Econometric Theory, **36**, 2020, 1167–1191.
doi:10.1017/S0266466619000409

QUANTILE TREATMENT EFFECTS IN REGRESSION KINK DESIGNS*

HENG CHEN
Bank of Canada

HAROLD D. CHIANG
Vanderbilt University

YUYA SASAKI
Vanderbilt University

The literature on regression kink designs develops identification results for average effects of continuous treatments (Nielsen et al., 2010, *American Economic Journal: Economic Policy* 2, 185–215; Card et al., 2015, *Econometrica* 83, 2453–2483), average effects of binary treatments (Dong, 2018, Jump or Kink? Identifying Education Effects by Regression Discontinuity Design without the Discontinuity), and quantile-wise effects of continuous treatments (Chiang and Sasaki, 2019, *Journal of Econometrics* 210, 405–433), but there has been no identification result for quantile-wise effects of binary treatments to date. In this article, we fill this void in the literature by providing an identification of quantile treatment effects in regression kink designs with binary treatment variables. For completeness, we also develop large sample theories for statistical inference, present a practical guideline on estimation and inference, conduct simulation studies, and provide an empirical illustration.

1. INTRODUCTION

Theories of identification in regression kink designs (RKDs) are advanced by a few papers in the recent literature. Nielsen et al. (2010) and Card et al. (2015) propose identification of average effects of continuous treatments. Dong (2018) proposes identification of average effects of binary treatments. Chiang and Sasaki (2019) propose identification of quantile-wise effects of continuous treatments. To date, no theory has been proposed for identification of quantile-wise effects of binary treatments in RKDs. This article aims to fill this void in the literature.

Specifically, in RKDs with binary treatments, we show that a local Wald ratio of derivatives of certain conditional expectation functions can be used to identify the conditional distribution functions of the potential outcomes given the event of local compliance. These conditional distribution functions can be used in turn to

*First arXiv date: March 15, 2017. We thank Peter Phillips (Editor), Arthur Lewbel (Co-Editor), three anonymous referees, Yingying Dong, Robert Moffitt, and participants at New York Camp Econometrics XIII for very useful comments and suggestions. All the remaining errors are ours. Address correspondence to Yuya Sasaki, Department of Economics, Vanderbilt University, VU Station B #351819, 2301 Vanderbilt Place, Nashville, TN 37235-1819, USA; e-mail: yuya.sasaki@vanderbilt.edu.

© Cambridge University Press 2020

https://doi.org/10.1017/S0266466619000409 Published online by Cambridge University Press

identify the quantile treatment effects given the event of local compliance. Our identification argument parallels that of Frandsen et al. (2012), who show that a local Wald ratio of certain conditional expectation functions can be used to identify the conditional distribution functions of potential outcomes given the event of local compliance in the context of regression discontinuity designs (RDDs). Because of the lack of discontinuity in our context of RKDs, however, our identification result entails the limit case of the event of local compliance, which amounts to the subpopulation to which the marginal treatment effects (Björklund and Moffitt, 1987; Heckman and Vytlacil, 1999, 2005) are relevant. This is analogous to, and provides a quantile counterpart of the identification result by Dong (2018).

Our identifying formula takes the form of local Wald ratios of *derivatives* of functions. Such a form is related to the identifying formulas of several papers in the existing literature. These papers include the following: Dong and Lewbel (2015)—also see Cerulli et al. (2017)—who use a local Wald ratio of derivatives of conditional expectation functions to identify the average effect of changing the threshold location in RDDs; Nielsen et al. (2010) and Card et al. (2015) who use a local Wald ratio of derivatives of conditional expectation functions to identify average effects of continuous treatments in RKDs; Dong (2018) who uses a local Wald ratio of derivatives of conditional expectation functions to identify average effects of binary treatments in RKDs; and Chiang and Sasaki (2019) who use a local Wald ratio of derivatives of conditional quantile functions to identify quantile-wise effects of continuous treatments in RKDs. In contrast to these papers, we use the difference of left-inverses of two local Wald ratios of derivatives of conditional expectation functions to identify quantile-wise effects of binary treatments in RKDs.

Although we motivate this article by quantile treatment effects, the identifying formulas we provide as the main result of this article can be also used to identity the distributional treatment effects. Therefore, this article also relates to Abadie (2002) who uses a form of Wald ratios to identify distributional treatment effects, and more closely relates to Shen and Zhang (2016) who consider distributional treatment effects in the context of RDDs.

In addition to the main identification result, we also provide methods of estimation and inference for quantile treatment effects based on analog estimators of our identifying formulas. Although our identification result is novel, estimation and inference results follow from an adaptation of existing approaches to our framework. Therefore, the main text focuses on the identification theory. Details of estimation and inference theories are found in the appendixes.

The rest of this article is organized as follows. In Section 2, we develop the main identification result. Section 3 presents a practical guideline on estimation and inference. Section 4 presents simulation studies. Section 5 illustrates a real data analysis. Appendix A presents formal theories for the method of inference. Appendix B presents additional practical considerations. Additional mathematical details are found in the Supplementary Material section of this article.

https://doi.org/10.1017/S0266466619000409 Published online by Cambridge University Press

2. IDENTIFICATION: THE MAIN RESULT

We model the random vector $(Y, D, X, U, V) : (\Omega^x, \mathscr{F}^x, \mathbb{P}^x) \to \mathscr{Y} \times \mathscr{D} \times \mathscr{X} \times \mathscr{U} \times \mathscr{V}$ through the following causal structure, where $\mathscr{Y} \subset \mathbb{R}$, $\mathscr{D} = \{0, 1\}$, $\mathscr{X} \subset \mathbb{R}$, $\mathscr{U} \subset \mathbb{R}^{d_U}$ for $d_U \in \mathbb{N}$, and $\mathscr{V} \subset \mathbb{R}$.

$$Y = g(D, X, U) \tag{2.1}$$

$$D = \mathbb{1}\{h(X) \geq V\} \tag{2.2}$$

In equation (2.1), the outcome variable Y is produced through function g by a binary treatment variable D, a continuous running variable or assignment variable X, and miscellaneous factors U. We let $Y^d = g(d, X, U)$ denote the potential outcome random variable that an individual with attributes (X, U) would produce under each hypothetical treatment choice $d \in \{0, 1\}$. The actual treatment choice D is determined by X and V through the threshold-crossing model (2.2). A researcher observes the joint distribution of Y, D, and X. However, a researcher cannot observe U or V. We do *not* impose any statistical independence condition in this model. Therefore, existing methods for instrumental variable quantile regression (e.g., Chernozhukov and Hansen, 2005) will not apply here. In particular, we do *not* assume statistical independence between the running variable X and the unobservables (U, V). Instead, we make the following assumption of the RKD.

Assumption 1 (RKD). *Let $x_0 = 0 \in \mathscr{X}$ be a designed kink location.*

(i) *h is continuously differentiable in a deleted neighborhood $I_X \backslash \{0\} \subset \mathscr{X}$ of $x_0 = 0$.*
(ii) *h is continuous at $x_0 = 0$.*
(iii) *$\lim_{x \downarrow 0} h'(x) \neq \lim_{x \uparrow 0} h'(x)$, where h' denotes dh/dx.*
(iv) *The conditional distribution of V given X is absolutely continuous with a continuously differentiable conditional density function $f_{V|X}(\cdot|\cdot)$.*
(v) *The conditional cumulative distribution function $F_{Y^d|VX}(y| \cdot, \cdot)$ is continuously differentiable for each $y \in \mathscr{Y}$ for each $d \in \{0, 1\}$.*
(vi) *$f_{V|X}(h(0)|0) > 0$.*

The research design as required by Assumption 1 consists of three broad components. First, the treatment assignment rule h has a kink at the designed location $x_0 = 0$, as formally stated in parts (ii) and (iii), but this assignment rule h is reasonably smooth elsewhere, as formally stated in part (i). Second, every other function is reasonably smooth, as formally stated in parts (iv) and (v). Third, there is sufficient data at the designed kink location $x_0 = 0$, as formally stated in part (vi). This assumption is analogous to that of Dong (2018) who analyzes average effects of binary treatments in the RKD. Under this design, we obtain the following identification result for conditional distributions of the potential outcomes Y^d given the event of $(V, X) = (h(0), 0)$.

THEOREM 1 (Identification). *Let Assumption 1 hold for the model (2.1) and (2.2). Then,*

https://doi.org/10.1017/S0266466619000409 Published online by Cambridge University Press

$$F_{Y^1|VX}(y|h(0),0)$$
$$=\frac{\lim_{x\downarrow 0}\frac{d}{dx}E[\mathbb{1}\{Y\le y\}\cdot D|X=x]-\lim_{x\uparrow 0}\frac{d}{dx}E[\mathbb{1}\{Y\le y\}\cdot D|X=x]}{\lim_{x\downarrow 0}\frac{d}{dx}E[D|X=x]-\lim_{x\uparrow 0}\frac{d}{dx}E[D|X=x]} \quad \textit{and}$$
$$F_{Y^0|VX}(y|h(0),0)$$
$$=\frac{\lim_{x\downarrow 0}\frac{d}{dx}E[\mathbb{1}\{Y\le y\}\cdot (1-D)|X=x]-\lim_{x\uparrow 0}\frac{d}{dx}E[\mathbb{1}\{Y\le y\}\cdot (1-D)|X=x]}{\lim_{x\downarrow 0}\frac{d}{dx}E[1-D|X=x]-\lim_{x\uparrow 0}\frac{d}{dx}E[1-D|X=x]}$$

hold for all $y\in\mathscr{Y}$.

Once the conditional cumulative distribution functions, $F_{Y^d|VX}(\cdot|h(0),0)$ for $d\in\{0,1\}$, are identified through the formulas presented in Theorem 1, the conditional quantile treatment effect is in turn identified by

$$\tau(\theta)=\inf\{y\in\mathscr{Y}: F_{Y^1|VX}(y|h(0),0)\ge\theta\}-\inf\{y\in\mathscr{Y}: F_{Y^0|VX}(y|h(0),0)\ge\theta\} \tag{2.3}$$

for $\theta\in(0,1)$. Theorem 1 also provides the identification of the distributional treatment effects of the form $F_{Y^1|VX}(\cdot|h(0),0)-F_{Y^0|VX}(\cdot|h(0),0)$, as in Abadie (2002) and Shen and Zhang (2016), which are useful to test important hypotheses such as first-order stochastic dominance.[1]

Proof of Theorem 1: By applying Leibniz rule under Assumption 1(i) and (iv), we have

$$\frac{d}{dx}\mathrm{E}[D|X=x]=\frac{d}{dx}\int_{-\infty}^{h(x)}f_{V|X}(v|x)dv=h'(x)\cdot f_{V|X}(h(x)|x)+\int_{-\infty}^{h(x)}\frac{\partial}{\partial x}f_{V|X}(v|x)dv$$

for all $x\in I_X\backslash\{0\}$. Similarly, by applying Leibniz rule again under Assumption 1(i), (iv), and (v), we have

$$\begin{aligned}\frac{d}{dx}\mathrm{E}[\mathbb{1}\{Y\le y\}\cdot D|X=x]&=\frac{d}{dx}\int_{-\infty}^{h(x)}\int_{u:g(1,x,u)\le y}F_{UV|X}(du,dv|x)\\&=\frac{d}{dx}\int_{-\infty}^{h(x)}f_{V|X}(v|x)\int_{u:g(1,x,u)\le y}F_{U|VX}(du|v,x)dv\\&=\frac{d}{dx}\int_{-\infty}^{h(x)}f_{V|X}(v|x)\cdot F_{Y^1|VX}(y|v,x)dv\\&=h'(x)\cdot f_{V|X}(h(x)|x)\cdot F_{Y^1|VX}(y|h(x),x)\\&\quad+\int_{-\infty}^{h(x)}\frac{d}{dx}\left[f_{V|X}(v|x)\cdot F_{Y^1|VX}(y|v,x)dv\right]dv\end{aligned}$$

[1] We remark that, with our identifying formulas provided in Theorem 1, $F_{Y^1|VX}(\cdot|h(0),0)-F_{Y^0|VX}(\cdot|h(0),0)$ can be simply expressed as a single Wald ratio: $\frac{\lim_{x\downarrow 0}\frac{d}{dx}\mathrm{E}[\mathbb{1}\{Y\le y\}|X=x]-\lim_{x\uparrow 0}\frac{d}{dx}\mathrm{E}[\mathbb{1}\{Y\le y\}|X=x]}{\lim_{x\downarrow 0}\frac{d}{dx}\mathrm{E}[D|X=x]-\lim_{x\uparrow 0}\frac{d}{dx}\mathrm{E}[D|X=x]}$.

https://doi.org/10.1017/S0266466619000409 Published online by Cambridge University Press

for all $(x, y) \in (I_X \backslash \{0\}) \times \mathscr{Y}$. Therefore, by Assumption 1(ii) and (iv), we can write

$$\lim_{x \downarrow 0} \frac{d}{dx} \mathrm{E}[D|X = x] - \lim_{x \uparrow 0} \frac{d}{dx} \mathrm{E}[D|X = x] = \left[h'(0^+) - h'(0^-)\right] \cdot f_{V|X}(h(0)|0),$$

and, by Assumption 1(ii), (iv), and (v), we can write

$$\begin{aligned} &\lim_{x \downarrow 0} \frac{d}{dx} \mathrm{E}[\mathbb{1}\{Y \leq y\} \cdot D|X = x] - \lim_{x \uparrow 0} \frac{d}{dx} \mathrm{E}[\mathbb{1}\{Y \leq y\} \cdot D|X = x] \\ &= \left[h'(0^+) - h'(0^-)\right] \cdot f_{V|X}(h(0)|0) \cdot F_{Y^1|VX}(y|h(0), 0) \end{aligned}$$

for all $y \in \mathscr{Y}$. Taking the ratio of these expressions under Assumption 1(iii) and (vi) yields

$$\begin{aligned} &\frac{\lim_{x \downarrow 0} \frac{d}{dx} \mathrm{E}[\mathbb{1}\{Y \leq y\} \cdot D|X = x] - \lim_{x \uparrow 0} \frac{d}{dx} \mathrm{E}[\mathbb{1}\{Y \leq y\} \cdot D|X = x]}{\lim_{x \downarrow 0} \frac{d}{dx} \mathrm{E}[D|X = x] - \lim_{x \uparrow 0} \frac{d}{dx} \mathrm{E}[D|X = x]} \\ &= F_{Y^1|VX}(y|h(0), 0) \end{aligned}$$

for all $y \in \mathscr{Y}$. Similar lines of arguments yield

$$\begin{aligned} &\frac{\lim_{x \downarrow 0} \frac{d}{dx} \mathrm{E}[\mathbb{1}\{Y \leq y\} \cdot (1 - D)|X = x] - \lim_{x \uparrow 0} \frac{d}{dx} \mathrm{E}[\mathbb{1}\{Y \leq y\} \cdot (1 - D)|X = x]}{\lim_{x \downarrow 0} \frac{d}{dx} \mathrm{E}[1 - D|X = x] - \lim_{x \uparrow 0} \frac{d}{dx} \mathrm{E}[1 - D|X = x]} \\ &= F_{Y^0|VX}(y|h(0), 0) \end{aligned}$$

for all $y \in \mathscr{Y}$. □

Discussions of Theorem 1: In the context of RDD where $h(0^-) < h(0^+)$, Frandsen et al. (2012) show that similar local Wald ratios identify the conditional distribution of the potential outcomes given the event

$$C_{\mathrm{RDD}} = \{\omega \in \Omega : X(\omega) = 0, h(0^-) < V(\omega) \leq h(0^+)\}$$

of local compliance. In our context of the RKD, where $h(0^-) = h(0^+)$, Theorem 1 shows that local Wald ratios of the derivatives identify the conditional distributions of the potential outcomes given the event

$$C_{\mathrm{RKD}} = \{\omega \in \Omega : X(\omega) = 0, V(\omega) = h(0)\},$$

which may be considered as a limit of the event C_{RDD} for RDD as $\left|h(0^+) - h(0^-)\right|$ approaches 0. In this sense, our causal interpretation result is similar to that of marginal treatment effects (Björklund and Moffitt, 1987; Heckman and Vytlacil, 1999, 2005). This interpretation is analogous to the identification result by Dong (2018) who analyzes average effects of binary treatments in the RKD. △

Relation to the Case of Continuous Treatments: As discussed earlier concerning RKDs, this article focuses on binary treatments, while Chiang and Sasaki (2019) focus on continuous treatments in RKDs. With $Q_{Y|X}(\cdot|x)$ and b denoting the conditional quantile function of Y given $X = x$ and a policy function, respectively,

https://doi.org/10.1017/S0266466619000409 Published online by Cambridge University Press

Chiang and Sasaki (2019) show that the Wald ratio of the form

$$\frac{\lim_{x\downarrow 0}\frac{d}{dx}Q_{Y|X}(\theta|x)-\lim_{x\uparrow 0}\frac{d}{dx}Q_{Y|X}(\theta|x)}{\lim_{x\downarrow 0}\frac{d}{dx}b(x)-\lim_{x\uparrow 0}\frac{d}{dx}b(x)}$$

identifies a weighted average of structural causal effects for the subpopulation of individuals at a designed kink location $X=0$ and at the θth conditional quantile of Y given $X=0$, by applying the technique of Sasaki (2015)—also see Kato and Sasaki (2017). In contrast, this article shows that the differences of the left-side functional inverses of the maps

$$y\mapsto\frac{\lim_{x\downarrow 0}\frac{d}{dx}\mathrm{E}[\mathbb{1}\{Y\le y\}\cdot D|X=x]-\lim_{x\uparrow 0}\frac{d}{dx}\mathrm{E}[\mathbb{1}\{Y\le y\}\cdot D|X=x]}{\lim_{x\downarrow 0}\frac{d}{dx}\mathrm{E}[D|X=x]-\lim_{x\uparrow 0}\frac{d}{dx}\mathrm{E}[D|X=x]}\quad\text{and}$$

$$y\mapsto\frac{\lim_{x\downarrow 0}\frac{d}{dx}\mathrm{E}[\mathbb{1}\{Y\le y\}\cdot(1-D)|X=x]-\lim_{x\uparrow 0}\frac{d}{dx}\mathrm{E}[\mathbb{1}\{Y\le y\}\cdot(1-D)|X=x]}{\lim_{x\downarrow 0}\frac{d}{dx}\mathrm{E}[1-D|X=x]-\lim_{x\uparrow 0}\frac{d}{dx}\mathrm{E}[1-D|X=x]}$$

identify the conditional quantile treatment effects given the event of the aforementioned local compliance $V=h(0)$ and the designed kink location $X=0$. This article thus focuses on different sorts of treatment parameters as well as using different identification strategies. Namely, the treatment parameter of Chiang and Sasaki (2019) is a functional of the derivative of a structural function, while the treatment parameter in this article is the quantile treatment effect. Note that the derivative of a structural function makes sense for continuous treatments, but it would not make sense for binary treatments as in this article. The identification strategies also differ as the above formulas suggest. Although Chiang and Sasaki (2019) use a local Wald ratio of derivatives involving a conditional quantile function, this article in contrast uses the difference of the left inverses of local Wald ratios of derivatives not involving conditional quantile functions but involving only conditional expectation functions. As a byproduct of our identification strategy, we can also identify the conditional distributions of potential outcomes which Chiang and Sasaki (2019) do not. These contrasts arise from different innate characteristics of binary and continuous treatments, rather than an artifact of merely taking different identification strategies. Binary treatments under monotonicity (2.2) generally allow one to identify detailed distributional characteristics of potential outcomes focusing on the subpopulation of compliers—see for example Abadie (2003, Thm. 3.1(b) and (c)). Extending this framework to local Wald ratios of derivatives, we demonstrate that one can identify detailed distributional characteristics of potential outcomes on the subpopulation of "limit" compliers. △

3. ESTIMATION AND INFERENCE: A PRACTICAL GUIDELINE

Although the main contribution of this article lies in our new identification result presented in Section 2, we also develop a theory and method of estimation and inference for completeness. Since the estimation and inference strategies are standard, we relegate most of the details to the Appendix. In this section, we present

https://doi.org/10.1017/S0266466619000409 Published online by Cambridge University Press

a practical guideline on estimation and inference for the conditional quantile treatment effects $\tau(\theta)$. A formal theory is presented in Appendix A. We also present additional practical considerations in Appendix B. Auxiliary lemmas and proofs are found in the online supplementary material available at Cambridge Journals Online (journals.cambridge.org/ect).

3.1. Estimation

The local Wald ratios proposed in Theorem 1 as identifying formulas can be succinctly rewritten as

$$F_{Y^d|VX}(y|h(0),0) = \frac{\mu_1'(0^+,y,d) - \mu_1'(0^-,y,d)}{\mu_2'(0^+,d) - \mu_2'(0^-,d)}, \tag{3.1}$$

where $\mu_1'(x,y,d)$ and $\mu_2'(x,d)$ are the partial derivatives with respect to x of $\mu_1(x,y,d)$ and $\mu_2(x,d)$ defined by

$$\mu_1(x,y,d) = E[\mathbb{1}\{Y \le y\} \cdot \mathbb{1}\{D = d\}|X = x] \qquad \text{and}$$
$$\mu_2(x,d) = E[\mathbb{1}\{D = d\}|X = x],$$

respectively. We estimate the components of (3.1) by the one-sided local cubic estimators

$$\hat{\mu}_1'(0^\pm,y,d)h_n = e_1^\top \arg\min_{\alpha\in\mathbb{R}^4} \sum_{i=1}^n \left[\mathbb{1}\{Y_i \le y\}\mathbb{1}\{D_i = d\} - r_3^\top\left(\frac{X_i}{h_n}\right)\alpha\right]^2 K\left(\frac{X_i}{h_n}\right)\delta_i^\pm \text{ and} \tag{3.2}$$

$$\hat{\mu}_2'(0^\pm,d)h_n = e_1^\top \arg\min_{\alpha\in\mathbb{R}^4} \sum_{i=1}^n \left[\mathbb{1}\{D_i = d\} - r_3^\top\left(\frac{X_i}{h_n}\right)\alpha\right]^2 K\left(\frac{X_i}{h_n}\right)\delta_i^\pm, \tag{3.3}$$

where K is a kernel function, h_n is a bandwidth parameter, $e_1 = (0,1,0,0)^\top$, $r_3(u) = (1,u,u^2,u^3)^\top$, $\delta_i^+ = \mathbb{1}\{X_i \ge 0\}$, and $\delta_i^- = \mathbb{1}\{X_i < 0\}$. A plug-in estimator for (3.1) is given by

$$\widehat{F}_{Y^d|VX}(y|h(0),0) = \frac{\hat{\mu}_1'(0^+,y,d) - \hat{\mu}_1'(0^-,y,d)}{\hat{\mu}_2'(0^+,d) - \hat{\mu}_2'(0^-,d)}.$$

The motivation for our using the local cubic polynomial is to account for the manual bias correction from local quadratic estimators. By considering the asymptotic distribution for the higher-order local polynomial, we effectively account for bias estimation in the asymptotic distribution from the lower-order one, thus allowing for robustness in inference against large bandwidths—see Calonico et al. (2014, Rem. 7) and Rem. S.A.7 in their supplementary material.

We can in turn estimate the conditional quantile treatment effect $\tau(\theta)$ by

$$\hat{\tau}(\theta) = \inf\left\{y \in \mathscr{Y} : \widehat{F}_{Y^1|VX}(y|h(0),0) \ge \theta\right\} - \inf\left\{y \in \mathscr{Y} : \widehat{F}_{Y^0|VX}(y|h(0),0) \ge \theta\right\}$$
$$= \widehat{Q}_{Y^1|VX}(\theta) - \widehat{Q}_{Y^0|VX}(\theta).$$

https://doi.org/10.1017/S0266466619000409 Published online by Cambridge University Press

The local Wald estimator $\widehat{F}_{Y^d|VX}(\cdot|h(0),0)$ is not always monotone increasing in finite samples. For ease of implementing the CDF inversion, we monotonize the estimated CDFs by rearrangements following Chernozhukov et al. (2010). This does not affect the asymptotic properties of the estimators, while allowing for inversion of the CDF estimators. Frandsen et al. (2012) also use this technology in the context of the RDD.

3.2. Multiplier Bootstrap

Let $\Gamma^{\pm} = \int_{\mathbb{R}_+} r_3(u) r_3^{\top}(u) K(u) du$. Under the assumptions stated in Appendix A, we obtain the following Uniform Bahadur Representations (UBR) for the local slope estimators (3.2) and (3.3).[2]

$$\begin{aligned} \nu_n^{\pm}(y,d,1) &= \sqrt{nh_n^3}[\hat{\mu}_1'(0^{\pm},y,d) - \mu_1'(0^{\pm},y,d) + O_p\left(h_n^3\right)] \\ &= \frac{1}{\sqrt{nh_n} f_X(0)} \sum_{i=1}^{n} e_1^{\top} (\Gamma^{\pm})^{-1} r_3\Big(\frac{X_i}{h_n}\Big) \\ &\quad \times \Big[\mathbb{1}\{Y_i \le y\}\mathbb{1}\{D_i = d\} - \mu_1(X_i,y,d)\Big] K\Big(\frac{X_i}{h_n}\Big)\delta_i^{\pm} \end{aligned} \tag{3.4}$$

$$\begin{aligned} \nu_n^{\pm}(y,d,2) &= \sqrt{nh_n^3}[\hat{\mu}_2'(0^{\pm},d) - \mu_2'(0^{\pm},d) + O_p\left(h_n^3\right)] \\ &= \frac{1}{\sqrt{nh_n} f_X(0)} \sum_{i=1}^{n} e_1^{\top} (\Gamma^{\pm})^{-1} r_3\Big(\frac{X_i}{h_n}\Big) \\ &\quad \times \Big[\mathbb{1}\{D_i = d\} - \mu_2(X_i,d)\Big] K\Big(\frac{X_i}{h_n}\Big)\delta_i^{\pm} \end{aligned} \tag{3.5}$$

We note that $\nu_n^{\pm}(y,d,2)$ are trivial functions of y.

Covariance functions for the limit processes are often cumbersome to approximate in practice. Qu and Yoon (2018) propose a simulation method to approximate limit processes under sharp designs—also see Qu and Yoon (2015)—but this method is not applicable to fuzzy designs. Thus, we propose to use the multiplier bootstrap method to approximate the asymptotic distributions of these UBR. Draw a random sample $\xi_1, \ldots, \xi_n$ from the standard normal distribution independently from the data $\{Y_i, D_i, X_i\}_{i=1}^n$. Replacing the unknowns μ_1, μ_2, and $f_X(0)$ in the UBR by their uniformly consistent estimators $\tilde{\mu}_1, \tilde{\mu}_2$, and $\hat{f}_X(0)$, respectively, we define the following Estimated Multiplier Processes (EMPs):

[2] UBR for conditional quantile estimators are studied by Kong et al. (2010) and Guerre and Sabbah (2012) among others.

https://doi.org/10.1017/S0266466619000409 Published online by Cambridge University Press

$$\hat{\nu}^{\pm}_{\xi,n}(y,d,1) = \frac{1}{\sqrt{nh_n}\hat{f}_X(0)} \sum_{i=1}^{n} \xi_i e_1^\top (\Gamma^{\pm})^{-1} r_3\Big(\frac{X_i}{h_n}\Big) \times \Big[\mathbb{1}\{Y_i \le y\}\mathbb{1}\{D_i = d\} - \tilde{\mu}_1(X_i,y,d)\Big] K\Big(\frac{X_i}{h_n}\Big)\delta_i^{\pm} \tag{3.6}$$

$$\hat{\nu}^{\pm}_{\xi,n}(y,d,2) = \frac{1}{\sqrt{nh_n}\hat{f}_X(0)} \sum_{i=1}^{n} \xi_i e_1^\top (\Gamma^{\pm})^{-1} r_3\Big(\frac{X_i}{h_n}\Big) \times \Big[\mathbb{1}\{D_i = d\} - \tilde{\mu}_2(X_i,d)\Big] K\Big(\frac{X_i}{h_n}\Big)\delta_i^{\pm}, \tag{3.7}$$

where the estimated components, $\hat{f}_X(0)$, $\tilde{\mu}_1(X_i,y,d)$, and $\tilde{\mu}_2(X_i,d)$, of these EMPs can be obtained by following the additional practical guidelines presented in Appendix B.2.

Under the assumptions to be stated in Appendix A, we show that the EMP can be used to uniformly approximate the asymptotic distribution of the UBR. Consequently, by the functional delta method, the asymptotic distribution of

$$\sqrt{nh_n^3}[\hat{\tau}(\cdot) - \tau(\cdot)]$$

can be approximated uniformly on $\Theta = [a, 1-a]$ for $a \in (0, 1/2)$ by the estimated process

$$\widehat{\Xi}(\cdot) = -\left[\frac{\hat{Z}_{\xi,n}(\hat{Q}_{Y^1|VX}(\cdot),1)}{\hat{f}_{Y^1|VX}(\hat{Q}_{Y^1|VX}(\cdot)|h(0),0)} - \frac{\hat{Z}_{\xi,n}(\hat{Q}_{Y^0|VX}(\cdot),0)}{\hat{f}_{Y^0|VX}(\hat{Q}_{Y^0|VX}(\cdot)|h(0),0)}\right],$$

where

$$\hat{Z}_{\xi,n}(y,d) = \frac{\begin{array}{c}[\hat{\mu}_2'(0^+,d) - \hat{\mu}_2'(0^-,d)][\hat{\nu}^+_{\xi,n}(y,d,1) - \hat{\nu}^-_{\xi,n}(y,d,1)] \\ -[\hat{\mu}_1'(0^+,y,d) - \hat{\mu}_1'(0^-,y,d)][\hat{\nu}^+_{\xi,n}(y,d,2) - \hat{\nu}^-_{\xi,n}(y,d,2)]\end{array}}{[\hat{\mu}_2'(0^+,d) - \hat{\mu}_2'(0^-,d)]^2}.$$

Once we obtain these approximations to the asymptotic distributions, we may conduct various tests of quantile functions following Koenker and Xiao (2002) and Chernozhukov and Fernández-Val (2005). For example, for the test of treatment nullity

$$H_0 : \tau(\theta) = 0 \quad \forall \theta \in \Theta \tag{3.8}$$

against the alternative of treatment significance, we use the test statistic

$$T^{TS} = \sup_{\theta\in\Theta}\left|\sqrt{nh_n^3}\hat{\tau}(\theta)\right|,$$

where $\Theta = [a, 1-a]$ for some $a \in (0, 1/2)$. We can approximate the asymptotic distribution of T^{TS} by

$$\sup_{\theta\in\Theta}\left|\widehat{\Xi}(\theta)\right|.$$

https://doi.org/10.1017/S0266466619000409 Published online by Cambridge University Press

Similarly, for the test of treatment homogeneity

$$H_0 : \tau(\theta) = \tau(\theta') \quad \forall \theta, \theta' \in \Theta \tag{3.9}$$

against the alternative of treatment heterogeneity, we use the test statistic

$$T^{TH} = \sup_{\theta \in \Theta} \left| \sqrt{nh_n^3} \left(\hat{\tau}(\theta) - \int_\Theta \hat{\tau}(\vartheta) d\vartheta \right) \right|.$$

We can approximate the asymptotic distribution of T^{TH} by

$$\sup_{\theta \in \Theta} \left| \left(\widehat{\Xi}(\theta) - \int_\Theta \widehat{\Xi}(\vartheta) d\vartheta \right) \right|.$$

3.3. Step-By-Step Procedure

We summarize the above estimation and multiplier bootstrap procedures as the 14-step algorithm presented below. For ease of following the procedure, recall some shorthand notations: K is a kernel function; h_n is a bandwidth parameter; $e_1 = (0,1,0,0)^\top$; $r_3(u) = (1, u, u^2, u^3)^\top$; $\delta_i^+ = \mathbb{1}\{X_i \geq 0\}$; $\delta_i^- = \mathbb{1}\{X_i < 0\}$; and $\Gamma^\pm = \int_{\mathbb{R}_\pm} r_3(u) r_3^\top(u) K(u) du$.

Algorithm 1 (Estimation and inference)**.**

1. Set grids $\mathcal{Y}$ and Θ of quantiles and outcome values, respectively. Fix a number B of bootstrap iterations.
2. Let $\left[\widehat{\mu}_1(0^\pm, y, d), \widehat{\mu}_1'(0^\pm, y, d) h_n, \widehat{\mu}_1''(0^\pm, y, d) h_n^2/2!, \widehat{\mu}_1'''(0^\pm, y, d) h_n^3/3!\right]^\top$ be given by

$$\arg\min_{\alpha \in \mathbb{R}^4} \sum_{i=1}^n \left[\mathbb{1}\{Y_i \leq y\} \mathbb{1}\{D_i = d\} - r_3^\top\left(\frac{X_i}{h_n}\right)\alpha \right]^2 K\left(\frac{X_i}{h_n}\right) \delta_i^\pm$$

for all $y \in \mathcal{Y}$ and $d \in \{0,1\}$, and $\left[\widehat{\mu}_2(0^\pm, d), \widehat{\mu}_2'(0^\pm, d) h_n, \widehat{\mu}_2''(0^\pm, d) h_n^2/2!, \widehat{\mu}_2'''(0^\pm, d) h_n^3/3!\right]^\top$ be given by

$$\arg\min_{\alpha \in \mathbb{R}^4} \sum_{i=1}^n \left[\mathbb{1}\{D_i = d\} - r_3^\top\left(\frac{X_i}{h_n}\right)\alpha \right]^2 K\left(\frac{X_i}{h_n}\right) \delta_i^\pm \quad \text{for all } d \in \{0,1\}$$

where the choice of h_n is given in Appendix B.3.
3. Using the estimates $\hat{\mu}_1'(0^\pm, \cdot, \cdot)$ and $\hat{\mu}_2'(0^\pm, \cdot)$ from Step 3, set

$$\widehat{F}_{Y^d|VX}(y|h(0),0) = \frac{\hat{\mu}_1'(0^+, y, d) - \hat{\mu}_1'(0^-, y, d)}{\hat{\mu}_2'(0^+, d) - \hat{\mu}_2'(0^-, d)} \quad \text{for all } y \in \mathcal{Y} \text{ and } d \in \{0,1\}.$$

https://doi.org/10.1017/S0266466619000409 Published online by Cambridge University Press

Optionally, monotonically rearrange $\{\widehat{F}_{Y^d|VX}(y|h(0),0)\}_{y\in\mathcal{Y}}$ for each $d \in \{0,1\}$.

4. Using the estimates $\{\widehat{F}_{Y^d|VX}(\cdot|h(0),0) : d \in \{0,1\}\}$ from Step 3, set

$$\widehat{Q}_{Y^d|VX}(\theta) = \inf\left\{y \in \mathscr{Y} : \widehat{F}_{Y^d|VX}(y|h(0),0) \geq \theta\right\} \quad \text{for all } \theta \in \Theta \text{ and } d \in \{0,1\}.$$

5. Using the estimates $\{\widehat{Q}_{Y^d|VX}(\cdot) : d \in \{0,1\}\}$ from Step 4, set

$$\hat{\tau}(\theta) = \widehat{Q}_{Y^1|VX}(\theta) - \widehat{Q}_{Y^0|VX}(\theta) \quad \text{for all } \theta \in \Theta.$$

6. Using the estimates from Step 2, set

$$\begin{aligned}\tilde{\mu}_1(x,y,d) = &\left[\widehat{\mu}_1(0^+,y,d) + \widehat{\mu}_1'(0^+,y,d)x + \widehat{\mu}_1''(0^+,y,d)\frac{x^2}{2}\right.\\ &\left.+\widehat{\mu}_1'''(0^+,y,d)\frac{x^3}{3!}\right]\delta_x^+\\ &+\left[\widehat{\mu}_1(0^-,y,d) + \widehat{\mu}_1'(0^-,y,d)x + \widehat{\mu}_1''(0^-,y,d)\frac{x^2}{2}\right.\\ &\left.+\widehat{\mu}_1'''(0^-,y,d)\frac{x^3}{3!}\right]\delta_x^- \quad \text{and}\\ \tilde{\mu}_2(x,d) = &\left[\widehat{\mu}_2(0^+,d) + \widehat{\mu}_2'(0^+,d)x + \widehat{\mu}_2''(0^+,d)\frac{x^2}{2} + \widehat{\mu}_2'''(0^+,d)\frac{x^3}{3!}\right]\delta_x^+\\ &+\left[\widehat{\mu}_2(0^-,d) + \widehat{\mu}_2'(0^-,d)x + \widehat{\mu}_2''(0^-,d)\frac{x^2}{2}\right.\\ &\left.+\widehat{\mu}_2'''(0^-,d)\frac{x^3}{3!}\right]\delta_x^-\end{aligned}$$

for all $x \in \mathcal{X}$, $y \in \mathcal{Y}$ and $d \in \{0,1\}$.

7. Let $\breve{\mu}'(0^\pm,y,d)$ be given by

$$a_n^{-1}e_1^\top \arg\min_{\alpha\in\mathbb{R}^4} \sum_{i=1}^n \left[\frac{1}{b_n}K\left(\frac{Y_i-y}{b_n}\right)\mathbb{1}\{D_i = d\} - r_3^\top\left(\frac{X_i}{a_n}\right)\alpha\right]^2 K\left(\frac{X_i}{a_n}\right)\delta_i^\pm$$

for all $y \in \mathcal{Y}$ and $d \in \{0,1\}$, where the choices of a_n and b_n are given in Appendix B.3.

8. Let $\hat{f}_X(0) = \frac{1}{nc_n}\sum_{i=1}^n K(X_i/c_n)$, where the choice of c_n is given in Appendix B.3.

9. Using the estimates $\hat{\mu}_2'(0^\pm,\cdot)$ from Step 2 and the estimates $\breve{\mu}'(0^\pm,\cdot,\cdot)$ from Step 7, set

https://doi.org/10.1017/S0266466619000409 Published online by Cambridge University Press

$$\hat{f}_{Y^d|VX}(y|h(0),0) = \frac{\check{\mu}'(0^+,y,d) - \check{\mu}'(0^-,y,d)}{\hat{\mu}_2'(0^+,d) - \hat{\mu}_2'(0^-,d)} \quad \text{for all } y \in \mathcal{Y} \text{ and } d \in \{0,1\}.$$

10. Generate the multipliers $\{\{\xi_{ib}\}_{i=1}^n\}_{b=1}^B \overset{i.i.d}{\sim} N(0,1)$ independently from the data.
11. Using the estimates $\tilde{\mu}_1(\cdot,\cdot,\cdot)$ and $\tilde{\mu}_2(\cdot,\cdot)$ from Step 6, the estimate $\hat{f}_X(0)$ from Step 8, and the multipliers $\{\{\xi_{ib}\}_{i=1}^n\}_{b=1}^B \overset{i.i.d}{\sim} N(0,1)$ from Step 0, obtain $\{\hat{v}^{\pm}_{\xi,n,b}(y,d,1) : y \in \mathcal{Y}, d \in \{0,1\}\}_{b=1}^B$ and $\{\hat{v}^{\pm}_{\xi,n,b}(y,d,2) : y \in \mathcal{Y}, d \in \{0,1\}\}_{b=1}^B$, where

$$\hat{v}^{\pm}_{\xi,n,b}(y,d,1) = \frac{1}{\sqrt{nh_n}\hat{f}_X(0)} \sum_{i=1}^n \xi_{ib} e_1^\top (\Gamma^{\pm})^{-1} r_3\Big(\frac{X_i}{h_n}\Big) \times \Big[\mathbb{1}\{Y_i \le y\}\mathbb{1}\{D_i = d\} - \tilde{\mu}_1(X_i,y,d)\Big] K\Big(\frac{X_i}{h_n}\Big)\delta_i^{\pm}$$

$$\hat{v}^{\pm}_{\xi,n,b}(y,d,2) = \frac{1}{\sqrt{nh_n}\hat{f}_X(0)} \sum_{i=1}^n \xi_{ib} e_1^\top (\Gamma^{\pm})^{-1} r_3\Big(\frac{X_i}{h_n}\Big) \times \Big[\mathbb{1}\{D_i = d\} - \tilde{\mu}_2(X_i,d)\Big] K\Big(\frac{X_i}{h_n}\Big)\delta_i^{\pm}$$

with the choice of h_n given in Appendix B.3.
12. Using the estimates $\hat{\mu}_1'(0^{\pm},\cdot,\cdot)$ and $\hat{\mu}_2'(0^{\pm},\cdot)$ from Step 2 and $\{\hat{v}^{\pm}_{\xi,n,b}(y,d,1) : y \in \mathcal{Y}, d \in \{0,1\}\}_{b=1}^B$ and $\{\hat{v}^{\pm}_{\xi,n,b}(y,d,2) : y \in \mathcal{Y}, d \in \{0,1\}\}_{b=1}^B$ from Step 11, obtain $\{\hat{Z}_{\xi,n,b}(y,d) : y \in \mathcal{Y}, d \in \{0,1\}\}_{b=1}^B$, where

$$\hat{Z}_{\xi,n,b}(y,d) = \frac{[\hat{\mu}_2'(0^+,d) - \hat{\mu}_2'(0^-,d)][\hat{v}^{+}_{\xi,n,b}(y,d,1) - \hat{v}^{-}_{\xi,n,b}(y,d,1)]}{[\hat{\mu}_2'(0^+,d) - \hat{\mu}_2'(0^-,d)]^2} [\hat{\mu}_1'(0^+,y,d) - \hat{\mu}_1'(0^-,y,d)] - \frac{[\hat{v}^{+}_{\xi,n,b}(y,d,2) - \hat{v}^{-}_{\xi,n,b}(y,d,2)]}{[\hat{\mu}_2'(0^+,d) - \hat{\mu}_2'(0^-,d)]^2}.$$

13. Using the estimates $\{\hat{Q}_{Y^d|VX}(\cdot) : d \in \{0,1\}\}$ from Step 4, the estimates $\{\hat{f}_{Y^d|VX}(\cdot|h(0),0) : d \in \{0,1\}\}$ from Step 9, and $\{\hat{Z}_{\xi,n,b}(y,d) : y \in \mathcal{Y}, d \in \{0,1\}\}_{b=1}^B$ from Step 12, obtain $\{\widehat{\Xi}_b(\cdot)\}_{b=1}^B$, where

$$\widehat{\Xi}_b(\cdot) = -\left[\frac{\hat{Z}_{\xi,n,b}(\hat{Q}_{Y^1|VX}(\cdot),1)}{\hat{f}_{Y^1|VX}(\hat{Q}_{Y^1|VX}(\cdot)|h(0),0)} - \frac{\hat{Z}_{\xi,n,b}(\hat{Q}_{Y^0|VX}(\cdot),0)}{\hat{f}_{Y^0|VX}(\hat{Q}_{Y^0|VX}(\cdot)|h(0),0)}\right].$$

https://doi.org/10.1017/S0266466619000409 Published online by Cambridge University Press

14. Let the process of $\{\widehat{\Xi}_b(\cdot)\}_{b=1}^B$ from Step 13 approximate the asymptotic process of $\sqrt{nh_n^3}[\hat{\tau}(\cdot) - \tau(\cdot)]$.

Although we presented a practical guideline on estimation and inference for the conditional quantile treatment effects $\tau(\theta)$ in this section, we refer interested readers to Appendix A for a formal theory which justifies the above procedure. Furthermore, Appendix B presents additional details of the practical guideline. These appendix sections provide a formal justification for every step in the above 14-step procedure.

4. SIMULATION STUDIES

In this section, we conduct Monte Carlo simulations to evaluate finite-sample performances of the inference method proposed above. Following earlier studies on econometric methods for regression discontinuity and kink designs, we use a realistic data generating model as described below.

Consider the following data generating process:

$$Y = m(X) + \beta_1 \cdot D + (\gamma_0 + \gamma_1 \cdot D) \cdot U$$
$$D = \mathbb{1}\{2 \cdot \mathbb{1}\{X \geq 0\} \cdot X - X - 1 \geq V\},$$

where (X, U, V) is generated according to

$$\begin{pmatrix} X \\ U \\ V \end{pmatrix} \overset{i.i.d.}{\sim} N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_X^2 & \rho_{XU} \cdot \sigma_X \cdot \sigma_U & \rho_{XV} \cdot \sigma_X \cdot \sigma_V \\ \rho_{XU} \cdot \sigma_X \cdot \sigma_U & \sigma_U^2 & \rho_{UV} \cdot \sigma_V \cdot \sigma_U \\ \rho_{XV} \cdot \sigma_X \cdot \sigma_V & \rho_{UV} \cdot \sigma_U \cdot \sigma_V & \sigma_V^2 \end{pmatrix} \right)$$

with the parameter values given by

$$\gamma_0 = 1.00, \sigma_X^2 = 0.1781742^2, \sigma_U^2 = 0.1295^2, \sigma_V^2 = 0.5^2,$$
$$\rho_{XU} = 0.25, \rho_{XV} = 0.00, \text{and } \rho_{UV} = 0.25.$$

These variance parameter values are chosen to match those of Calonico et al. (2014) and related subsequent studies. The treatment level parameter β_1 and the treatment heterogeneity parameter γ_1 will be varied throughout simulations. Notice that the treatment assignment probability has a kink at $x = 0$ in this data generating model.

Also following the simulation setting of Calonico et al. (2014), the polynomial part $m(\cdot)$ of the outcome equation is defined due to the estimates in the actual empirical studies by Lee (2008) and Ludwig and Miller (2007), less linear terms.[3] Specifically, the polynomial model due to Lee (2008) less linear terms reads

$$m(X) = \begin{cases} 7.18X^2 + 20.21X^3 + 21.54X^4 + 7.33X^5 & \text{if } x < 0 \\ -3.00X^2 + 7.99X^3 - 9.01X^4 + 3.56X^5 & \text{if } x \geq 0 \end{cases}, \tag{4.1}$$

[3] We remove the linear terms since our design is a regression kink rather than a regression discontinuity. Specifically, we make this modification to satisfy our Assumption 1(v).

https://doi.org/10.1017/S0266466619000409 Published online by Cambridge University Press

and the polynomial model due to Ludwig and Miller (2007) less linear terms reads

$$m(X) = \begin{cases} 3.28X^2 + 1.45X^3 + 0.23X^4 + 0.03X^5 & \text{if } x < 0 \\ -54.81X^2 + 74.30X^3 - 45.02X^4 + 9.83X^5 & \text{if } x \geq 0 \end{cases}. \tag{4.2}$$

Precisely following the step-by-step procedure presented in Algorithm 1 with the number of multiplier bootstrap resampling set to $B = 500$, we run simulations to assess finite-sample accuracy of frequencies of the uniform coverage of the quantile treatment effects on the quantile set $\Theta = [0.20, 0.80]$ for each of the sample sizes $n \in \{200, 400, 600, 800\}$, for each of the alternative parameter values of $\beta_1 \in \{0.00, 0.02, 0.04, 0.06, 0.08\}$ and $\gamma_1 \in \{0.0, 0.2, 0.4, 0.6, 0.8\}$, and for each of the polynomial models (4.1) due to Lee (2008) and (4.2) due to Ludwig and Miller (2007). The number of Monte Carlo replications is set to 5,000 for each set of simulations.

Table 1 shows simulated frequencies of uniform coverage with the nominal probability of 95%. The reported frequencies are far from the nominal probability when the sample size is as small as $n = 200$ under the polynomial model (4.1), while they are close to the nominal probability even at such a small sample size as $n = 200$ under the polynomial model (4.2) for all the parameter values. The reported frequencies become reasonably close to the nominal probability under both of the polynomial models (4.1) and (4.2) for all the parameter values when the sample size is $n = 600$ or 800. These results demonstrate that, for models designed similarly to Calonico et al. (2014) and calibrated to those of representative empirical studies such as Lee (2008) and Ludwig and Miller (2007), our proposed method of inference will yield fairly accurate uniform coverage probabilities if n is large enough, for example, $n \geq 500$.

5. AN EMPIRICAL ILLUSTRATION

In this section, we apply our proposed method to real data. Specifically, applying the proposed method to the Russian Longitudinal Monitoring Survey (RLMS),[4] we analyze quantile treatment effects of compulsory military service in the Russian Army on earnings at later time in life. To this end, we follow the ideas of preceding papers by Card and Yakovlev (2014) and Dong (2018) who use the same dataset and a RKD. With a sharp demilitarization at the end of the Cold War, the graph of the percentage of Russian males who served in the army as a function of calendar time at which they turn age 18 exhibits a kink in 1989. Card and Yakovlev (2014) and Dong (2018) use this feature as a RKD, and analyze health and labor effects of compulsory military service based on this research design. Although they focus on conditional average effects, we extend their empirical methods and make inference

[4]The Official source name is "Russia Longitudinal Monitoring Survey (RLMS-HSE)." Data locations http://www.cpc.unc.edu/projects/rlms-hse and http://www.hse.ru/org/hse/rlms.

https://doi.org/10.1017/S0266466619000409 Published online by Cambridge University Press

TABLE 1. Monte Carlo simulation results of the frequencies of uniform coverage of the quantile treatment effects on the set $\Theta = [0.20, 0.80]$ of quantiles with the nominal probability of 95%

Polynomial model			$m(X) =$ equation (4.1)				$m(X) =$ equation (4.2)			
			Sample size (n)				Sample size (n)			
Parameters	β_1	γ_1	200	400	600	800	200	400	600	800
	0.02	0.00	0.858	0.929	0.939	0.951	0.933	0.935	0.941	0.947
	0.04	0.00	0.855	0.920	0.944	0.955	0.919	0.939	0.945	0.946
	0.06	0.00	0.862	0.920	0.945	0.955	0.931	0.947	0.948	0.943
	0.08	0.00	0.871	0.931	0.950	0.956	0.927	0.939	0.942	0.948
Polynomial model			$m(X) =$ equation (4.1)				$m(X) =$ equation (4.2)			
			Sample size (n)				Sample size (n)			
Parameters	β_1	γ_1	200	400	600	800	200	400	600	800
	0.00	0.20	0.851	0.920	0.939	0.948	0.930	0.938	0.937	0.940
	0.00	0.40	0.852	0.919	0.940	0.948	0.929	0.935	0.943	0.944
	0.00	0.60	0.853	0.912	0.938	0.951	0.925	0.937	0.942	0.945
	0.00	0.80	0.844	0.926	0.935	0.946	0.932	0.936	0.945	0.945

Note: The results are displayed for each of the sample sizes $n \in \{200, 400, 600, 800\}$, for each of the alternative parameter values of $\beta_1 \in \{0.00, 0.02, 0.04, 0.06, 0.08\}$ and $\gamma_1 \in \{0.0, 0.2, 0.4, 0.6, 0.8\}$, and for each of the polynomial models (4.1) due to Lee (2008) and (4.2) due to Ludwig and Miller (2007). The number of Monte Carlo replications is set to 5,000 for each set of simulations.

for conditional quantile treatment effects for analysis of heterogeneous effects of compulsory military service.

Similarly to the previous work, we set the running and treatment variables as follows. The running variable X is the calendar time when a man turns age 18, with the kink location (year 1989) normalized to zero. The treatment variable D is the binary indicator that a man serves in the army. As a continuous outcome variable Y for the analysis of quantile treatment effects, we consider monthly earnings as in Dong (2018).[5] We refer interested readers to Card and Yakovlev (2014) and Dong (2018) for more details about the data, descriptive statistics, figures of the first-stage kink, and smoothness checks for design validation. Applying the same procedure as the one used for our simulation studies in Section 4, we draw estimates and confidence bands for the quantile treatment effects.

Figure 1 illustrates the estimates with uniform 90% and 95% confidence bands on the quantile set $\Theta = [0.20, 0.80]$. Notice that the estimates are negative atall

[5]Card and Yakovlev (2014) and Dong (2018) also consider other qualitative outcome variables, but we choose this continuous outcome variable as our objects of interest are the quantile treatment effects.

https://doi.org/10.1017/S0266466619000409 Published online by Cambridge University Press

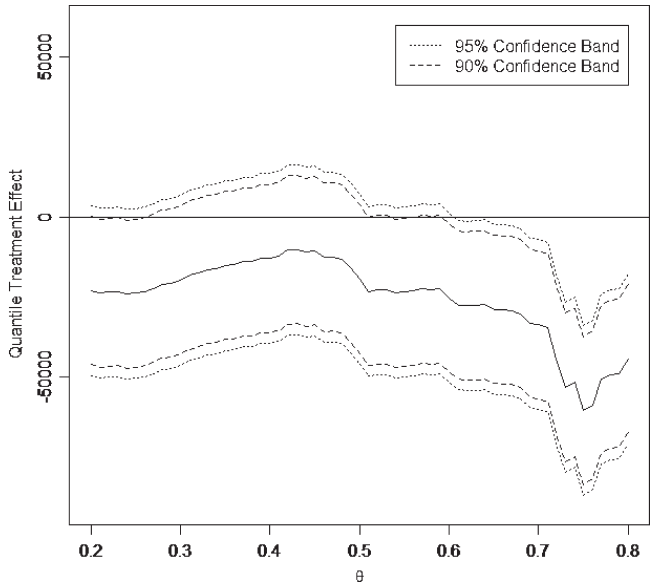


Fig. 1- B/W online, B/W in print

FIGURE 1. The quantile treatment effects of compulsory military service on earnings. The displayed curves indicate the quantile treatment effects on the quantile set $\Theta = [0.20, 0.80]$ with uniform 90% and 95% confidence bands.

the quantiles. The negative sign is consistent with the earlier study by Dong (2018), where compulsory military service in the Russian Army was concluded to have negative labor effects on average. Furthermore, we also emphasize that the confidence bands exhibit heterogeneous characteristics across quantile levels. Namely, while we fail to see significant quantile treatment effects for those men below the 60th percentile of monthly earnings, we do find significantly negative effects for those men above the 60th percentile. Furthermore, the point estimates exhibit larger negative values for those men above the 60th percentile than those men below this percentile. These results imply that compulsory military service indeed has negative effects, and the negative effects are particularly larger for the subpopulation of potential higher income earners.

We conclude this empirical analysis with statistical inference for the hypotheses of treatment significance and treatment heterogeneity. At the level of 95%, we do reject the null hypothesis (3.8) of treatment nullity in favor of the alternative hypothesis of treatment significance. At the level of 95%, we do reject the null hypothesis (3.9) of treatment homogeneity in favor of the alternative hypothesis of treatment heterogeneity. These results illustrate the utility of quantile treatment effects as a device for the analysis of heterogeneous treatment effects, and provide findings about more detailed features of treatment effects in addition to those discovered in previous studies.

https://doi.org/10.1017/S0266466619000409 Published online by Cambridge University Press

6. SUMMARY

The existing literature on identification in RKDs includes the following three results. Nielsen et al. (2010) and Card et al. (2015) propose identification of average effects of continuous treatments. Dong (2018) proposes identification of average effects of binary treatments. Chiang and Sasaki (2019) propose identification of quantile-wise effects of continuous treatments. This literature presently lacks an identification result for quantile-wise effects of binary treatments. To complete this literature on identification, this article proposes identification of quantile-wise effects of binary treatments in RKDs.

Specifically, we show that a local Wald ratio of derivatives of certain conditional expectation functions identifies the conditional distribution functions of potential outcomes given the event of local compliance. Furthermore, taking the difference of the left-inverses of these identified conditional distribution functions, we also identify the conditional quantile treatment effects given the event of local compliance. Although this article focuses on quantile treatment effects, the identified conditional distribution functions of potential outcomes can be also used for analysis of distributional treatment effects.

Although the main contribution of this article is the identification result presented in Section 2, we also develop a theory and method of estimation and inference for completeness (Section 3—also see the appendix for further details). Simulation studies in Section 4 demonstrate that the proposed methods of estimation and inference work well in finite samples. Finally, the empirical illustration in Section 5 shows a case in which the quantile treatment effects reveal heterogeneous treatment effects in real data.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S0266466619000409

REFERENCES

Abadie, A. (2002) Bootstrap tests for distributional treatment effects in instrumental variable models. *Journal of the American Statistical Association* 97, 284–292.

Abadie, A. (2003) Semiparametric instrumental variable estimation of treatment response models. *Journal of Econometrics* 113, 2, 231–263.

Björklund, A. & R. Moffitt (1987) The estimation of wage and welfare gains in self-selection models. *Review of Economics and Statistics* 69, 42–49.

Calonico, S., M.D. Cattaneo, & M. Farrell (2018a) On the effect of bias estimation on coverage accuracy in nonparametric inference. *Journal of the American Statistical Association* 113, 522, 767–779.

Calonico, S., M.D. Cattaneo, & M.H. Farrell (2018b) Optimal bandwidth choice for robust bias corrected inference in regression discontinuity designs. arXiv preprint arXiv:1809.00236.

Calonico, S., M.D. Cattaneo, & M.H. Farrell (2019) Coverage error optimal confidence intervals for local polynomial regression. arXiv preprint arXiv:1808.01398.

https://doi.org/10.1017/S0266466619000409 Published online by Cambridge University Press

Calonico, S., M.D. Cattaneo, & R. Titiunik (2014) Robust nonparametric confidence intervals for regression discontinuity designs. *Econometrica* 82, 2295–2326.
Card, D., D. Lee, Z. Pei, & A. Weber (2015) Inference on causal effects in a generalized regression kink design. *Econometrica* 83, 2453–2483.
Card, D. & E. Yakovlev, E (2014) The causal effect of serving in the army on health: Evidence from regression kink design and Russian data.
Cerulli, G., Y. Dong, A. Lewbel, & A. Poulsen (2017) Testing stability of regression discontinuity models. In M.D. Cattaneo and J.C. Escanciano (eds.) *Advances in Econometrics,* vol. 38, Regression Discontinuity Designs: Theory and Applications, pp. 317–339. Emerald Publishing.
Chernozhukov, V. & I. Fernández-Val (2005) Subsampling inference on quantile regression processes. *Sankhya: The Indian Journal of Statistics* 67, 253–276.
Chernozhukov, V., I. Fernández-Val, & A. Galichon (2010) Quantile and probability curves without crossing. *Econometrica* 78, 1093–1125.
Chernozhukov, V. & C. Hansen (2005) An IV model of quantile treatment effects. *Econometrica* 73, 245–261.
Chiang, H. & Y. Sasaki (2019) Causal inference by quantile regression kink designs. *Journal of Econometrics* 210, 405–433.
Chiang, H., Y.-C. Hsu, & Y. Sasaki (2019) Robust uniform inference for quantile treatment effects in regression discontinuity designs. *Journal of Econometrics* 211, 589–618.
Dong, Y. (2018) Jump or kink? Identifying education effects by regression discontinuity design without the discontinuity. Working Paper.
Dong, Y. & A. Lewbel (2015) Identifying the effect of changing the policy thresholds in regression discontinuity models. *Review of Economics and Statistics* 97, 1081–1092.
Frandsen, B.R., M. Frölich, & B. Melly (2012) Quantile treatment effects in the regression discontinuity design. *Journal of Econometrics* 168, 382–395.
Guerre, E. & C. Sabbah (2012) Uniform bias study and Bahadur representation for local polynomial estimators of the conditional quantile function. *Econometric Theory* 28, 87–129.
Heckman, J.J. & E.J. Vytlacil (1999) Local instrumental variables and latent variable models for identifying and bounding treatment effects. *Proceedings of the National Academy of Sciences* 96, 4730–4734.
Heckman, J.J. & E.J. Vytlacil (2005) Structural equations, treatment effects, and econometric policy evaluation." *Econometrica* 73, 669–738.
Kato, R. & Y. Sasaki (2017) On using linear quantile regressions for causal inference. *Econometric Theory* 33, 664–690.
Koenker, R. & Z. Xiao (2002) Inference on the quantile regression process. *Econometrica* 70, 1583–1612.
Kong, E., O. Linton, & Y. Xia (2010) Uniform Bahadur representation for local polynomial estimates of M-regression and its application to the additive model. *Econometric Theory* 26, 1529–1564.
Lee, D.S. (2008) Randomized experiments from non-random selection in U.S. House elections. *Journal of Econometrics* 142, 675–697.
Li, Q. & J.S. Racine (2008) *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.
Ludwig, J. & D.L. Miller (2007) Does head start improve children's life chances? Evidence from a regression discontinuity design." *Quarterly Journal of Economics* 122, 159–208.
Nielsen, H.S., T. Sørensen, & C. Taber (2010) Estimating the effect of student aid on college enrollment: evidence from a government grant policy reform. *American Economic Journal: Economic Policy* 2, 185–215.
Qu, Z. & J. Yoon (2015) Nonparametric estimation and inference on conditional quantile processes. *Journal of Econometrics* 185, 1–19.
Qu, Z. & J. Yoon (2018) Uniform inference on quantile effects under sharp regression discontinuity designs. *Journal of Business & Economic Statistics* 37, 625–647.
Sasaki, Y. (2015) What do quantile regressions identify for general structural functions? *Econometric Theory* 31, 1102–1116.
Shen, S. & X. Zhang (2016) Distributional tests for regression discontinuity: Theory and empirical examples. *Review of Economics and Statistics* 98, 685–700.

https://doi.org/10.1017/S0266466619000409 Published online by Cambridge University Press

APPENDIX

A. ESTIMATION AND INFERENCE: FORMAL THEORY

We use the following set of assumptions for the UBR, the bootstrap validity, and consistent conditional density and first-stage estimations. Fix $a \in (0, 1/2)$ and $\epsilon > 0$, denote

$$\mathscr{Y}_1 = [Q_{Y^1|VX}(a) - \epsilon, Q_{Y^1|VX}(1-a) + \epsilon] \cup [Q_{Y^0|VX}(a) - \epsilon, Q_{Y^0|VX}(1-a) + \epsilon].$$

We will write $a \lesssim b$ if there exists a universal constant C such that $a \leq Cb$. Denote

$$P_{D|X}(d|x) = \mathbb{P}^x(D = d|X = x).$$

We define the following objects for all $y_1, y_2 \in \mathscr{Y}_1$, $d_1, d_2 \in \mathscr{D}$:

$$\begin{aligned}
\sigma_{11}((y_1,d_1),(y_2,d_2)|x) &= E[(\mathbb{1}\{Y \leq y_1, D = d_1\} - \mu_1(X, y_1, d_1))\cdot \\
&\qquad (\mathbb{1}\{Y \leq y_2, D = d_2\} - \mu_1(X, y_2, d_2))|X = x], \\
\sigma_{22}((y_1,d_1),(y_2,d_2)|x) &= E[(\mathbb{1}\{D = d_1\} \\
&\qquad - \mu_2(X, d_1))(\mathbb{1}\{D = d_2\} - \mu_2(X, d_2))|X = x], \qquad \text{and} \\
\sigma_{12}((y_1,d_1),(y_2,d_2)|x) &= E[(\mathbb{1}\{Y \leq y_1, D = d_1\} - \mu_1(X, y_1, d_1))(\mathbb{1}\{D = d_2\} \\
&\qquad - \mu_2(X, d_2))|X = x].
\end{aligned}$$

Assumption A.1. Let $[\underline{x}, \bar{x}]$ be a compact interval containing 0 in its interior. Let $a \in (0, 1/2)$.

(i) (a) $\{Y_i, D_i, X_i\}_{i=1}^n$ are n independent copies of random vector (Y, D, X) with support $\mathscr{Y} \times \mathscr{D} \times \mathscr{X}$ defined on a probability space $(\Omega^x, \mathscr{F}^x, \mathbb{P}^x)$. (b) X has a continuously differentiable density function f_X with $0 < f_X(0) < \infty$. (c) $f_{YD|X}(y, d|x)$ is well-defined on $\mathscr{Y}_1 \times \mathscr{D} \times ([\underline{x}, \bar{x}] \setminus \{0\})$ and $|f_{YD|X}(y, d|0^+) - f_{YD|X}(y, d|0^-)| > m > 0$ on $\mathscr{Y}_1 \times \mathscr{D}$.

(ii) (a) Conditional density $f_{Y|XD}$ is Lipschitz continuous on $\mathscr{Y}_1 \times [\underline{x}, 0)$ and $\mathscr{Y}_1 \times (0, \bar{x}]$ for each d and is four-time partially differentiable in x and twice partially differentiable in y for each d. $\frac{\partial^j}{\partial x^j} \frac{\partial^k}{\partial y^k} f_{Y|XD}(\cdot|\cdot, d)$ is continuous and uniformly bounded on $\mathscr{Y}_1 \times [\underline{x}, 0)$ and $\mathscr{Y}_1 \times (0, \bar{x}]$ for each d for all j, $k \in \mathbb{N}$, $j + k \leq 4$. (b) $P_{D|X}(d|\cdot)$ is Lipschitz continuous in x, four-time differentiable on $[\underline{x}, 0)$ and $(0, \bar{x}]$ for each d. $\frac{\partial^4}{\partial x^4} P_{D|X}(d|\cdot)$ is continuous and uniformly bounded on $[\underline{x}, 0)$ and $(0, \bar{x}]$ for each d. (c) For any y_1, $y_2 \in \mathscr{Y}_1$, d_1, $d_2 \in \mathscr{D}$, we have $\sigma_{11}((y_1, d_1), (y_2, d_2)|\cdot)$, $\sigma_{12}((y_1, d_1), (y_2, d_2)|\cdot)$, and $\sigma_{22}((y_1, d_1), (y_2, d_2)|\cdot) \in \mathcal{C}^1([\underline{x}, \bar{x}] \setminus \{0\})$, where $\mathcal{C}^1$ is the collection of continuously differentiable functions.

(iii) The bandwidths satisfy $h_n \to 0$, $nh_n^3 \to \infty$, $nh_n^9 \to 0$, $0 < h_n \leq h_0$ for some finite h_0.

(iv) (a) $K : [-1, 1] \to \mathbb{R}_+$ is bounded and $\int_{\mathbb{R}} K(u)du = 1$. (b) $\{K(\cdot/h) : h > 0\}$ is of VC type. (c) $\Gamma^{\pm} = \int_{\mathbb{R}_\pm} r_3(u) r_3^\top(u) K(u) du$ are positive definite.

(v) $\hat{f}_X(0)$ is a consistent estimator for $f_X(0)$. For $d = 0, 1$, $\hat{f}_{Y^d|VX}(\cdot|h(0), 0)$ are uniformly consistent estimators for $f_{Y^d|VX}(\cdot|h(0), 0)$. $\tilde{\mu}_1(x, y, d)\mathbb{1}\{|x/h_n| \leq 1\}$ and $\tilde{\mu}_2(x, d)\mathbb{1}\{|x/h_n| \leq 1\}$ are uniformly consistent estimators for $\mu_1(x, y, d)\mathbb{1}\{|x/h_n| \leq 1\}$ and $\mu_2(x, d)\mathbb{1}\{|x/h_n| \leq 1\}$ on $\mathscr{X} \times \mathscr{Y}_1 \times \mathscr{D}$.

(vi) $\{\xi_1, \ldots, \xi_n\}$ are n independent and identically distributed copies of a standard normal random variable ξ defined on a probability space $(\Omega^\xi, \mathscr{F}^\xi, \mathbb{P}^\xi)$ that is independent of $(\Omega^x, \mathscr{F}^x, \mathbb{P}^x)$.

https://doi.org/10.1017/S0266466619000409 Published online by Cambridge University Press

Part (i) concerns the sampling procedure and the distribution of data. Part (ii) requires smoothness of the conditional expectation functions on a deleted neighborhood of $x_0 = 0$. Part (iii) regulates the rate at which bandwidth decreases, which is consistent with examples of common choice rules to be presented in Appendix B.3. For example, the MSE-optimal bandwidth for the local quadratic estimator (e.g., $nh_n^7 \to \infty$) is allowed. Although we write a single bandwidth h_n across all values of $y \in \mathcal{Y}$, and thus across all quantiles $\theta \in \Theta$, we can allow for heterogeneous bandwidths across them in light of Assumption 1(iii) in Chiang et al. (2019). Namely, if the dependence of the bandwidths $h_n(y)$ on y takes the form of $h_n(y) = h_n c(y)$ for some bounded Lipschitz function c, then we continue to obtain the same result of the validity of the multiplier bootstrap as stated below. Even in this case of varying bandwidths, we will hereafter suppress the notation of the dependence of $h_n(y)$ on y for notational simplicity, and will write them as h_n throughout. Part (iv) is satisfied by common kernel functions, such as uniform, triangular, biweight, triweight, and Epanechnikov kernels, for example. Part (v) is a high-level assumption of (uniformly) consistent estimation of the first-stage estimators. Although we keep this high-level statement for the current section, Appendix B.2 proposes concrete examples of such uniformly consistent estimators. Part (vi) requires the multiplier random sample to be drawn independently of the data $\{Y_i, D_i, X_i\}_{i=1}^n$. We remark that Part (vi) implies that all (uniformly) consistent estimators with respect to $\mathbb{P}^x$ are also (uniformly) consistent with respect to $\mathbb{P}^{x \times \xi}$.

Under Assumption A.1(i), (ii)(a), (ii)(b), (iii), (iv), an application of Lem. 1 of Chiang et al. (2019) gives the UBR as in equations (3.4) and (3.5). The following theorem establishes (i) (a) the asymptotic distribution of the UBR; (i) (b) the asymptotic distribution of the local Wald estimators; (i) (c) the asymptotic distribution of the conditional quantile treatment effect estimator; and (ii) the bootstrap validity. A proof is provided in Appendix C.2 in the online supplementary material available at Cambridge Journals Online (journals.cambridge.org/ect).

THEOREM A.1 (Asymptotic distributions and bootstrap validity). *Suppose Assumptions 1 and A.1 hold, then there exists a zero-mean Gaussian process* $\mathbb{G} : \Omega^x \mapsto \ell^\infty(\{\mathscr{Y}_1 \times \mathscr{D} \times \{1,2\}\})$, *where* l^∞ *is the collection of all bounded real valued functions, such that*

(i) (a) $v_n^+ - v_n^- \rightsquigarrow \mathbb{G}$.

(i) (b) $\sqrt{nh_n^3}[\widehat{F}_{Y^d|VX}(\cdot|h(0),0) - F_{Y^d|VX}(\cdot|h(0),0)] \rightsquigarrow \mathbb{G}_F(\cdot,d)$ *holds, where* $\mathbb{G}_F(\cdot,d)$ *is given by*

$$\mathbb{G}_F(y,d) = \frac{\begin{array}{c}[\mu_2'(0^+,d) - \mu_2'(0^-,d)]\mathbb{G}(y,d,1) \\ -[\mu_1'(0^+,y,d) - \mu_1'(0^-,y,d)]\mathbb{G}(y,d,2)\end{array}}{[\mu_2'(0^+,d) - \mu_2'(0^-,d)]^2}.$$

(i) (c) $\sqrt{nh_n^3}[\hat{\tau} - \tau] \rightsquigarrow \mathbb{G}_\tau$ *holds, where* $\mathbb{G}_\tau$ *is given for each* $\theta \in \Theta = [a, 1-a]$ *by*

$$\mathbb{G}_\tau(\theta) = -\left[\frac{\mathbb{G}_F(Q_{Y^1|VX}(\theta|h(0),0),1)}{f_{Y^1|VX}(Q_{Y^1|VX}(\theta|h(0),0)|h(0),0)} - \frac{\mathbb{G}_F(Q_{Y^0|VX}(\theta|h(0),0),1)}{f_{Y^0|VX}(Q_{Y^0|VX}(\theta|h(0),0)|h(0),0)}\right].$$

https://doi.org/10.1017/S0266466619000409 Published online by Cambridge University Press

(ii) We have

$$\widehat{\Xi}(\cdot) = -\left[\frac{\hat{Z}_{\xi,n}(\hat{Q}_{Y^1|VX}(\cdot|h(0),0),1)}{\hat{f}_{Y^1|VX}(\hat{Q}_{Y^1|VX}(\cdot|h(0),0)|h(0),0)} - \frac{\hat{Z}_{\xi,n}(\hat{Q}_{Y^0|VX}(\cdot|h(0),0),0)}{\hat{f}_{Y^0|VX}(\hat{Q}_{Y^0|VX}(\cdot|h(0),0)|h(0),0)}\right] \overset{p}{\underset{\xi}{\rightsquigarrow}} \mathbb{G}_\tau(\cdot)$$

Remark A.1. By considering the asymptotic distribution for the local cubic local polynomial above, we effectively account for bias estimation in the asymptotic distribution from the local quadratic kernel estimate—see Calonico et al. (2014, Rem. 7) and Rem. S.A.7 in their supplementary material. Therefore, the proposed theory and bootstrap allow for robust inference under the MSE-optimal bandwidth from the local quadratic kernel estimate.

Remark A.2. $\hat{\mu}_1'(0^\pm,y,d)$, $\hat{\mu}_2'(0^\pm,d)$, and Theorem A.1 are developed for the unconstrained estimators, that is, without imposing continuity in the conditional expectation of $\mathbb{1}\{Y_i \le y\}\mathbb{1}\{D_i = d\}$ and $\mathbb{1}\{D_i = d\}$. On the other hand, for example, consider the constrained version with the restriction with $\mu_1(0^+,y,d) = \mu_1(0^-,y,d)$: the estimates can be obtained by solving the "pooled" least squares problem

$$\underset{\{\alpha,b^+,b^-\}\in\mathbb{R}^7}{\arg\min} \sum_{i=1}^{n}\left[\mathbb{1}\{Y_i \le y\}\mathbb{1}\{D_i = d\} - \alpha - \delta_i^+ r_{3\backslash 0}^\top\left(\frac{X_i}{h_n}\right)b^+ - \delta_i^- r_{3\backslash 0}^\top\left(\frac{X_i}{h_n}\right)b^-\right]^2 K\left(\frac{X_i}{h_n}\right),$$

where $r_{3\backslash 0}(u) = \left(u,u^2,u^3\right)$ and $b^\pm \in \mathbb{R}^6$ denoting the first/second/third left (right) derivatives. As shown in Appendix C.5 in the online supplementary material available at Cambridge Journals Online (journals.cambridge.org/ect), when a uniform kernel and symmetric bandwidths are used, the constrained estimators have the same asymptotic distributions as the unconstrained ones, thus our previous results still hold under the constrained estimates.

APPENDIX B. ADDITIONAL PRACTICAL CONSIDERATIONS

In order to compute the uniform consistent conditional density $f_{Y^d|VX}(\cdot|h(0),0)$ in Appendix B.1, and $\mu_1(x,y,d)\mathbb{1}\{|x/h_n| \le 1\}$ and $\mu_2(x,d)\mathbb{1}\{|x/h_n| \le 1\}$ in Appendix B.2, we continue to use the local cubic kernel models so the single MSE-optimal bandwidth from the local quadratic regression can be used throughout.

B.1. A Conditional Density Estimator

The statement of Theorem A.1 presumes that the densities $f_{Y^d|VX}(\cdot|h(0),0)$ are unknown. In order to simulate the multiplier process, we need to replace them by their uniformly consistent estimators. Note that the identifying formulas in Theorem 1 suggest

$$f_{Y^d|VX}(y|h(0),0) = \frac{\partial}{\partial y}F_{Y^d|VX}(y|h(0),0) = \frac{\frac{\partial}{\partial y}\mu_1'(0^+,y,d) - \frac{\partial}{\partial y}\mu_1'(0^-,y,d)}{\mu_2'(0^+,d) - \mu_2'(0^-,d)}.$$

Equation (3.3) gives uniformly consistent estimators for the two terms in the denominator. The two terms in the numerator can be written as

$$\frac{\partial}{\partial y}\mu_1'(0^\pm,y,d) = \frac{\partial}{\partial y}\frac{\partial}{\partial x}E[\mathbb{1}\{Y \le y\}\mathbb{1}\{D = d\}|X = 0^\pm]. \tag{B.1}$$

https://doi.org/10.1017/S0266466619000409 Published online by Cambridge University Press

With the bandwidth parameter b_n, we represent $\frac{\partial}{\partial y}\mu_1(0^{\pm},y,d)$ by the limit of the regularized approximation

$$\mu(0^{\pm},y,d) = \lim_{n\to\infty} E\Big[\frac{1}{b_n}K\Big(\frac{Y_i-y}{b_n}\Big)\mathbb{1}\{D_i=d\}\Big|X=0^{\pm}\Big], \tag{B.2}$$

and we estimate it by the local cubic polynomial regression

$$\breve{\mu}'(0^{\pm},y,d)a_n = e_1^{\top}\arg\min_{\alpha\in\mathbb{R}^4}\sum_{i=1}^{n}\Big[\frac{1}{b_n}K\Big(\frac{Y_i-y}{b_n}\Big)\mathbb{1}\{D_i=d\} - r_3^{\top}\Big(\frac{X_i}{a_n}\Big)\alpha\Big]^2 K\Big(\frac{X_i}{a_n}\Big)\delta_i^{\pm}. \tag{B.3}$$

This estimate $\breve{\mu}'(0^{\pm},y,d)$ is used for (B.1). Therefore, $\widehat{f}_{Y^d|VX}(y|h(0),0)$ is estimated by

$$\hat{f}_{Y^d|VX}(y|h(0),0) = \frac{\breve{\mu}'(0^{+},y,d)-\breve{\mu}'(0^{-},y,d)}{\hat{\mu}_2'(0^{+},d)-\hat{\mu}_2'(0^{-},d)}.$$

We make the following assumption about the bandwidth parameters a_n and b_n.

Assumption B.1. The bandwidth parameters a_n and b_n satisfy $a_n\to 0$, $b_n\to 0$, $na_n\to\infty$, and $na_n^2b_n^2\to\infty$ and $\frac{b_n}{a_n}\to 0$.

The following lemma shows that the first-order derivative of the kernel regularization (B.2) with respect to x are equivalent to the objects (B.1) of interest. We may thus use the estimates of $\frac{\partial}{\partial x}\mu(0^{\pm},y,d)$ to approximate $\frac{\partial}{\partial y}\mu_1'(0^{\pm},y,d)$.

LEMMA B.1. *Let Assumptions A.1(i) (b), (ii) (a) (b), (iv) (a), and B.1 hold. For each* $(y,d,x)\in\mathscr{Y}\times\mathscr{D}\times([\underline{x},\bar{x}]\setminus\{0\})$, $\frac{\partial}{\partial x}\mu(0^{\pm},y,d) = \frac{\partial}{\partial y}\mu_1'(0^{\pm},y,d)$.

A proof is provided in Appendix C.3 in the online supplementary material available at Cambridge Journals Online (journals.cambridge.org/ect). To show the uniform consistency of $\widehat{f}_{Y^d|VX}(\cdot|h(0),0)$, $d\in\{0,1\}$, it suffices to show $\sup_{(y,d)\in\mathscr{Y}\times\mathscr{D}}|\tilde{\mu}'(0^{\pm},y,d)-\mu'(0^{\pm},y,d)|\xrightarrow[x\times\xi]{p}0$. The following lemma establishes this point.

LEMMA B.2. *Under Assumptions A.1(i), (ii) (a) (b), (iv) (a) (b), and B.1, it holds that*

$$\sup_{(y,d)\in\mathscr{Y}\times\mathscr{D}}|\tilde{\mu}'(0^{\pm},y,d)-\mu'(0^{\pm},y,d)|\xrightarrow[x\times\xi]{p}0.$$

A proof is provided in Appendix C.4 in the online supplementary material available at Cambridge Journals Online (journals.cambridge.org/ect).

B.2. First-Stage Estimators

We will now give some examples of uniformly consistent estimators that satisfy the high-level condition in Assumption A.1(v). First, the density function of X can be estimated by

https://doi.org/10.1017/S0266466619000409 Published online by Cambridge University Press

$$\hat{f}_X(0) = \frac{1}{nc_n}\sum_{i=1}^{n} K(X_i/c_n).$$

This can be shown to be consistent if $c_n \to 0$ and $nc_n \to \infty$, f_X is three-time differentiable and $\frac{\partial^2}{\partial x^2} f_X(0) < \infty$—see Thm. 1.1 of Li and Racine (2008).

We now propose first-stage estimators $\tilde{\mu}_1(x,y,d)\mathbb{1}\{|x/h_n| \le 1\}$ and $\tilde{\mu}_2(x,d)\mathbb{1}\{|x/h_n| \le 1\}$ that are used in the EMP. Denote $\delta_x^+ = \mathbb{1}\{x \ge 0\}$ and $\delta_x^- = \mathbb{1}\{x < 0\}$. We reuse the local cubic estimates from equations (3.2) and (3.3) without requiring to solve an additional optimization problem. We define the first-stage estimators by

$$\begin{aligned}\tilde{\mu}_1(x,y,d) =& \Big[\widehat{\mu}_1(0^+,y,d) + \widehat{\mu}_1'(0^+,y,d)x + \widehat{\mu}_1''(0^+,y,d)\frac{x^2}{2} + \widehat{\mu}_1'''(0^+,y,d)\frac{x^3}{3!}\Big]\delta_x^+ \\ &+ \Big[\widehat{\mu}_1(0^-,y,d) + \widehat{\mu}_1'(0^-,y,d)x + \widehat{\mu}_1''(0^-,y,d)\frac{x^2}{2} + \widehat{\mu}_1'''(0^-,y,d)\frac{x^3}{3!}\Big]\delta_x^- \qquad \text{and} \\ \tilde{\mu}_2(x,d) =& \Big[\widehat{\mu}_2(0^+,d) + \widehat{\mu}_2'(0^+,d)x + \widehat{\mu}_2''(0^+,d)\frac{x^2}{2} + \widehat{\mu}_2'''(0^+,d)\frac{x^3}{3!}\Big]\delta_x^+ \\ &+ \Big[\widehat{\mu}_2(0^-,d) + \widehat{\mu}_2'(0^-,d)x + \widehat{\mu}_2''(0^-,d)\frac{x^2}{2} + \widehat{\mu}_2'''(0^-,d)\frac{x^3}{3!}\Big]\delta_x^-,\end{aligned}$$

where

$$\begin{aligned}&\Big[\widehat{\mu}_1(0^\pm,y,d), \widehat{\mu}_1'(0^\pm,y,d)h_n, \widehat{\mu}_1''(0^\pm,y,d)h_n^2/2!, \widehat{\mu}_1'''(0^\pm,y,d)h_n^3/3!\Big]^\top \\ &= \arg\min_{\alpha\in\mathbb{R}^4} \sum_{i=1}^{n}\Big[\mathbb{1}\{Y_i \le y\}\mathbb{1}\{D_i = d\} - r_3^\top\Big(\frac{X_i}{h_n}\Big)\alpha\Big]^2 K\Big(\frac{X_i}{h_n}\Big)\delta_i^\pm \\ &\Big[\widehat{\mu}_2(0^\pm,d), \widehat{\mu}_2'(0^\pm,d)h_n, \widehat{\mu}_2''(0^\pm,d)h_n^2/2!, \widehat{\mu}_2'''(0^\pm,d)h_n^3/3!\Big]^\top \\ &= \arg\min_{\alpha\in\mathbb{R}^4} \sum_{i=1}^{n}\Big[\mathbb{1}\{D_i = d\} - r_3^\top\Big(\frac{X_i}{h_n}\Big)\alpha\Big]^2 K\Big(\frac{X_i}{h_n}\Big)\delta_i^\pm.\end{aligned}$$

The uniform consistency of these first-stage estimators, required as the high-level condition in Assumption A.1(v), follows from Lem. 7 of Chiang et al. (2019), which is applicable under our Assumption A.1(i)–(iv).

B.3. Bandwidths

Another practical consideration is about a rule for selecting bandwidths in finite samples. We propose to start with the MSE-optimal bandwidths for local quadratic kernel smoothers, and then to apply the rule-of-thumb (ROT) correction for coverage optimality. This procedure is rationalized by the series of studies by Calonico et al. (2018b, 2019)—see also Calonico et al. (2018a). Calonico et al. (2019) develop a new uniform Edgeworth expansion framework to study the optimal worst-case coverage error for confidence intervals for local polynomial regressions. Based on this framework, Calonico et al. (2018b) further extend the results to study coverage optimal bandwidths for regression discontinuity estimands and provide an easy-to-implement ROT correction for MSE optimal bandwidths. Here, we follow the ROT correction from Sect. 4.1 of Calonico et al. (2018b) to obtain our main

https://doi.org/10.1017/S0266466619000409 Published online by Cambridge University Press

bandwidth for the case of robust bias-correction with $\rho = h/b = 1$—see p. 10 of Calonico et al. (2018b).

To keep the implementation simple, we use a single bandwidth h_n that is based on minimizing the sum of MSEs of $\overline{\mu}_1'(0^+,y,1) - \mu_1'(0^-,y,1)$ and $\overline{\mu}_1'(0^+,y,0) - \mu_1'(0^-,y,0)$, where both $\overline{\mu}_1'(0^+,y,1)$ and $\overline{\mu}_1'(0^+,y,0)$ are from local quadratic estimation problems. We first introduce shorthand notations. Let $\Psi^{\pm} = \int_{\mathbb{R}_{\pm}} r_3(u) r_3^{\top}(u) K^2(u) du$ and $\Lambda^{\pm} = \int_{\mathbb{R}_{\pm}} u^2 r_3(u) K(u) du$.

For the kernel density estimator $\hat{f}_X(0)$, we make use of Silverman's rule-of-thumb

$$c_n = 1.06 \hat{\sigma}_X n^{-1/5},$$

where $\hat{\sigma}_X$ is the sample standard deviation of $\{X_i\}_{i=1}^n$.

For the main bandwidth h_n, we first choose

$$h_{0,n}(y,d) = \left(\frac{1}{2} \frac{V_0(y,d)}{B_0^2(y,d)} \right)^{1/7} n^{-1/7},$$

where the leading bias and variance terms are given by

$$B_0(y,d) = e_1^{\top} \left[\frac{(\Gamma^+)^{-1}\Lambda^+}{3!} \overline{\overline{\mu}}_+'''(y,d) - \frac{(\Gamma^-)^{-1}\Lambda^-}{3!} \overline{\overline{\mu}}_-'''(y,d) \right] \quad \text{and}$$

$$V_0(y,d) = \frac{e_1^{\top} [\overline{\overline{\sigma}}_+^2(y,d)(\Gamma^+)^{-1}\Psi^+(\Gamma^+)^{-1} + \overline{\overline{\sigma}}_-^2(y,d)(\Gamma^-)^{-1}\Psi^-(\Gamma^-)^{-1}] e_1}{\hat{f}_X(0)},$$

respectively, with $\overline{\overline{\mu}}_{\pm}'''(y,d)$ and $\overline{\overline{\sigma}}_{\pm}^2(y,d)$ given by global cubic parametric regressions of $\mu_1'''(x,y,d)\delta_x^{\pm}$ and $\sigma^2(y,d|x)\delta_x^{\pm}$, respectively, evaluated at $0^{\pm}$. One can either simply consider to obtain this preliminary bandwidth evaluated at certain (y,d), or consider to obtain one for each (y,d). In either case, we will hereafter suppress the notation of its dependence on (y,d), and will simply write $h_{0,n}$.

With the first-stage bandwidth $h_{0,n}$ having been selected, we can solve

$$\left[\check{\mu}_1(0^{\pm},y,d), \check{\mu}_1'(0^{\pm},y,d) h_{0,n}, \check{\mu}_1''(0^{\pm},y,d) h_{0,n}^2/2!, \check{\mu}_1'''(0^{\pm},y,d) h_{0,n}^3/3! \right]^{\top} =$$

$$\arg\min_{\alpha \in \mathbb{R}^4} \sum_{i=1}^n \left[\mathbb{1}\{Y_i \le y\}\mathbb{1}\{D_i = d\} - r_3^{\top}\left(\frac{X_i}{h_{0,n}}\right)\alpha \right]^2 K(X_i/h_{0,n})\delta_i^{\pm},$$

and thus compute our first-stage level estimate

$$\check{\mu}_1(x,y,d) = \left[\check{\mu}_1(0^+,y,d) + \check{\mu}_1'(0^+,y,d)x + \check{\mu}_1''(0^+,y,d)\frac{x^2}{2} + \check{\mu}_1'''(0^+,y,d)\frac{x^3}{3!} \right]\delta_x^+$$
$$+ \left[\check{\mu}_1(0^-,y,d) + \check{\mu}_1'(0^-,y,d)x + \check{\mu}_1''(0^-,y,d)\frac{x^2}{2} + \check{\mu}_1'''(0^-,y,d)\frac{x^3}{3!} \right]\delta_x^-.$$

We next define the variance estimator by

$$\hat{\sigma}(y,d|0^{\pm}) = \left(\frac{\sum_{i=1}^n (\mathbb{1}\{Y_i \le y, D_i = d\} - \check{\mu}_1(X_i,y,d))^2 K(\frac{X_i}{h_{0,n}})\delta_i^{\pm}}{\sum_{i=1}^n K(\frac{X_i}{h_{0,n}})\delta_i^{\pm}} \right)^{1/2},$$

where $\check{\mu}_1(\cdot,y,d)$ is the first-stage level estimator given above.

https://doi.org/10.1017/S0266466619000409 Published online by Cambridge University Press

Finally, the main bandwidth selector h_n is defined by

$$h_n(y,d) = \left(\frac{1}{2}\frac{V(y,d)}{B^2(y,d)}\right)^{1/7} n^{-1/7},$$

where the leading bias and variance terms are given by

$$B(y,d) = e_1^\top [\frac{(\Gamma^+)^{-1}\Lambda^+}{3!}\breve{\mu}_1'''(0^+,y,d) - \frac{(\Gamma^-)^{-1}\Lambda^-}{3!}\breve{\mu}_1'''(0^-,y,d)] \qquad \text{and}$$

$$V(y,d) = \frac{e_1^\top [\hat{\sigma}(y,d|0^+)(\Gamma^+)^{-1}\Psi^+(\Gamma^+)^{-1} + \hat{\sigma}(y,d|0^-)(\Gamma^-)^{-1}\Psi^-(\Gamma^-)^{-1}]e_1}{\hat{f}_X(0)}.$$

Similarly to the preliminary bandwidth, one can either simply consider to obtain this preliminary bandwidth evaluated at certain (y,d), or consider to obtain one for each (y,d). In either case, we will hereafter suppress the notation of its dependence on (y,d), and will simply write h_n.

For coverage optimality, we can further modify h_n by the ROT correction $h_n^{ROT} = n^{-2/35}h_n$.

For the bandwidth parameters a_n and b_n used for the conditional density estimator $\widehat{f}_{Y^d|VX}(y|h(0),0)$ in Appendix B.1, we follow the choice rules proposed in the end of Appendix C in Frandsen et al. (2012), and propose to set $a_n = h_n$ and $b_n = h_n^2$.

https://doi.org/10.1017/S0266466619000409 Published online by Cambridge University Press