

A Unified Framework for Identification and Inference of Local Treatment Effects in Sharp Regression Kink Designs*

ZHIXIN WANG

School of Economics, Shanghai University of Finance and Economics

ZHENGYU ZHANG

School of Economics, Shanghai University of Finance and Economics

Abstract

This paper develops a unified framework for the identification, estimation, and uniform inference of local treatment effects (LTEs) in sharp regression kink designs (RKDs). These LTEs quantify the effect of a marginal change in the treatment at the kink point on various features of the outcome distribution. The identification strategy applies to Hadamard-differentiable functionals of the outcome distribution—including means, quantiles, and inequality measures—and encompasses several existing RKD estimands as special cases. For estimation, we categorize the corresponding estimands into two general classes and implement their estimation via local polynomial constrained regression. We establish the asymptotic theory for this framework and provide a valid resampling procedure for uniform inference. The method is applied to examine the effect of unemployment insurance on unemployment durations, focusing on the policy’s impact on the distribution and inequality of durations, as a complement to existing empirical evidence.

Keywords: Continuous treatments; regression kink design; nonparametric identification; nonseparable models; multiplier bootstrap.

*The two authors contributed equally and are listed in alphabetical order. Address correspondence to Zhengyu Zhang, e-mail: zy.zhang@mail.shufe.edu.cn.

1 INTRODUCTION

The regression kink design (RKD) exploits the change in the slope of the relationship between an endogenous treatment variable and a covariate. This quasi-experimental method has become a popular approach in empirical economics, particularly for studying the labor supply effects of social insurance (Landais, 2015; Card et al., 2015a; Kolsrud et al., 2018; Landais and Spinnewijn, 2021). Formal econometric frameworks for the RKD are provided by Card et al. (2015b) and Chiang and Sasaki (2019). These frameworks establish that conventional RKD estimands, such as those for means and quantiles, identify a weighted average of structural derivatives, with weights that vary across (sub)populations. However, the causal interpretation of these estimands within the potential outcomes framework (Rubin, 2005; Imbens and Rubin, 2015) remains unclear. Furthermore, the identification and inference of more general causal effects in this setting remain relatively underexplored.

This paper develops a unified framework for the identification and estimation of local treatment effects in sharp regression kink designs. Unlike the prior literature, which primarily seeks causal interpretations for specific RKD estimands, our approach begins by defining a broad class of causal parameters of interest and then derives the corresponding strategies for their identification and estimation. In a sharp RKD, the treatment B is a deterministic function of the running variable X , given by $B = b(X)$. The general causal effect at the treatment level $b_0 := b(x_0)$ is captured by the following parameter:

$$\Delta_\phi := \frac{\partial}{\partial b} \phi(F_{Y(b)|X=x_0}) \Big|_{b=b_0} := \lim_{\delta \rightarrow 0} \frac{\phi(F_{Y(b_0+\delta)|X=x_0}) - \phi(F_{Y(b_0)|X=x_0})}{\delta},$$

where ϕ is a functional on the space of one-dimensional distribution functions, and $Y(b)$ denotes the potential outcome under treatment level $b \in \mathbb{R}$. Typically, x_0 represents the *kink* point in this design, defined as a threshold where the derivative of $b(x)$ is discontinuous: $\lim_{x \rightarrow x_0^+} \frac{d}{dx} b(x) \neq \lim_{x \rightarrow x_0^-} \frac{d}{dx} b(x)$. The parameter Δ_ϕ measures the effect of an infinitesimal change in the treatment variable on a feature of the conditional potential outcome distribution $F_{Y(b)|X=x_0}$, evaluated at $b = b_0$. For instance, when ϕ is the mean functional, i.e., $\phi(F) = \mu(F) := \int w dF(w)$, Δ_ϕ corresponds to the “local average response” parameter identified in Card et al. (2015b).

We refer to the parameter Δ_ϕ as the *local treatment effect* (LTE). This terminology is analogous to that of Florens et al. (2008), who study the identification of the average treatment effect (ATE) in models with a continuous endogenous treatment. In their framework, the ATE at treatment level $b \in \mathbb{R}$ is defined as

$$\Delta^{\text{ATE}}(b) := \frac{\partial}{\partial b} \mu(F_{Y(b)}) := \lim_{\delta \rightarrow 0} \frac{E[Y(b+\delta)] - E[Y(b)]}{\delta}.$$

In standard settings, the unconditional distribution of a potential outcome, $F_{Y(b)}$, can often be identified for any $b \in \mathbb{R}$ using methods like instrumental variables or control functions (Florens

et al., 2008; Imbens and Newey, 2009). This distribution then serves as a foundation for identifying causal parameters such as $\Delta^{\text{ATE}}(b)$. However, such approaches are not suitable for the sharp RKD, where the treatment is a deterministic function of the running variable, precluding the existence of a valid external instrument. Instead, identification in a sharp RKD relies on the change in the slope of the treatment assignment function $b(x)$ at the kink x_0 . While the deterministic nature of treatment assignment is a known restriction, a challenge in identifying the derivative effect Δ_ϕ arises because the conditional distribution of the potential outcome $F_{Y(b_0+\delta)|X=x_0}$ remains unidentified under the kink assumption alone. Consequently, functionals of this conditional distribution, such as its mean or quantiles, cannot be directly recovered.

This paper proposes an identification strategy for the general local treatment effect Δ_ϕ in the sharp RKD, applicable to any Hadamard differentiable functional ϕ . The strategy utilizes standard assumptions from the RKD literature, primarily smoothness conditions on relevant structural functions and on the conditional distributions of the unobserved disturbance ε and the outcome Y given the running variable X (Card et al., 2015b; Chiang and Sasaki, 2019; Qu and Yoon, 2019). As a result, we derive a unified identification formula for Δ_ϕ and formally establish the relationship between this general class of LTEs and their corresponding conventional RKD estimands.

This paper develops a unified estimation and inference framework for two new classes of RKD estimands, which correspond to the LTEs for various smooth functionals. Motivated by the local polynomial constrained quantile smoother of Chiang and Sasaki (2019), we employ a local polynomial constrained regression RKD estimator and establish its asymptotic properties. This approach has two notable features. First, it explicitly imposes the continuity of the conditional distribution $F_{Y|X=x}$ at the kink point x_0 as a constraint. Second, unlike methods that require separate local polynomial regressions on each side of the threshold (Calonico et al., 2014), this estimator computes the required left- and right-derivatives in a single step, which can improve computational efficiency. Moreover, we develop uniform inference procedures for the proposed distributional RKD estimators using a multiplier bootstrap method. This inference framework applies to various LTEs. By integrating pivotal methods developed for quantiles by Chiang and Sasaki (2019), our framework covers effects on the mean, the distribution, quantiles, and inequality measures such as the Lorenz curve. To complete the methodology, we also provide a bandwidth selector based on mean squared error (MSE) minimization.

To illustrate the applicability of our framework, we analyze the effect of unemployment insurance (UI) benefits on the distribution of unemployment durations using the Continuous Wage and Benefit History (CWBH) data. This analysis extends previous empirical studies, which primarily focused on mean or quantile effects. In particular, we estimate the local distributional treatment effect (LDTE) to assess the marginal impact on the probability distribution, and the local Lorenz treatment effect (LLTE) to examine the effects on the dispersion and inequality of unemployment durations.

The remainder of this paper is organized as follows. The next subsection outlines our contri-

butions to the literature. Section 2 introduces the model and our parameters of interest. Section 3 establishes our main identification results. Section 4 provides further discussion and a detailed comparison with related literature. Section 5 develops the estimation and inference framework and derives its asymptotic properties. Section 6 presents an empirical application and evaluates the finite-sample performance of the estimators via Monte Carlo simulations. Section 7 concludes. Appendix A contains the technical assumptions and auxiliary lemmas, while Appendix B details the bandwidth selection procedure. All mathematical proofs are provided in the online Supplementary Material.

Contribution to the Literature

This paper contributes to the theoretical literature on regression kink designs (Card et al., 2015b; Dong, 2018; Chiang and Sasaki, 2019; Chen and Fan, 2019; Chen et al., 2020). The main contribution of this paper is the development of a unified framework that clarifies the causal interpretation of a broad class of sharp RKD estimands within the potential outcomes model. This framework serves two key purposes. First, it provides a unified identification formula for the general causal parameter Δ_ϕ . Second, by nesting existing estimands as special cases, it refines and elucidates their causal interpretations. To illustrate the latter, consider the conventional quantile regression kink design (QRKD) estimand:

$$\begin{aligned} \text{QRKD}(\tau) &:= \frac{\lim_{x \rightarrow x_0^+} \frac{\partial}{\partial x} Q_{Y|X}(\tau|x) - \lim_{x \rightarrow x_0^-} \frac{\partial}{\partial x} Q_{Y|X}(\tau|x)}{b'(x_0^+) - b'(x_0^-)} \\ &= \frac{\lim_{x \rightarrow x_0^+} \frac{\partial}{\partial x} Q_\tau(F_{Y(b(x))|X=x}) - \lim_{x \rightarrow x_0^-} \frac{\partial}{\partial x} Q_\tau(F_{Y(b(x))|X=x})}{\lim_{x \rightarrow x_0^+} \frac{d}{dx} b(x) - \lim_{x \rightarrow x_0^-} \frac{d}{dx} b(x)}. \end{aligned}$$

It is known that $\text{QRKD}(\tau)$ lacks a direct interpretation as a standard quantile treatment effect (QTE). Chiang and Sasaki (2019) show that $\text{QRKD}(\tau)$ identifies a weighted average of marginal effects across a subpopulation on the boundary $\partial\mathcal{V}(y_\tau, x_0) := \{e \in \mathbb{R}^{d_\varepsilon} : g(b_0, x_0, e) = y_\tau\}$, where $y_\tau := Q_{Y|X}(\tau|x_0)$. This paper shows that, under a continuity condition on the conditional density of the outcome Y given the running variable X (Assumption S4), the QRKD estimand identifies the local quantile treatment effect (LQTE) for the subpopulation at the kink point $X = x_0$:

$$\text{QRKD}(\tau) = \Delta_Q(\tau) := \lim_{\delta \rightarrow 0} \frac{Q_{Y(b_0+\delta)|X}(\tau|x_0) - Q_{Y(b_0)|X}(\tau|x_0)}{\delta}.$$

This continuity assumption, while commonly used for developing asymptotic theory in the RKD and related literature (e.g., Chiang and Sasaki, 2019; Qu and Yoon, 2019), has not previously been applied for identification purposes. The identifying conditions are discussed in greater detail in Section 4.

Second, our estimation and inference framework, based on local polynomial constrained regres-

sion, extends the framework of [Chiang et al. \(2019\)](#) for local Wald estimands. Their approach cannot directly accommodate certain causal parameters, particularly those defined by smooth inequality functionals such as the Lorenz curve and the Gini coefficient, because these parameters lack a local Wald structure. In contrast, our framework addresses this limitation. We show that such LTEs can be expressed as smooth transformations of foundational distributional and quantile RKD estimands. Based on this result, they can be estimated by applying the corresponding transformations to the foundational estimators. Furthermore, the uniform Bahadur representations established for the foundational estimators, combined with the functional Delta method, provide the basis for valid uniform inference on the class of resulting LTEs.

Finally, this paper contributes to the literature on identifying continuous causal effects with endogenous treatments ([Florens et al., 2008](#); [Imbens and Newey, 2009](#); [Kasy, 2014](#); [Hoderlein et al., 2017](#)). Within this literature, identification often requires recovering the full conditional distribution of potential outcomes. However, this step typically relies on structural assumptions, such as separability ([Florens et al., 2008](#)), the existence of control functions ([Imbens and Newey, 2009](#)), or monotonicity conditions ([Kasy, 2014](#); [Hoderlein et al., 2017](#)). Our framework provides an alternative identification approach. We establish a structural equivalence between the LTE and the sharp RKD analogue of the local average structural derivative (LASD) proposed by [Hoderlein and Mammen \(2007, 2009\)](#), as formalized in Lemma 1. This approach is analogous to that of [Chernozhukov et al. \(2015\)](#), who exploit a similar connection to a panel version of the LASD to identify quantile derivative effects in nonseparable panel models. Taken together, our work and these related studies suggest that even when identification of the full distribution of potential outcomes is not readily available, causal derivative effects can often be identified by establishing a link to an appropriate version of the LASD.

2 MODEL AND TARGET PARAMETERS

We observe a random sample $\{(Y_i, B_i, X_i)\}_{i=1}^n$, where Y_i is the outcome, B_i is the treatment, and X_i is the running variable. We assume the data are drawn independently and identically distributed (i.i.d.) from a population where these variables are continuous. For notational simplicity, we drop the subscript i when discussing the population model. Let $\mathcal{Y} \subset \mathbb{R}$, $\mathcal{B} \subset \mathbb{R}$, and $\mathcal{X} \subset \mathbb{R}$ denote the supports of Y , B , and X , respectively.

Assumption 1 (Nonseparable Model). *Let $g : \mathbb{R} \times \mathcal{X} \times \mathcal{E} \rightarrow \mathbb{R}$ be a measurable function.*

(i) *The potential outcome $Y(b)$ for any treatment level $b \in \mathbb{R}$ is given by*

$$Y(b) = g(b, X, \varepsilon)$$

where $\varepsilon \in \mathcal{E} \subset \mathbb{R}^{d_\varepsilon}$ is a vector representing unobserved individual heterogeneity.

(ii) *Let the Stable Unit Treatment Value Assumption (SUTVA) hold, which implies $Y = Y(B)$. Given the deterministic treatment assignment rule $B = b(X)$, the observed outcome is thus*

determined by

$$Y = g(b(X), X, \varepsilon).$$

Let I_{x_0} be a closed interval containing the kink point x_0 , and define $I_{x_0}^o := I_{x_0} \setminus \{x_0\}$.

Assumption 2 (Sharp Kink Characterization in the First Stage). *The treatment assignment rule $b(\cdot)$ is continuous on I_{x_0} and continuously differentiable on $I_{x_0}^o$. This implies continuity at the kink point, $b(x_0^+) = b(x_0^-) = b(x_0)$, but a discontinuity in the first-order derivative, $b'(x_0^+) \neq b'(x_0^-)$, where $b'(x_0^\pm) := \lim_{x \rightarrow x_0^\pm} \frac{d}{dx} b(x)$.*

Assumptions 1 and 2 define the nonparametric, nonseparable structural model that forms the basis of our analysis. Assumption 1(i) specifies the general nonseparable form of the potential outcome. Assumption 1(ii) imposes a standard condition from the causal inference literature: the observed outcome Y is determined by the treatment B , which itself is generated by a deterministic treatment selection rule. This setup is consistent with the canonical RKD framework (Card et al., 2015b; Chiang and Sasaki, 2019). Assumption 2 formally imposes the sharp kink on the treatment selection function, $b(\cdot)$. This requires that $b(x)$ is continuous at the kink point x_0 , while its first derivative, $b'(x)$, exhibits a jump discontinuity at that point.

Within this model, we focus on the following causal parameter.

Definition 1 (Local Treatment Effect). Let \mathcal{F} denote the space of all one-dimensional distribution functions. For any functional $\phi : \mathcal{F} \rightarrow \mathbb{R}$, the *local treatment effect at the kink* (LTE) is defined as

$$\Delta_\phi := \frac{\partial}{\partial b} \phi(F_{Y(b)|X=x_0}) \Big|_{b=b_0} := \lim_{\delta \rightarrow 0} \frac{\phi(F_{Y(b_0+\delta)|X=x_0}) - \phi(F_{Y(b_0)|X=x_0})}{\delta}$$

provided the limit exists, where $b_0 := b(x_0)$.

The LTE parameter Δ_ϕ has a clear causal interpretation. Under the structural model defined by Assumptions 1–2, its definition is equivalent to

$$\lim_{\delta \rightarrow 0} \frac{\phi(F_{g(b_0+\delta, x_0, \varepsilon)|X=x_0}) - \phi(F_{g(b_0, x_0, \varepsilon)|X=x_0})}{\delta}.$$

This derivative captures the marginal effect of the treatment B at its baseline level b_0 on the feature of the potential outcome distribution characterized by the functional ϕ . The fundamental identification problem, however, is that the key component of this definition—the counterfactual distribution $F_{Y(b_0+\delta)|X=x_0}$ —is unobserved, since the structural function g is unknown.

We conclude this section with several key examples of the LTE Δ_ϕ .

Example 1 (Average Effect). When ϕ is the mean functional, $\phi(F) = \mu(F) := \int w dF(w)$, Δ_ϕ

becomes the local average treatment effect (LATE):

$$\Delta_\mu := \frac{\partial}{\partial b} \mu(F_{Y(b)|X=x_0})|_{b=b_0} := \lim_{\delta \rightarrow 0} \frac{E[Y(b_0 + \delta)|X = x_0] - E[Y(b_0)|X = x_0]}{\delta}.$$

Example 2 (Distributional and Quantile Effects). When ϕ is the identity mapping at a point y , which we denote by $\phi(F) = Id_y(F) := F(y)$, Δ_ϕ becomes the local distributional treatment effect (LDTE). This measures the marginal effect on the cumulative distribution function:

$$\Delta_{Id(y)} := \frac{\partial}{\partial b} Id_y(F_{Y(b)|X=x_0})|_{b=b_0} := \lim_{\delta \rightarrow 0} \frac{F_{Y(b_0+\delta)|X}(y|x_0) - F_{Y(b_0)|X}(y|x_0)}{\delta}.$$

When ϕ is the τ th quantile functional, $\phi(F) = Q_\tau(F) := \inf\{w : F(w) \geq \tau\}$ for $\tau \in (0, 1)$, Δ_ϕ becomes the local quantile treatment effect (LQTE):

$$\Delta_Q(\tau) := \frac{\partial}{\partial b} Q_\tau(F_{Y(b)|X=x_0})|_{b=b_0} := \lim_{\delta \rightarrow 0} \frac{Q_{Y(b_0+\delta)|X}(\tau|x_0) - Q_{Y(b_0)|X}(\tau|x_0)}{\delta},$$

where $Q_{Y(b)|X}(\tau|x)$ is a shorthand for $Q_\tau(F_{Y(b)|X=x})$.

Example 3 (Lorenz Effect). Our framework also covers inequality measures. For instance, if ϕ corresponds to the Lorenz curve, $L_\tau(F) := \int_0^\tau Q_u(F) du / \mu(F)$, then Δ_ϕ is the local Lorenz treatment effect (LLTE):

$$\Delta_L(\tau) := \frac{\partial}{\partial b} L_\tau(F_{Y(b)|X=x_0})|_{b=b_0} := \lim_{\delta \rightarrow 0} \frac{L_\tau(F_{Y(b_0+\delta)|X=x_0}) - L_\tau(F_{Y(b_0)|X=x_0})}{\delta}.$$

3 IDENTIFICATION

The definition of the LTE Δ_ϕ as a derivative makes the class of Hadamard differentiable functionals a natural choice for our analysis. This class is particularly relevant as it includes most statistical measures used in policy evaluation, such as means, quantiles, and various inequality indices. To formalize this concept, let $\ell^\infty(\mathbb{T})$ denote the space of all bounded functions defined on a set \mathbb{T} , and let \mathbb{B} be a Banach space. A functional $\phi : \mathcal{F} \subset \ell^\infty(\mathbb{R}) \rightarrow \mathbb{B}$ is called *Hadamard differentiable* at $F \in \mathcal{F}$ tangentially to a set $\mathbb{D}_0 \subseteq \ell^\infty(\mathbb{R})$, if there exists a continuous linear map $\phi'_F : \ell^\infty(\mathbb{R}) \rightarrow \mathbb{B}$ such that

$$\lim_{\delta \rightarrow 0} \left\| \frac{\phi(F + \delta h_\delta) - \phi(F)}{\delta} - \phi'_F(h) \right\|_{\mathbb{B}} = 0$$

holds for all sequences $\{h_\delta\}$ satisfying $h_\delta \rightarrow h \in \mathbb{D}_0$ and $F + \delta h_\delta \in \mathcal{F}$ for all sufficiently small δ . The map ϕ'_F is the Hadamard derivative of ϕ at F tangentially to $\mathbb{D}_0 \subseteq \ell^\infty(\mathbb{R})$. Building on this definition, we impose the following assumption on the functional ϕ .

Assumption S1 (Smooth Functional). *The functional $\phi : \mathcal{F} \rightarrow \mathbb{R}$ is Hadamard differentiable at $F_{Y|X=x_0}$, with its Hadamard derivative denoted by $\phi'_{F_{Y|X=x_0}}$.*

The smooth structural model is specified by the following assumption.

Assumption S2 (Smooth Structural Functions). *The structural function $g(\cdot, \cdot, e)$ is continuously differentiable for all $e \in \mathcal{E}$, with its partial derivatives with respect to the first and second arguments denoted by $g_1(\cdot, \cdot, e)$ and $g_2(\cdot, \cdot, e)$, respectively.*

Assumption S2, which aligns with Assumptions 1(ii) and 2 in Card et al. (2015b), imposes smoothness on the structural function, requiring that the outcome varies continuously with both the treatment and the running variable. Notably, this assumption is weaker than Assumption 2(i) of Chiang and Sasaki (2019), as we do not impose continuous differentiability with respect to the unobserved heterogeneity ε . Under this smoothness condition, the partial derivative of the function $h(x, e) := g(b(x), x, e)$ with respect to x is given by

$$\frac{\partial}{\partial x} h(x, e) = b'(x)g_1(b(x), x, e) + g_2(b(x), x, e).$$

A key implication follows: although the structural partial derivatives, $x \mapsto g_1(b(x), x, e)$ and $x \mapsto g_2(b(x), x, e)$, are continuous at x_0 , the function $x \mapsto h(x, e)$ is not continuously differentiable at this point because $b'(x)$ has a jump discontinuity.

Our identification strategy requires the following smoothness conditions on the conditional distributions given the running variable.

Assumption S3 (Smooth Disturbance Distributions). *The conditional distribution of ε given X is absolutely continuous with respect to Lebesgue measure. Its conditional density, $f_{\varepsilon|X}(e|\cdot)$, is continuously differentiable on the interval I_{x_0} for all $e \in \mathcal{E}$.*

Assumption S4 (Smooth Outcome Distribution). *The conditional distribution of Y given X is absolutely continuous with respect to Lebesgue measure. Furthermore, for each $\tau \in (0, 1)$, the conditional density $f_{Y|X}(Q_{Y|X}(\tau|\cdot)|\cdot)$ is continuous and strictly positive on I_{x_0} .*

Assumption S3 is used to address selection bias in the identification strategy and is analogous to Assumption 2(iii) in Chiang and Sasaki (2019). A set of sufficient conditions for Assumption S3 is provided in Assumption S3' (Section 4), which follows the formulation used by Card et al. (2015b).

Assumption S4 facilitates the derivation of the unified identification formula. A contribution of this paper relates to the use of this assumption for identification purposes. In the existing regression discontinuity design (RDD) and RKD literature, the continuity of the conditional density $f_{Y|X}$ has been used primarily for inference—specifically, to derive the asymptotic properties of estimators (e.g., Qu and Yoon, 2019; Chiang and Sasaki, 2019). In contrast, our framework is the first to exploit this condition to achieve point identification. Finally, it is important to note that these two assumptions are not redundant, as the continuity of $f_{\varepsilon|X}$ (Assumption S3) does not, in general, imply the continuity of $f_{Y|X}$.

The following lemma forms the cornerstone of our identification strategy, establishing a key link between the LTE and the causal parameter defined by the structural model. All regularity conditions, prefixed with ‘R’, are collected in Appendix A.

Lemma 1 (Structural Representation). *Suppose Assumptions 1–2, S1–S2, S4, and R1(i)–(ii) hold. Then,*

$$\Delta_\phi = \phi'_{F_{Y|X=x_0}} \left(E \left[-f_{Y|X}(\cdot|x_0) g_1(b_0, x_0, \varepsilon) | Y = \cdot, X = x_0 \right] \right).$$

Lemma 1 establishes that the LTE Δ_ϕ can be expressed as the Hadamard derivative of ϕ applied to a conditional expectation. The term inside this expectation is the structural derivative of interest, g_1 , weighted by the negative conditional density of the outcome, $-f_{Y|X}$. This representation provides the link to identification. While the LTE is defined using unobserved potential outcomes, the lemma expresses it in terms of two key components: (i) an estimable weight function, and (ii) the conditional expectation of the structural derivative, $E[g_1(b_0, x_0, \varepsilon) | Y = \cdot, X = x_0]$. The second component, which is the object of identification, is analogous to the *local average structural derivative* (LASD) proposed by Hoderlein and Mammen (2007, 2009).

Building on Lemma 1, the following theorem establishes our main identification result for the LTE.

Theorem 1 (Identification of the LTE). *Under Assumptions 1–2, S1–S4, and R1, the local treatment effect is identified as:*

$$\Delta_\phi = \phi'_{F_{Y|X=x_0}} (\text{DRKD}(\cdot)),$$

where $\text{DRKD}(\cdot)$ is the *distributional RKD (DRKD) estimand*, defined by:

$$\text{DRKD}(y) := \frac{\frac{\partial}{\partial x} F_{Y|X}(y|x_0^+) - \frac{\partial}{\partial x} F_{Y|X}(y|x_0^-)}{b'(x_0^+) - b'(x_0^-)}.$$

Theorem 1 shows that the LTE, Δ_ϕ , is identified by applying the Hadamard derivative of the functional ϕ in the direction of the DRKD estimand. This DRKD estimand is a local Wald-type ratio constructed from the derivatives of the observable conditional distribution on either side of the kink. Furthermore, this result provides a causal interpretation for a specific case of the generalized local Wald ratio estimand proposed by Chiang et al. (2019), defined as

$$\begin{aligned} & \tau^{(v)}(\theta''; \mathcal{R}, \varphi, \psi) \\ &:= \mathcal{R} \left(\frac{\varphi \left(\lim_{x \rightarrow x_0^+} \frac{\partial^v E[h_1(Y, \cdot) | X=x]}{\partial x^v} \right) (\cdot) - \varphi \left(\lim_{x \rightarrow x_0^-} \frac{\partial^v E[h_1(Y, \cdot) | X=x]}{\partial x^v} \right) (\cdot)}{\psi \left(\lim_{x \rightarrow x_0^+} \frac{\partial^v E[h_2(B, \cdot) | X=x]}{\partial x^v} \right) (\cdot) - \psi \left(\lim_{x \rightarrow x_0^-} \frac{\partial^v E[h_2(B, \cdot) | X=x]}{\partial x^v} \right) (\cdot)} \right) (\theta''). \end{aligned}$$

To establish the connection, setting the tuning parameters to $v = 1$, $(\mathcal{R}, \varphi, \psi) = (\phi'_{F_{Y|X=x_0}}, Id, Id)$,

$h_1(Y, \cdot) = \mathbb{1}(Y \leq \cdot)$, and $h_2(B, \cdot) = B = b(X)$ yields the following form for their generalized estimand:

$$\tau^{(1)}(\theta''; \phi'_{F_{Y|X=x_0}}, Id, Id) = \phi'_{F_{Y|X=x_0}} \left(\frac{\frac{\partial}{\partial x} F_{Y|X}(\cdot|x_0^+) - \frac{\partial}{\partial x} F_{Y|X}(\cdot|x_0^-)}{b'(x_0^+) - b'(x_0^-)} \right) (\theta'')$$

Therefore, under the conditions of our theorem, the estimand $\tau^{(1)}(\theta''; \phi'_{F_{Y|X=x_0}}, Id, Id)$ admits a causal interpretation as the local treatment effect, Δ_ϕ .

We conclude this section by applying Theorem 1 to the functionals introduced earlier. For detailed derivations, we refer interested readers to Appendix S.1 in the supplementary material.

Example 1 (Average Effect, Continued). The mean functional, $\mu(F) = \int w dF(w)$, is linear. Consequently, its Hadamard derivative is the functional itself: $\mu'_F(h) = \int w dh(w)$. Substituting this into the general formula from Theorem 1 and applying integration by parts yields the familiar RKD estimand for the mean effect:

$$\Delta_\mu = \frac{\frac{d}{dx} E[Y|X = x_0^+] - \frac{d}{dx} E[Y|X = x_0^-]}{b'(x_0^+) - b'(x_0^-)} =: \text{MRKD}. \quad (1)$$

This result is consistent with the causal interpretation for the MRKD estimand in [Card et al. \(2015b\)](#), which we compare in detail in Section 4.1.

Example 2 (Distributional and Quantile Effects, Continued). For the identity mapping $\phi = Id_y$, Theorem 1 directly yields the identification of the local distributional treatment effect (LDTE):

$$\Delta_{Id}(y) = \frac{\frac{\partial}{\partial x} F_{Y|X}(y|x_0^+) - \frac{\partial}{\partial x} F_{Y|X}(y|x_0^-)}{b'(x_0^+) - b'(x_0^-)} =: \text{DRKD}(y)$$

for all $y \in \mathcal{Y}_{x_0}$, where \mathcal{Y}_x denotes the support of the conditional distribution $F_{Y|X=x}$. This result provides a direct causal interpretation for the DRKD estimand: it identifies the marginal effect of the treatment on the potential outcome's conditional distribution $F_{Y(b)|X=x_0}$ at the baseline treatment level b_0 .

For the quantile functional $\phi = Q_\tau$, Hadamard differentiability is guaranteed under certain regularity conditions (see, e.g., Lemma 21.4 of [Van der Vaart \(2000\)](#)). Applying Theorem 1 then yields the identification of the local quantile treatment effect (LQTE):

$$\Delta_Q(\tau) = \frac{\frac{\partial}{\partial x} Q_{Y|X}(\tau|x_0^+) - \frac{\partial}{\partial x} Q_{Y|X}(\tau|x_0^-)}{b'(x_0^+) - b'(x_0^-)} =: \text{QRKD}(\tau)$$

for all $\tau \in (0, 1)$. This result is a key contribution of our paper. It establishes that $\text{QRKD}(\tau)$ has a direct and intuitive causal interpretation as the marginal effect on the potential outcome quantile. This contrasts with the complex weighted average interpretation in [Chiang and Sasaki \(2019\)](#), which we discuss in detail in Section 4.2.

Example 3 (Lorenz Effect, Continued). Our framework extends naturally to inequality measures. Consider the Lorenz curve functional, $L_\tau(F) := \int_0^\tau Q_u(F) du / \mu(F)$. Under standard regularity conditions (e.g., Proposition 2 in [Bhattacharya \(2007\)](#)), its Hadamard derivative is given by:

$$[L_\tau]'_F(h) = \frac{\int_0^\tau [Q_u]'_F(h) du}{\mu(F)} - \frac{L_\tau(F)}{\mu(F)} \cdot \mu(h).$$

Substituting this derivative into the general identification result from Theorem 1 yields the local Lorenz treatment effect (LLTE):

$$\Delta_L(\tau) = \frac{1}{\mu_0} \left(\int_0^\tau \text{QRKD}(u) du - L_{Y|X}(\tau|x_0) \cdot \text{MRKD} \right) \quad (2)$$

for all $\tau \in (0, 1)$, where $\mu_0 := E[Y|X = x_0]$ and $L_{Y|X}(\tau|x_0) := L_\tau(F_{Y|X=x_0})$ is the baseline conditional Lorenz curve. This outcome illustrates the structure of the framework: the treatment effect on the Lorenz curve (LLTE) is identified as a function involving the effects on the mean (MRKD) and quantiles (the integrated QRKD).

4 COMPARISON AND FURTHER DISCUSSION

This section presents a detailed comparison between our main identification results and the existing literature on RKD designs. We focus in particular on the foundational contributions of [Card et al. \(2015b\)](#) and [Chiang and Sasaki \(2019\)](#), and conclude by discussing several extensions of our identification framework.

4.1 Average Effect

[Card et al. \(2015b\)](#) propose a set of sufficient conditions for the causal interpretation of the mean RKD estimand. To facilitate a comparison with our framework, we now summarize their key identifying assumptions—specifically, those required in addition to our shared Assumptions 1–2 and S2—along with their main identification result.

Assumption S3'.

- (i) The support $\mathcal{E} \subset I_\varepsilon$ is bounded for some large compact set $I_\varepsilon \subset \mathbb{R}^{d_\varepsilon}$.
- (ii) The set $\mathcal{A}_\varepsilon := \{e \in \mathbb{R}^{d_\varepsilon} : f_{X|\varepsilon}(x|e) > 0, \forall x \in I_{x_0}\}$ has a positive measure under ε : $\int_{\mathcal{A}_\varepsilon} dF_\varepsilon(e) > 0$.
- (iii) The conditional density $f_{X|\varepsilon}(\cdot|e)$ is continuously differentiable on I_{x_0} for all $e \in I_\varepsilon$.

Lemma 2 (Card et al., 2015b, Proposition 1). *Suppose Assumptions 1–2, S2, and S3' hold. Then,*

$$\text{MRKD} = E[g_1(b_0, x_0, \varepsilon) | X = x_0] = \int \omega_{x_0}(e) \cdot g_1(b_0, x_0, \varepsilon) dF_\varepsilon(e),$$

where $\omega_x(e) := f_{X|\varepsilon}(x|e)/f_X(x)$.

We first note that our main conclusion is consistent with Card et al. (2015b). Under standard regularity conditions that permit interchanging limits and expectations, our LTE for the mean simplifies to:

$$\Delta_\mu = E[g_1(b_0, x_0, \varepsilon) | X = x_0].$$

This is precisely the parameter identified by Card et al. (2015b), confirming that both frameworks target the same causal object for the mean effect.

Despite this consistency, our identifying assumptions differ in two key respects. First, embedding the mean case within our unified framework requires Assumptions S1 and S4. While not strictly necessary for identifying the mean effect in isolation, these assumptions are needed for the generality of a framework covering all Hadamard differentiable functionals. Second, our framework utilizes Assumption S3—which directly requires the continuous differentiability of the conditional density $x \mapsto f_{\varepsilon|X}(e|x)$ —to establish the causal interpretation of MRKD as the LATE, Δ_μ . While Assumption S3' provides sufficient conditions for S3 (as shown in their Lemma 2), it is specifically employed by Card et al. (2015b) to enable the population-weighted average interpretation by addressing the measure-zero conditioning problem. In particular, S3' allows one to rigorously handle conditioning on the measure-zero event $\{X = x_0\}$ and to express Δ_μ as a weighted average of the structural derivative over the entire distribution of unobserved heterogeneity. The resulting weight function, $\omega_{x_0}(e)$, is thus well defined and reflects the relative likelihood that an individual with type $\varepsilon = e$ is located at the kink point (Card et al., 2015b, Remark 1).

4.2 Quantile Effect

We now provide a detailed comparison between our framework and that of Chiang and Sasaki (2019). The two approaches differ significantly in their starting point. Our analysis addresses a “forward problem”: we first define a causal effect of interest (Δ_Q) and then derive an identification strategy for it. In contrast, Chiang and Sasaki (2019) address an “inverse problem”: they start with the conventional estimand $\text{QRKD}(\tau)$, and investigate what causal interpretation, if any, it admits. To facilitate this comparison, we first summarize their main assumptions and results

The identification strategy of Chiang and Sasaki (2019) relies on geometric concepts. Let $h(x, e) := g(b(x), x, e)$ be the reduced-form outcome function. They define a volume in the

space of unobservables as $\mathcal{V}(y, x) := \{e \in \mathbb{R}^{d_\varepsilon} : h(x, e) \leq y\}$, with its boundary denoted by $\partial\mathcal{V}(y, x) := \{e \in \mathbb{R}^{d_\varepsilon} : h(x, e) = y\}$. Using the Hausdorff probability measure developed in [Sasaki \(2015\)](#), they derive a reduced-form expression for the key object of their analysis: the conditional quantile partial derivative, $\frac{\partial}{\partial x} Q_{Y|X}(\tau|x)$. The Hausdorff probability measure on the boundary $\partial\mathcal{V}(y, x)$ is defined differently depending on the dimension M of the unobserved heterogeneity. For $M > 1$, the measure is defined as the ratio of weighted surface areas:

$$P_{y,x}^{M-1}(S) := \frac{\int_S \frac{1}{\|\nabla_e h(x, e)\|} f_{\varepsilon|X}(e|x) dH^{M-1}(e)}{\int_{\partial\mathcal{V}(y,x)} \frac{1}{\|\nabla_e h(x, e)\|} f_{\varepsilon|X}(e|x) dH^{M-1}(e)}$$

for any set S in the collection of Borel subsets of $\partial\mathcal{V}(y, x)$. For $M = 1$, the boundary $\partial\mathcal{V}(y, x)$ is typically a discrete set of points. The Hausdorff measure H^0 becomes the counting measure, yielding the discrete probability measure:

$$P_{y,x}^0(\{e\}) := \frac{\frac{1}{\|\nabla_e h(x, e)\|} f_{\varepsilon|X}(e|x)}{\sum_{e \in \partial\mathcal{V}(y,x)} \frac{1}{\|\nabla_e h(x, e)\|} f_{\varepsilon|X}(e|x)}$$

for each point $e \in \partial\mathcal{V}(y, x)$.

Assumption S4' ([Chiang and Sasaki, 2019](#), Assumptions 2–3).

- (i) *The boundary $\partial\mathcal{V}(y, x)$ is a smooth manifold that can be parameterized by a mapping $\Pi_{y,x} : \Sigma \rightarrow \partial\mathcal{V}(y, x)$ satisfying Assumption 3 of [Chiang and Sasaki \(2019\)](#), where Σ is an $(M-1)$ -dimensional rectangle.*
- (ii) *$e \mapsto h(x, e)$ is continuously differentiable for all $x \in I_{x_0}$, and $\|\nabla_e h(x, e)\| \neq 0$ on $\partial\mathcal{V}(y, x)$ for all $(y, x) \in \mathcal{Y}_x \times I_{x_0}$*
- (iii) *$\int_{\partial\mathcal{V}(y,x)} f_{\varepsilon|X}(e|x) dH^{M-1}(e) > 0$ for all $(y, x) \in \mathcal{Y}_x \times I_{x_0}$.*

We now relate Assumption S4' to the conditions in [Chiang and Sasaki \(2019\)](#). Assumption S4'(i) is identical to their Assumption 3, while Assumptions S4'(ii) and (iii) correspond to parts of their Assumption 2. The remaining conditions in their Assumption 2 are already encompassed by our earlier Assumptions S2 and S3. It is worth noting that S4'(ii) is stronger than Assumption S2 in one respect, as it implies the existence and continuous differentiability of the function $e \mapsto g(b(x), x, e)$ —a condition not required under S2.

Under these conditions, [Chiang and Sasaki \(2019\)](#) exploit the sharp kink to derive a reduced-form expression for the QRKD estimand that is free from selection bias. This result is stated in the following lemma.

Lemma 3 ([Chiang and Sasaki, 2019](#), Theorem 1 and Corollary 1). *Suppose Assumptions 1–2, S2–S3, and S4' hold, and that the probability measure $P_{y,x}^{M-1}$ on $\partial\mathcal{V}(y, x)$ is well-defined for*

$(y, x) \in \mathcal{Y}_x \times I_{x_0}$.¹ Then,

$$\text{QRKD}(\tau) = E_{P_{y_\tau, x_0}^{M-1}} [g_1(b(x_0), x_0, \varepsilon)] = \int_{\partial\mathcal{V}(y_\tau, x_0)} g_1(b(x_0), x_0, e) dP_{y_\tau, x_0}^{M-1}(e),$$

where $y_\tau := Q_{Y|X}(\tau|x_0)$. If ε is a random scalar and $\partial\mathcal{V}(y_\tau, x_0)$ a singleton, then

$$\text{QRKD}(\tau) = g_1(b(x_0), x_0, \varepsilon(y_\tau, x_0)),$$

where $\varepsilon(y_\tau, x_0)$ denotes the sole element of $\partial\mathcal{V}(y_\tau, x_0)$.

The causal interpretation provided by [Chiang and Sasaki \(2019\)](#) differs significantly from ours. Their result shows that the QRKD estimand identifies a complex weighted average of the structural derivative, $E_{P_{y_\tau, x_0}^{M-1}} [g_1(b_0, x_0, \varepsilon)]$, where the weighting depends on the geometry of the structural function. In contrast, our Theorem 1 establishes that the same estimand identifies the local quantile treatment effect, $\Delta_Q(\tau)$, a direct marginal effect of the treatment on the potential outcome's conditional quantile $Q_{Y(b)|X}(\tau|x_0)$ at the baseline level b_0 . The divergence in results stems from a difference in the underlying identification assumptions. Interpreting $\text{QRKD}(\tau)$ as the local quantile treatment effect $\Delta_Q(\tau)$ requires Assumption S4. This assumption is less restrictive than Assumption S4', which is used in approaches addressing the measure-zero conditioning problem. Notably, Assumption S4 does not require certain conditions employed in those contexts, such as smoothness of the function $e \mapsto h(x, e)$ or conditions ensuring the Hausdorff probability measure is well-defined. The ability to obtain a direct causal interpretation under these less restrictive conditions is an advantage of our framework.

To clarify the relationship between Assumption S4 and more primitive conditions such as Assumption S4', we provide two sets of sufficient conditions under which S4 holds. These conditions involve smoothness restrictions on the conditional density $f_{\varepsilon|X}$ and the structural function g .

Assumption S5. Define

$$\tilde{f}(y, x) := \int_{\partial\mathcal{V}(y, x)} \frac{1}{\|\nabla_e h(x, e)\|} f_{\varepsilon|X}(e|x) dH^{M-1}(e).$$

Suppose $\tilde{f}(y_\tau(\cdot), \cdot)$ is continuous on I_{x_0} for all $\tau \in (0, 1)$.

Assumption S6.

- (i) The unobserved heterogeneity ε can be decomposed into two components, $\varepsilon = (\varepsilon_1, A)$, satisfying the following conditions: (a) The scalar component ε_1 is absolutely continuous given (A, X) . Its conditional density, $f_{\varepsilon_1|A, X}(e_1|a, x)$ is continuous in both e_1 and $x \in I_{x_0}$,

¹We omit the regularity assumptions that ensure $P_{y, x}^{M-1}$ is a well-defined probability measure and that justify the use of the dominated convergence theorem. Interested readers may refer to [Chiang and Sasaki \(2019\)](#) and [Sasaki \(2015\)](#) for details.

and is strictly positive for all a . (b) The vector component A is absolutely continuous given X . Its conditional density $f_{A|X}(a|x)$ is continuous in $x \in I_{x_0}$, and is strictly positive for all a .

- (ii) The function $h(x, e_1, a) := g(b(x), x, e_1, a)$ is continuously differentiable with respect to its second argument e_1 for all a and $x \in I_{x_0}$. Its partial derivative $\partial_{e_1} h$ satisfies $\partial_{e_1} h(x, e_1, a) \geq c > 0$ for some constant $c > 0$. Furthermore, $x \mapsto \partial_{e_1} h(x, e_1, a)$ is continuous for all e_1 and a .

Proposition 1. Suppose Assumptions 1–2, and S2 hold. Then, Assumption S4 is satisfied if either of the following two conditions holds: (a) Assumption S4' and S5 hold. (b) Assumption S6 holds.

Proposition 1 provides two distinct sets of primitive conditions under which Assumption S4 holds. The first set, Condition (a), follows the approach of Chiang and Sasaki (2019). It requires the structural function to be smooth with respect to the entire vector of unobserved heterogeneity but does not impose monotonicity. The second set, Condition (b), utilizes an alternative set of assumptions similar to those in Chernozhukov et al. (2015) and Hoderlein et al. (2017). It requires the existence of at least one continuously distributed scalar component within the unobserved heterogeneity vector, and that the structural function is smooth and monotonic with respect to this single scalar component, though not necessarily with respect to others.

4.2.1 An Alternative Weighted Interpretation

Our structural representation in Lemma 1 establishes a connection between the LQTE and a sharp RKD analogue of the local average structural derivative (LASD) from Hoderlein and Mammen (2007, 2009). Applying Lemma 1 with $\phi = Q_\tau$ yields:

$$\Delta_Q(\tau) = E \left[g_1(b_0, x_0, \varepsilon) | Y = y_\tau, X = x_0 \right].$$

While this provides a direct interpretation as a structural marginal effect for a specific subpopulation $\{Y = y_\tau, X = x_0\}$, this interpretation involves conditioning on an event of Lebesgue measure zero. Handling such conditioning rigorously presents known theoretical difficulties (Sasaki, 2015).

Two approaches can address the issue of conditioning on a measure-zero event. The first approach, following Chiang and Sasaki (2019), involves replacing our Assumption S4 with Assumption S4'. This leads to their weighted-average interpretation on the boundary $\partial\mathcal{V}(y_\tau, x_0)$, as given by Lemma 3. However, integrating this approach within our unified framework would necessitate imposing the additional Assumption S5 to ensure that Assumption S4 still holds. We propose an alternative approach, showing that the measure-zero issue can be addressed by replacing Assumption S4 with the following condition instead.

Assumption S7.

- (i) The set $\mathcal{A}_\varepsilon(\tau) := \{e \in \mathbb{R}^{d_\varepsilon} : f_{Y,X|\varepsilon}(y_\tau(x), x|e) > 0, \forall x \in I_{x_0}\}$ has a positive measure under F_ε for all $\tau \in (0, 1)$: $\int_{\mathcal{A}_\varepsilon(\tau)} dF_\varepsilon(e) > 0$.
- (ii) The conditional density $f_{Y,X|\varepsilon}(\cdot, \cdot|e)$ is continuous on $\mathcal{Y}_x \times I_{x_0}$ for all $e \in I_\varepsilon$.

Assumption S7 imposes two conditions on the conditional density of the observables (Y, X) given the unobserved heterogeneity ε . Part (i) requires this density to be strictly positive in a neighborhood of the kink point for a non-trivial subpopulation. Part (ii) requires this density to be continuous in the same neighborhood. Notably, Assumption S7, in conjunction with Assumption S3', provides sufficient conditions for Assumptions S3 and S4, which enable the derivation of the unified identification formula. The following proposition provides an alternative weighted causal interpretation for the QRKD estimand.

Proposition 2. Suppose Assumptions 1–2, S2, S3', S7 and R1 hold. Then,

$$\text{QRKD}(\tau) = E[g_1(b_0, x_0, \varepsilon)|Y = y_\tau, X = x_0] = \int \omega_{y_\tau, x_0}(e) \cdot g_1(b_0, x_0, e) dF_\varepsilon(e)$$

for all $\tau \in (0, 1)$, where the weight function is $\omega_{y, x}(e) := f_{Y, X|\varepsilon}(y, x|e)/f_{Y, X}(y, x)$.

This proposition provides two results. The first equality shows that the QRKD estimand corresponds to the sharp RKD version of the local average structural derivative. The second equality addresses the issue of conditioning on the measure-zero event $\{Y = y_\tau, X = x_0\}$ by expressing the LASD as a weighted average of the structural derivative, g_1 , over the entire population of unobserved heterogeneity. The weight function, $\omega_{y_\tau, x_0}(e)$, is well-defined under Assumptions S3' and S7 and represents the relative likelihood of observing an individual with type $\varepsilon = e$ having characteristics $\{Y = y_\tau, X = x_0\}$.

Remark 4.1 (The Role of Monotonicity). If the structural function $g(b, x, e)$ is monotonic in a scalar error term e , then for any given (y, x) , the boundary $\partial\mathcal{V}(y, x)$ consists of the unique point $h^{-1}(y, x)$. In this case, the conditional expectation simplifies to the structural derivative evaluated at this point:

$$\begin{aligned} E[g_1(b_0, x_0, \varepsilon)|Y = y_\tau, X = x_0] &= E[g_1(b_0, x_0, \varepsilon)|h(x_0, \varepsilon) = y_\tau, X = x_0] \\ &= g_1(b_0, x_0, h^{-1}(y_\tau, x_0)). \end{aligned}$$

This result is consistent with the special case in Chiang and Sasaki (2019), where the unobserved heterogeneity is one-dimensional ($M = 1$) and the boundary $\partial\mathcal{V}(y_\tau, x_0)$ is a singleton.

The following table summarizes the causal interpretations of conventional RKD estimands as established in the prior literature and in this paper.

Table 1: Causal Interpretations of Classic RKD Estimands

Estimands	Causal Interpretations	Conditions	Related Literature
DRKD(y)	$\Delta_{Id}(y)$	Assumptions 1–2, S2, S3, S4	This paper
MRKD	$\Delta_\mu = E[\omega_{x_0}(\varepsilon) \cdot g_1(b_0, x_0, \varepsilon)]$	Assumptions 1–2, S2, S3'	Card et al. (2015b)
QRKD(τ)	$\Delta_Q(\tau)$	Assumptions 1–2, S2, S3, S4	This paper
	$E_{P_{y_\tau, x_0}^{M-1}}[g_1(b_0, x_0, \varepsilon)]$	Assumptions 1–2, S2, S3, S4'	Chiang and Sasaki (2019)
	$E[\omega_{y_\tau, x_0}(\varepsilon) \cdot g_1(b_0, x_0, \varepsilon)]$	Assumptions 1–2, S2, S3', S7	This paper

¹ Role of Assumption S1 (Hadamard Differentiability): This assumption enables the derivation of the general identification formula in Theorem 1, applicable to any Hadamard differentiable functional. However, it may not be strictly required when analyzing a specific smooth functional if a direct identification argument is available for that particular case.

² Assumption Hierarchy: Under our baseline model for the sharp RKD (Assumptions 1–2 and S2), the following relationships hold between the various smoothness conditions discussed: (i) Assumption S3' \implies Assumption S3. (ii) Assumption S4' & S5 \implies Assumption S4. (iii) Assumptions S3' & S7 \implies S3 & S4.

5 ESTIMATION AND INFERENCE

5.1 Estimation Strategy

Our identification results in Section 3 show that local treatment effects can be expressed either as direct Wald-type ratios or as smooth transformations thereof. This section develops a unified estimation and inference framework to handle both cases. To this end, we define two general classes of RKD estimands.

Type 1. (Local Wald Ratio Estimand): The first class of estimands takes the form of a local Wald ratio of derivatives:

$$\begin{aligned} \text{RKD}_m(\theta) &:= \frac{m^{(1)}(\theta, x_0^+) - m^{(1)}(\theta, x_0^-)}{b'(x_0^+) - b'(x_0^-)} \\ &:= \frac{\lim_{x \rightarrow x_0^+} \frac{\partial}{\partial x} E[\varphi(Y, \theta) | X = x] - \lim_{x \rightarrow x_0^-} \frac{\partial}{\partial x} E[\varphi(Y, \theta) | X = x]}{\lim_{x \rightarrow x_0^+} \frac{d}{dx}(x) - \lim_{x \rightarrow x_0^-} \frac{d}{dx}(x)}, \quad \theta \in \Theta. \end{aligned}$$

Here, $m(\theta, x) := E[\varphi(Y, \theta) | X = x]$ is the conditional expectation of a known transformation $\varphi(\cdot, \theta)$ of the outcome, and $m^{(1)}(\theta, x)$ denotes its first-order derivative with respect to x . This foundational class subsumes many standard sharp RKD estimands. For instance, setting $\varphi(y, \theta) = \mathbb{1}(y \leq \theta)$ yields the DRKD estimand ($\text{RKD}_F = \text{DRKD}$) for identifying the LDTE. Similarly, setting $\varphi(y, \theta) = y$ yields the mean RKD estimand ($\text{RKD}_\mu = \text{MRKD}$) for identifying the LATE.

Type 2. (Composite Estimand): The second class of estimands consists of composite functionals of the Type 1 estimands and the quantile RKD estimand:

$$\text{RKD}_{\psi|m}(\theta') := \psi(\text{RKD}_m(\cdot), \text{QRKD}(\cdot))(\theta'), \quad \theta' \in \Theta'.$$

Here, ψ is a functional defined on a product of function spaces. This class is designed to handle

LTes that are identified via transformations of simpler effects. The local Lorenz treatment effect is a key example. By defining ψ as

$$\psi_L(g, h) := \frac{1}{\mu_0} \left(\int_0^\tau h(u) du - L_{Y|X}(\tau|x_0) \cdot g \right), \quad (3)$$

and choosing $g = \text{MRKD}$ and $h = \text{QRKD}$, the resulting estimand $\text{RKD}_{\psi_L|\mu}$ identifies the LLTE, Δ_L , where $\theta' = \tau$ and $\Theta' = (0, 1)$.

Under our assumptions, the function $x \mapsto E[\varphi(Y, \theta)|X = x]$ is continuous at the kink point $x = x_0$. This continuity holds for cases such as the conditional CDF $F_{Y|X}$. This continuity motivates a convenient estimation strategy that jointly estimates the function's level and its left- and right-hand derivatives within a single procedure. We therefore propose a p th-order local polynomial constrained regression estimator. For each θ in a compact set $\bar{\Theta} \subset \Theta$, we solve the following minimization problem:

$$\hat{\alpha}(\theta) = \arg \min_{\alpha \in \mathbb{R}^{2p+1}} \sum_{i=1}^n \left(\varphi(Y_i, \theta) - \bar{r}_p(X_i - x_0)^\top \alpha \right)^2 K \left(\frac{X_i - x_0}{h_{n,\theta}} \right), \quad \theta \in \bar{\Theta}, \quad (4)$$

where $K(\cdot)$ is a kernel function. The vector of regressors, $\bar{r}_p(u) := (1, u\delta_u^+, u\delta_u^-, \dots, u^p\delta_u^+, u^p\delta_u^-)^\top \in \mathbb{R}^{2p+1}$, is constructed to impose the continuity constraint, where $\delta_u^+ := \mathbb{1}(u \geq 0)$, and $\delta_u^- := \mathbb{1}(u < 0)$. It includes a single intercept but allows for different coefficients on the polynomial terms to the left ($u < 0$) and right ($u \geq 0$) of the kink. Then, we obtain

$$\hat{\alpha}(\theta) = \left(\hat{m}(\theta, x_0), \frac{\hat{m}^{(1)}(\theta, x_0^+)}{1!}, \frac{\hat{m}^{(1)}(\theta, x_0^-)}{1!}, \dots, \frac{\hat{m}^{(p)}(\theta, x_0^+)}{p!}, \frac{\hat{m}^{(p)}(\theta, x_0^-)}{p!} \right)^\top,$$

The required function and its left- and right-hand derivatives can thus be obtained as $\hat{m}(\theta, x_0) = \iota_1^\top \hat{\alpha}(\theta)$, $\hat{m}^{(1)}(\theta, x_0^+) = \iota_2^\top \hat{\alpha}(\theta)$, and $\hat{m}^{(1)}(\theta, x_0^-) = \iota_3^\top \hat{\alpha}(\theta)$, where ι_j is the j th standard basis vector (e.g., $\iota_2 = (0, 1, 0, \dots, 0)^\top$).

To estimate the quantile derivatives that form the numerator of the QRKD estimand, we follow [Chiang and Sasaki \(2019\)](#) and use a local p th-order polynomial constrained quantile regression. In our notation, the estimator is:

$$\hat{\beta}(\tau) = \arg \min_{\beta \in \mathbb{R}^{2p+1}} \sum_{i=1}^n \rho_\tau \left(Y_i - \bar{r}_p(X_i - x_0)^\top \beta \right) K \left(\frac{X_i - x_0}{h_{n,\tau}} \right), \quad \tau \in \mathcal{T}, \quad (5)$$

where $\mathcal{T} \subset (0, 1)$ is a closed interval; $\rho_\tau(u) = (\tau - \mathbb{1}(u < 0))u$ is the standard check function. The vector of regressors $\bar{r}_p(u)$ is the same as that used for the mean regression, thereby imposing the continuity of the conditional quantile function $x \mapsto Q_{Y|X}(\tau|x)$ at the kink. Let $Q_{Y|X}^{(\nu)}(\tau|x) := \frac{\partial^\nu}{\partial x^\nu} Q_{Y|X}(\tau|x)$ for an integer $1 \leq \nu \leq p$. Analogous to the mean regression case, the estimated coefficient vector $\hat{\beta}(\tau)$ provides estimates of the conditional quantile function and its scaled

derivatives at the kink:

$$\hat{\beta}(\tau) = \left(\hat{Q}_{Y|X}(\tau|x_0), \frac{\hat{Q}_{Y|X}^{(1)}(\tau|x_0^+)}{1!}, \frac{\hat{Q}_{Y|X}^{(1)}(\tau|x_0^-)}{1!}, \dots, \frac{\hat{Q}_{Y|X}^{(p)}(\tau|x_0^+)}{p!}, \frac{\hat{Q}_{Y|X}^{(p)}(\tau|x_0^-)}{p!} \right)^\top,$$

The required quantile function and its left- and right-hand derivatives can thus be obtained as $\hat{Q}_{Y|X}(\tau|x_0) = \iota_1^\top \hat{\beta}(\tau)$, $\hat{Q}_{Y|X}^{(1)}(\tau|x_0^+) = \iota_2^\top \hat{\beta}(\tau)$, and $\hat{Q}_{Y|X}^{(1)}(\tau|x_0^-) = \iota_3^\top \hat{\beta}(\tau)$.

Remark 5.1 (On the Constrained Regression Approach). Our constrained, one-step estimation approach presents several advantages over the common practice of fitting separate local polynomial regressions on each side of the kink (e.g., [Calonico et al., 2014](#); [Card et al., 2015b](#)). First, it is computationally efficient, particularly when estimating quantile effects over a large grid of quantiles. Second, it directly incorporates the known continuity of the underlying function (e.g., $m(\theta, x)$ or $Q_{Y|X}(\tau|x)$) at the kink point—the same continuity property used in the asymptotic theory. The concept of a constrained estimation approach has been noted in other contexts, for example by [Chen et al. \(2020\)](#) for quantile treatment effects in a fuzzy RKD setting.

It is important, however, to recognize the limitation of this approach. Our constrained estimators are designed specifically for regression kink designs, where the underlying conditional functions are continuous. They are not applicable to standard RDDs, as the primary identifying assumption in RDDs involves a discontinuity in these functions at the threshold (i.e., $m(\theta, x_0^+) \neq m(\theta, x_0^-)$).

Remark 5.2. The conditional quantile function is theoretically strictly increasing in τ . This follows from its derivative, $\frac{\partial}{\partial \tau} Q_{Y|X}(\tau|x_0) = 1/f_{Y|X}(Q_{Y|X}(\tau|x_0)|x_0)$, which is positive under Assumption S4. While the true quantile function is monotonic, its empirical counterpart, $\hat{Q}_{Y|X}(\tau|x_0)$, may fail to exhibit monotonicity over the full range of τ due to sampling variation. To ensure a properly specified estimate, we apply the standard rearrangement procedure of [Chernozhukov et al. \(2010, Equations \(2.1\)–\(2.2\)\)](#). This procedure transforms the potentially non-monotonic estimate into a monotonically increasing function that preserves the quantile properties.

We now detail the application of the general estimation framework developed above to the specific local treatment effects discussed previously.

Example 1 (Average Effect, Continued). To estimate the LATE, Δ_μ , we apply our Type 1 estimator, the local polynomial constrained regression defined in (4). The transformation function is the identity, $\varphi(Y, \theta) = Y$, and the parameter θ is degenerate. The conditional moment function simplifies to $m(x, \theta) = E[Y|X = x] =: \mu_{Y|X}(x)$. The estimated coefficient vector, $\hat{\alpha}$, thus provides estimates of the conditional mean and its derivatives at the kink point:

$$\hat{\alpha} = \left(\hat{\mu}_0, \frac{\hat{\mu}_{Y|X}^{(1)}(x_0^+)}{1!}, \frac{\hat{\mu}_{Y|X}^{(1)}(x_0^-)}{1!}, \dots, \frac{\hat{\mu}_{Y|X}^{(p)}(x_0^+)}{p!}, \frac{\hat{\mu}_{Y|X}^{(p)}(x_0^-)}{p!} \right)^\top.$$

The required left- and right-hand derivatives can be obtained from $\hat{\alpha}$, yielding the estimator for the LATE:

$$\hat{\Delta}_\mu = \frac{\hat{\mu}_{Y|X}^{(1)}(x_0^+) - \hat{\mu}_{Y|X}^{(1)}(x_0^-)}{b'(x_0^+) - b'(x_0^-)} = \widehat{\text{RKD}}_\mu.$$

Example 2 (Distributional and Quantile Effects, Continued). We now detail the estimation of the distributional and quantile effects. First, the LQTE, $\Delta_Q(\tau)$, is estimated directly using the local polynomial constrained quantile regression from (5). As shown previously, the estimated coefficient vector $\hat{\beta}(\tau)$ yields the required left- and right-hand derivatives, which are then used to form the QRKD estimand:

$$\hat{\Delta}_Q(\tau) = \frac{\hat{Q}_{Y|X}^{(1)}(\tau|x_0^+) - \hat{Q}_{Y|X}^{(1)}(\tau|x_0^-)}{b'(x_0^+) - b'(x_0^-)} = \widehat{\text{QRKD}}(\tau).$$

Second, the LDTE, $\Delta_{Id}(y)$, is estimated using our Type 1 estimator from (4), with the transformation function set to the indicator $\varphi(Y, y) = \mathbb{1}(Y \leq y)$. Since the LDTE, $\Delta_{Id}(y)$, is identified pointwise for $y \in \mathcal{Y}_{x_0}$ by $\text{RKD}_F(y)$, a common application involves evaluating this effect at specific quantiles of interest, $y_\tau = Q_{Y|X}(\tau|x_0)$. This yields the estimator for the LDTE:

$$\hat{\Delta}_{Id}(\hat{y}_\tau) = \frac{\hat{F}_{Y|X}^{(1)}(\hat{y}_\tau|x_0^+) - \hat{F}_{Y|X}^{(1)}(\hat{y}_\tau|x_0^-)}{b'(x_0^+) - b'(x_0^-)} = \widehat{\text{RKD}}_F(\hat{y}_\tau).$$

A feature of the constrained estimation approach is that the necessary components for this evaluation are obtained jointly. The local polynomial quantile regression used to estimate the LQTE also yields an estimate of the quantile level itself, $\hat{y}_\tau = \hat{Q}_{Y|X}(\tau|x_0)$. The LDTE evaluated at the estimated τ -th quantile is then estimated by substituting \hat{y}_τ into the DRKD estimator.

Example 3 (Lorenz Effect, Continued). The identification result for the LLTE, $\Delta_L(\tau)$, expresses it as a function of components related to the mean and quantile effects. Specifically, constructing an estimator for $\Delta_L(\tau)$ requires estimates of: the conditional mean (μ_0), the mean RKD estimand ($\widehat{\text{MRKD}}$), the conditional Lorenz curve ($L_{Y|X}(\tau|x_0)$), and the quantile RKD estimand ($\widehat{\text{QRKD}}$). Each of these components is available from the estimation procedures described previously. The mean-related terms, $\hat{\mu}_0$ and $\widehat{\text{MRKD}}$, are obtained from the local polynomial regression detailed in Example 1. The quantile-related terms—namely $\widehat{\text{QRKD}}(u)$ needed for the integral and $\hat{Q}_{Y|X}(u|x_0)$ needed to construct the estimated Lorenz curve $\hat{L}_{Y|X}(\tau|x_0)$ —are obtained from the local polynomial quantile regression detailed in Example 2. A plug-in estimator for the LLTE is therefore constructed by substituting these previously obtained estimates into the identification formula:

$$\hat{\Delta}_L(\tau) = \frac{1}{\hat{\mu}_0} \left(\int_0^\tau \widehat{\text{QRKD}}(u) du - \hat{L}_{Y|X}(\tau|x_0) \cdot \widehat{\text{RKD}}_\mu \right) = \widehat{\text{RKD}}_{\psi_L|\mu}(\tau). \quad (6)$$

The integral in this expression can be computed using standard numerical methods.

5.2 Asymptotic Theory

This section establishes the asymptotic properties of our Type 1 and Type 2 estimators, $\widehat{\text{RKD}}_m(\cdot)$ and $\widehat{\text{RKD}}_{\psi|m}(\cdot)$. We begin by defining the necessary notation. Define the following kernel-dependent constant matrices and vectors: $\bar{L}_p := \int_{\mathbb{R}} \bar{r}_p(u) \bar{r}_p(u)^\top K(u) du$, $\bar{\vartheta}_{p,q}^\pm := \int_{\mathbb{R}_\pm} \bar{r}_p(u) u^q K(u) du$, and $\bar{\Psi}_p^\pm := \int_{\mathbb{R}_\pm} \bar{r}_p(u) \bar{r}_p(u)^\top K^2(u) du$. Let $\varepsilon^m(Y, X, \theta) = \varphi(Y, \theta) - m(\theta, X)$ be the error term from the conditional expectation. We define its conditional covariance function as $\sigma_{\varepsilon^m}(\theta_1, \theta_2 | x) := E[\varepsilon^m(Y, X, \theta_1) \varepsilon^m(Y, X, \theta_2) | X = x]$.

Assumption 3.

- (i) (a) The data $\{(Y_i, X_i)\}_{i=1}^n$ are i.i.d. copies of random vector (Y, X) on the probability space (Ω, \mathcal{F}, P) . (b) The density $f_X(\cdot)$ is continuously differentiable on I_{x_0} and satisfies $0 < f_X(x_0) < \infty$.
- (ii) (a) $m(\theta, x)$ is Lipschitz continuous at $x = x_0$ and $m^{(\nu)}(\theta, \cdot)$ is Lipschitz continuous on $I_{x_0} \setminus \{x_0\}$ for each $\theta \in \bar{\Theta}$ and $\nu \in \{1, \dots, p+1\}$. (b) The classes of real-valued functions $\{y \mapsto \varphi(y, \theta) : \theta \in \bar{\Theta}\}$ and $\{x \mapsto m(\theta, x) : \theta \in \bar{\Theta}\}$ are of Vapnik-Cervonenkis (VC) type with a common integrable envelope $\chi : \mathcal{Y} \times \mathcal{X} \rightarrow \mathbb{R}$ satisfying $\int |\chi(y, x)|^{2+\eta} dF_{Y,X}(y, x) < \infty$ for some $\eta > 0$. (c) The conditional covariance $\sigma_{\varepsilon^m}(\theta_1, \theta_2 | \cdot) \in \mathcal{C}^1(I_{x_0} \setminus \{x_0\})$ and $\sigma_{\varepsilon^m}(\theta_1, \theta_2 | x_0^\pm) < \infty$ for each $\theta_1, \theta_2 \in \bar{\Theta}$.
- (iii) (a) The kernel function $K : [-1, 1] \rightarrow \mathbb{R}_+$ is bounded and continuous, and the class $\{x \mapsto K((x - x_0)/h) : h > 0\}$ is of VC type. (b) \bar{L} is positive definite. (c) $\bar{\Psi}^+$, $\bar{\Psi}^-$, and $\int_{\mathbb{R}_\pm} |\bar{r}_p(u) u^q K(u)| du$ are finite.
- (iv) The bandwidth $h_{n,\theta}$ is specified as $h_{n,\theta} = \varsigma(\theta) h_n$, where $\varsigma : \Theta \rightarrow (0, \infty)$ is a bounded, Lipschitz continuous function. The baseline bandwidth sequence $\{h_n\}$ is required to satisfy the following conditions as $n \rightarrow \infty$: (a) $h_n \rightarrow 0$ and $nh_n^3 \rightarrow \infty$; (b) $nh_n^{2p+3} \rightarrow 0$.

Assumption 3 collects the regularity conditions required to establish the uniform asymptotic expansion and weak convergence of our estimators. We briefly discuss each component. For Condition (i), Part (a) is the standard i.i.d. sampling assumption; Part (b) imposes smoothness on the density of the running variable, f_X . This is a standard condition used to rule out endogenous sorting around the kink point, and can be derived from the more primitive Assumption S3'. For Condition (ii), Part (a) imposes smoothness on the conditional moment function, $m(\theta, x)$. Continuity at the kink x_0 and differentiability away from the kink are the definitional feature of the RKD. Parts (b) and (c) are technical conditions on the complexity (i.e., VC-type) of the relevant function classes, and are analogous to Assumptions (ii)(a) and (ii)(c) in Chiang et al. (2019). Condition (iii) imposes standard requirements on the kernel function, such as boundedness, which ensures a finite asymptotic variance. It is analogous to Assumption 1(iv) of Chiang

et al. (2019) and notably rules out the use of the Gaussian kernel. Condition (iv) specifies the admissible rates for the bandwidth sequence h_n . Part (a) imposes a rate consistent with the RKD literature (Card et al., 2015b; Chiang and Sasaki, 2019) and standard theory for local polynomial estimation of first-order derivatives. Part (b) is a standard undersmoothing requirement, which ensures that the bias term of the estimator is asymptotically negligible relative to the variance term.

Lemma 4 (Asymptotic Properties of $\hat{m}^{(1)}(\cdot, x_0^\pm)$).

(i) (Uniform Bahadur Representation) Suppose Assumptions 3(i)–(iii), and (iv)(a) hold. Then,

$$\begin{aligned} & \left[\sqrt{nh_{n,\theta}^3} \left(\hat{m}^{(1)}(\theta, x_0^+) - m^{(1)}(\theta, x_0^+) - h_{n,\theta}^p \frac{\iota_2^\top \bar{\Gamma}_p^{-1} (m^{(p+1)}(\theta, x_0^+) \bar{v}_{p,p+1}^+ + m^{(p+1)}(\theta, x_0^-) \bar{v}_{p,p+1}^-)}{(p+1)!} \right) \right] \\ & \left[\sqrt{nh_{n,\theta}^3} \left(\hat{m}^{(1)}(\theta, x_0^-) - m^{(1)}(\theta, x_0^-) - h_{n,\theta}^p \frac{\iota_3^\top \bar{\Gamma}_p^{-1} (m^{(p+1)}(\theta, x_0^+) \bar{v}_{p,p+1}^+ + m^{(p+1)}(\theta, x_0^-) \bar{v}_{p,p+1}^-)}{(p+1)!} \right) \right] \\ & = \begin{bmatrix} \mathbb{Z}_n^m(\theta, 2) \\ \mathbb{Z}_n^m(\theta, 3) \end{bmatrix} + o_P(1) \end{aligned}$$

as $n \rightarrow \infty$ uniformly in $\theta \in \bar{\Theta}$, where

$$\mathbb{Z}_n^m(\theta, k) := \sum_{i=1}^n \frac{\iota_k \bar{\Gamma}_p^{-1} \bar{r}_p \left(\frac{X_i - x_0}{h_{n,\theta}} \right) K \left(\frac{X_i - x_0}{h_{n,\theta}} \right) \varepsilon^m(Y_i, X_i, \theta)}{\sqrt{nh_{n,\theta}} f_X(x_0)}, \quad k \in \{2, 3\}.$$

(ii) (Weak Convergence) Suppose Assumption 3 holds. Then,

$$\begin{bmatrix} \mathbb{Z}_n^m(\cdot, 2) \\ \mathbb{Z}_n^m(\cdot, 3) \end{bmatrix} \rightsquigarrow \begin{bmatrix} \mathbb{Z}^m(\cdot, 2) \\ \mathbb{Z}^m(\cdot, 3) \end{bmatrix},$$

where $\mathbb{Z}^m : \Omega \rightarrow \ell^\infty(\bar{\Theta} \times \{2, 3\})$ is a tight zero-mean Gaussian process with covariance function

$$E[\mathbb{Z}^m(\theta_1, k_1) \mathbb{Z}^m(\theta_2, k_2)] = \frac{\iota_{k_1}^\top \bar{\Gamma}_p^{-1} \bar{\Xi}_p^m(\theta_1, \theta_2) \bar{\Gamma}_p^{-1} \iota_{k_2}}{f_X(x_0)}$$

for all $\theta_1, \theta_2 \in \bar{\Theta}$ and $k_1, k_2 \in \{2, 3\}$. The matrix $\bar{\Xi}_p^m$ in the covariance is given by $\bar{\Xi}_p^m(\theta_1, \theta_2) := \bar{\Psi}_p^+(\theta_1, \theta_2) \sigma_{\varepsilon^m}(\theta_1, \theta_2 | x_0^+) + \bar{\Psi}_p^-(\theta_1, \theta_2) \sigma_{\varepsilon^m}(\theta_1, \theta_2 | x_0^-)$, where

$$\bar{\Psi}_p^\pm(\theta_1, \theta_2) := \frac{1}{\sqrt{\varsigma(\theta_1) \varsigma(\theta_2)}} \int_{\mathbb{R}_\pm} \bar{r}_p \left(\frac{u}{\varsigma(\theta_1)} \right) \bar{r}_p \left(\frac{u}{\varsigma(\theta_2)} \right)^\top K \left(\frac{u}{\varsigma(\theta_1)} \right) K \left(\frac{u}{\varsigma(\theta_2)} \right) du.$$

Part (i) of the lemma establishes the uniform Bahadur representation for our estimator and provides the explicit expression for the asymptotic bias. In contrast to standard local p th-order polynomial estimators (e.g., Chiang et al., 2019, Lemma 1), the bias term exhibits two

specific characteristics. First, it depends on the $(p + 1)$ th-order derivatives from both sides of the kink. Second, the associated constant matrices $\bar{\Gamma}_p$ and vectors $\bar{v}_{p,p+1}^\pm$ are constructed using the regressor vector $\bar{r}_p(u) \in \mathbb{R}^{2p+1}$ rather than the standard polynomial basis $r_p(u) \in \mathbb{R}^{p+1}$. Part (ii) of the lemma establishes the weak convergence of the normalized processes. This result requires the use of an undersmoothing bandwidth satisfying $nh_n^{2p+3} = o(1)$. This condition ensures that the higher-order bias term from Part (i) is asymptotically negligible. Consequently, the normalized process converges weakly to a zero-mean Gaussian process, free of asymptotic bias.

We now introduce the assumptions required to establish the asymptotic properties of the quantile-based and composite estimands, $\widehat{\text{QRKD}}$ and $\widehat{\text{RKD}}_{\psi|m}$.

Assumption 4 (Chiang and Sasaki, 2019, Assumption 6).

- (i) Assumption 3(i).
- (ii) (a) $f_{Y|X}(Q_{Y|X}(\cdot|x_0)|x_0)$ is Lipschitz continuous on \mathcal{T} . (b) There exists finite constants $f_L > 0$, $f_U > 0$, and $\delta > 0$ such that for all $\tau \in \mathcal{T}$, $|\eta| \leq \delta$ and $x \in I_{x_0}$, $f_{Y|X}(Q_{Y|X}(\tau|x) + \eta|x) \in [f_L, f_U]$.
- (iii) (a) $Q_{Y|X}(\cdot|x_0)$, $\partial Q_{Y|X}(\cdot|x_0^\pm)/\partial \tau$ exist and are Lipschitz continuous on \mathcal{T} . (b) $Q_{Y|X}(\tau|\cdot)$ is continuous at x_0 ; the function $(x, \tau) \mapsto \partial^v Q_{Y|X}(\tau|x)/\partial x^v$ exists and is Lipschitz continuous on $(x, \tau) \in (I_{x_0} \setminus \{x_0\}) \times \mathcal{T}$ for all $v \in \{0, 1, \dots, p+1\}$.
- (iv) The kernel K is compactly supported, nonnegative, and satisfies $K'(u) < \infty$, $\int K(u) du = 1$ and $\int uK(u) du = 0$. The matrix $\bar{\Gamma}_p$ is positive definite.
- (v) The bandwidths $h_{n,\tau}$ satisfies $h_{n,\tau} = c(\tau)h_n$, where $nh_n^3 \rightarrow \infty$ and $nh_n^{2p+3} \rightarrow 0$ as $n \rightarrow \infty$ and $c(\tau)$ is Lipschitz continuous satisfying $0 < \underline{c} \leq c(\tau) \leq \bar{c} < \infty$ for all $\tau \in \mathcal{T}$.

Assumption 5.

- (i) For each $(g, h) \in \ell^\infty(\bar{\Theta}) \times \ell^\infty(\mathcal{T})$, $\hat{\psi}(g, h)(\cdot) = \psi(g, h)(\cdot) + o_P(1)$ uniformly on $\bar{\Theta}'$.
- (ii) ψ is Hadamard differentiable at $(\text{RKD}_m, \text{QRKD})$ with derivative $\psi'_{(\text{RKD}_m, \text{QRKD})}(\cdot, \cdot)$.

Assumption 4 is required to establish the asymptotic properties of $\widehat{\text{QRKD}}$. This assumption is consistent with Assumption 6 of Chiang and Sasaki (2019), and a result that relies on it is restated for completeness as Lemma A1 in Appendix A.

Assumption 5 provides the high-level conditions required for the asymptotic analysis of our Type 2 composite estimators, $\widehat{\text{RKD}}_{\psi|m}$. Part (i) requires the uniform consistency of any ‘nuisance’ estimators that appear in the definition of the functional ψ . For example, when estimating the LLTE, the functional ψ_L depends on the baseline mean μ_0 and the conditional quantile function $Q_{Y|X}(\cdot|x_0)$. The consistent estimators for these nuisance components are conveniently provided by the same constrained regressions, (4) and (5), used for the main effects. Part (ii)

imposes the standard Hadamard differentiability condition on the functional ψ . This condition enables the application of the functional Delta method to derive the asymptotic distribution of the composite estimator (e.g., [Kosorok, 2008](#), Theorem 2.8).

Theorem 2 (Weak Convergence of RKD Estimators).

(i) Suppose Assumptions 2 and 3 hold. Then,

$$\sqrt{nh_{n,\theta}^3} \left(\widehat{\text{RKD}}_m(\theta) - \text{RKD}_m(\theta) \right) \rightsquigarrow \mathbb{G}^m(\theta) := \frac{\mathbb{Z}^m(\theta, 2) - \mathbb{Z}^m(\theta, 3)}{b'(x_0^+) - b'(x_0^-)}$$

uniformly in $\theta \in \bar{\Theta}$. The limiting process $\mathbb{G}^m(\cdot)$ is a zero-mean Gaussian process with covariance function

$$E[\mathbb{G}^m(\theta_1)\mathbb{G}^m(\theta_2)] = \frac{(\iota_2 - \iota_3)^\top \bar{\Gamma}_p^{-1} \bar{\Xi}_p^m(\theta_1, \theta_2) \bar{\Gamma}_p^{-1} (\iota_2 - \iota_3)}{(b'(x_0^+) - b'(x_0^-))^2 f_X(x_0)},$$

where the terms $\mathbb{Z}^m(\cdot, \cdot)$ and $\bar{\Xi}_p^m(\cdot, \cdot)$ are as defined in Lemma 4.

(ii) Suppose Assumptions 2, 3, 4, and 5 hold, Then

$$\sqrt{nh_n^3} \left(\widehat{\text{RKD}}_{\psi|m}(\theta') - \text{RKD}_{\psi|m}(\theta') \right) \rightsquigarrow \mathbb{G}^{\psi|m}(\theta') := \psi'_{(\text{RKD}_m, \text{QRKD})} \left(\frac{\mathbb{G}^m(\cdot)}{\sqrt{\varsigma^3(\cdot)}}, \frac{\mathbb{G}^Q(\cdot)}{\sqrt{c^3(\cdot)}} \right) (\theta')$$

uniformly in $\theta' \in \bar{\Theta}'$, where $\mathbb{G}^{\psi|m}(\theta')$ is a zero-mean Gaussian process with covariance function $E[\mathbb{G}^{\psi|m}(\theta'_1)\mathbb{G}^{\psi|m}(\theta'_2)]$, the form of which depends on ψ . The limiting process $\mathbb{G}^Q : \Omega \rightarrow \ell^\infty(\mathcal{T})$ is the Gaussian process for the quantile estimator, as defined in Lemma A1.

Theorem 2(i) provides the weak convergence result for the Type 1 estimators. This result includes the distributional RKD estimator and the mean RKD estimator as special cases. While asymptotic theory for local Wald ratios exists (e.g., the general, higher-order case in [Chiang et al., 2019](#)), the result here applies specifically to the first-order derivative and is derived for the constrained estimation approach used in this paper. Theorem 2(ii) establishes the weak convergence for the Type 2 (composite) estimators. One application is the local Lorenz effect estimator. The limiting process $\mathbb{G}^{\psi|m}$ for these estimators is a composite of the limiting processes \mathbb{G}^m and \mathbb{G}^Q for the Type 1 and quantile estimators. Because this limiting process is not a local Wald ratio, the inference framework of [Chiang et al. \(2019\)](#), though general in other contexts, cannot be directly applied to these Type 2 estimators.

5.3 Multiplier Bootstrap and Pivotal Methods

This section develops a resampling method to approximate the asymptotic distributions derived in Theorem 2 and to conduct uniform inference.

For the Type 1 estimator, $\widehat{\text{RKD}}_m(\cdot)$, we employ a multiplier bootstrap to approximate the distribution of the limiting process $\mathbb{G}^m(\cdot)$. The bootstrap process is constructed based on the leading term of the uniform Bahadur representation established in Lemma 4(i). Let $\{\xi_i\}_{i=1}^n$ be a sequence of i.i.d. standard normal random variables, drawn independently of the original data. We define the estimated multiplier process (EMP) as:

$$\widehat{\mathbb{G}}_\xi^m(\theta) := \frac{1}{\sqrt{nh_{n,\theta}}} \sum_{i=1}^n \xi_i \frac{(\iota_2 - \iota_3)^\top \bar{F}_p^{-1} \bar{r}_p \left(\frac{X_i - x_0}{h_{n,\theta}} \right) K \left(\frac{X_i - x_0}{h_{n,\theta}} \right) \hat{\varepsilon}^m(Y_i, X_i, \theta)}{(b'(x_0^+) - b'(x_0^-)) \hat{f}_X(x_0)} \quad (7)$$

for each $\theta \in \bar{\Theta}$. This process requires estimators for the unknown components in the asymptotic representation. The residuals are computed using the fitted values from the constrained regression, $\hat{\varepsilon}^m(y, x, \theta) := (\varphi(y, \theta) - \bar{r}_p(x - x_0) \hat{\alpha}(\theta)) \mathbb{1}(|x - x_0| \leq h_{n,\theta})$, whose uniform consistency is established in Lemma A2. The density $f_X(x_0)$ is estimated using a standard kernel density estimator $\hat{f}_X(x_0) = \frac{1}{nv_n} \sum_{i=1}^n K((X_i - x_0)/v_n)$.

Approximating the asymptotic distribution for the Type 2 estimator $\widehat{\text{RKD}}_{\psi|m}$ requires to generate valid bootstrap analogues for both. The process \mathbb{G}^m is simulated using the multiplier bootstrap $\widehat{\mathbb{G}}_\xi^m$ as described previously. For the quantile component \mathbb{G}^Q , we employ the pivotal method developed by Chiang and Sasaki (2019) and Qu and Yoon (2015). Let $\{U_i\}_{i=1}^n$ be a sequence of i.i.d. Uniform(0, 1) random variables, drawn independently of the data. The estimated pivotal process for \mathbb{G}^Q is then constructed as:

$$\widehat{\mathbb{G}}_u^Q(\tau) := \frac{1}{\sqrt{nh_{n,\tau}}} \sum_{i=1}^n \frac{(\iota_2 - \iota_3)^\top \bar{F}_p^{-1} \bar{r}_p \left(\frac{X_i - x_0}{h_{n,\tau}} \right) K \left(\frac{X_i - x_0}{h_{n,\tau}} \right) (\tau - \mathbb{1}(U_i \leq \tau))}{(b'(x_0^+) - b'(x_0^-)) \hat{f}_X(x_0) \hat{f}_{Y|X}(\hat{Q}_{Y|X}(\tau|x_0)|x_0)}.$$

The bootstrap analogue for the composite limiting process $\mathbb{G}^{\psi|m}$ is then constructed by combining these two simulated processes using the estimated Hadamard derivative of the functional ψ :

$$\widehat{\mathbb{G}}_{\xi,u}^{\psi|m}(\theta') := \hat{\psi}'_{(\text{RKD}_m, \text{QRKD})} \left(\frac{\widehat{\mathbb{G}}_\xi^m(\cdot)}{\sqrt{c(\cdot)}}, \frac{\widehat{\mathbb{G}}_u^Q(\cdot)}{\sqrt{c(\cdot)}} \right) (\theta'). \quad (8)$$

The following high-level assumption is required to establish the uniform validity of our resampling procedures.

Assumption 6 (First Stage Estimation).

(i) $\hat{f}_X(x_0) = f_X(x_0) + o_P(1)$, and $\hat{f}_{Y|X}(y|x_0) = f_{Y|X}(y|x_0) + o_P(1)$ uniformly in $y \in \mathcal{Y}_{x_0}$.

(ii) For each $(g, h) \in \ell^\infty(\bar{\Theta}) \times \ell^\infty(\mathcal{T})$, the estimated Hadamard derivative satisfies

$$\hat{\psi}'_{(\text{RKD}_m, \text{QRKD})}(g, h)(\cdot) = \psi'_{(\text{RKD}_m, \text{QRKD})}(g, h)(\cdot) + o_P(1)$$

uniformly on $\bar{\Theta}'$

Assumption 6 ensures the consistency of the nuisance components used in constructing $\hat{\mathbb{G}}_\xi^m$ and $\hat{\mathbb{G}}_{\xi,u}^{\psi|m}$. The role of Part (ii) can be illustrated using the Lorenz effect Δ_L as an example. For this effect, the functional ψ_L is bilinear, and its Hadamard derivative, $[\psi_L]'$, depends on the nuisance parameters μ_0 and the function $Q_{Y|X}(\cdot|x_0)$. In this context, Assumption 6(ii) is a high-level condition requiring uniform consistency of the estimators for these nuisance components, i.e., $\hat{\mu}_0 \xrightarrow{P} \mu_0$ and $\sup_{\tau \in \mathcal{T}} |\hat{Q}_{Y|X}(\tau|x_0) - Q_{Y|X}(\tau|x_0)| \xrightarrow{P} 0$.

Let $\xrightarrow[M]{P}$ denote the conditional weak convergence, where the subscript M indicates conditional expectation over the weights M given the remaining data (Kosorok, 2008, Page. 20).

Theorem 3 (Conditional Weak Convergence). *Suppose Assumptions 2–6 hold. Then:*

- (i) $\hat{\mathbb{G}}_\xi^m(\theta) \xrightarrow[\xi]{P} \mathbb{G}^m(\theta)$ uniformly in $\theta \in \bar{\Theta}$.
- (ii) $\hat{\mathbb{G}}_{\xi,u}^{\psi|m}(\theta') \xrightarrow[\xi \times u]{P} \mathbb{G}^{\psi|m}(\theta')$ uniformly in $\theta' \in \bar{\Theta}'$.

Theorem 3 establishes that the empirical distributions of the simulated processes, $\hat{\mathbb{G}}_\xi^m(\theta)$ and $\hat{\mathbb{G}}_{\xi,u}^{\psi|m}(\theta')$, validly approximate the asymptotic distributions of the normalized processes of $\widehat{\text{RKD}}_m(\theta)$ and $\widehat{\text{RKD}}_{\psi|m}(\theta')$, respectively. This result enables the construction of asymptotically valid uniform confidence bands and hypothesis tests in practice.

5.4 Hypothesis Tests and Uniform Confidence Bands

Building on the asymptotic results of Theorems 2 and 3, this section details the procedure for conducting uniform inference on the local treatment effects, including uniform hypothesis testing and the construction of uniform confidence bands. Our goal is to conduct inference on the functions $\theta \mapsto \Delta_m(\theta)$ and $\theta' \mapsto \Delta_{\psi|m}(\theta')$, which represent our *Type 1* and *Type 2 LTEs*, respectively.

To test for the uniform significance of the treatment effect, we consider the null hypothesis that the LTE function is identically zero over the region of interest. For our Type 1 and Type 2 effects, respectively, the hypotheses are:

$$\begin{aligned} \mathcal{H}_{0,m}^S : \Delta_m(\theta) &= 0 \quad \text{for all } \theta \in \bar{\Theta}, \\ \text{and } \mathcal{H}_{0,\psi|m}^S : \Delta_{\psi|m}(\theta') &= 0 \quad \text{for all } \theta' \in \bar{\Theta}', \end{aligned}$$

where $\bar{\Theta} \subset \Theta$ and $\bar{\Theta}' \subset \Theta'$ are compact spaces of interests. We test these hypotheses using Kolmogorov–Smirnov type test statistics, which measure the maximum deviation of the estimated process from zero:

$$W_m^S(\bar{\Theta}) = \sup_{\theta \in \bar{\Theta}} \left| \sqrt{nh_{n,\theta}^3} \widehat{\text{RKD}}_m(\theta) \right|,$$

$$\text{and } W_{\psi|m}^S(\bar{\Theta}') = \sup_{\theta' \in \bar{\Theta}'} \left| \sqrt{nh_n^3} \widehat{\text{RKD}}_{\psi|m}(\theta') \right|.$$

To test for the homogeneity of the treatment effect, we consider the null hypothesis that the LTE function is constant, though not necessarily zero. The hypotheses are:

$$\begin{aligned} \mathcal{H}_{0,m}^H : \Delta_m(\theta_1) &= \Delta_m(\theta_2) \quad \text{for all } \theta_1, \theta_2 \in \bar{\Theta}, \\ \text{and } \mathcal{H}_{0,\psi|m}^H : \Delta_{\psi|m}(\theta'_1) &= \Delta_{\psi|m}(\theta'_2) \quad \text{for all } \theta'_1, \theta'_2 \in \bar{\Theta}', \end{aligned}$$

These hypotheses can be tested by measuring the deviation from the mean over the region of interest. The specific statistics are defined as:

$$\begin{aligned} W_m^H(\bar{\Theta}) &= \sup_{\theta \in \bar{\Theta}} \left| \sqrt{nh_{n,\theta}^3} \left(\widehat{\text{RKD}}_m(\theta) - \frac{1}{|\bar{\Theta}|} \int_{\bar{\Theta}} \widehat{\text{RKD}}_m(\theta) d\theta \right) \right|, \\ \text{and } W_{\psi|m}^H(\bar{\Theta}') &= \sup_{\theta' \in \bar{\Theta}'} \left| \sqrt{nh_n^3} \left(\widehat{\text{RKD}}_{\psi|m}(\theta') - \frac{1}{|\bar{\Theta}'|} \int_{\bar{\Theta}'} \widehat{\text{RKD}}_{\psi|m}(\theta') d\theta' \right) \right|. \end{aligned}$$

The following corollary, which is a direct consequence of Theorem 2, establishes the asymptotic distribution of our test statistics under their respective null hypotheses.

Corollary 1. *Suppose Assumptions 2–6 hold. Then,*

- (i) $W_m^S(\bar{\Theta}) \rightsquigarrow \sup_{\theta \in \bar{\Theta}} |\mathbb{G}^m(\theta)|$ under the null hypothesis $\mathcal{H}_{0,m}^S$;
 $W_m^H(\bar{\Theta}') \rightsquigarrow \sup_{\theta \in \bar{\Theta}'} \left| \mathbb{G}^m(\theta) - \frac{1}{|\bar{\Theta}'|} \int_{\bar{\Theta}'} \mathbb{G}^m(\theta) d\theta \right|$ under the null hypothesis $\mathcal{H}_{0,m}^H$.
- (ii) $W_{\psi|m}^S(\bar{\Theta}) \rightsquigarrow \sup_{\theta \in \bar{\Theta}} |\mathbb{G}^{\psi|m}(\theta)|$ under the null hypothesis $\mathcal{H}_{0,\psi|m}^S$;
 $W_{\psi|m}^H(\bar{\Theta}') \rightsquigarrow \sup_{\theta' \in \bar{\Theta}'} \left| \mathbb{G}^{\psi|m}(\theta') - \frac{1}{|\bar{\Theta}'|} \int_{\bar{\Theta}'} \mathbb{G}^{\psi|m}(\theta') d\theta' \right|$ under the null hypothesis $\mathcal{H}_{0,\psi|m}^H$.

In practice, the asymptotic distributions of the test statistics in Corollary 1 can be simulated based on the estimated limiting processes \hat{G}_ξ^m and $\hat{G}_{\xi,u}^{\psi|m}$, respectively.

Building on the bootstrap validity established in Theorem 3, we now detail the construction of uniform confidence bands for our LTEs. The procedure for a generic LTE function, $\Delta(\cdot)$, follows a standard two-step process:

Step 1: Estimate the Critical Value. We simulate the distribution of the supremum norm of the relevant bootstrap process. For $\lambda \in (0, 1)$, the estimated $(1 - \lambda)$ critical value, $\hat{c}(1 - \lambda)$, is then obtained as the $(1 - \lambda)$ sample quantile of this simulated distribution.

Step 2: Construct the Band. The $100(1 - \lambda)\%$ uniform confidence band for the function $\Delta(\cdot)$ is then formed by:

$$\left[\hat{\Delta}(\cdot) \pm \hat{c}(1 - \lambda) / \sqrt{nh_{n,\cdot}^3} \right],$$

where $\hat{\Delta}(\cdot)$ is the corresponding point estimator $\widehat{\text{RKD}}_m(\theta)$ or $\widehat{\text{RKD}}_{\psi|m}(\theta')$. Specifically, for the Type 1 LTE Δ_m , we use the estimated bootstrap process \hat{G}_ξ^m to compute $\hat{c}_m(1 - \lambda)$, and for the Type 2 LTE $\Delta_{\psi|m}$, we use the estimated composite process $\hat{G}_{\xi,u}^{\psi|m}$ to compute $\hat{c}_{\psi|m}(1 - \lambda)$.

This section concludes by detailing the implementation of our inference procedures for the specific LTEs discussed previously. Bandwidth selection procedures for each effect are provided in Appendix B. Let B be the number of bootstrap replications.

Example 1 (Average Effect, Continued). The LATE Δ_μ is estimated by $\widehat{\text{RKD}}_\mu$. For each replication $b = 1, \dots, B$, we generate an i.i.d. sequence of $N(0, 1)$ random weights $\{\xi_i^b\}_{i=1}^n$, drawn independently of the data. Then, the distribution of $\sqrt{nh_n^3}(\widehat{\text{RKD}}_\mu - \Delta_\mu)$ is approximated by the empirical distribution of $\{\widehat{\mathbb{G}}_{\xi^b}^\mu\}_{b=1}^B$, where

$$\widehat{\mathbb{G}}_{\xi^b}^\mu := \sum_{i=1}^n \xi_i^b \frac{(\iota_2 - \iota_3)^\top \bar{I}_p^{-1} \bar{r}_p \left(\frac{X_i - x_0}{h_n} \right) K \left(\frac{X_i - x_0}{h_n} \right) \hat{\varepsilon}^\mu(Y_i, X_i)}{(b'(x_0^+) - b'(x_0^-)) \sqrt{nh_n} \hat{f}_X(x_0)}.$$

The random error is estimated by $\hat{\varepsilon}^\mu(y, x) = (y - \bar{r}_p(x - x_0)\hat{\alpha}) \mathbb{1}(|x - x_0| \leq h_n)$. To test the significance hypothesis, $\Delta_\mu = 0$, we first compute the test statistic $W_\mu^S = |\sqrt{nh_n^3} \widehat{\text{RKD}}_\mu|$. The $(1 - \lambda)$ critical value, $\hat{c}_\mu(1 - \lambda)$, is the $(1 - \lambda)$ sample quantile of the simulated distribution $\{|\widehat{\mathbb{G}}_{\xi^b}^\mu|\}_{b=1}^B$. We reject the null if $W_\mu^S > \hat{c}_\mu(1 - \lambda)$. The bootstrap p -value is calculated as $\hat{p}_\mu^S = \frac{1}{B} \sum_{b=1}^B \mathbb{1}(|\widehat{\mathbb{G}}_{\xi^b}^\mu| > W_\mu^S)$. Finally, the uniform confidence band is constructed as $[\widehat{\text{RKD}}_\mu \pm \hat{c}_\mu(1 - \lambda)/\sqrt{nh_n^3}]$.

Example 2 (Distributional and Quantile Effects, Continued).

The LDTE $\Delta_{Id}(\cdot)$ is estimated by $\widehat{\text{RKD}}_F(\cdot)$. A common application is to test properties of the LDTE Δ_{Id} over a range of quantiles, i.e., to conduct inference on the process $\{\Delta_{Id}(y_\tau)\}_{\tau \in \mathcal{T}}$. For each replication $b = 1, \dots, B$, generate an i.i.d. sequence of $N(0, 1)$ weights $\{\xi_i^b\}_{i=1}^n$. We construct the b th draw of the bootstrap process $\widehat{\mathbb{G}}_{\xi^b}^{Id}(\cdot)$ as

$$\widehat{\mathbb{G}}_{\xi^b}^{Id}(\theta) := \sum_{i=1}^n \xi_i^b \frac{(\iota_2 - \iota_3)^\top \bar{I}_p^{-1} \bar{r}_p \left(\frac{X_i - x_0}{h_{n,\theta}} \right) K \left(\frac{X_i - x_0}{h_{n,\theta}} \right) \hat{\varepsilon}^F(Y_i, X_i, \theta)}{(b'(x_0^+) - b'(x_0^-)) \sqrt{nh_{n,\theta}} \hat{f}_X(x_0)},$$

where the residual is computed as $\hat{\varepsilon}^F(y, x, \theta) = (\mathbb{1}(y \leq \theta) - \bar{r}_p(x - x_0)\hat{\alpha}) \mathbb{1}(|x - x_0| \leq h_n)$. For each replication b , evaluate the bootstrap process $\widehat{\mathbb{G}}_{\xi^b}^{Id}(\cdot)$ at the estimated quantile points $\{\hat{y}_\tau\}_{\tau \in \mathcal{T}}$ to obtain the simulated process $\{\widehat{\mathbb{G}}_{\xi^b}^{Id}(\hat{y}_\tau)\}_{\tau \in \mathcal{T}}$.

To test the hypothesis of uniform treatment significance $\Delta_{Id}(y_\tau) = 0$ for all $\tau \in \mathcal{T}$, we first compute the test statistic $W_{Id}^S := \sup_{\tau \in \mathcal{T}} |\sqrt{nh_{n,y_\tau}^3} \widehat{\text{RKD}}_F(\hat{y}_\tau)|$. Then, we use $\{\widehat{\mathbb{G}}_{Id,b}^S := \sup_{\tau \in \mathcal{T}} |\widehat{\mathbb{G}}_{\xi^b}^{Id}(\hat{y}_\tau)|\}_{b=1}^B$ to approximate the distribution of W_{Id}^S . The $(1 - \lambda)$ critical value, $\hat{c}_{Id}(1 - \lambda)$, is the $(1 - \lambda)$ sample quantile of $\{\widehat{\mathbb{G}}_{Id,b}^S\}_{b=1}^B$. The bootstrap p -value is given by $\hat{p}_{Id}^S := \frac{1}{B} \sum_{b=1}^B \mathbb{1}(\widehat{\mathbb{G}}_{Id,b}^S > W_{Id}^S)$.

To test the hypothesis of uniform treatment homogeneity $\Delta_{Id}(y_{\tau_1}) = \Delta_{Id}(y_{\tau_2})$ for all $\tau_1, \tau_2 \in \mathcal{T}$, we first compute the test statistic

$$W_{Id}^H := \sup_{\tau \in \mathcal{T}} \left| \sqrt{nh_{n,y_\tau}^3} \left(\widehat{\text{RKD}}_F(\hat{y}_\tau) - \frac{1}{|\mathcal{T}|} \int_{\mathcal{T}} \widehat{\text{RKD}}_F(\hat{y}_\tau) d\tau \right) \right|.$$

Then, we use $\{\widehat{\mathbb{G}}_{Id,b}^H := \sup_{\tau \in \mathcal{T}} |\widehat{\mathbb{G}}_{\xi^b}^{Id}(\hat{y}_\tau) - \frac{1}{|\mathcal{T}|} \int_{\mathcal{T}} \widehat{\mathbb{G}}_{\xi^b}^{Id}(\hat{y}_\tau) d\tau|\}_{b=1}^B$ to approximate the distribution of W_{Id}^H . The critical value and p -value are calculated analogously to the significance test. Finally, the uniform confidence band is constructed as

$$\left[\widehat{\text{RKD}}_F(\hat{y}_\tau) \pm \hat{c}_{Id}(1 - \lambda) / \sqrt{nh_{n,y_\tau}^3} \right].$$

Inference for the LQTE, $\Delta_Q(\cdot)$, which is estimated by $\widehat{\text{QRKD}}(\cdot)$, is conducted using the pivotal method detailed in Section 5. This involves generating i.i.d. $\text{Uniform}(0, 1)$ random variables to construct the pivotal process $\{\widehat{\mathbb{G}}_{u^b}^Q(\tau)\}_{\tau \in \mathcal{T}}$ for each replication b . The subsequent steps for constructing test statistics and obtaining critical values are analogous to those for the LDTE. For the sake of brevity, we omit the detailed formulas and refer readers to the comprehensive procedure outlined in Appendix C.2 of [Chiang and Sasaki \(2019\)](#), which our method follows directly.

Example 3 (Lorenz Effect, Continued). By Equation (6), the LLTE $\Delta_L(\tau)$ is estimated by $\widehat{\text{RKD}}_{\psi_L|\mu}(\tau)$. For each bootstrap iteration $b \in \{1, \dots, B\}$, we draw the i.i.d. sequence $\{\xi_i^b\}_{i=1}^n$ from $N(0, 1)$, and the i.i.d. sequence $\{u_i^b\}_{i=1}^n$ from $\text{Uniform}(0, 1)$, respectively. Then, we approximate the distribution of $\sqrt{n(h_{n,\tau}^L)^3}(\text{RKD}_{\psi_L|\mu}(\tau) - \Delta_L(\tau))$ by the empirical distribution of $\{\widehat{\mathbb{G}}_{\xi^b, u^b}^L(\tau)\}_{b=1}^B$ for each $\tau \in \mathcal{T}$, where

$$\widehat{\mathbb{G}}_{\xi^b, u^b}^L(\tau) := \frac{1}{\hat{\mu}_0} \left(\int_0^\tau \widehat{\mathbb{G}}_{u^b}^Q(u) du - \hat{L}_{Y|X}(\tau|x_0) \cdot \widehat{\mathbb{G}}_{\xi^b}^\mu \right).$$

It is important to note the bandwidth choices used in the bootstrap procedures. For the mean effect bootstrap, $\widehat{\mathbb{G}}_{\xi^b}^\mu$, the single baseline bandwidth h_n is employed. In contrast, for the quantile effect bootstrap, $\widehat{\mathbb{G}}_{u^b}^Q(\tau)$, the pointwise bandwidth $h_{n,\tau}$ is used. As established by Lemma A1, using the pointwise bandwidth $h_{n,\tau}$ ensures that this bootstrap process correctly targets the desired limiting process $\mathbb{G}^Q(\tau)$, rather than the scaled version.

To test the hypothesis of uniform treatment significance $\Delta_L(\tau) = 0$ for all $\tau \in \mathcal{T}$, we first compute the test statistic $W_L^S := \sup_{\tau \in \mathcal{T}} |\sqrt{n(h_{n,\tau}^L)^3} \widehat{\text{RKD}}_{\psi_L|\mu}(\tau)|$. Then, we use $\{\widehat{\mathbb{G}}_{L,b}^S := \sup_{\tau \in \mathcal{T}} |\widehat{\mathbb{G}}_{\xi^b, u^b}^L(\tau)|\}_{b=1}^B$ to approximate the distribution of W_L^S . The $(1 - \lambda)$ th critical value, $\hat{c}_L(1 - \lambda)$, is $(1 - \lambda)$ sample quantile of $\{\widehat{\mathbb{G}}_{L,b}^S\}_{b=1}^B$. The p -value is given by $\hat{p}_L^S := \frac{1}{B} \sum_{b=1}^B \mathbb{1}(\widehat{\mathbb{G}}_{L,b}^S > W_L^S)$.

To test the hypothesis of uniform treatment homogeneity $\Delta_L(\tau_1) = \Delta_L(\tau_2)$ for all $\tau_1, \tau_2 \in \mathcal{T}$, we first compute the test statistic

$$W_L^H := \sup_{\tau \in \mathcal{T}} \left| \sqrt{n(h_{n,\tau}^L)^3} \left(\widehat{\text{RKD}}_{\psi_L|\mu}(\tau) - \frac{1}{|\mathcal{T}|} \int_{\mathcal{T}} \widehat{\text{RKD}}_{\psi_L|\mu}(\tau) d\tau \right) \right|.$$

Then, we use $\{\widehat{\mathbb{G}}_{L,b}^H := \sup_{\tau \in \mathcal{T}} |\widehat{\mathbb{G}}_{\xi^b, u^b}^L(\tau) - \frac{1}{|\mathcal{T}|} \int_{\mathcal{T}} \widehat{\mathbb{G}}_{\xi^b, u^b}^L(\tau) d\tau|\}_{b=1}^B$ to approximate the distribution of W_L^H . The critical value is $(1 - \lambda)$ sample quantile of $\{\widehat{\mathbb{G}}_{L,b}^H\}_{b=1}^B$. The p -value of the treatment

homogeneity test is given by $\hat{p}_L^H = \frac{1}{B} \sum_{b=1}^B \mathbb{1}(\hat{\mathbb{G}}_{L,b}^H > W_L^H)$. Finally, the uniform confidence band is constructed as

$$\left[\widehat{\text{RKD}}_{\psi_L|\mu}(\tau) \pm \hat{c}_L(1 - \lambda)/\sqrt{n(h_{n,\tau}^L)^3} \right].$$

6 NUMERICAL ILLUSTRATIONS

6.1 Empirical Application

In this section, we apply our framework to a classic dataset from the Continuous Wage and Benefit History (CWBH) to analyze the causal effects of unemployment insurance (UI) benefits on unemployment durations. The data cover unemployed individuals in Louisiana over two periods: September 1981–September 1982 and September 1982–December 1983. Our analysis builds upon the work of [Landais \(2015\)](#), to which we refer readers for detailed descriptive statistics and institutional background. The dataset contains the following variables: (i) Running variable (X): The highest quarterly wage in an individual’s base period. (ii) Treatment (B): The weekly UI benefit amount received. (iii) Outcome (Y): The duration of unemployment, measured in two ways: (a) the number of weeks benefits were claimed (UI Claimed), and (b) the number of weeks benefits were paid (UI Paid). This setting is a canonical example of a sharp regression kink design. The UI benefit amount B is a deterministic, piecewise-linear function of the base wage X : $b(x) = \min\{0.04 \cdot x, b_{max}\}$. The maximum benefit level b_{max} , which defines the location of the kink point x_0 , changed between the two periods. For the 1981–1982 period, b_{max} was \$4,575, whereas for the 1982–1983 period, it increased to \$5,125.

Building on the theoretical framework, this analysis examines the local distributional (LDTE) and local Lorenz (LLTE) treatment effects of UI benefits on unemployment durations. This focus extends the existing literature, which has primarily examined the average effect ([Landais, 2015](#)) and the quantile effect ([Chiang and Sasaki, 2019](#)). We estimate the LDTE, $\Delta_{Id}(y_\tau)$, and the LLTE, $\Delta_L(\tau)$, over a grid of quantiles τ . The full estimated effect functions for both time periods are displayed graphically in Figures 1 and 2. For specific numerical results, Tables 2 and 3 present point estimates and standard errors for these effects at selected quantiles $\tau \in \{0.1, 0.2, \dots, 0.9\}$, along with p -values from the corresponding uniform hypothesis tests. All estimation and inference procedures follow the methods detailed in Section 5. Consistent with [Chiang and Sasaki \(2019\)](#), we employ a tricube kernel function, $K(u) = \frac{70}{81}(1 - |u|^3)^3 \mathbb{1}(|u| \leq 1)$, and set the local polynomial order to $p = 2$ for bias-correction. Selection of the MSE-optimal bandwidths follows the procedures detailed in Algorithms 1 and 2 in Appendix B.

Panel A of the tables presents the results for the LDTE. The findings are summarized below. First, the estimated LDTE is negative across both time periods examined. This implies that a marginal increase in UI benefits reduces the cumulative probability of an unemployment spell ending within any given number of weeks for individuals at the kink wage. Second, the magnitude of the estimated effect varies with duration and changes between periods. In the 1981–1982 period, the effect’s magnitude is smaller at shorter durations (e.g., the 10th percentile) and

increases at longer durations. This pattern appears reversed in the 1982–1983 period, where the effect is largest in magnitude at short durations (e.g., the 20th percentile) and smaller at longer durations. Third, the LDTE is statistically significant at most quantile levels in both periods. Furthermore, formal tests for homogeneity are strongly rejected, confirming that the distributional effect is not constant. The results are also robust across the two different measures of unemployment duration (UI Claimed vs. UI Paid).

Panel B presents the results for the LLTE, which measures the effect of UI benefits on the inequality of the duration distribution. The estimated LLTE is positive at all quantiles considered in both periods. This indicates that a marginal increase in UI benefits increases the dispersion (inequality) of unemployment durations for individuals at the kink wage. This finding is consistent with the quantile effect patterns reported in [Chiang and Sasaki \(2019\)](#). Their observation—that benefit increases have a larger effect on longer unemployment spells—implies an increase in dispersion, which the LLTE directly measures. Similar to the LDTE, the estimated Lorenz effects are statistically significant in both periods. Tests also reject the null hypothesis of homogeneity. The findings remain consistent across the two measures of unemployment duration.

6.2 Simulation Experiment

This section evaluates the finite-sample performance of the proposed estimation and inference methods using Monte Carlo simulations. We generate i.i.d. samples $\{(Y_i, B_i, X_i)\}_{i=1}^n$ from the following data generating process:

$$\begin{aligned} Y_i &= 1 + 0.5B_i + X_i + 0.1X_i^2 + 1.5B_iX_i + (1 + 2B_i)\varepsilon_i \\ B_i &= X_i(2 \cdot \mathbb{1}(X_i \geq 0) - 1), \\ \begin{pmatrix} X_i \\ \varepsilon_i \end{pmatrix} &\sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_x^2 & \rho \cdot \sigma_x \cdot \sigma_\varepsilon \\ \rho \cdot \sigma_x \cdot \sigma_\varepsilon & \sigma_\varepsilon^2 \end{bmatrix}\right). \end{aligned}$$

The structure is similar to that in [Chiang and Sasaki \(2019\)](#) but is extended to include conditional heteroskedasticity through the $(1 + 2B_i)\varepsilon_i$ term. The kink point is normalized to $x_0 = 0$. The distributional parameters are chosen to match those in [Calonico et al. \(2014\)](#), with $\sigma_x = 0.1781742$, $\sigma_\varepsilon = 0.1295$, and correlation $\rho = 0.25$.

We evaluate the finite-sample performance of estimators for four local treatment effects: the average effect (Δ_μ), the distributional effect (Δ_{Id}), the quantile effect (Δ_Q), and the Lorenz effect (Δ_L). The LATE Δ_μ is estimated at a single point. The LQTE $\Delta_Q(\tau)$ and LLTE $\Delta_L(\tau)$ are evaluated over the quantile grid $\mathcal{T} = \{0.1, 0.2, \dots, 0.9\}$. The LDTE $\Delta_{Id}(y)$ is evaluated at points corresponding to the quantiles in \mathcal{T} , i.e., at $y = Q_{Y|X}(\tau|x_0)$ for $\tau \in \mathcal{T}$. Estimation and inference follow the procedures detailed in Section 5. For the LLTE estimator, required numerical integrals are computed using a finer grid, $\mathcal{U} = \{0.01, \dots, 0.99\}$. Nuisance parameters (e.g., μ_0 , $Q_{Y|X}$) are estimated as part of the main procedures. We consider sample sizes $n \in \{500, 1000, 2000, 4000\}$

Table 2: Estimation and Inference of the Local Distributional and Lorenz Treatment Effects, September 1981 – September 1982

September 1981–September 1982					
Dependent Variable		Ui Claimed		Ui Paid	
Panel A. LDTE $\Delta_{Id}(y)$	$y = y_{0.1}$	−0.004	(0.003)	−0.004	(0.003)
	$y = y_{0.2}$	−0.016	(0.004)	−0.014	(0.003)
	$y = y_{0.3}$	−0.016	(0.004)	−0.014	(0.004)
	$y = y_{0.4}$	−0.014	(0.004)	−0.013	(0.004)
	$y = y_{0.5}$	−0.011	(0.004)	−0.013	(0.004)
	$y = y_{0.6}$	−0.015	(0.003)	−0.013	(0.003)
	$y = y_{0.7}$	−0.015	(0.002)	−0.014	(0.002)
	$y = y_{0.8}$	−0.017	(0.002)	−0.016	(0.002)
	$y = y_{0.9}$	−0.012	(0.002)	−0.010	(0.002)
Test of significance	p -Value	0.000		0.000	
Test of Homogeneity	p -Value	0.010		0.001	
Panel B. LLTE $\Delta_L(\tau)$	$\tau = 0.10$	0.000	(0.000)	0.000	(0.000)
	$\tau = 0.20$	0.000	(0.000)	0.001	(0.000)
	$\tau = 0.30$	0.001	(0.001)	0.002	(0.001)
	$\tau = 0.40$	0.002	(0.001)	0.003	(0.001)
	$\tau = 0.50$	0.003	(0.002)	0.004	(0.002)
	$\tau = 0.60$	0.004	(0.002)	0.005	(0.002)
	$\tau = 0.70$	0.005	(0.002)	0.006	(0.003)
	$\tau = 0.80$	0.005	(0.003)	0.006	(0.003)
	$\tau = 0.90$	0.007	(0.002)	0.008	(0.002)
Test of significance	p -Value	0.002		0.001	
Test of Homogeneity	p -Value	0.010		0.018	

Note: The numbers in parentheses denote the estimated standard errors.

and set the number of Monte Carlo replications to 5,000 for each simulation design. In each replication, we employ both the multiplier bootstrap and the pivotal method, each based on 2,500 bootstrap draws, to construct the uniform confidence bands.

Figure 3 plots the estimated effects across simulations against the true parameter values for different sample sizes. Table 4 reports quantitative measures of finite-sample performance, including the absolute bias ratio ($|\hat{\Delta} - \Delta|/|\Delta|$), root mean square error (RMSE), and empirical coverage probability for 95% uniform confidence bands. The results show a general improvement as the sample size N increases. Specifically, the RMSE consistently decreases with N for all estimators. The bias ratios also tend to decrease, although not always monotonically across all quantile points examined. This overall pattern indicates satisfactory finite-sample performance and suggests convergence of the estimators. Furthermore, the empirical coverage rates of the 95% uniform confidence bands approach the nominal 95% level as the sample size grows, supporting the validity of the resampling inference procedure.

Table 3: Estimation and Inference of the Local Distributional and Lorenz Treatment Effects, September 1982 – December 1983

September 1982–December 1983					
Dependent Variable		Ui Claimed		Ui Paid	
Panel A. LDTE $\Delta_{Id}(y)$	$y = y_{0.1}$	−0.018	(0.001)	−0.018	(0.001)
	$y = y_{0.2}$	−0.023	(0.002)	−0.022	(0.002)
	$y = y_{0.3}$	−0.013	(0.002)	−0.017	(0.002)
	$y = y_{0.4}$	−0.013	(0.003)	−0.016	(0.002)
	$y = y_{0.5}$	−0.009	(0.002)	−0.015	(0.002)
	$y = y_{0.6}$	−0.012	(0.002)	−0.013	(0.002)
	$y = y_{0.7}$	−0.007	(0.002)	−0.011	(0.002)
	$y = y_{0.8}$	−0.008	(0.001)	−0.003	(0.001)
	$y = y_{0.9}$	−0.007	(0.001)	−0.008	(0.001)
Test of significance	p -Value	0.000		0.000	
Test of Homogeneity	p -Value	0.000		0.000	
Panel B. LLTE $\Delta_L(\tau)$	$\tau = 0.10$	0.000	(0.000)	0.000	(0.000)
	$\tau = 0.20$	0.001	(0.000)	0.001	(0.000)
	$\tau = 0.30$	0.003	(0.000)	0.003	(0.001)
	$\tau = 0.40$	0.004	(0.001)	0.004	(0.001)
	$\tau = 0.50$	0.006	(0.001)	0.006	(0.001)
	$\tau = 0.60$	0.006	(0.001)	0.006	(0.001)
	$\tau = 0.70$	0.006	(0.001)	0.006	(0.001)
	$\tau = 0.80$	0.005	(0.001)	0.005	(0.001)
	$\tau = 0.90$	0.003	(0.000)	0.001	(0.000)
Test of significance	p -Value	0.000		0.001	
Test of Homogeneity	p -Value	0.000		0.000	

Note: The numbers in parentheses denote the estimated standard errors.

7 CONCLUSION

This paper develops a unified framework for the identification, estimation, and uniform inference of a class of local treatment effects (LTEs) in the sharp regression kink design. The identification strategy applies to Hadamard differentiable functionals of the outcome distribution and utilizes a continuity assumption on the conditional density of the outcome variable at the kink point. This approach yields a unified identification formula and establishes a connection between the LTE and the local average structural derivative.

For estimation and inference, we consider two classes of local polynomial constrained regression estimators and develop their asymptotic theory, including a resampling method for uniform inference. The framework covers parameters such as effects on the mean, the distribution, and inequality measures. We apply the methods to re-examine the effect of unemployment insurance on unemployment durations, estimating the local distributional and Lorenz treatment effects.

This analysis provides information on the impacts of the policy on the shape and dispersion of the duration distribution. An extension of this framework to the fuzzy regression kink design could be a direction for future research, potentially broadening its applicability.

A REGULARITY CONDITIONS AND AUXILIARY LEMMAS

Recall the notation: Let $I_{x_0}^o := I_{x_0} \setminus \{x_0\}$ be a deleted neighborhood of the kink x_0 . Define $b_0 := b(x_0)$ and $y_\tau := Q_{Y|X}(\tau|x_0)$. Define $h(x, e) := g(b(x), x, e)$, and $h_x(x, e) := \frac{\partial}{\partial x} h(x, e) = b'(x)g_1(b(x), x, e) + g_2(b(x), x, e)$. Let $\mathcal{V}(y, x) := \{e \in \mathbb{R}^{d_\varepsilon} : h(x, e) \leq y\}$ be the volume of unobservables yielding an outcome less than or equal to y , and $\partial\mathcal{V}(y, x) := \{e \in \mathbb{R}^{d_\varepsilon} : h(x, e) = y\}$ denote its boundary.

Assumption R1 (Regularity Conditions for Identification).

(i) $g_1(b_0, x_0, \cdot)$ is measurable and satisfies

$$P \left[|g(b_0 + \delta, x_0, \varepsilon) - g(b_0, x_0, \varepsilon) - \delta g_1(b_0, x_0, \varepsilon)| \geq c|\delta| \mid X = x_0 \right] = o(\delta)$$

as $\delta \rightarrow 0$ for all $c > 0$.

(ii) The conditional distribution of $(Y, g_1(b_0, x_0, \varepsilon))$ given $X = x_0$ is absolutely continuous with respect to Lebesgue measure, and its density $f_{Y, g_1|X}(y, y'|x_0)$ is continuous in y for all y' . Moreover, there exists a Lebesgue integrable function $\varpi : \mathbb{R} \rightarrow \mathbb{R}$ with $\int |y' \varpi(y')| dy' < \infty$ such that for all (y, y') , $f_{Y, g_1|X}(y, y'|x_0) \leq |\varpi(y')|$.

(iii) The conditional distribution $x \mapsto F_{Y|X}(y|x)$ is continuously differentiable on $I_{x_0}^o$ for all $y \in \mathcal{Y}_x$. The left and right limits of this derivative at x_0 , $y \mapsto \partial F_{Y|X}(y|x_0^-)/\partial x$ and $y \mapsto \partial F_{Y|X}(y|x_0^+)/\partial x$, are continuous on \mathcal{Y}_{x_0} . Moreover, for all $x \in I_{x_0}^o$ and $t > 0$ such that $x + t \in I_{x_0}^o$, $h(x, \cdot)$ is measurable and satisfies

$$P \left[|h(x + t, \varepsilon) - h(x, \varepsilon) - t h_x(x, \varepsilon)| \geq ct \mid X = x \right] = o(t)$$

as $t \rightarrow 0^+$ for all $c > 0$.

(iv) The conditional distribution of $(Y, h_x(x, \varepsilon))$ given $X = x$ is absolutely continuous with respect to the Lebesgue measure for each $x \in I_{x_0}^o$. Its conditional density, $f_{Y, h_x|X}(y, y'|x)$, satisfies the following conditions: (a) The function $(y, y') \mapsto f_{Y, h_x|X}(y, y'|x)$ is continuous for all $x \in I_{x_0}^o$. (b) There exists a Lebesgue integrable function $\varphi : \mathbb{R} \rightarrow \mathbb{R}$ such that $\int |y' \varphi(y')| dy' < \infty$ and $f_{Y, h_x|X}(y, y'|x) \leq |\varphi(y')|$ for all (y, y') and $x \in I_{x_0}^o$.

(v) The conditional distribution of ε given (Y, X) is absolutely continuous with respect to the Lebesgue measure. Its conditional density, $f_{\varepsilon|Y, X}(e|y, x)$, satisfies the following conditions: (a) The function $(y, x) \mapsto f_{\varepsilon|Y, X}(e|y, x)$ is continuous on $\mathcal{Y}_x \times I_{x_0}$ for all $e \in \mathcal{E}$. (b) There

exists a Lebesgue integrable function $\chi : \mathcal{E} \rightarrow \mathbb{R}$ and positive constants C_1, C_2 such that for all $(e, y, x) : f_{\varepsilon|Y,X}(e|y, x) \leq C_1|\chi(e)|$ for $(y, x) \in \mathcal{Y}_x \times I_{x_0}$, and $|\partial f_{\varepsilon|X}(e|x)/\partial x| \leq C_2|\chi(e)|$ for $x \in I_{x_0}$.

Assumption R1 collects several technical regularity conditions. Parts (i) and (iii), analogous to Assumption A3 in [Hoderlein and Mammen \(2007\)](#), impose smoothness on the structural marginal effects, g_1 and g_2 . Parts (ii), (iv), and (v) are standard moment and boundedness conditions required to justify the application of the Lebesgue dominated convergence theorem in our proofs.

We now restate a result from [Chiang and Sasaki \(2019\)](#) concerning the asymptotic properties of the QRKD estimator. This result provides a basis for the asymptotic theory of our composite RKD estimator.

Lemma A1 ([Chiang and Sasaki, 2019](#), Corollary 2). *Suppose Assumptions 2 and 4 hold. Then,*

$$\sqrt{nh_{n,\tau}^3} \left(\widehat{\text{QRKD}}(\tau) - \text{QRKD}(\tau) \right) \rightsquigarrow \mathbb{G}^Q(\tau) := \frac{\mathbb{Z}^Q(\tau, 2) - \mathbb{Z}^Q(\tau, 3)}{b'(x_0^+) - b'(x_0^-)},$$

uniformly in $\tau \in \mathcal{T} \subset (0, 1)$, where $\mathbb{Z}^Q(\cdot, 2)$ and $\mathbb{Z}^Q(\cdot, 3)$ are tight zero-mean Gaussian processes of the normalized quantile derivative processes. $\mathbb{G}^Q : \Omega \rightarrow \ell^\infty(\mathcal{T})$ is a tight zero-mean Gaussian process with covariance function

$$E \left[\mathbb{G}^Q(\tau_1) \mathbb{G}^Q(\tau_2) \right] = \frac{(\tau_1 \wedge \tau_2 - \tau_1 \tau_2) \cdot (\iota_2 - \iota_3)^\top \bar{I}_p^{-1} T(\tau_1, \tau_2) \bar{I}_p^{-1} (\iota_2 - \iota_3)}{(b'(x_0^+) - b'(x_0^-))^2 f_X(x_0) f_{Y|X}(y_{\tau_1}|x_0) f_{Y|X}(y_{\tau_2}|x_0)}$$

for all $\tau_1, \tau_2 \in \mathcal{T}$, where $y_\tau := Q_{Y|X}(\tau|x_0)$, and

$$T(\tau_1, \tau_2) := \frac{1}{\sqrt{c(\tau_1)c(\tau_2)}} \int_{\mathbb{R}} \bar{r}_p \left(\frac{u}{c(\tau_1)} \right) \bar{r}_p \left(\frac{u}{c(\tau_2)} \right)^\top K \left(\frac{u}{c(\tau_1)} \right) K \left(\frac{u}{c(\tau_2)} \right) du.$$

Lemma A2 (First Stage Consistency). *Let*

$$\hat{\varepsilon}^m(y, x, \theta) := (\varphi(y, \theta) - \bar{r}_p(x - x_0) \hat{\alpha}(\theta)) \mathbb{1}(|x - x_0| \leq h_{n,\theta}).$$

Under Assumption 3, $\hat{\varepsilon}^m(y, x, \theta) = \varepsilon^m(y, x, \theta) \mathbb{1}(|x - x_0| \leq h_{n,\theta}) + o_P(1)$ uniformly in $\theta \in \bar{\Theta}$.

B BANDWIDTH SELECTION

We first recall the necessary notation. Let $\bar{r}_p(u) := (1, u\delta_u^+, u\delta_u^-, \dots, u^p\delta_u^+, u^p\delta_u^-)^\top$ be the $(2p+1)$ -dimensional vector of regressors, where $\delta_u^+ := \mathbb{1}(u \geq 0)$ and $\delta_u^- := \mathbb{1}(u < 0)$. The following kernel-dependent constants are used in the derivation: $\bar{I}_p := \int_{\mathbb{R}} \bar{r}_p(u) \bar{r}_p(u)^\top K(u) du$; $\bar{\vartheta}_{p,q}^\pm := \int_{\mathbb{R}_\pm} \bar{r}_p(u) u^q K(u) du$; $\bar{\Psi}_p^\pm := \int_{-\pm} \bar{r}_p(u) \bar{r}_p(u)^\top K^2(u) du$; $\bar{\Psi}_p := \int_{\mathbb{R}} \bar{r}_p(u) \bar{r}_p(u)^\top K^2(u) du$.

B.1 Bandwidth Selector for Estimating RKD_m

This section details the procedure for selecting a pointwise mean squared error (MSE)-optimal bandwidth for estimating the difference in derivatives, $m^{(1)}(\theta, x_0^+) - m^{(1)}(\theta, x_0^-)$, using our constrained regression estimator. Under Assumption 3(iv), the asymptotic MSE (AMSE) of $\hat{m}^{(1)}(\theta, x_0^+) - \hat{m}^{(1)}(\theta, x_0^-)$ is given by

$$\begin{aligned} \text{AMSE} \left[\hat{m}^{(1)}(\theta, x_0^+) - \hat{m}^{(1)}(\theta, x_0^-) \right] &= \left(\text{ABias} \left[\hat{m}^{(1)}(\theta, x_0^+) - \hat{m}^{(1)}(\theta, x_0^-) \right] \right)^2 \\ &\quad + \text{AVar} \left[\hat{m}^{(1)}(\theta, x_0^+) - \hat{m}^{(1)}(\theta, x_0^-) \right] \\ &= h_{n,\theta}^{2p} \mathbf{B}_{1,p}^2(\theta) + \frac{1}{nh_{n,\theta}^3} \mathbf{V}_{1,p}(\theta), \end{aligned}$$

where for every $\nu, p \in \mathbb{N}_+$ with $1 \leq \nu \leq p$,

$$\begin{aligned} \mathbf{B}_{\nu,p}(\theta) &:= \frac{\iota_{2\nu}^\top \bar{\Gamma}_p^{-1} \left[m^{(p+1)}(\theta, x_0^+) \bar{\vartheta}_{p,p+1}^+ + m^{(p+1)}(\theta, x_0^-) \bar{\vartheta}_{p,p+1}^- \right]}{(p+1)!} \\ &\quad - \frac{\iota_{2\nu+1}^\top \bar{\Gamma}_p^{-1} \left[m^{(p+1)}(\theta, x_0^+) \bar{\vartheta}_{p,p+1}^+ + m^{(p+1)}(\theta, x_0^-) \bar{\vartheta}_{p,p+1}^- \right]}{(p+1)!} \end{aligned} \quad (\text{A.1})$$

$$\begin{aligned} \mathbf{V}_{\nu,p}(\theta) &:= \frac{\iota_{2\nu}^\top \bar{\Gamma}_p^{-1} \left[\sigma_{\varepsilon^m}(\theta, \theta|x_0^+) \bar{\Psi}_p^+ + \sigma_{\varepsilon^m}(\theta, \theta|x_0^-) \bar{\Psi}_p^- \right] \bar{\Gamma}_p^{-1} \iota_{2\nu}}{f_X(x_0)} \\ &\quad + \frac{\iota_{2\nu+1}^\top \bar{\Gamma}_p^{-1} \left[\sigma_{\varepsilon^m}(\theta, \theta|x_0^+) \bar{\Psi}_p^+ + \sigma_{\varepsilon^m}(\theta, \theta|x_0^-) \bar{\Psi}_p^- \right] \bar{\Gamma}_p^{-1} \iota_{2\nu+1}}{f_X(x_0)} \\ &\quad - \frac{2\iota_{2\nu}^\top \bar{\Gamma}_p^{-1} \left[\sigma_{\varepsilon^m}(\theta, \theta|x_0^+) \bar{\Psi}_p^+ + \sigma_{\varepsilon^m}(\theta, \theta|x_0^-) \bar{\Psi}_p^- \right] \bar{\Gamma}_p^{-1} \iota_{2\nu+1}}{f_X(x_0)}. \end{aligned} \quad (\text{A.2})$$

Following the general approach for MSE-optimal bandwidth selection in local polynomial regression (e.g., [Calonico et al. \(2014\)](#)), we can derive the optimal bandwidth for our constrained estimator. For a general problem of estimating a ν th-order derivative ratio with a p th-order local polynomial, the MSE-optimal bandwidth takes the form:

$$\begin{aligned} h_{\nu,p}(\theta) &= \arg \min_{h>0} \left\{ h^{2(p+1-\nu)} \mathbf{B}_{\nu,p}^2(\theta) + \frac{1}{nh^{1+2\nu}} \mathbf{V}_{\nu,p}(\theta) \right\} \\ &= \left\{ \frac{1+2\nu}{2(p+1-\nu)} \frac{\mathbf{V}_{\nu,p}(\theta)}{\mathbf{B}_{\nu,p}^2(\theta)} \right\}^{1/(2p+3)} n^{-1/(2p+3)}. \end{aligned}$$

Here, $\mathbf{V}_{\nu,p}(\theta)$ and $\mathbf{B}_{\nu,p}(\theta)$ represent the leading terms of the asymptotic variance and bias, respectively. Our specific estimand, $\text{RKD}_m(\theta)$, is a ratio of first-order derivatives, which corresponds to the case $\nu = 1$. Therefore, the pointwise MSE-optimal bandwidth for our p th-order estimator is given by $h_{n,\theta} = h_{1,p}(\theta)$. Implementing this choice in practice requires a feasible procedure to estimate the unknown variance and bias terms. The following algorithm details these steps.

Algorithm 1 (Plug-in Bandwidth Selector for Estimating RKD_m).

Fix positive integers p, q such that $p < q$.

Step 1. (Pilot Bandwidths) The pilot bandwidth $\hat{b}_{n,\theta}$ is given by

$$\hat{b}_{n,\theta} = \left\{ \frac{2p+3}{2(q-p)} \frac{\tilde{V}_{p+1,q}(\theta)}{\tilde{B}_{p+1,q}^2(\theta)} \right\}^{1/(2q+3)} n^{-1/(2q+3)}, \quad \theta \in \bar{\Theta}.$$

Calculating this requires an estimate of $\tilde{B}_{p+1,q}$ and $\tilde{V}_{p+1,q}$, which in turn depends on the $(q+1)$ -th order derivatives, $\tilde{m}^{(q+1)}(\theta, x_0^\pm)$, and the conditional variance of the regression error, $\sigma_{\varepsilon^m}(\theta, \theta|x_0^\pm)$, respectively. To obtain these high-order derivative estimates, we fit a $(q+1)$ -th order global polynomial regression using our constrained specification:

$$\tilde{\gamma}_{q+1}(\theta) = \arg \min_{\gamma \in \mathbb{R}^{2q+3}} \sum_{i=1}^n (\varphi(Y_i, \theta) - \bar{r}_{q+1}(X_i - x_0)^\top \gamma)^2.$$

The required derivative estimates are then extracted from the resulting coefficient vector: $\tilde{m}^{(q+1)}(\theta, x_0^+) = \iota_{2q+2}^\top \tilde{\gamma}_{q+1}(\theta)/(q+1)!$ and $\tilde{m}^{(p+1)}(\theta, x_0^-) = \iota_{2q+3}^\top \tilde{\gamma}_{q+1}(\theta)/(q+1)!$.

Next, we estimate this conditional variance $\sigma_{\varepsilon^m}(\theta, \theta|x_0^\pm)$ using a two-stage procedure. First, we obtain the fitted residuals from the global polynomial regression in the previous step: $\tilde{\varepsilon}^m(y, x, \theta) = \varphi(y, \theta) - \bar{r}_{q+1}(x - x_0)^\top \tilde{\gamma}_{q+1}(\theta)$. Second, we estimate the conditional mean at x_0 from the left and right by running separate local linear regressions:

$$\tilde{\sigma}_{\varepsilon^m}(\theta, \theta|x_0^\pm) = \iota_1^\top \arg \min_{\kappa \in \mathbb{R}^3} \sum_{i=1}^n \delta_i^\pm ([\tilde{\varepsilon}^m(Y_i, X_i, \theta)]^2 - (1, X_i - x_0)^\top \kappa)^2.$$

The density $f_X(x_0)$ is estimated by

$$\hat{f}_X(x_0) = \frac{1}{nv_n} \sum_{i=1}^n K\left(\frac{X_i - x_0}{v_n}\right),$$

where $v_n = 2.576 \cdot \min\{sd(X), IQR(X)/1.349\} \cdot n^{-1/5}$; sd denotes the sample standard error and IQR denotes the interquartile range.

Step 2. (Main Bandwidths) The main bandwidth $\hat{h}_{n,\theta}$ is given by

$$\hat{h}_{n,\theta} = \hat{h}_{1,p}(\theta) = \left\{ \frac{3}{2p} \frac{\hat{V}_{1,p}(\theta)}{\hat{B}_{1,p}^2(\theta)} \right\}^{1/(2p+3)} n^{-1/(2p+3)}, \quad \theta \in \bar{\Theta}.$$

To estimate the $(p+1)$ th-order derivatives needed for the bias term, we fit a $(p+1)$ th-order local polynomial constrained regression, using the pilot bandwidth $\hat{b}_{n,\theta}$:

$$\hat{\gamma}_{p+1}(\theta) = \arg \min_{\gamma \in \mathbb{R}^{2p+3}} \sum_{i=1}^n (\varphi(Y_i, \theta) - \bar{r}_{p+1}(X_i - x_0)^\top \gamma)^2 K\left(\frac{X_i - x_0}{\hat{b}_{n,\theta}}\right).$$

The required derivatives, $\hat{m}^{(p+1)}(\theta, x_0^\pm)$, are then extracted from the coefficient vector: $\hat{m}^{(p+1)}(\theta, x_0^+) = \iota_{2p+2}^\top \hat{\gamma}_{p+1}(\theta)/(p+1)!$, and $\hat{m}^{(p+1)}(\theta, x_0^-) = \iota_{2p+3}^\top \hat{\gamma}_{p+1}(\theta)/(p+1)!$.

To estimate the term $\sigma_{\varepsilon^m}(\theta, \theta|x_0^\pm)$ in $\hat{V}_{1,p}$, we first obtain fitted residuals from the regression in the step above: $\hat{\varepsilon}^m(y, x, \theta) = \varphi(y, \theta) - \bar{r}_{p+1}(x - x_0)^\top \hat{\gamma}_{p+1}(\theta)$. We then estimate the conditional error variance, $\hat{\sigma}_{\varepsilon^m}(\theta, \theta|x_0^\pm)$, by running separate local linear regressions of the squared residuals on each side of the kink, again using the pilot bandwidth $\hat{b}_{n,\theta}$:

$$\hat{\sigma}_{\varepsilon^m}(\theta, \theta|x_0^\pm) = \iota_1^\top \arg \min_{\kappa \in \mathbb{R}^3} \sum_{i=1}^n \delta_i^\pm ([\hat{\varepsilon}^m(Y_i, X_i, \theta)]^2 - (1, X_i - x_0)^\top \kappa)^2 K\left(\frac{X_i - x_0}{\hat{b}_{n,\theta}}\right).$$

The estimator of the density $\hat{f}_X(x_0)$ is the same as that used in Step 1.

B.2 Bandwidth Selector for Estimating the LLTE Δ_L

Deriving the fully optimal bandwidth for the composite local Lorenz treatment effect estimator is complex. We therefore adopt a practical approximation based on the standard MSE-optimal formula for a first-order derivative:

$$h_{n,\tau}^L = \left\{ \frac{3}{2p} \frac{V_{1,p}^L(\tau)}{[B_{1,p}^L(\tau)]^2} \right\}^{1/(2p+3)} n^{-1/(2p+3)}, \quad \tau \in \mathcal{T},$$

where, for $\nu, p \in \mathbb{N}_+$ with $1 \leq \nu \leq p$,

$$\begin{aligned} B_{\nu,p}^L(\tau) &:= \frac{1}{\mu_0} \left(\int_0^\tau B_{\nu,p}^Q(u) du - L_{Y|X}(\tau|x_0) B_{\nu,p}^\mu \right), \\ V_{\nu,p}^L(\tau) &:= \frac{1}{\mu_0^2} \left(\int_0^\tau V_{\nu,p}^Q(u) du + L_{Y|X}^2(\tau|x_0) V_{\nu,p}^\mu \right). \end{aligned}$$

The terms $B_{1,p}^L$ and $V_{1,p}^L$ are constructed from the asymptotic bias and variance components of the underlying mean and quantile estimators. Specifically, the components related to the mean effect, $B_{1,p}^\mu$ and $V_{1,p}^\mu$, are as defined in Equations (A.1) and (A.2) for the case where $m(x, \theta) = E[Y|X = x]$. The components related to the quantile effect, $B_{1,p}^Q$ and $V_{1,p}^Q$, represent the leading bias and variance terms, respectively, of the estimated difference in quantile derivatives, $\hat{Q}_{Y|X}^{(1)}(\tau|x_0^+) - \hat{Q}_{Y|X}^{(1)}(\tau|x_0^-)$. They are given by:

$$\begin{aligned} B_{\nu,p}^Q(\tau) &= \frac{\iota_{2\nu}^\top \bar{F}_p^{-1} \left(Q_{Y|X}^{(p+1)}(\tau|x_0^+) \bar{\vartheta}_{p,p+1}^+ + Q_{Y|X}^{(p+1)}(\tau|x_0^-) \bar{\vartheta}_{p,p+1}^- \right)}{(p+1)!} \\ &\quad - \frac{\iota_{2\nu+1}^\top \bar{F}_p^{-1} \left(Q_{Y|X}^{(p+1)}(\tau|x_0^+) \bar{\vartheta}_{p,p+1}^+ + Q_{Y|X}^{(p+1)}(\tau|x_0^-) \bar{\vartheta}_{p,p+1}^- \right)}{(p+1)!}, \end{aligned}$$

$$V_{\nu,p}^Q(\tau) = \frac{\tau(1-\tau) [\iota_{2v}^\top \bar{\Gamma}_p^{-1} \bar{\Psi}_p \bar{\Gamma}_p^{-1} \iota_{2v} + \iota_{2v+1}^\top \bar{\Gamma}_p^{-1} \bar{\Psi}_p \bar{\Gamma}_p^{-1} \iota_{2v+1} - 2\iota_{2v}^\top \bar{\Gamma}_p^{-1} \bar{\Psi}_p \bar{\Gamma}_p^{-1} \iota_{2v+1}]}{f_X(x_0)f_{Y|X}(y_\tau|x_0)}.$$

The following algorithm details the bandwidth selection procedure for the local Lorenz treatment effect.

Algorithm 2 (Plug-in Bandwidth Selector for Estimating Δ_L).

Fix positive integers p, q such that $p < q$. Define a fine grid of quantile levels, e.g., $\mathcal{U} = \{0.01, 0.02, \dots, 0.99\}$.

Step 1. (Pilot Bandwidths) The pilot bandwidths are given by

$$\hat{b}_n = \left\{ \frac{2p+3}{2(q-p)} \frac{\hat{V}_{p+1,q}^\mu}{[\hat{B}_{p+1,q}^\mu]^2} \right\}^{1/(2q+3)} n^{-1/(2q+3)},$$

$$\text{and } \hat{b}_{n,u} = \left\{ \frac{2p+3}{2(q-p)} \frac{\hat{V}_{p+1,q}^Q(u)}{[\hat{B}_{p+1,q}^Q(u)]^2} \right\}^{1/(2q+3)} n^{-1/(2q+3)}, \quad u \in \mathcal{U}.$$

The terms $\hat{B}_{p+1,q}^\mu$ and $\hat{V}_{p+1,q}^\mu$ are independent of u and are computed once following the procedure for the conditional mean in Algorithm 1.

To estimate the $(q+1)$ th-order derivatives $\hat{Q}_{Y|X}^{(q+1)}(u|x_0^\pm)$ needed for the quantile bias term $\hat{B}_{p+1,q}^Q$. We fit a $(q+1)$ th-order global constrained polynomial quantile regression to obtain the coefficient vector:

$$\hat{\gamma}_{q+1}(u) = \arg \min_{\gamma \in \mathbb{R}^{2q+3}} \sum_{i=1}^n \rho_u(Y_i - \bar{r}_{q+1}(X_i - x_0)^\top \gamma).$$

Then, we extract the derivative estimates: $\hat{y}_u = \iota_1^\top \hat{\gamma}_{q+1}(u)$, $\hat{Q}_{Y|X}^{(q+1)}(u|x_0^+) = \iota_{2q+2}^\top \hat{\gamma}_{q+1}(u)/(q+1)!$, and $\hat{Q}_{Y|X}^{(q+1)}(u|x_0^-) = \iota_{2q+3}^\top \hat{\gamma}_{q+1}(u)/(q+1)!$.

Next, the (conditional) densities f_X and $f_{Y|X}$ in $\hat{V}_{p+1,q}^Q(u)$ are estimated as follows:

$$\hat{f}_X(x_0) = \frac{1}{nv_n} \sum_{i=1}^n K\left(\frac{X_i - x_0}{v_n}\right),$$

$$\text{and } \hat{f}_{Y|X}(y|x_0) = \frac{1}{h_{1,n}} \frac{\sum_{i=1}^n K\left(\frac{Y_i - y}{h_{1,n}}\right) K\left(\frac{X_i - x_0}{h_{2,n}}\right)}{\sum_{i=1}^n K\left(\frac{X_i - x_0}{h_{2,n}}\right)}, \quad (\text{A.3})$$

where v_n is given by Step 1 of Algorithm 1. We select the bandwidths $h_{1,n}$ and $h_{2,n}$ in $\hat{f}_{Y|X}$ following the rule-of-thumb procedure proposed by [Bashtannyk and Hyndman \(2001, Section 2.2\)](#):

$$h_{2,n} = \left\{ \frac{16cR_K^2\sigma_Y^5[288\pi^9\sigma_X^{58}\lambda^2(c)]^{1/8}}{\varrho_K^4 b^{5/2} v^{3/4}(c) \{v^{1/2}(c) + b[18\pi\sigma_X^{10}\lambda^2(c)]^{1/4}\}} \right\}^{1/6} n^{-1/6},$$

$$\text{and } h_{1,n} = \left\{ \frac{b^2 v(c)}{3\sqrt{2\pi}\sigma_X^5 \lambda(c)} \right\}^{1/4} h_{2,n},$$

where σ_Y and σ_X denote the sample standard deviation; $\lambda(c) = \int_{-c}^c \varphi(t) dt$; $v(c) = \sqrt{2\pi}\sigma_X^3(3q\sigma_X^2 + 8\sigma_Y^2)\lambda(c) - 16c\sigma_X^2\sigma_Y^2 e^{-k^2/2}$; q is the slope of an OLS regression of Y_i on $[1, X_i]$. $R_K = \int K^2(u) du$ and $\varrho_K = \int u^2 K(u) du$. Typically, constant c can be set to 2 or 3.

Step 2. (Main Bandwidths) The main bandwidth is given by

$$\hat{h}_{n,\tau}^L = \left\{ \frac{3}{2p} \frac{\hat{V}_{1,p}^L(\tau)}{[\hat{B}_{1,p}^L(\tau)]^2} \right\}^{1/(2p+3)} n^{-1/(2p+3)}, \quad \tau \in \mathcal{T}$$

where $\hat{L}_{Y|X}(\tau|x_0) = \int_0^\tau \hat{g}_u du / \hat{\mu}_0$;

$$\begin{aligned} \hat{B}_{1,p}^L(\tau) &= \frac{1}{\mu_0} \left(\int_0^\tau \hat{B}_{1,p}^Q(u) du - \hat{L}_{Y|X}(\tau|x_0) \hat{B}_{1,p}^\mu \right), \\ \text{and } \hat{V}_{1,p}^L(\tau) &= \frac{1}{\mu_0^2} \left(\int_0^\tau \hat{V}_{1,p}^Q(u) du + \hat{L}_{Y|X}^2(\tau|x_0) \hat{V}_{1,p}^\mu \right). \end{aligned}$$

The terms related to the mean effect— $\hat{\mu}_0$, $\hat{B}_{1,p}^\mu$, and $\hat{V}_{1,p}^\mu$ —are independent of the quantile level u . They are computed once as described in Algorithm 1, Step 2, using the pilot bandwidth \hat{b}_n .

The following components are computed for each $u \in \mathcal{U}$: To estimate the $(p+1)$ -th order derivatives, we fit a $(p+1)$ -th order local constrained polynomial quantile regression, using the corresponding pilot bandwidth $\hat{b}_{n,u}$:

$$\hat{\gamma}_{p+1}(u) = \arg \min_{\gamma \in \mathbb{R}^{2p+3}} \sum_{i=1}^n \rho_\tau(Y_i - \bar{r}_{p+1}(X_i - x_0)^\top \gamma) K\left(\frac{X_i - x_0}{\hat{b}_{n,u}}\right).$$

The intercept and derivative estimates are then extracted from the coefficient vector: $\hat{g}_u = \iota_1^\top \hat{\gamma}_{p+1}(u)$, $\hat{Q}_{Y|X}^{(p+1)}(u|x_0^+) = \iota_{2p+2}^\top \hat{\gamma}_{p+1}(u)/(p+1)!$, and $\hat{Q}_{Y|X}^{(p+1)}(u|x_0^-) = \iota_{2p+3}^\top \hat{\gamma}_{p+1}(u)/(p+1)!$. The conditional densities, f_X and $f_{Y|X}$, required for the term $\hat{V}_{1,p}^Q(u)$ are estimated using standard kernel methods as detailed in Equation (A.3).

Remark B.1. The procedure in Algorithm 2 can be straightforwardly adapted to find the MSE-optimal bandwidth for the standard QRKD estimator, $\widehat{\text{QRKD}}(\tau)$. The steps are analogous. For each quantile of interest $\tau \in \mathcal{T}$: (1) A pilot bandwidth, $\hat{b}_{n,\tau}$, is computed as described in Step 1 of the algorithm. (2) The bias and variance components, $\hat{B}_{1,p}^Q(\tau)$ and $\hat{V}_{1,p}^Q(\tau)$, are computed as described in Step 2, using the pilot bandwidth $\hat{b}_{n,\tau}$ from the previous step. These components

are then plugged into the standard formula for the main optimal bandwidth:

$$\hat{h}_{n,\tau} = \left\{ \frac{3}{2p} \frac{\hat{V}_{1,p}^Q(\tau)}{[\hat{B}_{1,p}^Q(\tau)]^2} \right\}^{1/(2p+3)} n^{-1/(2p+3)}, \quad \tau \in \mathcal{T},$$

This provides a MSE-optimal bandwidth for the p th-order local polynomial constrained quantile regression estimator in (5). Our result generalizes the bandwidth selector in Appendix C.1 of [Chiang and Sasaki \(2019\)](#), which focuses on the local linear ($p = 1$) case.

REFERENCES

- Bashtannyk, D. M. and Hyndman, R. J. (2001). Bandwidth selection for kernel conditional density estimation. *Computational Statistics & Data Analysis*, 36(3):279–298.
- Bhattacharya, D. (2007). Inference on inequality from household survey data. *Journal of Econometrics*, 137(2):674–707.
- Calonico, S., Cattaneo, M. D., and Titiunik, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, 82(6):2295–2326.
- Card, D., Johnston, A., Leung, P., Mas, A., and Pei, Z. (2015a). The effect of unemployment benefits on the duration of unemployment insurance receipt: New evidence from a regression kink design in missouri, 2003–2013. *American Economic Review*, 105(5):126–130.
- Card, D., Lee, D. S., Pei, Z., and Weber, A. (2015b). Inference on causal effects in a generalized regression kink design. *Econometrica*, 83(6):2453–2483.
- Chen, H., Chiang, H. D., and Sasaki, Y. (2020). Quantile treatment effects in regression kink designs. *Econometric Theory*, 36(6):1167–1191.
- Chen, H. and Fan, Y. (2019). Identification and wavelet estimation of weighted ate under discontinuous and kink incentive assignment mechanisms. *Journal of Econometrics*, 212(2):476–502.
- Chernozhukov, V., Fernández-Val, I., and Galichon, A. (2010). Quantile and probability curves without crossing. *Econometrica*, 78(3):1093–1125.
- Chernozhukov, V., Fernandez-Val, I., Hoderlein, S., Holzmann, H., and Newey, W. (2015). Nonparametric identification in panels using quantiles. *Journal of Econometrics*, 188(2):378–392.
- Chiang, H. D., Hsu, Y.-C., and Sasaki, Y. (2019). Robust uniform inference for quantile treatment effects in regression discontinuity designs. *Journal of Econometrics*, 211(2):589–618.
- Chiang, H. D. and Sasaki, Y. (2019). Causal inference by quantile regression kink designs. *Journal of Econometrics*, 210(2):405–433.

- Dong, Y. (2018). Jump or kink? regression probability jump and kink design for treatment effect evaluation. *Unpublished manuscript*.
- Florens, J.-P., Heckman, J. J., Meghir, C., and Vytlacil, E. (2008). Identification of treatment effects using control functions in models with continuous, endogenous treatment and heterogeneous effects. *Econometrica*, 76(5):1191–1206.
- Hoderlein, S., Holzmann, H., Kasy, M., and Meister, A. (2017). Corrigendum: Instrumental variables with unrestricted heterogeneity and continuous treatment. *The Review of Economic Studies*, 84(2):964–968.
- Hoderlein, S. and Mammen, E. (2007). Identification of marginal effects in nonseparable models without monotonicity. *Econometrica*, 75(5):1513–1518.
- Hoderlein, S. and Mammen, E. (2009). Identification and estimation of local average derivatives in non-separable models without monotonicity. *The Econometrics Journal*, 12(1):1–25.
- Imbens, G. W. and Newey, W. K. (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica*, 77(5):1481–1512.
- Imbens, G. W. and Rubin, D. B. (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge university press.
- Kasy, M. (2014). Instrumental variables with unrestricted heterogeneity and continuous treatment. *The Review of Economic Studies*, 81(4):1614–1636.
- Kolsrud, J., Landais, C., Nilsson, P., and Spinnewijn, J. (2018). The optimal timing of unemployment benefits: Theory and evidence from sweden. *American Economic Review*, 108(4-5):985–1033.
- Kosorok, M. R. (2008). *Introduction to empirical processes and semiparametric inference*. Springer.
- Landais, C. (2015). Assessing the welfare effects of unemployment benefits using the regression kink design. *American Economic Journal: Economic Policy*, 7(4):243–278.
- Landais, C. and Spinnewijn, J. (2021). The value of unemployment insurance. *The Review of Economic Studies*, 88(6):3041–3085.
- Qu, Z. and Yoon, J. (2015). Nonparametric estimation and inference on conditional quantile processes. *Journal of Econometrics*, 185(1):1–19.
- Qu, Z. and Yoon, J. (2019). Uniform inference on quantile effects under sharp regression discontinuity designs. *Journal of Business & Economic Statistics*, 37(4):625–647.

- Rubin, D. B. (2005). Causal inference using potential outcomes: Design, modeling, decisions. *Journal of the American statistical Association*, 100(469):322–331.
- Sasaki, Y. (2015). What do quantile regressions identify for general structural functions? *Econometric Theory*, 31(5):1102–1116.
- Van der Vaart, A. W. (2000). *Asymptotic statistics*, volume 3. Cambridge university press.

Figure 1: Estimates of Local Distributional and Lorenz Treatment Effects, September 1981 – September 1982

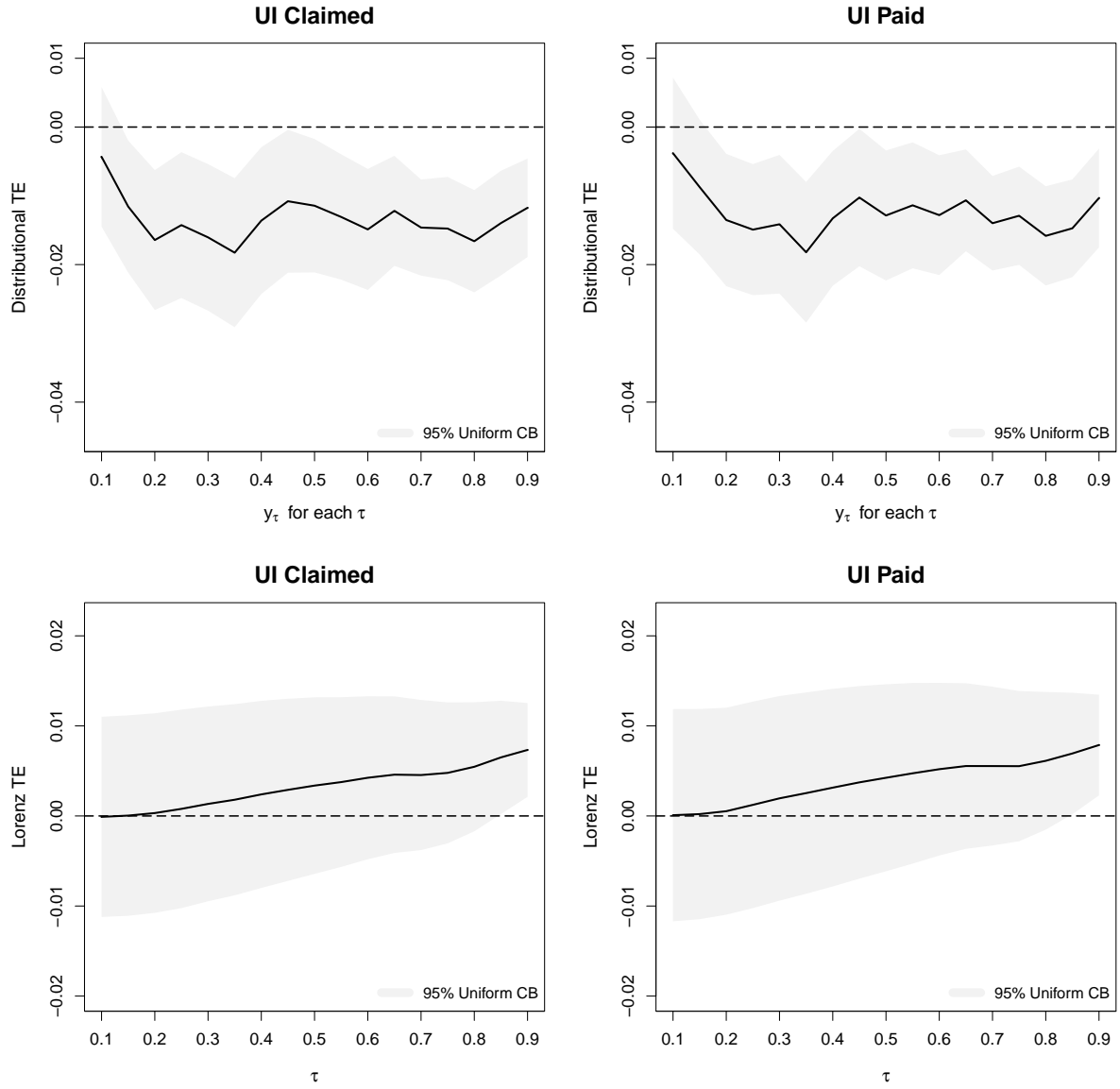


Figure 2: Estimates of Local Distributional and Lorenz Treatment Effects, September 1982 – December 1983

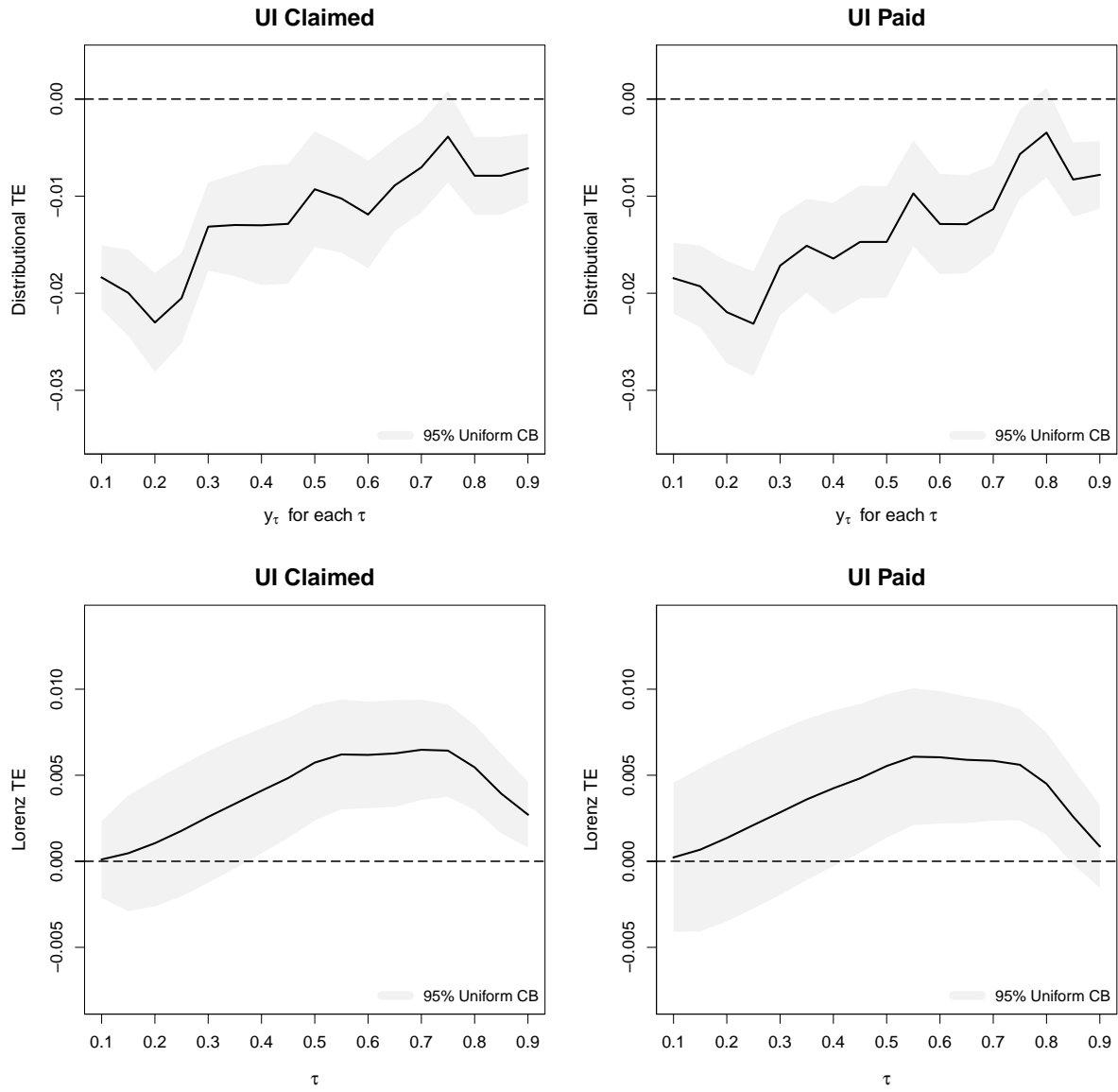
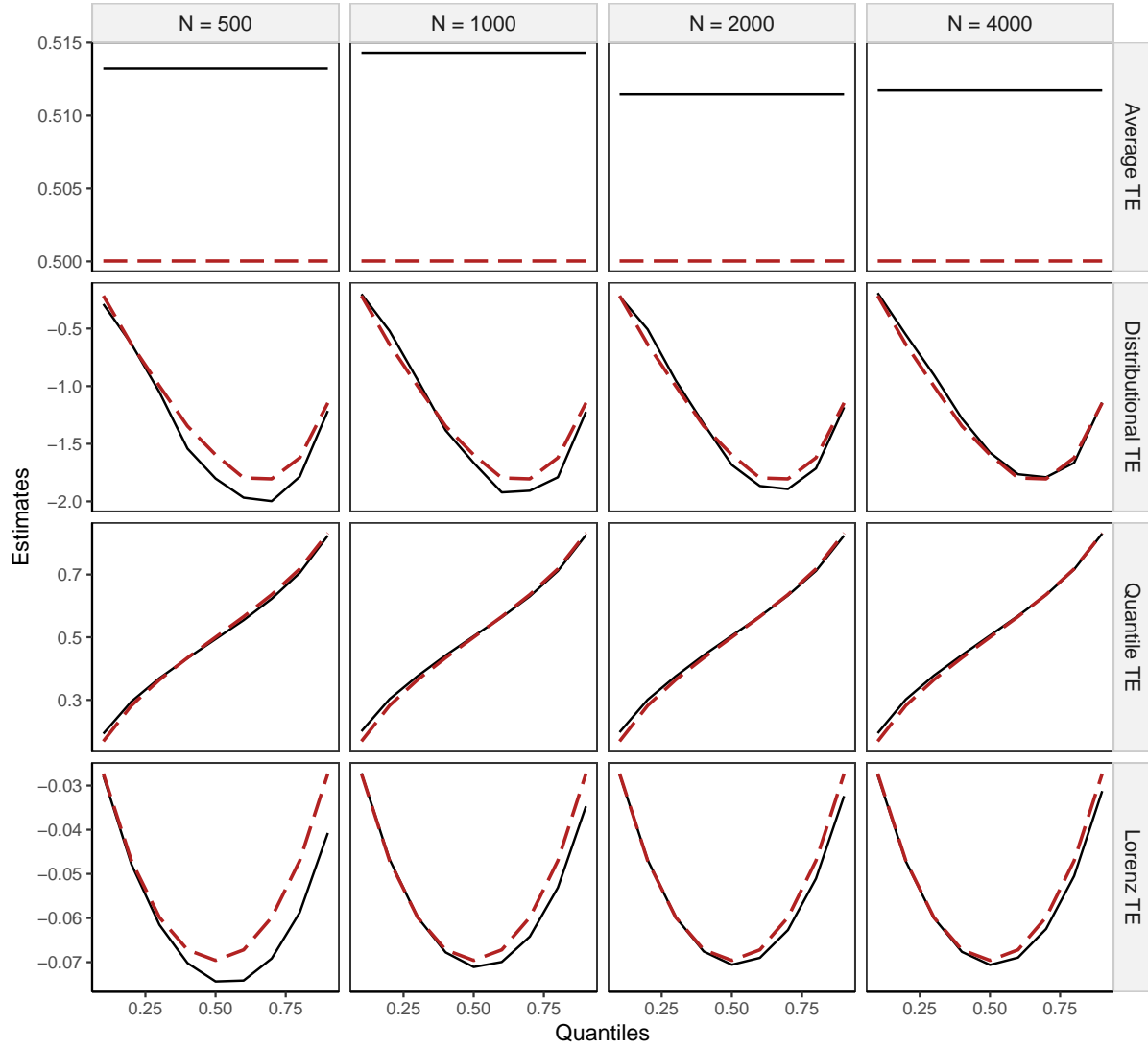


Table 4: Simulated Statistics of the Local Treatment Effects

$N =$		Bias Ratio			RMSE			Unif Cvg		
		1000	2000	4000	1000	2000	4000	1000	2000	4000
Δ_μ		0.029	0.023	0.023	0.118	0.084	0.061	0.953	0.953	0.946
$\Delta_{Id}(y)$	$y = y_{0.1}$	0.069	0.022	0.111	3.595	2.820	2.248	0.977	0.979	0.978
	$y = y_{0.2}$	0.183	0.204	0.132	4.087	3.225	2.605			
	$y = y_{0.3}$	0.054	0.049	0.098	4.577	3.569	2.877			
	$y = y_{0.4}$	0.027	0.014	0.049	4.947	3.800	3.126			
	$y = y_{0.5}$	0.044	0.055	0.012	5.137	4.081	3.384			
	$y = y_{0.6}$	0.070	0.039	0.018	5.228	4.258	3.481			
	$y = y_{0.7}$	0.057	0.049	0.008	5.026	4.129	3.388			
	$y = y_{0.8}$	0.104	0.057	0.027	4.459	3.696	2.980			
	$y = y_{0.9}$	0.068	0.033	0.002	2.956	2.442	2.029			
$\Delta_Q(\tau)$	$\tau = 0.10$	0.190	0.172	0.154	0.182	0.131	0.096	0.892	0.938	0.967
	$\tau = 0.20$	0.073	0.066	0.069	0.169	0.123	0.089			
	$\tau = 0.30$	0.033	0.031	0.037	0.165	0.120	0.086			
	$\tau = 0.40$	0.018	0.019	0.020	0.164	0.118	0.086			
	$\tau = 0.50$	0.007	0.010	0.011	0.162	0.117	0.086			
	$\tau = 0.60$	0.003	0.001	0.004	0.162	0.118	0.087			
	$\tau = 0.70$	0.009	0.004	0.001	0.164	0.117	0.086			
	$\tau = 0.80$	0.009	0.009	0.002	0.170	0.122	0.086			
	$\tau = 0.90$	0.007	0.009	0.002	0.190	0.132	0.091			
$\Delta_L(\tau)$	$\tau = 0.10$	0.003	0.005	0.014	0.014	0.010	0.007	0.871	0.917	0.945
	$\tau = 0.20$	0.006	0.003	0.003	0.023	0.016	0.011			
	$\tau = 0.30$	0.000	0.000	0.002	0.030	0.021	0.014			
	$\tau = 0.40$	0.009	0.006	0.007	0.036	0.025	0.017			
	$\tau = 0.50$	0.022	0.014	0.015	0.041	0.029	0.019			
	$\tau = 0.60$	0.041	0.027	0.026	0.046	0.031	0.021			
	$\tau = 0.70$	0.073	0.048	0.044	0.054	0.034	0.022			
	$\tau = 0.80$	0.130	0.086	0.074	0.056	0.036	0.023			
	$\tau = 0.90$	0.271	0.186	0.145	0.059	0.037	0.023			

Figure 3: Simulated Estimates of The Local Treatment Effects across Sample Sizes



Note: The red dashed lines correspond to the true values of each local treatment effect, whereas the black solid lines represent their estimates.