conditional quantile of the dependent variable in nonparametric and semiparametric quantile regression models. The 'check function' approach of Koenker and Bassett (1978) has been extended to estimating these models as well, see e.g., Truong (1989), Chaudhuri (1991), He, Ng, and Portnoy (1998), and He and Ng (1999) for nonparametric estimation of conditional quantiles; Chaudhuri, Doksum, and Samarov (1997) for nonparametric average derivative quantile estimation; Fan, Hu and Truong (1994), Yu and Jones (1998) and Guerre and Sabbah (2012) for local polynomial estimation of regression quantiles; Lee (2003) and Song, Ritov, and Hardle (2012) for partial linear quantile regression models; Wu, Yu, and Yu (2010) and Kong and Xia (2012) for single index quantile regression models; and Chen and Khan (2001) for partially linear censored regression models.[1]

For nonparametric quantile regression models, an alternative estimation approach to the 'check function' approach is taken in Stute (1986), Bhattacharya and Gangopadhyay (1990), Cai (2002), Fan and Liu (2011), and Li and Racine (2008), among others. In this approach, the conditional distribution function of Y, the dependent variable, given the covariate X is estimated first and the generalized inverse of this estimator at a given quantile level $p \in (0, 1)$ is taken as an estimator of the p-th conditional quantile. Stute (1986) and Bhattacharya and Gangopadhyay (1990) focused on univariate covariate and estimated the conditional distribution function by k-NN method, while Fan and Liu (2011) and Li and Racine (2008) allowed for multivariate covariate and adopted respectively k-NN and kernel estimators of the conditional distribution function.

Under regularity conditions, existing work establish asymptotic normality of the conditional quantile estimators which is the basis for the Wald-type inference, i.e., using the t statistic to test hypotheses or form confidence intervals for the true conditional quantiles. Regardless of the

[1] Conditional quantile function also plays an important role in the recent structural econometrics literature, see e.g., Chesher (2003) for non-separable models, Chernozhukov and Hansen (2005) for quantile treatment effect models using IV, Holderlein and Mammen (2007) for analyzing marginal effects in non-separable models without assuming monotonicity, Echenique and Komunjer (2009) for models involving multiple equilibria, and Chernozhukov, Fernandez-Val, and Kowalski (2011) for models with censoring and endogeneity.

approach used to estimate the conditional quantile in parametric, semiparametric, or nonparametric quantile regression models, one common feature of the asymptotic distributions of the conditional quantile estimators is that their asymptotic variances depend on the conditional quantile density function of Y given $X = x$ and some even depend on the density function of X, see e.g., Horowitz (1998), Khan (2001), Koenker and Xiao (2002), Li and Racine (2008), Hardle and Song (2010), and Song, Ritov, and Hardle (2012), among others. As a result, inference procedures for the conditional quantiles based on the asymptotic distributions of these estimators require consistent estimators of the conditional quantile density function of Y given $X = x$ and/or the density of X both involving bandwidth choice. Numerical evidence presented in De Angelis, Hall, and Young (1993), Buchinsky (1995), Horowitz (1998), and Kocherginsky, He, and Mu (2005) shows that although asymptotically valid, these inference procedures are sensitive in finite samples to the choice of smoothing parameter used to estimate the conditional quantile density function.

Various alternative approaches have been proposed in the current literature to improve on the finite sample performance of Wald-type inferences. Most of these are developed for linear or parametric conditional quantile regression models. First, Goh and Knight (2009) propose a different scale statistic to standardize the estimator of the model parameter in linear quantile regression models resulting in a nonstandard inference procedure; Second, Zhou and Portnoy (1996) construct confidence intervals/bands directly from pairs of estimators of conditional quantiles in the location-scale forms of linear quantile regression models extending the direct or order statistics approach for sample quantiles in Thompson (1936), see also Serfling (1980), Csörgő and Réveséz (1984), and van der Vaart (1998); Third, Gutenbrunner and Jureckova (1992) and Gutenbrunner, Jureckova, Koenker, and Portnoy (1993) employ rank scores to test a class of linear hypotheses; Fourth, Whang (2006) and Otsu (2008) apply the empirical likelihood approach to parametric quantile regression models; Lastly, MCMC related approaches have been proposed to improve standard resampling or simulation paradigms: He and Hu (2002) resample estimators from the marginal estimating equation along the generated Markov chain; Chernozhukov, Hansen, and Janssen (2009) develop finite sample inference procedures based on conditional pivotal statistics in parametric quantile regression models. A nice survey of various inference procedures targeted at linear quantile regression models could be found in Kocherginsky, He, and Mu (2005).

Compared with parametric quantile regression models, inference in nonparametric and semiparametric quantile regression models is still in its infancy. The only alternative approach to the Wald-type and bootstrap inferences that is currently available is the empirical likelihood procedure in Xu (2012) for nonparametric quantile regression models. In semiparametric quantile regression models including partial linear and single index models, only Wald-type and bootstrap inferences are available. Although the empirical likelihood approach in Xu (2012) avoids estimation of the conditional quantile density function and performs better than the Wald-type inference procedures, it is known to be computationally costly. Among existing approaches to inference in parametric quantile regression models, the direct approach is the simplest to implement and least costly computationally—it only requires computing pairs of the quantile estimate. In addition, it does not rely on estimation of the conditional quantile density function and exhibits superior finite sample performance compared with the Wald-type inference, see Zhou and Portnoy (1996). However, as discussed in Portnoy (2012), it appears that the direct approach in Zhou and Portnoy (1996) has theoretical justification only under location-scale forms of linear quantile regression models, severely limiting its applicability—in numerous cases the data do not satisfy the location-scale paradigm, see Koenker (2005).

This paper aims at bridging this gap. Specifically, it makes two main contributions to inference in quantile regression models. First, we develop direct inference procedures including confidence intervals/bands for conditional quantiles in nonparametric and semiparametric quantile regression models allowing for the covariate to affect the conditional quantile of the dependent variable in general ways. Compared with the currently available Wald-type inference procedures in these models, our new confidence intervals/bands avoid the estimation of the conditional quantile density function of Y given $X = x$ and/or the density function of the covariate X. In fact, our confidence intervals/bands do not even require the covariate to have a density function. The underlying idea is easily explained in a nonparametric quantile regression model. To avoid assuming the existence of a density function for the covariate, we adopt the symmetrized k-NN estimator of the conditional distribution function and take the conditional quantile estimator as the generalized inverse of the symmetrized k-NN estimator. Instead of relying on the asymptotic normality of the k-NN estimator, we construct confidence intervals/bands directly from our k-NN quantile estimator evaluated at two appropriately chosen quantile levels. Like the empirical likelihood confidence interval in Xu (2012), our confidence interval for nonparametric quantiles internalizes the conditional quantile density estimation of Y given X and the covariate density estimation and is not necessarily symmetric. Compared with Xu (2012), our confidence interval is easier to compute and does not require optimization. In addition, we also construct confidence intervals/bands for conditional quantiles in partial linear and single index quantile regressions. Compared with Hardle and Song (2010), and Song, Ritov, and Hardle (2012), our confidence intervals/bands share the same features as those in nonparametric quantile regression models—easy to compute and density-free. A small scale simulation study demonstrates the advantages and feasibility of our confidence intervals/bands over existing ones in practically relevant model set-ups.

Second, under a high level assumption on the conditional quantile function and the original quantile estimator, we present a generic confidence interval for conditional quantiles using the rearranged quantile curves that is asymptotically valid for any quantile regression (parametric, nonparametric, or semiparametric), any method of estimation, and any data structure. We verify our high level assumption for three examples: nonparametric quantile regression with multivariate covariate estimated by the standard asymmetric k-NN quantile estimator; Censored nonparametric quantile regression; and nonparametric quantile regression for time series. Interestingly pairs of the standard k-NN asymmetric quantile estimate in our first example correspond to pairs of order statistics of the induced order statistics of Y, so our generic confidence interval in this case shares the elegance and simplicity of the confidence interval for unconditional quantiles based on order statistics originally proposed in Thompson (1936), see also van der Vaart (1998).

The rest of this paper is organized as follows. Section 2 considers the nonparametric quantile regression with a univariate covariate. It introduces our conditional quantile estimator, constructs a new confidence interval, and a new confidence band. Section 3 extends the confidence intervals/bands developed in Section 2 to two popular semiparametric models, partial linear and single index quantile regression models. Section 4 presents a generic confidence interval, shows its asymptotic validity under a high level assumption on the original quantile estimator, and verifies the high level assumption in several examples. Section 5 provides a simulation study comparing the finite sample performance of our new confidence intervals with Wald-type confidence intervals and two bootstrap versions for nonparametric and partial linear quantile regressions. We conclude in the last section. All the proofs are collected in the Appendices.

2 Nonparametric Quantile Regression

Let $\{X_i, Y_i\}_{i=1}^n$ denote an i.i.d. copy of the bivariate random vector $\{X, Y\}$, with marginal distribution functions $F_X(x)$, $F_Y(y)$ respectively, where[2] $x \in \mathcal{X} \subset \mathcal{R}$ and $y \in \mathcal{Y} \subset \mathcal{R}$. Further, let $F_{Y|X}(\cdot|x)$ denote the conditional distribution function of Y given $X = x$ with density function $f_{Y|X}(\cdot|x)$ and $\xi_p(x) = \xi(p|x) = F_{Y|X}^{-1}(p|x)$ denote the p-th conditional quantile[3] of Y given $X = x$, where $0 < p < 1$ and $x \in \mathcal{X}$.

The nonparametric quantile regression can be written as

$$Y_i = \xi_p(X_i) + \varepsilon_i, \text{ with } \Pr[\varepsilon_i \leq 0 | X_i = x] = p \text{ and } i = 1, ..., n.$$

Let $x_0 \in \mathcal{X}$ denote a fixed covariate value. We now introduce our estimator of $\xi_p(x_0)$. Let $F_n(\cdot)$ denote the empirical distribution function of $\{X_i\}_{i=1}^n$ and $\widehat{F}_n(y|x_0)$ denote the symmetrized k-NN estimator of $F_{Y|X}(y|x)$ introduced in Yang (1980) and further studied in Stute (1984b, 1986):

$$\widehat{F}_n(y|x_0) = \frac{\sum_{i=1}^n 1\{Y_i \leq y\} K\left(\frac{F_n(x_0) - F_n(X_i)}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{F_n(x_0) - F_n(X_i)}{h_n}\right)}, \tag{1}$$

[2] To focus on the main idea, we consider nonparametric quantile regression with a univariate covariate in this section. Section 4 allows for nonparametric quantile regression with multivariate covariate.

[3] In the sequel, we will use $\xi_p(x)$, $\xi(p|x)$, and $F_{Y|X}^{-1}(p|x)$ interchangably.

where $K(\cdot)$ is a kernel function and $h_n \to 0$ is a bandwidth. The estimator of $\xi_p(x_0)$ based on $\widehat{F}_n(y|x_0)$ is defined as

$$\widehat{\xi}(p|x_0) = \widehat{\xi}_p(x_0) = \widehat{F}_n^{-1}(p|x_0). \tag{2}$$

In the rest of this section, we introduce a new confidence interval for the conditional quantile at a fixed covariate value and then construct a confidence band that is uniformly valid over a range of covariate values.

2.1 A New Confidence Interval

Our new level $(1-\alpha)$-confidence interval for $\xi_p(x_0)$ takes the following form:

$$\mathrm{CIN}_{1-\alpha} = \left(\widehat{F}_n^{-1}\left(p - z_{\alpha/2}\sigma_{np}(K)\,|x_0\right), \widehat{F}_n^{-1}\left(p + z_{\alpha/2}\sigma_{np}(K)\,|x_0\right)\right], \tag{3}$$

where $z_{\alpha/2}$ denotes the $\alpha/2$-th quantile of the standard normal distribution and

$$\sigma_{np}(K) = \sqrt{\frac{R(K)\,p(1-p)}{nh_n}} \tag{4}$$

in which $R(K) = \int K^2(u)\,du$.

Let $q_p(x) = 1/f_{Y|X}\left(\xi_p(x)\,|x\right)$ denote the conditional quantile density function of Y given X. It is obvious from (3) that our new confidence interval (CI), $\mathrm{CIN}_{1-\alpha}$, has several advantages over existing CIs. First, compared with Wald-type confidence intervals, our new confidence interval, $\mathrm{CIN}_{1-\alpha}$, does not require either a consistent estimator of the density function of X or the conditional quantile density function of Y given $X = x_0$, $q_p(x_0)$. Second, compared with the CI based on the empirical likelihood approach in Xu (2012), our CI is much easier to implement; there is no optimization involved and it only requires evaluating our conditional quantile estimator $\widehat{\xi}_p(x_0)$ at two specific quantile levels, $p - z_{\alpha/2}\sigma_{np}(K)$ and $p + z_{\alpha/2}\sigma_{np}(K)$.

Below we provide a list of sufficient conditions for the asymptotic validity of $\mathrm{CIN}_{1-\alpha}$.

Assumption (S). Let $H(y|u) = F_{Y|X}\left(y|F_X^{-1}(u)\right)$.

(i) Assume that

$$\sup_{|t-s|\le\tau}\left|H\left(F_Y^{-1}(t)\,|u\right) - H\left(F_Y^{-1}(s)\,|u\right)\right| = o\left(\left(\ln\tau^{-1}\right)^{-1}\right) \text{ as } \tau \to 0$$

uniformly in a neighborhood of $u_0 = F_X(x_0)$;

(ii) Uniformly in y, $H(y|\cdot)$ belongs to the second order Holder class at $u_0 \in (0,1)$, i.e., for any y, $H(y|u)$ is differentiable w.r.t u at u_0 and there exists a neighborhood of u_0 such that for any u_1, u_2 in this neighborhood, we have that

$$\left|H'(y|u_1) - H'(y|u_2)\right| \le L|u_1 - u_2|$$

holds uniformly in y, where $H'(y|u) = \partial H(y|u)/\partial u$ and $L < \infty$.

Assumption (H). The bandwidth satisfies $h_n = n^{-\delta}$ for some $\delta \in (1/5, 1/3)$, i.e., it satisfies: $nh_n^5 \to 0$ and $nh_n^3 \to \infty$ as $n \to \infty$.

Assumption (K). The kernel function $K(\cdot)$ is a twice continuously differentiable density function with zero mean, compact support and bounded second order derivative.

Assumption (X). X has continuous distribution function $F_X(x)$.

Assumption (S) is chosen in accordance with Assumptions (A), (B) in Stute (1986). For (S) (i), we added the corresponding quantile transformation since Stute (1986) directly works with (X, Y) with uniform marginal distributions. (S) (ii) is written slightly differently from Assumption (B) in Stute (1986) as it does not require second order differentiability of $H(y|u)$, but achieves the same purpose in controlling the bias term. The so-called uniform Holder class is adapted from Tsybakov (2008), Korostelev and Korosteleva (2011), see also Guerre and Sabbah (2012). Assumption (X) spells out this asymptotic distribution freeness advocated by Stute (1984b), as by the elementary fact $F_X(X_i) \sim U[0, 1]$. The requirement on the bandwidth is standard, with one added condition $nh_n^3 \to \infty$, which is necessary in dealing with the asymptotic variance term as demonstrated in Stute (1984b). Assumption (K) ensures that our quantile estimator $\widehat{F}_n^{-1}(p|x_0)$ is monotone in $p \in (0, 1)$ so $\mathrm{CIN}_{1-\alpha}$ is non-empty. Kernel functions satisfying Assumption (K) include Bisquare and Triweight kernels.

THEOREM 2.1 *Suppose Assumptions (X), (S), (K), and (H) hold and x_0 is an interior point not on the flat part of F_X. In addition, assume $F_{Y|X}(y|x_0)$ is continuously differentiable in a neighborhood of $\xi_p(x_0)$ corresponding to $[p_1, p_2]$ containing p with strictly positive derivative and $0 < p_1 < p_2 < 1$. For $0 < \alpha < 1$, we get:* $\Pr\left(\xi_p(x_0) \in CIN_{1-\alpha}\right) \to 1 - \alpha$ *as $n \to \infty$.*

The Lemma below demonstrates the critical role played by the symmetrized k-NN estimator $\widehat{F}_n(y|x_0)$ in our new confidence interval which not only avoids the estimation of the conditional quantile density function of Y given $X = x_0$ but also the estimation of the density function of X.

Lemma 2.2 *Suppose the conditions of Theorem 2.1 hold. Then*

(i) $\sqrt{nh_n}\left[\widehat{F}_n(\cdot|x_0) - F_{Y|X}(\cdot|x_0)\right] \Longrightarrow B_0(\cdot)$, *where $B_0(\cdot)$ is the Brownian Bridge with the following covariance structure:*

$$Cov(B_0(y_1), B_0(y_2)) = R(K)\left[F_{Y|X}(y_1 \wedge y_2|x_0) - F_{Y|X}(y_1|x_0)F_{Y|X}(y_2|x_0)\right];$$

(ii) Moreover, the conditional density function of Y given X is strictly positive on the interval: $\left[F_{Y|X}^{-1}(p_1|x_0) - \epsilon, F_{Y|X}^{-1}(p_2|x_0) + \epsilon\right]$ *for some $\epsilon > 0$. Then*

$$\left\{\sqrt{nh_n}\left[\widehat{F}_n^{-1}(p|x_0) - F_{Y|X}^{-1}(p|x_0)\right] : p \in [p_1, p_2]\right\} \Longrightarrow q_p(x_0)\, B_0(F_{Y|X}^{-1}(p|x_0)).$$

Lemma 2.2 (i) is restated from Stute (1986). It makes clear that in contrast to the commonly used Nadaraya-Watson estimator or the local polynomial estimator of the conditional distribution function, the asymptotic variance of $\widehat{F}_n(y|x_0)$ does not depend on the density of the covariate X. In fact, Lemma 2.2 does not even require that X has a density. It is this "density-free" feature of $\widehat{F}_n(y|x_0)$ that enables us to dispense with the density of X in our new confidence interval.

Lemma 2.2 (ii) follows from Lemma 2.2 (i), Lemma 21.3 in van der Vaart (1998), and the functional Delta method. It implies that for a fixed $p \in [p_1, p_2]$,

$$\sqrt{nh_n}\left[\widehat{F}_n^{-1}(p|x_0) - F_{Y|X}^{-1}(p|x_0)\right] \Longrightarrow N\left(0, \sigma^2\right)$$

with $\sigma^2 = R(K)\, p(1-p)\, q_p^2(x_0)$. So even though the use of $\widehat{F}_n(y|x_0)$ frees us from estimating the density of X, the asymptotic variance of $\widehat{F}_n^{-1}(p|x_0)$ still depends on the conditional quantile density $q_p(x_0)$. As a result, Wald-type inference procedures based on the asymptotic normality of $\widehat{F}_n^{-1}(p|x_0)$ would still require a consistent estimator of $q_p(x_0)$ or $f_{Y|X}\left(\xi_p(x_0)|x_0\right)$ which our new confidence interval avoids as well.

2.2 A New Confidence Band

In many applications, uniformly valid confidence bands over a range of covariate values may be desirable, see Hardle and Song (2010), Song, Ritov, and Hardle (2012) for interesting empirical applications in labor economics. Below we extend our confidence interval $\text{CIN}_{1-\alpha}$ to confidence bands over a range of covariate values.

Let

$$\text{CBN}_{1-\alpha} = \left[\widehat{F}_n^{-1}\left(p - c_{n\delta}(\alpha, K)\,\sigma_{np}(K)\,|x\right), \widehat{F}_n^{-1}\left(p + c_{n\delta}(\alpha, K)\,\sigma_{np}(K)\,|x\right)\right], \tag{5}$$

where $\sigma_{np}(K)$ is defined in (4) and

$$c_{n\delta}(\alpha, K) = \frac{c(\alpha)}{(2\delta \log n)^{1/2}} + d_n \tag{6}$$

in which $c(\alpha) = \log 2 - \log|\log(1-\alpha)|$ and

$$d_n = (2\delta \log n)^{1/2} + (2\delta \log n)^{-1/2} \log\left[\frac{\int \left(K'(u)\right)^2 du}{4\pi R(K)}\right]. \tag{7}$$

Note that like our confidence interval $\text{CIN}_{1-\alpha}$, our confidence band, $\text{CBN}_{1-\alpha}$, is easy to compute and shares the remarkable density-free feature.

Below we provide additional conditions under which we show the uniform asymptotic validity of our confidence band. Let $\mathcal{J} \subset \mathcal{X}$ denote an inner compact subset of $\mathcal{X}$.

Assumption ($\tilde{\mathbf{S}}$). Assumption (S) holds uniformly for $x \in \mathcal{J}$.

Assumption ($\tilde{\mathbf{X}}$). Assumption (X) plus the compactness of $\mathcal{J}$ gives uniform continuity of $F_X(\cdot)$. We list this rather redundant assumption for easy reference.

Assumption (B). (i) $h_n^{-3}\log n\int_{|y|>a_n} f_Y(y)dy = O(1)$, where $f_Y(y)$ is the marginal density of Y and $(a_n)_{n=1}^{\infty}$ is a sequence of constants tending to infinity as $n\to\infty$; (ii) $\inf_{x\in\mathcal{J}} f_{Y|X}(\xi_p(x)|x) > 0$; (iii) $\sup_y \sup_{x\in\mathcal{J}} f_{Y|X}(y|x) < \infty$; (iv) Y has Lipschitz continuous distribution function $F_Y(\cdot)$ and (X,Y) has uniformly bounded copula density function $c(x,y)$.

Assumption (B) (i), (ii) are added in accordance with the strong approximation result in Hardle and Song (2010). Since we base our analysis on the covariate X after (empirical) probability integral transform, some of the assumptions in Hardle and Song (2010) will be satisfied automatically here such as their (A5) and (A6). Also notice that our Assumption (H) on the bandwidth implies Assumption (A2) in Hardle and Song (2010) and our Assumption (K) implies their assumption (A1). Assumption (B) (iii), (iv) will be needed to establish the uniform Bahadur representation. Specifically Assumption (B)(iii) aims to control the bias term in the local oscillation uniformly, and with the help of (B)(iv) we could utilize certain nice maximal inequality in Stute (1984a) to bound the local oscillation of copula process within a shrinking rectangle. Details could be found in our Lemmas A6 and A7.

THEOREM 2.3 *Suppose Assumptions (B), ($\tilde{S}$), ($\tilde{X}$), (K), and (H) hold. Then the confidence band $CBN_{1-\alpha}$ is asymptotically valid with coverage probability $1-\alpha$ uniformly over $x\in\mathcal{J}$.*

Compared with our confidence interval, our confidence band replaces $z_{\alpha/2}$ with $c_{n\delta}(\alpha, K)$. The Lemma below explains why.

Lemma 2.4 *Under Assumptions (B), ($\tilde{S}$), ($\tilde{X}$), (K), and (H), it holds that*

$$\begin{aligned}&\Pr\left((2\delta\log n)^{1/2}\left[\sigma_{np}^{-1}(K)\sup_{x\in\mathcal{J}}\left\{f_{Y|X}\left(\xi_p(x)|x\right)|\widehat{F}_n^{-1}(p|x) - F^{-1}(p|x)|\right\} - d_n\right]\le z\right)\\ \to\;& \exp(-2\exp(-z)) \;\; \text{as } n\to\infty.\end{aligned}$$

3 Semiparametric Quantile Regression Models

In most applications, the covariate X is multivariate. Semiparametric quantile regression models are introduced in the literature to alleviate the curse of dimensionality associated with fully non-parametric models and at the same time are more robust than fully parametric regression models. Commonly used semiparametric quantile regression models include partial linear and single index quantile regression models. Although most work in the literature concern root-n estimation of the finite dimensional parameters, Song, Ritov and Hardle (2012) have constructed uniform confidence bands for partial linear quantile regressions. Their confidence bands, however, require the estimation of both the conditional quantile density and the density function of the covariate.

In the next two subsections, we extend our confidence interval/band for univariate nonparametric quantile regression in Section 2 to both partial linear and single index quantile regressions.

3.1 Partial Linear Quantile Regression Model

Consider the following partial linear quantile regression model with a univariate covariate X and multivariate covariate Z having support $\mathcal{Z} \subset \mathcal{R}^d$:

$$Y_i = Z_i' \beta_0 + g(X_i) + \varepsilon_i, \; i = 1, ..., n,$$

where $\{Y_i, X_i, Z_i\}_{i=1}^n$ is a random sample and $\Pr[\varepsilon_i \leq 0|X_i = x, Z_i = z] = p$ for all $x \in \mathcal{X}$ and $z \in \mathcal{Z}$. Notice that $\Pr[\varepsilon_i \leq 0|X_i = x] = p$ holds as well.

Root-n consistent estimators of β_0 are available in Lee (2003) and Song, Ritov, and Hardle (2012). Semiparametric efficient estimation of the above model has been studied by Lee (2003). Let $\hat{\beta}$ denote a root-n consistent estimator of β_0. For $x_0 \in \mathcal{X}$ and $z_0 \in \mathcal{Z}$, let

$$\widehat{F}_{n,\mathrm{PL}}(y|x_0) = \frac{\sum_{i=1}^n 1\{Y_i - Z_i'\hat{\beta} \leq y\} K\left(\frac{F_n(x_0) - F_n(X_i)}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{F_n(x_0) - F_n(X_i)}{h_n}\right)}. \tag{8}$$

Our CI for the conditional quantile $[z_0'\beta_0 + g(x_0)]$ is defined as:

$$\mathrm{CIPL}_{1-\alpha} = \left(z_0'\hat{\beta} + \widehat{F}_{n,\mathrm{PL}}^{-1}\left(p - z_{\alpha/2}\sigma_{np}(K)|x_0\right), z_0'\hat{\beta} + \widehat{F}_{n,\mathrm{PL}}^{-1}\left(p + z_{\alpha/2}\sigma_{np}(K)|x_0\right)\right], \tag{9}$$

where $\sigma_{np}(K)$ is defined in (4).

We now introduce two assumptions.

Assumption (Z1). Let $\widetilde{Y} = Y - Z'\beta_0$. Assumptions (S) and (X) hold for $\left(\widetilde{Y}, X\right)$.

Assumption (PL). Z_i has a finite conditional (on X_i) second moment and

$$E\left[\left|1\{Y_i - Z_i'\beta_1 \leq y\} - 1\{Y_i - Z_i'\beta_2 \leq y\}\right| |X_i\right] \leq M|\beta_1 - \beta_2|$$

holds uniformly in y, where M is a positive constant.

THEOREM 3.1 *Suppose $\hat{\beta} - \beta_0 = O_p\left(n^{-1/2}\right)$ and Assumptions (Z1), (PL), (K) and (H) hold. Then the confidence interval, $CIPL_{1-\alpha}$, achieves the nominal level $(1-\alpha)$ asymptotically.*

Similarly, the new confidence band is defined as

$$\mathrm{CBPL}_{1-\alpha} = \left[z'\hat{\beta} + \widehat{F}_{n,\mathrm{PL}}^{-1}\left(p - c_{n\delta}(\alpha, K)\sigma_{np}(K)|x\right), z'\hat{\beta} + \widehat{F}_{n,\mathrm{PL}}^{-1}\left(p + c_{n\delta}(\alpha, K)\sigma_{np}(K)|x\right)\right], \tag{10}$$

where $\sigma_{np}(K)$ is defined in (4) and $c_{n\delta}(\alpha, K)$ is defined in (6).

Once we strengthen our assumptions to handle various uniformity issues, we get the asymptotic validity of the new confidence band for the partial linear quantile model.

Assumption ($\tilde{\mathbf{Z}}$1). Let $\widetilde{Y} = Y - Z'\beta_0$. For Z taking values restricted to a given compact set $\mathcal{K} \subset \mathcal{Z}$, Assumptions ($\tilde{\mathrm{S}}$), ($\tilde{\mathrm{X}}$), and (B) hold for $\left(\widetilde{Y}, X\right)$.

THEOREM 3.2 *Suppose* $\hat{\beta} - \beta_0 = O_p\left(n^{-1/2}\right)$ *and Assumptions ($\tilde{Z}$1), (PL), (K) and (H) hold. Then the confidence band, $CBPL_{1-\alpha}$, is asymptotically valid with coverage probability* $(1-\alpha)$ *uniformly over* $x \in \mathcal{J}$ *and* $z \in \mathcal{K}$.

3.2 Single Index Quantile Regression Model

Consider the single index model with multivariate covariate Z below:

$$Y_i = g\left(Z_i'\beta_0\right) + \varepsilon_i,\ i = 1, ..., n,$$

where $\{Y_i, Z_i\}_{i=1}^n$ is a random sample and $\Pr\left[\varepsilon_i \leq 0 | Z_i = z\right] = p$ for all $z \in \mathcal{Z}$.

Let $\widehat{\beta}$ denote a consistent estimator of β_0 such as that in Wu, Yu, and Yu (2010) or Kong and Xia (2012), based on structural adaptive estimation methods. Those authors find that their estimators improve upon the two step M-estimators as in Chen and Pouzo (2009), Ichimura and Lee (2010) in terms of both computation time and accuracy. For $z_0 \in \mathcal{Z}$, let

$$\widehat{F}_{n,\mathrm{SI}}\left(y|z_0\right) = \frac{\sum_{i=1}^n 1\{Y_i \leq y\} K\left(\frac{\widehat{F}_n(z_0'\hat{\beta}) - \widehat{F}_n(Z_i'\hat{\beta})}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{\widehat{F}_n(z_0'\hat{\beta}) - \widehat{F}_n(Z_i'\hat{\beta})}{h_n}\right)}, \tag{11}$$

where $\widehat{F}_n$ is the empirical distribution function of $\left\{Z_i'\hat{\beta}\right\}_{i=1}^n$. Our CI for $g\left(z_0'\beta_0\right)$ is defined as:

$$\mathrm{CISI}_{1-\alpha} = \left(\widehat{F}_{n,\mathrm{SI}}^{-1}\left(p - z_{\alpha/2}\sigma_{np}\left(K\right)|z_0\right), \widehat{F}_{n,\mathrm{SI}}^{-1}\left(p + z_{\alpha/2}\sigma_{np}\left(K\right)|z_0\right)\right], \tag{12}$$

where $\sigma_{np}\left(K\right)$ is defined in (4).

We make the following assumptions.

Assumption (Z2). Let $\tilde{X} = Z'\beta_0$. Assumptions (S) and (X) hold for $\left(Y,\ \tilde{X}\right)$. Moreover, $E||Z||^\gamma < \infty$ for some $\gamma \geq 4$.

Assumption (HS). In addition to Assumption (H), the bandwidth also satisfies:

$$h_n^{-1/2} n^{-1/4+1/2\gamma}\sqrt{\ln n} = o(1) \text{ and } h_n^{-5/2} n^{-1+1/\gamma} \ln n = o(1).$$

THEOREM 3.3 *Suppose* $\hat{\beta} - \beta_0 = O_p\left(n^{-1/2}\right)$ *and Assumptions (Z2), (K), and (HS) hold. Then the confidence interval, $CISI_{1-\alpha}$, achieves the nominal level* $(1-\alpha)$ *asymptotically.*

Remark 3.1. The root-n asymptotic normality of the estimator $\hat{\beta}$ in Wu, Yu, and Yu (2010), Kong and Xia (2012) actually requires much stronger assumptions than what we assume here. For the restriction on bandwidth, Assumption (H) is maintained, letting $h_n = n^{-\delta}$. A suitable δ could be chosen from $(1/5, 1/3)$ ensuring $\gamma \geq 4$, which is a rather mild restriction.

Again once we strengthen our assumptions in accordance with various uniformity issues, we get the asymptotic validity of the new confidence band defined in (13) below for the single index quantile model.

Assumption ($\tilde{\mathbf{Z}}$2). Let $\tilde{X} = Z'\beta_0$. In addition to the moment restriction in (Z2), Assumptions ($\tilde{\mathrm{S}}$), ($\tilde{\mathrm{X}}$), and (B) hold for $\left(Y,\ \tilde{X}\right)$ uniformly in a compact set $\mathcal{K} \subset \mathcal{Z}$.

THEOREM 3.4 *Suppose $\hat{\beta} - \beta_0 = O_p\left(n^{-1/2}\right)$ and Assumptions ($\tilde{Z}$2), (K), and (HS) hold. Then the confidence band below is asymptotically valid with coverage probability $(1-\alpha)$ uniformly over $z \in \mathcal{K}$:*

$$CBSI_{1-\alpha} = \left[\widehat{F}_{n,SI}^{-1}\left(p - c_{n\delta}\left(\alpha, K\right)\sigma_{np}\left(K\right)|z\right), \widehat{F}_{n,SI}^{-1}\left(p + c_{n\delta}\left(\alpha, K\right)\sigma_{np}\left(K\right)|z\right)\right], \tag{13}$$

where $\sigma_{np}\left(K\right)$ is defined in (4) and $c_{n\delta}\left(\alpha, K\right)$ is defined in (6).

4 A Generic Confidence Interval Based on Rearranged Quantile Curves

In this section we present a generic confidence interval based on a consistent estimator of the conditional quantile function using the direct approach. In Subsection 4.1, we show that under a high level assumption, our generic confidence interval is asymptotically valid regardless of the specification of the quantile function (parametric, nonparametric, or semiparametric), the method of estimation, and data structure. In Subsection 4.2, we verify the high level assumption in three examples and obtain a novel confidence interval based on order statistics for conditional quantiles, a new confidence interval for censored nonparametric quantile regression, and a new confidence interval for nonparametric time series quantile regression.

4.1 A Generic Confidence Interval

Let $X \in \mathcal{X} \subseteq \mathcal{R}^d$ denote the covariate of dimension $d \geq 1$. Throughout, we will use $\xi_p\left(x_0\right)$, $p \in (0, 1)$, to denote the conditional quantile of interest, where $x_0 \in \mathcal{X}$ is fixed and $\xi_p\left(x\right)$ could be a parametric, nonparametric, or semiparametric function of x. Let $\widehat{\xi}\left(p|x_0\right) \equiv \widehat{\xi}_p\left(x_0\right)$ denote a consistent estimator of $\xi_p\left(x_0\right)$, where $\widehat{\xi}\left(p|x_0\right)$ may or may not be monotone in $p \in (0, 1)$. If it is not monotone, we adopt the rearranged version of $\widehat{\xi}\left(p|x_0\right)$ in Chernozhukov, Fernandez-Val, and Galichon (2010).[4] The monotonicity issue has to be tackled here since we need to make sure the

[4] Besides Chernozhukov, Fernandez-Val, and Galichon (2010), alternative monotone rearrangement plans have been carried out by He (1997), Yu and Jones (1998), Dette and Volgushev (2008). In the last two references, additional

nonemptyness of the new confidence interval based on two ordered pairs of the estimated quantile, see (14) below.

Let $\widehat{\xi}^*(p|x_0)$ denote the rearranged version of $\widehat{\xi}(p|x_0)$ proposed by Chernozhukov, Fernandez-Val, and Galichon (2010):

$$\widehat{\xi}^*(p|x_0) = \inf\left[y : \int_0^1 1\{\widehat{\xi}(\tau|x_0) \le y\}d\tau \ge p\right].$$

We will use $\widehat{\xi}^*(p|x_0)$ to construct our confidence interval for $\xi_p(x_0)$ based on the direct approach and show its asymptotic validity under the following high level assumption on the quantile function $\xi_p(x_0)$ and its estimator $\widehat{\xi}(p|x_0)$.

Assumption (G)

(i) Let $\xi(p|x) = \xi_p(x)$. Then $\xi(p|x_0)$ is a continuously differentiable function in $p \in (0,1)$ and for fixed $p \in (0,1)$, $\xi(p|x)$ is continuously differentiable at $x = x_0$;

(ii) Let $q_p(x_0) = \frac{\partial}{\partial p}\xi(p|x_0)$. Then $q_p(x_0) > 0$ for $p \in (0,1)$;

(iii) The quantile estimator $\widehat{\xi}(\cdot|x_0)$ takes its values in the space of bounded measurable functions defined on $(0,1)$ and in $l^\infty((0,1))$,

$$b_n\left(\widehat{\xi}(p|x_0) - \xi(p|x_0)\right) \Longrightarrow q_p(x_0)\, B(p|x_0),\ p \in (0,1),$$

as a stochastic process indexed by $p \in (0,1)$, where $B(p|x_0), p \in (0,1)$ is a Gaussian process whose variance is $Var[B(p|x_0)] = p(1-p)\varpi_{x_0}^2$, for some positive constant ϖ_{x_0} depending on x_0 and b_n is a sequence of positive constants such that $b_n \to \infty$ as $n \to \infty$.

Assumption (G) (i) and (ii) are taken directly from Chernozhukov, Fernandez-Val, and Galichon (2010). Assumption (G) (iii) is a special case of Assumption 2 in Chernozhukov, Fernandez-Val, and Galichon (2010). It imposes a specific structure on the asymptotic variance of the quantile estimator $\widehat{\xi}(p|x_0)$ which ensures the asymptotic validity of the following confidence interval obtained from the direct approach:

$$\text{CI-G}_{1-\alpha} = \left[\widehat{\xi}^*\left(p - \frac{z_{\alpha/2}\hat{\varpi}_{x_0}\sqrt{p(1-p)}}{b_n}|x_0\right), \widehat{\xi}^*\left(p + \frac{z_{\alpha/2}\hat{\varpi}_{x_0}\sqrt{p(1-p)}}{b_n}|x_0\right)\right], \tag{14}$$

where $\hat{\varpi}_{x_0}$ is a consistent estimator of ϖ_{x_0}.

THEOREM 4.1 *Suppose Assumption (G) holds. Then CI-G$_{1-\alpha}$ is asymptotically valid with coverage probability equal to* $(1-\alpha)$.

Proof. Resorting to Corollary 3 in Chernozhukov, Fernandez-Val, and Galichon (2010) which asserts that the rearranged estimator $\widehat{\xi}^*(p|x_0)$ has the same first order asymptotic properties as

smoothing parameters are introduced and under mild regularity conditions, the monotone estimators have the same first order asymptotics as the original estimators (see Theorem 2 in Yu and Jones, 1998; Remark 6 in Dette and Volgushev, 2008).

$\widehat{\xi}(p|x_0)$. In particular, Assumption (G)(iii) implies that

$$b_n\left(\widehat{\xi}^*(p|x_0) - \xi(p|x_0)\right) \Longrightarrow q_p(x_0)\, B(p|x_0).$$

The rest of the proof follows verbatim the proof of our Theorem 2.1, making use of stochastic equicontinuity of the process $b_n\left(\widehat{\xi}^*(p|x_0) - \xi(p|x_0)\right), p \in (0,1)$ and the simple fact that under Assumption (G)(i),

$$b_n\left[\xi\left(p \pm \frac{z_{\alpha/2}\hat{\varpi}_{x_0}\sqrt{p(1-p)}}{b_n}|x_0\right) - \xi(p|x_0)\right] = q_p(x_0)\, z_{\alpha/2}\varpi_{x_0}\sqrt{p(1-p)} + o_p(1).$$

Q.E.D

The above proof makes it clear that Assumption (G) (iii) is crucial to the asymptotic validity of our generic confidence interval CI-G$_{1-\alpha}$ defined in (14). It is worth noting, however, that Assumption (G)(iii) is satisfied by many quantile regression estimators regardless of the model and data structure. For example, the special class of location scale forms of linear quantile regression models in Zhou and Portnoy (1996) ensures that under standard regularity conditions, the quantile estimator of Koenker and Bassett (1978) satisfies Assumption G (iii), see e.g., Gutenbrunner and Jureckova (1992), Koenker and Xiao (2005), and Portnoy (2012), and thus the asymptotic validity of our generic confidence interval CI-G$_{1-\alpha}$ defined in (14). In fact, for location scale forms of the linear quantile regression models, our generic confidence interval CI-G$_{1-\alpha}$ is essentially the confidence interval in Zhou and Portnoy (1996) except that Zhou and Portnoy (1996) uses the original estimator of Koenker and Bassett (1978) instead of its monotone rearranged version. Assumption (G) (iii) is also satisfied by the nonparametric and semiparametric quantile regression models and the symmetric k-NN estimators studied in Sections 2 and 3 in this paper. Since we directly take the generalized inverse of the estimated conditional distribution function as the estimator for the conditional quantile, our quantile estimator is automatically monotone and is identical to its rearranged version.

4.2 Applications of the Generic Confidence Interval

To demonstrate the broad applicability of the confidence interval, CI-G$_{1-\alpha}$, defined in (14), we present three examples in this subsection. These include a novel confidence interval for nonparametric conditional quantiles based on order statistics/induced order statistics in Example 4.1; a new confidence interval for nonparametric censored quantile regression in Example 4.2; and a new confidence interval for nonparametric time series quantile regression in Example 4.3.

Example 4.1. (A Novel Order Statistic Approach): The direct approach when applied to sample quantiles leads to confidence intervals for unconditional quantiles based on pairs of order statistics of $\{Y_i\}_{i=1}^n$, see Thompson (1936) and van der Vaart (1998). The generic confidence interval in (14) when applied to the standard asymmetric k-NN estimator of the conditional quantile leads to a novel confidence interval for conditional quantiles based on pairs of order statistics of an appropriately chosen set of induced order statistics of $\{Y_i\}_{i=1}^n$. To introduce it, let $R_i = ||X_i - x_0||$, for $i = 1, \cdots, n$, where $||\cdot||$ is the standard Euclidean norm in $\mathcal{R}^d$, and $(Y_{n,i})_{i=1}^n$ denote the collection of induced order statistics by rank $(R_i)_{i=1}^n$, i.e., $Y_j = Y_{n,i}$ iff $R_j = R_{(i)}$ and $R_{(i)}$ is the i-th order statistic of $(R_i)_{i=1}^n$. For $k \leq n$, the standard asymmetric k-NN estimator of the distribution function of Y given $X = x_0$ is defined as

$$\widehat{F}_{n,k}\left(y|x_0\right) = k^{-1}\sum_{i=1}^{k} I\left(Y_{n,i} \leq y\right)$$

and the asymmetric k-NN estimator of $\xi_p\left(x_0\right)$ is given by

$$\begin{aligned} \widehat{\xi}\left(p|x_0\right) &\equiv \widehat{\xi}_p\left(x_0\right) = \inf\left\{y : \widehat{F}_{n,k}\left(y|x_0\right) \geq \frac{[kp]}{k}\right\} \\ &= \text{the } [kp]\text{-th order statistic of } Y_{n,1}, Y_{n,2}, \cdots, Y_{n,k}, \end{aligned} \tag{15}$$

where $k \equiv k_n$ is a sequence of constants such that $k_n \to \infty$ and $k_n = o\left(n^{\frac{4}{4+d}}\right)$. Assuming (G) (i) and (ii), the asymptotic validity of CI-G$_{1-\alpha}$ based on the asymmetric k-NN estimator relies on Assumption (G) (iii). For a random sample $\{X_i, Y_i\}_{i=1}^n$, Dabrowska (1987) provides primitive conditions under which the standard k-NN estimator of the conditional distribution function converges weakly to a Gaussian process which can be used to show that Assumption (G) (iii) holds for $\widehat{\xi}\left(p|x_0\right)$ in (15) with $b_n = \sqrt{k_n}$ and $\varpi_{x_0}^2 = \pi^{d/2}/\Gamma\left(d/2+1\right)$. We refer the reader to Section 3.3 and the proof of Proposition 3.4 in Dabrowska (1987) for further details including the primitive conditions. Using $\widehat{\xi}\left(p|x_0\right)$ in (15), the confidence interval (14) reduces to:

$$\begin{aligned} \text{CI-O}_{1-\alpha} &= \left[\widehat{\xi}\left(p - z_{\alpha/2}\sigma_{kn}|x_0\right), \widehat{\xi}\left(p + z_{\alpha/2}\sigma_{kn}|x_0\right)\right] \\ &= \left[Y_{n,\left(\left[k\left(p - z_{\alpha/2}\sigma_{kn}\right)\right]\right)}, Y_{n,\left(\left[k\left(p + z_{\alpha/2}\sigma_{kn}\right)\right]\right)}\right], \end{aligned} \tag{16}$$

where $\sigma_{kn} = \sqrt{p(1-p)\varpi_{x_0}^2/k_n}$ and $Y_{n,(i)}$ denotes the i-th order statistic of $\{Y_{n,i}\}_{i=1}^k$. Notice that σ_{kn} involves no covariates' density and the constant factor $\varpi_{x_0}^2$ is the volume of the unit ball in $\mathcal{R}^d$, which appears in the asymptotic variance of the standard asymmetric k-NN estimator, see Mack (1981). The new confidence interval CI-O$_{1-\alpha}$ defined in (16) for conditional quantiles shares the elegance and simplicity of the confidence interval for unconditional quantiles based on order statistics.

Example 4.2. (Nonparametric Quantile Regression With Censoring): Consider a nonparametric censored quantile regression model where the dependent variable Y_i is subject to conditional random censoring C_i. So instead of observing $\{X_i, Y_i\}_{i=1}^n$, we observe a random sample $\left(\min\left(Y_i, C_i\right), \delta_i, X_i\right)_{i=1}^n$, where $\delta_i = 1\{Y_i \leq C_i\}$, Y_i and C_i are independent of each other conditional on X_i. Dabrowska (1987) extends various nonparametric quantile regression estimators for random samples including the kernel estimator, the symmetric k-NN estimator, and the asymmetric k-NN estimator to the above censored case. Under primitive conditions, she establishes weak convergence of the associated quantile processes which ensures Assumption (G) (iii). Thus our generic confidence interval defined in (14) is asymptotically valid.

Example 4.3. (Nonparametric Quantile Regression for Time Series): Suppose $\{X_i, Y_i\}_{i=1}^n$ is a realization of a stationary α-mixing process. Xu (2012) constructs an asymptotically valid confidence interval for $\xi(p|x_0)$ via the empirical likelihood approach. The generic confidence interval (14) shares the advantages of the empirical likelihood based confidence interval and is less involved computationally. Let $\widehat{\xi}(p|x_0)$ denote a local polynomial quantile regression estimator or a generalized inverse of a kernel estimator of $F_{Y|X}(y|x_0)$. Under primitive conditions, Su and White (2011) and Polonik and Yao (2002) establish respectively Bahadur representation[5] for $\widehat{\xi}(p|x_0)$ valid uniformly over $p \in (0,1)$, where the linear representation is proportional to $q_p(x_0)$. So Assumption (G) (iii) is satisfied under their conditions, where $b_n = \sqrt{nh_n^d}$ with h_n the bandwidth. For unconditional quantiles, Wu (2005) establishes a uniform Bahadur representation for sample quantiles for a wide class of processes which can be used to justify CI-G$_{1-\alpha}$ for sample quantiles.[6]

5 Simulation

In this section, we investigate the finite sample performance of our new confidence intervals for nonparametric and partially linear quantile regressions and compare them with Wald-type confidence intervals and two bootstrap confidence intervals. In order to see the separate effects of estimating $f_X(x)$ and $q_p(x)$, we use both the Nadaraya-Watson estimator and Stute's symmetric k-NN estimator of the conditional distribution function of Y given X. In sum, we compare four asymptotic confidence intervals in our simulation. For the nonparametric quantile regression, they take the following forms:

$$\text{W-NW}_{1-\alpha} = \left(\hat{\xi}_{p,NW}(x) - \frac{z_{\frac{\alpha}{2}}\sigma_{np}(K)\,\widehat{q}_{p,NW}(x)}{\sqrt{\hat{f}_X(x)}}, \hat{\xi}_{p,NW}(x) + \frac{z_{\frac{\alpha}{2}}\sigma_{np}(K)\,\widehat{q}_{p,NW}(x)}{\sqrt{\hat{f}_X(x)}}\right], \quad (17)$$

$$\text{W-S}_{1-\alpha} = \left(\hat{\xi}_p(x) - z_{\alpha/2}\sigma_{np}(K)\,\widehat{q}_p(x), \hat{\xi}_p(x) + z_{\alpha/2}\sigma_{np}(K)\,\widehat{q}_p(x)\right], \quad (18)$$

$$\text{CI-NW}_{1-\alpha} = \left(\widehat{F}_{n,NW}^{-1}\left(p - \frac{z_{\alpha/2}\sigma_{np}(K)}{\sqrt{\hat{f}_X(x)}}|x\right), \widehat{F}_{n,NW}^{-1}\left(p + \frac{z_{\alpha/2}\sigma_{np}(K)}{\sqrt{\hat{f}_X(x)}}|x\right)\right], \quad (19)$$

$$\text{CIN}_{1-\alpha} = \left(\widehat{F}_n^{-1}\left(p - z_{\alpha/2}\sigma_{np}(K)\,|x\right), \widehat{F}_n^{-1}\left(p + z_{\alpha/2}\sigma_{np}(K)\,|x\right)\right],$$

[5] To save space, we refer the reader to Su and White (2011) and Polonik and Yao (2002) for details.
[6] There is no covariate for sample quantiles.

where $\widehat{q}_p(x) = 1/\hat{f}_{Y|X}\left(\hat{\xi}_p(x)|x\right)$, $\widehat{q}_{p,NW}(x) = 1/\hat{f}_{Y|X}\left(\hat{\xi}_{p,NW}(x)|x\right)$, $\hat{\xi}_{p,NW}(x) = \widehat{F}^{-1}_{n,NW}(p|x)$ in which $\widehat{F}^{-1}_{n,NW}(p|x)$ is the generalized inverse of $\widehat{F}_{n,NW}(y|x)$ defined as

$$\widehat{F}_{n,NW}(y|x) = \frac{\sum_{i=1}^n 1\{Y_i \leq y\} K\left(\frac{x-X_i}{h_{n,NW}}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_{n,NW}}\right)}, \tag{20}$$

and

$$\hat{f}_X(x) = \frac{1}{nh_X}\sum_{i=1}^n K\left(\frac{x-X_i}{h_X}\right),\ \hat{f}_{Y|X}(y|x) = \frac{\sum_{i=1}^n K\left(\frac{y-Y_i}{h_{C,Y}}\right) K\left(\frac{x-X_i}{h_{C,X}}\right)}{h_{C,Y}\sum_{i=1}^n K\left(\frac{x-X_i}{h_{C,X}}\right)}, \tag{21}$$

in which h_X, $h_{C,X}$, $h_{C,Y}$, and $h_{n,NW}$ are all bandwidths that need to be chosen.

While the first two confidence intervals, W-NW$_{1-\alpha}$ and W-S$_{1-\alpha}$, are both Wald-type confidence intervals relying on a consistent estimator of the conditional quantile density function, W-S$_{1-\alpha}$ does not require a consistent estimator of the covariate density function $f_X(x)$. The two new confidence intervals,[7] CIN$_{1-\alpha}$and CI-NW$_{1-\alpha}$, make use of the conditional quantile estimators directly. They differ in the quantile estimators being used the consequence of which is that CIN$_{1-\alpha}$ does not depend on any density estimation, but CI-NW$_{1-\alpha}$ depends on a consistent estimator of the covariate density $f_X(x)$.

Throughout the simulation, we used the Bisquare Kernel function, $K(u) = \frac{15}{16}\left(1-u^2\right)^2 I\{|u| \leq 1\}$. The choice of bandwidths is delicate and will be discussed below. Among these four confidence intervals, our new confidence interval, CIN$_{1-\alpha}$, is the least demanding in terms of bandwidth choice, as it only requires choosing one bandwidth which is needed to estimate the conditional quantile function. In sharp contrast, the Wald-type confidence interval, W-NW$_{1-\alpha}$, is the most demanding, as there are four bandwidths involved.[8]

As the Wald-type inference is known to be poor in linear models (see Kocherginsky, He, and Mu, 2005), we also compared our confidence intervals with the following two bootstrap competitors:

$$\text{Boot-Norm}_{1-\alpha} = \left(\hat{\xi}_p(x) - z_{\alpha/2}\sigma_{\text{Boot}}, \hat{\xi}_p(x) + z_{\alpha/2}\sigma_{\text{Boot}}\right],$$

$$\text{Boot-Perc}_{1-\alpha} = \left(\hat{\xi}_p(x) - z_{\text{Boot},1-\alpha/2}, \hat{\xi}_p(x) + z_{\text{Boot},\alpha/2}\right], \tag{22}$$

where σ_{Boot} is the bootstrap standard deviation for $\hat{\xi}_p(x)$ and $z_{\text{Boot},\alpha/2}$ is the bootstrap percentile.

In the tables below, we denote these confidence intervals, W-NW$_{1-\alpha}$, W-S$_{1-\alpha}$, CI-NW$_{1-\alpha}$, CIN$_{1-\alpha}$, Boot-Norm$_{1-\alpha}$ and Boot-Perc$_{1-\alpha}$ as 'Asy NW', 'Asy CI', 'New NW', 'New CI', 'BootNm', and 'BootPerc' respectively.

[7] Section 2 establishes the asymptotic validity of CIN$_{1-\alpha}$. The asymptotic validity of CI-NW$_{1-\alpha}$ can be established using Theorem 4.1.

[8] The results in Tables 1-7 reveal the best performance of the Wald-type confidence intervals when these bandwidths are different and chosen carefully and the worst performance when these bandwiths are chosen to be the same.

5.1 Nonparametric Quantile Regression

The first two designs are taken from Yu and Jones (1998). Model 1 gives curvy quantile with homoskedasticity while Model 2 exhibits almost linear quantile with heteroskedasticity:

$$\begin{aligned} \text{Model 1} &: \quad Y_i = 2.5 + \sin(2X_i) + 2\exp\left(-16X_i^2\right) + 0.5\varepsilon_i \text{ and} \\ \text{Model 2} &: \quad Y_i = \sin(0.75X_i) + 1 + 0.3\sqrt{(\sin(0.75X_i) + 1)}\varepsilon_i, \end{aligned}$$

where X_i and ε_i are independent bivariate normal with standard normal marginal distributions.

We computed the coverage rates of six confidence intervals based 5,000 simulations with sample size n varying from 200, 500 to 1000 and nominal size equal to 95%. The bootstrap replication[9] is set to be 500. The confidence interval, W-NW$_{1-\alpha}$, involves four bandwidths: (i) the bandwidth $h_{n,NW}$ in the quantile estimator is chosen to be $n^{-1/20}h_{YJ}$, where h_{YJ} is the rule of thumb bandwidth in Yu and Jones (1998) based on a preliminary Ruppert-Sheather-Wand bandwidth. The presence of the factor $n^{-1/20}$ reflects the slightly undersmoothing requirement in our Assumption (H); (ii) the bandwidth h_X in $\hat{f}_X(x)$ is chosen to be the Sheather-Jones bandwidth with Silverman's rule of thumb as the pilot estimate; (iii) the two bandwidths $(h_{C,Y}, h_{C,X})$ in the conditional quantile density estimator are chosen by the 'normal-reference' rule in Racine's *np* package. The bandwidths involved in the remaining three confidence intervals, W-S$_{1-\alpha}$, CI-NW$_{1-\alpha}$, and CIN$_{1-\alpha}$, are chosen in the same way. However it is worth mentioning that h_n and $h_{n,NW}$ are different as the first one is based on the sample $(Y_i, F_n(X_i))$ after we transform X_i using its empirical distribution function.[10] The results are presented in Tables 1-3 for different sample sizes.

Insert Tables 1-3 here

Several observations follow immediately from Tables 1-3. First, the performance of the two new CIs based on pairs of quantile estimates is very stable across models, quantile levels, and sample sizes, especially our new CI using the symmetric k-NN estimator—its performance is comparable to the computationally more extensive Bootstrap percentile method and in many cases better with finite sample coverage rate very close to the nominal level even for sample size 200; Second, the performance of the two Wald-type CIs is not as stable. For small sample sizes, their coverage rates at most covariate points for both models are not close to the nominal level. Even at sample size 1000, the coverage rates of the two Wald-type CIs could be far away from the nominal level, e.g.,

[9] To ease the computational burden, we fixed the bandwidth for the bootstrap sample.

[10] We need to truncate the support of X in order to avoid the crash of computation of Ruppert-Sheather-Wand bandwidth for the Nadaraya-Watson type estimators. In particular, for Model 1, we restrict the computation of the R-S-W bandwidth only for those points whose covariate values are in $[-1.65, 1.65]$ and for Model 2, the restricted range is $[-2, 2]$. When it comes to the small sample with 200 observations, we always truncate at $[-0.75, 0.75]$ for both models. In contrast, the empirical probability integral transformation prevents this crash due to the equal spacing of sample points.

0.991, 0.9896 for Model 1 when $x = 0$ and $p = 0.5$ and 0.9302, 0.9254 for Model 2 when $x = 1.5$ and $p = 0.25$; For the two bootstrap confidence intervals, the one based on normal approximation is biased towards undercovering even in relatively large samples, while the one based on the percentile approach is much more accurate, but showing some variability in small samples.

To see the sensitivity of Wald-type confidence intervals to the choice of bandwidths, we also computed their coverage rates using one bandwidth only, the bandwidth in the conditional quantile estimate. Table 4 presents the results for sample size 1000. For comparison purposes, we also presented the coverage rates for the two new confidence intervals, $\text{CIN}_{1-\alpha}$ and $\text{CI-NW}_{1-\alpha}$.[11] The coverage rates for $\text{CIN}_{1-\alpha}$ are the same as in Table 3. Interestingly we observe that the coverage rate of $\text{CI-NW}_{1-\alpha}$ does not change much, but the performance of the two Wald-type intervals is very poor for Model 1.

Insert Table 4 here

Overall these results reveal the superior performance of our new confidence interval, $\text{CIN}_{1-\alpha}$, and the sensitivity of Wald-type confidence intervals to the choice of bandwidths in the estimation of the conditional quantile density function.

5.2 Partial Linear Quantile Regression

The design is adapted from Song, Ritov, and Hardle (2012) and the finite dimensional parameter β was estimated by the method proposed in Song, Ritov, and Hardle (2012):

Model 3: $Y_i = 2Z_i + {X_i}^2 + \varepsilon_i$,

where X_i, Z_i, and ε_i are independent of each other, $X_i \sim U(0,1)$, $Z_i \sim U(0,2)$, and ε_i is standard normal.

Our new confidence interval, $\text{CIPL}_{1-\alpha}$, is presented in (9). Modifications will be required to the other three types of confidence intervals for partial linear models. Specifically, we need to replace Y_i with $Y_i - Z_i\hat{\beta}$ in computing $\widehat{F}_{n,NW}(p|x)$, $\widehat{F}_n(p|x)$, and $\hat{f}_{Y|X}\left(\hat{\xi}_p(x)|x\right)$ and also add $z'\hat{\beta}$ to both end points of the intervals in (17), (18), and (19). The bandwidths are chosen in the same way as in the nonparametric model. Tables 5 and 6 report results for $n = 500, 1000$.

Insert Tables 5 and 6 here

Like in the nonparametric case, the two new confidence intervals based on pairs of estimated quantiles perform remarkably well across covariate values, quantile levels, and sample sizes. Their

[11] The two bootstrap confidence intervals also require only one bandwidth from estimating the conditional quantile, hence the results would not change from Table 3 and we will not replicate that part.

performance is comparable and sometimes better than the Boot-Norm$_{1-\alpha}$ which performs better than Boot-Perc$_{1-\alpha}$ for the partial linear model. In contrast the two Wald-type intervals do not perform well even when the sample size is 1000.

We also computed the coverage rates of the first three confidence intervals using one bandwidth only, the bandwidth in the conditional quantile estimate. Table 7 presents the results for sample size 1000. Again the performance of the Wald-type intervals deteriorates dramatically.

Insert Table 7 here

6 Concluding Remarks

In this paper, we have constructed "density-free" confidence intervals and bands for conditional quantiles based on estimated conditional quantiles evaluated at two appropriately chosen quantile levels. In contrast to Wald-type confidence intervals or bands based on the asymptotic distributions of estimators of the conditional quantiles, our confidence intervals and bands circumvent the need to estimate the density of the covariate and the conditional quantile density of the response variable, thus freeing practitioners from choosing bandwidths involved in estimating the covariate density and the conditional quantile density. A small Monte Carlo study reveals the superior finite sample performance of our new CIs compared with the Wald-type CIs that are sensitive to the choice of bandwidth needed to estimate the conditional quantile density function and two bootstrap CIs.

We have also presented a generic confidence interval for conditional quantiles using the rearranged quantile curves that is asymptotically valid for any quantile regression (parametric, nonparametric, or semiparametric), any method of estimation, and any data structure, provided that the conditional quantile function satisfies some mild smoothness assumptions and the original quantile estimator is such that its associated quantile process converges weakly to a Gaussian process with a covariance kernel proportional to the conditional quantile density function.

As far as we know, this paper is the first paper presenting a systematic investigation of the direct approach to inference in nonparametric and semiparametric quantile regression models. Given the simplicity and superior performance of this approach compared with existing approaches, it would be worthwhile investigating its applicability in other contexts. One example is inference on the finite dimensional parameter in semiparametric models. This paper has focused exclusively on inference for the conditional quantile function. In semiparametric models, the finite dimensional parameter might be of interest. It would also be interesting to extend our generic confidence interval to a generic confidence band across the covariates' support. This is more challenging, since the confidence bands for parametric and nonparametric models differ substantially. For the location-scale forms of linear quantile models, Zhou and Portnoy (1996) construct Scheffe type confidence band using the direct approach (see their Proposition 3.1) based on chi-square asymptotics, while we construct our confidence bands based on the extreme-value asymptotics for nonparametric and semiparametric models.

7 Appendix A. Technical Proofs For Section 2

Throughout the proofs, M denotes an unspecified positive constant and its value does not depend on n and typically does not depend on $x \in \mathcal{J}$ and y either (This will be clear in specific context that M is used); Δ denotes an intermediate value in the Taylor series expansion. The values of both M and Δ may vary from line to line. Also the limits are taken as $n \to \infty$ unless stated otherwise.

We define $\widehat{f}_U(x)$ and $\widetilde{f}_U(x)$ below which appear frequently in the proofs:

$$\widehat{f}_U(x) = \frac{1}{nh_n}\sum\nolimits_{i=1}^{n} K\left(\frac{F_n(x) - F_n(X_i)}{h_n}\right) \text{ and } \widetilde{f}_U(x) = \frac{1}{nh_n}\sum\nolimits_{i=1}^{n} K\left(\frac{F_X(x) - F_X(X_i)}{h_n}\right).$$

Proof of Theorem 2.1. First, we show that

$$\begin{aligned}
\sqrt{nh_n}\left(\widehat{F}_n^{-1}\left(p + z_{\alpha/2}\sigma_{np}(K)\,|x_0\right) - \widehat{F}_n^{-1}\left(p\,|x_0\right)\right) &= \frac{z_{\alpha/2}\sqrt{R(K)\,p(1-p)}}{f_{Y|X}\left(\xi_p(x_0)\,|x_0\right)} + o_p(1) \text{ and} \\
\sqrt{nh_n}\left(\widehat{F}_n^{-1}\left(p - z_{\alpha/2}\sigma_{np}(K)\,|x_0\right) - \widehat{F}_n^{-1}\left(p\,|x_0\right)\right) &= -\frac{z_{\alpha/2}\sqrt{R(K)\,p(1-p)}}{f_{Y|X}\left(\xi_p(x_0)\,|x_0\right)} + o_p(1).
\end{aligned}$$

We provide a detailed proof of the first result. The second can be proved similarly. It follows from Lemma 2.2 (ii) that the conditional quantile process: $\left\{\sqrt{nh_n}\left[\widehat{F}_n^{-1}(p|x_0) - F_{Y|X}^{-1}(p|x_0)\right] : p \in [p_1, p_2]\right\}$ converges weakly in $l^\infty(0,1)$. This implies that

$$\begin{aligned}
&\sqrt{nh_n}\left[\widehat{F}_n^{-1}\left(p + z_{\alpha/2}\sigma_{np}(K)\big|\, x_0\right) - F_{Y|X}^{-1}\left(p + z_{\alpha/2}\sigma_{np}(K)\big|\, x_0\right)\right] \\
&-\sqrt{nh_n}\left[\widehat{F}_n^{-1}(p|x_0) - F_{Y|X}^{-1}(p|x_0)\right] \\
=\;& o_p(1).
\end{aligned}$$

That is,

$$\begin{aligned}
&\sqrt{nh_n}\left(\widehat{F}_n^{-1}\left(p + z_{\alpha/2}\sigma_{np}(K)\,|x_0\right) - \widehat{F}_n^{-1}\left(p\,|x_0\right)\right) \\
=\;& \sqrt{nh_n}\left[F_{Y|X}^{-1}\left(p + z_{\alpha/2}\sigma_{np}(K)\,|x_0\right) - F_{Y|X}^{-1}(p|x_0)\right] + o_p(1) \\
=\;& \frac{z_{\alpha/2}\sqrt{R(K)\,p(1-p)}}{f_{Y|X}\left(\xi_p(x_0)\,|x_0\right)} + o_p(1).
\end{aligned}$$

Finally, we obtain:

$$\begin{aligned}
&\Pr\left(\xi_p(x_0) \in (\widehat{F}_n^{-1}\left(p - z_{\alpha/2}\sigma_{np}(K)\,|x_0\right), \widehat{F}_n^{-1}\left(p + z_{\alpha/2}\sigma_{np}(K)\,|x_0\right)]\right) \\
=\;& \Pr\left(\widehat{F}_n^{-1}(p|x_0) - \frac{z_{\alpha/2}\sigma_{np}(K)}{f_{Y|X}\left(\xi_p(x_0)\,|x_0\right)} < \xi_p(x_0) \le \widehat{F}_n^{-1}(p|x_0) + \frac{z_{\alpha/2}\sigma_{np}(K)}{f_{Y|X}\left(\xi_p(x_0)\,|x_0\right)}\right) + o(1) \\
=\;& \Pr\left(\sqrt{nh_n}\left|\widehat{F}_n^{-1}(p|x_0) - \xi_p(x_0)\right| \le z_{\alpha/2}\sigma\right) + o(1) \\
=\;& (1-\alpha) + o(1).
\end{aligned}$$

Q.E.D

We now present a few lemmas used in the proofs of Theorem 2.3 and Lemma 2.4. Let

$$B_n(x, X_i) = F_n(x) - F_n(X_i) - F_X(x) + F_X(X_i).$$

Lemma A.1 *(Stute, 1982) Under our Assumptions (H), (K), and (X),*

(i) for any given $x_0 \in \mathcal{X}$*, we have:* $\sqrt{nh_n^{-1}} \sup_{|F_X(x_0) - F_X(X_i)| \leq Mh_n} |B_n(x_0, X_i)| = O_p(1)$*;*

(ii) uniformly over $x \in \mathcal{X}$*, we have:* $\sqrt{n(h_n \log n)^{-1}} \sup_{|F_X(x) - F_X(X_i)| \leq Mh_n} |B_n(x, X_i)| = O_p(1)$*.*

Proof. Those bounds actually hold almost surely. The proof could be found in the proofs of Lemma 2.4 and Theorem 2.14 in Stute (1982). **Q.E.D**

Lemma A.2 *Given Assumptions* $(\tilde{S})$ *(ii) and (K), for any interior point* $x \in \mathcal{X}$ *and any* $y \in \mathcal{Y}$*, when* $h_n \to 0$*, we have*

$$\left| \frac{1}{h_n} \int F_{Y|X}(y|X_i) K\left(\frac{F_X(x) - F_X(X_i)}{h_n}\right) dF_X(X_i) - F_{Y|X}(y|x) \right| \leq Mh_n^2,$$

where M *is independent of* y *and* x *in* $\mathcal{J} \subset \mathcal{X}$*.*

Proof. Let $u = F_X(x)$. Then

$$\begin{aligned} & \left| \frac{1}{h_n} \int F_{Y|X}(y|X_i) K\left(\frac{F_X(x) - F_X(X_i)}{h_n}\right) dF_X(X_i) - F_{Y|X}(y|x) \right| \\ = & \left| \int F_{Y|X}\left(y|F_X^{-1}(u - Uh_n)\right) K(U) dU - F_{Y|X}\left(y|F_X^{-1}(u)\right) \right| \\ = & \left| \int \left[H(y|u - Uh_n) - H(y|u)\right] K(U) dU \right|. \end{aligned}$$

The first equality is obtained by a change of variables and the second one is just rewritten in terms of $H(y|u)$. Notice that $U \in [0,1]$. By Assumption (S̃) (ii) and Lemma 8.5 in Korostelev and Korosteleva (2011), we have

$$H(y|u - Uh_n) - H(y|u) = Uh_n H'(y|u) + \rho(u, U)$$

with $|\rho(u, U)| \leq \frac{Lh_n^2}{2}$, for an L independent of u or y. The result follows immediately from Assumption (K). **Q.E.D**

Let

$$\tilde{F}_n(\cdot|x) = \frac{\frac{1}{nh_n} \sum_{i=1}^n 1\{Y_i \leq \cdot\} K\left(\frac{F_X(x) - F_X(X_i)}{h_n}\right)}{\frac{1}{nh_n} \sum_{i=1}^n K\left(\frac{F_X(x) - F_X(X_i)}{h_n}\right)}.$$

The lemma below shows that $\tilde{F}_n(\cdot|\cdot)$ and $\widehat{F}_n(\cdot|\cdot)$ are uniformly close to each other. This also demonstrates that instead of treating $\widehat{F}_n(\cdot|\cdot)$ as the nearest neighbor estimator based on ranks, it could also be viewed as a feasible kernel estimator after probability integral transformation.

Lemma A.3 *Under our Assumptions (H), (K), and ($\tilde{X}$), we have:*

$$\sup_{y\in\mathcal{Y}}\sup_{x\in\mathcal{J}}\sqrt{\frac{nh_n}{\log n}}\left|\widehat{F}_n(y|x)-\tilde{F}_n(y|x)\right| = o_p(1). \tag{A.1}$$

Proof. Decompose $\widehat{F}_n(y|x)-\tilde{F}_n(y|x)$ as in the following standard way,

$$\begin{aligned}
&\widehat{F}_n(y|x)-F_{Y|X}(y|x)+F_{Y|X}(y|x)-\tilde{F}_n(y|x)\\
=\;& \frac{1}{nh_n\widehat{f}_U(x)}\sum_{i=1}^{n}\left[1\{Y_i\le y\}-F_{Y|X}(y|x)\right]K\left(\frac{F_n(x)-F_n(X_i)}{h_n}\right)\\
&-\frac{1}{nh_n\widetilde{f}_U(x)}\sum_{i=1}^{n}\left[1\{Y_i\le y\}-F_{Y|X}(y|x)\right]K\left(\frac{F_X(x)-F_X(X_i)}{h_n}\right)\\
=\;& \frac{1}{nh_n\widehat{f}_U(x)}\sum_{i=1}^{n}\left[1\{Y_i\le y\}-F_{Y|X}(y|x)\right]\left[K\left(\frac{F_n(x)-F_n(X_i)}{h_n}\right)-K\left(\frac{F_X(x)-F_X(X_i)}{h_n}\right)\right]+\\
&\frac{1}{nh_n}\sum_{i=1}^{n}\left[1\{Y_i\le y\}-F_{Y|X}(y|x)\right]K\left(\frac{F_X(x)-F_X(X_i)}{h_n}\right)\left(\frac{1}{\widehat{f}_U(x)}-\frac{1}{\widetilde{f}_U(x)}\right).
\end{aligned}$$

In order to handle the denominator and the term in parenthesis in the above decomposition, we need to bound the difference $\widehat{f}_U(x)-\widetilde{f}_U(x)$:

$$\begin{aligned}
\widehat{f}_U(x)-\widetilde{f}_U(x) &= \frac{1}{nh_n}\sum\left[K\left(\frac{F_n(x)-F_n(X_i)}{h_n}\right)-K\left(\frac{F_X(x)-F_X(X_i)}{h_n}\right)\right]\\
=\;& \frac{1}{nh_n^2}\sum K'\left(\frac{F_X(x)-F_X(X_i)}{h_n}\right)\left[F_n(x)-F_X(x)-F_n(X_i)+F_X(X_i)\right]\\
&+\frac{1}{nh_n^3}\sum K''(\Delta)\left[B_n(x,X_i)\right]^2.
\end{aligned}$$

It would be clear in a moment that the above difference could be shown as $O_p\left(\sqrt{\frac{\log n}{nh_n}}\right)$ uniformly. For the present purpose, it suffices that the difference is uniformly $o_p(1)$. Hence we could just focus on the first term's numerator.

Similarly, the numerator for the first term admits the following decomposition,

$$\begin{aligned}
&\frac{1}{nh_n}\sum\left[1\{Y_i\le y\}-F_{Y|X}(y|x)\right]\left[K\left(\frac{F_n(x)-F_n(X_i)}{h_n}\right)-K\left(\frac{F_X(x)-F_X(X_i)}{h_n}\right)\right]\\
=\;& \frac{1}{nh_n^2}\sum\left[1\{Y_i\le y\}-F_{Y|X}(y|x)\right]K'\left(\frac{F_X(x)-F_X(X_i)}{h_n}\right)\left[F_n(x)-F_X(x)-F_n(X_i)+F_X(X_i)\right]\\
+\;& \frac{1}{nh_n^3}\sum\left[1\{Y_i\le y\}-F_{Y|X}(y|x)\right]K''(\Delta)\left[B_n(x,X_i)\right]^2\\
\doteq\;& I_n+II_n
\end{aligned}$$

For II_n, as argued in Lemma 1 in Stute (1984b), for any x, we only need to consider those sample points for which $|F_n(x)-F_n(X_i)|\le h_n$ and by the Kvorezky-Kiefer-Wolfowitz bound, we have $\sup_x|F(x)-F_n(x)|\le Cn^{-1/2}$. Therefore we only need to consider the oscillation restricted by $\sup_x|F_X(x)-F_X(X)|\le Ch_n$, so

$$II_n\le O_p\left(\frac{h_n\log n}{n}\right)\frac{|K''(\Delta)|}{nh_n^3}\sum\left|1\{Y_i\le y\}-F_{Y|X}(y|x)\right| = O_p\left(\frac{\log n}{nh_n^2}\right).$$

To handle I_n, we first show that it could be written as a scaled U-statistic plus some smaller order term, and then we characterize the approximation order of the U-statistic by its Hajek projection. Finally we end the derivation by showing that the Hajek projection is $o_p\left(\sqrt{\frac{\log n}{nh_n}}\right)$.

Let $F_{n-1}^{i}(x)$ be the leave-one-out empirical distribution function and define $B_{n-1}^{i}(x, X_i)$ similarly. Proceeding as Lemma 2 in Stute (1984b), we have

$$F_n(x) = F_{n-1}^{i}(x) - n^{-1}F_{n-1}^{i}(x) + n^{-1}1\{X_i \leq x\}.$$

Therefore, $B_n(x, X_i) = B_{n-1}^{i}(x, X_i) + O_p\left(n^{-1}\right)$, where the residual term's order is uniform w.r.t. x by standard Glivenko-Cantelli result.

Now it suffices to consider the following U-process indexed by x: $II_n = h_n^{-2}U_2^n h_{y,x} + s.o.$, where

$$\begin{aligned} & U_2^n h_{y,x} \\ = \ & \frac{2}{n(n-1)}\sum_{i\neq j}\frac{1}{2}\left(\begin{array}{c} \left[1\{Y_i \leq y\} - F_{Y|X}(y|x)\right] K'\left(\frac{F_X(x)-F_X(X_i)}{h_n}\right)\left[\begin{array}{c} 1\{X_j \leq x\} - F_X(x) \\ -1\{X_j \leq X_i\} + F_X(X_i) \end{array}\right] \\ + \left[1\{Y_j \leq y\} - F_{Y|X}(y|x)\right] K'\left(\frac{F_X(x)-F_X(X_j)}{h_n}\right)\left[\begin{array}{c} 1\{X_i \leq x\} - F_X(x) \\ -1\{X_i \leq X_j\} + F_X(X_j) \end{array}\right] \end{array}\right). \end{aligned}$$

Consider the function class:

$$\mathcal{F}_{y,x} = \{1\{Y_i \leq y\}K'\left(\frac{F_X(x) - F_X(X_i)}{h_n}\right)\left[1\{X_j \leq x\} - F_X(x) - 1\{X_j \leq X_i\} + F_X(X_i)\right] | y \in \mathcal{R}, x \in \mathcal{J}\}.$$

Because $K'(\cdot)$ has bounded variation, $1\{Y_i \leq y\}$ and $[1\{X_j \leq x\} - F_X(x)]$ ($F_X(x)$ is uniformly continuous by Assumption ($\tilde{\mathrm{X}}$)) are both VC classes, we have

$$\log N_p(\epsilon, \mathcal{F}_{y,x}, Q) \leq M \log\left(\frac{1}{\epsilon}\right), \text{ for } p \in (0, \infty),$$

for any probability measure Q. Therefore, by Lemma A.1 in Ghosal, Sen, and van der Vaart (2000), we can approximate $U_2^n h_{y,x}$ by its Hajek Projection with an error of order n^{-1}, i.e.,

$$\frac{1}{h_n^2}U_2^n h_{y,x} = \frac{1}{h_n^2}U_1^n \Pi_1 h_{y,x} + O_p\left(\frac{1}{nh_n^2}\right).$$

Next, we compute the projection explicitly. Let $U_j = F_X(X_j)$, $U_i = F_X(X_i)$, and $u = F_X(x)$.

Then

$$\frac{1}{h_n^2} U_1^n \Pi_1 h_{y,x} = \frac{1}{nh_n^2} \sum_{j=1}^{n} \int \left[F_{Y|X}(y|X_i) - F_{Y|X}(y|x)\right] K'\left(\frac{F_X(x) - F_X(X_i)}{h_n}\right) \times$$

$$[1\{F_X(X_j) \le F_X(x)\} - F_X(x) - 1\{F_X(X_j) \le F_X(X_i)\} + F_X(X_i)]\, dF_X(X_i)$$

$$= \frac{1}{nh_n} \sum_{j=1}^{n} \int [H(y|U_i) - H(y|u)]\,[1\{U_j \le u\} - u - 1\{U_j \le U_i\} + U_i]\, dK\left(\frac{u - U_i}{h_n}\right)$$

$$= \frac{1}{n} \sum_{j=1}^{n} \int [H(y|u - vh_n) - H(y|u)]\,[1\{U_j \le u\} - u - 1\{U_j \le u - vh_n\} + u - vh_n]\, dK(v)$$

$$\le Mh_n \int \sup_{|u-v| \le h_n} \left| \frac{1}{n} \sum_{j=1}^{n} 1\{U_j \le u\} - u - 1\{U_j \le u - vh_n\} + u - vh_n \right| d|K(v)|$$

$$= O_p\left(h_n \sqrt{\frac{\log n}{nh_n}}\right) = o_p\left(\sqrt{\frac{\log n}{nh_n}}\right),$$

where $u = F_X(x)$. Notice that $|H(y|U) - H(y|u)| \le Mh_n$ for U satisfying $|U - u| \le \frac{h_n}{2}$. The term after the sup in the above inequality is nothing but the local oscillation of the uniform empirical process, whose order is given in Lemma A.1. Also $K(v)$ is of bounded variation, hence the integral term is of order $O_p\left(\sqrt{\frac{\log n}{nh_n}}\right)$.

In sum, by our Assumption (H), we have

$$\sup_{y \in \mathcal{Y}} \sup_{x \in \mathcal{J}} |\widehat{F}_n(y|x) - \tilde{F}_n(y|x)|$$

$$= \left[O_p\left(h_n \sqrt{\frac{\log n}{nh_n}}\right) + O_p\left(\frac{\log n}{nh_n^2}\right)\right] + O_p\left(\sqrt{\frac{\log n}{nh_n}}\right) o_p(1) = o_p\left(\sqrt{\frac{\log n}{nh_n}}\right).$$

Q.E.D

Lemma A.4 *Under our Assumptions (H), (K), ($\tilde{X}$), ($\tilde{S}$)(ii), and (B)(iii), it holds that for any* $c_n = O\left(\sqrt{\frac{\log n}{nh_n}}\right)$, $\sup_{x \in \mathcal{J}} \left|\widehat{F}_n^{-1}(p + c_n|x) - F^{-1}(p|x)\right| = O_p\left(\sqrt{(nh_n)^{-1} \log n}\right)$.

Proof. First of all we have

$$\sup_{y \in \mathcal{Y}} \sup_{x \in \mathcal{J}} \left|\widetilde{F}_n(y|x) - F_{Y|X}(y|x)\right| = O_p\left(\sqrt{\frac{\log n}{nh_n}}\right).$$

This follows directly from Theorem 3 in Einmahl and Mason (2005). Actually it is even easier, because the transformation makes the covariate uniformly distributed, and there is no denominator of any kernel function. Note that we always use an undersmoothing bandwidth to kill the bias (uniformly over x) as shown in Lemma A.2.

It follows from Lemmas A.3 and A.4 that

$$\sup_{y \in \mathcal{Y}} \sup_{x \in \mathcal{J}} \left|\widehat{F}_n(y|x) - F_{Y|X}(y|x)\right| = O_p\left(\sqrt{\frac{\log n}{nh_n}}\right). \tag{A.2}$$

Hence,

$$\begin{aligned}
&\Pr\left[\widehat{F}_n^{-1}\left(p+c_n|x\right)-F^{-1}\left(p|x\right)>M\sqrt{\frac{nh_n}{\log n}}\right] \\
=\ &\Pr\left[p+c_n>\widehat{F}_n\left(F^{-1}\left(p|x\right)+M\sqrt{\frac{nh_n}{\log n}}|x\right)\right] \\
=\ &\Pr\left[\begin{array}{c} p-F\left(F^{-1}\left(p|x\right)+M\sqrt{\frac{nh_n}{\log n}}|x\right) \\ >c_n+\widehat{F}_n\left(F^{-1}\left(p|x\right)+M\sqrt{\frac{nh_n}{\log n}}|x\right)-F\left(F^{-1}\left(p|x\right)+M\sqrt{\frac{nh_n}{\log n}}|x\right)\end{array}\right] \\
=\ &\Pr\left[\begin{array}{c} -f_{Y|X}\left(\Delta|x\right)M> \\ \sqrt{\frac{\log n}{nh_n}}\left(c_n+\widehat{F}_n\left(F^{-1}\left(p|x\right)+M\sqrt{\frac{nh_n}{\log n}}|x\right)-F\left(F^{-1}\left(p|x\right)+M\sqrt{\frac{nh_n}{\log n}}|x\right)\right)\end{array}\right].
\end{aligned}$$

Therefore we obtain

$$\lim_{M\to\infty}\limsup_n \Pr\left[\widehat{F}_n^{-1}\left(p+c_n|x\right)-F^{-1}\left(p|x\right)\geq M\sqrt{\frac{nh_n}{\log n}}\right]=0$$

by the requirement on c_n and (A.2). Analogous argument shows that

$$\lim_{M\to\infty}\limsup_n \Pr\left[\widehat{F}_n^{-1}\left(p+c_n|x\right)-F^{-1}\left(p|x\right)<-M\sqrt{\frac{nh_n}{\log n}}\right]=0$$

and the conclusion follows. **Q.E.D**

Lemma A.5 *Under our Assumptions (H), (K), ($\tilde{X}$), and ($\tilde{S}$)(ii), uniformly in y and $x\in\mathcal{J}$ and for any $a_n=O_p\left(\sqrt{\frac{\log n}{nh_n}}\right)$, it holds that*

$$\left|\widehat{F}_n\left(y+a_n|x\right)-\widehat{F}_n\left(y|x\right)-\tilde{F}_n\left(y+a_n|x\right)+\tilde{F}_n\left(y|x\right)\right|=o_p\left(\frac{\log n}{nh_n}\right).$$

Proof. The proof follows that of Lemma A.3 closely, except that we have $1\{Y_i\leq y\}$ replaced by $1\{y<Y_i\leq y+a_n\}$ (say $a_n\geq 0$ w.l.o.g.). Everything works through straightforwardly up to the Hajek projection. Now the projection becomes

$$\begin{aligned}
&\frac{1}{h_n^2}U_1^n\Pi_1 h_{y,x}=\frac{1}{nh_n}\sum_{j=1}^n\int\left[H\left(y+a_n|U\right)-H\left(y|U\right)\right]\times \\
&\left[1\{U_j\leq u\}-u-1\{U_j\leq U_i\}+U_i\right]dK\left(\frac{u-U_i}{h_n}\right)+s.o. \\
\leq\ &a_nM\sup\int\left|\frac{1}{n}\sum_{j=1}^n 1\{U_j\leq u\}-u-1\{U_j\leq u-vh_n\}+u-vh_n\right|d|K\left(v\right)| \\
=\ &o_p\left(\frac{\log n}{nh_n}\right),
\end{aligned}$$

where the integral term in the above inequality is handled similarly as the proof of Lemma A.3.

For II_n, this $a_n M$ term could also be factored out:

$$\begin{aligned} II_n &= O_p\left(\frac{h_n \log n}{n}\right) \frac{K''(\Delta)}{nh_n^3} \sum \left[1\{y < Y_i \le y + a_n\} - F_{Y|X}(y + a_n|x) + F_{Y|X}(y|x)\right] \\ &= O_p\left(\frac{h_n \log n}{n}\right) \frac{K''(\Delta)}{h_n^3} \Pr\{y < Y_i \le y + a_n\} - a_n f_{Y|X}(y|x) \\ &= O_p\left(\frac{a_n \log n}{nh_n^2}\right). \end{aligned}$$

Q.E.D

Lemma A.6 *Under our Assumptions (H), (K), ($\tilde{X}$), ($\tilde{S}$)(ii), and (B)(iii)(iv),*

$$\sup_{x\in\mathcal{J}} \sup_{|y'|\le a_n} \left|\left[\widetilde{F}_n\left(F^{-1}(p|x) + y'|x\right) - \widetilde{F}_n\left(F^{-1}(p|x)|x\right)\right] - \left[F\left(F^{-1}(p|x) + y'|x\right) - F\left(F^{-1}(p|x)|x\right)\right]\right|$$
$$= O_p\left(\left(\frac{\log n}{nh_n}\right)^{3/4}\right),$$

where a_n is any positive sequence of order $O\left(\sqrt{\frac{\log n}{nh_n}}\right)$.

Proof. The standard decomposition shows that the problem could be simplified a bit:

$$\begin{aligned} &\left|\left[\widetilde{F}_n\left(F^{-1}(p|x) + y'|x\right) - \widetilde{F}_n\left(F^{-1}(p|x)|x\right)\right] - \left[F\left(F^{-1}(p|x) + y'|x\right) - F\left(F^{-1}(p|x)|x\right)\right]\right| \\ =& \frac{1}{nh_n \widetilde{f}_U(x)} \sum_{i=1}^n \left[\begin{array}{c} 1\{Y_i \le F^{-1}(p|x) + y'\} - 1\{Y_i \le F^{-1}(p|x)\} \\ -\left[F\left(F^{-1}(p|x) + y'|x\right) - F\left(F^{-1}(p|x)|x\right)\right] \end{array}\right] K\left(\frac{F_X(x) - F_X(X_i)}{h_n}\right) \\ =& \frac{1}{\widetilde{f}_U(x)} \left(\begin{array}{c} \frac{1}{nh_n}\sum_{i=1}^n \left[1\{Y_i \le F^{-1}(p|x) + y'\} - 1\{Y_i \le F^{-1}(p|x)\}\right] K\left(\frac{F_X(x) - F_X(X_i)}{h_n}\right) \\ -\left[F\left(F^{-1}(p|x) + y'|x\right) - F\left(F^{-1}(p|x)|x\right)\right] \end{array}\right) \\ &+ \frac{\left[F\left(F^{-1}(p|x) + y'|x\right) - F\left(F^{-1}(p|x)|x\right)\right]}{\widetilde{f}_U(x)} \times \left[1 - \widetilde{f}_U(x)\right]. \end{aligned}$$

It is clear from the above decomposition that we only need to work with the first term's numerator. Another simplification is that we could modulo the bias term along the derivation. By the standard kernel technique one could get (uniformly)

$$\begin{aligned} &E\left[\widetilde{F}_n\left(F^{-1}(p|x) + y'|x\right) - \widetilde{F}_n\left(F^{-1}(p|x)|x\right)\right] - \left[F\left(F^{-1}(p|x) + y'|x\right) - F\left(F^{-1}(p|x)|x\right)\right] \\ =& O_p\left(\sqrt{\frac{\log n}{nh_n}} \times h_n^2\right) = o_p\left(\left(\frac{\log n}{nh_n}\right)^{3/4}\right), \end{aligned}$$

by Taylor expansion w.r.t x, ($\tilde{S}$)(ii) and applying Assumption (B)(iii) to getting linearization w.r.t y. Hence it suffices to characterize the stochastic order of the following term uniformly:

$$\begin{aligned} \omega(x, y; h_n, a_n) &= \frac{1}{nh_n}\sum_{i=1}^n \left[1\{Y_i \le y + y'\} - 1\{Y_i \le y\}\right] K\left(\frac{F_X(x) - F_X(X_i)}{h_n}\right) \\ &- E\left(\frac{1}{nh_n}\sum_{i=1}^n \left[1\{Y_i \le y + y'\} - 1\{Y_i \le y\}\right] K\left(\frac{F_X(x) - F_X(X_i)}{h_n}\right)\right). \end{aligned}$$

Notice that by Lipschitz continuity of $F_Y(\cdot)$, the shrinkage along y axis could be translated to $F_Y(\cdot)$ upon multiplying some finite Lipschitz constant L in front:

$$
\begin{aligned}
& |\omega(x,y;h_n,a_n)| \\
\le\; & \frac{M}{h_n}\int_{F_Y(y)\wedge\left(F_Y(y)+Ly'\right)}^{F_Y(y)\vee\left(F_Y(y)+Ly'\right)} \int \left| K\left(\frac{F_X(x)-F_X(X)}{h_n}\right)\right| d\left|\mathcal{C}_n\left(F_X(X),F_Y(Y)\right)-\mathcal{C}\left(F_X(X),F_Y(Y)\right)\right| \\
\le\; & \frac{M}{h_n}\int_{F_Y(y)\wedge\left(F_Y(y)+Ly'\right)}^{F_Y(y)\vee\left(F_Y(y)+Ly'\right)} \int_{F_X(x)-Mh_n}^{F_X(x)+Mh_n} d\left|\mathcal{C}_n(u,v)-\mathcal{C}(u,v)\right|,
\end{aligned}
$$

where $\mathcal{C}_n$ and $\mathcal{C}$ denote the empirical and population copula function between (Y,X) respectively. The double integral term corresponds to the multivariate local oscillation of empirical process within a shrinking rectangle studied by Stute (1984b). By Theorem 1.5 or Theorem 3.1 in Stute (1984a) and existence and boundedness of the copula density we have

$$
\begin{aligned}
& \sup_{y,x}\int_{F_Y(y)\wedge\left(F_Y(y)+Ly'\right)}^{F_Y(y)\vee\left(F_Y(y)+Ly'\right)} \int_{F_X(x)-Mh_n}^{F_X(x)+Mh_n} d\left|\mathcal{C}_n(u,v)-\mathcal{C}(u,v)\right| \\
=\; & O_p\left(\frac{\sqrt{h_n a_n}\sqrt{\log\left(h_n a_n\right)^{-1}}}{\sqrt{n}}\right) = O_p\left(\frac{\sqrt{h_n}\sqrt{\log n}}{\sqrt{n}}\times\left(\frac{\log n}{nh_n}\right)^{1/4}\right).
\end{aligned}
$$

Hence overall, we get:

$$
\sup|\omega(x,y;h_n,a_n)| \le O_p\left(\frac{1}{h_n}\times\frac{\sqrt{h_n}\sqrt{\log n}}{\sqrt{n}}\times\left(\frac{\log n}{nh_n}\right)^{1/4}\right) = O_p\left(\left(\frac{\log n}{nh_n}\right)^{3/4}\right).
$$

Q.E.D

Lemma A.7 *Under Assumptions (H), (K), ($\tilde{X}$), ($\tilde{S}$)(ii), and (B)(ii)(iii)(iv), we have*

$$
\widehat{F}_n^{-1}(p+c_n|x) - F^{-1}(p|x) = \frac{1}{f_{Y|X}\left(\xi_p(x)|x\right)}\left[p+c_n-\widehat{F}_n^{-1}(p|x)\right] + R_n(x)
$$

for any $c_n = O\left(\sqrt{\frac{\log n}{nh_n}}\right)$, *where* $R_n(x)$ *satisfies:* $\sup_{x\in\mathcal{J}}|R_n(x)| = O_p\left(\left((nh_n)^{-1}\log n\right)^{3/4}\right)$ *given uniform (w.r.t* $x\in\mathcal{J}$*) second order differentiability of* $F_{Y|X}(y|x)$ *at* $y=\xi_p(x)$.

Proof. Setting $a_n = \widehat{F}_n^{-1}(p+c_n|x) - F^{-1}(p|x)$, we have the following successive approximations,

$$
\begin{aligned}
& \widehat{F}_n\left(F^{-1}(p|x)+a_n|x\right) - \widehat{F}_n\left(F^{-1}(p|x)|x\right) \\
=\; & \tilde{F}_n\left(F^{-1}(p|x)+a_n|x\right) - \tilde{F}_n\left(F^{-1}(p|x)|x\right) + o_p\left(\frac{\log n}{nh_n}\right) \\
=\; & \left[F\left(F^{-1}(p|x)+a_n|x\right) - F\left(F^{-1}(p|x)|x\right)\right] + \Delta_n(x) + o_p\left(\frac{\log n}{nh_n}\right) \\
=\; & f_{Y|X}\left(\xi_p(x)|x\right)a_n + \Delta_n(x) + \Delta_n'(x) + o_p\left(\frac{\log n}{nh_n}\right) \qquad \text{(A.3)}
\end{aligned}
$$

where the first equality follows from Lemma A.5, the second from Lemma A.6, and

$$\begin{aligned}
\Delta_n(x) &= \tilde{F}_n\left(F^{-1}(p|x)+a_n|x\right)-\tilde{F}_n\left(F^{-1}(p|x)|x\right)-\left[F\left(F^{-1}(p|x)+a_n|x\right)-F\left(F^{-1}(p|x)|x\right)\right], \\
\Delta_n^{'}(x) &= \left[F\left(F^{-1}(p|x)+a_n|x\right)-F\left(F^{-1}(p|x)|x\right)\right]-f_{Y|X}\left(\xi_p(x)|x\right)a_n.
\end{aligned}$$

Thus we have $\sup_{x\in\mathcal{J}}|\Delta_n(x)| = O_p\left(\left(\frac{\log n}{nh}\right)^{3/4}\right)$ and $\sup_{x\in\mathcal{J}}|\Delta_n^{'}(x)| = O_p\left(\frac{\log n}{nh}\right)$ given uniform (w.r.t $x \in \mathcal{J}$) second order differentiability of $F_{Y|X}(y|x)$ when $y = \xi_p(x)$(without the second order differentiability $\sup_{x\in\mathcal{J}}|\Delta_n^{'}(x)| = o_p\left(\sqrt{\frac{\log n}{nh}}\right)$, which does not affect the asymptotic validity of our inference procedure whatsoever. This assumption is merely imposed in accordance with usual Bahadur Representation, see Theorem 2.5.1 in Serfling, 1980). Overall $O_p\left(\left(\frac{\log n}{nh}\right)^{3/4}\right)$ is the dominating term. The result follows from noting that the LHS expression in (A.3) becomes $\left[p+c_n-\widehat{F}_n\left(F^{-1}(p|x)|x\right)\right]$. **Q.E.D**

Proof of Theorem 2.3. The following string of equalities shall be self-explaining:

$$\begin{aligned}
&\Pr\left[F^{-1}(p|x) \le \widehat{F}_n^{-1}\left(p+c_{n\delta}(\alpha,K)\sigma_{np}(K)|x\right) \text{ for all } x\in\mathcal{J}\right] \\
=& \Pr\left[F^{-1}(p|x)-\widehat{F}_n^{-1}(p|x) \le \widehat{F}_n^{-1}\left(p+c_{n\delta}(\alpha,K)\sigma_{np}(K)|x\right)-\widehat{F}_n^{-1}(p|x) \text{ for all } x\in\mathcal{J}\right] \\
=& \Pr\left[F^{-1}(p|x)-\widehat{F}_n^{-1}(p|x) \le \frac{1}{f_{Y|X}\left(\xi_p(x)|x\right)}c_{n\delta}(\alpha,K)\sigma_{np}(K)+O_p\left(\left(\frac{\log n}{nh}\right)^{3/4}\right) \text{ for all } x\in\mathcal{J}\right] \\
=& \Pr\left[F^{-1}(p|x)-\tilde{F}_n^{-1}(p|x) \le \frac{1}{f_{Y|X}\left(\xi_p(x)|x\right)}c_{n\delta}(\alpha,K)\sigma_{np}(K)+o_p\left(\sqrt{\frac{\log n}{nh}}\right) \text{ for all } x\in\mathcal{J}\right] \\
=& \Pr\left[(2\delta\log n)^{1/2}\left[\sup_{x\in\mathcal{J}} f_{Y|X}\left(\xi_p(x)|x\right)\sigma_{np}^{-1}(K)\left(F^{-1}(p|x)-\tilde{F}_n^{-1}(p|x)\right)-d_n\right] \le c(\alpha)\right].
\end{aligned}$$

Similarly,

$$\begin{aligned}
&\Pr\left[F^{-1}(p|x) \ge \widehat{F}_n^{-1}\left(p-c_{n\delta}(\alpha,K)\sigma_{np}(K)|x\right) \text{ for all } x\in\mathcal{J}\right] \\
=& \Pr\left[(2\delta\log n)^{1/2}\left[\sup_{x\in\mathcal{J}} f_{Y|X}\left(\xi_p(x)|x\right)\sigma_{np}^{-1}(K)(\left(\tilde{F}_n^{-1}(p|x)-F^{-1}(p|x)\right)-d_n\right] \le c(\alpha)\right].
\end{aligned}$$

Hence the result follows from Lemma 2.4. **Q.E.D**

Proof of Lemma 2.4. The proof follows from Lemma A.3 and Lemma A.7 with $c_n = 0$. Therefore uniformly over $x \in \mathcal{J}$, we have

$$\begin{aligned}
&\widehat{F}_n^{-1}(p|x)-F^{-1}(p|x) \\
=& \frac{1}{f_{Y|X}\left(\xi_p(x)|x\right)}\left[p-\frac{1}{nh_n}\sum 1\{Y_i \le F^{-1}(p|x)\}K\left(\frac{F_X(x)-F_X(X_i)}{h_n}\right)\right]+o_p\left(\left(\frac{\log n}{nh_n}\right)^{1/2}\right).
\end{aligned}$$

Hence we could apply the strong approximation result in Hardle and Song (2010). For completeness we give sketch on the successive approximation steps in Appendix C. **Q.E.D**

Appendix B. Technical Proofs For Section 3

We first present a lemma used in the proof of Theorem 3.1.

Lemma B.1 *Under Assumptions (H), (K), and (PL), the following class of functions indexed by* $s = (\beta, y)$ *is P-Donsker, where* $\beta \in \mathcal{B} \subset \mathcal{R}^d$ *and* $y \in \mathcal{Y} \subset \mathcal{R}$*:*

$$\mathcal{F}_{n.s} = \left\{ f_{n,s} | f_{n,s} = 1\{Y_i - Z_i^{'}\beta \leq y\} \frac{1}{\sqrt{h_n}} K\left(\frac{F_X(x_0) - F_X(X_i)}{h_n}\right) : \beta \in \mathcal{B} \text{ and } y \in \mathcal{Y} \right\}.$$

Proof. We denote $f_{n,s} = f_s \frac{1}{\sqrt{h_n}} K\left(\frac{F_X(x_0) - F_X(X_i)}{h_n}\right)$, with $f_s = 1\{Y_i - Z_i^{'}\beta \leq y\}$. Define $\mathcal{F}_s = \left\{ f_s | f_s = 1\{Y_i - Z_i^{'}\beta \leq y\} : \beta \in \mathcal{B} \text{ and } y \in \mathcal{Y} \right\}$. Note that the family of sets $\mathcal{D}_s = \{(Y, Z) : Y - Z^{'}\beta \leq y, (\beta, y) \in \mathcal{R}^{d+1}\}$ is a VC class with VC dimension $V(\mathcal{D}_s)$ bounded up by $d + 2$ (see Chapter 2 in Pollard, 1984). Hence f_s belongs to VC subgraph class and by Theorem 2.6.7 in Van der Vaart and Wellner (1996), we could have the following bound for the entropy (w.r.t. $L_p(Q)$ norm) for any probability measure Q:

$$\log N_P(\epsilon, \mathcal{F}_s, Q) \leq M(V(\mathcal{D}_s), p) \log\left(\frac{1}{\epsilon}\right) \text{ for } p \in (0, \infty),$$

where M is a universal finite constant depending only on $V(\mathcal{D}_s)$ and p.

Now we are ready to use Theorem 2.11.22 in Van der Vaart and Wellner (1996) to prove that $\mathcal{F}_{n.s}$ is P-Donsker. We begin by verifying three conditions in (2.11.21).

(i) The envelope function is $F_n = \frac{1}{\sqrt{h_n}} K\left(\frac{F_X(x_0) - F_X(X_i)}{h_n}\right)$ satisfying $PF_n^2 = \int K^2(u)\, du < \infty$;

(ii) $PF_n^2 1\{F_n > \eta\sqrt{n}\} \leq \int_{K(u) > \eta\sqrt{nh_n}} K^2(u)\, du \to 0, \forall \eta > 0$, as $n \to \infty$;

(iii) Let $\rho(s, t)$ be the usual Euclidean norm in $\mathcal{R}^{d+1}$, further denote $s = (\beta, y)$ and $t = \left(\beta^{'}, y^{'}\right)$ and the conditional measure as $dQ_{\cdot|X}$. Then

$$\begin{aligned} P(f_{n,s} - f_{n,t})^2 &= \iint (f_s - f_t)^2 dQ_{\cdot|X} \frac{1}{h_n} K^2\left(\frac{F_X(x_0) - F_X(X)}{h_n}\right) dF_X \\ &\leq \iint M_d \left[\left(y - y^{'}\right)^2 + Z_1^2(\beta_1 - \beta_1^{'})^2 + \cdots + Z_d^2(\beta_d - \beta_d^{'})^2\right] dQ_{\cdot|X} \\ &\quad \frac{1}{h_n} K^2\left(\frac{F_X(x_0) - F_X(X)}{h_n}\right) dF_X \\ &\leq M \cdot R(K) \rho^2(s, t) \end{aligned}$$

where we have used the fact that $|f_s - f_t| \leq |y - y^{'}| + |Z_1(\beta_1 - \beta_1^{'})| + \cdots + |Z_d(\beta_d - \beta_d^{'})|$ and M_d is a finite constant depending on d. The last equality follows from assumption (PL) as Z has finite conditional (on X) second moment. Therefore $\sup_{\rho(s,t) < \delta_n} P(f_{n,s} - f_{n,t})^2 \to 0$ as $\delta_n \to 0$.

When it comes to the $L_2(Q)$ entropy, for any probability measure Q, we have

$$\log N_P(\epsilon ||K||_2, \mathcal{F}_{n,s}, Q) \leq \log N_P\left(\epsilon, \mathcal{F}_s, Q_{\cdot|X}\right) \leq M(V(\mathcal{D}_s), p) \log\left(\frac{1}{\epsilon}\right)$$

by the simple fact that $\int (f_{n,s} - f_{n,t})^2 \, dQ = \iint (f_s - f_t)^2 \, dQ_{\cdot|X} \frac{1}{h_n} K^2 \left(\frac{F_X(x_0) - F_X(X)}{h_n} \right) dF_X$.

In sum, the conditions in Theorem 2.11.22 in Van der Vaart and Wellner (1996) is satisfied for $\mathcal{F}_{n.s}$. **Q.E.D**

Proof of Theorem 3.1. Let

$$F_{n,\text{PL}}(y|x_0) = \frac{\sum_{i=1}^n 1\{Y_i - Z_i'\beta_0 \le y\} K\left(\frac{F_n(x_0) - F_n(X_i)}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{F_n(x_0) - F_n(X_i)}{h_n}\right)}.$$

We will complete the proof in two steps:

Step 1. We show that $\sqrt{nh_n}\left[\widehat{F}_{n,\text{PL}}(\cdot|x_0) - F_{(Y-Z'\beta_0)|X}(\cdot|x_0)\right]$ converges weakly to the same Gaussian process as $\sqrt{nh_n}\left[F_{n,\text{PL}}(\cdot|x_0) - F_{(Y-Z'\beta_0)|X}(\cdot|x_0)\right]$;

Step 2. We show that $(\widehat{F}_{n,\text{PL}}^{-1}\left(p - z_{\alpha/2}\sigma_{np}(K)|x_0\right) + z_0'\widehat{\beta}_0, \widehat{F}_{n,\text{PL}}^{-1}\left(p + z_{\alpha/2}\sigma_{np}(K)|x_0\right) + z_0'\widehat{\beta}_0]$ is an asymptotically valid confidence interval for $\left[z_0'\beta_0 + g(x_0)\right]$ with confidence level $1-\alpha$.

Proof of Step 1. As the denominator will converge to 1 in probability as in Stute (1986), it is sufficient to show that

$$\frac{1}{\sqrt{nh_n}} \sum_{i=1}^n \left[1\{Y_i - Z_i'\hat{\beta} \le y\} - 1\{Y_i - Z_i'\beta_0 \le y\}\right] K\left(\frac{F_n(x_0) - F_n(X_i)}{h_n}\right) = o_p(1). \quad \text{(B.1)}$$

Again taking the second order Taylor expansion, the left hand side of (B.1) becomes:

$$\begin{aligned}
&\frac{1}{\sqrt{nh_n}} \sum_{i=1}^n [1\{Y_i - Z_i'\hat{\beta} \le y\} - 1\{Y_i - Z_i'\beta_0 \le y\}] K\left(\frac{F_X(x_0) - F_X(X_i)}{h_n}\right) \\
&+ \frac{1}{\sqrt{nh_n}h_n} \sum_{i=1}^n [1\{Y_i - Z_i'\hat{\beta} \le y\} - 1\{Y_i - Z_i'\beta_0 \le y\}] K'\left(\frac{F_X(x_0) - F_X(X_i)}{h_n}\right) B_n(x_0, X_i) \\
&+ \frac{1}{\sqrt{nh_n}h_n^2} \sum_{i=1}^n [1\{Y_i - Z_i'\hat{\beta} \le y\} - 1\{Y_i - Z_i'\beta_0 \le y\}] K''(\Delta) B_n^2(x_0, X_i) \\
&= A_{n1} + A_{n2} + A_{n3}. \qquad \text{(B.2)}
\end{aligned}$$

Again A_{n3} is the easiest term to handle. It follows from the argument in Lemma 1 in Stute (1984b) that as the kernel is of bounded support, we can first restrict on considering those X_i's s.t. $|F_n(x_0) - F_n(X_i)| \le Mh_n$, for some constant M. By DKW bound, this implies that $|F_X(x_0) - F_X(X_i)| \le Mh_n$ almost surely. Hence we could apply Lemma B.1 to bound the local oscillation of the empirical processes.

The third term in (B.2) could be bounded up by

$$|A_{n3}| \le \left(\frac{n}{h_n} \sup_{|F_X(x_0) - F_X(X_i)| \le Mh_n} B_n^2(x_0, X_i)\right) \frac{2|K''(\Delta)|}{\sqrt{n}h_n^{3/2}} = o_p(1).$$

It converges to zero in probability following our assumption on the bandwidth and boundedness of the second order derivatives of the kernel function.

Similar as the proof of Lemma A.3, we write the rescaled A_{n2} as a U-statistics plus a smaller order term,

$$\frac{1}{\sqrt{nh_n}} A_{n2} = \frac{1}{h_n^2} U_2^n \left[h_{(\hat{\beta},y)} - h_{(\beta,y)}\right] + O_p\left(\frac{1}{nh_n^2}\right)$$

with symmetric kernel function

$$
\begin{aligned}
& h_{(\beta,y)}(\cdot,\cdot) \\
= \quad & \frac{1}{2}\left(\begin{array}{c} \left[1\{Y_i - Z_i'\hat{\beta} \le y\} - 1\{Y_i - Z_i'\beta_0 \le y\}\right] K'\left(\frac{F_X(x)-F_X(X_i)}{h_n}\right)\left[\begin{array}{c} 1\{X_j \le x_0\} - F_X(x_0) \\ -1\{X_j \le X_i\} + F_X(X_i) \end{array}\right] \\ + \left[1\{Y_j - Z_j'\hat{\beta} \le y\} - 1\{Y_j - Z_j'\beta_0 \le y\}\right] K'\left(\frac{F_X(x)-F_X(X_j)}{h_n}\right)\left[\begin{array}{c} 1\{X_i \le x_0\} - F_X(x_0) \\ -1\{X_i \le X_j\} + F_X(X_j) \end{array}\right] \end{array}\right).
\end{aligned}
$$

Similar as Lemma 3.1 in Ghosal, Sen, and van der Vaart (2000), we first approximate the U-process $U_2^n h_{(\beta,y)}$ by its projection uniformly for (β, y). We only need to consider one part in the summation of $h_{(\beta,y)}(\cdot,\cdot)$. Define the following classes ($\mathcal{F}_2$ only serves as a scaling factor when we fix $x = x_0$, its entropy would come into play when we consider uniformity issue when x varies across $\mathcal{J}$),

$$
\begin{aligned}
\mathcal{F}_1 &= \{1\{Y_i - Z_i'\beta_0 \le y\} : (\beta, y) \in \mathcal{R}^{d+1}\}, \\
\mathcal{F}_2 &= \{K'\left(\frac{F_X(x) - F_X(X_i)}{h_n}\right), x \in \mathcal{J}\}, \\
\mathcal{F}_3 &= \{[1\{X_i \le X_j\} - F_X(X_j)]\}.
\end{aligned}
$$

By Lemma A.1 in Ghosal, Sen, and van der Vaart (2000) and the entropy bound in our Lemma B.1 for $\mathcal{F}_1$, we have: $\sup_Q \log N\left(\epsilon||K'||_2, \mathcal{F}_1\mathcal{F}_2\mathcal{F}_3, L_2(Q)\right) \le M \log\left(\frac{1}{\epsilon}\right)$. Hence it follows that

$$
nE\left[\sup_{(\beta,y)} |U_2^n h_{(\beta,y)} - 2U_1^n \Pi_1 h_{(\beta,y)}|\right] \le M \int_0^1 \log\left(\frac{1}{\epsilon}\right) d\epsilon = O(1).
$$

Therefore $\sup_{(\beta,y)} |U_2^n h_{(\beta,y)} - 2U_1^n \Pi_1 h_{(\beta,y)}| = O_p\left(\frac{1}{n}\right)$ and $\sup_{(\beta,y)} \frac{1}{h_n^2}|U_2^n h_{(\beta,y)} - 2U_1^n \Pi_1 h_{(\beta,y)}| = O_p\left(\frac{1}{nh_n^2}\right) = o_p(1)$ under Assumption (H).

Now we work with the projection term:

$$
\begin{aligned}
& \frac{1}{h_n^2} 2U_1^n \left[\Pi_1 h_{(\hat{\beta},y)} - \Pi_1 h_{(\beta,y)}\right] \\
= \quad & \frac{1}{nh_n^2}\sum_{j=1}^n \int E\left[[1\{Y_i - Z_i'\hat{\beta} \le y\} - 1\{Y_i - Z_i'\beta_0 \le y\}]|X_i\right] K'\left(\frac{F_X(x_0) - F_X(X_i)}{h_n}\right) \times \\
& [1\{F_X(X_j) \le F_X(x_0)\} - F_X(x_0) - 1\{F_X(X_j) \le F_X(X_i)\} + F_X(X_i)]\, dF_X(X_i) \\
= \quad & \frac{1}{nh_n}\sum_{j=1}^n \int E\left[[1\{Y_i - Z_i'\hat{\beta} \le y\} - 1\{Y_i - Z_i'\beta_0 \le y\}]|X_i\right] \times \\
& [1\{U_j \le u_0\} - u_0 - 1\{U_j \le U_i\} + U_i]\, dK\left(\frac{u_0 - U_i}{h_n}\right) \\
= \quad & \frac{1}{nh_n}\sum_{j=1}^n \int E\left[[1\{Y_i - Z_i'\hat{\beta} \le y\} - 1\{Y_i - Z_i'\beta_0 \le y\}]|X_i\right] \times \\
& [1\{U_j \le u_0\} - u_0 - 1\{U_j \le u_0 - vh_n\} + u_0 - vh_n]\, dK(v) \\
= \quad & O_p\left(\frac{1}{\sqrt{n}}\right) o_p\left(\frac{1}{\sqrt{nh_n}}\right).
\end{aligned}
$$

When it comes to A_{n1}, we make use of Lemma B.2 which states that the class of functions $\mathcal{F}_n$ below is Donsker:

$$\mathcal{F}_{n,s} = \left\{ f_{n,s} | f_{n,s} = 1\{Y_i - Z_i^{'}\beta \le y\} \frac{1}{\sqrt{h_n}} K\left(\frac{F_X(x_0) - F_X(X_i)}{h_n} \right) : \beta \in \mathcal{B} \text{ and } y \in \mathcal{Y} \right\}.$$

Let

$$\begin{aligned} \hat{f}_n &= 1\{Y_i - Z_i^{'}\hat{\beta} \le y\} \frac{1}{\sqrt{h_n}} K\left(\frac{F_X(x_0) - F_X(X_i)}{h_n} \right) \text{ and} \\ f_0 &= 1\{Y_i - Z_i^{'}\beta_0 \le y\} \frac{1}{\sqrt{h_n}} K\left(\frac{F_X(x_0) - F_X(X_i)}{h_n} \right). \end{aligned}$$

Because $\hat{\beta} - \beta_0 = o_p(1)$ and $E\left[(\hat{f}_n - f_0)^2 \right] \to 0$, by Lemma 19.24 in van der Vaart (1998), we have $\mathcal{G}_n(\hat{f}_n - f_0) \to_p 0$, where $\mathcal{G}_n$ denotes the empirical process operator.

$$\begin{aligned} & A_{n1} = \mathcal{G}_n(\hat{f}_n - f_0) + \\ & \sqrt{n} E\left[\left(1\{Y_i - Z_i^{'}\hat{\beta} \le y\} - 1\{Y_i - Z_i^{'}\beta_0 \le y\} \right) \frac{1}{\sqrt{h_n}} K\left(\frac{F_X(x_0) - F_X(X_i)}{h_n} \right) \right] \\ = \; & E\left\{ E\left[\sqrt{n} [1\{Y_i - Z_i^{'}\hat{\beta} \le y\} - 1\{Y_i - Z_i^{'}\beta_0 \le y\}] | X_i \right] \frac{1}{\sqrt{h_n}} K\left(\frac{F_X(x_0) - F_X(X_i)}{h_n} \right) \right\} + o_P(1) \\ = \; & O_p(1) E\left[\frac{1}{\sqrt{h_n}} K\left(\frac{F_X(x_0) - F_X(X_i)}{h_n} \right) \right] + o_P(1) = o_P(1), \end{aligned}$$

where the third equality follows from $E\left[\sqrt{n} [1\{Y_i - Z_i^{'}\hat{\beta} \le y\} - 1\{Y_i - Z_i^{'}\beta_0 \le y\}] | X_i \right] = O_p(1)$ as $\hat{\beta} - \beta_0 = O_p\left(\frac{1}{\sqrt{n}} \right)$ and we assume $E\left[1\{Y_i - Z_i^{'}\beta \le y | Z_i\} \right]$ has bounded derivative w.r.t β and the last equality follows from the fact that $E\left[\frac{1}{\sqrt{h_n}} K\left(\frac{F_X(x_0) - F_X(X_i)}{h_n} \right) \right] = O(\sqrt{h_n})$.

Proof of Step 2. It follows from the same proof as that of Theorem 2.1 that

$$\Pr\left(g(x_0) \in (\widehat{F}_{n,\text{PL}}^{-1}\left(p - z_{\alpha/2}\sigma_{np}(K) | x_0 \right), \widehat{F}_{n,\text{PL}}^{-1}\left(p + z_{\alpha/2}\sigma_{np}(K) | x_0 \right)] \right) \to 1 - \alpha.$$

So

$$\begin{aligned} & \Pr\left(z_0^{'}\beta_0 + g(x_0) \in (z_0^{'}\beta_0 + \widehat{F}_{n,\text{PL}}^{-1}\left(p - z_{\alpha/2}\sigma_{np}(K) | x_0 \right), z_0^{'}\beta_0 + \widehat{F}_{n,\text{PL}}^{-1}\left(p - z_{\alpha/2}\sigma_{np}(K) | x_0 \right) \right) \\ = \; & \Pr\left(z_0^{'}\beta_0 + g(x_0) \in (z_0^{'}\hat{\beta} + \widehat{F}_{n,\text{PL}}^{-1}\left(p - z_{\alpha/2}\sigma_{np}(K) | x_0 \right), z_0^{'}\beta + \widehat{F}_{n,\text{PL}}^{-1}\left(p - z_{\alpha/2}\sigma_{np}(K) | x_0 \right) \right) + o(1) \\ \to \; & 1 - \alpha, \end{aligned}$$

where in the first equality above, we have replaced β_0 with its root-n consistent estimator. This is valid since the length of the interval is of order $(nh_n)^{-1/2}$, wider than $n^{-1/2}$. **Q.E.D**

The following lemma is used in the proof of Theorem 3.3.

Lemma B.2 *(Stute and Zhu, 2005) Referring to the notation in Section 3.2, given Assumptions (Z2), (HS) and a root-n consistent estimator $\hat{\beta}$ in the single index model, uniformly for any z,*

$$\sup_{n^{1/2}||\hat{\beta} - \beta_0|| \le M, n^{1/2-1/\gamma}|Z_i^{'}\hat{\beta} - Z_i^{'}\beta_0| \le M} |F_n(z^{'}\hat{\beta}) - F_n(Z_i^{'}\hat{\beta}) - F(z^{'}\beta_0) + F(Z_i^{'}\beta_0)| = O_p\left(n^{-3/4+1/2\gamma}\sqrt{\ln n} \right).$$

Proof. We refer the readers to Lemma 4.2 and its proof in Stute and Zhu (2005). **Q.E.D**

Proof of Theorem 3.3. We will prove the result focusing on the estimator without the denominator, as it would follow along the proof that the denominator converges to 1 in probability.

First we claim that

$$\frac{1}{\sqrt{nh_n}}\left[\sum\nolimits_{i=1}^{n}1\{Y_i\leq y\}\left(K\left(\frac{F_n(z_0'\hat{\beta})-F_n(Z_i'\hat{\beta})}{h_n}\right)-K\left(\frac{F(z_0'\beta_0)-F(Z_i'\beta_0)}{h_n}\right)\right)\right]=o_p\left(1\right). \tag{B.3}$$

Given (B.3), after normalizing, our conditional empirical process converges to the same Brownian Bridge as $\frac{1}{\sqrt{nh_n}}\sum_{i=1}^{n}\left[1\{Y_i\leq y\}K\left(\frac{F(z_0'\beta_0)-F(Z_i'\beta_0)}{h_n}\right)-F_{Y|\tilde{X}}(\cdot|\tilde{x}_0)\right]$ does, where $\tilde{x}_0=z_0'\beta_0$. Going over the proofs of Theorems 2.1 and 3.1, we have $(\widehat{F}_{n,\mathrm{SI}}^{-1}\left(p-z_{\alpha/2}\sigma_{np}\left(K\right)|z_0\right),\widehat{F}_{n,\mathrm{SI}}^{-1}\left(p+z_{\alpha/2}\sigma_{np}\left(K\right)|z_0\right)]$ as the confidence interval for $g(\tilde{x}_0)$ with asymptotic nominal size $1-\alpha$.

Now we show the claim in (B.3). Taking a second order Taylor expansion, we obtain:

$$\begin{aligned}
&\frac{1}{\sqrt{nh_n}}\left[\sum\nolimits_{i=1}^{n}K\left(\frac{F_n(z_0'\hat{\beta})-F_n(Z_i'\hat{\beta})}{h_n}\right)-K\left(\frac{F(z_0'\beta_0)-F(Z_i'\beta_0)}{h_n}\right)\right]\\
=\;&\frac{1}{\sqrt{nh_n}}\sum\nolimits_{i=1}^{n}\frac{1}{h_n}K'\left(\frac{F(z_0'\beta_0)-F(Z_i'\beta_0)}{h_n}\right)\left[F_n(z_0'\hat{\beta})-F_n(Z_i'\hat{\beta})-F(z_0'\beta_0)+F(Z_i'\beta_0)\right]\\
&+\frac{1}{\sqrt{nh_n}}\sum\nolimits_{i=1}^{n}\frac{1}{h_n^2}K''\left(\Delta\right)\left[F_n(z_0'\hat{\beta})-F_n(Z_i'\hat{\beta})-F(z_0'\beta_0)+F(Z_i'\beta_0)\right]^2
\end{aligned}$$

Recall Lemma B.3, under our assumptions we have,

$$\sup_{n^{1/2}||\hat{\beta}-\beta_0||\leq M,\,n^{1/2-1/\gamma}|Z_i'\hat{\beta}-Z_i'\beta_0|\leq M}|F_n(z_0'\hat{\beta})-F_n(Z_i'\hat{\beta})-F(z_0'\beta_0)+F(Z_i'\beta_0)|=O_p\left(n^{-3/4+1/2\gamma}\sqrt{\ln n}\right).$$

Hence,

$$\begin{aligned}
&\frac{1}{\sqrt{nh_n}}\sum\nolimits_{i=1}^{n}\left|\frac{1}{h_n}K'\left(\frac{F(z_0'\beta_0)-F(Z_i'\beta_0)}{h_n}\right)\right|\cdot|F_n(z_0'\hat{\beta})-F_n(Z_i'\hat{\beta})-F(z_0'\beta_0)+F(Z_i'\beta_0)|\\
=\;&\frac{n^{-1/4+1/2\gamma}\sqrt{\ln n}}{\sqrt{h_n}}\left[\frac{1}{nh_n}\sum\nolimits_{i=1}^{n}|K'\left(\frac{F(z_0'\beta_0)-F(Z_i'\beta_0)}{h_n}\right)|\right]=h_n^{-1/2}n^{-1/4+1/2\gamma}\sqrt{\ln n}O_p\left(1\right)\\
=\;&o_p\left(1\right),
\end{aligned}$$

where the second equality use the fact that $\frac{1}{nh_n}\sum_{i=1}^{n}\left|K'\left(\frac{F(z_0'\beta_0)-F(Z_i'\beta_0)}{h_n}\right)\right|=O_p\left(1\right)$ by standard kernel convergence result.

The last equality follows from the assumption of the bandwidth.

$$\begin{aligned}
&\frac{1}{\sqrt{nh_n}}\sum\nolimits_{i=1}^{n}\frac{1}{h_n^2}K''\left(\Delta\right)\left[F_n(z_0'\hat{\beta})-F_n(Z_i'\hat{\beta})-F(z_0'\beta_0)+F(Z_i'\beta_0)\right]^2\\
=\;&\left(\frac{1}{\sqrt{n}h_n^{5/2}}n^{-3/2+1/\gamma}\ln n\right)O_p\left(1\right)=h_n^{-5/2}n^{-1+1/\gamma}\ln n=o_p\left(1\right).
\end{aligned}$$

Hence the claim is indeed satisfied.

In the above proof, take $y = \infty$, we also get the desired convergence (to 1 in probability) for the denominator. **Q.E.D**

Proofs of Theorems 3.2 and 3.4. We only give a sketch of the changes needed here. It suffices to show that $\widehat{F}_{n,\text{PL}}(y|x)$ and $\widehat{F}_{n,\text{SI}}(y|z)$ can be uniformly approximated well by the corresponding $\widetilde{F}_{n,\text{PL}}(y|x)$ and $\widetilde{F}_{n,\text{SI}}(y|z)$. Then the results would follow after going over Lemmas A.2-A.7**.** Notice that the sup-norm convergence rate is in fact slower by a factor of $\sqrt{\log n}$, which corresponds to the compensating factor along these uniform approximations.

Similar to $\widehat{f}_U(x)$ and $\widetilde{f}_U(x)$ defined earlier, we introduce the following notations:

$$\begin{aligned}
\widehat{f}_{U,\text{SI}}(z) &= \frac{1}{nh_n}\sum\nolimits_{i=1}^{n} K\left(\frac{F_n(z'\hat{\beta}) - F_n(Z_i'\hat{\beta})}{h_n}\right) \text{ and} \\
\widetilde{f}_{U,\text{SI}}(z) &= \frac{1}{nh_n}\sum\nolimits_{i=1}^{n} K\left(\frac{F(z'\beta_0) - F(Z_i'\beta_0)}{h_n}\right).
\end{aligned}$$

For Theorem 3.4, we have:

$$\begin{aligned}
&\widehat{F}_{n,\text{SI}}(y|z) - \widetilde{F}_{n,\text{SI}}(y|z) \\
= &\frac{\frac{1}{nh_n}\sum_{i=1}^{n}\left[1\{Y_i \le y\} - F_{Y|\tilde{X}}(y|\tilde{x})\right]\left[K\left(\frac{F_n(z'\hat{\beta})-F_n(Z_i'\hat{\beta})}{h_n}\right) - K\left(\frac{F(z'\beta_0)-F(Z_i'\beta_0)}{h_n}\right)\right]}{\widehat{f}_{U,\text{SI}}(z)} + \\
&\frac{1}{nh_n}\sum\nolimits_{i=1}^{n}\left[1\{Y_i \le y\} - F_{Y|\tilde{X}}(y|\tilde{x})\right] K\left(\frac{F(z'\beta_0) - F(Z_i'\beta_0)}{h_n}\right)\left(\frac{1}{\widehat{f}_{U,\text{SI}}(z)} - \frac{1}{\widetilde{f}_{U,\text{SI}}(z)}\right)
\end{aligned}$$

with $\tilde{x} = z'\beta_0$, The proof about switching from $F_n(z'\hat{\beta})$ to $F(z'\beta_0)$ follows directly, since when we characterize the two smaller terms, the bound in Lemma B.2 holds uniformly in $\tilde{x}$. Also the denominator converges to 1 with a rate $O_p\left(\sqrt{\frac{\log n}{nh_n}}\right)$. The rest would be the same.

For Theorem 3.2, we have:

$$
\begin{aligned}
& \widehat{F}_{n,\mathrm{PL}}(y|x) - \widetilde{F}_{n,\mathrm{PL}}(y|x) \\
= & \frac{\frac{1}{nh_n}\sum_{i=1}^{n}\left[1\{Y_i - Z_i'\hat{\beta} \le y\} - F_{\tilde{Y}|X}(\tilde{y}|x)\right]\left[K\left(\frac{F_n(x)-F_n(X_i)}{h_n}\right) - K\left(\frac{F(x)-F(X_i)}{h_n}\right)\right]}{\widehat{f}_U(x)} + \\
& \frac{1}{nh_n}\sum_{i=1}^{n}\left[1\{Y_i - Z_i'\hat{\beta} \le y\} - F_{\tilde{Y}|X}(\tilde{y}|x)\right] K\left(\frac{F(x)-F(X_i)}{h_n}\right)\left(\frac{1}{\widehat{f}_U(x)} - \frac{1}{\widetilde{f}_U(x)}\right) \\
= & \frac{\frac{1}{nh_n}\sum_{i=1}^{n}\left[1\{Y_i - Z_i'\hat{\beta} \le y\} - 1\{Y_i - Z_i'\beta_0 \le y\}\right]\left[K\left(\frac{F_n(x)-F_n(X_i)}{h_n}\right) - K\left(\frac{F(x)-F(X_i)}{h_n}\right)\right]}{\widehat{f}_U(x)} \\
& + \frac{\frac{1}{nh_n}\sum_{i=1}^{n}\left[1\{Y_i - Z_i'\beta_0 \le y\} - F_{\tilde{Y}|X}(\tilde{y}|x)\right]\left[K\left(\frac{F_n(x)-F_n(X_i)}{h_n}\right) - K\left(\frac{F(x)-F(X_i)}{h_n}\right)\right]}{\widehat{f}_U(x)} \\
& + \frac{1}{nh_n}\sum_{i=1}^{n}\left[1\{Y_i - Z_i'\hat{\beta} \le y\} - 1\{Y_i - Z_i'\beta_0 \le y\}\right] K\left(\frac{F(x)-F(X_i)}{h_n}\right)\left(\frac{1}{\widehat{f}_U(x)} - \frac{1}{\widetilde{f}_U(x)}\right) \\
& + \frac{1}{nh_n}\sum_{i=1}^{n}\left[1\{Y_i - Z_i'\beta_0 \le y\} - F_{\tilde{Y}|X}(\tilde{y}|x)\right] K\left(\frac{F(x)-F(X_i)}{h_n}\right)\left(\frac{1}{\widehat{f}_U(x)} - \frac{1}{\widetilde{f}_U(x)}\right) \\
= & P_{n1} + P_{n2} + P_{n3} + P_{n4}
\end{aligned}
$$

where $\tilde{y} = y - z^{'}\beta_0$. The terms P_{n2} and P_{n4} could be dealt with just as in the univariate nonparametric case.

When it comes to P_{n1} and P_{n3}, the only change occurs at the first order Taylor expansion term where the approximation of the U-statistics by the Hajek projection holds uniformly in x, i.e., we need to incorporate the class $\mathcal{F}_2 = \{K^{'}\left(\frac{F_X(x)-F_X(X_i)}{h_n}\right), x \in \mathcal{J}\}$ indexed by x now. As $K^{''}$ exists and is bounded, hence $K^{'}$ has bounded variation, the conclusion in Ghosal, Sen, and van der Vaart (2000) still holds. For the $\mathcal{F}_{n.s}$ introduced in Lemma B.1, the additional index x could be handled as in Einmahl and Mason (2005) once we incorporate the additional factor $\sqrt{\log n}$. **Q.E.D**

8 Appendix C. Strong Approximation Results

The strong approximation used in this paper follows from Hardle and Song (2010) upon changing X to $F_X(X)$ and removing the X's density $f_X(\cdot)$. For completeness we sketch the successive approximation steps according to our notation, and refer the readers to Hardle and Song (2010) for a detailed proof.

Recall our conditional quantile estimator admits the following linear representation uniformly over $x \in \mathcal{J}$, after replacing $F_n(\cdot)$ with $F_X(\cdot)$ inside the kernel function and applying Bahadur representation:

$$\begin{aligned}
&\widehat{F}_n^{-1}(p|x) - F^{-1}(p|x) \\
= &\frac{1}{f_{Y|X}(\xi_p(x)|x)} \left[\begin{array}{c} \frac{1}{nh_n}\sum_{i=1}^n \left[p - 1\{Y_i \le \xi_p(x)\}\right] K\left(\frac{F_X(x)-F_X(X_i)}{h_n}\right) - \\ E\left[p - 1\{Y_i \le \xi_p(x)\}\right] \frac{1}{h_n} K\left(\frac{F_X(x)-F_X(X_i)}{h_n}\right) \end{array} \right] + o_p\left(\sqrt{\frac{\log n}{nh_n}}\right).
\end{aligned}$$

Define the dominating linear term times $\sqrt{\frac{nh_n}{p(1-p)}}$ as $Y_n(u)$, with $u = F_X(x)$. Also let $T(v,y) = \left[F_{U|y}(v|y), F_Y(y)\right]$ be the Rosenblatt transformation and $\psi(s) = p - 1\{s \le 0\}$. Now we have the following successive approximating processes:

$$\begin{aligned}
Y_{0,n}(u) &= \frac{1}{\sqrt{h_n g(u)}} \iint_{\Gamma_n} K\left(\frac{u-v}{h_n}\right) \psi\left(y - \xi_p\left(F_X^{-1}(u)\right)\right) dZ_n(v,y), \\
Y_{1,n}(u) &= \frac{1}{\sqrt{h_n g(u)}} \iint_{\Gamma_n} K\left(\frac{u-v}{h_n}\right) \psi\left(y - \xi_p\left(F_X^{-1}(u)\right)\right) dB_n[T(v,y)], \\
Y_{2,n}(u) &= \frac{1}{\sqrt{h_n g(u)}} \iint_{\Gamma_n} K\left(\frac{u-v}{h_n}\right) \psi\left(y - \xi_p\left(F_X^{-1}(u)\right)\right) dW_n[T(v,y)], \\
Y_{3,n}(u) &= \frac{1}{\sqrt{h_n g(u)}} \iint_{\Gamma_n} K\left(\frac{u-v}{h_n}\right) \psi\left(y - \xi_p\left(F_X^{-1}(v)\right)\right) dW_n[T(v,y)], \\
Y_{4,n}(u) &= \frac{\sqrt{p(1-p)}}{\sqrt{h_n g(u)}} \int K\left(\frac{u-v}{h_n}\right) dW(v), \\
Y_{5,n}(u) &= \frac{1}{\sqrt{h_n}} \int K\left(\frac{u-v}{h_n}\right) dW(v),
\end{aligned}$$

where $\Gamma_n = \{|y| \le a_n\}$ and $g(u) = E\left[\psi\left(y - \xi_p\left(F_X^{-1}(u)\right)\right) \times 1\{|y| \le a_n\} | U = u\right]$. $Z_n(\cdot,\cdot)$ denotes bivariate empirical processes, $\{B_n\}$ being a sequence of Brownian bridges, $\{W_n\}$ being a sequence of Wiener processes and $W(\cdot)$ being the Wiener process.

The proof goes by approximating the linear term by $Y_{0,n}(u)$ up to $Y_{3,n}(u)$, confirming $Y_{3,n}(u)$ and $Y_{4,n}(u)$ having the same distribution, and finally approximating $Y_{4,n}(u)$ by $Y_{5,n}(u)$. The limiting distribution and normalizing and centering sequences are from Bickel and Rosenblatt (1973). Claeskens and Van Keilegom (2003) obtain similar strong approximation result in local likelihood models without truncating the range of Y, however their results rely on stronger differentiability assumption on the sample objective function which is not directly applicable in our context.

References

[1] Bhattacharya, P.K. and A.K. Gangopadhyay (1990), "Kernel and Nearest-Neighbor estimation of a conditional quantile," *The Annals of Statistics* 18, 1400-1415.

[2] Bickel, P. and M. Rosenblatt (1973), "On some global measures of the deviation of density function estimators," *The Annals of Statistics* 1, 1071-1095.

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

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

[5] Buchinsky, M. and J. Hahn (1998), "An alternative estimator for the censored quantile regression model," *Econometrica* 66, 653-671.

[6] Cai, Z. (2002), "Regression quantiles for time series data," *Econometric Theory* 18, 169-192.

[7] Chen, S. and S. Khan (2001), "Semiparametric efficient estimation of a partially linear censored regression model," *Econometric Theory* 17, 567-590.

[8] Chen, X. and D. Pouzo (2009), "Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals," *Journal of Econometrics* 152, 46-60.

[9] Chernozhukov, V., Fernandez-Val, I., and A. Galichon (2010), "Quantile and probability curves without crossing," *Econometrica* 78, 1093-1125.

[10] Chernozhukov, V., I. Fernandez-Val, and A. Kowalski (2011), "Quantile regression with censoring and endogeneity," working paper.

[11] Chernozhukov, V. and C. Hansen (2005), "An IV model of quantile treatment effects," *Econometrica* 73, 245-261.

[12] Chernozhukov, V., C. Hansen, and M. Jansson (2009), "Finite sample inference for quantile regression models," *Journal of Econometrics* 152, 93-103.

[13] Chernozhukov, V. and H. Hong (2002), "3-step censored quantile regression and extramarital affairs," *Journal of the American Statistical Association* 97, 872-882.

[14] Chesher, A. (2003), "Identification in nonseparable models," *Econometrica* 71, 1405-1441.

[15] Claeskens, G. and I.V. Keilegom (2003), "Bootstrap confidence bands for regression curves and their derivatives," *The Annals of Statistics* 31, 1852-1884.

[16] Chaudhuri, P. (1991), "Nonparametric estimates of regression quantiles and their local Bahadur representation," *The Annals of Statistics* 19, 760-777.

[17] Chaudhuri, P., K. Doksum, and A. Samarov (1997), "On average derivative quantile regression," *The Annals of Statistics* 25, 715-744.

[18] Csorgo, M. and P. Revesez (1984), "Two approaches to constructing simultaneous confidence bounds for quantiles," *Probability and Mathematical Statistics* 4, 221-236.

[19] Dabrowska, D.M. (1987), "Non-parametric regression with censored survival time data," *Scandiavian Journal of Statistics* 14, 181-197.

[20] De Angelis, D., P. Hall, and G.A. Young (1993), "Analytical and bootstrap approximations to estimator distributions in L1 regressions," *Journal of the American Statistical Association* 88, 1310-1316.

[21] Dette, H. and S. Volgushev (2008), "Non-crossing non-parametric estimates of quantile curves," *Journal of the Royal Statistical Society*, Series B 70, 609-627.

[22] Echenique, F. and I. Komunjer (2009), "Testing models with multiple equilibria by quantile methods," *Econometrica* 77, 1281-1297.

[23] Einmahl, U. and D.M. Mason (2005), "Uniform in bandwidth consistency of kernel-type function estimators," *The Annals of Statistics* 33, 1380-1403.

[24] Fan, J., T.-C. Hu, and Y.K. Truong (1994), "Robust non-parametric function estimation," *Scandianvian Journal of Statistics* 21, 433-446.

[25] Fan, Y. and R. Liu (2011), "Symmetrized multivariate k-NN estimators," forthcoming in *Econometric Reviews.*

[26] Gutenbrunner, C. and J. Jureckova (1992), "Regression quantile and regression rank score process in the linear model and derived statistics," *The Annals of Statistics* 20, 305-330.

[27] Gutenbrunner, C., J. Jureckova, R. Koenker, and S. Portnoy (1993), "Test of Linear Hypotheses Based on Regression Rank Scores," *Journal of Nonparametric Statistics* 2, 307-331

[28] Ghosal, S., A. Sen and A. W. van der Vaart (2000), "Testing monotonicity of regression," *The Annals of Statistics* 28, 1054-1082.

[29] Goh, S.C. and K. Knight (2009), "Nonstandard quantile-regression inference," *Econometric Theory* 25, 1415-1432.

[30] Guerre, E. and C. Sabbah (2012), "Uniform bias study and Bahadur representation for local polynomial estimators of the conditional quantile function," *Econometric Theory* 28, 87-129.

[31] Hardle W.K. and S. Song (2010), "Confidence bands in quantile regression," *Econometric Theory* 26, 1180-1200.

[32] Holderlein, S. and E. Mammen (2007), "Identification of marginal effects in nonseparable models without monotonicity," *Econometrica* 75, 1513-1518.

[33] Honoré, B., S. Khan, and J. Powell (2002), "Quantile regression under random censoring," *Journal of Econometrics* 64, 241-278.

[34] He, X. (1997), "Quantile curves without crossing," *The American Satistician* 51, 186-191.

[35] He, X. and F. Hu (2002), "Markov chain marginal bootstrap," *Journal of American Statistical Association* 97, 783-795.

[36] He, X. and P. Ng (1999), "Quantile splines with several covariates," *Journal of Statistical Planning and Inference* 75, 343-352.

[37] He, X., P. Ng, and S. Portnoy (1998), "Bivariate quantile smoothing splines," *Journal of the Royal Statistical Society*, Series B 60, 537-550.

[38] Horowitz, J. (1998), "Bootstrap methods for median regression models," *Econometrica* 66, 1327-1352.

[39] Ichimura, H. and S. Lee (2010), "Characterization of the asymptotic distribution of semiparametric M-estimators," *Journal of Econometrics* 159, 252-266.

[40] Khan, S. (2001), "Two stage rank Estimation of quantile index models," *Journal of Econometrics* 100, 319-355.

[41] Khan, S. and J.L. Powell (2001), "Two-step estimation of semiparametric censored regression models," *Journal of Econometrics* 103, 73-110.

[42] Kocherginsky, M., X. He, and Y. Mu (2005), "Practical confidence intervals for regression quantiles," *Journal of Computational and Graphical Statistics* 14, 41-55.

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

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

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

[46] Koenker, R. and Z. Xiao (2004), "Unit root quantile autoregression inference," *Journal of American Statistical Association* 99, 775-787.

[47] Kong, E. and Y. Xia (2012), "A single-index quantile regression model and its estimation," forthcoming in *Econometric Theory.*

[48] Korostelev, A. and O. Korosteleva (2011), *Mathematical Statistics: Asymptotic Minimax Theory,* American Mathematical Society.

[49] Lee, S. (2003), "Efficient semiparametric estimation of a partially linear quantile regression model," *Econometric Theory* 19, 1-31.

[50] Li, Q. and J.S. Racine (2008), "Nonparametric Estimation of Conditional CDF and Quantile Functions with Mixed Categorical and Continuous Data," *Journal of Business and Economic Statistics* 26, 423-434.

[51] Mack, Y.P. (1981), "Local properties of k-NN regression estimates," *SIAM Journal on Algebraic and Discrete Methods* 2, 211-232.

[52] Ostu, T. (2008), "Conditional empirical likelihood estimation and inference for quantile regression models," *Journal of Econometrics* 142, 508–538.

[53] Pollard, D. (1984), *Convergence of Stochastic Processes,* Springer, New York.

[54] Polonik, W. and Q. Yao (2002), "Asymptotics of set-indexed conditional empirical processes based on dependent data," *Journal of Multivariate Analysis* 80, 234-255.

[55] Portnoy, S. (2012), "Nearly root-n approximation of regression quantile processes," *The Annals of Statistics* 40, 1714-1736.

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

[57] Serfling, R.J. (1980), *Approximation Theorems of Mathematical Statistics,* Wiley, New York.

[58] Song, S., Y. Ritov, and W.K. Hardle (2012), "Partial linear quantile regression and bootstrap confidence bands," *Journal of Multivariate Analysis* 107, 244-262.

[59] Stute, W. (1982), "The oscillation behavior of empirical processes," *The Annals of Probability* 10, 86-107.

[60] Stute, W. (1984a), "The oscillation behavior of empirical processes: the multivariate case," *The Annals of Probability* 12, 361-379.

[61] Stute, W. (1984b), "Asymptotic normality of nearest neighbor regression function estimates," *The Annals of Statistics* 12, 917-926.

[62] Stute, W. (1986), "Conditional empirical processes, " *The Annals of Statistics* 14, 638-647.

[63] Stute, W. and L.-X. Zhu (2005), "Nonparametric checks for single-index models," *The Annals of Statistics* 33, 1048-1083.

[64] Su, L. and H. White (2011), "Conditional independence specification testing for dependent processes with local polynomial quantile regression," forthcoming in *Advances in Econometrics.*

[65] Truong, Y.K. (1989), "Asymptotic properties of kernel estimators based on local medians," *The Annals of Statistics* 17, 606-617.

[66] Thompson, W.R. (1936), "On confidence ranges for the median and other Expectation distributions for populations of unknown distribution form," *The Annals of Mathematical Statistics* 3, 122-128.

[67] Tsybakov, A. (2008), *Introduction to Nonparametric Estimation,* Springer, New York.

[68] van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge University Press.

[69] van der Vaart, A.W. and J.A. Wellner (1996), *Weak Convergence and Empirical Processes: With Application to Statistics,* Springer-Verlag, New York.

[70] Whang, Y.J. (2006), "Smoothed empirical likelihood methods for quantile regression models," *Econometric Theory* 22, 173-205.

[71] Wu, W. B. (2005), "On the Bahadur representation of sample quantiles for dependent sequences," *The Annals of Statistics* 33, 1934-1963.

[72] Wu, T.Z., K. Yu, and Y. Yu (2010), "Single-index quantile regression," *Journal of Multivariate Analysis* 101, 1607-1621.

[73] Xu, K.-L. (2012), "Nonparametric inference for conditional quantiles of time series," forthcoming in *Econometric Theory.*

[74] Yu, K. and M.C. Jones (1998), "Local linear quantile regression," *Journal of the American Statistical Association* 93, 228-237.

[75] Zhou, K.Q. and S. Portnoy (1996), "Direct use of regression quantiles to construct confidence sets in linear models," *The Annals of Statistics* 24, 287-306.

Table 1: Coverage Rate in Nonparametric Models ($n = 200$)

Q-Level	$p = 0.25$			$p = 0.5$			$p = 0.75$		
x	0	0.75	1.5	0	0.75	1.5	0	0.75	1.5
Model-1	Curvy Homo								
Asy NW	0.9674	0.9514	0.9358	0.9834	0.9894	0.9846	0.9692	0.9316	0.9556
Asy CI	0.9832	0.9546	0.9350	0.9884	0.9918	0.9552	0.9738	0.9380	0.9398
New NW	0.9232	0.9630	0.9386	0.9434	0.9498	0.9678	0.9598	0.9632	0.9074
New CI	0.9372	0.9656	0.9678	0.9508	0.9558	0.9566	0.9682	0.9654	0.9436
Boot Nm	0.8824	0.8670	0.8586	0.8938	0.8844	0.884	0.8602	0.856	0.879
Boot Perc	0.9584	0.9466	0.9320	0.9096	0.9600	0.8968	0.8586	0.931	0.9584
Model-2	Linear Hetero								
Asy NW	0.9556	0.9424	0.9068	0.9650	0.9532	0.9254	0.9550	0.9294	0.8920
Asy CI	0.9610	0.9452	0.8918	0.9756	0.9582	0.9084	0.9532	0.9382	0.8850
New NW	0.9518	0.9550	0.9606	0.9546	0.9566	0.9554	0.9594	0.9584	0.9568
New CI	0.9534	0.9582	0.9358	0.9566	0.9582	0.9454	0.9646	0.9598	0.9524
Boot Nm	0.8952	0.9066	0.8958	0.8964	0.9026	0.8988	0.8884	0.8860	0.8808
Boot Perc	0.9552	0.9576	0.9532	0.9534	0.9552	0.9486	0.9390	0.9456	0.9246

Table 2: Coverage Rate in Nonparametric Models ($n = 500$)

Q-Level	$p = 0.25$			$p = 0.5$			$p = 0.75$		
x	0	0.75	1.5	0	0.75	1.5	0	0.75	1.5
Model-1	Curvy Homo								
Asy NW	0.9664	0.9440	0.9430	0.9870	0.9780	0.9890	0.9830	0.9432	0.9536
Asy CI	0.9814	0.9516	0.9474	0.9892	0.9808	0.9646	0.9836	0.9466	0.9402
New NW	0.9382	0.9580	0.9634	0.9492	0.9546	0.9422	0.9584	0.9570	0.9598
New CI	0.9548	0.9570	0.9594	0.9580	0.9584	0.9582	0.9608	0.9626	0.9526
Boot Nm	0.8944	0.8944	0.8810	0.9030	0.8986	0.8908	0.8940	0.8986	0.8998
Boot Perc	0.9626	0.9532	0.9418	0.9608	0.9660	0.9544	0.9322	0.9534	0.9618
Model-2	Linear Hetero								
Asy NW	0.9514	0.9476	0.9278	0.9642	0.9540	0.9486	0.9606	0.9460	0.9170
Asy CI	0.9544	0.9462	0.9148	0.9672	0.9612	0.9362	0.9574	0.9452	0.9090
New NW	0.9396	0.9420	0.9542	0.9546	0.9484	0.9610	0.9520	0.9584	0.9652
New CI	0.9528	0.9558	0.9428	0.9568	0.9586	0.9532	0.9538	0.9610	0.9522
Boot Nm	0.907	0.9136	0.903	0.9086	0.9158	0.8986	0.9088	0.9086	0, 8998
Boot Perc	0.953	0.9556	0.943	0.9568	0.9612	0.9446	0.9506	0.9562	0.9336

Table 3: Coverage Rate in Nonparametric Models ($n = 1000$)

Q-Level	$p = 0.25$			$p = 0.5$			$p = 0.75$		
x	0	0.75	1.5	0	0.75	1.5	0	0.75	1.5
Model-1	Curvy Homo								
Asy NW	0.9602	0.9494	0.9508	0.9896	0.9704	0.9832	0.9864	0.9536	0.9574
Asy CI	0.9714	0.9518	0.9572	0.9910	0.9748	0.9700	0.9838	0.9536	0.9554
New NW	0.9512	0.9598	0.9634	0.9512	0.9570	0.9474	0.9578	0.9560	0.9598
New CI	0.9540	0.9618	0.9612	0.9538	0.9582	0.9620	0.9632	0.9600	0.9554
Boot Nm	0.9092	0.8894	0.8916	0.9068	0.9140	0.9052	0.9000	0.9066	0.9094
Boot Perc	0.9614	0.9554	0.9480	0.9566	0.9614	0.9596	0.9444	0.9572	0.9628
Model-2	Linear Hetero								
Asy NW	0.9602	0.9518	0.9302	0.9522	0.9642	0.9540	0.9622	0.9526	0.9398
Asy CI	0.9642	0.9552	0.9254	0.9572	0.9672	0.9612	0.9602	0.9554	0.9280
New NW	0.9526	0.9504	0.9602	0.9602	0.9586	0.9630	0.9586	0.9572	0.9606
New CI	0.9548	0.9516	0.9460	0.9604	0.9602	0.9518	0.9600	0.9616	0.9488
Boot Nm	0.9186	0.9170	0.9025	0.9146	0.9178	0.9098	0.9138	0.9166	0.9004
Boot Perc	0.9633	0.9546	0.9418	0.9576	0.9560	0.9488	0.9516	0.9538	0.9434

Table 4: Coverage Rate in Nonparametric Models ($n = 1000$, One Bandwidth)

Q-Level	$p = 0.25$			$p = 0.5$			$p = 0.75$		
x	0	0.75	1.5	0	0.75	1.5	0	0.75	1.5
Model-1	Curvy Homo								
Asy NW	0.7858	0.7488	0.5634	0.8000	0.7748	0.5906	0.7852	0.7464	0.5560
Asy CI	0.7728	0.7470	0.6980	0.7894	0.7778	0.7260	0.7772	0.7484	0.7084
New NW	0.9522	0.9626	0.9672	0.9518	0.9598	0.9572	0.9582	0.9594	0.9672
New CI	0.9540	0.9618	0.9612	0.9538	0.9582	0.9620	0.9632	0.9600	0.9554
Model-2	Linear Hetero								
Asy NW	0.9522	0.9416	0.9162	0.9518	0.9442	0.9264	0.9518	0.9408	0.9116
Asy CI	0.9508	0.9414	0.9204	0.9570	0.9480	0.9234	0.9472	0.9448	0.9140
New NW	0.9516	0.9502	0.9610	0.9612	0.9586	0.9644	0.9588	0.9578	0.9602
New CI	0.9548	0.9516	0.9460	0.9604	0.9602	0.9518	0.9600	0.9616	0.9488

Table 5: Coverage Rate in Partial Linear Model ($n = 500$)

Q-Level	$p = 0.25$			$p = 0.5$			$p = 0.75$		
$\binom{x}{z}$	$\binom{0.25}{0.5}$	$\binom{0.5}{1}$	$\binom{0.75}{1.5}$	$\binom{0.25}{0.5}$	$\binom{0.5}{1}$	$\binom{0.75}{1.5}$	$\binom{0.25}{0.5}$	$\binom{0.5}{1}$	$\binom{0.75}{1.5}$
Asy NW	0.9180	0.9334	0.9172	0.9278	0.9384	0.9272	0.9164	0.9288	0.9196
Asy CI	0.9192	0.9320	0.9224	0.9262	0.9422	0.9276	0.9152	0.9286	0.9174
New NW	0.9486	0.9612	0.9482	0.9506	0.9582	0.9512	0.9502	0.9568	0.9462
New CI	0.9482	0.9600	0.9460	0.9502	0.9558	0.9506	0.9510	0.9560	0.9494
Boot Nm	0.9370	0.9448	0.9484	0.9366	0.9490	0.9478	0.9290	0.9486	0.9412
Boot Perc	0.9822	0.9800	0.9708	0.9816	0.9796	0.9714	0.9804	0.9798	0.9712

Table 6: Coverage Rate in Partial Linear Model ($n = 1000$)

Q-Level	$p = 0.25$			$p = 0.5$			p=0.75		
$\binom{x}{z}$	$\binom{0.25}{0.5}$	$\binom{0.5}{1}$	$\binom{0.75}{1.5}$	$\binom{0.25}{0.5}$	$\binom{0.5}{1}$	$\binom{0.75}{1.5}$	$\binom{0.25}{0.5}$	$\binom{0.5}{1}$	$\binom{0.75}{1.5}$
Asy NW	0.9278	0.9382	0.9282	0.9324	0.9410	0.9382	0.9274	0.9374	0.9290
Asy CI	0.9262	0.9414	0.9284	0.9328	0.9434	0.9368	0.9282	0.9344	0.9302
New NW	0.9536	0.9530	0.9564	0.9502	0.9604	0.9514	0.9484	0.9598	0.9482
New CI	0.9496	0.9532	0.9534	0.9506	0.9566	0.9492	0.9512	0.9570	0.9462
Boot Nm	0.9394	0.9536	0.947	0.9418	0.954	0.9470	0.9410	0.9506	0.9464
Boot Perc	0.9800	0.9782	0.972	0.9762	0.978	0.9722	0.9766	0.9778	0.9716

Table 7: Coverage Rate in Partial Linear Model ($n = 1000$, One Bandwidth)

Q-Level	$p = 0.25$			$p = 0.5$			$p = 0.75$		
$\binom{x}{z}$	$\binom{0.25}{0.5}$	$\binom{0.5}{1}$	$\binom{0.75}{1.5}$	$\binom{0.25}{0.5}$	$\binom{0.5}{1}$	$\binom{0.75}{1.5}$	$\binom{0.25}{0.5}$	$\binom{0.5}{1}$	$\binom{0.75}{1.5}$
Asy NW	0.7972	0.8040	0.7946	0.8142	0.8266	0.8168	0.7982	0.8068	0.7962
Asy CI	0.8060	0.8014	0.7940	0.8164	0.8276	0.8168	0.8002	0.8096	0.7980
New NW	0.9526	0.9534	0.9562	0.9512	0.9598	0.9524	0.9498	0.9594	0.9470
New CI	0.9496	0.9532	0.9534	0.9506	0.9566	0.9492	0.9512	0.9570	0.9462