Econometric Theory, **0**, 2023, 1–46.
doi:10.1017/S026646662300021X

IDENTIFICATION AND INFERENCE IN A QUANTILE REGRESSION DISCONTINUITY DESIGN UNDER RANK SIMILARITY WITH COVARIATES

ZEQUN JIN
School of Economics, Shanghai University of Finance and Economics

YU ZHANG
School of Economics, Shanghai University of Finance and Economics

ZHENGYU ZHANG
School of Economics, Shanghai University of Finance and Economics

YAHONG ZHOU
School of Economics, Shanghai University of Finance and Economics

This study investigates the identification and inference of quantile treatment effects (QTEs) in a fuzzy regression discontinuity (RD) design under *rank similarity*. Unlike Frandsen et al. (2012, *Journal of Econometrics* 168, 382–395), who focus on QTEs only for the compliant subpopulation, our approach can identify QTEs and average treatment effect for the whole population at the threshold. We derived a new set of moment restrictions for the RD model by imposing a local rank similarity condition, which restricts the evolution of individual ranks across treatment status in a neighborhood around the threshold. Based on the moment restrictions, we derive closed-form solutions for the estimands of the potential outcome cumulative distribution functions for the whole population. We demonstrate the functional central limit theorems and bootstrap validity results for the QTE estimators by explicitly accounting for observed covariates. In particular, we develop a multiplier bootstrap-based inference method with robustness against large bandwidths that applies to uniform inference by extending the recent work of Chiang et al. (2019, *Journal of Econometrics* 211, 589–618). We also propose a test for the local rank similarity assumption. To illustrate the estimation approach and its properties, we provide a simulation study and estimate the impacts of India's 40-billion-dollar national rural road construction program on the reallocation of labor out of agriculture.

The four authors contributed to the paper equally. We are grateful to the Editor (Peter C.B. Phillips), the Co-Editor, and two anonymous referees for their very useful and insightful comments. This research is supported by the National Science Foundation of China (Grant Nos. 72103126, 71873080, 72273076, and 71833004) and the Fundamental Research Funds for the Central Universities (Grant No. 2023110077). Address correspondence to Zhengyu Zhang, Shanghai University of Finance and Economics, Shanghai, China; e-mail: zy.zhang@mail.shufe.edu.cn.

1. INTRODUCTION

Thistlethwaite and Campbell (1960) introduced the concept of regression discontinuity (RD), which is now widely used in studies including those on labor markets, political economy, health, environment, and development. Imbens and Lemieux (2008), van der Klaauw (2008), and Lee and Lemieux (2010) surveyed the applied and theoretical literature on RD. Most of these studies focus on estimating average treatment effects (ATEs); however, the mean treatment effect is not the only object of interest in many contexts. We might be interested in the distribution of effects for outcomes such as student achievement or earnings. Frandsen et al. (2012; FFM hereafter) provided the first extension to quantile treatment effects (QTEs) in the RD framework.

FFM studied the identification and estimation of quantiles of the potential outcomes in an RD design based on the local average treatment effect (LATE) framework (Imbens and Angrist, 1994; Abadie, Angrist, and Imbens, 2002). In the LATE framework, identification is usually achieved through a certain monotonicity assumption on the selection equation, and only the QTE for the subpopulation called "local compliers" are identified. Local compliers are the units that switch treatment status at the discontinuity, and monotonicity means that all units that switch treatment at the discontinuity do it in the same direction. The goal and main contribution of this study are to develop a new nonparametric approach for estimating QTEs for an entire population at the threshold. Instead of maintaining a monotonicity condition, we impose on the outcome equation a *local rank similarity* condition that restricts the evolution of individual ranks across treatment status in a neighborhood around the threshold.

The statistical framework we developed in this study and the local quantile treatment effect (LQTE) framework in FFM are generally non-nested. Neither one is more general than the other. Our model imposes restrictions mainly on the outcome equation, such as *rank similarity* (Assumption 2.3) and a full-rank Jacobian matrix condition (Assumption 3.2.1(ii)) while leaving the structure of the selection equation free. However, FFM works with a general outcome equation but at the cost of a restricted selection equation by imposing a monotonicity condition.[1] In some applications of the RD, assuming local rank similarity may be reasonable. For example, the disturbance in the earnings equation is often considered an innate ability. Local rank similarity requires that the distributions of people's innate abilities do not change across treatment status within some neighborhood of the cutoff point while not ruling out a certain degree of noisy variations of ability across persons. Moreover, having a rich set of covariates makes rank similarity a more plausible approximation, which we explicitly recognize in our estimation and inference developed later. As another example, in the study of the effect of retirement on household expenditures (e.g., Li, Shi, and Wu, 2016), rank similarity means that the distribution of unobserved individual consumption preferences does

[1] Appendix C provides a detailed discussion of the key identifying assumptions in this study.

not change systematically before and after retirement, which is likely to hold if retirement is a predictable event.

Chernozhukov and Hansen (2005; CH05 hereafter) propose exploiting rank similarity to identify QTEs in the instrumental variable quantile regression (IVQR) model. Rank similarity does not imply nor is implied by the assumptions of the LQTE framework. Moreover, the two frameworks have different estimands. While FFM's approach can identify QTEs for the compliant subpopulation, our approach can identify QTEs and ATE for the whole population at the threshold. In particular, instead of forming the estimator directly based on the moment conditions originated in CH05 (see Theorem 1 in CH05), we propose a new representation of the moment restrictions implied by the model. We derive closed-form expressions for the cumulative distribution functions (CDFs) of the potential outcomes based on these moment restrictions. These closed-form expressions are compositions of identifiable conditional CDFs and probabilities and can thus be estimated using a plug-in approach.

We then develop a systematic inferential procedure for the proposed QTE estimands. Recall that RD studies routinely employ local linear estimators to construct confidence intervals (e.g., in FFM). However, as Calonico, Cattaneo, and Titiunik (2014) point out, the performance of these confidence intervals may be seriously hampered by their sensitivity to the bandwidth employed. The available bandwidth selectors typically yield a large bandwidth, leading to potentially biased data-driven confidence intervals that may be biased. Here, we propose a multiplier bootstrap-based inference method with robustness against large bandwidths that applies to uniform inference by extending the recent work of Chiang, Hsu, and Sasaki (2019). In particular, we develop the limit process of a specific and more complicated local Wald estimator, which can be seen as a generalization to the class of local Wald estimands analyzed by Chiang et al. (2019). Our analysis differs from Chiang et al. (2019) in two regards. First, our estimator cannot be expressed in the form of the class of local Wald estimators in Chiang et al. (2019, Sect. 4, eqn. (4.1)). Second, we explicitly account for the presence of covariates in a fully nonparametric manner. In particular, to establish robust uniform inference with large bandwidth, one should employ the local polynomial kernel regression to estimate unconditional CDFs by taking the average conditional CDFs over the distribution of covariates, which is novel in this study.[2]

Our idea partially builds on Wuthrich's (2019) insight after deriving closed-form solutions for CH05's IVQR estimands with a binary endogenous variable and binary instruments. As Hahn, Todd, and van der Klaauw (2001) suggest, the mean RD estimator is closely related to instrumental variables (IV) type estimators. This insight suggests that the closed-form estimators Wuthrich (2019) derive for the IVQR model might have an analog in the RD context. However, extending Wuthrich's (2019) estimator to the RD context is not straightforward. Compared with Wuthrich (2019), our study has three novelties. First, we do not directly work

[2]See Remark 5.1.1 for more discussion.

with the moment condition in Chernozhukov and Hansen (2005, 2006). Instead, we derive a new set of moment restrictions (Lemma 3.1) based on which we can form closed-form estimators. Second, our plug-in estimators are fully nonparametric, whereas Wuthrich specifies parametric models for the conditional CDFs and the conditional probabilities. Third, we systematically handle the uniform inference of QTEs by extending the idea of Calonico et al. (2014) and Chiang et al. (2019). We also note that Guiteras (2008), in his unpublished work, proposes an inverse quantile regression (IQR) estimator in the spirit of Chernozhukov and Hansen (2006; CH06 hereafter) for the RD model. In Guiteras (2008), the RD model is specified as linear-in-parameters. Moreover, Guiteras (2008) did not derive the limiting distribution of his estimator, and does not address the issue of uniform inference. In contrast, we consider the quantile RD model (with covariates) in a fully nonparametric way and propose a complete inferential procedure.

In Appendix B, we also introduce a test of local rank similarity as one of the key identifying conditions of our model. As an application, we estimate the impacts of India's 40-billion-dollar national rural road construction program on the reallocation of labor out of agriculture using the data in Asher and Novosad (2020; hereafter AN20). We find that constructing a new road has heterogeneous effects in facilitating workers' movement out of the agricultural sector. Such effects are significant for villages in the middle of the agricultural production index distribution. In contrast, the effect is negligible for villages at the lowest and highest ends of agricultural production distribution.

The remainder of this article is organized as follows: Section 2 sets up the statistical framework and discusses the key identifying assumptions. We establish the identification results in Section 3. Section 4 describes the estimation procedure. Section 5.1 derives the asymptotic distribution of the proposed QTE estimator. Section 5.2 develops a multiplier bootstrap-based inference method with robustness against large bandwidths that applies to uniform inference.

Section 6 reports the Monte Carlo simulation results. Section 7 presents the application. Section 8 concludes. We provide the proofs of main theorems in Appendix A and propose the test for local rank similarity in Appendix B. Appendix C provides further discussion of the main assumptions used for identification. The proofs of lemmas in Appendix A, additional Monte Carlo simulation results, and the discussion on bandwidths selection in applications are presented in the Supplementary Material.

2. THE MODEL

We are interested in the causal effect of a binary treatment D on an absolutely continuous outcome variable Y. We observe n units, indexed by $i = 1, \ldots, n$, drawn randomly and independently from a large population. Let Y_{1i} and Y_{0i} be the potential outcomes of individual i under treatment and no treatment, respectively, so the observed outcome is $Y_i = Y_{0i}(1 - D_i) + Y_{1i}D_i$. According to CH05, by the Skorohod representation of random variables, we can represent the potential outcomes as

$$Y_1 = q_1\big(R, X, U_1\big), \quad Y_0 = q_0\big(R, X, U_0\big), \tag{2.1}$$

where R is a running variable that influences the probability of treatment discontinuously and R is not necessarily independent of (U_0, U_1). X is a vector of covariates with dimension L. U_d determines the relative ranking of observationally equivalent individuals in the distribution of potential outcomes, and we thus refer to it as the rank variable in CH05. One may think of U_d as representing some unobserved characteristic, for example, ability or proneness. Like CH05, we maintain the following strict monotonicity. Let S_A be the support of a generic random variable A.

Assumption 2.1. $q_1(r,x,u)$ and $q_0(r,x,u)$ are strictly increasing with respect to u for each $(r,x) \in S_R \times S_X$.

Assumption 2.2.

(i) For any $x \in S_X$,

$$\lim_{r \to r_0^+} P\big(D = 1 \big| R = r, X = x\big) \neq \lim_{r \to r_0^-} P\big(D = 1 \big| R = r, X = x\big). \tag{2.2}$$

(ii) Treatment status is determined by

$$D = \rho(\mathbb{1}\{R > r_0\}, R, X, V), \tag{2.3}$$

where ρ is an indicator function and V is the unobserved random vector. Moreover, $\rho(1,r,x,v)$ is right-continuous in r at r_0 and $\rho(0,r,x,v)$ is left-continuous in r at $r = r_0$.

Assumption 2.2(i) is the defining feature of an RD design: the probability of treatment changes discontinuously at the threshold value of the running variable. The so-called sharp RD design has one difference in the probability of treatment across the threshold: treatment status is completely determined by location relative to the threshold. In the fuzzy RD design, the difference in treatment probability is less than one but still strictly positive, so other factors influence selection into treatment besides the running variable. We focus on the more general fuzzy design, treating the sharp design as a special case. The intuition behind Assumption 2.2(ii) is that once we control the side of the cutoff on which the unit falls, the probability of receiving treatment relies on the value of the running variable in a neighborhood around the threshold. We note here that we impose no structural restriction in equation (2.3). For example, V can be multidimensional, and monotonicity with respect to V is not required.

Assumption 2.3 (Rank similarity). For any positive sequence $\delta_s \to 0$, U_1 and U_0 are identically distributed conditional on $R \in (r_0 - \delta_s, r_0 + \delta_s)$, X and V;

https://doi.org/10.1017/S026646662300021X Published online by Cambridge University Press

namely, $\lim_{r\to r_0} F_{U_1|R,X,V}(u,r,x,v) = \lim_{r\to r_0} F_{U_0|R,X,V}(u,r,x,v)$ for any $(u,x,v) \in S_U \times S_X \times S_V$.[3]

Assumption 2.3 is the principal identifying assumption of our model, which restricts the evolution of individual ranks across treatment status in a neighborhood around the threshold. The idea is that each unit possesses an underlying proneness or ability, for example, to die early, learn fast, or grow taller, which does not change with the treatment. In some applications, such rank preservation (RP) is natural because it seems unlikely that the treatment makes weak units robust and strong units weak. CH05 show that RP can be weakened to rank similarity (RS). Formally, RS requires that conditional on observed covariates and the disturbance in the selection equation, U_1 and U_0 are identically distributed. Consider studies on the effects of summer school on educational achievement, where students who did not meet the end-of-year test score threshold were required to attend summer school (Jacob and Lefgren, 2004). Imagine two students, A and B, with the same end-of-year test, score R, but student A is more able than student B. RP predicts that student A will outperform student B on an end-of-summer test whether both do or do not attend summer school, while RS indicates that if we expect student A to outperform student B on an end-of-summer test if neither attends summer school, we would also expect student A to outperform student B if both attend summer school. It is important to note that this prediction can still be satisfied if student A has a bad day on the end-of-summer test date, that is, outperformed by student B in the end. In other words, RS allows for unsystematic slippages from an individual's rank level. For in-depth discussions of the RS condition, interested readers can refer to CH05 and the reviews by Chernozhukov and Hansen (2013) and Chernozhukov, Hansen, and Wuthrich (2017).

Remark 2.1. Assumption 2.3 is analogous to the RS condition in an IV setting (CH05). There, the RS condition is

$$U_1 \sim^d U_0, \text{ conditional on } Z, X \text{ and } V, \tag{2.4}$$

where Z is an instrumental variable independent of (U_1, U_0). Clearly, Assumption 2.3 is stated in parallel with (2.4) but differs in two aspects. First, (2.4) is conditioning on Z, X, and V. However, the RD design has no instrument: R can be arbitrarily correlated with (U_0, U_1). Second, Assumption 2.3 is a local version of (2.4) in that it is required to hold only within a small neighborhood around $R = r_0$, which is weaker than to hold for each $R = r$, $r \in S_R$.

3. IDENTIFICATION

Our identification proceeds in two steps. First, we present a set of nonlinear conditional moment restrictions containing the parameters of interest. Next, we

[3]According to Assumption 3.1.1 in the next section, we conclude that $S_U = (0,1)$.

https://doi.org/10.1017/S026646662300021X Published online by Cambridge University Press

show, under a full rank condition of the Jacobian of the moment functions, the potential outcome CDFs, namely,

$$F_{Y_d|R,X}(y,r_0,x) = P\left(Y_d \le y \middle| R = r_0, X = x\right) \tag{3.1}$$

are identified for $d = 0,1$. Based on these results, we can also identify many other smooth functionals of the potential outcome CDFs, such as conditional/unconditional QTEs, average treatment effects, and distributional treatment effects.

3.1. Moment Restrictions

Besides those introduced in Section 2, we have the following additional assumptions for identification.

Assumption 3.1.1. The random variables U_0 and U_1 conditional on $R = r_0$ and $X = x$ are uniformly distributed on $(0,1)$.

Remark 3.1.1. Assumption 3.1.1 is a normalization condition, under which $F_{Y_d|R,X}(y,r_0,x)$ is the inverse function of $q_d(r_0,x,u)$ with respect to u, because for any $\tau \in (0,1)$,

$$\begin{aligned} &F_{Y_d|R,X}\left(q_d\left(r_0,x,\tau\right),r_0,x\right) = P\left(Y_d \le q_d\left(r_0,x,\tau\right)\middle|R = r_0,X = x\right) \\ =&P\left(q_d\left(r_0,x,U_d\right) \le q_d\left(r_0,x,\tau\right)\middle|R = r_0,X = x\right) \\ =&P\left(U_d \le \tau \middle| R = r_0,X = x\right) = \tau, \end{aligned}$$

where the last equality is by Assumption 3.1.1.

Assumption 3.1.2. (i) $f_{R|VX}(r|v,x)$ is continuous in r at $r = r_0$ for each $(v,x) \in S_V \times S_X$. (ii) $q_0(r,x,u)$ and $q_1(r,x,u)$ are continuous in r at $r = r_0$ for each $(x,u) \in S_X \times S_U$. (iii) For $d = 0,1$, the distribution of U_d conditional on $R = r$, $X = x$ and $V = v$, namely, $F_{U_d|RXV}(u,r,x,v)$ is continuous in u and r at r_0, for each $(x,v) \in S_X \times S_V$.

Assumption 3.1.2 consists of a set of smoothness conditions, analogous to Assumption I2 in FFM. Intuitively, these smooth conditions ensure that after controlling smoothly for the running variable R, differences in the distribution of outcomes on either side of the threshold occur because of the change in the treatment status D. We emphasize here that these smoothness conditions are essential to proving the moment conditions below (Lemma 3.1). This setting contrasts with the IV quantile model in CH05, which requires no smoothness condition to obtain the moment condition (Theorem 1 in CH05).

LEMMA 3.1. *Suppose that (2.1), (2.3), and Assumptions 2.1–2.3, 3.1.1, and 3.1.2 hold; then, for each $\tau \in (0,1)$ and $x \in S_X$,*

$$\lim_{\epsilon \to 0^+} P\left(Y \le q_D(r_0,x,\tau)\middle| r_0 < R < r_0 + \epsilon, X = x\right) = \tau \tag{3.1.2a}$$

https://doi.org/10.1017/S026646662300021X Published online by Cambridge University Press

and

$$\lim_{\epsilon \to 0^+} P\left(Y \leq q_D(r_0, x, \tau) \middle| r_0 - \epsilon < R < r_0, X = x \right) = \tau, \tag{3.1.2b}$$

where $q_D(r, x, \tau) = Dq_1(r, x, \tau) + (1 - D)q_0(r, x, \tau)$.

Remark 3.1.2. Compared with the moment condition described in Guiteras (2008, eqn. (9)), Lemma 3.1 here is novel in two aspects. First, (3.1.2a) and (3.1.2b) explicitly contain the parameters we intend to identify; that is, $(q_1(r_0, x, \tau), q_0(r_0, x, \tau))$ (via $q_D(r_0, x, \tau) = Dq_1(r_0, x, \tau) + (1 - D)q_0(r_0, x, \tau)$). Recasting the moment restriction as (3.1.2a) and (3.1.2b) facilitates a closed-form representation of the potential outcome CDF. Second, we provide a new and more formal proof of these moment conditions. Compared with CH05's Theorem 1 and Guiteras (2008), the proof of Lemma 3.1 highlights the role of the local RS and the smoothness conditions in the RD context.

3.2. Closed-Form Solution for CDFs

Moment conditions (3.1.2a) and (3.1.2b) alone do not point-identify the potential outcome CDF without additional assumptions. For example, to ensure point identification, CH05 imposes a full-rank condition on the Jacobian matrix of the moment functions (Theorem 2 in CH05). To introduce a similar full-rank-type condition for the RD design, define

$$\begin{aligned}
p(r_0^+, x) &= \lim_{\epsilon \to 0^+} P(D = 1 | r_0 < R < r_0 + \epsilon, X = x) = \lim_{r \to r_0^+} P(D = 1 | R = r, X = x), \\
p(r_0^-, x) &= \lim_{\epsilon \to 0^-} P(D = 1 | r_0 - \epsilon < R < r_0, X = x) = \lim_{r \to r_0^-} P(D = 1 | R = r, X = x), \\
F_{Y|DRX}(y, d, r_0^+, x) &= \lim_{r \to r_0^+} P(Y \leq y | D = d, R = r, X = x), \\
F_{Y|DRX}(y, d, r_0^-, x) &= \lim_{r \to r_0^-} P(Y \leq y | D = d, R = r, X = x).
\end{aligned}$$

Define the moment functions

$$\Pi(y_1, y_0, x) = \lim_{\epsilon \to 0+} \begin{bmatrix} P\left(Y < Dy_1 + (1 - D)y_0 \middle| r_0 < R < r_0 + \epsilon, X = x\right) \\ P\left(Y < Dy_1 + (1 - D)y_0 \middle| r_0 - \epsilon < R < r_0, X = x\right) \end{bmatrix}.$$

For notational simplicity, we suppress the dependence of $q_d(r, x, \tau)$ on r when it is evaluated at $r = r_0$. Thus, we let $q_d(x, \tau) = q_d(r_0, x, \tau)$ for $d = 0, 1$. It follows from Lemma 3.1 that for each $\tau \in (0, 1)$ and $x \in S_X$,

$$\Pi\left(q_1\left(x, \tau\right), q_0\left(x, \tau\right), x\right) = 0,$$

because

$$\begin{aligned}
&\lim_{\epsilon \to 0^+} P\left(Y < Dy_1 + (1 - D)y_0 \middle| r_0 < R < r_0 + \epsilon, X = x\right) \\
=&F_{Y|DRX}\left(y_1, 1, r_0^+, x\right) p\left(r_0^+, x\right) + F_{Y|DRX}\left(y_0, 0, r_0^+, x\right)\left(1 - p\left(r_0^+, x\right)\right),
\end{aligned}$$

https://doi.org/10.1017/S026646662300021X Published online by Cambridge University Press

the Jacobian of $\Pi(y_1, y_0, x)$ with respect to (y_1, y_0), is

$$\Pi'(y_1, y_0, x) = \begin{bmatrix} f_{Y|DRX}\left(y_1, 1, r_0^+, x\right) p\left(r_0^+, x\right) & f_{Y|DRX}\left(y_0, 0, r_0^+, x\right)\left(1 - p\left(r_0^+, x\right)\right) \\ f_{Y|DRX}\left(y_1, 1, r_0^-, x\right) p\left(r_0^-, x\right) & f_{Y|DRX}\left(y_0, 0, r_0^-, x\right)\left(1 - p\left(r_0^-, x\right)\right) \end{bmatrix}.$$

We make an assumption similar to the full-rank and completeness conditions given by CH05 (the conditions in CH05's Theorem 2.)

Assumption 3.2.1.

(i) There exists a positive$\delta > 0$ such that $\Pi'(y_1, y_0, x)$ is continuous in (y_1, y_0) in the support of $(Y_1, Y_0)\big|R \in (r_0 - \delta, r_0 + \delta), X = x$.
(ii) There exists a positive $\delta > 0$ such that $\Pi'(y_1, y_0, x)$ is of full-rank for all (y_1, y_0) in the support of $(Y_1, Y_0)\big|R \in (r_0 - \delta, r_0 + \delta), X = x$.

Remark 3.2.1. In Appendix A, we prove that the discontinuity condition (equation (2.2)) is necessary for the full-rank Jacobian condition (Assumption 3.2.1(ii)) to hold. Specifically, we show that if combining $p(r_0^+, x) = p(r_0^-, x) = p(r_0, x)$ with the smoothness conditions in Assumption 3.1.2, the determinant of $\Pi'(y_1, y_0, x)$ is zero.

Remark 3.2.2. The full-rank Jacobian matrix condition does not rule out the existence of defiers. In Appendix B of the Supplementary Material, we provide a numerical example in which the monotonicity is violated, but the full-rank condition still holds. Furthermore, Assumption 3.2.1(ii) is equivalent to that the derivative of $\widetilde{F}_1(y, x)$ and $\widetilde{F}_0(y, x)$ with respect to y are nonzero for any $(y, x) \in S_Y \times S_X$, which we define in (3.2.1) and (3.2.2). Appendix C provides further discussions and sufficient interpretable conditions for this full rank condition.

The next lemma establishes an important relationship between the parameters of interest $(q_1(x, \tau), q_0(x, \tau))$, based on which we can obtain a closed-form representation of $(q_1(x, \tau), q_0(x, \tau))$.

LEMMA 3.2. *Under the same assumptions as Lemma 3.1 and Assumption 3.2.1, we have:*

(i) *For each $\tau \in (0, 1)$ and $x \in S_X$,*

$$\widetilde{F}_1\left(q_1\left(x, \tau\right), x\right) = \widetilde{F}_0\left(q_0\left(x, \tau\right), x\right),$$

where

$$\widetilde{F}_1(y, x) = F_{Y|DRX}\left(y, 1, r_0^+, x\right) p\left(r_0^+, x\right) - F_{Y|DRX}\left(y, 1, r_0^-, x\right) p\left(r_0^-, x\right) \tag{3.2.1}$$

and

$$\widetilde{F}_0(y, x) = F_{Y|DRX}\left(y, 0, r_0^-, x\right)\left(1 - p\left(r_0^-, x\right)\right) - F_{Y|DRX}(y, 0, r_0^+, x)(1 - p(r_0^+, x)). \tag{3.2.2}$$

(ii) $\widetilde{F}_1(y, x)$ *and* $\widetilde{F}_0(y, x)$ *are strictly monotone with respect to y.*

https://doi.org/10.1017/S026646662300021X Published online by Cambridge University Press

For notational simplicity, we suppress the dependence of $F_{Y_d|RX}(y,r_0,x)$ on r when evaluated at $r=r_0$. Thus, we write $F_{d|X}(y,x)=F_{Y_d|RX}(y,r_0,x)$ for $d=0,1$.

THEOREM 1. *Under the same assumptions as Lemma 3.1 and Assumption 3.2.1,*

$$F_{1|X}(y|x)=F_{Y|DRX}\left(y,1,r_0^+,x\right)p\left(r_0^+,x\right) + F_{Y|DRX}\left(\widetilde{q}_0\left(\widetilde{F}_1\left(y,x\right),x\right),0,r_0^+,x\right)\left(1-p\left(r_0^+,x\right)\right), \quad \textbf{(3.2.3)}$$

$$F_{0|X}(y,x)=F_{Y|DRX}\left(\widetilde{q}_1\left(\widetilde{F}_0\left(y,x\right),x\right),1,r_0^-,x\right)p\left(r_0^-,x\right) + F_{Y|DRX}\left(y,0,r_0^-,x\right)\left(1-p\left(r_0^-,x\right)\right), \quad \textbf{(3.2.4)}$$

where $\widetilde{q}_d(\cdot,x)$ denotes the inverse function of $\widetilde{F}_d(\cdot,x)$ for $d=0,1$, which is well defined by Lemma 3.2(ii).

Assumption 3.2.2. $f_{X|R}(x|r)$ is continuous in r at $r=r_0$.[4]

COROLLARY 3.1. *Under the same assumptions as Theorem 1 and Assumption 3.2.2,*

$$F_1(y)=\int F_{1|X}(y|x)f_{X|R}(x|r_0)dx,$$

$$F_0(y)=\int F_{0|X}(y|x)f_{X|R}(x|r_0)dx.$$

Remark 3.2.3. By similar reasoning, we can show the distributional treatment effects (DTEs, denoted by $\delta_{DTE}(y)$) and the ATE (denoted by δ_{ATE}) are also identifiable. Specifically, $\delta_{DTE}(y)=F_1(y)-F_0(y)$ and $\delta_{ATE}=\int_{S_Y} yd\left(F_1(y)-F_0(y)\right)$.

4. ESTIMATION

In this section, we propose a plug-in estimation approach based on the closed-form solutions derived from Theorem 1 and Corollary 3.1. We focus on the estimation of the unconditional QTE at a particular quantile, namely

$$\delta(\tau)=q_1(\tau)-q_0(\tau)=F_1^{-1}(\tau)-F_0^{-1}(\tau),$$

for $\tau\in(0,1)$. The estimation procedure is straightforward: we first estimate $F_1(y)$ and $F_0(y)$ by replacing the CDFs and conditional probabilities with their empirical counterparts, and then invert $F_1(y)$ and $F_0(y)$ to obtain the estimate of $\delta(\tau)$.

[4] Assumption 3.2.2 is not necessary for the identification of $F_1(y)$, and we assume $F_0(y)$ for notation simplicity. We can follow Frölich and Huber (2019), in the absence of this continuity assumption, to establish the identification of $F_1(y)$ and $F_0(y)$ by weighting the conditional density of X at r_0. (In the limit, the density of X conditional on R being within a symmetric neighborhood around r_0 is $\left(f_{X|R}\left(x|r_0^+\right)+f_{X|R}\left(x|r_0^-\right)\right)/2$.) Moreover, Assumption 3.2.2 is testable because both X and R are observable.

https://doi.org/10.1017/S026646662300021X Published online by Cambridge University Press

The local linear estimator is commonly used to estimate conditional expectations in RD designs. To make an inference with the local linear estimation, one often chooses rather large bandwidths, such as $h \propto n^{-1/5}$. However, one consequence of such bandwidth selectors is that they often lead to a non-negligible bias in the distributional approximation of the estimator, resulting in confidence intervals with empirical coverage well below their nominal target. To overcome this disadvantage of local linear estimation, following Chiang et al. (2019; CHS hereafter), we use local quadratic regression techniques to estimate the CDFs and conditional probabilities at the discontinuity threshold, which effectively accounts for the second-order bias estimation.

For any generic random variable W, let $m^+(W|x)$ (or $m^-(W|x)$) signify the conditional mean of W at $R = r_0$ and $X = x$ from above (or below), namely,

$$m^+(W|x) = E\left(W\middle|R = r_0^+, X = x\right), \quad m^-(W|x) = E\left(W\middle|R = r_0^-, X = x\right).$$

Let $\widehat{m}^+(W|x)$ (or $\widehat{m}^-(W|x)$) denote a local quadratic estimate of $m^+(W|x)$ (or $m^-(W|x)$). That is, $\widehat{m}^+(W|x)$ is estimated by the first component of a that solves

$$\arg\min_a \sum_{i=1}^n \left(W_i - \nu_i(r_0,x)'a\right)^2 K\left(\frac{X_i - x}{h_x}\right) K\left(\frac{R_i - r_0}{h}\right) \mathbb{1}\{R_i > r_0\}, \tag{4.1.1}$$

where $\nu_i(r_0,x) = \left(1, (R_i - r_0)/h, (R_i - r_0)^2/h^2, (X_i - x)/h_x, (X_i - x)^2/h_x^2, (R_i - r_0)(X_i - x)/hh_x\right)'$, and $\widehat{m}^-(W|x)$ is estimated by the first component of a that solves[5]

$$\arg\min_a \sum_{i=1}^n \left(W_i - \nu_i(r_0,x)'a\right)^2 K\left(\frac{X_i - x}{h_x}\right) K\left(\frac{R_i - r_0}{h}\right) \mathbb{1}\{R_i < r_0\}. \tag{4.1.2}$$

According to Theorem 1, $F_{1|X}(y|x)$ can be rewritten as

$$\begin{aligned} F_{1|X}(y|x) &= E(\mathbb{1}\{Y < y\}D|R = r_0^+, X = x) \\ &\quad + E(\mathbb{1}\{Y < \widetilde{q}_0(\widetilde{F}_1(y,x),x)\}(1-D)|R = r_0^+, X = x), \end{aligned} \tag{4.1.3}$$

and thus it can be estimated by

$$\widehat{F}_{1|X}(y|x) = \widehat{m}^+(\mathbb{1}\{Y < y\}D|x) + \widehat{m}^+(\mathbb{1}\{Y < \widehat{\widetilde{q}}_0(\widehat{\widetilde{F}}_1(y,x),x)\}(1-D)|x), \tag{4.1.4}$$

where

$$\widehat{\widetilde{q}}_0(\tau,x) = \inf\{a : \widehat{\widetilde{F}}_0(a,x) \geq \tau\},$$

$\widehat{\widetilde{F}}_0(y,x)$ and $\widehat{\widetilde{F}}_1(y,x)$ are consistent estimators of $\widetilde{F}_0(y,x)$ and $\widetilde{F}_1(y,x)$, given by (3.2.1) and (3.2.2), respectively. Since $\widetilde{F}_1(y,x)$ can be rewritten as

[5] If X has discrete components, such that $X = (X^c, X^d)$, we replace $K\left(\frac{X_i - x}{h_x}\right)$ with $K\left(\frac{X_i^c - x^c}{h_x}\right)\mathbb{1}\{X_i^d = x^d\}$, and $\nu_i(r_0,x)$ with $\left(1, (R_i - r_0)/h, (R_i - r_0)^2/h^2, (X_i^c - x^c)/h_x, (X_i^c - x^c)^2/h_x^2, (R_i - r_0)(X_i^c - x^c)/hh_x\right)'$. Similar modifications apply to the multiplier bootstrap procedure.

https://doi.org/10.1017/S026646662300021X Published online by Cambridge University Press

$$\widetilde{F}_1(y,x) = E\left(\mathbb{1}\{Y<y\}D\middle|R=r_0^+,X=x\right) - E\left(\mathbb{1}\{Y<y\}D\middle|R=r_0^-,X=x\right), \tag{4.1.5}$$

a consistent estimator of $\widehat{\widetilde{F}}_1(y,x)$ is

$$\widehat{\widetilde{F}}_1(y,x) = \widehat{m}^+\left(\mathbb{1}\left\{Y<y\right\}D\middle|x\right) - \widehat{m}^-\left(\mathbb{1}\left\{Y<y\right\}D\middle|x\right). \tag{4.1.6}$$

Similarly, a consistent estimator of $\widehat{\widetilde{F}}_0(y,x)$ is

$$\widehat{\widetilde{F}}_0(y,x) = \widehat{m}^-\left(\mathbb{1}\left\{Y<y\right\}(1-D)\middle|x\right) - \widehat{m}^+\left(\mathbb{1}\left\{Y<y\right\}(1-D)\middle|x\right). \tag{4.1.7}$$

Applying the same reasoning, $F_{0|X}(y|x)$ can be estimated by

$$\widehat{F}_{0|X}(y|x) = \widehat{m}^-\left(\mathbb{1}\left\{Y<\widehat{q}_1\left(\widehat{\widetilde{F}}_0(y,x),x\right)\right\}D\middle|x\right) + \widehat{m}^-\left(\mathbb{1}\left\{Y<y\right\}(1-D)\middle|x\right). \tag{4.1.8}$$

We then estimate $F_1(y)$ by averaging $\widehat{F}_{1|X}(y|X_i)$ over X_i given $R=r_0$, that is, $F_1(y)$ is estimated by the first component of a that solves

$$\arg\min_a \sum_{i=1}^n \left(\widehat{F}_{1|X}(y|X_i) - \widetilde{\nu}_i(r_0)'a\right)^2 K\left(\frac{R_i-r_0}{h_r}\right). \tag{4.1.9}$$

Similarly, $F_0(y)$ is estimated by the first component of a that solves and

$$\arg\min_a \sum_{i=1}^n \left(\widehat{F}_{0|X}(y|X_i) - \widetilde{\nu}_i(r_0)'a\right)^2 K\left(\frac{R_i-r_0}{h_r}\right), \tag{4.1.10}$$

where $\widetilde{\nu}_i(r_0) = \left(1, \left(R_i-r_0\right)/h_r, \left(R_i-r_0\right)^2/h_r^2\right)'$.

Remark 4.1.1. The basic purpose of (4.1.9) and (4.1.10) is to take the average over X_i given $R=r_0$ through the local quadratic method to reduce bias as long as $f_{X|R}(x|r)$ is continuous at r_0.[6] CHS use this idea, establishing the bias reduction in RD designs without X. Frölich and Huber (2019) used a boundary kernel to reduce the bias and improve the convergence rate if X is discontinuously distributed at the threshold. Compared with Frölich and Huber (2019), the local quadratic regression approach has at least two advantages. First, it reduces bias without affecting the optimal convergence rate achieved by the estimator, which facilitates the development of test statistics. Second, it provides a unified framework for researchers to compute the bias of higher orders by including higher-order polynomials in the regressors.

[6]Note that (4.1.9) and (4.1.10) still consistently estimate $F_1(y)$ and $F_0(y)$ if Assumption 3.2.2 is violated. See equations (7) and (13) in Frölich and Huber (2019) for more details.

Based on the estimates of $\widehat{F}_0(y)$ and $\widehat{F}_1(y)$, we estimate $\delta(\tau)$ by

$$\widehat{\delta}(\tau) = \widehat{q}_1(\tau) - \widehat{q}_0(\tau), \tag{4.1.11}$$

where

$$\widehat{q}_1(\tau) = \inf\left\{a : \widehat{F}_1(a) \geq \tau\right\}, \quad \widehat{q}_0(\tau) = \inf\left\{a : \widehat{F}_0(a) \geq \tau\right\}.$$

Remark 4.1.2. The estimation procedure described above involves inverting several CDFs to obtain the quantile functions. In finite samples, these estimated conditional CDFs can be non-monotonic step functions. To refine the finite-sample performance, we follow the same procedure as in FFM: we monotonize the estimated distribution functions by rearrangement. As Chernozhukov, Fernandez-Val, and Galichon (2010) suggest, this rearrangement procedure will not affect the asymptotic properties of the estimators, and we keep them implicit throughout the study.

Remark 4.1.3. By similar reasoning, we can also propose the estimators of the DTEs and ATE, respectively. For the DTEs, we can estimate $\delta_{DTE}(y)$ by $\widehat{\delta}_{DTE}(y) = \widehat{F}_1(y) - \widehat{F}_0(y)$, where $\widehat{F}_1(y)$ and $\widehat{F}_0(y)$ are the first components of a that solves (4.1.9) and (4.1.10), respectively. For the ATE, we construct a grid of y values: $y_0 < y_1 < y_2 < \cdots < y_{K-1} < y_K$.[7] The estimator of ATE is $\widehat{\delta}_{ATE} = \sum_{k=1}^{K} y_k \left[\left(\widehat{F}_1(y_k) - \widehat{F}_0(y_k)\right) - \left(\widehat{F}_1(y_{k-1}) - \widehat{F}_0(y_{k-1})\right)\right]$.

5. ASYMPTOTIC PROPERTIES AND INFERENCE

This section provides the asymptotic properties and inference for the QTE estimator. Section 5.1 establishes the uniform Gaussianity for the QTE estimator based on a set of regular conditions. In Section 5.2, we apply the multiplier bootstrap method to construct uniform confidence bands for the QTE estimator. We summarize the algorithm detailing the comprehensive procedure for practitioners, including estimation, inference, and bandwidth selection, in Section 5.3.

5.1. Limiting Distribution for QTE

In this section, we study the asymptotic properties of $\widehat{\delta}(\tau)$ in (4.1.11). We impose the following regularity assumptions.

Assumption 5.1. $\left\{Y_i, D_i, R_i, X_i\right\}_{i=1}^{n}$ are n i.i.d. copies of random vector (Y, D, R, X) defined on a probability space $(\Omega^{\upsilon}, \mathcal{F}^{\upsilon}, \mathbb{P}^{\upsilon})$.

Assumption 5.2. The density $f_{RX}(r,x)$ is continuously differentiable and bounded away from zero and infinity (i) with respect to the continuous component of x at r_0; (ii) with respect to r in an interval around r_0.

[7] Similar to Remark 3.1 in Chernozhukov, Fernandez-Val, and Melly (2013), the number of grid points K and the observations n should satisfy $\sqrt{n}/K \to 0$.

https://doi.org/10.1017/S026646662300021X Published online by Cambridge University Press

Assumption 5.3. For $d \in \{0,1\}$, $j \in \{0,1,2,3\}$, the left and right limits of the functions $\frac{\partial^j}{\partial r^j} E\left[\mathbb{1}\left\{Y \leq y, D = d\right\} | R = r, X = x\right]$ and $\frac{\partial^j}{\partial r^j} E\left[\mathbb{1}\left\{D = d\right\} | R = r, X = x\right]$ are Lipschitz in r at r_0 and $x \in S_X$, respectively.

Assumption 5.4. Let $\Delta p = \lim_{r \to r_0^+} E\left(D \middle| R = r\right) - \lim_{r \to r_0^-} E\left(D \middle| R = r\right)$. Δp is strictly positive.

Assumption 5.5. $K(\cdot)$ is Borel measurable, bounded, continuous, symmetric, nonnegative-valued on $[-1,1]$, and integrates to one. $\left\{K(\cdot/h) : h > 0\right\}$ is of VC type. Define

$$\widetilde{\Gamma} = \int (1\ u\ u^2)' \cdot K(u) \cdot (1\ u\ u^2) du,$$

$$\Gamma^+ = \int (1\ u\ u^2\ s\ s^2\ us)' \cdot K(u)K(t) \cdot (1\ u\ u^2\ s\ s^2\ us) \mathbb{1}\left\{u > 0\right\} duds,$$

$$\Gamma^- = \int (1\ u\ u^2\ s\ s^2\ us)' \cdot K(u)K(t) \cdot (1\ u\ u^2\ s\ s^2\ us) \mathbb{1}\left\{u < 0\right\} duds.$$

$\widetilde{\Gamma}$, Γ^+, and Γ^- are positive definite. Let $h/h_r = \gamma^2$ with $0 < \gamma < \infty$. Moreover, the bandwidth should satisfy $nhh_x^L / \log n \to \infty$ and $\sqrt{nh} \max\{h^3, h_x^3\} \to 0$.[8]

Assumptions 5.1–5.5 are standard regularity conditions that ensure the functional central limit theorems can apply to the conditional expectations involved in the estimation procedure. Most of the prior work includes Assumption 5.1. Assumption 5.2 ensures that the joint distribution of the running variable and covariates is well behaved near the threshold. Assumption 5.3 ensures that the underlying conditional distributions of potential outcomes are sufficiently smooth at discontinuity. Assumption 5.4 requires that the probability of the treatment changes discretely at the threshold. Assumption 5.5 imposes standard conditions on the kernel function, which are satisfied by many frequently used kernel functions, such as uniform, triangular, and Epanechnikov kernels.

Moreover, Assumption 5.5 specifies admissible rates of bandwidths. Suppose that h and h_x depend on the sample size, as follows:

$$h \propto n^{-\zeta} \quad \text{and} \quad h_x \propto n^{-\zeta_x},$$

then, the bandwidth conditions in Assumption 5.5 require that

$$\zeta + L\zeta_x < 1, \quad \zeta + 6\min\{\zeta, \zeta_x\} > 1,$$

in which the first and second inequalities restrict the rate of variance and bias, respectively. We can choose $\zeta = \zeta_x = 1/5$ when X contains at most three continuous components, and the resulting convergence rates are consistent with common MSE minimizing selectors of bandwidths (e.g., Imbens and Kalyanaraman, 2012;

[8]Compared with the results in Frölich and Huber (2019), the local polynomial regression simultaneously reduces the bias through h and h_x, while the higher-order kernel only reduces the bias generated by h_x.

https://doi.org/10.1017/S026646662300021X Published online by Cambridge University Press

Calonico et al., 2014, 2016, 2018; Arai and Ichimura, 2016, 2018; Frölich and Huber, 2019). If $4 \leq L \leq 5$, then we can still choose $\zeta = 1/5$, but h_x must converge to zero at a slower rate than h, such that

$$\frac{2}{15} < \zeta_x < \frac{4}{5L}.$$

If X contains more than six continuous components, then one should use local polynomial regression estimators with higher order. Following CHS, applying general pth order local polynomial regression requires

$$\zeta + L\zeta_x < 1, \quad \zeta + 2(1+p)\min\{\zeta, \zeta_x\} > 1.$$

Then, we choose $\zeta = 1/5$ as usual and a slower rate of ζ_x satisfying

$$\frac{2}{5(1+p)} < \zeta_x < \frac{4}{5L}.$$

In practice, one can set the bandwidths as $h_r = c_r n^{-\zeta}$, $h = cn^{-\zeta}$, and $h_x = c_x n^{-\zeta_x}$, where the choice rule of ζ and ζ_x is given above. The remaining step is to select the constants c_r, c, and c_x, which can be determined by a grid search. For example, we construct a grid of c_r values: $\underline{c}_r < c_r^1 < c_r^2 < \cdots < c_r^L < \overline{c}_r$. Then, we choose c_r^* in this grid that minimizes objective functions (4.1.9) or (4.1.10), where we replace a by its estimator. For c_x, we let $c_x = \widetilde{c}\widehat{\sigma}_x$, where $\widehat{\sigma}_x$ is a vector of standard deviation of X, and $\widetilde{c}$ is a constant to be determined. Similarly, with the choice of c_r, we can determine $\widetilde{c}$ and c simultaneously by conducting a two-dimensional grid search.

Assumption 5.6. For $d \in \{0,1\}$, $\widetilde{F}_d(y,x)$ ((3.2.1) and (3.2.2)) admits a nonzero derivative $\widetilde{f}_d(y,x)$ with respect to y, which is uniformly bounded and uniformly continuous on $S_Y \times S_X$.

FFM and CHS do not require Assumption 5.6. It strengthens the condition of the full-rank Jacobian (Assumption 3.2.1), which implies that $\widetilde{f}_d(y,x)$ is nonzero to guarantee Hadamard differentiability of the closed-form solutions in Theorem 1. Asymptotic normality of $\widehat{\delta}(\tau)$ follows from the weak convergence as a process of the potential outcome CDF estimators, and the functional delta method (van der Vaart, 1998).

THEOREM 2. *Under the same assumptions as Theorem 1 and Assumptions 5.1–5.6,*

$$\sqrt{nh_r}\begin{pmatrix} \widehat{q}_1(\tau) - q_1(\tau) \\ \widehat{q}_0(\tau) - q_0(\tau) \end{pmatrix} \to^d \begin{pmatrix} \mathbb{Z}_{q_1}(\tau) \\ \mathbb{Z}_{q_0}(\tau) \end{pmatrix},$$

where $\mathbb{Z}_{q_1}$ and $\mathbb{Z}_{q_0}$ are tight zero-mean Gaussian processes in $\ell^\infty(\mathcal{T})^2$, in which $\mathcal{T} \subset (0,1)$ is a compact interval. The covariance functions $\Sigma^q(\tau,\widetilde{\tau})$ are, for $j,k \in \{1,2\}$,

$$\Sigma_{jk}^q(\tau,\widetilde{\tau}) = \Sigma_{jk}^F\left(q_{2-j}(\tau), q_{2-k}(\widetilde{\tau})\right) \Big/ f_{2-j}\left(q_{2-j}(\tau)\right) f_{2-k}\left(q_{2-k}(\widetilde{\tau})\right), \quad \textbf{(5.1.1)}$$

https://doi.org/10.1017/S026646662300021X Published online by Cambridge University Press

where

$$\Sigma^F_{jk}(y,\widetilde{y}) = c'\sigma_{jk}(y,\widetilde{y})\,c$$

and $c = (1,1,1,1,1)'$. *Moreover, let*

$$\widetilde{\lambda} = \widetilde{e}_1'\widetilde{\Gamma}^{-1}\widetilde{\Gamma}_2\widetilde{\Gamma}^{-1}\widetilde{e}_1,$$
$$\bar{\lambda} = \widetilde{e}_1'\widetilde{\Gamma}^{-1}(1,0,\mu_2)'\cdot(1,0,\mu_2)\,\widetilde{\Gamma}^{-1}\widetilde{e}_1,$$
$$\Delta^+ = \int\left[\int e_1'\left(\Gamma^+\right)^{-1}\left(1,t,t^2,u,u^2,tu\right)K(u)du\right]^2 K^2(t)\mathbb{1}\{t>0\}dt,$$
$$\Delta^- = \int\left[\int e_1'\left(\Gamma^-\right)^{-1}\left(1,t,t^2,u,u^2,tu\right)K(u)du\right]^2 K^2(t)\mathbb{1}\{t<0\}dt.$$

$$v_1^+(y,x) = \Big(1+\omega_0^+(y,x),1+\omega_0^+(y,x)\Big),\quad v_1^-(y,x) = \Big(-\omega_0^+(y,x),-\omega_0^+(y,x)\Big),$$

and

$$v_0^+(y,x) = \Big(-\omega_1^-(y,x),-\omega_1^-(y,x)\Big),\quad v_0^-(y,x) = \Big(1+\omega_1^-(y,x),1+\omega_1^-(y,x)\Big)$$

where $e_1 = (1,0,0,0,0,0)'$, $\widetilde{e}_1 = (1,0,0)'$, $\mu_2 = \int t^2K(t)dt$,

$$\widetilde{\Gamma}_2 = \int (1\ u\ u^2)'\cdot K^2(u)\cdot(1\ u\ u^2)du,$$

and

$$\omega_0^+(y,x) = f_{Y|DRX}\left(\widetilde{q}_0\left(\widetilde{F}_1(y,x),x\right),0,r_0^+,x\right)\left(1-p\left(r_0^+,x\right)\right)\Big/\widetilde{f}_0\left(\widetilde{q}_0\left(\widetilde{F}_1(y,x),x\right),x\right),$$
$$\omega_1^-(y,x) = f_{Y|DRX}\left(\widetilde{q}_1\left(\widetilde{F}_0(y,x),x\right),1,r_0^-,x\right)p\left(r_0^-,x\right)\Big/\widetilde{f}_1\left(\widetilde{q}_1\left(\widetilde{F}_0(y,x),x\right),x\right).$$

For $j,k\in\{0,1\}$, *also define*

$$\sigma_{jk}(y,\widetilde{y}) = \frac{1}{f_R(r_0)}\begin{bmatrix}\widetilde{\lambda}Cov\left(F_{2-j|X}(y,X),F_{2-k|X}(\widetilde{y},X)|R=r_0\right) & 0\\ 0 & \widetilde{\sigma}_{jk}(y,\widetilde{y})\end{bmatrix},$$

where

$$\widetilde{\sigma}_{jk}(y,\widetilde{y}) = \frac{\bar{\lambda}}{\gamma^2}\begin{bmatrix}\widetilde{\sigma}^+_{jk}(y,\widetilde{y}) & 0\\ 0 & \widetilde{\sigma}^-_{jk}(y,\widetilde{y})\end{bmatrix},$$

with

$$\widetilde{\sigma}^+_{jk}(y,\widetilde{y}) = \Delta^+E\big[v^{+\prime}_{2-j}(y,X)Cov\left(W_{2-j|X}(y,X),W_{2-k|X}(\widetilde{y},X)|R,X\right)v^+_{2-k}(y,X)|R=r_0^+\big]$$

https://doi.org/10.1017/S026646662300021X Published online by Cambridge University Press

and

$$\widetilde{\sigma}_{jk}^{-}(y,\widetilde{y}) \\ = \Delta^{-}E\left[v_{2-j}^{-\prime}(y,X)Cov\left(W_{2-j|X}(y,X),W_{2-k|X}(\widetilde{y},X)|R,X\right)v_{2-k}^{-}(y,X)|R=r_0^{-}\right],$$

in which

$$W_1=\begin{pmatrix}\mathbb{1}\left\{Y\le y\right\}D\\ \mathbb{1}\left\{Y\le \widetilde{q}_0\left(\widetilde{F}_1(y,x),x\right)\right\}(1-D)\end{pmatrix},\quad W_0=\begin{pmatrix}\mathbb{1}\left\{Y\le \widetilde{q}_1\left(\widetilde{F}_0(y,x),x\right)\right\}D\\ \mathbb{1}\left\{Y\le y\right\}(1-D)\end{pmatrix}.$$

From Theorem 2, we can easily establish the limiting process for the QTE estimator.

COROLLARY 5.1. *Under the same assumptions as Theorem 2, we have*

$$\sqrt{nh_r}\left(\widehat{\delta}(\tau)-\delta(\tau)\right)\to\mathbb{Z}_\delta=\mathbb{Z}_{q_1}(\tau)-\mathbb{Z}_{q_0}(\tau),$$

where $\delta(\tau)=q_1(\tau)-q_0(\tau)$. *The covariance function of* $\mathbb{Z}_\delta$ *is*

$$\Sigma^{\delta}(\tau,\widetilde{\tau})=\Sigma_{11}^{q}(\tau,\widetilde{\tau})+\Sigma_{22}^{q}(\tau,\widetilde{\tau})-\Sigma_{12}^{q}(\tau,\widetilde{\tau})-\Sigma_{21}^{q}(\tau,\widetilde{\tau}),$$

with $\Sigma_{jk}^{q}(\tau,\widetilde{\tau})$ *given by (5.1.1).*

Remark 5.1.1. Generally, asymptotic analysis by explicitly accounting for the presence of covariates is more complicated than CHS in three aspects. First, we must address the "local empirical process" when controlling for $X=x$. Technically, one cannot establish the Gaussian process uniformly over x because the uniform convergence rate of the standard kernel estimator should be adjusted by $\sqrt{\log n}$. See Chernozhukov, Chetverikov, and Kato (2014) for more discussion. Second, including covariates complicates the choice of optimal bandwidth. Third, averaging over X_i in (4.1.9) and (4.1.10) gives rise to U-statistics of order 2. Related techniques, such as, U-statistic projection, are applied to represent each as a degenerate U-statistic, contributing to the asymptotic variance.

5.2. Multiplier Bootstrap

In practice, inference based on estimating the asymptotic variance of the limit process can be overly complicated. In such cases, bootstrap methods can effectively construct uniform confidence bands. We propose to approximate the limit process $\mathbb{Z}_\delta=\mathbb{Z}_{q_1}(\tau)-\mathbb{Z}_{q_0}(\tau)$ derived in Corollary 5.1 by the multiplier bootstrap, analogously to the wild bootstrap approximation of Bartalotti, Calhoun, and He (2017). CHS establishes an approximation procedure based on a general framework of local Wald estimands. In particular, we develop the limit process of a specific and more complicated local Wald estimator, which can be seen as a generalization and an extension to the class of local Wald estimands analyzed by CHS.

Let $\{\xi\}_{i=1}^{n}$ be a random sample drawn from the standard normal distribution defined on $\left(\Omega^{\xi},\mathcal{F}^{\xi},\mathbb{P}^{\xi}\right)$, that is independently of the data $\{Z_i\}_{i=1}^{n}=\{Y_i,D_i,R_i,X_i\}_{i=1}^{n}$ defined on $(\Omega^{\upsilon},\mathcal{F}^{\upsilon},\mathbb{P}^{\upsilon})$. To simplify notation, we choose $h=h_r$, which is permitted by the bandwidth selection in Assumption 5.5. We define the estimated multiplier processes for $F_1(y)$ and $F_0(y)$ by

$$\widehat{\nu}_1^{\xi}(y)=\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\xi_i\widetilde{e}_1'\left(\widetilde{\Gamma}\widehat{f}_R(r_0)\right)^{-1}\widehat{\mathcal{Q}}_1(Z_i,y)K\left(\frac{R_i-r_0}{h}\right)$$

and

$$\widehat{\nu}_0^{\xi}(y)=\frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\xi_i\widetilde{e}_1'\left(\widetilde{\Gamma}\widehat{f}_R(r_0)\right)^{-1}\widehat{\mathcal{Q}}_0(Z_i,y)K\left(\frac{R_i-r_0}{h}\right),$$

where $\widehat{f}_R(r_0)$ estimates $f_R(r_0)$, and

$$\begin{aligned}\widehat{\mathcal{Q}}_1(Z_i,y)=&\,\widehat{\widetilde{\epsilon}}_i(y)\widetilde{\nu}_i(r_0)+\widehat{\mathcal{Q}}_{11}(Z_i,y)\mathbb{1}\left\{R_i>r_0\right\}+\widehat{\mathcal{Q}}_{12}(Z_i,y)\mathbb{1}\left\{R_i>r_0\right\}\\&+\widehat{\mathcal{Q}}_{13}(Z_i,y)\mathbb{1}\left\{R_i<r_0\right\}+\widehat{\mathcal{Q}}_{14}(Z_i,y)\mathbb{1}\left\{R_i<r_0\right\},\\\widehat{\mathcal{Q}}_0(Z_i,y)=&\,\widehat{\widetilde{\varepsilon}}_i(y)\widetilde{\nu}_i(r_0)+\widehat{\mathcal{Q}}_{01}(Z_i,y)\mathbb{1}\left\{R_i>r_0\right\}+\widehat{\mathcal{Q}}_{02}(Z_i,y)\mathbb{1}\left\{R_i>r_0\right\}\\&+\widehat{\mathcal{Q}}_{03}(Z_i,y)\mathbb{1}\left\{R_i<r_0\right\}+\widehat{\mathcal{Q}}_{04}(Z_i,y)\mathbb{1}\left\{R_i<r_0\right\}.\end{aligned}$$

Let $\bar{\nu}_i(r_0,u)=\left(1,\left(R_i-r_0\right)/h,\left(R_i-r_0\right)^2/h^2,u,u^2,\left(R_i-r_0\right)u/h\right)'$ and

$$\begin{aligned}\bar{\Delta}_i^{+}&=\int e_1'\left(\Gamma^{+}\right)^{-1}\bar{\nu}_i(r_0,u)K(u)du\cdot(1,0,\mu_2)',\\\bar{\Delta}_i^{-}&=\int e_1'\left(\Gamma^{-}\right)^{-1}\bar{\nu}_i(r_0,u)K(u)du\cdot(1,0,\mu_2)'.\end{aligned}$$

Moreover, let

$$\begin{aligned}\widetilde{\epsilon}_i(y)&=F_{1|X}(y|X_i)-E\left[F_{1|X}(y|X)|R=R_i\right],\\\widetilde{\varepsilon}_i(y)&=F_{0|X}(y|X_i)-E\left[F_{0|X}(y|X)|R=R_i\right],\\\epsilon_i(y)&=\mathbb{1}\{Y_i\le y\}D_i-E[\mathbb{1}\{Y\le y\}D|R=R_i,X=X_i],\\\varepsilon_i(y)&=\mathbb{1}\{Y_i\le y\}(1-D_i)-E[\mathbb{1}\{Y\le y\}(1-D)|R=R_i,X=X_i],\end{aligned}$$

and

$$\begin{aligned}\mathcal{Q}_{11}(Z_i,y)&=\epsilon_i(y)\left(1+\omega_0^{+}(y,X_i)\right)\bar{\Delta}_i^{+},\\\mathcal{Q}_{12}(Z_i,y)&=\varepsilon_i\left(\widetilde{q}_0\left(\widetilde{F}_1(y,X_i),X_i\right)\right)\left(1+\omega_0^{+}(y,X_i)\right)\bar{\Delta}_i^{+},\\\mathcal{Q}_{13}(Z_i,y)&=-\epsilon_i(y)\omega_0^{+}(y,X_i)\bar{\Delta}_i^{-},\\\mathcal{Q}_{14}(Z_i,y)&=-\varepsilon_i\left(\widetilde{q}_0\left(\widetilde{F}_1(y,X_i),X_i\right)\right)\omega_0^{+}(y,X_i)\bar{\Delta}_i^{-},\end{aligned}$$

$$\mathcal{Q}_{01}(Z_i,y) = -\epsilon_i\left(\widetilde{q}_1\left(\widetilde{F}_0(y,X_i),X_i\right)\right)\omega_1^-(y,X_i)\,\bar{\Delta}_i^+,$$
$$\mathcal{Q}_{02}(Z_i,y) = -\varepsilon_i(y)\,\omega_1^-(y,X_i)\,\bar{\Delta}_i^+,$$
$$\mathcal{Q}_{03}(Z_i,y) = \epsilon_i\left(\widetilde{q}_1\left(\widetilde{F}_0(y,X_i),X_i\right)\right)\left(1+\omega_1^-(y,X_i)\right)\bar{\Delta}_i^-,$$
$$\mathcal{Q}_{04}(Z_i,y) = \varepsilon_i(y)\left(1+\omega_1^-(y,X_i)\right)\bar{\Delta}_i^-.$$

Thus, for $(j,k)\in\{1,2\}\times\{1,2,3,4\}$, $\widehat{\widetilde{\epsilon}}_i(y)$, $\widehat{\widetilde{\varepsilon}}_i(y)$, $\widehat{\epsilon}_i(y)$, $\widehat{\varepsilon}_i(y)$ and $\widehat{\mathcal{Q}}_{jk}(Z_i,y)$ estimate $\widetilde{\epsilon}_i(y)$, $\widetilde{\varepsilon}_i(y)$, $\epsilon_i(y)$, $\varepsilon_i(y)$, and $\mathcal{Q}_{jk}(Z_i,y)$, respectively. By the Hadamard derivative, we construct the approximate estimated multiplier process

$$\widehat{\mathbb{Z}}_\delta(\tau) = \widehat{\mathbb{Z}}_{q_1}(\tau) - \widehat{\mathbb{Z}}_{q_1}(\tau) = \widehat{\mathbb{Z}}_0\big(\widehat{q}_0(\tau)\big)\big/\widehat{f}_0\big(\widehat{q}_0(\tau)\big) - \widehat{\mathbb{Z}}_1\big(\widehat{q}_1(\tau)\big)\big/\widehat{f}_1\big(\widehat{q}_1(\tau)\big),$$

where $\widehat{\mathbb{Z}}_1(y) = \widehat{\nu}_1^\xi(y)$ and $\widehat{\mathbb{Z}}_0(y) = \widehat{\nu}_0^\xi(y)$.

In the following, we discuss the estimation procedure for the multiplier bootstrap method. Specifically, we propose to use the kernel density estimator for $\widehat{f}_R(r_0)$:

$$\widehat{f}_R(r_0) = \frac{1}{nh_{f_R}}\sum_{i=1}^n K\left(\frac{R_i - r_0}{h_{f_R}}\right).$$

For $d\in\{0,1\}$, the conditional expectation $p_d(y,R_i) = E\left[F_{d|X}(y|X)|R=R_i\right]$ can be estimated by

$$\widehat{p}_d(y,R_i) = \sum_{j=1}^n \widehat{F}_{d|X}(y|X_j)K\left(\frac{R_j-R_i}{h_p}\right)\Bigg/\sum_{j=1}^n K\left(\frac{R_j-R_i}{h_p}\right).$$

Now, we consider $\widetilde{p}(y,d,R_i,X_i) = E[\mathbb{1}\{Y\le y\}\mathbb{1}\{D=d\}|R=R_i, X=X_i]$ for $d\in\{0,1\}$. Notice that $\widetilde{p}(y,d,r,x)$ is discontinuous at $r=r_0$, we should jointly estimate

$$\widehat{\widetilde{p}}(y,d,R_i,X_i)\mathbb{1}\{R_i>r_0\}$$
$$= \mathbb{1}\{R_i>r_0\}\frac{\sum_{j=1}^n \mathbb{1}\{Y_j\le y\}\mathbb{1}\{D_j=d\}\mathbb{1}\{R_j>r_0\}K\left(\frac{R_j-R_i}{h_{\widetilde{p}}}\right)K\left(\frac{X_j-X_i}{h_{\widetilde{p}}}\right)}{\sum_{j=1}^n \mathbb{1}\{R_j>r_0\}K\left(\frac{R_j-R_i}{h_{\widetilde{p}}}\right)K\left(\frac{X_j-X_i}{h_{\widetilde{p}}}\right)}$$

and

$$\widehat{\widetilde{p}}(y,d,R_i,X_i)\mathbb{1}\{R_i<r_0\}$$
$$= \mathbb{1}\{R_i<r_0\}\frac{\sum_{j=1}^n \mathbb{1}\{Y_j\le y\}\mathbb{1}\{D_j=d\}\mathbb{1}\{R_j<r_0\}K\left(\frac{R_j-R_i}{h_{\widetilde{p}}}\right)K\left(\frac{X_j-X_i}{h_{\widetilde{p}}}\right)}{\sum_{j=1}^n \mathbb{1}\{R_j<r_0\}K\left(\frac{R_j-R_i}{h_{\widetilde{p}}}\right)K\left(\frac{X_j-X_i}{h_{\widetilde{p}}}\right)}.$$

The remaining terms can be estimated as follows:

$$\widehat{\omega}_0^+(y,x) = \widehat{f}_{YD|RX}\left(\widehat{q}_0\left(\widehat{\widehat{F}}_1(y,x),x\right),0,r_0^+,x\right) \Big/ \widehat{\widehat{f}}_0\left(\widehat{q}_0\left(\widehat{\widehat{F}}_1(y,x),x\right),x\right),$$

$$\widehat{\omega}_1^-(y,x) = \widehat{f}_{YD|RX}\left(\widehat{q}_1\left(\widehat{\widehat{F}}_0(y,x),x\right),1,r_0^-,x\right) \Big/ \widehat{\widehat{f}}_1\left(\widehat{q}_1\left(\widehat{\widehat{F}}_0(y,x),x\right),x\right),$$

$$\widehat{f}_{1|X}(y|x) = \widehat{f}_{YD|RX}\left(y,1,r_0^+,x\right) + \widehat{f}_{YD|RX}\left(\widehat{q}_0\left(\widehat{\widehat{F}}_1(y,x),x\right),0,r_0^+,x\right),$$

$$\widehat{f}_{0|X}(y|x) = \widehat{f}_{YD|RX}\left(y,0,r_0^-,x\right) + \widehat{f}_{YD|RX}\left(\widehat{q}_1\left(\widehat{\widehat{F}}_0(y,x),x\right),1,r_0^-,x\right),$$

$$\widehat{\widehat{f}}_1(y,x) = \widehat{f}_{YD|RX}\left(y,1,r_0^+,x\right) - \widehat{f}_{YD|RX}\left(y,1,r_0^-,x\right),$$

$$\widehat{\widehat{f}}_0(y,x) = \widehat{f}_{YD|RX}\left(y,0,r_0^-,x\right) - \widehat{f}_{YD|RX}\left(y,0,r_0^+,x\right),$$

and

$$\widehat{f}_1(y) = \sum_{i=1}^{n} \widehat{f}_{1|X}(y|X_i) K\left(\frac{R_i - r_0}{h_f}\right) \Big/ \sum_{i=1}^{n} K\left(\frac{R_i - r_0}{h_f}\right),$$

$$\widehat{f}_0(y) = \sum_{i=1}^{n} \widehat{f}_{0|X}(y|X_i) K\left(\frac{R_i - r_0}{h_f}\right) \Big/ \sum_{i=1}^{n} K\left(\frac{R_i - r_0}{h_f}\right),$$

$$\widehat{f}_{YD|RX}\left(y,d,r_0^+,x\right) = \frac{h_{fx}^{-1}\sum_{i=1}^{n} K\left(\frac{Y_i-y}{h_{fx}}\right) K\left(\frac{R_i-r_0}{h_{fx}}\right) K\left(\frac{X_i-x}{h_{fx}}\right) \mathbb{1}\{D_i = d\}\mathbb{1}\{R_i > r_0\}}{\sum_{i=1}^{n} K\left(\frac{R_i-r_0}{h_{fx}}\right) K\left(\frac{X_i-x}{h_{fx}}\right)\mathbb{1}\{R_i > r_0\}},$$

$$\widehat{f}_{YD|RX}\left(y,d,r_0^-,x\right) = \frac{h_{fx}^{-1}\sum_{i=1}^{n} K\left(\frac{Y_i-y}{h_{fx}}\right) K\left(\frac{R_i-r_0}{h_{fx}}\right) K\left(\frac{X_i-x}{h_{fx}}\right) \mathbb{1}\{D_i = d\}\mathbb{1}\{R_i < r_0\}}{\sum_{i=1}^{n} K\left(\frac{R_i-r_0}{h_{fx}}\right) K\left(\frac{X_i-x}{h_{fx}}\right)\mathbb{1}\{R_i < r_0\}}.$$

Under suitable conditions, with probability approaching one, the process $\widehat{\mathbb{Z}}_\delta(\tau)$ weakly converges to the limit process, $\mathbb{Z}_\delta(\tau)$, of interest conditionally on the data $\{Y_i, D_i, R_i, X_i\}_{i=1}^n$. Similarly, in the CHS model, we could apply this result to test several hypotheses, such as uniform treatment nullity and treatment homogeneity across quantiles. In our case, we can use the approximate estimated multiplier process to construct uniform confidence bands for the QTE. In summary, we provide a step-by-step procedure below.

Algorithm 1. (Practical guidelines for constructing uniform confidence bands for QTE).

Step 1. Pick a finite set Y$\subseteq S_Y$ of grid points of outcome values and a finite set T $\subseteq [a, 1-a]$ of grid points of quantiles for some $0 < a < 1/2$. Estimate $F\hat{}_1(y)$ and $F\hat{}_0(y)$ for all $y \in$Y.

Step 2. Calculate $q\hat{}_1(\tau)$ and $q\hat{}_0(\tau)$ for each $\tau \in$T, and then compute $\delta\hat{}(\tau)$ $=q\hat{}_1(\tau) - q\hat{}_0(\tau)$ for each $\tau \in$T.

Step 3. Compute $f\hat{}_R(r_0)$, $f\hat{}_1(q\hat{}_1(\tau))$, $f\hat{}_0(q\hat{}_0(\tau))$, $\epsilon\tilde{}\hat{}_i(y)$, $\varepsilon\tilde{}\hat{}_i(y)$, $\epsilon\hat{}_i(y)$, $\varepsilon\hat{}_i(y)$, and Q$\hat{}_{jk}(Z_i, y)$ for $(j, k) \in \{1, 2\} \times \{1, 2, 3, 4\}$.

https://doi.org/10.1017/S026646662300021X Published online by Cambridge University Press

Step 4. For each bootstrap iteration $b = 1, 2, \ldots, B$, generate independent standard normal $\xi^b = \{\xi_i^{\,b}\}_{i=1}{}^n$ independently from the data, and compute $\nu\hat{}_{1,\,b}{}^{\xi}(y)$ and $\nu\hat{}_{0,\,b}{}^{\xi}(y)$ for each $y \in$Y.

Step 5. Construct $\mathbb{Z}\,\hat{}_{\delta,\,b}(\tau)$ for each $\tau \in$T:

$$\widehat{\mathbb{Z}}_{\delta,b}(\tau) = \widehat{\nu}^{\xi}_{0,b}\left(\widehat{q}_0(\tau)\right)/\widehat{f}_0\left(\widehat{q}_0(\tau)\right) - \widehat{\nu}^{\xi}_{1,b}\left(\widehat{q}_1(\tau)\right)/\widehat{f}_1\left(\widehat{q}_1(\tau)\right).$$

Step 6. Set $C^{\wedge B}(a,\ 1-a;l)$ equal to the $(1-l)$th quantile of $\{\max_{\tau\in T}|\mathbb{Z}\,\hat{}_{\delta,\,b}(\tau)|\}_{b=1}{}^B$, and construct an asymptotically valid $100(1-l)\%$ uniform confidence band over $[a, 1-a]$ by

$$\left[\widehat{\delta}(\tau) \pm \frac{1}{\sqrt{nh}}\widehat{C}^B(a,1-a;l) : \tau \in \mathcal{T}\right].$$

Before formally establishing the validity of the multiplier bootstrap, we first introduce some useful notations. Let $\rightarrow^p_\bullet$ denote the convergence in probability with respect to the probability measure $\mathbb{P}^\bullet$, let $E_{\xi|\upsilon}$ denote the expectation with respect to the probability measure $\mathbb{P}^\upsilon \times \mathbb{P}^\xi$ given the events in $\mathcal{F}^\upsilon$, and let E_υ denote the expectation with respect to the probability measure $\mathbb{P}^\upsilon$. We define the conditional weak convergence in probability, or convergence of the conditional limit laws of bootstraps, denoted by $Z_n \rightsquigarrow^p_\xi \mathbb{Z}$, by $\sup_{T\in BL_1}\left|E_{\xi|\upsilon}T(Z_n) - ET(\mathbb{Z})\right| \rightarrow^p_\upsilon 0$, where BL_1 is the set of functions with Lipschitz constant and supremum norm bounded by 1.

Assumption 5.7. $\{\xi\}_{i=1}^n$ is a random sample drawn from the standard normal distribution defined on $\left(\Omega^\xi, \mathcal{F}^\xi, \mathbb{P}^\xi\right)$, that is independently of the data $\{Y_i, D_i, R_i, X_i\}_{i=1}^n$ defined on $(\Omega^\upsilon, \mathcal{F}^\upsilon, \mathbb{P}^\upsilon)$.

Assumption 5.8. The bandwidths should satisfy:

1. $\widetilde{h} \rightarrow 0$ for $\widetilde{h} \in \{h_{f_R}, h_p, h_{\widetilde{p}}, h_f, h_{f_x}\}$.
2. $nh_{f_R} \rightarrow \infty$, $nh_f \rightarrow \infty$, $\frac{\log n}{nh_p} \rightarrow 0$, $\frac{\log n}{nh_x^{1+L}} \rightarrow 0$, and $\frac{\log n}{nh_{\widetilde{p}}^{2+L}} \rightarrow 0$.

Assumption 5.7 is the standard assumption for the multiplier, score, and wild bootstrap methods (cf. Kosorok, 2003, 2008). Assumption 5.8 imposes standard bandwidth conditions to ensure the first-stage estimators' uniform consistency, which are required to derive the uniform validity of the multiplier bootstrap.

THEOREM 3. *Under the same assumptions as Theorem 1 and Assumptions 5.1–5.8,*

$$\begin{pmatrix} \widehat{\nu}_1^{\xi}(y) \\ \widehat{\nu}_0^{\xi}(y) \end{pmatrix} \rightsquigarrow^p_\xi \begin{pmatrix} \mathbb{Z}_1(y) \\ \mathbb{Z}_0(y) \end{pmatrix}$$

and

$$\begin{pmatrix} \widehat{\mathbb{Z}}_{q_1}(\tau) \\ \widehat{\mathbb{Z}}_{q_0}(\tau) \end{pmatrix} \rightsquigarrow^p_\xi \begin{pmatrix} \mathbb{Z}_{q_1}(\tau) \\ \mathbb{Z}_{q_1}(\tau) \end{pmatrix}.$$

COROLLARY 5.2. *Under the same assumptions as Theorem 1 and Assumptions 5.1–5.8,*

$$\widehat{\mathbb{Z}}_\delta(\tau) \rightsquigarrow^p_\xi \mathbb{Z}_\delta(\tau).$$

Corollaries 5.1 and 5.2 show that we can use the estimated multiplier process $\widehat{\mathbb{Z}}_\delta(\tau)$ to approximate the limit process of $\sqrt{nh}\left(\widehat{\delta}(\tau)-\delta(\tau)\right)$ in practice.

Remark 5.2.1. We can similarly construct uniform confidence bands for the DTEs and ATE through the multiplier bootstrap. For the DTEs, Steps 1–4 are identical, and Step 5 can be discarded. Step 6 changes as follows: we replace $\widehat{\mathbb{Z}}_{\delta,b}(\tau)$ and $\widehat{\delta}(\tau)$ with $\widehat{\nu}^\xi_{1,b}(y)-\widehat{\nu}^\xi_{0,b}(y)$ and $\widehat{F}_1(y)-\widehat{F}_0(y)$, respectively. For the ATE, set $\mathcal{Y}=\{y_0,y_1,\ldots,y_K\}$, where the grids are defined in Remark 4.1.3. Steps 1–4 are identical. We calculate

$$\widehat{\mathbb{Z}}_b=\sum_{k=1}^K y_k\left[\left(\widehat{\nu}^\xi_{1,b}(y_k)-\widehat{\nu}^\xi_{0,b}(y_k)\right)-\left(\widehat{\nu}^\xi_{1,b}(y_{k-1})-\widehat{\nu}^\xi_{0,b}(y_{k-1})\right)\right],$$

for $b=1,2,\ldots,B$. Let $\mathrm{std}\left(\widehat{\mathbb{Z}}_b\right)$ be the standard error of $\left\{\widehat{\mathbb{Z}}_b\right\}_{b=1}^B$. The $100(1-l)\%$ confidence interval of $\widehat{\delta}_{ATE}$ is then given by

$$\left[\widehat{\delta}_{ATE}\pm\mathrm{std}\left(\widehat{\mathbb{Z}}_b\right)\Phi^{-1}\left(1-\frac{l}{2}\right)\right],$$

where Φ^{-1} is the inverse of Φ, and Φ is the CDF of standard normal distribution.

5.3. Detailed Algorithm for Estimation and Inference

Sections 4 and 5.2 describe the estimation and the multiplier bootstrap procedure for constructing uniform confidence bands for QTE. This subsection summarizes practical guidelines on estimation and inference for QTEs. Additionally, we use this procedure for both simulation studies and empirical applications.

Algorithm 2 (Practical guidelines on estimation and inference for QTE).

Step 1. Pick a finite set $\mathrm{Y}\subseteq S_Y$ of grid points of outcome values and a finite set $\mathrm{T}\subseteq[a,1-a]$ of grid points of quantiles for some $0<a<1/2$. For any $y\in\mathrm{Y}$ and $W\in\left\{\mathbb{1}\{Y<y\}D,\mathbb{1}\{Y<y\}(1-D)\right\}$, estimate $m^+(W|X_i)$ and $m^-(W|X_i)$ by (4.1.1) and (4.1.2), respectively. The bandwidth h and h_x are chosen as $h=cn^{-\zeta}$ and $h_x=c_xn^{-\zeta_x}$,respectively, where $\zeta=1/5$, and

$$\zeta_x=\begin{cases}\frac{1}{5}, & \text{if } L\le 3,\\ \left(\frac{2}{15},\frac{4}{5L}\right), & \text{if } 4\le L\le 5,\\ \left(\frac{2}{5(1+p)},\frac{4}{5L}\right), & \text{if } L\ge 6.\end{cases}$$

The constants c and c_x are determined by a grid search, formally stated before Assumption 5.6.

https://doi.org/10.1017/S026646662300021X Published online by Cambridge University Press

Step 2. For any $y \in$Y, calculate $F\tilde{}\hat{}_1(y, X_i)$ and $F\tilde{}\hat{}_0(y, X_i)$ by (4.1.6), (4.1.7), and Step 1. For any $\tau \in$T, calculate $q\tilde{}\hat{}_1(\tau, X_i)$ and $q\tilde{}\hat{}_0(\tau, X_i)$ by inverting $F\tilde{}\hat{}_1(y, X_i)$ and $F\tilde{}\hat{}_0(y, X_i)$ with respect to y.

Step 3. For any $y \in$Y, calculate $F\hat{}_{1|X}(y|X_i)$ and $F\hat{}_{0|X}(y|X_i)$ by (4.1.4) and (4.1.8), respectively.

Step 4. For any $y \in$Y, estimate $F_1(y)$ and $F_0(y)$ by (4.1.9) and (4.1.10), respectively. The bandwidth h_r is chosen as $h_r = c_r n^{-\zeta_r}$, where $\zeta_r = 1/5$, and c_r is determined by a grid search.

Step 5. For any $\tau \in$T, calculate $q\hat{}_1(\tau)$ and $q\hat{}_0(\tau)$ by inverting $F\hat{}_1(y)$ and $F\hat{}_0(y)$, respectively.

Step 6. For any $\tau \in$T, calculate $\delta\hat{}(\tau)$ by (4.1.11).

Step 7. Construct uniform confidence bands for $\delta\hat{}(\tau)$ by Algorithm 1. The bandwidths h_{f_R}, h_p, $h_{p\sim}$, h_f, and h_{f_x} are determined by rule of thumb, for example, Section 1.7 in Li and Racine (2007, p. 26).

6. MONTE CARLO SIMULATION

In this section, we conduct a set of Monte Carlo simulations to evaluate the finite-sample performance of the proposed estimator and the effectiveness of the uniform confidence bands. We consider the following data-generating process (DGP):

$$\begin{cases} Y_i = \mu(R_i) + \beta_1 D_i + (1+\gamma_1 D_i)\cdot U_i + 0.1X_i + 0.1X_i^2, \\ D_i = \mathbb{1}\{\alpha \cdot \mathbb{1}\{R_i \geq 0\} - 1 \geq V_i\}, \end{cases}$$

where β_1=0.2 and γ_1=1. We define the functional form of $\mu(r)$ as in Lee (2008):

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

We generate (R_i, X_i, U_i, V_i) by

$$\begin{pmatrix} R_i \\ X_i \\ U_i \\ V_i \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 0.1782^2 & 0 & 0 & 0 \\ 0 & 0.5^2 & 0 & 0 \\ 0 & 0 & 0.1295^2 & 0 \\ 0 & 0 & 0 & 0.5^2 \end{pmatrix}\right).$$

Note that the threshold r_0 is 0 under this DGP. The sample size n, and the jump in the probability of treatment at the threshold, denoted as Δp, are two key parameters that influence the estimator's performance. The parameter α controls the jump in the probability:

$$\Delta p = \Phi(\alpha - 1) - \Phi(-1).$$

We conduct the simulations for several designs based on 2,500 multiplier bootstrap replications and 500 Monte Carlo replications for each sample size $n \in \{2,000, b3,000, 5,000\}$. We investigate the performance of the estimator with $\alpha = 2$

https://doi.org/10.1017/S026646662300021X Published online by Cambridge University Press

Table 1. Uniform coverage probability of the true quantile treatment effects by the uniform confidence bands

	$n = 2,000$	$n = 3,000$	$n = 5,000$
$\alpha = 2$	0.948	0.954	0.956
$\alpha = 4$	0.948	0.952	0.952

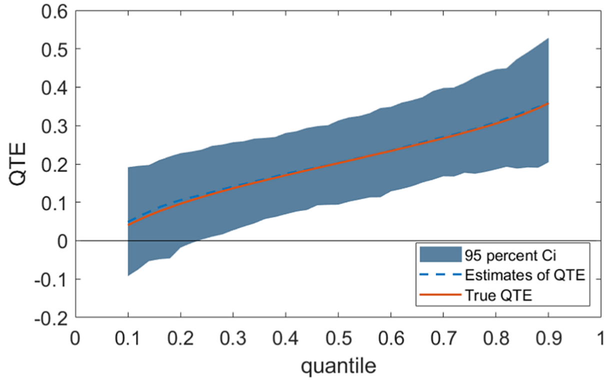
as small Δp, and $\alpha = 4$ as large Δp. We use a uniform kernel, and the nominal acceptance probability is 95%.

Figures 1 and 2 show the estimation results for each design, as well as the 95% confidence interval, spanned by the 2.5th and 97.5th percentile estimates. Figure 1 plots the results for the small Δp, which imply a jump in the probability of treatment at the threshold of about 68.27%. We observe that the confidence intervals are somewhat wider than other cases in Figure 2, but the bias is small enough to be neglected. Figure 2 plots the results for the large Δp, which imply a jump in the probability of treatment at the threshold of about 84%. The bias remains negligible and appears to be quite precise. Moreover, the confidence intervals are substantially narrower than small Δp. We can see that the estimators become more accurate, and the confidence intervals become narrower as the sample size increases. Overall, the estimator performs better for the case with a large sample size and large jump at the threshold than the case with a small sample size and small jump.

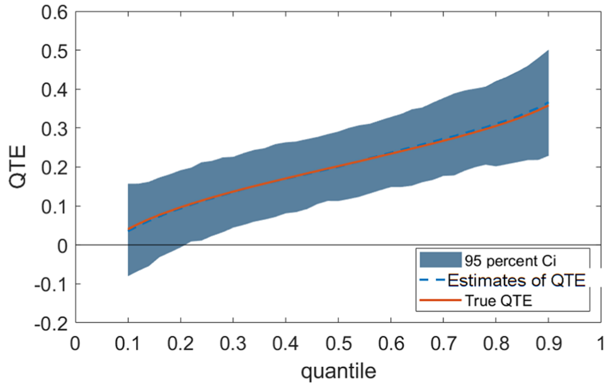
Table 1 summarizes the results for the uniform coverage probability across sample sizes of $n \in \{2,000, 3,000, 5,000\}$ and $\alpha \in \{2,4\}$. The uniform coverage probability almost keeps the nominal size 95% across all cases. This confirms the effectiveness of the uniform confidence bands.

7. APPLICATION: THE PRIME MINISTER'S VILLAGE ROAD PROGRAM IN INDIA

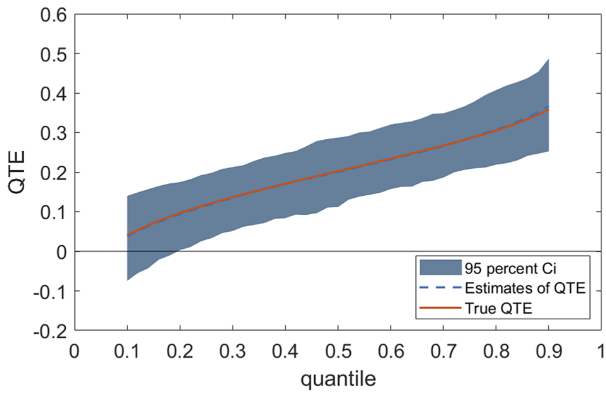
Rural road construction promotes farm and nonfarm economic growth, as well as poverty reduction. However, the causal impact of rural roads is difficult to assess because of the endogeneity of road placement: the high costs and potentially large benefits of infrastructure investments mean that the placement of new roads is typically correlated with both the economic and political characteristics of locations. To overcome this endogeneity challenge, AN20 exploit the implementation rule in the Indian Prime Minister's Village Road Program, or PMGSY, which targeted roads at villages with population exceeding certain thresholds to identify the causal impact of rural roads using a fuzzy RD design. The main finding of AN20 is that there exist no large effects of this program on agricultural outcomes, income, or assets in villages; the main effect of new feeder roads is to facilitate the movement of workers out of agriculture. In other words, the new rural roads make it easier for workers to access non-agricultural opportunities. In this section, we will employ



(a) n=2000



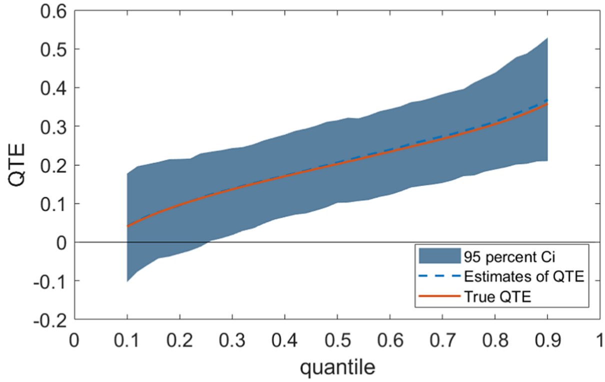
(b) n=3000



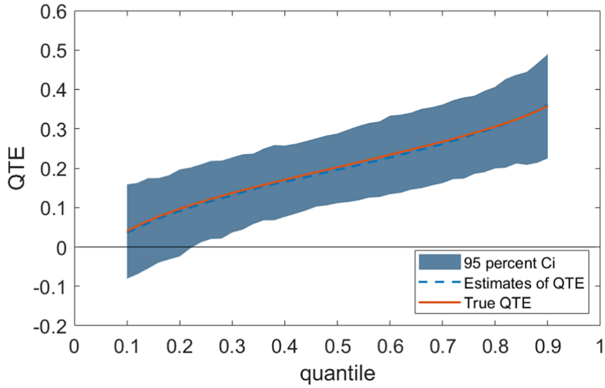
(c) n=5000

FIGURE 1. The figures show estimators and 95% confidence intervals from Monte Carlo simulations with a small Δp.

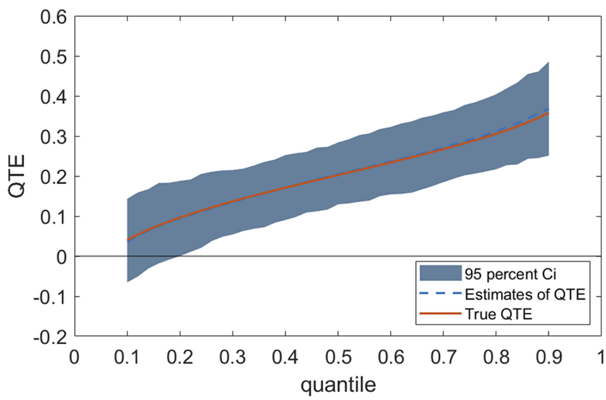
https://doi.org/10.1017/S026646662300021X Published online by Cambridge University Press



(a) n=2000



(b) n=3000



(c) n=5000

FIGURE 2. The figures show estimators and 95% confidence intervals from Monte Carlo simulations with a large Δp .

https://doi.org/10.1017/S026646662300021X Published online by Cambridge University Press

TABLE 2. Summary statistics and balance tests

	Mean	SD	RD estimate	p-value
Normalized population	−24.052	128.816		
New road	0.356	0.479		
Agricultural occupation index	−0.013	0.988		
Control variables				
Literate (share)	0.452	0.155	−0.008	0.724
Scheduled caste (share)	0.141	0.171	−0.022	0.467
Subsistence agriculture (share)	0.437	0.266	0.021	0.609
HH > INR 250 (share)	0.755	0.279	−0.026	0.570
Medical center	0.170	0.376	−0.091	0.152

Note: Normalized village population is the reported population minus the threshold, either 500 or 1,000. An indicator for a new road would equal one if the village had access roads built before 2012. The agricultural occupation index comprises the share of workers in agriculture and the opposite of the share of workers in manual labor generated, according to Anderson (2008). The regression discontinuity estimators use a uniform kernel and a bandwidth of 84 with heteroskedasticity robust standard errors.

our method to complement the findings by AN20 and explore the heterogeneous effects of new road construction on the reallocation of labor.

Until 2001, approximately 49% of Indian villages were still not accessible to all-season roads. To remedy this, the Government of India launched the PMGSY in 2000. The program envisaged connecting all habitations with a population of 1,000 or more by 2003, a population of 500 or more by 2007, and a population of 250 or more after that. This population-based eligibility standard was lower in some states, such as hill states, tribal areas, desert areas, and districts affected by left-wing extremism. Moreover, they allowed the implementation of a few other rules to help decide on the allocation of roads at the same time. The program's guidelines indicate a discontinuous increase in the probability of new road construction at the population threshold, making it possible to perform the analysis under an RD framework.

The dataset combines village-level variables from the PMGSY program, household and individual characteristics from the 2011 Population Census based on names, and microdata from the Socioeconomic and Caste Census of 2012. Following AN20, we restrict the sample to villages in Chhattisgarh, Gujarat, Madhya Pradesh, Maharashtra, Orissa, and Rajasthan, with population thresholds of 500 or 1,000. The final dataset comprises a sample of 29,426 villages that did not have an all-weather road in 2001 and were matched across all primary datasets. The main variables include village population (running variable), an indicator for new road construction (treatment variable), sectoral allocation of labor measured by agricultural occupation index (outcome variable), and village-level variables (control variable), including the literacy rate, share of inhabitants that belong to a scheduled caste, share of households who are subsistence farmers, share of households

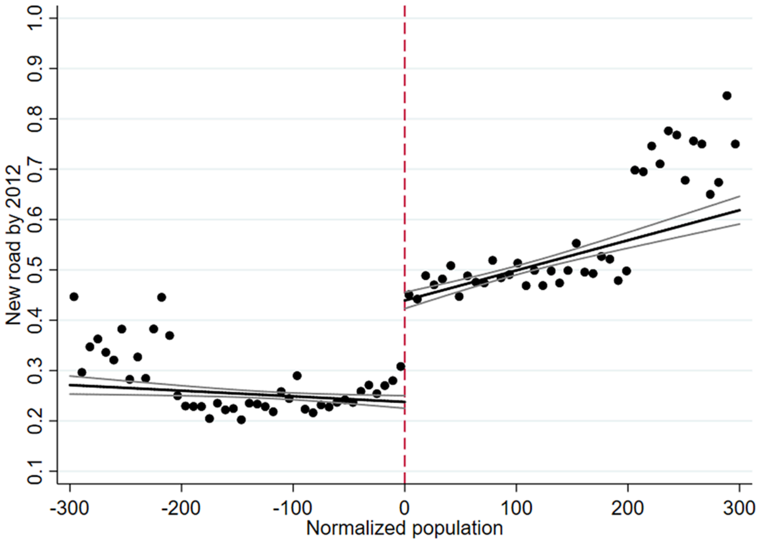


FIGURE 3. This figure presents the impact of population on new road construction. It shows the discontinuity in the probability of receiving a new road at the treatment threshold. We find a discrete large jump occurring at the threshold by 21.35 percentage points (from 23.06% to 44.41%).

earning over 250 INR cash per month, and an indicator for a medical center. The agricultural occupation index comprises the share of workers in agriculture and the opposite of the share of workers in manual labor generated, according to Anderson (2008). Table 2 reports the summary statistics and balancing tests of the variables used in our empirical study. The balancing tests are based on running RD estimations with each village-level variable as the outcome variable, respectively, while controlling for the remaining village-level variables. We can see that all covariates are balanced around the threshold.

Figure 3 shows the discontinuity in the probability of new road construction before 2012 at the treatment threshold. We obtain the curve by regressing the dummy variable for receiving a new road on the dummy variable for villages of the population over the threshold while controlling for the piecewise linear function of the population relative to the threshold. There is a discrete large jump occurring at the threshold by 21.35 percentage points (from 23.06% to 44.41%). The sudden jump at the threshold shows that the village population can strongly predict the probability of receiving new roads.

We first estimate the LATE with the following equation using 2SLS with an indicator for villages with the populations above the threshold as the instrument.

$$\begin{aligned}&\textit{Agricultural occupation index} = \gamma_0 + \gamma_1 \textit{New road} + \gamma_2 \textit{Normalized pop}\\&\quad + \gamma_3 \mathbb{1}\{\textit{Normalized pop} \geq 0\} \times \textit{Normalized pop} + \gamma_4 \textit{Control} + \varepsilon.\end{aligned}$$

https://doi.org/10.1017/S026646662300021X Published online by Cambridge University Press

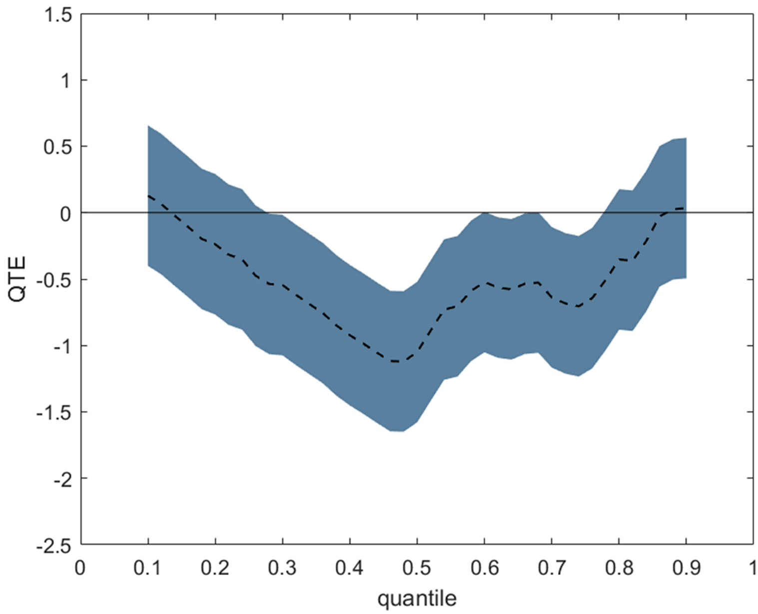


FIGURE 4. This figure plots the estimated quantile treatment effects *without* covariates and 90% confidence intervals (based on multiplier bootstrap) of new road construction on the distribution of all employment shares in India. A uniform kernel and bandwidths ranging from 51 to 300 were used.

The optimal bandwidth, according to Imbens and Kalyanaraman's (2012) method, is 84. The LATE (captured by γ_1) is -0.339, with a heteroskedasticity robust standard error of 0.181. This result means receiving a new road leads to a significant reallocation of workers away from agriculture. We then use QTE to examine the heterogeneity of the treatment effect as a function of subgroups located differently on the outcome distribution. Figure 4 presents the estimated QTEs without covariates using a uniform kernel and MSE-minimizing bandwidths ranging from 51 to 300. The optimal bandwidths are selected according to the procedure described in Appendix C of the Supplementary Material. The figure also plots the 90% uniform confidence band based on the multiplier bootstrap method. The effects of rural roads on agricultural employment share as a function of the percentile of the outcome distribution show a U-shaped relationship. The figure indicates a significant negative effect of new road construction in the central parts of the agricultural employment share distribution. However, there is no evidence of a significant effect of new road construction in the lower and upper parts of the agricultural employment share distribution. Figure 5 graphs the QTE estimates with covariates using a uniform kernel and bandwidths ranging from 92 to 172, which yields a similar pattern across quantiles.

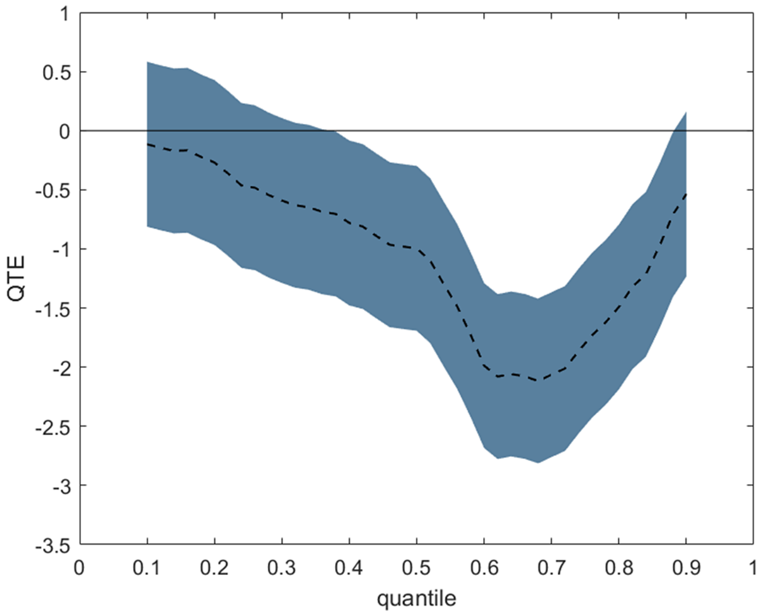


FIGURE 5. This figure plots the estimated quantile treatment effects *with* covariates and 90% confidence intervals (based on multiplier bootstrap) of new road construction on the distribution of all employment shares in India. A uniform kernel and bandwidths ranging from 92 to 172 were used.

AN20 provide evidence in their Appendix Table A5 that participation in non-agricultural occupations is lower for laborers in households with more land. Cultural obligation (the expectation of occupational succession) and transaction costs in the market for land may lead to persistent work in agriculture owning more land (Fernando 2018). To understand the reasons for the heterogeneous effects of new road construction, we divide the total sample into five subgroups based on the empirical quantiles of the agricultural occupation index (outcome variable). We calculate the average proportion of households owning more than 2 acres of agricultural land in each subsample. In Table 3, we see that households in the villages with a higher share of agricultural employment own relatively more agricultural land on average. Thus, we argue that workers in villages with a high share of agriculture employment are less likely to move out of the agricultural sector because of their greater landholdings and high transfer costs. On the other hand, according to Fernando (2018), laborers with small landholdings are those with the highest returns for their labor in the non-agricultural sector. They are likely to have left of the village before the program was implemented. Thus, the probability of moving out of agriculture for laborers in villages with a low share of agricultural employment will decline, even if the new roads lower the transport cost.

Table 3. Mean of the proportion of households owning more than 2 acres of agricultural land of different agricultural occupation index levels

Quantile intervals	Proportion of household owning more than 2 acres of agricultural land (%)
0–20%	17.403
20%–40%	23.602
40%–60%	29.059
60%–80%	33.266
80%–1	34.614

Note: Intervals are sorted by the level of agricultural occupation index in ascending order.

8. CONCLUSION

This study proposes a new and constructive approach for estimating QTEs in the RD model under a local RS condition, which restricts the evolution of individual ranks across treatment status in a neighborhood around the threshold. The feature that distinguishes our study from the prior work is that our approach directly estimates QTEs and the ATE for a whole population instead of for only the compliant subpopulation at the threshold. We derive closed-form solutions for the estimands of the potential outcome CDFs for the whole population, which are compositions of conditional CDFs and probabilities of observed variables, and can thus be estimated with a plug-in approach.

We demonstrate the functional central limit theorems and bootstrap validity results for the QTE estimators by explicitly accounting for observed covariates. In particular, we develop a multiplier bootstrap-based inference method with robustness against large bandwidths that applies to uniform inference. We also propose a test for the local RS assumption. To illustrate the estimation approach and its properties, we provide a simulation study and estimate the impacts of India's 40-billion-dollar national rural road construction program on the reallocation of labor out of agriculture.

SUPPLEMENTARY MATERIAL

Zequn Jin, Yu Zhang, Zhengyu Zhang, Yahong Zhou (2023): Supplement to "Identification and Inference in a Quantile Regression Discontinuity Design under Rank Similarity with Covariates," Econometric Theory Supplementary Material. To view, please visit: https://doi.org/10.1017/S026646662300021X

Appendix A. Proofs

Proof of Lemma 3.1. We only prove the first moment condition; the second is similar. For $d = 0,1$, let $\eta_d(r,x,y)$ denote the inverse of $q_d(r,x,u)$ with respect to u, namely,

$\eta_d(r,x,q_d(r,x,u)) = u$ for each r, x, and u.

$$\lim_{\epsilon\to 0^+} P\left(Y \le q_D(r_0,x,\tau)\middle| r_0 < R < r_0+\epsilon, X=x\right)$$

$$=^{(1)} \lim_{\epsilon\to 0^+} P\left(q_D(R,x,U_D) \le q_D(r_0,x,\tau)\middle| r_0 < R < r_0+\epsilon, X=x\right)$$

$$=^{(2)} \lim_{\epsilon\to 0^+} P\left(U_D \le \eta_D(R,x,q_D(r_0,x,\tau))\middle| r_0 < R < r_0+\epsilon, X=x\right)$$

$$=^{(3)} \lim_{\epsilon\to 0^+} \frac{\int_{-\infty}^{+\infty}\int_{r_0}^{r_0+\epsilon} P\left(U_D \le \eta_D(R,x,q_D(r_0,x,\tau))\middle| R=r, X=x, V=v\right) f_{RV|X}(r,v|x)drdv}{P(r_0 < R < r_0+\epsilon|X=x)}$$

$$=^{(4)} \lim_{\epsilon\to 0^+} \frac{\int_{-\infty}^{+\infty}\int_{r_0}^{r_0+\epsilon} P\left(U_{\rho(1,r,x,v)} \le \eta_{\rho(1,r,x,v)}\left(r,x,q_{\rho(1,r,x,v)}(r_0,x,\tau)\right)|R=r, X=x, V=v\right) f_{RV|X}(r,v|x)drdv}{P(r_0 < R < r_0+\epsilon|X=x)}$$

$$=^{(5)} \lim_{\epsilon\to 0^+} \frac{\int_{-\infty}^{+\infty}\int_{r_0}^{r_0+\epsilon} P\left(U_0 \le \eta_{\rho(1,r,x,v)}\left(r,x,q_{\rho(1,r,x,v)}(r_0,x,\tau)\right)\middle| R=r, X=x, V=v\right) f_{RV|X}(r,v|x)drdv}{P(r_0 < R < r_0+\epsilon|X=x)}$$

$$=^{(6)} \int_{-\infty}^{+\infty} \lim_{\epsilon\to 0^+} \frac{\int_{r_0}^{r_0+\epsilon} P\left(U_0 \le \eta_{\rho(1,r,x,v)}\left(r,x,q_{\rho(1,r,x,v)}(r_0,x,\tau)\right)\middle| R=r, X=x, V=v\right) f_{RV|X}(r,v|x)dr}{\int_{r_0}^{r_0+\epsilon} f_{R|X}(r|x)dr} dv$$

$$=^{(7)} \int_{-\infty}^{+\infty} \lim_{\epsilon\to 0^+} \frac{P\left(U_0 \le \eta_{\rho(1,\bar{r},x,v)}\left(\bar{r},x,q_{\rho(1,\bar{r},x,v)}(r_0,x,\tau)\right)\middle| R=\bar{r}, X=x, V=v\right) f_{RV|X}(\bar{r},v|x)}{f_{R|X}(\check{r}|x)} dv$$

$$=^{(8)} \frac{\int_{-\infty}^{+\infty} P\left(U_0 \le \eta_{\rho(1,r_0,x,v)}\left(r_0,x,q_{\rho(1,r_0,x,v)}(r_0,x,\tau)\right)\middle| R=r_0, X=x, V=v\right) f_{RV|X}(r_0,v|x)dv}{f_{R|X}(r_0|x)}$$

$$=^{(9)} \frac{\int_{-\infty}^{+\infty} P\left(U_0 \le \tau\middle| R=r_0, X=x, V=v\right) f_{RV|X}(r_0,v|x)dv}{f_{R|X}(r_0|x)}$$

$$=^{(10)} \int_{-\infty}^{+\infty} P\left(U_0 \le \tau\middle| R=r_0, X=x, V=v\right) f_{V|RX}(v|r_0,x)dv =^{(11)} P\left(U_0 \le \tau\middle| R=r_0, X=x\right) =^{(12)} \tau,$$

where (1) by DGP; (2) is by the definition of $\eta_d(r,x,y)$; (3) is by the formula of conditional probability; (4) is by equation (3); (5) is by Assumption 2.3; (6) is by the dominated convergence theorem; (7) is by the integral mean value theorem, where $\bar{r}$ and $\check{r}$ lie between r_0 and $r_0+\epsilon$; (8) is by a number of smooth conditions, that is, Assumptions 2.2(ii) and 3.1.2; (9) is by $\eta_d(r,x,q_d(r,x,u)) = u$ for each d, r, x, and u; (10) and (11) are by the formula of the conditional probability; and (12) is by Assumption 3.1.1. □

Proof of Remark 3.2.1. Notice that $p(r_0^+,x) = p(r_0^-,x) = p(r_0,x)$ implies that

$$\det\left(\Pi'(y_1,y_0,x)\right) = f_{Y|DRX}\left(y_1,1,r_0^+,x\right)p\left(r_0^+,x\right)f_{Y|DRX}\left(y_0,0,r_0^-,x\right)\left(1-p\left(r_0^-,x\right)\right)$$
$$-f_{Y|DRX}\left(y_0,0,r_0^+,x\right)\left(1-p\left(r_0^+,x\right)\right)f_{Y|DRX}\left(y_1,1,r_0^-,x\right)p\left(r_0^-,x\right)$$
$$=\left[f_{Y|DRX}\left(y_1,1,r_0^+,x\right)f_{Y|DRX}\left(y_0,0,r_0^-,x\right) - f_{Y|DRX}\left(y_0,0,r_0^+,x\right)f_{Y|DRX}\left(y_1,1,r_0^-,x\right)\right]$$
$$\times p(r_0,x)(1-p(r_0,x)).$$

By definition, we have

$$f_{Y|DRX}\left(y_1,1,r_0^+,x\right) = f_{U_1|DRX}\left(\eta_1\left(r_0^+,x,y\right),1,r_0^+,x\right)\frac{\partial}{\partial y}\eta_1\left(r_0^+,x,y\right).$$

Notice that if R and X are given, D is determined by V from equation (3). Thus, according to Assumption 3.1.2, we have that both $f_{U_1|DRX}(u,d,r,x)$ and $\eta_1(r,x,y)$ are continuous in r at r_0, which leads to

$$f_{Y|DRX}\left(y_1,1,r_0^+,x\right)=f_{U_1|D^{(1)}RX}\left(\eta_1\left(r_0,x,y\right),1,r_0,x\right)\frac{\partial}{\partial y}\eta_1\left(r_0,x,y\right),$$

where $D^{(1)}=\rho(1,r_0,x,V)$. By similar reasoning,

$$f_{Y|DRX}\left(y_1,1,r_0^-,x\right)=f_{U_1|D^{(0)}RX}\left(\eta_1\left(r_0,x,y\right),1,r_0,x\right)\frac{\partial}{\partial y}\eta_1\left(r_0,x,y\right).$$

Moreover, notice that $D^{(1)}=D^{(0)}$ if there is no jump. Then

$$f_{Y|DRX}\left(y_1,1,r_0^+,x\right)=f_{Y|DRX}\left(y_1,1,r_0^-,x\right).$$

Similarly,

$$f_{Y|DRX}\left(y_0,0,r_0^+,x\right)=f_{Y|DRX}\left(y_0,0,r_0^-,x\right).$$

Combining these results, we know $\det\left(\Pi'(y_1,y_0,x)\right)=0$, which contradicts the full-rank condition. □

Proof of Lemma 3.2. By the formula of total probability, (3.1.2a) is equivalent to

$$F_{Y|DRX}\left(q_1(x,\tau),1,r_0^+,x\right)p\left(r_0^+,x\right)+F_{Y|DRX}\left(q_0(x,\tau),0,r_0^+,x\right)\left(1-p\left(r_0^+,x\right)\right)=\tau.$$

Similarly, (3.1.2b) is equivalent to

$$F_{Y|DRX}\left(q_1(x,\tau),1,r_0^-,x\right)p\left(r_0^-,x\right)+F_{Y|DRX}\left(q_0(x,\tau),0,r_0^-,x\right)\left(1-p\left(r_0^-,x\right)\right)=\tau.$$

Rearranging terms yields

$$\begin{aligned}&F_{Y|DRX}\big(q_1(x,\tau),1,r_0^+,x\big)p\big(r_0^+,x\big)-F_{Y|DRX}\big(q_1(x,\tau),1,r_0^-,x\big)p\big(r_0^-,x\big)\\&=F_{Y|DRX}\big(q_0(x,\tau),0,r_0^-,x\big)\big(1-p\big(r_0^-,x\big)\big)-F_{Y|DRX}\big(q_0(x,\tau),0,r_0^+,x\big)\big(1-p\big(r_0^+,x\big)\big),\end{aligned}$$

which proves (i).

Substituting $\tau=F_{0|X}(y_0,x)$ in $\widetilde{F}_1\left(q_1\left(x,\tau\right),x\right)=\widetilde{F}_0\left(q_0\left(x,\tau\right),x\right)$ gives

$$\widetilde{F}_1\left(q_1\left(x,F_{0|X}(y_0,x)\right),x\right)=\widetilde{F}_0\left(q_0\left(x,F_{0|X}(y_0,x)\right),x\right)=\widetilde{F}_0\left(y_0,x\right).$$

Taking the derivative of both sides of the above equation with respect to y_0 gives

$$\widetilde{f}_1\left(q_1\left(x,F_{0|X}(y_0,x)\right),x\right)\cdot\left(q_1\circ F_0\right)'\left(y_0,x\right)=\widetilde{f}_0\left(y_0,x\right),$$

where $\widetilde{f}_d$ and $(q_1\circ F_0)'$ are understood as the derivative of $\widetilde{F}_d$ and $(q_1\circ F_0)$ with respect to y_0. Because $q_1(x,\cdot)$ and $F_{0|X}(\cdot,x)$ are strictly increasing for any given x, the above implies that $\widetilde{f}_1\left(q_1\left(x,F_{0|X}(y_0,x)\right),x\right)$ and $\widetilde{f}_0\left(y_0,x\right)$ must be either positive, negative, or zero simultaneously for any given y_0 and x.

Next, we show $\widetilde{f}_1\left(q_1\left(x,F_{0|X}(y_0,x)\right),x\right)\neq 0$ and $\widetilde{f}_0\left(y_0,x\right)\neq 0$ for any given y_0 and x by contradiction. Suppose not; then there exists y_0^* such that $\widetilde{f}_1\left(q_1\left(x,F_{0|X}(y_0^*,x)\right),x\right)=0$ and $\widetilde{f}_0\left(y_0^*,x\right)=0$ for some given x. Direct computation of the derivative of $\widetilde{F}_1(y,x)$ and $\widetilde{F}_0(y,x)$

https://doi.org/10.1017/S026646662300021X Published online by Cambridge University Press

with respect to y gives

$$\widetilde{f}_1(y,x) = f_{Y|DRX}(y,1,r_0^+,x)p(r_0^+,x) - f_{Y|DRX}(y,1,r_0^-,x)p(r_0^-,x)$$

and

$$\widetilde{f}_0(y,x) = f_{Y|DRX}(y,0,r_0^-,x)(1-p(r_0^-,x)) - f_{Y|DRX}(y,0,r_0^+,x)(1-p(r_0^+,x)).$$

Then

$$f_{Y|DRX}(y_0^*,0,r_0^-,x)(1-p(r_0^-,x)) = f_{Y|DRX}(y_0^*,0,r_0^+,x)(1-p(r_0^+,x)),$$

and

$$f_{Y|DRX}(q_1\left(x,F_{0|X}(y_0^*,x)\right),1,r_0^+,x)p(r_0^+,x) = f_{Y|DRX}(q_1\left(x,F_{0|X}(y_0^*,x)\right),1,r_0^-,x)p(r_0^-,x),$$

which implies that

$$\det(\Pi(q_1\left(x,F_{0|X}(y_0^*,x)\right),y_0^*,x)) = 0,$$

and contradicts Assumption 3.2.1(ii). Moreover, by the continuity condition, namely Assumption 3.2.1(i), $\widetilde{f}_1(q_1\left(x,F_{0|X}(y,x)\right),x)$ and $\widetilde{f}_0(y,x)$ must be positive or negative for all $y \in \left\{q_0\left(x,\tau\right), x \in S_X, \tau \in (0,1)\right\}$, which completes the proof of (ii). □

Proof of Theorem 1. Substituting $F_{1|X}(y|x) = \tau$ in Lemma 3.2(i) gives

$$\widetilde{F}_1(y,x) \equiv \widetilde{F}_1\left(q_1\left(x,F_{1|X}(y,x)\right),x\right) = \widetilde{F}_0\left(q_0\left(x,F_{1|X}(y|x)\right),x\right),$$

or equivalently

$$\widetilde{q}_0\left(x,\widetilde{F}_1(y,x)\right) = q_0\left(x,F_{1|X}(y|x)\right).$$

Moreover, substituting $F_{1|X}(y|x) = \tau$ in

$$F_{Y|DRX}\left(q_1(x,\tau),1,r_0^+,x\right)p\left(r_0^+,x\right) + F_{Y|DRX}\left(q_0(x,\tau),0,r_0^+,x\right)\left(1-p\left(r_0^+,x\right)\right) = \tau$$

gives

$$\begin{aligned}
F_{1|X}(y|x) &= F_{Y|DRX}\left(q_1\left(x,F_{1|X}(y|x)\right),1,r_0^+,x\right)p\left(r_0^+,x\right) \\
&\quad + F_{Y|DRX}\left(q_0\left(x,F_{1|X}(y|x)\right),0,r_0^+,x\right)\left(1-p\left(r_0^+\right)\right) \\
&= F_{Y|DRX}\left(y,1,r_0^+,x\right)p\left(r_0^+,x\right) + F_{Y|DRX}\left(q_0\left(x,F_{1|X}(y|x)\right),0,r_0^+,x\right)\left(1-p\left(r_0^+\right)\right) \\
&= F_{Y|DRX}\left(y,1,r_0^+,x\right)p\left(r_0^+,x\right) + F_{Y|DRX}\left(\widetilde{q}_0\left(x,\widetilde{F}_1(y,x)\right),0,r_0^+,x\right)\left(1-p\left(r_0^+\right)\right).
\end{aligned}$$

Similar reasoning applies to deriving the expression of $F_{0|X}(y|x)$.

Without loss of generality, we assume that X is a single continuous variable with compact support S_X to simplify notation. □

https://doi.org/10.1017/S026646662300021X Published online by Cambridge University Press

LEMMA A.1. *Under the same assumptions as Theorem 2, then for any given $x \in S_X$,*

$$\sqrt{nhh_x}\begin{pmatrix} \widehat{m}^{+}\left(\mathbb{1}\left\{Y<y\right\}D\middle|x\right)-m^{+}\left(\mathbb{1}\left\{Y<y\right\}D\middle|x\right) \\ \widehat{m}^{+}\left(\mathbb{1}\left\{Y<y\right\}(1-D)\middle|x\right)-m^{+}\left(\mathbb{1}\left\{Y<y\right\}(1-D)\middle|x\right) \\ \widehat{m}^{-}\left(\mathbb{1}\left\{Y<y\right\}D\middle|x\right)-m^{-}\left(\mathbb{1}\left\{Y<y\right\}D\middle|x\right) \\ \widehat{m}^{-}\left(\mathbb{1}\left\{Y<y\right\}(1-D)\middle|x\right)-m^{-}\left(\mathbb{1}\left\{Y<y\right\}(1-D)\middle|x\right) \end{pmatrix} \to^{d} \begin{pmatrix} \mathbb{Z}_{m_D^+}(y|x) \\ \mathbb{Z}_{m_{1-D}^+}(y|x) \\ \mathbb{Z}_{m_D^-}(y|x) \\ \mathbb{Z}_{m_{1-D}^-}(y|x) \end{pmatrix},$$

where $\mathbb{Z}_{m_D^+}$, $\mathbb{Z}_{m_{1-D}^+}$, $\mathbb{Z}_{m_D^-}$, and $\mathbb{Z}_{m_{1-D}^-}$ are tight zero-mean Gaussian processes in $\ell^\infty(S_Y)^4$.

LEMMA A.2. *Under the same assumptions as Theorem 2, then for any given $x \in S_X$,*

$$\sqrt{nhh_x}\begin{pmatrix} \widehat{F}_{1|X}(y|x)-F_{1|X}(y|x) \\ \widehat{F}_{0|X}(y|x)-F_{0|X}(y|x) \end{pmatrix} \to^{d} \begin{pmatrix} \mathbb{Z}_{F_1}(y|x) \\ \mathbb{Z}_{F_0}(y|x) \end{pmatrix},$$

where $\mathbb{Z}_{F_1}$ and $\mathbb{Z}_{F_0}$ are tight zero-mean Gaussian processes in $\ell^\infty(S_Y)^2$.

LEMMA A.3. *Assume $h/h_r = \gamma^2$ with $0<\gamma<\infty$. Under the same assumptions as Theorem 2, then*

$$\sqrt{nh_r}\begin{pmatrix} \widehat{F}_1(y)-F_1(y) \\ \widehat{F}_0(y)-F_0(y) \end{pmatrix} \to^{d} \begin{pmatrix} \mathbb{Z}_1(y) \\ \mathbb{Z}_0(y) \end{pmatrix},$$

where $\mathbb{Z}_1$ and $\mathbb{Z}_0$ are tight zero-mean Gaussian processes in $\ell^\infty(S_Y)^2$.

Proof of Theorem 2. Using the results established in Lemma A.3 and noting that the quantile operator is Hadamard differentiable, it follows from the functional delta method that

$$\sqrt{nh_r}\begin{pmatrix} \widehat{q}_1(\tau)-q_1(\tau) \\ \widehat{q}_0(\tau)-q_0(\tau) \end{pmatrix} \to^{d} \begin{pmatrix} \mathbb{Z}_{q_1}(\tau) \\ \mathbb{Z}_{q_0}(\tau) \end{pmatrix},$$

where $\mathbb{Z}_{q_1}$ and $\mathbb{Z}_{q_0}$ are tight zero-mean Gaussian processes in $\ell^\infty(\mathcal{T})^2$, and

$$\mathbb{Z}_{q_1}(\tau) = -\mathbb{Z}_{F_1}\big(q_1(\tau)\big)/f_1\big(q_1(\tau)\big),$$
$$\mathbb{Z}_{q_0}(\tau) = -\mathbb{Z}_{F_0}\big(q_0(\tau)\big)/f_0\big(q_0(\tau)\big),$$

with $f_d(\cdot)$ denoting the derivative of $F_d(\cdot)$. Moreover, the covariance function $\Sigma^q(\tau,\widetilde{\tau})$ can be written as, for $j,k \in \{1,2\}$,

$$\Sigma^q_{jk}(\tau,\widetilde{\tau}) = \Sigma^F_{jk}\left(q_{2-j}(\tau),q_{2-k}(\widetilde{\tau})\right)\Big/ f_{2-j}\big(q_{2-j}(\tau)\big)f_{2-k}\big(q_{2-k}(\widetilde{\tau})\big). \qquad \square$$

Proof of Corollary 5.1. The result directly follows from Theorem 2 and the functional delta method. □

From now on, we provide detailed proof of Theorem 3 and Corollary 5.2 accompanied by several preliminary lemmas. Lemmas A.4–A.8 show that the first-stage estimators are uniformly consistent. Lemma A.9 establishes the relationship between convergence in probability in supremum norm and convergence in probability concerning the semi-metric $\mathbb{T}$ induced by the limiting Gaussian process. Specifically, we use this lemma to ensure that

https://doi.org/10.1017/S026646662300021X Published online by Cambridge University Press

the asymptotically equicontinuous process, $\widehat{\nu}_1^{\xi}(y)$ and $\widehat{\nu}_0^{\xi}(y)$, evaluating $\widehat{q}_1(\tau)$ and $\widehat{q}_0(\tau)$ can nicely approximate this process evaluating $q_1(\tau)$ and $q_0(\tau)$, respectively.

LEMMA A.4. *Under the same assumptions as Theorem 3,*

$$\widehat{f}_R(r_0) - f_R(r_0) = o_p^{\upsilon}(1).$$

LEMMA A.5. *Under the same assumptions as Theorem 3, then for $d \in \{0,1\}$,*

$$\sup_{(y,r)\in S_Y\times S_R} |\widehat{p}_d(y,r) - p_d(y,r)| = o_p^{\upsilon}(1).$$

LEMMA A.6. *Under the same assumptions as Theorem 3, for $d \in \{0,1\}$,*

$$\sup_{(y,x)\in S_Y\times S_X} \left|\widehat{f}_{YD|RX}\left(y,d,r_0^+,x\right) - f_{YD|RX}\left(y,d,r_0^+,x\right)\right| = o_p^{\upsilon}(1),$$
$$\sup_{(y,x)\in S_Y\times S_X} \left|\widehat{f}_{YD|RX}\left(y,d,r_0^-,x\right) - f_{YD|RX}\left(y,d,r_0^-,x\right)\right| = o_p^{\upsilon}(1)$$

and

$$\sup_{(y,r,x)\in S_Y\times S_R\times S_X} \left|\widehat{\widetilde{p}}(y,d,r,x)\mathbb{1}\{r>r_0\} - \widetilde{p}(y,d,r,x)\mathbb{1}\{r>r_0\}\right| = o_p^{\upsilon}(1),$$
$$\sup_{(y,r,x)\in S_Y\times S_R\times S_X} \left|\widehat{\widetilde{p}}(y,d,r,x)\mathbb{1}\{r<r_0\} - \widetilde{p}(y,d,r,x)\mathbb{1}\{r<r_0\}\right| = o_p^{\upsilon}(1).$$

LEMMA A.7. *Under the same assumptions as Theorem 3, then (i)*

$$\sup_{(y,x)\in S_Y\times S_X} \left|\widehat{\widetilde{f}}_1(y,x) - \widetilde{f}_1(y,x)\right| = o_p^{\upsilon}(1),$$
$$\sup_{(y,x)\in S_Y\times S_X} \left|\widehat{\widetilde{f}}_0(y,x) - \widetilde{f}_0(y,x)\right| = o_p^{\upsilon}(1),$$

(ii)

$$\sup_{(y,x)\in S_Y\times S_X} \left|\widehat{f}_{1|X}(y|x) - f_{1|X}(y|x)\right| = o_p^{\upsilon}(1),$$
$$\sup_{(y,x)\in S_Y\times S_X} \left|\widehat{f}_{0|X}(y|x) - f_{0|X}(y|x)\right| = o_p^{\upsilon}(1),$$

and (iii)

$$\sup_{(y,x)\in S_Y\times S_X} \left|\widehat{\omega}_0^+(y,x) - \omega_0^+(y,x)\right| = o_p^{\upsilon}(1),$$
$$\sup_{(y,x)\in S_Y\times S_X} \left|\widehat{\omega}_1^-(y,x) - \omega_1^-(y,x)\right| = o_p^{\upsilon}(1).$$

LEMMA A.8. *Under the same assumptions as Theorem 3, then for $d \in \{0,1\}$,*

$$\sup_{y\in S_Y} \left|\widehat{f}_d(y) - f_d(y)\right| = o_p^{\upsilon}(1).$$

https://doi.org/10.1017/S026646662300021X Published online by Cambridge University Press

LEMMA A.9. *Define*

$$\mathbb{T}(t_1(\tau), t_2(\tau)) = \lim_{n\to\infty}\left(\sum_{i=1}^{n} E\left|f_{ni}\big(t_1(\tau)\big) - f_{ni}\big(t_2(\tau)\big)\right|^2\right)^{1/2},$$

where

$$f_{ni}(y) = \frac{1}{\sqrt{nh}}\widetilde{e}_1'\left(\widetilde{\Gamma} f_R(r_0)\right)^{-1}\mathcal{Q}_1(Z_i, y)K\left(\frac{R_i - r_0}{h}\right)$$

or

$$f_{ni}(y) = \frac{1}{\sqrt{nh}}\widetilde{e}_1'\left(\widetilde{\Gamma} f_R(r_0)\right)^{-1}\mathcal{Q}_0(Z_i, y)K\left(\frac{R_i - r_0}{h}\right).$$

Under the same assumptions as Theorem 3, then

$$\sup_{\tau\in\mathcal{T}}\left|t_1(\tau) - t_2(\tau)\right| \to_{\upsilon}^{p} 0$$

implies

$$\sup_{\tau\in\mathcal{T}}\mathbb{T}(t_1(\tau), t_2(\tau)) \to_{\upsilon}^{p} 0.$$

Proof of Theorem 3. For the first part, we will only show that $\widehat{\nu}_1^{\xi}(y) \rightsquigarrow_{\xi}^{p} \mathbb{Z}_1(y)$, and the proof of the remaining term is similar. First, define

$$\nu_1^{\xi}(y) = \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}\xi_i\widetilde{e}_1'\left(\widetilde{\Gamma} f_R(r_0)\right)^{-1}\mathcal{Q}_1(Z_i, y)K\left(\frac{R_i - r_0}{h}\right).$$

Applying Theorem 2 of Kosorok (2003) (which is also the same as Theorem 11.19 of Kosorok (2008)), we obtain

$$\nu_1^{\xi}(y) \rightsquigarrow_{\xi}^{p} \mathbb{Z}_1(y).$$

According to Lemma 2 in CHS, we can conclude

$$\widehat{\nu}_1^{\xi}(y) \rightsquigarrow_{\xi}^{p} \mathbb{Z}_1(y)$$

if the condition

$$\sup_{y\in S_Y}\left|\widehat{\nu}_1^{\xi}(y) - \nu_1^{\xi}(y)\right| \to_{\upsilon\times\xi}^{p} 0 \qquad \textbf{(A.1)}$$

https://doi.org/10.1017/S026646662300021X Published online by Cambridge University Press

holds. Thus, we need to verify condition (A.1). Note that

$$
\begin{aligned}
&\widehat{\nu}_1^{\xi}(y) - \nu_1^{\xi}(y) \\
=&\frac{1}{f_R(r_0)\widehat{f}_R(r_0)}\sum_{i=1}^{n}\xi_i\frac{\widetilde{e}_1'\left(\widetilde{\Gamma}\right)^{-1}}{\sqrt{nh}}K\left(\frac{R_i-r_0}{h}\right)\left\{\widehat{\mathcal{Q}}_1(Z_i,y)f_R(r_0)-\mathcal{Q}_1(Z_i,y)\widehat{f}_R(r_0)\right\} \\
=&\frac{f_R(r_0)}{f_R^2(r_0)+o_p^{\upsilon\times\xi}(1)}\sum_{i=1}^{n}\xi_i\frac{\widetilde{e}_1'\left(\widetilde{\Gamma}\right)^{-1}}{\sqrt{nh}}K\left(\frac{R_i-r_0}{h}\right)\left\{\widehat{\mathcal{Q}}_1(Z_i,y)-\mathcal{Q}_1(Z_i,y)\right\} \\
&-\frac{o_p^{\upsilon\times\xi}(1)}{f_R^2(r_0)+o_p^{\upsilon\times\xi}(1)}\sum_{i=1}^{n}\xi_i\frac{\widetilde{e}_1'\left(\widetilde{\Gamma}\right)^{-1}}{\sqrt{nh}}K\left(\frac{R_i-r_0}{h}\right)\mathcal{Q}_1(Z_i,y) \\
=&\dagger_1-\dagger_2,
\end{aligned}
$$

where the second equality holds by the fact that $\widehat{f}_R(r_0)-f_R(r_0)=o_p^{\upsilon\times\xi}(1)$ in Lemma A.4. It can be shown following the same procedures in the proof of Lemma A.1 that

$$
\begin{aligned}
&\sum_{i=1}^{n}\xi_i\frac{\widetilde{e}_1'\left(\widetilde{\Gamma}\right)^{-1}}{\sqrt{nh}}K\left(\frac{R_i-r_0}{h}\right)\to^d\mathbb{Z}_1^{\xi}, \\
&\sum_{i=1}^{n}\xi_i\frac{\widetilde{e}_1'\left(\widetilde{\Gamma}\right)^{-1}}{\sqrt{nh}}K\left(\frac{R_i-r_0}{h}\right)\mathcal{Q}_1(Z_i,y)\to^d\mathbb{Z}_1^{\xi}(y),
\end{aligned}
$$

for some zero-mean Gaussian processes in $\mathbb{R}\times\ell^{\infty}(S_Y)$. By Prohorov's theorem, weak convergence implies asymptotic tightness and therefore implies that

$$
\begin{aligned}
&\sum_{i=1}^{n}\xi_i\frac{\widetilde{e}_1'\left(\widetilde{\Gamma}\right)^{-1}}{\sqrt{nh}}K\left(\frac{R_i-r_0}{h}\right)=O_p^{\upsilon\times\xi}(1), \\
&\sum_{i=1}^{n}\xi_i\frac{\widetilde{e}_1'\left(\widetilde{\Gamma}\right)^{-1}}{\sqrt{nh}}K\left(\frac{R_i-r_0}{h}\right)\mathcal{Q}_1(Z_i,y)=O_p^{\upsilon\times\xi}(1)
\end{aligned}
$$

uniformly over y. Thus,

$$
\dagger_2=\frac{o_p^{\upsilon\times\xi}(1)}{f_R^2(r_0)+o_p^{\upsilon\times\xi}(1)}O_p^{\upsilon\times\xi}(1)=o_p^{\upsilon\times\xi}(1)
$$

uniformly over y. We then consider $\dagger_1$. Lemmas A.4–A.7 imply that

$$
\widehat{\mathcal{Q}}_1(Z_i,y)-\mathcal{Q}_1(Z_i,y)=o_p^{\upsilon\times\xi}(1)
$$

uniformly over Z_i and y. This property implies

$$
\begin{aligned}
\dagger_1=&\frac{f_R(r_0)}{f_R^2(r_0)+o_p^{\upsilon\times\xi}(1)}\sum_{i=1}^{n}\xi_i\frac{\widetilde{e}_1'\left(\widetilde{\Gamma}\right)^{-1}}{\sqrt{nh}}K\left(\frac{R_i-r_0}{h}\right)\times o_p^{\upsilon\times\xi}(1) \\
=&\frac{f_R(r_0)}{f_R^2(r_0)+o_p^{\upsilon\times\xi}(1)}O_p^{\upsilon\times\xi}(1)\times o_p^{\upsilon\times\xi}(1)=o_p^{\upsilon\times\xi}(1)
\end{aligned}
$$

https://doi.org/10.1017/S026646662300021X Published online by Cambridge University Press

uniformly over y. Consequently, we conclude that condition (A.1) holds. Applying Lemma 2 in CHS, we have

$$\widehat{\nu}_1^{\xi}(y) \rightsquigarrow_{\xi}^{p} \mathbb{Z}_1(y).$$

Applying similar arguments, we can also derive

$$\widehat{\nu}_0^{\xi}(y) \rightsquigarrow_{\xi}^{p} \mathbb{Z}_0(y).$$

Next, we will show that

$$\begin{pmatrix} \widehat{\mathbb{Z}}_{q_1}(\tau) \\ \widehat{\mathbb{Z}}_{q_0}(\tau) \end{pmatrix} \rightsquigarrow_{\xi}^{p} \begin{pmatrix} \mathbb{Z}_{q_1}(\tau) \\ \mathbb{Z}_{q_1}(\tau) \end{pmatrix}.$$

By the functional delta method for bootstrap (Theorem 2.9 of Kosorok (2008)),

$$\begin{pmatrix} \widetilde{\mathbb{Z}}_{q_1}(\tau) \\ \widetilde{\mathbb{Z}}_{q_0}(\tau) \end{pmatrix} = \begin{pmatrix} -\widehat{\nu}_1^{\xi}\big(q_1(\tau)\big)/f_1\big(q_1(\tau)\big) \\ -\widehat{\nu}_0^{\xi}\big(q_0(\tau)\big)/f_0\big(q_0(\tau)\big) \end{pmatrix} \rightsquigarrow_{\xi}^{p} \begin{pmatrix} \mathbb{Z}_{q_1}(\tau) \\ \mathbb{Z}_{q_0}(\tau) \end{pmatrix}.$$

Moreover, we can conclude that

$$\sup_{\tau\in\mathcal{T}} \big|\widehat{\mathbb{Z}}_{q_1}(\tau) - \widetilde{\mathbb{Z}}_{q_1}(\tau)\big| \to_{\upsilon\times\xi}^{p} 0$$

and

$$\sup_{\tau\in\mathcal{T}} \big|\widehat{\mathbb{Z}}_{q_0}(\tau) - \widetilde{\mathbb{Z}}_{q_0}(\tau)\big| \to_{\upsilon\times\xi}^{p} 0,$$

which are true by the asymptotic $\mathbb{T}$-equicontinuity of $\widehat{\nu}_1^{\xi}$ and $\widehat{\nu}_0^{\xi}$ in Lemma A.9, the uniform consistency of $\widehat{q}_d$ and $\widehat{f}_d$ for $d \in \{0,1\}$ in Theorem 2 and Lemma A.8. Again, by applying Lemma 2 in CHS,

$$\begin{pmatrix} \widehat{\mathbb{Z}}_{q_1}(\tau) \\ \widehat{\mathbb{Z}}_{q_0}(\tau) \end{pmatrix} \rightsquigarrow_{\xi}^{p} \begin{pmatrix} \mathbb{Z}_{q_1}(\tau) \\ \mathbb{Z}_{q_0}(\tau) \end{pmatrix}. \qquad \square$$

Proof of Corollary 5.2. The result follows from Theorem 3 and the functional delta method for the bootstrap. □

Appendix B. Testing for Local RS

Local RS (Assumption 2.3) is a critical identifying assumption of this study. In this section, we propose a simple test for this assumption in the RD context. While a failure to reject RS may justify the use of our method, a rejection implies that, at a minimum, the model should include more control variables if researchers must have RS hold for the dataset using the method in our study.

The intuition underlying our testing procedure is similar to Frandsen and Lefgren (2018), who test for RS using pre-determined covariates in the context of a binary endogenous treatment. Our testing procedure builds on their work and generalizes it to the RD context. The equivalence between an RD design estimator and an IV estimator, using an indicator for exceeding a threshold in the running variable as the instrument, is already well known in the literature (Hahn et al., 2001). Our test procedure exploits this unique feature in the RD design and proposes a weighted 2SLS method to compute the regression-based statistic.

Assume that we can observe a pre-determined variable S in addition to any covariates X required for identification. In other words, S predicts outcomes but is uncorrelated with the treatment status conditional on X. For example, Deshpande (2016) studies the distributional effect of removing a cash welfare program on earnings using the estimator proposed by FFM. In that case, prior earnings are a good candidate for the S variable because it is correlated with potential outcomes (earnings) in the absence of treatment. For clarity, we still suppress X in the following discussion. Our test has the power to detect the violation of the following statement: $\mathbf{H}_0$: *U_1 and U_0 are identically distributed, conditional on $R = r_0$ and $S = s$ for each s in the support of S.*

As $U_d = F_{Y_d|R}\left(q_d(r_0, U_d), r_0\right)$ for $d = 0, 1$, an individual's rank in the untreated and treated distributions conditional on $R = r_0$ is U_0 and U_1, respectively. By normalization, the marginal distributions of U_0 and U_1 conditional on $R = r_0$ are uniform. However, the conditional distributions of U_0 and U_1 may not be uniform if S predicts ranks. Thus, the rank-shifting variable S imposes testable restrictions on observed data: it means that conditional on S, D does not affect the distribution of ranks. Our test directly examines this implication via the following procedure.

Step 1. Estimate the potential outcome CDFs $\widehat{F}_1(y)$ and $\widehat{F}_0(y)$ for $y \in \mathcal{S}_Y$, conditional on $R = r_0$.

Step 2. Construct sample ranks

$$\widehat{U}_i = D_i\widehat{F}_1(Y_i) + (1 - D_i)\widehat{F}_0(Y_i). \tag{B.1}$$

Step 3. Estimate the following linear specification by an appropriate local estimation method (using only the observations in the h-neighborhood of the cutoff):

$$\widehat{U}_i = \gamma_0 + \gamma_1(R_i - r_0)\mathbb{1}\{R \geq r_0\} + \gamma_2(R_i - r_0)\mathbb{1}\{R < r_0\} + \alpha_1 D_i + S_i\alpha_2 + D_iS_i\eta + \varepsilon_i. \tag{B.2}$$

Remark B.1. The above regression (B.2) is flexible to allow testing violations of similarity for any feature of the rank distribution. For example, to test for differences in the conditional expectation, we can estimate $\beta = (\gamma_0, \gamma_1, \gamma_2, \alpha_1, \alpha_2, \eta)'$ by weighted 2SLS with

$$\left(\mathbb{1}\left\{R_i \geq r_0\right\}, \mathbb{1}\left\{R_i \geq r_0\right\}S_i\right)$$

as the excluded instrument for $\left(D_i, D_iS_i\right)$ and weights $K\left(\frac{R-r_0}{h}\right)$. Let

$$P_i = \left(1, \mathbb{1}\left\{R_i \geq r_0\right\}(R_i - r_0), \mathbb{1}\left\{R_i < r_0\right\}(R_i - r_0), D_i, S_i, D_iS_i\right)',$$

$$Z_i = \left(1, \mathbb{1}\left\{R_i \geq r_0\right\}(R_i - r_0), \mathbb{1}\left\{R_i < r_0\right\}(R_i - r_0), \mathbb{1}\left\{R_i \geq r_0\right\}, S_i, \mathbb{1}\left\{R_i \geq r_0\right\}S_i\right)'.$$

The weighted 2SLS estimator of β can be written as

$$\widehat{\beta} = \left[\sum_{i=1}^{n} Z_iP_i'K\left(\frac{R_i - r_0}{h}\right)\right]^{-1}\left[\sum_{i=1}^{n} Z_i\widehat{U}_iK\left(\frac{R_i - r_0}{h}\right)\right]. \tag{B.3}$$

To test for differences in other parts of the rank distribution, one may estimate (B.2) using quantile regression methods for a range of quantile indices, for example, the IQR estimator Guiteras (2008) proposes for the RD context. Hence, the proposed test maintains power against any departure from RS.

Step 4. Perform a Student's t test of $\eta = 0$. Specifically, we compute the test statistic

$$\widehat{\Delta} = nh\widehat{\beta}' e \left(e'\widehat{V}e\right)^{-1} e'\widehat{\beta}, \tag{B.4}$$

where $e = (0, \ldots, 0, 1)'$ and $\widehat{V}$ is the estimated asymptotic variance–covariance matrix of $\widehat{\beta}$.

Remark B.2. In (B.2), the parameters α_1 and α_2 reflect different normalizations of each treatment category, and each subpopulation with $S = s$. Therefore, α_1 and α_2 are not free parameters and are excluded from the test statistic.

Remark B.3. As Frandsen and Lefgren (2018) note, the testing procedure does not require (B.2) to correspond to a correctly specified structural mean or quantile function of U_i's. The least-squares and quantile regression estimators are well known to converge to a population quantity that linearly approximates the underlying conditional mean or quantile functions. Under the null hypothesis of RS, these population quantities will be identical to U_1 and U_0. The η parameter in (B.2), which corresponds to the treatment–control difference in those population quantities, will therefore be zero even under misspecification.

Applying an argument similar to Frandsen and Lefgren's (2018) Theorem 1, we can show $\widehat{\Delta}$ as a quadratic form of an asymptotically normal estimator, which thus converges to a $\chi^2(1)$ random variable. Specifically, the limiting variance–covariance matrix of $\widehat{\beta}$ consists of an adjustment term resulting from the first-step estimators $\widehat{F}_0(y)$ and $\widehat{F}_1(y)$, whose asymptotic properties can be found in Section 5.1. As the asymptotic variance of $\widehat{\beta}$ is complicated owing to the effect of estimating $\widehat{F}_0(y)$ and $\widehat{F}_1(y)$, in practice, we recommend using the nonparametric bootstrap to obtain $\widehat{V}$. As expected, the power of the test grows with the informativeness of the rank shifter S, that is, the degree of correlation between S and the outcome. Moreover, as Frandsen and Lefgren (2018) note, in contexts with an endogenous treatment, the power of the test also grows with the instrument's strength. In other words, the power of our test grows with the size of $\Delta p = \lim_{r \to r_0^+} E\left(D \middle| R = r\right) - \lim_{r \to r_0^-} E\left(D \middle| R = r\right)$.

Appendix C. Further Discussion of the Full-Rank Jacobian Matrix Condition

This section consists of two parts. First, we discuss the implication of the full-rank condition (Assumption 3.2.1(ii)) thoroughly and demonstrate the relationship between the full-rank condition and Assumption 5.6. Second, we provide sufficient and more interpretable conditions for the full-rank condition to hold in practice.

To set the idea, we suppress the dependence of notation on X. Recall that

$$\Pi'(y_1, y_0) = \begin{bmatrix} f_{Y|DR}\left(y_1, 1, r_0^+\right) p\left(r_0^+\right) & f_{Y|DR}\left(y_0, 0, r_0^+\right)\left(1 - p\left(r_0^+\right)\right) \\ f_{Y|DR}\left(y_1, 1, r_0^-\right) p\left(r_0^-\right) & f_{Y|DR}\left(y_0, 0, r_0^-\right)\left(1 - p\left(r_0^-\right)\right) \end{bmatrix}$$

https://doi.org/10.1017/S026646662300021X Published online by Cambridge University Press

and

$$\widetilde{F}_1(y) = F_{Y|DR}\left(y,1,r_0^+\right)p\left(r_0^+\right) - F_{Y|DR}\left(y,1,r_0^-\right)p\left(r_0^-\right),$$
$$\widetilde{F}_0(y) = F_{Y|DR}\left(y,0,r_0^-\right)\left(1-p\left(r_0^-\right)\right) - F_{Y|DR}\left(y,0,r_0^+\right)\left(1-p\left(r_0^+\right)\right).$$

Then, the two following statements are equivalent.

C1. There exists a positive $\delta > 0$ such that $\Pi'(y_1, y_0)$ is of full rank for all (y_1, y_0) in the support of $(Y_1, Y_0)\big|R \in (r_0-\delta, r_0+\delta)$.

C2. For $d \in \{0,1\}$, $\widetilde{F}_d(y)$ admits a nonzero derivative $\widetilde{f}_d(y)$ with respect to y.

We have the sufficiency from the proof of Lemma 3.2 in Appendix A. Now, we prove the necessity. By the proof of Lemma 3.2, we conclude that $\widetilde{f}_1(y)$ and $\widetilde{f}_0(y)$ have the same sign. Consider the case where $\widetilde{f}_1(y)$ and $\widetilde{f}_0(y)$ are both strictly positive. Thus,

$$\widetilde{f}_1(y) = f_{Y|DR}\left(y,1,r_0^+\right)p\left(r_0^+\right) - f_{Y|DR}\left(y,1,r_0^-\right)p\left(r_0^-\right) > 0,$$
$$\widetilde{f}_0(y) = f_{Y|DR}\left(y,0,r_0^-\right)\left(1-p\left(r_0^-\right)\right) - f_{Y|DR}\left(y,0,r_0^+\right)\left(1-p\left(r_0^+\right)\right) > 0.$$

Consequently, we have

$$\begin{aligned}&f_{Y|DR}\left(y,1,r_0^+\right)p\left(r_0^+\right)f_{Y|DR}\left(y,0,r_0^-\right)\left(1-p\left(r_0^-\right)\right)\\&> f_{Y|DR}\left(y,1,r_0^-\right)p\left(r_0^-\right)f_{Y|DR}\left(y,0,r_0^+\right)\left(1-p\left(r_0^+\right)\right),\end{aligned}$$

implying that $\det\left(\Pi'(y_1,y_0)\right) \neq 0$. The case when $\widetilde{f}_1(y)$ and $\widetilde{f}_0(y)$ are both strictly negative is similar and thus omitted.

Let the limit of D as r approaches r_0 from below (above) be denoted by D_0 (D_1). According to FFM, individuals can be categorized into four different types: never takers (N, $D_1 = D_0 = 0$), always takers (A, $D_1 = D_0 = 1$), compliers (C, $D_1 = 1$, $D_0 = 0$), and defiers (F, $D_1 = 0$, $D_0 = 1$). Notice that

$$\begin{aligned}\widetilde{F}_1(y) &= E\left[1\{Y \le y\}D\middle|R = r_0^+\right] - E\left[1\{Y \le y\}D\middle|R = r_0^-\right]\\&= E\left[1\{Y_1 \le y\}D_1\middle|R = r_0^+\right] - E\left[1\{Y_1 \le y\}D_0\middle|R = r_0^-\right]\\&= E\left[1\{Y_1 \le y\}D_1\middle|R = r_0\right] - E\left[1\{Y_1 \le y\}D_0\middle|R = r_0\right]\\&= E\left[1\{Y_1 \le y\}\left(D_1 - D_0\right)\middle|R = r_0\right]\\&= E\left[1\{Y_1 \le y\}\middle|C, R = r_0\right]\pi_C - E\left[1\{Y_1 \le y\}\middle|F, R = r_0\right]\pi_F,\end{aligned} \tag{C.1}$$

where π_C and π_F denote the fractions of compliers and defiers conditional on $R = r_0$, respectively. Similarly,

$$\widetilde{F}_0(y) = E\left[1\{Y_0 \le y\}\middle|C, R = r_0\right]\pi_C - E\left[1\{Y_0 \le y\}\middle|F, R = r_0\right]\pi_F. \tag{C.2}$$

From equations (C.1) and (C.2), for $d \in \{0,1\}$, we know that we can express $\widetilde{F}_d(y)$ as some mixture distribution of compliers and defiers.

https://doi.org/10.1017/S026646662300021X Published online by Cambridge University Press

- Suppose that the monotonicity condition holds; that is, $\pi_F = 0$. Then, $\widetilde{F}_d(y) = F_{Y_d|C,R}(y,r_0)\pi_C$. Moreover, note that the fraction of compliers $\pi_C > 0$ by the discontinuity assumption. Thus, in this case, **D2** is equivalent to

$$f_{Y_d|C,R}(y,r_0) > 0,$$

for $d \in \{0,1\}$. In other words, under the monotonicity assumption, the nonzero derivative assumption and the full-rank condition are equivalent to full support of the potential outcome distributions of compliers.

- However, the monotonicity condition is not always plausible in practice. Let us consider the implication of the full-rank condition when monotonicity fails (in the presence of defiers). In this case, we can follow the idea of De Chaisemartin (2017) to provide some weaker assumptions to show the plausibility of **D2**.

CD. There exists a subpopulation of compliers C_F that satisfies:
(i) $\pi_{C_F} = \pi_F$, where π_{C_F} represents the fraction of C_F conditional on $R = r_0$;
(ii) for $d \in \{0,1\}$, $F_{Y_d|C_F,R}(y,r_0) = F_{Y_d|F,R}(y,r_0)$ for any given y.

CD is satisfied if a subpopulation of compliers accounts for the same percentage of the population as defiers and has the same potential outcome distribution. De Chaisemartin (2017) states that **CD** is weaker than the monotonicity assumption. In the absence of defiers, one can find a zero-probability subset of compliers with the same potential outcome distribution as defiers. If compliers and defiers have the same potential outcome distribution, and then we can randomly choose π_F/π_C of compliers and denote them as C_F.

Let $C_V = C \backslash C_F$ and $\pi_{C_V} = \pi_C - \pi_{C_F}$. If **CD** holds, then we can show that

$$\begin{aligned}
\widetilde{F}_d(y) =& E\left[1\{Y_d \le y\}\middle| C, R = r_0\right]\pi_C - E\left[1\{Y_d \le y\}\middle| F, R = r_0\right]\pi_F \\
=& \left(\frac{\pi_{C_V}}{\pi_C} E\left[1\{Y_d \le y\}\middle| C_V, R = r_0\right] + \frac{\pi_{C_F}}{\pi_C} E\left[1\{Y_d \le y\}\middle| C_F, R = r_0\right]\right)\pi_C \\
& - E\left[1\{Y_d \le y\}\middle| F, R = r_0\right]\pi_F \\
=& (\pi_C - \pi_F) E\left[1\{Y_d \le y\}\middle| C_V, R = r_0\right].
\end{aligned}$$

Again, by the discontinuity condition, $\pi_C - \pi_F > 0$. Thus, under **CD**, the nonzero derivative assumption and the full-rank condition are equivalent to full support of the potential outcome distributions of C_V, which is the subpopulation of compliers.

CD is just an abstract condition and hard to interpret. Furthermore, similar to De Chaisemartin (2017), we propose a sufficient condition for **CD**, which is more interpretable. Let $\pi_{C|Y_d}$ ($\pi_{F|Y_d}$) denote the fraction of compliers (defiers) conditional on Y_d and $R = r_0$, for $d \in \{0,1\}$.

MC. For $d \in \{0,1\}$ and any $y \in S_{Y_d}$, $\pi_{F|Y_d} \le \pi_{C|Y_d}$.

Condition **MC** requires that each subgroup of the population Y_d comprises more compliers than defiers. This condition is automatically satisfied if there are no defiers. See De Chaisemartin (2017) for more discussion. Next, we will show the relationship between these two conditions. We work with the following lemma to prove that **MC** implies **CD**.

LEMMA C.1. *A subpopulation of compliers C_F satisfies* **CD** *if there is a real-valued function g defined on S_{Y_d} such that for $d \in \{0,1\}$*

$$0 \le g(\delta, r_0) \le f_{Y_d|C,R}(\delta, r_0)\pi_C \ \textit{for almost every} \ \delta \in S_{Y_d} \tag{C.3}$$

$$\int_{S_{Y_d}} g(\delta, r_0) d\delta = \pi_F, \tag{C.4}$$

$$\int_{S_{Y_d}} 1\{\delta \le y\} \frac{g(\delta, r_0)}{\pi_F} d\delta = F_{Y_d|F,R}(y, r_0). \tag{C.5}$$

The proof of Lemma C.1 is similar to Lemma B.1(i) of De Chaisemartin (2017); we omit here. According to Lemma D.1, the argument that **MC** implies **CD** can be proved by finding a real-valued function g satisfying (C.3)–(C.5) under **MC**. Let $g_1 = f_{Y_d|F,R} \cdot \pi_F$. Under **MC**, it is easy to verify that g_1 satisfies (C.3)–(C.5), which completes the proof.

To summarize, we demonstrated that the full-rank condition holds if each subgroup of the population Y_d comprises more compliers than defiers for any given r in the neighborhood of r_0, which seems reasonable in practice. To provide a concrete example, let us consider the DGP, namely,

$$Y_1 = R + \omega U_1, \quad Y_0 = R + U_0,$$

and

$$D = \mathbb{1}\{R > 0\} \cdot \mathbb{1}\{V_1 > 0\} + \mathbb{1}\{R \le 0\} \cdot \mathbb{1}\{V_0 > 1\}.$$

By assumption, $(V_1, V_0)|U_d \sim^d (V_0, V_1)|U_d$ for $d \in \{0,1\}$; thus,

$$\begin{aligned}\pi_{F|Y_d=y,X=x} &= P\left\{V_1 \le 0, V_0 > 1 \middle| Y_d = y, X = x, R = 0\right\} \\ &= P\left\{V_1 \le 0, V_0 > 1 \middle| U_d = \frac{y-x}{d(\omega-1)+1}\right\}\end{aligned}$$

and

$$\begin{aligned}\pi_{C|Y_d=y,X=x} &= P\left\{V_1 > 0, V_0 \le 1 \middle| Y_d = y, X = x, R = 0\right\} \\ &= P\left\{V_1 > 0, V_0 \le 1 \middle| U_d = \frac{y-x}{d(\omega-1)+1}\right\} \\ &= P\left\{V_1 \le 1, V_0 > 0 \middle| U_d = \frac{y-x}{d(\omega-1)+1}\right\},\end{aligned}$$

which satisfies **MC**.

REFERENCES

Abadie, A., J.D. Angrist, & G. Imbens (2002) Instrumental variables estimates of the effect of subsidized training on the quantiles of trainee earnings. *Econometrica* 70(1), 91–117.

Anderson, M.L. (2008) Multiple inference and gender differences in the effects of early intervention: A reevaluation of the Abecedarian, Perry preschool, and early training projects. *Journal of the American Statistical Association* 103(484), 1481–1495.

Arai, Y. & H. Ichimura (2016) Optimal bandwidth selection for the fuzzy regression discontinuity estimator. *Economics Letters* 141, 103–106.

Arai, Y. & H. Ichimura (2018) Simultaneous selection of optimal bandwidths for the sharp regression discontinuity estimator. *Quantitative Economics* 9(1), 441–482.

Asher, S. & P. Novosad (2020) Rural roads and local economic development. *American Economic Review* 110(3), 797–823.

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

Calonico, S., M.D. Cattaneo, & M.H. Farrell (2016) Coverage Error Optimal Confidence Intervals for Regression Discontinuity Designs. University of Michigan, Working paper.

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

Calonico, S., M.D. Cattaneo, & R. Titiunik (2014) Robust nonparametric confidence intervals for regression discontinuity designs. *Econometrica* 82(6), 2295–2326.

Chernozhukov, V., D. Chetverikov, & K. Kato (2014) Gaussian approximation of suprema of empirical processes. *Annals of Statistics* 42(4), 1564–1597.

Chernozhukov, V., I. Fernandez-Val, & A. Galichon (2010) Quantile and probability curves without crossing. *Econometrica* 78(3), 1093–1125.

Chernozhukov, V., I. Fernandez-Val, & B. Melly (2013) Inference on counterfactual distributions. *Econometrica* 81(6), 2205–2268.

Chernozhukov, V. & C. Hansen (2005) An IV model of quantile treatment effects. *Econometrica* 73(1), 245–261.

Chernozhukov, V. & C. Hansen (2006) Instrumental quantile regression inference for structural and treatment effect models. *Journal of Econometrics* 132(2), 491–525.

Chernozhukov, V. & C. Hansen (2013) Quantile models with endogeneity. *Annual Review of Economics* 5(1), 57–81.

Chernozhukov, V., C. Hansen, & K. Wuthrich (2017) Instrumental variable quantile regression. In X. He, R. Koenker, & L. Peng (eds.), *Chernozhukov, V. Handbook of Quantile Regression*, pp. 119–143. CRC Chapman-Hall.

Chiang, H.D., Y.C. Hsu, & Y. Sasaki (2019) Robust uniform inference for quantile treatment effects in regression discontinuity designs. *Journal of Econometrics* 211(2), 589–618.

De Chaisemartin, C. (2017) Tolerating defiance? Local average treatment effects without monotonicity. *Quantitative Economics* 8(2), 367–396.

Deshpande, M. (2016) Does welfare inhibit success? The long-term effects of removing low-income youth from the disability rolls. *American Economic Review* 106(11), 3300–3330.

Fernando, A.N. (2018) Shackled to the Soil: The Long-Term Effects of Inherited Land on Labor Mobility and Consumption. Harvard University, Working paper.

Frandsen, B.R., M. FroLich, & B. Melly (2012) Quantile treatment effects in the regression discontinuity design. *Journal of Econometrics* 168(3638), 382–395.

Frandsen, B.R. & L.J. Lefgren (2018) Testing rank similarity. *Review of Economics and Statistics* 100(1), 86–91.

Frölich, M. & M. Huber (2019) Including covariates in the regression discontinuity design. *Journal of Business and Economic Statistics* 37(4), 736–748.

Guiteras, R. (2008) *Estimating Quantile Treatment Effects in a Regression Discontinuity Design*. MIT Press.

Hahn, J., P. Todd, & W. van der Klaauw (2001) Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica* 69(1), 201–209.

Imbens, G. & J.D. Angrist (1994) Identification and estimation of local average treatment effects. *Econometrica* 62(2), 467–475.

Imbens, G. & K. Kalyanaraman (2012) Optimal bandwidth choice for the regression discontinuity estimator. *Review of Economic Studies* 79(3), 933–959.

Imbens, G.W. & T. Lemieux (2008) Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 142(2), 615–635.

Jacob, B.A. & L. Lefgren (2004) Remedial education and student achievement: A regression discontinuity analysis. *Review of Economics and Statistics* 86(1), 226–244.

Kosorok, M.R. (2003) Bootstraps of sums of independent but not identically distributed stochastic processes. *Journal of Multivariate Analysis* 84(2), 299–318.

Kosorok, M.R. (2008) *Introduction to Empirical Processes and Semiparametric Inference*. Springer.

Lee, D.S. (2008) Randomized experiments from non-random selection in US house elections. *Journal of Econometrics* 142(2), 675–697.

Lee, D.S. & T. Lemieux (2010) Regression discontinuity designs in economics. *Journal of Economic Literature* 48(2), 281–355.

Li, H., X. Shi, & B. Wu (2016) The retirement consumption puzzle revisited: Evidence from the mandatory retirement policy in China. *Journal of Comparative Economics* 44(3), 623–637.

Li, Q. & J.S. Racine (2007) *Nonparametric Econometrics: Theory and Practice*. Princeton University Press.

Mason, D.M. (2004) A uniform functional law of the logarithm for the local empirical process. *Annals of Probability* 32(2), 1391–1418.

Thistlethwaite, D.L. & D.T. Campbell (1960) Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology* 51(6), 309.

van der Klaauw, W. (2008) Regression-discontinuity analysis: A survey of recent developments in economics. *Labour* 22(2), 219–245.

van der Vaart, A. (1998) *Asymptotic Statistics*. Cambridge University Press.

Wuthrich, K. (2019) A closed-form estimator for quantile treatment effects with endogeneity. *Journal of Econometrics* 210(2), 219–235.