Robust uniform inference for quantile treatment effects in regression discontinuity designs☆

Harold D. Chiang [a], Yu-Chin Hsu [b,c,d], Yuya Sasaki [a,*]

[a] *Department of Economics, Vanderbilt University, VU Station B #351819, 2301 Vanderbilt Place, Nashville, TN 37235-1819, USA*
[b] *Institute of Economics, Academia Sinica, 128 Academia Road, Section 2, Nankang, Taipei, 115, Taiwan*
[c] *Department of Finance, National Central University, NO.300 Jung-da Road, Jung-li City, Taoyuan, 320, Taiwan*
[d] *Department of Economics, National Chengchi University, 64, Chih-nan Road, Sec. 2, Wenshan, Taipei, 11623, Taiwan*

ARTICLE INFO

ABSTRACT

The practical importance of inference with robustness against large bandwidths for causal effects in regression discontinuity and kink designs is widely recognized. Existing robust methods cover many cases, but do not handle uniform inference for CDF and quantile processes in fuzzy designs. In this light, this paper extends the literature by developing a unified framework of inference with robustness against large bandwidths that applies to uniform inference for quantile treatment effects in fuzzy designs, as well as all the other cases. We present Monte Carlo simulation studies and an empirical application for evaluations of the Oklahoma pre-K program.

© 2019 Elsevier B.V. All rights reserved.

1. Introduction

Empirical researchers have used various versions of local Wald estimators. In widest use are the local Wald estimators for the regression discontinuity design (RDD). More recently, researchers have also used local Wald ratios of derivative estimators for the regression kink design (RKD). Furthermore, local Wald ratios of conditional cumulative distribution functions and their variants are used for estimation of quantile treatment effects. In all of these variants of local Wald estimators, researchers often choose large bandwidths by alternative data-driven selectors in practice. As such, ideal estimators and inference procedures need to be robust against large bandwidths.

Proposal of inference methods which are robust against large bandwidths has occurred relatively recently in the literature, given the history of local Wald estimators. The framework of Calonico et al. (2014) covers robust point-wise inference for mean effects in such designs as the sharp mean RDD, fuzzy mean RDD, sharp mean RKD, and fuzzy mean RKD. The robust wild bootstrap method proposed by Bartalotti et al. (2017) covers the sharp mean RDD. The pivotal method proposed by Qu and Yoon (2018) covers robust uniform inference for the sharp quantile RDD, and is also extensible to the sharp quantile RKD (Chiang and Sasaki, 2019).

To our knowledge, these methods proposed in the existing literature do not cover another important case, namely robust uniform inference[1] for quantile treatment effects in the fuzzy RDD (Frandsen et al., 2012), despite its frequent use in the recent literature in applied economics, e.g., Clark and Martorell (2014) and Deshpande (2016), to list a few.[2] In this light, this paper proposes a new general robust inference method and construction of bias-corrected uniform confidence bands that cover the fuzzy quantile RDD in particular. Instead of proposing a robust inference method which specifically applies to the fuzzy quantile RDD, however, we propose one generic framework that uniformly applies to most, if not all, versions of the local Wald estimators including the sharp mean RDD, the fuzzy mean RDD, the sharp mean RKD, the fuzzy mean RKD, the sharp quantile RDD, the fuzzy quantile RDD, the sharp quantile RKD, and the fuzzy quantile RKD, to list a few most popular examples used in empirical research. We focus on the case of the fuzzy quantile RDD for most parts of this paper, as the applicability to this particular case is new in the literature.

After assessing the performance of the proposed method through Monte Carlo simulation studies, we apply it to real data and study causal effects on test outcomes of the Oklahoma pre-K program, following the earlier work by Gormley Jr. et al. (2005) and Frandsen et al. (2012). While Frandsen et al. (2012) provide point-wise confidence intervals for the quantile treatment effects in this application, we follow up and complement their earlier analysis by further providing uniform confidence bands with robustness against large data-driven bandwidths.

The rest of this paper is organized as follows. Section 2 discusses the related literature. Section 3 provides an overview of the method. Section 4 presents the main theoretical results. Section 5 demonstrates Monte Carlo simulation studies. Section 6 presents an empirical illustration. Section 7 presents extended results, including cluster robust inference (Section 7.1), inference with robustness against no or weak jumps (Section 7.2), and inference for models with covariates (Section 7.3). Section 8 concludes. All mathematical proofs and additional details are delegated to the appendix.

2. Relation to the literature

In this section, we overview the most relevant parts of the existing literature. Because of the breadth of the related literature, what we write below is far from being exhaustive.

Literature on Local Designs: The idea of the RDD is introduced by Thistlethwaite and Campbell (1960). There is a vast literature on the RDD. Instead of enumerating all papers, we refer the readers to a seminal paper by Hahn et al. (2001) and a collection of surveys, including Cook (2008) contained in the special issue of *Journal of Econometrics* edited by Imbens and Lemieux (2008), Imbens and Wooldridge (2009, Section 6.4), Lee and Lemieux (2010), and Volume 38 of *Advances in Econometrics* edited by Cattaneo and Escanciano (2017), as well as the references cited therein. For technical matters, we mainly refer to Porter (2003) in deriving the general Bahadur representation for higher-order local polynomial mean regression. While it mostly evolved around the RDD, recent additions to this local design literature include the RKD (e.g., Nielsen et al., 2010; Chen and Fan, 2011; Landais, 2015; Simonsen et al., 2016; Card et al., 2015; Dong, 2016), quantile extensions (e.g., Frandsen et al., 2012; Qu and Yoon, 2018), and their combination (Chiang and Sasaki, 2019). While we focus on the fuzzy quantile RDD for most parts of this paper, we note that all these different frameworks are uniformly encompassed by the general framework developed in this paper.

Literature on Robust Inference: Calonico et al. (2014) introduce bias correction to achieve the robustness of asymptotic inference against large bandwidths. This innovation paves the way for empirical practitioners to obtain valid standard errors for their estimates under popular data-driven methods of bandwidths.[3] Bartalotti et al. (2017) adapt this idea of bias correction to a wild bootstrap method of inference for the sharp mean RDD. Qu and Yoon (2018) adapt this idea of bias correction to a simulation method of uniform inference for the sharp quantile RDD.

Our approach is closely related to Calonico et al. (2014), Qu and Yoon (2015, 2018), and Bartalotti et al. (2017). Calonico et al. (2014) analytically develop the asymptotic distribution accounting for effects of bias estimation. We also analytically develop the limit processes for CDF and quantile processes accounting for effects of bias estimation, which can be seen as a uniform extension to the analytic asymptotic distribution of Calonico et al. (2014). Bartalotti et al. (2017) come up with the idea of approximating the asymptotic distribution of Calonico et al. (2014) via wild bootstrap. We propose to

---

[1] Throughout this paper, we use the phrase "uniform inference" to refer to inference based on weak convergence in the uniform normed linear space ℓ^∞.

[2] We also refer readers to Shigeoka (2014), Ito (2015), Deshpande (2016), and Bernal et al. (2017) for empirical applications with quantile treatment effects in the sharp RDD.

[3] Examples include Imbens and Kalyanaraman (2012), Calonico et al. (2014), Arai and Ichimura (2016), Calonico et al. (2016, 2018a), and Arai and Ichimura (2018). Calonico et al. (2016, 2018a) propose a coverage-probability optimal bandwidth selector and provide a rule of thumb adjustment method to convert MSE-optimal bandwidths into the coverage-probability optimal ones.

approximate the limit process by the multiplier bootstrap, analogously to the wild bootstrap approximation of Bartalotti et al. (2017). Qu and Yoon (2015, 2018) analytically develop the limit processes for local quantile processes via higher-order local polynomials, which effectively account for effects of bias estimation (Calonico et al., 2014, Remark 7). We also analytically develop the limit process for a more general classes of local Wald estimators, which can be seen as a generalization and an extension to Qu and Yoon (2015, 2018). While Qu and Yoon (2015, 2018) – Chiang and Sasaki (2019) likewise – propose a simulation method of approximating the limits of conditional quantile processes for sharp designs by exploiting a pivotal property of quantile regression, we propose to approximate the limit processes via the multiplier bootstrap because it applies to fuzzy designs as well where the joint process of the numerator and the denominator in the local Wald ratio are concerned. In summary, our contribution relies on the ideas developed in these three previous papers.

Literature on Uniform Bahadur Representation: The Bahadur representation is a key to asymptotic distributional results. To cover uniform inference for local polynomial estimators over a general index set with bias correction of any arbitrary order, uniform validity of Bahadur representation over the set is essential. For classes of nonparametric kernel regressions on which our method relies, Masry (1996), Kong et al. (2010) and Fan and Liu (2016) develop uniform Bahadur representations over regressors. Furthermore, Guerre and Sabbah (2012), Qu and Yoon (2015), Lee et al. (2015), Fan and Guerre (2016) develop uniform validity over quantiles as well. We take advantage of this existing idea. In order to deal with a general class of complexity, we use a new maximal inequality (van der Vaart and Wellner, 2011; Chernozhukov et al., 2014).

Literature on Multiplier Bootstrap: In the broad literature, the multiplier bootstrap for Donsker and other weakly convergent classes is first studied by Ledoux and Talagrand (1988) and Giné and Zinn (1990). To our knowledge, use of the multiplier bootstrap in econometrics dates back to Hansen (1996). The multiplier bootstrap for different parametric models has been extensively studied in the literature − it is sometimes referred to as the score bootstrap. For nonparametric CDF estimators, Barrett and Donald (2003) and Donald et al. (2012) use the multiplier bootstrap for uniform inference on unconditional and conditional CDFs, respectively, using the exact solution of their estimators. Chernozhukov et al. (2014) demonstrate the validity of the multiplier bootstrap for inference on suprema of certain non-Donsker processes without using an extreme value limit distribution. Due to the unique nature of our local Wald estimators, our results are based on a multiplier central limit theorem developed more lately by Kosorok (2003, 2008), along with our uniform Bahadur representation.

3. An overview

In this section, we present an overview of the main result, focusing on the case of the fuzzy quantile RDD, which has not been covered by the existing literature of robust inference yet despite its use in the recent literature on empirical microeconomics, such as Clark and Martorell (2014) and Deshpande (2016). A formal and general treatment will follow in Section 4.

Suppose that we observe a random sample of (Y^*, D^*, X), where X is the running variable or the forcing variable, D^* is the binary treatment indicator, and Y^* is the outcome of interest. A researcher faces a fuzzy regression discontinuity design where the cutoff location is normalized to $X = 0$ without loss of generality. Frandsen et al. (2012) identify the conditional CDF of the potential outcome Y_i^d under each treatment status $d \in \{0, 1\}$ given the event C of compliance locally at $X = 0$ by

$$F_{Y^d|C}(y) = \frac{\lim_{x\downarrow 0} E[\mathbb{1}\{Y_i^* \le y\} \cdot \mathbb{1}\{D_i^* = d\}|X_i = x] - \lim_{x\uparrow 0} E[\mathbb{1}\{Y_i^* \le y\} \cdot \mathbb{1}\{D_i^* = d\}|X_i = x]}{\lim_{x\downarrow 0} E[\mathbb{1}\{D_i^* = d\}|X_i = x] - \lim_{x\uparrow 0} E[\mathbb{1}\{D_i^* = d\}|X_i = x]}, \tag{3.1}$$

where we omit the conditioning argument $X = 0$ from our notation and thus $F_{Y^d|C}$ succinctly denotes $F_{Y^d|C,X=0}$. Consequently, the local θ-th quantile treatment effect is identified by

$$\tau(\theta) =: Q_{Y^1|C}(\theta) - Q_{Y^0|C}(\theta); \qquad \text{where } Q_{Y^d|C}(\theta) := \inf\{y : F_{Y^d|C}(y) \ge \theta\}, \tag{3.2}$$

where we again note that the conditioning argument $X = 0$ is omitted from our notations. Frandsen et al. (2012) develop methods of inference for τ based on the exact solutions of local linear estimation of the components of the local Wald ratio (3.1).

In order to make an inference with the local linear estimation, one would need to choose an under-smoothing bandwidth parameter h_n. However, commonly available procedures choose rather large bandwidths, e.g., $h_n \propto n^{-1/5}$. To accommodate these common procedures in the framework of Frandsen et al. (2012), we need to estimate higher-order bias and to develop the limit process accounting for this bias estimation, as in Calonico et al. (2014). In this section, we present how to make uniform inference for τ specifically based on local quadratic estimation of the components of the local Wald ratio (3.1), effectively accounting for the second-order bias estimation − see Remark 7 of Calonico et al. (2014). Consequently, the uniform inference turns robust against large bandwidths as in the commonly available procedures, e.g., $h_n \propto n^{-1/5}$.

The right-hand limit, $\lim_{x\downarrow 0} E[\mathbb{1}\{Y_i^* \leq y\} \cdot \mathbb{1}\{D_i^* = d\}|X_i = x]$, in the numerator of the local Wald ratio (3.1) can be estimated by $\hat{\mu}_1(0^+, y, d)$ in the local quadratic estimator

$$\left(\hat{\mu}_1(0^+, y, d), \hat{\mu}_1'(0^+, y, d), \hat{\mu}_1''(0^+, y, d)\right) =$$

$$\arg\min_{(\mu,\mu',\mu'')} \sum_{i:X_i>0} \left(\mathbb{1}\left\{Y_i^* \leq y\right\} \cdot \mathbb{1}\left\{D_i^* = d\right\} - \left\{\mu + \mu'X_i + \frac{\mu''}{2!}X_i^2\right\}\right)^2 \cdot K\left(\frac{X_i}{h_n}\right),$$

where K is a kernel function and h_n is a bandwidth parameter. The left-hand limit, $\lim_{x\uparrow 0} E[\mathbb{1}\{Y_i^* \leq y\} \cdot \mathbb{1}\{D_i^* = d\}|X_i = x]$, in the numerator of the local Wald ratio (3.1) can be similarly estimated by $\hat{\mu}_1(0^-, y, d)$ using the observations $\{i : X_i < 0\}$. Likewise, the right-hand limit, $\lim_{x\downarrow 0} E[\mathbb{1}\{D_i^* = d\}|X_i = x]$, in the denominator of (3.1) can be estimated by $\hat{\mu}_2(0^+, d)$ in the local quadratic estimator

$$\left(\hat{\mu}_2(0^+, d), \hat{\mu}_2'(0^+, d), \hat{\mu}_2''(0^+, d)\right) =$$

$$\arg\min_{(\mu,\mu',\mu'')} \sum_{i:X_i>0} \left(\mathbb{1}\left\{D_i^* = d\right\} - \left\{\mu + \mu'X_i + \frac{\mu''}{2!}X_i^2\right\}\right)^2 \cdot K\left(\frac{X_i}{h_n}\right),$$

where we can use the same bandwidth h_n as above for simplicity here. The left-hand limit, $\lim_{x\uparrow 0} E[\mathbb{1}\{D_i^* = d\}|X_i = x]$, in the denominator of the local Wald ratio (3.1) can be similarly estimated by $\hat{\mu}_2(0^-, d)$ using the observations $\{i : X_i < 0\}$. With these component estimates, the estimand (3.2) for the identified local quantile treatment effect may be estimated by

$$\hat{\tau}(\theta) = \hat{Q}_{Y^1|C}(\theta) - \hat{Q}_{Y^0|C}(\theta),$$

where

$$\hat{Q}_{Y^d|C}(\theta) = \inf\left\{y : \frac{\hat{\mu}_1(0^+, y, d) - \hat{\mu}_1(0^-, y, d)}{\hat{\mu}_2(0^+, d) - \hat{\mu}_2(0^-, d)} \geq \theta\right\} \quad \text{for each } d \in \{0, 1\}.$$

Under suitable conditions, there exists a zero mean Gaussian process $\mathbb{G}'$ such that

$$\sqrt{nh_n}[\hat{\tau}(\cdot) - \tau(\cdot)] \rightsquigarrow \mathbb{G}'(\cdot) \quad \text{as } n \to \infty. \tag{3.3}$$

See Corollary 1(i) ahead for formal and general arguments. This result establishes the asymptotic distribution result for the fuzzy quantile RDD process with robustness against large bandwidth, e.g., $h_n \propto n^{-1/5}$. In practice, it will be somewhat easier to approximate the limit process $\mathbb{G}'$ by the multiplier bootstrap procedure outlined below.

Let $\{\xi_i\}_{i=1}^n$ be a random sample drawn from the standard normal distribution independently from the data $\{(Y_i^*, D_i^*, X_i)\}_{i=1}^n$. Letting $\Gamma_2^{\pm} = \int_{\mathbb{R}_{\pm}}(1\ u\ u^2)' \cdot K(u) \cdot (1\ u\ u^2)du$ be a 3×3 matrix, we define the estimated multiplier processes for $\hat{\mu}_1(0^+, y, d_1)$ and $\hat{\mu}_2(0^+, d_2)$ by

$$\hat{v}_{\xi,n}^+(y, d, 1) = \sum_{i:X_i>0} \xi_i \frac{(1\ 0\ 0) \cdot (\Gamma_2^{\pm})^{-1} \cdot \left(1\ \frac{X_i}{h_n}\ \frac{X_i^2}{h_n^2}\right)' [\mathbb{1}\{Y_i^* \leq y, D_i^* = d\} - \tilde{\mu}_1(X_i, y, d)]K\left(\frac{X_i}{h_n}\right)}{\sqrt{nh_n}\hat{f}_X(0)}$$

$$\text{and} \quad \hat{v}_{\xi,n}^+(d, 2) = \sum_{i:X_i>0} \xi_i \frac{(1\ 0\ 0) \cdot (\Gamma_2^{\pm})^{-1} \cdot \left(1\ \frac{X_i}{h_n}\ \frac{X_i^2}{h_n^2}\right)' [\mathbb{1}\{D_i^* = d\} - \tilde{\mu}_2(X_i, d)]K\left(\frac{X_i}{h_n}\right)}{\sqrt{nh_n}\hat{f}_X(0)},$$

respectively, where $\hat{f}_X(0)$ estimates $f_X(0)$, $\tilde{\mu}_1(X_i, y, d)$ estimates $E\left[\mathbb{1}\left\{Y_i^* \leq y\right\} \cdot \mathbb{1}\left\{D_i^* = d\right\} |X_i\right]$, and $\tilde{\mu}_2(X_i, d)$ estimates $E\left[\mathbb{1}\left\{D_i^* = d\right\} |X_i\right]$. Specifically, one can choose any consistent kernel density estimator for $\hat{f}_X(0)$, and concrete examples of the estimators, $\tilde{\mu}_1(X_i, y, d)$ and $\tilde{\mu}_2(X_i, d)$, are provided in Appendix A.6. The estimated multiplier processes for $\hat{\mu}_1(0^-, y, d_1)$ and $\hat{\mu}_2(0^-, d_2)$ are similarly defined by $\hat{v}_{\xi,n}^-(y, d, 1)$ and $\hat{v}_{\xi,n}^-(d, 2)$ using the observations $\{i : X_i < 0\}$. By the Hadamard derivative, we construct the approximate estimated multiplier process

$$\widehat{\mathbb{G}}_n'(\theta) = \frac{\hat{\mu}_2(0^+, 1) - \hat{\mu}_2(0^-, 1)}{\hat{f}_{Y^1|C}(\hat{Q}_{Y^1|C}(\theta))[\hat{\mu}_2(0^+, 1) - \hat{\mu}_2(0^-, 1)]^2} \cdot \left\{\hat{v}_{\xi,n}^+(\hat{Q}_{Y^1|C}(\theta), 1, 1) - \hat{v}_{\xi,n}^-(\hat{Q}_{Y^1|C}(\theta), 1, 1)\right\}$$

$$- \frac{\hat{\mu}_1(0^+, \hat{Q}_{Y^1|C}(\theta), 1) - \hat{\mu}_1(0^-, \hat{Q}_{Y^1|C}(\theta), 1)}{\hat{f}_{Y^1|C}(\hat{Q}_{Y^1|C}(\theta))[\hat{\mu}_2(0^+, 1) - \hat{\mu}_2(0^-, 1)]^2} \cdot \left\{\hat{v}_{\xi,n}^+(1, 2) - \hat{v}_{\xi,n}^-(1, 2)\right\}$$

$$- \frac{\hat{\mu}_2(0^+, 0) - \hat{\mu}_2(0^-, 0)}{\hat{f}_{Y^0|C}(\hat{Q}_{Y^0|C}(\theta))[\hat{\mu}_2(0^+, 0) - \hat{\mu}_2(0^-, 0)]^2} \cdot \left\{\hat{v}_{\xi,n}^+(\hat{Q}_{Y^0|C}(\theta), 0, 1) - \hat{v}_{\xi,n}^-(\hat{Q}_{Y^0|C}(\theta), 0, 1)\right\}$$

$$+ \frac{\hat{\mu}_1(0^+, \hat{Q}_{Y^0|C}(\theta), 0) - \hat{\mu}_1(0^-, \hat{Q}_{Y^0|C}(\theta), 0)}{\hat{f}_{Y^0|C}(\hat{Q}_{Y^0|C}(\theta))[\hat{\mu}_2(0^+, 0) - \hat{\mu}_2(0^-, 0)]^2} \cdot \left\{\hat{v}_{\xi,n}^+(0, 2) - \hat{v}_{\xi,n}^-(0, 2)\right\}. \tag{3.4}$$

Under suitable conditions, with probability approaching one, this process $\widehat{\mathbb{G}}'_n$ weakly converges to the limit process, $\mathbb{G}'$, of interest conditionally on the data $\{(Y_i^*, D_i^*, X_i)\}_{i=1}^n$, i.e.,

$$\widehat{\mathbb{G}}'_n \overset{p}{\underset{\xi}{\rightsquigarrow}} \mathbb{G}' \quad \text{as } n \to \infty. \tag{3.5}$$

See Corollary 1 (ii) ahead for formal and general arguments. From (3.3) and (3.5), therefore, we may use $\widehat{\mathbb{G}}'_n$ to approximate the limit process of $\sqrt{nh_n}[\hat{\tau} - \tau]$.

One of the most relevant practical applications of this result is to test the null hypothesis of uniform treatment nullity:

$$H_0 : \tau(\theta) = 0 \quad \text{for all } \theta \in [a, 1-a]$$

for some $\alpha \in (0, 1/2)$. To test this hypothesis, we can use $\sup_{\theta \in [a, 1-a]} \sqrt{nh_n} \left| \hat{\tau}(\theta) \right|$ as the test statistic, and use

$$\sup_{\theta \in [a, 1-a]} \left| \widehat{\mathbb{G}}'_n(\theta) \right|$$

to simulate its asymptotic distribution.

Another of the most relevant practical applications of the above corollary is to test the null hypothesis of treatment homogeneity across quantiles:

$$H_0 : \tau(\theta) = \tau(\theta') \quad \text{for all } \theta, \theta' \in [a, 1-a].$$

To test this hypothesis, we can use $\sup_{\theta \in [a, 1-a]} \sqrt{nh_n} \left| \hat{\tau}(\theta) - (1-2a)^{-1} \int_{[a, 1-a]} \hat{\tau}(\vartheta) d\vartheta \right|$ as the test statistic, and use

$$\sup_{\theta \in [a, 1-a]} \left| \widehat{\mathbb{G}}'_n(\theta) - \frac{1}{1-2a} \int_{[a, 1-a]} \widehat{\mathbb{G}}'_n(\vartheta) d\vartheta \right|$$

to simulate its asymptotic distribution.

Finally, we can use the approximate estimated multiplier process to construct uniform confidence bands for the quantile treatment effects. To this end, we compute

$$\hat{\mathcal{C}}_n(a, 1-a; \lambda) = \text{the } (1-\lambda)\text{-th quantile of } \sup_{\theta \in [a, 1-a]} \left| \widehat{\mathbb{G}}'_n(\theta) \right|.$$

The band of the form

$$\left[\hat{\tau}(\theta) \pm \frac{1}{\sqrt{nh_n}} \hat{\mathcal{C}}_n(a, 1-a; \lambda) \ : \ \theta \in [a, 1-a] \right]$$

constitutes a $100(1-\lambda)$ percent uniform confidence band for the quantile treatment effects τ. In summary, we provide a step-by-step procedure below.

Algorithm 1 (*Practical Guideline on Constructing Uniform Confidence Bands*)**.**

Step 1. Pick a finite set $\mathcal{Y}^* \subset \mathcal{Y}$ of grid points of outcome values and a finite set $\mathcal{T}^* \subset [a, 1-a]$ of grid points of quantiles.

Estimate $\hat{\mu}_1(0^\pm, y, d)$ and $\hat{\mu}_2(0^\pm, d)$ for all $y \in \mathcal{Y}^*$, $d \in \{0, 1\}$.

Step 2. Calculate $\hat{Q}_{Y^d|C}(\theta)$ for each $\theta \in \mathcal{T}^*$ by

$$\hat{Q}_{Y^d|C}(\theta) = \inf \left\{ y \in \mathcal{Y}^* : \frac{\hat{\mu}_1(0^+, y, d) - \hat{\mu}_1(0^-, y, d)}{\hat{\mu}_2(0^+, d) - \hat{\mu}_2(0^-, d)} \geq \theta \right\}$$

for $d \in \{0, 1\}$, and then compute $\hat{\tau} = \hat{Q}_{Y^1|C}(\theta) - \hat{Q}_{Y^1|C}(\theta)$ for each $\theta \in \mathcal{T}^*$.

Step 3. Compute $\hat{f}_X(0)$ and $\hat{f}_{Y^d|C}(\hat{Q}_{Y^d|C}(\theta))$ – see Appendix A.6 for formulas.

Step 4. For each bootstrap iteration $b = 1, \ldots, B$, generate independent standard normal $\xi^b = \{\xi_i^b\}_{i=1}^n$ independently from data, and compute $\hat{v}_{\xi^b, n}(\theta, d, 1)$ and $\hat{v}_{\xi^b, n}(d, 2)$ for each $\theta \in \mathcal{T}^*$, $d \in \{0, 1\}$. (Note in each iteration b, we compute these values for different θ and d using the same ξ^b.)

Step 5. Construct $\hat{\mathbb{G}}'_{n,b}$ for each $\theta \in \mathcal{T}^*$:

$$\widehat{\mathbb{G}}'_{n,b}(\theta) = \frac{\hat{\mu}_2(0^+, 1) - \hat{\mu}_2(0^-, 1)}{\hat{f}_{Y^1|C}(\hat{Q}_{Y^1|C}(\theta))[\hat{\mu}_2(0^+, 1) - \hat{\mu}_2(0^-, 1)]^2} \cdot \left\{ \hat{v}_{\xi^b, n}^+(\hat{Q}_{Y^1|C}(\theta), 1, 1) - \hat{v}_{\xi^b, n}^-(\hat{Q}_{Y^1|C}(\theta), 1, 1) \right\}$$

$$- \frac{\hat{\mu}_1(0^+, \hat{Q}_{Y^1|C}(\theta), 1) - \hat{\mu}_1(0^-, \hat{Q}_{Y^1|C}(\theta), 1)}{\hat{f}_{Y^1|C}(\hat{Q}_{Y^1|C}(\theta))[\hat{\mu}_2(0^+, 1) - \hat{\mu}_2(0^-, 1)]^2} \cdot \left\{ \hat{v}_{\xi^b, n}^+(1, 2) - \hat{v}_{\xi^b, n}^-(1, 2) \right\}$$

$$-\frac{\hat{\mu}_2(0^+, 0) - \hat{\mu}_2(0^-, 0)}{\hat{f}_{Y^0|C}(\hat{Q}_{Y^0|C}(\theta))[\hat{\mu}_2(0^+, 0) - \hat{\mu}_2(0^-, 0)]^2} \cdot \left\{ \hat{v}^+_{\xi^b,n}(\hat{Q}_{Y^0|C}(\theta), 0, 1) - \hat{v}^-_{\xi^b,n}(\hat{Q}_{Y^0|C}(\theta), 0, 1) \right\}$$

$$+\frac{\hat{\mu}_1(0^+, \hat{Q}_{Y^0|C}(\theta), 0) - \hat{\mu}_1(0^-, \hat{Q}_{Y^0|C}(\theta), 0)}{\hat{f}_{Y^0|C}(\hat{Q}_{Y^0|C}(\theta))[\hat{\mu}_2(0^+, 0) - \hat{\mu}_2(0^-, 0)]^2} \cdot \left\{ \hat{v}^+_{\xi^b,n}(0, 2) - \hat{v}^-_{\xi^b,n}(0, 2) \right\}.$$

Step 6. Set $\hat{C}^B_n(a, 1 - a; \lambda) = $ the $(1 - \lambda)$-th quantile of $\left\{ \max_{\theta \in \mathcal{T}^*} \left| \widehat{\mathbb{G}}'_{n,b}(\theta) \right| \right\}^B_{b=1}$, and construct an asymptotically valid $100(1 - \lambda)$ percent uniform confidence band over $[a, 1 - a]$ by

$$\left[\hat{\tau}(\theta) \pm \frac{1}{\sqrt{nh_n}} \hat{C}^B_n(a, 1 - a; \lambda) \ : \ \theta \in \mathcal{T}^* \right].$$

Remark 1. Steps 1–5 also give the bootstrapped distribution $\{\widehat{\mathbb{G}}'_{n,b}\}^B_{b=1}$, which can be used to construct critical values for tests of uniform treatment nullity and treatment homogeneity as well. For the null hypothesis of uniform treatment nullity, use $\max_{\theta \in [a,1-a]} \sqrt{nh_n} |\hat{\tau}(\theta)|$ as the test statistic, and use $(1 - \lambda)$-th quantile of $\left\{ \max_{\theta \in [a,1-a]} \left| \widehat{\mathbb{G}}'_{n,b}(\theta) \right| \right\}^B_{b=1}$ as the critical value. For the null hypothesis of treatment homogeneity, let $\max_{\theta \in [a,1-a]} \sqrt{nh_n} \left| \hat{\tau}(\theta) - (1 - 2a)^{-1} \int_{[a,1-a]} \hat{\tau}(\vartheta) d\vartheta \right|$ be the test statistic, and the $(1-\lambda)$-th quantile of $\left\{ \max_{\theta \in [a,1-a]} \left| \widehat{\mathbb{G}}'_{n,b}(\theta) - \frac{1}{1-2a} \int_{[a,1-a]} \widehat{\mathbb{G}}'_{n,b}(\vartheta) d\vartheta \right| \right\}^B_{b=1}$ be the critical value for the test.

Remark 2. In practice we may set $|\mathcal{T}^*| = 50$ and $|\mathcal{Y}^*| = 5000$. We tried a local polynomial mean regression for each of the 5000 grid points of $y \in \mathcal{Y}^*$, and found that it is not computationally burdensome in general, since local polynomial mean regressions are smooth convex problems and gradient decent algorithms can solve them efficiently.

4. A unified framework

In this section, we present a generalized framework for a broad class of local Wald estimands that encompasses not only the case of the fuzzy quantile RDD, but also the cases of the sharp mean RDD, the fuzzy mean RDD, the sharp mean RKD, the fuzzy mean RKD, the sharp CDF discontinuity design, the fuzzy CDF discontinuity design, the sharp quantile RDD, the sharp quantile RKD, and the fuzzy quantile RKD. We revisit the case of the fuzzy quantile RDD in Section 4.5 to provide a formal justification of the overview in Section 3. All the other examples are relegated to Appendices B and C.

4.1. The general framework

Let (Y, D, X) be a random vector defined on a probability space $(\Omega^x, \mathcal{F}^x, \mathbb{P}^x)$, where Y is a random vector containing an outcome and possibly other variables, D is a random vector containing a treatment indicator and possibly others, and X is a running variable or an assignment variable. We denote their supports as $\mathcal{Y}$, $\mathcal{D}$ and $\mathcal{X}$, respectively. Suppose that a researcher observes n i.i.d. copies $\{(Y_i, D_i, X_i)\}^n_{i=1}$ of (Y, D, X). Consider some subsets of some finite dimensional Euclidean spaces Θ_1, Θ_2, Θ'_1, Θ'_2, and Θ''. We will use them to denote sets of indices. Let $\Theta = \Theta_1 \times \Theta_2$, and let $g_1 : \mathcal{Y} \times \Theta_1 \to \mathbb{R}$ and $g_2 : \mathcal{D} \times \Theta_2 \to \mathbb{R}$ be functions to be defined in various contexts of empirical research designs — concrete examples are suggested in the subsections below and in the Supplementary Appendix. When we discuss the continuity of g_k in θ_k, $k \in \{1, 2\}$, we consider Θ_1 and Θ_2 with the topologies they inherit from the finite dimensional Euclidean spaces they reside in. We write $\mu_1(x, \theta_1) = E[g_1(Y_i, \theta_1)|X_i = x]$ and $\mu_2(x, \theta_2) = E[g_2(D_i, \theta_2)|X_i = x]$. Their vth order partial derivatives with respect to x are denoted by $\mu_1^{(v)} = \frac{\partial^v}{\partial x^v} \mu_1$ and $\mu_2^{(v)} = \frac{\partial^v}{\partial x^v} \mu_2$. For a set T, we denote $\mathcal{C}^1(T)$ as the collection of all real-valued functions on T that are continuously differentiable, and $\ell^\infty(T)$ is the collection of all bounded real-valued functions on T. With suitable operators $\phi : \ell^\infty(\Theta_1) \to \ell^\infty(\Theta'_1)$, $\psi : \ell^\infty(\Theta_2) \to \ell^\infty(\Theta'_2)$, and $\Upsilon : \ell^\infty(\Theta'_1 \times \Theta'_2) \to \ell^\infty(\Theta'')$, a general class of local Wald estimands can be expressed in the form of

$$\tau(\theta'') = \Upsilon \left(\frac{\phi \left(\lim_{x \downarrow 0} \mu_1^{(v)}(x, \cdot) \right)(\cdot) - \phi \left(\lim_{x \uparrow 0} \mu_1^{(v)}(x, \cdot) \right)(\cdot)}{\psi \left(\lim_{x \downarrow 0} \mu_2^{(v)}(x, \cdot) \right)(\cdot) - \psi \left(\lim_{x \uparrow 0} \mu_2^{(v)}(x, \cdot) \right)(\cdot)} \right)(\theta''). \tag{4.1}$$

for all $\theta'' \in \Theta''$. This class of local Wald estimands encompasses a wide array of design-based estimands used by empirical practitioners. In all examples, v is either 0 (for RDD) or 1 (for RKD), and setting the order of local polynomial estimator $p = v + 2$ would generally suffice in practice. We list two examples below: Example 1 illustrates the case of the fuzzy mean RDD; Example 2 illustrates the case of the fuzzy quantile RDD for which the existing literature has proposed no robust uniform inference methods — also overviewed in Section 3. See Appendix C for additional examples. For convenience of writing, we introduce the notation for the intermediate local Wald estimand:

$$W = \frac{\phi(\mu_1^{(v)}(0^+, \cdot)) - \phi(\mu_1^{(v)}(0^-, \cdot))}{\psi(\mu_2^{(v)}(0^+, \cdot)) - \psi(\mu_2^{(v)}(0^-, \cdot))}.$$

Example 1 (*Fuzzy Mean RDD*). We do not need index sets for fuzzy mean RDD, so let $\Theta_1 = \Theta_2 = \Theta_1' = \Theta_2' = \Theta'' = \{0\}$ for simplicity. Set $g_1(Y_i, \theta_1) = Y_i$ and $g_2(D_i, \theta_2) = D_i$. Note that $\mu_1(x, \theta_1) = \mathrm{E}[g_1(Y_i, \theta_1)|X_i = x] = \mathrm{E}[Y_i|X_i = x]$ and $\mu_2(x, \theta_2) = \mathrm{E}[g_2(D_i, \theta_2)|X_i = x] = \mathrm{E}[D_i|X_i = x]$. Let ϕ and ψ be the identity operators, and for $W \in \ell^\infty(\Theta_1' \times \Theta_2')$ let the operator Υ be $\Upsilon(W)(\theta'') = W(\theta'', \theta'') \, \forall \theta'' \in \Theta''$. The local Wald estimand (4.1) with $v = 0$ in this setting becomes

$$\tau(\theta'') = \frac{\lim_{x\downarrow 0} \mathrm{E}[Y_i|X_i = x] - \lim_{x\uparrow 0} \mathrm{E}[Y_i|X_i = x]}{\lim_{x\downarrow 0} \mathrm{E}[D_i|X_i = x] - \lim_{x\uparrow 0} \mathrm{E}[D_i|X_i = x]} \tag{4.2}$$

for all $\theta'' \in \Theta'' = \{0\}$. This estimand $\tau(0)$ will be denoted by τ_{FMRD} for Fuzzy Mean RD design. □

Example 2 (*Fuzzy Quantile RDD*). Consider the estimand (3.1) proposed by Frandsen et al. (2012) to identify the conditional CDF of potential outcome Y_i^d under each treatment status $d \in \{0, 1\}$ given the event C of compliance, and the quantile treatment effect (3.2) given this event C. These estimands also fit in the general framework (4.1). We first fix an $a \in (0, 1/2)$, $\varepsilon > 0$ and let $\mathscr{Y}_1 = [Q_{Y^1|C}(a) - \varepsilon, Q_{Y^1|C}(1 - a) + \varepsilon] \cup [Q_{Y^0|C}(a) - \varepsilon, Q_{Y^0|C}(1 - a) + \varepsilon]$. Let $\Theta_1 = \Theta_1' = \mathscr{Y}_1 \times \mathscr{D}$ and $\Theta_2 = \Theta_2' = \mathscr{D}$ for $\mathscr{D} = \{0, 1\}$, and let $\Theta'' = [a, 1 - a]$ for a constant $a \in (0, 1/2)$. Let $Y_i = (Y_i^*, D_i^*)$ and $D_i = D_i^*$. Set $g_1((Y_i^*, D_i^*), (y, d)) = \mathbb{1}\{Y_i^* \le y\} \cdot \mathbb{1}\{D_i^* = d\}$ and $g_2(D_i^*, d) = \mathbb{1}\{D_i^* = d\}$. Note that $\mu_1(x, y, d) = \mathrm{E}[g_1((Y_i^*, D_i^*), y, d)|X_i = x] = \mathrm{E}[\mathbb{1}\{Y_i^* \le y\} \cdot \mathbb{1}\{D_i^* = d\}|X_i = x]$ and $\mu_2(x, d) = \mathrm{E}[g_2(D_i^*, d)|X_i = x] = \mathrm{E}[\mathbb{1}\{D_i^* = d\}|X_i = x]$. Let ϕ and ψ be the identity operators, and define Υ for each $W \in \ell^\infty(\Theta_1' \times \Theta_2') = \ell^\infty(\mathscr{Y} \times \mathscr{D}^2)$ by $\Upsilon(W)(\theta'') = \inf\{y \in \mathscr{Y} : W(y, 1, 1) \ge \theta''\} - \inf\{y \in \mathscr{Y} : W(y, 0, 0) \ge \theta''\}$. The local Wald estimand (4.1) with $v = 0$ in this setting becomes (3.2), i.e.,

$$\tau(\theta'') = Q_{Y^1|C}(\theta'') - Q_{Y^0|C}(\theta'')$$

for all $\theta'' \in \Theta'' = [a, 1 - a]$, where $Q_{Y^d|C}(\theta'') := \inf\{y \in \mathscr{Y} : F_{Y^d|C}(y) \ge \theta''\}$ for a short-hand notation, and $F_{Y^d|C}(y)$ is given in (3.1) for all $(y, d) \in \mathscr{Y} \times \mathscr{D}$. This estimand τ will be denoted by τ_{FQRD} for Fuzzy Quantile RD design. □

We introduce some short-hand notations for conservation of space. Let $\mathcal{E}_k(y, d, x, \theta) = g_k(y, \theta_k) - \mu_k(x, \theta_k)$ for $(\theta_1, \theta_2) \in \Theta$, $k \in \{1, 2\}$, $y \in \mathscr{Y}$, $d \in \mathscr{D}$, and $x \in \mathscr{X}$. Let $\sigma_{kl}(\theta, \vartheta|x) = E[\mathcal{E}_k(Y_i, D_i, X_i, \theta) \cdot \mathcal{E}_l(Y_i, D_i, X_i, \vartheta)|X_i = x]$ denote the conditional covariance of residuals for each $\theta = (\theta_1, \theta_2)$, $\vartheta = (\vartheta_1, \vartheta_2) \in \Theta$, and $k, l \in \{1, 2\}$. Also define the product space $\mathbb{T} = \Theta \times \{1, 2\} = (\Theta_1 \times \Theta_2) \times \{1, 2\}$. We will also use the following short-hand notations for functions at right- and left-hand limits: $\mu_k^{(v)}(0^+, \theta) = \lim_{x\downarrow 0} \mu_k^{(v)}(x, \theta)$ and $\mu_k^{(v)}(0^-, \theta) = \lim_{x\uparrow 0} \mu_k^{(v)}(x, \theta)$ for $k = \{1, 2\}$. The composite notation $\mu_k^{(v)}(0^\pm, \theta)$ is used to collectively refer to $\mu_k^{(v)}(0^+, \theta)$ and $\mu_k^{(v)}(0^-, \theta)$. Let $r_p(x) = (1, x, \ldots, x^p)'$. Let K denote a kernel function, and let $(h_{1,n}(\theta_1), h_{2,n}(\theta_2))$ denote bandwidth parameters that depend on $\theta = (\theta_1, \theta_2) \in \Theta$ and the sample size $n \in \mathbb{N}$. For $p \in \mathbb{N}$, let e_v denote the vth standard basis element of $\mathbb{R}^p$. We write $\Gamma_p^\pm = \int_{\mathbb{R}_\pm} K(u) r_p(u) r_p'(u) du$ and $\Lambda_{p,q}^\pm = \int_{\mathbb{R}_\pm} u^q K(u) r_p(u) du$. We use the notation $\rightsquigarrow$ to denote weak convergence, and the notation $\overset{p}{\underset{\xi}{\rightsquigarrow}}$ for conditional weak convergence as defined in Section 2.2.3 of Kosorok (2008) – see Appendix A.2 for more details. Let $v, p, q \in \mathbb{N}_+$ with $v \le p$. We will use v for the order of derivative of interest as in (4.1), and p stands for the order of local polynomial fitting to estimate (4.1).

4.2. The local wald estimator

In this section, we develop an estimator for the nonparametric components $\mu_k^{(v)}(0^\pm, \cdot)$, $k \in \{1, 2\}$, of the local Wald estimand (4.1) based on local polynomial fitting with the bias correction approach proposed by Calonico et al. (2014) – for a comprehensive treatment for local polynomial models, see Fan and Gijbels (1996). Under proper smoothness assumptions to be formally stated below, the pth order approximations

$$\mu_k(x, \theta_k) \approx \mu_k(0^+, \theta_k) + \mu_k^{(1)}(0^+, \theta_k)x + \cdots + \frac{\mu_k^{(p)}(0^+, \theta_k)}{p!}x^p = r_p(x/h)'\alpha_{k+,p}(\theta_k) \qquad x > 0$$

$$\mu_k(x, \theta_k) \approx \mu_k(0^-, \theta_k) + \mu_k^{(1)}(0^-, \theta_k)x + \cdots + \frac{\mu_k^{(p)}(0^-, \theta_k)}{p!}x^p = r_p(x/h)'\alpha_{k-,p}(\theta_k) \qquad x < 0$$

hold for each $k \in \{1, 2\}$, where $\alpha_{k\pm,p}(\theta) = [\mu_k(0^\pm, \theta_k)/0!, \mu_k^{(1)}(0^\pm, \theta_k)h/1!, \ldots, \mu_k^{(p)}(0^\pm, \theta_k)h^p/p!]$ and $h > 0$. To estimate $\alpha_{k\pm,p}(\theta_k)$, we solve the one-sided local weighted least squares problems

$$\hat{\alpha}_{1\pm,p}(\theta_1) = \underset{\alpha \in \mathbb{R}^{p+1}}{\arg\min} \sum_{i=1}^n \delta_i^\pm \left(g_1(Y_i, \theta_1) - r_p\left(\frac{X_i}{h_{1,n}(\theta_1)}\right)'\alpha\right)^2 K\left(\frac{X_i}{h_{1,n}(\theta_1)}\right) \tag{4.3}$$

$$\hat{\alpha}_{2\pm,p}(\theta_2) = \underset{\alpha \in \mathbb{R}^{p+1}}{\arg\min} \sum_{i=1}^n \delta_i^\pm \left(g_2(D_i, \theta_2) - r_p\left(\frac{X_i}{h_{2,n}(\theta_2)}\right)'\alpha\right)^2 K\left(\frac{X_i}{h_{2,n}(\theta_2)}\right), \tag{4.4}$$

where $\delta_i^+ = \mathbb{1}\{X_i \ge 0\}$ and $\delta_i^- = \mathbb{1}\{X_i \le 0\}$. We let the coordinates of these estimates be written by

$$\hat{\alpha}_{k\pm,p}(\theta_k) = [\hat{\mu}_{k,p}(0^\pm, \theta_k)/0!, \hat{\mu}_{k,p}^{(1)}(0^\pm, \theta_k)h_{k,n}(\theta_k)/1!, \ldots, \hat{\mu}_{k,p}^{(p)}(0^\pm, \theta)h_{k,n}^p(\theta_k)/p!]'.$$

With these component estimates, the local Wald estimand (4.1) is in turn estimated by the plug-in estimator.

$$\hat{\tau}(\theta'') = \Upsilon \left(\frac{\phi\left(\hat{\mu}_{1,p}^{(v)}(0^+, \cdot)\right)(\cdot) - \phi\left(\hat{\mu}_{1,p}^{(v)}(0^-, \cdot)\right)(\cdot)}{\psi\left(\hat{\mu}_{2,p}^{(v)}(0^+, \cdot)\right)(\cdot) - \psi\left(\hat{\mu}_{2,p}^{(v)}(0^-, \cdot)\right)(\cdot)} \right)(\theta'') \qquad \text{for each } \theta'' \in \Theta''. \tag{4.5}$$

4.3. Weak convergence

In this section, we establish the weak convergence result for the process $\sqrt{nh_n^{1+2v}}[\hat{\tau} - \tau]$ for some bandwidth h_n. To this end we state the following set of assumptions.

Assumption 1 (*Uniform Bahadur Representation*). Let $\underline{x} < 0 < \overline{x}$. (i) (a) $\{(Y_i, D_i, X_i)\}_{i=1}^n$ are n i.i.d. copies of random vector (Y, D, X) defined on a probability space $(\Omega^x, \mathcal{F}^x, \mathbb{P}^x)$; (b) X has a density function f_X which is continuously differentiable on $[\underline{x}, \overline{x}]$, and $0 < f_X(0) < \infty$. (ii) For each $k = 1, 2$, (a) the collections of real-valued functions, $\{x \mapsto \mu_k(x, \theta_k) : \theta_k \in \Theta_k\}$, $\{y \mapsto g_1(y, \theta_1) : \theta_1 \in \Theta_1\}$, and $\{d \mapsto g_2(d, \theta_2) : \theta_2 \in \Theta_2\}$, are of VC type with a common integrable envelope $F_{\mathcal{E}}$ such that $\int_{\mathscr{Y} \times \mathscr{D} \times [\underline{x}, \overline{x}]} |F_{\mathcal{E}}(y, d, x)|^{2+\epsilon} d\mathbb{P}^x(y, d, x) < \infty$ for some $\epsilon > 0$; (b) $\mu_k^{(j)}$ is Lipschitz on $[\underline{x}, 0) \times \Theta_k$ and $(0, \overline{x}] \times \Theta_k$ for $j = 0, 1, 2, \ldots, p + 1$; (c) For any $(\theta, k), (\vartheta, l) \in \mathbb{T}$, we have $\sigma_{kl}(\theta, \vartheta \mid \cdot) \in \mathcal{C}^1([\underline{x}, \overline{x}] \setminus \{0\})$ with bounded derivatives in x and $\sigma_{kl}(\theta, \vartheta \mid 0^{\pm}) < \infty$; (d) For each $y \in \mathscr{Y}$, $g_1(y, \cdot)$ is left- or right-continuous in each dimension. Similarly, for each $d \in \mathscr{D}$, $g_2(d, \cdot)$ is left- or right-continuous in each dimension. (iii) There exist bounded Lipschitz functions $c_1 : \Theta_1 \to [\underline{c}, \overline{c}] \subset (0, \infty)$ and $c_2 : \Theta_2 \to [\underline{c}, \overline{c}] \subset (0, \infty)$ such that $h_{1,n}(\theta_1) = c_1(\theta_1)h_n$ and $h_{2,n}(\theta_2) = c_2(\theta_2)h_n$ hold for baseline bandwidth h_n satisfying $h_n \to 0$, $nh_n^2 \to \infty$ and $nh_n^{2p+3} \to 0$ for some $h_0 < \infty$. (iv) (a) $K : [-1, 1] \to \mathbb{R}^+$ is bounded and continuous; (b) $\{K(\cdot/h) : h > 0\}$ is of VC type. (c) $\Gamma_p^{\pm}$ is positive definite.

Condition (i) requires a random sampling of (Y, D, X) and sufficient data around $X = 0$. The i.i.d. condition (i) (a) is shared by most of the prior work on regression discontinuity and kink designs. Exceptions are Bartalotti and Brummet (2017) and Calonico et al. (2018b), which relax the assumption of identical distribution and study cluster-robust inference — we will also present a method of cluster-robust inference in Section 7.1 by extending our baseline results. Versions of the smoothness and nonzero requirements in condition (i) (b) are shared by the assumptions made in the prior work, and relate to the absence of endogenous sorting. This assumption is analogous to Assumption 1.1. in the closely related benchmark paper by Bartalotti et al. (2017). Regarding condition (ii), a sufficient condition for $\{x \mapsto \mu_k(x, \theta_k) : \theta_k \in \Theta_k\}$ to be of VC type class is, for example, the existence of some non-negative function $M_k : \mathscr{X} \to \mathbb{R}_+$ such that $|\mu_k(x, \bar{\theta}_k) - \mu_k(x, \theta_k)| \le M_k(x)|\bar{\theta}_k - \theta_k|$ for all $\bar{\theta}_k, \theta_k \in \Theta_k$ for each $k = 1, 2$. Analogous remarks apply to $\{y \mapsto g_1(y, \theta_1) : \theta_1 \in \Theta_1\}$ and $\{d \mapsto g_2(d, \theta_2) : \theta_2 \in \Theta_2\}$ as well. Another sufficient condition is the case when a class of functions is of variations bounded by one, e.g. in the case of CDF estimation, $\{y \mapsto \mathbb{1}\{y \le y'\} : y' \in \mathscr{Y}\}$ satisfies the VC type condition. Also notice that the common integrable envelope $F_{\mathcal{E}}$ in condition (ii) is satisfied if all the classes of functions are uniformly bounded, but does not rule out some cases that some of these classes of functions are unbounded. In Appendix C, we will check these high-level assumptions with primitive sufficient assumptions for each of the ten specific examples presented in Examples 1–2 and Appendix B.1–B.8. Parts (ii) (b) and (ii) (c) are analogous to Assumption 1.3. and Assumption 1.4., respectively, in Bartalotti et al. (2017). Condition (iii) specifies admissible rates of bandwidths, which are consistent with common choice rules (e.g., Imbens and Kalyanaraman, 2012; Calonico et al., 2014; Arai and Ichimura, 2016; Calonico et al., 2016, 2018a; Arai and Ichimura, 2018; for regression discontinuity designs) – see Appendix F. This condition is analogous to Assumption 2 in Bartalotti et al. (2017) for the special case of $h \sim b$ in their notation. Condition (iv) is satisfied by common kernel functions, such as uniform, triangular, biweight, triweight, and Epanechnikov kernels to list a few examples, while the normal kernel is obviously ruled out.

We will show below the weak convergence in $\ell^\infty(\mathbb{T})$ of the Bahadur Representation (BR)

$$\nu_n^{\pm}(\theta, k) = v! \sum_{i=1}^n \frac{e_v'(\Gamma_p^{\pm})^{-1}\mathcal{E}_k(Y_i, D_i, X_i, \theta)r_p(\frac{X_i}{h_{k,n}(\theta_k)})K(\frac{X_i}{h_{k,n}(\theta_k)})\delta_i^{\pm}}{\sqrt{nh_{k,n}(\theta_k)}f_X(0)} \qquad (\theta, k) \in \mathbb{T}.$$

We also write $\nu_n(\cdot) = \nu_n^+(\cdot) - \nu_n^-(\cdot)$. By the functional delta method, the weak convergence translate into the asymptotic distribution of the process $\sqrt{nh_n^{1+2v}}[\hat{\tau}(\cdot) - \tau(\cdot)]$ for the local Wald estimator (4.5). We set the following additional assumption. Primitive conditions for it will be discussed specifically for the fuzzy quantile RDD in Section 4.5, and for each of the nine examples in Appendix C.

Assumption 2 (*Conditional Weak Convergence*). (i) ψ, ϕ and Υ are Hadamard differentiable at $\mu_1^{(v)}(0^{\pm}, \cdot)$, $\mu_2^{(v)}(0^{\pm}, \cdot)$, and W, respectively, tangentially to some subspaces of their domains, with their Hadamard derivatives denoted by $\phi'_{\mu_1^{(v)}(0^{\pm}, \cdot)}$, $\psi'_{\mu_2^{(v)}(0^{\pm}, \cdot)}$, and Υ'_W, respectively. (ii) $\inf_{\theta_2' \in \Theta_2'}|\psi(\mu_2^{(v)}(0^+, \cdot))(\theta_2') - \psi(\mu_2^{(v)}(0^-, \cdot))(\theta_2')| > 0$. (iii) $nh_n^{1+2v} \to \infty$ as $n \to \infty$.

Condition (ii) requires the existence of a jump or a kink, which is assumed in most of the prior work as the key identification condition. Exceptions are Otsu et al. (2015) and Feir et al. (2016), which provide weak-identification-robust

methods of inference. We later use this idea to relax condition (ii) of this assumption in Section 7.2. Condition (iii) restricts admissible rates of bandwidths, which are consistent with common choice rules (e.g., Imbens and Kalyanaraman, 2012; Calonico, Cattaneo and Titiunik, 2014; Arai and Ichimura, 2016; Calonico, Cattaneo and Farrell, 2016, 2018; Arai and Ichimura, 2018; for regression discontinuity designs) – see Appendix F. The next theorem states the weak convergence result for the process $\sqrt{nh_n^{1+2v}}[\hat{\tau} - \tau]$.

Theorem 1 (*Weak Convergence*). *Under Assumptions 1 and 2, we have* $v_n^\pm \rightsquigarrow \mathbb{G}_{H\pm}$, *where* $\mathbb{G}_{H\pm}$ *are zero mean Gaussian processes* $\mathbb{G}_{H\pm} : \Omega^x \mapsto \ell^\infty(\mathbb{T})$ *with covariance function*

$$H^\pm((\theta, k), (\vartheta, l)) = \frac{\sigma_{kl}(\theta, \vartheta | 0^\pm) e_v'(\Gamma_p^\pm)^{-1}\Psi_p^\pm((\theta, k), (\vartheta, l))(\Gamma_p^\pm)^{-1}e_v}{\sqrt{c_k(\theta_k)c_l(\vartheta_l)}f_X(0)}$$

where $\Psi_p^\pm((\theta, k), (\vartheta, l)) = \int_{\mathbb{R}_\pm} r_p(u/c_k(\theta_k))r_p'(u/c_l(\vartheta_l))K(\frac{u}{c_k(\theta_1)})K(\frac{u}{c_l(\vartheta_l)})du$ *for each* $\theta = (\theta_1, \theta_2)$, $\vartheta = (\vartheta_1, \vartheta_2) \in \Theta$. *Therefore,*

$$\sqrt{nh_n^{1+2v}}[\hat{\tau} - \tau] \rightsquigarrow \Upsilon_W'\left(\frac{[\psi(\mu_2^{(v)}(0+, \cdot)) - \psi(\mu_2^{(v)}(0^-, \cdot))]\mathbb{G}'(\cdot, 1) - [\phi(\mu_1^{(v)}(0^+, \cdot)) - \phi(\mu_1^{(v)}(0^-, \cdot))]\mathbb{G}'(\cdot, 2)}{[\psi(\mu_2^{(v)}(0+, \cdot)) - \psi(\mu_2^{(v)}(0^-, \cdot))]^2}\right),$$

where $\mathbb{G}' : \Omega^x \mapsto \ell^\infty(\mathbb{T})$ *is defined as*

$$\begin{bmatrix} \mathbb{G}'(\cdot, 1) \\ \mathbb{G}'(\cdot, 2) \end{bmatrix} = \begin{bmatrix} \phi'_{\mu_1^{(v)}(0^+, \cdot)}\left(\mathbb{G}_{H+}(\cdot, 1)/\sqrt{c_1^{1+2v}(\cdot)}\right)(\cdot) - \phi'_{\mu_1^{(v)}(0^-, \cdot)}\left(\mathbb{G}_{H-}(\cdot, 1)/\sqrt{c_1^{1+2v}(\cdot)}\right)(\cdot) \\ \psi'_{\mu_2^{(v)}(0^+, \cdot)}\left(\mathbb{G}_{H+}(\cdot, 2)/\sqrt{c_2^{1+2v}(\cdot)}\right)(\cdot) - \psi'_{\mu_2^{(v)}(0^-, \cdot)}\left(\mathbb{G}_{H-}(\cdot, 2)/\sqrt{c_2^{1+2v}(\cdot)}\right)(\cdot) \end{bmatrix}.$$

See Appendix A.3 for a proof. For the sharp mean RDD, the fuzzy mean RDD, the sharp mean RKD, and the fuzzy mean RKD, computation of the limit process is straightforward in practice. On the other hand, for CDF and the quantile process, it is somewhat easier to approximate the limit process through the multiplier bootstrap. The following subsection presents this additional practical consideration.

4.4. Multiplier bootstrap

To simulate the limiting process of the BR, i.e., Theorem 1, we use the pseudo random sample $\{\xi_i\}_{i=1}^n$ drawn from the standard normal distribution, independently from the data $\{(Y_i, D_i, X_i)\}_{i=1}^n$. Precisely, $\{\xi_i\}_{i=1}^n$ is defined on $(\Omega^\xi, \mathcal{F}^\xi, \mathbb{P}^\xi)$, a probability space that is independent of $(\Omega^x, \mathcal{F}^x, \mathbb{P}^x)$ – this condition will be formally stated in Assumption 3. With this pseudo random sample, define the multiplier processes (MP)

$$v_{\xi,n}^\pm(\theta, k) = v! \sum_{i=1}^n \xi_i \frac{e_v'(\Gamma_p^\pm)^{-1}\mathcal{E}_k(Y_i, D_i, X_i, \theta)r_p(\frac{X_i}{h_{k,n}(\theta_k)})K(\frac{X_i}{h_{k,n}(\theta_k)})\delta_i^\pm}{\sqrt{nh_{k,n}(\theta_k)}f_X(0)}.$$

We also write $v_{\xi,n}(\cdot) = v_{\xi,n}^+(\cdot) - v_{\xi,n}^-(\cdot)$.

In practice, we need to replace $\mathcal{E}_k$ and f_X with their estimates. Let $\hat{f}_X$ be an estimate of f_X. For estimation of $\mathcal{E}_k$, since every component in the BR is multiplied by the kernel K supported on $[-1, 1]$, we only need to consider $\mathcal{E}_k(Y_i, D_i, X_i, \theta)\mathbb{1}\{|X_i/h_{k,n}(\theta_k)| \leq 1\}$. Write its estimate by $\hat{\mathcal{E}}_k(Y_i, D_i, X_i, \theta)\mathbb{1}\{|X_i/h_{k,n}(\theta_k)| \leq 1\}$, which has $\mu_{k,p}$ replaced by some estimate $\tilde{\mu}_{k,p}$ of $\mu_{k,p}$. Appendix A.6 discusses the effects of these first-stage estimates. Substituting these estimated components in the MP, we define the estimated multiplier processes (EMP)

$$\hat{v}_{\xi,n}^\pm(\theta, k) = v! \sum_{i=1}^n \xi_i \frac{e_v'(\Gamma_p^\pm)^{-1}\hat{\mathcal{E}}_k(Y_i, D_i, X_i, \theta)r_p(\frac{X_i}{h_{k,n}(\theta_k)})K(\frac{X_i}{h_{k,n}(\theta_k)})\delta_i^\pm}{\sqrt{nh_{k,n}(\theta_k)}\hat{f}_X(0)}.$$

We also write $\hat{v}_{\xi,n}(\cdot) = \hat{v}_{\xi,n}^+(\cdot) - \hat{v}_{\xi,n}^-(\cdot)$.

In order to establish the uniform validity of the multiplier bootstrap, we invoke the following two sets of assumptions.

Assumption 3 (*Multiplier*). $\{\xi_i\}_{i=1}^n$ *is an independent standard normal random sample defined on* $(\Omega^\xi, \mathcal{F}^\xi, \mathbb{P}^\xi)$, *a probability space that is independent of* $(\Omega^x, \mathcal{F}^x, \mathbb{P}^x)$.

Assumption 4 (*First Stage Estimation*). $\tilde{\mu}_{k,p}(x, \theta_k)\mathbb{1}\{|x/h_{k,n}(\theta_k)| \leq 1\}$ *is uniformly consistent for* $\mu_k(x, \theta_k)\mathbb{1}\{|x/h_{k,n}(\theta_k)| \leq 1\}$ *on* $([\underline{x}, \overline{x}]\backslash\{0\}) \times \mathbb{T}$. $\hat{f}_X(0)$ *is consistent for* $f_X(0)$.

We remark that Assumption 3 is the standard assumption for multiplier, score, and wild bootstrap methods, cf. Kosorok (2003, 2008). At the current level of generality, we state these high-level conditions for the first stage estimation in Assumption 4. However, we will propose a concrete $\tilde{\mu}_k$ in Appendix A.6 that satisfies Assumption 4 under Assumptions 1 and 2. The following theorem, together with Theorem 1, establishes the uniform validity of the multiplier bootstrap.

Theorem 2 (*Conditional Weak Convergence*)**.** *Under* [Assumption](#) 1, 2, 3, *and* 4, *we have* $\hat{v}^{\pm}_{\xi,n} \overset{p}{\underset{\xi}{\rightsquigarrow}} \mathbb{G}_{H\pm}$, *and therefore*

$$\Upsilon'_W\Bigg(\frac{[\psi(\mu_2^{(v)}(0^+,\cdot)) - \psi(\mu_2^{(v)}(0^-,\cdot))]\widehat{\mathbb{X}}'_n(\cdot,1) - [\phi(\mu_1^{(v)}(0^+,\cdot)) - \phi(\mu_1^{(v)}(0^-,\cdot))]\widehat{\mathbb{X}}'_n(\cdot,2)}{(\psi(\mu_2^{(v)}(0^+,\cdot)) - \psi(\mu_2^{(v)}(0^-,\cdot)))^2}\Bigg)$$

$$\overset{p}{\underset{\xi}{\rightsquigarrow}}\Upsilon'_W\Bigg(\frac{[\psi(\mu_2^{(v)}(0^+,\cdot)) - \psi(\mu_2^{(v)}(0^-,\cdot))]\mathbb{G}'(\cdot,1) - [\phi(\mu_1^{(v)}(0^+,\cdot)) - \phi(\mu_1^{(v)}(0^+,\cdot))]\mathbb{G}'(\cdot,2)}{(\psi(\mu_2^{(v)}(0^+,\cdot)) - \psi(\mu_2^{(v)}(0^-,\cdot)))^2}\Bigg),$$

where

$$\begin{bmatrix}\widehat{\mathbb{X}}'_n(\cdot,1)\\\widehat{\mathbb{X}}'_n(\cdot,2)\end{bmatrix} = \begin{bmatrix}\phi'_{\mu_1^{(v)}(0^+,\cdot)}\Big(\hat{v}^+_{\xi,n}(\cdot,1)/\sqrt{c_1^{1+2v}(\cdot)}\Big)(\cdot) - \phi'_{\mu_1^{(v)}(0^-,\cdot)}\Big(\hat{v}^-_{\xi,n}(\cdot,1)/\sqrt{c_1^{1+2v}(\cdot)}\Big)(\cdot)\\\psi'_{\mu_2^{(v)}(0^+,\cdot)}\Big(\hat{v}^+_{\xi,n}(\cdot,2)/\sqrt{c_2^{1+2v}(\cdot)}\Big)(\cdot) - \psi'_{\mu_2^{(v)}(0^-,\cdot)}\Big(\hat{v}^-_{\xi,n}(\cdot,2)/\sqrt{c_2^{1+2v}(\cdot)}\Big)(\cdot)\end{bmatrix}.$$

See [Appendix A.4](#) for a proof. [Theorems](#) 1 and 2 show that the estimated multiplier process

$$\Upsilon'_W\Bigg(\frac{[\psi(\mu_2^{(v)}(0^+,\cdot)) - \psi(\mu_2^{(v)}(0^-,\cdot))]\widehat{\mathbb{X}}'_n(\cdot,1) - [\phi(\mu_1^{(v)}(0^+,\cdot)) - \phi(\mu_1^{(v)}(0^-,\cdot))]\widehat{\mathbb{X}}'_n(\cdot,2)}{(\psi(\mu_2^{(v)}(0^+,\cdot)) - \psi(\mu_2^{(v)}(0^-,\cdot)))^2}\Bigg)$$

can be used to approximate the limit process of $\sqrt{nh_n^{1+2v}}[\hat{\tau} - \tau]$ in practice.

4.5. Fuzzy quantile RDD ([Example](#) 2) revisited

In this section, we apply the main general results, namely [Theorems 1](#) and [2](#), to the fuzzy quantile RDD introduced in [Example 2](#). What we present in this section provides a formal justification of the overview in Section [3](#). We present our assumptions for the case of $p = 2$ as we did in the overview. We remark that, however, using a different order p of local polynomial fitting is also possible by similar arguments.

Consider Θ_1, Θ_2, Θ'_1, Θ'_2, Θ'', g_1, g_2, ϕ, ψ, and Υ defined in [Example 2](#). Recall that we denote the local Wald estimand (4.1) with $v = 0$ in this setting by τ_{FQRD}. We also denote the analog estimator (4.5) with $v = 0$ in this setting by

$$\hat{\tau}_{FQRD}(\theta'') = \Upsilon(\hat{F}_{Y\cdot|C})(\theta''),$$

where

$$\hat{F}_{Y^1|C}(y) = \frac{\hat{\mu}_{1,2}(0^+,y,1) - \hat{\mu}_{1,2}(0^-,y,1)}{\hat{\mu}_{2,2}(0^+,1) - \hat{\mu}_{2,2}(0^-,1)} \quad \text{and} \quad \hat{F}_{Y^0|C}(y) = \frac{\hat{\mu}_{1,2}(0^+,y,0) - \hat{\mu}_{1,2}(0^-,y,0)}{\hat{\mu}_{2,2}(0^+,0) - \hat{\mu}_{2,2}(0^-,0)}.$$

By van der Vaart and Wellner (1996; Lemma 3.9.23) and van der Vaart (1998; Theorem 20.9), Υ is Hadamard differentiable at $(F_{Y\cdot|C})$ tangentially to $C(\mathscr{Y}_1 \times \mathscr{D}^2)$, and the Hadamard derivative is a map that takes each $g \in C(\mathscr{Y}_1 \times \mathscr{D}^2)$ to

$$\Upsilon'_W(g)(\cdot) =: \Upsilon'_{F_{Y\cdot|C}}(g)(\cdot) = -\frac{g(Q_{Y^1|C}(\cdot),1,1)}{f_{Y^1|C}(Q_{Y^1|C}(\cdot))} + \frac{g(Q_{Y^0|C}(\cdot),0,0)}{f_{Y^0|C}(Q_{Y^0|C}(\cdot))}$$

under the assumptions to be stated below. Sufficient conditions for the assumptions required for the general result, tailored to the current example, are stated as follows.

Assumption S. (a) $\{(Y_i, D_i, X_i)\}_{i=1}^n$ are n i.i.d. copies of the random vector (Y, D, X) defined on a probability space $(\Omega^x, \mathcal{F}^x, \mathbb{P}^x)$. (b) X has a density function f_X which is continuously differentiable on $[\underline{x}, \bar{x}]$ that contains 0 in its interior, and $0 < f_X(0) < \infty$.

Assumption K. (a) $K : [-1, 1] \to \mathbb{R}^+$ is bounded and continuous. (b) $\{K(\cdot/h) : h > 0\}$ is of VC type. (c) $\Gamma_p^{\pm}$ is positive definite.

Assumption M. $\{\xi_i\}_{i=1}^n$ are independent standard normal random variables defined on $(\Omega^\xi, \mathcal{F}^\xi, \mathbb{P}^\xi)$, a probability space that is independent of $(\Omega^x, \mathcal{F}^x, \mathbb{P}^x)$.

Assumption FQRD. (i) $(x, y, d) \mapsto \frac{\partial^j}{\partial x^j}E[\mathbb{1}\{Y_i^* \leq y, D_i^* = d\}|X_i = x]$ is Lipschitz in x on $[\underline{x}, 0) \times \Theta_1$ and $(0, \bar{x}] \times \Theta_1$ for $j = 0, 1, 2, 3$, and $(x, d) \mapsto \frac{\partial^j}{\partial x^j}E[\mathbb{1}\{D_i^* = d\}|X_i = x]$ is Lipschitz in x on $[\underline{x}, 0) \times \Theta_2$ and $(0, \bar{x}] \times \Theta_2$ for $j = 0, 1, 2, 3$. (ii) The baseline bandwidth $\bar{h}_n$ satisfies $h_n \to 0$ and $nh_n^2 \to \infty$, $nh_n^7 \to 0$. There exist bounded constants $0 < c_1$, $c_2 < \infty$ such that $h_{1,n} = c_1 h_n$ and $h_{2,n} = c_2 h_n$. (iii) $|\mathbb{P}^x(D_i = 1|X_i = 0^+) - \mathbb{P}^x(D_i = 1|X_i = 0^-)| > 0$. (iv) $F_{Y^1|C}, F_{Y^0|C} \in \mathcal{C}^1(\mathscr{Y}_1)$, and $f_{Y^1|C}$ and $f_{Y^0|C}$ are bounded away from 0 on $\mathscr{Y}_1$. (v) There exists $\hat{f}_{Y|XD^*}(y|0^{\pm}, d)$ such that $\sup_{(y,d)\in\mathscr{Y}_1\times\mathscr{D}}|\hat{f}_{Y|XD^*}(y|0^{\pm}, d) - f_{Y|XD^*}(y|0^{\pm}, d)| = o_p^x(1)$.

Assumption S (a) is assumed in most of the prior work, including the closely related benchmark by Frandsen et al. (2012). As emphasized after "Assumptions 1 (i) (a)", a part of the literature has relaxed the assumption of identical distribution and studies cluster-robust inference — we will also present a method of cluster-robust inference in Section 7.1 by extending our baseline results. Assumption S (b) is analogous to Assumption E3 of Frandsen et al. (2012). Assumption K is analogous to Assumption E4 of Frandsen et al. (2012). Assumption M is new to our paper due to our use of the multiplier bootstrap, which is not used by prior work such as Frandsen et al. (2012). Assumption FQRD (i), (ii), (iii), and (iv) are analogous to Assumptions E1, E5, E2, and Q, respectively, of Frandsen et al. (2012). We state Assumptions FQRD (v) at this high level in order to accommodate a number of alternative estimators. In Lemma 14 in Appendix E.1.2, however, we propose one such concrete estimator which satisfies part (v). All the other parts of this assumption are immediately interpretable.

Define the EMP by

$$\hat{v}_{\xi,n}^{\pm}(y, d_1, d_2, 1) = \sum_{i=1}^{n} \xi_i \frac{e_0'(\Gamma_2^{\pm})^{-1}[\mathbb{1}\{Y_i^* \leq y, D_i^* = d_1\} - \tilde{\mu}_{1,2}(X_i, y, d_1)] r_2(\frac{X_i}{h_n}) K(\frac{X_i}{h_n}) \delta_i^{\pm}}{\sqrt{n h_n} \hat{f}_X(0)}$$

$$\hat{v}_{\xi,n}^{\pm}(y, d_1, d_2, 2) = \sum_{i=1}^{n} \xi_i \frac{e_0'(\Gamma_2^{\pm})^{-1}[\mathbb{1}\{D_i^* = d_2\} - \tilde{\mu}_{2,2}(X_i, d_2)] r_2(\frac{X_i}{h_n}) K(\frac{X_i}{h_n}) \delta_i^{\pm}}{\sqrt{n h_n} \hat{f}_X(0)}$$

$$\widehat{\mathbb{X}}_n'(y, d_1, d_2, k) = \hat{v}_{\xi,n}^{+}(y, d_1, d_2, k)/\sqrt{c_k} - \hat{v}_{\xi,n}^{-}(y, d_1, d_2, k)/\sqrt{c_k}$$

for each $(y, d_1, d_2, k) \in \mathbb{T} = \mathscr{Y}_1 \times \mathscr{D} \times \mathscr{D} \times \{1, 2\}$. Define the following estimated process through the Hadamard derivative.

$$\widehat{\Upsilon}_W'(\widehat{\mathbb{Y}}_n)(\theta'')$$
$$= \frac{[\hat{\mu}_{2,2}(0^+, 1) - \hat{\mu}_{2,2}(0^-, 1)]\widehat{\mathbb{X}}_n'(\hat{Q}_{Y^1|C}(\theta''), 1, 1, 1) - [\hat{\mu}_{1,2}(0^+, \hat{Q}_{Y^1|C}(\theta''), 1) - \hat{\mu}_{1,2}(0^-, \hat{Q}_{Y^1|C}(\theta''), 1)]\widehat{\mathbb{X}}_n'(\hat{Q}_{Y^1|C}(\theta''), 1, 1, 2)}{\hat{f}_{Y^1|C}(\hat{Q}_{Y^1|C}(\theta''))[\hat{\mu}_{2,2}(0^+, 1) - \hat{\mu}_{2,2}(0^-, 1)]^2}$$
$$- \frac{[\hat{\mu}_{2,2}(0^+, 0) - \hat{\mu}_{2,2}(0^-, 0)]\widehat{\mathbb{X}}_n'(\hat{Q}_{Y^0|C}(\theta''), 0, 0, 1) - [\hat{\mu}_{1,2}(0^+, \hat{Q}_{Y^0|C}(\theta''), 0) - \hat{\mu}_{1,2}(0^-, \hat{Q}_{Y^0|C}(\theta''), 0)]\widehat{\mathbb{X}}_n'(\hat{Q}_{Y^0|C}(\theta''), 0, 0, 2)}{\hat{f}_{Y^0|C}(\hat{Q}_{Y^0|C}(\theta''))[\hat{\mu}_{2,2}(0^+, 0) - \hat{\mu}_{2,2}(0^-, 0)]^2}$$

where

$$\widehat{\mathbb{Y}}_n(y, d_1, d_2) = \frac{[\hat{\mu}_{2,2}(0^+, d_2) - \hat{\mu}_{2,2}(0^-, d_2)]\widehat{\mathbb{X}}_n'(y, d_1, d_2, 1) - [\hat{\mu}_{1,2}(0^+, y, d_1) - \hat{\mu}_{1,2}(0^-, y, d_1)]\widehat{\mathbb{X}}_n'(y, d_1, d_2, 2)}{[\hat{\mu}_{2,2}(0^+, d_2) - \hat{\mu}_{2,2}(0^-, d_2)]^2}$$

for $(y, d_1, d_2) \in \mathscr{Y}_1 \times \mathscr{D}^2$. With these preparations, our general result applied to the current case yields the following corollary.

Corollary 1 (*Example: Fuzzy Quantile RDD*)**.** *Suppose that Assumptions K, M, S and FQRD hold. (i) There exists a zero mean Gaussian process* $\mathbb{G}_{FQRD}' : \Omega^x \mapsto \ell^{\infty}([a, 1-a])$ *such that* $\sqrt{n h_n}[\hat{\tau}_{FQRD} - \tau_{FQRD}] \rightsquigarrow \mathbb{G}_{FQRD}'$*. (ii) Furthermore, with probability approaching one,* $\widehat{\Upsilon}_W'(\widehat{\mathbb{Y}}_n) \overset{p}{\underset{\xi}{\rightsquigarrow}} \mathbb{G}_{FQRD}'$*.*

A proof is provided in Appendix E.2.1. This result justifies the overview in Section 3. Specifically, the estimated multiplier process $\widehat{\Upsilon}_W'(\widehat{\mathbb{Y}}_n)$ – denoted by $\widehat{\mathbb{G}}_{FQRD,n}'$ in Section 3 for simplicity – can be used to approximate the limit process of $\sqrt{n h_n}[\hat{\tau}_{FQRD} - \tau_{FQRD}]$. We could consider a set of functions defined on $(0, 1)$ instead of $[a, 1-a]$ by making a stronger assumption that the potential outcomes are compactly supported with their conditional density functions bounded away from zero, although we may not want to make such a strong assumption in general for typical applications. This tradeoff between the globalization of the domain and the strength of the assumption is even true for simple unconditional quantile processes, e.g., Lemma 21.4 (i) versus Lemma 21.4 (ii) in van der Vaart (1998).

Remark 3. Our bias correction is conducted on the mean regression estimates $\hat{\mu}_{1,2}$ and $\hat{\mu}_{2,2}$ in light of Remark 7 of Calonico et al. (2014). Some calculation shows the biases of local Wald-ratios $\hat{F}_{Y^d|C}$ also have the same bias order as the regression estimates above. Finally, since the quantile estimate based on a left inverse of a CDF has the same order of bias as the CDF estimate, we can achieve bias correction for fuzzy quantile RDD estimate.

Remark 4. We emphasize that our one-step bias correction as well as Remark 7 of Calonico et al. (2014) relies on choosing the same bandwidth for both main estimators and higher order bias estimators. According to the simulation studies in Calonico et al. (2014), such bandwidth choice works rather well in their various DGPs. While studies of optimal coverage probability for CDF and quantile estimation would be both useful and important, it is out of the scope of the current paper.

It is also worth noting that our framework can accommodate the case when two bandwidths are different as well, since we can acquire the uniform Bahadur representations for both the main estimator and the bias estimator via our uniform Bahadur representation in Appendix A.1. The multiplier bootstrap can be applied to the difference of these two processes, which would be similar to Lemma 3 and Proposition 2 in Qu and Yoon (2018).

5. Simulation studies

We conduct simulation studies for the fuzzy quantile RDD − see Sections 3 and 4.5. We also conduct simulation studies for the other cases − see Appendix D. We follow the procedure outlined in Appendix F for choices of bandwidths in finite samples. The kernel function that we use is the Epanechnikov kernel. All the other procedures exactly follow the guideline in Section 4.5.

To systematically evaluate the hypothesis testing for the hypotheses of uniform treatment nullity and treatment homogeneity presented in Section 3, we consider the following data generating process. We generate an i.i.d. sample $\{(Y_i, D_i, X_i)\}_{i=1}^n$ through the following data generating process:

$$\begin{cases} Y_i = \mu(X_i) + \beta_1 D_i + (1 + \gamma_1 D_i) \cdot U_i, \\ D_i = \mathbb{1}\{2 \cdot \mathbb{1}\{X_i \geq 0\} - 1 \geq V_i\}, \end{cases} \qquad (X_i, U_i, V_i)' \sim N(0, \Sigma),$$

where β_1 and γ_1 are to be varied across simulation sets, $\Sigma_{11} = \sigma_X^2 = 0.1781742^2$, $\Sigma_{22} = \sigma_U^2 = 0.1295^2$, $\Sigma_{33} = \sigma_V^2 = 0.5^2$, $\Sigma_{12} = \rho_{XU} \cdot \sigma_X \cdot \sigma_U = 0.25 \cdot 1.0^2$, $\Sigma_{13} = \rho_{XV} \cdot \sigma_X \cdot \sigma_V = 0.0 \cdot 1.0 \cdot 0.5$, and $\Sigma_{23} = \rho_{UV} \cdot \sigma_U \cdot \sigma_V = 0.25 \cdot 1.0 \cdot 0.5$. We choose the numbers, $\sigma_X^2 = 0.1781742^2$ and $\sigma_U^2 = 0.1295^2$, so that these variances match the corresponding variances in the data generating processes in the simulation studies by Calonico et al. (2014). Furthermore, the polynomial part $\mu(X_i)$ of the outcome equation is defined due to Lee (2008):

$$\mu(x) = \begin{cases} 1.27x + 7.18x^2 + 20.21x^3 + 21.54x^4 + 7.33x^5 & \text{if } x < 0 \\ 0.84x - 3.00x^2 + 7.99x^3 - 9.01x^4 + 3.56x^5 & \text{if } x \geq 0, \end{cases}$$

also following the simulation studies by Calonico et al. (2014). Observe under the above data generating process that

θ-th conditional quantile treatment effect at $x = 0$ is $\beta_1 + \gamma_1 F_{U|X}^{-1}(\theta|0)$.

We set $\Theta'' = [a, 1 - a] = [0.20, 0.80]$ as the set of quantiles on which we conduct uniform inference. We use a grid with the interval size of 0.02 to approximate the continuum Θ'' for numerical evaluation of functions defined on Θ''. Similarly, we use a grid with interval size of 0.02 to approximate the continuum $\mathscr{Y}$ for numerical evaluation of functions defined on $\mathscr{Y}$. First, we simulate the 95% test for the null hypothesis $H_0 : \tau_{FQRD}(\theta'') = 0 \; \forall \theta'' \in [a, 1 - a]$ of uniform treatment nullity using the procedure described in Section 4.5. Second, we simulate the 95% test for the null hypothesis $H_0 : \tau_{FQRD}(\theta'') = \tau_{FQRD}(\theta''') \; \forall \theta'', \theta''' \in [a, 1 - a]$ of treatment homogeneity using the procedure described in Section 4.5. Third, we simulate the uniform coverage probability of the true quantile treatment effects $\tau_{FQRD}(\cdot)$ using the uniform confidence band constructed in Section 4.5.

Table 1 shows simulated acceptance and coverage probabilities based on 2,500 multiplier bootstrap replications for 2,500 Monte Carlo replications for each of the sample sizes $n = 500, 1,000, 1,500,$ and $2,000$. Panel (A) reports results for the test of uniform treatment nullity, panel (B) shows results for the test of treatment homogeneity, and panel (C) shows results for the uniform coverage probability. The left column groups (I) present results across alternative values of $\beta_1 \in \{0.00, 0.05, 0.10, 0.15, 0.20\}$ while fixing $\gamma_1 = 0$. The case of $\beta_1 = 0.00$ evaluates the size in (A), whereas the cases of $\beta_1 \in \{0.05, 0.10, 0.15, 0.20\}$ evaluate the power in (A). All of the cases of $\beta_1 \in \{0.00, 0.05, 0.10, 0.15, 0.20\}$ evaluate the coverage in (B) and (C). The right column groups (II) present results across alternative values of $\gamma_1 \in \{0.00, 0.25, 0.50, 0.75, 1.00\}$ while fixing $\beta_1 = 0$. The case of $\gamma_1 = 0.00$ evaluates the coverage in (A) and (B), whereas the cases of $\gamma_1 \in \{0.25, 0.50, 0.75, 1.00\}$ evaluate the power in (A) and (B). All of the cases of $\beta_1 \in \{0.00, 0.25, 0.50, 0.75, 1.00\}$ evaluate the coverage in (C). The nominal acceptance probability is 95%.

In view of the columns for $\beta_1 = 0.00$ and $\gamma_1 = 0.00$, we confirm the correct size. Furthermore, the simulated and coverage probabilities under $\beta_1 = 0$ and $\gamma_1 = 0$ approach the nominal probability as the sample size increases. The acceptance probability in panel (A) decreases as β_1 or γ_1 deviates away from zero, which is consistent with the fact that the joint treatment nullity is violated by $\beta_1 \neq 0$ or $\gamma_1 \neq 0$. The acceptance probability in panel (B) stays roughly constant as β_1 deviates away from zero, but it decreases as γ_1 deviates away from zero. This result is consistent with the fact that the treatment homogeneity is retained for any value of β_1, but is violated for $\gamma_1 \neq 0$. The uniform coverage probability in panel (C) keeps the nominal size across all the values of β_1 and γ_1, which evidences the effectiveness of the uniform confidence bands across alternative data generating processes in the presence of nontrivial and/or heterogeneous treatment effects.

To assess the sensitivity of the proposed inference procedure to weak jumps, we conduct additional simulations with a sequence of values of $\Sigma_{33} \in \{2^0, 2^1, 2^2, 2^3, 2^4, 2^5, 2^6\}$ while fixing $\beta_1 = \gamma = 1 = 0$. Table 2 shows simulated acceptance and coverage probabilities based on 2,500 multiplier bootstrap replications for 2,500 Monte Carlo replications for each of the sample sizes $n = 500, 1,000, 1,500,$ and $2,000$. Panel (A) reports results for the test of uniform treatment nullity, and panel (B) shows results for the test of treatment homogeneity. Results for (C) the uniform coverage probability are omitted because they are the same as the results of (A) under $\beta_1 = \gamma_1 = 0$. Notice that the size becomes smaller and deviates away from the nominal size as the jump becomes weak (i.e., as Σ_{33} increases). This observation of size distortions under weak jumps is consistent with prior study, e.g., Feir et al. (2016). Motivated by these results, we present an extended theory of inference with robustness against the possibility of no or weak jumps in Section 7.2.

Table 1
(A) Simulated acceptance probabilities for uniform treatment nullity, (B) simulated acceptance probabilities for treatment homogeneity under the fuzzy quantile RDD, and (C) uniform coverage probability of the true quantile treatment effects by the uniform confidence bands. The left column groups (I) present results across alternative values of $\beta_1 \in \{0.00, 0.05, 0.10, 0.15, 0.20\}$ while fixing $\gamma_1 = 0$. The right column groups (II) present results across alternative values of $\gamma_1 \in \{0.00, 0.25, 0.50, 0.75, 1.00\}$ while fixing $\beta_1 = 0$. The nominal acceptance probability under the null hypothesis is 95%.

(I) (A) Joint Treatment Nullity						(II) (A) Joint Treatment Nullity					
n	β_1					n	γ_1				
	0.00	0.05	0.10	0.15	0.20		0.00	0.25	0.50	0.75	1.00
500	0.969	0.874	0.709	0.532	0.362	500	0.969	0.956	0.942	0.934	0.922
1000	0.968	0.823	0.559	0.302	0.131	1000	0.968	0.941	0.916	0.867	0.800
1500	0.957	0.784	0.422	0.149	0.045	1500	0.957	0.944	0.877	0.802	0.708
2000	0.957	0.725	0.318	0.082	0.014	2000	0.957	0.927	0.852	0.747	0.617

(I) (B) Treatment Homogeneity						(II) (B) Treatment Homogeneity					
n	β_1					n	γ_1				
	0.00	0.05	0.10	0.15	0.20		0.00	0.25	0.50	0.75	1.00
500	0.989	0.984	0.988	0.986	0.987	500	0.989	0.966	0.928	0.873	0.828
1000	0.980	0.979	0.979	0.979	0.985	1000	0.980	0.939	0.842	0.729	0.591
1500	0.968	0.973	0.972	0.977	0.977	1500	0.968	0.913	0.770	0.606	0.438
2000	0.963	0.962	0.966	0.970	0.967	2000	0.963	0.884	0.705	0.503	0.311

(I) (C) Uniform Coverage Probability						(II) (C) Uniform Coverage Probability					
n	β_1					n	γ_1				
	0.00	0.05	0.10	0.15	0.20		0.00	0.25	0.50	0.75	1.00
500	0.969	0.966	0.969	0.966	0.970	500	0.969	0.962	0.962	0.969	0.970
1000	0.968	0.965	0.965	0.969	0.976	1000	0.968	0.953	0.953	0.952	0.960
1500	0.957	0.967	0.966	0.962	0.967	1500	0.957	0.958	0.952	0.944	0.951
2000	0.957	0.953	0.965	0.968	0.971	2000	0.957	0.950	0.950	0.941	0.944

Table 2
(A) Simulated acceptance probabilities for uniform treatment nullity, (B) simulated acceptance probabilities for treatment homogeneity under the fuzzy quantile RDD for a sequence of weakening jumps. The results are displayed for $\Sigma_{33} \in \{2^0, 2^1, 2^2, 2^3, 2^4, 2^5, 2^6\}$. The parameter values are fixed at $\beta_1 = \gamma_1 = 0$. The nominal acceptance probability under the null hypothesis is 95%. The results for (C) uniform coverage probability of the true quantile treatment effects by the uniform confidence bands are omitted because they produce the same numbers as (A) under $\beta_1 = \gamma_1 = 0$.

(A) Joint Treatment Nullity							
n	Σ_{33}						
	2^0	2^1	2^2	2^3	2^4	2^5	2^6
500	0.968	0.984	0.984	0.985	0.986	0.987	0.989
1000	0.962	0.969	0.979	0.984	0.984	0.983	0.988
1500	0.956	0.962	0.967	0.976	0.980	0.982	0.980
2000	0.957	0.955	0.957	0.972	0.977	0.979	0.981

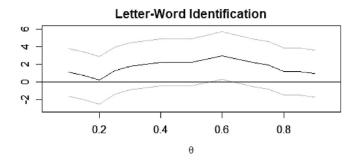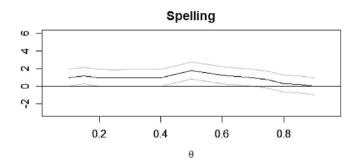| (B) Treatment Homogeneity							
n	Σ_{33}						
	2^0	2^1	2^2	2^3	2^4	2^5	2^6
500	0.990	0.996	0.992	0.990	0.988	0.990	0.992
1000	0.982	0.991	0.995	0.993	0.992	0.991	0.995
1500	0.974	0.992	0.994	0.992	0.991	0.993	0.990
2000	0.977	0.986	0.991	0.993	0.993	0.991	0.993

6. Empirical illustration

In this section, we apply our method of robust uniform inference for the fuzzy quantile RDD (Sections 3 and 4.5). Using RDD with a birth-day cutoff eligibility rule for the Oklahoma pre-K program, Gormley Jr. et al. (2005) find significant positive effects of cognitive development on average test scores. They also find that the average effects are positive among sub-samples of students in lower socio-economic status. Following up the latter finding, Frandsen et al. (2012) provide additional evidence that these effects are positive among the lower end of the distribution via estimated quantile treatment effects. We apply our method of constructing robust uniform confidence bands to complement the findings by Frandsen et al. (2012).
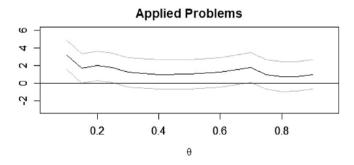
Fig. 1. The estimated local quantile treatment effects of the pre-K programs on scores on the three sub-tests of the Woodcock–Johnson tests, and their 90% uniform confidence bands.

The data consist of a sample of 4,710 incoming Tulsa Public Schools kindergartners and pre-K participants for the 2003–2004 school year. The main variables used in this data are the birth date (X), an indicator for participation in the pre-K program in the previous year (D^*), and scores on the Woodcock–Johnson sub-tests (Y^*): Letter-word, Spelling, and Applied Problems. The implementation procedure follows the guideline provided in Section 4.5 – or the overview in Section 3 – as well as the additional first-stage estimators suggested in Appendix A.6 and the bandwidth selection procedure suggested in Appendix F.

Fig. 1 plots the estimated local quantile treatment effects of the pre-K programs on scores on the three sub-tests of the Woodcock–Johnson tests. The figure shows the point estimates for each quantile θ by black curves. It also shows the 90% uniform confidence bands based on our proposed procedure.

The qualitative patterns of our results resemble those of Frandsen et al. (2012) for each of the three sub-tests. Not surprisingly, our 90% confidence bands are wider than the 90% point-wise confidence intervals obtained by Frandsen et al. (2012), and statistical significance vanishes at some quantile indices. Nonetheless, despite the generally greater widths of confidence bands than confidence intervals, the statistical significance remains for relevant quantile indices. In particular, as in Frandsen et al. (2012), we continue to conclude that the program succeeded in significantly raising the lower end of the distribution of test scores, especially for the Applied Problems sub-test, which is consistent with Gormley Jr. et al.'s (2005) finding that estimated average effects are larger for children with disadvantaged socio-economic status.

7. Extensions

In this section, we discuss three directions for extending the baseline model and method presented in Section 4. These extensions include cluster-robust inference (Section 7.1), inference with robustness against no or weak jumps or kinks (Section 7.2), and augmented models with observed covariates (Section 7.3).

7.1. Cluster robust inference

In applications, researchers may encounter situations where data are cluster sampled and thus the i.i.d. assumption may be implausible. Cluster-robust standard errors for sharp/fuzzy mean RDD are studied by Bartalotti and Brummet (2017) and Calonico et al. (2018b). In this section, we show that Theorems 1 and 2 can be generalized to the cases of cluster sampled data. Suppose that a researcher observes $\{(Y_i, D_i, X_i) : i \in C_g, g \in \{1, \ldots, G\}\}$, where each $C_g \subset \{1, \ldots, n\}$ is an index set with cardinality $|C_g| \leq \bar{N}$ for an $\bar{N} \in \mathbb{N}$ independent of G and g, $\sum_{g=1}^{G} |C_g| = n$, and $C_g \cap C_{g'} = \emptyset$ whenever $g \neq g'$. Observations within the same cluster g can be arbitrarily dependent, while any two observations across different clusters are independent. Denote $\bar{f}_X = \frac{1}{G} \sum_{g=1}^{G} \sum_{i \in C_g} f_{X_i}$, where f_{X_i} is the density function of X_i. To keep our writing and analysis simple, we let $h_{1,G}(\theta_1) = h_{2,G}(\theta_2) = h_G$. The local Wald estimate $\hat{\tau}$ is calculated according to (4.4) and (4.5) in Section 4 with the bandwidth h_G. Note that $\frac{1}{n} \sum_{i=1}^{n}$ can also be written as $\frac{1}{n} \sum_{g=1}^{G} \sum_{i \in C_g}$. Since $\frac{c}{\sqrt{G}} \leq \frac{1}{\sqrt{n}} \leq \frac{1}{\sqrt{G}}$ for some $c > 0$, we can rescale estimates by $\frac{n}{G}$ without loss of generality. We now define the cluster-robust EMP as

$$\hat{v}_{\xi,G}^{\pm}(\theta, k) = v! \sum_{g=1}^{G} \xi_g \sum_{i \in C_g} \frac{e_v'(\Gamma_p^{\pm})^{-1} \hat{\mathcal{E}}_k(Y_i, D_i, X_i, \theta) r_p(\frac{X_i}{h_G}) K(\frac{X_i}{h_G}) \delta_i^{\pm}}{\sqrt{G h_G} \hat{\bar{f}}_X(0)},$$

where $\hat{\mathcal{E}}_k = g_k - \tilde{\mu}_k$ with $\tilde{\mu}_k$ and $\hat{\bar{f}}_X$ denoting first-stage estimators of μ_k and $\bar{f}_X$, respectively, satisfying the assumption below. Also denote $\hat{v}_{\xi,G}(\cdot) = \hat{v}_{\xi,G}^{+}(\cdot) - \hat{v}_{\xi,G}^{-}(\cdot)$.

Assumption 5 (*Cluster Sampling*). Let $\underline{x} < 0 < \bar{x}$. (i) (a) For each $G \in \mathbb{N}$, $\{(Y_i, D_i, X_i) : i \in C_1\}, \ldots, \{(Y_i, D_i, X_i) : i \in C_G\}$ are independent random vectors defined on a probability space $(\Omega^x, \mathcal{F}^x, \mathbb{P}^x)$; (b) For each $G \in \mathbb{N}$ and for each $i \in C_1 \cup \cdots \cup C_G$, X_i has a density function f_{X_i} which is continuously differentiable and 1-Lipschitz, and satisfies $0 < \underline{f} \leq f_{X_i} \leq \bar{f} < \infty$ on $[\underline{x}, \bar{x}]$. (ii) For each $k = 1, 2$: (a) the collections of real-valued functions, $\{x \mapsto \mu_k(x, \theta_k) : \theta_k \in \Theta_k\}$, $\{y \mapsto g_1(y, \theta_1) : \theta_1 \in \Theta_1\}$, and $\{d \mapsto g_2(d, \theta_2) : \theta_2 \in \Theta_2\}$, are of VC type and are uniformly bounded by $\bar{M} < \infty$; (b) $\mu_k^{(j)}$ is Lipschitz on $[\underline{x}, 0) \times \Theta_k$ and $(0, \bar{x}] \times \Theta_k$ for $j = 0, 1, 2, \ldots, p+1$; (c) for each $y \in \mathscr{Y}$, $g_1(y, \cdot)$ is left- or right-continuous in each dimension; and similarly, for each $d \in \mathscr{D}$, $g_2(d, \cdot)$ is left- or right-continuous in each dimension. (d) For each $G \in \mathbb{N}$, $i \in \{1, \ldots, n\}$, $\theta \in \Theta$, it holds that $E[\mathcal{E}_k(Y_i, D_i, X_i, \theta)|(X_j : j \in C_g, i \in C_g)] = E[\mathcal{E}_k(Y_i, D_i, X_i, \theta)|X_i] = 0$. (iii) $h_G \to 0$, $G h_G^2 \to \infty$ and $G h_G^{2p+3} \to 0$ for some $h_0 < \infty$. (iv) (a) $K : [-1, 1] \to \mathbb{R}^+$ is bounded and continuous; (b) $\{K(\cdot/h) : h > 0\}$ is of VC type. (c) $\Gamma_p^{\pm}$ is positive definite. (v) There exist uniformly bounded functions $\Sigma^{\pm} : (\Theta \times \{1, 2\})^2 \mapsto \mathbb{R}$ such that

$$\frac{1}{G} \sum_{g=1}^{G} E\left[\left(\sum_{i \in C_g} \frac{e_v'(\Gamma_p^{\pm})^{-1} r_p(\frac{X_i}{h_G}) \mathcal{E}_k(Y_i, D_i, X_i, \theta) K(\frac{X_i}{h_G}) \delta_i^{\pm}}{\sqrt{h_G} \bar{f}_X(0)}\right)\left(\sum_{i \in C_g} \frac{e_v'(\Gamma_p^{\pm})^{-1} r_p(\frac{X_i}{h_G}) \mathcal{E}_k(Y_i, D_i, X_i, \vartheta) K(\frac{X_i}{h_G}) \delta_i^{\pm}}{\sqrt{h_G} \bar{f}_X(0)}\right)'\right] = \Sigma^{\pm}((\theta, k), (\vartheta, l)) + O_p(h_G).$$

(vi)(a) ϕ and ψ are identity mappings while Υ is Hadamard differentiable at W tangentially to a subspace of its domain, with its Hadamard derivative denoted by Υ_W'. (vii) $\{\xi_g\}_{g=1}^{G}$ is an independent standard normal random sample defined on $(\Omega^{\xi}, \mathcal{F}^{\xi}, \mathbb{P}^{\xi})$, a probability space that is independent of $(\Omega^x, \mathcal{F}^x, \mathbb{P}^x)$. (b) $\inf_{\theta_2 \in \Theta_2} |\mu_2^{(v)}(0^+, \theta_2) - \mu_2^{(v)}(0^-, \theta_2)| > 0$. (viii) There exist first stage estimators $\tilde{\mu}_{k,p}$ and $\hat{\bar{f}}_X$ such that $\tilde{\mu}_{k,p}(x, \theta_k) \mathbb{1}\{|x/h_G| \leq 1\}$ is uniformly consistent for $\mu_k(x, \theta_k) \mathbb{1}\{|x/h_G| \leq 1\}$ on $([\underline{x}, \bar{x}] \backslash \{0\}) \times \mathbb{T}$. $\hat{\bar{f}}_X(0)$ is consistent for $\bar{f}_X(0)$.

This Assumption 5 for cluster sampling corresponds to Assumption 1, 2, 3 and 4 for the i.i.d. setting. In part (i)(b) of Assumption 5, we restrict the density functions to be 1-Lipschitz, but this particular scale is imposed only for the sake of concise writings in proofs. This number can be replaced by any finite constant. Part (vi)(a) of Assumption 5 imposes more restrictions (namely, the identity ϕ and the identity ψ) than the baseline case, but this assumption is made only for the sake of simplicity and concise writing in proofs, and it can be relaxed. We make this simplifying assumption because it is already satisfied by most of the important cases anyway, including the case of FQRD, which is the main focus of this paper, as well as the cases of sharp/fuzzy mean RDD studied in prior work (Bartalotti and Brummet, 2017; Calonico et al., 2018b). Under this assumption for cluster sampling, we obtain the weak convergence and the conditional weak convergence results as stated in the following corollary.

Corollary 2. *Suppose Assumption 5 holds, then there exists a zero mean Gaussian process $\mathbb{G} : \Omega^x \mapsto \ell^{\infty}(\mathbb{T})$ such that*

$$\sqrt{G h_G^{1+2v}} [\hat{\tau}(\cdot) - \tau(\cdot)] \rightsquigarrow \Upsilon_W' \left(\frac{[\mu_2^{(v)}(0^+, \cdot) - \mu_2^{(v)}(0^-, \cdot)] \mathbb{G}(\cdot, 1) - [\mu_1^{(v)}(0^+, \cdot) - \mu_1^{(v)}(0^-, \cdot)] \mathbb{G}(\cdot, 2)}{(\mu_2^{(v)}(0^+, \cdot) - \mu_2^{(v)}(0^-, \cdot))^2}\right)(\cdot).$$

Furthermore, for the EMP, we have

$$\Upsilon_W' \left(\frac{[\mu_2^{(v)}(0^+, \cdot) - \mu_2^{(v)}(0^-, \cdot)] \hat{v}_{\xi, G}(\cdot, 1) - [\mu_1^{(v)}(0^+, \cdot) - \mu_1^{(v)}(0^-, \cdot)] \hat{v}_{\xi, G}(\cdot, 2)}{(\mu_2^{(v)}(0^+, \cdot) - \mu_2^{(v)}(0^-, \cdot))^2} \right)(\cdot)$$

$$\overset{p}{\underset{\xi}{\rightsquigarrow}} \Upsilon_W' \left(\frac{[\mu_2^{(v)}(0^+, \cdot) - \mu_2^{(v)}(0^-, \cdot)] \mathbb{G}(\cdot, 1) - [\mu_1^{(v)}(0^+, \cdot) - \mu_1^{(v)}(0^-, \cdot)] \mathbb{G}(\cdot, 2)}{(\mu_2^{(v)}(0^+, \cdot) - \mu_2^{(v)}(0^-, \cdot))^2} \right)(\cdot).$$

A proof is found in Appendix A.7.

7.2. Inference with robustness against no or weak jump or kink

Another practically relevant issue to consider as an extension to our baseline model is the issue of no and weak jumps in regression discontinuity designs. Feir et al. (2016) report significant size distortions in the event of weak jumps in the mean RDD, similarly to weak instrument problems. In addition, we also find size distortions in the event of weak jumps in the FQRD designs — see Section 5. Otsu et al. (2015) and Feir et al. (2016) provide weak-identification-robust methods of inference in regression discontinuity designs. In the spirit of Anderson and Rubin (1949) – also see Kleibergen (2002), Moreira (2003), and Andrews et al. (2006) – a null restriction can provide valid standard errors robustly against the possibility of weak and no jumps (Feir et al., 2016). Adopting this idea, we apply our baseline results of the weak convergence and the conditional weak convergence to non- and weak-identification-robust inference in this section.

Under the null hypothesis $H_0 : \tau = \tau^*$, we can rewrite (4.1) as

$$\phi \left(\lim_{x \downarrow 0} \mu_1^{(v)}(x, \cdot) \right)(\cdot) - \phi \left(\lim_{x \uparrow 0} \mu_1^{(v)}(x, \cdot) \right)(\cdot) = \left(\Upsilon^{-1} \tau^* \right)(\cdot) \left[\psi \left(\lim_{x \downarrow 0} \mu_2^{(v)}(x, \cdot) \right)(\cdot) - \psi \left(\lim_{x \uparrow 0} \mu_2^{(v)}(x, \cdot) \right)(\cdot) \right],$$

provided that the operator Υ is invertible. Notice that this characterization allows for a statistical inference without assuming that a nonzero jump or kink exists. To see this, note that $\psi \left(\lim_{x \downarrow 0} \mu_2^{(v)}(x, \cdot) \right)(\cdot) - \psi \left(\lim_{x \uparrow 0} \mu_2^{(v)}(x, \cdot) \right)(\cdot)$ no longer appears as a denominator, unlike the original expression (4.1). We use this equality restriction for a statistical inference. A sample-analog statistic can be defined as

$$\hat{\Xi}_\tau(\cdot) = \phi \left(\hat{\mu}_{1,p}^{(v)}(0^+, \cdot) \right)(\cdot) - \phi \left(\hat{\mu}_{1,p}^{(v)}(0^-, \cdot) \right)(\cdot) - \left(\Upsilon^{-1} \tau \right)(\cdot) \left[\psi \left(\hat{\mu}_{2,p}^{(v)}(0^+, \cdot) \right)(\cdot) - \psi \left(\hat{\mu}_{2,p}^{(v)}(0^-, \cdot) \right)(\cdot) \right].$$

Our identification-robust inference procedure is based on the idea that $\hat{\Xi}_{\tau^*}$ should be uniformly close to zero under the null hypothesis $H_0 : \tau = \tau^*$.

Assumption 2′ (*Conditional Weak Convergence*). (i) ψ and ϕ are Hadamard differentiable at $\mu_1^{(v)}(0^\pm, \cdot)$ and $\mu_2^{(v)}(0^\pm, \cdot)$, respectively, tangentially to some subspaces of their domains, with their Hadamard derivatives denoted by $\phi'_{\mu_1^{(v)}(0^\pm, \cdot)}$ and $\psi'_{\mu_2^{(v)}(0^\pm, \cdot)}$, respectively. Υ is invertible. (ii) $nh_n^{1+2v} \to \infty$ as $n \to \infty$.

Compared with Assumption 2, this assumption is weaker. Specifically, part (ii) of Assumption 2 has been dropped, and hence we do allow for a possibility of no jump or no kink. Similar lines of arguments to those of Theorems 1 and 2 yield the following weak and conditional weak convergence results robustly against possible non-identification due to a lack of jumps or kinks.

Corollary 3 (*Weak and Conditional Weak Convergence*). *Under Assumptions 1 and 2′ and the null hypothesis $H_0 : \tau = \tau^*$, we have $\sqrt{nh_n^{1+2v}} \hat{\Xi}_{\tau^*} \rightsquigarrow \mathbb{G}'(\cdot, 1) - (\Upsilon^{-1} \tau^*)(\cdot) \mathbb{G}'(\cdot, 2)$, where $\mathbb{G}'$ is defined in the statement of Theorem 1. If Assumptions 3 and 4 hold in addition, then $\hat{\mathbb{X}}_n'(\cdot, 1) - (\Upsilon^{-1} \tau^*)(\cdot) \hat{\mathbb{X}}_n'(\cdot, 2) \overset{p}{\underset{\xi}{\rightsquigarrow}} \mathbb{G}'(\cdot, 1) - (\Upsilon^{-1} \tau^*)(\cdot) \mathbb{G}'(\cdot, 2)$, where $\hat{\mathbb{X}}_n'$ is defined in the statement of Theorem 2.*

7.3. Covariates

Many applied researchers augment their RDD estimating equations with additional predetermined covariates. Formal justification and asymptotic properties of such a practice for mean sharp/fuzzy RDD/RKD are shown in Calonico et al. (2018b). In this section, we demonstrate that such a practice is also justified in the case of FQRD designs, and then provide a generalization of Corollary 1 for models with covariates.

Suppose that we have a d-dimensional vector of covariates denoted by Z_i. Define

$$\left(\check{\mu}_1(0^+, y, d), \check{\mu}_1'(0^+, y, d), \check{\mu}_1''(0^+, y, d), \check{\gamma}_{1+}(y, d) \right) =$$

$$\arg \min_{(\mu, \mu', \mu'', \gamma) \in \mathbb{R}^{3+d}} \sum_{i: X_i > 0} \left(\mathbb{1} \left\{ Y_i^* \leq y, D_i^* = d \right\} - \left\{ \mu + \mu' X_i + \frac{\mu''}{2!} X_i^2 + Z_i' \gamma \right\} \right)^2 \cdot K \left(\frac{X_i}{h_n} \right),$$

$$\left(\check{\mu}_2(0^+, d), \check{\mu}_2'(0^+, d), \check{\mu}_2''(0^+, d), \check{\gamma}_{2+}(d)\right) =$$

$$\arg \min_{(\mu, \mu', \mu'', \gamma) \in \mathbb{R}^{3+d}} \sum_{i:X_i>0} \left(\mathbb{1}\left\{D_i^* = d\right\} - \left\{ \mu + \mu'X_i + \frac{\mu''}{2!}X_i^2 + Z_i'\gamma \right\} \right)^2 \cdot K\left(\frac{X_i}{h_n}\right), \text{ and}$$

$$\check{F}_{Y^d|C}(y) = \frac{\check{\mu}_{1+}(y, d) - \check{\mu}_{1-}(y, d)}{\check{\mu}_{2+}(d) - \check{\mu}_{2-}(d)}.$$

The left-hand-side counterparts of them are defined analogously. The FQRD estimator based on these covariate-augmented local linear estimators is defined by

$$\check{\tau}_{FQRD}(\theta) = \check{Q}_{Y^1|C}(\theta) - \check{Q}_{Y^0|C}(\theta),$$

where

$$\check{Q}_{Y^d|C}(\theta) = \inf \left\{ y : \check{F}_{Y^d|C}(y) \geq \theta \right\} \quad \text{for each } d \in \{0, 1\}.$$

We next introduce the short-hand notations: $\gamma_{1+}(y, d) = \sigma_{Z+}^{-1}E[(Z_i - \mu_{Z+}(X_i))\mathbb{1}\{Y_i \leq y, D_i = d\}|X_i = 0^+]$, $\gamma_{2+}(d) = \sigma_{Z+}^{-1}E[(Z_i - \mu_{Z+}(X_i))\mathbb{1}\{D_i = d\}|X_i = 0^+]$, $\mu_{Z+}(x) = [\mu_{Z_1+}(x), \ldots, \mu_{Z_d+}(x)]'$, $\mu_{Z_l+}(x) = E[Z_{il}|X_i = x]$, and $\sigma_{Z+}^2 = V(Z_i|X_i = 0^+)$. Their left-hand-side counterparts are also defined analogously. Consider the following assumptions.

Assumption C. (i) $\gamma_1 = \gamma_{1+} = \gamma_{1-}$ and $\gamma_2 = \gamma_{2+} = \gamma_{2-}$ on $\mathscr{Y}_1 \times \{0, 1\}$. (ii) $\mu_{Z+} = \mu_{Z-}$ on $(\underline{x}, 0]$ and $[0, \overline{x})$. (iii) For all $(y, d) \in \mathscr{Y}_1 \times \{0, 1\}$, $E[Z_i\mathbb{1}\{Y_i \leq y, D_i = d\}, D_i]'|X = \cdot]$ are continuously differentiable on $(\underline{x}, 0]$ and $[0, \overline{x})$. (iv) $\mu_{Z+}(\cdot)$ and $\mu_{Z-}(\cdot)$ are three-times continuously differentiable on $(\underline{x}, 0]$ and $[0, \overline{x})$. (v) $\sigma_{Z\pm}^2(\cdot)$ are continuous and invertible on $(\underline{x}, 0]$ and $[0, \overline{x})$.

Assumption 4′ (*First Stage Estimation*). $\ddot{\mu}_{k,p}(x, \theta_k)\mathbb{1}\{|x/h_{k,n}(\theta_k)| \leq 1\}$ is uniformly consistent for $\mu_k(x, \theta_k)\mathbb{1}\{|x/h_{k,n}(\theta_k)| \leq 1\}$ on $([\underline{x}, \overline{x}]\setminus\{0\}) \times \mathbb{T}$. $\hat{f}_X(0)$ is consistent for $f_X(0)$.

With a first-stage estimator satisfying this assumption, we define $\check{U}_{1i} = \mathbb{1}\{Y_i^* \leq y, D_i^* = d\} - \check{\gamma}_{1+}'(y, d)\hat{\mu}_{Z+}(X_i) - \ddot{\mu}_1(X_i, y, d)$ and $\check{U}_{2i} = \mathbb{1}\{D_i^* = d\} - \check{\gamma}_{2+}'(d)\hat{\mu}_{Z+}(X_i) - \ddot{\mu}_2(X_i, d)$. Define the estimated multiplier processes

$$\check{v}_{\xi,n}^+(y, d, 1) = \sum_{i:X_i>0} \xi_i \frac{(1\ 0\ 0) \cdot (\Gamma_2^\pm)^{-1} \cdot \left(1\ \frac{X_i}{h_n}\ \frac{X_i^2}{h_n^2}\right)' \check{U}_{1i}K\left(\frac{X_i}{h_n}\right)}{\sqrt{nh_n}\hat{f}_X(0)}$$

and

$$\check{v}_{\xi,n}^+(d, 2) = \sum_{i:X_i>0} \xi_i \frac{(1\ 0\ 0) \cdot (\Gamma_2^\pm)^{-1} \cdot \left(1\ \frac{X_i}{h_n}\ \frac{X_i^2}{h_n^2}\right)' \check{U}_{2i}K\left(\frac{X_i}{h_n}\right)}{\sqrt{nh_n}\hat{f}_X(0)}.$$

We now define $\check{\mathbb{X}}_n'$, $\check{\mathbb{Y}}_n$ and $\check{\Upsilon}_W'$ as $\widehat{\mathbb{X}}_n'$, $\widehat{\mathbb{Y}}_n$ and $\widehat{\Upsilon}_W'$, respectively, but with $\check{v}_{\xi,n}^\pm$ in place of $\hat{v}_{\xi,n}^\pm$, $\check{Q}_{Y^d|C}$ in place of $\hat{Q}_{Y^d|C}$, and $\check{\mu}_{k,2}$ in place of $\hat{\mu}_{k,2}$.

Corollary 4. *Suppose that Assumptions S, K, M, FQRD, C and 4′ hold. Corollary 1 holds with $\check{\tau}_{FQRD}$ and $\check{\Upsilon}_W'(\check{\mathbb{Y}}_n)$ in place of $\hat{\tau}_{FQRD}$ and $\widehat{\Upsilon}_W'(\widehat{\mathbb{Y}}_n)$, respectively.*

A proof is found in Appendix A.9.

Remark 5. While many cases covered in our general framework can also be extended to covariate-augmented estimating equations *à la* Calonico et al. (2018b), not all the cases can be. The covariate extension is not compatible with the designs in which the operators, $\phi(\cdot)$ and $\psi(\cdot)$, are not linear, and hence we did not work with the general unified framework in this subsection. Since the cases of mean sharp/fuzzy RDD/RKD are covered in prior work (Calonico et al., 2018b), we focused on the case of the FQRD design, which is the main design of interest in this paper.

8. Summary

The existing literature on robust inference for causal effects in RDD and RKD covers major important cases, including the sharp mean RDD, the fuzzy mean RDD, the sharp mean RKD, the sharp fuzzy RKD, the sharp quantile RDD, and the sharp quantile RKD. These existing methods, however, do not cover uniform inference for CDF and quantile cases in fuzzy designs. Particularly, the existing methods are not able to handle robust uniform inference for quantile treatment effects in the fuzzy RDD, despite its practical relevance, e.g., Clark and Martorell (2014) and Deshpande (2016). In this light, this paper proposes a new general robust inference method that covers the fuzzy quantile RDD in particular, but also all the

other cases covered by the existing methods of robust inference. We provide high-level statements for the general result, and also provide more primitive conditions and detailed discussions focusing on the fuzzy quantile RDD.

Monte Carlo simulation studies confirm the theoretical properties for data generating processes calibrated to match real applications. Applying the proposed method to real data, we study causal effects on test outcomes of the Oklahoma pre-K program, following the earlier work by Gormley Jr. et al. (2005) and Frandsen et al. (2012). Despite the generally larger lengths of uniform confidence bands than point-wise confidence intervals, we obtain results qualitatively similar to those of Frandsen et al. (2012).

Finally, we conclude this paper with a guide for practitioners. Our method applies to most, if not all, of the commonly used local Wald estimators, but practitioners may want to know which method they should consider in empirical research. We are not aware yet of theoretical benefits of the multiplier bootstrap method over the analytic method – which is left for future research – and hence analytic methods may be preferred for their computational advantage for applications where such an analytic method is applicable. As such, the analytic method proposed by Calonico et al. (2014) is probably a superior option if a practitioner is interested in mean-regression-based designs. As their method is limited to mean-regression-based designs, one may need to seek alternative methods for quantile-regression-based designs. The pivotal methods of Qu and Yoon (2018) and Chiang and Sasaki (2019) are probably superior to our method if a practitioner is interested in sharp quantile RDD and sharp quantile RKD, respectively. The bootstrap method proposed in this paper, to the best of our knowledge, is the only option if a practitioner is interested in the fuzzy quantile RDD and the fuzzy CDF discontinuity design.

Appendix A. Mathematical appendix

The relation $a \lesssim b$ means that there exists C, $0 < C < \infty$, such that $a \leq Cb$. For the definitions and notations of what follows, we refer the reader to van der Vaart and Wellner (1996), Kosorok (2003), and Giné and Nickl (2015). For an arbitrary semimetric space (T, d), define the covering number $N(\epsilon, T, d)$ to be the minimal number of closed d-balls of radius ϵ required to cover T. For any function $f : T \times \Omega \mapsto \mathbb{R}$ and any $w \in \Omega$, denote $\|f(\cdot, w)\|_T = \sup_{t \in T} |f(t, w)|$. We further define the Uniform Entropy Integral $J(\delta, \mathscr{F}, F) = \sup_Q \int_0^\delta \sqrt{1 + \log N(\epsilon \|F\|_{Q,2}, \mathscr{F}, \|\cdot\|_{Q,2})} d\epsilon$, where $\|\cdot\|_{Q,2}$ is the $L2$ norm with respect to measure Q and the supremum is taken over all probability measure over $(\Omega^x, \mathcal{F}^x)$. A class of measurable functions $\mathscr{F}$ is called a VC (Vapnik–Chervonenkis) type with respect to a measurable envelope F of $\mathscr{F}$ if there exist finite constants $A \geq 1$, $V \geq 1$ such that for all probability measures Q on $(\Omega^x, \mathcal{F}^x)$, we have $N(\epsilon \|F\|, \mathscr{F}, \|\cdot\|_{Q,2}) \leq (\frac{A}{\epsilon})^V$, for $0 < \epsilon \leq 1$. We use C, C_1, C_2,..., to denote constants that are positive and independent of n. The values of C may change at each appearance but C_1, C_2,..., are fixed.

A.1. Uniform Bahadur representation

We obtain the uniform BR of the local polynomial estimators $\hat{\mu}_{1,p}^{(v)}(0^\pm, \cdot)$ and $\hat{\mu}_{2,p}^{(v)}(0^\pm, \cdot)$ presented in the following lemma — see Appendix A.1 for a proof.

Lemma 1 (*Uniform Bahadur Representation*)**.** *Under Assumption 1, we have:*

$$\sqrt{nh_{k,n}^{1+2v}(\theta_k)}\left(\hat{\mu}_{k,p}^{(v)}(0^\pm, \theta_k) - \mu_k^{(v)}(0^\pm, \theta_k) - h_{k,n}^{p+1-v}(\theta_k)\frac{e_v'(\Gamma_p^\pm)^{-1}\Lambda_{p,p+1}^\pm}{(p+1)!}\mu_k^{(p+1)}(0^\pm, \theta_k)\right)$$
$$= v! \sum_{i=1}^n \frac{e_v'(\Gamma_p^\pm)^{-1}\mathcal{E}_k(Y_i, D_i, X_i, \theta)r_p(\frac{X_i}{h_{k,n}(\theta_k)})K(\frac{X_i}{h_{k,n}(\theta_k)})\delta_i^\pm}{\sqrt{nh_{k,n}(\theta_k)}f_X(0)} + o_p^x(1)$$

uniformly for all $\theta_k \in \Theta_k$ for each $k \in \{1, 2\}$.

Notice that the leading bias terms on the left-hand side of the equations in this lemma are of the $(p+1)$-st order. Thus, the asymptotic distributions of the Bahadur representation take into account the pth order bias reduction. This property is the key to develop a method of inference which is robust against large bandwidth parameter choices.

Remark 6. It is worth remarking on the relation between this lemma and Theorems 2 and 3 in Frandsen et al. (2012). The proofs of Theorem 2 and 3 in Frandsen et al. (2012) rely on the exact solution of the local linear smoother (see their equation (11)), while ours makes use of the uniform Bahadur representation derived in our Lemma 1. As illustrated in Section 4.2 of Dony et al. (2006), if we let $\tilde{f}_{n,h,j} = \frac{1}{nh}\sum_{i=1}^n \left(\frac{X_i-x}{h}\right)^j K\left(\frac{x-X_i}{h}\right)$ and $\tilde{r}_{n,h,j} = \frac{1}{nh}\sum_{i=1}^n Y_i\left(\frac{X_i-x}{h}\right)^j K\left(\frac{x-X_i}{h}\right)$, then the exact solution for the local linear mean regression estimator ($p = 1$) has the form $\hat{\mu}(x) = \frac{\tilde{f}_{n,h,2}\tilde{r}_{n,h,0} - \tilde{f}_{n,h,1}\tilde{r}_{n,h,1}}{\tilde{f}_{n,h,0}\tilde{f}_{n,h,2} - \tilde{f}_{n,h,1}^2}$. For the local quadratic mean regression estimator ($p = 2$), $\hat{\mu}(x) = \frac{(\tilde{f}_{n,h,2}\tilde{f}_{n,h,4} - \tilde{f}_{n,h,3}^2)\tilde{r}_{n,h,0} + (\tilde{f}_{n,h,2}\tilde{f}_{n,h,3} - \tilde{f}_{n,h,1}\tilde{f}_{n,h,4})\tilde{r}_{n,h,1} + (\tilde{f}_{n,h,1}\tilde{f}_{n,h,3} - \tilde{f}_{n,h,2}^2)\tilde{r}_{n,h,2}}{\tilde{f}_{n,h,0}\tilde{f}_{n,h,2}\tilde{f}_{n,h,4} - \tilde{f}_{n,h,0}\tilde{f}_{n,h,3}^2 - \tilde{f}_{n,h,1}^2\tilde{f}_{n,h,4} + \tilde{f}_{n,h,1}\tilde{f}_{n,h,2}\tilde{f}_{n,h,3} - \tilde{f}_{n,h,2}^2}$ is the exact solution. Therefore, for kernel estimators with the squared loss, even though it is in principle possible to establish results by working with the exact solution for any order of polynomial, the expression of exact solution becomes rather

complex as the order p increases due to the presence of an inverse factor, whose determinant gives the denominator terms of the above expressions. Further, this illustration is only for mean regressions, and it would add even more complications if we have an index θ to keep track of in this inverse factor. By working with its limit $\Gamma_p^{\pm}$, our uniform Bahadur representation holds with the same unified general expression for any arbitrary order, with any index set that satisfies Assumption 1 (ii) (a). This helps us to allow for robust bias correction of any order in a unified manner, in comparison with the exact solution approach taken by Frandsen et al. (2012).

Proof. In this proof, we will show the first result for the case of $k = 1$ and $\pm = +$. All the other results can be shown by similar lines of proof. As in Section 1.6 of Tsybakov (2008), the solution to (4.4) can be computed explicitly as $\hat{\alpha}_{1+,p}(\theta_1) =$
$[\hat{\mu}_{1,p}(0^{\pm}, \theta_1), \hat{\mu}_{1,p}^{(1)}(0^{\pm}, \theta_1)h_{1,n}(\theta_1), \ldots, \hat{\mu}_{1,p}^{(p)}(0^{\pm}, \theta_1)h_{1,n}^p(\theta_1)/p!]' = \left[\frac{1}{nh_{1,n}(\theta_1)} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})r_p'(\frac{X_i}{h_{1,n}(\theta_1)}) \right]^{-1} \cdot$
$\left[\frac{1}{nh_{1,n}(\theta_1)} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})g_1(Y_i, \theta_1) \right]$ For each data point $X_i > 0$, a mean value expansion by Assumption 1 (ii)(b) gives $g_1(Y_i, \theta_1) = \mu_1(X_i, \theta) + \mathcal{E}_1(Y_i, D_i, X_i, \theta) = \mu_1(0, \theta) + \mu_1^{(1)}(0^+, \theta_1)X_i + \cdots + \mu_1^{(p)}(0^+, \theta_1)\frac{X_i^p}{p!} + \mu_1^{(p+1)}(x_{ni}^*, \theta_1)\frac{X_i^{(p+1)}}{(p+1)!} +$
$\mathcal{E}_1(Y_i, D_i, X_i, \theta) = r_p'(\frac{X_i}{h_{1,n}(\theta_1)})\alpha_{1+,p}(\theta_1) + \mu_1^{(p+1)}(x_{ni}^*, \theta_1)h_{1,n}^{p+1}(\theta_1)\frac{(\frac{X_i}{h_{1,n}(\theta_1)})^{(p+1)}}{(p+1)!} + \mathcal{E}_1(Y_i, D_i, X_i, \theta)$ for an $x_{ni}^* \in (0, X_i]$. Substituting this expansion in the equation above and multiplying both sides by $\sqrt{nh_{1,n}(\theta_1)}e_v'$, we obtain

$$\sqrt{nh_{1,n}(\theta_1)}\hat{\mu}_1^{(v)}(0^+, \theta_1)h_{1,n}^v(\theta_1)/v! = \sqrt{nh_{1,n}(\theta_1)}\mu_1^{(v)}(0^+, \theta_1)h_{1,n}^v(\theta_1)/v! + (a) + (b)$$

where

$$(a) = e_v'\left[\frac{1}{nh_{1,n}(\theta_1)} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})r_p'(\frac{X_i}{h_{1,n}(\theta_1)}) \right]^{-1} \cdot$$
$$\frac{1}{\sqrt{nh_{1,n}(\theta_1)}} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})\mu_1^{(p+1)}(x_{ni}^*, \theta_1)h_{1,n}^{p+1}(\theta_1)\frac{(\frac{X_i}{h_{1,n}(\theta_1)})^{p+1}}{(p+1)!}$$

and

$$(b) = e_v'\left[\frac{1}{nh_{1,n}(\theta_1)} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})r_p'(\frac{X_i}{h_{1,n}(\theta_1)}) \right]^{-1} \cdot$$
$$\frac{1}{\sqrt{nh_{1,n}(\theta_1)}} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})\mathcal{E}_1(Y_i, D_i, X_i, \theta)$$

We will show stochastic limits of the (a) and (b) terms above.

Step 1 First, we consider their common inverse factor. Specifically, we show that

$$[\frac{1}{nh_{1,n}(\theta_1)} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})r_p'(\frac{X_i}{h_{1,n}(\theta_1)})]^{-1} \xrightarrow{p} (\Gamma_p^+)^{-1}/f_X(0) \tag{A.1}$$

uniformly in θ_1. Note that by Minkowski's inequality

$$\left| [\frac{1}{nh_{1,n}(\theta_1)} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})r_p'(\frac{X_i}{h_{1,n}(\theta_1)})]^{-1} - (\Gamma_p^+)^{-1}/f_X(0) \right|_{\Theta_1}$$
$$\leq \left| [\frac{1}{nh_{1,n}(\theta_1)} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})r_p'(\frac{X_i}{h_{1,n}(\theta_1)})]^{-1} \right.$$
$$\left. - E[[\frac{1}{nh_{1,n}(\theta_1)} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})r_p'(\frac{X_i}{h_{1,n}(\theta_1)})]^{-1}] \right|_{\Theta_1}$$
$$+ \left| E[[\frac{1}{nh_{1,n}(\theta_1)} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})r_p'(\frac{X_i}{h_{1,n}(\theta_1)})]^{-1}] - (\Gamma_p^+)^{-1}/f_X(0) \right|_{\Theta_1}$$

where the first term on the right hand side is stochastic, while the second term is deterministic. First, regarding the deterministic part, $E[\frac{1}{nh_{1,n}(\theta_1)} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})r_p'(\frac{X_i}{h_{1,n}(\theta_1)})] = f_X(0)\Gamma_p^+ + O(h_{1,n}(\theta_1))$ holds uniformly by Assumption 1(i), (iii), and (iv). For the stochastic part, we will show that each entry of such a matrix converges in probability with respect to $\mathbb{P}^x$ uniformly. We may write $\mathscr{F}_r = \{x \mapsto \mathbb{1}\{x \geq 0\}K(ax)(ax)^r\mathbb{1}\{ax \in [-1, 1]\} : a > 1/h_0\}$ and $\mathscr{F}_{n,r} = \{x \mapsto \mathbb{1}\{x \geq 0\}K(x/h_{1,n}(\theta_1))(x/h_{1,n}(\theta_1))^r : \theta_1 \in \Theta_1\}$ for each integer r such that $0 \leq r \leq 2p$. By Lemma 6, each $\mathscr{F}_r$ is of VC type (Euclidean) with envelope $F = \|K\|_\infty$ under Assumption 1 (iii) and (iv)(a) and (b), i.e., there exist constants $k, v < \infty$ such that $\sup_Q \log N(\epsilon \|F\|_{Q,2}, \mathscr{F}_r, \|\cdot\|_{Q,2}) \leq (\frac{k}{\epsilon})^v$ for $0 < \epsilon \leq 1$ and for all probability measures Q supported on

$[\underline{x}, \overline{x}]$. This implies $J(1, \mathscr{F}_r, F) = \sup_Q \int_0^1 \sqrt{1 + \log N(\epsilon \|F\|_{Q,2}, \mathscr{F}_r, \|\cdot\|_{Q,2})} d\epsilon < \infty$. Since $F \in L_2(P)$, we can apply Theorem 5.2 of Chernozhukov et al. (2014) to obtain $E\left[\sup_{f \in \mathscr{F}_r} \left|\frac{1}{\sqrt{n}} \sum_{i=1}^n \left(f(X_i) - Ef\right)\right|\right] \leq C\{J(1, \mathscr{F}_r, F) \|F\|_{P,2} + \frac{\|K\|_\infty J^2(1, \mathscr{F}_r, F)}{\delta^2 \sqrt{n}}\} < \infty$ for a universal constant $C > 0$. Note that $\mathscr{F}_{n,r} \subset \mathscr{F}_r$ for all $n \in \mathbb{N}$. Multiplying both sides by $[\sqrt{n} h_{1,n}(\theta_1)]^{-1}$ yields

$$E\left[\sup_{\theta_1 \in \Theta_1} \left|\frac{1}{nh_{1,n}(\theta_1)} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)})(\frac{X_i}{h_{1,n}(\theta_1)})^s - \frac{1}{nh_{1,n}(\theta_1)} \sum_{i=1}^n E[\delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)})(\frac{X_i}{h_{1,n}(\theta_1)})^s]\right|\right]$$

$$\leq \frac{1}{\sqrt{n} h_{1,n}(\theta_1)} C\{J(1, \mathscr{F}_r, F) \|F\|_{P,2} + \frac{BJ^2(1, \mathscr{F}_r, F)}{\delta^2 \sqrt{n}}\} = O(\frac{1}{\sqrt{n} h_n})$$

The last line goes to zero uniformly under Assumption 1(iii). Finally, Markov's inequality gives the uniform convergence of the stochastic part at the rate $O_p^x(\frac{1}{\sqrt{n} h_n})$. Consequently, we have the uniform convergence in probability for each $r \in \{0, \ldots, 2p\}$. Assumption 1 (iv)(c) and the continuous mapping theorem concludes (A.1).

Step 2 For term (a), we may again use Minkowski's inequality under the supremum norm as in Step 1 to decompose

$$\left|\frac{1}{\sqrt{n} h_{1,n}(\theta_1)} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)}) r_p(\frac{X_i}{h_{1,n}(\theta_1)}) \mu_1^{(p+1)}(x_{ni}^*, \theta_1) h_{1,n}^{p+1}(\theta_1) \frac{(\frac{X_i}{h_{1,n}(\theta_1)})^{p+1}}{(p+1)!} - 0\right|_{\Theta_1}$$

$$\leq \left|\frac{1}{\sqrt{n} h_{1,n}(\theta_1)} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)}) r_p(\frac{X_i}{h_{1,n}(\theta_1)}) \mu_1^{(p+1)}(x_{ni}^*, \theta_1) h_{1,n}^{p+1}(\theta_1) \frac{(\frac{X_i}{h_{1,n}(\theta_1)})^{p+1}}{(p+1)!}\right.$$

$$\left. - E[\frac{1}{\sqrt{n} h_{1,n}(\theta_1)} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)}) r_p(\frac{X_i}{h_{1,n}(\theta_1)}) \mu_1^{(p+1)}(x_{ni}^*, \theta_1) h_{1,n}^{p+1} \frac{(\frac{X_i}{h_{1,n}(\theta_1)})^{p+1}}{(p+1)!}]\right|_{\Theta_1}$$

$$+ \left|E[\frac{1}{\sqrt{n} h_{1,n}(\theta_1)} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)}) r_p(\frac{X_i}{h_{1,n}(\theta_1)}) \mu_1^{(p+1)}(x_{ni}^*, \theta_1) h_{1,n}^{p+1}(\theta_1) \frac{(\frac{X_i}{h_{1,n}(\theta_1)})^{p+1}}{(p+1)!}] - 0\right|_{\Theta_1}$$

Under Assumption 1(i),(ii)(b),(iii),(iv)(a), standard calculations show that the deterministic part

$$E[\frac{1}{\sqrt{n} h_{1,n}(\theta_1)} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)}) r_p(\frac{X_i}{h_{1,n}(\theta_1)}) \mu_1^{(p+1)}(x_{ni}^*, \theta_1) h_{1,n}^{p+1}(\theta_1) \frac{(\frac{X_i}{h_{1,n}(\theta_1)})^{p+1}}{(p+1)!}]$$

$$= \frac{h_{1,n}^{p+1}(\theta_1) \Lambda_{p,p+1}^+}{\sqrt{n} h_{1,n}(\theta_1)(p+1)!} f_X(0) \mu_1^{(p+1)}(0^+, \theta_1) + O(\frac{h_n^{p+2}}{\sqrt{n} h_{1,n}}) = O(\sqrt{\frac{h_n^{2p+1}}{n}})$$

uniformly in θ_1.

As for the stochastic part, first note that under Assumption 1(ii), we know that for a Lipschitz constant L such that $0 \leq L < \infty$, it holds uniformly in θ_1 that

$$\left|\frac{1}{\sqrt{n} h_{1,n}(\theta_1)} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)}) r_p(\frac{X_i}{h_{1,n}(\theta_1)}) h_{1,n}^{p+1}(\theta_1) \frac{(\frac{X_i}{h_{1,n}(\theta_1)})^{p+1}}{(p+1)!} [\mu_1^{(p+1)}(x_{ni}^*, \theta_1) - \mu_1^{(p+1)}(0^+, \theta_1)]\right|$$

$$\lesssim n \max_{1 \leq i \leq n} \left|\frac{1}{\sqrt{n} h_{1,n}(\theta_1)} \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)}) r_p(\frac{X_i}{h_{1,n}(\theta_1)}) h_{1,n}^{p+1}(\theta_1) \frac{(\frac{X_i}{h_{1,n}(\theta_1)})^{p+1}}{(p+1)!} [\mu_1^{(p+1)}(x_{ni}^*, \theta_1) - \mu_1^{(p+1)}(0^+, \theta_1)]\right|$$

$$\lesssim \frac{n \max_{1 \leq i \leq n} |\mu_1^{(p+1)}(x_{ni}^*, \theta_1) - \mu_1^{(p+1)}(0^+, \theta_1)| h_{1,n}^{p+1}(\theta_1)}{\sqrt{n} h_n} \leq \frac{nLh_n^{p+2}}{\sqrt{n} h_n} = O_p^x(\sqrt{n h_n^{2p+3}}).$$

The second inequality holds since $\delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)}) r_p(\frac{X_i}{h_{1,n}(\theta_1)}) \frac{(\frac{X_i}{h_{1,n}(\theta_1)})^{p+1}}{(p+1)!}$ is bounded under Assumption 1(iii) and (iv)(a), while the third one is due to Assumption 1(ii). It is then sufficient to consider the asymptotic behavior of $\frac{1}{\sqrt{n} h_{1,n}(\theta_1)} \sum_{i=1}^n \delta_i^+ K$

$(\frac{X_i}{h_{1,n}(\theta_1)}) r_p(\frac{X_i}{h_{1,n}(\theta_1)}) \mu_1^{(p+1)}(0^+, \theta_1) h_{1,n}^{p+1}(\theta_1) \frac{(\frac{X_i}{h_{1,n}(\theta_1)})^{p+1}}{(p+1)!}$. Note that $r_p(x/h)(x/h)^{p+1} = [(x/h)^{p+1}, \ldots, (x/h)^{2p+1}]'$. So we may let

$$\mathscr{F}_s = \{x \mapsto \mathbb{1}\{x \geq 0\} K(ax)(ax)^{s+p+1} \mu_1^{p+1}(0^+, \theta_1) \mathbb{1}\{ax \in [-1, 1]\} : a \geq 1/h_0, \theta_1 \in \Theta_1\} \quad \text{and}$$

$$\mathscr{F}_{n,s} = \{x \mapsto \mathbb{1}\{x \geq 0\} K(x/h_{1,n}(\theta_1))(x/h_{1,n}(\theta_1))^{s+p+1} \mu_1^{p+1}(0^+, \theta_1) : \theta_1 \in \Theta_1\}$$

for each integer s such that $0 \leq s \leq p$. Since $(ax)^{s+p+1} \mathbb{1}\{ax \in [-1, 1]\}$ is Lipschitz continuous for each $a \geq 1/h_0$ and bounded by 1, it is of VC type by Lemma 5. We then apply Lemma 6 to show that for each $0 \leq s \leq p$, $\mathscr{F}_s$ is a VC type class

with envelope $F_s(x) = \|K\|_\infty \int_{\mathcal{Y} \times \mathcal{D}} F_\epsilon(y, d', x) d\mathbb{P}^x(y, D = d'|x)$, which is integrable under Assumption 1(ii)(b) and (iv)(a). By Assumption 1 (iii) and (iv)(a), $\mathscr{F}_{n,s} \subset \mathscr{F}_s$ for all $n \in \mathbb{N}$, thus an argument similar to the one above with Theorem 5.2 of Chernozhukov et al. (2014) shows that for each $0 \le s \le p$

$$E[\sup_{f \in \mathscr{F}_s} |\frac{1}{\sqrt{n}} \sum_{i=1}^n (f(X_i) - Ef(X_i))|] = O(1) \tag{A.2}$$

To prove the uniform convergence of the part of (a) outside the inverse sign, it suffices to show that for each s $\sup_{f \in \mathscr{F}_s} |\frac{1}{\sqrt{nh_n}} \sum_{i=1}^n (f(X_i)h_n^{p+1} - E[f(X_i)h_n^{p+1}])| \xrightarrow{P}_x 0$. Multiplying both sides of Eq. (A.2) by $h_n^{p+1}/\sqrt{h_n}$ and applying Markov's inequality as in Step 1, we have

$$\sup_{f \in \mathscr{F}_s} |\frac{1}{\sqrt{nh_n}} \sum_{i=1}^n (f(X_i)h_n^{p+1} - Ef(X_i)h_n^{p+1})| = O_p^x(\sqrt{\frac{h_n^{2p+1}}{n}}) \tag{A.3}$$

which converges to zero in probability ($\mathbb{P}^x$) under Assumption 1(iii). To conclude, we have shown

$$\frac{1}{\sqrt{nh_{1,n}(\theta_1)}} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)}) r_p(\frac{X_i}{h_{1,n}(\theta_1)}) \mu_1^{(p+1)}(x_{ni}^*, \theta_1) h_{1,n}^{p+1}(\theta_1) \frac{(X_i/h_{l,n}(\theta_1))^{p+1}}{(p+1)!}$$

$$= O(\sqrt{\frac{h_n^{2p+1}}{n}}) + O_p^x(\sqrt{nh_n^{2p+3}}) + O_p^x(\sqrt{\frac{h_n^{2p+1}}{n}})$$

uniformly in θ_1. Finally, the continuous mapping theorem gives $(a) \xrightarrow{p} \left(O(h_n) + O_p^x(\frac{1}{\sqrt{nh_n}}) \right) \left(O(\sqrt{\frac{h_n^{2p+1}}{n}}) + O_p^x(\sqrt{nh_n^{2p+3}}) + O_p^x(\sqrt{\frac{h_n^{2p+1}}{n}}) \right) = o_p^x(1)$ uniformly in θ_1.

Step 3 For term (b), Minkowski's inequality under the supremum norm implies

$$\left| \frac{1}{\sqrt{nh_{1,n}(\theta_1)}} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)}) r_p(\frac{X_i}{h_{1,n}(\theta_1)}) \mathcal{E}_1(Y_i, D_i, X_i, \theta) \right|_{\Theta_1}$$

$$\le \left| \frac{1}{\sqrt{nh_{1,n}(\theta_1)}} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)}) r_p(\frac{X_i}{h_{1,n}(\theta_1)}) \mathcal{E}_1(Y_i, D_i, X_i, \theta) \right.$$

$$\left. - E[\frac{1}{\sqrt{nh_{1,n}(\theta_1)}} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)}) r_p(\frac{X_i}{h_{1,n}(\theta_1)}) \mathcal{E}_1(Y_i, D_i, X_i, \theta)] \right|_{\Theta_1}$$

$$+ \left| E[\frac{1}{\sqrt{nh_{1,n}(\theta_1)}} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)}) r_p(\frac{X_i}{h_{1,n}(\theta_1)}) \mathcal{E}_1(Y_i, D_i, X_i, \theta)] - 0 \right|_{\Theta_1}.$$

By construction, local polynomial regression satisfies $E[\mathcal{E}_1(Y_i, D_i, X_i, \theta)|X] = E[g(Y_i, \theta_1) - \mu_1(X_i, \theta_1)|X] = 0$, thus by the law of iterated expectations, we have

$$E[\frac{1}{\sqrt{nh_{1,n}(\theta_1)}} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)}) r_p(\frac{X_i}{h_{1,n}(\theta_1)}) \mathcal{E}_1(Y_i, D_i, X_i, \theta)]$$

$$= E[\frac{1}{\sqrt{nh_{1,n}(\theta_1)}} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)}) r_p(\frac{X_i}{h_{1,n}(\theta_1)}) E[\mathcal{E}_1(Y_i, D_i, X_i, \theta)|X]] = 0.$$

Therefore, in light of (A.1), in order to show $(b) \xrightarrow{p} \sum_{i=1}^n \frac{(\Gamma_p^+)^{-1} \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)}) r_p(\frac{X_i}{h_{1,n}(\theta_1)}) \mathcal{E}_1(Y_i, D_i, X_i, \theta)}{\sqrt{nh_{1,n}(\theta_1)} f_X(0)}$ uniformly in θ_1, it remains to be shown that for each coordinate $0 \le s \le p$

$$\sup_{\theta_1 \in \Theta_1} \left| \frac{e_s'}{\sqrt{nh_{1,n}(\theta_1)}} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)}) r_p(\frac{X_i}{h_{1,n}(\theta_1)}) \mathcal{E}_1(Y_i, D_i, X_i, \theta) \right| = O_p^x(1).$$

First note that, under Assumptions 1(i), (ii) (c), (iii), (iv),

$$
\sup_{\theta_1 \in \Theta_1} E[\frac{e_s'}{\sqrt{nh_{1,n}(\theta_1)}} \sum_{i=1}^{n} \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)}) r_p(\frac{X_i}{h_{1,n}(\theta_1)}) \mathcal{E}_1(Y_i, D_i, X_i, \theta)]^2
$$

$$
= \sup_{\theta_1 \in \Theta_1} \int_{\mathbb{R}_+} \sigma_{11}(\theta_1, \theta_1|0^+) K^2(u) e_s' r_p(u) r_p'(u) e_s f_X(0^+) du + O_p^x(h_n)
$$

$$
\leq f_X(0) e_s' \Psi_p^+ e_s \sup_{\theta_1 \in \Theta_1} \sigma_{11}(\theta_1, \theta_1|0^+) + O_p^x(h_n) \lesssim f_X(0) e_s' \Psi_p^+ e_s + O_p^x(h_n),
$$

where the right hand side is bounded and does not depend on θ_1. Using Markov's inequality, we know that for some constant that does not depend on n, for each $M > 0$

$$
\mathbb{P}^x\Big(\Big|\frac{e_s'}{\sqrt{nh_{1,n}(\theta_1)}} \sum_{i=1}^{n} \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)}) r_p(\frac{X_i}{h_{1,n}(\theta_1)}) \mathcal{E}_1(Y_i, D_i, X_i, \theta)\Big| > M\Big) \leq \frac{C[f_X(0) e_s' \Psi_p^+ e_s + h_n]}{M^2}.
$$

Since the right hand side can be made arbitrarily small by choosing M, it is bounded in probability uniformly in θ_1, which concludes the proof. □

Remark 7. As an anonymous referee pointed out, it is worth discussing the proof of Lemma 1 in comparison to Einmahl and Mason (2000). The main difference between their results and our convergence of stochastic parts is that the rates are different; the framework of Einmahl and Mason (2000) is uniform in both x and a set of functions. Therefore, their rate has an extra $\log h_n$ factor and thus the estimator based on their result does not converge weakly to a Gaussian process.

In addition, their results are about the uniform convergence of a kernel type estimator to its expectation, which can be considered counterparts of the stochastic parts in Steps 2 and 3 in our proof of Lemma 1. This is insufficient, and we also need to handle the asymptotic bias terms since our uniform Bahadur representation is only an approximation. Note that much of our proof for Lemma 1 is about establishing the uniform validity of such an approximation.

A.2. Auxiliary lemmas for the general result

Since we are working on two probability spaces, $(\Omega^x, \mathcal{F}^x, \mathbb{P}^x)$ and $(\Omega^\xi, \mathcal{F}^\xi, \mathbb{P}^\xi)$, we use the following notations to clarify the sense of various modes of convergence and expectations. We let $\overset{p}{\underset{\bullet}{\to}}$ denote the convergence in probability with respect to the probability measure $\mathbb{P}^\bullet$, let $E_{\xi|x}$ denote the conditional expectation with respect to the product probability measure $\mathbb{P}^x \times \mathbb{P}^\xi$ given the events in $\mathcal{F}^x$, and let E_x denote the expectation with respect to the probability measure $\mathbb{P}^x$. Following Section 1.13 of van der Vaart and Wellner (1996), we define the conditional weak convergence in probability, or convergence of the conditional limit laws of bootstraps, denoted by $X_n \overset{p}{\underset{\xi}{\rightsquigarrow}} X$, by $\sup_{h \in BL_1} |E_{\xi|x} h(X_n) - Eh(X)| \overset{p}{\underset{x}{\to}} 0$, where BL_1 is the set of functions with Lipschitz constant and supremum norm bounded by 1. We state and prove the following lemma, which can be seen as a conditional weak convergence analogy of Theorem 18.10 (iv) of van der Vaart (1998).

The following lemma is used for the purpose of bounding estimation errors in the EMP to approximate the MP.

Lemma 2. *Let $(\Omega^x \times \Omega^\xi, \mathcal{F}^x \otimes \mathcal{F}^\xi, \mathbb{P}^{x \times \xi})$ be the product probability space of $(\Omega^x, \mathcal{F}^x, P^x)$ and $(\Omega^\xi, \mathcal{F}^\xi, P^\xi)$, where $\mathcal{F}^x \otimes \mathcal{F}^\xi$ stands for the product sigma field of $\mathcal{F}^x$ and $\mathcal{F}^\xi$. For a metric space $(\mathbb{T}, d)$, consider $X_n, Y_n, X : \Omega^x \times \Omega^\xi \to \mathbb{T}$, $n = 1, 2, \ldots$. If $X_n \overset{p}{\underset{\xi}{\rightsquigarrow}} X$ and $d(Y_n, X_n) \overset{p}{\underset{x \times \xi}{\to}} 0$, then $Y_n \overset{p}{\underset{\xi}{\rightsquigarrow}} X$.*

Proof. For each $h \in BL_1$, we can write $|E_{\xi|x} h(Y_n) - Eh(X)| \leq |E_{\xi|x} h(Y_n) - E_{\xi|x} h(X_n)| + |E_{\xi|x} h(X_n) - Eh(X)|$. For the second term on the right-hand side, $|E_{\xi|x} h(X_n) - Eh(X)| \overset{p}{\underset{\xi}{\to}} 0$ by the assumption $X_n \overset{p}{\underset{\xi}{\rightsquigarrow}} X$ and the definition of BL_1. To analyze the first term on the right-hand side, note that for any $\varepsilon \in (0, 1)$, we have $|E_{\xi|x} h(Y_n) - E_{\xi|x} h(X_n)| \leq \varepsilon E_{\xi|x} \mathbb{1}\{d(X_n, Y_n) \leq \varepsilon\} + 2E_{\xi|x} \mathbb{1}\{d(X_n, Y_n) > \varepsilon\}$. The first part can be set arbitrarily small by letting $\varepsilon \to 0$. To bound the second part, note first that the assumption of $d(Y_n, X_n) \overset{p}{\underset{x \times \xi}{\to}} 0$ yields $\lim_{n \to \infty} \mathbb{P}^{x \times \xi}(d(X_n, Y_n) > \varepsilon) = \lim_{n \to \infty} E[\mathbb{1}\{d(X_n, Y_n) > \varepsilon\}] = 0$. By the law of iterated expectations and the dominated convergence theorem, we obtain $\lim_{n \to \infty} E[\mathbb{1}\{d(X_n, Y_n) > \varepsilon\}] = \lim_{n \to \infty} E_x[E_{\xi|x}[\mathbb{1}\{d(X_n, Y_n) > \varepsilon\}]] = E_x[\lim_{n \to \infty} E_{\xi|x}[\mathbb{1}\{d(X_n, Y_n) > \varepsilon\}]] = 0$ In other words, $\lim_{n \to \infty} E_{\xi|x}[\mathbb{1}\{d(X_n, Y_n) > \varepsilon\}] = 0$ $\mathbb{P}^x$-almost surely. □

The following lemma will be used for deriving the weak convergence of Wald type statistics from joint weak convergence results for the numerator and denominator processes. It can be easily checked by using definition of Hadamard differentiation, and so we omit a proof.

Lemma 3. *Let $(A(\cdot), B(\cdot)) \in \ell^\infty(\Theta) \times \ell^\infty(\Theta)$. If $B(\cdot) > C > 0$ on Θ, then $(F, G) \overset{\Phi}{\mapsto} F/G$ is Hadamard differentiable at (A, B) tangentially to $\ell^\infty(\Theta)$ with the Hadamard derivative $\Phi'_{(A,B)}$ given by $\Phi'_{(A,B)}(g, h) = (Bg - Ah)/B^2$.*

We restate the Functional Central Limit Theorem of Pollard (1990) as the following lemma, which plays a pivotal role in the proof of our main Theorem. To cope with some measurability issues, we present the version with sufficient conditions for measurability by Kosorok (2003, Lemma1). See also Theorem 10.6 of Pollard (1990).

Lemma 4 (*Pollard, 1990; Kosorok, 2003*)**.** *Denote outer expectation, as defined in Section 1.2 of* van der Vaart and Wellner *(1996), by E^*. Let a triangular array of almost measurable Suslin (AMS) stochastic processes $\{f_{ni}(t) : i = 1, \ldots, n, t \in T\}$ be row independent, and define $v_n(t) = \sum_{i=1}^{n}[f_{ni}(t) - Ef_{ni}(\cdot, t)]$. Define $\rho_n(s, t) = (\sum_{i=1}^{n}[f_{ni}(s) - f_{ni}(t)]^2)^{1/2}$. Suppose that the following conditions are satisfied. 1. The $\{f_{ni}\}$ are manageable, with envelope $\{F_{ni}\}$ which are also independent within rows; 2. $H(s, t) = \lim_{n \to \infty} E v_n(s)v_n(t)$ exists for every $s, t \in T$; 3. $\limsup_{n \to \infty} \sum_{i=1}^{n} E^* F_{ni}^2 < \infty$; 4. $\lim_{n \to \infty} \sum_{i=1}^{n} E^* F_{ni}^2 \mathbb{1}\{F_{ni} > \epsilon\} = 0$ for each $\epsilon > 0$; 5. $\rho(s, t) = \lim_{n \to \infty} \rho_n(s, t)$ exists for every $s, t \in T$, and for all deterministic sequences $\{s_n\}$ and $\{t_n\}$ in T, if $\rho(s_n, t_n) \to 0$ then $\rho_n(s_n, t_n) \to 0$. Then T is totally bounded under the ρ pseudometric and X_n converges weakly to a tight mean zero Gaussian process X concentrated on $\{z \in \ell^\infty(T) : z$ is uniformly ρ-continuous$\}$, with covariance $H(s, t)$.*

Remark 8. The AMS condition is technical and thus we refer the readers to Kosorok (2003). In this paper, we will make use of the following separability as a sufficient condition for AMS (Lemma 2; Kosorok (2003)):

Denote $\mathbb{P}^*$ as outer probability, as defined in Section 1.2 of van der Vaart and Wellner (1996). A triangular array of stochastic processes $\{f_{ni}(t) : i = 1, \ldots, n, t \in T\}$ is said to be separable if for every $n \geq 1$, there exists a countable subset $T_n \subset T$ such that $\mathbb{P}^*\left(\sup_{t \in T} \inf_{s \in T_n} \sum_{i=1}^{n}(f_{ni}(s) - f_{ni}(t))^2 > 0\right) = 0$.

Checking the manageability in condition 1 above is usually not straightforward. In practice, we use VC type as a sufficient condition. We state Proposition 3.6.12 of Giné and Nickl (2015) as a lemma below, which is used for establishing the VC type of functions we encounter.

Lemma 5. *Let f be a function of bounded p-variation, $p \geq 1$. Then, the collection $\mathscr{F}$ of translations and dilation of f, $\mathscr{F} = \{x \mapsto f(tx - s) : t > 0, s \in \mathbb{R}\}$ is of VC type.*

We also cite some results of Chernozhukov et al. (2014) as the following lemma, which shows the stability of VC type classes under element-wise addition and multiplication.

Lemma 6. *Let $\mathscr{F}$ and $\mathscr{G}$ be of VC type with envelopes F and G respectively. Then the collection of element-wise sums $\mathscr{F} + \mathscr{G}$ and the collection of element-wise products $\mathscr{F}\mathscr{G}$ are of VC type with envelope $F + G$ and FG, respectively.*

The first one is a special case of Lemma A.6 of Chernozhukov et al. (2014). The second one is proven in Corollary A.1 of Chernozhukov et al. (2014).

A.3. Proof of Theorem 1 (the main result)

Proof. Part (i) For $(\theta, k) \in \mathbb{T}$, we define $f_{ni}(\theta, k) = \frac{e_v'(\Gamma_p^+)^{-1}r_p(\frac{X_i}{h_{k,n}(\theta_k)})}{\sqrt{nh_{k,n}(\theta_k)f_X(0)}}\mathcal{E}_k(Y_i, D_i, X_i, \theta)K(\frac{X_i}{h_{k,n}(\theta_k)})\delta_i^+ = \frac{a_0 + a_1(\frac{X_i}{c_k(\theta_k)h_n}) + \cdots + a_p(\frac{X_i}{c_k(\theta_k)h_n})^p}{\sqrt{nc_k(\theta_k)h_nf_X(0)}}$
$\mathcal{E}_k(Y_i, D_i, X_i, \theta)K(\frac{X_i}{c_k(\theta_k)h_n})\delta_i^+$ and $v_n^+(\theta, k) = \sum_{i=1}^{n}[f_{ni}(\theta, k) - Ef_{ni}(\theta, k)]$. By Assumption 1(i)(a), the triangular array $\{f_{ni}(\theta, k)\}$ is row independent. The separability follows from the same argument as in the proof of Theorem 4 in Kosorok (2003) and the left or right continuity (in both θ_1 and θ_2) of the process $f_{ni}(\theta, k)$, which follows from Assumption 1 (ii)(d),(iii) and (iv). We claim that it satisfies the conditions required by Lemma 4.

For condition 1, we note that $\mathcal{E}_k(Y_i, X_i, \cdot)$ is a VC type (Euclidean) class with envelope $2F_{\mathcal{E}}$ by Assumption 1(ii)(a) and Lemma 6. Notice that for a fixed n, denote $\delta_x^+ = \mathbb{1}\{x > 0\}$. Both $\left\{x \mapsto \frac{(a_0 + a_1(\frac{x}{c_k(\theta_k)h_n}) + \cdots + a_p(\frac{x}{c_k(\theta_k)h_n})^p)\mathbb{1}\{|x| \leq \bar{c}h_n\}\delta_x^+}{\sqrt{nc_k(\theta_k)h_nf_X(0)}} : (\theta, k) \in \mathbb{T}\right\}$
and $\left\{x \mapsto K(\frac{x}{c_k(\theta_k)h_n}) : (\theta, k) \in \mathbb{T}\right\}$ are of VC type with envelopes $\frac{C_1}{\sqrt{nh_n}}\mathbb{1}\{|x| \leq \bar{c}h_n\}$ and $\mathbb{1}\{|x| \leq \bar{c}h_n\}\|K\|_\infty$, respectively, under Assumptions 1(i),(iii) and (iv) and Lemma 5. By Lemma 6, their product is a VC type class with envelope $F_{ni}(y, d, x) = \frac{C_3}{\sqrt{nh_n}}F_{\mathcal{E}}(y, d, x)\mathbb{1}\{C_2\frac{x}{h_n} \in [-1, 1]\}$. Applying Lemma 9.14 (iii) and Theorem 9.15 of Kosorok (2008), we obtain that $\{f_{ni}\}$ is a bounded uniform entropy integral class with row independent envelopes F_{ni}. Theorem 1 of Andrews (1994) then implies that $\{f_{ni}\}$ is manageable with respect to the envelope $\{F_{ni}\}$, and therefore condition 1 is satisfied.

To check condition 2, notice that $E[v_n^+(\theta, k)v_n^+(\vartheta, l)] = \sum_{i=1}^{n} Ef_{ni}(\theta, k)f_{ni}(\vartheta, l) - (\sum_{i=1}^{n} Ef_{ni}(\theta, k)) \cdot (\sum_{i=1}^{n} Ef_{ni}(\vartheta, l))$. Under Assumptions 1(i)(b),(ii)(c),(iii),(iv)(a) we can write

$$\sum_{i=1}^{n} Ef_{ni}(\theta, k)f_{ni}(\vartheta, l)$$
$$= \int_{\mathbb{R}^+} \frac{e_v'(\Gamma_p^+)^{-1}r_p(u/c_k(\theta_1))r_p'(u/c_l(\vartheta_l))(\Gamma_p^+)^{-1}e_v}{\sqrt{c_k(\theta_k)c_l(\vartheta_l)}f_X^2(0)}\sigma_{kl}(\theta, \vartheta | uh_n)K(\frac{u}{c_k(\theta_k)})K(\frac{u}{c_l(\vartheta_l)})f_X(uh_n)du$$
$$= \frac{\sigma_{kl}(\theta, \vartheta | 0^+)e_v'(\Gamma_p^+)^{-1}\Psi_p^+((\theta, k), (\vartheta, l))(\Gamma_p^+)^{-1}e_v}{\sqrt{c_k(\theta_k)c_l(\vartheta_l)}f_X(0)} + O(h_n).$$

$\Psi_p^+((\theta, k), (\vartheta, l))$ exists under Assumptions 1(iii) and (iv)(a). All entries in the matrix part are bounded under Assumption 1(iii),(iv)(a)(c). In the last line, n enters only through $O(h_n)$. Therefore, by Assumption 1(iii), the limit exists and is finite. Thus, $\lim_{n\to\infty} \sum_{i=1}^n Ef_{ni}(\theta_1, l_1)f_{ni}(\theta_2, l_2)$ exists. Since $Ef_{ni}(\theta_1, l_1) = 0$ implies $\lim_{n\to\infty}(\sum_{i=1}^n Ef_{ni}(\theta, k))(\sum_{i=1}^n Ef_{ni}(\vartheta, l)) = 0$, and condition 2 is satisfied.

Under Assumption 1(i)(a), (ii)(a), (iii), and (iv)(a), it is clear that $\sum_{i=1}^n E^* F_{ni}^2 = \sum_{i=1}^n EF_{ni}^2 \lesssim \int_{\mathscr{Y}\times\mathscr{D}\times\mathscr{X}} F_{\mathcal{E}}^2(y, d, uh_n)\mathbb{1}\{C_2 u \in [-1, 1]\}d\mathbb{P}^x(y, d, uh_n) + o(h_n) < \infty$ as $n \to \infty$. This shows condition 3.

To show condition 4, note that for any $\epsilon > 0$

$$\lim_{n\to\infty} \sum_{i=1}^n E^* F_{ni}^2 \mathbb{1}\{F_{ni} > \epsilon\} = \lim_{n\to\infty} \sum_{i=1}^n EF_{ni}^2 \mathbb{1}\{F_{ni} > \epsilon\}$$

$$\lesssim \lim_{n\to\infty} \int_{\mathscr{Y}\times\mathscr{D}\times\mathscr{X}} F_{\mathcal{E}}^2(y, d, uh_n)\mathbb{1}\{\frac{C_3}{\sqrt{nh_n}}F_{\mathcal{E}}(y, d, uh_n)\mathbb{1}\{C_2 u \in [-1, 1]\} > \epsilon\}d\mathbb{P}^x(y, d, uh_n)$$

$$= \int_{\mathscr{Y}\times\mathscr{D}\times\mathscr{X}} F_{\mathcal{E}}^2(y, d, uh_n) \lim_{n\to\infty} \mathbb{1}\{\frac{C_3}{\sqrt{nh_n}}F_{\mathcal{E}}(y, d, uh_n)\mathbb{1}\{C_2 u \in [-1, 1]\} > \epsilon\}d\mathbb{P}^x(y, d, uh_n) = 0$$

by the dominated convergence theorem under Assumption 1(ii)(a), (iii).

To show condition 5, note that we can write $\rho_n^2((\theta, k), (\vartheta, l)) = \sum_{i=1}^n E[f_{ni}(\theta, k) - f_{ni}(\vartheta, l)]^2 = nE[f_{ni}^2(\theta, k) + f_{ni}^2(\vartheta, l) - 2f_{ni}(\theta, k)f_{ni}(\vartheta, l)]$. From our calculations on the way to show condition 2, we know that each term exists on the right-hand side. Since n enters the expression only through the $O(h_n)$ part, for all deterministic sequences $\{s_n\}$ and $\{t_n\}$ in $\mathbb{T}$, $\rho^2(s_n, t_n) \to 0$ implies $\rho_n^2(s_n, t_n) \to 0$.

Now, applying Lemma 4, we have $v_n^+(\cdot)$ converging weakly to a tight mean-zero Gaussian process $\mathbb{G}_{H^+}(\cdot)$ with covariance function $H^+((\theta, k), (\vartheta, l)) = \frac{\sigma(\theta, \vartheta|0)e_v'(\Gamma_p^+)^{-1}\Psi^+((\theta, k), (\vartheta, l))(\Gamma_p^+)^{-1}e_v}{\sqrt{c_k(\theta_k)c_l(\vartheta_l)}f_X(0)}$. Slutsky's Theorem and Assumption 1(iv) then give

$$\sqrt{nh_n^{1+2v}}\begin{bmatrix} \hat{\mu}_{1,p}^{(v)}(0^+, \cdot) - \mu_1^{(v)}(0^+, \cdot) \\ \hat{\mu}_{2,p}^{(v)}(0^+, \cdot) - \mu_2^{(v)}(0^+, \cdot) \end{bmatrix} \rightsquigarrow \begin{bmatrix} \mathbb{G}_{H^+}(\cdot, 1)/\sqrt{c_1^{1+2v}(\cdot)} \\ \mathbb{G}_{H^+}(\cdot, 2)/\sqrt{c_2^{1+2v}(\cdot)} \end{bmatrix}$$

Applying the functional delta method under Assumption 2(i), we then have

$$\sqrt{nh_n^{1+2v}}\begin{bmatrix} \phi(\hat{\mu}_{1,p}^{(v)}(0^+, \cdot))(\cdot) - \phi(\mu_1^{(v)}(0^+, \cdot))(\cdot) \\ \psi(\hat{\mu}_{2,p}^{(v)}(0^+, \cdot))(\cdot) - \psi(\mu_2^{(v)}(0^+, \cdot))(\cdot) \end{bmatrix} \rightsquigarrow \begin{bmatrix} \phi'_{\mu_1^{(v)}(0^+, \cdot)}\Big(\mathbb{G}_{H^+}(\cdot, 1)/\sqrt{c_1^{1+2v}(\cdot)}\Big)(\cdot) \\ \psi'_{\mu_2^{(v)}(0^+, \cdot)}\Big(\mathbb{G}_{H^+}(\cdot, 2)/\sqrt{c_2^{1+2v}(\cdot)}\Big)(\cdot) \end{bmatrix}$$

All arguments above can be replicated for the left limit objects, and thus by Assumption 1(i)(a), we obtain

$$\sqrt{nh_n^{1+2v}}\begin{bmatrix} \Big(\phi(\hat{\mu}_{1,p}^{(v)}(0^+, \cdot)) - \phi(\hat{\mu}_{1,p}^{(v)}(0^-, \cdot))\Big)(\cdot) - \Big(\phi(\mu_1^{(v)}(0^+, \cdot)) - \phi(\mu_1^{(v)}(0^-, \cdot))\Big)(\cdot) \\ \Big(\psi(\hat{\mu}_{2,p}^{(v)}(0^+, \cdot)) - \psi(\hat{\mu}_{2,p}^{(v)}(0^-, \cdot))\Big)(\cdot) - \Big(\psi(\mu_2^{(v)}(0^+, \cdot)) - \psi(\mu_2^{(v)}(0^-, \cdot))\Big)(\cdot) \end{bmatrix}$$

$$\rightsquigarrow \begin{bmatrix} \phi'_{\mu_1^{(v)}(0^+, \cdot)}\Big(\mathbb{G}_{H^+}(\cdot, 1)/\sqrt{c_1^{1+2v}(\cdot)}\Big)(\cdot) - \phi'_{\mu_1^{(v)}(0^-, \cdot)}\Big(\mathbb{G}_{H^-}(\cdot, 1)/\sqrt{c_1^{1+2v}(\cdot)}\Big)(\cdot) \\ \psi'_{\mu_2^{(v)}(0^+, \cdot)}\Big(\mathbb{G}_{H^+}(\cdot, 2)/\sqrt{c_2^{1+2v}(\cdot)}\Big)(\cdot) - \psi'_{\mu_2^{(v)}(0^-, \cdot)}\Big(\mathbb{G}_{H^-}(\cdot, 2)/\sqrt{c_2^{1+2v}(\cdot)}\Big)(\cdot) \end{bmatrix} = \begin{bmatrix} \mathbb{G}'(\cdot, 1) \\ \mathbb{G}'(\cdot, 2) \end{bmatrix}.$$

Finally, by another application of the functional delta method, the chain rule for the functional delta method (Lemma 3.9.3 of van der Vaart and Wellner (1996)), and Lemma 3 under Assumption 2(i) and (ii), we obtain

$$\sqrt{nh_n^{1+2v}}[\hat{\tau}(\cdot) - \tau(\cdot)] = \sqrt{nh_n^{1+2v}}[\Upsilon\Big(\frac{\phi(\hat{\mu}_{1,p}^{(v)}(0^+, \cdot)) - \phi(\hat{\mu}_{1,p}^{(v)}(0^-, \cdot))}{\psi(\hat{\mu}_{2,p}^{(v)}(0^+, \cdot)) - \psi(\hat{\mu}_{2,p}^{(v)}(0^-, \cdot))}\Big)(\cdot) - \Upsilon\Big(\frac{\phi(\mu_1^{(v)}(0^+, \cdot)) - \phi(\mu_1^{(v)}(0^-, \cdot))}{\psi(\mu_2^{(v)}(0^+, \cdot)) - \psi(\mu_2^{(v)}(0^-, \cdot))}\Big)(\cdot)]$$

$$\rightsquigarrow \Upsilon_W'\Big(\frac{[\psi(\mu_2^{(v)}(0^+, \cdot)) - \psi(\mu_2^{(v)}(0^-, \cdot))]\mathbb{G}'(\cdot, 1) - [\phi(\mu_1^{(v)}(0^+, \cdot)) - \phi(\mu_1^{(v)}(0^-, \cdot))]\mathbb{G}'(\cdot, 2)}{(\psi(\mu_2^{(v)}(0^+, \cdot)) - \psi(\mu_2^{(v)}(0^-, \cdot)))^2}\Big)(\cdot). \quad \square$$

A.4. Proof of Theorem 2

Proof. Part (ii) We introduce the following notations.

$$v_{\xi,n}^+(\theta, k) = \sum_{i=1}^n \xi_i \frac{e_v'(\Gamma_p^+)^{-1}r_p(\frac{X_i}{h_{k,n}(\theta_k)})}{\sqrt{nc_k(\theta_k)h_n}f_X(0)}\mathcal{E}_k(Y_i, D_i, X_i, \theta)K(\frac{X_i}{c_k(\theta_k)h_n})\delta_i^+$$

$$\hat{v}_{\xi,n}^+(\theta, k) = \sum_{i=1}^n \xi_i \frac{e_v'(\Gamma_p^+)^{-1}r_p(\frac{X_i}{h_{k,n}(\theta_k)})}{\sqrt{nc_k(\theta_k)h_n}\hat{f}_X(0)}\hat{\mathcal{E}}_k(Y_i, D_i, X_i, \theta)K(\frac{X_i}{c_k(\theta_k)h_n})\delta_i^+$$

Applying Theorem 2 of Kosorok (2003) (which is also the same as Theorem 11.19 of Kosorok (2008)), we have $v_{\xi,n}^+ \overset{p}{\underset{\xi}{\rightsquigarrow}} \mathbb{G}_{H+}$. In order to apply Lemma 2, we need to show $\sup_{(\theta,k)\in\mathbb{T}}|v_{\xi,n}^+(\theta,k) - \hat{v}_{\xi,n}^+(\theta,k)| \overset{p}{\underset{x\times\xi}{\rightarrow}} 0$.

We will focus on the case of $k = 1$, and same argument applies to the case of $k = 2$. Note that under Assumption 4, $|\hat{f}_X(0) - f_X(0)| = o_p^{x\times\xi}(1)$. Thus under Assumption 1(i)(b),

$$v_{\xi,n}^+(\theta, 1) - \hat{v}_{\xi,n}^+(\theta, 1)$$

$$= \frac{1}{f_X(0)\hat{f}_X(0)} \sum_{i=1}^n \xi_i \frac{e_v'(\Gamma_p^+)^{-1}r_p(\frac{X_i}{h_{1,n}(\theta_1)})}{\sqrt{nc_1(\theta_1)h_n}} K(\frac{X_i}{c_1(\theta_1)h_n})\delta_i^+[\hat{\mathcal{E}}_1(Y_i, D_i, X_i, \theta)f_X(0) - \mathcal{E}_1(Y_i, D_i, X_i, \theta)\hat{f}_X(0)]$$

$$= \frac{1}{f_X^2(0) + o_p^{x\times\xi}(1)} \sum_{i=1}^n Z_i(\theta_1)[\hat{\mathcal{E}}_1(Y_i, D_i, X_i, \theta) - \mathcal{E}_1(Y_i, D_i, X_i, \theta)]f_X(0)$$

$$+ \frac{o_p^{x\times\xi}(1)}{f_X^2(0) + o_p^{x\times\xi}(1)} \sum_{i=1}^n Z_i(\theta_1)\mathcal{E}_1(Y_i, D_i, X_i, \theta) = (1) + (2)$$

where $Z_i(\theta_1) = \xi_i \frac{e_v'(\Gamma_p^+)^{-1}r_p(\frac{X_i}{h_{1,n}(\theta_1)})}{\sqrt{nc_1(\theta_1)h_n}} K(\frac{X_i}{c_1(\theta_1)h_n})\delta_i^+$. It can be shown following the same procedures in the proof of Theorem 1 that under Assumption 1, 2, $\sum_{i=1}^n Z_i(\theta_1) \rightsquigarrow \mathbb{G}_1$ and $\sum_{i=1}^n Z_i(\theta_1)\mathcal{E}_1(Y_i, D_i, X_i, \theta) \rightsquigarrow \mathbb{G}_2$ for some zero mean Gaussian processes $\mathbb{G}_1, \mathbb{G}_2 : \Omega^x \times \Omega^\xi \mapsto \ell^\infty(\Theta)$. By Prohorov's Theorem, the weak convergence implies asymptotic tightness and therefore implies that $\sum_{i=1}^n Z_i(\theta_1) = O_p^{x\times\xi}(1)$ uniformly on Θ and $\sum_{i=1}^n Z_i(\theta_1)\mathcal{E}_1(Y_i, D_i, X_i, \theta) = O_p^{x\times\xi}(1)$ uniformly on Θ. Thus (2) $= \frac{o_p^{x\times\xi}(1)}{f_X^2(0)+o_p^{x\times\xi}(1)}$ uniformly on Θ. We then control (1). Assumption 4 implies $\sum_{i=1}^n Z_i(\theta_1)[\hat{\mathcal{E}}_1(Y_i, D_i, X_i, \theta) - \mathcal{E}_1(Y_i, D_i, X_i, \theta)]f_X(0)$ $= \sum_{i=1}^n Z_i(\theta_1)[o_p^{x\times\xi}(1)]f_X(0) = f_X(0)[O_p^{x\times\xi}(1)]\sum_{i=1}^n Z_i(\theta_1)$ uniformly on Θ. Therefore, we have $\sup_{(\theta,k)\in\mathbb{T}}|v_{\xi,n}^+(\theta,k) - \hat{v}_{\xi,n}^+(\theta,k)| \overset{p}{\underset{x\times\xi}{\rightarrow}} 0$, and thus we can apply Lemma 2 to conclude $\hat{v}_{\xi,n}^+ \overset{p}{\underset{\xi}{\rightsquigarrow}} \mathbb{G}_{H+}$. With similar arguments, we can derive that $\hat{v}_{\xi,n}^- \overset{p}{\underset{\xi}{\rightsquigarrow}} \mathbb{G}_{H-}$.

The continuous mapping theorem for bootstrap (Kosorok, 2008, Proposition 10.7) and the continuity of the Hadamard derivatives imply

$$\begin{bmatrix} \widehat{\mathbb{X}}_n'(\cdot, 1) \\ \widehat{\mathbb{X}}_n'(\cdot, 2) \end{bmatrix} = \begin{bmatrix} \phi'_{\mu_1^{(v)}(0^+,\cdot)}\left(\hat{v}_{\xi,n}^+(\cdot, 1)/\sqrt{c_1^{1+2v}(\cdot)}\right)(\cdot) - \phi'_{\mu_1^{(v)}(0^-,\cdot)}\left(\hat{v}_{\xi,n}^+(\cdot, 1)/\sqrt{c_1^{1+2v}(\cdot)}\right)(\cdot) \\ \psi'_{\mu_2^{(v)}(0^+,\cdot)}\left(\hat{v}_{\xi,n}^+(\cdot, 2)/\sqrt{c_2^{1+2v}(\cdot)}\right)(\cdot) - \psi'_{\mu_2^{(v)}(0^-,\cdot)}\left(\hat{v}_{\xi,n}^+(\cdot, 2)/\sqrt{c_2^{1+2v}(\cdot)}\right)(\cdot) \end{bmatrix} \overset{p}{\underset{\xi}{\rightsquigarrow}} \begin{bmatrix} \mathbb{G}'(\cdot, 1) \\ \mathbb{G}'(\cdot, 2) \end{bmatrix}$$

Recursively applying Functional Delta for Bootstrap (Theorem 2.9 of Kosorok (2008)) then gives

$$\Upsilon_W'\left(\frac{[\psi(\mu_2^{(v)}(0^+, \cdot)) - \psi(\mu_2^{(v)}(0^-, \cdot))]\widehat{\mathbb{X}}_n'(\cdot, 1) - [\phi(\mu_1^{(v)}(0^+, \cdot)) - \phi(\mu_1^{(v)}(0^-, \cdot))]\widehat{\mathbb{X}}_n'(\cdot, 2)}{[\psi(\mu_2^{(v)}(0^+, \cdot)) - \psi(\mu_2^{(v)}(0^+, \cdot))]^2}\right)$$

$$\overset{p}{\underset{\xi}{\rightsquigarrow}} \Upsilon_W'\left(\frac{[\psi(\mu_2^{(v)}(0^+, \cdot)) - \psi(\mu_2^{(v)}(0^+, \cdot))]\mathbb{G}'(\cdot, 1) - [\phi(\mu_1^{(v)}(0^+, \cdot)) - \phi(\mu_1^{(v)}(0^-, \cdot))]\mathbb{G}'(\cdot, 2)}{[\psi(\mu_2^{(v)}(0^+, \cdot)) - \psi(\mu_2^{(v)}(0^-, \cdot))]^2}\right)$$

This completes the proof. □

A.5. Uniform consistency

Theorem 1 implies the following useful asymptotic identities. They state the uniform consistency of the numerator and the denominator of Wald ratios. Because these implications themselves prove useful when we apply the main theorem to the specific example, we state them as corollaries below for convenience of reference.

Corollary 5. *Under Assumptions 1 and 2, $\hat{\mu}_{l,p}^{(v)}(0^\pm, \cdot) - \mu_k^{(v)}(0^\pm, \cdot) \overset{p}{\underset{x}{\rightarrow}} 0$ uniformly.*

Corollary 6. *Under Assumptions 1 and 2, $\phi(\hat{\mu}_{l,p}^{(v)}(0^\pm, \cdot)) - \phi(\mu_k^{(v)}(0^\pm, \cdot)) \overset{p}{\underset{x}{\rightarrow}} 0$ uniformly.*

Assumption 2(iv) further implies that the mode $\overset{p}{\underset{x}{\rightarrow}}$ of convergence in the above corollaries can be replaced by the mode $\overset{p}{\underset{x\times\xi}{\rightarrow}}$ of convergence.

A.6. First stage estimators

To estimate MP, we replace $\mu_k(x, \theta)\mathbb{1}\{|x/h_{k,n}(\theta_k)| \leq 1\}$ by its estimate $\tilde{\mu}_{k,p}(x, \theta)\mathbb{1}\{|x/h_{k,n}(\theta_k)| \leq 1\}$, which is uniformly consistent across (x, θ). Lemma 7 proposes such a uniformly consistent estimator without requiring to solve an additional optimization problem; by a mean-value expansion and uniform boundedness of $\mu_k^{(p+1)}$, we can reuse the first stage local polynomial estimates of $\hat{\mu}_{l,p}^{(v)}(0^{\pm}, \theta)$ for all $v \leq p$. This auxiliary result will prove useful when we apply Theorems 1 and 2 to specific examples. In fact, we will prove a more general result that allows us to use any first r terms of our pth order polynomial estimator for a t such that $0 \leq t \leq p$.

Lemma 7. *Fix an integer t such that $0 \leq t \leq p$. Suppose that Assumptions 1 and 2 are satisfied. Let $\mathcal{E}_1(y, d, x, \theta) = g_1(y, \theta_1) - \mu_1(x, \theta_1)$, $\mathcal{E}_2(y, d, x, \theta) = g_2(d, \theta_2) - \mu_2(d, \theta_2)$, $\delta_x^+ = \mathbb{1}\{x \geq 0\}$ and $\delta_x^- = \mathbb{1}\{x \leq 0\}$. Define $\tilde{\mu}_{1,t}(x, \theta_1) = r_t(x/h_{1,n}(\theta_1))'\hat{\alpha}_{1+,t}(\theta_1)\delta_x^+ + r_t(x/h_{1,n}(\theta_1))'\hat{\alpha}_{1-,t}(\theta_1)\delta_x^-$ and $\tilde{\mu}_{2,t}(x, \theta_2) = r_t(x/h_{2,n}(\theta_2))'\hat{\alpha}_{2+,t}(\theta_2)\delta_x^+ + r_t(x/h_{2,n}(\theta_2))'\hat{\alpha}_{2-,t}(\theta_2)\delta_x^-$. Then we have $\hat{\mathcal{E}}_1(y, d, x, \theta) = [g_1(y, \theta_1) - \tilde{\mu}_{1,t}(x, \theta_1)]\mathbb{1}\{|x/h_{1,n}(\theta_1)| \leq 1\}$ and $\hat{\mathcal{E}}_2(y, d, x, \theta) = [g_2(d, \theta_2) - \tilde{\mu}_{2,t}(x, \theta_2)]\mathbb{1}\{|x/h_{2,n}(\theta_2)| \leq 1\}$ are uniformly consistent for $\mathcal{E}_1(y, d, x, \theta)\mathbb{1}\{|x/h_{1,n}(\theta_1)| \leq 1\}$ and $\mathcal{E}_2(y, d, x, \theta)\mathbb{1}\{|x/h_{2,n}(\theta_2)| \leq 1\}$ on $[\underline{x}, \bar{x}] \times \mathcal{Y} \times \mathcal{D}$, respectively.*

Proof. We will show the 1+ part only, since the other parts can be shown similarly. Recall that $\mathcal{E}_1(y, d, x, \theta) = g_1(y, \theta_1) - \mu_1(x, \theta_1)$. If $x > 0$, $\mu_1(x, \theta_1)\mathbb{1}\{|x/h_{1,n}(\theta_1)| \leq 1\} = \left(\mu_1(0^+, \theta_1) + \mu_1^{(1)}(0^+, \theta_1)x + \cdots + \mu_1^{(t)}(0^+, \theta_1)\frac{x^t}{t!} + \mu_1^{(t+1)}(x_{ni}^*, \theta_1)\frac{x^{(t+1)}}{(t+1)!}\right)\mathbb{1}\{|x/c_1(\theta)h_n| \leq 1\}$. By Corollary 5, $\hat{\mu}_{1,t}^{(v)}(0^{\pm}, \theta_1)$ is uniformly consistent for $\mu_1^{(v)}(0^{\pm}, \theta_1)$, $v = 0, 1, \ldots, t$. Thus, $[\hat{\mathcal{E}}_1(y, d, x, \theta) - \mathcal{E}_1(y, d, x, \theta)]\mathbb{1}\{|x/h_{1,n}(\theta)| \leq 1\} = \left(\hat{\mu}_{1,t}(0^+, \theta_1) + \hat{\mu}_{1,t}^{(1)}(0^+, \theta_1)x + \cdots + \hat{\mu}_{1,t}^{(t)}(0^+, \theta_1)\frac{x^t}{t!}\right)\mathbb{1}\{|x/h_{1,n}(\theta)| \leq 1\} - \left(\mu_1(0^+, \theta_1) + \mu_1^{(1)}(0^+, \theta_1)x + \cdots + \mu_1^{(t)}(0^+, \theta_1)\frac{x^t}{t!} + \mu_1^{(t+1)}(x^*, \theta_1)\frac{x^{(t+1)}}{(t+1)!}\right)\mathbb{1}\{|x/h_{1,n}(\theta)| \leq 1\}$

$= o_p^x(1) - \mu_1^{(t+1)}(x^*, \theta_1)\frac{x^{(t+1)}}{(t+1)!}\mathbb{1}\{|x| \leq h_{1,n}(\theta)\} = o_p^x(1) + O(h_n)$, where the last equality is by Assumption 1(iii) and by the uniform boundedness of $\mu^{(t+1)}$ under Assumption 1 (ii)(a). $\square$

Finally, we also provide consistent estimators, $\hat{f}_X(0)$ and $\hat{f}_{Y^d|C}(\cdot)$. For $f_X(0)$, we propose to use the kernel density estimator $\hat{f}_X(0) = \frac{1}{nb_n}\sum_{i=1}^n K\left(\frac{X_i}{b_n}\right)$ with $b_n \to 0$ and $nb_n \to \infty$. For $f_{Y^d|C}(\cdot)$, we propose

$$\hat{f}_{Y^1|C}(y) = \frac{\hat{f}_{Y^*|XD^*}(y|x_0^+, 1)\hat{P}(D_i^* = 1|X_i = x_0^+) - \hat{f}_{Y^*|XD^*}(y|x_0^-, 1)\hat{P}(D_i^* = 1|X_i = x_0^-)}{\hat{\mu}_{2,2}(0^+, 1) - \hat{\mu}_{2,2}(0^-, 1)} \quad \text{and}$$

$$\hat{f}_{Y^0|C}(y) = \frac{\hat{f}_{Y^*|XD^*}(y|x_0^+, 0)\hat{P}(D_i^* = 0|X_i = x_0^+) - \hat{f}_{Y^*|XD^*}(y|x_0^-, 0)\hat{P}(D_i^* = 0|X_i = x_0^-)}{-(\hat{\mu}_{2,2}(0^+, 0) - \hat{\mu}_{2,2}(0^-, 0))},$$

where $\hat{f}_{Y^*|XD^*}(y|x_0^{\pm}, 1) = \frac{\frac{1}{na_n^2}\sum_{i=1}^n K(\frac{X_i}{a_n})K(\frac{Y_i^*-y}{a_n})D_i^*\delta_i^{\pm}}{\frac{1}{na_n}\sum_{i=1}^n K(\frac{X_i}{a_n})D_i^*\delta_i^{\pm}}$, $\hat{f}_{Y^*|XD^*}(y|x_0^{\pm}, 0) = \frac{\frac{1}{na_n^2}\sum_{i=1}^n K(\frac{X_i}{a_n})K(\frac{Y_i^*-y}{a_n})(1-D_i^*)\delta_i^{\pm}}{\frac{1}{na_n}\sum_{i=1}^n K(\frac{X_i}{a_n})(1-D_i^*)\delta_i^{\pm}}$, $\hat{P}(D_i^* = 1|X_i = x_0^{\pm}) = \frac{\sum_{i=1}^n K(\frac{X_i}{c_n})D_i^*\delta_i^{\pm}}{\sum_{i=1}^n K(\frac{X_i}{c_n})\delta_i^{\pm}}$, and $\hat{P}(D_i^* = 0|X_i = x_0^{\pm}) = \frac{\sum_{i=1}^n K(\frac{X_i}{c_n})(1-D_i^*)\delta_i^{\pm}}{\sum_{i=1}^n K(\frac{X_i}{c_n})\delta_i^{\pm}}$.

Lemma 8. *Suppose Assumptions K, M, S and FQRD (i)–(iv) hold. Assume that the conditional CDF $f_{Y^*|XD^*}(\cdot|0^+, d)$ is continuously differentiable and its derivative is uniformly bounded on $\mathcal{Y}_1$ for $d \in \{0, 1\}$. If a_n satisfies $a_n \to 0$, $na_n^2/|\log a_n| \to \infty$, $a_n \leq ca_{2n}$ for some $c > 0$ and $|\log a_n|/\log \log n \to \infty$ and c_n satisfies $c_n \to 0$ and $nc_n \to \infty$, then $\sup_{y \in \mathcal{Y}_1}|\hat{f}_{Y^1|C}(y) - f_{Y^1|C}(y)| = o_p^x(1)$ and $\sup_{y \in \mathcal{Y}_1}|\hat{f}_{Y^0|C}(y) - f_{Y^0|C}(y)| = o_p^x(1)$.*

Proof. We will prove this lemma for $f_{Y^1|C}$, and the corresponding results for $f_{Y^0|C}(y)$ follow similarly. From (3.1) and Assumption FQRD (i), (vi) $f_{Y^1|C}(y) = \frac{\partial}{\partial y}F_{Y^1|C}(y) = \frac{\frac{\partial}{\partial y}\lim_{x \downarrow 0}E[\mathbb{1}\{Y_i^* \leq y\}\cdot\mathbb{1}\{D_i^*=d\}|X_i=x] - \frac{\partial}{\partial y}\lim_{x \uparrow 0}E[\mathbb{1}\{Y_i^* \leq y\}\cdot\mathbb{1}\{D_i^*=d\}|X_i=x]}{\lim_{x \downarrow 0}E[\mathbb{1}\{D_i^*=d\}|X_i=x] - \lim_{x \uparrow 0}E[\mathbb{1}\{D_i^*=d\}|X_i=x]}$. Theorem 1 then implies that the denominator terms can be consistently estimated by $\hat{\mu}_{2,2}(0^+, 1) - \hat{\mu}_{2,2}(0^-, 1)$ at the rate of $O_p^x(1/\sqrt{nh_n})$ and the limit is bounded away from zero by Assumption FQRD (iii).

As for the numerator terms, the boundedness of the integrand ensures that we can interchange the expectation and limit. Thus, we can write $\frac{\partial}{\partial y}E[\mathbb{1}\{Y_i^* \leq y\}\cdot\mathbb{1}\{D_i^*=d\}|X_i = 0^+] = \frac{\partial}{\partial y}E[\mathbb{1}\{Y_i^* \leq y\}|X_i = 0^+, D_i^* = 1]\mathbb{P}^x(D_i^* = 1|X_i = 0^+) + 0 = \frac{\partial}{\partial y}F_{Y^*|XD^*}(y|0^+, 1)\mathbb{P}^x(D_i^* = 1|X_i = 0^+) = f_{Y^*|XD^*}(y|0^+, 1)\mathbb{P}^x(D_i^* = 1|X_i = 0^+)$. The standard point-wise convergence result under the stated assumptions for Nadaraya–Watson estimator and Assumption FQRD (i) imply the consistency of $\hat{P}(D_i^* = 1|X_i = 0^+)$ for $\mathbb{P}^x(D_i^* = 1|X_i = 0^+)$. The uniform consistency of $\hat{f}_{Y^*|XD^*}(\cdot|0^+, 1)$ follows from Assumption FQRD (i), (ii), continuous differentiability, uniform boundedness of derivatives of $\hat{f}_{Y^*|XD^*}(\cdot|0^+, 1)$ on $\mathcal{Y}_1$ and Theorem 2.3 of Giné and Guillou (2002), which is applicable under the above bandwidth assumptions for a_n and Assumption K. $\square$

A.7. Proof of Corollary 2

Proof. We first show that Lemma 1 holds under cluster sampling. Note that Steps 1 and 2 of the Proof for Lemma 1 follow through. To see this, note that for the inverse factor in Step 1, the deterministic part is now $E[\frac{1}{Gh_G}\sum_{g=1}^{G}\sum_{i\in C_g}\delta_i^+ K(\frac{X_i}{h_G})r_p(\frac{X_i}{h_G})r_p'(\frac{X_i}{h_G})]$ $= \frac{1}{G}\sum_{g=1}^{G}\sum_{i\in C_g}E[\frac{1}{h_G}\delta_i^+ K(\frac{X_i}{h_G})r_p(\frac{X_i}{h_G})r_p'(\frac{X_i}{h_G})]$. Assumption 5(i), (iii), and (iv) imply that the above becomes $\frac{1}{G}\sum_{g=1}^{G}\sum_{i\in C_g}f_{X_i}(0)\left(\int_{\mathbb{R}^+}K(u)r_p(u)r_p'(u)+O(h_G)\right) = \bar{f}_X(0)\Gamma_p^+ + O(h_G)$. For the stochastic part $\frac{1}{G}\sum_{g=1}^{G}$ $\left(\frac{1}{h_G}\sum_{i\in C_g}\delta_i^+ K(\frac{X_i}{h_G})r_p(\frac{X_i}{h_G})r_p'(\frac{X_i}{h_G})\right) - E\left[\frac{1}{G}\sum_{g=1}^{G}\left(\frac{1}{h_G}\sum_{i\in C_g}\delta_i^+ K(\frac{X_i}{h_G})r_p(\frac{X_i}{h_G})r_p'(\frac{X_i}{h_G})\right)\right] = o_p^x(1)$ follows from the WLLN for the row-wise independent triangular array since nothing depends on θ under Assumption 5(ii)(a).

For Step 2, the deterministic part follows the same argument under Assumption 5(i), (ii)(b), (iii), and (iv)(a), and thus we have $E\left[\frac{1}{\sqrt{Gh_G}}\sum_{g=1}^{G}\sum_{i\in C_g}\delta_i^+ K(\frac{X_i}{h_G})r_p(\frac{X_i}{h_G})\mu_1^{(p+1)}(x_{ni}^*,\theta_1)h_G^{p+1}\frac{(\frac{X_i}{h_G})^{p+1}}{(p+1)!}\right] = O(\sqrt{\frac{h_G^{2p+1}}{G}})$ uniformly in θ_1. For the stochastic part, by the same Lipschitz argument under Assumption 5 (ii), (iii) and (iv)(a), it suffices to consider $\frac{1}{\sqrt{Gh_G}}\sum_{g=1}^{G}\sum_{i\in C_g}\delta_i^+ K(\frac{X_i}{h_G})r_p(\frac{X_i}{h_G})\mu_1^{(p+1)}(0^+,\theta_1)h_G^{p+1}\frac{(\frac{X_i}{h_G})^{p+1}}{(p+1)!}$. For each cluster g, we relabel the subscript i as tg, where $t \in \{1,\ldots,n_g\}$ indicates the index in cluster g. Write $W_g = (W_{1g},\ldots,W_{\bar{N}g})$, where $W_{tg} = (Y_{tg}, D_{tg}, X_{tg})$ if $t \leq n_g$ and $W_{tg} = 0$ if $t > n_g$. Let $\mathscr{F}_s^t = \{W_g \mapsto \mathbb{1}\{X_{tg} > 0\}K(aX_{tg})(aX_{tg})^{s+p+1}\mu_1^{p+1}(0^+,\theta_1)\mathbb{1}\{X_{tg} \in [-1,1]\} : a \geq 1/h_0, \theta_1 \in \Theta_1\}$ and $\mathscr{F}_{G,s} = \{W_g \mapsto \sum_{t=1}^{\bar{N}}\mathbb{1}\{X_{tg} > 0\}K(X_{tg}/h_G)(X_{tg}/h_G)^{s+p+1}\mu_1^{p+1}(0^+,\theta_1) : \theta_1 \in \Theta_1\}$ for each integer s such that $0 \leq s \leq p$. Note that $\mathbb{1}\{X_{tg} > 0\} = 0$ for $t > n_g$. Under Assumption 5 (ii), (iii), and (iv), $\mathscr{F}_{G,s} \subset \mathscr{F}_s := \sum_{t=1}^{\bar{N}}\mathscr{F}_s^t$ for all G. $\mathscr{F}_s$ is a VC type class with an integrable envelope $F_s = \bar{N}\bar{M}\|K\|_\infty$ by Lemma 6. Thus, applying Lemma 7 of Chiang (2018) in place of Theorem 5.2 of Chernozhukov et al. (2014) yields $\sup_{f\in\mathscr{F}_s}\left|\frac{1}{\sqrt{Gh_G}}\sum_{g=1}^{G}(f(W_g)-Ef(W_g))h_G^{p+1}\right| = O_p^x(\frac{h_G^{2p+1}}{G})$, and thus the same conclusion can be made.

For Step 3, under Assumption 5(i)(a), (ii)(d) and (iv)(a), its deterministic part now becomes

$$E[\frac{1}{\sqrt{Gh_G}}\sum_{g=1}^{G}\sum_{i\in C_g}\delta_i^+ K(\frac{X_i}{h_G})r_p(\frac{X_i}{h_G})\mathcal{E}_1(Y_i,D_i,X_i,\theta)]$$
$$=E[\frac{1}{\sqrt{Gh_G}}\sum_{g=1}^{G}\sum_{i\in C_g}\delta_i^+ K(\frac{X_i}{h_G})r_p(\frac{X_i}{h_G})E[\mathcal{E}_1(Y_i,D_i,X_i,\theta)|X_i]] = 0.$$

The stochastic part follows from Markov's inequality, Assumption 5(i), (ii)(c), (iii), (iv) and (v), and the calculation:

$$\sup_{\theta_1\in\Theta_1}E[\frac{e_s'}{\sqrt{Gh_G}}\sum_{g=1}^{G}\sum_{i\in C_g}\delta_i^+ K(\frac{X_i}{h_G})r_p(\frac{X_i}{h_G})\mathcal{E}_1(Y_i,D_i,X_i,\theta)]^2$$
$$= \sup_{\theta_1\in\Theta_1}\sum_{g=1}^{G}\frac{1}{G}E\left[\frac{e_v'}{h_G}\sum_{i\in C_g}\delta_i^+ K(\frac{X_i}{h_G})r_p(\frac{X_i}{h_G})\mathcal{E}_1(Y_i,D_i,X_i,\theta)\right]^2 \leq \bar{N}^2\bar{M}^2\|K\|_\infty^2 = O(1)$$

for each $1 \leq v \leq p$. Combining all three steps, we have $\sqrt{Gh_G^{1+2v}}\left(\hat{\mu}_{k,p}^{(v)}(0^\pm,\theta_k) - \mu_k^{(v)}(0^\pm,\theta_k) - h_G^{p+1-v}\frac{e_v'(\Gamma_p^\pm)^{-1}\Lambda_{p,p+1}^\pm}{(p+1)!}\mu_k^{(p+1)}\right.$ $(0^\pm,\theta_k)) = v!\sum_{g=1}^{G}\sum_{i\in C_g}\frac{e_v'(\Gamma_p^\pm)^{-1}\mathcal{E}_k(Y_i,D_i,X_i,\theta)r_p(\frac{X_i}{h_G})K(\frac{X_i}{h_G})\delta_i^\pm}{\sqrt{nh_G}\bar{f}_X(0)} + o_p^x(1)$ uniformly for all $\theta_k \in \Theta_k$ for each $k \in \{1,2\}$.

Second, we show that Theorem 1 holds under cluster sampling. This is straightforward since Lemma 4 holds for non-identically distributed row-wise independent processes. Specifically, for $(\theta,k) \in \mathbb{T}$, we define $f_{Gg}(\theta,k) = \sum_{i\in C_g}\frac{e_v'(\Gamma_p^+)^{-1}r_p(\frac{X_i}{h_G})}{\sqrt{Gh_G}\bar{f}_X(0)}\mathcal{E}_k(Y_i,D_i,X_i,\theta)K(\frac{X_i}{h_G})\delta_i^+$ and $v_G^+(\theta,k) = \sum_{g=1}^{G}[f_{Gg}(\theta,k) - Ef_{Gg}(\theta,k)]$. Then, using the closure-under-summation property for VC type classes from Lemma 6 and following the same argument as in proof of Theorem 1, Condition 1 required by Lemma 4 can be established. Condition 2 required by Lemma 4 holds under Assumption 5(v). Conditions 3, 4 and 5 follow from similar arguments to those in the proof of Theorem 1. By an application of Lemma 4, v^+ converges weakly to a tight mean-zero Gaussian process $\mathbb{G}_{\Sigma^+}$ with covariance function Σ^+ defined in Assumption 5(v). Finally, by an application of the functional delta method, the chain rule for the functional delta method (Lemma 3.9.3 of van der Vaart and Wellner (1996)), and Lemma 3 under Assumption 2 (i) and (ii), we obtain

$$\sqrt{Gh_G^{1+2v}}[\hat{\tau}(\cdot) - \tau(\cdot)] = \sqrt{Gh_G^{1+2v}}[\Upsilon\left(\frac{\hat{\mu}_{1,p}^{(v)}(0^+,\cdot)-\hat{\mu}_{1,p}^{(v)}(0^-,\cdot)}{\hat{\mu}_{2,p}^{(v)}(0^+,\cdot)-\hat{\mu}_{2,p}^{(v)}(0^-,\cdot)}\right)(\cdot) - \Upsilon\left(\frac{\mu_1^{(v)}(0^+,\cdot)-\mu_1^{(v)}(0^-,\cdot)}{\mu_2^{(v)}(0^+,\cdot)-\mu_2^{(v)}(0^-,\cdot)}\right)(\cdot)]$$
$$\rightsquigarrow \Upsilon_W'\left(\frac{[\mu_2^{(v)}(0^+,\cdot)-\mu_2^{(v)}(0^-,\cdot)]\mathbb{G}(\cdot,1) - [\mu_1^{(v)}(0^+,\cdot)-\mu_1^{(v)}(0^-,\cdot)]\mathbb{G}(\cdot,2)}{(\mu_2^{(v)}(0^+,\cdot)-\mu_2^{(v)}(0^-,\cdot))^2}\right)(\cdot),$$

where $\mathbb{G} := \mathbb{G}_{\Sigma^+} - \mathbb{G}_{\Sigma^-}$.

Finally, we argue that Theorem 2 holds under cluster sampling. This is also straightforward since Theorem 2 of Kosorok (2003) only requires row-wise independence as well. Let us denote $\nu_{\xi,G}^+(\theta,k) = \sum_{g=1}^G \xi_g \sum_{i\in C_g} \frac{e_v'(\Gamma_p^+)^{-1} r_p(\frac{X_i}{h_G})}{\sqrt{Gh_G \hat{f}_X(0)}} \mathcal{E}_k(Y_i, D_i, X_i, \theta)$ $K(\frac{X_i}{h_G})\delta_i^+$. Applying Theorem 2 of Kosorok (2003), we have $\nu_{\xi,G}^+ \overset{p}{\underset{\xi}{\rightsquigarrow}} \mathbb{G}_{\Sigma^+}$ and similar results hold that $\nu_{\xi,G}^- \overset{p}{\underset{\xi}{\rightsquigarrow}} \mathbb{G}_{\Sigma^-}$. To apply Lemma 2, note that it follows from Assumption 5 (viii) that $\sup_{(\theta,k)\in\mathbb{T}} |\nu_{\xi,n}^+(\theta,k) - \hat{\nu}_{\xi,n}^+(\theta,k)| \overset{p}{\underset{x\times\xi}{\rightarrow}} 0$. The result then follows from the proof of Theorem 2. □

A.8. Auxiliary lemma for the extended result with covariates

Lemma 9. *Suppose Assumptions C, K, M, S and FQRD hold, then* $\sup_{(y,d)\in\mathscr{Y}_1\times\{0,1\}} |\check{\gamma}_{1+}(y,d) - \gamma_{1+}(y,d)| + \sup_{(y,d)\in\mathscr{Y}_1\times\{0,1\}} |\check{\gamma}_{1-}(y,d) - \gamma_{1-}(y,d)| = o_p(1)$ *and* $\sup_{(y,d)\in\mathscr{Y}_1\times\{0,1\}} |\check{F}_{Y^d|C}(y) - F_{Y^d|C}(y)| = o_p(1)$.

Proof. Section 2.1 and Lemma SA-14 of the Supplementary Appendix of Calonico et al. (2018b) imply that for each fixed (y,d), under Assumptions FQRD(i)(iii) and C, one has $\frac{\check{\mu}_1(0^+,y,d) - \check{\mu}_1(0^-,y,d)}{\check{\mu}_2(0^+,d) - \check{\mu}_2(0^-,d)} \overset{p}{\rightarrow}$ $\frac{\mu_1(0^+,y,d) - \mu_1(0^-,y,d) - [\gamma_{1+}'(y,d)\mu_Z(0^+) - \gamma_{1-}'(y,d)\mu_Z(0^-)]}{\mu_{2+}(d) - \mu_{2-}(d) - [\gamma_{2+}'(d)\mu_Z(0^+) - \gamma_{2-}'(d)\mu_Z(0^-)]}$. Assumptions C (i) and C (ii) further imply that for all $(y,d)\in\mathscr{Y}_1\times\{0,1\}$, it holds that $\gamma_{1+}'(y,d)\mu_Z(0^+) - \gamma_{1-}'(y,d)\mu_Z(0^-) = 0$ and $\gamma_{2+}'(d)\mu_Z(0^+) - \gamma_{2-}'(d)\mu_Z(0^-) = 0$. Thus

$$\frac{\check{\mu}_1(0^+,y,d) - \check{\mu}_1(0^-,y,d)}{\check{\mu}_2(0^+,d) - \check{\mu}_2(0^-,d)} \overset{p}{\rightarrow} F_{Y^d|C}(y). \tag{A.4}$$

This convergence can be made uniformly over $\mathscr{Y}_1\times\{0,1\}$. To see this, note that Section 8 of Supplementary Appendix of Calonico et al. (2018b) implies that $\check{\gamma}_{1+}(y,d) = \left[\mathbf{Z}'\mathbf{K}_+\mathbf{Z}/n - \Upsilon_{Z+}'\Gamma_+^{-1}\Upsilon_{Z+}\right]^{-1}\left[\mathbf{Z}'\mathbf{K}_+\mathbf{Y}(y,d)/n - \Upsilon_{Z+}'\Gamma_+^{-1}\Upsilon_{Y+}(y,d)\right]$, where $\mathbf{Z} = [Z_1, \ldots, Z_n]'$, $\mathbf{K}_+ = \text{diag}\{K(X_1/h_n)\delta_1^+, \ldots, K(X_n/h_n)\delta_n^+\}$, $\mathbf{Y}(y,d) = [\mathbb{1}\{Y_1 \le y, D_1 = d\}, \ldots, \mathbb{1}\{Y_n \le y, D_n = d\}]'$, $\Upsilon_{Z+} = \frac{1}{n}\sum_{i=1}^n r(X_i/h_n)K(X_i/h_n)\delta_i^+ Z_i$, and $\Upsilon_{Y+}(y,d) = \frac{1}{n}\sum_{i=1}^n r(X_i/h_n)K(X_i/h_n)\delta_i^+ \mathbb{1}\{Y_i \le y, D_i = d\}$. Note also that $\mathbf{Z}'\mathbf{K}_+\mathbf{Z}/n$ and Υ_{Z+} converge to their respective probability limits following Lemmas SA-2 and SA-3 of Supplementary Appendix of Calonico et al. (2018b). Convergence of Γ_+^{-1} follows from proof of Lemma 1, and uniform convergence of $\mathbf{Y}(y,d)$ and $\Upsilon_{Y+}(y,d)$ follows the same arguments as those for the convergence of Γ_+^{-1} in Assumption 1. Therefore, $\check{\gamma}_{1+}(y,d)$ converges uniformly in probability to its probability limit $\gamma_{1+}(y,d)$. This shows that (A.4) holds uniformly over $\mathscr{Y}_1\times\{0,1\}$. □

A.9. Proof of Corollary 4

Proof. Let $\check{g}_1(Y_i,\theta) = \mathbb{1}\{Y_i^* \le y, D_i^* = d\} - Z_i'\check{\gamma}_{1+}(y,d)$ and $\check{g}_2(D_i,\theta) = \mathbb{1}\{D_i^* = d\} - Z_i'\check{\gamma}_{1+}(d)$. Then, Lemma 9 implies that $g_1(Y_i,\theta) = \mathbb{1}\{Y_i^* \le y, D_i^* = d\} - Z_i'\gamma_1(y,d) + o_p(1)$ and $g_2(D_i,\theta) = \mathbb{1}\{D_i^* = d\} - Z_i'\gamma_1(d) + o_p(1)$ uniformly over $\mathscr{Y}_1\times\{0,1\}$. Lemma 1 holds for such $\check{g}_1(Y_i,\theta)$ and $\check{g}_2(D_i,\theta)$ and using $\check{g}_1$ and $\check{g}_2$ in place leads to asymptotically equivalent expressions. Theorem 1 thus follows. Furthermore, the proof of Theorem 2 follows by replacing γ_1 by $\check{\gamma}_{1\pm}$ and replacing γ_2 by $\check{\gamma}_{2\pm}$. From these results follows Corollary 1. □

Appendix B. Supplementary data

The supplementary material includes additional examples, additional proofs, and additional simulation results. Supplementary material related to this article can be found online at https://doi.org/10.1016/j.jeconom.2019.03.006.

References

Anderson, T.W., Rubin, Herman, 1949. Estimation of the parameters of a single equation in a complete system of stochastic equations. Ann. Math. Stat. 20 (1), 46–63.

Andrews, Donald W.K., 1994. Empirical process methods in econometrics(chapter 37), vol.4. In: Handbook of Econometrics, Elsevier, pp. 2247–2294.

Andrews, Donald W.K., Moreira, Marcelo J., Stock, James H., 2006. Optimal two-sided invariant similar tests for instrumental variables regression. Econometrica 74 (3), 715–752.

Arai, Yoichi, Ichimura, Hidehiko, 2016. Optimal bandwidth selection for the fuzzy regression discontinuity estimator. Econom. Lett. 141, 103–106.

Arai, Yoichi, Ichimura, Hidehiko, 2018. Simultaneous selection of optimal bandwidths for the sharp regression discontinuity estimator. Quant. Econ. 9 (1), 441–482.

Barrett, Garry F., Donald, Stephen G., 2003. Consistent tests for stochastic dominance. Econometrica 71 (1), 71–104.

Bartalotti, Otávio, Brummet, Quentin, 2017. Regression discontinuity designs with clustered data. In: Cattaneo, M.D., Escanciano, J.C. (Eds.), Advances in Econometrics: Regression Discontinuity Designs: Theory and Applications, Vol. 38. Emerald Publishing.

Bartalotti, Otávio, Calhoun, Gray, He, Yang, 2017. Bootstrap confidence intervals for sharp regression discontinuity designs with the uniform kernel. In: Cattaneo, M.D., Escanciano, J.C. (Eds.), Advances in Econometrics: Regression Discontinuity Designs: Theory and Applications, vol. 38. Emerald Publishing.

Bernal, Noelia, Carpio, Miguel A., Klein, Tobias J., 2017. The effects of access to health insurance: evidence from a regression discontinuity design in Peru. J. Public Econ. 154, 122–136.

Calonico, Sebastian, Cattaneo, Matias D., Farrell, Max H., Titiunik, Rocio, 2018b. Regression discontinuity designs using covariates. Rev. Econ. Stat. (forthcoming).

Calonico, Sebastian, Cattaneo, Matias D., Max, H., 2016. Coverage error optimal confidence intervals for regression discontinuity designs. In: Working Paper.

Calonico, Sebastian, Cattaneo, Matias D., Max, H., 2018a. On the effect of bias estimation on coverage accuracy in nonparametric inference. J. Amer. Statist. Assoc. 113 (522), 767–779.

Calonico, Sebastian, Cattaneo, Matias D., Titiunik, Rocio, 2014. Robust nonparametric confidence intervals for regression discontinuity designs. Econometrica 82 (6), 2295–2326.

Card, David, Lee, David S., Pei, Zhuan, Weber, Andrea, 2015. Inference on causal effects in a generalized regression kink design. Econometrica 83 (6), 2453–2483.

Cattaneo, Matias D., Escanciano, Juan Carlos, 2017. Regression discontinuity designs: theory and applications. In: Advances in Econometrics, vol. 38. Emerald Publishing.

Chen, Heng, Fan, Yanqin, 2011. Identification and wavelet estimation of LATE in a class of switching regime models. In: Working Paper.

Chernozhukov, Victor, Chetverikov, Denis, Kato, Kengo, 2014. Gaussian Approximation of suprema of empirical processes. Ann. Statist. 42 (4), 1564–1597.

Chiang, Harold D., 2018. Cluster-robust simultaneous inference for many average partial effects in binary/fractional models. ArXiv preprint arXiv: 1812.09397.

Chiang, Harold D., Sasaki, Yuya, 2019. Causal inference by quantile regression kink designs. J. Econometrics (forthcoming).

Clark, Damon, Martorell, Paco, 2014. The signaling value of a high school diploma. J. Political Economy 122 (2), 282–318.

Cook, Thomas D., 2008. Waiting for life to arrive: a history of the regression-discontinuity design in psychology, statistics and economics. J. Econometrics 142 (2), 636–654.

Deshpande, Manasi, 2016. Does welfare inhibit success? The long-term effects of removing low-income youth from the disability rolls. Amer. Econ. Rev. 106 (11), 3300–3330.

Donald, Stephen G., Hsu, Yu-Chin, Barrett, Garry F., 2012. Incorporating covariates in the measurement of welfare and inequality: methods and applications. Econom. J. 15 (1), C1–C30.

Dong, Yingying, 2016. Jump or kink? Identification of binary treatment regression discontinuity design without the discontinuity. In: Working Paper.

Dony, Julia, Einmahl, Uwe, Mason, David M., 2006. Uniform in bandwidth consistency of local polynomial regression function estimators. Austrian J. Stat. 35 (2 & 3), 105–120.

Einmahl, Uwe., Mason, David M., 2000. An empirical process approach to the uniform consistency of kernel-type function estimators. J. Theoret. Prob. 13 (1), 1–37.

Fan, Jianqing, Gijbels, Irene, 1996. Local Polynomial Modelling and Its Applications: Monographs on Statistics and Applied Probability, vol. 66, CRC Press.

Fan, Yanqin, Guerre, Emmanuel, 2016. Multivariate local polynomial estimators: uniform boundary properties and asymptotic linear representation. In: Hill, R.C., Gonzalez-Rivera, G., Lee, T.-H. (Eds.), Advances in Econometrics, vol. 36. Emerald Publishing, pp. 489–537.

Fan, Yanqin, Liu, Ruixuan, 2016. A direct approach to inference in nonparametric and semiparametric quantile models. J. Econometrics 191 (1), 196–216.

Feir, Donna, Lemieux, Thomas, Marmer, Vadim, 2016. Weak identification in fuzzy regression discontinuity designs. J. Bus. Econom. Statist. 34 (2), 185–196.

Frandsen, Brigham R., Frölich, Markus, Melly, Blaise, 2012. Quantile treatment effects in the regression discontinuity design. J. Econometrics 168 (2), 382–395.

Giné, Evarist, Guillou, Armelle, 2002. Rates of strong uniform consistency for multivariate kernel density estimators. Annales de l'Institut Henri Poincaré Probabilités et Statistiques 38 (6), 907–921.

Giné, Evarist, Nickl, Richard, 2015. Mathematical Foundations of Infinite-Dimensional Statistical Models. Cambridge University Press.

Giné, Evarist, Zinn, Joel, 1990. Bootstrapping general empirical measures. Ann. Probab. 18 (2), 851–869.

Gormley Jr., William T., Gayer, Ted, Phillips, Deborah, Dawson, Brittany, 2005. The effects of universal pre-K on cognitive development. Developmental Psychology 41 (6), 872–884.

Guerre, Emmanuel, Sabbah, Camille, 2012. Uniform bias study and bahadur representation for local polynomial estimators of the conditional quantile function. Econom. Theory 28 (1), 87–129.

Hahn, Jinyong, Todd, Petra, van der Klaauw, Wilbert, 2001. Identification and estimation of treatment effects with a regression-discontinuity design. Econometrica 69 (1), 201–209.

Hansen, Bruce E., 1996. Inference when a nuisance parameter is not identified under the null hypothesis. Econometrica 64 (2), 413–430.

Imbens, Guido, Kalyanaraman, Karthik, 2012. Optimal bandwidth choice for the regression discontinuity estimator. Rev. Econom. Stud. 79 (3), 933–959.

Imbens, Guido, Lemieux, Thomas, 2008. Special issue editors' introduction: the regression discontinuity design – theory and applications. J. Econometrics 142 (2), 611–614.

Imbens, Guido W., Wooldridge, Jeffrey M., 2009. Recent developments in the econometrics of program evaluation. J. Econ. Lit. 47 (1), 5–86.

Ito, Koichiro, 2015. Asymmetric incentives in subsidies: evidence from a large-scale electricity rebate program. Amer. Econ. J.: Econom. Policy 7 (3), 209–237.

Kleibergen, Frank, 2002. Pivotal statistics for testing structural parameters in instrumental variables regression. Econometrica 70 (5), 1781–1803.

Kong, Efang, Linton, Oliver, Xia, Yingcun, 2010. Uniform bahadur representation for local polynomial estimates of M-regression and its application to the additive model? Econom. Theory 26 (5), 1529–1564.

Kosorok, Michael R., 2003. Bootstraps of sums of independent but not identically distributed stochastic processes. J. Multivariate Anal. 84 (2), 299–318.

Kosorok, Michael R., 2008. Introduction To Empirical Processes and Semiparametric Inference. Springer.

Landais, Camille, 2015. Assessing the welfare effects of unemployment benefits using the regression kink design. Amer. Econ. J.: Econom. Policy 7 (4), 243–278.

Ledoux, Michel, Talagrand, Michel, 1988. Un Critere sur les Petites Boules dans le Théoreme Limite Central. Probab. Theory Related Fields 77 (1), 29–47.

Lee, David S., 2008. Randomized experiments from non-random selection in U.S. house elections. J. Econometrics 142 (2), 675–697.

Lee, David S., Lemieux, Thomas, 2010. Regression discontinuity designs in economics. J. Econ. Lit. 48 (2), 281–355.

Lee, Sokbae, Song, Kyungchul, Whang, Yoon-Jae, 2015. Uniform asymptotics for nonparametric quantile regression with an application to testing monotonicity. ArXiv preprint arXiv:1506.05337.

Masry, Elias, 1996. Multivariate local polynomial regression for time series: Uniform strong consistency and rates. J. Time Series Anal. 17 (6), 571–599.

Moreira, Marcelo J., 2003. A conditional likelihood ratio test for structural models. Econometrica 71 (4), 1027–1048.

Nielsen, Helena Skyt, Sørensen, Torben, Taber, Christopher, 2010. Estimating the effect of student aid on college enrollment: Evidence from a government grant policy reform. Amer. Econ. J.: Econom. Policy 2 (2), 185–215.

Otsu, Taisuke, Xu, Ke-Li, Matsushita, Yukitoshi, 2015. Empirical likelihood for regression discontinuity design. J. Econometrics 186 (1), 94–112.

Pollard, David, 1990. Empirical Processes: Theory and Applications. In: NSF-CBMS Regional Conference Series in Probability and Statistics, vol. 2.

Porter, Jack, 2003. Estimation in the Regression Discontinuity Model. Department of Economics, University of Wisconsin at Madison, (unpublished).

Qu, Zhongjun, Yoon, Jungmo, 2015. Nonparametric estimation and inference on conditional quantile processes. J. Econometrics 185 (1), 1–19.

Qu, Zhongjun, Yoon, Jungmo, 2018. Uniform inference on quantile effects under sharp regression discontinuity designs. J. Bus. Econom. Statist. (forthcoming).

Shigeoka, Hitoshi, 2014. The effect of patient cost sharing on utilization, health, and risk protection. Amer. Econ. Rev. 104 (7), 2152–2184.

Simonsen, Marianne, Skipper, Lars, Skipper, Niels, 2016. Price sensitivity of demand for prescription drugs: exploiting a regression kink design. J. Appl. Econometrics 31 (2), 320–337.

Thistlethwaite, Donald L., Campbell, Donald T., 1960. Regression-discontinuity analysis: an alternative to the ex post facto experiment. J. Edu. Psych. 51 (6), 309–317.

Tsybakov, Alexandre B., 2008. Introduction to Nonparametric Estimation. In: Springer Series in Statistics.

van der Vaart, Aad W., 1998. Asymptotic Statistics. Cambridge University Press.

van der Vaart, Aad W., Wellner, Jon A., 1996. Weak Convergence and Empirical Processes. Springer-Verlag.

van der Vaart, Aad, Wellner, Jon A., 2011. A local maximal inequality under uniform entropy. Electron. J. Stat. 5, 192–203.