Causal Inference by Quantile Regression Kink Designs

Harold D. Chiang∗ Yuya Sasaki†‡
Johns Hopkins University

August 30, 2016

Abstract

The quantile regression kink design (QRKD) is proposed by empirical researchers as a potential method to assess heterogeneous treatment effects under suitable research designs, but its causal interpretation remains unknown. We propose causal interpretations of the QRKD estimand. Under flexible heterogeneity and endogeneity, the sharp and fuzzy QRKD estimands measure weighted averages of heterogeneous marginal effects at respective conditional quantiles of outcome given a designed kink point. In addition, we develop weak convergence results for the QRKD estimator as a local quantile process for the purpose of conducting statistical inference of heterogeneous treatment effects using the QRKD. Applying our methods to the Continuous Wage and Benefit History Project (CWBH) data, we find significantly heterogeneous positive moral hazard effects of unemployment insurance benefits on unemployment durations in Louisiana between 1981 and 1982. We find that these effects are larger for individuals with longer unemployment durations.

Keywords: causal inference, heterogeneous treatment effects, identification, regression kink design, quantile regression, unemployment duration.

∗hchiang7@jhu.edu. Department of Economics, Johns Hopkins University.
†sasaki@jhu.edu. Department of Economics, Johns Hopkins University.
‡We would like to thank Patty Anderson and Bruce Meyer for kindly agreeing to our use of the CWBH data. We benefited from very useful comments by Andrew Chesher, Antonio Galvao, Emmanuel Guerre, Blaise Melly, Jungmo Yoon, seminar participants at Academia Sinica, National Chengchi University, National University of Singapore, Otaru University of Commerce, Penn State University, Queen Mary University of London, University of Pittsburgh and University of Surrey, and conference participants at AMES 2016, CeMMAP-SNU-Tokyo Conference on Advances in Microeconometrics, ESEM 2016, New York Camp Econometrics XI. All remaining errors are ours.
1 Introduction

Some recent empirical research papers, including Nielsen, Sørensen and Taber (2010), Landais (2015), Simonsen, Skipper and Skipper (2015), Card, Lee, Pei and Weber (2016), and Dong (2016), conduct causal inference via the regression kink design (RKD). A natural extension of the RKD with a flavor of unobserved heterogeneity is the quantile RKD (QRKD), which is the object that we explore in this paper. Specifically, consider the quantile derivative Wald ratio of the form

\[
QRKD(\tau) = \frac{\lim_{x \downarrow x_0} \partial x \partial Q_Y|X(\tau | x) - \lim_{x \uparrow x_0} \partial x \partial Q_Y|X(\tau | x)}{\lim_{x \downarrow x_0} \frac{d}{dx} b(x) - \lim_{x \uparrow x_0} \frac{d}{dx} b(x)},
\]  

(1.1)

at a design point \(x_0\) of a running variable \(x\), where \(Q_Y|X(\tau | x) := \inf\{y : F(y|x) \geq \tau\}\) denotes the \(\tau\)-th conditional quantile function of \(Y\) given \(X = x\), and \(b\) is a policy function. Note that it is analogous to the RKD estimand of Card, Lee, Pei and Weber (2016):

\[
RKD = \frac{\lim_{x \downarrow x_0} \partial x \partial E[Y | X = x] - \lim_{x \uparrow x_0} \partial x \partial E[Y | X = x]}{\lim_{x \downarrow x_0} \frac{d}{dx} b(x) - \lim_{x \uparrow x_0} \frac{d}{dx} b(x)},
\]  

(1.2)

except that the conditional expectations in the numerator are replaced by the corresponding conditional quantiles. While the QRKD estimand (1.1) is of potential interest in the empirical literature for assessment of heterogeneous treatment effects (e.g., Landais (2011) considers (1.1)), little seems known about theories of identification, estimation, and inference. This paper develops causal interpretation (identification) and estimation theories for the QRKD estimand (1.1). Consequently, we also propose methods of inference for heterogeneous treatment effects based on the QRKD.

Causal analysis for types of estimands in the Wald-ratio form, like (1.1) and (1.2), dates back to the seminal paper by Imbens and Angrist (1994). Making causal interpretations of the QRKD estimand (1.1) is perhaps more challenging than the mean RKD estimand (1.2) because the differentiation operator \(\frac{d}{dx}\) and the conditional quantile do not ‘swap.’ For the mean RKD estimand (1.2), the interchangeability of the differentiation operator and the expectation (integration) operator allows each term of the numerator in (1.2) to be additively decomposed into two parts, namely the causal effects and the endogeneity effects. Taking the difference of two terms in the numerator then cancels out
the endogeneity effects, leaving only the causal effects. This trick allows the mean RKD estimand (1.2) to have causal interpretations in the presence of endogeneity. Due to the lack of such interchangeability for the case of quantiles, this trick is not straightforwardly inherited by the quantile counterpart (1.1).

Having said this, we show in Section 2.1 and more formally in Section 2.2 that a similar decomposition is possible for the QRKD estimand (1.1), and therefore argue that its causal interpretations are possible even under the lack of monotonicity. Specifically, we show that the QRKD estimand corresponds to the quantile marginal effect under monotonicity and to a weighted average of marginal effects under non-monotonicity and/or with fuzzy design in a similar manner to Card, Lee, Pei and Weber (2016).

For estimation of the causal effects, we propose a sample-counterpart estimator for the QRKD estimand (1.1) in Sections 3. To derive its asymptotic properties, we take advantage of the existing literature on uniform Bahadur representations for quantile-type loss functions, including Kong, Linton and Xia (2010), Guerre and Sabbah (2012), Sabbah (2014), and Qu and Yoon (2015a). Qu and Yoon (2015b) apply the results of Qu and Yoon (2015a) to develop methods of statistical inference with quantile regression discontinuity designs (QRDD), which are closely related to our QRKD framework. We take a similar approach with suitable modifications to derive asymptotic properties of our QRKD estimator. Weak convergence results for the estimator as a quantile process are derived. Applying the weak convergence results, we propose procedures for testing treatment significance and treatment heterogeneity following Koenker and Xiao (2002), Chernozhukov and Fernández-Val (2005) and Qu and Yoon (2015b). Simulation studies presented in Section 4 support the theoretical properties.

Literature: The method studied in this paper falls in the broad framework of design-based causal inference, including RDD and RKD. There is an extensive body of literature on RDD by now – see for example the special issue of *Journal of Econometrics* edited by Imbens and Lemieux (2008) and the literature reviews by Imbens and Wooldridge (2009; Sec. 6.4) and Lee and Lemieux (2010). The first extension to quantile treatment effects in the RDD framework was made by Frandsen, Frölich and Melly (2012). More recently, Qu and Yoon (2015b) develop uniform inference methods
with QRDD that empirical researchers can use to test a variety of important empirical questions on heterogeneous treatment effects. While the RDD has a rich set of empirical and theoretical results including the quantile extensions, the RKD method which developed more recently does not have a quantile counterpart in the literature yet, despite potential demands for it by empirical researchers (e.g., Landais, 2011). Our paper can be seen as either a quantile extension to Card, Lee, Pei and Weber (2016) or a RKD counterpart of Frandsen, Frölich and Melly (2012) and Qu and Yoon (2015b).

## 2 Causal Interpretation of the QRKD Estimand

In this section, we develop some causal interpretations of the QRKD estimand (1.1). For the purpose of illustration, we first present a simple case with rank invariance in Section 2.1. It is followed by a formal argument for general cases in Section 2.2.

### 2.1 Illustration: Causal Interpretation under Rank Invariance

The causal relation of interest is represented by the structural equation

\[ y = g(b, x, \epsilon). \]

The outcome \( y \) is determined through the structural function \( g \) by two observed factors, \( b \in \mathbb{R} \) and \( x \in \mathbb{R} \), and a scalar unobserved factor, \( \epsilon \in \mathbb{R} \). We assume that \( g \) is monotone increasing in \( \epsilon \), effectively imposing the rank invariance; causal interpretations in a more general setup with non-monotone \( g \) and/or multivariate \( \epsilon \) is established in Section 2.2. The factor \( b \) is a treatment input, and is in turn determined by the running variable \( x \) through the structural equation

\[ b = b(x) \]

for a known policy function \( b \). We say that \( b \) has a kink at \( x_0 \) if

\[ b'(x_0^+) := \lim_{x \to x_0^+} \frac{db(x)}{dx} \neq \lim_{x \to x_0^-} \frac{db(x)}{dx} =: b'(x_0^-) \]

is true, where \( x \to x_0^+ \) and \( x \to x_0^- \) mean \( x \downarrow x_0 \) and \( x \uparrow x_0 \), respectively.
Throughout this paper, we assume that the location, \( x_0 \), of the kink is known from a policy-based research design, as is the case with Card, Lee, Pei and Weber (2016).

**Assumption 1.** \( b'(x_0^+) \neq b'(x_0^-) \) holds, and \( b \) is continuous on \( \mathbb{R} \) and differentiable on \( \mathbb{R} \setminus \{x_0\} \).

The structural partial effects are \( g_1(b, x, \epsilon) := \frac{\partial}{\partial b} g(b, x, \epsilon), g_2(b, x, \epsilon) := \frac{\partial}{\partial x} g(b, x, \epsilon) \) and \( g_3(b, x, \epsilon) := \frac{\partial}{\partial \epsilon} g(b, x, \epsilon) \). In particular, a researcher is interested in \( g_1 \) which measures heterogeneous partial effects of the treatment intensity \( b \) on an outcome \( y \). While the structural partial effect \( g_1 \) is of interest, it is not clear if the QRKD estimand (1.1) provides any information about \( g_1 \). In this section, we argue that (1.1) does have a causal interpretation in the sense that it measures the structural causal effect \( g_1(b(x_0), x_0, \epsilon) \) at the \( \tau \)-th conditional quantile of \( \epsilon \) given \( X = x_0 \).

Under regularity conditions (to be discussed in Section 2.2 in detail), some calculations yield the decomposition

\[
\frac{\partial}{\partial x} Q_{Y|X}(\tau \mid x) = g_1(b(x), x, \epsilon) \cdot b'(x) + g_2(b(x), x, \epsilon) - \int_{-\infty}^{\epsilon} \frac{\partial}{\partial x} f_{\epsilon|X}(e \mid x) de \cdot g_3(b(x), x, \epsilon),
\]

(2.1)

where \( \tau = F_{\epsilon|X}(\epsilon \mid x) \). The first term on the right-hand side is the partial effect of the running variable \( x \) on the outcome \( y \) through the policy function \( b \). The second term is the direct partial effect of the running variable on the outcome \( y \). The third term measures the effect of endogeneity in the running variable \( x \). We can see that this third term is zero under exogeneity, \( \frac{\partial}{\partial x} f_{\epsilon|X} = 0 \). In order to get the causal effect \( g_1(b(x), x, \epsilon) \) of interest through the QRKD estimand (1.1), therefore, we want to remove the last two terms in (2.1).

Suppose that the designed kink condition of Assumption 1 is true, but all the other functions, \( g_1, g_2, g_3, 1/f_{\epsilon|X} \) and \( \frac{\partial}{\partial x} f_{\epsilon|X} \), in the right-hand side of (2.1) are continuous in \( (b, x) \) at \( (b(x_0), x_0) \). Then, (2.1) yields

\[
\frac{\partial}{\partial x} Q_{Y|X}(\tau \mid x_0^+) - \frac{\partial}{\partial x} Q_{Y|X}(\tau \mid x_0^-) = g_1(b(x_0), x_0, \epsilon),
\]

showing that the QRKD estimand (1.1) measures the structural causal effect \( g_1(b(x_0), x_0, \epsilon) \) of \( b \) on \( y \) for the subpopulation of individuals at the \( \tau \)-th conditional quantile of \( \epsilon \) given \( X = x_0 \). This
section provides only an informal argument for ease of exposition, but Section 2.2 provides a formal
mathematical argument under a general setup without the rank invariance assumption. Furthermore,
we provide a result for the case of fuzzy QRKD in Section A.1 in the supplementary appendix.

2.2 General Result: Causal Interpretation without Rank Invariance

In this section, we continue to use the basic settings from Section 2.1 except that the unobserved
factors \( \epsilon \) are now allowed to be \( M \)-dimensional, as opposed to be a scalar. As such, we can consider
general structural functions \( g \) without the rank invariance. Define the lower contour set of \( \epsilon \) evaluated
by \( g(b(x), x, \cdot) \) below a given level of \( y \) as follows:

\[
V(y, x) = \{ \epsilon \in \mathbb{R}^M | g(b(x), x, \epsilon) \leq y \}.
\]

Its boundary is denoted by \( \partial V(y, x) \). Furthermore, the velocities of the boundary \( \partial V(y, x) \) at \( u \)
with respect to a change in \( y \) and a change in \( x \) are denoted by \( \partial v(y, x; u) / \partial y \) and \( \partial v(y, x; u) / \partial x \),
respectively. \( \Sigma \) denotes an \( (M - 1) \)-dimensional rectangle. For a short hand notation, we write
\( h(x, \epsilon) = g(b(x), x, \epsilon) \) and \( h_1(x, \epsilon) = \partial h(x, \epsilon) / \partial x \). Let \( m^M \) and \( H^{M-1} \)
denote the Lebesgue measure on \( \mathbb{R}^M \) and the Hausdorff measure on \( \partial V(y, x) \), respectively.\(^1\) Letting \( X = \text{supp}(X) \), we make the following assumptions.

Assumption 2. (i) \( h(\cdot, \epsilon) \) is continuously differentiable on \( X \setminus \{x_0\} \) for all \( \epsilon \in \mathcal{E} \) and \( h(x, \cdot) \) is
continuously differentiable for all \( x \in X \). (ii) \( \| \nabla \epsilon h(x, \cdot) \| \neq 0 \) on \( \partial V(y, x) \) for all \( (x, y) \in X \times Y \).
(iii) The conditional distribution of \( \epsilon \) given \( X \) is absolutely continuous with respect to \( m^M \), \( f_{\epsilon|X} \) is
continuously differentiable, and \( f_{\epsilon|X} \in C^1(X; L^1(\mathbb{R}^M)) \) is true. (iv) \( \int_{\partial V(y, x)} f_{\epsilon|X}(\epsilon | x) dH^{M-1}(\epsilon) > 0 \)
for all \( (x, y) \in X \times Y \). (v) \( \partial V(y, \cdot) \in C^1(\Sigma \times X; \mathbb{R}^M) \) holds for all \( y \in Y \) and \( \partial V(\cdot, x) \in C^1(\Sigma \times Y; \mathbb{R}^M) \)
holds for all \( x \in X \). (vi) \( \partial v(y, \cdot; \cdot) / \partial x \in C^1(\Sigma; L^1(\Sigma)) \) holds for all \( y \in Y \) and \( \partial v(\cdot, x; \cdot) / \partial y \in
\]

\(^1\)We obtain the \( (M - 1) \)-dimensional Hausdorff measure by the restriction of the function \( H^{M-1} : 2^{\mathbb{R}^M} \to \mathbb{R} \) defined
by \( H^{M-1}(S) = \sup_{\delta > 0} \inf \left\{ \sum_{i=1}^\infty (\text{diam}U_i)^{M-1} \mid \bigcup_{i=1}^\infty S_i \supset S, \ \text{diam}S_i < \delta \right\} \) to the collection of Carathéodory-measurable sets.
$C^1(\mathcal{Y}; L^1(\Sigma))$ holds for all $x \in \mathcal{X}$.

**Assumption 3.** Let $\gamma(x, \epsilon) := \|\nabla h(x, \epsilon)\|^{-1}$. There exist values $p \geq 1$ and $q \geq 1$ satisfying $p^{-1} + q^{-1} = 1$ such that $\|\gamma(x, \cdot)\|_{L^p(\partial \mathcal{V}(y, x), H^{M-1})} < \infty$ and $\|f_\epsilon\|_{L^q(\partial \mathcal{V}(y, x), H^{M-1})} < \infty$ hold for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$.

**Assumption 4.** There exists a function $w \in L^1(\partial \mathcal{V}(y, x), H^{M-1})$ such that $|\gamma(x, \epsilon) h(x, \epsilon) f_\epsilon x| \leq w(\epsilon)$ for all $\epsilon \in \times \mathcal{V}(y, x)$ for all $x \in \mathcal{X}$.

**Assumption 5.** $\lim_{x \to x^+} \frac{\partial}{\partial x} Q_Y|X(\tau|x)$ and $\lim_{x \to x^-} \frac{\partial}{\partial x} Q_Y|X(\tau|x)$ exist.

Assumptions 2 and 3 are used to derive a structural decomposition of the quantile partial derivative – see Sasaki (2015) for detailed discussions of these assumptions. The regularity conditions in Assumptions 4 and 5 together facilitate the dominated convergence theorem to make a structural sense of the QRKD estimand (1.1). With $\mathcal{B}(y, x)$ denoting the collection of Borel subsets of $\partial \mathcal{V}(y, x)$, we define the function $\mu_{y, x} : \mathcal{B}(y, x) \to \mathbb{R}$ by

$$\mu_{y, x}(S) := \frac{\int_{S} \frac{1}{\|\nabla h(x, \epsilon)\|} f_\epsilon x dH^{M-1}(\epsilon)}{\int_{\partial \mathcal{V}(y, x)} \frac{1}{\|\nabla h(x, \epsilon)\|} f_\epsilon x dH^{M-1}(\epsilon)}$$

for all $S \in \mathcal{B}(y, x)$.

The next theorem claims that this is a probability measure and gives weights with respect to which the QRKD estimand (1.1) measures the average structural causal effect of the treatment intensity $b$ on an outcome $y$ for those individuals at the $\tau$-th conditional quantile of $Y$ given $X = x_0$.

**Theorem 1.** Suppose that Assumptions 1, 2, 3, 4 and 5 hold. Then, $\mu_{y, x}$ is a probability measure on $\partial \mathcal{V}(y, x)$ for all $(x, y) \in \mathcal{X} \times \mathcal{Y}$, and we have

$$QRKD(\tau) = \int g_1(b(x_0), x_0, \epsilon) d\mu_{y, x_0}(\epsilon) = E_{\mu_{y, x_0}}[g_1(b(x_0), x_0, \epsilon)]$$

where $\tau = F_{Y|X}(y \mid x_0)$.

A proof is provided below at the end of the current section. We may derive a similar causal interpretation for the case of fuzzy QRKD – see Section A.1 in the supplementary appendix. As is often the
case in the treatment literature (e.g., Angrist and Imbens, 1995), this theorem shows a causal interpretation in terms of a weighted average. Specifically, (2.2) shows that the QRKD estimand (1.1) measures a weighted average of the heterogeneous causal effects $g_1(b(x_0), x_0, \epsilon)$ displayed on the right-hand side of (2.2). The weights given in the definition of $\mu_{y,x_0}$ are proportional to $f_{\epsilon|x}(\epsilon|x_0)/\|\nabla h(x_0, \epsilon)\|$ which is positive on the conditional support of $\epsilon$ given $X = x_0$. In other words, the QRKD estimand measures a strict convex combination of the ceteris paribus causal effects of $b$ on $y$ for those individuals at the $\tau$-th conditional quantile of $Y$ given $X = x_0$. One may worry about the obscurity of the causal interpretations under the ‘weighted’ averages. There are two special cases where the QRKD estimand allows for causal interpretations in terms of purely unweighted averages, i.e., $\|\nabla h(x_0, \epsilon)\|$ is constant in $\epsilon$. One example is the case where the structural function $g(b, x, \cdot)$ is monotone in a scalar unobservable $\epsilon$, which is the special case discussed in Section 2.1. The other example is the more general case where the structure exhibits partial additivity, e.g., $g(b, x, \epsilon) = \sum_{m=1}^{M} \epsilon_m g^m(b, x)$. When an empirical practitioner is reluctant to make either of these assumptions, the QRKD estimand can be still interpreted as a weighted average measurement of the treatment effects among the subpopulation of individuals at the $\tau$-th conditional quantile of $Y$ given $X = x_0$. In either of these cases, heterogeneity in values of the QRKD estimand across quantiles $\tau$ can be used as evidence for heterogeneity in treatment effects. Therefore, we can still conduct statistical inference for heterogeneous treatment effects based on the weak convergence results obtained in Section 3 and Section B.5 in the supplementary appendix.

**Proof of Theorem 1:** The first result that $\mu_{y,x}$ is a probability measure on $\partial V(y, x)$ follows from Lemma 2 of Sasaki (2015) under Assumption 3. Next, by Lemma 1 of Sasaki (2015) under Assumptions 2 and 3, the QPD $\frac{\partial}{\partial x} Q_{Y|X}(\tau | x)$ exists and

$$
\frac{\partial}{\partial x} Q_{Y|X}(\tau | x) = \frac{\int_{\partial V(y,x)} h_x(x, \epsilon) \frac{f_{\epsilon|x}(\epsilon|x) \cdot M \pi(M-1)^{1/2}}{2^{M-1} \Gamma(\frac{M+1}{2})} dH_{M-1}(\epsilon) - \int_{V(y,x)} \frac{\partial}{\partial x} f_{\epsilon|x}(\epsilon| x) dm^{M}(\epsilon)}{\int_{\partial V(y,x)} \frac{1}{\|\nabla_h h(x, \epsilon)\|} f_{\epsilon|x}(\epsilon| x) \cdot M \pi(M-1)^{1/2}} dH_{M-1}(\epsilon) \times \frac{2^{M-1} \Gamma(\frac{M+1}{2})}{dH_{M-1}(\epsilon)}}
$$

$$
= E_{\mu_{y,x}}[h_x(x, \epsilon)] - A(y, x),
$$
where $\Gamma$ is the Gamma function and $A$ is defined by

$$A(y, x) := \frac{\int_{V(y, x)} \frac{\partial}{\partial x} f_{\epsilon | X}(\epsilon | x) dm^M(\epsilon)}{\int_{\partial V(y, x)} \frac{1}{\|\nabla h(x, \epsilon)\|} \frac{\int_{|X|} f_{\epsilon | X}(\epsilon | x) M_\epsilon^{M-1} / 2}{\Gamma(M-1) / 2} dH^{M-1}(\epsilon)}$$

Note that $g_2 = \frac{\partial g}{\partial x}$ is continuous in $x$ by Assumption 2 (i). Also, $\mu_{y,x}(\epsilon)$ is continuous in $x$ for each fixed $y$ according to parts (i), (ii) and (iii) of Assumption 2. Furthermore, Assumption 2 (i), (ii), (iii) and (iv) imply that $A(y, x)$ is well-defined and is continuous in $x$ for all $y \in \mathcal{Y}$. Therefore, applying the dominated convergence theorem under Assumptions 4 and 5 yields

$$\lim_{x \to x_0^+} \frac{\partial}{\partial x} Q_{y \mid X}(\tau \mid x) = \lim_{x \to x_0^+} \int \left\{ h_x(x, \epsilon) \right\} d\mu_{y,x}(\epsilon) - \lim_{x \to x_0^+} A(y, x)$$

$$= \int \lim_{x \to x_0^+} \frac{\partial}{\partial x} \{g(b(x), x, \epsilon)\} d\mu_{y,x}(\epsilon) - A(y, x_0)$$

$$= \int \lim_{x \to x_0^+} \{g_1(b(x), x, \epsilon) b'(x) + g_2(b(x), x, \epsilon)\} d\mu_{y,x}(\epsilon) - A(y, x_0)$$

$$= \int \{g_1(b(x_0), x_0, \epsilon) b'(x_0^+) + g_2(b(x_0), x_0, \epsilon)\} d\mu_{y,x_0}(\epsilon) - A(y, x_0)$$

Similarly, taking the limit from the left, we have

$$\lim_{x \to x_0^-} \frac{\partial}{\partial x} Q_{y \mid X}(\tau \mid x) = \int \{g_1(b(x_0), x_0, \epsilon) b'(x_0^-) + g_2(b(x_0), x_0, \epsilon)\} d\mu_{y,x_0}(\epsilon) - A(y, x_0).$$

Taking the difference of the right and left limits eliminates $\int g_2(b(x_0), x_0, \epsilon) d\mu_{y,x_0}(\epsilon) - A(y, x_0)$, and thus produces

$$\lim_{x \to x_0^+} \frac{\partial}{\partial x} Q_{y \mid X}(\tau \mid x) - \lim_{x \to x_0^-} \frac{\partial}{\partial x} Q_{y \mid X}(\tau \mid x) = [b'(x_0^+) - b'(x_0^-)] E_{\mu_{y,x_0}} [g_1(b(x_0), x_0, \epsilon)] .$$

Finally, note that Assumption 1 has $b'(x_0^+) - b'(x_0^-) \neq 0$, and hence we can divide both sides of the above equality by $b'(x_0^+) - b'(x_0^-)$. This gives the desired result.

\[\square\]

### 3 Estimation and Inference

#### 3.1 The Estimator and Its Asymptotic Distribution

We propose to estimate the QRKD estimand (1.1) by the sample counterpart

$$\hat{Q}_{RKD}(\tau) = \frac{\hat{\beta}^+(\tau) - \hat{\beta}^-(\tau)}{\lim_{x \downarrow x_0} b'(x) - \lim_{x \uparrow x_0} b'(x)} \quad (3.1)$$

\[9\]
with the two terms in the numerator given by the one-sided local linear quantile smoothers

\[ \hat{\beta}(\tau) = \nu_2 \arg \min_{\alpha,\beta} \sum_{i=1}^{n} \left( d_i^+ K\left( \frac{x_i - x_0}{h_{n,\tau}} \right) \rho_{\tau}(y_i - \alpha - \beta(x_i - x_0)) \right) \]

\[ \hat{\beta}^-(\tau) = \nu_2 \arg \min_{\alpha,\beta} \sum_{i=1}^{n} \left( d_i^- K\left( \frac{x_i - x_0}{h_{n,\tau}} \right) \rho_{\tau}(y_i - \alpha - \beta(x_i - x_0)) \right) \]

for \( \tau \in T \), where \( T \subset (0,1) \) is a closed interval, \( K \) is a kernel function, \( \rho_{\tau}(u) = u(\tau - 1 \{ u < 0 \}) \), \( d_i^+ = 1\{x_i \geq x_0\} \), \( d_i^- = 1\{x_i \leq x_0\} \), and \( \nu_2 = [0,1]' \). A researcher observing a sample \( \{y_i, x_i\}_{i=1}^{n} \) of \( n \) observations can compute (3.1) explicitly to estimate (1.1). In the remainder of this section, weak convergence results are developed for the quantile processes of \( \hat{\beta}^+(\tau) \) and \( \hat{\beta}^-(\tau) \), which in turn yield weak convergence results for the quantile process of the QRKD estimator of treatment effects. Furthermore, using the weak convergence results, we propose some tests of hypotheses concerning heterogeneous treatment effects. With the kernel-dependent constant matrices \( N^+(\tau) = \int_{0}^{\infty} (1, u)'(1, u)K(u)du \) and \( N^-(\tau) = \int_{-\infty}^{0} (1, u)'(1, u)K(u)du \), we make the following assumptions.

**Assumption 6.** There exist \( \bar{x} > x_0 \) and \( \underline{x} < x_0 \) such that the following conditions are satisfied:

(i) (a) The density function \( f_X(\cdot) \) exists and is continuously differentiable in a neighborhood of \( x_0 \) and \( 0 < f_X(x_0) < \infty \). (b) \( \{(y_i, x_i)\}_{i=1}^{n} \) is an i.i.d. sample of \( n \) observations of the bivariate random vector \( (Y, X) \).

(ii) (a) \( f_{Y|X}(y|x) \) is Lipschitz continuous on \( \inf_{(\tau,x)\in T\times(x_0,\bar{x})} Q(\tau|x) \leq x \leq \sup_{(\tau,x)\in T\times(x_0,\bar{x})} Q(\tau|x) \) \times (x_0, \bar{x}) \) and \( \inf_{(\tau,x)\in T\times[\underline{x},x_0]} Q(\tau|x) \leq x \leq \sup_{(\tau,x)\in T\times[\underline{x},x_0]} Q(\tau|x) \) \times [\underline{x}, x_0) \). (b) There exist finite constants \( f_L > 0 \), \( f_U > 0 \) and \( \epsilon > 0 \), such that \( f_{Y|X}(Q(\tau|x) + \eta|x) \) lies between \( f_L \) and \( f_U \) for all \( \tau \in T \), \( |\eta| \leq \epsilon \) and \( x \in [\underline{x}, \bar{x}] \).

(iii) (a) \( \partial Q(\tau|x_0^+)\), \( \partial Q(\tau|x_0^-) \), and \( \partial^2 Q(\tau|x_0^-) \) exist and are Lipschitz continuous in \( \tau \) on \( T \). (b) \( \partial Q(\tau|x)/\partial x \) and \( \partial Q^2(\tau|x)/\partial x^2 \) exist and are Lipschitz continuous on \( \{(x, \tau)|x \in (x_0, \bar{x}], \tau \in T\} \) and \( \{(x, \tau)|x \in [\underline{x}, x_0), \tau \in T\} \).

(iv) The kernel \( K \) is compactly supported, Lipschitz, differentiable, and satisfying \( K(\cdot) \geq 0 \), \( \int K(u)du = 1 \), \( \int uK(u)du = 0 \). Also, \( \int_{0}^{\infty} u^k K(u)du \) and \( \int_{-\infty}^{0} u^k K(u)du \) are finite for \( k = 1, 2, 3 \). The matrices
and $N^-$ are positive definite.

(v) The bandwidths satisfy $h_{n,\tau} = c(\tau)h_n$, where $nh_n^3 \to \infty$ and $h_n = o(n^{-1/5})$ as $n \to \infty$, and $c(\cdot)$ is Lipschitz continuous satisfying $0 < c \leq c(\tau) \leq \tau < \infty$ for all $\tau \in T$.

Part (i) (a) requires smoothness of the density of the running variable. This can be interpreted as the design requirement for absence of endogenous sorting across the kink point $x_0$. The i.i.d assumption in part (i) (b) is usually considered to be satisfied for micro data of random samples. Part (ii) concerns about regularities of the conditional density function of $Y$ given $X$. It requires sufficient smoothness, but does not rule out quantile regression kinks at $x_0$, which is the main crucial assumption for our identification argument. Part (iii) concerns about regularities of the conditional quantile function of $Y$ given $X$. Like part (ii), it does not rule out quantile regression kinks at $x_0$. Part (iv) prescribes requirements for kernel functions to be chosen by users. In Section 4 for simulation studies, we propose an example of such a choice to satisfy this requirement. Finally, part (v) specifies the rate at which the bandwidth parameters diminish as the sample size becomes large. It obeys the standard rate for a first-order derivative estimation, but we also require its uniformity over quantiles $\tau$ in $T$. While $h_n = o(n^{-1/5})$ is required for a valid inference without bias reduction, this requirement can be relaxed to $h_n = O(n^{-1/5})$ if one is willing to make additional assumptions and to take an additional step of bias reduction – see Section B.5 in the supplementary appendix. Under this set of assumptions, we obtain uniform Bahadur representations for the component estimators, $\hat{\beta}^+(\tau)$ and $\hat{\beta}^-(\tau)$, of our interest – see Lemma 3 in Section B.2 in the supplementary appendix.

Applying Lemma 3, we now establish weak convergence results for our component estimators, $\hat{\beta}^+$ and $\hat{\beta}^-$. We focus on $\hat{\beta}^+$, but a similar result follows for $\hat{\beta}^-$. 

**Theorem 2.** Under Assumptions 6, we have the weak convergence

$$
\sqrt{nh_{n,\tau}^3 f_X(x_0) f_Y(Q(\tau|x_0^+)|x_0^+)} \times
($$
\hat{\beta}^+(\tau) - \frac{\partial Q(\tau|x_0^+)}{\partial x} - h_{n,\tau} \frac{\partial^2 Q(\tau|x_0^+)}{2\partial x^2}(1,u)'K(u)du \Rightarrow G^+(\tau),
$$)$$

11
for the zero mean Gaussian process $G^+(\tau)$ defined over $T$ with covariance function

$$E(G^+(r)G^+(s)) = (\kappa(r)\kappa(s))^{-1/2}(r \wedge s - rs)\ell_2^2(N^+)^{-1}T^+(r, s)(N^+)^{-1}_{r, s},$$

for each $r, s \in T$, where $T^+(r, s) = \int_0^\infty \left[ 1 - \frac{u}{\kappa(\tau)} \right]' K\left( \frac{u}{\kappa(\tau)} \right) \left[ 1 - \frac{u}{\kappa(s)} \right] K\left( \frac{u}{\kappa(s)} \right) du$ with $\kappa(\tau) = h_{n, \tau}/h_{n, 1/2} = \frac{c(\tau)}{c(1/2)} \geq (\zeta/\tau) > 0$. A similar result follows for $\hat{\beta}^-(\tau)$.

A proof is provided in Section B.3 in the supplementary appendix. While we write the above weak convergence result explicitly accounting for the finite-sample bias term $h_{n, \tau} \frac{\ell_2^2(N^+)^{-1}}{2} \int_0^\infty u^2 \frac{\partial^2 Q(\tau|x_0^+)}{\partial x^2}(1, u)\ K(u) du$, it goes away in large sample as $h_{n, \tau}$ goes to zero uniformly in $\tau \in T$. In other words, this bias term can be considered to be absent in the above equation. With this said, in case one wish to reduce this finite-sample bias at the cost of additional assumptions, we propose in Section B.5 in the supplementary appendix how to estimate this bias term.

We are now ready to present a weak convergence result for our QRKD estimator (3.1). By Section 2.1 of Giné and Nickl (2015), the sum of two independent Gaussian processes is a Gaussian process with the mean (respectively, the covariance) being the sum of the means (respectively, the covariances).

**Corollary 1.** Under Assumptions 1 and 6 we have the weak convergence

$$\sqrt{nh_{n, \tau}^2} \left( Q_{\text{RKD}}(\tau) - \text{QRKD}(\tau) \right) \Rightarrow Y(\tau) = \frac{1}{\sqrt{f_X(x_0)}(b'(x_0^+) - b'(x_0^-))} \left[ \frac{G^+(\tau)}{f_Y|X(Q(\tau|x_0^+)|x_0^+)} - \frac{G^-(\tau)}{f_Y|X(Q(\tau|x_0^-)|x_0^-)} \right].$$

The random process $Y(\cdot)$ has mean zero, as $G^+(\tau)$ and $G^-(\tau)$ do. For any pair $r, s \in T$ of quantiles, the covariance can be computed by:

$$E[Y(r)Y(s)] = \frac{1}{f_X(x_0)(b'(x_0^+) - b'(x_0^-))^2} \times \left[ \frac{EG^+(r)G^+(s)}{f_Y|X(Q(r|x_0^+)|x_0^+)f_Y|X(Q(s|x_0^+)|x_0^+)} + \frac{EG^-(r)G^-(s)}{f_Y|X(Q(r|x_0^-)|x_0^-)f_Y|X(Q(s|x_0^-)|x_0^-)} \right].$$

This uniform convergence result is applicable to many purposes, such as to compute uniform confidence bands for the QRKD. Of particular interest may be the empirical tests discussed in Section 3.2.
3.2 Testing for Heterogeneous Treatment Effects

Researchers are often interested in the following hypotheses regarding heterogeneous treatment effects.

Treatment Significance  \( H_0^S : \quad QRKD(\tau) = 0 \) for all \( \tau \in T \).

Treatment Heterogeneity  \( H_0^H : \quad QRKD(\tau) = QRKD(\tau') \) for all \( \tau, \tau' \in T \).

They are both considered in Koenker and Xiao (2002), Chernozhukov and Fernández-Val (2005) and Qu and Yoon (2015b), among others. Following the approach of these preceding papers, the two hypotheses, \( H_0^S \) and \( H_0^H \), may be tested using the statistics

\[
WS_n(T) = \sqrt{nh^3_{n,\tau}} \sup_{\tau \in T} |\hat{QRKD}(\tau)| \quad \text{and} \quad
WH_n(T) = \sqrt{nh^3_{n,\tau}} \sup_{\tau \in T} \left| \hat{QRKD}(\tau) - \int_T \hat{QRKD}(\tau') d\tau' \right|,
\]

respectively. While, for the latter statistic, we may also use a mean RKD estimator in place of \( \int_T \hat{QRKD}(\tau') d\tau' \), we use the above definition for its simplicity as a functional only of \( \hat{QRKD}(\cdot) \).

Consequence of Corollary 1 are the following asymptotic distributions of these test statistics, a proof of which is provided in Section B.4 in the supplementary appendix.

**Corollary 2.** Under Assumptions 1 and 6, we have

(i) \( WS_n(T) \Rightarrow \sup_{\tau \in T} |Y(\tau)| \) under the null hypothesis \( H_0^S \); and

(ii) \( WH_n(T) \Rightarrow \sup_{\tau \in T} |\phi'_{QRKD}(Y)(\tau)| \) under the null hypothesis \( H_0^H \), where \( \phi'_{QRKD}(g)(\tau) = g(\tau) - \int_T g(\tau') d\tau' \) for all \( g \in L^\infty_m(T) \), the space of all bounded, measurable, real-valued functions defined on \( T \).

3.3 Remarks about Bias Reduction

In concluding the current section on estimation and inference, we remark on the issue of bias reduction that is relevant in practice. Under our assumption of under-smoothing, the proposed asymptotic theory is indeed correct. In finite sample, however, bias reduction would help if one is willing to make
additional smoothness assumptions. While this issue itself is an important research topic, it is out of the scope of this paper. In the context of quantile RDD, Qu and Yoon (2015b) propose a procedure of bias reduction motivated by Calonico, Cattaneo and Titiunik (2014). We directly use their approach in our context. While it is not our original product, we describe the procedure in Section B.5 in the supplementary appendix for convenience of the readers.

4 Simulation Studies

Consider the following policy function with a kink at \( x_0 = 0 \).

\[
b(x) = \begin{cases} 
2x & \text{if } x \leq 0 \\
0.5x & \text{if } x > 0 
\end{cases}
\]

For convenience of assessing the performance of our estimator for homogeneous treatment effects and heterogeneous treatment effects, we consider the following three outcome structures.

Structure 0: \( g(b, x, \epsilon) = 0.0b + 0.5x + 0.05x^2 + \epsilon \)

Structure 1: \( g(b, x, \epsilon) = 0.5b + 0.5x + 0.05x^2 + \epsilon \)

Structure 2: \( g(b, x, \epsilon) = 0.5[0.5 + F_\epsilon(\epsilon)]b + 0.5x + 0.05x^2 + \epsilon \)

where \( F_\epsilon \) denotes the CDF of \( \epsilon \). Note that Structures 0 and 1 entail homogeneous treatment effects, while Structure 2 entails heterogeneous treatment effects across quantiles \( \tau \) as follows.

Structure 0: \( g_1(b, x, Q_\epsilon(\tau)) = 0.0 \)

Structure 1: \( g_1(b, x, Q_\epsilon(\tau)) = 0.5 \)

Structure 2: \( g_1(b, x, Q_\epsilon(\tau)) = 0.5[0.5 + \tau] \)

To allow for endogeneity, we generate the primitive data according to

\[
\begin{pmatrix} X_i \\ \epsilon_i \end{pmatrix} \sim_{i.i.d.} N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho \sigma_X \sigma_\epsilon \\ \rho \sigma_X \sigma_\epsilon & \sigma_\epsilon^2 \end{pmatrix} \right),
\]

where \( \sigma_X \) and \( \sigma_\epsilon \) are the standard deviations of \( X \) and \( \epsilon \), and \( \rho \) is the correlation coefficient between \( X \) and \( \epsilon \).
where $\sigma_X = \sigma = \rho = 0.5$. For estimation, we use the tricube kernel defined by

$$K(u) = \frac{70}{81} (1 - |u|)^3 1\{|u| < 1\}.$$  

The bandwidths are selected with our choice rule based on the MSE minimization – see Sections B.5 and C in the supplementary appendix for details.

Figures 1 and 2 show simulated distributions of the QRKD estimates under Structure 1 and Structure 2, respectively. In each of the two figures, the left column lists basic results, and the right column lists results with bias reduction (based on Section B.5 in the supplementary appendix). The top row, the middle row and the bottom row list results for the sample sizes $N = 1,000, 2,000$ and $4,000$, respectively. In each graph, the horizontal axis measures quantiles $\tau$, while the vertical axis measures the QRKD. The true QRKD are indicated by solid gray lines. The other broken curves indicate the 5-th, 10-th, 50-th, 90-th, and 95-th percentiles of the simulated distributions of our estimates based on 5,000 iterations. We observe that the displayed distribution shrinks for each structure at each quantile $\tau$ as the sample size $N$ increases. The results in the right column with bias reduction indeed improve the biases, although this procedure requires additional assumptions – see Section B.5 in the supplementary appendix. These aspects of the results confirm that our methods work as in theory.

In order to more quantitatively analyze the finite sample pattern, we summarize some basic statistics for the simulated distributions in Tables 1 and 2 for Structure 1 and Structure 2, respectively. In each table, the four column groups list the absolute biases (Bias), the standard deviations (SD), root mean squared errors (RMSE), and the rejection frequencies for point-wise 5%-level t-tests for the null hypotheses of the true QRKD values (5% Size). The upper half of each table displays statistics for the basic results, and the lower half of each table displays statistics for the results with bias reduction. For each structure at each quantile $\tau$, we again observe that SD and RMSE decrease as the sample size $N$ increases. These patterns are consistent with our previous discussions on Figures 1 and 2. Observe that the 5% sizes are reasonably accurate at each quantile for each structure with bias reduction, while they can increase without bias reduction as the sample size increases. We remark that the latter
fact does not contradict with our theory, because the practical choice of optimal bandwidth based on an approximate MSE fails to under-smooth the estimates while our theory of asymptotic distribution without bias reduction requires an under-smoothing.

While these sizes concern about point-wise inference, we also provide uniform inference results. Figure 3 shows rejection probabilities for the 95% level uniform test of significance (panel A) and the 95% level uniform test of heterogeneity (panel B) based on 1,000 iterations. Panel A shows that the rejection probability for the test of the null hypothesis of insignificance does not increase in the sample size for Structure 0, while it increases in the sample size for each of Structure 1 and Structure 2. Panel B shows that the rejection probability for the test of the null hypothesis of homogeneity does not increase in the sample size for Structure 0 and Structure 1, while it increases in the sample size for Structure 2. The above observations are consistent with the construction of Structure 0, Structure 1, and Structure 2.

5 An Empirical Illustration

In labor economics, causal effects of the unemployment insurance (UI) benefits on the duration of unemployment are of interest from policy perspectives. The elasticity of labor supply with respect to changes in unemployment insurance is an intertwining result of two endogenous forces – the liquidity effects and the moral hazard effects. Landais (2015) demonstrates a reinterpretation of these forces in terms of the traditional framework of dynamic labor supply, and shows how the moral hazard effects of UI on search efforts can be explained by the Frisch elasticity concept, i.e., responses of search efforts to changes in benefits keeping the marginal utility of wealth constant. He then proposes an empirical strategy using the RKD to identify the moral hazard effects of UI. Using the data set of the Continuous Wage and Benefit History Project (CWBH – see Moffitt, 1985), Landais estimates the effects of benefit amounts on the duration of unemployment. In this section, we apply our QRKD methods, and aim to discover potential heterogeneity in these causal effects. Using quantiles in this application also has an
Figure 1: Simulated distributions of QRKD estimates under Structure 1.

<table>
<thead>
<tr>
<th>Basic Result; $N = 1,000$</th>
<th>With BR; $N = 1,000$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Basic Result; $N = 2,000$</th>
<th>With BR; $N = 2,000$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Basic Result; $N = 4,000$</th>
<th>With BR; $N = 4,000$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Figure 2: Simulated distributions of QRKD estimates under Structure 2.

Basic Result; $N = 1,000$

Basic Result; $N = 2,000$

Basic Result; $N = 4,000$

With BR; $N = 1,000$

With BR; $N = 2,000$

With BR; $N = 4,000$
<table>
<thead>
<tr>
<th>Basic</th>
<th>Bias</th>
<th>SD</th>
<th>RMSE</th>
<th>5% Size</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>N =</td>
<td>1000</td>
<td>2000</td>
<td>4000</td>
<td></td>
</tr>
<tr>
<td>τ = 0.10</td>
<td>0.05</td>
<td>0.06</td>
<td>0.07</td>
<td>0.33 0.26 0.22</td>
</tr>
<tr>
<td>τ = 0.20</td>
<td>0.04</td>
<td>0.05</td>
<td>0.05</td>
<td>0.24 0.19 0.15</td>
</tr>
<tr>
<td>τ = 0.30</td>
<td>0.04</td>
<td>0.05</td>
<td>0.05</td>
<td>0.21 0.16 0.13</td>
</tr>
<tr>
<td>τ = 0.40</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.19 0.15 0.12</td>
</tr>
<tr>
<td>τ = 0.50</td>
<td>0.05</td>
<td>0.05</td>
<td>0.04</td>
<td>0.19 0.14 0.11</td>
</tr>
<tr>
<td>τ = 0.60</td>
<td>0.05</td>
<td>0.05</td>
<td>0.04</td>
<td>0.19 0.14 0.11</td>
</tr>
<tr>
<td>τ = 0.70</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.19 0.15 0.11</td>
</tr>
<tr>
<td>τ = 0.80</td>
<td>0.06</td>
<td>0.06</td>
<td>0.05</td>
<td>0.21 0.16 0.13</td>
</tr>
<tr>
<td>τ = 0.90</td>
<td>0.07</td>
<td>0.07</td>
<td>0.06</td>
<td>0.25 0.19 0.16</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>With BR</th>
<th>Bias</th>
<th>SD</th>
<th>RMSE</th>
<th>5% Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>N =</td>
<td>1000</td>
<td>2000</td>
<td>4000</td>
<td></td>
</tr>
<tr>
<td>τ = 0.10</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.36 0.28 0.23</td>
</tr>
<tr>
<td>τ = 0.20</td>
<td>0.00</td>
<td>0.01</td>
<td>0.01</td>
<td>0.28 0.21 0.16</td>
</tr>
<tr>
<td>τ = 0.30</td>
<td>0.01</td>
<td>0.01</td>
<td>0.01</td>
<td>0.25 0.19 0.14</td>
</tr>
<tr>
<td>τ = 0.40</td>
<td>0.01</td>
<td>0.01</td>
<td>0.02</td>
<td>0.24 0.18 0.13</td>
</tr>
<tr>
<td>τ = 0.50</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.24 0.17 0.13</td>
</tr>
<tr>
<td>τ = 0.60</td>
<td>0.02</td>
<td>0.02</td>
<td>0.02</td>
<td>0.24 0.17 0.12</td>
</tr>
<tr>
<td>τ = 0.70</td>
<td>0.03</td>
<td>0.03</td>
<td>0.02</td>
<td>0.24 0.18 0.13</td>
</tr>
<tr>
<td>τ = 0.80</td>
<td>0.04</td>
<td>0.04</td>
<td>0.03</td>
<td>0.26 0.19 0.14</td>
</tr>
<tr>
<td>τ = 0.90</td>
<td>0.07</td>
<td>0.05</td>
<td>0.05</td>
<td>0.30 0.22 0.18</td>
</tr>
</tbody>
</table>

Table 1: Simulated finite-sample statistics of the QRKD estimates under Structure 1.
<table>
<thead>
<tr>
<th>Basic</th>
<th>Bias</th>
<th>SD</th>
<th>RMSE</th>
<th>5% Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>N =</td>
<td>1000</td>
<td>2000</td>
<td>4000</td>
<td>1000</td>
</tr>
<tr>
<td>( \tau = 0.10 )</td>
<td>0.07</td>
<td>0.08</td>
<td>0.08</td>
<td>0.29</td>
</tr>
<tr>
<td>( \tau = 0.20 )</td>
<td>0.11</td>
<td>0.12</td>
<td>0.12</td>
<td>0.22</td>
</tr>
<tr>
<td>( \tau = 0.30 )</td>
<td>0.14</td>
<td>0.14</td>
<td>0.14</td>
<td>0.20</td>
</tr>
<tr>
<td>( \tau = 0.40 )</td>
<td>0.16</td>
<td>0.16</td>
<td>0.14</td>
<td>0.20</td>
</tr>
<tr>
<td>( \tau = 0.50 )</td>
<td>0.17</td>
<td>0.17</td>
<td>0.15</td>
<td>0.19</td>
</tr>
<tr>
<td>( \tau = 0.60 )</td>
<td>0.18</td>
<td>0.17</td>
<td>0.15</td>
<td>0.19</td>
</tr>
<tr>
<td>( \tau = 0.70 )</td>
<td>0.18</td>
<td>0.17</td>
<td>0.16</td>
<td>0.19</td>
</tr>
<tr>
<td>( \tau = 0.80 )</td>
<td>0.18</td>
<td>0.17</td>
<td>0.15</td>
<td>0.20</td>
</tr>
<tr>
<td>( \tau = 0.90 )</td>
<td>0.17</td>
<td>0.16</td>
<td>0.15</td>
<td>0.21</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>With BR</th>
<th>Bias</th>
<th>SD</th>
<th>RMSE</th>
<th>5% Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>N =</td>
<td>1000</td>
<td>2000</td>
<td>4000</td>
<td>1000</td>
</tr>
<tr>
<td>( \tau = 0.10 )</td>
<td>0.03</td>
<td>0.01</td>
<td>0.01</td>
<td>0.33</td>
</tr>
<tr>
<td>( \tau = 0.20 )</td>
<td>0.01</td>
<td>0.02</td>
<td>0.04</td>
<td>0.27</td>
</tr>
<tr>
<td>( \tau = 0.30 )</td>
<td>0.03</td>
<td>0.04</td>
<td>0.05</td>
<td>0.25</td>
</tr>
<tr>
<td>( \tau = 0.40 )</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.24</td>
</tr>
<tr>
<td>( \tau = 0.50 )</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.23</td>
</tr>
<tr>
<td>( \tau = 0.60 )</td>
<td>0.06</td>
<td>0.05</td>
<td>0.05</td>
<td>0.24</td>
</tr>
<tr>
<td>( \tau = 0.70 )</td>
<td>0.05</td>
<td>0.05</td>
<td>0.05</td>
<td>0.23</td>
</tr>
<tr>
<td>( \tau = 0.80 )</td>
<td>0.03</td>
<td>0.04</td>
<td>0.04</td>
<td>0.24</td>
</tr>
<tr>
<td>( \tau = 0.90 )</td>
<td>0.01</td>
<td>0.02</td>
<td>0.02</td>
<td>0.26</td>
</tr>
</tbody>
</table>

Table 2: Simulated finite-sample statistics of the QRKD estimates under Structure 2.
(A) Rejection Probabilities for the 95% Level Test of Significance

Basic Result

With Bias Reduction

(B) Rejection Probabilities for the 95% Level Test of Heterogeneity

Basic Result

With Bias Reduction

Figure 3: Rejection probabilities for the 95% level uniform test of significance (panel A) and the 95% level uniform test of heterogeneity (panel B) based on 1,000 replications.
advantage of informing a likely direction of the selection bias of the mean RKD estimator that stems from not observing the mass of employed individuals at the low quantile \((y = 0)\).

In all of the states in the United States, a compensated unemployed individual receives a weekly benefit amount \(b\) that is determined as a fraction \(\tau_1\) of his or her highest earning quarter \(x\) in the base period (the last four completed calendar quarters immediately preceding the start of the claim) up to a fixed maximum amount \(b_{\text{max}}\), i.e. \(b = \min\{\tau_1 \cdot x, \ b_{\text{max}}\}\). The both parameters, \(\tau_1\) and \(b_{\text{max}}\), of the policy rule vary from state to state. Furthermore, the ceiling level \(b_{\text{max}}\) changes over time within a state. For these reasons, empirical analysis needs to be conducted for each state for each restricted time period. The potential duration of benefits is determined in a somewhat more complicated manner. Yet, it also can be written as a piecewise linear and kinked function of a fraction of a running variable \(x\) in the CWBH data set.

Following Landais (2015), we make our QRKD empirical illustration by using the CWBH data for Louisiana. The data cleaning procedure is conducted in the same manner as in Landais. As a result of the data processing, we obtain the same descriptive statistics (up to deflation) as those in Landais for those variables that we use in our analysis. For the dependent variable \(y\), we consider both the claimed number of weeks of UI and the actually paid number of weeks. For the running variable \(x\), we use the highest quarter wage in the based period. The treatment intensity \(b\) is computed by using the formula \(b(x) = \min\{(1/25) \cdot x, \ b_{\text{max}}\}\), with a kink where the maximum amount is \(b_{\text{max}} = $4,575\) for the period between September 1981 and September 1982 and \(b_{\text{max}} = $5,125\) for the period between September 1982 and December 1983.

Table 3 summarizes empirical results for the time period between September 1981 and September 1982. Table 4 summarizes empirical results for the time period between September 1982 and December 1983. In each table, we display the RKD results by Landais (2015) for a reference. In the following rows, the QRKD estimates are reported with respective standard errors in parentheses for quantiles \(\tau \in \{0.10, \cdots, 0.90\}\). At the bottom of each table, we report the p-values for the test of significance.
<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>UI Claimed</th>
<th>UI Paid</th>
</tr>
</thead>
<tbody>
<tr>
<td>RKD (Landais, 2015)</td>
<td>0.038 (0.009)</td>
<td>0.040 (0.009)</td>
</tr>
<tr>
<td>QRKD</td>
<td>$\tau = 0.10$</td>
<td>0.019 (0.030)</td>
</tr>
<tr>
<td></td>
<td>$\tau = 0.20$</td>
<td>0.036 (0.037)</td>
</tr>
<tr>
<td></td>
<td>$\tau = 0.30$</td>
<td>0.054 (0.041)</td>
</tr>
<tr>
<td></td>
<td>$\tau = 0.40$</td>
<td>0.067 (0.044)</td>
</tr>
<tr>
<td></td>
<td>$\tau = 0.50$</td>
<td>0.081 (0.044)</td>
</tr>
<tr>
<td></td>
<td>$\tau = 0.60$</td>
<td>0.109 (0.043)</td>
</tr>
<tr>
<td></td>
<td>$\tau = 0.70$</td>
<td>0.115 (0.041)</td>
</tr>
<tr>
<td></td>
<td>$\tau = 0.80$</td>
<td>0.161 (0.037)</td>
</tr>
<tr>
<td></td>
<td>$\tau = 0.90$</td>
<td>0.167 (0.030)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test of Significance</th>
<th>p-Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test of Significance</td>
<td>0.000</td>
</tr>
<tr>
<td>Test of Heterogeneity</td>
<td>0.004</td>
</tr>
</tbody>
</table>

Table 3: Empirical estimates and inference for the causal effects of UI benefits on unemployment durations based on the RKD and QRKD. The period of data is from September 1981 to September 1982. The numbers in parentheses indicate standard errors.

We can observe the following patterns in these result tables. First, the estimated causal effects have positive signs throughout all the quantiles, implying that higher benefit amounts cause longer unemployment durations consistently across all the outcome levels. Second, these causal effects are smaller and insignificant at lower quantiles, while they are larger and significantly different from zero at middle and higher quantiles. This pattern implies that unemployed individuals who have longer unemployment durations tend to have larger unemployment elasticities with respect to benefit levels.

This result is consistent with a simple intuitive story; individuals with longer unemployment durations
<table>
<thead>
<tr>
<th>Dependent Variable</th>
<th>UI Claimed</th>
<th>UI Paid</th>
</tr>
</thead>
<tbody>
<tr>
<td>RKD (Landais, 2015)</td>
<td>0.046 (0.006)</td>
<td>0.042 (0.006)</td>
</tr>
<tr>
<td>QRKD ( \tau = 0.10 )</td>
<td>0.023 (0.028)</td>
<td>0.024 (0.028)</td>
</tr>
<tr>
<td>( \tau = 0.20 )</td>
<td>0.049 (0.034)</td>
<td>0.053 (0.034)</td>
</tr>
<tr>
<td>( \tau = 0.30 )</td>
<td>0.067 (0.038)</td>
<td>0.065 (0.038)</td>
</tr>
<tr>
<td>( \tau = 0.40 )</td>
<td>0.086 (0.040)</td>
<td>0.080 (0.040)</td>
</tr>
<tr>
<td>( \tau = 0.50 )</td>
<td>0.108 (0.041)</td>
<td>0.107 (0.041)</td>
</tr>
<tr>
<td>( \tau = 0.60 )</td>
<td>0.092 (0.040)</td>
<td>0.097 (0.040)</td>
</tr>
<tr>
<td>( \tau = 0.70 )</td>
<td>0.111 (0.038)</td>
<td>0.110 (0.038)</td>
</tr>
<tr>
<td>( \tau = 0.80 )</td>
<td>0.074 (0.034)</td>
<td>0.082 (0.034)</td>
</tr>
<tr>
<td>( \tau = 0.90 )</td>
<td>0.073 (0.027)</td>
<td>0.070 (0.027)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Test of Significance</th>
<th>p-Value</th>
<th>0.026</th>
<th>0.021</th>
</tr>
</thead>
<tbody>
<tr>
<td>Test of Heterogeneity</td>
<td>p-Value</td>
<td>0.265</td>
<td>0.276</td>
</tr>
</tbody>
</table>

Table 4: Empirical estimates and inference for the causal effects of UI benefits on unemployment durations based on the RKD and QRKD. The period of data is from September 1982 to December 1983. The numbers in parentheses indicate standard errors.
are associated with lower abilities and are therefore more likely to show moral hazard. The extent of this increase of the causal effects in quantiles is more prominent for the results in Table 3 (1981–1982) than in Table 4 (1982–1983). Under the assumption of rank invariance, this result unambiguously implies that the moral hazard effects are heterogeneous. Without the rank invariance, one may want to argue that the heterogeneous quantile treatment effects can be attributed to just heterogeneous weights even without nonseparable heterogeneity. However, in the absence of nonseparability, the weights would be also constant. Hence, our results show that there is nonseparable heterogeneity in the causal structure even without the rank invariance. Third, the causal effects are very similar between the results for claimed UI as the outcome and the results for paid UI as the outcome variable. The respective standard errors are almost the same between these two outcome variables, but they are not exactly the same. Fourth, the uniform tests show that the causal effects are significantly different from zero for the both time periods. Lastly, the uniform tests show that the causal effects are significantly heterogeneous in Table 3 (1981–1982), while the heterogeneity is insignificant in Table 4 (1982–1983). The asymmetry in the last results may stem from different macro-economic environments between the two time periods: the U.S. economy experienced a recession in most of the period 1981–1982 while it was in recovery during the period 1982–1983.

6 Summary

Economists have taken advantage of policy irregularities to assess causal effects of endogenous treatment intensities. A new approach along this line is the regression kink design (RKD) used by recent empirical papers, including Nielsen, Sørensen and Taber (2010), Landais (2015), Simonsen, Skipper and Skipper (2015), Card, Lee, Pei and Weber (2016), and Dong (2016). While the prototypical framework is only able to assess the average treatment effect at the kink point, inference for heterogeneous treatment effects using the RKD is of potential interest by empirical researchers (e.g., Landais (2011) considers it). In this light, this paper develops causal analysis and methods for the quantile regression
kink design (QRKD).

We first develop causal interpretations of the QRKD estimand. It is shown that the QRKD estimand measures the marginal effect of the treatment variable on the outcome variable at the conditional quantile of the outcome given the design point of the running variable provided that the causal structure exhibits rank invariance. This result is generalized to the case of no rank invariance, where the QRKD estimand is shown to measure a weighted average of the marginal effects of the treatment variable on the outcome variable at the conditional quantile of the outcome given the design point of the running variable. Second, we propose a sample counterpart QRKD estimator, and develop its asymptotic properties for statistical inference of heterogeneous treatment effects. Using uniform Bahadur representations, we obtain a weak consistency result for the QRKD estimator. Applying the weak consistency result, we propose procedures for statistical tests of treatment significance and treatment heterogeneity. Simulation studies support our theoretical results. Applying our methods to the Continuous Wage and Benefit History Project (CWBH) data, we find significantly heterogeneous causal effects of unemployment insurance benefits on unemployment durations in the state of Louisiana for the period between September 1981 and September 1982. Finally, while the main text mostly focuses on the sharp QRKD that is relevant to our empirical illustration, we remark that identification and estimation results for the fuzzy QRKD are also available in Section A in the supplementary appendix for completeness.

References


Supplementary Appendix to
“Causal Inference by Quantile Regression Kink Designs”

Harold D. Chiang       Yuya Sasaki
Johns Hopkins University

August 30, 2016

This supplementary appendix provides extensions to and details of the contents presented in the main text. The appendix consists of three sections. First, we extend our identification and estimation results for sharp quantile regression kink designs (QRKD) to fuzzy QRKD in Section A. Second, Section B contains mathematical appendix including auxiliary lemmas, their proofs, and proofs for the results presented in the main text. Third, we present practical procedures in Section C.

A  Extension to Fuzzy QRKD

A.1  Causal Interpretation of the Fuzzy QRKD

While Section 2.2 in the main text focuses on the case of sharp QRKD, our identification result (Theorem 1) can be extended to the case of fuzzy QRKD in an analogous manner to the corresponding extension in Card, Lee, Pei and Weber (2016). In the current appendix section, we show the causal interpretation for the fuzzy QRKD.

The fuzzy QRKD estimand reads as

\[
\begin{align*}
\lim_{x \downarrow x_0} \frac{\partial}{\partial x} Q_{Y \mid X}(\tau \mid x) - \lim_{x \uparrow x_0} \frac{\partial}{\partial x} Q_{Y \mid X}(\tau \mid x) \\
\lim_{x \downarrow x_0} \frac{\partial}{\partial x} E[B \mid X = x] - \lim_{x \uparrow x_0} \frac{\partial}{\partial x} E[B \mid X = x]
\end{align*}
\]
where $B$ denotes the random variable for the treatment intensity. Unlike the sharp case, it is not deterministically controlled by the running variable. With the $M$-dimensional unobservables $\epsilon = (\epsilon_1, \epsilon_2)$ decomposed into two parts, we specify the relevant causal structure by

$$
y = g(b, x, \epsilon_2)$$

$$
b = b(x, \epsilon_1)
$$

For short-hand notations, we write $b_1(x, \epsilon_1) = \frac{\partial}{\partial x} b(x, \epsilon_1)$ and $h(x, \epsilon) = g(b(x, \epsilon_1), x, \epsilon_2)$. With these notations under the above setup, we make the following assumption.

**Assumption 7.**

(a) $b(\cdot, \epsilon_1)$ is continuous on $\mathcal{X}$ and continuously differentiable on $\mathcal{X} \setminus \{x_0\}$ for all $\epsilon_1$.

(b) There exist an absolutely integrable function that envelops $b_1(x, \cdot)$ for all $x$. (c) $E[(b_1(x_0^+, \epsilon_1) - b_1(x_0^-, \epsilon_1))|X = x_0]$ and $\int (b_1(x_0^+, \epsilon_1) - b_1(x_0^-, \epsilon_1))d\mu_{y,x_0}(\epsilon)$ exist and are finite and nonzero, where $\mu_{y,x_0}$ is defined as in Section 2.2.

Under this assumption, together with the basic assumptions from Section 2.2, we obtain the following causal interpretation of the fuzzy QRKD estimand by similar lines of proof to those of Theorem 1.

**Theorem 3.** Suppose that Assumptions 2 (with the modified definitions of $g$, $h$ and $\epsilon$ in the current appendix section), 3, 4, 5 and 7 are satisfied. For each $y \in \mathcal{Y}$, we have

$$
\lim_{x \downarrow x_0} \frac{\partial}{\partial x} Q_{Y|X}(\tau \mid x) - \lim_{x \uparrow x_0} \frac{\partial}{\partial x} Q_{Y|X}(\tau \mid x) = \chi \int g_1(b(x_0, \epsilon_1), x_0, \epsilon_2)d\psi_{y,x_0}(\epsilon_1, \epsilon_2) = \chi E_{\psi_{y,x_0}}[g_1(b(x_0, \epsilon_1), x_0, \epsilon_2)]
$$

where $\tau = F_{Y|X}(y \mid x_0)$, $\chi = \int_{b_1(x_0^+, \epsilon_1) - b_1(x_0^-, \epsilon_1)}d\mu_{y,x_0}(\epsilon)$, and $\psi_{y,x_0}(S) = \int_{b_1(x_0^+, \epsilon_1) - b_1(x_0^-, \epsilon_1)}d\mu_{y,x_0}(\epsilon)$ for all $S \in \mathcal{B}(y, x)$.

This result shows that the fuzzy QRKD estimand equals a constant $\chi$ times a weighted average of the marginal effects. The presence of the constant term $\chi$ somewhat obscures the causal interpretation.
of the estimand. In case where the kink direction is uniform, i.e., \( b_1(x_0^+, \epsilon_1) - b_1(x_0^-, \epsilon_1) \) has the same sign for all \( \epsilon_1 \), the sign of \( \chi \) is positive, and so the sign of the fuzzy QRKD can be considered to measure the sign of the weighted average marginal effects. A working paper version of Sasaki (2015) implies that \( \mu_{y,x_0} = F_\epsilon \) holds if \( g \) takes the following semi-separable form

\[
g(b(x, \epsilon_1), x, \epsilon_2) = g_I(b(x, \epsilon_1), x) + g_{II}(x, \epsilon_2).
\]

In this case, we have \( \chi = 1 \) and therefore the fuzzy QRKD can be interpreted as a weighted average of marginal effects without worrying about the multiplying constant.

### A.2 Estimation and Asymptotics for the Fuzzy QRKD

The main text focuses on the sharp QRKD. In this appendix, we provide an estimator for the fuzzy QRKD estimand developed in Appendix A.1 and its asymptotic properties. The conditional mean of the policy with errors, \( b(X, \epsilon_1) \), is written as \( m(x) = E[b(X, \epsilon_1)|X = x] \). The regression is also represented by the canonical decomposition \( B = m(X) + U \), where the error \( U \) satisfies \( E[U|X = 0] = 0 \) and \( V(U|X = x) = \sigma^2(x) \). The fuzzy QRKD estimand can then be estimated by

\[
\hat{QRKD}_f(\tau) = \hat{\beta}^+(\tau) - \hat{\beta}^-(\tau) / \hat{m}'(x_0^+) - \hat{m}'(x_0^-),
\]

where the numerator is the same as the one in Section 3 in the main text. For denominator, we use the local derivative estimator

\[
\hat{m}'(x_0^+) = - \frac{1}{nh_n^2} \sum_{i=1}^n b_i K'(\frac{x_i-x_0}{h_n})d_i^+ - \hat{m}(x_0^+) \left( - \frac{1}{nh_n^2} \sum_{i=1}^n K'(\frac{x_i-x_0}{h_n})d_i^+ \right)
\]

where \( \hat{m}(x_0^+) = \sum_{i=1}^n b_i K(\frac{x_i-x_0}{h_n})d_i^+ / \sum_{i=1}^n K(\frac{x_i-x_0}{h_n})d_i^+ \), as in Equation (4.14) of Pagan and Ullah (1999). The left counterpart, \( \hat{m}'(x_0^-) \), can be defined analogously. Notice that we are using \( h_n \) as the bandwidth for \( \hat{m}'(x_0^+) \). This is reasonable for the asymptotic argument because, as we will see, \( \hat{m}'(x_0^+) \) and \( \hat{\beta}^+(\tau) \) have the same rate of convergence. We make the following assumptions.

**Assumption 8.**

(i) The partial derivatives of \( f_X \), \( m \) and \( \int b^2 f_{BX}(b, \cdot) \) db exist up to the third order and are bounded.
(ii) There exists a $\delta > 0$ such that $E[|U|^{2+\delta} \mid X = x_0] < \infty$ and $\int |K(u)|^{2+\delta} < \infty$.

(iii) $\|K'\|_{\infty} < \infty$.

(iv) $f_{YXB}$ exists and is continuous in $x$ for each $(y, b) \in \mathbb{R}^2$. Also, there exists $a > 0$ such that $|m'(x_0^+) - m'(x_0^-)| > a$.

(v) There exists an $\bar{x}$ such that $Q(\tau|\cdot)$ is monotone on $(x_0, \bar{x})$.

**Theorem 4.** Under Assumptions 6, 7, and 8, we have

$$
\sqrt{nh_n^2} (Q \hat{R} K D f(\tau) - Q R K D f(\tau)) \Rightarrow \frac{(m'(x_0^+) - m'(x_0^-))G_\Delta(\tau) - (\beta^+(\tau) - \beta^-(\tau))G_\Delta(b)}{(m'(x_0^+) - m'(x_0^-))^2},
$$

where $G_\Delta$ is a Gaussian process with mean zero and covariance function as following: for any given $r, s, \tau \in T$ and let $b$ stand for the dimension of $m'(x)$, $\text{Cov}(G_\Delta(b), G_\Delta(b)) = \sigma_{b,b}^+ + \sigma_{b,b}^-$,

$$
\text{Cov}(G_\Delta(\tau), G_\Delta(b)) = \sigma_{\tau,b}^+ + \sigma_{\tau,b}^-, \text{Cov}(G_\Delta(r), G_\Delta(s)) = \sigma_{r,s}^+ + \sigma_{r,s}^-,
$$

where

$$
\sigma_{b,b}^+ = \frac{\sigma^2(x_0)}{f_X(x_0)} \int K'(v)^2 dv,
$$

$$
\sigma_{\tau,b}^+ = \frac{\iota_2'(N^+)^{-1}}{c(\tau) f_X(x_0)^2 f_Y|X(Q(\tau|x_0^+)|x_0^+)} \int \int (\tau - 1 \{y \leq Q(\tau|x_0)\})(1, \frac{v}{c(\tau)})' \times \int K(\frac{v}{c(\tau)})K'(v)(b - E[b(X, \mathcal{E}_1)|X = x_0])f_{YXB}(y, x_0 + h_n v, b) dv dy db,
$$

$$
\sigma_{r,s}^+ = \frac{c(1/2)}{c(r)c(s) f_X(x_0) f_Y|X(Q(\tau|x_0^+)|x_0) f_Y|X(Q(\tau|x_0^+)|x_0)} (r \wedge s - rs) \iota_2'(N^+)^{-1} T^+(r, s)(N^+)^{-1} \iota_2,
$$

and the left counterparts are defined analogously.

A proof of this theorem is provided in Section B.8.

**B Mathematical Appendix**

In addition to those notations introduced in the main text, we define the linear extrapolation error

$$
e_i = \left[Q(\tau|x_i^+) + (x_i - x_0) \frac{\partial Q(\tau|x_i^+)}{\partial x} \right] - Q(\tau|x_i) \text{ and the estimation errors } \hat{\phi}(\tau) = \sqrt{nh_{n,\tau}}[\hat{\alpha}^+(\tau) - Q(\tau|x_0^+), h_{n,\tau} \hat{\beta}^+(\tau) - \frac{\partial Q(\tau|x_0^+)}{\partial x}]'.
$$

Although they are not direct objects of interest, the level estimators are denoted by $\hat{\alpha}^+(\tau) = \iota_1' \arg \min_{\alpha, \beta} \sum_{i=1}^n d_i^+ K(\frac{x_i - x_0}{h_{n,\tau}}) p_{\tau}(y_i - \alpha - \beta(x_i - x_0))$ and $\hat{\alpha}^-(\tau) = \iota_1' \arg \min_{\alpha, \beta} \sum_{i=1}^n d_i^- K(\frac{x_i - x_0}{h_{n,\tau}}) p_{\tau}(y_i - \alpha - \beta(x_i - x_0))$. 


\[ \ell_1 \arg \min_{\alpha, \beta} \sum_{i=1}^{n} d_i^{-} K(\frac{x_i-x_0}{h_{n, \tau}}) \rho_{\tau}(y_i - \alpha - \beta(x_i - x_0)), \text{ where } \ell_1 = [1, 0]'. \] We also use short-hand notations
\[ z_{i,n,\tau}' = (1,(x_i-x_0)/h_{n,\tau}) \text{ and } K_{i,n,\tau} = K(\frac{x_i-x_0}{h_{n,\tau}}). \]

### B.1 Auxiliary Lemmas: Uniform Consistency

In this appendix section, we develop the following auxiliary results that show uniform convergences of some useful local sample moments over \( T \). While it is stated for right observations only, we remark that similar results hold for left observations too.

**Lemma 1.** Under Assumption 6, we have

\( (i) (nh_{n,\tau})^{-1} \sum_{i=1}^{n} K_{i,n,\tau} z_{i,n,\tau} z_{i,n,\tau}' d_{i}^{+} \overset{P}{\to} f_X(x_0) N^+ \text{ uniformly in } \tau \in T; \)

\( (ii) (nh_{n,\tau})^{-1} \sum_{i=1}^{n} f_{Y|X}(\tilde{y}_i|x_i) K_{i,n,\tau} z_{i,n,\tau} z_{i,n,\tau}' d_{i}^{+} \overset{P}{\to} f_{Y|X}(Q(\tau|x_0^+)|x_0^+) f_X(x_0) N^+ \text{ uniformly in } \tau \in T \)

with any \( \tilde{y}_i \) lying between \( Q(\tau|x_i) \) and \( Q(\tau|x_i) + e_i(\tau) + (nh_{n,\tau})^{-1/2} z_{i,n,\tau} d_{i}^{+} \hat{\phi}(\tau) \) for each \( i; \)

\( (iii) (nh_{n,\tau})^{-1} \sum_{i=1}^{n} \left( \frac{x_i-x_0}{h_{n,\tau}} \right)^2 \frac{\partial^2 Q(\tau|x_0^+)}{\partial x^2} h_{n,\tau}^2 z_{i,n,\tau} K_{i,n,\tau} d_{i}^{+} \overset{P}{\to} \int_{0}^{\infty} u^2 \frac{\partial^2 Q(\tau|x_0^+)}{\partial x^2}(1, u)' K(u) du \text{ uniformly in } \tau \in T. \)

**Proof.** (i): We claim \( E \left[ (nh_{n,\tau})^{-1} \sum_{i=1}^{n} K_{i,n,\tau} z_{i,n,\tau} z_{i,n,\tau}' d_{i}^{+} \right] \to f_X(x_0) N^+ \text{ uniformly in } \tau \in T \) and
\( (nh_{n,\tau})^{-1} \sum_{i=1}^{n} K_{i,n,\tau} z_{i,n,\tau} z_{i,n,\tau}' d_{i}^{+} \overset{P}{\to} E \left[ (nh_{n,\tau})^{-1} \sum_{i=1}^{n} K_{i,n,\tau} z_{i,n,\tau} z_{i,n,\tau}' d_{i}^{+} \right] \text{ uniformly in } \tau \in T. \) We provide a proof for only \( \ell_2^{(nh_{n,\tau})^{-1}} \sum_{i=1}^{n} K_{i,n,\tau} z_{i,n,\tau} z_{i,n,\tau}' d_{i}^{+} \overset{P}{\to} \int_{0}^{\infty} u^2 K(u) \{ f_X(x_0) + f'(x_0) u h_{n,\tau} + o(u h_{n,\tau}) \} du. \)

Assumption 6 (i) (a) implies that \( f'_{X}(x) \) is bounded in a neighbourhood of \( x_0 \). Assumption 6 (iv) and (v) then implies that the right-hand side in the equation above is \( f_X(x_0) \ell_2^{(nh_{n,\tau})^{-1}} + O(h_{n,\tau}) \). Since \( O(h_{n,\tau}) = O(c(\tau)h_{n}) = O(\dot{c} h_{n}) \), the convergence is uniform in \( \tau \in T. \)

In the remainder, we show the uniform convergence of the stochastic part using empirical process theories. We denote \( \sup \{ supp(K) \} = \bar{k} \). Let \( \mathcal{F} = \{ x_i \to \frac{a^2(x_i-x_0)^2}{c(\tau)^3} \mathbb{1} \{ K(\frac{a(x_i-x_0)}{c(\tau)}) > 0 \} K(\frac{a(x_i-x_0)}{c(\tau)}) \mathbb{1} \{ x_i \geq \} \}

\]
\( x_0 \) : \( (\tau, a) \in T \times [0, \infty) \). Because each \( f \in \mathcal{F} \) is right continuous under (iv) of Assumption 6, this family is a point-wise measurable class – see Einmahl and Mason (2005) and Section 2.3 of van der Vaart and Wellner (1996). For each \((\tau, a) \in \{ T \times [0, \infty) , x_i \mapsto \frac{a^2(x_i - x_0)^2}{c(\tau)^2} \mathbb{1}\{K(\frac{a(x_i - x_0)}{c(\tau)}) > 0\} \) is monotone on its support and bounded by \( \bar{k}^2 \) under (iv) of Assumption 6. Meanwhile, \( c(\tau) \) is finite and bounded away from 0 uniformly in \( \tau \in T \) under (v) of Assumption 6, and so is \( (1/c(\tau)) \). Therefore, \( x_i \mapsto \frac{a^2(x_i - x_0)^2}{c(\tau)^2} \mathbb{1}\{K(\frac{a(x_i - x_0)}{c(\tau)}) > 0\} \) is of bounded variation. \( x_i \mapsto \mathbb{1}\{x_i \geq x_0\} \) is trivially of bounded variation. Putting them together, we have that each element in \( \mathcal{F} \) is of bounded variation with a measurable envelope \( F(x_i) = \frac{\bar{k}^2\|K\|_{\infty}}{\varepsilon} \), where the constant \( \varepsilon \) is from (v) of Assumption 6. Since \( F \) is a finite constant, \( \|F\|^2_{p,2} = \int |F|^2 f_X(x_i) dx_i = P|F|^2 < \infty \). Without loss of generality as it is bounded, we can assume \( F \leq 1 \). By Theorem 3.6.12 of Giné and Nickl (2016), \( \mathcal{F} \) is of VC-type (Euclidean), that is to say that there exists constant \( k, v < \infty \) such that \( \sup_Q \log N(\epsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q)) \leq (\frac{k}{\varepsilon})^v \) for \( 0 < \epsilon \leq 1 \) and for all probability measures \( Q \) supported on \( \text{supp}(X) \).

This implies the uniform entropy integral \( J(1, \mathcal{F}, F) = \sup_Q \int_0^1 \sqrt{1 + \log N(\epsilon \|F\|_{Q,2}, \mathcal{F}, L_2(Q))} d\epsilon < \infty \). Since \( F \in L_2(P) \), we can apply Theorem 5.2 of Chernozhukov, Chetverikov and Kato (2014) to obtain

\[
E \left[ \sup_{f \in \mathcal{F}} \left| \frac{1}{\sqrt{n}} \sum_{i=1}^n f(X_i) - \int f dP \right| \right] \leq C \{ J(1, \mathcal{F}, F) \|F\|_{p,2} + BJ^2(1, \mathcal{F}, F) \frac{\delta v^2}{\sqrt{n}} \} < \infty
\]

with \( B = \sqrt{\max_{1 \leq i \leq n} F(X_i)} < \infty \) and an universal constant \( C > 0 \).

Multiplying both sides by \( (\sqrt{nh_n})^{-1} \) yields

\[
E \left[ \sup_{\tau \in T} \left( nh_n, \tau \right)^{-1} \sum_{i=1}^n \left( \frac{x_i - x_0}{h_n, \tau} \right)^2 K(\frac{x_i - x_0}{h_n, \tau}) d_i^+ - E \left[ (nh_n, \tau)^{-1} \sum_{i=1}^n \left( \frac{x_i - x_0}{h_n, \tau} \right)^2 K(\frac{x_i - x_0}{h_n, \tau}) d_i^+ \right] \right] \leq \frac{1}{\sqrt{nh_n, \tau} C \{ J(1, \mathcal{F}, F) \|F\|_{p,2} + BJ^2(1, \mathcal{F}, F) \frac{\delta v^2}{\sqrt{n}} \} }.
\]

Thus, the right hand side goes to 0 if \( nh_n, \tau \to \infty \) as \( n \to \infty \). Finally, Markov inequality gives the desired result.

(ii): As in the proof of part (i), we show \( E \left[ (nh_n, \tau)^{-1} \sum_{i=1}^n f_Y|X(y_i|x_i) K_i, n, \tau z_i, n, \tau z_i'_{i, n, \tau} d_i^+ \right] \to f_Y|X(Q|\tau z_0^+) f_X(x_0) N^+ \) uniformly in \( \tau \in T \), and then we also show \( (nh_n, \tau)^{-1} \sum_{i=1}^n f_Y|X(y_i|x_i) \to f_Y|X(Q|\tau x_0^+) f_X(x_0) N^+ \).
\[ K_{i,n,\tau} z_{i,n,\tau} z_{i,n,\tau}' d_i^+ \to E \left[ (nh_{n,\tau})^{-1} \sum_{i=1}^n f_{Y|X}(\tilde{y}_i|x_i)K_{i,n,\tau} z_{i,n,\tau} z_{i,n,\tau}' d_i^+ \right] \] uniformly in \( \tau \in T \).

First we bound \( |Q(\tau|x_i) - (Q(\tau|x_i) + e_i(\tau) + (nh_{n,\tau})^{-1/2} z_{i,n,\tau}' \hat{\phi}(\tau))| \{ K(\frac{z_{i,n,\tau}'}{nh_{n,\tau}}) > 0 \} = |e_i(\tau) + (nh_{n,\tau})^{-1/2} z_{i,n,\tau}' \hat{\phi}(\tau)| \{ K(\frac{z_{i,n,\tau}'}{nh_{n,\tau}}) > 0 \} \). Following part 1 of the proof of Theorem 1 in Qu and Yoon (2015a), and using part (i) of our Lemma 1, we have \( \hat{\phi}(\tau) \leq \sqrt{\log nh_n} \) with probability approaching one uniformly in \( \tau \in T \). Therefore, under Assumption 6 (iv) and (v), we have \( (nh_{n,\tau})^{-1/2} z_{i,n,\tau}' \hat{\phi}(\tau) \{ K(\frac{z_{i,n,\tau}'}{nh_{n,\tau}}) > 0 \} = O_p(\sqrt{\log nh_n}) \) uniformly in \( i \) and \( \tau \in T \).

We next bound \( e_i(\tau) \{ K(\frac{z_{i,n,\tau}'}{nh_{n,\tau}}) > 0 \} = \{ [Q(\tau|x_0^+ + (x_i - x_0) \frac{\partial Q(x_0^+)}{\partial x}] - Q(\tau|x_i) \} \{ K(\frac{z_{i,n,\tau}'}{nh_{n,\tau}}) > 0 \} \). By the mean value expansion of \( Q(\tau|x_i) \) at \( x = x_0 + \delta \) and let \( x \to x_0^+ \), we have \( e_i(\tau) \{ K(\frac{z_{i,n,\tau}'}{nh_{n,\tau}}) > 0 \} \leq M \| (\tau,x_0) - (\tau,\tilde{x}) \| (x_i - x_0) \{ K(\frac{z_{i,n,\tau}'}{nh_{n,\tau}}) > 0 \} \leq M \| (\max T, x_0) - (\min T, \tilde{x}) \| O(h_n) = O(h_n) \) uniformly in \( \tau \in T \) for some constant \( M \) by Lipschitz continuity and properties of bandwidth from Assumption 6 (iii) (a), (iii) (b) and (iv).

Similarly, we can bound \( |Q(\tau|x_0^+ - Q(\tau|x_i)| \{ K(\frac{z_{i,n,\tau}'}{nh_{n,\tau}}) > 0 \} \leq M \| (x_0, \tau) - (x_i, \tau) \| = O(h_n) \) uniformly in \( \tau \in T \) for some constant \( M \).

Combining the auxiliary results above, we have \( |Q(\tau|x_i) - \tilde{y}_i| \{ K(\frac{z_{i,n,\tau}'}{nh_{n,\tau}}) > 0 \} \leq |Q(\tau|x_i) - (Q(\tau|x_i) + e_i(\tau) + (nh_{n,\tau})^{-1/2} z_{i,n,\tau}' \hat{\phi}(\tau))| \{ K(\frac{z_{i,n,\tau}'}{nh_{n,\tau}}) > 0 \} = O_p(\sqrt{\log nh_n}) + O(h_n) \) and \( |Q(\tau|x_0^+ - Q(\tau|x_i)| \{ K(\frac{z_{i,n,\tau}'}{nh_{n,\tau}}) > 0 \} \leq 1 \} = O(h_n) \) uniformly in \( i \) and in \( \tau \in T \). By the triangle inequality, \( |Q(\tau|x_0^+ - \tilde{y}_i| \{ K(\frac{z_{i,n,\tau}'}{nh_{n,\tau}}) > 0 \} \leq O_p(\sqrt{\log nh_n}) + O(h_n) \) uniformly in \( i \) and \( \tau \in T \).

Using Assumption 6 (i) (a), (i) (b), (ii) (a) and (iv) along with the asymptotic bounds obtained above,

\[ E\tilde{c}'_2(nh_{n,\tau})^{-1} \sum_{i=1}^n f_{Y|X}(\tilde{y}_i|x_i)K_{i,n,\tau} z_{i,n,\tau} z_{i,n,\tau}' d_i^+ \]

\[ = f_{Y|X}(Q(\tau|x_0^+)|x_0^+), f_X(x_0)\tilde{c}_2 N^+ + O_p((\log nh_n/nh_n)^{1/2}) + O(h_n) \]

holds uniformly in \( \tau \in T \). Convergence of the other entries follows similarly.

For the second part, let \( \mathcal{F} = \{ x_i \to \frac{a(x_i-x_0)^2 f_{Y|X}(b|x_i)}{c(\tau)^2} \{ K(\frac{a(x_i-x_0)}{c(\tau)}) > 0 \} K(\frac{a(x_i-x_0)}{c(\tau)}) \{ x_i \geq \} \}

\]
\( x_0 \) : \((\tau, a, b) \in T \times [0, \infty) \times [\inf_{(\tau, x) \in T \times (x_0, \bar{x})} Q(\tau | x), \sup_{(\tau, x) \in T \times (x_0, \bar{x})} Q(\tau | x)] \). Notice that the interval of infimum and supremum is bounded by Assumption 6 (iii) (b). An argument similar to the proof of (i) shows that each element in \( \mathcal{F} \) is of bounded variation with a measurable envelope \( F(x_i) = \frac{k^2||K||_\infty}{\xi} \sup_{(y, x)} f_Y(x) \), where the supremum in the numerator is taken over \((x, y) \in (x_0, \bar{x}] \times [\inf_{(\tau, x) \in T \times (x_0, \bar{x})} Q(\tau | x), \sup_{(\tau, x) \in T \times (x_0, \bar{x})} Q(\tau | x)]\), under parts (ii), (iv) and (v) of Assumption 6. Using the VC-type class argument and applying the same inequality from Chernozhukov, Chetverikov and Kato (2014) and Markov inequality then gives the desired result.

(iii): We focus on the entry \( \phi_2(nh_{n, \tau}^3)^{-1} \sum_{i=1}^n \frac{1}{2} \left( \frac{x_i-x_0}{nh_{n, \tau}} \right)^2 \frac{\partial^2 Q(\tau | x_i^+)}{\partial x^2} \bar{h}_{n, \tau}^3 z_i, n, \tau^+ K_i, n, \tau^+ d_i^+ \phi_2. \) Similar arguments apply to the other entries. The process is similar to (i). The deterministic part can be shown by computing the expectation. For the stochastic part, let \( \mathcal{F} = \{ x_i \mapsto a^{-3}(x_i-x_0)^3 \frac{\partial^2 Q(\tau | x_i^+)}{\partial x^2} \}, \) \( K(\frac{a(x_i-x_0)}{c(\tau)}) > 0 \} K(\frac{a(x_i-x_0)}{c(\tau)}) \} \{ x_i \geq x_0 \} : (\tau, a) \in T \times [0, \infty) \} \) is a VC type class with a measurable envelope \( F(x_i) = \frac{k^2m||K||_\infty}{\xi} \) where \( m = \sup_T \frac{\partial^2 Q(\tau | x_0^+)}{\partial x^2} < \infty \) under Assumption 6 (iii)(b) and \( T \) is compact. Then the same uniform consistency argument applies under parts (iii) (a) (b), (iv) and (v) of Assumption 6. The same inequality from Chernozhukov, Chetverikov and Kato (2014) and Markov inequality give the desired result.

Following similar reasoning, we also have the corresponding results for uniform convergences of some local moments for local quadratic regression.

**Lemma 2.** Under Assumption 6, 9, we have

(i) \((nh_{n, \tau})^{-1} \sum_{i=1}^n \bar{K}_{i, n, \tau} \hat{z}_{i, n, \tau}^+ d_i^+ \xrightarrow{P} f_X(x_0) \hat{N}^+ \) uniformly in \( \tau \in T \);

(ii) \((nh_{n, \tau})^{-1} \sum_{i=1}^n f_Y(x_i) \bar{K}_{i, n, \tau} \hat{z}_{i, n, \tau}^+ d_i^+ \xrightarrow{P} f_Y(x) (Q(\tau | x_0^+) | x_0^+) f_X(x_0) \hat{N}^+ \) uniformly in \( \tau \in T \) with any \( \hat{y}_i \) lying between \( Q(\tau | x) \) and \( Q(\tau | x) + \varepsilon_i(\tau) + (nh_{n, \tau})^{-1/2} \hat{z}_{i, n, \tau}^+ \hat{\phi}(\tau) \) for each \( i \), where \( \hat{\phi}(\tau) = \sqrt{nh_{n, \tau}[\hat{a}^+(\tau) - Q(\tau | x_0^+) \bar{h}_{n, \tau}^\beta(\hat{\beta}^+(\tau) - \frac{\partial Q(\tau | x_0^+)}{\partial x})], \] and \( \bar{h}_{n, \tau}^\beta(\hat{\beta}^+(\tau) - \frac{\partial Q(\tau | x_0^+)}{\partial x}) \] and \( \varepsilon_i(\tau) = [Q(\tau | x_0^+) + (x_i - x_0) \frac{\partial Q(\tau | x_0^+)}{\partial x} + (x_i - x_0) \frac{\partial^2 Q(\tau | x_0^+)}{\partial x^2}] - Q(\tau | x) \); \( \hat{\phi}(\tau) \)

(iii) \((nh_{n, \tau})^{-1} \sum_{i=1}^n \tilde{d}_i^+ \frac{\partial Q(\tau | x_0^+)}{\partial x} \bar{h}_{n, \tau}^\beta \hat{z}_{i, n, \tau} \bar{K}_{i, n, \tau} \xrightarrow{P} f_X(x_0) \int_{0}^{\infty} u^3 \frac{\partial^2 Q(\tau | x_0^+)}{\partial x^2} [1, u, u^2] K(u) du \)
uniformly in $\tau \in T$.

### B.2 Uniform Bahadur Representation

**Lemma 3.** Under Assumption 6, we have

\[
\sqrt{nh_{n,\tau}^3} \left( \hat{\beta}^+(\tau) - \frac{\partial Q(\tau|x_0^+)}{\partial x} - h_{n,\tau} \frac{\partial^2 Q(\tau|x_0^+)}{\partial x^2} \right) \int_0^\infty u^2 \frac{\partial^2 Q(\tau|x_0^+)}{\partial x^2}(1, u) K(u) du
\]

\[
= \frac{\partial^2 Q(\tau|x_0^+)}{\partial x^2} \left( \sum_{i=1}^n (\tau - 1) \{y_i \leq Q(\tau|x_i)\} z_{i, n, \tau} K_{i, n, \tau} d_i^+ \right)
\]

uniformly in $\tau \in T$. Similar results hold for $\hat{\beta}^-(\tau)$.

A proof is provided below – auxiliary results of uniform consistency that are used to prove this theorem are also provided in Section B.1 in the supplementary appendix. The Bahadur representation obtained in this lemma is uniform in quantiles $\tau \in T$ for a fixed location of the running variable $x$. Kong, Linton and Xia (2010) derived a Bahadur representation that is uniform in $x$ for a fixed quantile level. Guerre and Sabbah (2012) derived a result that is uniform in both $\tau$ and $x$ for interior points – see also Sabbah (2014). Since we are interested in a representation at the boundary point $x_0$ of the truncated distribution, and since we did not require the uniformity in $x$, and we developed our approach more closely following Qu and Yoon (2015a).

**Proof.** For this lemma, we mostly follow the proof of Theorem 1 in Qu and Yoon (2015a). The major difference is that we focus on the second coordinate of $\hat{\phi}$ instead of the first one. By step 1 of the proof of Theorem 1 in Qu and Yoon (2015a) which is applicable under parts (i), (ii) (a), (ii) (b), (iii) (a), (iii) (b), (iv) and (v) of our Assumption 6, we have $\sup_{\tau \in T} \left\| \hat{\phi}(\tau) \right\| \leq (\log nh_n)^{1/2}$ with probability approaching one as $n \to \infty$. Asymptotically, therefore, we only need to focus on studying the behavior of the subgradient

\[
\text{(subgradient)} = \sum_{i=1}^n \{\tau - 1(u_i^0(\tau) \leq e_i(\tau) + (nh_{n,\tau})^{-1/2} z_{i, n, \tau} \hat{\phi}(\tau))\} z_{i, n, \tau} K_{i, n, \tau} d_i^+
\]

on the set $\Phi_n = \{ (\tau, \phi(\tau)) : \tau \in T, \|\phi(\tau)\| \leq \log^{1/2}(nh_n) \}$, where $u_i^0(\tau) = y_i - Q(\tau|x_i)$ and $e_i(\tau) = \ldots$
\[
[Q(\tau|x_0^\tau) + (x_i - x_0) \frac{\partial Q(\tau|x_0^\tau)}{\partial x}] - Q(\tau|x_i).
\]

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

Theorem 2.1 of Koenker (2005) and Assumption 6 (iv) imply \((nh_n)^{-1/2}(\text{subgradient}) = O_P((nh_n)^{-1/2})\) uniformly in \(\tau \in T\).

Following Qu and Yoon (2015a), we can rewrite the subgradient (scaled by \((nh_n)^{-1/2}\)) as
\[
(nh_n)^{-1/2} \sum_{i=1}^{n} \{\tau - \mathbb{1}(u_i^0(\tau) \leq e_i(\tau) + (nh_n,\tau)^{-1/2}z_{i,n,\tau}' \phi(\tau))\}z_{i,n,\tau}K_{i,n,\tau}d_i^+
= \{S_n(\tau, \hat{\phi}(\tau), e_i(\tau)) - S_n(\tau, 0, e_i(\tau))\} + \{S_n(\tau, 0, e_i(\tau)) - S_n(\tau, 0, 0)\} + S_n(\tau, 0, 0)
+ (nh_n)^{-1/2} \sum_{i=1}^{n} \{\tau - P((u_i^0(\tau) \leq e_i(\tau) + (nh_n,\tau)^{-1/2}z_{i,n,\tau}' \phi(\tau)|x_i))\}z_{i,n,\tau}K_{i,n,\tau}d_i^+
\]

The differences inside the first two pairs of curly brackets are of order \(o_p(1)\) on the set \(\Phi_n\) by Lemma B5 of Qu and Yoon (2015a), which is applicable under parts (i), (ii) (a), (ii) (b), (iii) (a), (iii) (b), (iv) and (v) of our Assumption 6. The \(S_n(\tau, 0, 0)\) term is \(O_p(1)\) under Assumption 6 (i) (a), (iv), (v). The conditional probability in the last term is a conditional CDF of \(Y|X\). Applying the first order mean value expansion to the last term at \(y = Q(\tau|x_i)\) yields
\[
(nh_n)^{-1/2} \sum_{i=1}^{n} \{\tau - P((u_i^0(\tau) \leq e_i(\tau) + (nh_n,\tau)^{-1/2}z_{i,n,\tau}' \phi(\tau)|x_i))\}z_{i,n,\tau}K_{i,n,\tau}d_i^+
= - (nh_n)^{-1/2} \sum_{i=1}^{n} f_{Y|X}(\bar{y}_i|x_i)e_i(\tau)z_{i,n,\tau}K_{i,n,\tau}d_i^+
- (nh_n)^{-1/2}(nh_n,\tau)^{-1/2}(\sum_{i=1}^{n} f_{Y|X}(\bar{y}_i|x_i)K_{i,n,\tau}z_{i,n,\tau}z_{i,n,\tau}' d_i^+)\hat{\phi}(\tau),
\]

where \(\bar{y}_i\) lies between \(Q(\tau|x_i)\) and \(Q(\tau|x_i) + e_i(\tau) + (nh_n,\tau)^{-1/2}z_{i,n,\tau}' d_i^+\). The \(\hat{\phi}(\tau)\).

Taking the above auxiliary results together, we can now rewrite subgradient (scaled by \((nh_n)^{-1/2}\)) as
\[
S_n(\tau, 0, 0) - (nh_n)^{-1/2} \sum_{i=1}^{n} f_{Y|X}(\bar{y}_i|x_i)e_i(\tau)z_{i,n,\tau}K_{i,n,\tau}d_i^+
- (nh_n)^{-1/2}(nh_n,\tau)^{-1/2}(\sum_{i=1}^{n} f_{Y|X}(\bar{y}_i|x_i)K_{i,n,\tau}z_{i,n,\tau}z_{i,n,\tau}' d_i^+)\hat{\phi}(\tau).
\]
Recall that this subgradient (scaled by \((nh_n)^{-1/2}\)) is \(o_p(1)\) uniformly in \(\tau \in T\).

By Lemma 1 (ii), \((nh_n,\tau)^{-1}\sum_{i=1}^{n} f_{Y \mid X}(\tilde{y}_i|x_i)K_{i,n,\tau}z_{i,n,\tau}z_{i,n,\tau}d_i^+ \overset{P}{\to} f_{Y \mid X}(Q(\tau|x_0^+)|x_0^+|x_0^+)f_{X}(x_0)N^+\)
uniformly in \(\tau \in T\) and so

\[
S_n(\tau, 0, 0) - (nh_n)^{-1/2}\sum_{i=1}^{n} f_{Y \mid X}(\tilde{y}_i|x_i)e_i(\tau)z_{i,n,\tau}K_{i,n,\tau}d_i^+ = \left(\frac{h_n}{h_n,\tau}\right)^{1/2}[f_{Y \mid X}(Q(\tau|x_0^+)|x_0^+)f_{X}(x_0)N^++o_p(1)] \hat{\phi}(\tau) + o_p(1)
\]

uniformly in \(\tau \in T\). Since \(N^+\) is positive definite and \(f_{Y \mid X}(Q(\tau|x_0^+)|x_0^+)f_{X}(x_0) > 0\) by parts (i), (ii) (b) and (iv) of Assumption 6, we obtain

\[
\hat{\phi}(\tau) = (f_X(x_0)f_{Y \mid X}(Q(\tau|x_0^+)|x_0^+)N^++o_p(1))^{-1} \times \left(\left[\left(\frac{h_n}{h_n,\tau}\right)^{1/2}S_n(\tau, 0, 0) - (nh_{n,\tau})^{-1/2}\sum_{i=1}^{n} e_i(\tau)z_{i,n,\tau}K_{i,n,\tau}d_i^+ + o_p(1)\right]\right) \tag{B.1}
\]

uniformly in \(\tau \in T\).

Under Assumption 6 (iii) (a), (iii) (b) the Taylor expansion and \(e_i(\tau) = \left[Q(\tau|x_0^+)+(x_i-x_0)\frac{\partial Q(\tau|x_0^+)}{\partial x}\right]-Q(\tau|x_i)\) suggest that for any \(x_i \geq x_0\) such that \((x_i-x_0)/h_{n,\tau} \in \text{supp}(K)\), we have

\[
e_i(\tau) = -\frac{1}{2}\left(\frac{x_i-x_0}{h_{n,\tau}}\right)^2\frac{\partial^2 Q(\tau|x_0^+)}{\partial x^2}h_{n,\tau}^2 + o(\frac{h_{n,\tau}}{h_{n,\tau}})
\]

uniformly in \(\tau \in T\). By Lemma 1 (iii), we have \(-((h_{n,\tau})^{-2}(nh_{n,\tau})^{-1}\sum_{i=1}^{n} e_i(\tau)z_{i,n,\tau}K_{i,n,\tau}d_i^+ \overset{P}{\to} f_X(x_0)\int_0^{\infty} u^2\frac{\partial^2 Q(\tau|x_0^+)}{\partial x^2}(1,u)'K(u)du\)
uniformly in \(\tau \in T\). Substitute this and \(S_n(\tau, 0, 0)\) with its definition into equation (B.1), we have

\[
\hat{\phi}(\tau) = \frac{(N^+)^{-1}(nh_{n,\tau})^{-\frac{1}{2}}\sum_{i=1}^{n}(\tau - 1\{y_i \leq Q(\tau|x_i)\})z_{i,n,\tau}K_{i,n,\tau}d_i^+}{f_X(x_0)f_{Y \mid X}(Q(\tau|x_0^+)|x_0^+)} + (nh_{n,\tau}^{-1})^{\frac{1}{2}}(N^+)^{-1}\int_0^{\infty} u^2\frac{\partial^2 Q(\tau|x_0^+)}{\partial x^2}(1,u)'K(u)du + o_p(1)
\]

uniformly in \(\tau \in T\). \(\square\)

### B.3 Proof of Theorem 2

**Proof.** By Theorem 18.14 in van der Vaart (1998), it suffices to show the finite dimensional convergence in distribution and the tightness. The tightness follows from the boundedness assumptions of \(f_X\) and
For finite dimensional convergence in distribution, we introduce a couple of additional short-hand notations. For each \( \tau \in T \), let

\[
Z_{n,i}^+(\tau) = \frac{1}{\sqrt{n}} \left( \ell_t \right)_2^{-1}(N^+) \left( r - \mathbb{1}\{y_i \leq Q(r | x_i)\} \right) z_{n,r} d_i^+ K_{i,n,\tau}.
\]

For any finite set \( \{\tau_1, ..., \tau_k\} \subset T \) of quantiles, we write \( W_{n,i}^+(\tau_1, ..., \tau_k) = (Z_{n,i}^+(\tau_1), ..., Z_{n,i}^+(\tau_k))' \).

Note that

\[
E \left( \left( \ell_t \right)_2^{-1}(N^+) \left( r - \mathbb{1}\{y_i \leq Q(r | x_i)\} \right) z_{i,n,r} d_i^+ K_{i,n,r} \right) = E \left( \left( \ell_t \right)_2^{-1}(N^+) \left( r - \mathbb{1}\{y_i \leq Q(r | x_i)\} \right) z_{i,n,r} d_i^+ K_{i,n,r} \right)
\]

holds for each \( n \in \mathbb{N} \) and \( r \in T \). Since it is \( n \)-invariant, let \( \sum_{i=1}^n \text{Cov} W_{n,i}^+(\tau_1, ..., \tau_k) = \Sigma \{\tau_1, ..., \tau_k\} \). The entry of the covariance matrix \( \Sigma \{\tau_1, ..., \tau_k\} \) corresponding to the pair \( r, s \in T \) of quantiles is given by

\[
E \left( \left( \ell_t \right)_2^{-1}(N^+) \left( r - \mathbb{1}\{y_i \leq Q(r | x_i)\} \right) z_{i,n,r} d_i^+ K_{i,n,r} \right) \left( \left( \ell_t \right)_2^{-1}(N^+) \left( s - \mathbb{1}\{y_i \leq Q(s | x_i)\} \right) z_{i,n,s} d_i^+ K_{i,n,s} \right)'
\]

\[
= \frac{1}{f_X(x_0) \sqrt{h_{n,r} h_{n,s}}} \left( \ell_t \right)_2^{-1} z_{i,n,r} K_{i,n,s} z_{i,n,s}' \left( \ell_t \right)_2^{-1} d_i^+ (r \wedge s - r s)
\]

\[
= (\kappa(r) \kappa(s))^{-1/2} (r \wedge s - r s) \left( \ell_t \right)_2^{-1} \int_0^\infty \frac{1}{u} K\left( \frac{u}{\kappa(r)} \right) K\left( \frac{u}{\kappa(s)} \right) du (N^+)^{-1} d_i^2.
\]

We remark that the last line is invariant from \( h_{n,\tau} = c(\tau) h_n \) in Assumption 6 (v). This is finite because Assumption 6 (iv) and (v) imply

\[
|\int_0^\infty u^k K\left( \frac{u}{\kappa(r)} \right) K\left( \frac{u}{\kappa(s)} \right) du| \leq \|K\|_\infty \int_0^\infty u^k K\left( \frac{u}{\kappa(r)} \right) du < \infty \quad \text{for } k = 0, 1, 2 \text{ and all other parts are finite.}
\]

Secondly, we show that the moment condition of Lindeberg-Feller is satisfied. Write \( (N^+)^{-1} = [a \ b \ c] \), fix any finite set \( \{\tau_1, ..., \tau_k\} \subset T \) of quantiles. Under Assumptions 6 (i)(a)(b), (iv) and (v), we
have for any $\epsilon > 0$,

$$
\sum_{i=1}^{n} E \left[ \left\| W_{n,i}^{+}(\tau_{1}, ..., \tau_{k}) \right\|^2 \mathbb{1}\left( \left\| W_{n,i}^{+}(\tau_{1}, ..., \tau_{k}) \right\| > \epsilon \right) \right]
\leq n \sum_{i=1}^{n} E \left[ \sum_{j=1}^{k} Z_{n,i}^{+}(\tau_{j})^2 \mathbb{1}\left( \sum_{j=1}^{k} Z_{n,i}^{+}(\tau_{j})^2 > \epsilon^2 \right) \right]
\leq n \sum_{i=1}^{n} E \left[ \sum_{j=1}^{k} \left( \frac{\nu^{'}(N^{+})^{-1}(r - \mathbb{1}\{y_{i} \leq Q(\tau_{j}|x_{i})\}) z_{i,n,\tau_{j}} d_{i}^{+} K_{i,n,\tau_{j}}}{\sqrt{fx(0) nh_{n,\tau_{j}}}} \right)^2 \right.
\times \left. \mathbb{1}\left( \sum_{j=1}^{k} \left( \frac{\nu^{'}(N^{+})^{-1}(r - \mathbb{1}\{y_{i} \leq Q(\tau_{j}|x_{i})\}) z_{i,n,\tau_{j}} d_{i}^{+} K_{i,n,\tau_{j}}}{\sqrt{fx(0) nh_{n,\tau_{j}}}} \right)^2 > \epsilon^2 \right) \right]
\leq E \left[ k \sup_{\tau \in \mathbb{T}} \left( \frac{[b + c(\frac{x_{i}-x_{0}}{h_{n,r}})]^2 d_{i}^{+} K\left(\frac{x_{i}-x_{0}}{h_{n,r}}\right)}{fx(0) h_{n,\tau}} \right) \mathbb{1}\left\{ \sqrt{\frac{[b + c(\frac{x_{i}-x_{0}}{h_{n,r}})]^2 d_{i}^{+} K\left(\frac{x_{i}-x_{0}}{h_{n,r}}\right)}{fx(0) h_{n,\tau}}} > \epsilon^2 / m \right\} \right]
\leq \int_{0}^{\infty} m \left( \frac{b + \bar{k}}{m} \right) \|K\|_{\infty} K(u) \mathbb{1}\left\{ \frac{b + \bar{k}}{m} \|K\|_{\infty} > \epsilon^2 / m \right\} \left( fx(0) + O(uh_{n}) \right) du
= \int_{0}^{\infty} m_{1} K(u) \mathbb{1}\left\{ \frac{1}{nh_{n}} > m_{2} \epsilon^2 \right\} du fx(0) + O(h_{n})
$$

for some finite non-negative constant $m$, $m_{1}$, $m_{2}$ and $\bar{k} = \sup supp(K)$. By applying Dominated Convergence Theorem, the last equation goes to zero since $nh_{n} \to \infty$ implies the indicator goes to zero and all other terms are finite.

Therefore, by Lindeberg-Feller’s Central Limit Theorem, we have

$$
\left[ \sqrt{nh_{n}^{3} fx(0) f_{Y}|x(0) (Q(\tau|x_{i})|x_{0}^{+})} \left( \beta^{+} (\tau) - \frac{\partial Q(\tau|x_{i})}{\partial x} - h_{n,r} \nu^{'}(N^{+})^{-1} \frac{1}{2} \int_{0}^{\infty} u^2 \frac{\partial^2 Q(\tau|x_{i})}{\partial x^2} (1, u)^{T} K(u) du \right) \right]_{\tau \in (\tau_{1}, ..., \tau_{k})} \to N(0, \Sigma_{(\tau_{1}, ..., \tau_{k})})
$$

as $n \to \infty$. 

\[ \square \]

**B.4 Proof of Corollary 2**

*Proof.* The first part of the corollary follows immediately from Corollary 1. The second part follows by an application of the functional delta method (van der Vaart, 1998; Theorem 20.8). It suffices to show that the linear functional $\phi : g \mapsto g - \int_{T} gd\tau$ is Hadamard differentiable at $QRKD$ tangentially to $L_{m}^{\infty}(T)$. The linearity of $\phi^{'}_{QRKD}$ is trivial and the continuity is implied by its boundedness as
\[ \| \phi'_{QRKD}(g) \| \leq \| g \| |1 + \text{diam}(T)| \text{ for all } g \in L^\infty_m(T). \] We want to show that for \( g_n \to g \in L^\infty_m(T) \) and \( t_n \to 0 \)

\[ \frac{\phi(QRD + t_n g_n) - \phi(QRD)}{t_n} - \phi'_{QRKD}(g) \to 0 \text{ in } L^\infty_m(T). \]

The left hand side is equal to \( g_n - \int_T g_n d\tau - \phi'_{QRKD}(g) \). By the bounded convergence theorem, it converges to 0.

\[ \square \]

**B.5 Bias Reduction and Bandwidth Selection**

While the bias term \( h_{n,\tau}  \frac{\varepsilon(N^+)}{2} \int_0^\infty u^2 \frac{\partial^2 Q(\tau|x^+_0)}{\partial x^2} K(u) du \) in Theorem 2 is asymptotically negligible, some users may wish to mitigate this finite sample bias by explicitly estimating it. Such a reduction may make a difference especially when the underlying quantile regressions exhibit large curvatures. The second derivative \( \frac{\partial^2 Q(\tau|x^+_0)}{\partial x^2} \) can be consistently estimated by the one-sided local quadratic quantile smoother

\[ \hat{\lambda}^+ = \lambda^+ \) (0, 0, 1). The left version of the estimator \( \lambda^- \) can be similarly defined. To ensure effective bias correction with these estimators, we first obtain a uniform Bahadur representation for the second derivative estimator \( \hat{\lambda}^+ \) in a similar manner to Lemma 3, and then derive its asymptotic properties in a similar manner to Theorem 2. To this goal, we introduce additional short-hand notations. Some transformations of data points are denoted by \( z_{i,n,\tau} = (1, \frac{(x_i - x_0)}{\bar{h}_{n,\tau}}, \frac{(x_i - x_0)^2}{\bar{h}^2_{n,\tau}}) \) and \( \bar{K}_{i,n,\tau} = K((x_i - x_0)/\bar{h}_{n,\tau}) \). We let \( \bar{N}^+ \) denote the 3-by-3 matrix with the \((i,j)\)-th element given by \( \mu_{i+j-2}^- = \int_0^\infty u^{i+j-2} K(u) du \). With these notations, we make the following assumption.

**Assumption 9.**

(i) \( \partial Q^3(\tau|x)/\partial x^3 \) is finite and Lipschitz continuous on \( T \times ([\underline{x}, \bar{x}] \setminus \{x_0\}) \).

(ii) \( \partial Q^3(\tau|x^-_0)/\partial x^3 \) and \( \partial Q^3(\tau|x^-_0)/\partial x^3 \) are finite and Lipschitz continuous in \( \tau \in T \).
(iii) The bandwidths satisfy $h_{n,\tau} = c(\tau)\hat{h}_n$, where $nh_n^5 \to \infty$ and $\hat{h}_n = o(n^{-1/2})$ as $n \to \infty$, and $c(\cdot)$ is Lipschitz continuous with $0 < c \leq c(\tau) \leq \bar{c} < \infty$ for all $\tau \in T$.

(iv) $h_n^3\hat{h}_n^{-5} \to c \in [0, \infty]$.

Following a similar argument to the proof of Lemma 3 by Qu and Yoon (2015b) and using our Lemma 2 in Section B.1 yield the following Bahadur representation result for $\hat{\lambda}^+$. 

**Lemma 4.** Under Assumptions 6 and 9 (i)–(iii), we have

$$
\sqrt{nh_n^5}(\hat{\lambda}^+(\tau) - \frac{1}{2} \frac{\partial Q^2(\tau|x_0^+)}{\partial x^2})
= \nu_3(N^+)^{-1}(nh_{n,\tau})^{-1/2} \sum_{i=1}^n (\tau - 1\{y_i \leq Q(\tau|x_i)\})\bar{z}_{i,n,\tau}d_i^+K_{i,n,\tau} + o_p(1)
$$

uniformly in $\tau \in T$. A similar result holds for $\hat{\lambda}^-(\tau)$.

With a similar reasoning to the proof of Theorem 2, this representation yields the following asymptotic property. From Lemmas 3 and 4, we can write

$$
\sqrt{nh_n^3}(\hat{\beta}^+(\tau) - \hat{\lambda}^+(\tau)h_{n,\tau}\nu_2(N^+)^{-1}R^+ - \frac{\partial Q(\tau|x_0^+)}{\partial x})
= A_{n,\tau}^+ + B_{n,\tau}^+ + o_p(1),
$$

where

$$
A_{n,\tau}^+ = \nu_3(N^+)^{-1}(nh_{n,\tau})^{-1/2} \sum_{i=1}^n (\tau - 1\{y_i \leq Q(\tau|x_i)\})\bar{z}_{i,n,\tau}d_i^+K_{i,n,\tau} \frac{f_X(x_0)f_Y|X(\tau|x_0^+)|x_0^+}{f_X(x_0)f_Y|X(\tau|x_0^+)|x_0^+},
$$

$$
B_{n,\tau}^+ = - (\frac{h_n^3}{h_{n,\tau}^5})^{1/2} \nu_3(N^+)^{-1}(nh_{n,\tau})^{-1/2} \sum_{i=1}^n (\tau - 1\{y_i \leq Q(\tau|x_i)\})\bar{z}_{i,n,\tau}d_i^+K_{i,n,\tau} \frac{f_X(x_0)f_Y|X(\tau|x_0^+)|x_0^+}{f_X(x_0)f_Y|X(\tau|x_0^+)|x_0^+},
$$

and $R^+ = \int_{\hat{u}}^\infty u^2(1,u)K(u)du$.

Define $A_{n,\tau}^-$ and $B_{n,\tau}^-$ similarly to $A_{n,\tau}^+$ and $B_{n,\tau}^+$, but with $d_i^-$ and $(N^-)^{-1}$ used in place of $d_i^+$ and $(N^+)^{-1}$, respectively. Also, let $R^- = \int_{-\infty}^0 u^2(1,u)K(u)du$. We can define a version of the QRKD estimator with bias reduction by

$$
\hat{Q}RKDBR(\tau) = \frac{\hat{\beta}^+(\tau) - \hat{\lambda}^+(\tau)h_{n,\tau}\nu_2(N^+)^{-1}R^+ - (\hat{\beta}^-(\tau) - \hat{\lambda}^-)(\tau)h_{n,\tau}\nu_2(N^-)^{-1}R^-}{b'(x_0^+) - b'(x_0^-)}.
$$
The derivation above gives

$$\sqrt{nh^3_{n,\tau}}[\hat{Q}RKD_{BR}(\tau) - QRKD(\tau)] = G^{BR}_*(\tau) + o_p(1),$$

where \(G^{BR}_*(\tau) = \frac{(A^+_{n,\tau} + B^+_{n,\tau}) - (A^-_{n,\tau} + B^-_{n,\tau})}{(v^*(x_0^n) - v^*(x_0))}.$$

As pointed out in Section 6 of Qu and Yoon (2015a) and Remark 2 of Qu and Yoon (2015b), the distribution of the process \(G^{BR}_*(\tau)\) is conditionally pivotal and the randomness of \(A^\pm_{n,\tau}\) and \(B^\pm_{n,\tau}\) comes only from the same source of \(\{\tau - 1 \{y_i \leq Q(\tau|x_i)\}\}_{i=1}^n\). Therefore, we can simulate the distribution of \(G^{BR}_*(\tau)\) by the following two-step methods. First, generate \(\{u_i\}_{i=1}^n \stackrel{i.i.d.}{\sim} \text{Uniform}(0,1)\) and evaluate \(\{\tau - 1 \{u_i \leq \tau\}\}_{i=1}^n\) in place of \(\{\tau - 1 \{y_i \leq Q(\tau|x_i)\}\}_{i=1}^n\) for all four terms of the form \(A^\pm_{n,\tau}\) and \(B^\pm_{n,\tau}\). Second, repeat step one for a large number of iterations.

With this procedure, we can perform the tests of significance and heterogeneity as in Section 3.2 via simulating the supremum of \(G^{BR}_*(\tau)\). Define

$$WS^{BR}_n(T) = \sqrt{nh^3_{n,\tau}} \sup_{\tau \in T} |\hat{Q}RKD^{BR}_*(\tau)| \quad \text{and} \quad WH^{BR}_n(T) = \sqrt{nh^3_{n,\tau}} \sup_{\tau \in T} \left|\hat{Q}RKD^{BR}_*(\tau) - \int_T \hat{Q}RKD^{BR}_*(\tau')d\tau'\right|$$

Then, with the same null hypotheses, the following can be shown by similar steps to the proof of Corollary 2:

**Corollary 3.** Under Assumptions 1, 6 and 9, we have

(i) \(WS^{BR}_n(T) - \sup_{\tau \in T} |G^{BR}_*(\tau)| = o_p(1)\) under the null hypothesis \(H^S_0\); and

(ii) \(WH^{BR}_n(T) - \sup_{\tau \in T} |\phi_{QKD}(G^{BR}_*(\tau))| = o_p(1)\) under the null hypothesis \(H^H_0\), where \(\phi_{QKD}(g)(\tau) = g(\tau) - \int_T g(\tau')d\tau'\) for all \(g \in L^\infty_m(T)\), the space of all bounded, measurable, real-valued functions defined on \(T\).

Finally, we discuss bandwidth choices. We derive the first-order optimal bandwidths in terms of mean square errors (MSE) in finite sample.
Corollary 4. Under Assumption 6, the approximate optimal choice of $h_{n,\tau}$ is
\[ h^*_n,\tau = \left( \frac{6}{\nu_2(N^+)^{-1}D^+} \frac{\tau(1 - \tau)\nu_2'(N^+)^{-1}T^+(N^+)^{-1}T_2}{f_X(x_0)(f_{Y|X}(Q(\tau|x_0^+)|x_0^+))^2} \right)^{\frac{1}{2}} n^{-\frac{1}{5}}, \]
where $D^+ = \int_{0}^{\infty} u^2 \frac{\partial^2 Q(\tau|x_0^+)}{\partial u^2} (1, u)'K(u)du$, $T^+ = \int_{0}^{\infty} (1, u)'(1, u)K(u)^2 du$.

Corollary 5. Under Assumptions 6 and 9, the approximate optimal choice of $h^*_{n,\tau}$ is
\[ h^*_n,\tau = \left( \frac{90}{\nu_3(N^+)^{-1}\bar{D}^+} \frac{\tau(1 - \tau)\nu_3'(\bar{N}^+)^{-1}\bar{T}^+(\bar{N}^+)^{-1}T_3}{f_X(x_0)(f_{Y|X}(Q(\tau|x_0^+)|x_0^+))^2} \right)^{\frac{1}{2}} n^{-\frac{1}{7}}, \]
where $\bar{N} = \int_{0}^{\infty} [1, u, u^2]'[1, u, u^2]K(u)du$, $\bar{T}^+ = \int_{0}^{\infty} [1, u, u^2]'[1, u, u^2]K(u)^2 du$, and $\bar{D}^+ = \int_{0}^{\infty} u^3 \frac{\partial^2 Q(\tau|x_0^+)}{\partial u^3} [1, u, u^2]'K(u)du$.

Proofs are provided in Sections B.6 and B.7. Note that these two corollaries prescribing the approximate MSE-optimal bandwidth choices involve unknown densities, $f_X$ and $f_{Y|X}$, as well as the unknown conditional quantile function $Q$. We suggest to plug-in preliminary estimates, $\hat{f}_X$, $\hat{f}_{Y|X}$ and $\hat{Q}$, where bandwidth choices for these preliminary estimates in turn can be conducted by existing rule-of-thumb or data-driven methods. See Section C for a guide to practice in a bandwidth choice procedure.

It is worthy of noting that Imbens and Kalyanaraman (2012) and Calonico, Cattaneo and Titiunik (2014) provide fully data-driven optimal bandwidth selection algorithms that are tailored to (mean) regression discontinuity design. By considering the minimization of the asymptotic MSE of the jump size, they are more suitable for this setting since regression discontinuity design focuses on level change at a given threshold instead of concerning the entire support. The latter paper further provides a new variance estimator that accounts for the additional variability brought in by the bias correction procedure, and is more well-behaved in finite sample. We conjecture that it is possible to derive corresponding results for our QRKD estimator following similar footsteps. Derivation of such quantile extensions would be long and out of the scope of this paper, and so we will leave it for future research.
B.6 Proof of Corollary 4

Proof. By Theorem 2, we have

\[ \text{MSE}(h) = \frac{h^2}{4} \left( \left( \frac{1}{2} \right) (N^+)^{-1} D^+ \right)^2 + \frac{(1 - \tau)\nu^2(N^+)^{-1} T^+(N^+)^{-1} \nu^2}{nh^3 f_X(x_0)(f_Y|x_x(Q(\tau|x_0^+)|x_0^+))^2} + o_p\left( \frac{1}{nh^3} \right). \]

By the first order condition with the two leading terms, we obtain the desired result. \qed

B.7 Proof of Corollary 5

Proof. Following similar steps of the proof of Theorem 2, we have

\[ \text{MSE}(\bar{h}) = \frac{\bar{h}^2}{36} \left( \left( \frac{1}{2} \right) (\bar{N}^+)^{-1} D^+ \right)^2 + \frac{(1 - \tau)\nu^2(\bar{N}^+)^{-1} T^+(\bar{N}^+)^{-1} \nu^2}{nh^3 f_X(x_0)(f_Y|x_x(Q(\tau|x_0^+)|x_0^+))^2} + o_p\left( \frac{1}{nh^3} \right). \]

By the first order condition with the two leading terms, we obtain the desired result. \qed

B.8 Proof of Theorem 4

Proof. From the proof of theorem 4.2 of Pagan and Ullah (1999),

\[ \sqrt{nh^3_n}(\hat{\beta}(x_0^+) - E[\hat{\beta}(X)|X = x_0^+]) = \sum_{i=1}^{n} \frac{K'_{n,i}u_i d_i^+}{\sqrt{nh_n f_X(x_0)}} + o_p(1) \]

where \( K'_{n,i} = \frac{\partial}{\partial x} K(u) \bigg|_{v=x_i-x_0} \). We denote \( T_{n,i}^+ = \frac{K'_{n,i}u_i d_i^+}{\sqrt{nh_n f_X(x_0)}} \). We define additional shorthand notations for this proof: let \( W_{n,i}^+(\tau_1, ... \tau_k) = (T_{n,i}^+, Z_{n,i}^+(\tau_1), ..., Z_{n,i}^+(\tau_k)) \), where \( Z_{n,i}^+(\tau) \) is defined as in the proof of Theorem 2. Define

\[ H_n^+ = \left( \sqrt{nh^3_n}(\hat{\beta}(x_0^+) - E[\hat{\beta}(X)|X = x_0^+]), \left\{ \sqrt{nh^3_n}(\hat{\beta}(\tau) - \beta_n^+(\tau)) : \tau \in T \right\} \right) = (b_n^+, A_n^+) \]

where \( \beta_n^+(\tau) = \frac{\partial Q(\tau|x_0^+)}{\partial x} + h_n, \frac{\nu^2(N^+)^{-1}}{2} \int_0^\infty u^2 \frac{\partial^2 Q(\tau|x_0^+)}{\partial x^2}(1, u)'K(u) du \). We also write \( A_n^+(\tau) \) for \( \sqrt{nh^3_n}(\hat{\beta}(\tau) - \beta_n^+(\tau)) \).

First we show the finite dimensional convergence of the process. The covariance of any combination of coordinates that does not involve \( b_n^+ \) is the same as the one in Theorem 2. Under Assumption 8...
(iv), (v), the covariance of a coordinate of $A^+_n$ with $\tau \in T$ and $b^+_n$ is

$$
\sum_{i=1}^{n} Cov\left( \frac{t_2'(N^+)^{-1} \sum_{i=1}^{n} (\tau - \mathbb{1}\{y_i \leq Q(\tau|x_i)\}) \tau_i \cdot \mathbb{1}\{y_i \leq Q(\tau|x_i)\} z_{i,n,\tau} d_{i}^{+}, \frac{K'_{i,n,u_i}}{\sqrt{nh_n f_X(x_0)}} \right) = \sigma_{r,b}^+ \\
\rightarrow \sigma_{r,b}^+ = \int_{0}^{1} \int_{R} \int_{(0,\infty)} (\tau - \mathbb{1}\{y \leq Q(\tau|x_0)\}) (1, \frac{\nu}{c(\tau)})^T \times K(\frac{\nu}{c(\tau)}) K'(\nu)(b - E[b(X, \mathcal{E}_1)|X = x_0]) f_{Y|X}(y, x_0, b) dv dy db
$$

as $n \to \infty$ by the dominated convergence theorem. This is finite under Assumption 8 (ii). Finally, as in Theorem 4.2 of Pagan and Ullah (1999), the asymptotic variance of $b^+_n$ is

$$
\sigma^2(x_0) = \int \frac{f_X(x_0)}{f_X(x_0)} K'(\nu)^2 dv < \infty.
$$

Thus the covariance matrix is finite for any given finite dimensions of $A^+_n$.

We now show that the moment condition of Lindeberg-Feller is satisfied. For any finite set \{\tau_1, ..., \tau_k\} \subset T of quantiles. Under Assumptions 6 (i) (a), (i) (b), (iv), and (v), and Assumption 8 (iii), we have

$$
\sum_{i=1}^{n} E\left\|W^+_n(\tau_1, ..., \tau_k)\right\|^2 I\left(\left\|W^+_n(\tau_1, ..., \tau_k)\right\| > \epsilon\right) = \sum_{i=1}^{n} E\left[\sum_{j=1}^{k} Z^+_n(\tau_j)^2 + (T^+_n)^2\right] I\left\{\sum_{j=1}^{k} Z^+_n(\tau_j)^2 + (T^+_n)^2 > \epsilon^2\right\} \leq \int (m_1 K(v) + m_2 K'(v) u^2) I\{u^2 > nh_n \epsilon^2 + m_3\} dF_{U|X}(u, x_0 + vh_n)
$$

For some constants, $m_1$, $m_2$ and $m_3$, for any $\epsilon > 0$ given a fixed $n$. Applying Fubini’s theorem under Assumptions 6 (iv) and Assumption 8 (iii), the last line above becomes

$$
\int_{(0,\infty)} \int_{R} (m_1 K(v) + m_2 K'(v) u^2) I\{u^2 > nh_n \epsilon^2 + m_3\} dF_{U|X}(u|X = x_0 + vh_n) f_{X}(x_0 + vh_n) dv
$$

We denote the first and second terms of the above expression by (1) and (2), respectively. Hölder’s inequality implies for any $1 < p, q < \infty$ such that $\frac{1}{p} + \frac{1}{q} = 1$,

$$
(2) \leq \int_{(0,\infty)} m_2 K(v) E[U^{2p}|X = x_0 + vh_n]^{1/p} P(U^2 > nh_n \epsilon^2 + m_3|X = x_0 + vh_n)^{1/q} f_{X}(x_0 + vh_n) dv \\
\leq \text{ess sup}_{z} E[U^{2p}|Z]^{1/p} \left( \int_{(0,\infty)} m_1 K(v) P(U^2 > nh_n \epsilon^2 + m_3|X = x_0 + vh_n)^{1/q} f_{X}(x_0) dv + O(h_n) \right)
$$

19
Assumption 8 (ii) implies the essential supremum term is finite. Since \( P(U^2 > nh_n \epsilon^2 + m_3|X = x_0 + \nu h_n) \to 0 \) as \( n \to \infty \), applying the dominated convergence theorem gives us that (2) \( \to 0 \) as \( n \to \infty \). It can be shown that (1) \( \to 0 \) following a similar reasoning. This shows that the moment condition of Lindeberg-Feller is satisfied. Together with the covariance condition, we have established the finite dimensional convergence of the process \( H_n^+ \).

The tightness of all but \( m'_{x_0^+}(x) \) dimensions is shown by the proof of Theorem 2 and \( m'_{x_0^+}(x) \) is trivially tight since it’s one dimensional. By Lemma 1.4.3 of van der Vaart and Wellner (1996), \( H_n^+ \), the product of them, is tight. Applying theorem 18.14 of van der Vaart (1998), we now have \( H_n^+ \Rightarrow G_f^+ \), a Gaussian process with zero mean and covariance function as specified above. Using a similar argument, we also have \( H_n^- \Rightarrow G_f^- \).

Since \( H_n^+ \) and \( H_n^- \) are based on an i.i.d. sample from two different sides of the kink, the continuous mapping theorem implies
\[
\sqrt{nh_n} \left( \begin{bmatrix} \hat{m}'(x_0^+) - \hat{m}'(x_0^-) \\ \hat{\beta}^+(\tau) - \hat{\beta}^-(\tau) \end{bmatrix} - \begin{bmatrix} m'(x_0^+) - m'(x_0^-) \\ \beta^+(\tau) - \beta^-(\tau) \end{bmatrix} \right)_{\tau \in T} \Rightarrow G_\Delta = G_f^+ - G_f^-
\]

Finally, to derive the asymptotic distribution of \( \hat{QRKD} f(\tau) \), we apply the uniform version of functional delta method – see theorem 3.9.5 of van der Vaart and Wellner (1996). This version required here because \( \beta^+(\tau) - \beta^- (\tau) \) depends on \( n \). Note \( \beta^+(\tau) = \lim_n \beta^+_{n}(\tau) \) and \( \beta^-(\tau) = \lim_n \beta^-_{n}(\tau) \), the existence is implied by (iii) and (v) of Assumption 6. Define \( \Phi : L_m^\infty(T) \times [a, \infty) \to L_m^\infty(T), a > 0 \), by \( \Phi(A(\tau), b) = \frac{A(\tau)}{b} \). We show that \( \Phi \) Hadamard differentiable at \( (A,b) \) tangentially to \( L_m^\infty(T) \times (a, \infty) \).

Since for any \( (g_n, c_n) \in L_m^\infty(T) \times [a, \infty) \) such that \( g_n \to g \) and \( c_n \to c \) and any \( t_n \to 0 \),
\[
\frac{\Phi(A + t_n g_n, b + t_n c_n) - \Phi(A, b)}{t_n} \to \Phi'_{(A,b)}(g, c) = \frac{bg - cA}{b^2}.
\]

The linearity of \( \Phi'_{(A,b)}(g, c) \) is obvious, and its continuity is implied by its boundedness as \( \| \Phi'_{(A,b)}(g, c) \| \leq M \max\{|g|_\infty, |c|\} = M \| (g, c) \| \). Since such Hadamard derivative exists at every \( (A,b) \in L_m^\infty(T) \times (a, \infty) \), \( \Phi \) is uniformly differentiable.
By Theorem 3.9.5 of van der Vaart and Wellner (1996), we have the weak convergence result for the fuzzy QRKD estimator

\[
\sqrt{nh_n^3}(Q_{\tilde{R}KD}(\tau) - Q_{KD}(\tau)) \Rightarrow \frac{(m'(x_0^+ - m'(x_0^-))G_\Delta(\tau) - (\beta^+(\tau) - \beta^-(\tau))G_\Delta(b))}{(m'(x_0^+) - m'(x_0^-))^2},
\]

as desired.

C Practical Recipe on Bandwidth Choice

This section provides a guide to practice in bandwidth choices. Corollaries 4 and 5 prescribe the approximate MSE-optimal bandwidth choices. The two corollaries suggest

\[
h_{n,\tau}^* = \left( \frac{6}{(\tau_2(N^+)^{-1} + 1)^2} \right)^{1/5} n^{-1/5},
\]

where \( D^+ = \int_0^\infty u^2 \frac{\partial Q(r|x_0^+)}{\partial x^+} (1, u)^T K(u) du \), and

\[
\bar{h}_{n,\tau}^* = \left( \frac{90}{(\tau_3(N^+)^{-1} + 1)^2} \right)^{1/5} n^{-1/5},
\]

where \( \bar{N} = \int_0^\infty [1, u, u^2]'[1, u, u^2] K(u) du \), \( \bar{T} = \int_0^\infty [1, u, u^2]'[1, u, u^2] K(u) du \), and \( \bar{D}^+ = \int_0^\infty u^3 \frac{\partial^3 Q(r|x_0^+)}{\partial x^+} (1, u)^T K(u) du \). Qu and Yoon (2015a) use a very large value for \( \bar{h}_{n,\tau}^* \), which is effectively assuming to have \( \frac{\partial^3 Q}{\partial x^+} = 0 \). Once we compute \( \hat{\lambda}^+(\tau) \) based on the choice of \( \bar{h}_{n,\tau}^* \), we can in turn substitute \( 2\hat{\lambda}^+(\tau) \) for \( \frac{\partial^2 Q(r|x_0^+)}{\partial x^+} \) in the definition of \( D^+ \) in order to choose \( h_{n,\tau}^* \).

In the above formulas, the unknown densities, \( f_X \) and \( f_{Y|X} \), and the unknown conditional quantile function \( Q \) need to be replaced by the respective non-parametric estimates \( \hat{f}_X \), \( \hat{f}_{Y|X} \) and \( \hat{Q} \). Bandwidth choices for the preliminary estimates, \( \hat{f}_X \), \( \hat{f}_{Y|X} \) and \( \hat{Q} \), in turn can be conducted by existing rule-of-thumb or data-driven methods. Because we are only using the observations to right of the kink point, we confine ourselves to the observations \( \{(y_j, x_j)\}_{j \in I^+} \) with \( I^+ = \{j \in \{1, 2, ..., n\} : x_j \geq x_0\} \). Write \( n^+ = |I^+| \), the number of observations to the right of the kink. Write the bandwidths used for estimating \( \hat{f}_X \), \( \hat{f}_{Y|X} \) and \( \hat{Q} \) as \( h_{n,\tau}^x \), \( h_{n,\tau}^y \), \( h_{n,\tau}^q \)' and \( h_{n,+,\tau}^x \) respectively.

First, for the standard kernel density estimator, \( h_n^x \) may be obtained by minimizing approximate mean integrated square errors: \( h_n^x = \left( \int u^2 K(u) du \right)^{-2/5} \left( \int K(u)^2 du \right)^{1/5} \left( \frac{3}{8v_\Omega} \sigma_X \right)^{-1/5} n^{-1/5} \), where \( \sigma_X \)
can be estimated by sample variance of $X$. See session 3.3 and 3.4 of Silverman’s (1986). Second, for the standard kernel conditional density estimator, Bashtannyk and Hyndman (2001) suggest that, based on normal approximation of the marginal of $X$ and heteroskedasticity of $Y|X$, $(\tilde{h}_{n+}^{y}, \tilde{h}_{n+}^{x})'$ may be obtained by

\[
(\tilde{h}_{n+}^{y}, \tilde{h}_{n+}^{x})' = \left( \left( \frac{d^{2}v}{2.85 \sqrt{2\pi} \sigma_{X+}^{2}} \right)^{1/4} \tilde{h}_{n+}^{x}, \left( \frac{32R^{2}(K)\sigma_{Y+}^{5}(260\pi^{9}\sigma_{X+}^{58})^{1/8}}{n^{+}\sigma_{K}^{4}d^{25/2}v^{3/4}[v^{1/2} + d(16.25\pi\sigma_{X+}^{10})^{1/4}]} \right)^{1/6} \right)'
\]

where $R(K) = \int K^{2}(u)du$, $v = 0.95\sqrt{2\pi} \sigma_{X+}^{3}(3d^{2}\sigma_{X+}^{2} + 8\sigma_{Y+}^{2}) - 32\sigma_{X+}^{2}\sigma_{Y+}^{2}e^{-2}$, and $d$ is the slope of an OLS of $y_{i}$ on $[1, x_{i}]$ computed with observations $i \in I^{+}$. $\sigma_{X+}^{2}$ and $\sigma_{Y+}^{2}$ can be computed by sample variances of $x_{i}$ and $y_{i}$ with $i \in I^{+}$. $\sigma_{K}^{2}$ is the variance of the kernel $K$. Third, for the local linear conditional function estimator, the bandwidth $h_{n+}^{q, \tau}$ for $Q$ can be set by the Yu and Jones’ (1998) rule of thumb based on the normality assumption of $f_{Y|X}$:

\[
h_{n+}^{q, \tau} = \left[ 2\pi^{-1}\tau(1 - \tau)\phi(\Phi^{-1}(\tau))^{-2} \right]^{1/5} h_{n+}^{q, \tau, 1/2},
\]

where $\phi$ and $\Phi$ denote the PDF and CDF for the standard normal distribution, respectively, and $h_{n+}^{q, \tau, 1/2}$ can be set to be equal to $\tilde{h}_{n+}^{x, \text{mean}} = \left( \frac{\int K(u)^{2}du\sigma^{2}(x)}{n^{+}(\int u^{2}K(u)^{2}du)^{2}m''(x)} \right)^{1/5}$. The functions, $m(x)$ and $\sigma^{2}(x)$, denote the conditional mean and the conditional variance of $Y$ given $X$. The second-derivative $m''(x)$ can be estimated by the coefficient of the square term of the OLS $y_{i}$ on $[1, x_{i}, x_{i}^{2}]$ with $i \in I^{+}$. The skedastic function $\sigma^{2}(x)$ can be estimated by the sample counterpart of $E[Y^{2}|X] - (E[Y|X])^{2}$ that can be computed by using the OLS of $y_{i}^{2}$ on $[1, x_{i}]$ and $y_{i}$ on $[1, x_{i}]$ with $i \in I^{+}$.

References


