arXiv:1702.04430v2 [stat.ME] 24 Feb 2017

A Unified Robust Bootstrap Method for Sharp/Fuzzy Mean/Quantile Regression Discontinuity/Kink Designs

Harold D. Chiang* Yu-Chin Hsu† Yuya Sasaki‡

February 28, 2017

Abstract

Computation of asymptotic distributions is known to be a nontrivial and delicate task for the regression discontinuity designs (RDD) and the regression kink designs (RKD). It is even more complicated when a researcher is interested in joint or uniform inference across heterogeneous subpopulations indexed by covariates or quantiles. Hence, bootstrap procedures are often preferred in practice. This paper develops a robust multiplier bootstrap method for a general class of local Wald estimators. It applies to 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 examples, as well as covariate-indexed versions of them. In addition to its generic applicability to a wide variety of local Wald estimators, our method also enjoys robustness against large bandwidths commonly used in practice. This robustness is achieved through a bias correction approach incorporated into our multiplier bootstrap framework. We demonstrate the generic applicability of our theory through ten examples of local Wald estimators including those listed above, and show by simulation studies that it indeed performs well, robustly, and uniformly across different examples. All the code files are available upon request.

Keywords: bias correction, local Wald estimator, multiplier bootstrap, quantile, regression discontinuity design, regression kink design, robustness

JEL Codes: C01, C14, C21

*Harold D. Chiang: hchiang7@jhu.edu. Department of Economics, Johns Hopkins University.

†Yu-Chin Hsu: ychsu@econ.sinica.edu.tw. Institute of Economics, Academia Sinica.

‡Yuya Sasaki (Corresponding Author): sasaki@jhu.edu. Department of Economics, Johns Hopkins University.

1 Introduction

Empirical researchers have used many different versions of local Wald estimators. Of mostly wide use are the local Wald estimators for the regression discontinuity design (RDD). More recently, researchers also use 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. For all of these different versions of local Wald estimators, computation of asymptotic distributions – and thus computation of standard errors – is known to be a nontrivial and delicate task to varying extents. Because of this undesired feature, bootstrap methods that circumvent the necessity to calculate complicated covariance matrices/functions are preferred in practice, but their availability is surprisingly limited in the existing literature. In this paper, we propose a general multiplier bootstrap framework, which uniformly applies to all of the commonly used versions of local Wald estimators. In addition to its generic applicability, our method enjoys robustness against practically used bandwidth choice rules. Therefore, the proposed method is readily applicable to most, if not all, types of empirical research based on local Wald estimators.

We are not the first to develop a bootstrap- or simulation-based method of inference for local Wald estimators, but no existing method shares the generic applicability and the robustness that our method enjoys. A couple of important references are Qu and Yoon (2015b) and Bartalotti, Calhoun, and He (2016). Qu and Yoon (2015b) propose a simulation method with bias corrections to achieve robustness, specifically for the sharp quantile RDD. Bartalotti, Calhoun, and He (2016) propose a wild bootstrap method with bias corrections to achieve robustness, specifically for the sharp mean RDD. To our knowledge, on the other hand, there is no robust bootstrap method proposed for any of the other versions of local Wald estimators, such as the fuzzy mean RDD, the sharp mean RKD, the fuzzy mean RKD, the fuzzy quantile RDD, the sharp quantile RKD, and the fuzzy quantile RKD. Instead of proposing a robust bootstrap method which specifically applies to each single version of local Wald estimators, we propose one generic multiplier bootstrap 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.

Our contributions consist of the robustness and the generic applicability as argued above. We achieve the robustness against practically used bandwidth choice rules through higher-order bias corrections following Calonico, Cattaneo, and Titiunik (2014), as Qu and Yoon (2015b) and Bartalotti, Calhoun, and He (2016) do in their respective contexts. We achieve the generic applicability by taking advantage of the fact that all of the commonly used versions of local Wald estimators share a common basic form of Bahadur representations (BR). Notably, the insight of Porter (2003; Proof of Theorem 3) allows us to derive the BR for higher-order local polynomials which we require for robustness, and we further develop the uniform validity of the BR similar to Qu and Yoon (2015a; Theorem 1). We first establish the validity of the multiplier bootstrap with this common basic form of BR following Kosorok (2003, 2008), and then apply the functional delta method to both of the BR and the multiplier bootstrap processes. Thereby, we achieve the generic applicability through a single unified framework, with only the variable definitions and the Hadamard derivatives having to be changed across different versions of the local Wald estimators.

This paper is organized as follows. After reviewing the literature in Section 2, we first present a general framework of local Wald estimands in Section 3, with ten examples (Sections 3.1–3.10). Section 4 proposes general multiplier bootstrap processes, and establishes the validity of the general multiplier bootstrap method. We then demonstrate its applications to the aforementioned ten examples in Section 5, and conduct simulation studies in Section 6. Proofs and mathematical details are delegated to Section A in the appendix. We provide some practical guide for bandwidth choice in Section B in 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 today. Instead of enumerating all papers, we refer the readers to a collection of surveys, including Cook (2008), the special issue of *Journal of Econometrics* edited by Imbens and Lemieux (2008), Imbens and Wooldridge (2009; Sec. 6.4), Lee and Lemieux (2010), and Volume 38 of *Advances in Econometrics* edited by Cattaneo and Escanciano (2016), as well as the references cited therein. For technical matters, we mainly refer to Porter (2003) for its convenience in deriving the general BR for higher-order local polynomials. While it mostly evolved around the RDD, recent additions to this local design literature include the RKD (e.g., Nielsen, Sørensen, and Taber, 2010; Landais, 2015; Simonsen, Skipper, and Skipper, 2015; Card, Lee, Pei, and Weber, 2016; Dong, 2016) and quantile extensions (e.g., Frandsen, Frölich, and Melly, 2012; Qu and Yoon, 2015b). We emphasize once again that all these different frameworks are uniformly encompassed by the general framework developed in this paper.

Literature on Bandwidth and Robustness: Calonico, Cattaneo, and Titiunik (2014) introduce bias correction for the purpose of achieving 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 (e.g., Imbens and Kalyanaraman, 2012; Calonico, Cattaneo and Titiunik, 2014; Arai and Ichimura, 2016; Calonico, Cattaeneo, and Farrell, 2016ab). Bartalotti, Calhoun, and He (2016) adopt this idea of bias correction in a wild bootstrap method of inference for the sharp mean RDD. Qu and Yoon (2015b) adopt this idea of bias correction in a simulation-based method of inference for the sharp quantile RDD. We adopt this idea of bias correction in a multiplier bootstrap method of inference that uniformly applies to the sharp mean RDD (Section 3.2), the fuzzy mean RDD (Section 3.1), the sharp mean RKD (Section 3.4), the fuzzy mean RKD (Section 3.3), the sharp quantile RDD (Section 3.6), the fuzzy quantile RDD (Section 3.7), the sharp quantile RKD (Section 3.9), and the fuzzy quantile RKD (Section 3.8), among others. Furthermore, Calonico, Cattaneo, and Farrell (2016ab) 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. We also take advantage of these recent innovations.

Literature on Uniform Bahadur Representation: The Bahadur representation is a key to asymptotic distributional results. As we aim to cover uniform inference over a general index set, uniform validity of Bahadur representation over the set is essential for our method. For classes of nonparametric kernel regressions on which our method relies, Masry (1996) and Kong, Linton, and Xia (2010) develop uniform Bahadur representations over regressors. Furthermore, Guerre and Sabbah (2012) and Qu and Yoon (2015a) develop uniform validity over quantiles as well. We take advantage of and extend this existing idea in deriving the uniform validity of our Bahadur representation. In order to deal with a more general class of complexity rather than just quantiles, we use a new maximal inequality (van der Vaart and Wellner, 2011; Chernozhukov, Chetverikov, and Kato, 2014).

Literature on Multiplier Bootstrap: The multiplier bootstrap for Donsker and other weakly convergent classes is first studied by Ledoux and Talagrand (1988) and Giné and Zinn (1990). Our results are based on a more general version developed more lately by Kosorok (2003, 2008), that works for triangular arrays of row-independent non-identically distributed stochastic processes. This version provides a conditional-weak-convergence counterpart of the functional central limit theorem for row-independent arrays (Pollard, 1990). To our knowledge, use of the multiplier bootstrap in econometrics dates back to Hansen (1996). This method proves particularly useful for uniform inference on CDFs, for its convenience of not having to estimate a gigantic covariance matrix to approximate a covariance function. Barrett and Donald (2003) use the multiplier bootstrap for uniform inference on unconditional CDFs. More recently, Donald, Hsu, and Barrett (2012) use the multiplier bootstrap with kernel weights for uniform inference on conditional CDFs. In this paper, we use the multiplier bootstrap with one-sided kernel weights for uniform inference on a generalized class of local Wald estimators. In particular, our examples include uniform inference on conditional CDFs of potential outcomes (Section 3.5) as well as uniform inference on conditional quantile treatment effects via inversions of conditional CDFs (Sections 3.6–3.7). With all these preceding papers on the multiplier bootstrap, we want to emphasize on our own theoretical contributions. We face an extra challenge that the existing researches did not face. Namely, we need to allow for a more general assumption to uniformly deal with complexities of processes across different estimators to accommodate all the relevant examples. In addition, we need to take into account the variability of the estimates for unknowns in the operators of local Wald ratios in many examples that we consider. To deal with these issues, we develop new proofs for the validity of the multiplier bootstrap in a different way from those of the existing literature – we elaborate on this point in Section 4.4.

3 The General Framework and Ten Examples

Let (Y, D, X) be a random vector defined on a probability space $(\Omega^x, \mathcal{F}^x, \mathbb{P}^x)$, where Y contains an outcome vector, D contains a treatment vector, and X is a running variable or an assignment variable. We denote their supports as $\mathscr{Y}$, $\mathscr{D}$ and $\mathscr{X}$, respectively. Suppose that a researcher observes n independent copies $\{(Y_i, D_i, X_i)\}_{i=1}^n$ 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 $g_1 : \mathscr{Y} \times \Theta_1 \to \mathbb{R}$ and $g_2 : \mathscr{D} \times \Theta_2 \to \mathbb{R}$ be functions to be defined in various contexts of empirical research designs – concrete examples are suggested in the ten subsections below. When we discuss the continuity of g_k in θ_k, 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 v-th order partial derivatives with respect to x are denoted by $\mu_1^{(v)} = \frac{\partial^v}{\partial x^v}\mu_1$ and $\mu_2^{(v)} = \frac{\partial^v}{\partial x^v}\mu_2$. We also denote $b' = \frac{d}{dx}b$ for a function $b : \mathscr{X} \mapsto \mathbb{R}$. 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

$$\tau(\theta'') = \Upsilon\left(\frac{\phi\left(\lim_{x\downarrow 0} \mu_1^{(v)}(x,\cdot)\right)(\cdot) - \phi\left(\lim_{x\uparrow 0} \mu_1^{(v)}(x,\cdot)\right)(\cdot)}{\psi\left(\lim_{x\downarrow 0} \mu_2^{(v)}(x,\cdot)\right)(\cdot) - \psi\left(\lim_{x\uparrow 0} \mu_2^{(v)}(x,\cdot)\right)(\cdot)} \right)(\theta''). \tag{3.1}$$

for all $\theta'' \in \Theta''$. This class of local Wald estimands encompasses a wide array of design-based estimands used by empirical practitioners. We list examples in the ten subsections below (Sections 3.1–3.10). Throughout this paper, we normalize the cutoff value of the running variable to $x = 0$.

3.1 Example: Fuzzy Mean RDD

We do not need index sets for fuzzy RDD, and 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 let $\Upsilon\left(\frac{\mu_1(x,\cdot)}{\mu_2(x,\cdot)}\right)(\theta'') = \frac{\mu_1(x,\theta'')}{\mu_2(x,\theta'')}$ $\forall \theta'' \in \Theta''$. The local Wald estimand (3.1) with $v = 0$ in this setting becomes

$$\tau(\theta'') \quad = \quad \frac{\lim_{x\downarrow 0} \mathrm{E}[Y_i|X_i = x] - \lim_{x\uparrow 0} \mathrm{E}[Y_i|X_i = x]}{\lim_{x\downarrow 0} \mathrm{E}[D_i|X_i = x] - \lim_{x\uparrow 0} \mathrm{E}[D_i|X_i = x]} \tag{3.2}$$

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

3.2 Example: Sharp Mean RDD

Sharp RDD is a special case of fuzzy RDD, where $D_i = \mathbb{1}\{X_i \geq 0\}$. Thus, we can write $\mu_2(x, \theta_2) = \mathrm{E}[D_i|X_i = x] = \mathbb{1}\{x \geq 0\}$, and the local Wald estimand (3.1) of the form (3.2) further reduces to

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

for all $\theta'' \in \Theta'' = \{0\}$. This estimand $\tau(0)$ will be denoted by τ_{SMRD} for Sharp Mean RD design. For this estimand, Bartalotti, Calhoun, and He (2016) propose a robust bootstrap method of inference, and hence we are not the first to propose a robust bootstrap method for τ_{SMRD}. The benefit of our method is its applicability not only to τ_{SMRD}, but also to many other estimands of the form (3.1).

3.3 Example: Fuzzy Mean RKD

Define $\Theta_1, \Theta_2, \Theta_1', \Theta_2', \Theta''$, g_1, g_2, ϕ, ψ, and Υ as in Section 3.1. The local Wald estimand (3.1) with $v = 1$ in this setting becomes

$$\tau(\theta'') \quad = \quad \frac{\lim_{x\downarrow 0} \frac{\partial}{\partial x} \mathrm{E}[Y_i|X_i = x] - \lim_{x\uparrow 0} \frac{\partial}{\partial x} \mathrm{E}[Y_i|X_i = x]}{\lim_{x\downarrow 0} \frac{\partial}{\partial x} \mathrm{E}[D_i|X_i = x] - \lim_{x\uparrow 0} \frac{\partial}{\partial x} \mathrm{E}[D_i|X_i = x]} \tag{3.3}$$

for all $\theta'' \in \Theta'' = \{0\}$. This estimand $\tau(0)$ will be denoted by τ_{FMRK} for Fuzzy Mean RK design. See Card, Lee, Pei, and Weber (2016) for a causal interpretation of this estimand.

3.4 Example: Sharp Mean RKD

Sharp RKD is a special case of fuzzy RKD where the treatment is defined by $E[g_2(D_i, \theta_2)|X_i] = b(X_i)$ through a known function b. Thus, the local Wald estimand (3.1) of the form (3.3) further reduces to

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

for all $\theta'' \in \Theta'' = \{0\}$. This estimand $\tau(0)$ will be denoted by τ_{SMRK} for Sharp Mean RK design. See Card, Lee, Pei, and Weber (2016) for a causal interpretation of this estimand.

3.5 Example: CDF Discontinuity and Test of Stochastic Dominance

Let $\Theta_1 = \Theta_1' = \Theta'' = \mathscr{Y}$ for $\mathscr{Y} \subset \mathbb{R}$, and let $\Theta_2 = \Theta_2' = \{0\}$. Set $g_1(Y_i, \theta_1) = \mathbb{1}\{Y_i \leq \theta_1\}$ and $g_2(D_i, \theta_2) = D_i$, where $D_i = \mathbb{1}\{X_i \geq 0\}$ holds under the sharp RD design. Note that $\mu_1(x, \theta_1) = \mathrm{E}[g_1(Y_i, \theta_1)|X_i = x] = F_{Y|X}(\theta_1|x)$ and $\mu_2(x, \theta_2) = \mathrm{E}[g_2(D_i, \theta_2)|X_i = x] = \mathrm{E}[D_i|X_i = x] = \mathbb{1}\{x \geq 0\}$. Let ϕ and ψ be the identity operators, and let $\Upsilon\left(\frac{\mu_1(x,\cdot)}{\mu_2(x,\cdot)}\right)(\theta'') = \frac{\mu_1(x,\theta'')}{\mu_2(x,0)}$ $\forall \theta'' \in \Theta''$. The local Wald estimand (3.1) with $v = 0$ in this setting becomes

$$\tau(\theta'') \quad = \quad \lim_{x\downarrow 0} F_{Y|X}(\theta''|x) - \lim_{x\uparrow 0} F_{Y|X}(\theta''|x)$$

for all $\theta'' \in \Theta'' = \mathscr{Y} \subset \mathbb{R}$. This estimand τ will be denoted by τ_{SCRD} for Sharp CDF RD design. This estimand may be useful to test the hypothesis of stochastic dominance ($\tau(\theta'') \leq 0$ for all $\theta'' \in \Theta''$). See Shen and Zhang (2016) for this estimand and hypothesis testing.

3.6 Example: Sharp Quantile RDD

Denote $Q_{Y|X}(\theta'') := \inf\{y \in \mathscr{Y} : F_{Y|X}(y) \geq \theta''\}$, fix an $a \in (0, 1/2)$, $\varepsilon > 0$ and let $\mathscr{Y}_1 = [Q_{Y|X}(a|0^-) - \varepsilon, Q_{Y|X}(1-a|0^-) + \varepsilon] \cup [Q_{Y|X}(a|0^+) - \varepsilon, Q_{Y|X}(1-a|0^+) + \varepsilon]$.

Let $\Theta_1 = \mathscr{Y}_1$, $\Theta_1' = \Theta'' = [a, 1-a]$ and $\Theta_2 = \Theta_2' = \{0\}$. Set $g_1(Y_i, \theta_1) = \mathbb{1}\{Y_i \leq \theta_1\}$ and $g_2(D_i, \theta_2) = D_i$, where $D_i = \mathbb{1}\{X_i \geq 0\}$ holds under the sharp RDD. Note that $\mu_1(x, \theta_1) = \mathrm{E}[g_1(Y_i, \theta_1)|X_i = x] = F_{Y|X}(\theta_1|x)$ and $\mu_2(x, \theta_2) = \mathrm{E}[g_2(D_i, \theta_2)|X_i = x] = \mathrm{E}[D_i|X_i = x] = \mathbb{1}\{x \geq 0\}$. Let $\phi(F_{Y|X}(\cdot|x))(\theta') = \inf\{\theta_1 \in \Theta_1 : F_{Y|X}(\theta_1|x) \geq \theta'\}$ $\forall \theta' \in \Theta_1'$, $\psi(\mathbb{1}\{x \geq 0\})(\theta') = \mathbb{1}\{x \geq 0\}$ $\forall \theta' \in \Theta_2' = \{0\}$, and let $\Upsilon\left(\frac{\mu_1(x,\cdot)}{\mu_2(x,\cdot)}\right)(\theta'') = \frac{\mu_1(x,\theta'')}{\mu_2(x,0)}$ $\forall \theta'' \in \Theta''$. The local Wald estimand (3.1) with $v = 0$ in this setting becomes

$$\tau(\theta'') = \lim_{x \downarrow 0} Q_{Y|X}(\theta''|x) - \lim_{x \uparrow 0} Q_{Y|X}(\theta''|x)$$

for all $\theta'' \in \Theta'' = [a, 1-a]$, where $Q_{Y|X}(\theta''|x) := \inf\{\theta_1 \in \Theta_1 : F_{Y|X}(\theta_1|x) \geq \theta''\}$ for a short-hand notation. This estimand τ will be denoted by τ_{SQRD} for Sharp Quantile RD design. For this estimand, Qu and Yoon (2015b) propose a method of uniform inference based on uniform random sampling, and hence we are not the first to propose a bootstrap method for τ_{SQRD}. Our method adds the property of robustness to this existing method, besides the main result that it applies to other estimands too.

3.7 Example: Fuzzy Quantile RDD

Frandsen, Frölich and Melly (2012) propose the following estimand 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.

$$F_{Y^d|C}(y) = \frac{\lim_{x\downarrow 0} E[\mathbb{1}\{Y_i^* \leq y\} \cdot \mathbb{1}\{D_i^* = d\}|X_i = x] - \lim_{x\uparrow 0} E[\mathbb{1}\{Y_i^* \leq y\} \cdot \mathbb{1}\{D_i^* = d\}|X_i = x]}{\lim_{x\downarrow 0} E[\mathbb{1}\{D_i^* = d\}|X_i = x] - \lim_{x\uparrow 0} E[\mathbb{1}\{D_i^* = d\}|X_i = x]} \tag{3.4}$$

Further, they propose to identify the local quantile treatement effect by $F^{-1}_{Y^1|C} - F^{-1}_{Y^0|C}$. This estimand also fits in the general framework (3.1). Denote $Q_{Y^d|C}(\theta'') = \inf\{y \in \mathscr{Y} : F_{Y^d|C}(y) \geq \theta''\}$. 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 $g_1((Y_i^*, D_i^*), (y, d)) = \mathbb{1}\{Y_i^* \leq y\} \cdot \mathbb{1}\{D_i^* = d\}$ and $g_2(D_i^*, d) = \mathbb{1}\{D_i^* = d\}$. Note that $\mu_1(x, y, d) = \mathrm{E}[g_1((Y_i^*, D_i^*), y, d)|X_i = x] = \mathrm{E}[\mathbb{1}\{Y_i^* \leq y\} \cdot \mathbb{1}\{D_i^* = d\}|X_i = x]$ 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 let Υ be defined by $\Upsilon(F_{Y^\cdot|C})(\theta'') = \inf\{y \in \mathscr{Y} : F_{Y^1|C}(y) \geq \theta''\} - \inf\{y \in \mathscr{Y} : F_{Y^0|C}(y) \geq \theta''\}$. The local Wald estimand (3.1) with $v = 0$ in this setting becomes

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

for all $\theta'' \in \Theta'' = [a, 1-a]$, where $Q_{Y^d|C}(\theta'') := \inf\{y \in \mathscr{Y} : F_{Y^d|C}(y) \geq \theta''\}$ for a short-hand notation, and $F_{Y^d|C}(y)$ is given in (3.4) for all $(y, d) \in \mathscr{Y} \times \mathscr{D}$. This estimand τ will be denoted by τ_{FQRD} for Fuzzy Quantile RD design.

3.8 Example: Fuzzy Quantile RKD

Let $\mathscr{Y}_1 = [Q_{Y|X}(a|0) - \varepsilon, Q_{Y|X}(1-a|0) + \varepsilon]$, $\Theta_1 = \mathscr{Y}_1$, $\Theta_1' = \Theta'' = [a, 1-a]$ for a constant $a \in (0, 1/2)$, and $\Theta_2 = \Theta_2' = \{0\}$. We set $g_1(Y_i, \theta_1) = \mathbb{1}\{Y_i \leq \theta_1\}$ and $g_2(D_i, \theta_2) = D_i$. Note that $\mu_1(x, \theta_1) = \mathrm{E}[g_1(Y_i, \theta_1)|X_i = x] = F_{Y|X}(\theta_1|x)$ and $\mu_2(x, \theta_2) = \mathrm{E}[g_2(D_i, \theta_2)|X_i = x] = \mathrm{E}[D_i|X_i = x]$. With the short-hand notations $f_{Y|X} = \frac{\partial}{\partial y} F_{Y|X}$ and $F_{Y|X}^{(1)} := \frac{\partial}{\partial x} F_{Y|X}$, let

$$\phi(F_{Y|X}'(\cdot|x))(\theta') := -\frac{F_{Y|X}^{(1)}(\inf\{\theta \in \Theta_1 : F_{Y|X}(\theta|0) \geq \theta'\}|x)}{f_{Y|X}(\inf\{\theta \in \Theta_1 : F_{Y|X}(\theta|0) \geq \theta'\}|x)} \qquad \forall \theta' \in \Theta_1',$$

let ψ be the identity operator, and let $\Upsilon\left(\frac{\mu_1(x,\cdot)}{\mu_2(x,\cdot)}\right)(\theta'') = \frac{\mu_1(x,\theta'')}{\mu_2(x,0)}$ $\forall \theta'' \in \Theta''$. We emphasize that $F_{Y|X}^{(1)}(\cdot|x)$ does *not* map to $f_{Y|X}(\cdot|0)$ or $F_{Y|X}(\cdot|0)$ in the definition of ϕ; instead $f_{Y|X}(\cdot|0)$ and $F_{Y|X}(\cdot|0)$ are embedded in the definition of ϕ. It will be shown that that $\phi(F_{Y|X}^{(1)}(\cdot|x))(\theta'') = \frac{\partial}{\partial x} Q_{Y|X}(\theta''|x)$. The local Wald estimand (3.1) with $v = 1$ in this setting becomes

$$\tau(\theta'') \quad = \quad \frac{\lim_{x \downarrow 0} \frac{\partial}{\partial x} Q_{Y|X}(\theta''|x) - \lim_{x \uparrow 0} \frac{\partial}{\partial x} Q_{Y|X}(\theta''|x)}{\lim_{x \downarrow 0} \frac{d}{dx} \mathrm{E}[D_i|X_i = x] - \lim_{x \uparrow 0} \frac{d}{dx} \mathrm{E}[D_i|X_i = x]} \tag{3.5}$$

for all $\theta'' \in \Theta'' = [a, 1-a]$. This estimand τ will be denoted by τ_{FQRK} for Fuzzy Quantile RK design. See Chiang and Sasaki (2016) for its causal interpretation.

3.9 Example: Sharp Quantile RKD

Sharp quantile RKD is a special case of fuzzy quantile RKD where the treatment is defined by $D_i = b(X_i)$ as a known function b of X_i. Thus, the local Wald estimand (3.1) of the form (3.5) further reduces to

$$\tau(\theta'') \quad = \quad \frac{\lim_{x\downarrow 0} \frac{\partial}{\partial x} Q_{Y|X}(\theta''|x) - \lim_{x\uparrow 0} \frac{\partial}{\partial x} Q_{Y|X}(\theta''|x)}{\lim_{x\downarrow 0} b'(x) - \lim_{x\uparrow 0} b'(x)}$$

for all $\theta'' \in \Theta'' = [a, 1-a]$. This estimand τ will be denoted by τ_{SQRK} for Sharp Quantile RK design. See Chiang and Sasaki (2016) for its causal interpretation.

3.10 Example: Group Covariate and Test of Heterogeneous Treatment Effects

Suppose that a researcher wants to make a joint inference for the average causal effects across observed heterogeneous groups G_i taking categorical values in $\Theta_1 = \Theta_2 = \Theta_1' = \Theta_2' = \Theta'' = \{1, \cdots, K\}$ by RDD. Let $Y_i = (Y_i^*, G_i)$ and $D_i = (D_i^*, G_i)$. Set $g_1((Y_i^*, G_i), \theta_1) = Y_i^* \cdot \mathbb{1}\{G_i = \theta_1\}$ and $g_2((D_i^*, G_i), \theta_2) = D_i^* \cdot \mathbb{1}\{G_i = \theta_2\}$. Note that $\mu_1(x, \theta_1) = \mathrm{E}[g_1((Y_i^*, G_i), \theta_1)|X_i = x] = \mathrm{E}[Y_i^* \cdot \mathbb{1}\{G_i = \theta_1\}|X_i = x]$ and $\mu_2(x, \theta_2) = \mathrm{E}[g_2((D_i^*, G_i), \theta_2)|X_i = x] = \mathrm{E}[D_i^* \cdot \mathbb{1}\{G_i = \theta_2\}|X_i = x]$. Let ϕ and ψ be the identity operators, and let $\Upsilon \left(\frac{\mu_1(x,\cdot)}{\mu_2(x,\cdot)} \right)(\theta'') = \frac{\mu_1(x,\theta'')}{\mu_2(x,\theta'')}$ $\forall \theta'' \in \Theta''$. The local Wald estimand (3.1) with $v = 0$ in this setting becomes

$$\tau(\theta'') \quad = \quad \frac{\lim_{x\downarrow 0} E[Y_i^* \cdot \mathbb{1}\{G_i = \theta''\}|X_i = x] - \lim_{x\uparrow 0} E[Y_i^* \cdot \mathbb{1}\{G_i = \theta''\}|X_i = x]}{\lim_{x\downarrow 0} E[D_i^* \cdot \mathbb{1}\{G_i = \theta''\}|X_i = x] - \lim_{x\uparrow 0} E[D_i^* \cdot \mathbb{1}\{G_i = \theta''\}|X_i = x]}$$

for all $\theta'' \in \Theta'' = \{1, \cdots, K\}$. This estimand τ will be denoted by τ_{GFMRD} for Group Fuzzy Mean RD design. This estimand may be useful to test the hypotheses of heterogeneous treatment effects ($\tau(\theta_1'') \neq \tau(\theta_2'')$ for some $\theta_1'', \theta_2'' \in \Theta''$) or unambiguous treatment significance ($\tau(\theta'') > 0$ for all $\theta'' \in \Theta''$). While we introduced this group estimand for the fuzzy mean regression discontinuity design, we remark that a similar estimand can be developed for any combinations of sharp/fuzzy mean/quantile regression discontinuity/kink designs.

3.11 Notations

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

$$\sigma_{kl}(\theta, \vartheta|x) = E[\mathcal{E}_k(Y_i, D_i, X_i, \theta)\ \mathcal{E}_l(Y_i, D_i, X_i, \vartheta)|X_i = x]$$

denote the conditional covariance of residuals for each $\theta = (\theta_1, \theta_2)$, $\vartheta = (\vartheta_1, \vartheta_2) \in \Theta$. 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 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 $r_p(x) = (1, x, ..., x^p)'$. 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 $n \in \mathbb{N}$. We write $\Gamma_p^\pm = \int_{\mathbb{R}_\pm} K(u) r_p(u) r_p'(u) du$ and $\Lambda_{p,q}^\pm = \int_{\mathbb{R}_\pm} u^q K(u) r_p(u) du$.

The notation $\rightsquigarrow$ will be used to denote weak convergence and $\underset{\xi}{\overset{p}{\rightsquigarrow}}$ for the notation of conditional weak convergence (convergence of the conditional limit laws of bootstraps), as defined in Section 2.2.3 of Korosok (2008). See Section A.2 in the Mathematical Appendix for more details. 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. Let v, p, $q \in \mathbb{T}_+$ with $v \le p < q$. We will use v for the order of derivative of interest as in (3.1), and p stands for the order of local polynomial fitting to estimate (3.1).

4 The General Results

4.1 The Local Wald Estimator

We first develop an estimator for the nonparametric components $\mu_k^{(v)}(0^\pm, \cdot)$, $k \in \{1, 2\}$, of the local Wald estimand (3.1) based on local polynomial fitting with the bias correction approach proposed by

Calonico, Cattaneo and Titiunik (2014). Under proper smoothness assumptions to be formally stated below, the p-th order approximations

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

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

$$\hat{\alpha}_{1\pm,p}(\theta_k) = \underset{\alpha\in\mathbb{R}^{p+1}}{\arg\min} \sum_{i=1}^n \delta_i^\pm \Big(g_1(Y_i,\theta_1) - r_p(\frac{X_i}{h_{1,n}(\theta_1)})'\alpha\Big)^2 K\left(\frac{X_i}{h_{1,n}(\theta_1)}\right) \qquad (4.1)$$
$$\hat{\alpha}_{2\pm,p}(\theta_2) = \underset{\alpha\in\mathbb{R}^{p+1}}{\arg\min} \sum_{i=1}^n \delta_i^\pm \Big(g_2(D_i,\theta_2) - r_p(\frac{X_i}{h_{2,n}(\theta_2)})'\alpha\Big)^2 K\left(\frac{X_i}{h_{2,n}(\theta_2)}\right) \qquad (4.2)$$

for each $k \in \{1,2\}$, where $\delta_i^+ = \mathbb{1}\{X_i \le 0\}$ and $\delta_i^- = \mathbb{1}\{X_i \ge 0\}$. We let the coordinates of these estimates be written by

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

With these component estimates, the local Wald estimand (3.1) is in turn estimated by plug in.

$$\hat{\tau}(\theta'') = \Upsilon\left(\frac{\phi\left(\hat{\mu}_{1,p}^{(v)}(0^+,\cdot)\right)(\cdot) - \phi\left(\hat{\mu}_{1,p}^{(v)}(0^-,\cdot)\right)(\cdot)}{\psi\left(\hat{\mu}_{2,p}^{(v)}(0^+,\cdot)\right)(\cdot) - \psi\left(\hat{\mu}_{2,p}^{(v)}(0^-,\cdot)\right)(\cdot)}\right)(\theta'') \qquad \text{for each } \theta'' \in \Theta''. \qquad (4.3)$$

The rest of this section develops a method to approximate the asymptotic distribution of the process $\sqrt{nh_n^{1+2v}}\,[\hat{\tau}(\cdot) - \tau(\cdot)]$ for the local Wald estimator. In order to translate the nonparametric estimators, $\hat{\mu}_{1,p}^{(v)}(0^\pm,\cdot)$ and $\hat{\mu}_{2,p}^{(v)}(0^\pm,\cdot)$, into a bootstrap framework, we first derive uniform Bahadur representations (BR) of them in Section 4.2. In Section 4.3, we define estimated multiplier processes (EMP) based on the uniform BR. The validity of the bootstrap is then established in Section 4.4 by the weak convergence of the BR and the conditional weak convergence of the EMP. Finally, the functional delta method and functional chain rule establish the bootstrap validity for the process $\sqrt{nh_n^{1+2v}}\,[\hat{\tau}(\cdot) - \tau(\cdot)]$ of interest.

4.2 Uniform Bahadur Representations

The following set of assumptions is used to develop uniform BR.

Assumption 1 (Uniform Bahadur Representation)**.** *Let* $\underline{x} < 0 < \overline{x}$*,*

(i) (a) $\{(Y_i, D_i, X_i)\}_{i=1}^n$ *are* n *independent 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* $j = 0, 1, 2, ..., p+1$*; (c) For any* (θ, k)*,* $(\vartheta, l) \in \mathbb{T}$*, we have* $\sigma_{kl}(\theta, \vartheta | \cdot\,) \in \mathcal{C}^1([\underline{x}, \overline{x}] \setminus \{0\})$ *with bounded derivatives in* x *and* $\sigma_{kl}(\theta, \vartheta | 0^{\pm}) < \infty$*; (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* 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 at $X = 0$. 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 $|\mu_k(x, \bar{\theta}_k) - \mu_k(x, \theta_k)| \leq M_k(x) |\bar{\theta}_k - \theta_k|$ 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, $\{y \mapsto \mathbb{1}\{y \leq y'\} : y' \in \mathscr{Y}\}$ 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 Section 5, we will check these high-level assumptions with primitive sufficient assumptions for each of the ten specific example presented in Sections 3.1–3.10. Condition (iii) specifies admissible rates of bandwidths, which are consistent with common choice rules to be presented in Section B. 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. Under this set of assumptions, we obtain the uniform BR of the local polynomial estimators $\hat{\mu}^{(v)}_{1,p}(0^{\pm},\cdot)$ and $\hat{\mu}^{(v)}_{2,p}(0^{\pm},\cdot)$ presented in the following lemma – see Section A.1 for a proof.

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

$$\sqrt{nh^{1+2v}_{k,n}(\theta_k)}\big(\hat{\mu}^{(v)}_{k,p}(0^{\pm},\theta_k)-\mu^{(v)}_k(0^{\pm},\theta_k)-h^{p+1-v}_{k,n}(\theta_k)\frac{e_v'(\Gamma^{\pm}_p)^{-1}\Lambda^{\pm}_{p,p+1}}{(p+1)!}\mu^{(p+1)}_k(0^{\pm},\theta_k)\big)$$
$$=v!\sum_{i=1}^{n}\frac{e_v'(\Gamma^{\pm}_p)^{-1}\mathcal{E}_k(Y_i,D_i,X_i,\theta)r_p(\frac{X_i}{h_{k,n}(\theta_k)})K(\frac{X_i}{h_{k,n}(\theta_k)})\delta^{\pm}_i}{\sqrt{nh_{k,n}(\theta_k)}f_X(0)}+o^x_p(1)$$

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 p-th order bias reduction. This property is the key to develop a method of inference which is robust against large bandwidth parameter choices.

4.3 Multiplier Bootstrap: Multiplier Processes

We now propose the multiplier bootstrap processes. Write the BR in Lemma 1 by

$$\nu^{\pm}_n(\theta,k)=v!\sum_{i=1}^{n}\frac{e_v'(\Gamma^{\pm}_p)^{-1}\mathcal{E}_k(Y_i,D_i,X_i,\theta)r_p(\frac{X_i}{h_{k,n}(\theta_k)})K(\frac{X_i}{h_{k,n}(\theta_k)})\delta^{\pm}_i}{\sqrt{nh_{k,n}(\theta_k)}f_X(0)}$$

for each $(\theta,k)\in\mathbb{T}$. We also write $\nu_n(\cdot)=\nu^+_n(\cdot)-\nu^-_n(\cdot)$.

To simulate the limiting process of the BR, we use the pseudo random sample $\{\xi_i\}^n_{i=1}$ drawn from the standard normal distribution, independently from the data $\{(Y_i,D_i,X_i)\}^n_{i=1}$. Precisely, $\{\xi_i\}^n_{i=1}$ 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 2 (v) below. With this pseudo random sample, define the multiplier processes (MP)

$$\nu^{\pm}_{\xi,n}(\theta,k) = v! \sum_{i=1}^{n} \xi_i \frac{e'_v(\Gamma^{\pm}_p)^{-1}\mathcal{E}_k(Y_i,D_i,X_i,\theta) r_p(\frac{X_i}{h_{k,n}(\theta_k)}) K(\frac{X_i}{h_{k,n}(\theta_k)})\delta^{\pm}_i}{\sqrt{nh_{k,n}(\theta_k)} f_X(0)}$$

for each $(\theta, k) \in \mathbb{T}$. We also write $\nu_{\xi,n}(\cdot) = \nu^{+}_{\xi,n}(\cdot) - \nu^{-}_{\xi,n}(\cdot)$.

In practice, we need to replace $\mathcal{E}_k$ and f_X by 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 $\mathcal{E}_k(Y_i, D_i, X_i, \theta)\mathbb{1}\{|X_i/h_{k,n}(\theta_k)| \le 1\}$. Write its estimate by $\hat{\mathcal{E}}_k(Y_i, D_i, X_i, \theta)\mathbb{1}\{|X_i/h_{k,n}(\theta_k)| \le 1\}$ which has $\mu_{k,p}$ replaced by some estimate $\tilde{\mu}_{k,p}$ of $\mu_{k,p}$. Section A.4 discusses the effects of these first-stage estimates. Substituting these estimated components in the MP, we define the estimated multiplier processes (EMP)

$$\hat{\nu}^{\pm}_{\xi,n}(\theta,k) = v! \sum_{i=1}^{n} \xi_i \frac{e'_v(\Gamma^{\pm}_p)^{-1}\hat{\mathcal{E}}_k(Y_i,D_i,X_i,\theta) r_p(\frac{X_i}{h_{k,n}(\theta_k)}) K(\frac{X_i}{h_{k,n}(\theta_k)})\delta^{\pm}_i}{\sqrt{nh_{k,n}(\theta_k)} \hat{f}_X(0)}$$

for each $(\theta, k) \in \mathbb{T}$. We also write $\hat{\nu}_{\xi,n}(\cdot) = \hat{\nu}^{+}_{\xi,n}(\cdot) - \hat{\nu}^{-}_{\xi,n}(\cdot)$.

4.4 Multiplier Bootstrap: Uniform Validity

In this section, we derive the asymptotic distribution of the BR, $\nu^{\pm}_n(\cdot)$, and show that the EMP, $\hat{\nu}^{\pm}_{\xi,n}(\cdot)$, can be used to approximate this asymptotic distribution, i.e., the uniform validity of the multiplier bootstrap is established. By the functional delta method, these asymptotic distributions translate into the asymptotic distribution of the process $\sqrt{nh_n^{1+2v}}[\hat{\tau}(\cdot) - \tau(\cdot)]$ for the local Wald estimator. For convenience of writing, we introduce the notation

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

for the intermediate local Wald estimand. We invoke the following two sets of assumptions.

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* $n\to\infty$.

(iv) $\{\xi_i\}_{i=1}^n$ *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 3 (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* $f_X(0)$.

Primitive conditions for Assumption 2 will be discussed for each of the ten examples in Section 5. We will propose $\tilde{\mu}_k$ in Section A.4 to satisfy Assumption 3 under Assumptions 1 and 2. At the current level of generality, we state these high-level assumptions. The following theorem states the main general results of this paper.

Theorem 1. *Under Assumptions 1 and 2(i)-(iii), we have the following results.*

(i) $\nu_n^\pm\rightsquigarrow\mathbb{G}_{H^\pm}$, *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,*

$$\begin{aligned}&\sqrt{nh_n^{1+2v}}[\hat{\tau}-\tau]\\ &\rightsquigarrow\Upsilon'_W\Big(\frac{[\psi(\mu_2^{(v)}(0+,\cdot))-\psi(\mu_2^{(v)}(0^-,\cdot))]\mathbb{G}'(\cdot,1)-[\phi(\mu_1^{(v)}(0^+,\cdot))-\phi(\mu_1^{(v)}(0^-,\cdot))]\mathbb{G}'(\cdot,2)}{[\psi(\mu_2^{(v)}(0+,\cdot))-\psi(\mu_2^{(v)}(0^-,\cdot))]^2}\Big),\end{aligned}$$

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

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

(ii) If Assumptions 2(iv) and 3 also hold in addition, then $\hat{\nu}^{\pm}_{\xi,n}(\cdot) \overset{p}{\underset{\xi}{\rightsquigarrow}} \mathbb{G}_{H^\pm}(\cdot)$*, and therefore*

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

where

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

See Section A.3 for a proof. In part (i), we first derive the asymptotic distributions of the BR, $\nu^{\pm}_n(\cdot)$. By the functional delta method, these asymptotic distributions translate into the asymptotic distribution of the process $\sqrt{nh_n^{1+2v}}[\hat{\tau}(\cdot) - \tau(\cdot)]$ for the local Wald estimator. In part (ii), we show that the EMP, $\hat{\nu}^{\pm}_{\xi,n}(\cdot)$, can be used to approximate this asymptotic distribution, i.e., the uniform validity of the multiplier bootstrap is established. Therefore, using the functional delta method again, we propose to let the transformed EMP approximate the asymptotic distribution of the process $\sqrt{nh_n^{1+2v}}[\hat{\tau}(\cdot) - \tau(\cdot)]$ for the local Wald estimator.

Remark 1. *As mentioned in Section 2, we develop proofs differently from those of the existing literature. We elaborate on two pieces of our theoretical innovations here. We need to allow for a more general assumption to uniformly deal with complexities of processes across different estimators to accommodate all the relevant examples. To deal with this issue, we show that Vapnik-Chervonenkis (VC) type class provides a convenient notion of complexity that is sufficient in most examples. In addition, we need to take into account the variability of the estimates for unknowns in the operators of local Wald ratios in many examples that we consider. To deal with this issue, we combine Theorem 2 of Kosorok (2003) and our Lemma 4 to show that the estimated multiplier process conditional on the* *data can accurately characterizes the covariance structure of sample paths and thus can be used for inference on the limiting Gaussian processes of interest provided that the unknowns can be uniformly consistently estimated.*

Part (i) of the Theorem further implies the following useful asymptotic identities. Because these implications themselves prove useful when we apply Theorem 1 to the ten aforementioned examples, we state them as corollaries below for convenience of later references.

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

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

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

5 Applications of the General Results to the Ten Examples

In this section, we apply the general results of Section 4 to the ten examples introduced in Sections 3.1–3.9. Throughout this section, we present our assumptions for the case of $p = 2$. We remark that, however, using a different order p of local polynomial fitting is also possible by similar arguments.

5.1 Example: Sharp Mean RDD

Consider Θ_1, Θ_2, Θ_1', Θ_2', Θ'', g_1, g_2, ϕ, ψ, and Υ defined in Section 3.2. Recall that we denote the local Wald estimand (3.1) with $v = 0$ in this setting by τ_{SMRD}. We also denote the analog estimator (4.3) with $v = 0$ in this setting by

$$\hat{\tau}_{SMRD} = \hat{\mu}_{1,2}(0^+, 0) - \hat{\mu}_{1,2}(0^-, 0).$$

For this application, we consider the following set of assumptions.

Assumption SMRD.

(i) (a) $\{(Y_i, X_i)\}_{i=1}^n$ are n independent copies of the random vector (Y, 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 $[\underline{x}, \overline{x}]$ that contains 0 in its interior and satisfies $0 < f_X(0) < \infty$.

(ii) (a) $E[|Y|^{2+\epsilon}|X = \cdot] < \infty$ on $[\underline{x}, \overline{x}]\backslash\{0\}$ for some $\epsilon > 0$. (b) $\frac{\partial^j}{\partial x^j} E[Y|X = \cdot]$ is Lipschitz on $[\underline{x}, 0)$ and $(0, \overline{x}]$ for $j = 0, 1, 2, 3$.

(iii) h_n satisfies $h_n \to 0$, $nh_n^7 \to 0$ and $nh_n^2 \to \infty$.

(iv) (a) $K : [-1, 1] \to \mathbb{R}^+$ is bounded and continuous. (b) $\{K(\cdot/h) : h > 0\}$ is of VC type. (c) $\Gamma_2^\pm$ is positive definite.

(v) $V(Y|X = \cdot) \in \mathcal{C}^1([\underline{x}, \overline{x}] \setminus \{0\})$ with bounded derivative in x and $0 < V(Y|X = 0^\pm) < \infty$

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

Define the EMP

$$\hat{\nu}_{\xi,n}^\pm = \sum_{i=1}^n \xi_i \frac{e_0'(\Gamma_2^\pm)^{-1}[Y_i - \tilde{\mu}_{1,2}(X_i, 0)] r_2(\frac{X_i}{h_n}) K(\frac{X_i}{h_n}) \delta_i^\pm}{\sqrt{nh_n} \hat{f}_X(0)},$$

where $\tilde{\mu}_{1,2}$ is defined in the statement of Lemma 9. Our general result applied to the current case yields the following corollary.

Corollary 3 (Example: Sharp Mean RDD)**.** *Suppose that Assumption SMRD holds.*

(i) There exists $\sigma_{SMRD} > 0$ such that

$$\sqrt{nh_n}[\hat{\tau}_{SMRD} - \tau_{SMRD}] \rightsquigarrow N(0, \sigma_{SMRD}^2)$$

(ii) Furthermore, with probability approaching to one,

$$\hat{\nu}_{\xi,n}^+ - \hat{\nu}_{\xi,n}^- \underset{\xi}{\overset{p}{\rightsquigarrow}} N(0, \sigma_{SMRD}^2).$$

A proof is provided in Section A.6.1. Perhaps the most practically relevant application of this corollary is the test of the null hypothesis of treatment nullity:

$$H_0 : \tau_{SMRD} = 0.$$

To test this hypothesis, we can use $\sqrt{nh_n}\,|\hat{\tau}_{SMRD}|$ as the test statistic, and use $\left|\hat{\nu}^+_{\xi,n} - \hat{\nu}^-_{\xi,n}\right|$ to simulate its asymptotic distribution.

5.2 Example: Sharp Mean RKD

Consider Θ_1, Θ_2, Θ_1', Θ_2', Θ'', g_1, g_2, ϕ, ψ, and Υ defined in Section 3.4. Recall that we denote the local Wald estimand (3.1) with $v = 1$ in this setting by τ_{SMRK}. We also denote the analog estimator (4.3) with $v = 1$ in this setting by

$$\hat{\tau}_{SMRK} = \frac{\hat{\mu}^{(1)}_{1,2}(0^+, 0) - \hat{\mu}^{(1)}_{1,2}(0^-, 0)}{b'(0^+) - b'(0^-)}.$$

For this application, we consider the following set of assumptions.

Assumption SMRK.

(i) (a) $\{(Y_i, X_i)\}_{i=1}^n$ are n independent copies of the random vector (Y, 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 $[\underline{x}, \overline{x}]$ that contains 0 in its interior and satisfies $0 < f_X(0) < \infty$.

(ii) (a) $E[|Y|^{2+\epsilon}|X = \cdot] < \infty$ on $[\underline{x}, \overline{x}]\backslash\{0\}$ for some $\epsilon > 0$. (b) $\frac{\partial^j}{\partial x^j} E[Y|X = \cdot]$ is Lipschitz on $[\underline{x}, 0)$ and $(0, \overline{x}]$ for $j = 0, 1, 2, 3$.

(iii) h_n satisfies $h_n \to 0$, $nh_n^3 \to \infty$ and $nh_n^7 \to 0$.

(iv) (a) $K : [-1, 1] \to \mathbb{R}^+$ is bounded and continuous. (b) $\{K(\cdot/h) : h > 0\}$ is of VC type. (c) $\Gamma_2^{\pm}$ is positive definite.

(v) $V(Y|X = x) \in \mathcal{C}^1([\underline{x}, \overline{x}] \setminus \{0\})$ with bounded derivative in x and $0 < V(Y|X = 0^{\pm}) < \infty$

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

(vii) b is continuously differentiable on $I \setminus \{0\}$ and $b'(0^+) - b'(0^-) \neq 0$.

Define the EMP

$$\hat{\nu}^{\pm}_{\xi,n} = \sum_{i=1}^n \xi_i \frac{e_1'(\Gamma_2^{\pm})^{-1}[Y_i - \tilde{\mu}_{1,2}(X_i, 0)] r_2(\frac{X_i}{h_n}) K(\frac{X_i}{h_n})}{\sqrt{nh_n}\hat{f}_X(0)},$$

where $\tilde{\mu}_{1,2}$ is defined in the statement of Lemma 9. Our general result applied to the current case yields the following corollary.

Corollary 4 (Example: Sharp Mean RKD)**.** *Suppose that Assumption SMRK holds.*

(i) There exists $\sigma_{SMRK} > 0$ *such that*

$$\sqrt{nh_n^3}[\hat{\tau}_{SMRK} - \tau_{SMRK}] \rightsquigarrow N(0, \sigma_{SMRK}^2).$$

(ii) Furthermore, with probability approaching to one,

$$\frac{\hat{\nu}_{\xi,n}^+ - \hat{\nu}_{\xi,n}^-}{b'(0^+) - b'(0^-)} \underset{\xi}{\overset{p}{\rightsquigarrow}} N(0, \sigma_{SMRK}^2).$$

This corollary can be proved similarly to Corollary 3. Perhaps the most practically relevant application of this corollary is the test of the null hypothesis of treatment nullity:

$$H_0 : \tau_{SMRK} = 0.$$

To test this hypothesis, we can use $\sqrt{nh_n^3}\,|\hat{\tau}_{SMRK}|$ as the test statistic, and use $\left|\frac{\hat{\nu}_{\xi,n}^+ - \hat{\nu}_{\xi,n}^-}{b'(0^+) - b'(0^-)}\right|$ to simulate its asymptotic distribution.

5.3 Example: Fuzzy Mean RDD

Consider Θ_1, Θ_2, Θ_1', Θ_2', Θ'', g_1, g_2, ϕ, ψ, and Υ defined in Section 3.1. Recall that we denote the local Wald estimand (3.1) with $v = 1$ in this setting by τ_{FMRD}. We also denote the analog estimator (4.3) with $v = 1$ in this setting by

$$\hat{\tau}_{FMRD} = \frac{\hat{\mu}_{1,2}(0^+, 0) - \hat{\mu}_{1,2}(0^-, 0)}{\hat{\mu}_{2,2}(0^+, 0) - \hat{\mu}_{2,2}(0^-, 0)}.$$

For this application, we consider the following set of assumptions.

Assumption FMRD.

(i) (a) $\{(Y_i, X_i, D_i)\}_{i=1}^n$ are n independent copies of the random vector (Y, X, D) 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

$[\underline{x}, \overline{x}]$ that contains 0 in its interior and satisfies $0 < f_X(0) < \infty$.

(ii) (a) $E[|Y|^{2+\epsilon}|X = \cdot] < \infty$ on $[\underline{x}, \overline{x}]\backslash\{0\}$ for some $\epsilon > 0$. (b) $\frac{\partial^j}{\partial x^j}E[Y|X = \cdot]$ and $\frac{\partial^j}{\partial x^j}E[D|X = \cdot]$ are Lipschitz on $[\underline{x}, 0)$ and $(0, \overline{x}]$ for $j = 0, 1, 2, 3$. (d) $E[D|X = 0^+] \neq E[D|X = 0^-]$.

(iii) The baseline bandwidth h_n satisfies $h_n \to 0$, $nh_n^2 \to \infty$, $nh_n^7 \to 0$. There exist constants c_1, c_2 such that $h_{1,n} = c_1 h_n$ and $h_{2,n} = c_2 h_n$.

(iv) (a) $K : [-1, 1] \to \mathbb{R}^+$ is bounded and continuous. (b) $\{K(\cdot/h) : h > 0\}$ is of VC type. (c) $\Gamma_2^{\pm}$ is positive definite.

(v) $V(Y|X = \cdot)$, $V(D|X = \cdot) \in \mathcal{C}^1([\underline{x}, \overline{x}] \setminus \{0\})$ with bounded derivatives in x and $0 < V(Y|X = 0^{\pm}) < \infty$

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

For $k \in \{1, 2\}$, define

$$\widehat{\mathbb{X}}_n'(0, k) = \frac{1}{\sqrt{c_k}}[\hat{\nu}_{\xi,n}^+(0, k) - \hat{\nu}_{\xi,n}^-(0, k)],$$

where the EMP is given by

$$\hat{\nu}_{\xi,n}^{\pm}(0, 1) = \sum_{i=1}^n \xi_i \frac{e_0'(\Gamma_2^{\pm})^{-1}[Y_i - \tilde{\mu}_{1,2}(0^{\pm}, 0)]r_2(\frac{X_i}{h_{1,n}})K(\frac{X_i}{h_{1,n}})}{\sqrt{nh_{1,n}}\hat{f}_X(0)}$$
$$\hat{\nu}_{\xi,n}^{\pm}(0, 2) = \sum_{i=1}^n \xi_i \frac{e_0'(\Gamma_2^{\pm})^{-1}[D_i - \tilde{\mu}_{2,2}(0^{\pm}, 0)]r_2(\frac{X_i}{h_{2,n}})K(\frac{X_i}{h_{2,n}})}{\sqrt{nh_{2,n}}\hat{f}_X(0)}$$

and $\tilde{\mu}_{k,2}(0^{\pm}, 0)$ is defined in Lemma 9. Our general result applied to the current case yields the following corollary.

Corollary 5 (Example: Fuzzy Mean RDD)**.** *Suppose that Assumption FMRD holds.*

(i) There exists $\sigma_{FMRD} > 0$ such that

$$\sqrt{nh_n}[\hat{\tau}_{FMRD} - \tau_{FMRD}] \rightsquigarrow N(0, \sigma_{FMRD}^2).$$

(ii) Furthermore, with probability approaching to one,

$$\frac{(\hat{\mu}_{2,2}(0^+, 0) - \hat{\mu}_{2,2}(0^-, 0))\widehat{\mathbb{X}}_n'(0, 1) - (\hat{\mu}_{1,2}(0^+, 0) - \hat{\mu}_{1,2}(0^-, 0))\widehat{\mathbb{X}}_n'(0, 2)}{(\hat{\mu}_{2,2}(0^+, 0) - \hat{\mu}_{2,2}(0^-, 0))^2} \overset{p}{\underset{\xi}{\rightsquigarrow}} N(0, \sigma_{FMRD}^2).$$

A proof is provided in Section A.6.2. Perhaps the most practically relevant application of this corollary is the test of the null hypothesis of treatment nullity:

$$H_0 : \tau_{FMRD} = 0.$$

To test this hypothesis, we can use $\sqrt{nh_n}\,|\hat{\tau}_{FMRD}|$ as the test statistic, and use

$$\left|\frac{(\hat{\mu}_{2,2}(0^+,0) - \hat{\mu}_{2,2}(0^-,0))\widehat{\mathbb{X}}'_n(0,1) - (\hat{\mu}_{1,2}(0^+,0) - \hat{\mu}_{1,2}(0^-,0))\widehat{\mathbb{X}}'_n(0,2)}{(\hat{\mu}_{2,2}(0^+,0) - \hat{\mu}_{2,2}(0^-,0))^2}\right|$$

to simulate its asymptotic distribution.

5.4 Example: Fuzzy Mean RKD

Consider Θ_1, Θ_2, Θ'_1, Θ'_2, Θ'', g_1, g_2, ϕ, ψ, and Υ defined in Section 3.3. Recall that we denote the local Wald estimand (3.1) with $v = 1$ in this setting by τ_{FMRK}. We also denote the analog estimator (4.3) with $v = 1$ in this setting by

$$\hat{\tau}_{FMRK} = \frac{\hat{\mu}^{(1)}_{1,2}(0^+,0) - \hat{\mu}^{(1)}_{1,2}(0^-,0)}{\hat{\mu}^{(1)}_{2,2}(0^+,0) - \hat{\mu}^{(1)}_{2,2}(0^-,0)}$$

For this application, we consider the following set of assumptions.

Assumption FMRK.

(i) (a) $\{(Y_i, X_i, D_i)\}_{i=1}^n$ are n independent copies of the random vector (Y, X, D) 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 $[\underline{x}, \overline{x}]$ that contains 0 in its interior and satisfies $0 < f_X(0) < \infty$.

(ii) (a) $E[|Y|^{2+\epsilon}|X = \cdot] < \infty$ and $E[|D|^{2+\epsilon}|X = \cdot] < \infty$ on $[\underline{x}, \overline{x}]\backslash\{0\}$ for some $\epsilon > 0$. (b) $\frac{\partial^j}{\partial x^j}E[Y|X = \cdot]$ and $\frac{\partial^j}{\partial x^j}E[D|X = \cdot]$ are Lipschitz on $[\underline{x}, 0)$ and $(0, \overline{x}]$ for $j = 0, 1, 2, 3$. (c) $\frac{\partial}{\partial x}E[D|X = 0^+] \neq \frac{\partial}{\partial x}E[D|X = 0^-]$.

(iii) The baseline bandwidth h_n satisfies $h_n \to 0$, $nh_n^3 \to \infty$, $nh_n^7 \to 0$. There exist constant c_1, c_2 such that $h_{1,n} = c_1 h_n$ and $h_{2,n} = c_2 h_n$.

(iv) (a) $K : [-1, 1] \to \mathbb{R}^+$ is bounded and continuous. (b) $\{K(\cdot/h) : h > 0\}$ is of VC type. (c) $\Gamma_2^{\pm}$ is positive definite.

(v) $V(Y|X=\cdot)$, $V(D|X=\cdot) \in \mathcal{C}^1([\underline{x},\overline{x}] \setminus \{0\})$ with bounded derivatives in x and $0 < V(Y|X = 0^{\pm}) < \infty$

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

For $k \in \{1,2\}$, define

$$\widehat{\mathbb{X}}'_n(0,k) = \frac{1}{\sqrt{c_k^3}}[\hat{\nu}^+_{\xi,n}(0,k) - \hat{\nu}^-_{\xi,n}(0,k)],$$

where the EMP is given by

$$\hat{\nu}^{\pm}_{\xi,n}(0,1) = \sum_{i=1}^n \xi_i \frac{e_1'(\Gamma_2^{\pm})^{-1}[Y_i - \tilde{\mu}_{1,2}(0^{\pm},0)]r_2(\frac{X_i}{h_{1,n}})K(\frac{X_i}{h_{1,n}})}{\sqrt{nh_{1,n}}\hat{f}_X(0)}$$
$$\hat{\nu}^{\pm}_{\xi,n}(0,2) = \sum_{i=1}^n \xi_i \frac{e_1'(\Gamma_2^{\pm})^{-1}[D_i - \tilde{\mu}_{2,2}(0^{\pm},0)]r_2(\frac{X_i}{h_{2,n}})K(\frac{X_i}{h_{2,n}})}{\sqrt{nh_{2,n}}\hat{f}_X(0)}$$

and $\tilde{\mu}_{k,2}(0^{\pm},0)$ is defined in Lemma 9. Our general result applied to the current case yields the following corollary.

Corollary 6 (Example: Fuzzy Mean RKD)**.** *Suppose that Assumption FMRK holds.*

(i) There exists $\sigma_{FMRK} > 0$ *such that*

$$\sqrt{nh_n^3}[\hat{\tau}_{FMRK} - \tau_{FMRK}] \rightsquigarrow N(0, \sigma^2_{FMRK}).$$

(ii) Furthermore, with probability approaching to one,

$$\frac{(\hat{\mu}^{(1)}_{2,2}(0^+,0) - \hat{\mu}^{(1)}_{2,2}(0^-,0))\widehat{\mathbb{X}}'_n(0,1) - (\hat{\mu}^{(1)}_{1,2}(0^+,0) - \hat{\mu}^{(1)}_{1,2}(0^-,0))\widehat{\mathbb{X}}'_n(0,2)}{(\hat{\mu}^{(1)}_{2,2}(0^+,0) - \hat{\mu}^{(1)}_{2,2}(0^-,0))^2} \underset{\xi}{\overset{p}{\rightsquigarrow}} N(0, \sigma^2_{FMRK}).$$

This corollary can be proved similarly to Corollary 5. Perhaps the most practically relevant application of this corollary is the test of the null hypothesis of treatment nullity:

$$H_0 : \tau_{FMRK} = 0.$$

To test this hypothesis, we can use $\sqrt{nh_n^3}\,|\hat{\tau}_{FMRK}|$ as the test statistic, and use

$$\left|\frac{(\hat{\mu}^{(1)}_{2,2}(0^+,0)-\hat{\mu}^{(1)}_{2,2}(0^-,0))\widehat{\mathbb{X}}'_n(0,1)-(\hat{\mu}^{(1)}_{1,2}(0^+,0)-\hat{\mu}^{(1)}_{1,2}(0^-,0))\widehat{\mathbb{X}}'_n(0,2)}{(\hat{\mu}^{(1)}_{2,2}(0^+,0)-\hat{\mu}^{(1)}_{2,2}(0^-,0))^2}\right|$$

to simulate its asymptotic distribution.

5.5 Example: Group Covariate and Test of Heterogeneous Treatment Effects

Consider Θ_1, Θ_2, Θ'_1, Θ'_2, Θ'', g_1, g_2, ϕ, ψ, and Υ defined in Section 3.10. Recall that we denote the local Wald estimand (3.1) with $v=1$ in this setting by τ_{GFMRD}. We also denote the analog estimator (4.3) with $v=1$ in this setting by $\hat{\tau}_{GFMRD}$.

$$\hat{\tau}_{GFMRD}(\theta'')=\frac{\hat{\mu}_{1,2}(0^+,\theta'')-\hat{\mu}_{1,2}(0^-,\theta'')}{\hat{\mu}_{2,2}(0^+,\theta'')-\hat{\mu}_{2,2}(0^-,\theta'')}$$

For this application, we consider the following set of assumptions.

Assumption GFMRD.

(i) (a) $\{(Y_i,X_i,D_i)\}_{i=1}^n$ are n independent copies of the random vector (Y,X,D) defined on a probability space $(\Omega^x,\mathcal{F}^x,\mathbb{P}^x)$ and supported in $\mathscr{Y}\times\mathscr{X}\times\mathscr{D}\subset\mathbb{R}^5$. (b) X has a density function f_X which is continuously differentiable on $[\underline{x},\overline{x}]$ that contains 0 in its interior, and $0<f_X(0)<\infty$.

(ii) (a) $E[|Y|^{*2+\epsilon}\cdot\mathbb{1}\{G=\theta''\}|X=\cdot]<\infty$ on $[\underline{x},\overline{x}]\backslash\{0\}$ for some $\epsilon>0$ for $\theta''\in\{1,...,K\}$. (b) $\frac{\partial^j}{\partial x^j}E[Y^*\cdot\mathbb{1}\{G=\theta''\}|X=\cdot]<C$ and $\frac{\partial^j}{\partial x^j}E[D^*\cdot\mathbb{1}\{G=\theta''\}|X=\cdot]$ are Lipschitz on $[\underline{x},0)$ and $(0,\overline{x}]$ for $j=0,1,2,3$. (c) $E[D^*\cdot\mathbb{1}\{G=\theta''\}|X=0^+]\neq E[D^*\cdot\mathbb{1}\{G=\theta''\}|X=0^-]$ for $\theta''\in\{1,...,K\}$. (d) $E[\mathbb{1}\{G=\theta''\}|X=\cdot]>0$ for all $\theta''\in\Theta''$

(iii) The baseline bandwidth h_n satisfies $h_n\to 0$, $nh_n^2\to\infty$, $nh_n^7\to 0$. There exist functions c_1, $c_2:\{1,...,K\}\to[\underline{c},\overline{c}]\subset(0,\infty)$ such that $h_{1,n}(\theta''_1)=c_1(\theta''_1)h_n$ and $h_{2,n}(\theta''_2)=c_2(\theta''_2)h_n$.

(iv) (a) $K:[-1,1]\to\mathbb{R}^+$ is bounded and continuous. (b) $\{K(\cdot/h):h>0\}$ is of VC type. (c) $\Gamma_2^{\pm}$ is positive definite.

(v) $V(Y^*\cdot\mathbb{1}\{G=\theta''\}|X=x)$, $V(D^*\cdot\mathbb{1}\{G=\theta''\}|X=x)\in\mathcal{C}^1([\underline{x},\overline{x}]\setminus\{0\})$ and $0<V(Y^*\cdot\mathbb{1}\{G=\theta''\}|X=0^{\pm})$ with derivatives in x all bounded on $[\underline{x},\overline{x}]\backslash\{0\}$. Furthermore, $0<V(D^*\cdot\mathbb{1}\{G=\theta''\}|X=$

$0^{\pm}$) for $\theta'' \in \{1, ..., K\}$.

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

For $k \in \{1, 2\}$ and $\theta'' \in \{1, ..., K\}$, define

$$\widehat{\mathbb{X}}_n'(\theta'', k) = \frac{1}{\sqrt{c_k}}[\hat{\nu}_{\xi,n}^{+}(\theta'', k) - \hat{\nu}_{\xi,n}^{-}(\theta'', k)],$$

where the EMP is given by

$$\hat{\nu}_{\xi,n}^{\pm}(0,1) = \sum_{i=1}^n \xi_i \frac{e_0'(\Gamma_2^{\pm})^{-1}[Y_i - \tilde{\mu}_{1,2}(0^{\pm}, \theta_1'')]r_2(\frac{X_i}{h_{1,n}(\theta_1'')})K(\frac{X_i}{h_{1,n}(\theta_1'')})}{\sqrt{nh_{1,n}(\theta_1'')}\hat{f}_X(0)}$$
$$\hat{\nu}_{\xi,n}^{\pm}(0,2) = \sum_{i=1}^n \xi_i \frac{e_0'(\Gamma_2^{\pm})^{-1}[D_i - \tilde{\mu}_{2,2}(0^{\pm}, \theta_2'')]r_2(\frac{X_i}{h_{2,n}(\theta_2'')})K(\frac{X_i}{h_{2,n}(\theta_2'')})}{\sqrt{nh_{2,n}(\theta_2'')}\hat{f}_X(0)}$$

and $\tilde{\mu}_{k,2}(0^{\pm}, \theta_k'')$ are defined in Lemma 9. Our general result applied to the current case yields the following corollary.

Corollary 7 (Example: Group Covariate)**.** *Suppose that Assumption GFMRD holds.*

(i) There exists a symmetric positive definite K-by-K matrix Σ_{GFMRD} such that

$$\sqrt{nh_n}[\hat{\tau}_{GFMRD} - \tau_{GFMRD}] \rightsquigarrow N(0, \Sigma_{GFMRD})$$

(ii) Furthermore, with probability approaching one,

$$\frac{(\hat{\mu}_{2,2}(0^+,\cdot) - \hat{\mu}_{2,2}(0^-,\cdot))\widehat{\mathbb{X}}_n'(\cdot,1) - (\hat{\mu}_{1,2}(0^+,\cdot) - \hat{\mu}_{1,2}(0^-,\cdot))\widehat{\mathbb{X}}_n'(\cdot,2)}{(\hat{\mu}_{2,2}(0^+,\cdot) - \hat{\mu}_{2,2}(0^-,\cdot))^2} \underset{\xi}{\overset{p}{\rightsquigarrow}} N(0, \Sigma_{GFMRD}).$$

A proof is provided in Section A.6.3. One of the practically most relevant applications of this corollary is the test of the null hypothesis of joint treatment nullity:

$$H_0 : \tau_{FMRD}(\theta'') = 0 \quad \text{for all } \theta'' \in \{1, ..., K\}.$$

To test this hypothesis, we can use $\max_{\theta'' \in \{1,...,K\}} \sqrt{nh_n}\, |\hat{\tau}_{GFMRD}(\theta'')|$ as the test statistic, and use

$$\max_{\theta'' \in \{1,...,K\}} \left| \frac{(\hat{\mu}_{2,2}(0^+,\theta'') - \hat{\mu}_{2,2}(0^-,\theta''))\widehat{\mathbb{X}}_n'(\theta'',1) - (\hat{\mu}_{1,2}(0^+,\theta'') - \hat{\mu}_{1,2}(0^-,\theta''))\widehat{\mathbb{X}}_n'(\theta'',2)}{(\hat{\mu}_{2,2}(0^+,\theta'') - \hat{\mu}_{2,2}(0^-,\theta''))^2} \right|$$

to simulate its asymptotic distribution.

Another one of the practically most relevant applications of the above corollary is the test of the null hypothesis of treatment homogeneity across the covariate-index groups:

$$H_0 : \tau_{FMRD}(\theta'') = \tau_{FMRD}(\theta''') \quad \text{for all } \theta'', \theta''' \in \{1, ..., K\}.$$

To test this hypothesis, we can use $\max_{\theta'' \in \{1,...,K\}} \sqrt{nh_n} \left| \hat{\tau}_{GFMRD}(\theta'') - K^{-1} \sum_{\theta'''=1}^{K} \hat{\tau}_{GFMRD}(\theta''') \right|$ as the test statistic, and use

$$\max_{\theta'' \in \{1,...,K\}} \left| \frac{(\hat{\mu}_{2,2}(0^+, \theta'') - \hat{\mu}_{2,2}(0^-, \theta''))\widehat{\mathbb{X}}'_n(\theta'', 1) - (\hat{\mu}_{1,2}(0^+, \theta'') - \hat{\mu}_{1,2}(0^-, \theta''))\widehat{\mathbb{X}}'_n(\theta'', 2)}{(\hat{\mu}_{2,2}(0^+, \theta'') - \hat{\mu}_{2,2}(0^-, \theta''))^2} \right.$$
$$\left. - \frac{1}{K} \sum_{\theta'''=1}^{K} \frac{(\hat{\mu}_{2,2}(0^+, \theta''') - \hat{\mu}_{2,2}(0^-, \theta'''))\widehat{\mathbb{X}}'_n(\theta''', 1) - (\hat{\mu}_{1,2}(0^+, \theta''') - \hat{\mu}_{1,2}(0^-, \theta'''))\widehat{\mathbb{X}}'_n(\theta''', 2)}{(\hat{\mu}_{2,2}(0^+, \theta''') - \hat{\mu}_{2,2}(0^-, \theta'''))^2} \right|$$

to simulate its asymptotic distribution.

5.6 Example: CDF Discontinuity and Test of Stochastic Dominance

Consider Θ_1, Θ_2, Θ'_1, Θ'_2, Θ'', g_1, g_2, ϕ, ψ, and Υ defined in Section 3.5. Recall that we denote the local Wald estimand (3.1) with $v = 1$ in this setting by τ_{SCRD}. We also denote the analog estimator (4.3) with $v = 1$ in this setting by

$$\hat{\tau}_{SCRD}(\theta'') = \hat{F}_{Y|X}(\theta''|0^+) - \hat{F}_{Y|X}(\theta''|0^-) = \hat{\mu}_{1,2}(0^+, \theta'') - \hat{\mu}_{1,2}(0^-, \theta'').$$

For this application, we consider the following set of assumptions.

Assumption SCRD.

(i) (a) $\{(Y_i, X_i)\}_{i=1}^n$ are n independent copies of the random vector (Y, 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 $[\underline{x}, \overline{x}]$ that contains 0 in its interior, and $0 < f_X(0) < \infty$.

(ii) $\frac{\partial^j}{\partial x^j} F_{Y|X}$ is Lipschitz in x on $\mathscr{Y} \times [\underline{x}, 0)$ and $\mathscr{Y} \times (0, \overline{x}]$ for $j = 0, 1, 2, 3$.

(iii) h_n satisfies $h_n \to 0$, $nh_n^7 \to 0$, and $nh_n^2 \to \infty$.

(iv) (a) $K : [-1, 1] \to \mathbb{R}^+$ is bounded and continuous. (b) $\{K(\cdot/h) : h > 0\}$ is of VC type. (c) $\Gamma_2^{\pm}$ is positive definite.

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

Let the EMP be given by

$$\hat{\nu}^{\pm}_{\xi,n}(\theta'') = \sum_{i=1}^n \xi_i \frac{e_0'(\Gamma_2^{\pm})^{-1}[\mathbb{1}\{Y_i \le \theta''\} - \tilde{F}_{Y|X}(\theta''|X_i)] r_2(\frac{X_i}{h_n}) K(\frac{X_i}{h_n}) \delta_i^{\pm}}{\sqrt{nh_n}\hat{f}_X(0)},$$

where $\tilde{F}_{Y|X}(\theta''|x) = \tilde{\mu}_{1,2}(x,\theta'')\mathbb{1}\{|x/h_n| \in [-1,1]\}$, and $\tilde{\mu}_{1,2}$ is defined in the statement of Lemma 9. Our general result applied to the current case yields the following corollary. A proof is provided in Section A.6.4.

Corollary 8 (Example: CDF Discontinuity)**.** *Suppose that Assumption SCRD holds.*

(i) There exists a zero mean Gaussian process $\mathbb{G}'_{SCRD} : \Omega^x \mapsto \ell^\infty(\mathscr{Y})$ *such that*

$$\sqrt{nh_n}[\hat{\tau}_{SCRD} - \tau_{SCRD}] \rightsquigarrow \mathbb{G}'_{SCRD}.$$

(ii) Furthermore, with probability approaching one,

$$\hat{\nu}^{+}_{\xi,n} - \hat{\nu}^{-}_{\xi,n} \overset{p}{\underset{\xi}{\rightsquigarrow}} \mathbb{G}'_{SCRD}.$$

One of the most common applications of weak convergence results for CDFs as stated in this corollary is the test of the stochastic dominance:

$$H_0 : \tau_{SCRD}(\theta'') \le 0 \qquad \forall \theta'' \in \Theta''.$$

See McFadden (1989). To test this hypothesis, we can use $\sup_{\theta'' \in \Theta''} \sqrt{nh_n} \max\{\hat{\tau}_{SCRD}(\theta''), 0\}$ as the test statistic, and use $\sup_{\theta'' \in \Theta''} \sqrt{nh_n} \max\{\hat{\nu}^{+}_{\xi,n}(\theta'') - \hat{\nu}^{-}_{\xi,n}(\theta''), 0\}$ to simulate its asymptotic distribution.

5.7 Example: Sharp Quantile RDD

Consider Θ_1, Θ_2, Θ_1', Θ_2', Θ'', g_1, g_2, ϕ, ψ, and Υ defined in Section 3.6. Recall that we denote the local Wald estimand (3.1) with $v = 1$ in this setting by τ_{SQRD}. We also denote the analog estimator (4.3) with $v = 1$ in this setting by

$$\hat{\tau}_{SQRD}(\theta'') = \hat{Q}_{Y|X}(\theta''|0^+) - \hat{Q}_{Y|X}(\theta''|0^-) = \phi(\hat{\mu}_{1,p}(0^+,\cdot))(\theta'') - \phi(\hat{\mu}_{1,p}(0^-,\cdot))(\theta'')$$

for $\theta'' \in [a, 1-a] \subset (0,1)$. For this application, we consider the following set of assumptions.

Assumption SQRD.

(i) (a) $\{(Y_i, X_i)\}_{i=1}^n$ are n independent copies of the random vector (Y, 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 $[\underline{x}, \overline{x}]$ that contains 0 in its interior, and $0 < f_X(0) < \infty$.

(ii) (a) $\frac{\partial^j}{\partial x^j} F_{Y|X}$ is Lipschitz in x on $\mathscr{Y}_1 \times [\underline{x}, 0)$ and $\mathscr{Y}_1 \times (0, \overline{x}]$ for $j = 0, 1, 2, 3$. (b) $f_{Y|X}(y|x)$ is Lipschitz in x and $0 < C < f_{Y|X} < C' < \infty$ on $\mathscr{Y}_1 \times [\underline{x}, 0)$ and $\mathscr{Y}_1 \times (0, \overline{x}]$.

(iii) h_n satisfies $h_n \to 0$, $nh_n^7 \to 0$, and $nh_n^2 \to \infty$.

(iv) (a) $K : [-1,1] \to \mathbb{R}^+$ is bounded and continuous. (b) $\{K(\cdot/h) : h > 0\}$ is of VC type. (c) $\Gamma_2^{\pm}$ is positive definite.

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

(vi) There exists $\hat{f}_{Y|X}(y|0^\pm)$ such that $\sup_{y \in \mathscr{Y}_1} |\hat{f}_{Y|X}(y|0^\pm) - f_{Y|X}(y|0^\pm)| = o_p^x(1)$.

We state (vi) as a high level assumption to accommodate a number of alternative estimators. An example and sufficient conditions for (vi) is given above Lemma 13 in section A.5.2 in the Mathematical Appendix.

Define

$$\widehat{\phi}'_{F_{Y|X}(\cdot|0^\pm)}(\hat{\nu}^{\pm}_{\xi,n})(\theta'') = -\frac{\hat{\nu}^{\pm}_{\xi,n}(\hat{Q}_{Y|X}(\theta''|0^\pm))}{\hat{f}_{Y|X}(\hat{Q}_{Y|X}(\theta''|0^\pm)|0^\pm)}$$
$$= -\sum_{i=1}^{n}\xi_i \frac{e_0'(\Gamma_2^+)^{-1}[\mathbb{1}\{Y_i \le \hat{Q}_{Y|X}(\theta''|0^\pm)\} - \tilde{F}_{Y|X}(\hat{Q}_{Y|X}(\theta''|0^\pm)|X_i)]r_2(\frac{X_i}{h_n})K(\frac{X_i}{h_n})}{\sqrt{nh_n}\hat{f}_X(0)\hat{f}_{Y|X}(\hat{Q}_{Y|X}(\theta''|0^\pm)|0^\pm)}$$

where $\tilde{F}_{Y|X}(y|x) = \tilde{\mu}_{1,2}(x,y)\mathbb{1}\{|x/h_n| \le 1\}$, and $\tilde{\mu}_{1,2}$ is defined in the statement of Lemma 9. Our general result applied to the current case yields the following corollary.

Corollary 9 (Example: Sharp Quantile RDD)**.** *Suppose that Assumption SQRD holds.*

(i) There exists a zero mean Gaussian process $\mathbb{G}'_{SQRD} : \Omega^x \mapsto \ell^\infty([a, 1-a])$ *such that*

$$\sqrt{nh_n}[\hat{\tau}_{SQRD} - \tau_{SQRD}] \rightsquigarrow \mathbb{G}'_{SQRD}.$$

(ii) Furthermore, with probability approaching one,

$$\widehat{\phi}'_{F_{Y|X}(\cdot|0+)}(\hat{\nu}^{+}_{\xi,n}) - \widehat{\phi}'_{F_{Y|X}(\cdot|0-)}(\hat{\nu}^{-}_{\xi,n}) \underset{\xi}{\overset{p}{\rightsquigarrow}} \mathbb{G}'_{SQRD}.$$

A proof is provided in Section A.6.5. One of the practically most relevant applications of this corollary is the test of the null hypothesis of uniform treatment nullity:

$$H_0 : \tau_{SQRD}(\theta'') = 0 \quad \text{for all } \theta'' \in \Theta'' = [a, 1-a].$$

See Koenker and Xiao (2002), Chernozhukov and Fernández-Val (2005), and Qu and Yoon (2015b). To test this hypothesis, we can use $\sup_{\theta'' \in [a,1-a]} \sqrt{nh_n}\,|\hat{\tau}_{SQRD}(\theta'')|$ as the test statistic, and use

$$\sup_{\theta'' \in [a,1-a]} \left| \widehat{\phi}'_{F_{Y|X}(\cdot|0+)}(\hat{\nu}^{+}_{\xi,n})(\theta'') - \widehat{\phi}'_{F_{Y|X}(\cdot|0-)}(\hat{\nu}^{-}_{\xi,n})(\theta'') \right|$$

to simulate its asymptotic distribution.

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

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

We again refer to the list of references in the previous paragraph. To test this hypothesis, we can use $\sup_{\theta'' \in [a,1-a]} \sqrt{nh_n} \left| \hat{\tau}_{SQRD}(\theta'') - (1-2a)^{-1} \int_{[a,1-a]} \hat{\tau}_{SQRD}(\theta''')d\theta''' \right|$ as the test statistic, and use

$$\sup_{\theta'' \in [a,1-a]} \Bigg| \widehat{\phi}'_{F_{Y|X}(\cdot|0^+)}(\hat{\nu}^+_{\xi,n})(\theta'') - \widehat{\phi}'_{F_{Y|X}(\cdot|0^-)}(\hat{\nu}^-_{\xi,n})(\theta'') - \frac{1}{1-2a} \int_{[a,1-a]} \left(\widehat{\phi}'_{F_{Y|X}(\cdot|0^+)}(\hat{\nu}^+_{\xi,n})(\theta''') - \widehat{\phi}'_{F_{Y|X}(\cdot|0^-)}(\hat{\nu}^-_{\xi,n})(\theta''') \right) d\theta''' \Bigg|$$

to simulate its asymptotic distribution.

Remark 2. *It may happen that $\hat{F}_{Y|X}(\cdot|0^{\pm})$ is not monotone increasing in finite sample. We may monotonize the estimated CDFs by re-arrangements following Chernozhukov, Fernandez-Val, Galichon (2010). This does not affect the asymptotic properties of the estimators, while allowing for inversion of the CDF estimators.*

5.8 Example: Sharp Quantile RKD

Consider Θ_1, Θ_2, Θ_1', Θ_2', Θ'', g_1, g_2, ϕ, ψ, and Υ defined in Section 3.9. Recall that the operator ϕ is defined by

$$\phi(F^{(1)}_{Y|X}(\cdot|x))(\theta'') = -\frac{F^{(1)}_{Y|X}(Q_{Y|X}(\theta''|x)|x)}{f_{Y|X}(Q_{Y|X}(\theta''|x)|x)} \tag{5.1}$$

where $F^{(1)}_{Y|X}(y|x) = \frac{\partial}{\partial x} F(y|x)$ and $Q_{Y|X}(\theta'|x) = \inf\{\theta \in \Theta_1 : F_{Y|X}(\theta|x) \geq \theta'\}$. Also recall that the local Wald estimand (3.1) with $v = 1$ in this setting is denoted by τ_{SQRK}. We denote the analog 'intermediate' estimator (4.3) with $v = 1$ in this setting by

$$\tilde{\tau}_{SQRK}(\theta'') = \frac{\phi(\hat{F}^{(1)}_{Y|X}(\cdot|0^+))(\theta'') - \phi(\hat{F}^{(1)}_{Y|X}(\cdot|0^-))(\theta'')}{b'(0^+) - b'(0^-)}$$

for $\theta'' \in [a, 1-a] \subset (0,1)$. This is intermediate because the operator ϕ contains unknowns, $f_{Y|X}(\cdot|x)$ and $Q_{Y|X}(\cdot|x)$. In practice, we need to also estimate this operator ϕ by replacing these unknowns by uniformly consistent estimators. Thus, a feasible analog estimator is denoted by

$$\hat{\tau}_{SQRK}(\theta'') := \frac{\widehat{\phi}(\hat{F}^{(1)}_{Y|X}(\cdot|0^+))(\theta'') - \widehat{\phi}(\hat{F}^{(1)}_{Y|X}(\cdot|0^-))(\theta'')}{b'(0^+) - b'(0^-)}$$

where

$$\widehat{\phi}(\hat{F}^{(1)}_{Y|X}(\cdot|0^{\pm}))(\theta'') = -\frac{\hat{F}^{(1)}_{Y|X}(\hat{Q}_{Y|X}(\theta''|0)|0^{\pm})}{\hat{f}_{Y|X}(\hat{Q}_{Y|X}(\theta''|0)|0)}$$

for $\hat{Q}_{Y|X}(\theta''|x) = \inf\{\theta \in \Theta_1 : \hat{F}_{Y|x}(\theta|x) \geq \theta''\}$ and $\widehat{\phi}(F^{(1)}_{Y|X}(\cdot|x))(\theta'') := -\frac{F^{(1)}_{Y|X}(\hat{Q}_{Y|X}(\theta''|x)|x)}{\hat{f}_{Y|X}(\hat{Q}_{Y|X}(\theta''|x)|x)}$. For this application, we consider the following set of assumptions.

Assumption SQRK.

(i) (a) $\{(Y_i, X_i)\}_{i=1}^n$ are n independent copies of the random vector (Y, 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 $[\underline{x}, \overline{x}]$ that contains 0 in its interior, and $0 < f_X(0) < \infty$.

(ii) (a) $\frac{\partial^j}{\partial x^j} F_{Y|X}$ is Lipschitz in x on $\mathscr{Y}_1 \times [\underline{x}, 0)$ and $\mathscr{Y}_1 \times (0, \overline{x}]$ for $j = 0, 1, 2, 3$. (b) $f_{Y|X}$ is Lipschitz in x and $0 < C < f_{Y|X}(y|x) < C' < \infty$ on $\mathscr{Y}_1 \times [\underline{x}, \overline{x}]$.

(iii) h_n satisfies $h_n \to 0$, $nh_n^7 \to 0$, and $nh_n^3 \to \infty$.

(iv) (a) $K : [-1, 1] \to \mathbb{R}^+$ is bounded and continuous. (b) $\{K(\cdot/h) : h > 0\}$ is of VC type. (c) $\Gamma_2^{\pm}$ is positive definite.

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

(v) b' is continuous on $[\underline{x}, \overline{x}]\backslash\{0\}$ and $\lim_{x\downarrow 0} b'(x) \neq \lim_{x\uparrow 0} b'(x)$.

(vi) $\sup_{y \in \mathscr{Y}_1} |\sqrt{nh_n^3}[\hat{f}_{Y|X}(y|0) - f_{Y|X}(y|0)]| \underset{x}{\overset{p}{\to}} 0$

We state part (vi) of this assumption at this high-level in order to accommodate a number of alternative estimators. Above Lemma 14 in Section A.5.2, we propose one particular such estimator which satisfies (vi).

We defined the operator ϕ for τ_{SQRK} with ϕ in (5.1) without a formal justification. Now that Assumption SQRK is stated, we can now provide the following lemma for a justification.

Lemma 2. *Suppose that Assumption SQRK (i) and (ii) hold. Then, we have*

$$\frac{\partial}{\partial x} Q_{Y|X}(\theta''|0^{\pm}) = -\frac{F^{(1)}_{Y|X}(Q_{Y|X}(\theta''|0)|0^{\pm})}{f_{Y|X}(Q_{Y|X}(\theta''|0)|0)} = \phi(F^{(1)}_{Y|X}(\cdot|0^{\pm})).$$

To state the result of this subsection, define the following objects.

$$\phi'_{F_{Y|X}(\cdot|0^\pm)}(\hat{\nu}^{\pm}_{\xi,n})(\theta'') = -\frac{\hat{\nu}^{\pm}_{\xi,n}(Q_{Y|X}(\theta''|0^\pm))}{f_{Y|X}(Q_{Y|X}(\theta''|0)|0)}$$
$$= -\sum_{i=1}^{n} \xi_i \frac{e_1'(\Gamma_2^{\pm})^{-1}[\mathbb{1}\{Y_i \leq Q_{Y|X}(\theta''|0)\} - \tilde{F}_{Y|X}(Q_{Y|X}(\theta''|0)|X_i)] r_2(\frac{X_i}{h_n})K(\frac{X_i}{h_n})\delta_i^{\pm}}{\sqrt{nh_n}\hat{f}_X(0) f_{Y|X}(Q_{Y|X}(\theta''|0)|0)}$$
$$\widehat{\phi}'_{F_{Y|X}(\cdot|0^\pm)}(\hat{\nu}^{\pm}_{\xi,n})(\theta'') = -\frac{\hat{\nu}^{\pm}_{\xi,n}(\hat{Q}_{Y|X}(\theta''|0^\pm))}{\hat{f}_{Y|X}(\hat{Q}_{Y|X}(\theta''|0)|0)}$$
$$= -\sum_{i=1}^{n} \xi_i \frac{e_1'(\Gamma_2^{\pm})^{-1}[\mathbb{1}\{Y_i \leq \hat{Q}_{Y|X}(\theta''|0)\} - \tilde{F}_{Y|X}(\hat{Q}_{Y|X}(\theta''|0)|X_i)] r_2(\frac{X_i}{h_n})K(\frac{X_i}{h_n})\delta_i^{\pm}}{\sqrt{nh_n}\hat{f}_X(0) \hat{f}_{Y|X}(\hat{Q}_{Y|X}(\theta''|0)|0)}$$

where $\tilde{F}(y|x) = \tilde{\mu}_{1,2}(x,y)\mathbb{1}\{|x/h_n| \leq 1\}$ with $\tilde{\mu}_{1,2}$ defined in the statement of Lemma 9. Our general result applied to the current case yields the following corollary.

Corollary 10 (Example: Sharp Quantile RKD)**.** *Suppose that Assumption SQRK holds.*

(i) There exists a zero mean Gaussian process $\mathbb{G}'_{SQRK} : \Omega^x \mapsto \ell^\infty([a, 1-a])$ *such that*

$$\sqrt{nh_n^3}[\tilde{\tau}_{SQRK} - \tau_{SQRK}] \rightsquigarrow \mathbb{G}'_{SQRK},$$

and thus

$$\sqrt{nh_n^3}[\hat{\tau}_{SQRK} - \tau_{SQRK}] \rightsquigarrow \mathbb{G}'_{SQRK}.$$

(ii) Furthermore, with probability approaching one,

$$\frac{\phi'_{F_{Y|X}(\cdot|0^+)}(\hat{\nu}^{+}_{\xi,n}) - \phi'_{F_{Y|X}(\cdot|0^-)}(\hat{\nu}^{-}_{\xi,n})}{b'(0^+) - b'(0^-)} \overset{p}{\underset{\xi}{\rightsquigarrow}} \mathbb{G}'_{SQRK},$$

and thus

$$\frac{\widehat{\phi}'_{F_{Y|X}(\cdot|0^+)}(\hat{\nu}^{+}_{\xi,n}) - \widehat{\phi}'_{F_{Y|X}(\cdot|0^-)}(\hat{\nu}^{-}_{\xi,n})}{b'(0^+) - b'(0^-)} \overset{p}{\underset{\xi}{\rightsquigarrow}} \mathbb{G}'_{SQRK}.$$

A proof is provided in Section A.6.6. One of the practically most relevant applications of this corollary is the test of the null hypothesis of uniform treatment nullity:

$$H_0 : \tau_{SQRK}(\theta'') = 0 \quad \text{for all } \theta'' \in \Theta'' = [a, 1-a].$$

To test this hypothesis, we can use $\sup_{\theta'' \in [a,1-a]} \sqrt{nh_n^3}\, |\hat{\tau}_{SQRK}(\theta'')|$ as the test statistic, and use

$$\sup_{\theta'' \in [a,1-a]} \left| \frac{\widehat{\phi}'_{F_{Y|X}(\cdot|0^+)}(\hat{\nu}^{+}_{\xi,n})(\theta'') - \widehat{\phi}'_{F_{Y|X}(\cdot|0^-)}(\hat{\nu}^{-}_{\xi,n})(\theta'')}{b'(0^+) - b'(0^-)} \right|$$

to simulate its asymptotic distribution.

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

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

To test this hypothesis, we can use $\sup_{\theta'' \in [a,1-a]} \sqrt{nh_n^3} \left| \hat{\tau}_{SQRK}(\theta'') - (1-2a)^{-1} \int_{[a,1-a]} \hat{\tau}_{SQRK}(\theta''') d\theta''' \right|$ as the test statistic, and use

$$\sup_{\theta'' \in [a,1-a]} \left| \frac{\widehat{\phi}'_{F_{Y|X}(\cdot|0^+)}(\hat{\nu}^+_{\xi,n})(\theta'') - \widehat{\phi}'_{F_{Y|X}(\cdot|0^-)}(\hat{\nu}^-_{\xi,n})(\theta'')}{b'(0^+) - b'(0^-)} \right.$$
$$\left. - \frac{1}{1-2a} \int_{[a,1-a]} \frac{\widehat{\phi}'_{F_{Y|X}(\cdot|0^+)}(\hat{\nu}^+_{\xi,n})(\theta''') - \widehat{\phi}'_{F_{Y|X}(\cdot|0^-)}(\hat{\nu}^-_{\xi,n})(\theta''')}{b'(0^+) - b'(0^-)} d\theta''' \right|$$

to simulate its asymptotic distribution.

5.9 Example: Fuzzy Quantile RDD

Consider Θ_1, Θ_2, Θ_1', Θ_2', Θ'', g_1, g_2, ϕ, ψ, and Υ defined in Section 3.7. Recall that we denote the local Wald estimand (3.1) with $v = 1$ in this setting by τ_{FQRD}. We also denote the analog estimator (4.3) with $v = 1$ in this setting by

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

where

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

By van der Vaart and Wellner (1996; Lemma 3.9.23) and van der Vaart (1998; Theorem 20.9), Υ is Hadamard differentiable at $(F_{Y^{\cdot}|C})$, and the Hadamard derivative is

$$\Upsilon'_W(g.)(\cdot) =: \Upsilon'_{F_{Y^{\cdot}|C}}(g.)(\cdot) = -\frac{g_1(Q_{Y^1|C}(\cdot))}{f_{Y^1|C}(Q_{Y^1|C}(\cdot))} + \frac{g_0(Q_{Y^0|C}(\cdot))}{f_{Y^0|C}(Q_{Y^0|C}(\cdot))}$$

tangential to $C(\mathscr{Y}_1 \times \mathscr{D})$ under the assumptions to be stated below. For this application, we consider the following set of assumptions.

Assumption FQRD.

(i) (a) $\{(Y_i^*, D_i^*, X_i)\}_{i=1}^n$ are n independent 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 $[\underline{x}, \overline{x}]$ that contains 0 in its interior, and $0 < f_X(0) < \infty$.

(ii) $(x, y, d) \mapsto \frac{\partial^j}{\partial x^j} E[\mathbb{1}\{Y_i^* \leq y, D_i^* = d\}|X_i = x]$ is Lipschitz in x on $[\underline{x}, 0) \times \Theta_1$ and $(0, \overline{x}] \times \Theta_1$ 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 $(0, \overline{x}] \times \Theta_2$ for $j = 0, 1, 2, 3$.

(iii) The baseline bandwidth 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 $h_{2,n} = c_2 h_n$.

(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) $|\mathbb{P}^x(D_i = 1|X_i = 0^+) - \mathbb{P}^x(D_i = 1|X_i = 0^-)| > 0$.

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

(vii) $F_{Y^1|C}, F_{Y^0|C} \in \mathcal{C}^1(\mathscr{Y}_1)$, and $f_{Y^1|C}$ and $f_{Y^0|C}$ are bounded away from 0 on $\mathscr{Y}_1$.

(viii) There exists $\hat{f}_{Y|XD}(y|0^{\pm}, d)$ such that $\sup_{(y,d) \in \mathscr{Y}_1 \times \mathscr{D}} |\hat{f}_{Y|XD}(y|0^{\pm}, d) - f_{Y|XD}(y|0^{\pm}, d)| = o_p^x(1)$.

We state part (viii) of this assumption at this high-level in order to accommodate a number of alternative estimators. In Lemma 15 above Section A.5.2, we propose one particular such estimator which satisfies (viii).

The next lemma, which follows from Theorem 1 (i), shows the weak convergence in probability of the estimators of the local conditional quantiles of the potential outcomes. A proof of the lemma is provided in Section A.5.4.

Lemma 3. *Suppose that Assumption FQRD holds. Then,* $\sup_{\mathscr{Y}_1} |\hat{Q}_{Y^d|C} - Q_{Y^d|C}| \underset{x \times \xi}{\overset{p}{\to}} 0$ *for* $d = 1, 0$.

This lemma will be used in proving the result of this subsection below. Define the EMP by

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

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

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

for each $(y,d_1,d_2,k)\in\mathbb{T}=\mathscr{Y}_1\times\mathscr{D}\times\mathscr{D}\times\{1,2\}$. An estimated Hadarmard derivative is given by

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

Our general result applied to the current case yields the following corollary.

Corollary 11 (Example: Fuzzy Quantile RDD)**.** *Suppose that Assumptions FQRD holds.*

(i) There exists a zero mean Gaussian process $\mathbb{G}'_{FQRD}:\Omega^x\mapsto\ell^\infty([a,1-a])$ *such that*

$$\sqrt{nh_n}[\hat{\tau}_{FQRD}-\tau_{FQRD}]\rightsquigarrow\mathbb{G}'_{FQRD}.$$

(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 Section A.6.7. One of the practically most relevant applications of this corollary is the test of the null hypothesis of uniform treatment nullity:

$$H_0:\tau_{FQRD}(\theta'')=0\quad\text{for all }\theta''\in\Theta''=[a,1-a].$$

To test this hypothesis, we can use $\sup_{\theta''\in[a,1-a]}\sqrt{nh_n}\,|\hat{\tau}_{FQRD}(\theta'')|$ as the test statistic, and use

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

to simulate its asymptotic distribution.

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

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

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

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

to simulate its asymptotic distribution.

Remark 3. *In practice, we may encounter the situation that for a fixed n, $\hat{f}_{Y^d|C}(y) < 0$ at some y. We may make use of a sequence $l_n > 0$, $l_n \to 0$ as $n \to \infty$ and define the estimator as $\tilde{f}_{Y^d|C}(y) = \max\{\hat{f}_{Y^d|C}(y), l_n\}$ to avoid this problem.*

5.10 Example: Fuzzy Quantile RKD

Consider Θ_1, Θ_2, Θ'_1, Θ'_2, Θ'', g_1, g_2, ϕ, ψ, and Υ defined in Section 3.8. Recall that we denote the local Wald estimand (3.1) with $v = 1$ in this setting by τ_{FQRK}. We also denote the analog estimator (4.3) with $v = 1$ in this setting by

$$\hat{\tau}_{FQRK}(\theta'') = \frac{\widehat{\phi}(\hat{F}^{(1)}_{Y|X}(\cdot|0^+))(\theta'') - \widehat{\phi}(\hat{F}^{(1)}_{Y|X}(\cdot|0^-))(\theta'')}{\hat{\mu}^{(1)}_{2,2}(0^+,0) - \hat{\mu}^{(1)}_{2,2}(0^-,0)},$$

where $\widehat{\phi}(\hat{F}^{(1)}_{Y|X}(\cdot|0^\pm))(\theta'') := F^{(1)}_{Y|X}(\hat{Q}_{Y|X}(\cdot|0)|0^\pm)/\hat{f}_{Y|X}(\hat{Q}_{Y|X}(\cdot|0)|0)$. We further define

$$\begin{aligned}
&\widehat{\phi}'_{F^{(1)}_{Y|X}(\cdot|0^\pm)}(\hat{\nu}^\pm_{\xi,n}(\cdot,1)/\sqrt{c_1^3})(\cdot) \\
&\qquad = -\sum_{i=1}^n \xi_i \frac{e_1'(\Gamma_2^\pm)^{-1}[\mathbb{1}\{Y_i \le \hat{Q}_{Y|X}(\cdot|0)\} - \tilde{F}_{Y|X}(\hat{Q}_{Y|X}(\cdot|0)|X_i)] r_2(\frac{X_i}{h_{1,n}}) K(\frac{X_i}{h_{1,n}}) \delta_i^\pm}{\sqrt{c_1^3 n h_{1,n}} \hat{f}_X(0) \hat{f}_{Y|X}(\hat{Q}_{Y|X}(\cdot|0)|0)}, \\
&\widehat{\psi}'_{\mu_2^{(1)}(0^\pm,0)}((\cdot,2)/\sqrt{c_2^3})(\cdot) = \sum_{i=1}^n \xi_i \frac{e_1'(\Gamma_2^\pm)^{-1}[D_i - \tilde{\mu}_2(X_i,0)] r_2(\frac{X_i}{h_{2,n}}) K(\frac{X_i}{h_{2,n}}) \delta_i^\pm}{\sqrt{c_2^3 n h_n} \hat{f}_X(0)}, \\
&\widehat{\mathbb{X}}'_n(\cdot,1) = \widehat{\phi}'_{F_{Y|X}(\cdot|0^+)}(\hat{\nu}^+_{\xi,n}(\cdot,1)/\sqrt{c_1^3})(\cdot) - \widehat{\phi}'_{F_{Y|X}(\cdot|0^-)}(\hat{\nu}^-_{\xi,n}(\cdot,1)/\sqrt{c_1^3})(\cdot), \qquad \text{and} \\
&\widehat{\mathbb{X}}'_n(\cdot,2) = \widehat{\psi}'_{\mu_2^{(1)}(0^+,0)}(\hat{\nu}^+_{\xi,n}(\cdot,2)/\sqrt{c_2^3})(\cdot) - \widehat{\psi}'_{\mu_2^{(1)}(0^-,0)}(\hat{\nu}^-_{\xi,n}(\cdot,2)/\sqrt{c_2^3})(\cdot),
\end{aligned}$$

where $\tilde{F}_{Y|X}(y|x) = \tilde{\mu}_{1,2}(x,y)\mathbb{1}\{|x/h_{1,n}| \leq 1\}$. For this application, we consider the following set of assumptions.

Assumption FQRK.

(i) (a) $\{(Y_i, D_i, X_i)\}_{i=1}^n$ are n independent 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 $[\underline{x}, \overline{x}]$ that contains 0 in its interior, and $0 < f_X(0) < \infty$.

(ii)(a) $\frac{\partial^j}{\partial x^j} F_{Y|X}$ is Lipschitz on $\mathscr{Y}_1 \times [\underline{x}, 0)$ and $\mathscr{Y}_1 \times (0, \overline{x}]$ in x for $j = 0, 1, 2, 3$. $\frac{\partial^j}{\partial x^j} E[D|X = \cdot]$ is Lipschitz continuous on $[\underline{x}, 0)$ and $(0, \overline{x}]$ in x for $j = 0, 1, 2, 3$. (b) $f_{Y|X}$ is Lipschitz in x and $0 < C < f_{Y|X}(y|x) < C' < \infty$ on $\mathscr{Y}_1 \times ([\underline{x}, \overline{x}] \setminus \{0\})$.

(iii) The baseline bandwidth h_n satisfies $h_n \to 0$, $nh_n^7 \to 0$, and $nh_n^3 \to \infty$, and there exist constants c_1, $c_2 \in (0, \infty)$ such that $h_{k,n} = c_k h_n$.

(iv) (a) $K : [-1, 1] \to \mathbb{R}^+$ is bounded and continuous. (b) $\{K(\cdot/h) : h > 0\}$ is of VC type. (c) $\Gamma_2^{\pm}$ is positive definite.

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

(v) $E[D|X = 0^+] \neq E[D|X = 0^-]$.

(vi) $\sup_{y \in \mathscr{Y}_1} |\sqrt{nh_n^3}[\hat{f}_{Y|X}(y|0) - f_{Y|X}(y|0)]| \underset{x}{\overset{p}{\to}} 0$

Our general result applied to the current case yields the following corollary.

Corollary 12 (Example: Fuzzy Quantile RKD)**.** *Suppose that Assumptions FQRK holds.*

(i) There exists a zero mean Gaussian process $\mathbb{G}'_{FQRK} : \Omega^x \mapsto \ell^\infty([a, 1-a])$ *such that*

$$\sqrt{nh_n^3}[\hat{\tau}_{FQRK} - \tau_{FQRK}] \rightsquigarrow \mathbb{G}'_{FQRK}$$

(ii) Furthermore, with probability approaching one,

$$\frac{[\hat{\mu}_{2,2}^{(1)}(0^+, 0) - \hat{\mu}_{2,2}^{(1)}(0^-, 0)]\widehat{\mathbb{X}}'_n(\cdot, 1) - [\widehat{\phi}(\hat{F}_{Y|X}^{(1)}(\cdot|0^+))(\cdot) - \widehat{\phi}(\hat{F}_{Y|X}^{(1)}(\cdot|0^-))(\cdot)]\widehat{\mathbb{X}}'_n(\cdot, 2)}{[\hat{\mu}_{2,2}^{(1)}(0^+, 0) - \hat{\mu}_{2,2}^{(1)}(0^-, 0)]^2} \underset{\xi}{\overset{p}{\rightsquigarrow}} \mathbb{G}'_{FQRK}(\cdot)$$

This corollary can be proved similarly to Corollary 10. One of the practically most relevant applications of this corollary is the test of the null hypothesis of uniform treatment nullity:

$$H_0 : \tau_{FQRK}(\theta'') = 0 \quad \text{for all } \theta'' \in \Theta'' = [a, 1-a].$$

To test this hypothesis, we can use $\sup_{\theta'' \in [a,1-a]} \sqrt{nh_n^3}\, |\hat{\tau}_{FQRK}(\theta'')|$ as the test statistic, and use

$$\sup_{\theta'' \in [a,1-a]} \left| \frac{[\hat{\mu}^{(1)}_{2,2}(0^+,0) - \hat{\mu}^{(1)}_{2,2}(0^-,0)]\widehat{\mathbb{X}}'_n(\theta'',1) - [\widehat{\phi}(\hat{F}^{(1)}_{Y|X}(\cdot|0^+))(\cdot) - \widehat{\phi}(\hat{F}^{(1)}_{Y|X}(\cdot|0^-))(\cdot)]\widehat{\mathbb{X}}'_n(\theta'',2)}{[\hat{\mu}^{(1)}_{2,2}(0^+,0) - \hat{\mu}^{(1)}_{2,2}(0^-,0)]^2} \right|$$

to simulate its asymptotic distribution.

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

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

To test this hypothesis, we can use $\sup_{\theta'' \in [a,1-a]} \sqrt{nh_n^3} \left| \hat{\tau}_{SQRK}(\theta'') - (1-2a)^{-1} \int_{[a,1-a]} \hat{\tau}_{SQRK}(\theta''')d\theta''' \right|$

as the test statistic, and use

$$\sup_{\theta'' \in [a,1-a]} \left| \frac{[\hat{\mu}^{(1)}_{2,2}(0^+,0) - \hat{\mu}^{(1)}_{2,2}(0^-,0)]\widehat{\mathbb{X}}'_n(\theta'',1) - [\widehat{\phi}(\hat{F}^{(1)}_{Y|X}(\cdot|0^+))(\cdot) - \widehat{\phi}(\hat{F}^{(1)}_{Y|X}(\cdot|0^-))(\cdot)]\widehat{\mathbb{X}}'_n(\theta'',2)}{[\hat{\mu}^{(1)}_{2,2}(0^+,0) - \hat{\mu}^{(1)}_{2,2}(0^-,0)]^2} \right.$$
$$\left. - \frac{1}{1-2a} \int_{[a,1-a]} \frac{[\hat{\mu}^{(1)}_{2,2}(0^+,0) - \hat{\mu}^{(1)}_{2,2}(0^-,0)]\widehat{\mathbb{X}}'_n(\theta''',1) - [\widehat{\phi}(\hat{F}^{(1)}_{Y|X}(\cdot|0^+))(\cdot) - \widehat{\phi}(\hat{F}^{(1)}_{Y|X}(\cdot|0^-))(\cdot)]\widehat{\mathbb{X}}'_n(\theta''',2)}{[\hat{\mu}^{(1)}_{2,2}(0^+,0) - \hat{\mu}^{(1)}_{2,2}(0^-,0)]^2} d\theta''' \right|$$

to simulate its asymptotic distribution.

6 Simulation Studies

We conduct simulation studies to demonstrate the robustness and unified applicability of our general multiplier bootstrap method. Each of the ten examples covered Sections 3.1–3.10 and Sections 5.1–5.10 are tested, except for the example of CDF discontinuity – this example is omitted because it is a less complicated form of the sharp Quantile RDD without CDF inversions. The concrete bootstrap procedures outlined in Section 5 are used in the respective subsections below. We follow the procedure outlined in Section B for choices of bandwidths in finite sample. Other details are discussed in each example subsection below.

6.1 Example: Sharp Mean RDD

Consider the case of sharp RDD presented in Sections 3.2 and 5.1. We generate an i.i.d. sample $\{(Y_i, D_i, X_i)\}_{i=1}^n$ through the following data generating process:

$$Y_i = \alpha_0 + \alpha_1 X_i + \alpha_2 X_i^2 + \beta_1 D_i + U_i,$$

$$D_i = \mathbb{1}\{X_i \geq 0\},$$

$$(X_i, U_i)' \sim N(0, \Sigma),$$

where $\alpha_0 = 1.00$, $\alpha_1 = 0.10$, $\alpha_2 = 0.01$, β_1 is to be varied across simulation sets, $\Sigma_{11} = \sigma_X^2 = 1.0^2$, $\Sigma_{22} = \sigma_U^2 = 1.0^2$, and $\Sigma_{12} = \rho_{XU} \cdot \sigma_X \cdot \sigma_U = 0.5 \cdot 1.0^2$. In this setup, we have

$$\text{Treatment Effect} = \beta_1.$$

We simulate the 95% test for the null hypothesis $H_0 : \tau_{SMRD} = 0$ of treatment nullity using the procedure described in the last paragraph of Section 5.1. Table 1 shows simulated acceptance probabilities based on 2,500 multiplier bootstrap replications for 2,500 Monte Carlo replications for each of the sample sizes $n = 1,000$, $2,000$, and $4,000$. The first column under $\beta_1 = 0.00$ shows that simulated acceptance probabilities are close to the designed nominal probability, 95%. The next four columns show that the acceptance probability decreases in β_1, and the rate of decrease is higher for the larger sample sizes. These results evidence the power as well as the size correctness.

6.2 Example: Sharp Mean RKD

Consider the case of sharp RKD presented in Sections 3.4 and 5.2. We generate an i.i.d. sample $\{(Y_i, D_i, X_i)\}_{i=1}^n$ through the following data generating process:

$$Y_i = \alpha_0 + \alpha_1 X_i + \alpha_2 X_i^2 + \beta_1 D_i + U_i,$$

$$D_i = X_i \cdot (2 \cdot \mathbb{1}\{X_i \geq 0\} - 1),$$

$$(X_i, U_i)' \sim N(0, \Sigma),$$

where $\alpha_0 = 1.00$, $\alpha_1 = 0.10$, $\alpha_2 = 0.01$, β_1 is to be varied across simulation sets, $\Sigma_{11} = \sigma_X^2 = 1.0^2$, $\Sigma_{22} = \sigma_U^2 = 1.0^2$, and $\Sigma_{12} = \rho_{XU} \cdot \sigma_X \cdot \sigma_U = 0.5 \cdot 1.0^2$. In this setup, we have

$$\text{Treatment Effect} = \beta_1.$$

We simulate the 95% test for the null hypothesis $H_0 : \tau_{SMRK} = 0$ of treatment nullity using the procedure described in the last paragraph of Section 5.2. Table 2 shows simulated acceptance probabilities based on 2,500 multiplier bootstrap replications for 2,500 Monte Carlo replications for each of the sample sizes $n = 1,000$, $2,000$, and $4,000$. The results, exhibiting the same qualitative features as those in the previous subsection, evidence the power as well as the size correctness.

6.3 Example: Fuzzy Mean RDD

Consider the case of fuzzy RDD presented in Sections 3.1 and 5.3. We generate an i.i.d. sample $\{(Y_i, D_i, X_i)\}_{i=1}^n$ through the following data generating process:

$$Y_i = \alpha_0 + \alpha_1 X_i + \alpha_2 X_i^2 + \beta_1 D_i + U_i,$$

$$D_i = \mathbb{1}\{2 \cdot \mathbb{1}\{X_i \geq 0\} - 1 \geq V_i\},$$

$$(X_i, U_i, V_i)' \sim N(0, \Sigma),$$

where $\alpha_0 = 1.00$, $\alpha_1 = 0.10$, $\alpha_2 = 0.01$, β_1 is to be varied across simulation sets, $\Sigma_{11} = \sigma_X^2 = 1.0^2$, $\Sigma_{22} = \sigma_U^2 = 1.0^2$, $\Sigma_{33} = \sigma_V^2 = 0.5^2$, $\Sigma_{12} = \rho_{XU} \cdot \sigma_X \cdot \sigma_U = 0.5 \cdot 1.0^2$, $\Sigma_{13} = \rho_{XV} \cdot \sigma_X \cdot \sigma_V = 0.0 \cdot 1.0 \cdot 0.5$, and $\Sigma_{23} = \rho_{UV} \cdot \sigma_U \cdot \sigma_V = 0.5 \cdot 1.0 \cdot 0.5$. In this setup, we have

$$\text{Treatment Effect} = \beta_1.$$

We simulate the 95% test for the null hypothesis $H_0 : \tau_{FMRD} = 0$ of treatment nullity using the procedure described in the last paragraph of Section 5.3. Table 3 shows simulated acceptance probabilities based on 2,500 multiplier bootstrap replications for 2,500 Monte Carlo replications for each of the sample sizes $n = 1,000$, $2,000$, and $4,000$. The results, exhibiting the same qualitative features as those in the previous subsections, evidence the power as well as the size correctness.

6.4 Example: Fuzzy Mean RKD

Consider the case of fuzzy RKD presented in Sections 3.3 and 5.4. We generate an i.i.d. sample $\{(Y_i, D_i, X_i)\}_{i=1}^n$ through the following data generating process:

$$Y_i = \alpha_0 + \alpha_1 X_i + \alpha_2 X_i^2 + \beta_1 D_i + U_i,$$

$$D_i = X_i \cdot (2 \cdot \mathbb{1}\{X_i \geq 0\} - 1) + V_i,$$

$$(X_i, U_i, V_i)' \sim N(0, \Sigma),$$

where $\alpha_0 = 1.00$, $\alpha_1 = 0.10$, $\alpha_2 = 0.01$, β_1 is to be varied across simulation sets, $\Sigma_{11} = \sigma_X^2 = 1.0^2$, $\Sigma_{22} = \sigma_U^2 = 1.0^2$, $\Sigma_{33} = \sigma_V^2 = 0.1^2$, $\Sigma_{12} = \rho_{XU} \cdot \sigma_X \cdot \sigma_U = 0.5 \cdot 1.0^2$, $\Sigma_{13} = \rho_{XV} \cdot \sigma_X \cdot \sigma_V = 0.0 \cdot 1.0 \cdot 0.1$, and $\Sigma_{23} = \rho_{UV} \cdot \sigma_U \cdot \sigma_V = 0.5 \cdot 1.0 \cdot 0.1$. In this setup, we have

$$\text{Treatment Effect} = \beta_1.$$

We simulate the 95% test for the null hypothesis $H_0 : \tau_{FMRK} = 0$ of treatment nullity using the procedure described in the last paragraph of Section 5.4. Table 4 shows simulated acceptance probabilities based on 2,500 multiplier bootstrap replications for 2,500 Monte Carlo replications for each of the sample sizes $n = 1,000$, $2,000$, and $4,000$. The results, exhibiting the same qualitative features as those in the previous subsections, evidence the power as well as the size correctness.

6.5 Example: Group Covariate and Test of Heterogeneous Treatment Effects

Consider the case of fuzzy RDD with heterogeneous groups presented in Sections 3.10 and 5.5. We generate an i.i.d. sample $\{(Y_i^*, D_i^*, G_i, X_i)\}_{i=1}^n$ through the following data generating process:

$$Y_i^* = \alpha_0 + \alpha_1 X_i + \alpha_2 X_i^2 + \beta_1 D_i^* \cdot \mathbb{1}\{G_i = 1\} + \beta_2 D_i^* \cdot \mathbb{1}\{G_i = 2\} + U_i,$$

$$D_i^* = \mathbb{1}\{2 \cdot \mathbb{1}\{X_i \geq 0\} - 1 \geq V_i\},$$

$$G_i \sim Bernoulli(\pi) + 1,$$

$$(X_i, U_i, V_i)' \sim N(0, \Sigma),$$

where $\alpha_0 = 1.00$, $\alpha_1 = 0.10$, $\alpha_2 = 0.01$, β_1 or β_2 is to be varied across simulation sets, $\pi = 0.5$, $\Sigma_{11} = \sigma_X^2 = 1.0^2$, $\Sigma_{22} = \sigma_U^2 = 1.0^2$, $\Sigma_{33} = \sigma_V^2 = 0.5^2$, $\Sigma_{12} = \rho_{XU} \cdot \sigma_X \cdot \sigma_U = 0.5 \cdot 1.0^2$, $\Sigma_{13} = \rho_{XV} \cdot \sigma_X \cdot \sigma_V = 0.0 \cdot 1.0 \cdot 0.5$, and $\Sigma_{23} = \rho_{UV} \cdot \sigma_U \cdot \sigma_V = 0.5 \cdot 1.0 \cdot 0.5$. In this setup, we have

$$\text{Treatment Effect} = \begin{cases} \beta_1 & \text{if } G_i = 1 \\ \\ \beta_2 & \text{if } G_i = 2 \end{cases}.$$

First, we simulate the 95% test for the null hypothesis $H_0 : \tau_{GFMRD}(1) = \tau_{GFMRD}(2) = 0$ of joint treatment nullity using the procedure described in the second to the last paragraph of Section 5.5. Part (A) of Table 5 shows simulated acceptance probabilities based on 2,500 multiplier bootstrap replications for 2,500 Monte Carlo replications for each of the sample sizes $n = 1,000$, $2,000$, and $4,000$. The first column under $\beta_1 = 0.00$ shows that simulated acceptance probabilities are close to the designed nominal probability, 95%. The next four columns show that the acceptance probability decreases in β_1, and the rate of decrease is higher for the larger sample sizes. These results evidence the power as well as the size correctness for the test of joint treatment nullity.

We next simulate the 95% test for the null hypothesis $H_0 : \tau_{GFMRD}(1) = \tau_{GFMRD}(2)$ of treatment homogeneity using the procedure described in the last paragraph of Section 5.5. Part (B) of Table 5 shows simulated acceptance probabilities based on 2,500 multiplier bootstrap replications for 2,500 Monte Carlo replications for each of the sample sizes $n = 1,000$, $2,000$, and $4,000$. The first column under $\beta_1 = 0.00$ shows that simulated acceptance probabilities are close to the designed nominal probability, 95%. The next four columns show that the acceptance probability decreases in β_1, and the rate of decrease is higher for the larger sample sizes. These results evidence the power as well as the size correctness for the test of treatment homogeneity.

6.6 Example: Sharp Quantile RDD

Consider the case of sharp quantile RDD presented in Sections 3.6 and 5.7. We generate an i.i.d. sample $\{(Y_i, D_i, X_i)\}_{i=1}^n$ through the following data generating process:

$$Y_i = \alpha_0 + \alpha_1 X_i + \alpha_2 X_i^2 + \beta_1 D_i + (\gamma_0 + \gamma_1 D_i) \cdot U_i,$$

$$D_i = \mathbb{1}\{X_i \geq 0\},$$

$$(X_i, U_i)' \sim N(0, \Sigma),$$

where $\alpha_0 = 1.00$, $\alpha_1 = 0.10$, $\alpha_2 = 0.01$, β_1 is to be varied across simulation sets, $\gamma_0 = 1$, γ_1 is to be varied across simulation sets, $\Sigma_{11} = \sigma_X^2 = 1.0^2$, $\Sigma_{22} = \sigma_U^2 = 1.0^2$, and $\Sigma_{12} = \rho_{XU} \cdot \sigma_X \cdot \sigma_U = 0.5 \cdot 1.0^2$. In this setup, we have

$$\theta\text{-th Conditional Quantile Treatment Effect at } x = 0 = \beta_1 + \gamma_1 F_{U|X}^{-1}(\theta|0).$$

We set $\Theta'' = [a, 1-a] = [0.20, 0.80]$ as the set of quantiles on which we conduct 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''. First, we simulate the 95% test for the null hypothesis $H_0 : \tau_{SQRD}(\theta'') = 0 \ \forall \theta'' \in [a, 1-a]$ of uniform treatment nullity using the procedure described in the second to the last paragraph of Section 5.7. Next, we simulate the 95% test for the null hypothesis $H_0 : \tau_{SQRD}(\theta'') = \tau_{SQRD}(\theta''') \ \forall \theta'', \theta''' \in [a, 1-a]$ of treatment homogeneity using the procedure described in the last paragraph of Section 5.7.

Table 6 show simulated acceptance probabilities based on 2,500 multiplier bootstrap replications for 2,500 Monte Carlo replications for each of the sample sizes $n = 1,000$, $2,000$, and $4,000$. Part (A) reports results for the test of uniform treatment nullity and part (B) shows results for the test of treatment homogeneity. The top panel (I) presents results across alternative values of $\beta_1 \in \{0.00, 0.25, 0.50, 0.75, 1.00\}$ while fixing $\gamma_1 = 0$. The bottom panel (II) presents 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 is 95%.

The first column of each part of the table shows that simulated acceptance probabilities are close to the designed nominal probability, 95%. The next four columns in parts (I) (A) and (II) (A) of the table show that the acceptance probability decreases in β_1 and γ_1, respectively, and the rate of decrease is higher for the larger sample sizes. These results evidence the power as well as the size correctness for the test of uniform treatment nullity. In part (I) (B), all the simulated acceptance probabilities are close to the designed nominal probability, 95%. This is consistent with the fact that β_1 does not contribute to treatment heterogeneity. On the other hand, part (II) (A) of the table shows that the acceptance probability decreases in γ_1, and the rate of decrease is higher for the larger sample sizes. These results evidence the power as well as the size correctness for the test of treatment homogeneity.

6.7 Example: Sharp Quantile RKD

Consider the case of sharp quantile RKD presented in Sections 3.9 and 5.8. We generate an i.i.d. sample $\{(Y_i, D_i, X_i)\}_{i=1}^n$ through the following data generating process:

$$
\begin{aligned}
Y_i &= \alpha_0 + \alpha_1 X_i + \alpha_2 X_i^2 + \beta_1 D_i + (\gamma_0 + \gamma_1 D_i) \cdot U_i, \\
D_i &= X_i \cdot (2 \cdot \mathbb{1}\{X_i \geq 0\} - 1), \\
(X_i, U_i)' &\sim N(0, \Sigma),
\end{aligned}
$$

where $\alpha_0 = 1.00$, $\alpha_1 = 0.10$, $\alpha_2 = 0.01$, β_1 is to be varied across simulation sets, $\gamma_0 = 1$, γ_1 is to be varied across simulation sets, $\Sigma_{11} = \sigma_X^2 = 1.0^2$, $\Sigma_{22} = \sigma_U^2 = 1.0^2$, and $\Sigma_{12} = \rho_{XU} \cdot \sigma_X \cdot \sigma_U = 0.5 \cdot 1.0^2$.

$$
\theta\text{-th Conditional Quantile Treatment Effect at } x = 0 = \beta_1 + \gamma_1 F_{U|X}^{-1}(\theta|0).
$$

We set $\Theta'' = [a, 1-a] = [0.20, 0.80]$ as the set of quantiles on which we conduct 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''. First, we simulate the 95% test for the null hypothesis $H_0 : \tau_{SQRK}(\theta'') = 0 \; \forall \theta'' \in [a, 1-a]$ of uniform treatment nullity using the procedure described in the second to the last paragraph of Section 5.8. Next, we simulate the 95% test for the null hypothesis $H_0 : \tau_{SQRK}(\theta'') = \tau_{SQRK}(\theta''') \ \forall \theta'', \theta''' \in [a, 1-a]$ of treatment homogeneity using the procedure described in the last paragraph of Section 5.8.

Table 7 show simulated acceptance probabilities based on 2,500 multiplier bootstrap replications for 2,500 Monte Carlo replications for each of the sample sizes $n = 1,000,\ 2,000,$ and $4,000$. Part (A) reports results for the test of uniform treatment nullity and part (B) shows results for the test of treatment homogeneity. The top panel (I) presents results across alternative values of $\beta_1 \in \{0.00, 0.25, 0.50, 0.75, 1.00\}$ while fixing $\gamma_1 = 0$. The bottom panel (II) presents 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 is 95%. The results, exhibiting the same qualitative features as those in the previous subsection, evidence the power as well as the size correctness for both of the tests of uniform treatment nullity and treatment homogeneity.

6.8 Example: Fuzzy Quantile RDD

Consider the case of fuzzy quantile RDD presented in Sections 3.7 and 5.9. We generate an i.i.d. sample $\{(Y_i, D_i, X_i)\}_{i=1}^n$ through the following data generating process:

$$Y_i = \alpha_0 + \alpha_1 X_i + \alpha_2 X_i^2 + \beta_1 D_i + (\gamma_0 + \gamma_1 D_i) \cdot U_i,$$

$$D_i = \mathbb{1}\{2 \cdot \mathbb{1}\{X_i \geq 0\} - 1 \geq V_i\},$$

$$(X_i, U_i, V_i)' \sim N(0, \Sigma),$$

where $\alpha_0 = 1.00$, $\alpha_1 = 0.10$, $\alpha_2 = 0.01$, β_1 is to be varied across simulation sets, $\gamma_0 = 1$, γ_1 is to be varied across simulation sets, $\Sigma_{11} = \sigma_X^2 = 1.0^2$, $\Sigma_{22} = \sigma_U^2 = 1.0^2$, $\Sigma_{33} = \sigma_V^2 = 0.5^2$, $\Sigma_{12} = \rho_{XU} \cdot \sigma_X \cdot \sigma_U = 0.5 \cdot 1.0^2$, $\Sigma_{13} = \rho_{XV} \cdot \sigma_X \cdot \sigma_V = 0.0 \cdot 1.0 \cdot 0.5$, and $\Sigma_{23} = \rho_{UV} \cdot \sigma_U \cdot \sigma_V = 0.5 \cdot 1.0 \cdot 0.5$.

$$\theta\text{-th Conditional Quantile Treatment Effect at } x = 0 = \beta_1 + \gamma_1 F_{U|X}^{-1}(\theta|0).$$

We set $\Theta'' = [a, 1-a] = [0.20, 0.80]$ as the set of quantiles on which we conduct 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''. 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 the second to the last paragraph of Section 5.9. Next, 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 the last paragraph of Section 5.9.

Table 8 show simulated acceptance probabilities based on 2,500 multiplier bootstrap replications for 2,500 Monte Carlo replications for each of the sample sizes $n = 1,000$, $2,000$, and $4,000$. Part (A) reports results for the test of uniform treatment nullity and part (B) shows results for the test of treatment homogeneity. The top panel (I) presents results across alternative values of $\beta_1 \in \{0.00, 0.25, 0.50, 0.75, 1.00\}$ while fixing $\gamma_1 = 0$. The bottom panel (II) presents 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 is 95%. The results, exhibiting the same qualitative features as those in the previous two subsections, evidence the power as well as the size correctness for both of the tests of uniform treatment nullity and treatment homogeneity.

6.9 Example: Fuzzy Quantile RKD

Consider the case of fuzzy quantile RKD presented in Sections 3.8 and 5.10. We generate an i.i.d. sample $\{(Y_i, D_i, X_i)\}_{i=1}^n$ through the following data generating process:

$$Y_i = \alpha_0 + \alpha_1 X_i + \alpha_2 X_i^2 + \beta_1 D_i + (\gamma_0 + \gamma_1 D_i) \cdot U_i,$$

$$D_i = X_i \cdot (2 \cdot \mathbb{1}\{X_i \geq 0\} - 1) + V_i,$$

$$(X_i, U_i, V_i)' \sim N(0, \Sigma),$$

where $\alpha_0 = 1.00$, $\alpha_1 = 0.10$, $\alpha_2 = 0.01$, β_1 is to be varied across simulation sets, $\gamma_0 = 1$, γ_1 is to be varied across simulation sets, $\Sigma_{11} = \sigma_X^2 = 1.0^2$, $\Sigma_{22} = \sigma_U^2 = 1.0^2$, $\Sigma_{33} = \sigma_V^2 = 0.1^2$,

$\Sigma_{12} = \rho_{XU} \cdot \sigma_X \cdot \sigma_U = 0.5 \cdot 1.0^2$, $\Sigma_{13} = \rho_{XV} \cdot \sigma_X \cdot \sigma_V = 0.0 \cdot 1.0 \cdot 0.1$, and $\Sigma_{23} = \rho_{UV} \cdot \sigma_U \cdot \sigma_V = 0.5 \cdot 1.0 \cdot 0.1$.

$$\theta\text{-th Conditional Quantile Treatment Effect at } x = 0 = \beta_1 + \gamma_1 F^{-1}_{U|X}(\theta|0).$$

We set $\Theta'' = [a, 1-a] = [0.20, 0.80]$ as the set of quantiles on which we conduct 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''. First, we simulate the 95% test for the null hypothesis $H_0 : \tau_{FQRK}(\theta'') = 0 \ \forall \theta'' \in [a, 1-a]$ of uniform treatment nullity using the procedure described in the second to the last paragraph of Section 5.10. Next, we simulate the 95% test for the null hypothesis $H_0 : \tau_{FQRK}(\theta'') = \tau_{FQRK}(\theta''') \ \forall \theta'', \theta''' \in [a, 1-a]$ of treatment homogeneity using the procedure described in the last paragraph of Section 5.10.

Table 9 show simulated acceptance probabilities based on 2,500 multiplier bootstrap replications for 2,500 Monte Carlo replications for each of the sample sizes $n = 1,000$, $2,000$, and $4,000$. Part (A) reports results for the test of uniform treatment nullity and part (B) shows results for the test of treatment homogeneity. The top panel (I) presents results across alternative values of $\beta_1 \in \{0.00, 0.25, 0.50, 0.75, 1.00\}$ while fixing $\gamma_1 = 0$. The bottom panel (II) presents 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 is 95%. The results, exhibiting the same qualitative features as those in the previous three subsections, evidence the power as well as the size correctness for both of the tests of uniform treatment nullity and treatment homogeneity.

7 Summary

The availability of bootstrap methods is desired for empirical research involving local Wald estimators due to the difficulty in computing asymptotic distributions for those estimators. In this paper, we develop a unified robust multiplier bootstrap method of inference which generically applies to a wide variety of local Wald estimators. Examples include the sharp mean RDD, the fuzzy mean RDD, the sharp mean RKD, the fuzzy mean RKD, the CDF discontinuity, the sharp quantile RDD, the fuzzy quantile RDD, the sharp quantile RKD, the fuzzy quantile RKD, and covariate-index versions of them. We achieve robustness against practically employed large bandwidths by incorporating a higher-order bias correction in our multiplier bootstrap framework. We achieve the generic applicability by taking advantage of the fact that all of the commonly used versions of local Wald estimators share a common basic form of BR. As such, the validity of the multiplier bootstrap is developed only for a single unified framework, with only the variable definitions and the Hadamard derivatives having to be changed across different versions of the local Wald estimators. We demonstrate applications of our unified theoretical results to ten examples including those listed above. The simulation studies demonstrate that the proposed method indeed performs well, robustly, and uniformly across the different examples. Finally, we are happy to share all the readers with the bootstrap code files which we wrote for each example – please contact the corresponding author.

A Mathematical Appendix

We use $a \lesssim b$ to denote the relation that there exists C, $0 < C < \infty$, such that $a \leq Cb$. We will borrow the following notations from empirical process theory. For more detail, we refer the reader to Giné and Nickl (2016), Kosorok (2003) and van der Vaart and Wellner (1996), to name a few references. 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. We further define the Uniform Entropy Integral $J(\delta, \mathscr{F}, F) = \sup_Q \int_0^\delta \sqrt{1 + \log N(\epsilon \left\|F\right\|_{Q,2}, \mathscr{F}, \left\|\cdot\right\|_{Q,2})} d\epsilon$, where $\left\|\cdot\right\|_{Q,2}$ 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 $N(\epsilon \left\|F\right\|, \mathscr{F}, \left\|\cdot\right\|_{Q,2}) \leq (\frac{A}{\epsilon})^V$, for $0 < \epsilon \leq 1$.

A.1 Proof of Lemma 1 (BR)

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 (2003), the solution to (4.2) can be computed explicitly as

$$\begin{aligned}\hat{\alpha}_{1+,p}(\theta_1) =&[\hat{\mu}_{1,p}(0^{\pm},\theta_1), \hat{\mu}_{1,p}^{(1)}(0^{\pm},\theta_1)h_{1,n}(\theta_1), ..., \hat{\mu}_{1,p}^{(p)}(0^{\pm},\theta_1)h_{1,n}^{p}(\theta_1)/p!]' \\ =&\Big[\frac{1}{nh_{1,n}(\theta_1)}\sum_{i=1}^{n}\delta_i^{+}K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})r_p'(\frac{X_i}{h_{1,n}(\theta_1)})\Big]^{-1}\cdot \\ &\Big[\frac{1}{nh_{1,n}(\theta_1)}\sum_{i=1}^{n}\delta_i^{+}K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})g_1(Y_i,\theta_1)\Big]\end{aligned}$$

For each data point $X_i > 0$, a mean value expansion by Assumption 1 (ii)(b) gives

$$\begin{aligned}g_1(Y_i,\theta_1) &= \mu_1(X_i,\theta) + \mathcal{E}_1(Y_i,D_i,X_i,\theta) \\ &= \mu_1(0,\theta) + \mu_1^{(1)}(0^{+},\theta_1)X_i + ... + \mu_1^{(p)}(0^{+},\theta_1)\frac{X_i^p}{p!} + \mu_1^{(p+1)}(x_{ni}^{*},\theta_1)\frac{X_i^{(p+1)}}{(p+1)!} + \mathcal{E}_1(Y_i,D_i,X_i,\theta) \\ &= r_p'(\frac{X_i}{h_{1,n}(\theta_1)})\alpha_{1+,p}(\theta_1) + \mu_1^{(p+1)}(x_{ni}^{*},\theta_1)h_{1,n}^{p+1}(\theta_1)\frac{(\frac{X_i}{h_{1,n}(\theta_1)})^{(p+1)}}{(p+1)!} + \mathcal{E}_1(Y_i,D_i,X_i,\theta)\end{aligned}$$

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

$$
\begin{aligned}
&\sqrt{nh_{1,n}(\theta_1)}\hat{\mu}_1^{(v)}(0^+,\theta_1)h_{1,n}^v(\theta_1)/v! \\
=&\sqrt{nh_{1,n}(\theta_1)}e_v'\Big[\frac{1}{nh_{1,n}(\theta_1)}\sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})r_p'(\frac{X_i}{h_{1,n}(\theta_1)})\Big]^{-1}\cdot \\
&\Big[\frac{1}{nh_{1,n}(\theta_1)}\sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})g_1(Y_i,t)\Big] \\
=&\sqrt{nh_{1,n}(\theta_1)}e_v'\Big[\frac{1}{nh_{1,n}(\theta_1)}\sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})r_p'(\frac{X_i}{h_{1,n}(\theta_1)})\Big]^{-1}\cdot \\
&\Big[\frac{1}{nh_{1,n}(\theta_1)}\sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})\Big(r_p'(\frac{X_i}{h_{1,n}(\theta_1)})\alpha_{1+,p}(\theta_1) \\
&+\mu_1^{(p+1)}(x_{ni}^*,\theta_1)h_{l,n}^{p+1}(\theta_1)\frac{(\frac{X_i}{h_{1,n}(\theta_1)})^{p+1}}{(p+1)!}+\mathcal{E}_1(Y_i,D_i,X_i,\theta)\Big)\Big] \\
=&\sqrt{nh_{1,n}(\theta_1)}e_v'\alpha_{1+,p}(\theta_p) \\
&+e_v'\Big[\frac{1}{nh_{1,n}(\theta_1)}\sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})r_p'(\frac{X_i}{h_{1,n}(\theta_1)})\Big]^{-1}\cdot \\
&\frac{1}{\sqrt{nh_{1,n}(\theta_1)}}\sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})\mu_1^{(p+1)}(x_{ni}^*,\theta_1)h_{l,n}^{p+1}(\theta_1)\frac{(\frac{X_i}{h_{1,n}(\theta_1)})^{p+1}}{(p+1)!} \\
&+e_v'\Big[\frac{1}{nh_{1,n}(\theta_1)}\sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})r_p'(\frac{X_i}{h_{1,n}(\theta_1)})\Big]^{-1}\cdot \\
&\frac{1}{\sqrt{nh_{1,n}(\theta_1)}}\sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})\mathcal{E}_1(Y_i,D_i,X_i,\theta) \\
=&\sqrt{nh_{1,n}(\theta_1)}\mu_1^{(v)}(0^+,\theta_1)h_{1,n}^v(\theta_1)/v!+(a)+(b)
\end{aligned}
$$

where

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

and

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

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

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

where the first term on the right hand side is stochastic while the second term is deterministic. First, regarding the deterministic part, we have

$$E[\frac{1}{nh_{1,n}(\theta_1)}\sum_{i=1}^{n}\delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})r_p'(\frac{X_i}{h_{1,n}(\theta_1)})]=f_X(0)\Gamma_p^+ + O(h_{1,n}(\theta_1))$$

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

$$\mathscr{F}_r=\{x\mapsto \mathbb{1}\{x\geq 0\}K(ax)(ax)^r\mathbb{1}\{ax\in[-1,1]\}:a>1/h_0\}$$

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

for each integer r such that $0\leq r\leq 2p$. By Lemma 8, 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 exists 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, Chetverikov and Kato (2014) to obtain

$$E\Big[\sup_{f \in \mathscr{F}_r} \Big| \frac{1}{\sqrt{n}} \sum_{i=1}^n \Big(f(X_i) - Ef\Big)\Big|\Big] \leq C\{J(1, \mathscr{F}_r, F) \|F\|_{P,2} + \frac{\|K\|_\infty J^2(1, \mathscr{F}_r, F)}{\delta^2 \sqrt{n}}\} < \infty$$

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

$$\begin{aligned}
&E\Big[\sup_{\theta_1 \in \Theta_1} \Big| \frac{1}{n h_{1,n}(\theta_1)} \sum_{i=1}^n \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)})(\frac{X_i}{h_{1,n}(\theta_1)})^s \\
&\qquad - \frac{1}{n h_{1,n}(\theta_1)} \sum_{i=1}^n E[\delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)})(\frac{X_i}{h_{1,n}(\theta_1)})^s)]\Big|\Big] \\
\leq& \frac{1}{\sqrt{n} h_{1,n}(\theta_1)} C\{J(1, \mathscr{F}_r, F) \|F\|_{P,2} + \frac{B J^2(1, \mathscr{F}_r, F)}{\delta^2 \sqrt{n}}\} \\
=& O(\frac{1}{\sqrt{n} h_n})
\end{aligned}$$

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 $r \in \{0, ..., 2p\}$. 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 supremum norm as in Step 1 to decompose

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

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

$$
\begin{aligned}
&E[\frac{1}{\sqrt{nh_{1,n}(\theta_1)}}\sum_{i=1}^{n}\delta_i^{+}K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})\mu_1^{(p+1)}(x_{ni}^{*},\theta_1)h_{1,n}^{p+1}(\theta_1)\frac{(\frac{X_i}{h_{1,n}(\theta_1)})^{p+1}}{(p+1)!}]\\
=&\frac{h_{1,n}^{p+1}(\theta_1)\Lambda_{p,p+1}^{+}}{\sqrt{nh_{1,n}(\theta_1)}(p+1)!}f_X(0)\mu_1^{(p+1)}(0^{+},\theta_1)+O(\frac{h_n^{p+2}}{\sqrt{nh_{1,n}}})\\
=&O(\sqrt{\frac{h_n^{2p+1}}{n}})
\end{aligned}
$$

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

$$
\begin{aligned}
&\Big|\frac{1}{\sqrt{nh_{1,n}(\theta_1)}}\sum_{i=1}^{n}\delta_i^{+}K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})h_{1,n}^{p+1}(\theta_1)\frac{(\frac{X_i}{h_{1,n}(\theta_1)})^{p+1}}{(p+1)!}[\mu_1^{(p+1)}(x_{ni}^{*},\theta_1)-\mu_1^{(p+1)}(0^{+},\theta_1)]\Big|\\
\lesssim& n\max_{1\leq i\leq n}\Big|\frac{1}{\sqrt{nh_{1,n}(\theta_1)}}\delta_i^{+}K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})h_{1,n}^{p+1}(\theta_1)\frac{(\frac{X_i}{h_{1,n}(\theta_1)})^{p+1}}{(p+1)!}[\mu_1^{(p+1)}(x_{ni}^{*},\theta_1)-\mu_1^{(p+1)}(0^{+},\theta_1)]\Big|\\
\lesssim&\frac{n\max_{1\leq i\leq n}|\mu_1^{(p+1)}(x_{ni}^{*},\theta_1)-\mu_1^{(p+1)}(0^{+},\theta_1)|h_{1,n}^{p+1}(\theta_1)}{\sqrt{nh_n}}\\
\leq&\frac{nLh_n^{p+2}}{\sqrt{nh_n}}=O_p^x(\sqrt{nh_n^{2p+3}})
\end{aligned}
$$

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), the third one is due to Assumption 1(ii). It is then sufficient to consider the asymptotic behavior of

$$
\frac{1}{\sqrt{nh_{1,n}(\theta_1)}}\sum_{i=1}^{n}\delta_i^{+}K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})\mu_1^{(p+1)}(0^{+},\theta_1)h_{1,n}^{p+1}(\theta_1)\frac{(\frac{X_i}{h_{1,n}(\theta_1)})^{p+1}}{(p+1)!}
$$

Note that $r_p(x/h)(x/h)^{p+1} = [(x/h)^{p+1},...,(x/h)^{2p+1}]'$. So we may let

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

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

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. Applying Lemma 7, it is also of VC type. We then apply Lemma 8 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 a similar argument as the one above with Theorem 5.2 of Chernozhukov, Chetverikov and Kato (2014) shows that for each $0\leq s\leq p$

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

To prove the uniform convergence of the part of (a) outside the inverse sign, it suffices to show that for each s

$$\sup_{f\in\mathscr{F}_s}|\frac{1}{\sqrt{nh_n}}\sum_{i=1}^n\Big(f(X_i)h_n^{p+1}-E[f(X_i)h_n^{p+1}]\Big)|\xrightarrow[x]{p}0$$

multiply both sides of equation (A.2) by $h_n^{p+1}/\sqrt{h_n}$ and apply 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

$$\begin{aligned}&\frac{1}{\sqrt{nh_{1,n}(\theta_1)}}\sum_{i=1}^n\delta_i^+K(\frac{X_i}{h_{1,n}(\theta_1)})r_p(\frac{X_i}{h_{1,n}(\theta_1)})\mu_1^{(p+1)}(x_{ni}^*,\theta_1)h_{1,n}^{p+1}(\theta_1)\frac{(X_i/h_{l,n}(\theta_1))^{p+1}}{(p+1)!}\\ =&O(\sqrt{\frac{h_n^{2p+1}}{n}})+O_p^x(\sqrt{nh_n^{2p+3}})+O_p^x(\sqrt{\frac{h_n^{2p+1}}{n}})\end{aligned}$$

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

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

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

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

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

Therefore, in the 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 show that for each coordinate $0 \le s \le p$

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

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

$$\begin{aligned}
&\sup_{\theta_1 \in \Theta_1} E[\frac{e_s'}{\sqrt{nh_{1,n}(\theta_1)}} \sum_{i=1}^{n} \delta_i^+ K(\frac{X_i}{h_{1,n}(\theta_1)}) r_p(\frac{X_i}{h_{1,n}(\theta_1)}) \mathcal{E}_1(Y_i, D_i, X_i, \theta)]^2 \\
=& \sup_{\theta_1 \in \Theta_1} E[\frac{1}{h_{1,n}(\theta_1)} \delta_i^+ E[\mathcal{E}_1^2(Y_i, D_i, X_i, \theta)|X_i] K^2(\frac{X_i}{h_{1,n}(\theta_1)}) e_s' r_p(\frac{X_i}{h_{1,n}(\theta_1)}) r_p'(\frac{X_i}{h_{1,n}(\theta_1)}) e_s] \\
=& \sup_{\theta_1 \in \Theta_1} \int_{\mathbb{R}_+} \sigma_{11}(\theta_1, \theta_1 | u h_{1,n}(\theta_1)) K^2(u) e_s' r_p(u) r_p'(u) e_s f_X(u h_{1,n}(\theta_1)) du \\
=& \sup_{\theta_1 \in \Theta_1} \int_{\mathbb{R}_+} \sigma_{11}(\theta_1, \theta_1 | 0^+) K^2(u) e_s' r_p(u) r_p'(u) e_s f_X(0^+) du + O_p^x(h_n) \\
\le& f_X(0) e_s' \Psi_p^+ e_s \sup_{\theta_1 \in \Theta_1} \sigma_{11}(\theta_1, \theta_1 | 0^+) + O_p^x(h_n) \\
\lesssim& f_X(0) e_s' \Psi_p^+ e_s + O_p^x(h_n)
\end{aligned}$$

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 doesn't depend on n, for each $M > 0$

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

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. □

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 4. *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

$$|E_{\xi|x} h(Y_n) - Eh(X)| \le |E_{\xi|x} h(Y_n) - E_{\xi|x} h(X_n)| + |E_{\xi|x} h(X_n) - Eh(X)|.$$

For the second term on the right-hand side, $|E_{\xi|x} h(X_n) - Eh(X)| \overset{p}{\underset{x}{\to}} 0$ 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

$$|E_{\xi|x} h(Y_n) - E_{\xi|x} h(X_n)| \le \varepsilon E_{\xi|x} \mathbb{1}\{d(X_n, Y_n) \le \varepsilon\} + 2E_{\xi|x} \mathbb{1}\{d(X_n, Y_n) > \varepsilon\}.$$

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 5. *Let $(A(\cdot), B(\cdot)) \in \ell^\infty(\Theta) \times \ell^\infty(\Theta)$, if $B(\theta) > C > 0$ 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 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 (Lemma 1; 2003). See also Theorem 10.6 of Pollard (1990).

Lemma 6 (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 $\{f_{ni}(t) : i = 1, ...n, t \in T\}$ be row independent and define $\nu_n(t) = \sum_{i=1}^n [f_{ni}(t) - Ef_{ni}(\cdot, t)]$. 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. *$H(s, t) = \lim_{n\to\infty} E\nu_n(s)\nu_n(t)$ 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* $\{z \in \ell^\infty(T) : z \text{ is uniformly } \rho\text{-continuous}\}$*, with covariance* $H(s,t)$*.*

Remark 4. *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* $\{f_{ni}(t) : i = 1, ...n, t \in T\}$ *is said to be separable if for every* $n \geq 1$*, there exists a countable subset* $T_n \subset T$ *such that*

$$\mathbb{P}^*\Big(\sup_{t\in T}\inf_{s\in T_n}\sum_{i=1}^{n}(f_{ni}(s) - f_{ni}(t))^2 > 0\Big) = 0$$

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 (2016) as a lemma below, which is used for establishing VC type of functions we encounter.

Lemma 7. *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, Chetverikov and Kato (2014) as the following lemma, which shows the stability of VC type classes under element-wise addition and multiplication.

Lemma 8. *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, Chetverikov and Kato (2014). The second one is proven in Corollary A.1 of Chernozhukov, Chetverikov and Kato (2014).

A.3 Proof of Theorem 1 (The Main Result)

Proof.

Part (i) For $(\theta, k) \in \mathbb{T}$, we define

$$\begin{aligned} f_{ni}(\theta,k) =& \frac{e_v'(\Gamma_p^+)^{-1} r_p(\frac{X_i}{h_{k,n}(\theta_k)})}{\sqrt{n h_{k,n}(\theta_k)} f_X(0)} \mathcal{E}_k(Y_i, D_i, X_i, \theta) K(\frac{X_i}{h_{k,n}(\theta_k)}) \delta_i^+, \\ =& \frac{a_0 + a_1(\frac{X_i}{c_k(\theta_k)h_n}) + \ldots + a_p(\frac{X_i}{c_k(\theta_k)h_n})^p}{\sqrt{n c_k(\theta_k) h_n} f_X(0)} \mathcal{E}_k(Y_i, D_i, X_i, \theta) K(\frac{X_i}{c_k(\theta_k)h_n}) \delta_i^+ \qquad \text{and} \\ \nu_n^+(\theta,k) =& \sum_{i=1}^n [f_{ni}(\theta,k) - E f_{ni}(\theta,k)]. \end{aligned}$$

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 followings from Assumption 1 (ii)(d),(iii) and (iv). We claim that it satisfies the conditions required by Lemma 6

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 8. Notice that for a fixed n, denote $\delta_x^+ = \mathbb{1}\{x > 0\}$, both

$$\{x \mapsto \frac{(a_0 + a_1(\frac{x}{c_k(\theta_k)h_n}) + \ldots + a_p(\frac{x}{c_k(\theta_k)h_n})^p)\mathbb{1}\{|x| \le \bar{c}h_n\}\delta_x^+}{\sqrt{n c_k(\theta_k) h_n} f_X(0)} : (\theta,k) \in \mathbb{T}\} \qquad \text{and}$$

$$\{x \mapsto K(\frac{x}{c_k(\theta_k)h_n}) : (\theta,k) \in \mathbb{T}\}$$

are of VC type with envelopes $\frac{C_1}{\sqrt{nh_n}} \mathbb{1}\{|x| \le \bar{c}h_n\}$ and $\mathbb{1}\{|x| \le \bar{c}h_n\} \|K\|_\infty$, respectively, under Assumptions 1(i),(iii) and (iv) and Lemma 7. By Lemma 8, 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

$$E[\nu_n^+(\theta,k)\nu_n^+(\vartheta,l)] = \sum_{i=1}^n Ef_{ni}(\theta,k)f_{ni}(\vartheta,l) - (\sum_{i=1}^n Ef_{ni}(\theta,k))(\sum_{i=1}^n Ef_{ni}(\vartheta,l)).$$

Under Assumptions 1(i)(b),(ii)(c),(iii),(iv)(a) we can write

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

$\Psi_p^+((\theta,k),(\vartheta,l))$ 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 exist 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

$$\sum_{i=1}^n E^*F_{ni}^2 = \sum_{i=1}^n EF_{ni}^2 \lesssim \int_{\mathscr{Y}\times\mathscr{D}\times\mathscr{X}} F_{\mathcal{E}}^2(y,d,uh_n)\mathbb{1}\{C_2u\in[-1,1]\}d\mathbb{P}^x(y,d,uh_n)+o(h_n)<\infty$$

as $n\to\infty$. This shows condition 3.

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

$$\begin{aligned}
&\lim_{n\to\infty}\sum_{i=1}^n E^*F_{ni}^2\mathbb{1}\{F_{ni}>\epsilon\}\\
=&\lim_{n\to\infty}\sum_{i=1}^n EF_{ni}^2\mathbb{1}\{F_{ni}>\epsilon\}\\
\lesssim&\lim_{n\to\infty}\int_{\mathscr{Y}\times\mathscr{D}\times\mathscr{X}} F_{\mathcal{E}}^2(y,d,uh_n)\mathbb{1}\{\frac{C_3}{\sqrt{nh_n}}F_{\mathcal{E}}(y,d,uh_n)\mathbb{1}\{C_2u\in[-1,1]\}>\epsilon\}d\mathbb{P}^x(y,d,uh_n)\\
=&\int_{\mathscr{Y}\times\mathscr{D}\times\mathscr{X}} F_{\mathcal{E}}^2(y,d,uh_n)\lim_{n\to\infty}\mathbb{1}\{\frac{C_3}{\sqrt{nh_n}}F_{\mathcal{E}}(y,d,uh_n)\mathbb{1}\{C_2u\in[-1,1]\}>\epsilon\}d\mathbb{P}^x(y,d,uh_n)\\
=&0
\end{aligned}$$

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 6, we have $\nu_n^+(\cdot)$ 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

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

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

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

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

Finally, by another application of the functional delta method, chain rule for the functional delta method (Lemma 3.9.3 of van der Vaart and Wellner(1996)), and Lemma 5 under Assumption 2(i) and (ii), we obtain

$$\begin{aligned}
&\sqrt{nh_n^{1+2v}}[\hat{\tau}(\cdot)-\tau(\cdot)] = \\
&\sqrt{nh_n^{1+2v}}[\Upsilon\Big(\frac{\phi(\hat{\mu}_{1,p}^{(v)}(0^+,\cdot))-\phi(\hat{\mu}_{1,p}^{(v)}(0^-,\cdot))}{\psi(\hat{\mu}_{2,p}^{(v)}(0^+,\cdot))-\psi(\hat{\mu}_{2,p}^{(v)}(0^-,\cdot))}\Big)(\cdot)-\Upsilon\Big(\frac{\phi(\mu_1^{(v)}(0^+,\cdot))-\phi(\mu_1^{(v)}(0^-,\cdot))}{\psi(\mu_2^{(v)}(0^+,\cdot))-\psi(\mu_2^{(v)}(0^-,\cdot))}\Big)(\cdot)] \\
\rightsquigarrow &\Upsilon'_W\Big(\frac{[\psi(\mu_2^{(v)}(0^+,\cdot))-\psi(\mu_2^{(v)}(0^-,\cdot))]\mathbb{G}'(\cdot,1)-[\phi(\mu_1^{(v)}(0^+,\cdot))-\phi(\mu_1^{(v)}(0^-,\cdot))]\mathbb{G}'(\cdot,2)}{(\psi(\mu_2^{(v)}(0^+,\cdot))-\psi(\mu_2^{(v)}(0^-,\cdot)))^2}\Big)(\cdot).
\end{aligned}$$

Part (ii) We introduce the following notations.

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

Applying Theorem 2 of Kosorok (2003) (same as Theorem 11.19 of Kosorok (2008)), we have $\nu_{\xi,n}^+ \underset{\xi}{\overset{p}{\rightsquigarrow}} B_H^+$. In order to apply Lemma 4, we need to show

$$\sup_{(\theta,k)\in\mathbb{T}} |\nu_{\xi,n}^+(\theta,k)-\hat{\nu}_{\xi,n}^+(\theta,k)| \underset{x\times\xi}{\overset{p}{\to}} 0.$$

We will show for $k=1$ and same argument applies to $k=2$ and thus the above holds. Note that under Assumption 3, $|\hat{f}_X(0) - f_X(0)| = o_p^{x\times\xi}(1)$. Thus under Assumption 1(i)(b),

$$\begin{aligned}
&\nu_{\xi,n}^+(\theta,1) - \hat{\nu}_{\xi,n}^+(\theta,1)\\
=&\frac{1}{f_X(0)\hat{f}_X(0)}\sum_{i=1}^n \xi_i \frac{e_v'(\Gamma_p^+)^{-1} r_p(\frac{X_i}{h_{1,n}(\theta_1)})}{\sqrt{nc_1(\theta_1)h_n}} K(\frac{X_i}{c_1(\theta_1)h_n})\delta_i^+ [\hat{\mathcal{E}}_1(Y_i,D_i,X_i,\theta) f_X(0) - \mathcal{E}_1(Y_i,D_i,X_i,\theta)\hat{f}_X(0)]\\
=&\frac{1}{f_X^2(0)+o_p^{x\times\xi}(1)}\sum_{i=1}^n \xi_i \frac{e_v'(\Gamma_p^+)^{-1} r_p(\frac{X_i}{h_{1,n}(\theta_1)})}{\sqrt{nc_1(\theta_1)h_n}} K(\frac{X_i}{c_1(\theta_1)h_n})\delta_i^+ [\hat{\mathcal{E}}_1(Y_i,D_i,X_i,\theta) f_X(0) - \mathcal{E}_1(Y_i,D_i,X_i,\theta)\hat{f}_X(0)]\\
=&\frac{1}{f_X^2(0)+o_p^{x\times\xi}(1)}\sum_{i=1}^n Z_i(\theta_1)[\hat{\mathcal{E}}_1(Y_i,D_i,X_i,\theta) f_X(0) - \mathcal{E}_1(Y_i,D_i,X_i,\theta)\hat{f}_X(0)]\\
=&\frac{1}{f_X^2(0)+o_p^{x\times\xi}(1)}\sum_{i=1}^n Z_i(\theta_1)[\hat{\mathcal{E}}_1(Y_i,D_i,X_i,\theta) f_X(0) - \mathcal{E}_1(Y_i,D_i,X_i,\theta) f_X(0)\\
&+\mathcal{E}_1(Y_i,D_i,X_i,\theta) f_X(0) - \mathcal{E}_1(Y_i,D_i,X_i,\theta)\hat{f}_X(0)]\\
=&\frac{1}{f_X^2(0)+o_p^{x\times\xi}(1)}\sum_{i=1}^n Z_i(\theta_1)[\hat{\mathcal{E}}_1(Y_i,D_i,X_i,\theta) - \mathcal{E}_1(Y_i,D_i,X_i,\theta)] f_X(0)\\
&+\frac{o_p^{x\times\xi}(1)}{f_X^2(0)+o_p^{x\times\xi}(1)}\sum_{i=1}^n Z_i(\theta_1)\mathcal{E}_1(Y_i,D_i,X_i,\theta)\\
=&(1)+(2)
\end{aligned}$$

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 Step 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 3 implies

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

uniformly on Θ. Therefore, we have

$$\sup_{(\theta,k)\in\mathbb{T}} |\nu^+_{\xi,n}(\theta,k) - \hat{\nu}^+_{\xi,n}(\theta,k)| \underset{x\times\xi}{\overset{p}{\to}} 0,$$

And thus we can apply Lemma 4 to conclude $\hat{\nu}^+_{\xi,n} \underset{\xi}{\overset{p}{\rightsquigarrow}} \mathbb{G}^+_H$. With similar arguments, we can derive that $\hat{\nu}^-_{\xi,n} \underset{\xi}{\overset{p}{\rightsquigarrow}} \mathbb{G}^-_H$.

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

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

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

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

This completes the proof. □

A.4 First Stage Estimators

To estimate MP, we replace $\mu_k(x,\theta)\mathbb{1}\{|x/h_{k,n}(\theta_k)| \le 1\}$ by its estimate $\tilde{\mu}_{k,p}(x,\theta)\mathbb{1}\{|x/h_{k,n}(\theta_k)| \le 1\}$ which is uniformly consistent across (x,θ). Lemma 9 below 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 $\hat{\mu}^{(v)}_{l,p}(0^\pm,\theta)$ for all $v \le p$. This auxiliary result will prove useful when we apply Theorem 1 to specific examples. In fact, we will prove a more general result that allows us to use any first r terms of our p-th order polynomial estimator for a t such that $0 \le t \le p$.

Lemma 9. *Fix an integer t such that $0 \le t \le p$. Suppose that Assumptions 1 and 2(i)-(iii) are satisfied. Let $\mathcal{E}_1(y,d,x,\theta) = g_1(y,\theta_1) - \mu_1(x,\theta_1)$, $\mathcal{E}_2(y,d,x,\theta) = g_2(d,\theta_2) - \mu_2(d,\theta_2)$, $\delta^+_x = \mathbb{1}\{x \ge 0\}$*

and $\delta_x^- = \mathbb{1}\{x \le 0\}$. *Define*

$$\tilde{\mu}_{1,t}(x,\theta_1) = r_t(x/h_{1,n}(\theta_1))'\hat{\alpha}_{1+,t}(\theta_1)\delta_x^+ + r_t(x/h_{1,n}(\theta_1))'\hat{\alpha}_{1-,t}(\theta_1)\delta_x^- \quad \text{and}$$

$$\tilde{\mu}_{2,t}(x,\theta_2) = r_t(x/h_{2,n}(\theta_2))'\hat{\alpha}_{2+,t}(\theta_2)\delta_x^+ + r_t(x/h_{2,n}(\theta_2))'\hat{\alpha}_{2-,t}(\theta_2)\delta_x^-$$

Then we have

$$\hat{\mathcal{E}}_1(y,d,x,\theta) = [g_1(y,\theta_1) - \tilde{\mu}_{1,t}(x,\theta_1)]\mathbb{1}\{|x/h_{1,n}(\theta_1)| \le 1\} \quad \text{and}$$

$$\hat{\mathcal{E}}_2(y,d,x,\theta) = [g_2(d,\theta_2) - \tilde{\mu}_{2,t}(x,\theta_2)]\mathbb{1}\{|x/h_{2,n}(\theta_2)| \le 1\}$$

are uniformly consistent for $\mathcal{E}_1(y,d,x,\theta)\mathbb{1}\{|x/h_{1,n}(\theta_1)| \le 1\}$ *and* $\mathcal{E}_2(y,d,x,\theta)\mathbb{1}\{|x/h_{2,n}(\theta_2)| \le 1\}$ *on* $[\underline{x},\overline{x}] \times \mathscr{Y} \times \mathscr{D}$, *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 $x > 0$,

$$\begin{aligned}
&\mu_1(x,\theta_1)\mathbb{1}\{|x/h_{1,n}(\theta_1)| \le 1\} = \\
&\Big(\mu_1(0^+,\theta_1) + \mu_1^{(1)}(0^+,\theta_1)x + ... + \mu_1^{(t)}(0^+,\theta_1)\frac{x^t}{t!} + \mu_1^{(t+1)}(x_{ni}^*,\theta_1)\frac{x^{(t+1)}}{(t+1)!}\Big)\mathbb{1}\{|x/c_1(\theta)h_n| \le 1\}.
\end{aligned}$$

By Corollary 1, $\hat{\mu}_{1,t}^{(v)}(0^\pm,\theta_1)$ is uniformly consistent for $\mu_1^{(v)}(0^\pm,\theta_1)$, $v = 0,1,...,t$. Thus,

$$\begin{aligned}
&[\hat{\mathcal{E}}_1(y,d,x,\theta) - \mathcal{E}_1(y,d,x,\theta)]\mathbb{1}\{|x/h_{1,n}(\theta)| \le 1\} \\
=&\Big(\hat{\mu}_{1,t}(0^+,\theta_1) + \hat{\mu}_{1,t}^{(1)}(0^+,\theta_1)x + ... + \hat{\mu}_{1,t}^{(t)}(0^+,\theta_1)\frac{x^t}{t!}\Big)\mathbb{1}\{|x/h_{1,n}(\theta)| \le 1\} - \\
&\Big(\mu_1(0^+,\theta_1) + \mu_1^{(1)}(0^+,\theta_1)x + ... + \mu_1^{(t)}(0^+,\theta_1)\frac{x^t}{t!} + \mu_1^{(t+1)}(x^*,\theta_1)\frac{x^{(t+1)}}{(t+1)!}\Big)\mathbb{1}\{|x/h_{1,n}(\theta_1)| \le 1\} \\
=&o_p^x(1) - \mu_1^{(t+1)}(x^*,\theta_1)\frac{x^{(t+1)}}{(t+1)!}\mathbb{1}\{|x| \le h_{1,n}(\theta)\} = o_p^x(1) + O(h_n),
\end{aligned}$$

where the last equality is by Assumption 1(iii) and by the uniform boundedness of $\mu^{(t+1)}$ under Assumption 1 (ii)(a). □

A.5 Auxiliary Lemmas for the Ten Examples

A.5.1 Lemmas Related to Asymptotic Equicontinuity

The following Lemma establish the relationship between convergence in probability in supremum norm and convergence in probability with respect to the semi-metric ρ induced by the limiting Gaussian process. We use this lemma in the corollary below to ensure that the asymptotically equicontinuous process, $\hat{\nu}^{\pm}_{\xi,n}$, evaluating $\hat{Q}_{Y|X}(\cdot|0^{\pm})$ can nicely approximate this process evaluating $Q_{Y|X}(\cdot|0^{\pm})$.

Lemma 10. *Suppose that Assumption SQRD holds. Define*

$$\rho(\hat{Q}_{Y|X}(\theta''|0^{\pm}), Q_{Y|X}(\theta''|0^{\pm})) = \lim_{n\to\infty} \big(\sum_{i=1}^{n} E|f_{ni}(\hat{Q}_{Y|X}(\theta''|0^{\pm})) - f_{ni}(Q_{Y|X}(\theta''|0^{\pm}))|^2\big)^{1/2},$$

where

$$f_{ni}(y) = \frac{e_0'(\Gamma_2^{\pm})^{-1} r_2(\frac{X_i}{h_n}) K(\frac{X_i}{h_n})[\mathbb{1}\{Y_i \leq y\} - F_{Y|X}(y|X_i)]\delta_i^{\pm}}{\sqrt{nh_n} f_X(0)}.$$

Then, $\left\|\hat{Q}_{Y|X}(\cdot|0^{\pm}) - Q_{Y|X}(\cdot|0^{\pm})\right\|_{[a,1-a]} \overset{p}{\underset{x}{\to}} 0$ *implies* $\sup_{\theta''\in[a,1-a]} \rho(\hat{Q}_{Y|X}(\theta''|0^{\pm}), Q_{Y|X}(\theta''|0^{\pm})) \overset{p}{\underset{x}{\to}} 0.$

Proof. We will show the claim for the + side only. The case of the − side can be similarly proved. Notice that by Law of Iterated Expectations and calculations under Assumption SQRD (i) (a), (i) (b), (iii), and (iv),

$$\begin{aligned}
&\rho^2(\hat{Q}_{Y|X}(\theta''|0^{+}), Q_{Y|X}(\theta''|0^{+}))\\
=&\lim_{n} \sum_{i=1}^{n} E[E[\Big(\frac{e_0'(\Gamma_2^{+})^{-1} r_2(\frac{X_i}{h_n}) K(\frac{X_i}{h_n})\delta_i^{+}}{\sqrt{nh_n} f_X(0)}\\
&[\mathbb{1}\{Y_i \leq \hat{Q}_{Y|X}(\theta''|0^{+})\} - F_{Y|X}(\hat{Q}_{Y|X}(\theta''|0^{+})|X_i) - \mathbb{1}\{Y_i \leq Q_{Y|X}(\theta''|0^{+})\} + F_{Y|X}(Q_{Y|X}(\theta''|0^{+})|X_i)]\Big)^2|X_i]]\\
=&\lim_{n} E[\frac{e_0'(\Gamma_2^{+})^{-1}\Psi^{+}(\Gamma_2^{+})^{-1}e_0}{f_X(0)} E[(\mathbb{1}\{Y_i \leq \hat{Q}_{Y|X}(\theta''|0^{+})\} - F_{Y|X}(\hat{Q}_{Y|X}(\theta''|0^{+})|X_i)\\
&\quad - \mathbb{1}\{Y_i \leq Q_{Y|X}(\theta''|0^{+})\} + F_{Y|X}(Q_{Y|X}(\theta''|0^{+})|X_i))^2|X_i]]
\end{aligned}$$

It then suffices to show that

$$E[(\mathbb{1}\{Y_i \leq \hat{Q}_{Y|X}(\theta''|0^{+})\} - F_{Y|X}(\hat{Q}_{Y|X}(\theta''|0^{+})|X_i) - \mathbb{1}\{Y_i \leq Q_{Y|X}(\theta''|0^{+})\} + F_{Y|X}(Q_{Y|X}(\theta''|0^{+})|X_i))^2|X_i] \overset{p}{\underset{x}{\to}} 0$$

uniformly in θ''. We write

$$
\begin{aligned}
&(\mathbb{1}\{Y_i \le \hat{Q}_{Y|X}(\theta''|0^+)\} - F_{Y|X}(\hat{Q}_{Y|X}(\theta''|0^+)|X_i) - \mathbb{1}\{Y_i \le Q_{Y|X}(\theta''|0^+)\} + F_{Y|X}(Q_{Y|X}(\theta''|0^+)|X_i))^2 \\
=&([\mathbb{1}\{Y_i \le \hat{Q}_{Y|X}(\theta''|0^+)\} - \mathbb{1}\{Y_i \le Q_{Y|X}(\theta''|0^+)\}] - [F_{Y|X}(\hat{Q}_{Y|X}(\theta''|0^+)|X_i) - F_{Y|X}(Q_{Y|X}(\theta''|0^+)|X_i)])^2 \\
=&[\mathbb{1}\{Y_i \le \hat{Q}_{Y|X}(\theta''|0^+)\} - \mathbb{1}\{Y_i \le Q_{Y|X}(\theta''|0^+)\}]^2 + [F_{Y|X}(\hat{Q}_{Y|X}(\theta''|0^+)|X_i) - F_{Y|X}(Q_{Y|X}(\theta''|0^+)|X_i)]^2 \\
&- 2[\mathbb{1}\{Y_i \le \hat{Q}_{Y|X}(\theta''|0^+)\} - \mathbb{1}\{Y_i \le Q_{Y|X}(\theta''|0^+)\}][F_{Y|X}(\hat{Q}_{Y|X}(\theta''|0^+)|X_i) - F_{Y|X}(Q_{Y|X}(\theta''|0^+)|X_i)] \\
=&(1) + (2) - (3).
\end{aligned}
$$

The conditional expectation of part (1) is

$$
\begin{aligned}
E[(1)|X_i] = F_{Y|X}(\hat{Q}_{Y|X}(\theta''|0^+)|X_i) + F_{Y|X}(Q_{Y|X}(\theta''|0^+)|X_i) - 2F_{Y|X}(\hat{Q}_{Y|X}(\theta''|0^+) \wedge Q_{Y|X}(\theta''|0^+)|X_i) \\
\xrightarrow[x]{p} 0
\end{aligned}
$$

uniformly by the uniform consistency $\left\|\hat{Q}_{Y|X}(\cdot|0^\pm) - Q_{Y|X}(\cdot|0^\pm)\right\|_{[a,1-a]} \xrightarrow[x]{p} 0$ and the continuous mapping theorem under Assumption SQRD (ii) (a). The uniform convergence in probability of other parts, (2) and (3), can be concluded similarly. □

Similar results hold in the cases of Sharp FQRK, Fuzzy FQRK and Fuzzy FQRD as the following

Lemma 11. *Suppose that Assumption SQRK or FQRK holds. Define*

$$
\rho(\hat{Q}_{Y|X}(\theta''|0), Q_{Y|X}(\theta''|0)) = \lim_{n\to\infty} \big(\sum_{i=1}^n E|f_{ni}(\hat{Q}_{Y|X}(\theta''|0)) - f_{ni}(Q_{Y|X}(\theta''|0))|^2\big)^{1/2},
$$

where

$$
f_{ni}(y) = \frac{e_1'(\Gamma_2^\pm)^{-1} r_2(\frac{X_i}{h_n}) K(\frac{X_i}{h_n})[\mathbb{1}\{Y_i \le y\} - F_{Y|X}(y|X_i)]\delta_i^\pm}{\sqrt{nh_n} f_X(0)}.
$$

Then, $\left\|\hat{Q}_{Y|X}(\cdot|0) - Q_{Y|X}(\cdot|0)\right\|_{[a,1-a]} \xrightarrow[x]{p} 0$ *implies* $\sup_{\theta''\in[a,1-a]} \rho(\hat{Q}_{Y|X}(\theta''|0), Q_{Y|X}(\theta''|0)) \xrightarrow[x]{p} 0.$

Lemma 12. *Suppose that Assumption FQRD holds. Define*

$$
\rho(\hat{Q}_{Y^d|C}(\theta''), Q_{Y^d|C}(\theta'')) = \lim_{n\to\infty} \big(\sum_{i=1}^n E|f_{ni}(\hat{Q}_{Y^d|C}(\theta''), d) - f_{ni}(Q_{Y^d|C}(\theta''), d)|^2\big)^{1/2},
$$

where

$$f_{ni}(y,d) = \frac{e_1'(\Gamma_2^{\pm})^{-1} r_2(\frac{X_i}{h_n}) K(\frac{X_i}{h_n})[\mathbb{1}\{Y_i^* \leq y, D_i^* = d\} - \mu_1(X_i,y,d)]\delta_i^{\pm}}{\sqrt{nh_n} f_X(0)}.$$

Then, $\left\| \hat{Q}_{Y^d|C}(\theta'') - Q_{Y^d|C}(\theta'') \right\|_{[a,1-a]\times\{0,1\}} \overset{p}{\underset{x}{\to}} 0$ *implies* $\sup_{(\theta'',d)\in[a,1-a]\times\{0,1\}} \rho(\hat{Q}_{Y^d|C}(\theta''), Q_{Y^d|C}(\theta'')) \overset{p}{\underset{x}{\to}} 0$.

Since the proofs are mostly identical to Lemma 10, we omit them.

A.5.2 Lemmas for Uniform Consistency of Estimators for Conditional Densities

Recall that in Section 5.7, the existence of an uniform consistency estimator is assumed in Assumption 5.7 (vi) An example of estimators $\hat{f}_{Y|X}(y, 0^{\pm})$ that satisfy Assumption 5.7 (vi) are

$$\hat{f}_{Y|X}(y|0^{\pm}) = \frac{\frac{1}{na_n^2}\sum_{i=1}^n K(\frac{Y_i - y}{a_n})K(\frac{X_i}{a_n})\delta_i^{\pm}}{\frac{1}{na_n}\sum_{i=1}^n K(\frac{X_i}{a_n})\delta_i^{\pm}} = \frac{\hat{g}_{YX}(y,0^{\pm})}{\hat{g}_X(0^{\pm})}$$

The following result gives the sufficient conditions for $f_{Y|X}(y|0^{\pm})$ to satisfy Assumption 5.7 (vi).

Lemma 13. *Assuming Assumption 5.7. In addition, let* f_{YX} *be continuously differentiable on* $\mathscr{Y}_1 \times [\underline{x}, 0)$*, and* $\mathscr{Y}_1 \times (0, \overline{x}]$ *with bounded partial derivatives and assume kernel* K *be symmetric. Let* a_n *be such that* $a_n \to 0$*,* $na_n \to \infty$*,* $\frac{na_n^2}{|\log a_n|} \to \infty$*,* $\frac{|\log a_n|}{\log\log a_n} \to \infty$*, and* $a_n^2 \leq ca_{2n}^2$ *for some* $c > 0$*. then* $\hat{f}_{Y|X}(y|0^{\pm})$ *such that* $\sup_{y\in\mathscr{Y}_1} |\hat{f}_{Y|X}(y|0^{\pm}) - f_{Y|X}(y|0^{\pm})| = o_p^x(1)$.

Proof. We will show only for 0^+. The other side follows similarly. Note the denominator $\hat{g}_X(0^+) = \frac{1}{2}f_X(0^+) + o_p^x(1)$ and $f_X(0)$ is bounded away from 0 under Assumption 1(ii).

Thus, it suffices to show $\sup_{y\in\mathscr{Y}_1} |\hat{g}_{YX}(y,0^+) - \frac{1}{2}f_{YX}(y,0^+)| = o_p^x(1)$. Note $|\hat{g}_{YX}(y,0^+) - \frac{1}{2}f_{YX}(y,0^+)| \leq |\hat{g}_{YX}(y,0^+) - E\hat{g}_{YX}(y,0^+)| + |E\hat{g}_{YX}(y,0^+) - \frac{1}{2}f_{YX}(y,0^+)|$. To control the stochastic part, Theorem 2.3 of Giné and Guillou (2002) suggests that

$$\sup_{y\in\mathscr{Y}_1} |\hat{g}_{YX}(y,0^+) - E\hat{g}_{YX}(y,0^+)| = O_{a.s.}^x(\sqrt{\frac{\log\frac{1}{a_n}}{na_n^2}}).$$

For the deterministic part, under the smoothness assumptions made on f_{YX}, a mean value expansion gives

$$
\begin{aligned}
&E[\hat{g}_{YX}(y,0^+)] \\
=&\int_{\mathbb{R}}\int_{\mathbb{R}_+} K(u)K(v)(f_{YX}(y,0^+) + \frac{\partial}{\partial y} f_{YX}(y^*,x^*)ua_n + \frac{\partial}{\partial x} f_{YX}(y^*,x^*)va_n)dudv \\
=&\frac{1}{2}f_{YX}(y,0^+) + O(a_n)
\end{aligned}
$$

uniformly on $\mathscr{Y}_1$, where (y^*,x^*) is a linear combination of $(y,0)$ and $(y+uh_n, vh_n)$. This concludes the proof. ☐

In Section 5.8, we assumed the existence of an estimator that satisfies Assumption 5.8. Define

$$
\hat{f}_{Y|X}(y|0) = \frac{\frac{1}{na_n^2}\sum_{i=1}^n K(\frac{Y_i-y}{a_n})K(\frac{X_i}{a_n})}{\frac{1}{na_n}\sum_{i=1}^n K(\frac{X_i}{a_n})} = \frac{\hat{f}_{YX}(y,0)}{\hat{f}_X(0)}
$$

The following lemma provides such estimator.

Lemma 14. *Assume Assumption SQRK (i) (a), (i) (b), and (ii) (b) hold. In addition, if the joint density $f_{YX}(y,x)$ is twice continuously differentiable on $\mathscr{Y}_2 \times [\underline{x},\overline{x}]$ with bounded partial derivatives for some open set $\mathscr{Y}_2$ such that $\mathscr{Y}_1 \subset \mathscr{Y}_2 \subset \mathscr{Y}$. Let K be a bounded kernel function such that $\{x \mapsto K(x-x'/a) : a > 0, x' \in \mathbb{R}\}$ forms a VC type class (e.g. Epanechnikov) and $\int_{\mathbb{R}} uK(u)du = 0$, $\int_{\mathbb{R}} u^2K^2(u)du < 0$. The bandwidth $a_n \to 0$ satisfies $\frac{na_n^2}{|\log a_n|} \to \infty$, $\frac{|\log a_n|}{\log\log a_n} \to \infty$, and $a_n^2 \le ca_{2n}^2$ for some $c > 0$, $\frac{h_n^3 \log\frac{1}{a_n}}{a_n^2} \to 0$ and $nh_n^3a_n^4 \to 0$. Then, $\sup_{y\in\mathscr{Y}_1}\sqrt{nh_n^3}|\hat{f}_{Y|X}(y|0) - f_{Y|X}(y|0)| \xrightarrow[x]{p} 0$. Thus, Assumption SQRK (vi) is satisfied.*

We remark that the condition $\frac{h_n^3 \log\frac{1}{a_n}}{a_n^2} \to 0$ is easily satisfied by the bandwidth selectors $h_{1,n}^{MSE} \propto n^{-1/5}$ and $h_{1,n}^{ROT} \propto n^{-1/4}$ proposed in Section B along with a Silverman's rule of thumb $a_n \propto n^{-1/6}$.

Proof. First note that the denominator is essentially $f_X(0) + o_p^x(1)$ and $f_X(0) > C > 0$. It then suffices to show that the numerator $\hat{f}_{YX}(y,0)$ satisfies $\sup_{y\in\mathscr{Y}_1}\sqrt{nh_n}|\hat{f}_{YX}(y,0) - f_{YX}(y,0)| = o_p^x(1)$. Write $|\hat{f}_{YX}(y,0) - f_{YX}(y,0)| \le |\hat{f}_{YX}(y,0) - E\hat{f}_{YX}(y,0)| + |E\hat{f}_{YX}(y,0) - f_{YX}(y,0)|$. To control the stochastic part, Theorem 2.3 of Giné and Guillou (2002) suggests that

$$\sup_{y\in\mathscr{Y}_1}|\hat{f}_{YX}(y,0)-E\hat{f}_{YX}(y,0)|=O^x_{a.s.}(\sqrt{\frac{\log\frac{1}{a_n}}{na_n^2}}).$$

For the deterministic part, a mean value expansion gives

$$\begin{aligned}
&E[\hat{f}_{YX}(y,0)]\\
=&\int_{\mathbb{R}}\int_{\mathbb{R}}K(u)K(v)(f_{YX}(y,0)+\frac{\partial}{\partial y}f_{YX}(y,0)ua_n+\frac{\partial}{\partial x}f_{YX}(y,0)va_n\\
&+\frac{\partial}{\partial y}\frac{\partial}{\partial y}f_{YX}(y^*,x^*)uva_n^2+\frac{1}{2}\frac{\partial^2}{\partial y^2}f_{YX}(y^*,x^*)u^2a_n^2+\frac{1}{2}\frac{\partial^2}{\partial x^2}f_{YX}(y^*,x^*)v^2a_n^2)dudv\\
=&f_{YX}(y,0)+O(a_n^2)
\end{aligned}$$

uniformly on $\mathscr{Y}_1$, where (y^*,x^*) is a linear combination of $(y,0)$ and $(y+uh_n,vh_n)$. Thus, the conclusion follows from $\frac{\log\frac{1}{a_n}h_n^3}{a_n^2}\to 0$ and $nh_n^3a_n^4\to 0$. □

In Section 5.9, the densities in the denominator can be estimated in the following manner. Define

$$\hat{f}_{Y^d|C}(y)=\frac{\hat{f}_{Y|XD}(y|0^+,d)\hat{\mu}_{2,2}(0^+,d)-\hat{f}_{Y|XD}(y|0^-,d)\hat{\mu}_{2,2}(0^-,d)}{\hat{\mu}_{2,2}(0^+,d)-\hat{\mu}_{2,2}(0^-,d)},$$

where $\hat{f}_{Y|XD}(y|0^\pm,1)=\frac{\frac{1}{na_n^2}\sum_{i=1}^nK(\frac{X_i}{a_n})K(\frac{Y_i-y}{a_n})D_i\delta_i^\pm}{\frac{1}{na_n}\sum_{i=1}^nK(\frac{X_i}{a_n})D_i\delta_i^\pm}$, $\hat{f}_{Y|XD}(y|0^\pm,0)=\frac{\frac{1}{na_n^2}\sum_{i=1}^nK(\frac{X_i}{a_n})K(\frac{Y_i-y}{a_n})(1-D_i)\delta_i^\pm}{\frac{1}{na_n}\sum_{i=1}^nK(\frac{X_i}{a_n})(1-D_i)\delta_i^\pm}$, with bandwidths a_n. The following lemma shows uniform consistency of these estimators.

Lemma 15. *Suppose that Assumption FQRD holds, in addition, for each $d=0$, 1, let $f_{YX|D}(\cdot,\cdot|d)$ be continuously differentiable on $\mathscr{Y}_1\times[\underline{x},0)$, and $\mathscr{Y}_1\times(0,\overline{x}]$ with bounded partial derivatives and assume kernel K be symmetric. Let a_n be such that $a_n\to 0$, $na_n\to\infty$, $\frac{na_n^2}{|\log a_n|}\to\infty$, $\frac{|\log a_n|}{\log\log a_n}\to\infty$, and $a_n^2\le ca_{2n}^2$ for some $c>0$. Then, $\sup_{y\in\mathcal{Y}}|\hat{f}_{Y^0|C}(y)-f_{Y^0|C}(y)|=o_p^x(1)$ and $\sup_{y\in\mathcal{Y}}|\hat{f}_{Y^1|C}(y)-f_{Y^1|C}(y)|=o_p^x(1)$.*

Proof. Note that

$$f_{Y^1|C}(y)=\frac{\partial}{\partial y}F_{Y^1|C}(y)=\frac{\frac{\partial}{\partial y}\mu_1(0^+,(y,1))-\frac{\partial}{\partial y}\mu_1(0^-,(y,1))}{\mu_2(0^+,1)-\mu_2(0^-,1)}$$

The numerator can be estimated consistently with the local quadratic regression defined above. For the denominator, we have

$$\begin{aligned}
\frac{\partial}{\partial y}\mu_1(0^+,(y,1)) =& \frac{\partial}{\partial y}E[\mathbb{1}\{Y_i \leq y\}\mathbb{1}\{D_i = 1\}|X_i = 0^+] \\
=& \frac{\partial}{\partial y}\Big(E[\mathbb{1}\{Y_i \leq y\}|X_i = 0^+, D_i = 1]\mathbb{P}^x(D_i = 1|X_i = 0^+) + 0\Big) \\
=& \frac{\partial}{\partial y}F_{Y^1|XD}(y|0^+,1)\mu_2(0^+,1) \\
=& f_{Y^1|XD}(y|0^+,1)\mu_2(0^+,1)
\end{aligned}$$

Uniform consistency of $\hat{f}_{Y^1|X,D}(y|0^\pm,1)$ can be shown by applying Theorem 2.3 of Giné and Guillou (2002), as in Lemma 13. Also $\hat{\mu}_{2,p}(0^\pm)$ is uniformly consistent by Corollary 1. Corresponding result for $\hat{f}_{Y^0|C}(y)$ can be shown similarly. □

A.5.3 Proof of Lemma 2

Proof. The Hadamard differentiability of the left-inverse operator and a mean value expansion give

$$\begin{aligned}
\frac{\partial}{\partial x}Q_{Y|X}(\theta''|0^+) &= \lim_{\delta\downarrow 0}\frac{Q_{Y|X}(\theta''|\delta) - Q_{Y|X}(\theta''|0^+)}{\delta} \\
&= \lim_{\delta\downarrow 0}\frac{\Phi\Big(F_{Y|X}(\cdot|\delta)\Big)(\theta'') - \Phi\Big(F_{Y|X}(\cdot|0^+)\Big)(\theta'')}{\delta} \\
&= \lim_{\delta\downarrow 0}\frac{\Phi\Big((F_{Y|X}(\cdot|0^+) + \delta\frac{\partial}{\partial x}F_{Y|X}(\cdot|x^*)\Big)(\theta'') - \Phi\Big(F_{Y|X}(\cdot|0^+)\Big)(\theta'')}{\delta} \\
&= \Phi'_{F_{Y|X}(\cdot|0)}\Big(\frac{\partial}{\partial x}F_{Y|X}(\cdot|0^+)\Big)(\theta'') \\
&= -\frac{F^{(1)}_{Y|X}(Q_{Y|X}(\theta''|0)|0^+)}{f_{Y|X}(Q_{Y|X}(\theta''|0)|0)} \\
&= \phi(F^{(1)}_{Y|X}(\cdot|0^+))(\theta'')
\end{aligned}$$

by the definition of Hadamard Derivative and then Lemma 3.9.23 (i) of van der Vaart and Wellner (1996). □

A.5.4 Proof of Lemma 3

Proof. Write $\Phi(F_{Y^d|C})(\theta'') := \inf\{y \in \mathscr{Y}_1 : F_{Y^d|C}(y) \geq \theta''\}$ for $\theta'' \in [a, 1-a]$. The weak convergence of $\nu_n^{\pm}$ from Theorem 1 (i), Lemma 3.9.23 (i) of van der Vaart and Wellner (1996), and the functional delta method yield

$$\sqrt{nh_n}[\hat{Q}_{Y^d|C}(\cdot) - Q_{Y^d|C}(\cdot)] = \sqrt{nh_n}[\Phi(\hat{F}_{Y^d|C})(\cdot) - \Phi(F_{Y^d|C})(\cdot)] = O_p^x(1)$$

The result follows from Assumption FQRD (vi) and Slutsky's lemma. □

A.6 Proofs of Corollaries for the Ten Examples

We use $\mathbb{G}$ and $\mathbb{G}_{\pm}$ to denote generic zero mean Gaussian processes that appear in some intermediate steps in the proofs without having to specify their covariance structure. They differ across different examples, but are fixed within each one.

A.6.1 Proof of Corollary 3

Proof. It suffices to show that Assumption SMRD implies Assumptions 1, 2, and 3. Most of these implications are direct. For Assumption 1 (ii)(a), note that $\{\mu_1(x,0) : [\underline{x}, \overline{x}] \mapsto \mathbb{R}\}$ is a singleton, and therefore forms a VC type class with its sole element serving as an envelope, which is integrable by Assumption SMRD (ii)(a). Assumption 3 follows from Lemma 9. □

A.6.2 Proof of Corollary 5

Proof. To check the Assumptions required by Theorem 1, for Assumption 1(ii)(a), notice that $\{\mu_2(x,0) : [\underline{x}, \overline{x}] \mapsto \mathbb{R}\}$ is a singleton, and therefore forms a VC type class with envelope 1. Assumption 3 follows from Lemma 9. Other conditions can be checked as before. Applying Theorem 1 gives

$$\frac{(\mu_2(0^+,0) - \mu_2(0^-,0))\widehat{\mathbb{X}}_n'(0,1) - (\mu_1(0^+,0) - \mu_1(0^-,0))\widehat{\mathbb{X}}_n'(0,2)}{(\mu_2(0^+,0) - \mu_2(0^-,0))^2} \rightsquigarrow N(0, \sigma_{FMRD}^2)$$

with probability approaching to one. According to Lemma 4, it remains to show

$$\frac{(\mu_2(0^+,0)-\mu_2(0^-,0))\widehat{\mathbb{X}}'_n(0,1)-(\mu_1(0^+,0)-\mu_1(0^-,0))\widehat{\mathbb{X}}'_n(0,2)}{(\mu_2(0^+,0)-\mu_2(0^-,0))^2}$$
$$-\frac{(\hat\mu_{2,2}(0^+,0)-\hat\mu_{2,2}(0^-,0))\widehat{\mathbb{X}}'_n(0,1)-(\hat\mu_{1,2}(0^+,0)-\hat\mu_1(0^-,0))\widehat{\mathbb{X}}'_n(0,2)}{(\hat\mu_{2,2}(0^+,0)-\hat\mu_{2,2}(0^-,0))^2}=o_p^{x\times\xi}(1).$$

This is the case due to the uniform consistency of $\hat\mu_{1,2}(0^\pm,0)$, $\hat\mu_{2,2}(0^\pm,0)$ that follows from Corollary 1, the independence between data and ξ_i under Assumption FMRD (vi), and the fact that $|\mu_2(0^+,0)-\mu_2(0^-,0)|>0$ under Assumption FMRD (ii) (d). □

A.6.3 Proof of Corollary 7

Proof. It is direct to show that Assumption GFMRD and Lemma 9 together imply the three assumptions required by Theorem 1. Note for $k=1,\ 2$, $\{\mu_k(\cdot,\theta):[\underline{x},\overline{x}]\mapsto\mathbb{R}:\theta\in\{1,...,K\}\}$ has finite elements and therefore is of VC-subgraph class with envelope $\max_{\theta\in\{1,\dots,K\}}\mu_k(x,\theta)$ and Assumption 1 (ii)(a) is satisfied. The theorem then gives

$$\frac{(\mu_2(0^+,\cdot)-\mu_2(0^-,\cdot))\widehat{\mathbb{X}}'_n(\cdot,1)-(\mu_1(0^+,\cdot)-\mu_1(0^-,\cdot))\widehat{\mathbb{X}}'_n(\cdot,2)}{(\mu_2(0^+,\cdot)-\mu_2(0^-,\cdot))^2}\rightsquigarrow N(0,\Sigma_{GFMRD})$$

with probability approaching to one. Lemmas 4 and 9 then give the desired result. □

A.6.4 Proof of Corollary 8

Proof. We check that Assumption SCRD implies the assumptions required by Theorem 1. Most are direct, and we only need to check the following three points. For Assumption 1 (ii) (a), $\{x\mapsto F_{Y|X}(\theta''|x):\theta''\in\mathscr{Y}_1\}$ and $\{y'\mapsto\mathbb{1}\{y'\le\theta''\}:\theta''\in\mathscr{Y}_1\}$ are increasing stochastic processes bounded by one, and they are of VC-subgraph classes according to Lemma 9.10 of Kosorok (2008), and thus of VC type. For Assumption 1 (ii)(c), note that $E[(\mathbb{1}\{Y_i\le y_1\}-F_{Y|X}(y_1|X_i))(\mathbb{1}\{Y_i\le y_2\}-F_{Y|X}(y_2|X_i))|X_i]=F_{Y|X}(y_1\wedge y_2|X_i)-F_{Y|X}(y_1|X_i)F_{Y|X}(y_2|X_i)$ and thus it follows from Assumption SCRD (ii). Assumption 1 (ii) (d) is implied by the right continuity of $y'\mapsto\mathbb{1}\{y\le y'\}$.

Assumption 3 is implied by Lemma 9. The result then follows from Theorem 1. □

A.6.5 Proof of Corollary 9

Proof. We need to check that Assumption SQRD implies all the three Assumptions required by Theorem 1. The only non-trivial ones are Assumptions 1 (ii) (a), (d) 2 (i), and 3. For Assumptions 1 (ii) (a) and 3 (i), note that $\{y' \mapsto \mathbb{1}\{y' \le y\} : y \in \mathscr{Y}_1\}$ and $\{x \mapsto F_{Y|X}(y|x) : y \in \mathscr{Y}_1\}$ are collections of increasing stochastic processes, and Lemma 9.10 of Kosorok (2008) suggests that they are of VC-class and thus VC type with envelope one. Assumption 1 (ii) (d) is implied by the right continuity of $y' \mapsto \mathbb{1}\{y \le y'\}$. Assumption 2 (i) follows from Lemma 3.9.23 (i) of van der Vaart and Wellner (1996) and Assumption 5.7 (ii). Assumption 3(b) is implied by Lemma 9.

By Theorem 1 (i), we have $\sqrt{nh_n}[\hat{\tau}_{SQRD}(\cdot) - \tau_{SQRD}(\cdot)] \rightsquigarrow \mathbb{G}'_{SQRD}$, where

$$\begin{aligned}\mathbb{G}'_{SQRD}(\cdot) =& \phi'_{F_{Y|X}(\cdot|0^+)}\Big(\mathbb{G}_{H+}\Big)(\cdot) - \phi'_{F_{Y|X}(\cdot|0^-)}\Big(\mathbb{G}_{H-}\Big)(\cdot) \\ =& -\frac{\mathbb{G}_{H+}(Q_{Y|X}(\cdot|0^+))}{f_{Y|X}(Q_{Y|X}(\cdot|0^+)|0^+)} + \frac{\mathbb{G}_{H-}(Q_{Y|X}(\cdot|0^-))}{f_{Y|X}(Q_{Y|X}(\cdot|0^-)|0^-)}.\end{aligned}$$

Also, by Theorem 1 (ii), $\phi'_{F_{Y|X}(\cdot|0^+)}(\hat{\nu}^+_{\xi,n}) - \phi'_{F_{Y|X}(\cdot|0^-)}(\hat{\nu}^-_{\xi,n}) \underset{\xi}{\overset{p}{\rightsquigarrow}} \mathbb{G}'_{SQRD}$, where

$$\phi'_{F_{Y|X}(\cdot|0^\pm)}(\hat{\nu}^\pm_{\xi,n})(\theta'') = -\frac{\hat{\nu}^\pm_{\xi,n}(Q_{Y|X}(\theta''|0^\pm))}{f_{Y|X}(Q_{Y|X}(\theta''|0^\pm)|0^\pm)}.$$

In the EMP, we replace $f_{Y|X}(\cdot|0^\pm)$, $Q_{Y|X}(\cdot|0^\pm)$ by their uniformly consistent estimators $\hat{f}_{Y|X}(\cdot|0^\pm)$, $\hat{Q}_{Y|X}(\cdot|0^\pm)$, where the uniform consistency of the former follows from Assumption 5.7 (vi) (see Lemma 10) and the uniform consistency of the latter follows from Corollary 2.

By Lemma 4, it suffices to show that $\sup_{\theta'' \in [a,1-a]} \Big| \Big(\widehat{\phi}'_{F_{Y|X}(\cdot|0^+)}(\hat{\nu}^+_{\xi,n})(\theta'') - \widehat{\phi}'_{F_{Y|X}(\cdot|0^-)}(\hat{\nu}^-_{\xi,n})(\theta'')\Big) - \Big(\phi'_{F_{Y|X}(\cdot|0^+)}(\hat{\nu}^+_{\xi,n})(\theta'') - \phi'_{F_{Y|X}(\cdot|0^-)}(\hat{\nu}^-_{\xi,n})(\theta'')\Big)\Big| \underset{x\times\xi}{\overset{p}{\to}} 0$. We first show

$$\begin{aligned}&\Big\|\widehat{\phi}'_{F_{Y|X}(\cdot|0^+)}(\hat{\nu}^+_{\xi,n}(\hat{Q}_{Y|X}(\cdot|0^+))) - \phi'_{F_{Y|X}(\cdot|0^+)}(\hat{\nu}^+_{\xi,n}(\hat{Q}_{Y|X}(\cdot|0^+)))\Big\|_{[a,1-a]} \\ =& \Big\| -\frac{\hat{\nu}^+_{\xi,n}(\hat{Q}_{Y|X}(\cdot|0^+))}{\hat{f}_{Y|X}(\hat{Q}_{Y|X}(\cdot|0^+)|0^+)} + \frac{\hat{\nu}^+_{\xi,n}(Q_{Y|X}(\cdot|0^+))}{f_{Y|X}(Q_{Y|X}(\cdot|0^+)|0^+)}\Big\|_{[a,1-a]} \underset{x\times\xi}{\overset{p}{\to}} 0.\end{aligned}$$

Lemma 10 and Corollary 2 along with the asymptotic equicontinuity of $\hat{\nu}^+_{\xi,n}$ implied by its weak convergence in Theorem 1 (i) suggest that $\hat{\nu}^+_{\xi,n}(\hat{Q}_{Y|X}(\cdot|0^+)) - \hat{\nu}^+_{\xi,n}(Q_{Y|X}(\cdot|0^+)) \underset{x\times\xi}{\overset{p}{\to}} 0$ uniformly. Assumption

SQRD (ii)(b) and the uniform consistency of both $\hat{Q}_{Y|X}(\cdot|0^+)$ and $\hat{f}_{Y|X}(\cdot|0^+)$ shows that $f_{Y|X}(\cdot|0^+)$ is bounded away from 0 uniformly, and with probability approaching one

$$\sup_{\theta''\in[a,1-a]} |\hat{f}_{Y|X}(\hat{Q}_{Y|X}(\theta''|0^+)|0^+) - f_{Y|X}(Q_{Y|X}(\theta''|0^+)|0^+)|$$

$$\leq \sup_{y\in\mathscr{Y}_1} |\hat{f}_{Y|X}(y|0^+) - f_{Y|X}(y|0^+)| + \sup_{\theta''\in[a,1-a]} L|\hat{Q}_{Y|X}(\theta''|0^+) - Q_{Y|X}(\theta''|0^+)| = o_p^x(1) + o_p^x(1)$$

for a Lipschitz constant $L > 0$. Thus, $\left\| -\frac{\hat{\nu}^+_{\xi,n}(\hat{Q}_{Y|X}(\cdot|0^+))}{\hat{f}_{Y|X}(\hat{Q}_{Y|X}(\cdot|0^+)|0^+)} + \frac{\hat{\nu}^+_{\xi,n}(Q_{Y|X}(\cdot|0^+))}{f_{Y|X}(Q_{Y|X}(\cdot|0^+)|0^+)} \right\|_{[a,1-a]} \underset{x\times\xi}{\overset{p}{\to}} 0.$

Similar lines show $\left\| \widehat{\phi}'_{F_{Y|X}(\cdot|0^-)}(\hat{\nu}^-_{\xi,n}(\hat{Q}_{Y|X}(\cdot|0^-))) - \phi'_{F_{Y|X}(\cdot|0^-)}(\hat{\nu}^-_{\xi,n}(\hat{Q}_{Y|X}(\cdot|0^-))) \right\|_{[a,1-a]} \underset{x\times\xi}{\overset{p}{\to}} 0$ as well. Therefore, we have

$$\sup_{\theta''\in[a,1-a]} \left| \left(\widehat{\phi}'_{F_{Y|X}(\cdot|0^+)}(\hat{\nu}^+_{\xi,n})(\theta'') - \widehat{\phi}'_{F_{Y|X}(\cdot|0^-)}(\hat{\nu}^-_{\xi,n})(\theta'')\right) - \left(\phi'_{F_{Y|X}(\cdot|0^+)}(\hat{\nu}^+_{\xi,n})(\theta'') - \phi'_{F_{Y|X}(\cdot|0^-)}(\hat{\nu}^-_{\xi,n})(\theta'')\right) \right| \underset{x\times\xi}{\overset{p}{\to}} 0.$$

We then apply Lemma 4 to conclude the proof. □

A.6.6 Proof of Corollary 10

Proof. For the unconditional weak convergence, the assumptions required by Theorem 1 can be checked as in Corollary 9 except for Assumption 2(i) now follows from Assumption 5.8 (ii)(a). Applying Theorem 1 (i), we have $\sqrt{nh_n^3}[\tilde{\tau}_{SQRK} - \tau_{SQRK}] \rightsquigarrow \mathbb{G}_{SQRD}$. It then suffices to show

$$\sqrt{nh_n^3}\,\|\tilde{\tau}_{SQRK} - \hat{\tau}_{SQRK}\|_{\Theta''} = o_p^x(1)$$

By definition, we only need to show

$$\begin{aligned} &\sqrt{nh_n^3}\left\|\widehat{\phi}(\hat{F}^{(1)}_{Y|X}(\cdot|0^\pm)) - \phi(\hat{F}^{(1)}_{Y|X}(\cdot|0^\pm))\right\|_{\Theta''} \\ =& -\sqrt{nh_n^3}\left\| \frac{\hat{F}^{(1)}_{Y|X}(\hat{Q}_{Y|X}(\theta''|0^\pm)|0)}{\hat{f}_{Y|X}(\hat{Q}_{Y|X}(\theta''|0)|0)} - \frac{\hat{F}^{(1)}_{Y|X}(Q_{Y|X}(\theta''|0^\pm)|0)}{f_{Y|X}(Q_{Y|X}(\theta''|0)|0)} \right\|_{\Theta''} = o_p^x(1). \end{aligned} \tag{A.4}$$

We first claim that

$$\sqrt{nh_n^3}[\hat{F}^{(1)}_{Y|X}(\hat{Q}_{Y|X}(\theta''|0)|0^\pm) - \hat{F}^{(1)}_{Y|X}(Q_{Y|X}(\theta''|0)|0^\pm)] = o_p^x(1)$$

respectively. Then by Assumption SQRK (ii) (b) and uniform super-consistency of $\hat{f}_{Y|X}(\cdot|0)$ and $\hat{Q}_{Y|X}(\cdot|0)$, $\sup_{\theta''\in[a,1-a]} |\hat{f}_{Y|X}(\hat{Q}_{Y|X}(\theta''|0)|0) - f_{Y|X}(Q_{Y|X}(\theta''|0)|0)| = o_p^x(1)$ and $\left|\frac{1}{f_{Y|X}(Q_{Y|X}(\theta''|0)|0)}\right| < C < \infty$. We can thus conclude that equation (A.4) is true.

To prove the claim, notice that

$$
\begin{aligned}
&\sqrt{nh_n^3}[\hat{F}^{(1)}_{Y|X}(\hat{Q}_{Y|X}(\theta''|0)|0^{\pm}) - \hat{F}^{(1)}_{Y|X}(Q_{Y|X}(\theta''|0)|0^{\pm})] \\
\leq &\sqrt{nh_n^3}[\hat{F}^{(1)}_{Y|X}(\hat{Q}_{Y|X}(\theta''|0)|0^{\pm}) - F^{(1)}_{Y|X}(\hat{Q}_{Y|X}(\theta''|0)|0^{\pm})] \\
&+ \sqrt{nh_n^3}[F^{(1)}_{Y|X}(\hat{Q}_{Y|X}(\theta''|0)|0^{\pm}) - F^{(1)}_{Y|X}(Q_{Y|X}(\theta''|0)|0^{\pm})] \\
&+ \sqrt{nh_n^3}[F^{(1)}_{Y|X}(Q_{Y|X}(\theta''|0)|0^{\pm}) - \hat{F}^{(1)}_{Y|X}(Q_{Y|X}(\theta''|0)|0^{\pm})] \\
= &(1) + (2) + (3)
\end{aligned}
$$

From Theorem 1(i) and the Hadamard differentiability of left inverse operator, we have $\sqrt{nh_n}[\hat{Q}_{Y|X}(\theta''|0) - Q_{Y|X}(\theta''|0)] = O_p^x(1)$ uniformly. An application of Slutsky's lemma implies $|\hat{Q}_{Y|X}(\theta''|0) - Q_{Y|X}(\theta''|0)| = O^x(\frac{1}{\sqrt{nh_n}})$. This and Assumption 5.8(ii)(a) imply $(2) = o_p^x(1)$ uniformly in θ''. From Lemma 1, $(3) = -\nu_n^{\pm}(Q_{Y|X}(\theta''|0)) + o_p^x(1)$ and $(1) = \nu_n^{\pm}(\hat{Q}_{Y|X}(\theta''|0)) + o_p^x(1) = \nu_n^{\pm}(Q_{Y|X}(\theta''|0)) + o_p^x(1)$ uniformly in θ'', where the last equality is due to Lemma 11 and asymptotic $\rho-$equicontinuity of $\nu_n^{\pm}$ implied by its weak convergence from Theorem 1(i). Thus we have $\sqrt{nh_n^3}[\hat{F}^{(1)}_{Y|X}(\hat{Q}_{Y|X}(\theta''|0)|0^{\pm}) - \hat{F}^{(1)}_{Y|X}(Q_{Y|X}(\theta''|0)|0^{\pm})] = o_p^x(1)$.

As for the conditional weak convergence part of the statement, Theorem 1(ii) shows

$$\phi'_{F_{Y|X}(\cdot|0^{\pm})}(\hat{\nu}^{\pm}_{\xi,n})(\cdot) = -\frac{\hat{\nu}^{\pm}_{\xi,n}(Q_{Y|X}(\cdot|0))}{f_{Y|X}(Q_{Y|X}(\cdot|0))} \underset{\xi}{\overset{p}{\rightsquigarrow}} -\frac{\mathbb{G}_{\pm}(Q_{Y|X}(\cdot|0))}{f_{Y|X}(Q_{Y|X}(\cdot|0)|0)}$$

Uniform consistency of $\hat{f}_{Y|X}(\cdot|0)$, $\hat{Q}_{Y|X}(\cdot|0)$, Assumption SQRK (ii), asymptotic $\rho-$equicontinuity of $\nu^{\pm}_{\xi,n}$, Lemmas 11, 4 and 14 then imply

$$\widehat{\phi}'_{F_{Y|X}(\cdot|0^{\pm})}(\hat{\nu}^{\pm}_{\xi,n})(\cdot) = -\frac{\hat{\nu}^{\pm}_{\xi,n}(\hat{Q}_{Y|X}(\cdot|0))}{\hat{f}_{Y|X}(\hat{Q}_{Y|X}(\cdot|0))} \underset{\xi}{\overset{p}{\rightsquigarrow}} -\frac{\mathbb{G}_{\pm}(Q_{Y|X}(\cdot|0))}{f_{Y|X}(Q_{Y|X}(\cdot|0)|0)}$$

and thus by Assumption 5.8 (i)(a) and the continuous mapping theorem, we conclude

$$\frac{\widehat{\phi}'_{F_{Y|X}(\cdot|0^+)}(\hat{\nu}^{+}_{\xi,n})(\cdot) - \widehat{\phi}'_{F_{Y|X}(\cdot|0^-)}(\hat{\nu}^{-}_{\xi,n})(\cdot)}{b'(0^+) - b'(0^-)} = \frac{1}{b'(0^+) - b'(0^-)}[-\frac{\hat{\nu}^{+}_{\xi,n}(\hat{Q}_{Y|X}(\cdot|0))}{\hat{f}_{Y|X}(\hat{Q}_{Y|X}(\cdot|0)|0)} + \frac{\hat{\nu}^{-}_{\xi,n}(\hat{Q}_{Y|X}(\cdot|0))}{\hat{f}_{Y|X}(\hat{Q}_{Y|X}(\cdot|0)|0)}]$$

$$\underset{\xi}{\overset{p}{\rightsquigarrow}} \mathbb{G}'_{SQRK}(\cdot)$$

□

A.6.7 Proof of Corollary 11

Proof. We first verify that the Assumptions required by Theorem 1 are satisfied.

To show Assumption 1(ii)(a), it is true that the subgraphs of $\{(y, d) \mapsto \mathbb{1}\{y \leq y'\} : y' \in \mathscr{Y}_1\}$ and $\{(y, d) \mapsto \mathbb{1}\{d = d'\} : d' \in \mathscr{D}\}$ can not shatter any two-point sets (for definition, see Section 9.1.1 of Kosorok(2008)); to see this, let (y_1, d_1, r_1), $(y_2, d_2, r_2) \in \mathscr{Y}_1 \times \mathscr{D} \times \mathbb{R}$ with $y_1 \leq y_2$. Notice that $\{(y_1, d_1, r_1)\}$ can never be picked out by any function in either of the classes. Therefore they are of VC-subgraph classes with VC index of 2. This implies that they are of VC type and Lemma 8 implies that the set of products $\{(t, y) \mapsto \mathbb{1}\{y \leq y'\}t : y' \in \mathscr{Y}_1\}$ is of VC type with square integrable constant function 1 as its envelope. On the other hand, $\{x \mapsto E[\mathbb{1}\{Y_i \leq y\}\mathbb{1}\{t = d\}|X_i = x] : (y, d) \in \mathscr{Y}_1 \times \mathscr{D}\}$ is of VC type with square integrable envelope 1 because of Example 19.7 of van der Vaart (1998) under boundedness of $\mathscr{Y}_1$ and Lipschitz continuity from Assumption FQRD (i) (a) and (ii). Assumption 1(ii)(c) follows from Assumption FQRD (ii) since for any (y_1, d_1), $(y_2, d_2) \in \mathscr{Y} \times \mathscr{D}$, if $d_1 = d_2 = d$ $E[\mathbb{1}\{Y \leq y_1, D = d\}\mathbb{1}\{Y \leq y_2, D = d\}|X = x] = E[\mathbb{1}\{Y \leq y_1 \wedge y_2, D = d\}|X = x]$ and It equals zero if $d_1 \neq d_2$. Assumption 1 (ii) (d) is implied by the right continuity of $(y', d') \mapsto \mathbb{1}\{y \leq y', d = d'\}$ on $\mathscr{Y}_1 \times \{0, 1\}$ in y for each d. Assumption 3 is implied by Lemma 9. Assumption FQRD (vii) and Lemma 3.9.23 of van der Vaart and Wellner (applicable under Assumption FQRD (ii), (v) and (vii)) implies Hadamard differentiability of Assumption 2(i).

The rest are directly implied by Assumption FQRD. Thus we may apply Theorem 1(i) and acquire $\nu_n \rightsquigarrow \mathbb{G} := \mathbb{G}^+ - \mathbb{G}^-$ for a zero mean Gaussian process $\mathbb{G}$, where

$$\begin{bmatrix} \nu_n(y, d_1, d_2, 1) \\ \nu_n(y, d_1, d_2, 2) \end{bmatrix} = \begin{bmatrix} \sqrt{nh_n}[(\hat{\mu}_{1,2}(0^+, y, d_1) - \hat{\mu}_{1,2}(0^-, y, d_1)) - (\mu_1(0^+, y, d_1) - \mu_1(0^-, y, d_1))] \\ \sqrt{nh_n}[(\hat{\mu}_{2,2}(0^+, d_2) - \hat{\mu}_{2,2}(0^-, d_2)) - (\mu_2(0^+, d_2) - \mu_2(0^-, d_2))] \end{bmatrix}$$

Theorem 1(i) further implies

$$
\begin{aligned}
&\sqrt{nh}(\hat{\tau}_{FQRD}(\cdot) - \tau_{FQRD}(\cdot)) \\
\rightsquigarrow&\mathbb{G}'_{FQRD}(\cdot) \\
:=&\Upsilon'_W\Big(\frac{[\mu_2(0^+,d)-\mu_2(0^-,d)]\mathbb{G}(\cdot,\cdot,\cdot,1)-[\mu_1(0^+,y,d)-\mu_1(0^-,y,d)]\mathbb{G}(\cdot,\cdot,\cdot,2)}{[\mu_2(0^+,d)-\mu_2(0^-,d)]^2}\Big)(\cdot) \\
=&-\frac{[\mu_2(0^+,d)-\mu_2(0^-,d)]\mathbb{G}(Q_{Y^1|C}(\cdot),1,1,1)-[\mu_1(0^+,Q_{Y^1|C}(\cdot),1)-\mu_1(0^-,Q_{Y^1|C}(\cdot),1)]\mathbb{G}(Q_{Y^1|C}(\cdot),1,1,2)}{f_{Y^1|C}(Q_{Y^1|C}(\cdot))[\mu_2(0^+,1)-\mu_2(0^-,1)]^2} \\
&+\frac{[\mu_2(0^+,d)-\mu_2(0^-,d)]\mathbb{G}(Q_{Y^0|C}(\cdot),0,0,1)-[\mu_1(0^+,Q_{Y^0|C}(\cdot),0)-\mu_1(0^-,Q_{Y^0|C}(\cdot),0)]\mathbb{G}(Q_{Y^1|C}(\cdot),0,0,2)}{f_{Y^0|C}(Q_{Y^0|C}(\cdot))[\mu_2(0^+,0)-\mu_2(0^-,0)]^2}
\end{aligned}
$$

For the conditional weak convergence part of the proof, define

$$
\begin{aligned}
&\Upsilon'_W(\widehat{\mathbb{Y}}_n)(\theta'') = \\
-&\frac{[\mu_2(0^+,1)-\mu_2(0^-,1)]\widehat{\mathbb{X}}'_n(Q_{Y^1|C}(\theta''),1,1,1)-[\mu_1(0^+,Q_{Y^1|C}(\theta''),1)-\mu_1(0^-,Q_{Y^1|C}(\theta''),1)]\widehat{\mathbb{X}}'_n(Q_{Y^1|C}(\theta''),1,1,2)}{f_{Y^1|C}(Q_{Y^1|C}(\theta''))[\mu_2(0^+,1)-\mu_2(0^-,1)]^2} \\
+&\frac{[\mu_2(0^+,0)-\mu_2(0^-,0)]\widehat{\mathbb{X}}'_n(Q_{Y^0|C}(\theta''),0,0,1)-[\mu_1(0^+,Q_{Y^0|C}(\theta''),0)-\mu_1(0^-,Q_{Y^0|C}(\theta''),0)]\widehat{\mathbb{X}}'_n(Q_{Y^0|C}(\theta''),0,0,2)}{f_{Y^0|C}(Q_{Y^0|C}(\theta''))[\mu_2(0^+,0)-\mu_2(0^-,0)]^2}
\end{aligned}
$$

Theorem 1(ii) then suggests that $\Upsilon'_W(\widehat{\mathbb{Y}}_n) \overset{p}{\underset{\xi}{\rightsquigarrow}} \mathbb{G}'_{FQRD}$. Thus it suffices to show

$$\sup_{\theta''\in[a,1-a]} |\widehat{\Upsilon}'_W(\widehat{\mathbb{Y}}_n)(\theta'') - \Upsilon'_W(\widehat{\mathbb{Y}}_n)(\theta'')| \underset{x\times\xi}{\overset{p}{\to}} 0$$

which is true by the asymptotic $\rho-$equicontinuity of $\mathbb{X}'_n$ (which is inherited from the conditional weak convergence of $\hat{\nu}_{\xi,n}$), the uniform consistency of $\hat{Q}_{Y^d|C}$ and $\hat{f}_{Y^d|C}$ for $d = 1, 0$ by Lemmas 15, 3 and 12. □

B Bandwidth Choice in Practice

While our theory prescribes asymptotic rates of bandwidths, empirical practitioners need to choose bandwidth for each finite n. In this section, we provide a guide for this matter. We emphasize that our robust inference procedure allows for large bandwidths such as the ones based on the MSE optimality. Following the bias-robust approach from Calonico, Cattaneo and Titiunik (2014), we increment the degree of local polynomial estimation by one or more to p for the purpose of bias correction while using the optimal bandwidths for the correct order s. For instance, if we are interested in the local linear model ($s=1$), then we run a local quadratic regression ($p=2$) while using the optimal bandwidths for the local linear model ($s=1$). Generally, when we want to estimate the v-th order derivative via a local s-th order polynomial estimation, we fix the degree p such that $0 \le v \le s \le p$, and additionally $s < p$ if one wants to implement a bias correction. We now remind the readers of the following short-hand notations: $\Psi_s = \int_{\mathbb{R}} r_s(u) r_s'(u) K^2(u) du$, $\Psi_s^{\pm} = \int_{\mathbb{R}_\pm} r_s(u) r_s'(u) K^2(u) du$, $\Gamma_s = \int_{\mathbb{R}} K(u) r_s(u) r_s'(u) du$, $\Gamma_s^{\pm} = \int_{\mathbb{R}_\pm} K(u) r_s(u) r_s'(u) du$, $\Lambda_{s,s+1} = \int_{\mathbb{R}} u^{s+1} r_s(u) K(u) du$, and $\Lambda^{\pm}_{s,s+1} = \int_{\mathbb{R}_\pm} u^{s+1} r_s(u) K(u) du$.

For the main local polynomial estimation, we first derive the oracle MSE-optimal bandwidths for the v-th order derivative based on a local polynomial estimation of the s-th degree. For the numerator, we have

$$h_{1,n}^{orac}(\theta_1|s,v) = \left(\frac{2v+1}{2s+2-2v} \frac{C'_{1,\theta_1,s,v}}{C^2_{1,\theta_1,s,v}} \right)^{1/(2s+3)} n^{-1/(2s+3)},$$

where $C'_{1,\theta_1,s,v}$ and $C^2_{1,\theta_1,s,v}$ are given by the leading terms of bias and variance, respectively:

$$C_{1,\theta_1,s,v} = \frac{Bias(\mu_1^{(v)}(0^+,\theta_1) - \mu_1^{(v)}(0^-,\theta_1))}{h_{1,n}^{s+1-v}(\theta_1)} = e_v' \left[\frac{(\Gamma_s^+)^{-1}\Lambda^+_{s,s+1}}{(s+1)!} \mu_1^{(s+1)}(0^+,\theta_1) - \frac{(\Gamma_s^-)^{-1}\Lambda^-_{s,s+1}}{(s+1)!} \mu_1^{(s+1)}(0^-,\theta_1) \right]$$

$$C'_{1,\theta_1,s,v} = n h_{1,n}^{2v+1}(\theta_1) Var(\mu_1^{(v)}(0^{\pm},\theta_1) - \mu_1^{(v)}(0^-,\theta_1)) =$$
$$\frac{e_v'[\sigma_{11}(\theta_1,\theta_1|0^+)(\Gamma_s^+)^{-1}\Psi_s^+((\theta_1,1)(\theta_1,1))(\Gamma_s^+)^{-1} + \sigma_{11}(\theta_1,\theta_1|0^-)(\Gamma_s^-)^{-1}\Psi_s^-((\theta_1,1)(\theta_1,1))(\Gamma_s^-)^{-1}]e_v}{f_X(0)}$$

Likewise, for the denominator, we have

$$h_{2,n}^{orac}(\theta_2|s,v) = \left(\frac{2v+1}{2s+2-2v} \frac{C'_{2,\theta_2,s,v}}{C^2_{2,\theta_2,s,v}} \right)^{1/(2s+3)} n^{-1/(2s+3)}$$

where $C'_{2,\theta_2,s,v}$ and $C^2_{2,\theta_2,s,v}$ are given by the leading terms of bias and variance, respectively:

$$C_{2,\theta_2,s,v} = \frac{Bias(\mu_2^{(v)}(0^+,\theta_2) - \mu_2^{(v)}(0^-,\theta_2))}{h_{2,n}^{s+1-v}(\theta_2)} = e_v' \left[\frac{(\Gamma_s^+)^{-1}\Lambda^+_{s,s+1}}{(s+1)!} \mu_2^{(s+1)}(0^+,\theta_2) - \frac{(\Gamma_s^-)^{-1}\Lambda^-_{s,s+1}}{(s+1)!} \mu_2^{(s+1)}(0^-,\theta_2) \right]$$

$$C'_{2,\theta_2,s,v} = n h_{2,n}^{2v+1}(\theta_2) Var(\mu_2^{(v)}(0^+,\theta_2) - \mu_2^{(v)}(0^-,\theta_2)) =$$
$$\frac{e_v'[\sigma_{22}(\theta_2,\theta_2|0^+)(\Gamma_s^+)^{-1}\Psi_s^+((\theta_2,2)(\theta_2,2))(\Gamma_s^+)^{-1} + \sigma_{22}(\theta_2,\theta_2|0^-)(\Gamma_s^-)^{-1}\Psi_s^-((\theta_2,2)(\theta_2,2))(\Gamma_s^-)^{-1}]e_v}{f_X(0)}$$

In practice, the unknowns in the above bandwidth selectors need to be replaced by their consistent estimates. We propose the following three-step procedure.

Step 1: Estimate $f_X(0)$ by the kernel density estimator

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

with the bandwidth c_n determined by Silverman's rule of thumb

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

where $\hat{\sigma}_X$ is the standard deviation of the sample $\{X_i\}_{i=1}^n$. We then compute the preliminary bandwidths for first-stage estimates, $\mu_k^{(v)}$, $k = 1, 2$, by

$$h_{1,n}^0 = (\frac{2v+1}{2s+2-v}\frac{C'_{1,0}}{C_{1,0}^2})^{1/5} n^{-1/5},$$
$$h_{2,n}^0 = (\frac{2v+1}{2s+2-v}\frac{C'_{2,0}}{C_{2,0}^2})^{1/5} n^{-1/5},$$

where the constant terms

$$C_{k,0} = e'_v[\frac{(\Gamma_s^+)^{-1}\Lambda_{s,s+1}^+}{(s+1!)}\bar{\mu}_{k,+}^{(s+1)} - \frac{(\Gamma_s^-)^{-1}\Lambda_{s,s+1}^-}{(s+1)!}\bar{\mu}_{k,-}^{(s+1)}]$$
$$C'_{k,0} = e'_v[\bar{\sigma}_{k,+}^2(\Gamma_s^+)^{-1}\Psi_s^+(\Gamma_s^+)^{-1} + \bar{\sigma}_{k,-}^2(\Gamma_s^-)^{-1}\Psi_s^-(\Gamma_s^-)^{-1}]e_v/\hat{f}_X(0)$$

depend on the preliminary estimates $\bar{\mu}_{k,\pm}^{(s+1)}$ and $\bar{\sigma}_{k,\pm}^2$ for $\mu_k^{(v)}$ and σ_{kk}, respectively. These preliminary estimates may be obtained by global parametric polynomial regressions of order greater or equal to $s+1$ and the sample variance of $\bar{\mu}_{k,\pm}^{(s+1)}$. Through simulations to be presented below, we find that simply setting $\bar{\mu}_{k,\pm}^{(s+1)}$ and $\bar{\sigma}_{k,\pm}^2$ to one in this first step also yields fine results, whereas $\hat{f}_X(0)$ should not be substituted by an arbitrary constant.

Step 2 Using the preliminary bandwidths obtained in Step 1, we next obtain the first stage estimates $[\check{\mu}_k(0^\pm,\theta_k), ..., \check{\mu}_k^{(s)}(0^\pm,\theta_k)]' = \check{\alpha}'_{k\pm,s}\text{diag}[1, 1!/h_{k,n}^0, .., s!/(h_{k,n}^0)^s]$ as follows. Solve

$$\check{\alpha}'_{1\pm,s} := \arg\min_{\alpha\in\mathbb{R}^{s+1}} \sum_{i=1}^{n} \delta_i^\pm(g_1(Y_i|\theta_1) - r_s(X_i/h_{0,n})'\alpha)K(\frac{X_i}{h_{1,n}^0}),$$
$$\check{\alpha}'_{2\pm,s} := \arg\min_{\alpha\in\mathbb{R}^{s+1}} \sum_{i=1}^{n} \delta_i^\pm(g_2(D_i,\theta_2) - r_s(X_i/h_{0,n})'\alpha)K(\frac{X_i}{h_{2,n}^0}).$$

Using these estimates of the local polynomial coefficients, we in turn compute first stage estimates based on s-th order expansion

$$\begin{aligned}\check{\mu}_k(x,\theta_k) =&[\mu_k(0^+,\theta_k)+\mu_k^{(1)}(0^+,\theta_k)x+...+\mu_k^{(s)}(0^+,\theta_k)\frac{x^s}{s!}]\delta_x^+\\ &+[\mu_k(0^-,\theta_k)+\mu_k^{(1)}(0^-,\theta_k)x+...+\mu_k^{(s)}(0^-,\theta_k)\frac{x^s}{s!}]\delta_x^-,\end{aligned}$$

for $k=1,2$. The covariance estimates are in turn computed by

$$\hat{\sigma}_{11}(\theta_1,\theta_1|0^\pm)=\Big(\frac{\sum_{i=1}^n(g_1(Y_i,\theta_1)-\check{\mu}_1(X_i,\theta_1))^2K(\frac{X_i}{h_{1,n}^0})\delta_i^\pm}{\sum_{i=1}^n K(\frac{X_i}{h_{1,n}^0})\delta_i^\pm}\Big)^{1/2},$$

$$\hat{\sigma}_{22}(\theta_2,\theta_2|0^\pm)=\Big(\frac{\sum_{i=1}^n(g_2(D_i,\theta_2)-\check{\mu}_2(X_i,\theta_2))^2K(\frac{X_i}{h_{2,n}^0})\delta_i^\pm}{\sum_{i=1}^n K(\frac{X_i}{h_{2,n}^0})\delta_i^\pm}\Big)^{1/2}.$$

The uniform consistency of $\check{\mu}_k(x,\theta_k)\mathbb{1}\{|x|\leq h_{k,n}^0\}$ in (x,θ_k) is implied by Lemma 9 with bandwidths $h_{1,n}(\theta_1)=h_{2,n}(\theta_2)=h_{0,n}$ selected in Step 1 under $r=s$. This further implies the uniform consistency of $\hat{\sigma}_{kk}$ for $k=1,2$.

Step 3: We are now ready to derive a feasible version of the main bandwidths. Let

$$h_{k,n}^{MSE}(\theta_1|s,v)=\left(\frac{2v+1}{2s+2-2v}\frac{\hat{C}'_{k,\theta_k,s,v}}{\hat{C}^2_{k,\theta_k,s,v}}\right)^{1/(2s+3)} n^{-1/(2s+3)}$$

where $C_{1,\theta_1,s,v}$, $C'_{1,\theta_1,s,v}$, $C_{2,\theta_2,s,v}$ and $C'_{2,\theta_2,s,v}$ are replaced by their estimates:

$$\begin{aligned}\hat{C}_{1,\theta_1,s,v}&=e_v'\left[\frac{(\Gamma_s^+)^{-1}\Lambda_{s,s+1}^+}{(s+1)!}\check{\mu}_{s,1}^{(s+1)}(0^+,\theta_1)-\frac{(\Gamma_s^-)^{-1}\Lambda_{s,s+1}^-}{(s+1)!}\check{\mu}_{s,1}^{(s+1)}(0^-,\theta_1)\right],\\ \hat{C}'_{1,\theta_1,s,v}&=\frac{e_v'[\hat{\sigma}_{11}(\theta_1,\theta_1|0^+)(\Gamma_s^+)^{-1}\Psi_s^+(\Gamma_s^+)^{-1}+\hat{\sigma}_{11}(\theta_1,\theta_1|0^-)(\Gamma_s^-)^{-1}\Psi_s^-(\Gamma_s^-)^{-1}]e_v}{\hat{f}_X(0)}\\ \hat{C}_{2,\theta_2,s,v}&=e_v'\left[\frac{(\Gamma_s^+)^{-1}\Lambda_{s,s+1}^+}{(s+1)!}\check{\mu}_{2}^{(s+1)}(0^+,\theta_2)-\frac{(\Gamma_s^-)^{-1}\Lambda_{s,s+1}^-}{(s+1)!}\check{\mu}_{2}^{(s+1)}(0^-,\theta_2)\right],\\ \hat{C}'_{2,\theta_2,s,v}&=\frac{e_v'[\hat{\sigma}_{22}(\theta_2,\theta_2|0^+)(\Gamma_s^+)^{-1}\Psi_s^+(\Gamma_s^+)^{-1}+\hat{\sigma}_{22}(\theta_2,\theta_2|0^-)(\Gamma_s^-)^{-1}\Psi_s^-(\Gamma_s^-)^{-1}]e_v}{\hat{f}_X(0)},\end{aligned}$$

respectively. To these feasible MSE-optimal choices, we further apply the rule of thumb (ROT) bandwidth algorithm for optimal coverage error following Calonico, Cattaneo, and Farrell (2016ab):

$$h_{1,n}^{ROT}(\theta_1|s,v)=h_{1,n}^{MSE}(\theta_1|s,v)n^{-s/(2s+3)(s+3)},$$

$$h_{2,n}^{ROT}(\theta_2|s,v)=h_{2,n}^{MSE}(\theta_2|s,v)n^{-s/(2s+3)(s+3)}.$$

References

Arai, Yoichi, and Hidehiko Ichimura (2016) "Optimal bandwidth selection for the fuzzy regression discontinuity estimator." *Economics Letters,* Vol 141, pp. 103–106.

Barrett, Garry F., and Stephen G. Donald (2003) "Consistent tests for stochastic dominance." Econometrica Vol. 71, No.1, pp. 71–104.

Bartalotti, Otvio C., Gray Calhoun, and Yang He (2016) "Bootstrap Confidence Intervals for Sharp Regression Discontinuity Designs with the Uniform Kernel." in Advances in Econometrics: Regression Discontinuity Designs: Theory and Applications, Vol. 38 (ed. M. Cattaneo and J. C. Escanciano).

Calonico, Sebastian, Matias D. Cattaneo, and Max Farrell (2016a) "Coverage Error Optimal Confidence Intervals for Regression Discontinuity Designs," *Working paper.*

Calonico, Sebastian, Matias D. Cattaneo, and Max Farrell (2016b) "On the Effect of Bias Estimation on Coverage Accuracy in Nonparametric Inference," *Journal of the American Statistical Association,* forthcoming.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik (2014) "Robust Nonparametric Confidence Intervals for Regression Discontinuity Designs," *Econometrica,* Vol. 82, No. 6, pp. 2295–2326.

Card, David, David S. Lee, Zhuan Pei, and Andrea Weber (2016) "Inference on causal effects in a generalized regression kink design." Econometrica Vol. 83, No. 6, pp. 2453–2483.

Cattaneo, Matias D., and Juan Carlos Escanciano (2016) "Regression Discontinuity Designs: Theory and Applications," *Advances in Econometrics,* Vol. 38 (Forthcoming)

Chernozhukov, Victor, Denis Chetverikov, and Kengo Kato (2014) "Gaussian Approximation of Suprema of Empirical Processes," *The Annals of Statistics,* Vol. 42, No. 4, pp. 1564–1597.

Chernozhukov, Victor and Iván Fernández-Val (2005) “Subsampling Inference on Quantile Regression Processes,” *Sankhya: The Indian Journal of Statistics,* Vol. 67, No. 2, pp. 253–276.

Chernozhukov, Victor, Iván Fernández-Val and Alfred Galichon (2010) “Quantile and Probability Curves without Crossing,” *Econometrica,* Vol. 78, No. 3, pp. 1093–1125.

Chiang, Harold D. and Yuya Sasaki (2016) “Quantile Regression Kink Designs,” *Working paper, Johns Hopkins University.*

Cook, Thomas D. (2008) “Waiting for life to arrive: a history of the regression-discontinuity design in psychology, statistics and economics.” *Journal of Econometrics,* Vol. 142, No. 2, pp. 636–654.

Donald, Stephen G., Yu-Chin Hsu, and Garry F. Barrett (2012) “Incorporating covariates in the measurement of welfare and inequality: methods and applications.” The Econometrics Journal Vol. 15, No. 1, pp. C1-C30.

Dong, Yingying (2016) “Jump or Kink? Identification of Binary Treatment Regression Discontinuity Design without the Discontinuity.” working paper.

Dony, Julia, Uwe Einmahl and David M. Mason (2006) “Uniform in Bandwidth Consistency of Local Polynomial Regression Function Estimators,” *Austrian Journal of Statistics,* Vol. 35, pp. 105–120.

Frandsen, Brigham R., Markus Frölich and Blaise Melly (2012) “Quantile Treatment Effects in the Regression Discontinuity Design,” *Journal of Econometrics,* Vol. 168, No.2 pp. 382-395.

Giné, Evarist, and Armelle Guillou (2002) “Rates of Strong Uniform Consistency for Multivariate Kernel Density Estimators,” *Annales de l’Institut Henri Poincaré, Probabilités et Statistiques,* Vol. 38. No. 6. pp. 907–921.

Giné, Evarist and Richard Nickl (2016) “Mathematical Foundations of Infinite-Dimensional Statistical Models,” Cambridge University Press: Cambridge.

Giné, Evarist, and Joel Zinn. (1990) "Bootstrapping general empirical measures," *The Annals of Probability,* Vol. 18, No. 2, pp. 851–869.

Guerre, Emmanuel and Camille Sabbah (2012) "Uniform Bias Study and Bahadur Representation for Local polynomial Estimators of the Conditional Quantile Function," *Econometric Thoery,* Vol. 26, No. 5, pp. 1529-1564.

Hansen, Bruce E. (1996) "Inference when a nuisance parameter is not identified under the null hypothesis." *Econometrica,* Vol. 64, No. 2, pp. 413–430.

Imbens, Guido, and Karthik Kalyanaraman. "Optimal bandwidth choice for the regression discontinuity estimator." *The Review of Economic Studies,* Vol. 79, No. 3, pp. 933–959.

Imbens, Guido and Thomas Lemieux (2008) "Special Issue Editors' Introduction: The Regression Discontinuity Design – Theory and Applications," *Journal of Econometrics,* Vol. 142, No. 2, pp. 611–614.

Imbens, Guido W., and Jeffrey M. Wooldridge (2009) "Recent developments in the econometrics of program evaluation." *Journal of Economic Literature,* Vol. 47, No. 1, pp. 5–86.

Koenker, Roger and Zhijie Xiao (2002) "Inference on the quantile regression process," *Econometrica,* Vol. 70, No. 4, pp.1583–1612.

Kong, Efang, Oliver B. Linton, and Yingcun Xia (2010) "Uniform Bahadur Representation for Local Polynomial Estimates of M-Regression and its Application to the Additive Model?," *Econometric Thoery,* Vol. 26, No. 5, pp. 1529–1564.

Kosorok, Michael R. (2003) "Bootstraps of sums of independent but not identically distributed stochastic processes." *Journal of Multivariate Analysis,* Vol 84, No. 2, pp. 299–318.

Kosorok, Michael R. (2008) "Introduction to Empirical Processes and Semiparametric Inference Springer: New York." (2008).

Landais, Camille (2015) "Assessing the Welfare Effects of Unemployment Benefits Using the Regression Kink Design," *American Economic Journal: Economic Policy*, Vol. 7, No. 4, pp. 243–278.

Ledoux, Michel, and Michel Talagrand. (1988) "Un critere sur les petites boules dans le thoreme limite central." *Probability Theory and Related Fields,* Vol. 77, No. 1, pp. 29–47.

Lee, David S., and Thomas Lemieux (2010) "Regression Discontinuity Designs in Economics," *Journal of Economic Literature,* Vol. 48, No. 2, pp. 281–355.

Masry, Elias (1996) "Multivariate local polynomial regression for time series: uniform strong consistency and rates." *Journal of Time Series Analysis,* Vol. 17, No. 6, pp. 571–599.

McFadden, Daniel (1989) "Testing for Stochastic Dominance," in *Studies in the Economics of Uncertainty: In Honor of Josef Hadar*, ed. by T. B. Fomby and T. K. Seo, Springer.

Nielsen, Helena Skyt, Torben Sørensen, and Christopher Taber (2010) "Estimating the Effect of Student Aid on College Enrollment: Evidence from a Government Grant Policy Reform," *American Economic Journal: Economic Policy,* Vol. 2, No. 2, pp. 185–215.

Pollard, David (1990) "Empirical Processes: Theory and Applications. *NSF-CBMS Regional Conference Series in Probability and Statistics* Vol. 2

Porter, Jack (2003). "Estimation in the regression discontinuity model." Unpublished Manuscript, Department of Economics, University of Wisconsin at Madison.

Qu, Zhongjun and Jungmo Yoon (2015a) "Nonparametric Estimation and Inference on Conditional Quantile Processes," *Journal of Econometrics,* Vol. 185, No.1, pp. 1–19.

Qu, Zhongjun and Jungmo Yoon (2015b) "Uniform Inference on Quantile Effects under Sharp Regression Discontinuity Designs," *Working Paper,* 2015.

Shen, Shu and Xiaohan Zhang (2016) "Distributional Tests for Regression Discontinuity: Theory and Empirical Examples," *Review of Economics and Statistics,* Vol. 98, No. 4, pp. 685–700.

Simonsen, Marianne, Lars Skipper, and Niels Skipper (2015) "Price sensitivity of demand for prescription drugs: Exploiting a regression kink design," *Journal of Applied Econometrics,* Forthcoming.

Tsybakov, Alexandre B. (2000) "Introduction to Nonparametric Estimation," Springer Series in Statistics.

van der Vaart, Aad W. (1998) "Asymptotic Statistics," Cambridge University Press: Cambridge.

van der Vaart, Aad W. and Jon A. Wellner (1996) "Weak Convergence and Empirical Processes," Springer-Verlag: New York.

Van Der Vaart, Aad, and Jon A. Wellner (2011). "A local maximal inequality under uniform entropy." *Electronic Journal of Statistics,* Vol. 5, pp. 192–203.

Tables

Table 1: Simulated acceptance probabilities for treatment nullity under the sharp RDD across alternative values of $\beta_1 \in \{0.00, 0.25, 0.50, 0.75, 1.00\}$. The nominal acceptance probability is 95%.

n	β_1				
	0.00	0.25	0.50	0.75	1.00
1000	0.939	0.842	0.641	0.366	0.190
2000	0.949	0.812	0.481	0.204	0.075
4000	0.955	0.738	0.304	0.082	0.019

Table 2: Simulated acceptance probabilities for treatment nullity under the sharp RKD across alternative values of $\beta_1 \in \{0.00, 0.25, 0.50, 0.75, 1.00\}$. The nominal acceptance probability is 95%.

n	β_1				
	0.00	0.25	0.50	0.75	1.00
1000	0.914	0.843	0.731	0.598	0.469
2000	0.917	0.812	0.641	0.465	0.324
4000	0.932	0.772	0.516	0.348	0.194

Table 3: Simulated acceptance probabilities for treatment nullity under the fuzzy RDD across alternative values of $\beta_1 \in \{0.00, 0.25, 0.50, 0.75, 1.00\}$. The nominal acceptance probability is 95%.

n	β_1				
	0.00	0.25	0.50	0.75	1.00
1000	0.934	0.876	0.686	0.403	0.210
2000	0.956	0.846	0.516	0.230	0.090
4000	0.949	0.762	0.336	0.103	0.017

Table 4: Simulated acceptance probabilities for treatment nullity under the fuzzy RKD across alternative values of $\beta_1 \in \{0.00, 0.25, 0.50, 0.75, 1.00\}$. The nominal acceptance probability is 95%.

n	β_1				
	0.00	0.25	0.50	0.75	1.00
1000	0.913	0.864	0.760	0.642	0.554
2000	0.925	0.838	0.686	0.551	0.414
4000	0.936	0.786	0.588	0.419	0.290

Table 5: Simulated acceptance probabilities for (A) joint treatment nullity and (B) treatment homogeneity under the fuzzy RDD with group covariate across alternative values of $\beta_1 \in \{0.00, 0.25, 0.50, 0.75, 1.00\}$ while fixing $\beta_2 = 0$. The nominal acceptance probability is 95%.

(A) Joint Treatment Nullity						(B) Treatment Homogeneity					
n	β_1					n	β_1				
	0.00	0.25	0.50	0.75	1.00		0.00	0.25	0.50	0.75	1.00
1000	0.980	0.971	0.957	0.940	0.885	1000	0.972	0.966	0.964	0.944	0.919
2000	0.977	0.959	0.942	0.881	0.803	2000	0.971	0.964	0.951	0.913	0.867
4000	0.972	0.959	0.902	0.786	0.601	4000	0.971	0.965	0.925	0.855	0.762

Table 6: Simulated acceptance probabilities for (A) uniform treatment nullity and (B) treatment homogeneity under the sharp quantile RDD. The top panel (I) presents results across alternative values of $\beta_1 \in \{0.00, 0.25, 0.50, 0.75, 1.00\}$ while fixing $\gamma_1 = 0$. The bottom panel (II) presents 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 is 95%.

(I) (A) Joint Treatment Nullity						(I) (B) Treatment Homogeneity					
n	β_1					n	β_1				
	0.00	0.25	0.50	0.75	1.00		0.00	0.25	0.50	0.75	1.00
1000	0.966	0.917	0.798	0.620	0.484	1000	0.967	0.967	0.965	0.954	0.963
2000	0.959	0.859	0.633	0.414	0.269	2000	0.959	0.959	0.966	0.955	0.958
4000	0.950	0.740	0.400	0.161	0.074	4000	0.950	0.946	0.958	0.948	0.947

(II) (A) Joint Treatment Nullity						(II) (B) Treatment Homogeneity					
n	γ_1					n	γ_1				
	0.00	0.25	0.50	0.75	1.00		0.00	0.25	0.50	0.75	1.00
1000	0.966	0.928	0.842	0.742	0.647	1000	0.967	0.920	0.808	0.698	0.574
2000	0.959	0.866	0.669	0.500	0.378	2000	0.959	0.855	0.625	0.446	0.327
4000	0.950	0.718	0.362	0.206	0.138	4000	0.950	0.693	0.327	0.170	0.116

Table 7: Simulated acceptance probabilities for (A) uniform treatment nullity and (B) treatment homogeneity under the sharp quantile RKD. The top panel (I) presents results across alternative values of $\beta_1 \in \{0.00, 0.25, 0.50, 0.75, 1.00\}$ while fixing $\gamma_1 = 0$. The bottom panel (II) presents 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 is 95%.

(I) (A) Joint Treatment Nullity						(I) (B) Treatment Homogeneity					
n	β_1					n	β_1				
	0.00	0.25	0.50	0.75	1.00		0.00	0.25	0.50	0.75	1.00
1000	0.930	0.928	0.906	0.881	0.851	1000	0.927	0.936	0.945	0.936	0.940
2000	0.941	0.917	0.883	0.850	0.837	2000	0.932	0.930	0.929	0.929	0.928
4000	0.941	0.902	0.850	0.835	0.829	4000	0.929	0.930	0.918	0.936	0.924

(II) (A) Joint Treatment Nullity						(II) (B) Treatment Homogeneity					
n	γ_1					n	γ_1				
	0.00	0.25	0.50	0.75	1.00		0.00	0.25	0.50	0.75	1.00
1000	0.930	0.954	0.955	0.943	0.936	1000	0.927	0.938	0.939	0.921	0.911
2000	0.941	0.931	0.936	0.926	0.902	2000	0.932	0.932	0.923	0.901	0.875
4000	0.941	0.931	0.928	0.905	0.879	4000	0.929	0.920	0.885	0.861	0.816

Table 8: Simulated acceptance probabilities for (A) uniform treatment nullity and (B) treatment homogeneity under the fuzzy quantile RDD. The top panel (I) presents results across alternative values of $\beta_1 \in \{0.00, 0.25, 0.50, 0.75, 1.00\}$ while fixing $\gamma_1 = 0$. The bottom panel (II) presents 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 is 95%.

(I) (A) Joint Treatment Nullity						(I) (B) Treatment Homogeneity					
n	β_1					n	β_1				
	0.00	0.25	0.50	0.75	1.00		0.00	0.25	0.50	0.75	1.00
1000	0.983	0.894	0.703	0.472	0.265	1000	0.987	0.985	0.992	0.994	0.996
2000	0.984	0.843	0.507	0.236	0.072	2000	0.985	0.990	0.990	0.995	0.992
4000	0.978	0.775	0.292	0.044	0.004	4000	0.978	0.979	0.993	0.989	0.992

(II) (A) Joint Treatment Nullity						(II) (B) Treatment Homogeneity					
n	γ_1					n	γ_1				
	0.00	0.25	0.50	0.75	1.00		0.00	0.25	0.50	0.75	1.00
1000	0.983	0.968	0.933	0.886	0.857	1000	0.987	0.949	0.876	0.790	0.715
2000	0.984	0.966	0.890	0.796	0.715	2000	0.985	0.960	0.776	0.593	0.438
4000	0.978	0.932	0.817	0.629	0.441	4000	0.978	0.874	0.601	0.345	0.166

Table 9: Simulated acceptance probabilities for (A) uniform treatment nullity and (B) treatment homogeneity under the fuzzy quantile RKD. The top panel (I) presents results across alternative values of $\beta_1 \in \{0.00, 0.25, 0.50, 0.75, 1.00\}$ while fixing $\gamma_1 = 0$. The bottom panel (II) presents 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 is 95%.

(I) (A) Joint Treatment Nullity						(I) (B) Treatment Homogeneity					
n	β_1					n	β_1				
	0.00	0.25	0.50	0.75	1.00		0.00	0.25	0.50	0.75	1.00
1000	0.945	0.929	0.903	0.868	0.842	1000	0.938	0.942	0.945	0.946	0.936
2000	0.941	0.911	0.873	0.836	0.815	2000	0.939	0.927	0.936	0.931	0.930
4000	0.935	0.904	0.846	0.799	0.802	4000	0.929	0.928	0.935	0.931	0.929

(II) (A) Joint Treatment Nullity						(II) (B) Treatment Homogeneity					
n	γ_1					n	γ_1				
	0.00	0.25	0.50	0.75	1.00		0.00	0.25	0.50	0.75	1.00
1000	0.945	0.956	0.950	0.952	0.928	1000	0.938	0.948	0.949	0.930	0.903
2000	0.941	0.944	0.936	0.919	0.921	2000	0.939	0.938	0.914	0.890	0.872
4000	0.935	0.941	0.928	0.905	0.888	4000	0.929	0.938	0.896	0.848	0.790