

A Direct Approach to Inference in Nonparametric and Semiparametric Quantile Models

Yanqin Fan and Ruixuan Liu*
University of Washington, Seattle

Working Paper no. 140
Center for Statistics and the Social Sciences
University of Washington

First version: December 2012
This version: September 2013

Abstract

This paper makes two main contributions to inference for conditional quantiles. First, we construct a generic confidence interval for a conditional quantile from any given estimator of the conditional quantile via the direct approach. Our generic confidence interval makes use of two estimates of the conditional quantile function evaluated at two appropriately chosen quantile levels. In contrast to the standard Wald type confidence interval, ours circumvents the need to estimate the conditional density function of the dependent variable given the covariate. We show that our new confidence interval is asymptotically valid for any quantile function (parametric, nonparametric, or semiparametric), any conditional quantile estimator (standard kernel, local polynomial or sieve estimates), and any data structure (random samples, time series, or censored data), provided that certain weak convergence of the conditional quantile process holds for the preliminary quantile estimator. In the same spirit, we also construct a generic confidence band for the conditional quantile function across a range of covariate values. Second, we use a specific estimator, the Yang-Stute (also known as the symmetrized k -NN) estimator for a nonparametric quantile function, and two popular semiparametric quantile functions to demonstrate that oftentimes by a judicious choice of the quantile estimator combined with the specific model structure, one may further take advantage of the flexibility and simplicity of the direct approach. For instance, by using the Yang-Stute estimator, we construct confidence intervals and bands for a nonparametric and two semiparametric quantile functions that are free from additional bandwidth choices involved in estimating not only the conditional but also the marginal density functions and that are very easy to implement. The advantages of our new confidence intervals are borne out in a simulation study.

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

JEL Codes: C12; C14; C21

*Department of Economics, University of Washington, Box 353330, Seattle, WA 98195. We thank Matias Cattaneo, Xuming He, Yu-Chin Hsu, Sang Soo Park, and seminar participants at Northwestern, Peking University, Simon Fraser University, Texas A&M, University of British Columbia, and the University of Michigan, as well as participants at 2013 Tsinghua International Conference in Econometrics and the 2013 North American Econometric Society Meetings 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), 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), and He, Ng, and Portnoy (1998) 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), Fan and Liu (2011), and Li and Racine (2008), among others. In this approach, the conditional distribution function of the dependent variable Y 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) focus on univariate covariate and estimate the conditional distribution function by k -NN method, while Fan and Liu (2011) and Li and Racine (2008) allow for multivariate covariate and adopt 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 (CI) for the true conditional quantiles. Regardless of the 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

¹Conditional quantile function also plays an important role in the non-separable structural econometrics literature, see e.g., Chesher (2003), Holderlein and Mammen (2007) and in the estimation of quantile treatment effects, see e.g., Firpo (2007) and Fan and Park (2011).

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 estimates 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), Csorgo (1983), 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; and 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 semi-parametric 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 any estimate 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.

This paper aims at bridging this gap. Specifically, it makes two main contributions to inference on conditional quantiles. First, we construct a generic confidence interval (CI) for a conditional quantile from any given estimator of the conditional quantile via the direct approach. Our generic confidence interval makes use of two estimates of the conditional quantile function evaluated at two appropriately chosen quantile levels. If the original quantile estimator is monotone in the quantile level $p \in (0, 1)$, then the two estimates are computed from this estimator; else the two estimates are computed from the monotone rearranged version of the original quantile estimator as proposed in Chernozhukov, Fernandez-Val, and Galichon (2010). In contrast to the standard Wald type confidence interval, ours circumvents the need to estimate the conditional density function of the dependent variable given the covariate. We show that our new confidence interval is asymptotically valid for any quantile function (parametric, nonparametric, or semiparametric), any conditional quantile estimator (standard kernel, local polynomial or sieve estimates), and any data structure (random samples, time series, or censored data), provided that certain weak convergence of the conditional quantile process holds for the preliminary quantile estimator. In the same spirit, we also construct a generic confidence band (CB) for the conditional quantile function across a range of covariate values focusing on the nonparametric setting and a class of quantile estimators obtained from inverting proper estimators of the conditional distribution function of Y given X . Since members of this class of quantile estimators are monotone by construction, monotone rearrangement is avoided. Second, we use a specific estimator, the Yang-Stute (also known as the symmetrized k -NN) estimator for a nonparametric quantile function, and two popular semiparametric quantile functions to demonstrate that oftentimes by a judicious choice of the quantile estimator combined with the specific model structure, one may further take advantage of the flexibility and simplicity of the direct approach. For instance, by using the Yang-Stute estimator, we construct confidence intervals and bands for a nonparametric and two semiparametric quantile functions that are free from additional bandwidth choices involved in estimating not only the conditional but also the marginal density functions and that are very easy to compute. The reason that we choose the Yang-Stute estimator is its simplicity and elegance; It inherits the so-called asymptotic distributional-free property (Stute, 1984b) and avoids estimating covariate’s marginal density function (unlike standard kernel estimators), so we are able to eliminate all unnecessary tuning parameters. Besides, as we directly invert conditional distribution functions, the resulting conditional quantile estimators are indeed monotone, so there is no need for monotone rearrangement. Of course, practitioners

are free to choose their favorite preliminary quantile estimators and under the mild high level assumptions below, our generic CIs/CBs would apply.

Like the empirical likelihood confidence interval for a nonparametric quantile function in Xu (2012), our confidence intervals/bands for nonparametric quantiles based on the Yang-Stute estimator internalize the conditional quantile density estimation of Y given X and the covariate density estimation and they are not necessarily symmetric. Compared with Xu (2012), our procedure is much easier to implement and does not require optimization. For conditional quantiles in partial linear and single index quantile regressions, a direct application of the generic CI and CB would require monotone rearrangement, but by making use of the model structures, we construct CIs and CBs that are easy to implement avoiding monotone rearrangement. A small scale simulation study demonstrates the advantages and feasibility of our confidence intervals/bands over existing ones in practically relevant model set-ups. Finally, we point out that there is an interesting connection between our generic CI and the well-known CI for unconditional quantiles based on order statistics originally proposed in Thompson (1936), see also Serfling (1980) and van der Vaart (1998). In fact by using pairs of the standard k -NN asymmetric quantile estimate, our generic CI employs pairs of order statistics of the induced order statistics of Y , so it shares the elegance and simplicity of the confidence interval for unconditional quantiles based on order statistics.

The rest of this paper is organized as follows. Section 2.1 presents our generic confidence interval and shows its asymptotic validity under a high level assumption on the preliminary quantile estimator. The high level assumption is verified in four examples including the asymmetric k -NN estimator, local polynomial quantile regression, a nonparametric quantile regression with censoring, and the class of conditional quantile estimators in Donald, Hsu, and Barrett (2012) which includes parametric quantile estimators as well. A generic CB is proposed in Section 2.2. Section 3.1 considers the nonparametric quantile regression with a univariate covariate. It constructs a new confidence interval and a new confidence band using the Yang-Stute estimator. Section 3.2 extends the confidence intervals/bands developed in Section 3.1 to two popular semiparametric models, partial linear and single index quantile regression models. Section 4 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 technical proofs are collected in the Appendices.

2 Generic Results on the Direct Approach to Quantile Inference

Consider the random vector $(X', Y)'$ with marginal distribution functions $F_X(x)$, $F_Y(y)$ respectively, where $x \in \mathcal{X} \subset \mathcal{R}^d$ and $y \in \mathcal{Y} \subset \mathcal{R}$. Let $F_{Y|X}(\cdot|x)$ denote the conditional distribution function of Y given $X = x$ with density function $f_{Y|X}(\cdot|x)$. For $0 < p < 1$, we are interested in conducting inference on the p -th conditional quantile of Y given $X = x$: $\xi(p|x) \equiv F_{Y|X}^{-1}(p|x)$ either

at a specific location $x = x_0 \in \mathcal{X}$ or for all x in a subset of the support of X .

Let $\widehat{\xi}(p|x)$ denote a consistent and asymptotically normally distributed estimator of $\xi(p|x)$. To introduce the direct approach to quantile inference, consider inference for $\xi(p|x_0)$ for a fixed $x_0 \in \mathcal{X}$. Suppose $\widehat{\xi}(p|x)$ is monotone in the quantile level $p \in (0, 1)$. Then the CI for $\xi(p|x_0)$ based on the direct approach takes the form of a closed interval with the end points given by $\widehat{\xi}(\cdot|x_0)$ evaluated at two appropriately chosen quantile levels, one smaller and one larger than p , see (3) below with $\widehat{\xi}^*(p|x_0)$ being replaced by $\widehat{\xi}(p|x_0)$. In contrast to Wald-type CIs, CIs based on the direct approach are not dependent on any estimate of the conditional density function of Y given $X = x_0$.

To ensure the validity of the resulting CI, it is essential that the quantile estimator being used is monotone in the quantile level $p \in (0, 1)$. It is well known that some commonly used quantile estimators including the linear quantile estimator of Koenker and Bassett (1978) and local polynomial quantile estimators are not monotone in the quantile level. This is known as the quantile crossing problem (He, 1997; Chernozhukov, Fernandez-Val, and Galichon, 2010). To rectify this issue, various methods of monotonization have been proposed in the literature including Chernozhukov, Fernandez-Val, and Galichon (2010) who propose a monotone rearranged quantile estimator from a given preliminary quantile estimator and Dette and Volgushev (2008) who propose smooth monotone quantile estimators from consistent estimators of the conditional distribution function $F_{Y|X}(\cdot|x)$. Of course quantile estimators constructed from inverting monotone estimators of the conditional distribution function $F_{Y|X}(\cdot|x)$ are also monotone.

In Section 2.1, we construct a CI for $\xi(p|x_0)$ from a preliminary consistent estimator $\widehat{\xi}(p|x_0)$ using the direct approach and refer to it as the generic CI. For non-monotone $\widehat{\xi}(p|x_0)$, our generic CI makes use of the rearranged version of $\widehat{\xi}(p|x_0)$ in Chernozhukov, Fernandez-Val, and Galichon (2010).² In Section 2.2, we construct a generic CB from the direct approach which is valid for all x in a compact subset of \mathcal{X} .

2.1 A Generic Confidence Interval

For any consistent estimator $\widehat{\xi}(p|x)$ of $\xi(p|x)$, the monotone version of $\widehat{\xi}(p|x)$ in Chernozhukov, Fernandez-Val, and Galichon (2010) is based on the fact that

$$F_{Y|X}(y|x) = \xi^{-1}(y|x) = \int_0^1 I\{\xi(u|x) \leq y\} du, \quad (1)$$

so we can replace $\xi(u|x)$ with $\widehat{\xi}(u|x)$ in the expression on the right hand side of (1) to get a monotone estimator of $F_{Y|X}(y|x)$. Since the resulting estimator of $F_{Y|X}(y|x)$ is monotone, its generalized

²We could use the smooth monotone estimators proposed in Dette and Volgushev (2008) as well. But since they involve the choice of an additional smoothing parameter, we find the method in Chernozhukov, Fernandez-Val, and Galichon (2010) more suitable for our purpose. Marmer and Shneyerov (2012) offer an alternative method for constructing a monotone quantile estimator from a preliminary quantile estimator, but the asymptotic properties of their monotone estimator are unknown.

inverse is a consistent and monotone estimator of $\xi(p|x)$. This is the monotone rearranged version of $\widehat{\xi}(p|x_0)$ proposed by Chernozhukov, Fernandez-Val, and Galichon (2010):

$$\widehat{\xi}^*(p|x) = \inf \left[y : \int_0^1 1\{\widehat{\xi}(u|x) \leq y\} du \geq p \right]. \quad (2)$$

We note that if the original estimator $\widehat{\xi}(p|x)$ is monotone in $p \in (0, 1)$, then $\widehat{\xi}^*(p|x) = \widehat{\xi}(p|x)$ for all $p \in (0, 1)$. So we will use $\widehat{\xi}^*(p|x)$ in this section to introduce our generic CI for $\xi(p|x_0)$.

Below we first provide assumptions on the quantile function $\xi(p|x_0)$ and a high level assumption on the original estimator $\widehat{\xi}(p|x_0)$ under which our generic CI is asymptotically valid and then verify the high level assumption for four examples in Section 2.1.1.

Assumption (GI)

(i) $\xi(p|x_0)$ 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 $q_p(x) \equiv \frac{\partial}{\partial p} \xi(p|x) = 1/f_{Y|X}(\xi(p|x)|x)$ denote the conditional quantile density function. Then $q_p(x_0) > 0$ for $p \in [p_1, p_2] \subset (0, 1)$;

(iii) The quantile estimator $\widehat{\xi}(\cdot|x_0)$ takes its values in the space of bounded measurable functions defined on $[p_1, p_2] \subset (0, 1)$, where $p \in [p_1, p_2]$, and in $l^\infty([p_1, p_2])$,

$$c_n \left(\widehat{\xi}(\cdot|x_0) - \xi(\cdot|x_0) \right) \implies q_\cdot(x_0) B(\cdot|x_0),$$

as a stochastic process indexed by $p \in [p_1, p_2]$, where $\{B(p|x_0), p \in [p_1, p_2]\}$ is a Gaussian process with variance $\sigma^2(p|x_0) \equiv \text{Var}[B(p|x_0)]$ which does not depend on $q_p(x_0)$ and c_n is a sequence of positive constants such that $c_n \rightarrow \infty$ as $n \rightarrow \infty$.

Assumption (GI) (i) and (ii) are taken directly from Chernozhukov, Fernandez-Val, and Galichon (2010). Assumption (GI) (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 generic confidence interval obtained from the direct approach:

$$\text{CI-G}_{1-\alpha} = \left[\widehat{\xi}^* \left(p - \frac{z_{\alpha/2} \widehat{\sigma}(p|x_0)}{c_n} \middle| x_0 \right), \widehat{\xi}^* \left(p + \frac{z_{\alpha/2} \widehat{\sigma}(p|x_0)}{c_n} \middle| x_0 \right) \right], \quad (3)$$

where $\alpha \in (0, 1)$, $\widehat{\sigma}(p|x_0)$ is a consistent estimator of $\sigma(p|x_0)$, and $z_{\alpha/2}$ is the upper quantile of standard normal random variable N , i.e. $\Pr\{N > z_{\alpha/2}\} = 1 - \alpha/2$. For many quantile estimators $\widehat{\xi}(p|x)$ regardless of the model and data structure, Assumption (GI) (iii) is either established in existing work or can be shown using results in existing work, see Section 2.1.1 and Section 3 for examples of parametric, nonparametric and semiparametric quantile estimators. Moreover, for many quantile estimators, $\sigma^2(p|x_0)$ takes the form of $p(1-p)\varpi_{x_0}^2$ for some positive constant ϖ_{x_0} depending on x_0 , see the first three examples in Section 2.1.1 and Section 3. Example 2.4 presents

an example of $\sigma^2(p|x_0)$ that does not take this form.³ In addition to the examples in Section 2.1.1 and Section 3, another (parametric) example of $\widehat{\xi}(p|x)$ is the quantile estimator of Koenker and Bassett (1978). Under standard regularity conditions, the quantile estimator of Koenker and Bassett (1978) satisfies Assumption (GI) (iii) for the special class of location scale forms of linear quantile regression models, see Zhou and Portnoy (1996), Gutenbrunner and Jureckova (1992), Koenker and Xiao (2005), and Portnoy (2012), so our generic confidence interval CI-G $_{1-\alpha}$ defined in (3) is asymptotically valid. In fact, for location scale forms of the linear quantile regression models, our generic confidence interval CI-G $_{1-\alpha}$ is just the confidence interval in Zhou and Portnoy (1996) using the original estimator of Koenker and Bassett (1978), due to the absence of quantile crossing problem as demonstrated by He (1997).

THEOREM 2.1 *Suppose Assumption (GI) 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 $\widehat{\xi}(p|x_0)$. In particular, Assumption (GI) (iii) implies that

$$c_n \left(\widehat{\xi}^*(p|x_0) - \xi(p|x_0) \right) \implies q_p(x_0) B(p|x_0).$$

Making use of stochastic equicontinuity of the process $\left\{ c_n \left(\widehat{\xi}^*(p|x_0) - \xi(p|x_0) \right), p \in (0, 1) \right\}$ and $c_n \rightarrow \infty$, we have

$$c_n \left[\widehat{\xi}^* \left(p \pm \frac{z_{\alpha/2} \widehat{\sigma}(p|x_0)}{c_n} | x_0 \right) - \widehat{\xi}^*(p|x_0) \right] = c_n \left[\xi \left(p \pm \frac{z_{\alpha/2} \widehat{\sigma}(p|x_0)}{c_n} | x_0 \right) - \xi(p|x_0) \right] + o_p(1).$$

Now use the simple fact that under Assumption (GI) (i) and consistency of $\widehat{\sigma}(p|x_0)$:

$$c_n \left[\xi \left(p \pm \frac{z_{\alpha/2} \widehat{\sigma}(p|x_0)}{c_n} | x_0 \right) - \xi(p|x_0) \right] = \pm q_p(x_0) z_{\alpha/2} \sigma(p|x_0) + o_p(1).$$

Hence the generic confidence interval is asymptotically valid:

$$\begin{aligned} & \lim_{n \rightarrow \infty} \Pr \{ \xi(p|x_0) \in \text{CI-G}_{1-\alpha} \} \\ &= \lim_{n \rightarrow \infty} \Pr \left\{ \widehat{\xi}^* \left(p - \frac{z_{\alpha/2} \widehat{\sigma}(p|x_0)}{c_n} | x_0 \right) \leq \xi(p|x_0) \leq \widehat{\xi}^* \left(p + \frac{z_{\alpha/2} \widehat{\sigma}(p|x_0)}{c_n} | x_0 \right) \right\} \\ &= \lim_{n \rightarrow \infty} \Pr \left\{ \widehat{\xi}^*(p|x_0) - \frac{z_{\alpha/2} \sigma(p|x_0)}{c_n} q_p(x_0) \leq \xi(p|x_0) \leq \widehat{\xi}^*(p|x_0) + \frac{z_{\alpha/2} \sigma(p|x_0)}{c_n} q_p(x_0) \right\} \\ &= 1 - \alpha. \end{aligned}$$

Q.E.D

³We thank Yu-Chin Hsu for suggesting this example.

Remark 2.1. Under Assumption (GI), the standard Wald type confidence interval is constructed by centering around $\widehat{\xi}(p|x_0)$ with the standard error multiplied by the normal critical value as

$$\left[\widehat{\xi}(p|x_0) - \frac{z_{\alpha/2}\widehat{\sigma}(p|x_0)}{\widehat{f}_{Y|X}(\widehat{\xi}(p|x_0)|x_0)c_n}, \widehat{\xi}(p|x_0) + \frac{z_{\alpha/2}\widehat{\sigma}(p|x_0)}{\widehat{f}_{Y|X}(\widehat{\xi}(p|x_0)|x_0)c_n} \right], \quad (4)$$

where $\widehat{f}_{Y|X}(y|x_0)$ is a consistent estimator of $f_{Y|X}(y|x_0)$ such as a kernel conditional density estimator. It is well known that the finite sample performance of the Wald CI in (4) is very sensitive to the choice of the smoothing parameter involved in the estimate $\widehat{f}_{Y|X}(y|x_0)$. Distinct from the Wald type interval, our new confidence interval in (3) avoids the estimation of the conditional density function $f_{Y|X}(y|x_0)$ and the two end points of the CI are not necessarily symmetric around $\widehat{\xi}(p|x_0)$.

2.1.1 Examples of the Generic Confidence Interval

Let $\{(X'_i, Y'_i)\}_{i=1}^n$ denote the sample information on $(X', Y)'$. It could be a random sample or a time series. To demonstrate the broad applicability of the confidence interval, CI-G $_{1-\alpha}$, defined in (3), we present four examples in this subsection. They include a novel confidence interval for nonparametric conditional quantiles based on order statistics/induced order statistics in Example 2.1; a new confidence interval for nonparametric quantile regression based on local polynomial estimator in Example 2.2; a new confidence interval for nonparametric censored quantile regression in Example 2.3, and finally new confidence intervals for the class of conditional quantile models in Donald, Hsu, and Barrett (2012). For the first three examples, $\sigma^2(p|x_0)$ takes the form: $p(1-p)\varpi_{x_0}^2$ for some positive constant ϖ_{x_0} depending on x_0 , while for the last example it doesn't have this specific form.

Example 2.1. (A Novel Order Statistic Approach):⁴ The generic confidence interval in (3) 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 $\{Y_i\}_{i=1}^n$. It extends confidence intervals for unconditional quantiles based on pairs of order statistics of $\{Y_i\}_{i=1}^n$, see Thompson (1936) or van der Vaart (1998) for random samples and Wu (2005) for time series observations.

To introduce it, let $R_i = \|X_i - x_0\|$, for $i = 1, \dots, n$, 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

⁴After finishing the first version of this paper, we came across Kaplan (2013), who also proposed similar inference procedures to this example by further taking linear combinations of those order statistics. Higher order properties have been given in Kaplan (2013) using Dirichlet process theory.

k -NN estimator of the distribution function of Y given $X = x_0$ is defined as

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

and the asymmetric k -NN estimator of $\xi(p|x_0)$ is given by

$$\begin{aligned} \widehat{\xi}(p|x_0) &= \inf \left\{ y : \widehat{F}_{n,k}(y|x_0) \geq \frac{[kp]}{k} \right\} \\ &= \text{the } [kp]\text{-th order statistic of } Y_{n,1}, Y_{n,2}, \dots, Y_{n,k}, \end{aligned} \quad (5)$$

where $k \equiv k_n$ is a sequence of constants such that $k_n \rightarrow \infty$ and $k_n = o\left(n^{\frac{4}{4+d}}\right)$. Assuming (GI) (i) and (ii), the asymptotic validity of CI-G $_{1-\alpha}$ based on the asymmetric k -NN estimator relies on Assumption (GI) (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 (GI) (iii) holds for $\widehat{\xi}(p|x_0)$ in (5) with $c_n = \sqrt{k_n}$ and $\varpi_{x_0}^2 = \pi^{d/2}/\Gamma(d/2 + 1)$. 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. Since by definition $\widehat{\xi}(p|x_0)$ in (5) is monotone in p , the confidence interval (3) reduces to:

$$\begin{aligned} \text{CI-O}_{1-\alpha} &= \left[\widehat{\xi}(p - z_{\alpha/2}\sigma_{kn}|x_0), \widehat{\xi}(p + z_{\alpha/2}\sigma_{kn}|x_0) \right] \\ &= \left[Y_{n,([k(p - z_{\alpha/2}\sigma_{kn})])}, Y_{n,([k(p + z_{\alpha/2}\sigma_{kn})])} \right], \end{aligned} \quad (6)$$

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. The new confidence interval CI-O $_{1-\alpha}$ defined in (6) for conditional quantiles shares the elegance and simplicity of the confidence interval for unconditional quantiles based on order statistics.

Example 2.2. (Local Polynomial Quantile Regression): A local polynomial estimator defined as the minimizer of a weighted check function is the natural nonparametric analog of the linear quantile estimator originated by Koenker and Bassett (1978). Recall that the check function is of the form: $\rho_p(t) = pt_+ + (1-p)t_-$, where subscripts $+$, $-$ stand for the positive and negative parts respectively. The local polynomial estimator of order s using kernel function $K(\cdot)$ is defined by $\widehat{\xi}(p|x_0) = e_1' \widehat{\theta}(p|x_0)$, with $e_1 = (1, 0, \dots, 0)$ and

$$\widehat{\theta}(p|x_0) = \arg \min_{\theta} \sum_{i=1}^n \rho_p \left(Y_i - \theta^T P_s \left(\frac{x_0 - X_i}{h_n} \right) \right) K \left(\frac{x_0 - X_i}{h_n} \right),$$

where $P_s(z)^T = \left(\frac{z^{\mathbf{v}}}{\mathbf{v}!}, |\mathbf{v}| \leq s \right)$ for $\mathbf{v} = (v_1, \dots, v_d)$ with $|\mathbf{v}| = v_1 + \dots + v_d$, $\mathbf{v}! = \prod_{i=1}^d v_i!$ and the vectors of $\mathbf{v} \in \mathcal{N}^d$ being ordered lexicographically, $z^{\mathbf{v}} = \prod_{i=1}^d z_i^{v_i}$ (see Chaudhuri, 1991; Fan, Hu

and Truong, 1994; Guerre and Sabbah, 2012). With i.i.d. observations, under primitive conditions Guerre and Sabbah (2012) establish a Bahadur representation 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 (GI) (iii) is satisfied under their conditions, where $\varpi_{x_0} = \|K\|_2 / \sqrt{f_X(x_0)}$ and $c_n = \sqrt{nh_n^d}$ with h_n the bandwidth. In the time series setting, Assumption (GI) (iii) is satisfied under the conditions in either Polonik and Yao (2002) or Su and White (2012) under proper mixing conditions. It is known that the local polynomial estimator $\widehat{\xi}(p|x_0)$ is not guaranteed to be monotone in p , so our generic CI employs the rearranged version of $\widehat{\xi}(p|x_0)$.

Example 2.3. (Nonparametric Quantile Regression With Censoring): Consider a nonparametric censored quantile regression model where the dependent variable Y_i is subject to conditional random censoring by C_i . So instead of observing $\{X_i, Y_i\}_{i=1}^n$, we observe a random sample $(\min(Y_i, C_i), \delta_i, X_i)_{i=1}^n$, 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 symmetrized k -NN estimator, and the standard k -NN estimator to the above censored case. Under primitive conditions, she establishes weak convergence of the associated quantile processes which ensures Assumption (GI) (iii) with standard nonparametric convergence rate and the factor ϖ_{x_0} now also involving the conditional cumulative hazard function and conditional sub-survival function (see Corollary 2.2 in Dabrowska, 1987). Thus our generic confidence interval defined in (3) is asymptotically valid. Also one could use local polynomial type estimators given the work of Kong, Linton and Xia (2013).

Example 2.4. (Inverting Estimators of the Conditional Distribution Function): Donald, Hsu, and Barrett (2012) consider various models for the conditional distribution function including fully parametric models of the form: $F_{Y|X}(y|x) \equiv F(y|x, \theta_0)$ for a known function F and an unknown parameter θ_0 and semiparametric models of the form: $F_{Y|X}(y|x) \equiv F(\log y - x'\theta_0)$ for an unknown (F, θ_0) . For each model, they provide primitive conditions under which Assumption (GI) (iii) is satisfied for random samples, but $\sigma^2(p|x_0)$ is not of the form: $p(1-p)\varpi_{x_0}^2$. We refer interested readers to Section 3 in their paper for details.

2.2 A Generic Confidence Band

Often it is of interest to conduct inference simultaneously on $\xi(p|x)$ for all x in a subset of the support of X . In this section, we construct a generic CB for a nonparametric quantile function⁵ by the direct approach again avoiding the conditional density estimation, when x varies in some compact set \mathcal{J} contained in the interior of the support of X . Unlike the generic CI in (3) which relies on the normal limiting distribution, the generic CB developed in this section will rely on the

⁵For 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.

extreme type limiting distribution of a stationary Gaussian process (Bickel and Rosenblatt, 1973; Leadbetter, Lindgren and Rootzen, 1983) obtained from characterizing the maximal deviation of the original nonparametric estimator from the true conditional quantile. Although the first order asymptotic properties of the monotone quantile estimator in Chernozhukov, Fernandez-Val and Galichon (2010) are the same as the original quantile estimator, this is not sufficient for the purpose of constructing a generic CB, see the discussion on Assumption GB (iii) below for a more detailed explanation. Because of this, we will adopt a special class of monotone quantile estimators obtained from inverting monotone estimators of the conditional distribution function of Y given X . Since these estimators are monotone by construction, the generic CI in (3) is applicable to them. For both practical and technical reasons, we focus on the case when X is univariate. For multivariate covariate, semiparametric models including the partially linear and single index quantile regression models are introduced in the literature to alleviate the curse of dimensionality associated with fully nonparametric models. In Section 3, we illustrate how CBs using the direct approach can be constructed for semiparametric quantile models.

Let $\{(X'_i, Y_i)'\}_{i=1}^n$ denote the sample information on $(X', Y)'$, random sample or time series data. Consider any first-step estimator $\hat{F}(y|x)$ of the conditional distribution function $F_{Y|X}(\cdot|x)$ taking the following linear form:

$$\hat{F}(y|x) = \sum_{i=1}^n W_{in}(x) I[Y_i \leq y], \quad (7)$$

where $\{W_{in}(x)\}_{i=1}^n$ is a sequence of non-negative weights summing up to one. Let $\hat{F}^{-1}(p|x)$ denote the generalized inverse of $\hat{F}(\cdot|x)$.

The linear form given in (7) nests almost all commonly used kernel type local smoothers⁶ such as Nadaraya-Watson estimator, Yang-Stute estimator (Yang, 1981; Stute, 1984b), Local Partitioned estimator (the local constant version in Chaudhuri, 1991), and Adjusted Nadaraya-Watson estimator (Hall, Wolff and Yao, 1999):⁷

$$W_{in}^{NW}(x) = \frac{K_{h_n}(x - X_i)}{\sum_{j=1}^n K_{h_n}(x - X_j)}, \quad (8)$$

$$W_{in}^{YS}(x) = \frac{K_{h_n}(F_n(x) - F_n(X_i))}{\sum_{j=1}^n K_{h_n}(F_n(x) - F_n(X_j))}, \quad (9)$$

$$W_{in}^P(x) = \frac{I[X_i \in P_n(x)]}{\sum_{j=1}^n I[X_j \in P_n(x)]}, \text{ and} \quad (10)$$

⁶Theorem 2.2 and the high level assumptions are written focusing on local smoothers. Upon change of notations, sieve type estimators could be incorporated as well, see Figueroa-Lopez (2011).

⁷As shown by Hall, Wolff and Yao (1999), this Adjusted Nadaraya-Watson estimator is asymptotically the same as local linear estimator up to the first order, hence it inherits the smaller bias property of local linear estimator, especially at the boundary. Meanwhile it is superior to local linear estimator as the weights are all positive and summing up to 1, thus the estimated CDF is a proper distribution function.

$$W_{in}^{ANW}(x) = \frac{p_i(x) K_{h_n}(x - X_i)}{\sum_{j=1}^n p_j(x) K_{h_n}(x - X_j)}, \quad (11)$$

where $h \equiv h_n \rightarrow 0$ is a bandwidth and $K_h(\cdot) = K(\cdot/h)/h$ with standard kernel density function K . Moreover in the Yang-Stute estimator, $F_n(x)$ is the empirical distribution function of $\{X_i\}_{i=1}^n$; in the Adjusted Nadaraya-Watson weights, we maximize $\prod_{j=1}^n p_j(x)$ subject to the following constraints: $p_j(x) \geq 0$, $\sum_{j=1}^n p_j(x) = 1$ and $\sum_{j=1}^n K_{h_n}(X_j - x)(X_j - x)p_j(x) = 0$; and in the local partitioned version, $\mathcal{P}_n = \{P_{1n}, P_{2n}, \dots\}$ stands for a partition of \mathcal{X} with $\max_j \text{Leb}(P_{jn})$ of order h_n , and $P_n(x) \in \mathcal{P}_n$, is the set containing point x .

The high level assumptions below are written for the local smoothers with tuning parameter h_n defined above. Implicitly below we select an under-smoothed h_n to kill the bias term.

Assumption (GB)

(i). A continuous and positive conditional density function $f_{Y|X}(y|x)$ exists uniformly over the interval $[F_{Y|X}^{-1}(p_1|x) - \epsilon, F_{Y|X}^{-1}(p_2|x) + \epsilon]$ for some $\epsilon > 0$, where $[p_1, p_2]$ contains p and belongs to $(0, 1)$ and for $x \in \mathcal{J}$.

(ii). The first step estimator for the conditional distribution function in (7) has the following convergence rate:

$$\sup_{x \in \mathcal{J}} \sup_{y \in \mathcal{Y}} C_n \left| \widehat{F}(y|x) - F_{Y|X}(y|x) \right| = O_p(1).$$

Moreover the uniform local oscillation could be bounded as follows:

$$\sup_{x \in \mathcal{J}} \sup_{|y_1 - y_2| \leq O(C_n^{-1})} C_n \left| \widehat{F}(y_1|x) - \widehat{F}(y_2|x) - F_{Y|X}(y_1|x) + F_{Y|X}(y_2|x) \right| = o_p(1),$$

and for all $x \in \mathcal{J}$ the local weights satisfy: $\max_i W_{in}(x) = o(C_n^{-1})$ almost surely.

(iii). Given $p \in (0, 1)$, we could find A_n, D_n and a deterministic function $\Psi(x)$ such that

$$\lim \Pr \left\{ \sup_{x \in \mathcal{J}} A_n \left[\Psi(x) \sqrt{nh_n} \left| \widehat{F}(\xi(p|x)|x) - F_{Y|X}(\xi(p|x)|x) \right| \right] - D_n \leq z \right\} = \exp(-2 \exp(-z)),$$

with $(nh_n)^{-1/2} (D_n - A_n^{-1}) = O(C_n^{-1})$ and $\Psi(x)$ being uniformly bounded away from zero and infinity.

For those local smoothers the uniform convergence rate is $C_n = \sqrt{nh_n (\log n)^{-1}}$ under standard regularity conditions and the magnitude of local weights are all of order $O((nh_n)^{-1})$. The local oscillation could be readily handled given the results in Stute (1984a) or Einmahl and Mason (2005). Apropos of the convergence to the Gumbel distribution in (GB) (iii), we shall rely on the bivariate Gaussian approximation and deduce that the limiting distribution is the maximum of a stationary Gaussian process, say $h^{-1/2} \int K\left(\frac{u-v}{h}\right) dB(v)$ for a Gaussian process $B(\cdot)$ and kernel function $K(\cdot)$ as demonstrated in Hardle (1989), Hardle and Song (2010) for random sample, Liu and Wu (2010) for time series data. Given the limiting distribution in (GB)(iii) and condition in

(GB)(ii) for Bahadur representation of the conditional quantile estimator (by directly inverting the conditional CDF estimator), one could obtain the limiting distribution for the maximal deviation of this conditional quantile estimator, also see Lemma 3.4 and its following remark. The difficulty with the estimator $\widehat{\xi}^*(p|x)$ in Chernozhukov, Fernandez-Val and Galichon (2010) by rearranging an arbitrary $\widehat{\xi}(p|x)$ (may not be monotone in p) is due to the lack of uniform asymptotic property of $\widehat{\xi}^*(p|x)$. To the best of our knowledge, the characterization of the maximal deviation of $\widehat{\xi}^*(p|x)$ remains to be an open question.

Our generic CB takes the following form:

$$\text{CB-G}_{1-\alpha} = \left[\widehat{F}^{-1}(\tau - \sigma_n(x; \alpha) | x), \widehat{F}^{-1}(\tau + \sigma_n(x; \alpha) | x) \right] \quad (12)$$

where

$$\sigma_n(x; \alpha) = \frac{1}{\sqrt{nh_n \widehat{\Psi}(x)}} \left(D_n + \frac{\log 2 - \log |\log(1 - \alpha)|}{A_n} \right)$$

and $\widehat{\Psi}(x)$ denote a uniformly consistent estimator of $\Psi(x)$.

THEOREM 2.2 *Suppose Assumption (GB) holds. Then the confidence band in (12) has the desired asymptotic size:*

$$\lim_{n \rightarrow \infty} \Pr \{ \xi(p|x) \in \text{CB-G}_{1-\alpha}, \forall x \in \mathcal{J} \} = 1 - \alpha.$$

Proof. In order to prove the validity of the generic confidence band, we first need the following type of Bahadur representation for any $\varepsilon_n = O(C_n^{-1})$ around p , uniformly over $x \in \mathcal{J}$:

$$\widehat{F}^{-1}(p + \varepsilon_n | x) - \xi(p|x) = \frac{1}{f_{Y|X}(\xi(p|x) | x)} \left[p + \varepsilon_n - \widehat{F}(\xi(p|x) | x) \right] + o_p(C_n^{-1}). \quad (13)$$

To see why the above result holds, set $\eta_n = \widehat{F}^{-1}(p + \varepsilon_n | x) - \xi(p|x)$. The following string of equalities hold in view of the order of its local oscillation in Assumption (GB) (ii):

$$\begin{aligned} & \widehat{F}(\xi(p|x) + \eta_n | x) - \widehat{F}(\xi(p|x) | x) \\ &= F_{Y|X}(\xi(p|x) + \eta_n | x) - F_{Y|X}(\xi(p|x) | x) + o_p(C_n^{-1}) = f_{Y|X}(\xi(p|x) | x) \eta_n + o_p(C_n^{-1}). \end{aligned} \quad (14)$$

Also due to the linear structure of $\widehat{F}(y|x)$, and the negligibility of the individual weights in Assumption (GB) (ii), we have

$$\widehat{F}(\xi(p|x) + \eta_n | x) = \widehat{F}(\widehat{F}^{-1}(p + \varepsilon_n | x) | x) + o_p(C_n^{-1}) = p + \varepsilon_n + o_p(C_n^{-1}).$$

Thereafter replace $\widehat{F}(\xi(p|x) + \eta_n | x)$ with $p + \varepsilon_n + o_p(C_n^{-1})$ at the LHS in (14), we have shown the first claim. Similar representations as in (13) and their proofs could be found in Serfling (1980) for marginal quantile and Zhou and Portnoy (1996) for a linear quantile regression.

Now with (13) in hand, the next string of derivations need no further explanation once we set $\varepsilon_n = \sigma_n(x; \alpha)$, and for notational simplicity we suppress the smaller order term $o_p(C_n^{-1})$ along the lines:

$$\begin{aligned}
& \Pr \left\{ \widehat{F}^{-1}(p + \sigma_n(x; \alpha) | x) - \xi(p|x) \geq 0, \forall x \in \mathcal{J} \right\} \\
= & \Pr \left\{ \frac{1}{f_{Y|X}(\xi(p|x) | x)} \left[p + \varepsilon_n - \widehat{F}(\xi(p|x) | x) \right] \geq 0, \forall x \in \mathcal{J} \right\} \\
= & \Pr \left\{ \left[\widehat{F}(\xi(p|x) | x) - p - \frac{D_n}{\sqrt{nh_n \widehat{\Psi}(x)}} \right] \leq \frac{\log 2 - \log |\log(1 - \alpha)|}{\sqrt{nh \widehat{\Psi}(x)} A_n}, \forall x \in \mathcal{J} \right\} \\
= & \Pr \left\{ \left[\Psi(x) \sqrt{nh_n} \left(\widehat{F}(\xi(p|x) | x) - F_{Y|X}(\xi(p|x) | x) \right) - D_n \right] \leq \log 2 - \log |\log(1 - \alpha)|, \forall x \in \mathcal{J} \right\}.
\end{aligned}$$

Similarly by omitting the smaller order terms, we get

$$\begin{aligned}
& \Pr \left\{ \widehat{F}^{-1}(p - \sigma_n(x; \alpha) | x) - \xi(p|x) \leq 0, \forall x \in \mathcal{J} \right\} \\
= & \Pr \left\{ \left[\Psi(x) \sqrt{nh_n} \left(\widehat{F}(\xi(p|x) | x) - F_{Y|X}(\xi(p|x) | x) \right) + D_n \right] \geq \log |\log(1 - \alpha)| - \log 2, \forall x \in \mathcal{J} \right\}.
\end{aligned}$$

Thus applying the extreme value type asymptotics in Assumption (GB) (iii), we get

$$\lim_{n \rightarrow \infty} \Pr \left\{ \xi(p|x) \in \text{CB-G}_{1-\alpha}, \forall x \in \mathcal{J} \right\} = 1 - \alpha.$$

Q.E.D

3 Inference Based On Yang-Stute Estimator

Let $\{(X_i, Y_i)'\}_{i=1}^n$ denote a random sample on $(X, Y)'$ for a univariate X . In this section, we give a detailed illustration of the direct approach using the conditional quantile estimator defined as the generalized inverse of particular estimator of $F_{Y|X}(\cdot|x)$ proposed by Yang (1981) and Stute (1984b, 1986). In the sequel we will proceed with the interchangeable notation $F_{Y|X}^{-1}(p|x) = \xi(p|x)$ to highlight this generalized inverse nature. We consider three quantile models: nonparametric, partially linear, and single index models. For each model, we construct a generic CI and a generic CB for the corresponding conditional quantile function and provide primitive conditions under which we show their asymptotic validity. We use partially linear and single index models here to demonstrate how CIs/CBs may be constructed for semiparametric quantile models when the covariate is multivariate.

It is worth repeating here that no monotone rearrangement is needed in this section, because the quantile estimator we adopt is monotone by construction. Compared with the other local smoothers in the previous section, including the (Adjusted) Nadaraya-Watson estimator, and Local Partitioned estimator, we show that the advantage of Yang-Stute estimator is that it allows us to construct CIs

and CBs for all three quantile models that do not require estimation of the covariate’s marginal density function (in fact the method does not even require the existence of covariate’s density function), thus achieving the so-called asymptotic distribution-freeness in Stute (1984b).

3.1 Univariate Nonparametric Quantile Function

We now introduce our estimator of $F_{Y|X}^{-1}(p|x)$ at a fixed value $x_0 \in \mathcal{X}$. Let $\widehat{F}_n(y|x_0)$ denote the estimator of $F_{Y|X}(y|x_0)$ introduced in Yang (1981) and further studied in Stute (1984b, 1986). It is of the form in (7) with weights defined in (9), so

$$\widehat{F}_n(y|x_0) = \frac{\sum_{i=1}^n 1\{Y_i \leq y\} K\left(\frac{F_n(x_0) - F_n(X_i)}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{F_n(x_0) - F_n(X_i)}{h_n}\right)}, \quad (15)$$

where $K(\cdot)$ is a kernel function and $h_n \rightarrow 0$ is a bandwidth. Notice that the local neighborhood around the point of interest is calibrated according to ranks instead of Euclidean distance, so $\widehat{F}_n(y|x_0)$ is also known as the symmetrized k -NN estimator⁸. A more intuitive view of Yang-Stute estimator is the following kernel estimator replacing $F_n(\cdot)$ with $F_X(\cdot)$:

$$\widetilde{F}_n(y|x_0) = \frac{\sum_{i=1}^n 1\{Y_i \leq y\} K\left(\frac{F_X(x_0) - F_X(X_i)}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{F_X(x_0) - F_X(X_i)}{h_n}\right)}.$$

In Appendix A, we show that the difference between $\widetilde{F}_n(\cdot|\cdot)$ and $\widehat{F}_n(\cdot|\cdot)$ is negligible for the inferential purposes considered in our paper, thus $\widehat{F}_n(\cdot|\cdot)$ could be viewed as a feasible version of $\widetilde{F}_n(\cdot|\cdot)$ where the probability integral transformation $F_X(\cdot)$ on the covariate makes its marginal density equal to 1 along its whole support.

The estimator of $\xi(p|x_0)$ based on $\widehat{F}_n(y|x_0)$ is defined as the generalized inverse of $\widehat{F}_n(\cdot|x_0)$:

$$\widehat{\xi}(p|x_0) = \widehat{F}_n^{-1}(p|x_0). \quad (16)$$

In the rest of this section, we provide primitive conditions under which Assumptions (GI) and (GB) hold for $\widehat{F}_n(y|x_0)$ or $\widehat{F}_n^{-1}(p|x_0)$ so the generic CI in (3) and the generic CB in (12) are applicable to $\widehat{F}_n^{-1}(p|x_0)$.

3.1.1 A New Confidence Interval

Our new level $(1 - \alpha)$ -confidence interval for $F_{Y|X}^{-1}(p|x_0)$ takes the following form:

$$\text{CIN}_{1-\alpha} = \left[\widehat{F}_n^{-1}(p - z_{\alpha/2}\sigma_{np}(K)|x_0), \widehat{F}_n^{-1}(p + z_{\alpha/2}\sigma_{np}(K)|x_0) \right], \quad (17)$$

⁸To illustrate its symmetry, suppose there are three observations, $X_1 = 1.5$, $X_2 = 2$, and $X_3 = 5$, and we are interested in estimating the conditional functional at $x_0 = 3$ using effectively two observations. Then the standard k -NN estimator would choose (X_1, X_2) according to Euclidean distance, whereas the symmetrized k -NN estimator would pick (X_2, X_3) based on rank.

where

$$\sigma_{np}(K) = \sqrt{\frac{R(K)p(1-p)}{nh_n}} \quad (18)$$

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

Recall that $q_p(x) = 1/f_{Y|X}(F_{Y|X}^{-1}(p|x)|x)$ is the conditional quantile density function of Y given X . It is obvious from (17) that our new confidence interval (CI), $\text{CIN}_{1-\alpha}$, has several advantages over existing CIs. First, compared with Wald-type confidence intervals, our new confidence interval, $\text{CIN}_{1-\alpha}$, 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 $\text{CIN}_{1-\alpha}$.

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

(i) Assume that

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

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

$$|H'(y|u_1) - H'(y|u_2)| \leq L|u_1 - u_2|$$

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

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

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). The conditional density function $f_{Y|X}(\cdot|x_0)$ exists and is continuous and positive on the interval $[F_{Y|X}^{-1}(p_1|x_0) - \epsilon, F_{Y|X}^{-1}(p_2|x_0) + \epsilon]$ for some $\epsilon > 0$, where $[p_1, p_2]$ contains p and belongs to $(0, 1)$, also X has continuous distribution function $F_X(x)$.

Assumption (S) is chosen in accordance with Assumptions (A), (B) in Stute (1986). For (S)(i)⁹ is used by Stute (1986) to show the tightness of the conditional empirical process. It is pretty

⁹To clarify some notation, we added the corresponding quantile transformation since Stute (1986) directly works with (X, Y) with uniform marginal distributions.

weak; if the term in absolute values could be bounded by any polynomial order in τ , it would imply (S)(i). In Fan and Liu (2011), we show that existence of copula density function of Y and X would imply this result. (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), 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 $nh_n^3 \rightarrow \infty$, 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 $CIN_{1-\alpha}$ is non-empty. Kernel functions satisfying Assumption (K) include Bisquare and Triweight kernels.

THEOREM 3.1 *Suppose Assumptions (X), (S), (K), and (H) hold and x_0 is an interior point not on the flat part of F_X . For $0 < \alpha < 1$, we get: $\Pr\left(F_{Y|X}^{-1}(p|x_0) \in CIN_{1-\alpha}\right) \rightarrow 1 - \alpha$ as $n \rightarrow \infty$.*

Given our generic result in Section 2, the proof of this theorem follows immediately after Lemma 3.2 below and as $\widehat{F}_n(y|x_0)$ is a proper distribution function by Assumption (K), the rearranging step could be skipped. 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 3.2 *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:

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

(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 3.2 (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 3.2 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 3.2 (ii) follows from Lemma 2.3 (i), Lemma 21.3 in van der Vaart (1998), and the functional Delta method. It implies that for a fixed $p \in [p_1, p_2]$,

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

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 $f_{Y|X}(F_{Y|X}^{-1}(p|x_0)|x_0)$ which our new confidence interval avoids as well.

3.1.2 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

$$\text{CBN}_{1-\alpha} = \left[\widehat{F}_n^{-1}(p - c_{n\delta}(\alpha, K) \sigma_{np}(K) | x), \widehat{F}_n^{-1}(p + c_{n\delta}(\alpha, K) \sigma_{np}(K) | x) \right], \quad (19)$$

where $\sigma_{np}(K)$ is defined in (18) and

$$c_{n\delta}(\alpha, K) = \frac{c(\alpha)}{(2\delta \log n)^{1/2}} + d_n \quad (20)$$

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

$$d_n = (2\delta \log n)^{1/2} + (2\delta \log n)^{-1/2} \log \left[\frac{\int (K'(u))^2 du}{4\pi R(K)} \right]. \quad (21)$$

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 ($\tilde{\mathbf{X}}$). The conditional density function $f_{Y|X}(y|x)$ has bounded derivative with respect to y uniformly for $x \in \mathcal{J}$. Notice Assumption (X) plus the compactness of \mathcal{J} would give 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 $n \rightarrow \infty$; (ii) $\inf_{x \in \mathcal{J}} f_{Y|X}(F_{Y|X}^{-1}(p|x)|x) >$

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 A7 and A8.

THEOREM 3.3 *Suppose Assumptions (B), (\hat{S}) , (\hat{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 3.4 *Under Assumptions (B), (\hat{S}) , (\hat{X}) , (K), and (H), it holds that*

$$\Pr \left((2\delta \log n)^{1/2} \left[\sigma_{np}^{-1}(K) \sup_{x \in \mathcal{J}} \left\{ f_{Y|X} \left(F_{Y|X}^{-1}(p|x) | x \right) \left| \hat{F}_n^{-1}(p|x) - F_{Y|X}^{-1}(p|x) \right| \right\} - d_n \right] \leq z \right) \rightarrow \exp(-2 \exp(-z)) \text{ as } n \rightarrow \infty.$$

Remark 3.1. The above lemma follows from Theorem 2.2 in Hardle and Song (2010) when the covariate is uniformly distributed between $[0, 1]$. The detailed proof consists of a characterization of maximal deviation of the conditional CDF estimator, as in our Assumption (GB)(iii) and a uniform Bahadur representation of the conditional quantile estimator in terms of the conditional CDF estimators evaluated at $F_{Y|X}^{-1}(p|x)$.

3.2 Semiparametric Quantile Models

In most applications, the covariate is multivariate. Semiparametric quantile models are introduced in the literature to alleviate the curse of dimensionality associated with fully nonparametric models and at the same time are more robust than fully parametric 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 in Section 3.1 to both partial linear and single index quantiles.

3.2.1 Partial Linear Quantile Model

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

$$\xi(p|Z_i, X_i) = Z_i' \beta_0(p) + g_p(X_i)$$

where $\{Y_i, X_i, Z_i\}_{i=1}^n$ is a random sample. We note that the p -th conditional (on $X_i = x$) quantile function of $(Y_i - Z_i' \beta_0(p))$ is $g_p(x)$.

Root- n consistent estimators of $\beta_0(p)$ are available in Lee (2003) and Song, Ritov, and Hardle (2012). Semiparametric efficient estimation of the above model has been studied by Lee (2003). Due to the root- n consistency of the finite dimensional parameter and additive structure, the nonparametric component will play the dominating role in the inference for the whole conditional quantile function. Denote the random variable $\tilde{Y} = Y - Z' \beta_0(p)$, and the Wald type inference requires estimation of the conditional density $f_{\tilde{Y}|X}(g_p(x)|x)$, see Song, Ritov and Hardle (2012). Utilizing the observation that $F_{\tilde{Y}|X}^{-1}(p|x) = g_p(x)$, we would use estimates of $F_{\tilde{Y}|X}^{-1}(p \pm \cdot |x)$ to construct a CI for $g_p(x)$ from the direct approach avoiding the estimation of $f_{\tilde{Y}|X}(g_p(x)|x)$. Specifically let $\hat{\beta}(p)$ denote a root- n consistent estimator of $\beta_0(p)$. For $x_0 \in \mathcal{X}$ and $z_0 \in \mathcal{Z}$, let

$$\hat{F}_{n,PL}(y|x_0; p) = \frac{\sum_{i=1}^n 1\{Y_i - Z_i' \hat{\beta}(p) \leq y\} K\left(\frac{F_n(x_0) - F_n(X_i)}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{F_n(x_0) - F_n(X_i)}{h_n}\right)}. \quad (22)$$

Our CI for $g_p(x_0)$ is of the form:

$$\left[\hat{F}_{n,PL}^{-1}(p - z_{\alpha/2} \sigma_{np}(K) | x_0; p), \hat{F}_{n,PL}^{-1}(p + z_{\alpha/2} \sigma_{np}(K) | x_0; p) \right]$$

and that for the conditional quantile $\xi(p|z_0, x_0) = [z_0' \beta_0 + g_p(x_0)]$ is defined as:

$$\text{CIPL}_{1-\alpha} = \left[\begin{array}{l} z_0' \hat{\beta}(p) + \hat{F}_{n,PL}^{-1}(p - z_{\alpha/2} \sigma_{np}(K) | x_0; p), \\ z_0' \hat{\beta}(p) + \hat{F}_{n,PL}^{-1}(p + z_{\alpha/2} \sigma_{np}(K) | x_0; p) \end{array} \right], \quad (23)$$

where $\sigma_{np}(K)$ is defined in (18).

We now introduce two assumptions.

Assumption (Z1). Assumptions (S) and (X) hold for (\tilde{Y}, X) , where $\tilde{Y} = Y - Z' \beta_0(p)$.

Assumption (PL). Z_i has a finite absolute conditional (on X_i) 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 3.5 Suppose $\widehat{\beta}(p) - \beta_0(p) = O_p(n^{-1/2})$ 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

$$CBPL_{1-\alpha} = \left[\begin{array}{l} z' \widehat{\beta}(p) + \widehat{F}_{n,PL}^{-1}(p - c_{n\delta}(\alpha, K) \sigma_{np}(K) | x; p), \\ z' \widehat{\beta}(p) + \widehat{F}_{n,PL}^{-1}(p + c_{n\delta}(\alpha, K) \sigma_{np}(K) | x; p) \end{array} \right], \quad (24)$$

where $\sigma_{np}(K)$ is defined in (18) and $c_{n\delta}(\alpha, K)$ is defined in (20).

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

THEOREM 3.6 Suppose $\widehat{\beta}(p) - \beta_0(p) = O_p(n^{-1/2})$ 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 $z \in \mathcal{K}$.

Remark 3.2. Let $\widehat{\xi}_{PL}(p|z_0, x_0) = z_0' \widehat{\beta}(p) + \widehat{F}_{n,PL}^{-1}(p|x_0; p)$. Under mild conditions, one can show that $\widehat{\xi}_{PL}(p|z_0, x_0)$ is a consistent estimator of $\xi(p|z_0, x_0)$, so the generic CI in (3) would apply. However since $\widehat{\xi}_{PL}(p|z_0, x_0)$ may not be monotone in $p \in (0, 1)$, monotone rearrangement is needed in order to apply the generic CI to $\widehat{\xi}_{PL}(p|z_0, x_0)$. In contrast, by making use of the partially linear structure of the quantile function and applying the direct approach to $g_p(x_0)$, we are able to construct a computationally simpler CI for $\xi(p|z_0, x_0)$ which does not rely on any estimate of the conditional quantile density function.

3.2.2 Single Index Quantile Model

Consider the single index model with multivariate covariate Z : $\xi(p|Z_i) = g_p(Z_i' \beta_0(p))$, where $\{Y_i, Z_i\}_{i=1}^n$ is a random sample. Let $\tilde{X} = Z' \beta_0(p)$ and observe $F_{Y|\tilde{X}}^{-1}(p|\tilde{x}) = g_p(z' \beta_0(p))$. In order to implement the Wald inference one has to estimate the conditional density function $f_{Y|\tilde{X}}(g_p(\tilde{x})|\tilde{x})$. In contrast, our CI and CB avoid the need to estimate the aforementioned conditional density function.

Let $\widehat{\beta}(p)$ denote a consistent estimator of $\beta_0(p)$ such as that in Wu, Yu, and Yu (2010) or Kong and Xia (2012), based on structural adaptive estimation methods. For $z_0 \in \mathcal{Z}$, let

$$\widehat{F}_{n,SI}(y|z_0; p) = \frac{\sum_{i=1}^n 1\{Y_i \leq y\} K\left(\frac{\widehat{F}_n(z_0' \widehat{\beta}(p)) - \widehat{F}_n(Z_i' \widehat{\beta}(p))}{h_n}\right)}{\sum_{i=1}^n K\left(\frac{\widehat{F}_n(z_0' \widehat{\beta}(p)) - \widehat{F}_n(Z_i' \widehat{\beta}(p))}{h_n}\right)}, \quad (25)$$

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

$$\text{CISI}_{1-\alpha} = \left[\widehat{F}_{n,\text{SI}}^{-1}(p - z_{\alpha/2} \sigma_{np}(K) | z_0; p), \widehat{F}_{n,\text{SI}}^{-1}(p + z_{\alpha/2} \sigma_{np}(K) | z_0; p) \right], \quad (26)$$

where $\sigma_{np}(K)$ is defined in (18).

We make the following assumptions.

Assumption (Z2). Assumptions (S) and (X) hold for (Y, \tilde{X}) with $\tilde{X} = Z' \beta_0(p)$. 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 3.7 *Suppose $\widehat{\beta}(p) - \beta_0(p) = O_p(n^{-1/2})$ and Assumptions (Z2), (K), and (HS) hold. Then the confidence interval, $\text{CISI}_{1-\alpha}$, achieves the nominal level $(1 - \alpha)$ asymptotically.*

Remark 3.3. The root- n asymptotic normality of the estimator $\widehat{\beta}(p)$ 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 (27) below for the single index quantile model.

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

THEOREM 3.8 *Suppose $\widehat{\beta}(p) - \beta_0(p) = O_p(n^{-1/2})$ and Assumptions ($\tilde{\text{Z}}2$), (K), and (HS) hold. Then the confidence band below is asymptotically valid with coverage probability $(1 - \alpha)$ uniformly over $z \in \mathcal{K}$:*

$$\text{CBSI}_{1-\alpha} = \left[\widehat{F}_{n,\text{SI}}^{-1}(p - c_{n\delta}(\alpha, K) \sigma_{np}(K) | z; p), \widehat{F}_{n,\text{SI}}^{-1}(p + c_{n\delta}(\alpha, K) \sigma_{np}(K) | z; p) \right], \quad (27)$$

where $\sigma_{np}(K)$ is defined in (18) and $c_{n\delta}(\alpha, K)$ is defined in (20).

Remark 3.2 for the partially linear model applies here, since $\widehat{\xi}_{\text{SI}}(p|z_0) \equiv \widehat{F}_{n,\text{SI}}^{-1}(p|z_0; p)$ may not be monotone in $p \in (0, 1)$.

4 Simulation

In this section, we investigate the finite sample performance of our new confidence intervals for non-parametric 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 Yang-Stute 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:

$$\text{W-NW}_{1-\alpha} = \left[\begin{array}{c} \widehat{F}_{n,NW}^{-1}(p|x) - \frac{z_{\frac{\alpha}{2}} \sigma_{np}(K) \widehat{q}_{p,NW}(x)}{\sqrt{\widehat{f}_X(x)}}, \\ \widehat{F}_{n,NW}^{-1}(p|x) + \frac{z_{\frac{\alpha}{2}} \sigma_{np}(K) \widehat{q}_{p,NW}(x)}{\sqrt{\widehat{f}_X(x)}} \end{array} \right], \quad (28)$$

$$\text{W-S}_{1-\alpha} = \left[\widehat{F}_n^{-1}(p|x) - z_{\alpha/2} \sigma_{np}(K) \widehat{q}_p(x), \widehat{F}_n^{-1}(p|x) + z_{\alpha/2} \sigma_{np}(K) \widehat{q}_p(x) \right], \quad (29)$$

$$\text{CI-NW}_{1-\alpha} = \left[\widehat{F}_{n,NW}^{-1} \left(p - \frac{z_{\alpha/2} \sigma_{np}(K)}{\sqrt{\widehat{f}_X(x)}} \middle| x \right), \widehat{F}_{n,NW}^{-1} \left(p + \frac{z_{\alpha/2} \sigma_{np}(K)}{\sqrt{\widehat{f}_X(x)}} \middle| x \right) \right], \quad (30)$$

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

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

$$\widehat{F}_{n,NW}(y|x) = \frac{\sum_{i=1}^n 1\{Y_i \leq y\} K\left(\frac{x-X_i}{h_{n,NW}}\right)}{\sum_{i=1}^n K\left(\frac{x-X_i}{h_{n,NW}}\right)}, \quad (31)$$

$$\widehat{f}_X(x) = \frac{1}{nh_X} \sum_{i=1}^n K\left(\frac{x-X_i}{h_X}\right), \text{ and } \widehat{f}_{Y|X}(y|x) = \frac{\sum_{i=1}^n K\left(\frac{y-Y_i}{h_{C,Y}}\right) K\left(\frac{x-X_i}{h_{C,X}}\right)}{h_{C,Y} \sum_{i=1}^n K\left(\frac{x-X_i}{h_{C,X}}\right)}, \quad (32)$$

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, $\text{W-NW}_{1-\alpha}$ and $\text{W-S}_{1-\alpha}$, are both Wald-type confidence intervals relying on a consistent estimator of the conditional quantile density function, $\text{W-S}_{1-\alpha}$ does not require a consistent estimator of the covariate density function $f_X(x)$. The two new confidence intervals,¹⁰ $\text{CIN}_{1-\alpha}$ and $\text{CI-NW}_{1-\alpha}$, make use of the conditional quantile estimators directly. They differ in the quantile estimators being used the consequence of which is that $\text{CIN}_{1-\alpha}$ does not depend on any density estimation, but $\text{CI-NW}_{1-\alpha}$ depends on a consistent estimator of the covariate density $f_X(x)$.

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

¹⁰Section 3 establishes the asymptotic validity of $\text{CIN}_{1-\alpha}$. The asymptotic validity of $\text{CI-NW}_{1-\alpha}$ can be established using Theorem 2.1.

function. In sharp contrast, the Wald-type confidence interval, $W\text{-}NW_{1-\alpha}$, is the most demanding, as there are four bandwidths involved.¹¹

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:

$$\begin{aligned} \text{Boot-Norm}_{1-\alpha} &= \left(\widehat{F}_{n,NW}^{-1}(p|x) - z_{\alpha/2}\sigma_{\text{Boot}}, \widehat{F}_{n,NW}^{-1}(p|x) + z_{\alpha/2}\sigma_{\text{Boot}} \right], \\ \text{Boot-Perc}_{1-\alpha} &= \left(\widehat{F}_{n,NW}^{-1}(p|x) - z_{\text{Boot},1-\alpha/2}, \widehat{F}_{n,NW}^{-1}(p|x) + z_{\text{Boot},\alpha/2} \right], \end{aligned} \quad (33)$$

where σ_{Boot} is the bootstrap standard deviation for $\widehat{F}_{n,NW}^{-1}(p|x)$ and $z_{\text{Boot},\alpha/2}$ is the bootstrap percentile.

In the tables below, we denote these confidence intervals, $W\text{-}NW_{1-\alpha}$, $W\text{-}S_{1-\alpha}$, $CI\text{-}NW_{1-\alpha}$, $CIN_{1-\alpha}$, $\text{Boot-Norm}_{1-\alpha}$ and $\text{Boot-Perc}_{1-\alpha}$ as ‘Asy NW’, ‘Asy CI’, ‘New NW’, ‘New CI’, ‘BootNm’, and ‘BootPerc’ respectively.

4.1 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:

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

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 replication¹² is set to be 500. The confidence interval, $W\text{-}NW_{1-\alpha}$, 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 $\widehat{f}_X(x)$ 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, $W\text{-}S_{1-\alpha}$, $CI\text{-}NW_{1-\alpha}$, 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

¹¹The 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 bandwidths are chosen to be the same.

¹²To ease the computational burden, we fixed the bandwidth for the bootstrap sample.

first one is based on the sample $(Y_i, F_n(X_i))$ after we transform X_i using its empirical distribution function.¹³ 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., 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, $CIN_{1-\alpha}$ and $CI-NW_{1-\alpha}$.¹⁴ The coverage rates for $CIN_{1-\alpha}$ are the same as in Table 3. Interestingly we observe that the coverage rate of $CI-NW_{1-\alpha}$ 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 direct approach and it is also worthwhile to avoid estimating covariate’s density function using $CIN_{1-\alpha}$. In comparison the sensitivity of Wald-type confidence intervals to the choice of bandwidths in the estimation of the conditional quantile density function is quite severe.

¹³We 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.

¹⁴The 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.

4.2 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):

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

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 (23). 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 $\hat{F}_{n,NW}(p|x)$, $\hat{F}_n(p|x)$, and $\hat{f}_{Y|X}(\hat{\xi}(p|x)|x)$ and also add $z'\hat{\beta}$ to both end points of the intervals in (28), (29), and (30). 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 performance is comparable and sometimes better than the $\text{Boot-Norm}_{1-\alpha}$ which performs better than $\text{Boot-Perc}_{1-\alpha}$ 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

5 Concluding Remarks

In this paper, we have constructed a generic confidence interval for the p -th conditional quantile from any preliminary conditional quantile estimator using the direct approach. We have shown that our generic CI is asymptotically valid for any quantile function (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. In the same spirit, we have also constructed generic confidence bands across a range of covariate values from conditional quantile estimators defined as the generalized inverse of conditional distribution estimators.

To further demonstrate the flexibility and simplicity of the direct approach, we have constructed complete “density-free” confidence intervals and bands for conditional quantiles based on the Yang-Stute estimator for nonparametric conditional quantile function with univariate covariate and two semiparametric conditional quantile functions with multivariate covariate. 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.

As far as we know, this paper is the first paper presenting a systematic study of the direct approach to inference in nonparametric and semiparametric quantile 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 quantiles. In semiparametric models, the finite dimensional parameter might be of interest, but this is beyond the scope of the current paper.

6 Appendix A. Technical Proofs For Section 3.1

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 $n \rightarrow \infty$ unless stated otherwise. We rely on the device to control the local oscillation of empirical process on various occasions, let

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

The following bounds hold almost surely in fact, but it is sufficient for our purpose in the present form.

Lemma A.1 (*Stute, 1982*) *Under 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: $\sqrt{n(h_n \log n)^{-1}} \sup_{|F_X(x) - F_X(X_i)| \leq Mh_n} |B_n(x, X_i)| = O_p(1)$.*

We will also make frequent use of (local) U-process theory, thus some notations and terminologies will be collected here from Nolan and Pollard (1987), Gine and Mason (2007) for easy reference.

We say a class of function \mathcal{F} is of VC type with respect to an envelope F if the covering number $N(\mathcal{F}, L_2(Q), \varepsilon)$, the smallest number of $L_2(Q)$ open balls of radius ε required to cover \mathcal{F} , satisfies

$$N(\mathcal{F}, L_2(Q), \varepsilon) \leq \left(\frac{M \|F\|_{L_2(Q)}}{\varepsilon} \right)^v, \text{ for } 0 < \varepsilon \leq 2 \|F\|_{L_2(Q)},$$

for some universal positive constant M, v for every probability measure Q on the underlying space.

For a kernel function f of k variables, we denote

$$U_k^n(f) = \frac{(n-k)!}{n!} \sum_{i \in I_n^k} f(X_{i_1}, \dots, X_{i_k}),$$

where $I_n^m = \{(i_1, \dots, i_m) : 1 \leq i_j \leq n, i_j \neq i_k \text{ if } j \neq k\}$. Now suppose f is symmetric in its entries, we have the well-known Hajek-Hoeffding decomposition stated as:

$$U_m^n(f) - Ef = \sum_{k=1}^m U_k^n(\Pi_k f),$$

where

$$\Pi_k f = (\delta_{x_1} - P) \times \dots \times (\delta_{x_k} - P) \times P^{m-k} f.$$

Moreover let σ^2 (which we call maximal variance) is any number satisfying

$$\|P^m f^2\|_{\mathcal{F}} \leq \sigma^2 \leq M^2.$$

The following lemma comes handy when we consider various local U-processes in the proofs, when $m = 1$ similar bound appears in Proposition 1 in Einmahl and Mason (2005) for local empirical process.

Lemma A.2 (*Gine and Mason, 2007, Theorem 8*) *Let \mathcal{F} be a collection of measurable symmetric functions $f : S^m \rightarrow \mathcal{R}$, bounded up by M in absolute values, and let P be any probability measure on (S, \mathcal{S}) . Assume \mathcal{F} is of VC type with envelope function $F \equiv M$ and with characteristics A and v . Then for every $m \in \mathcal{N}$, and $A \geq e^m, v \geq 1$, there exist constants C_1, C_2 , s.t. for any $k = 1, \dots, m$,*

$$n^k E \|U_k^n(\Pi_k f)\|_{\mathcal{F}}^2 \leq C_1 2^k \sigma^2 \left(\log \frac{A}{\sigma} \right)^k, \tag{A.1}$$

assuming $n\sigma^2 \geq C_2 \log \left(\frac{A}{\sigma} \right)$.

Recall

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

where we also define $\widehat{f}_U(x)$ and $\widetilde{f}_U(x)$ as

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

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 taking probability integral transformation on the covariate. In fact to show the validity of our new confidence band, we proceed by approximating $\widehat{F}_n(\cdot|\cdot)$ by $\widetilde{F}_n(\cdot|\cdot)$ uniformly and then resort to the results in Hardle (1989), Hardle and Song (2010) to get the desired limiting distribution. Before that, let us first characterize the bias term.

Lemma A.3 *Given Assumptions (\widetilde{S}) (ii) and (K), for any interior point $x \in \mathcal{X}$ and any $y \in \mathcal{Y}$, when $h_n \rightarrow 0$, 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

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

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 $U \in [0, 1]$. Now take the Taylor expansion and by Assumption (\widetilde{S}) (ii) 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**

The following sequence of lemmas is patterned after Serfling (1980) to get the uniform Bahadur representation in Lemma A.8, thereafter the validity of confidence band could be argued as when we present the generic results.

Lemma A.4 *Under Assumptions (H), (K), and (\widetilde{X}) , we have:*

$$\sup_{y \in \mathcal{Y}} \sup_{x \in \mathcal{J}} \sqrt{\frac{nh_n}{\log n}} \left| \widehat{F}_n(y|x) - \widetilde{F}_n(y|x) \right| = o_p(1). \quad (\text{A.2})$$

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

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

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

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

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,

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

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

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

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

Indeed the diagonal term is of smaller order, and now it suffices to consider the following U-process indexed by x : $II_n = h_n^{-2}U_2^n f_1 + s.o.$, where

$$\begin{aligned} & U_2^n f_1 \\ = & \frac{2}{n(n-1)} \sum_{i \neq j} \frac{1}{2} \left(\begin{aligned} & [1\{Y_i \leq y\} - F_{Y|X}(y|x)] K' \left(\frac{F_X(x) - F_X(X_i)}{h_n} \right) \begin{bmatrix} 1\{X_j \leq x\} - F_X(x) \\ -1\{X_j \leq X_i\} + F_X(X_i) \end{bmatrix} \\ & + [1\{Y_j \leq y\} - F_{Y|X}(y|x)] K' \left(\frac{F_X(x) - F_X(X_j)}{h_n} \right) \begin{bmatrix} 1\{X_i \leq x\} - F_X(x) \\ -1\{X_i \leq X_j\} + F_X(X_j) \end{bmatrix} \end{aligned} \right). \end{aligned}$$

Consider the following function class:

$$\mathcal{F}_1 = \left\{ \begin{aligned} & [1\{Y_i \leq y\} - F_{Y|X}(y|x)] K' \left(\frac{F_X(x) - F_X(X_i)}{h_n} \right) \times \\ & [1\{X_j \leq x\} - F_X(x) - 1\{X_j \leq X_i\} + F_X(X_i)] : y \in \mathcal{R}, x \in \mathcal{J}, h_n > 0 \end{aligned} \right\},$$

which could be written as produce of three sub-classes:

$$\begin{aligned} \mathcal{F}_{1,1} &= \{[1\{Y_i \leq y\} - F_{Y|X}(y|x)] : y \in \mathcal{R}, x \in \mathcal{J}\}, \\ \mathcal{F}_{1,2} &= \{K' \left(\frac{F_X(x) - F_X(X_i)}{h_n} \right) : x \in \mathcal{J}, h_n > 0\}, \\ \mathcal{F}_{1,3} &= \{[1\{X_j \leq x\} - F_X(x) - 1\{X_j \leq X_i\} + F_X(X_i)] : x \in \mathcal{J}\}. \end{aligned}$$

$\mathcal{F}_{1,1}$ is of VC type because $F_{Y|X}(y|x)$ is bounded and monotone along y -axis and Lipschitz continuous along x -axis over the compact set \mathcal{J} . Because $K'(\cdot)$ has bounded variation due to the bounded second order derivative, $\mathcal{F}_{1,1}$ is also of VC type by Nolan and Pollard (1987). So is $\mathcal{F}_{1,3}$, noting $F_X(x)$ is uniformly continuous by Assumption (\tilde{X}) . we have

$$\log N(\mathcal{F}, L_2(Q), \varepsilon) \leq M \log \left(\frac{1}{\varepsilon} \right), \text{ for any measure } Q,$$

for any probability measure Q by Theorem 2.10.20 in Van der Vaart and Wellner (1996). Therefore, as the expectation of the U-statistic is zero and by the moment bound in Lemma A.2, we can approximate $U_2^n f_1$ by its Hajek-Hoeffding Projection with an error of order n^{-1} , i.e.,

$$\frac{1}{h_n^2} U_2^n f_1 = \frac{1}{h_n^2} U_1^n \Pi_1 f_1 + O_p \left(\frac{1}{nh_n^2} \right).$$

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

$$\begin{aligned}
\frac{1}{h_n^2} U_1^n \Pi_1 f_1 &= \frac{1}{nh_n^2} \sum_{j=1}^n \int [F_{Y|X}(y|X_j) - F_{Y|X}(y|x)] K' \left(\frac{F_X(x) - F_X(X_j)}{h_n} \right) \times \\
& [1\{F_X(X_j) \leq F_X(x)\} - F_X(x) - 1\{F_X(X_j) \leq F_X(X_i)\} + F_X(X_i)] dF_X(X_i) \\
&= \frac{1}{nh_n} \sum_{j=1}^n \int [H(y|U_i) - H(y|u)] [1\{U_j \leq u\} - u - 1\{U_j \leq U_i\} + U_i] dK \left(\frac{u - U_i}{h_n} \right) \\
&= \frac{1}{n} \sum_{j=1}^n \int [H(y|u - vh_n) - H(y|u)] [1\{U_j \leq u\} - u - 1\{U_j \leq u - vh_n\} + u - vh_n] dK(v) \\
&\leq Mh_n \int \sup_{|u-v| \leq h_n} \left| \frac{1}{n} \sum_{j=1}^n 1\{U_j \leq u\} - u - 1\{U_j \leq u - vh_n\} + u - vh_n \right| d|K(v)| \\
&= O_p \left(h_n \sqrt{\frac{\log n}{nh_n}} \right) = o_p \left(\sqrt{\frac{\log n}{nh_n}} \right),
\end{aligned}$$

where $u = F_X(x)$. Notice that $|H(y|U) - H(y|u)| \leq Mh_n$ for U satisfying $|U - u| \leq \frac{h_n}{2}$. 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

$$\begin{aligned}
&\sup_{y \in \mathcal{Y}} \sup_{x \in \mathcal{J}} |\widehat{F}_n(y|x) - \widetilde{F}_n(y|x)| \\
&= \left[O_p \left(h_n \sqrt{\frac{\log n}{nh_n}} \right) + O_p \left(\frac{\log n}{nh_n^2} \right) \right] + O_p \left(\sqrt{\frac{\log n}{nh_n}} \right) o_p(1) = o_p \left(\sqrt{\frac{\log n}{nh_n}} \right).
\end{aligned}$$

Q.E.D

Lemma A.5 *Under Assumptions (H), (K), (\check{X}), (\check{S})(ii), and (B)(iii), it holds that for any $\varepsilon_n = O \left(\sqrt{\frac{\log n}{nh_n}} \right)$, $\sup_{x \in \mathcal{J}} \left| \widehat{F}_n^{-1}(p + \varepsilon_n|x) - F_{Y|X}^{-1}(p|x) \right| = O_p \left(\sqrt{(nh_n)^{-1} \log n} \right)$.*

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

$$\sup_{y \in \mathcal{Y}} \sup_{x \in \mathcal{J}} \left| \widehat{F}_n(y|x) - F_{Y|X}(y|x) \right| = O_p \left(\sqrt{\frac{\log n}{nh_n}} \right). \quad (\text{A.3})$$

Hence,

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

Therefore we obtain

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

by the requirement on c_n and (A.3). Analogous argument shows that

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

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

Lemma A.6 *Under Assumptions (H), (K), (\tilde{X}), and (\tilde{S})(ii), uniformly in y and $x \in \mathcal{J}$ and for any $\eta_n = O_p \left(\sqrt{\frac{\log n}{nh_n}} \right)$, it holds that*

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

Proof. The proof follows that of Lemma A.3 closely, except that we have $1\{Y_i \leq y\}$ replaced by $1\{y < Y_i \leq y + \eta_n\}$ (say $\eta_n \geq 0$ w.l.o.g.) and index this new functional class by $f_2 \in \mathcal{F}_2$ incorporating additional parameter η_n . Everything works through straightforwardly up to the Hajek-Hoeffding projection. Now the projection becomes

$$\begin{aligned} & \frac{1}{h_n^2} U_1^n \Pi_1 f_2 = \frac{1}{nh_n} \sum_{j=1}^n \int [H(y + \eta_n|U) - H(y|U)] \times \\ & [1\{U_j \leq u\} - u - 1\{U_j \leq U_i\} + U_i] dK \left(\frac{u - U_i}{h_n} \right) + s.o. \\ & \leq \eta_n M \sup \int \left| \frac{1}{n} \sum_{j=1}^n 1\{U_j \leq u\} - u - 1\{U_j \leq u - vh_n\} + u - vh_n \right| d|K(v)| \\ & = o_p \left(\frac{\log n}{nh_n} \right), \end{aligned}$$

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

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

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

Q.E.D

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

$$\begin{aligned}
&\sup_{x \in \mathcal{J}} \sup_{y \in \mathcal{R}, |\bar{y}| \leq \varepsilon_n} \left| [\tilde{F}_n(y + \bar{y}|x) - \tilde{F}_n(y|x)] - [F_{Y|X}(y + \bar{y}|x) - F_{Y|X}(y|x)] \right| \\
&= O_p \left(\left(\frac{\log n}{nh_n} \right)^{3/4} \right),
\end{aligned}$$

where ε_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:

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

noting $\tilde{f}_U(x)$ converges to 1 uniformly with a rate of $\sqrt{\log n / (nh_n)}$. 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 existence of second order derivative $F_{Y|X}(\cdot|\cdot)$ along y -axis:

$$\begin{aligned}
&E \left[\frac{1}{h_n} [1\{Y_i \leq y + \bar{y}\} - 1\{Y_i \leq y\}] K \left(\frac{F_X(x) - F_X(X_i)}{h_n} \right) \right] - [F_{Y|X}(y + \bar{y}|x) - F_{Y|X}(y|x)] \\
&= \frac{1}{h_n} \int [F_{Y|X}(y + \bar{y}|X_i) - F_{Y|X}(y|X_i)] K \left(\frac{F_X(x) - F_X(X_i)}{h_n} \right) dF_X(X_i) - [F_{Y|X}(y + \bar{y}|x) - F_{Y|X}(y|x)] \\
&= \frac{1}{h_n} \int f_{Y|X}(y|X_i) \bar{y} K \left(\frac{F_X(x) - F_X(X_i)}{h_n} \right) dF_X(X_i) - f_{Y|X}(y|x) \bar{y} + O \left(\frac{\log n}{nh_n} \right) \\
&= O_p \left(\sqrt{\frac{\log n}{nh_n}} \times h_n^2 \right) = o_p \left(\left(\frac{\log n}{nh_n} \right)^{3/4} \right),
\end{aligned}$$

Hence it suffices to characterize the stochastic order of the following term uniformly:

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

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, we assume \bar{y} is nonnegative w.l.o.g.

$$\begin{aligned} &|\omega(x, y; h_n, \varepsilon_n)| \\ &\leq \frac{M}{h_n} \int_{F_Y(y)}^{F_Y(y)+L\bar{y}} \int \left| K\left(\frac{F_X(x) - F_X(X)}{h_n}\right) \right| d|\mathcal{C}_n(F_X(X), F_Y(Y)) - \mathcal{C}(F_X(X), F_Y(Y))| \\ &\leq \frac{M}{h_n} \int_{F_Y(y)}^{F_Y(y)+L\bar{y}} \int_{F_X(x)-Mh_n}^{F_X(x)+Mh_n} d|\mathcal{C}_n(u, v) - \mathcal{C}(u, v)|, \end{aligned}$$

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

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

Hence overall, we get:

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

Q.E.D

Lemma A.8 *Under Assumptions (H), (K), (\tilde{X}), (\tilde{S})(ii), and (B)(ii)(iii)(iv), we have*

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

for any $\varepsilon_n = O\left(\sqrt{\frac{\log n}{nh_n}}\right)$, where $R_n(x)$ satisfies: $\sup_{x \in \mathcal{J}} |R_n(x)| = O_p\left(\left((nh_n)^{-1} \log n\right)^{3/4}\right)$.

Proof. Setting $\eta_n = \widehat{F}_n^{-1}(p + \varepsilon_n|x) - F_{Y|X}^{-1}(p|x)$, we have the following successive approximations similar as proving Theorem 2.2

$$\begin{aligned}
& \widehat{F}_n \left(F_{Y|X}^{-1}(p|x) + \eta_n|x \right) - \widehat{F}_n \left(F_{Y|X}^{-1}(p|x) |x \right) \\
&= \widetilde{F}_n \left(F_{Y|X}^{-1}(p|x) + \eta_n|x \right) - \widetilde{F}_n \left(F_{Y|X}^{-1}(p|x) |x \right) + o_p \left(\frac{\log n}{nh_n} \right) \\
&= \left[F_{Y|X} \left(F_{Y|X}^{-1}(p|x) + \eta_n|x \right) - F_{Y|X} \left(F_{Y|X}^{-1}(p|x) |x \right) \right] + \Delta_n(x) + o_p \left(\frac{\log n}{nh_n} \right) \\
&= f_{Y|X}(\xi_p(x)|x) \eta_n + \Delta_n(x) + \Delta'_n(x) + o_p \left(\frac{\log n}{nh_n} \right) \tag{A.4}
\end{aligned}$$

where the first equality follows from Lemma A.6, the second from Lemma A.7, and

$$\begin{aligned}
\Delta_n(x) &= \widetilde{F}_n \left(F_{Y|X}^{-1}(p|x) + \eta_n|x \right) - \widetilde{F}_n \left(F_{Y|X}^{-1}(p|x) |x \right) \\
&\quad - \left[F_{Y|X} \left(F_{Y|X}^{-1}(p|x) + \eta_n|x \right) - F_{Y|X} \left(F_{Y|X}^{-1}(p|x) |x \right) \right], \\
\Delta'_n(x) &= \left[F_{Y|X} \left(F_{Y|X}^{-1}(p|x) + \eta_n|x \right) - F_{Y|X} \left(F_{Y|X}^{-1}(p|x) |x \right) \right] - f_{Y|X} \left(F_{Y|X}^{-1}(p|x) |x \right) \eta_n.
\end{aligned}$$

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 $y = F_{Y|X}^{-1}(p|x)$ (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.4) becomes $\left[p + \varepsilon_n - \widehat{F}_n \left(F_{Y|X}^{-1}(p|x) |x \right) \right]$. **Q.E.D**

Proof of Theorem 3.3. The following string of equalities shall be self-explaining:

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

Similarly,

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

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

Proof of Lemma 3.4. The proof follows from Lemma A.4 and Lemma A.8 with $\varepsilon_n = 0$. Therefore uniformly over $x \in \mathcal{J}$, we have

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

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 3.2

Because the constructions of both confidence intervals and bands only involve the finite dimensional parameter fixed at level p , we would drop the argument in terms of p for notational brevity. We first present a lemma used in the proof of Theorem 3.5.

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 $y \in \mathcal{Y} \subset \mathcal{R}$:*

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

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 $f_s = 1\{Y_i - Z'_i\beta \leq y\}$ and $f_s \in \mathcal{F}_s = \left\{ 1\{Y_i - Z'_i\beta \leq y\} : \beta \in \mathcal{B} \text{ and } y \in \mathcal{Y} \right\}$. \mathcal{F}_s is uniformly bounded and of VC type because f_s could be written as the composition of a monotone function $1\{Y_i \leq \cdot\}$ and a linear function $Z'_i\beta + y$, applying Lemma 2.6.18 (viii) in Van der Vaart and Wellner (1996). Hence we have the following entropy bound for any probability measure Q :

$$N(\mathcal{F}_s, L_2(Q), \varepsilon) \leq M \left(\frac{1}{\varepsilon}\right)^v$$

where M is a universal finite constant.

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

(iii) Let $\rho(s, t)$ be the usual Euclidean norm in \mathcal{R}^{d+1} , further denote $s = (\beta, y)$ and $t = (\bar{\beta}, \bar{y})$ and the conditional measure as $dQ_{\cdot|X}$. Then

$$\begin{aligned} P(f_{n,s} - f_{n,t})^2 &= \iint (f_s - f_t)^2 dQ_{\cdot|X} \frac{1}{h_n} K^2 \left(\frac{F_X(x_0) - F_X(X)}{h_n} \right) dF_X \\ &\leq M \iint [|y - \bar{y}| + |Z_1| |\beta_1 - \bar{\beta}_1| + \cdots + |Z_d| |\beta_d - \bar{\beta}_d|] dQ_{\cdot|X} \\ &\quad \frac{1}{h_n} K^2 \left(\frac{F_X(x_0) - F_X(X)}{h_n} \right) dF_X \\ &\leq M \cdot R(K) \rho(s, t) \end{aligned}$$

where we have used the fact that $|f_s - f_t|^2 \leq O(|y - \bar{y}| + |Z_1| |\beta_1 - \bar{\beta}_1| + \cdots + |Z_d| |\beta_d - \bar{\beta}_d|)$ and M is a finite constant. The last equality follows from assumption (PL) as Z has finite conditional (on X) absolute moment. Therefore $\sup_{\rho(s,t) < \delta_n} P(f_{n,s} - f_{n,t})^2 \rightarrow 0$ as $\delta_n \rightarrow 0$.

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

$$\log N_P(\mathcal{F}_{n,s}, L_2(Q), \varepsilon \|K\|_2) \leq \log N(\mathcal{F}_s, L_2(Q_{\cdot|X}), \sqrt{\varepsilon}) \leq M \log\left(\frac{1}{\varepsilon}\right)$$

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 3.5. Let

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

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 $(\widehat{F}_{n,\text{PL}}^{-1}(p - z_{\alpha/2} \sigma_{np}(K) | x_0) + z'_0 \widehat{\beta}, \widehat{F}_{n,\text{PL}}^{-1}(p + z_{\alpha/2} \sigma_{np}(K) | x_0) + z'_0 \widehat{\beta})$ is an asymptotically valid confidence interval for $[z'_0 \beta_0 + g(x_0)]$ 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

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

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

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

Again A_{n3} is the easiest term to handle by applying the bound of local oscillation on empirical process in Lemma A.1:

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

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 the diagonal term which is of smaller order:

$$\frac{1}{\sqrt{nh_n}} A_{n2} = \frac{1}{h_n^2} U_2^n [\widehat{f}_3 - f_3] + O_p\left(\frac{1}{nh_n^2}\right)$$

with symmetric kernel function

$$\begin{aligned} & f_3(\cdot, \cdot) \\ = & \frac{1}{2} \left(\begin{aligned} & \left[1\{Y_i - Z'_i \beta \leq y\} - 1\{Y_i - Z'_i \beta_0 \leq y\} \right] K' \left(\frac{F_X(x_0) - F_X(X_i)}{h_n} \right) \begin{bmatrix} 1\{X_j \leq x_0\} - F_X(x_0) \\ -1\{X_j \leq X_i\} + F_X(X_i) \end{bmatrix} \\ & + \left[1\{Y_j - Z'_j \beta \leq y\} - 1\{Y_j - Z'_j \beta_0 \leq y\} \right] K' \left(\frac{F_X(x_0) - F_X(X_j)}{h_n} \right) \begin{bmatrix} 1\{X_i \leq x_0\} - F_X(x_0) \\ -1\{X_i \leq X_j\} + F_X(X_j) \end{bmatrix} \end{aligned} \right). \end{aligned}$$

and we define \widehat{f}_3 by plugging $\widehat{\beta}$ into β . To see the functional class \mathcal{F}_3 containing f_3 is of VC type it suffices to consider the following three subclasses:

$$\begin{aligned} \mathcal{F}_{3,1} &= \{1\{Y_i - Z'_i \beta \leq y\} - 1\{Y_i - Z'_i \beta_0 \leq y\} : (\beta, y) \in \mathcal{R}^{d+1}\}, \\ \mathcal{F}_{3,2} &= \left\{ K' \left(\frac{F_X(x_0) - F_X(X_i)}{h_n} \right), h_n > 0 \right\}, \\ \mathcal{F}_{3,3} &= \{[1\{X_i \leq X_j\} - F_X(X_j)]\}, \end{aligned}$$

which are indeed uniformly bounded and of VC type recalling what have been shown in Lemma A.4 and Lemma B.1

Now by the moment bound in Lemma A.2 we have $\sup \frac{1}{h_n^2} |U_2^n f_3 - 2U_1^n \Pi_1 f_3| = O_p\left(\frac{1}{nh_n^2}\right) = o_p(1)$ under Assumption (H). Proceed as in Lemma A.4 we work with the projection term:

$$\begin{aligned}
& \frac{1}{h_n^2} 2U_1^n \left[\Pi_1 \widehat{f}_3 - \Pi_1 f_3 \right] \\
&= \frac{1}{nh_n^2} \sum_{j=1}^n \int E \left[[1\{Y_i - Z_i' \widehat{\beta} \leq y\} - 1\{Y_i - Z_i' \beta_0 \leq y\}] | X_i \right] K' \left(\frac{F_X(x_0) - F_X(X_i)}{h_n} \right) \times \\
&\quad [1\{F_X(X_j) \leq F_X(x_0)\} - F_X(x_0) - 1\{F_X(X_j) \leq F_X(X_i)\} + F_X(X_i)] dF_X(X_i) \\
&= \frac{1}{nh_n} \sum_{j=1}^n \int E \left[[1\{Y_i - Z_i' \widehat{\beta} \leq y\} - 1\{Y_i - Z_i' \beta_0 \leq y\}] | X_i \right] \times \\
&\quad [1\{U_j \leq u_0\} - u_0 - 1\{U_j \leq U_i\} + U_i] dK \left(\frac{u_0 - U_i}{h_n} \right) \\
&= \frac{1}{nh_n} \sum_{j=1}^n \int E \left[[1\{Y_i - Z_i' \widehat{\beta} \leq y\} - 1\{Y_i - Z_i' \beta_0 \leq y\}] | X_i \right] \times \\
&\quad [1\{U_j \leq u_0\} - u_0 - 1\{U_j \leq u_0 - vh_n\} + u_0 - vh_n] dK(v) \\
&= O_p\left(\frac{1}{\sqrt{n}}\right) o_p\left(\frac{1}{\sqrt{nh_n}}\right).
\end{aligned}$$

When it comes to A_{n1} , we make use of Lemma B.1 which states that the class of functions $\mathcal{F}_{n,s}$ below is Donsker:

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

Let

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

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

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

where the third equality follows from $E \left[\sqrt{n} [1\{Y_i - Z_i' \widehat{\beta} \leq y\} - 1\{Y_i - Z_i' \beta_0 \leq y\}] | X_i \right] = O_p(1)$ as $\widehat{\beta} - \beta_0 = O_p\left(\frac{1}{\sqrt{n}}\right)$ by Assumption (PL) 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

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

So

$$\begin{aligned} & \Pr \left(z'_0 \beta_0 + g(x_0) \in \left(z'_0 \widehat{\beta} + \widehat{F}_{n,\text{PL}}^{-1} \left(p - z_{\alpha/2} \sigma_{np}(K) \mid x_0 \right), z'_0 \widehat{\beta} + \widehat{F}_{n,\text{PL}}^{-1} \left(p + z_{\alpha/2} \sigma_{np}(K) \mid x_0 \right) \right) \right) \\ = & \Pr \left(z'_0 \beta_0 + g(x_0) \in \left(z'_0 \widehat{\beta} + \widehat{F}_{n,\text{PL}}^{-1} \left(p - z_{\alpha/2} \sigma_{np}(K) \mid x_0 \right), z'_0 \widehat{\beta} + \widehat{F}_{n,\text{PL}}^{-1} \left(p + z_{\alpha/2} \sigma_{np}(K) \mid x_0 \right) \right) \right) + o(1) \\ \rightarrow & 1 - \alpha, \end{aligned}$$

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 $\widehat{\beta}$ in the single index model, uniformly for any z ,

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

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

Proof of Theorem 3.7. 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

$$\frac{1}{\sqrt{nh_n}} \left[\sum_{i=1}^n 1\{Y_i \leq y\} \left(K \left(\frac{F_n(z'_0 \widehat{\beta}) - F_n(Z'_i \widehat{\beta})}{h_n} \right) - K \left(\frac{F(z'_0 \beta_0) - F(Z'_i \beta_0)}{h_n} \right) \right) \right] = o_p(1). \quad (\text{B.3})$$

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 $(\widehat{F}_{n,\text{SI}}^{-1} \left(p - z_{\alpha/2} \sigma_{np}(K) \mid z_0 \right), \widehat{F}_{n,\text{SI}}^{-1} \left(p + z_{\alpha/2} \sigma_{np}(K) \mid z_0 \right))$ 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:

$$\begin{aligned} & \frac{1}{\sqrt{nh_n}} \left[\sum_{i=1}^n K \left(\frac{F_n(z'_0 \widehat{\beta}) - F_n(Z'_i \widehat{\beta})}{h_n} \right) - K \left(\frac{F(z'_0 \beta_0) - F(Z'_i \beta_0)}{h_n} \right) \right] \\ = & \frac{1}{\sqrt{nh_n}} \sum_{i=1}^n \frac{1}{h_n} K' \left(\frac{F(z'_0 \beta_0) - F(Z'_i \beta_0)}{h_n} \right) \left[F_n(z'_0 \widehat{\beta}) - F_n(Z'_i \widehat{\beta}) - F(z'_0 \beta_0) + F(Z'_i \beta_0) \right] \\ & + \frac{1}{\sqrt{nh_n}} \sum_{i=1}^n \frac{1}{h_n^2} K''(\Delta) \left[F_n(z'_0 \widehat{\beta}) - F_n(Z'_i \widehat{\beta}) - F(z'_0 \beta_0) + F(Z'_i \beta_0) \right]^2 \end{aligned}$$

According to Lemma 4.1 in Stute and Zhu (2005) and under the moment conditions in (Z2), it suffices to control the oscillation of the induced empirical process on $\left\{Z'_i\widehat{\beta}\right\}_{i=1}^n$ for observations locally around $z'_0\beta_0$ within a distance of order $n^{1/\gamma-1/2}$. Now use Lemma B.2 at point z_0 , under our assumptions we have,

$$\sup_{n^{1/2}\|\widehat{\beta}-\beta_0\|\leq M, n^{1/2-1/\gamma}|Z'_i\widehat{\beta}-z'_0\beta_0|\leq M} |F_n(z'_0\widehat{\beta})-F_n(Z'_i\widehat{\beta})-F(z'_0\beta_0)+F(Z'_i\beta_0)| = O_p\left(n^{-3/4+1/2\gamma}\sqrt{\ln n}\right).$$

Hence,

$$\begin{aligned} & \frac{1}{\sqrt{nh_n}} \sum_{i=1}^n \left| \frac{1}{h_n} K' \left(\frac{F(z'_0\beta_0) - F(Z'_i\beta_0)}{h_n} \right) \right| \cdot |F_n(z'_0\widehat{\beta}) - F_n(Z'_i\widehat{\beta}) - F(z'_0\beta_0) + F(Z'_i\beta_0)| \\ &= \frac{n^{-1/4+1/2\gamma}\sqrt{\ln n}}{\sqrt{h_n}} \left[\frac{1}{nh_n} \sum_{i=1}^n \left| K' \left(\frac{F(z'_0\beta_0) - F(Z'_i\beta_0)}{h_n} \right) \right| \right] = O_p\left(h_n^{-1/2}n^{-1/4+1/2\gamma}\sqrt{\ln n}\right) \\ &= o_p(1), \end{aligned}$$

where the second equality use the fact that $\frac{1}{nh_n} \sum_{i=1}^n \left| K' \left(\frac{F(z'_0\beta_0) - F(Z'_i\beta_0)}{h_n} \right) \right| = O_p(1)$ by standard kernel convergence result.

The last equality follows from the assumption of the bandwidth.

$$\begin{aligned} & \frac{1}{\sqrt{nh_n}} \sum_{i=1}^n \frac{1}{h_n^2} K''(\Delta) \left[F_n(z'_0\widehat{\beta}) - F_n(Z'_i\widehat{\beta}) - F(z'_0\beta_0) + F(Z'_i\beta_0) \right]^2 \\ &= O_p\left(\frac{1}{\sqrt{nh_n^{5/2}}} n^{-3/2+1/\gamma} \ln n\right) = O_p\left(h_n^{-5/2} n^{-1+1/\gamma} \ln n\right) = o_p(1). \end{aligned}$$

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**

We now sketch 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.3-A.8. 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.

Proof of Theorems 3.8. Similar to $\widehat{f}_U(x)$ and $\widetilde{f}_U(x)$ defined earlier, we introduce the following notations:

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

For Theorem 3.8, we have:

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

with $\widehat{x} = z'\beta_0$. The proof about switching from $F_n(Z'_i\widehat{\beta})$ to $F(Z'_i\beta_0)$ follows directly, since when we characterize the two smaller terms, the bound in Lemma B.2 holds uniformly in \widehat{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.

Proof of Theorem 3.6. First we take the following decomposition:

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

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

When it comes to P_{n3} , change occurs at the first order Taylor expansion term where the approximation of the U-statistic by the Hajek-Hoeffding projection holds uniformly in x , i.e., we need to incorporate the class $\mathcal{F}_{3,2} = \left\{ K' \left(\frac{F_X(x) - F_X(X_i)}{h_n} \right), x \in \mathcal{J}, h_n > 0 \right\}$ indexed by x as well now. As K'' exists and is bounded, hence K' has bounded variation, overall we still get the functional class of VC type and the U-statistic is approximated by Lemma A2 once we incorporate the additional factor $\sqrt{\log n}$. Now for the dominating term in P_{n1} and P_{n3} , everything boils down to show the

negligibility of

$$\begin{aligned} P_{n0} &= \frac{1}{h_n} \mathcal{P}_n f_{n,t} \\ &= \frac{1}{nh_n} \sum_{i=1}^n \left[1\{Y_i - Z_i' \beta \leq y\} - 1\{Y_i - Z_i' \beta_0 \leq y\} \right] K \left(\frac{F_X(x) - F_X(X_i)}{h_n} \right). \end{aligned}$$

when $\beta - \beta_0 = O(1/\sqrt{n})$. Unlike Lemma B.1 we no longer have Donsker property would not hold for the class $\mathcal{F}_{n,t}$ when x changes too

$$\mathcal{F}_{n,t} = \left\{ \begin{array}{l} f_{n,t} = \left[1\{Y_i - Z_i' \beta \leq y\} - 1\{Y_i - Z_i' \beta_0 \leq y\} \right] K \left(\frac{F_X(x) - F_X(X_i)}{h_n} \right) : \\ y \in \mathcal{R}, x \in \mathcal{J}, h_n > 0, \beta - \beta_0 = O(1/\sqrt{n}) \end{array} \right\}$$

Instead we resort to the Talagrand's inequality presented in Einmahl and Mason (2005). Let $\{\epsilon_i\}_{i=1}^n$ be a sequence of independent Rademacher r.v., independent of the sample $\{Y_i, Z_i, X_i\}_{i=1}^n$. Since $\sup E f_{n,t}^2 = O\left(\frac{1}{h_n \sqrt{n}}\right)$, hence Proposition in Einmahl and Mason (2005) or our Lemma A.2 gives

$$\frac{1}{nh_n} E \left\| \sum_{i=1}^n \epsilon_i f_{n,t}(Y_i, Z_i, X_i) \right\|_{\mathcal{F}_{n,t}} = O \left(\sqrt{\frac{\log n}{nh_n}} \times \frac{1}{n^{1/4}} \right).$$

What is more, $\frac{1}{h_n} \|(\mathcal{P}_n - \mathcal{P}) f_{n,t}\|_{\mathcal{F}_{n,t}}$ is of the same order of $\frac{1}{nh_n} E \|\sum_{i=1}^n \epsilon_i f_{n,t}(Y_i, Z_i, X_i)\|_{\mathcal{F}_{n,t}}$ by the concentration inequality in Einmahl and Mason (2005). Finally for any $f_{n,t}$, $\sqrt{\frac{nh_n}{\log n}} \mathcal{P} f_{n,t} = O\left(\sqrt{\frac{h_n}{\log n}}\right) = o(1)$ as in the proof of Theorem 3.5. Overall we have $P_{n0} = o_p\left(\sqrt{\frac{\log n}{nh_n}}\right)$ uniformly. **Q.E.D**

7 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:

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

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

following successive approximating processes:

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

where $\Gamma_n = \{|y| \leq a_n\}$ and $g(u) = E[\psi(y - \xi_p(F_X^{-1}(u))) \times 1\{|y| \leq a_n\} | U = u]$. $Z_n(\cdot, \cdot)$ 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), and Hardle (1989).

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

- [7] Chernozhukov, V., Fernandez-Val, I., and A. Galichon (2010), “Quantile and probability curves without crossing,” *Econometrica* 78, 1093-1125.
- [8] Chernozhukov, V., C. Hansen, and M. Jansson (2009), “Finite sample inference for quantile regression models,” *Journal of Econometrics* 152, 93-103.
- [9] Chesher, A. (2003), “Identification in nonseparable models,” *Econometrica* 71, 1405-1441.
- [10] Chaudhuri, P. (1991), “Nonparametric estimates of regression quantiles and their local Bahadur representation,” *The Annals of Statistics* 19, 760-777.
- [11] Chaudhuri, P., K. Doksum, and A. Samarov (1997), “On average derivative quantile regression,” *The Annals of Statistics* 25, 715-744.
- [12] Csorgo, M. (1983), *Quantile Processes with Statistical Applications*, SIAM, Philadelphia.
- [13] Dabrowska, D.M. (1987), “Non-parametric regression with censored survival time data,” *Scandinavian Journal of Statistics* 14, 181-197.
- [14] 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.
- [15] 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.
- [16] Donald, S. G., Yu-Chin Hsu, and G. F. Barrett (2012), “Incorporating covariates in the measurement of welfare and inequality: methods and applications,” *Econometrics Journal* 15, C1-C30.
- [17] Einmahl, U. and D.M. Mason (2005), “Uniform in bandwidth consistency of kernel-type function estimators,” *The Annals of Statistics* 33, 1380-1403.
- [18] Fan, J., T.-C. Hu, and Y.K. Truong (1994), “Robust non-parametric function estimation,” *Scandinavian Journal of Statistics* 21, 433-446.
- [19] Fan, Y. and R. Liu (2011), “Symmetrized multivariate k -NN estimators,” forthcoming in *Econometric Reviews*.
- [20] Fan, Y. and S. Park (2011), “Confidence intervals for the quantile of treatment effects in randomized experiments,” *Journal of Econometrics* 167, 330-344.
- [21] Figueroa-Lopez, J. (2011), “Sieve-based confidence intervals and bands for Levy densities,” *Bernoulli*, 17, 643-670.

- [22] Firpo, S. (2007), “Efficient semiparametric estimation of quantile treatment effects,” *Econometrica* 75, 259–276.
- [23] Gine, E., and D.M. Mason (2007), “On local U-statistic processes and the estimation of densities of functions of several sample variables,” *The Annals of Statistics* 35, 1105-1145.
- [24] 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.
- [25] 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
- [26] Goh, S.C. and K. Knight (2009), “Nonstandard quantile-regression inference,” *Econometric Theory* 25, 1415-1432.
- [27] 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.
- [28] Hall, P., Wolff, R., and Q. Yao (1999), “Methods of estimating a conditional distribution function,” *Journal of the American Statistical Association*, 94, 154-163.
- [29] Hardle W.K. (1989), “Asymptotic maximal deviation of M-smoothers,” *Journal of Multivariate Analysis* 29, 163–179.
- [30] Hardle W.K. and S. Song (2010), “Confidence bands in quantile regression,” *Econometric Theory* 26, 1180-1200.
- [31] Holderlein, S. and E. Mammen (2007), “Identification of marginal effects in nonseparable models without monotonicity,” *Econometrica* 75, 1513-1518.
- [32] Honoré, B., S. Khan, and J. Powell (2002), “Quantile regression under random censoring,” *Journal of Econometrics* 64, 241-278.
- [33] He, X. (1997), “Quantile curves without crossing,” *The American Statistician* 51, 186-191.
- [34] He, X. and F. Hu (2002), “Markov chain marginal bootstrap,” *Journal of American Statistical Association* 97, 783-795.
- [35] He, X., P. Ng, and S. Portnoy (1998), “Bivariate quantile smoothing splines,” *Journal of the Royal Statistical Society, Series B* 60, 537-550.
- [36] Horowitz, J. (1998), “Bootstrap methods for median regression models,” *Econometrica* 66, 1327-1352.

- [37] Kaplan, D.M. (2013), “IDEAL inference on conditional quantiles via interpolated duals of exact analytic L-statistics,” working paper.
- [38] Kocherginsky, M., X. He, and Y. Mu (2005), “Practical confidence intervals for regression quantiles,” *Journal of Computational and Graphical Statistics* 14, 41-55.
- [39] Koenker, R. (2005). *Quantile Regression*. Cambridge University Press.
- [40] Koenker, R. and G. Bassett (1978), “Regression quantiles,” *Econometrica* 46, 33-50.
- [41] Koenker, R. and Z. Xiao (2002), “Inference on the quantile regression process,” *Econometrica* 70, 1583-1612.
- [42] Koenker, R. and Z. Xiao (2004), “Unit root quantile autoregression inference,” *Journal of American Statistical Association* 99, 775-787.
- [43] Kong, E. and Y. Xia (2012), “A single-index quantile regression model and its estimation,” forthcoming in *Econometric Theory*.
- [44] Kong, E., O. Linton, and Y. Xia (2013), “Global Bahadur representation for nonparametric censored regression quantiles and its applications,” forthcoming in *Econometric Theory*.
- [45] Leadbetter, M.R., G. Lindgren and H. Rootzen (1983), *Extremes and Related Properties of Random Sequences and Processes*, Springer, New York.
- [46] Lee, S. (2003), “Efficient semiparametric estimation of a partially linear quantile regression model,” *Econometric Theory* 19, 1-31.
- [47] 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.
- [48] Liu, W. and W.B. Wu (2010), “Simultaneous nonparametric inference of time series,” *The Annals of Statistics* 38, 2388-2421.
- [49] Marmer, V. and A. Shneyerov (2012), “Quantile-Based Nonparametric Inference for First-Price Auctions,” *Journal of Econometrics* 167(2), 345-357.
- [50] Nolan, D. and D. Pollard (1987), “U-processes: rates of convergence,” *The Annals of Statistics* 15, 780-799.
- [51] Ostu, T. (2008), “Conditional empirical likelihood estimation and inference for quantile regression models,” *Journal of Econometrics* 142, 508–538.

- [52] Polonik, W. and Q. Yao (2002), “Asymptotics of set-indexed conditional empirical processes based on dependent data,” *Journal of Multivariate Analysis* 80, 234-255.
- [53] Portnoy, S. (2012), “Nearly root-n approximation of regression quantile processes,” *The Annals of Statistics* 40, 1714-1736.
- [54] Powell, J. L. (1986), “Censored regression quantiles,” *Journal of Econometrics* 32, 143-155.
- [55] Serfling, R.J. (1980), *Approximation Theorems of Mathematical Statistics*, Wiley, New York.
- [56] Song, S., Y. Ritov, and W.K. Hardle (2012), “Partial linear quantile regression and bootstrap confidence bands,” *Journal of Multivariate Analysis* 107, 244-262.
- [57] Stute, W. (1982), “The oscillation behavior of empirical processes,” *The Annals of Probability* 10, 86-107.
- [58] Stute, W. (1984a), “The oscillation behavior of empirical processes: the multivariate case,” *The Annals of Probability* 12, 361-379.
- [59] Stute, W. (1984b), “Asymptotic normality of nearest neighbor regression function estimates,” *The Annals of Statistics* 12, 917-926.
- [60] Stute, W. (1986), “Conditional empirical processes,” *The Annals of Statistics* 14, 638-647.
- [61] Stute, W. and L.-X. Zhu (2005), “Nonparametric checks for single-index models,” *The Annals of Statistics* 33, 1048-1083.
- [62] Su, L. and H. White (2011), “Conditional independence specification testing for dependent processes with local polynomial quantile regression,” forthcoming in *Advances in Econometrics*.
- [63] Truong, Y.K. (1989), “Asymptotic properties of kernel estimators based on local medians,” *The Annals of Statistics* 17, 606-617.
- [64] 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.
- [65] Tsybakov, A. (2008), *Introduction to Nonparametric Estimation*, Springer, New York.
- [66] van der Vaart, A. W. (1998), *Asymptotic Statistics*, Cambridge University Press.
- [67] van der Vaart, A.W. and J.A. Wellner (1996), *Weak Convergence and Empirical Processes: With Application to Statistics*, Springer-Verlag, New York.

- [68] Whang, Y.J. (2006), “Smoothed empirical likelihood methods for quantile regression models,” *Econometric Theory* 22, 173-205.
- [69] Wu, W. B. (2005), “On the Bahadur representation of sample quantiles for dependent sequences,” *The Annals of Statistics* 33, 1934-1963.
- [70] Wu, T.Z., K. Yu, and Y. Yu (2010), “Single-index quantile regression,” *Journal of Multivariate Analysis* 101, 1607-1621.
- [71] Xu, K.-L. (2012), “Nonparametric inference for conditional quantiles of time series,” forthcoming in *Econometric Theory*.
- [72] Yang, S.S. (1981), “Linear functions of concomitants of order statistics with application to nonparametric estimation of a regression function,” *Journal of the American Statistical Association* 76, 658-662.
- [73] Yu, K. and M.C. Jones (1998), “Local linear quantile regression,” *Journal of the American Statistical Association* 93, 228-237.
- [74] 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	0.9584	0.931	0.8586
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	$p = 0.25$			$p = 0.5$			$p = 0.75$		
$\begin{pmatrix} x \\ z \end{pmatrix}$	$\begin{pmatrix} 0.25 \\ 0.5 \end{pmatrix}$	$\begin{pmatrix} 0.5 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 0.75 \\ 1.5 \end{pmatrix}$	$\begin{pmatrix} 0.25 \\ 0.5 \end{pmatrix}$	$\begin{pmatrix} 0.5 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 0.75 \\ 1.5 \end{pmatrix}$	$\begin{pmatrix} 0.25 \\ 0.5 \end{pmatrix}$	$\begin{pmatrix} 0.5 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 0.75 \\ 1.5 \end{pmatrix}$
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	$p = 0.25$			$p = 0.5$			$p = 0.75$		
$\begin{pmatrix} x \\ z \end{pmatrix}$	$\begin{pmatrix} 0.25 \\ 0.5 \end{pmatrix}$	$\begin{pmatrix} 0.5 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 0.75 \\ 1.5 \end{pmatrix}$	$\begin{pmatrix} 0.25 \\ 0.5 \end{pmatrix}$	$\begin{pmatrix} 0.5 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 0.75 \\ 1.5 \end{pmatrix}$	$\begin{pmatrix} 0.25 \\ 0.5 \end{pmatrix}$	$\begin{pmatrix} 0.5 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 0.75 \\ 1.5 \end{pmatrix}$
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	$p = 0.25$			$p = 0.5$			$p = 0.75$		
$\begin{pmatrix} x \\ z \end{pmatrix}$	$\begin{pmatrix} 0.25 \\ 0.5 \end{pmatrix}$	$\begin{pmatrix} 0.5 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 0.75 \\ 1.5 \end{pmatrix}$	$\begin{pmatrix} 0.25 \\ 0.5 \end{pmatrix}$	$\begin{pmatrix} 0.5 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 0.75 \\ 1.5 \end{pmatrix}$	$\begin{pmatrix} 0.25 \\ 0.5 \end{pmatrix}$	$\begin{pmatrix} 0.5 \\ 1 \end{pmatrix}$	$\begin{pmatrix} 0.75 \\ 1.5 \end{pmatrix}$
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