Nonparametric Estimation and Inference on Conditional Quantile Processes*

Zhongjun Qu†  Jungmo Yoon‡
Boston University  Claremont McKenna College

March 24, 2011; this version: June 17, 2011

Abstract

We consider the estimation and inference about a nonparametrically specified conditional quantile process. For estimation, a two-step procedure is proposed. The first step utilizes local linear regressions and maintains quantile monotonicity through simple inequality constraints. The second step involves linear interpolation between adjacent quantiles. When computing the estimator, the bandwidth parameter is allowed to vary across quantiles to adapt to the data sparsity. The procedure is computationally simple to implement and is feasible even for relatively large data sets. For inference, we first obtain a uniform Bahadur-type representation for the conditional quantile process. Then, we show that the estimator converges weakly to a continuous Gaussian process, whose critical values can be estimated via simulations by exploiting the fact that it is conditionally pivotal. Next, we demonstrate how to compute the optimal bandwidth, construct uniform confidence bands and test hypotheses about the quantile process. Finally, we examine the performance of the bandwidth selection rule and the uniform confidence bands through simulations.

Keywords: nonparametric quantile regression, monotonicity constraint, uniform Bahadur representation, uniform inference.

---

*We thank Adam McCloskey, Denis Tkachenko, seminar participants at University of Iowa, Boston University and 2011 Econometric Society summer meeting for useful comments and suggestions.
†Department of Economics, Boston University, 270 Bay State Rd., Boston, MA, 02215 (qu@bu.edu).
‡Robert Day School of Economics and Finance, Claremont McKenna College, 500 East Ninth Street, Claremont, California, 91711 (jyoon@cmc.edu).


1 Introduction

Models for conditional quantiles play an important role in econometrics and statistics. They complement the conventional conditional mean models, offering a broader view of relationships between variables than the mean effects. The parametric approach to conditional quantile modeling was pioneered by Koenker and Bassett (1978), while the nonparametric approach was developed in a series of works including Bhattacharya (1963), Stone (1977), Chaudhuri (1991), Koenker, Portnoy and Ng (1994), Fan, Hu and Truong (1994) and Yu and Jones (1998). Reviews of the related literature can be found in Koenker (2005).

One is often interested in examining a full family of quantiles to obtain a complete analysis of the stochastic relationships between variables. This motivation underlies the consideration of conditional quantile processes. A seminal contribution to this analysis is Koenker and Portnoy (1987), which established a uniform Bahadur representation and serves as the foundation for further developments in this area. More recently, Koenker and Xiao (2002) considered the issue of testing composite hypotheses about quantile regression processes using Khmaladzation (Khmaladze, 1981). Chernozhukov and Fernández-Val (2005) considered the same issue and suggested resampling as an alternative approach. Angrist, Chernozhukov and Fernández-Val (2006) established inferential theory in misspecified models. Their results can be used to study a wide range of issues, including but not restricted to (i) testing alternative model specifications, (ii) testing stochastic dominance, and (iii) detecting treatment effect significance and heterogeneity.

A main focus of the above literature has been on quantile processes in parametric quantile models. However, there are frequent occasions when parametric specifications fail, making more flexible nonparametric methods desirable. Relaxing parametric assumptions is also often practically feasible in view of the increasing availability of rich data sets in both econometrics and statistics. Such considerations motivate us to consider estimation and inference about conditional quantile processes in a nonparametric setting. Our first goal is to propose a simple to implement estimator and the second is to provide a tractable inferential theory.
that is useful for constructing uniform confidence bands as well as conducting hypothesis tests.

We propose a two-step procedure to estimate the conditional quantile process. The first step applies local linear regressions (Fan, Hu and Truong, 1994 and Yu and Jones, 1998) to a grid of quantiles and maintains quantile monotonicity through a set of simple inequality constraints. The constraints depend neither on the data (such as their support) nor on the model (such as the number of covariates), but only on the number of quantiles entering the estimation. The second step involves linear interpolation between adjacent quantiles, returning an estimate for the quantile process. This procedure has two useful features. First, the bandwidth parameter is allowed to vary across quantiles to adapt to the data sparsity. This is important because data are typically sparse near the tails of the conditional distribution. Second, the optimization can be written as a linear programing problem with linear inequality constraints and can be solved using the algorithm in Koenker and Ng (2005) without essential modification. Our experimentation suggests that the procedure is feasible even for relatively large sample sizes.

For inference, we obtain three sets of results. First, we derive a uniform Bahadur-type representation, generalizing Theorem 2.1 in Koenker and Portnoy (1987) to the local linear setting. While being of independent interest, this forms a key step in proving the subsequent results. Second, we prove that, under a certain rate condition on the quantile grid, the proposed estimator has the same first-order asymptotic distribution as when the inequality constraints are absent. Therefore, the constraints serve as a finite sample correction, having no first-order effect asymptotically. The limiting distribution is a continuous Gaussian process, whose critical values can be estimated via simulations by exploiting the fact that it is conditionally pivotal, drawing on the insight of Parzen, Wei and Ying (1994) and Chernozhukov, Hansen and Jansson (2009). We discuss in detail how to compute the optimal bandwidth, construct uniform confidence bands and conduct hypothesis tests. Finally, we establish a connection with results in Chernozhukov, Fernández-Val and Galichon (2010). Specifically, we show that an alternative estimator, obtained from the same two-step pro-
procedure but without inequality constraints, satisfies a functional central limit theorem, and can therefore be used as the basis for rearrangement. The limiting distribution we obtain is applicable to this rearranged estimator. Therefore, our results broaden the application of rearrangement to the nonparametric local linear regression context.

We evaluate the proposed methods using simulations and briefly summarize the results below. First, the bandwidth selection rule performs well: it is able to deliver a small (large) bandwidth when the curvature of the conditional quantile function is high (low). Second, the proposed estimator, the conventional quantile-by-quantile local linear estimator (i.e., Fan, Hu and Truong, 1994) and its rearranged version all perform similarly in terms of the integrated mean squared error criterion. This result is in line with our theoretical finding that they all share the same first-order asymptotic distribution. Third, the confidence band can have undercoverage because the bias term in the estimator can be difficult to estimate, a well known fact in the literature. To address this issue, we suggest a simple modification that allows for a more flexible bias correction. The resulting confidence band is asymptotically conservative. Simulation evidence shows that it has adequate coverage, even with small sample sizes, and that it is only mildly wider than the confidence band under the conventional bias correction.

This paper is related to Belloni, Chernozhukov and Fernández-Val (2011). A key difference between the two works lies in the approach undertaken. Belloni, Chernozhukov and Fernández-Val (2011) consider a series-based framework, where the conditional quantile function is modeled globally with a large number of parameters. (Consequently, their framework is also useful for linear quantile regressions with many covariates.) The current paper is based on local linear regressions, where the quantile function is modeled locally by a few parameters and the modeling complexity is governed by the bandwidth. The second difference is how the quantile monotonicity is achieved. Belloni, Chernozhukov and Fernández-Val (2011) apply monotonization procedures to a preliminary series-based estimator, while in our framework monotonicity enters directly into estimation via inequality constraints.

When viewed from a methodological perspective, the current paper is related to the fol-
ollowing two strands of literature. First, the estimation procedure is related to the literature on estimating nonparametric regression relationships subject to monotonicity constraints, a majority of which has focused on monotonicity as a function of the covariate. For example, Mammen (1991) considered an estimator consisting of a kernel smoothing step and an isotonisation step. He and Shi (1998) and Koenker and Ng (2005) considered smoothing splines subject to linear inequality constraints as a way to enforce monotonicity. In the current paper, the monotonicity constraint is with regards to quantiles, giving rise to a different type of estimator than those discussed above, and requiring different techniques for studying its statistical properties. Note that He (1997), Dette and Volgushev (2008), Bondell, Reich and Wang (2010) and Chernozhukov, Fernández-Val and Galichon (2010) considered monotonicity with respect to quantiles. The connection with their works is discussed later in the paper. Second, there is an active literature that studies uniform confidence bands for nonparametric conditional quantile functions. The closely related works are Härdle and Song (2010) and Koenker (2010). The former considered kernel-based estimators and obtained confidence bands using strong approximations. The latter paper considered additive quantile models analyzed with total-variation penalties and obtained confidence bands using Hotelling’s (1939) tube formula. These results are uniform in covariates but pointwise in quantiles. Therefore, their results and ours complement each other and, when jointly applied in practice, allow one to probe a broad spectrum of topics.

The paper is organized as follows. Section 2 introduces the issue of interest. Section 3 presents the estimator while Section 4 discusses computational implementation. Section 5 establishes the asymptotic properties of the estimator. Section 6 discusses bandwidth selection. Section 7 shows how to construct uniform confidence bands and conduct hypothesis tests on the conditional quantile process. Section 8 establishes the connection to rearrangement. Section 9 includes some Monte Carlo experiments and Section 10 concludes. All proofs are included in the two appendices, with Appendix A containing proofs of the main results and Appendix B some auxiliary lemmas.

The following notation is used. The superscript 0 indicates the true value. For a real-
valued vector $z$, $\|z\|$ denotes its Euclidean norm. $[z]$ is the integer part of $z$. $1(\cdot)$ is the indicator function. $\text{supp}(f)$ stands for the support of the function $f$. $D_{[0,1]}$ stands for the set of functions on $[0,1]$ that are right continuous and have left limits, equipped with the Skorohod metric. The symbols “$\Rightarrow$” and “$\to_p$” denote weak convergence under the Skorohod topology and convergence in probability, and $O_p(\cdot)$ and $o_p(\cdot)$ is the usual notation for the orders of stochastic magnitude.

2 The issue of interest

Let $(X, Y)$ be an $\mathbb{R}^{d+1}$-valued random vector, where $Y$ is a scalar response variable and $X$ is an $\mathbb{R}^d$-valued explanatory variable. Denote their densities by $f_{Y|X}(\cdot)$ and $f_X(\cdot)$, respectively. Also, denote the conditional cumulative distribution of $Y$ given $X = x$ by $F_{Y|X}(\cdot|x)$ and the conditional quantile function at $\tau \in (0, 1)$ by $Q(\tau|x)$, i.e.,

$$Q(\tau|x) = F_{Y|X}^{-1}(\tau|x) = \inf \left\{ s : F_{Y|X}(s|x) \geq \tau \right\}. $$

In this paper, $Q(\tau|x)$ is allowed to be a general nonlinear function of $x$ and $\tau$. We fix $x$ and treat $Q(\tau|x)$ as a process in $\tau$. Our goal is in estimating and conducting inference on $Q(\tau|x)$ for all $\tau \in \mathcal{T}$, where $\mathcal{T} = [\lambda_1, \lambda_2]$ with $0 < \lambda_1 \leq \lambda_2 < 1$. Here, $\mathcal{T}$ lies strictly within the unit interval in order to allow the conditional distribution of $Y$ to have unbounded support. Throughout the paper, we assume the availability of $\{(x_i, y_i)\}_{i=1}^n$, a sample of $n$ observations that are i.i.d as $(X, Y)$.

We present two examples to motivate the above issue of interest and to illustrate the applicability of the methods we propose. More discussion follows in Section 7.

Example 1 (Quantile treatment effect, QTE). QTE measures the effect of a treatment (a policy) on the distribution of the potential outcomes. Specifically, let $X = (D, Z)$, where $D$ is a binary policy variable and $Z$ are some covariates. Let $Q(\tau|d, z)$ denote the $\tau$-th conditional quantile of the potential outcome conditioned on the treatment $D = d$ and
controls $Z = z$. Then the QTE is defined as

$$Q(\tau|1, z) - Q(\tau|0, z).$$

One may be interested in examining (i) whether the treatment has a significant effect at some quantile, i.e., testing $H_0$: $Q(\tau|1, z) = Q(\tau|0, z)$ for all $\tau \in T$ against $H_1$: $Q(\tau|1, z) \neq Q(\tau|0, z)$ for some $\tau \in T$, and (ii) whether the policy effect is homogeneous, i.e., testing $H_0$: $Q(\tau|1, z) - Q(\tau|0, z) = \delta(z)$ for some $\delta(z)$ and for all $\tau \in T$ against the hypothesis that the preceding difference is quantile dependent. Koenker and Xiao (2002) and Chernozhukov and Fernández-Val (2005) developed inferential procedures for above hypotheses for parametric conditional quantile models. Results in the current paper will allow us to analyze these issues under a nonparametric setting. Note that, in practice, the set $T$ can be flexibly chosen to reflect the issues of interest. For example, if the policy target is the low part of the distribution, then we can choose $T = [\varepsilon, 0.5]$ with $\varepsilon$ being a small positive number.

**Example 2 (Conditional stochastic dominance).** Stochastic dominance is an important concept for the study of poverty and income inequality. Although a large part of the literature has focused on unconditional dominance (see McFadden, 1989 and Barrett and Donald, 2003), conditional dominance has also received interest recently, see Koenker and Xiao (2002), Chernozhukov and Fernández-Val (2005) and Linton, Maasoumi and Whang (2005). Specifically, let $Q_1(\tau|x)$ and $Q_2(\tau|x)$ be two conditional quantile functions associated with two conditional distributions. Then, Distribution One (weakly) first-order stochastically dominates Two at $x$ if and only if

$$Q_1(\tau|x) \geq Q_2(\tau|x) \quad \text{for all } \tau \in (0, 1).$$

Distribution One (weakly) second-order stochastically dominates Two at $x$ if and only if

$$\int_0^\tau Q_1(s|x)ds \geq \int_0^\tau Q_2(s|x)ds \quad \text{for all } \tau \in (0, 1).$$

---

1 This concept traces back to Lehmann (1975) and Doksum (1974), and recent works include Heckman, Smith, and Clements (1997), Abadie, Angrist, and Imbens (2002), Chernozhukov and Hansen (2005) and Firpo (2007), among others.
One may be interested in testing the above null against non-dominance alternatives. This can be achieved by letting \( \delta(\tau|x) = Q_1(\tau|x) - Q_2(\tau|x) \) and considering one sided Kolmogorov-Smirnov tests based on \( \delta(\tau|x) \) and its indefinite integral process. Note that the above three papers considered parametric conditional quantiles/mean models, while results in this paper will allow us to examine these hypotheses under a nonparametric setting.

3 The estimator

The estimation procedure is based on a family of local linear regressions. We therefore start by a brief review of the local linear fit; more detailed discussions can be found in Chaudhuri (1991), Fan, Hu and Truong (1994) and Yu and Jones (1998). The method assumes that \( Q(\tau|x) \) is a smooth function of \( x \) for a given \( \tau \in (0, 1) \) and exploits the following first-order Taylor approximation:

\[
Q(\tau|x_i) \approx Q(\tau|x) + (x_i - x)' \frac{\partial Q(\tau|x)}{\partial x}.
\] (1)

The local linear estimator of \( Q(\tau|x) \), denoted by \( \hat{\alpha}(\tau) \), is determined via

\[
(\hat{\alpha}(\tau), \hat{\beta}(\tau)) = \arg \min_{\alpha(\tau), \beta(\tau)} \sum_{i=1}^n \rho_\tau(y_i - \alpha(\tau) - (x_i - x)'\beta(\tau)) K \left( \frac{x_i - x}{h_{n,\tau}} \right),
\] (2)

where \( \rho_\tau(u) = u (\tau - I(u < 0)) \) is the check function, \( K \) is a kernel function and \( h_{n,\tau} \) is a bandwidth parameter that depends on \( \tau \). As demonstrated in Fan, Hu and Truong (1994), the local linear fitting is simple to implement, yet has several advantages over the local constant fit. In particular, the bias of \( \hat{\alpha}(\tau) \) is not affected by the value of \( f_X'(x) \) and \( \partial Q(\tau|x)/\partial x \) and is of the same order as in the interior at the boundary, thus not requiring edge modification. In addition, plug-in data-driven bandwidth selection is free of estimating the derivatives of the marginal density, therefore is relatively simple to implement. As will be seen later, these three features continue to hold for our estimator. Note that the results in Fan, Hu and Truong (1994) and Yu and Jones (1998) are pointwise in \( \tau \).

We propose the following two-step procedure to estimate the process \( Q(\tau|x), \tau \in \mathcal{T} \).
STEP 1. Partition $\mathcal{T}$ into a grid of equally spaced quantiles $\tau_1, \ldots, \tau_m$. The precise condition on $m$ will be stated later. Consider the following constrained optimization problem

$$\min_{\{\alpha(\tau_j), \beta(\tau_j)\}_{j=1}^m} \sum_{j=1}^m \sum_{i=1}^n \rho_\tau (y_i - \alpha(\tau_j) - (x_i - x)'\beta(\tau_j)) K\left(\frac{x_i - x}{h_{n,\tau_j}}\right)$$

subject to

$$\alpha(\tau_j) \leq \alpha(\tau_{j+1}) \text{ for all } j = 1, \ldots, m - 1.$$ 

Denote the estimates by $\tilde{\alpha}(\tau_1), \ldots, \tilde{\alpha}(\tau_m)$.

STEP 2. Linearly interpolate between the estimates to obtain an estimate for the entire process, i.e., for any $\tau \in \mathcal{T}$, compute

$$\tilde{\alpha}^* (\tau) = \gamma_n (\tau) \tilde{\alpha}(\tau_j) + (1 - \gamma_n (\tau)) \tilde{\alpha}(\tau_{j+1}) \quad \text{if} \quad \tau \in [\tau_j, \tau_{j+1}],$$

where $\gamma_n (\tau) = (\tau_{j+1} - \tau)/(\tau_{j+1} - \tau_j)$.

Remark 1 The set of constraints $\alpha(\tau_j) \leq \alpha(\tau_{j+1})$ $(j = 1, \ldots, m - 1)$ forms a necessary and sufficient condition for the monotonicity of the conditional quantiles. They depend neither on the data (such as their support) nor on the model (such as the number of covariates), but only on the number of quantiles entering the estimation. This is due to the local nature of the problem. To our knowledge, this is the first time such a local feature is explicitly exploited to maintain quantile monotonicity. Clearly, the above procedure can be applied to a different $x$ without essential modification.

Remark 2 In practice, solving (3) is essentially a special case of Koenker and Ng (2005), who provided algorithms for computing quantile regression estimates subject to general linear inequality constraints, and a computationally efficient algorithm can be constructed with little cost. The detail is provided in the next Section.

Remark 3 The linear interpolation step is motivated by Neocleous and Portnoy (2008). It permits obtaining a tractable inferential theory as presented later.
Remark 4 There exist other methods for ensuring quantile monotonicity. He (1997) exploited the structure of a location-scale model. Bondell, Reich and Wang (2010) considered quantile smoothing splines and showed that non-crossing can be imposed via inequality constraints on the knots. Dette and Volgushev (2008) obtained monotonic quantile curves by first estimating the conditional distribution function and then inverting it to obtain the quantiles. However, their results are pointwise in quantiles, and it remains an interesting open question whether they can lead to tractable inferential theory for the quantile process. Finally, Chernozhukov, Fernández-Val and Galichon (2010) proposed a simple and elegant framework to address the non-crossing issue using rearrangement. The connection with their work is discussed later in the paper.

4 Computation

The joint estimation problem in (3) can be solved by the Frisch-Newton method developed in Portnoy and Koenker (1997) and Koenker and Ng (2005) without essential modifications. Below, we summarize the key formulations for the ease of implementation.

Let \( w_{ij} = K\left((x_i - x)/h_{n,\tau_j}\right) \). Because \( w\rho_\tau(z) = \rho_\tau(wz) \) for any \( w \geq 0 \), (3) can be rewritten as

\[
\min_{\{\alpha(\tau_j), \beta(\tau_j)\}} \sum_{j=1}^{m} \sum_{i=1}^{n} \rho_\tau(w_{ij}y_i - w_{ij}\alpha(\tau_j) - w_{ij}(x_i - x)'\beta(\tau_j))
\]

subject to

\[
\alpha(\tau_j) \leq \alpha(\tau_{j+1}) \quad \text{for all} \ j = 1, \ldots, m - 1.
\]

To have a linear programming formulation of the above minimization problem, we define a few notations. Let \( y \) denote an \( n \)-vector of the dependent variable and \( X \) denote an \( n \times (d + 1) \) matrix of explanatory variables whose first column is a vector of ones. Let \( b_j = (\alpha(\tau_j), \beta(\tau_j)')' \) denote a \((d + 1)\)-vector of parameters and \( e \) denote an \( n \)-vector of ones. We define a diagonal matrix \( W_j \) with \((w_{1j}, \ldots, w_{nj})\) along the diagonal and let \( y_j^* = W_j y \).
and $X^*_j = W_j \mathbf{X}$. Then, (5) can be rewritten as

$$\min_{\{u_j, v_j, b_j\}_{j=1}^m} \sum_{j=1}^m (\tau_j e' u_j + (1 - \tau_j) e' v_j)$$

subject to

$$\begin{pmatrix} u_1 \\ u_2 \\ \vdots \\ u_m \end{pmatrix} - \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_m \end{pmatrix} + \begin{pmatrix} X^*_1 \\ \vdots \\ X^*_m \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} = \begin{pmatrix} y^*_1 \\ \vdots \\ y^*_m \end{pmatrix} \in \{O_{mn}\},$$

where $b_j \in \mathbb{R}^{d+1}$, $(u'_j, v'_j) \in \mathbb{R}^{2n}$ for $j = 1, \ldots, m$ and $O_{mn}$ denote an $mn$-vector of zeros.

The inequality constraints (6) can be written as

$$\begin{pmatrix} -s' \\ 0 \\ \vdots \\ 0 \end{pmatrix} + \begin{pmatrix} s' \\ 0 \\ \vdots \\ 0 \end{pmatrix} - \begin{pmatrix} 0 \\ \vdots \\ \vdots \end{pmatrix} \begin{pmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{pmatrix} \in \mathbb{R}^{m-1}_+$$

where a $(d+1)$-vector $s$ is defined by $s = (1, 0, \ldots, 0)'$. This is the conventional $Rb \geq r$ form of constraints when $b = (b'_1, \ldots, b'_m)'$, $R$ is the matrix in front of $b$, and $r = O_{m-1}$, an $(m-1)$-vector of zeros. To further simplify the notation, let $\tau = (\tau_1, \ldots, \tau_m)'$, $u = (u'_1, \ldots, u'_m)'$, $v = (v'_1, \ldots, v'_m)'$, $y^* = (y''_1, \ldots, y''_m)'$, and $X^* = \text{diag}(X^*_j)_{j=1}^m$, a matrix whose diagonal blocks are given by $X^*_j$. Then, in matrix notation, the above primal problem can be written as

$$\min_{u,v,b} \{(\tau \otimes e)' u + ((1 - \tau) \otimes e)' v | u - v + X^* b = y^*, Rb \in \mathbb{R}^{m-1}_+, (u', v') \in \mathbb{R}^{2mn}_+\}$$

As highlighted in Koenker and Ng (2005), the dual formulation can greatly reduce the effective dimensionality of the problem. Note that the dual constraint is

$$\begin{pmatrix} \tau \otimes e \\ (1 - \tau) \otimes e \\ O_{m(d+1)} \end{pmatrix} - \begin{pmatrix} I_{mn} & 0 \\ -I_{mn} & 0 \\ X' & R' \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \lambda_2 \end{pmatrix} \in \mathbb{R}^{2mn}_+ \times \{O_{m(d+1)}\}.$$
Therefore, the dual problem is

$$\max_\lambda \{y^*\lambda_1 + r^*\lambda_2 | X^*\lambda_1 + R\lambda_2 = 0, (\tau - 1) \otimes e \leq \lambda_1 \leq \tau \otimes e, \lambda_2 \in \mathbb{R}^{m-1}_+\}. $$

By the change of variables $a_1 = (1 - \tau) \otimes e + \lambda_1$ and $a_2 = \lambda_2$, we have

$$\max_a \{y^*a_1 + r^*a_2 | X^*a_1 + R^*a_2 = X^*((\tau - 1) \otimes e), a_1 \in [0, 1]^{mn}, a_2 \in \mathbb{R}^{m-1}_+\}. \quad (7)$$

Koenker and Ng (2005) showed how to solve (7) using interior methods based on Frisch’s log-barrier approach. See their Section 3 for details. In practice, the R package quantreg with the aid of SparseM can be used to compute the estimator efficiently.

5 Asymptotic properties

We impose the following assumptions, where $x$ stands for the evaluation point and $\delta(x)$ for some small open neighborhood of $x$.

**Assumption 1** The evaluation point $x \in \text{supp}(f_X)$ with $f_X$ being continuously differentiable at $x$ and $0 < f_X(x) < \infty$.

**Assumption 2** (i) $f_{Y|X}((\tau|x)|x)$ is Lipschitz continuous in $\tau$ over $T$. (ii) There exist some finite constants $f_L > 0$, $f_U > 0$ and $\epsilon > 0$, such that

$$f_L \leq f_{Y|X}(Q(\tau|s) + \eta|s) \leq f_U$$

for all $|\eta| \leq \epsilon, s \in \delta(x) \cap \text{supp}(f_X)$ and $\tau \in T$.

**Assumption 3** (i) $Q(\tau|x)$ and $\partial Q(\tau|x)/\partial \tau$ are finite and Lipschitz continuous in $\tau$ over $T$. (ii) The elements of $\partial^2 Q(\tau|s)/\partial s \partial s'$ are finite and Lipschitz continuous over the set $\{(s, \tau): s \in \delta(x) \cap \text{supp}(f_X) \text{ and } \tau \in T\}$. (iii) tr $(\partial^2 Q(\tau|x)/\partial x \partial x') > 0$ for all $\tau \in T$.

**Assumption 4** The kernel $K$ is compactly supported, bounded, having finite first-order derivatives and satisfying

$$K(\cdot) \geq 0, \int K(u)du = 1, \int uK(u)du = 0, \int uu'K(u)du = \mu_2(K)I_d,$$

where $\mu_2(K)$ is a positive scalar and $I_d$ is the $d$-by-$d$ identity matrix.
Assumption 5 The bandwidth $h_{n,\tau}$ satisfies $h_{n,\tau} = c(\tau)h_n$ with $h_n = O(n^{-1/(4+d)})$ and $nh_n^d \to \infty$, and $c(\tau)$ is Lipschitz continuous satisfying $0 < \underline{c} \leq c(\tau) \leq \overline{c} < \infty$ for all $\tau \in T$.

Assumption 1 is fairly standard. Assumption 2 imposes restrictions on the conditional density. Its part (i) requires the conditional density at $x$ to have finite first-order derivative with respect to $\tau$ over $T$. Its part (ii) concerns some neighborhood surrounding $x$ and $T$ and is used to ensure that local expansions of (3) are well behaved, so that the estimator will have the usual rate of convergence. Assumption 3 imposes restrictions on the conditional quantile process. Part (i) requires it and its derivative in $\tau$ to vary smoothly with respect to $\tau$. Part (ii) ensures that the error in the Taylor expansion (1) is uniformly small. Part (iii) implies that the curvature of $Q(\tau|x)$ in $x$ is not zero, therefore a linear model is inadequate. This part is needed only for the purpose of determining the bandwidth parameter.

Note that a common theme of Assumptions 2 and 3 is that they are local: they impose restrictions only on neighborhoods surrounding $x$ and $T$. For example, if $T = [0.5, 0.8]$, then the lower part of the conditional distribution and the extreme quantiles are left essentially unrestricted. The conditional densities can be unbounded or zero in such regions.

Assumption 4 permits univariate as well as multivariate kernels, although it rules out the ones with unbounded support such as the Gaussian kernel and ones with unbounded derivatives. Assumption 5 imposes restrictions on the set of bandwidths, in particular requiring them to vary smoothly across quantiles. This is needed to ensure stochastic equicontinuity. It is not restrictive, and it is satisfied by the optimal bandwidth that minimizes the asymptotic MSE; see the section on bandwidth selection. Note that for a fixed $\tau$, we use the same bandwidth for every coordinate of $x$. This greatly simplifies the expressions for the bias and variance, especially when $x$ is a boundary point. More general formulas allowing for different bandwidths can be obtained by applying similar arguments.

The main challenge in establishing the asymptotic property is that, due to the constraints, the estimates $\hat{\alpha}(\tau_j)$ for different $j$ can no longer be studied separately. Because we allow $m$ to grow with the sample size, the problem asymptotically involves an infinite number of inequality constraints. Our analysis takes an indirect approach, by first studying the
unconstrained estimates from (2) and then linking them to the constrained estimates in (3). Specifically, we sequentially examine the properties of the following three quantities:

\[ \hat{\alpha}(\tau) : \text{the solution of (2) at a given } \tau \in \mathcal{T}, \]

\[ \hat{\alpha}^*(\tau) : \text{the solution from the two-step procedure but without imposing the constraints,} \]

\[ \tilde{\alpha}^*(\tau) : \text{the solution from the two-step procedure.} \]

The results obtained in the intermediate steps are of independent interest as discussed below.

The asymptotic properties of the estimators are different for \( x \) lying in the interior of \( \text{supp}(f_X) \) and for \( x \) lying near the boundary. Motivated by Ruppert and Wand (1994), let \( \mathcal{E}_{x,\tau} = \{ z : h_{n,\tau}^{-1}(x - z) \in \text{supp}(K) \} \) and call \( x \) an interior point if \( \mathcal{E}_{x,\tau} \subset \text{supp}(f_X) \) for all \( \tau \in \mathcal{T} \). Otherwise, \( x \) will be called a boundary point. If \( x \) is a fixed point in the interior of \( \text{supp}(f_X) \), then \( x \) is an interior point for all large \( n \). Therefore, to study the boundary point case, we consider a sequence of points \( x = x^n \) converging to a point \( x_\beta \) on the boundary of \( \text{supp}(f_X) \) sufficiently rapidly so that \( x \) is a boundary point for all \( n \). Formally, we suppose \( x \) satisfies

\[ x = x_\beta + h_{n,\tau} z \quad \text{for some } \tau \in \mathcal{T} \text{ and some fixed } z \in \text{supp}(K). \]

We also define the following set which serves as the domain of integration:

\[ \mathcal{D}_{x,\tau} = \{ z \in \mathbb{R}^d : (x + h_{n,\tau} z) \in \text{supp}(f_X) \} \cup \text{supp}(K). \]

The next Assumption is made to avoid degeneracies in the boundary point case and is the same as in Ruppert and Wand (1994).

**Assumption 6** There is a convex set \( C \) with nonnull interior and containing \( x_\beta \) such that

\[ \inf_{x \in C} f(x) > 0. \]

**Theorem 1 (Uniform Bahadur representation)** Let Assumptions 1-6 hold, then the following results hold uniformly in \( \tau \in \mathcal{T} \).
1. If \( x \) is an interior point, then

\[
\sqrt{nh_{n,\tau}^d} (\hat{\alpha}(\tau) - Q(\tau|x) - d_{\tau}h_{n,\tau}^2) = \frac{(nh_{n,\tau}^d)^{-1/2} \sum_{i=1}^{n} (\tau - 1(u_i^0(\tau) \leq 0) K_{i,\tau}}{f_X(x) f_{Y|x}(Q(\tau|x)|x)} + o_p(1),
\]

where

\[
d_{\tau} = \frac{1}{2} \mu_2(K) \text{tr} \left( \frac{\partial^2 Q(\tau|x)}{\partial x \partial x'} \right), K_{i,\tau} = K \left( \frac{x_i - x}{h_{n,\tau}} \right), u_i^0(\tau) = y_i - Q(\tau|x_i).
\]

2. If \( x \) is a boundary point, then

\[
\sqrt{nh_{n,\tau}^d} (\hat{\alpha}(\tau) - Q(\tau|x) - d_{b,\tau}h_{n,\tau}^2)
\]

\[
= \frac{\iota_1' N_x(\tau)^{-1} (nh_{n,\tau}^d)^{-1/2} \sum_{i=1}^{n} (\tau - 1(u_i^0(\tau) \leq 0) z_{i,\tau} K_{i,\tau}}{f_X(x) f_{Y|x}(Q(\tau|x))} + o_p(1),
\]

where \( \iota_1 = (1, 0_d)', u \in \mathbb{R}^d, \bar{u} = (1, u)', \)

\[
N_x(\tau) = \int_{D_{x,\tau}} \bar{u} u' K(u) du, d_{b,\tau} = \frac{1}{2} \iota_1' N_x(\tau)^{-1} \int_{D_{x,\tau}} u' \frac{\partial^2 Q(\tau|x)}{\partial x \partial x'} u \bar{u} K(u) du.
\]

The result generalizes Theorem 2.1 in Koenker and Portnoy (1987) to the nonparametric setting. The proof faces two difficulties that do not arise in the parametric context. First, because the quantile function is approximated using a Taylor expansion, the effect of the remainder term needs to be explicitly accounted for. Second, we allow the bandwidth to vary across quantiles, which poses challenges for establishing stochastic equicontinuity. The main ingredients used in the proof are the decomposition of Knight (1998) and a chaining argument as in Bickel (1975). The maximal inequalities developed in Bai (1996) and Oka and Qu (2011) are also useful for our analysis. We note that this result is of independent interest because of its direct connection to inference on conditional quantile processes.

The above Theorem implies that the limit of \( \hat{\alpha}(\tau) \) is driven by the leading term in the Bahadur representation. This paves the way for establishing the weak convergence of \( \hat{\alpha}^*(\tau) \), stated below.

**Theorem 2 (Weak convergence of the unconstrained estimator)** Let Assumptions 1-6 hold and let

\[
\hat{\alpha}^*(\tau) = \gamma_n(\tau) \hat{\alpha}(\tau_j) + (1 - \gamma_n(\tau)) \hat{\alpha}(\tau_{j+1}) \quad \text{if} \quad \tau \in [\tau_j, \tau_{j+1}],
\]

14
where $\gamma_n (\tau) = (\tau_{j+1} - \tau)/(\tau_{j+1} - \tau_j)$. Suppose $m/(nh_n^d)^{1/4} \to \infty$ as $n \to \infty$. Then:

1. If $x$ is an interior point, then
   \[
   \sqrt{n h_n^d} f_X (x) f_Y | X (Q(\tau| x)| x) \left( \hat{\alpha}^*(\tau) - Q(\tau| x) - d, h_n^2 \right) \Rightarrow G (\tau),
   \]
   where $G (\tau)$ is a zero mean Gaussian process defined over $T$ with continuous sample path and covariance given by
   \[
   E (G (r) G (s)) = (\kappa (r) \kappa (s))^{-d/2} (r \land s - rs) \int K \left( \frac{u}{\kappa (r)} \right) K \left( \frac{u}{\kappa (s)} \right) du,
   \]
   where
   \[
   \kappa (\tau) = \frac{h_{n,\tau}}{h_{n,1/2}} = \frac{c(\tau)}{c(1/2)}.
   \]

2. If $x$ is a boundary point, then
   \[
   \sqrt{n h_n^d} f_X (x) f_Y | X (Q(\tau| x)| x) \left( \hat{\alpha}^*(\tau) - Q(\tau| x) - d, h_n^2 \right) \Rightarrow G_b (\tau),
   \]
   where $G_b (\tau)$ is a zero mean Gaussian process defined over $T$ with continuous sample path and covariance given by
   \[
   E (G_b (r) G_b (s)) = (\kappa (r) \kappa (s))^{-d/2} (r \land s - rs) \iota_1 T_x (r,s) N_x (s)^{-1} \iota_1,
   \]
   where
   \[
   T_x (r,s) = \int_{D_x, h_n^{1/2}} \left[ \begin{array}{c}
   1 \\
   \frac{u}{\kappa (r)}
   \end{array} \right] K \left( \frac{u}{\kappa (r)} \right) \left[ \begin{array}{c}
   1 \\
   \frac{u'}{\kappa (r)}
   \end{array} \right] K \left( \frac{u}{\kappa (s)} \right) du
   \]
   and $\kappa (r)$ and $N_x (r)$ have the same definition as before.

**Remark 5** The scalar Gaussian process $G (\tau)$ depends only on the kernel and the bandwidth. In the extreme case where the bandwidths are equal across quantiles, it is simply the Brownian bridge (truncated to the interval $T$) multiplied by $(\kappa (r) \kappa (s))^{-d/2} \int K^2 (u) du$. Stute (1986, Theorem 1) obtained such a re-scaled Brownian bridge as the weak limit of a nearest-neighbor-type conditional empirical process. In the boundary case, the structure of the process is similar, except now it also depends on the location of $x$ relative to the boundary of the data support.
Remark 6 The rate condition $m/(nh_n^d)^{1/4} \to \infty$ ensures that the linear interpolation induces no loss of efficiency asymptotically, i.e., the limiting process is the same as if all quantiles were being estimated directly.

Remark 7 When formulating the result, we have purposely related the bandwidths at $\tau$ to the one at the median. This facilitates the discussion and computation of the bandwidth selection rule in the next Section.

The next result presents the asymptotic property of $\tilde{\alpha}^*(\tau)$.

Theorem 3 (Weak convergence of the constrained estimator) Let Assumptions 1-6 hold, then

1. $\sup_{1 \leq j \leq m} \sqrt{nh_{n,\tau_j}^d} |\tilde{\alpha}(\tau_j) - Q(\tau_j|x)| = O_p(1)$.

2. Assume $m/(nh_n^d)^{1/4} \to \infty$ but $m/(nh_n^d)^{1/2} \to 0$ as $n \to \infty$. Then,

$$\sqrt{nh_{n,\tau}^d f_X(x)f_{Y|X}(Q(\tau|x)|x)} \left( \tilde{\alpha}^*(\tau) - Q(\tau|x) - d_n\tau h_{n,\tau}^2 \right) \Rightarrow G(\tau)$$

if $x$ is an interior point and

$$\sqrt{nh_{n,\tau}^d f_X(x)f_{Y|X}(Q(\tau|x)|x)} \left( \tilde{\alpha}^*(\tau) - Q(\tau|x) - d_n\tau h_{n,\tau}^2 \right) \Rightarrow G_b(\tau)$$

if $x$ is a boundary point, where the relevant quantities have the same definitions as in the previous theorem.

Remark 8 The first result allows for arbitrary relationships between $m$ and $n$. The second result involves a new rate condition $m/(nh_n^d)^{1/2} \to 0$. This is used to ensure that the constraints (6) act as a finite sample correction, having no effect on the first-order limiting distribution. The adequacy of the limiting distributions under different $m$ will be evaluated using simulations. It turns out that the property of $\tilde{\alpha}^*(\tau)$ is rather insensitive to such choices, as long as $m$ is not too small. Motivated by this finding, we suggest the following simple rule

$$m = \max \left( 10, \sqrt{nh_n^d/\log(nh_n^d)} \right),$$

where the first argument safeguards against using too few quantiles when the sample size is relatively small and the second permits choosing a large number if the sample size is large.
6 Bandwidth selection

Theorems 1-3 imply that the estimator $\tilde{\alpha}^*(\tau)$ has the same limiting behavior as the estimator from the conventional quantile-by-quantile approach for any fixed $\tau \in \mathcal{T}$. Therefore, the same bandwidth selection rule can be applied. We state the result as a Corollary.

**Corollary 1 (Optimal bandwidth for an interior point).** Let Assumptions 1-5 and those stated in Theorem 3 hold. Then, the bandwidth that minimizes the (interior) asymptotic MSE of $\tilde{\alpha}^*(\tau)$ for any $\tau \in \mathcal{T}$ is given by

$$h_{n,\tau} = \left( \tau (1 - \tau) d \int K(v)^2 dv \left( \mu_2^2(K) f_X(x) f_{Y|X} (Q(\tau|x)|x) \right)^2 \left( \text{tr} \left( \frac{\partial^2 Q(\tau|x)}{\partial x \partial x'} \right) \right)^2 \right)^{1/(4+d)} n^{-1/(4+d)}. \quad (10)$$

In Appendix A, we also verify that the above bandwidth satisfies Assumption 5. The result implies that the bandwidth selection rule for estimating the conditional quantile process is conceptually no more difficult than in the conventional situation. The result also illustrates how the bandwidths change as $\tau$ shifts away from the center of the distribution. It typically widens as changes in $f_{Y|X} (Q(\tau|x)|x)$ often dominate other terms.

In practice, the challenges in computing the optimal bandwidth in the above Corollary are $\text{tr} (\partial^2 Q(\tau|x)/\partial x \partial x')$. In the simulations and empirical application, we have experimented with an approximation due to Yu and Jones (1998) which treats $\text{tr} (\partial^2 Q(\tau|x)/\partial x \partial x')$ as constant across quantiles. Specifically, under such an approximation, the optimal bandwidth at $\tau$ is related to the median via

$$\left( \frac{h_{n,\tau}}{h_{n,1/2}} \right)^{4+d} = 4\tau (1 - \tau) \left( \frac{f_{Y|X} (Q(\frac{1}{2}|x)|x)}{f_{Y|X} (Q(\tau|x)|x)} \right)^2. \quad (11)$$

Next, applying a Normal reference method (considering $f_{Y|X}$ to be a Gaussian density) as in Yu and Jones (1998), the above relationship simplifies to

$$\left( \frac{h_{n,\tau}}{h_{n,1/2}} \right)^{4+d} = \frac{2\tau (1 - \tau)}{\pi \phi(\Phi^{-1}(\tau))^2}, \quad (11)$$

where $\phi$ and $\Phi$ are the density and the cdf of a standard normal random variable. Finally, the bandwidth $h_{n,1/2}$ can be determined using the above Corollary.
7 Uniform confidence bands and hypothesis tests

7.1 Confidence bands

Corollary 2 (Uniform confidence band for the conditional quantile process). Let Assumptions 1-6 and those stated in Theorem 3 hold, then an asymptotic $(1 - p)$ confidence band for $Q(\tau|x)$ is given by

1. If $x$ is an interior point,

$$C_p = \left[ (\hat{\alpha}^*(\tau) - d_{\tau}h_{n,\tau}^2) - \sigma_{n,\tau}(x)Z_p, \quad (\hat{\alpha}^*(\tau) - d_{\tau}h_{n,\tau}^2) + \sigma_{n,\tau}Z_p \right],$$

where

$$\sigma_{n,\tau}(x) = \frac{1}{\sqrt{nh_{n,\tau}^d f_X(x)f_Y|X(Q(\tau|x)|x)}}$$

and $Z_p$ is the $(1-p)$-th percentile of $\sup_{\tau \in \mathcal{T}} |G(\tau)|$.

2. If $x$ is a boundary point,

$$C_p = \left[ (\hat{\alpha}^*(\tau) - db_{\tau}h_{n,\tau}^2) - \sigma_{n,\tau}(x)Z_p, \quad (\hat{\alpha}^*(\tau) - db_{\tau}h_{n,\tau}^2) + \sigma_{n,\tau}Z_p \right],$$

where $\sigma_{n,\tau}(x)$ is the same as above and $Z_p$ is now the $(1-p)$-th percentile of $\sup_{\tau \in \mathcal{T}} |G_b(\tau)|$.

The critical values of $\sup_{\tau \in \mathcal{T}} |G(\tau)|$ or $\sup_{\tau \in \mathcal{T}} |G_b(\tau)|$ can be consistently estimated via simulations. Consider the interior point case first. From Theorems 1 and 2,

$$\frac{(nh_{n,\tau}^d)^{-1/2} \sum_{i=1}^n (\tau - 1(u_i^0(\tau) \leq 0) K_{i,\tau}}{f_X(x)^{1/2}} \Rightarrow G(\tau).$$

Therefore, it suffices to simulate the left hand side process. The key observation is that, conditional on $\{x_i\}_{i=1}^n$, the numerator has the same finite sample distribution as

$$(nh_{n,\tau}^d)^{-1/2} \sum_{i=1}^n (\tau - 1(u_i - \tau \leq 0)) K_{i,\tau},$$

with $u_i$ being i.i.d. Uniform(0,1) and independent of $\{x_i\}_{i=1}^n$. This result follows from the Skorohod representation (letting $u_i = F_Y|X(y_i|x_i)$):

$$1(u_i^0(\tau) \leq 0) \Leftrightarrow 1(y_i - F_Y|X^{-1}(\tau|x_i) \leq 0) \Leftrightarrow 1(F_Y|X(x_i|x_i) - \tau \leq 0) \Leftrightarrow 1(u_i - \tau \leq 0)$$
and the independence of $\{(x_i, y_i)\}_{i=1}^n$. Let $\hat{f}_X(x)$ be a consistent estimate of $f_X(x)$, then

$$
\frac{(nh_{n,\tau}^d)^{-1/2} \sum_{i=1}^n (\tau - 1(u_i - \tau) \leq 0)) K_{i,\tau}}{f_X(x)^{1/2}} \Rightarrow G(\tau).
$$

(13)

This suggests the following simple procedure: (1) simulate the left hand side of (13) by drawing i.i.d. Uniform(0,1) random variables and keeping $K_{i,\tau}$ and $\hat{f}_X(x)$ fixed, (2) repeat the first step for a large number of replications, (3) compute the desired functionals and sort them to obtain critical values.

The same idea can be applied in the boundary point case to consistently estimate the critical values of $\sup_{\tau \in T} |G_b(\tau)|$. Specifically, from the second result in Theorems 1 and 2,

$$
\frac{t_1 N_x(\tau)^{-1} (nh_{n,\tau}^d)^{-1/2} \sum_{i=1}^n (\tau - 1(u_i^0(\tau) \leq 0)) z_{i,\tau} K_{i,\tau}}{f_X(x)^{1/2}} \Rightarrow G_b(\tau).
$$

Because (see Equation (A.11) in Appendix A)

$$
(nh_{n,\tau}^d)^{-1} \sum_{i=1}^n z_{i,\tau} z'_{i,\tau} K_{i,\tau} \Rightarrow f_X(x) N_x(\tau)
$$

(14)

and using the Skorohod representation, we have

$$
\frac{t_1 (nh_{n,\tau}^d)^{-1} \sum_{i=1}^n z_{i,\tau} z'_{i,\tau} K_{i,\tau}^{-1} (nh_{n,\tau}^d)^{-1/2} \sum_{i=1}^n (\tau - 1(u_i - \tau) \leq 0)) z_{i,\tau} K_{i,\tau}}{f_X(x)^{-1/2}} \Rightarrow G_b(\tau),
$$

(15)

where $\hat{f}_X(x)$ is a consistent estimate of $f_X(x)$. Therefore, we can simply simulate from the left hand side distribution. The resulting critical values are asymptotically valid even if $x$ is an interior point. Because (15) automatically accounts for the relative location of the evaluation point from the boundary, it allows us to easily handle situations where the support of $f_X$ may have a complicated shape.

The above procedure is inspired by Parzen, Wei and Ying (1994) and Chernozhukov, Hansen and Jansson (2009). The former paper exploited the conditional pivotal property to obtain a resampling method, while the latter used such a property to conduct finite sample inference for quantile regressions. In Belloni, Chernozhukov and Fernández-Val (2011), the above idea is generalized and applied to quantile models of increasing dimensions. The
application of the conditional pivotal property to the local regression setting and, more importantly, to provide inference for both interior and boundary points appears to be new.

Now consider estimating the bias. We examine directly the boundary case, as it again includes the interior point case as a special case. Note that from Theorem 1,

\[ d_{b,\tau} = \left( \frac{1}{\tau} \right) \{ f_X(x) N_x(\tau) \}^{-1} \left\{ \frac{1}{2} f_X(x) \int_{D_{x,\tau}} u' \frac{\partial^2 Q(\tau|x)}{\partial x \partial x'} u \left[ \frac{1}{u} \right] K(u) \, du \right\} \]

The term inside the first curly bracket can be consistently estimated by (14). The second curly bracket is the limit of (see Equation (A.14) in Appendix A)

\[ -h_{n,\tau}^{-2} (nh_{n,\tau}^{-d})^{-1} \sum_{i=1}^{n} e_i (\tau) z_i K_i, \]

where

\[ e_i (\tau) = -\frac{1}{2} \left( \frac{x_i - x}{h_{n,\tau}} \right)' \frac{\partial^2 Q(\tau|x)}{\partial x \partial x'} \left( \frac{x_i - x}{h_{n,\tau}} \right) h_{n,\tau}^2. \]

Therefore, \( d_{b,\tau} \) can be estimated using

\[ \left( (nh_{n,\tau}^{-d})^{-1} \sum_{i=1}^{n} z_i (\tau) z_i K_i \right)^{-1} \left( -h_{n,\tau}^{-2} (nh_{n,\tau}^{-d})^{-1} \sum_{i=1}^{n} e_i (\tau) z_i K_i \right). \]

In (16), the only unknown is \( \frac{\partial^2 Q(\tau|x)}{\partial x \partial x'} \). As suggested in the literature, this can be estimated using a global or local cubic regression, although caution needs to be exercised because the estimate can be imprecise. As before, (16) automatically accounts for the relative location of the evaluation point from the boundary, allowing us to easily handle situations where the support of \( f_X \) may have a complicated shape.

**Remark 9** We use the formulas (15) and (16) throughout in the simulation section when constructing the confidence band. Note that as a by-product, it is no longer necessary to estimate \( f_X(x) \). This is because it appears in \( \sigma_{n,\tau}(x) \) and (15) and cancels out. This feature is desirable because it can be challenging to directly estimate \( f_X(x) \) if the dimension of \( x \) is high.
Remark 10 It is well known that the bias term in the estimator is difficult to estimate. Consequently, the confidence bands can have undercoverage. Therefore, some modifications that reflect the estimation uncertainty are desirable. We suggest the following simple modified confidence band, where the idea is to allow for, but do not force, a bias adjustment: (Consider the interior point case; the boundary point case can be handled in the same way.)

\[
\left[ \left( \hat{\alpha}^*(\tau) - d^+ \sigma^2_n(z) \right) - \sigma_n(x)Z_p, \left( \hat{\alpha}^*(\tau) - d^- \sigma^2_n(z) \right) + \sigma_n(x)Z_p \right],
\]

(17)

where

\[
d^+ = \max(d, 0) \quad \text{and} \quad d^- = \min(d, 0).
\]

This modified band has the same or higher coverage as in the case of not making any bias adjustment (i.e., by setting \(d\) to zero, as often done in the literature), which is preferable when the bias is small. It also has the same or higher coverage than the conventional band (12), which is preferable when the bias is large. Therefore, it has the flavor of a hybrid procedure. Here, an important concern is whether the band will be too wide to be informative. Our simulation evidence in Section 9 suggests otherwise. This is because when the curvature of the conditional function (\(d\)) is high, the proposed bandwidth selection rule can deliver a small bandwidth, therefore the value of \(d, \sigma_n(z)\) remains modest. Consequently, (17) is typically only mildly wider than (12)

7.2 Hypothesis tests

We illustrate testing hypotheses on quantile processes using the two examples in Section 2. Without loss of generality, assume the evaluation points are interior points. To be consistent with the notation used there, we let \(\hat{Q}(\tau|x)\) denote \(\hat{\alpha}^*(\tau)\) when the evaluation point is \(x\).

Example 1 (continued). Treatment significance can be tested using a Kolmogorov–Smirnov (KS) type test:

\[
\sup_{\tau \in \mathcal{T}} \sqrt{n h_n^d} \left| \hat{Q}(\tau|1, z) - \hat{Q}(\tau|0, z) \right|
\]
Let $x_1 = (1, z)$, $x_2 = (0, z)$, and define

$$G_x^{(j)}(\tau) = \frac{G^{(j)}(\tau)}{\sqrt{f_X(x) f_{Y|x}(Q(\tau|x)|x)}}$$

with $G^{(j)}(\tau)$ being independent copies of $G(\tau)$. Assume Assumptions 1-5 holds at $x_1$ and $x_2$ and $\partial^2 Q(\tau|x_1)/\partial x \partial x' = \partial^2 Q(\tau|x_2)/\partial x \partial x'$. Then, the test has the following null limiting distribution

$$\sup_{\tau \in T} \left| G^{(1)}_{x_1}(\tau) - G^{(2)}_{x_2}(\tau) \right|,$$

whose critical values can be consistently estimated by applying the simulation algorithm described earlier to the first and second components. The treatment homogeneity hypothesis can be tested using

$$\sup_{\tau \in T} \sqrt{nh^{d}_{n,\tau}} \left| \hat{Q}(\tau |1, z) - \hat{Q}(\tau |0, z) - \int_T \left\{ \hat{Q}(\tau |1, z) - \hat{Q}(\tau |0, z) \right\} d\tau \right|,$$

whose null limiting distribution, under the same assumptions as stated above, is given by

$$\sup_{\tau \in T} \left| G^{(1)}_{x_1}(\tau) - G^{(2)}_{x_2}(\tau) - \int_T \left\{ G^{(1)}_{x_1}(\tau) - G^{(2)}_{x_2}(\tau) \right\} d\tau \right|,$$

whose critical values can be obtained in a similar manner.

**Example 2 (continued).** The null of first-order conditional stochastic dominance can be tested against a non-dominance alternative using a signed version of the KS test:

$$\sup_{\tau \in T} \sqrt{nh^{d}_{n,\tau}} \left| \hat{Q}(\tau |1, z) - \hat{Q}(\tau |0, z) \right|.$$

Assume Assumptions 1-5 hold for both samples, and $\partial^2 Q_1(\tau|x)/\partial x \partial x' = \partial^2 Q_2(\tau|x)/\partial x \partial x'$, then under the least favorable null hypothesis, the test has the following limiting distribution

$$\sup_{\tau \in T} \left| G^{(1)}_{x_1}(\tau) - G^{(2)}_{x_2}(\tau) \right|.$$

The second-order dominance can be tested via

$$\sup_{\tau \in T} \sqrt{nh^{d}_{n,\tau}} \left| \int_\tau \left( \hat{Q}_1(s|x)d\tau - \hat{Q}_2(s|x) \right) d\tau \leq 0 \right| \int_\tau \left( \hat{Q}_1(s|x)d\tau - \hat{Q}_2(s|x) \right) d\tau,$$
then, under the least favorable null hypothesis, the test converges to
\[
\sup_{\tau \in \mathcal{T}} \left| \int_\varepsilon^\tau \left( \frac{1}{n} \left( G_x^{(1)}(\tau) - G_x^{(2)}(\tau) \right) d\tau \right) + \int_\varepsilon^\tau \left( G_x^{(1)}(\tau) - G_x^{(2)}(\tau) \right) d\tau \right|.
\]
In the above construction, the lower limit is some positive constant \( \varepsilon \) instead of zero. This allows the conditional distributions to have unbounded support, at the cost of possibly sacrificing some power if the main differences between two distributions lies in the lower tails. Again, the critical values can be consistently estimated via simulations.

8 Connection to rearrangement

Chernozhukov, Fernández-Val and Galichon (2010) proposed a generic framework for estimating monotone probability and quantile curves. Their method uses some preliminary estimator that is not necessarily monotonic and applies rearrangement to ensure monotonicity. Importantly, they showed that if the preliminary estimator satisfies a functional central limit theorem, say
\[
a_n \left( \hat{Q}(\tau|x) - Q(\tau|x) \right) \Rightarrow G_x(\tau) \text{ over } \tau \in \mathcal{T}
\]
with \( a_n \to \infty \) as \( n \to \infty \) and \( G_x(\tau) \) being a continuous stochastic process, then the rearranged estimator \( \hat{Q}^R(\tau|x) \) satisfies
\[
a_n \left( \hat{Q}^R(\tau|x) - Q(\tau|x) \right) \Rightarrow \tilde{G}_x(\tau) \text{ over } \tau \in \mathcal{T}
\]
provided \( Q(\tau|x) \) is strictly monotonic in \( \tau \) and continuously differentiable in both arguments. The next result shows that \( \hat{Q}(\tau|x) = \hat{\alpha}^*(\tau) \) suffices as such an estimator.

**Corollary 3** Suppose Assumptions 1-6 hold and \( h_{n,\tau} = c(\tau)n^{-1/(4+d)} \). Let \( \hat{Q}^R(\tau|x) \) be the rearranged version of \( \hat{\alpha}^*(\tau) \) with the latter defined in Theorem 2. Then, the results in Theorem 2 hold with \( \hat{\alpha}^*(\tau) \) replaced by \( \hat{Q}^R(\tau|x) \).

9 Monte Carlo experiments

We focus on three issues: (1) the performance of the bandwidth selection rule, (2) the performance of the proposed estimator relative to other estimators including the standard
local linear estimator and its rearranged version, and (3) the property of the confidence band.

We consider the following three models, each expressed in terms of their conditional quantile functions:

**Model 1**: $Q(\tau|x) = x_1 - x_2 + (0.5x_1 + 0.3x_2)Q_{e_1}(\tau)$.

**Model 2**: $Q(\tau|x) = (0.5 + 2x_1 + \sin(2\pi x_1 - 0.5)) + x_2Q_{e_1}(\tau)$.

**Model 3**: $Q(\tau|x) = \log(x_1x_2) + 1/(1 + \exp(-x_1Q_{e_1}(\tau) - x_2Q_{e_2}(\tau))) + x_2Q_{e_1}(\tau)$.

The regressors $x_1$ and $x_2$ are i.i.d. $U(0, 1)$. The error terms $e_1$ and $e_2$ are i.i.d. $N(0, 1)$ and $U(0, 1)$ respectively, with quantile functions given by $Q_{e_1}(\tau) = \Phi^{-1}(\tau)$ and $Q_{e_2}(\tau) = \tau$. Model 1 is a linear location-scale model. Model 2 is a similar model but with nonlinearity in the location. Model 3 is a fairly complicated nonlinear model. The data are simulated by first obtaining $n$ independent draws $u_i$ ($1 \leq i \leq n$) from the $U(0,1)$ distribution and then using $Y_i = Q(u_i|x_i)$ to generate $Y_i$. For instance, in Model 3, the data are generated via $Y_i = Q(u_i|x_i) = \log(x_1x_2) + 1/(1 + \exp(-x_1Q_{e_1}(u_i) - x_2Q_{e_2}(u_i))) + x_2Q_{e_1}(u_i)$ with $u_i$ being i.i.d. $U(0,1)$. Figure 1 depicts $Q(0.5|x)$ and $Q(0.8|x)$ for Models 2 and 3. It shows that these two models exhibit significant curvature in $x$, which poses substantial challenges for estimation.

Other aspects of the simulation design are as follows. We examine three evaluation points, $x = (0.5, 0.5), (0.75, 0.75)$, and $(0.9, 0.9)$. The latter can be regarded as a boundary point, since in a typical simulation run the size of bandwidth at that point is greater than 0.1. The sample sizes are $n = 250, 500$. Given that the experiment involves estimating the quantile process nonparametrically at $x = (0.9, 0.9)$, $n = 250$ should be viewed as a very small sample size. We set $T = [0.2, 0.8]$ unless stated otherwise. The number of quantiles (m) in the quantile grid is set to 10, 20 and 30 to examine the result sensitivity. The kernel function is the product of univariate Epanechnikov kernels. We also experimented with the product Gaussian kernel and obtained qualitatively and quantitatively similar results. They are omitted here to save space. All subsequent results are based on 500 simulation runs.
9.1 Performance of the bandwidth selection rule

We first provide some details on estimation. The optimal bandwidths given in Corollary 1 are estimated in three steps.

**STEP 1.** We obtain a set of pilot bandwidths in order to calculate relevant quantities in (10). This is done by first obtaining a pilot bandwidth for the local median regression, which is determined via leave-one-out cross validation, and then relating it to the other quantiles using (11) following Yu and Jones (1998).

**STEP 2.** We estimate \( \text{tr} \left( \partial^2 Q(\tau|x) / \partial x \partial x' \right) \) and \( f_{Y|X}(\cdot|x) \). For the former, we apply a local cubic polynomial regression with the bandwidth determined via leave-one-out cross validation. For the latter, we use

\[
\hat{f}_{Y|X}(z|x) = \int \frac{1}{h_{yx}} K_1 \left( \frac{z - y}{h_{yx}} \right) d\hat{F}(y|x),
\]

where \( h_{yx} \) is the bandwidth parameter and \( K_1 \) is a univariate Epanechnikov kernel, and \( \hat{F}(y|x) = \sup \{ \tau \in (0, 1) | \hat{Q}(\tau|x) \leq y \} \) with \( \hat{Q}(\tau|x) \) equal to our proposed two-step estimator \( \tilde{\alpha}^*(\tau) \) computed using the bandwidth in Step 1. To compute (18), we take \( \hat{F}(y|x) \), draw random samples from it, and apply kernel smoothing with the generated random variates. The bandwidth \( \hat{h}_{yx} \) is determined by Silverman’s rule-of-thumb method. In doing so, we find that some degree of over-smoothing is useful. We therefore use bandwidth \( c \hat{h}_{yx} \) with \( c = 2 \) for all the reported results. Under suitable regularity conditions, it can be shown that \( \hat{f}_{Y|X}(\cdot|x) \) converges uniformly to \( f_{Y|X}(\cdot|x) \) due to the uniform convergence of \( \hat{Q}(\cdot|x) \) using the arguments in Portnoy and Koenker (1989, Lemma 3.2).

**STEP 3.** We use the estimated quantities from Step 2 to compute the optimal bandwidth for the median using (10) and relate it to the bandwidths for other quantiles using (11).

The selected bandwidths are summarized in Table 1 (for \( n = 250 \)) and 2 (for \( n = 500 \)). The results for both cases are qualitatively similar. The procedure is able to adapt to different degrees of curvatures in \( Q(\tau|x) \). It delivers larger bandwidths for Model 1 relative to Model 2 and 3, which is desirable since Model 1 is linear in \( x \). Within a given Model, it can deliver different bandwidths at different design points. For example, in Model 2, the
selected bandwidths at (0.5, 0.5) are greater than at (0.75, 0.75). This is consistent with 
\( \text{tr} \left( \partial^2 Q(\tau|x)/\partial x \partial x' \right) \) being of greater magnitude at the latter point. Similar results hold for 
Model 3.

9.2 The relative performance of different estimators

We contrast the finite sample performance of the proposed estimator (labelled Proposed) 
with three alternatives. The first alternative is the classical estimator in Koenker and Bassett 
(1978), labelled as QR. This comparison is used to illustrate the gain or loss from considering 
a nonparametric estimator. The second alternative is the standard quantile-by-quantile 
local linear estimator (labelled Local linear), which is included to illustrate the effect of the 
monotonicity constraint. The third alternative is the rearranged version of the local linear 
estimator (labelled Rearrangement), as discussed in Section 8. This is used to gauge whether 
one approach is superior to another. For all the estimators, except QR, we use the same 
bandwidth selection rule described previously.

We report the RMISE averaged over 500 simulation runs, with

\[
RMISE = \sqrt{ \frac{1}{m} \sum_{j=1}^{m} |\hat{Q}(\tau_j|x) - Q(\tau_j|x)|^2 },
\]

where \( m \) is the number of quantile to be estimated. Table 3 reports the results for \( m = 30 \). We obtained similar findings with \( m = 10 \) and \( 20 \). They are omitted here to save space.

First, consider the proposed estimator and QR. When the underlying model is actually 
linear (Model 1), the proposed estimator pays only modest price in terms of efficiency. This is 
because the bandwidth selection rule proposed earlier is able to deliver a wide bandwidth in 
this case. When nonlinearity (Models 2 and 3) is present, the nonparametric estimator shows 
substantially superior performance overall. Therefore, the results are quite encouraging.

Second, consider the proposed estimator and the standard quantile-by-quantile local lin-
ear estimator. Note that the fraction of quantile crossings is listed below the local linear 
estimator. Quantile crossing is relatively infrequent at interior points, however it occurs 
frequently when the design point is close to the boundary. The proposed estimator often has
smaller RMISE, but the difference is always small. This is consistent with Theorem 3 that these two estimators are first-order asymptotically equivalent. Therefore, in our simulations, the monotonicity constraint serves as a way to ensure a coherent estimate, but not a way to substantially improve the precision in finite samples.

Third, consider the proposed estimator and the rearranged estimator. They have very similar RMISE, confirming the result in Corollary 3 that these two estimators are first-order asymptotically equivalent. Interestingly, the RMISE of the rearranged estimator tends to be slightly smaller. However, the difference is too small to prefer one estimator to another.

One might wonder whether the above conclusion depends on the criterion function (RMISE). To this end, we repeated the analysis by replacing RMISE with the integrated absolute deviation and the findings are quantitatively very similar. They are omitted to save space. Another issue is whether the difference between the estimators is substantially greater when the quantile range is increased. To this end, we repeated the analysis with $T = [0.05, 0.95]$. However, the results remain qualitatively the same. They are omitted to save space.

9.3 Properties of the uniform confidence band

We examine the following two issues: whether the modified confidence band (17) shows a significant improvement over the conventional one (12), and whether the improvement comes at the cost of a substantially wider band.

To obtain the confidence band, we need critical values of $\sup_{\tau \in T} |G (\tau)|$ and $\sup_{\tau \in T} |G_b (\tau)|$ and the biases $d_\tau$ and $d_{b, \tau}$. These are estimated using the method outlined in Section 7, where the conditional density is recomputed using (18) but with the new $\tilde{\alpha}^*(\tau)$ estimated using the optimal bandwidth.

Tables 4 and 5 contrast the coverage ratios under the two scenarios. The modification delivers adequate coverage for all models, sample sizes, quantile grids and design points. This includes the challenging situation of estimating the quantile process at $x = (0.9, 0.9)$ with $n = 250$. The most significant improvement comes under Model 1. For this model,
the selected bandwidths are large, therefore the bias adjustment plays a critical role. Since
the true bias is zero, the conventional bias correction can only have detrimental effects on
the coverage ratio. For the other two models, the improvement is more important when the
sample size is small or when the $x$ is close to the boundary of the data support.

Tables 6 and 7 summarize the relative width of the two confidence bands at the two
representative quantiles $\tau = 0.5, 0.8$. For each simulation replication, we compute the ratios
and then report their means and standard deviations over the 500 replications. Table 6
corresponds to $n = 250$. For Model 1, the means of the ratios are between 1.095 and 1.195;
for Model 2, between 1.061 and 1.157; for Model 3, between 1.081 and 1.153. Table 7
corresponds to $n = 500$. The corresponding ratios are between 1.085 and 1.179 for Model 1,
1.055 and 1.152 for Model 2, and 1.051 and 1.152 for Model 3. Therefore, the modification
only mildly increases the widths of the bands. Overall, the simulation suggests that the
modification is useful.

10 Conclusion

We have considered the estimation and inference about a nonparametrically specified condi-
tional quantile process. The estimation method is computationally simple to implement and
is practically feasible even for relatively large data sets. We obtained a uniform Bahadur
representation, a functional central limit theorem and provided practical procedures for con-
structing uniform confidence bands and testing hypothesis about the quantile process. We
also proposed a modified procedure for bias adjustment, which performed well in the simu-
lations.
References


Appendix A. Proof of Main Results

We first define some notations to be used in the two appendices. Let

\[ u_i^0(\tau) = y_i - Q(\tau|x_i), \]
\[ u_i(\tau) = y_i - \alpha(\tau) - (x_i - x)'\beta(\tau), \]

where \( \alpha(\tau) \in \mathbb{R} \) and \( \beta(\tau) \in \mathbb{R}^d \) are some arbitrary parameter values. Let

\[ e_i(\tau) = Q(\tau|x) + (x_i - x)'\frac{\partial Q(\tau|x)}{\partial x} - Q(\tau|x_i), \]
\[ \phi(\tau) = \sqrt{nh_{n,\tau}}\left( \frac{\alpha(\tau) - Q(\tau|x)}{h_{n,\tau}} \beta(\tau) - \frac{\partial Q(\tau|x)}{\partial x} \right). \]

Then, \( e_i(\tau) \) is the error from approximating \( Q(\tau|x_i) \) with a first-order Taylor expansion and \( \phi(\tau) \) is the normalized difference between \( (\alpha(\tau), \beta(\tau))' \) and their values in the expansion. Using the above notation, \( u_i(\tau) \) can be represented as

\[ u_i(\tau) = u_i^0(\tau) - e_i(\tau) - (nh_{n,\tau})^{-1/2}z_i'\phi(\tau), \]

where we have defined

\[ z_i' = \left( 1, \frac{(x_i - x)'}{h_{n,\tau}} \right). \]

This decomposition is useful because it breaks \( u_i(\tau) \) into three components: the true residual \( u_i^0(\tau) \), the error due to the Taylor approximation and the error due to replacing unknown values in the approximation by some estimates.

In addition, let

\[ V_{n,\tau}(\phi(\tau)) = \sum_{i=1}^{n} \left\{ \rho_x \left( u_i^0(\tau) - e_i(\tau) - (nh_{n,\tau})^{-1/2}z_i'\phi(\tau) \right) - \rho_x \left( u_i^0(\tau) - e_i(\tau) \right) \right\} K_{i,\tau} \tag{A.1} \]

with

\[ K_{i,\tau} = K \left( \frac{x_i - x}{h_{n,\tau}} \right), \]

where \( \rho_x \) is the check function. The term \( \rho_x \left( u_i^0(\tau) - e_i(\tau) \right) \) is introduced to recenter the first term and has no effect on estimation. Consequently, \( V_{n,\tau}(\phi(\tau)) \) is the recentered version of (2), whose minimizer delivers the local linear estimator. Also, note that (A.1) is convex in \( \phi(\tau) \).

Finally, let

\[ S_n(\tau, \phi(\tau), e_i(\tau)) = (nh_{n,\tau}^d)^{-1/2} \sum_{i=1}^{n} \left\{ P \left( 1(u_i^0(\tau) \leq (nh_{n,\tau}^d)^{-1/2}z_i'\phi(\tau) + e_i(\tau)) \right) x_i \right\} \]
\[ - 1 \left( u_i^0(\tau) \leq (nh_{n,\tau}^d)^{-1/2}z_i'\phi(\tau) + e_i(\tau) \right) \}
\[ = z_i'K_{i,\tau}, \]

A-1
which have mean zero and are independent in \( i \) for given \( \tau \) and \( \phi(\tau) \). A key element in the proofs lies in analyzing the property of \( S_n(\tau, \phi(\tau), e_i(\tau)) \); the details can be found in Appendix B.

**Proof of Theorem 1.** The strategy of the proof is similar to Theorem 2.1 in Koenker and Portnoy (1987). It consists of three steps. Step 1 proves \( \sup_{\tau \in T} \| \hat{\phi}(\tau) \| \leq \log^{1/2}(nh_d^d) \) with probability tending to 1. Step 2 provides an approximation to the subgradient process that holds uniformly over the set

\[
\Phi = \left\{ (\tau, \phi(\tau)) : \tau \in T, \| \phi(\tau) \| \leq \log^{1/2}(nh_d^d) \right\}.
\]

(A.2)

Step 3 verifies the Bahadur representation.

**Step 1.** Let \( K_n = \log^{1/2}(nh_d^d) \). First, note that

\[
V_{n,\tau}(\hat{\phi}(\tau)) \leq 0
\]

always holds for each \( \tau \) and every \( n \). This is because \( V_{n,\tau}(0) = 0 \), which can not be smaller than the minimized value \( V_{n,\tau}(\hat{\phi}(\tau)) \). The subsequent proof is by contradiction, showing that if \( \sup_{\tau \in T} \| \hat{\phi}(\tau) \| \geq K_n \), then \( V_{n,\tau} \) will be strictly positive with probability tending to 1 for some \( \tau \in T \). To this end, it is sufficient to show that for any \( \epsilon > 0 \), there exist some finite constants \( N_0 \) and \( \eta > 0 \) independent of \( \tau \), such that if \( \| \phi(\tau) \| \geq K_n \) holds for some \( \tau \), then

\[
P \left( V_{n,\tau}(\phi(\tau)) > \eta K_n^2 \right) > 1 - \epsilon \text{ holds for all } n \geq N_0.
\]

(A.3)

A sufficient condition for (A.3) is

\[
P \left( \inf_{\| \phi \| \geq K_n} \inf_{\tau \in T} V_{n,\tau}(\phi) > \eta K_n^2 \right) > 1 - \epsilon \text{ for all } n \geq N_0.
\]

(A.4)

This formulation is useful because \( \phi \) is no longer quantile dependent. Because \( V_{n,\tau}(\phi) \) is convex in \( \phi \), we always have

\[
V_{n,\tau}(\gamma \phi) - V_{n,\tau}(0) \geq \gamma \left( V_{n,\tau}(\phi) - V_{n,\tau}(0) \right) \text{ for any } \gamma \geq 1.
\]

(A.5)

Therefore, a further sufficient condition for (A.4) is

\[
P \left( \inf_{\| \phi \| = K_n} \inf_{\tau \in T} V_{n,\tau}(\phi) > \eta K_n^2 \right) > 1 - \epsilon \text{ for all } n \geq N_0,
\]

(A.6)

which we will now establish.

Consider the following decomposition, due to Knight (1998):

\[
V_{n,\tau}(\phi) = W_{n,\tau}(\phi) + Z_{n,\tau}(\phi),
\]

(A.7)

where

\[
W_{n,\tau}(\phi) = -(nh_{n,\tau}^d)^{-1/2} \sum_{i=1}^{n} K_{i,\tau} \psi_\tau(u_i^0(\tau) - e_i(\tau))z_{i,\tau}^0 \phi \text{ with } \psi_\tau(u) = \tau - 1(u < 0),
\]

\[
Z_{n,\tau}(\phi) = \sum_{i=1}^{n} K_{i,\tau} \int_{0}^{(nh_{n,\tau}^d)^{-1/2}z_{i,\tau}^0 \phi} \{1(u_i^0(\tau) - e_i(\tau) \leq s) - 1(u_i^0(\tau) - e_i(\tau) \leq 0)\} ds.
\]
Applying this decomposition, we have
\[
\inf_{\|\phi\|=K_n} \sup_{\tau \in T} K_n^{-2} V_n,\tau(\phi) \geq \inf_{\|\phi\|=K_n} \inf_{\tau \in T} K_n^{-2} Z_{n,\tau}(\phi) - \sup_{\|\phi\|=K_n} \sup_{\tau \in T} \left| K_n^{-2} W_n,\tau(\phi) \right|. \quad (A.8)
\]
We bound the two term on right hand side of (A.8) separately. First consider the second term:

\[
\sup_{\|\phi\|=K_n} \sup_{\tau \in T} \left| K_n^{-2} W_n,\tau(\phi) \right| 
\leq K_n^{-1} \sup_{\tau \in T} \left\| (nh^d_{n,\tau})^{-1/2} \sum_{i=1}^{n} \psi_x(u_i^0(\tau) - e_i(\tau)) z_{i,\tau} K_{i,\tau} \right\| 
\leq K_n^{-1} \sup_{\tau \in T} \left\| (nh^d_{n,\tau})^{-1/2} \sum_{i=1}^{n} \left\{ \psi_x(u_i^0(\tau) - e_i(\tau)) - \psi_x(u_i^0(\tau)) \right\} z_{i,\tau} K_{i,\tau} \right\| \quad (L1)
+ K_n^{-1} \sup_{\tau \in T} \left\| (nh^d_{n,\tau})^{-1/2} \sum_{i=1}^{n} \psi_x(u_i^0(\tau)) z_{i,\tau} K_{i,\tau} \right\| \quad (L2)
\]

Apply Lemma B.6, we have (L1) $= O_p \left(K_n^{-1}\right) = o_p(1)$. For Term (L2), the quantity inside the norm is $O_p(1)$ for any fixed $\tau \in T$ by a central limit theorem and it is stochastic equicontinuous in $\tau$ by Lemma B.3. Therefore, (L2) $= O_p \left(K_n^{-1}\right) = o_p(1)$. Thus,
\[
\sup_{\|\phi\|=K_n} \sup_{\tau \in T} \left| K_n^{-2} W_n,\tau(\phi) \right| = o_p(1). \quad (A.9)
\]

Next, consider the first term in (A.8). We will show that it is strictly positive with probability tending to 1. To this end, note that the integral appearing in $Z_{n,\tau}(\phi)$ is always nonnegative and satisfies (c.f. Lemma A.1 in Oka and Qu, 2011)
\[
\int_0^{(nh^d_{n,\tau})^{-1/2} z_{i,\tau} \phi} \left\{ 1(u_i^0(\tau) - e_i(\tau) \leq s) - 1(u_i^0(\tau) - e_i(\tau) \leq 0) \right\} ds 
\geq (nh^d_{n,\tau})^{-1/2} \frac{z_{i,\tau} \phi}{2} \left\{ 1(u_i^0(\tau) - e_i(\tau) \leq (nh^d_{n,\tau})^{-1/2} \frac{z_{i,\tau} \phi}{2}) - 1(u_i^0(\tau) - e_i(\tau) \leq 0) \right\}.
\]

Thus,
\[
K_n^{-2} Z_{n,\tau}(\phi) 
\geq K_n^{-2} (nh^d_{n,\tau})^{-1/2} \left( \frac{\phi}{2} \right) \sum_{i=1}^{n} \left\{ 1(u_i^0(\tau) - e_i(\tau) \leq (nh^d_{n,\tau})^{-1/2} \frac{z_{i,\tau} \phi}{2}) - 1(u_i^0(\tau) - e_i(\tau) \leq 0) \right\} z_{i,\tau} K_{i,\tau} 
= K_n^{-2} \left( \frac{h^d_n}{h^d_{n,\tau}} \right)^{1/2} \left( \frac{\phi}{2} \right) \left\{ S_n(\tau, 0, e_i(\tau)) - S_n(\tau, \frac{\phi}{2}, e_i(\tau)) \right\} \quad (L3)
+ K_n^{-2} (nh^d_{n,\tau})^{-1/2} \left( \frac{\phi}{2} \right) \sum_{i=1}^{n} \left\{ P \left( u_i^0(\tau) - e_i(\tau) \leq (nh^d_{n,\tau})^{-1/2} \frac{z_{i,\tau} \phi}{2} \right| x_i \right\) 
- P(u_i^0(\tau) - e_i(\tau) \leq 0 \| x_i) \right\} z_{i,\tau} K_{i,\tau}. \quad (L4)
\]
We now analyze (L3) and (L4) separately.

\[(L3) = O_p \left( K_n^{-1} \right) = o_p (1) \quad (A.10)\]

because of Lemma (B.5), \(\| \phi \| = K_n\) and \(h_{n, \tau}^d/h_n^d = O(1)\) by Assumption 5. Apply a mean-value theorem (L4):

\[(L4) = \frac{1}{4} K_n^{-2} (nh_{n, \tau}^d)^{-1} \sum_{i=1}^{n} f_{Y \mid X} (\tilde{y}_i \mid x_i) K_{i, \tau} \phi' z_i z_i' \phi,\]

where \(\tilde{y}_i\) lies between \(Q(\tau \mid x_i) + e_i(\tau)\) and \(Q(\tau \mid x_i) + e_i(\tau) + (nh_{n, \tau}^d)^{-1/2} z_i \phi / 2\). Note that \(x_i\) is in a vanishing neighborhood of \(x\) and \(\tilde{y}_i\) approaches \(Q(\tau \mid x_i)\) as \(n \to \infty\). Therefore, Assumption 2 implies that \(f_{Y \mid X} (\tilde{y}_i \mid x_i) \geq f_L\) for large \(n\). Thus,

\[(L4) \geq \frac{1}{4} f_L^2 \phi' \left( \frac{1}{nh_{n, \tau}^d} \sum_{i=1}^{n} K_{i, \tau} z_i z_i' \right) \phi \text{ uniformly in } \tau \text{ for large } n.\]

The term in the parentheses satisfies

\[(nh_{n, \tau}^d)^{-1} \sum_{i=1}^{n} z_i z_i' K_{i, \tau} \to^p f_X (x) N_x (\tau) \text{ uniformly in } \tau,\]

where

\[N_x (\tau) = \begin{cases} \int \left[ \begin{array}{cc} 1 & 0 \\ 0 & u^2 \end{array} \right] K (u) \, du & \text{if } x \text{ is an interior point} \\ \int_{D_x, \tau} \left[ \begin{array}{c} 1 \\ u \end{array} \right] \left[ \begin{array}{cc} 1 & 0 \\ 0 & u' \end{array} \right] K (u) \, du & \text{if } x \text{ is a boundary point} \end{cases}\]

In both cases, \(N_x (\tau)\) is positive definite for all \(\tau\) by Assumption 6. Let \(\lambda_{\min} (\tau) > 0\) denote its minimum eigenvalue, then

\[(L4) \geq \frac{1}{4} f_L \lambda_{\min} (\tau) \text{ in probability uniformly in } \tau.\]

Combining the results (A.9), (A.10) and (A.12), we see that (A.12) is strictly positive and dominates (A.9) and (A.10) with probability tending to 1. We have therefore proved (A.6). In fact, using the same arguments as above, we can establish a stronger result: for any \(\epsilon > 0\), there exists \(K_0 > 0\) and \(\eta > 0\), such that if

\[\| \hat{\phi} (\tau) \| \geq K_0\]

holds for some \(\tau\), then

\[P \left( V_{n, \tau} (\hat{\phi} (\tau)) > \eta \right) > 1 - \epsilon \text{ holds for all } n \geq N_0.\]

(A.13)

The weaker result suffices for proving the current theorem and the stronger result is needed for proving Theorem 3.
Step 2. Consider the subgradient, normalized by \((nh_n^d)^{-1/2}\):

\[
(nh_n^d)^{-1/2} \sum_{i=1}^{n} \left( \tau - 1(u_i^0(\tau) \leq e_i(\tau) + (nh_n^d)^{-1/2}z_{i,\tau}'\hat{\phi}(\tau)) \right) z_{i,\tau}K_{i,\tau},
\]

which, by Theorem 2.1 in Koenker (2005), is \(O_p\) \((nh_n^d)^{-1/2}\) = \(o_p(1)\) uniformly in \(\tau\). Adding and subtracting terms, it can be rewritten as

\[
\left\{ S_n(\tau, \hat{\phi}(\tau), e_i(\tau)) - S_n(\tau, 0, e_i(\tau)) \right\} + \left\{ S_n(\tau, 0, e_i(\tau)) - S_n(\tau, 0, 0) \right\} + (nh_n^d)^{-1/2} \sum_{i=1}^{n} \left( \tau - P\left( u_i^0(\tau) \leq e_i(\tau) + (nh_n^d)^{-1/2}z_{i,\tau}'\hat{\phi}(\tau) \right| x_i \right) \right\} z_{i,\tau}K_{i,\tau}
\]

Because of Step 1, we can restrict our attention to the set \(\Phi\) defined in (A.2). The term in the first curly bracket is \(o_p(1)\) uniformly over this set by Lemma B.5. The term in the second curly bracket is uniformly \(o_p(1)\) implied by the proof of Lemma B.5. Consider the last term, and apply a first-order Taylor expansion. It equals to

\[
-(nh_n^d)^{-1/2} \sum_{i=1}^{n} f_{Y|x}(\tilde{y}_i|x_i) e_i(\tau) z_{i,\tau}K_{i,\tau} -(nh_n^d)^{-1/2}(nh_n^d)^{-1/2} \left( \sum_{i=1}^{n} f_{Y|x}(\tilde{y}_i|x_i) K_{i,\tau} z_{i,\tau}z_{i,\tau}' \right) \hat{\phi}(\tau),
\]

where \(\tilde{y}_i\) lies between \(Q(\tau|x_i)\) and \(Q(\tau|x_i) + (nh_n^d)^{-1/2}z_{i,\tau}'\hat{\phi}/2\). Combining the above results, we have

\[
\hat{\phi}(\tau) = \left( (nh_n^d)^{-1/2} \sum_{i=1}^{n} f_{Y|x}(\tilde{y}_i|x_i) K_{i,\tau}z_{i,\tau}z_{i,\tau}' \right)^{-1} \left\{ \left( \frac{h_n^d}{h_n^d} \right)^{1/2} S_n(\tau, 0, 0) - (nh_n^d)^{-1/2} \sum_{i=1}^{n} f_{Y|x}(\tilde{y}_i|x_i) e_i(\tau) z_{i,\tau}K_{i,\tau} \right\} + o_p(1)
\]

\[
= \left( f_{Y|x}(Q(\tau|x)) f_X(x) N_x(\tau) \right)^{-1} \left\{ \left( \frac{h_n^d}{h_n^d} \right)^{1/2} S_n(\tau, 0, 0) - (nh_n^d)^{-1/2} f_{Y|x}(Q(\tau|x)) \sum_{i=1}^{n} e_i(\tau) z_{i,\tau}K_{i,\tau} \right\} + o_p(1),
\]

where the second equality follows from (A.11) and Assumption 2.

Step 3. We now derive the limit of the preceding expression. Note that for any \(x_i\) satisfying \((x_i - x)/h_{n,\tau} \in \text{supp}(K(\cdot))\), we have

\[
e_i(\tau) = -\frac{1}{2} \left( \frac{x_i - x}{h_n^{d}} \right)' \frac{\partial^2 Q(\tau|x)}{\partial x \partial x'} \left( \frac{x_i - x}{h_n^{d}} \right) h_n^{d} + o(h_n^{2}),
\]

Below, we consider two cases separately.
Case 1: $x$ is a boundary point. Then,
\[
-h_{n,\tau}^{-2} \left( n h_{n,\tau}^d \right)^{-1} \sum_{i=1}^{n} e_i (\tau) z_{i,\tau} K_{i,\tau} \tag{A.14}
\]
\[
= \frac{1}{2} \int \left( \frac{x_i - x}{h_{n,\tau}} \right) \frac{\partial^2 Q(\tau|x)}{\partial x \partial x'} \left( \frac{x_i - x}{h_{n,\tau}} \right) \left[ \frac{1}{h_{n,\tau}} K \left( \frac{x_i - x}{h_{n,\tau}} \right) f_X(x_i) dx_i + o_p(1) \right]
\]
\[
\rightarrow \frac{1}{2} f_X(x) \int_{D_{x,\tau}} u' \frac{\partial^2 Q(\tau|x)}{\partial x \partial x'} u \left[ \frac{1}{u} \right] K(u) du.
\]

Let $\ell_1'$ be a row vector with the first element equal to 1 and the others zero, and let
\[
\frac{1}{2} \ell_1' N_x(\tau)^{-1} \int_{D_{x,\tau}} u' \frac{\partial^2 Q(\tau|x)}{\partial x \partial x'} u \left[ \frac{1}{u} \right] K(u) du = d_{b,\tau}
\]
as in Theorem 1, we have
\[
\sqrt{n h_{n,\tau}^d} \left( \hat{\alpha}(\tau) - Q(\tau|x) - d_{b,\tau} h_{n,\tau}^2 \right) = \ell_1' N_x(\tau)^{-1} \left( n h_{n,\tau}^d \right)^{-1/2} \sum_{i=1}^{n} \left( \tau - 1(u_i^0(\tau) \leq 0) \right) z_{i,\tau} K_{i,\tau} + o_p(1).
\]

Case 2: $x$ is an interior point. Then the above formula simplifies because
\[
N_x(\tau) = \begin{bmatrix} 1 & 0 \\ 0 & \mu_2(K) I_d \end{bmatrix}
\]
and
\[
h_{n,\tau}^{-2} (n h_{n,\tau}^d)^{-1} \ell_1' \sum_{i=1}^{n} e_i (\tau) z_{i,\tau} K_{i,\tau} = \frac{1}{2} \mu_2(K) \text{tr} \left( \frac{\partial^2 Q(\tau|x)}{\partial x \partial x'} \right) + o_p(1).
\]

Let
\[
d_{\tau} = \frac{1}{2} \mu_2(K) \text{tr} \left( \frac{\partial^2 Q(\tau|x)}{\partial x \partial x'} \right)
\]
as in Theorem 1. Then
\[
\sqrt{n h_{n,\tau}^d} \left( \alpha(\tau) - Q(\tau|x) - d_{\tau} h_{n,\tau}^2 \right) = \frac{(n h_{n,\tau}^d)^{-1/2} \sum_{i=1}^{n} \left( \tau - 1(u_i^0(\tau) \leq 0) \right) K_{i,\tau}}{f_X(x) f_X(x) (Q(\tau|x))} + o_p(1).
\]

**Proof of Theorem 2.** The proof consists of two steps. The first step shows that $\hat{\alpha}(\tau)$, obtained by solving (2) for all $\tau \in T$, converges weakly to the desired limit. The second step shows that $\hat{\alpha}^*(\tau)$ has the same weak limit as $\hat{\alpha}(\tau)$ over $\tau \in T$. We focus on the interior point case, as the boundary case can be proved similarly.

**Step 1.** It suffices to show that the leading term in the uniform Bahadur representation converges weakly to the desired limit. Its finite dimensional convergence is an immediate consequence of
results in Fan, Hu and Truong (1994), therefore the detail is omitted. To verify tightness, note that the denominator is finite and bounded away from 0 by Assumptions 1 and 2. The numerator is tight by Lemma B.3.

**Step 2.** We use similar arguments as in Neocleous and Portnoy (2008). Consider

\[ \hat{\alpha}^*(\tau) - \hat{\alpha}(\tau) = \gamma_n(\tau) \hat{\alpha}(\tau_j) - \gamma_n(\tau) \hat{\alpha}(\tau) + (1 - \gamma_n(\tau)) \hat{\alpha}(\tau_{j+1}) - (1 - \gamma_n(\tau)) \hat{\alpha}(\tau) \]

\[ = \gamma_n(\tau) \left( \hat{\alpha}(\tau_j) - Q(\tau_j|x) - d_{\tau_j} h_{n,\tau_j}^2 \right) - \gamma_n(\tau) \left( \hat{\alpha}(\tau) - Q(\tau|x) - d_{\tau} h_{n,\tau}^2 \right) \]

\[ + (1 - \gamma_n(\tau)) \left( \hat{\alpha}(\tau_{j+1}) - Q(\tau_{j+1}|x) - d_{\tau_{j+1}} h_{n,\tau_{j+1}}^2 \right) - (1 - \gamma_n(\tau)) \left( \hat{\alpha}(\tau) - Q(\tau|x) - d_{\tau} h_{n,\tau}^2 \right) \]

\[ + \gamma_n(\tau) \left( Q(\tau_j|x) + d_{\tau_j} h_{n,\tau_j}^2 \right) + (1 - \gamma_n(\tau)) \left( Q(\tau_{j+1}|x) + d_{\tau_{j+1}} h_{n,\tau_{j+1}}^2 \right) - (Q(\tau|x) + d_{\tau} h_{n,\tau}^2) \]

For Term (L5), applying Theorem 1, we have

\[ \sqrt{n} h_n^d \left( \hat{\alpha}(\tau_j) - Q(\tau_j|x) - d_{\tau_j} h_{n,\tau_j}^2 \right) - \sqrt{n} h_n^d \left( \hat{\alpha}(\tau) - Q(\tau|x) - d_{\tau} h_{n,\tau}^2 \right) \]

\[ = \sqrt{n} h_n^d \sum_{i=1}^n \left( \tau_j - 1(u_i^0(\tau_j) \leq 0) \right) K_{i,\tau} \left( \frac{f_X(x) f_{Y|X}(Q(\tau_j|x)|x)}{f_X(x) f_{Y|X}(Q(\tau|x)|x)} \right) + o_p(1) \]

\[ = \sqrt{n} h_n^d \sum_{i=1}^n \left( \tau_j - 1(u_i^0(\tau_j) \leq 0) \right) K_{i,\tau} \left( \frac{1}{f_Y|X(Q(\tau_j|x)|x)} - \frac{1}{f_Y|X(Q(\tau|x)|x)} \right) + o_p(1). \]

The first term on the right hand side is \(o_p(1)\) by the stochastic equicontinuity of the subgradient process, see Lemma B.3. The second term is also \(o_p(1)\) by the tightness of the subgradient process and Assumption 2. Term (L6) can be analyzed in the same way. For Term (L7), apply a second-order Taylor expansion as in Neocleous and Portnoy (2008, p.1228):

\[ Q(\tau_j|x) = Q(\tau|x) + \frac{\partial Q(\tau|x)}{\partial \tau} (\tau_j - \tau) + O((\tau_{j+1} - \tau_j)^2), \]

where the order of the remainder term is due to the Lipschitz continuity of \(\partial Q(\tau|x)/\partial \tau\). This implies

\[ \gamma_n(\tau)Q(\tau_j|x) + (1 - \gamma_n(\tau)) Q(\tau_{j+1}|x) = Q(\tau|x) + O((\tau_{j+1} - \tau_j)^2). \]

Similarly,

\[ d_{\tau_j} h_{n,\tau_j}^2 = d_{\tau} h_{n,\tau}^2 + \frac{\partial d_{\tau} h_{n,\tau}^2}{\partial \tau} (\tau_j - \tau) + O((\tau_{j+1} - \tau_j)^2), \]

implying

\[ \gamma_n(\tau)d_{\tau_j} h_{n,\tau_j}^2 + (1 - \gamma_n(\tau)) d_{\tau_{j+1}} h_{n,\tau_{j+1}}^2 = d_{\tau} h_{n,\tau}^2 + O((\tau_{j+1} - \tau_j)^2). \]

Therefore,

\[ (L7) = O((\tau_{j+1} - \tau_j)^2). \]
Because $\tau_{j+1} - \tau_j = O((nh_n^d)^{-1/4})$, it follows $\sqrt{nh_n^d} (\tau_j - \tau) \to 0$. This holds uniformly in $\tau \in T$.

**Proof of Theorem 3.** Consider the first result. Without loss of generality, assume the minimizers to the constrained and unconstrained problems are both unique. The subsequent proof is by contradiction. Suppose the stochastic order condition is violated. Then,

$$\sqrt{nh_n^d} |\alpha(\tau_j) - Q(\tau_j|x)| > K_0$$

for some $\tau_j \in T$ and some $n \geq N_0$ with positive probability. We will argue along such a sequence. Without loss of generality, assume

$$\alpha(\tau_j) > Q(\tau_j|x) + K_0.$$  \hspace{1cm} (A.15)

Then, we consider the following two possibilities separately: (1) the constraint is not binding at $\tau_j$ (i.e., $\alpha(\tau_j) > \alpha(\tau_{j-1})$), and (2) the constraint is in fact binding (i.e., $\alpha(\tau_j) = \alpha(\tau_{j-1})$).

In the first case, $\alpha(\tau_j)$ is identical to the unconstrained estimate $\hat{\alpha}(\tau_j)$, therefore

$$V_n,\tau_j(\hat{\phi}(\tau_j)) = V_n,\tau_j(\hat{\phi}(\tau_j)) \leq 0.$$  

However, this contradicts (A.13). Consider the second case. Let $\tau_l$ denote the lowest quantile in the grid for which $\alpha(\tau_l) = \alpha(\tau_j)$. By the monotonicity of the quantile function and (A.15), we must have

$$\alpha(\tau_l) > Q(\tau_l|x) + K_0.$$  

Therefore,

$$P \left( V_n,\tau_l(\hat{\phi}(\tau_l)) > \eta \right) > 1 - \epsilon \text{ holds for all } n \geq N_0.$$  

Now, consider decreasing the value of $\alpha(\tau_l)$ by some small amount. This will decrease the value of the objective function $V_n,\tau_l(\hat{\phi}(\tau_l))$ due to its convexity, see (A.5). This action does not violate any constraint because of the definition of $\tau_l$ (i.e., it is the lowest quantile for which $\alpha(\tau_l) = \alpha(\tau_j)$). Therefore, the new value remains admissible and yet returns a smaller objective function value. It contradicts that $\alpha(\tau_l)$ is the minimizer.

The second result holds because under the given condition on $m$, the quantiles will not cross with probability arbitrarily close to 1 in large samples, therefore the large sample distribution is the same as the unconstrained case, which is given in Theorem 2. The constraints serve as a finite sample correction, having no effect asymptotically.

**Proof of Corollary 1.** The mean squared error (MSE) for an interior point $x$ is

$$\frac{1}{4} \mu_2^2(K) \text{tr} \left( \frac{\partial^2 Q(\tau|x)}{\partial x \partial x'} \right)^2 h_n^d = \frac{\tau (1 - \tau) \int K(v)^2 dv}{n h_n^d f_X(x) f_{Y|X}(Q(\tau|x)|x)^2} + o_p(n h_n^d).$$  

Therefore, the optimal bandwidth that minimizes the (interior) asymptotic MSE is

$$h_{n,\tau} = \left( \frac{\tau (1 - \tau) \int K(v)^2 dv}{\mu_2^2(K) f_X(x) f_{Y|X}(Q(\tau|x)|x)^2 \text{tr} \left( \frac{\partial^2 Q(\tau|x)}{\partial x \partial x'} \right)^2} \right)^{1/(4+d)} n^{-1/(4+d)}.$$
We now briefly verify that this bandwidth rule satisfies Assumption 5. The factor in front of \( n^{-1/(4+d)} \) corresponds to \( c(\tau) \). It is clearly bounded under Assumptions 1-4. We now verify that it is Lipschitz continuous. It suffices to verify the three quantities \((\tau (1-\tau))^{1/(4+d)} \), \( f_{Y|X}(Q(\tau|x)|x)^{-2/(4+d)} \) and \( \text{tr} \left( \frac{\partial^2 Q(\tau|x)/\partial x\partial x'}{(4+d)} \right)^{-2/(4+d)} \) are individually Lipschitz continuous, or equivalently, to show they have bounded first derivatives. This is true for \((\tau (1-\tau))^{1/(4+d)} \) because \( T \subset (0,1) \). For the second term, apply a mean-value theorem.

\[
\begin{align*}
\left| f_{Y|X}(Q(\tau_1|x)|x)^{-2/(4+d)} - f_{Y|X}(Q(\tau_2|x)|x)^{-2/(4+d)} \right| &= \left| \frac{2}{4+d} f(Q(\tau|x)|x)^{-\frac{6+d}{4+d}} f'_{Y|X}(Q(\tau|x)|x) Q'(\tau|x) (\tau_2 - \tau_1) \right|,
\end{align*}
\]

where \( \tau \in T \). The terms inside the absolute value on the right side are all bounded over \( T \) by Assumptions 1-3. Finally, \( \text{tr} \left( \frac{\partial^2 Q(\tau|x)/\partial x\partial x'}{(4+d)} \right)^{-2/(4+d)} \) is Lipschitz continuous following the same argument.

**Proof of Corollary 2.** We have, for the interior point case,

\[
\begin{align*}
P( Q(\tau|x) \notin C_p \text{ for some } \tau \in T) &= P \left( \frac{1}{\sigma_{n,\tau}(x)} \left| Q(\tau|x) - \bar{\alpha}^*(\tau) + d_\tau h_n^2 \right| > Z_p \text{ for some } \tau \in T \right) \\
&= P \left( \sup_{\tau \in T} \frac{1}{\sigma_{n,\tau}(x)} \left| Q(\tau|x) - \bar{\alpha}^*(\tau) + d_\tau h_n^2 \right| > Z_p \right) \\
&\rightarrow p
\end{align*}
\]

by the continuous mapping theorem. The boundary point case can be proved using the same argument.

**Proof of Corollary 3.** We focus on the interior point case and verify the conditions in Chernozhukov, Fernández-Val and Galichon (2010, Corollary 3) are satisfied. First, Assumptions 1-3 in this paper imply their Assumption 1 with the domain \((0,1)\) replaced by \( T \). They also imply the strict monotonicity of \( Q(\tau|x) \) in \( \tau \) as well as its continuous differentiability in both arguments. To verify their Assumption 2, simply let

\[
a_n = \sqrt{nh_{n,\tau}^d} \quad \text{and} \quad G_x(\tau) = \frac{G(\tau)}{\sqrt{f_X(x)} f_{Y|X}(Q(\tau|x)|x)} + h(\tau)d_\tau,
\]

and apply the first result in Theorem 2.
Appendix B. Auxiliary Lemmas

The same notations as in Appendix A are used throughout. The next lemma is needed for analyzing the effect of the quantile dependant bandwidth on the asymptotic properties of the estimators.

Lemma B.1 Let Assumptions 1, 4 and 5 hold.

1. For any \( \gamma \geq 1 \) there exists a \( B > 0 \), such that for any \( \tau_1, \tau_2 \in T \) with \( \tau_1 \leq \tau_2 \)

\[
E \| z_{i,\tau_2} K_{i,\tau_2} - z_{i,\tau_1} K_{i,\tau_1} \|^{2\gamma} \leq B h_n^{d} (\tau_2 - \tau_1)^{2\gamma}.
\]

2. Let \( b_n = (nh_n^d)^{-1/2-\kappa} \) with \( 0 < \kappa < 1/2 \) being some arbitrary constant and \( h_n \) satisfying \( h_n \to 0 \) and \( nh_n^d \to \infty \) as \( n \to \infty \). Let \( \delta_n = \{(\tau_1, \tau_2) : \tau_1 \in T, \tau_2 \in T, \tau_1 \leq \tau_2 \leq \tau_1 + b_n\} \), then

\[
\sup_{(\tau_1, \tau_2) \in \delta_n} (nh_n^d)^{-1/2} \sum_{i=1}^{n} \| z_{i,\tau_2} K_{i,\tau_2} - z_{i,\tau_1} K_{i,\tau_1} \| = o_p(1).
\]

Proof. By Assumption 4, we can, without loss of generality, assume that the support of \( K(.) \) is contained in a compact set \( D = [-1, 1]^d \).

To show the first result, note that

\[
\| z_{i,\tau_2} K_{i,\tau_2} - z_{i,\tau_1} K_{i,\tau_1} \| \leq \| z_{i,\tau_2} (K_{i,\tau_2} - K_{i,\tau_1}) \| + \| K_{i,\tau_1} (z_{i,\tau_2} - z_{i,\tau_1}) \|,
\]

We now analyze the two terms on the right side separately. Apply the mean-value theorem to the first term:

\[
\| z_{i,\tau_2} (K_{i,\tau_2} - K_{i,\tau_1}) \| = \| z_{i,\tau_2} K'_{i,\tau} \left( \frac{x_i - x}{h_{n,\tau_2}} - \frac{x_i - x}{h_{n,\tau_1}} \right) \|,
\]

where \( \tau \in [\tau_1, \tau_2] \) and \( K'_{i,\tau} \) is the derivative of \( K_{i,\tau} \) with respect to \( (x_i - x)/h_{n,\tau} \), evaluated at \( \tau = \tilde{\tau} \). Because \( K'_{i,\tilde{\tau}} \) is bounded by Assumption 4, say \( \| K'_{i,\tilde{\tau}} \| \leq M < \infty \) within \( D \), and is zero elsewhere, the preceding display is bounded by

\[
M \| z_{i,\tau_2} \left( \frac{x_i - x}{h_{n,\tilde{\tau}}}, D \right) \| \left( \frac{x_i - x}{h_{n,\tau_2}} - \frac{x_i - x}{h_{n,\tau_1}} \right) \|
\leq M \| z_{i,\tau_2} \left( \frac{x_i - x}{h_{n,\tilde{\tau}}}, D \right) \| \left( \frac{x_i - x}{h_{n,\tau_2}} - \frac{x_i - x}{h_{n,\tau_1}} \right) \|
= M \| T_1 \| \| T_2 \|.
\]

We have

\[
\| T_1 \| = \| z_{i,\tau_2} \left( \frac{x_i - x}{h_{n,\tilde{\tau}}}, D \right) \| \leq \| z_{i,\tau_2} \left( \frac{x_i - x}{h_{n,\tau_2}} \in (\check{c}/\check{\gamma}) D \right) \| \leq C,
\]

where the first inequality follows from Assumption 5, the second uses the definition of \( z_{i,\tau_2} \), and the constant \( C \) depends only on \( (\check{c}/\check{\gamma}) D \). Also,

\[
\| T_2 \| \leq \| 1 \left( \frac{x_i - x}{h_{n,\tau_2}} \in (\check{c}/\check{\gamma}) D \right) \left( \frac{x_i - x}{h_{n,\tau_2}} \right) \| \left( \frac{h_{n,\tau_1} - h_{n,\tau_2}}{h_{n,\tau_1}} \right) \|
\]

(B.2)
due to the Cauchy-Schwarz inequality. The first norm on the right hand side of (B.2) is bounded by
\[ C \ast 1 (x_i - x \in h_n, \tilde{D}) \leq C \ast 1 (x_i - x \in \tilde{c}Dh_n), \]
while the second norm equals to
\[ \left| \frac{c(\tau_1) - c(\tau_2)}{c(\tau_1)} \right| \leq \left| \frac{c(\tau_1) - c(\tau_2)}{C} \right| \leq G(\tau_2 - \tau_1), \]
where the last inequality uses the Lipschitz property of \( c(\tau) \) and \( C \) and \( G \) are some finite constants (c.f. Assumption 5). Therefore, the first term in (B.1) is bounded by
\[ CMG \ast 1 (x_i - x \in \tilde{c}Dh_n) (\tau_2 - \tau_1) \]
Now, consider the second term in (B.1) and apply similar arguments as above:
\[
\| K_{i,\tau_1} (z_{i,\tau_2} - z_{i,\tau_1}) \| = \left\| K_{i,\tau_1} \left( \frac{x_i - x}{h_{n,\tau_2}} - \frac{x_i - x}{h_{n,\tau_1}} \right) \right\|
\leq \left\| K_{i,\tau_1} \left( \frac{x_i - x}{h_{n,\tau_1}} \right) \right\|
\leq KCG \ast 1 (x_i - x \in \tilde{c}Dh_n) (\tau_2 - \tau_1),
\]
where \( \tilde{K} \) is an upper bound for \( K(\cdot) \) and \( C \) and \( G \) have the same definition as before. Let \( \tilde{C} = CMG + KCG \), then (B.1) satisfies
\[
\| z_{i,\tau_2} K_{i,\tau_2} - z_{i,\tau_1} K_{i,\tau_1} \| \leq \tilde{C} \ast 1 (x_i - x \in \tilde{c}Dh_n) (\tau_2 - \tau_1). \tag{B.3}
\]
Consequently,
\[
E \| (z_{i,\tau_2} K_{i,\tau_2} - z_{i,\tau_1} K_{i,\tau_1}) \|^2 \leq \tilde{C}^{2\gamma} (\tau_2 - \tau_1)^{2\gamma} P (x_i - x \in \tilde{c}Dh_n). \]
Because the density of \( x_i \) is bounded by Assumption 1, there exists a finite constant \( A \) independent of \( \tau \) such that
\[ P (x_i - x \in \tilde{c}Dh_n) \leq Ah^d_n. \]
Letting \( B = \tilde{C}^{2\gamma} A \) completes the proof for the first result.
To prove the second result, we apply (B.3):
\[
\sup_{(\tau_1,\tau_2) \in \delta_n} (nh_n^d)^{-1/2} \sum_{i=1}^n \| z_{i,\tau_2} K_{i,\tau_2} - z_{i,\tau_1} K_{i,\tau_1} \|
\leq \sup_{(\tau_1,\tau_2) \in \delta_n} \tilde{C} (nh_n^d)^{1/2} (\tau_2 - \tau_1) \left\{ (nh_n^d)^{-1} \sum_{i=1}^n 1 (x_i - x \in \tilde{c}Dh_n) \right\},
\]
B-2
The quantity inside the curly brackets is independent of \( \tau \) and satisfies a weak law of large numbers. Therefore, it is of order \( O_p(1) \). Finally,

\[
\sup_{(\tau_1, \tau_2) \in \delta_n} \bar{C}(nh_n^d)^{1/2}(\tau_2 - \tau_1) \leq \bar{C}(nh_n^d)^{1/2}b_n = \bar{C}(nh_n^d)^{-\kappa} \to 0
\]

because \( \kappa > 0 \) and \( nh_n^d \to \infty \). ■

The next Lemma is needed to establish the stochastic equicontinuity of the process \( S_n(\tau, 0, 0) \).

**Lemma B.2** Let \( b_n = (nh_n^d)^{-1/2 - \kappa} \) with \( \kappa \in (0, 1/2) \) being some arbitrary constant and \( h_n \to 0 \) and \( nh_n^d \to \infty \) as \( n \to \infty \). Then, there exist \( \gamma > 1 \) and \( C < \infty \), such that for any \( \tau_1, \tau_2 \in T \) satisfying \( |\tau_2 - \tau_1| \geq b_n \), we have

\[
E \|S_n(\tau_2, 0, 0) - S_n(\tau_1, 0, 0)\|^{2\gamma} \leq C|\tau_2 - \tau_1|^\gamma.
\]

**Proof.** Without loss of generality, assume \( \tau_2 \geq \tau_1 \). It is equivalent to show

\[
(E \|S_n(\tau_2, 0, 0) - S_n(\tau_1, 0, 0)\|^{2\gamma})^{1/\gamma} \leq C^{1/\gamma}(\tau_2 - \tau_1).
\]

Let

\[
A_{1i} = \left\{ (\tau_2 - 1(u_0^0(\tau_2) \leq 0)) - (\tau_1 - 1(u_0^0(\tau_1) \leq 0)) \right\} z_i, \tau_2 K_{i, \tau_2} \\
A_{2i} = (\tau_1 - 1(u_0^0(\tau_1) \leq 0)) (z_i, \tau_2 K_{i, \tau_2} - z_i, \tau_1 K_{i, \tau_1}).
\]

Then, we can write

\[
S_n(\tau_2, 0, 0) - S_n(\tau_1, 0, 0) = (nh_n^d)^{-1/2} \sum_{i=1}^n (A_{1i} + A_{2i}),
\]

therefore

\[
\left( E \|S_n(\tau_2, 0, 0) - S_n(\tau_1, 0, 0)\|^{2\gamma} \right)^{1/\gamma} = \left\{ (nh_n^d)^{-\gamma} E \left[ \sum_{i=1}^n A_{1i} + A_{2i} \right]^{2\gamma} \right\}^{1/\gamma}
\]

\[
= \left\{ (nh_n^d)^{-\gamma} E \left[ \sum_{j=1}^{d+1} \sum_{i=1}^n A_{1i,j} + A_{2i,j} \right]^{2\gamma} \right\}^{1/\gamma}
\]

\[
\leq \sum_{j=1}^{d+1} \left\{ (nh_n^d)^{-\gamma} E \left[ \sum_{i=1}^n A_{1i,j} + A_{2i,j} \right]^{2\gamma} \right\}^{1/\gamma}, \quad (B.4)
\]

where \( A_{1i,j} \) and \( A_{2i,j} \) denote the j-th element of \( A_{1i} \) and \( A_{2i} \) respectively, and the last line follows from the Minkowski inequality. We now bound the term inside the curly brackets using arguments
similar to Bai (1996, Lemma A1).

\[
(nh_n^d)^{-\gamma} E \left| \sum_{i=1}^{n} A_{1i,j} + A_{2i,j} \right|^{2\gamma}
\]

\[
\leq C(nh_n^d)^{-\gamma} \left( \sum_{i=1}^{n} E|A_{1i,j} + A_{2i,j}|^2 \right)^{\gamma} + C(nh_n^d)^{-\gamma} \sum_{i=1}^{n} E|A_{1i,j} + A_{2i,j}|^{2\gamma}
\]

\[
\leq 2\gamma C(nh_n^d)^{-\gamma} \left( \sum_{i=1}^{n} EA_{1i,j}^2 + EA_{2i,j}^2 \right)^{\gamma} + 2\gamma C(nh_n^d)^{-\gamma} \sum_{i=1}^{n} E \left( \|A_{1i}\|^2 + \|A_{2i}\|^2 \right)^{\gamma}
\]

\[
\leq 2\gamma C(nh_n^d)^{-\gamma} \left( \sum_{i=1}^{n} E\|A_{1i}\|^2 + E\|A_{2i}\|^2 \right)^{\gamma} + 2\gamma C(nh_n^d)^{-\gamma} \sum_{i=1}^{n} E \left( \|A_{1i}\|^2 + \|A_{2i}\|^2 \right)^{\gamma}
\]

\[
\leq 2\gamma (nh_n^d)^{-\gamma} C \sum_{i=1}^{n} \left( \left( E\|A_{1i}\|^{2\gamma} \right)^{1/\gamma} + \left( E\|A_{2i}\|^{2\gamma} \right)^{1/\gamma} \right)^{\gamma}
\]

(T3)

where the first inequality is because of the Rosenthal inequality for independent random variables (Hall and Heyde, 1980, p. 23) with the constant C depending only on \( \gamma \), the second is because of the triangle inequality, the third is due to the simple fact that \( A_{1i,j}^2 \leq \|A_{1i}\|^2 \) and \( A_{2i,j}^2 \leq \|A_{2i}\|^2 \), and the last inequality is due to the Minkowski inequality. We now derive bounds for the summands in (T3) and (T4).

\[
E\|A_{1i}\|^{2\gamma} = E \left\{ E \left( \|A_{1i}\|^{2\gamma} \right| {x}_i \right\}
\]

\[
= E \left\{ E \left( (\tau_2 - 1(u_i^0(\tau_2) \leq 0) - \tau_1 + 1(u_i^0(\tau_1) \leq 0) \right)^{2\gamma} \right| {x}_i \right\} \|z_i,\tau_2 K_i,\tau_2\|^{2\gamma}
\]

\[
\leq E \left\{ E \left( (\tau_2 - 1(u_i^0(\tau_2) \leq 0) - \tau_1 + 1(u_i^0(\tau_1) \leq 0) \right)^{2\gamma} \right| {x}_i \right\} \|z_i,\tau_2 K_i,\tau_2\|^{2\gamma}
\]

\[
\leq (\tau_2 - \tau_1) E \|z_i,\tau_2 K_i,\tau_2\|^{2\gamma}.
\]

where the first inequality is because \( |\tau_2 - 1(u_i^0(\tau_2) \leq 0) - \tau_1 + 1(u_i^0(\tau_1) \leq 0)| \leq 1 \). Because \( K_i,\tau \) has a bounded support, using similar arguments as in Lemma B.1, there exists a constant \( C \) independent of \( \tau \), such that \( E\|z_i,\tau_2 K_i,\tau_2\|^{2\gamma} \leq Ch_n^d \), which implies

\[
E\|A_{1i}\|^{2\gamma} \leq Ch_n^d (\tau_2 - \tau_1).
\]

Meanwhile,

\[
E\|A_{2i}\|^{2\gamma} = E \left\{ (\tau_1 - 1(u_i^0(\tau_1) \leq 0)) (z_i,\tau_2 K_i,\tau_2 - z_i,\tau_1 K_i,\tau_1) \right\}^{2\gamma}
\]

\[
\leq E \left\{ (z_i,\tau_2 K_i,\tau_2 - z_i,\tau_1 K_i,\tau_1) \right\}^{2\gamma}
\]

\[
\leq Bh_n^d (\tau_2 - \tau_1)^{2\gamma},
\]

B-4
where the last inequality uses the first result in Lemma B.1. The terms $E \|A_{1i}\|^2$ and $E \|A_{2i}\|^2$ in (T3) can be bounded similarly, giving

$$E \|A_{1i}\|^2 \leq Ch_n^d (\tau_2 - \tau_1) \quad \text{and} \quad E \|A_{2i}\|^2 \leq Bh_n^d (\tau_2 - \tau_1)^2.$$  

These bounds imply

$$\text{(T3)} \quad \leq 2^\gamma C \left( (nh_n^d)^{-\gamma} \sum_{i=1}^{\gamma} (C (\tau_2 - \tau_1) h_n^d + B (\tau_2 - \tau_1) (2 h_n^d)^2) \right)^\gamma$$

$$\leq M (\tau_2 - \tau_1)^\gamma \quad \text{for some constant } M$$

and

$$\text{(T4)} = 2^\gamma C (nh_n^d)^{-\gamma} \sum_{i=1}^{\gamma} \left( (E \|A_{1i}\|^2)^{1/\gamma} + (E \|A_{2i}\|^2)^{1/\gamma} \right)^\gamma$$

$$\leq 2^\gamma C (nh_n^d)^{-\gamma} \sum_{i=1}^{\gamma} \left( (Ch_n^d (\tau_2 - \tau_1))^{1/\gamma} + (Bh_n^d (\tau_2 - \tau_1)^{2/\gamma})^{1/\gamma} \right)^\gamma$$

$$\leq M (nh_n^d)^{1-\gamma} (\tau_2 - \tau_1)$$

$$= M (nh_n^d (\tau_2 - \tau_1))^{1-\gamma} (\tau_2 - \tau_1)^\gamma \quad \text{for some constant } M.$$  

By the definition of $b_n$ in the Lemma, we have $\tau_2 - \tau_1 \geq b_n > (nh_n^d)^{-1}$, which implies $nh_n^d (\tau_2 - \tau_1) > 1$. Consequently, $M (nh_n^d (\tau_2 - \tau_1))^{1-\gamma} < M$ because $\gamma > 1$. Therefore,

$$\text{(T4)} \leq M (\tau_2 - \tau_1)^\gamma.$$  

Combining the (B.5) and (B.6), it follows that each term inside the curly brackets in (B.4) is bounded by $2M (\tau_2 - \tau_1)^\gamma$. Thus, (B.4) is bounded by $(d+1)(2M)^{1/\gamma} (\tau_2 - \tau_1)$. Let $\tilde{C} = (d+1)(2M)^{1/\gamma}$, the proof is complete. $\blacksquare$

The next lemma establishes the stochastic equicontinuity of the process $S_n(\tau,0,0)$.

**Lemma B.3** For any $\varepsilon > 0$ and $\eta > 0$, there exists a $\delta > 0$, such that for large $n$,

$$P \left( \sup_{\tau''',\tau'' \in T, |\tau'' - \tau'''| \leq \delta} \| S_n (\tau'',0,0) - S_n (\tau',0,0) \| > \varepsilon \right) < \eta.$$

**Proof.** Let $z_{i,\tau,j}$ denote the $j$-th component of $z_{i,\tau}$ ($j = 1, \ldots, d+1$). Note that $z_{i,\tau,j} \geq 0$ if and only if $z_{i,\tau,j} \geq 0$ ($j = 1, \ldots, d+1$) for any $\tau'', \tau' \in T$. Without loss of generality, assume that the elements of $z_{i,\tau,j}$ are all nonnegative. Otherwise, let

$$z_{i,\tau,j} = z_{i,\tau,j}^+ - z_{i,\tau,j}^- \equiv z_{i,\tau,j}^+ (z_{i,\tau,j} \geq 0) - (-z_{i,\tau,j})^+ (z_{i,\tau,j} \leq 0)$$

and let $z_{i,\tau}^+ = (z_{i,\tau,1}^+, \ldots, z_{i,\tau,d+1}^+)$ and $z_{i,\tau}^- = (z_{i,\tau,1}^-, \ldots, z_{i,\tau,d+1}^-)$. Then,

$$S_n (\tau,0,0) = (nh_n^d)^{-1/2} \sum_{i=1}^{n} (\tau - 1(u_n^0 (\tau) \leq 0)) z_{i,\tau}^+ K_i$$

$$- (nh_n^d)^{-1/2} \sum_{i=1}^{n} (\tau - 1(u_n^0 (\tau) \leq 0)) z_{i,\tau}^- K_i,$$

B-5
where the elements of \( z_{i,\tau}^+ \) and \( z_{i,\tau}^- \) are nonnegative and the two summations can be analyzed separately. Note that Lemmas B.1 and B.2, which are the only results needed here, still hold when \( z_{i,\tau} \) is replaced by \( z_{i,\tau}^+ \) or \( z_{i,\tau}^- \).

For a given \( \delta, T \) contains at most \( 1/\delta \) intervals of length \( \delta \). Therefore, it suffices to show that for any \( \varepsilon > 0, \eta > 0 \), there exists a \( \delta > 0 \), such that (see Billingsley 1968, p. 58, equation 8.12)

\[
P\left( \sup_{s \leq \tau \leq s + \delta} \| S_n (\tau, 0, 0) - S_n (s, 0, 0) \| > \varepsilon \right) < \delta \eta \text{ holds for all } s \in T \text{ when } n \text{ is large,} \quad (B.8)
\]

(with \( \tau \) in the supremum restricted to \( \tau \in T \)).

To show this, we use a chaining argument. Partition the interval \([s, \delta + s]\) into small intervals of equal sizes with \( 0 < \kappa < 1/2 \). Let \( \tau_j \) denote the lower limit of the \( j \)-th interval. Note that \( \tau_1 = s \). Then, by the triangle inequality

\[
\sup_{s \leq \tau \leq s + \delta} \| S_n (\tau, 0, 0) - S_n (s, 0, 0) \| \leq \sup_{1 \leq j \leq b_n} \sup_{\tau \in [\tau_j, \tau_{j+1}]} \| S_n (\tau, 0, 0) - S_n (\tau_j, 0, 0) \| + \sup_{1 \leq j \leq b_n} \| S_n (\tau_j, 0, 0) - S_n (s, 0, 0) \|. \quad (B.9)
\]

For the first term, we have

\[
S_n (\tau, 0, 0) - S_n (\tau_j, 0, 0) \geq (nh_n^d)^{-1/2} \sum_{i=1}^{n} \left( \tau_j - 1(u_i^0(\tau_{j+1}) < 0) \right) z_{i,\tau} K_{i,\tau} - S_n (\tau_j, 0, 0) = S_n (\tau_{j+1}, 0, 0) - S_n (\tau_j, 0, 0) + (nh_n^d)^{-1/2} \sum_{i=1}^{n} \left( \tau_{j+1} - 1(u_i^0(\tau_{j+1}) < 0) \right) \left( z_{i,\tau} K_{i,\tau} - z_{i,\tau_{j+1}} K_{i,\tau_{j+1}} \right) \quad (T5)
\]

\[
- (nh_n^d)^{-1/2} \sum_{i=1}^{n} (\tau_{j+1} - \tau_j) z_{i,\tau} K_{i,\tau} \quad (T6)
\]

where the inequality uses the non-negativeness of \( z_{i,\tau} K_{i,\tau} \) and that \( \tau_j \) and \( \tau_{j+1} \) are the lower and upper limits of the interval containing \( \tau \), and the equality uses the definition of \( S_n (\tau, 0, 0) \). The terms \( (T5) \) and \( (T6) \) are negligible because

\[
\| (T5) \| \leq (nh_n^d)^{-1/2} \sum_{i=1}^{n} \left\| z_{i,\tau} K_{i,\tau} - z_{i,\tau_{j+1}} K_{i,\tau_{j+1}} \right\| = o_p(1)
\]

by the second result in Lemma B.1, and

\[
\| (T6) \| \leq (nh_n^d)^{-\kappa} \left\{ (nh_n^d)^{-1} \sum_{i=1}^{n} \left\| z_{i,\tau} K_{i,\tau} \right\| \right\} = o_p(1)
\]

B-6
because $|r_{j+1} - r_j| \leq (nh_n^d)^{-1/2} - \kappa$. Therefore

$$S_n (\tau, 0, 0) - S_n (\tau_j, 0, 0) \geq S_n (\tau_{j+1}, 0, 0) - S_n (\tau_j, 0, 0) - \frac{\varepsilon}{5}$$

(B.11)

with probability no less than $1 - \delta \eta$ uniformly in $T$ for large $n$. Meanwhile, there is a reversed inequality for (B.10) given by

$$S_n (\tau, 0, 0) - S_n (\tau_j, 0, 0) \leq (nh_n^d)^{-1/2} \sum_{i=1}^{n} (\tau_{j+1} - 1(u_i^0(\tau_j) \leq 0)) z_{i_r} K_i, \tau - (nh_n^d)^{-1/2} \sum_{i=1}^{n} (\tau_j - 1(u_i^0(\tau_j) < 0)) z_{i_r} K_i, \tau$$

$$+ (nh_n^d)^{-1/2} \sum_{i=1}^{n} (\tau_{j+1} - \tau_j) z_{i_r} K_i, \tau$$

(T7)

$$-(nh_n^d)^{-1/2} \sum_{i=1}^{n} \tau_j (z_{i_r} K_i, \tau - z_{i_r} K_i, \tau)$$

(T9). The terms (T7) and (T9) can be analyzed in the same way as (T5) and are $o_p(1)$ uniformly in $T$. The term (T8) can be analyzed in the same way as (T6) and is also $o_p(1)$ uniformly in $T$. Therefore

$$S_n (\tau, 0, 0) - S_n (\tau_j, 0, 0) \leq \frac{\varepsilon}{5}$$

(B.12)

holds with probability no less than $1 - \delta \eta$ uniformly in $T$ for large $n$. Combining (B.11) and (B.12), along with (B.9), we have

$$\sup_{s \leq \tau \leq \delta + s} \|S_n (\tau, 0, 0) - S_n (s, 0, 0)\|$$

$$\leq \sup_{1 \leq j \leq b_n} \|S_n (\tau_{j+1}, 0, 0) - S_n (\tau_j, 0, 0)\| + \frac{2\varepsilon}{5} + \sup_{1 \leq j \leq b_n} \|S_n (\tau_j, 0, 0) - S_n (s, 0, 0)\|$$

$$\leq \sup_{1 \leq j \leq b_n} 3 \|S_n (\tau_j, 0, 0) - S_n (s, 0, 0)\| + \frac{2\varepsilon}{5},$$

which holds with probability no less than $1 - \delta \eta$ uniformly in $T$ for large $n$, where the second inequality is due to the triangle inequality:

$$\sup_{1 \leq j \leq b_n} \|S_n (\tau_{j+1}, 0, 0) - S_n (\tau_j, 0, 0)\| \leq \sup_{1 \leq j \leq b_n} 2 \|S_n (\tau_j, 0, 0) - S_n (s, 0, 0)\|.$$

Therefore, proving (B.8) boils down to showing

$$P \left( \sup_{1 \leq j \leq b_n} \|S_n (\tau_j, 0, 0) - S_n (\tau_1, 0, 0)\| > \frac{\varepsilon}{5} \right) < \delta \eta$$

for large $n$. To show this, we follow Bai (1996, p.613) and apply Theorem 12.2 in Billingsley (1968). This result allows us to bound the above probability by using the bounds on the moments of $S_n (\tau_{j+1}, 0, 0)$.
In our case, let
\[ S_n(\tau_j, 0, 0), \] established in the previous lemma. Specifically, it states that if there exists \( \beta \geq 0, \alpha > 1 \) and \( u_i \geq 0 \) (\( i = 1, ..., b_n \)) such that

\[
E \left( \| S_n(\tau_j, 0, 0) - S_n(\tau_i, 0, 0) \|^\beta \right) \leq \left( \sum_{i<j} u_i \right)^\alpha, \quad 0 \leq i \leq j \leq b_n,
\]

then
\[
P \left( \sup_{1 \leq j \leq b_n} \| S_n(\tau_j, 0, 0) - S_n(\tau_1, 0, 0) \| > \varepsilon \right) \leq \frac{C_{\beta, \alpha}}{\varepsilon^\alpha} \left( u_1 + ... + u_{b_n} \right)^\alpha.
\]

In our case, let \( \beta = 2\gamma, \alpha = \gamma, \) then Lemma B.2 implies
\[
E \| S_n(\tau_j, 0, 0) - S_n(\tau_i, 0, 0) \|^\beta \leq \tilde{C}(\tau_j - \tau_i)^\alpha, \quad 0 \leq i \leq j \leq b_n,
\]
therefore
\[
P \left( \sup_{1 \leq j \leq b_n} \| S_n(\tau_j, 0, 0) - S_n(\tau_1, 0, 0) \| > \frac{\varepsilon}{5} \right) \leq \frac{C_{\beta, \alpha}}{(\frac{\varepsilon}{5})^\alpha} \tilde{C}(\tau_{b_n} - \tau_1)^\alpha = \delta \left( \frac{C_{\beta, \alpha}}{(\frac{\varepsilon}{5})^\alpha} \tilde{C}\delta^{\alpha-1} \right).
\]

We can choose \( \delta \) such that
\[
\left( \frac{C_{\beta, \alpha}}{(\frac{\varepsilon}{5})^\alpha} \tilde{C}\delta^{\alpha-1} \right) \leq \eta.
\]

This completes the proof. ■

The next two lemmas are used to establish the relationship between \( S_n(\tau, \phi, e_i(\tau)) \) and \( S_n(\tau, 0, e_i(\tau)) \). They are needed to quantify the effect of the parameter estimation on the subgradient.

**Lemma B.4** Let \( b_n = (nh_n^d)^{1/2+\kappa} \) and consider a partition of \( T \) into \( b_n \) intervals of equal sizes. Let \( \tau_j \) denote the lower limit of the \( j \)th interval. Then, under Assumptions 1-6, we have

\[
\sup_{1 \leq j \leq b_n} \sup_{\tau_{j-1} \leq \tau \leq \tau_j} \| (nh_n^d)^{-1/2} \sum_{i=1}^n z_{i,\tau} K_{i,\tau} \left\{ 1(u_i^0(\tau_j) \leq e_i(\tau)) - 1(u_i^0(\tau_j) \leq e_i(\tau)) \right\} \| = o_p(1)
\]

and

\[
\sup_{1 \leq j \leq b_n} \sup_{\tau_{j-1} \leq \tau \leq \tau_j} \sup_{\| \phi \| \leq \log^{1/2}(nh_n^d, \tau)} \| (nh_n^d)^{-1/2} \sum_{i=1}^n z_{i,\tau} K_{i,\tau} \left\{ 1\left( u_i^0(\tau_j) \leq e_i(\tau) \right) + (nh_n^d)^{-1/2} z_{i,\tau} \phi \right\} - (nh_n^d)^{-1/2} \sum_{i=1}^n z_{i,\tau} K_{i,\tau} \left\{ u_i^0(\tau_j) \leq e_i(\tau) \right\} + (nh_n^d)^{-1/2} z_{i,\tau} \phi \| = o_p(1).
\]

**Proof.** Without loss of generality, assume \( z_{i,\tau} \) is a scalar and nonnegative, otherwise, we can repeat the argument as in (B.7) and analyze each element separately.

By the second result in Lemma B.1, the left hand side term in the first result has the same order as

\[
\sup_{1 \leq j \leq b_n} \sup_{\tau_{j-1} \leq \tau \leq \tau_j} \| (nh_n^d)^{-1/2} \sum_{i=1}^n z_{i,\tau} K_{i,\tau} \left\{ 1(u_i^0(\tau_j) \leq e_i(\tau)) - 1(u_i^0(\tau_j) \leq e_i(\tau)) \right\} \|. \quad \text{(B.13)}
\]
Meanwhile, for any $x_i$ satisfying $(x_i - x)/h_{n, \tau} \in \text{supp}(K(\cdot))$, 
\[ e_i(\tau) = -\frac{1}{2} \left( \frac{x_i - x}{h_{n, \tau}} \right)' \frac{\partial^2 Q(\tau|x)}{\partial x \partial x'} \left( \frac{x_i - x}{h_{n, \tau}} \right) h_{n, \tau}^2 + o(h_{n, \tau}^2), \]
where the $o(h_{n, \tau}^2)$ term is uniformly in $\tau$ because of the boundedness and Lipschitz continuity of $\partial^2 Q(\tau|x)/\partial x \partial x'$ stated in Assumption 3. Because $|\tau_j - \tau_{j-1}| \leq (nh_n^d)^{-1/2 - \kappa}$ and $h_{n, \tau} = O\left(n^{-1/(4+d)}\right)$ for all $\tau$, we have 
\[ e_i(\tau) - e_i(\tau_j) = o\left((nh_n^d)^{-1/2}\right) \leq \varepsilon(nh_n^d)^{-1/2}, \]
which holds in probability uniformly in $\tau \in T$, where $\varepsilon$ is a constant that can be made arbitrarily small. Thus, (B.13) is bounded from above by
\[
(T10) = \sup_{1 \leq j \leq b_n} \left\| \sum_{i=1}^{n} z_{i, \tau_j} K_{i, \tau_j} \xi_{i, \tau_j} \right\| \leq \varepsilon \sup_{1 \leq j \leq b_n} \left( e_i(\tau) - \varepsilon(nh_n^d)^{-1/2} \leq u_i(\tau_j) \leq e_i(\tau_j) + \varepsilon(nh_n^d)^{-1/2} \right).
\]
To further bound (T10), let 
\[
\xi_{i, \tau_j} \equiv 1 \left( e_i(\tau_j) - \varepsilon(nh_n^d)^{-1/2} \leq u_i(\tau_j) \leq e_i(\tau_j) + \varepsilon(nh_n^d)^{-1/2} \right)
- E \left\{ 1 \left( e_i(\tau_j) - \varepsilon(nh_n^d)^{-1/2} \leq u_i(\tau_j) \leq e_i(\tau_j) + \varepsilon(nh_n^d)^{-1/2} \right) | x_i \right\},
\]
which has mean zero and, for any $\gamma > 1$, satisfies 
\[
E \left( \|\xi_{i, \tau_j}\|^2 | x_i \right) \leq E \left( \|\xi_{i, \tau_j}\|^2 \right) \leq B(nh_n^d)^{-1/2}
\]
for some constant $B$. Apply triangle inequality to (T10):
\[
(T10) \leq \sup_{1 \leq j \leq b_n} (nh_n^d)^{-1/2} \left\| \sum_{i=1}^{n} z_{i, \tau_j} K_{i, \tau_j} \xi_{i, \tau_j} \right\|
+ \sup_{1 \leq j \leq b_n} (nh_n^d)^{1/2} \sum_{i=1}^{n} E \left\{ 1 \left( e_i(\tau_j) - \varepsilon(nh_n^d)^{-1/2} \leq u_i(\tau_j) \leq e_i(\tau_j) + \varepsilon(nh_n^d)^{-1/2} \right) | x_i \right\}.
\]
The second term is $o_p(1)$ by choosing a small $\varepsilon$. The first term satisfies 
\[
P \left( \sup_{1 \leq j \leq b_n} (nh_n^d)^{-1/2} \left\| \sum_{i=1}^{n} z_{i, \tau_j} K_{i, \tau_j} \xi_{i, \tau_j} \right\| > \eta \right) \leq \sum_{j=1}^{b_n} P \left( (nh_n^d)^{-1/2} \left\| \sum_{i=1}^{n} z_{i, \tau_j} K_{i, \tau_j} \xi_{i, \tau_j} \right\| > \eta \right).
\]
Apply the Rosenthal inequality with $\gamma > 1$ to the $j$-th term on the right side (Hall and Heyde, 1980, p. 23):
\[
P \left( (nh_n^d)^{-1/2} \left\| \sum_{i=1}^{n} z_{i, \tau_j} K_{i, \tau_j} \xi_{i, \tau_j} \right\| > \eta \right)
\leq C \eta^{-2\gamma(nh_n^d)^{-\gamma/2}} \left\{ E \left( (nh_n^d)^{-1/2} \sum_{i=1}^{n} E \left( \|\xi_{i, \tau_j}\|^2 | x_i \right) \right)^\gamma \right. 
+ \left. (nh_n^d)^{-\gamma/2} \sum_{i=1}^{n} E \left( \|z_{i, \tau_j} K_{i, \tau_j}\|^2 E \left( \|\xi_{i, \tau_j}\|^2 | x_i \right) \right) \right\}.
\]
B-9
The term inside the curly brackets is bounded by a constant $M$ independent of $\tau$. Therefore, the preceding display is bounded by $C \eta^{-2\gamma} (nh_n^d)^{-\gamma/2}$, whose summation over $\sum_{j=1}^{b_n}$ is bounded by

$$C \eta^{-2\gamma} (nh_n^d)^{-\gamma/2} (nh_n^d)^{1/2 + \kappa},$$

which converges to zero by choosing $\gamma > 1 + 2\kappa$. This establishes the first result.

For the second result, notice that the difference between $(nh_n^d)^{-1/2} x_i^j \phi$ and $(nh_n^d)^{-1/2} z_i^j \phi$ is of order lower than $(nh_n^d)^{-1/2}$ because of Assumption 5. Therefore, the same arguments as above can be used. The uniformity in $\phi$ can be shown in the same way as in the next Lemma. The detail is omitted. ■

**Lemma B.5** Under Assumptions 1-6, let $K_n = \log^{1/2} (nh_n^d)$, we have

$$\sup_{\tau \in T} \sup_{\|\phi\| \leq K_n} \|S_n (\tau, \phi, e_i (\tau)) - S_n (\tau, 0, e_i (\tau))\| = o_p (1).$$

**Proof.** The proof is similar to Lemma B.3. As in that lemma, assume that the elements of $z_{i,\tau}$ are all nonnegative. We apply a chaining argument. First, partition $T$ into $b_n = (nh_n^d)^{1/2 + \kappa}$ intervals as before and let $\tau_j$ denote the lower limit of the $j$-th interval. Then, as in (B.10), for any $\tau \in [\tau_j, \tau_{j+1}],$

$$S_n (\tau, \phi, e_i (\tau)) - S_n (\tau, 0, e_i (\tau))$$

$$\leq (nh_n^d)^{-1/2} \sum_{i=1}^{n} z_{i,\tau_j} K_{i,\tau_j} 1(u_i^0(\tau_{j+1}) \leq e_i (\tau))$$

$$- (nh_n^d)^{-1/2} \sum_{i=1}^{n} z_{i,\tau_j} K_{i,\tau_j} 1(u_i^0(\tau_j) \leq e_i (\tau) + (nh_n^d)^{-1/2} z_i^j \phi)$$

$$- (nh_n^d)^{-1/2} \sum_{i=1}^{n} z_{i,\tau} K_{i,\tau} E \{1(u_i^0(\tau_j) \leq e_i (\tau)) | x_i \}$$

$$+ (nh_n^d)^{-1/2} \sum_{i=1}^{n} z_{i,\tau} K_{i,\tau} E \{1(u_i^0(\tau_{j+1}) \leq e_i (\tau) + (nh_n^d)^{-1/2} z_i^j \phi) | x_i \}$$

Applying the results in Lemma B.4, we have, for any $\varepsilon > 0,$

$$S_n (\tau, \phi, e_i (\tau)) - S_n (\tau, 0, e_i (\tau))$$

$$\leq (nh_n^d)^{-1/2} \sum_{i=1}^{n} z_{i,\tau_{j+1}} K_{i,\tau_{j+1}} 1(u_i^0(\tau_{j+1}) \leq e_i (\tau_{j+1}))$$

$$- (nh_n^d)^{-1/2} \sum_{i=1}^{n} z_{i,\tau_j} K_{i,\tau_j} 1(u_i^0(\tau_j) \leq e_i (\tau_j) + (nh_n^d)^{-1/2} z_i^j \phi)$$

$$- (nh_n^d)^{-1/2} \sum_{i=1}^{n} z_{i,\tau_j} K_{i,\tau_j} E \{1(u_i^0(\tau_j) \leq e_i (\tau_{j+1})) | x_i \}$$

$$+ (nh_n^d)^{-1/2} \sum_{i=1}^{n} z_{i,\tau} K_{i,\tau} E \{1(u_i^0(\tau_{j+1}) \leq e_i (\tau_j) + (nh_n^d)^{-1/2} z_i^j \phi) | x_i \}$$

$$+ \varepsilon$$

uniformly in $T$ and $\|\phi\| \leq K_n.$
By adding and subtracting terms, the right hand side can be rewritten as

\[ S_n (\tau_j, \phi, e_i (\tau_j)) - S_n (\tau_{j+1}, 0, e_i (\tau_{j+1})) \]

\[ + (nh_n^d)^{-1/2} \sum_{i=1}^{n} z_{i,\tau_{j+1}} K_{i,\tau_{j+1}} E \left\{ 1 \left( u_i^0 (\tau_{j+1}) \leq e_i (\tau_{j+1}) \right) | x_i \right\} \]  \hspace{1cm} (T11)

\[ - (nh_n^d)^{-1/2} \sum_{i=1}^{n} z_{i,\tau_{j+1}} K_{i,\tau_{j+1}} E \left\{ 1 (u_i^0 (\tau_{j}) \leq e_i (\tau_{j+1})) | x_i \right\} \]  \hspace{1cm} (T12)

\[ + (nh_n^d)^{-1/2} \sum_{i=1}^{n} z_{i,\tau_{j}} K_{i,\tau_{j}} E \left\{ 1 \left( u_i^0 (\tau_{j+1}) \leq e_i (\tau_{j}) + (nh_n^d)^{-1/2} z_{i,\tau_{j}} \phi \right) | x_i \right\} \]  \hspace{1cm} (T13)

\[ - (nh_n^d)^{-1/2} \sum_{i=1}^{n} z_{i,\tau_{j}} K_{i,\tau_{j}} E \left\{ 1 \left( u_i^0 (\tau_{j}) \leq e_i (\tau_{j}) + (nh_n^d)^{-1/2} z_{i,\tau_{j}} \phi \right) | x_i \right\} \]  \hspace{1cm} (T14)

\[ + \varepsilon. \]

Apply a first-order Taylor expansion, we have (T11)–(T12) = $O_p \left((nh_n)^{-\kappa}\right) = o_p\left(1\right)$ and (T13)–(T14) = $O_p \left((nh_n)^{-\kappa}\right) = o_p\left(1\right)$ uniformly in $\tau$. Therefore, the preceding display is bounded by

\[ S_n (\tau, \phi, e_i (\tau)) - S_n (\tau, 0, e_i (\tau)) \leq S_n (\tau_j, \phi, e_i (\tau_j)) - S_n (\tau_{j+1}, 0, e_i (\tau_{j+1})) + 2\varepsilon \]

uniformly in $\tau$ and $\|\phi\| \leq K_n$ for large $n$.  

B-11
Meanwhile, a reversed inequality is given by
\[
S_n(\tau, \phi, e_i(\tau)) - S_n(\tau, 0, e_i(\tau))
\]
\[
\geq (nh_n^d)^{-1/2} \sum_{i=1}^{n} z_{i,\tau_j} K_{i,\tau_j} 1(u_0^i(\tau_j) \leq e_i(\tau_j))
\]
\[
-(nh_n^d)^{-1/2} \sum_{i=1}^{n} z_{i,\tau_j} K_{i,\tau_j} 1(\frac{u_0^i(\tau_j)}{z_{i,\tau_j}} \leq e_i(\tau_j)) (nh_n^d)^{-1/2} z_{i,\tau_j+1}^i \phi
\]
\[
-(nh_n^d)^{-1/2} \sum_{i=1}^{n} z_{i,\tau_j} K_{i,\tau_j} E \{ 1(\frac{u_0^i(\tau_j)}{z_{i,\tau_j}} \leq e_i(\tau_j)) \mid x_i \}
\]
\[
+(nh_n^d)^{-1/2} \sum_{i=1}^{n} z_{i,\tau_j+1} K_{i,\tau_j+1} E \{ 1(\frac{u_0^i(\tau_j)}{z_{i,\tau_j}} \leq e_i(\tau_j)) (nh_n^d)^{-1/2} z_{i,\tau_j+1}^i \phi \mid x_i \}
\]
\[+\varepsilon\]
\[= S_n(\tau_{j+1}, \phi, e_i(\tau_{j+1})) - S_n(\tau_{j+1}, 0, e_i(\tau_{j+1}))
\]
\[+ (nh_n^d)^{-1/2} \sum_{i=1}^{n} z_{i,\tau_j} K_{i,\tau_j} E \{ 1(\frac{u_0^i(\tau_j)}{z_{i,\tau_j}} \leq e_i(\tau_j)) \mid x_i \}
\]
\[\geq S_n(\tau_{j+1}, \phi, e_i(\tau_{j+1})) - S_n(\tau_{j+1}, 0, e_i(\tau_{j+1})) + 2\varepsilon\]

Combining the above results:
\[
\sup_{\tau \in T} \sup_{\|\phi\| \leq K_n} \|S_n(\tau, \phi, e_i(\tau)) - S_n(\tau, 0, e_i(\tau))\|
\]
\[
\leq \sup_{1 \leq j \leq b_n} \sup_{\|\phi\| \leq K_n} \|S_n(\tau_{j+1}, \phi, e_i(\tau_{j+1})) - S_n(\tau_{j+1}, 0, e_i(\tau_{j+1}))\|
\]
\[\quad \quad + \sup_{1 \leq j \leq b_n} \sup_{\|\phi\| \leq K_n} \|S_n(\tau_j, \phi, e_i(\tau_j)) - S_n(\tau_{j+1}, 0, e_i(\tau_{j+1}))\| + 4\varepsilon.
\]

Adding and subtracting terms, the first term on the right hand side equals to the sup of
\[
\{ S_n(\tau_{j+1}, \phi, e_i(\tau_{j+1})) - S_n(\tau_{j+1}, 0, e_i(\tau_{j+1})) \} + \{ S_n(\tau_{j+1}, 0, e_i(\tau_{j+1})) - S_n(\tau_{j+1}, 0, 0) \}
\]
\[+ \{ S_n(\tau_{j+1}, 0, 0) - S_n(\tau_{j+1}, 0, 0) \} + \{ S_n(\tau_j, 0, 0) - S_n(\tau_j, 0, e_i(\tau_j)) \},\]

while second term equals to the sup of
\[
\{ S_n(\tau_j, \phi, e_i(\tau_j)) - S_n(\tau_j, 0, e_i(\tau_j)) \} + \{ S_n(\tau_j, 0, e_i(\tau_j)) - S_n(\tau_j, 0, 0) \}
\]
\[+ \{ S_n(\tau_j, 0, 0) - S_n(\tau_j, 0, 0) \} + \{ S_n(\tau_{j+1}, 0, 0) - S_n(\tau_{j+1}, 0, e_i(\tau_{j+1})) \}.
\]
Thus

\[
\sup_{\tau \in T} \sup_{\|\phi\| \leq K_n} \|S_n(\tau, \phi, e_1(\tau)) - S_n(\tau, 0, e_1(\tau))\| \\
\leq 2 \sup_{1 \leq j \leq b_n + 1} \sup_{\|\phi\| \leq K_n} \|S_n(\tau_j, \phi, e_i(\tau_j)) - S_n(\tau_j, 0, e_i(\tau_j))\| \quad (T15)
\]

\[
+ 4 \sup_{1 \leq j \leq b_n + 1} \|S_n(\tau_j, 0, e_i(\tau_j)) - S_n(\tau_j, 0, 0)\| \quad (T16)
\]

\[
+ 4 \sup_{1 \leq j \leq b_n} \|S_n(\tau_j+1, 0, 0) - S_n(\tau_j, 0, 0)\| \quad (T17)
\]

First consider Term (T15). For any \(\delta > 0\), the set \(\{\phi : \|\phi\| \leq K_n\}\) can be partitioned into \(N(\delta)\) spheres such that the diameter of each sphere is less than or equal to \(\delta\). Note that \(N(\delta) = O(K_n^{d+1})\).

Denote the spheres by \(D_h\) with centers being \(\phi_h\) (\(h \in \{1, 2, \ldots, N(\delta)\}\)). For any \(\phi \in D_h\), we have

\[
z_{i,j}^\prime \phi_h - \|z_{i,j}\| \delta \leq z_{i,j}^\prime \phi \leq z_{i,j}^\prime \phi_h + \|z_{i,j}\| \delta.
\]

Therefore,

\[
\sup_{1 \leq j \leq b_n + 1} \sup_{1 \leq h \leq N(\delta)} \sup_{k=1,2,\phi \in D_h} \left(\frac{\sum_{i=1}^{N(\delta)} z_{i,j}^\prime K_{i,j} E \left\{1 \left(u_i^0(\tau_j) \leq e_i(\tau_j) + z_{i,j}^\prime \phi_h + (-1)^k \|z_{i,j}\| \delta\right) \right\} x_i}{\left(\frac{n h_n^d}{\|\phi\|} \right)^{-\frac{1}{2}} \sum_{i=1}^{N(\delta)} z_{i,j}^\prime K_{i,j} \xi_{i,j,h,k}}\right)
\]

where

\[
\xi_{i,j,h,k} = E \left\{1 \left(u_i^0(\tau_j) \leq e_i(\tau_j) + (n h_n^d)^{\frac{1}{2}} z_{i,j}^\prime \phi_h + (-1)^k \|z_{i,j}\| \delta\right) \right\} x_i
\]

\[
-1 \left(u_i^0(\tau_j) \leq e_i(\tau_j) + (n h_n^d)^{\frac{1}{2}} z_{i,j}^\prime \phi_h + (-1)^k \|z_{i,j}\| \delta\right) x_i
\]

\[
-E \left(1 \left(u_i^0(\tau_j) \leq e_i(\tau_j)\right) x_i\right) + 1 \left(u_i^0(\tau_j) \leq e_i(\tau_j)\right) x_i.
\]

Apply a Taylor expansion to the first term after the inequality in (B.15). It is of the same order as \(\delta\), which can be made arbitrarily small by choosing a small \(\delta\). To bound the second norm, it is sufficient to show that for any \(\epsilon > 0, 1 \leq j \leq b_n + 1, 1 \leq h \leq N(\delta)\) and \(k \in \{1, 2\},\)

\[
b_n N(\delta) P \left(\left(\frac{n h_n^d}{\|\phi\|} \right)^{-\frac{1}{2}} \sum_{i=1}^{N(\delta)} z_{i,j}^\prime K_{i,j} \xi_{i,j,h,k} \right) > \epsilon \right) \to 0.
\]

To verify this, first notice that for any \(\gamma > 1,\)

\[
E \left(\|\xi_{i,j,h,k}\|^{2\gamma} x_i\right) \leq E \left(\|\xi_{i,j,h,k}\|^2 x_i\right) \leq B(n h_n^d)^{-\frac{1}{2}} K_n.
\]

B-13
Then, apply the Rosenthal inequality,

\[ b_n N(\delta) P \left( \left\| (nh_n^d)^{-1/2} \sum_{i=1}^{n} z_{i,\tau_j} K_{i,\tau_j} \xi_{i,j,h,k} \right\| > \varepsilon \right) \]

\[ \leq b_n N(\delta) C \varepsilon^{-\gamma (nh_n^d)^{-1/2}} \left\{ E \left( (nh_n^d)^{-1/2} \sum_{i=1}^{n} E \left( \|z_{i,\tau_j} K_{i,\tau_j}\|^2 \right) \right)^{\gamma} \right. \]

\[ + \left( nh_n^d \right)^{-\gamma/2} \sum_{i=1}^{n} E \left( \|z_{i,\tau_j} K_{i,\tau_j}\|^{2\gamma} E \left( \|\xi_{i,j,h,k}\|^2 \right) \right) \}

\[ \leq M \varepsilon^{-2\gamma b_n N(\delta)(nh_n^d)^{-1/2} K_n^\gamma} \text{ for large } n. \]

Now, using the definition of \( b_n, N(\delta) \) and \( K_n \), the preceding line is less than or equal to

\[ M \varepsilon^{-2\gamma b_n N(\delta)(nh_n^d)^{-1/2} K_n^\gamma} \]

for some constant \( M \). Choosing \( \gamma = 1 + 2\kappa + c \), for some \( c > 0 \), the preceding display converges to 0.

(T16) can be analyzed similarly and is also \( o_p(1) \). (T17) = \( o_p(1) \) because of Lemma B.3. ■

**Lemma B.6** Under Assumptions 1-6,

\[ \sup_{\tau \in T} \left\| (nh_n^d)^{-1/2} \sum_{i=1}^{n} \left\{ \psi_{\tau}(u_i^0(\tau) - e_i(\tau)) - \psi_{\tau}(u_i^0(\tau)) \right\} z_{i,\tau} K_{i,\tau} \right\| = O_p(1). \]

**Proof.**

\[ (nh_n^d)^{-1/2} \sum_{i=1}^{n} \left\{ \psi_{\tau}(u_i^0(\tau) - e_i(\tau)) - \psi_{\tau}(u_i^0(\tau)) \right\} z_{i,\tau} K_{i,\tau} \]

\[ = S_n (\tau, 0, 0) + S_n (\tau_j, 0, e_i(\tau)) \]

\[ + (nh_n^d)^{-1/2} \sum_{i=1}^{n} z_{i,\tau} K_{i,\tau} \left\{ P \left( u_i^0(\tau) \leq 0 \right) - E \left( \left( u_i^0(\tau) \leq e_i(\tau) \right) | x_i \right) \right\} + o_p(1) \]

The second to the last line is \( o_p(1) \) as shown in the previous Lemma. The remaining term is \( O_p(1) \) upon a Taylor expansion using the fact that \( e_i(\tau) = O(h_n^2) \). ■
Table 1. The Quantile-Specific Bandwidth: $n = 250$

<table>
<thead>
<tr>
<th>Models</th>
<th>$m = 10$</th>
<th>$m = 20$</th>
<th>$m = 30$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\tau = 0.5$</td>
<td>$\tau = 0.8$</td>
<td>$\tau = 0.5$</td>
</tr>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.50, 0.50)</td>
<td>0.855 (0.622)</td>
<td>0.861 (0.514)</td>
<td>0.897 (0.573)</td>
</tr>
<tr>
<td></td>
<td>0.893 (0.650)</td>
<td>0.900 (0.537)</td>
<td></td>
</tr>
<tr>
<td>(0.75, 0.75)</td>
<td>0.967 (0.633)</td>
<td>0.997 (0.768)</td>
<td>1.000 (0.677)</td>
</tr>
<tr>
<td></td>
<td>1.010 (0.661)</td>
<td>1.042 (0.802)</td>
<td></td>
</tr>
<tr>
<td>(0.90, 0.90)</td>
<td>0.991 (1.006)</td>
<td>0.956 (0.617)</td>
<td>0.972 (0.603)</td>
</tr>
<tr>
<td></td>
<td>1.035 (1.051)</td>
<td>0.998 (0.645)</td>
<td></td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.50, 0.50)</td>
<td>0.285 (0.040)</td>
<td>0.288 (0.042)</td>
<td>0.287 (0.054)</td>
</tr>
<tr>
<td></td>
<td>0.297 (0.042)</td>
<td>0.301 (0.044)</td>
<td></td>
</tr>
<tr>
<td>(0.75, 0.75)</td>
<td>0.239 (0.045)</td>
<td>0.241 (0.046)</td>
<td>0.239 (0.040)</td>
</tr>
<tr>
<td></td>
<td>0.249 (0.047)</td>
<td>0.251 (0.048)</td>
<td></td>
</tr>
<tr>
<td>(0.90, 0.90)</td>
<td>0.290 (0.193)</td>
<td>0.283 (0.099)</td>
<td>0.299 (0.149)</td>
</tr>
<tr>
<td></td>
<td>0.302 (0.201)</td>
<td>0.296 (0.104)</td>
<td></td>
</tr>
<tr>
<td>Model 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.50, 0.50)</td>
<td>0.254 (0.028)</td>
<td>0.255 (0.030)</td>
<td>0.252 (0.029)</td>
</tr>
<tr>
<td></td>
<td>0.265 (0.029)</td>
<td>0.267 (0.031)</td>
<td></td>
</tr>
<tr>
<td>(0.75, 0.75)</td>
<td>0.421 (0.224)</td>
<td>0.404 (0.196)</td>
<td>0.410 (0.212)</td>
</tr>
<tr>
<td></td>
<td>0.440 (0.234)</td>
<td>0.422 (0.204)</td>
<td></td>
</tr>
<tr>
<td>(0.90, 0.90)</td>
<td>0.460 (0.224)</td>
<td>0.441 (0.250)</td>
<td>0.471 (0.375)</td>
</tr>
<tr>
<td></td>
<td>0.480 (0.234)</td>
<td>0.461 (0.262)</td>
<td></td>
</tr>
</tbody>
</table>

We report the means of the selected bandwidths at two representative quantiles and their standard deviations (in parenthesis). $n$ is the sample size and $m$ is the number of quantiles in the grid. The results are based on 500 replications.
Table 2. The Quantile-Specific Bandwidth: $n = 500$

<table>
<thead>
<tr>
<th>Models</th>
<th>$m = 10$</th>
<th></th>
<th>$m = 20$</th>
<th></th>
<th>$m = 30$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\tau = 0.5$</td>
<td>$\tau = 0.8$</td>
<td>$\tau = 0.5$</td>
<td>$\tau = 0.8$</td>
<td>$\tau = 0.5$</td>
<td>$\tau = 0.8$</td>
</tr>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.50, 0.50)</td>
<td>0.870 (0.475)</td>
<td>0.909 (0.496)</td>
<td>0.871 (0.734)</td>
<td>0.910 (0.767)</td>
<td>0.917 (0.572)</td>
<td>0.958 (0.597)</td>
</tr>
<tr>
<td>(0.75, 0.75)</td>
<td>0.979 (0.579)</td>
<td>1.002 (0.604)</td>
<td>1.027 (0.773)</td>
<td>1.072 (0.807)</td>
<td>0.958 (0.555)</td>
<td>1.001 (0.580)</td>
</tr>
<tr>
<td>(0.90, 0.90)</td>
<td>0.963 (0.661)</td>
<td>1.006 (0.691)</td>
<td>0.950 (0.541)</td>
<td>0.992 (0.566)</td>
<td>0.964 (0.594)</td>
<td>1.007 (0.620)</td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.50, 0.50)</td>
<td>0.251 (0.025)</td>
<td>0.263 (0.026)</td>
<td>0.252 (0.025)</td>
<td>0.263 (0.026)</td>
<td>0.252 (0.024)</td>
<td>0.263 (0.025)</td>
</tr>
<tr>
<td>(0.75, 0.75)</td>
<td>0.207 (0.022)</td>
<td>0.217 (0.023)</td>
<td>0.211 (0.023)</td>
<td>0.221 (0.024)</td>
<td>0.210 (0.026)</td>
<td>0.219 (0.027)</td>
</tr>
<tr>
<td>(0.90, 0.90)</td>
<td>0.223 (0.053)</td>
<td>0.233 (0.055)</td>
<td>0.223 (0.040)</td>
<td>0.233 (0.041)</td>
<td>0.223 (0.039)</td>
<td>0.233 (0.041)</td>
</tr>
<tr>
<td>Model 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.50, 0.50)</td>
<td>0.229 (0.020)</td>
<td>0.239 (0.020)</td>
<td>0.230 (0.020)</td>
<td>0.240 (0.021)</td>
<td>0.229 (0.019)</td>
<td>0.239 (0.020)</td>
</tr>
<tr>
<td>(0.75, 0.75)</td>
<td>0.408 (0.244)</td>
<td>0.427 (0.254)</td>
<td>0.387 (0.163)</td>
<td>0.404 (0.170)</td>
<td>0.410 (0.209)</td>
<td>0.428 (0.218)</td>
</tr>
<tr>
<td>(0.90, 0.90)</td>
<td>0.405 (0.370)</td>
<td>0.423 (0.386)</td>
<td>0.372 (0.155)</td>
<td>0.389 (0.162)</td>
<td>0.395 (0.253)</td>
<td>0.412 (0.265)</td>
</tr>
</tbody>
</table>

We report the means of the selected bandwidths at two representative quantiles and their standard deviations (in parenthesis). $n$ is the sample size and $m$ is the number of quantiles in the grid. The results are based on 500 replications.
Table 3. Root Mean Integrated Squared Error (RMISE)

<table>
<thead>
<tr>
<th>Models</th>
<th>$n=250$</th>
<th></th>
<th>$n=500$</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(0.5, 0.5)</td>
<td>(0.75, 0.75)</td>
<td>(0.9, 0.9)</td>
<td>(0.5, 0.5)</td>
</tr>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proposed</td>
<td>0.03760</td>
<td>0.07229</td>
<td>0.12406</td>
<td>0.02548</td>
</tr>
<tr>
<td>Local Linear</td>
<td>0.03760 (0.00000)</td>
<td>0.07230</td>
<td>0.12418</td>
<td>0.02548</td>
</tr>
<tr>
<td>(Crossing)</td>
<td></td>
<td>(0.03800)</td>
<td>(0.72200)</td>
<td>(0.00000)</td>
</tr>
<tr>
<td>Rearrangement</td>
<td>0.03760</td>
<td>0.07229</td>
<td>0.12397</td>
<td>0.02548</td>
</tr>
<tr>
<td>QR</td>
<td>0.03412</td>
<td>0.05786</td>
<td>0.07812</td>
<td>0.02361</td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proposed</td>
<td>0.15552</td>
<td>0.22889</td>
<td>0.29916</td>
<td>0.12569</td>
</tr>
<tr>
<td>Local linear</td>
<td>0.15552 (0.05000)</td>
<td>0.22890</td>
<td>0.29976</td>
<td>0.12569</td>
</tr>
<tr>
<td>(Crossing)</td>
<td></td>
<td>(0.11400)</td>
<td>(0.74200)</td>
<td>(0.00600)</td>
</tr>
<tr>
<td>Rearrangement</td>
<td>0.15552</td>
<td>0.22889</td>
<td>0.29862</td>
<td>0.12569</td>
</tr>
<tr>
<td>QR</td>
<td>0.48764</td>
<td>0.42703</td>
<td>0.21227</td>
<td>0.49118</td>
</tr>
<tr>
<td>Model 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Proposed</td>
<td>0.13344</td>
<td>0.15490</td>
<td>0.33980</td>
<td>0.10339</td>
</tr>
<tr>
<td>Local linear</td>
<td>0.13344 (0.08000)</td>
<td>0.15490</td>
<td>0.34043</td>
<td>0.10339</td>
</tr>
<tr>
<td>(Crossing)</td>
<td></td>
<td>(0.08200)</td>
<td>(0.91200)</td>
<td>(0.01400)</td>
</tr>
<tr>
<td>Rearrangement</td>
<td>0.13344</td>
<td>0.15490</td>
<td>0.33921</td>
<td>0.10339</td>
</tr>
<tr>
<td>QR</td>
<td>0.52311</td>
<td>0.15542</td>
<td>0.62234</td>
<td>0.52754</td>
</tr>
</tbody>
</table>

The domain of the conditional quantile process: $\mathcal{T} = [0.2, 0.8]$. The number of quantiles in the grid: $m = 30$. We report the mean of the RMISE over 500 simulation replications. "Proposed" denotes the proposed estimator. "Local linear" is the quantile-by-quantile application of the local linear estimator. "Rearrangement" is the rearranged version of "Local linear". "QR" is obtained from the conventional linear quantile regression. "Crossing" denotes the fraction of simulation runs in which "Local linear" has quantile-crossing.
Table 4. Coverage Ratio of the Modified Confidence Bands

<table>
<thead>
<tr>
<th>Models</th>
<th>$m = 10$</th>
<th></th>
<th>$m = 20$</th>
<th></th>
<th>$m = 30$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n = 250$</td>
<td>$n = 500$</td>
<td>$n = 250$</td>
<td>$n = 500$</td>
<td>$n = 250$</td>
</tr>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.50, 0.50)</td>
<td>0.968</td>
<td>0.980</td>
<td>0.982</td>
<td>0.988</td>
<td>0.974</td>
</tr>
<tr>
<td>(0.75, 0.75)</td>
<td>0.986</td>
<td>0.986</td>
<td>0.994</td>
<td>0.994</td>
<td>0.986</td>
</tr>
<tr>
<td>(0.90, 0.90)</td>
<td>0.962</td>
<td>0.982</td>
<td>0.962</td>
<td>0.988</td>
<td>0.990</td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.50, 0.50)</td>
<td>0.842</td>
<td>0.888</td>
<td>0.832</td>
<td>0.910</td>
<td>0.876</td>
</tr>
<tr>
<td>(0.75, 0.75)</td>
<td>0.904</td>
<td>0.938</td>
<td>0.918</td>
<td>0.966</td>
<td>0.928</td>
</tr>
<tr>
<td>(0.90, 0.90)</td>
<td>0.872</td>
<td>0.970</td>
<td>0.900</td>
<td>0.972</td>
<td>0.912</td>
</tr>
<tr>
<td>Model 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.50, 0.50)</td>
<td>0.956</td>
<td>0.968</td>
<td>0.970</td>
<td>0.982</td>
<td>0.968</td>
</tr>
<tr>
<td>(0.75, 0.75)</td>
<td>0.910</td>
<td>0.938</td>
<td>0.930</td>
<td>0.946</td>
<td>0.942</td>
</tr>
<tr>
<td>(0.90, 0.90)</td>
<td>0.866</td>
<td>0.932</td>
<td>0.880</td>
<td>0.948</td>
<td>0.892</td>
</tr>
</tbody>
</table>

Confidence Bands with the modified bias adjustment (90% nominal level). Results are based on 500 Replications. $m$ is the number of quantiles in the grid and $n$ is the sample size.

Table 5. Coverage Ratio of the Conventional Confidence Bands

<table>
<thead>
<tr>
<th>Models</th>
<th>$m = 10$</th>
<th></th>
<th>$m = 20$</th>
<th></th>
<th>$m = 30$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$n = 250$</td>
<td>$n = 500$</td>
<td>$n = 250$</td>
<td>$n = 500$</td>
<td>$n = 250$</td>
</tr>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.50, 0.50)</td>
<td>0.640</td>
<td>0.680</td>
<td>0.706</td>
<td>0.694</td>
<td>0.670</td>
</tr>
<tr>
<td>(0.75, 0.75)</td>
<td>0.876</td>
<td>0.918</td>
<td>0.912</td>
<td>0.938</td>
<td>0.940</td>
</tr>
<tr>
<td>(0.90, 0.90)</td>
<td>0.612</td>
<td>0.640</td>
<td>0.612</td>
<td>0.668</td>
<td>0.666</td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.50, 0.50)</td>
<td>0.834</td>
<td>0.884</td>
<td>0.816</td>
<td>0.904</td>
<td>0.860</td>
</tr>
<tr>
<td>(0.75, 0.75)</td>
<td>0.882</td>
<td>0.916</td>
<td>0.898</td>
<td>0.960</td>
<td>0.916</td>
</tr>
<tr>
<td>(0.90, 0.90)</td>
<td>0.840</td>
<td>0.940</td>
<td>0.866</td>
<td>0.966</td>
<td>0.856</td>
</tr>
<tr>
<td>Model 3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.50, 0.50)</td>
<td>0.902</td>
<td>0.924</td>
<td>0.894</td>
<td>0.938</td>
<td>0.934</td>
</tr>
<tr>
<td>(0.75, 0.75)</td>
<td>0.750</td>
<td>0.746</td>
<td>0.770</td>
<td>0.778</td>
<td>0.848</td>
</tr>
<tr>
<td>(0.90, 0.90)</td>
<td>0.742</td>
<td>0.870</td>
<td>0.778</td>
<td>0.892</td>
<td>0.778</td>
</tr>
</tbody>
</table>

Confidence Bands with the conventional bias adjustment (90% nominal level). Results are based on 500 Replications. $m$ is the number of quantiles in the grid and $n$ is the sample size.
Table 6. The Relative Length of Two Confidence Bands: $n = 250$

<table>
<thead>
<tr>
<th>Models</th>
<th>$m = 10$</th>
<th>$m = 20$</th>
<th>$m = 30$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\tau = 0.5$</td>
<td>$\tau = 0.8$</td>
<td>$\tau = 0.5$</td>
</tr>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(0.50, 0.50)$</td>
<td>1.155</td>
<td>1.158</td>
<td>1.136</td>
</tr>
<tr>
<td></td>
<td>(0.103)</td>
<td>(0.119)</td>
<td>(0.088)</td>
</tr>
<tr>
<td></td>
<td>1.105</td>
<td>1.100</td>
<td>1.096</td>
</tr>
<tr>
<td></td>
<td>(0.062)</td>
<td>(0.073)</td>
<td>(0.061)</td>
</tr>
<tr>
<td></td>
<td>1.149</td>
<td>1.195</td>
<td>1.137</td>
</tr>
<tr>
<td></td>
<td>(0.137)</td>
<td>(0.215)</td>
<td>(0.117)</td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(0.50, 0.50)$</td>
<td>1.157</td>
<td>1.134</td>
<td>1.147</td>
</tr>
<tr>
<td></td>
<td>(0.052)</td>
<td>(0.064)</td>
<td>(0.044)</td>
</tr>
<tr>
<td></td>
<td>1.150</td>
<td>1.125</td>
<td>1.141</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.062)</td>
<td>(0.034)</td>
</tr>
<tr>
<td></td>
<td>1.092</td>
<td>1.074</td>
<td>1.075</td>
</tr>
<tr>
<td></td>
<td>(0.182)</td>
<td>(0.159)</td>
<td>(0.075)</td>
</tr>
<tr>
<td>Model 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$(0.50, 0.50)$</td>
<td>1.153</td>
<td>1.134</td>
<td>1.141</td>
</tr>
<tr>
<td></td>
<td>(0.032)</td>
<td>(0.048)</td>
<td>(0.029)</td>
</tr>
<tr>
<td></td>
<td>1.111</td>
<td>1.101</td>
<td>1.101</td>
</tr>
<tr>
<td></td>
<td>(0.052)</td>
<td>(0.089)</td>
<td>(0.045)</td>
</tr>
<tr>
<td></td>
<td>1.115</td>
<td>1.105</td>
<td>1.097</td>
</tr>
<tr>
<td></td>
<td>(0.302)</td>
<td>(0.255)</td>
<td>(0.285)</td>
</tr>
</tbody>
</table>

We report means of the ratios. The standard deviations are in parenthesis. $m$ is the number of quantiles in the grid and $n$ is the sample size. The results are based on 500 replications.
Table 7. The Relative Length of Two Confidence Bands: $n = 500$

<table>
<thead>
<tr>
<th>Models</th>
<th>$m = 10$</th>
<th>$m = 20$</th>
<th>$m = 30$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$\tau = 0.5$</td>
<td>$\tau = 0.8$</td>
<td>$\tau = 0.5$</td>
</tr>
<tr>
<td>Model 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.50, 0.50)</td>
<td>1.147 (0.095)</td>
<td>1.153 (0.119)</td>
<td>1.141 (0.090)</td>
</tr>
<tr>
<td>(0.75, 0.75)</td>
<td>1.103 (0.062)</td>
<td>1.099 (0.081)</td>
<td>1.098 (0.057)</td>
</tr>
<tr>
<td>(0.90, 0.90)</td>
<td>1.144 (0.115)</td>
<td>1.179 (0.182)</td>
<td>1.133 (0.123)</td>
</tr>
<tr>
<td>Model 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.50, 0.50)</td>
<td>1.150 (0.027)</td>
<td>1.131 (0.043)</td>
<td>1.142 (0.030)</td>
</tr>
<tr>
<td>(0.75, 0.75)</td>
<td>1.152 (0.029)</td>
<td>1.127 (0.054)</td>
<td>1.141 (0.023)</td>
</tr>
<tr>
<td>(0.90, 0.90)</td>
<td>1.087 (0.116)</td>
<td>1.069 (0.090)</td>
<td>1.073 (0.027)</td>
</tr>
<tr>
<td>Model 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(0.50, 0.50)</td>
<td>1.152 (0.024)</td>
<td>1.133 (0.037)</td>
<td>1.140 (0.023)</td>
</tr>
<tr>
<td>(0.75, 0.75)</td>
<td>1.106 (0.048)</td>
<td>1.100 (0.083)</td>
<td>1.098 (0.048)</td>
</tr>
<tr>
<td>(0.90, 0.90)</td>
<td>1.063 (0.142)</td>
<td>1.070 (0.210)</td>
<td>1.057 (0.145)</td>
</tr>
</tbody>
</table>

We report means of the ratios. The standard deviations are in parenthesis. $m$ is the number of quantiles in the grid and $n$ is the sample size. The results are based on 500 replications.
Figure 1. Surface Plots of Conditional Quantile Functions.

Model 2, $\tau = 0.5$

Model 2, $\tau = 0.8$

Model 3, $\tau = 0.5$

Model 3, $\tau = 0.8$