

Nonlinear and Nonseparable Structural Functions in Fuzzy Regression Discontinuity Designs*

Haitian Xie[†]

October 12, 2022

[Latest Version](#)

Abstract

Many empirical examples of regression discontinuity (RD) designs concern a continuous treatment variable, but the theoretical aspects of such models are less studied. This study examines the identification and estimation of the structural function in fuzzy RD designs with a continuous treatment variable. The structural function fully describes the causal impact of the treatment on the outcome. We show that the nonlinear and nonseparable structural function can be nonparametrically identified at the RD cutoff under shape restrictions, including monotonicity and smoothness conditions. Based on the nonparametric identification equation, we propose a three-step semiparametric estimation procedure and establish the asymptotic normality of the estimator. The semiparametric estimator achieves the same convergence rate as in the case of a binary treatment variable. As an application of the method, we estimate the causal effect of sleep time on health status by using the discontinuity in natural light timing at time zone boundaries.

Keywords: Causal Inference, Continuous Treatment, Dual Monotonicity, Nonparametric Identification, Semiparametric Estimation, Asymptotic Normality.

*The author is indebted to his advisors Graham Elliott and Yixiao Sun for their constant support on this paper. For helpful comments, the author also thanks Wei-Lin Chen, Gordon Dahl, Xinwei Ma, Kaspar Wüthrich, and participants at UC San Diego econometrics seminar, California econometrics conference, EcoSta, AMES China meeting, SETA, AMES East and South-East meeting, and YES.

[†]Department of Economics, University of California, San Diego. Address: 9500 Gilman Dr. La Jolla, CA 92093. Email: hax082@ucsd.edu.

1 Introduction

The *regression discontinuity* (RD) design is one of the most important approaches to causal inference in non-experimental settings. In an RD design, the researcher is interested in the effect of a treatment T on some outcome Y . The basic idea is that there is an observed running variable R (also called score or index or forcing variable) such that the treatment varies discontinuously when the running variable crosses some cutoff (also called threshold) value \bar{r} . By utilizing this discontinuity, the researcher has the power to identify and estimate the causal impact of interest.

Most theoretical studies of the RD design assume that the treatment is a binary intervention. However, in empirical studies of RD design, researchers may be interested in a continuous treatment that takes value inside an interval. Such examples include sleep time (Giuntella and Mazzonna, 2019), air pollution level (Chen et al., 2013; Ebenstein et al., 2017), and medical expense for infants (Almond et al., 2010; Barreca et al., 2011). See Table 1 in Section 1.1 for more details. This study aims to provide methods for examining the causal effect of a continuous treatment variable in an RD setting.

With a binary treatment, the sharp design is the case where the treatment can be predicted with certainty by whether or not a running variable R is above or below some cutoff (Figure 1(a)). The fuzzy design refers to the case where the treatment probability is a function of the running variable R and changes discretely at the cutoff (Figure 1(b)). Regression discontinuity methods exploit this predicted variation in the treatment above and below the cutoff to identify the effect of the treatment.

With a continuous treatment, where all observations are treated to some extent, the distribution of the treatment is a function of the running variable R , and the distribution of treatments shifts discretely at the cutoff. It is this variation, as in the binary treatment case, that allows the identification of the effect of the treatment at the cutoff.

The essence of the idea is shown in Figure 2. For any R , we have a distribution of the treatment T . Suppose that T is increasing in R , as illustrated in Figure 2(a). As R increases, the level of treatment becomes larger, as shown by the increasing quantiles of T . At $R = \bar{r}$, the distribution of treatment levels increases discretely, as shown by the discontinuous jump in the quantiles. This is the discontinuity that will be exploited

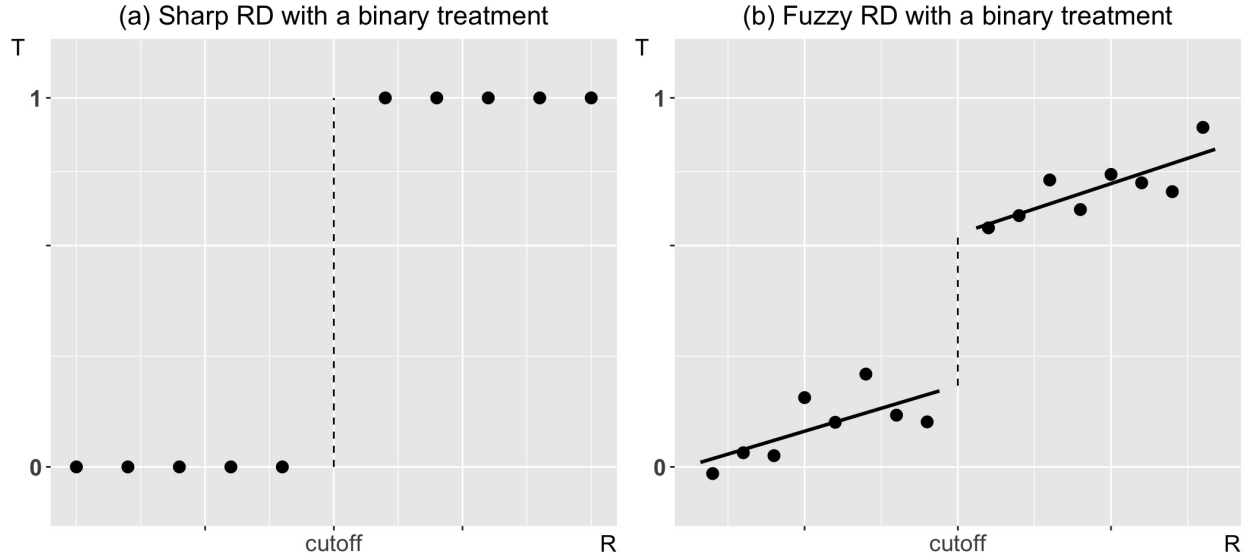


Figure 1: Demonstration of RD designs with a binary treatment.

Graph (a) demonstrates the sharp RD design using a raw scatter plot. Graph (b) demonstrates the fuzzy RD design using a binscatter plot, where each dot represents the treatment probability in the respective bin.

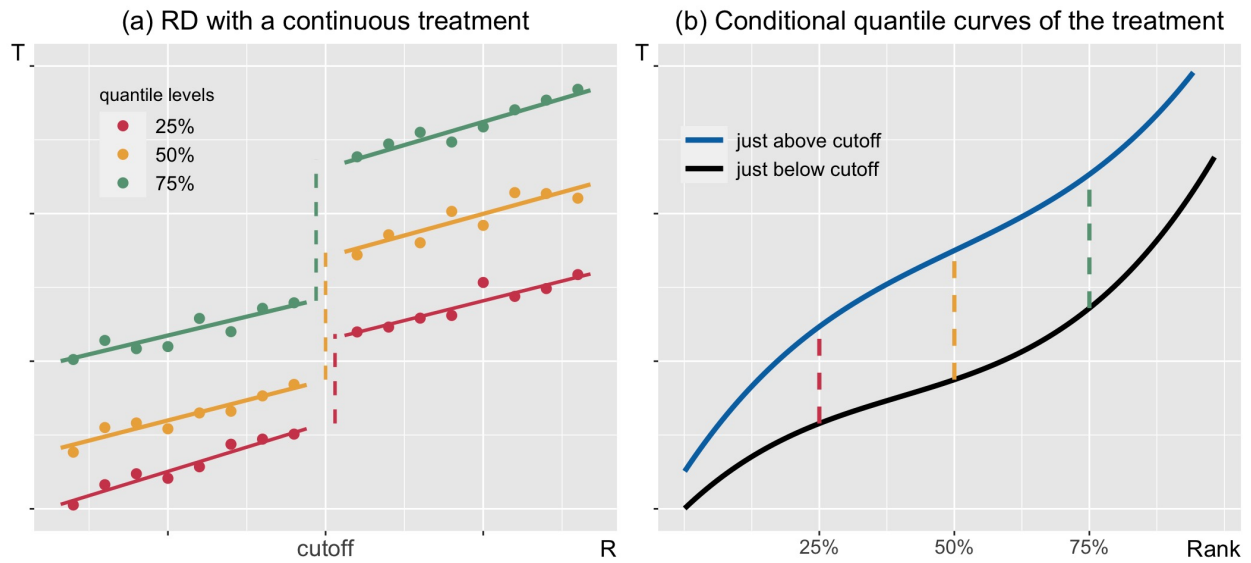


Figure 2: Demonstration of RD designs with a continuous treatment.

Graph (a) demonstrates the regression discontinuities of a continuous treatment variable at different quantile levels. The plot is a binscatter plot, where each dot represents the corresponding quantile treatment level in the respective bin. Different quantile regressions bring different discontinuities. Graph (b) plots the conditional quantile curve of the treatment from just below and just above the cutoff. The horizontal axis specifies which quantile level we are looking at. The entire difference between these two curves constitutes the content of RD of a continuous treatment variable. If there is no regression discontinuity, then the two quantile curves would completely overlap. Notice that we use the same color to denote the corresponding jumps between the two plots.

for estimating the effect of the treatment on the outcome variable. Consider now the distribution of treatments just below and just above the cutoff. Figure 2(b) shows that for R just below the cutoff, the distribution of treatment is below that of the distribution of treatment immediately above the cutoff. Thus the treatment levels differ, and we can separate levels of treatment when there is a small variation in R around this cutoff point.

Consider our empirical study for concreteness: we study the causal impact of sleep time T on health status Y by exploiting the discontinuity in the timing of natural light at time zone boundaries. In this empirical study, the running variable R is the distance to the time zone boundary, and the cutoff \bar{r} is at the time zone boundary. The causal identification is based on the following fact: Individuals living on the late sunset side of the time zone boundary tend to go to bed at a later time, while in the morning, everyone gets up and goes to work at 8 am. Consequently, there is an exogenous variation in sleep time across the time zone boundary. Figure 3(a) shows the histogram of sleep time based on the American Time Use Survey (ATUS) and demonstrates that sleep time is indeed a continuous treatment variable. Figure 3(b) shows the nonparametric estimates of the conditional quantiles of sleep time. We can see that the distribution of the sleep time for individuals living on the early sunset side first-order stochastically dominates the distribution on the late sunset side, which is evidence supporting the above identification strategy. This RD design based on the time zone system is first proposed by (Giuntella and Mazzonna, 2019). In Section 4, we present the novel empirical results based on the econometric methods developed in this paper.

Our paper aims to utilize the continuum of regression discontinuities presented in Figure 2(b) to identify the causal effect of interest. To fully express the causal effect of the treatment T on the outcome Y , we introduce the structural function

$$Y = g^*(T, R, \varepsilon), \quad (1)$$

where ε contains unobserved causal factors (for easy reference, ε will be called the error term hereafter). The structural function g^* specifies how the treatment T determines the outcome Y together with the running variable R and error term ε . More specifically, for any two treatment levels t_1 and t_2 , we can express the individual treatment effect in terms

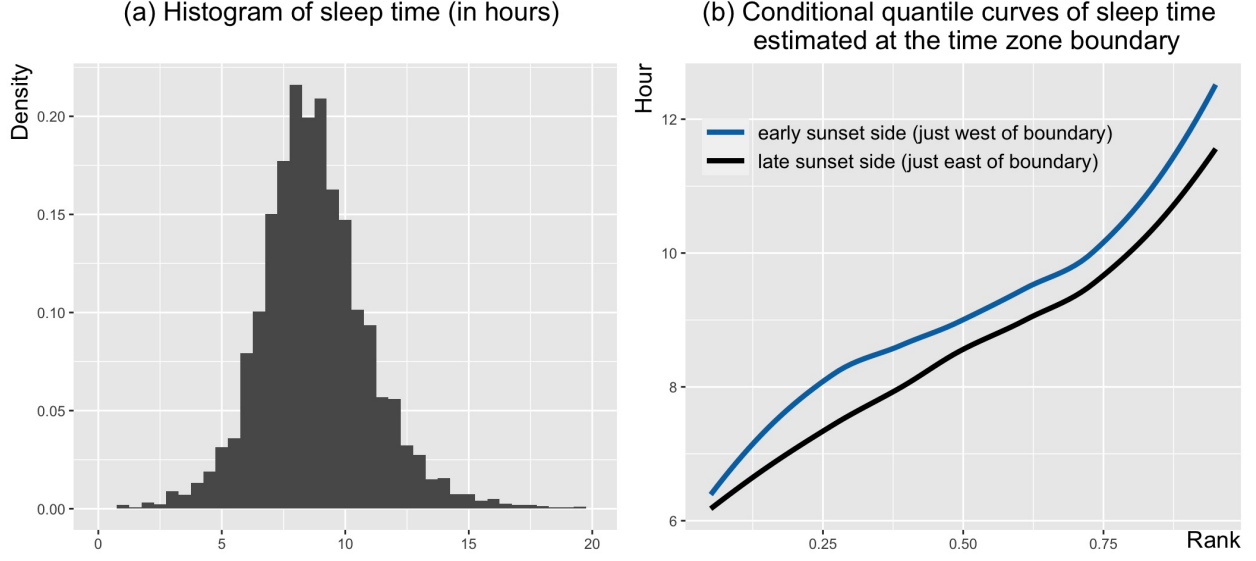


Figure 3: Empirical illustration of RD designs with a continuous treatment.

Graph (a) shows the histogram of sleep time. Evidently, this variable is better modeled as continuous rather than discrete. Graph (b) shows the estimated conditional quantile curves of sleep time given that the geographical location is just west and east of the time zone boundary. The nonparametric estimator used here is the local constant quantile regression. The RD is clearly observed as the blue curve first-order stochastically dominates the black curve, a similar situation as demonstrated in Figure 2(b).

of the structural function: $g^*(t_2, \bar{r}, \varepsilon) - g^*(t_1, \bar{r}, \varepsilon)$. This paper treats the structural function as the causal object of interest.

When the treatment is binary, the information contained in the structural function can be reduced to a scalar treatment effect

$$g^*(1, \bar{r}, \varepsilon) - g^*(0, \bar{r}, \varepsilon),$$

which is the difference in the outcome when the treatment is manipulated from 0 to 1. However, when the treatment is continuous, there are infinite treatment levels to consider. In that case, the structural function is an infinite-dimensional object and is much harder to identify.

In practice, empirical studies often use the *two-stage least squares* (TSLS) method to estimate the following *Wald ratio* around the cutoff:

$$\text{Wald ratio} = \frac{\lim_{r \uparrow \bar{r}} \mathbb{E}[Y|R = r] - \lim_{r \downarrow \bar{r}} \mathbb{E}[Y|R = r]}{\lim_{r \uparrow \bar{r}} \mathbb{E}[T|R = r] - \lim_{r \downarrow \bar{r}} \mathbb{E}[T|R = r]},$$

There are two motivations behind this procedure. First, in the binary treatment case, the Wald ratio would identify the treatment effect.¹ Second, in the continuous treatment case, if the structural function is linear and separable in the treatment, that is, if the structural function can be decomposed as

$$g^*(T, R, \varepsilon) = \beta T + \tilde{g}(R, \varepsilon),$$

then the Wald ratio would identify the slope coefficient β of the treatment.² However, the Wald ratio cannot identify the structural function in general because the structural function is infinite-dimensional while the Wald ratio only provides one-dimensional summary information. The Wald ratio is a weighted average effect of the treatment T on the outcome Y at the cutoff $R = \bar{r}$. Important information in the original structural function, for example, the optimal level of treatment, is lost when the infinite-dimensional function is condensed into a scalar weighted average.

The general identification of the structural function in RD designs remains an unsolved issue. It is desirable to know whether the structural function (at the cutoff) can be identified without the aforementioned linearity and separability conditions. This issue is a practical concern. For instance, in the time zone example, there are reasons for one to believe that the structural function is nonlinear and nonseparable in the treatment.³ Furthermore, a nonlinear structural function is required if the researcher wants to find the optimal sleep time for policy recommendations. A linear structural function would either recommend sleeping 0 hours or 24 hours every day, which is unsuitable for this scenario.

From the theoretical perspective, it is desirable to achieve identification in the non-parametric sense and avoid functional form restrictions such as linearity and separability. These functional form restrictions are usually not supported by any economic theory. As an advantage, the more flexible specification encompassing nonlinearity and nonsepara-

¹As shown in [Hahn et al. \(2001\)](#), the Wald ratio identifies the average treatment effect in the sharp design and the local average treatment effect (for the compliers) in the fuzzy design.

²Such a linear specification of RD design with a continuous treatment can be found in Section 3.4.2 of [Lee and Lemieux \(2010\)](#).

³The nonlinearity can be due to the fact that both undersleeping and oversleeping are harmful to health. The nonseparability can be due to the effect heterogeneity caused by unobserved eating habits, which affect both sleep time and health.

bility allows the treatment effect to be heterogeneous across different levels of the treatment and outcome.

The current study aims to precisely tackle the identification and estimation of the possibly nonlinear and nonseparable structural function. The nonparametric identification result is established based on shape restrictions, including monotonicity and smoothness conditions. The idea behind the identification result is that we are using the infinite set of regression discontinuities in Figure 2(b) to identify the infinite-dimensional structural function. The monotonicity condition restricts the structural function g^* to be strictly increasing in the error term ε . This condition requires the error term to be one-dimensional, which is the potential restriction of the model. However, this condition is common in the nonparametric identification literature (e.g., [Matzkin, 2003](#)) and is satisfied by most, if not all, parametric models used in practice. The smoothness and other regularity conditions imposed in this paper are common in the RD literature.

We develop a three-step semiparametric estimation procedure based on the nonparametric identification result. The structural function is parametrized while nonlinearity and nonseparability are maintained. One such parametrization could be

$$g_\gamma(T, \bar{r}, \varepsilon) = \gamma_1 T + \gamma_2 T^2 + \gamma_3 T \varepsilon + \varepsilon,$$

with structural parameter $\gamma = (\gamma_1, \gamma_2, \gamma_3)$. The relationship between the treatment and the running variable is left to be nonparametric. Under appropriate conditions, we show that the semiparametric estimator of the structural parameter γ is $n^{-2/5}$ -consistent and asymptotically normal. This is a formidable task, as we have two continuous random variables, T and R , and the estimation is performed for the local point $R = \bar{r}$. As an interesting finding, the convergence rate of the semiparametric estimator, $n^{-2/5}$, is the same as in the binary treatment case. There is no loss in terms of convergence rate when extending the RD design from the binary treatment case to the continuous case. The faster convergence rate is due to the integral smoothing in the estimation of the criterion function constructed from the identification equation. To understand this phenomenon, one can consider the analogy in regular semiparametric estimation theory, where the first step is nonparametric while the second step recovers the parametric rate.

The rest of the paper is organized as follows. The remaining part of this section discusses the literature. Section 2 introduces the RD model with a continuous treatment and presents the nonparametric identification result. Section 3 proposes the semiparametric estimation procedure and derives its asymptotic properties. Section 4 presents the empirical study and simulation results. The technical proofs for the identification and estimation results are collected in Appendices A and B, respectively.

1.1 Relation to the literature

The RD method is first introduced by [Thistlethwaite and Campbell \(1960\)](#) into the literature. [Hahn et al. \(2001\)](#) establish the theoretical foundation of RD designs by using the potential outcome framework and show that the RD Wald ratio can be interpreted as the *local average treatment effect* (LATE) for compliers local to the cutoff. Early reviews of the RD design can be found in [Imbens and Lemieux \(2008\)](#) and [Lee and Lemieux \(2010\)](#). For more recent reviews, see [Cattaneo and Escanciano \(2017\)](#) and [Cattaneo and Titiunik \(2021\)](#).

There are many empirical papers that study the causal effect of a continuous treatment in an RD design, some of which are given in Table 1. As explained earlier, these studies apply the TSLS method to estimate the Wald ratio. Hence, there is room for potential improvement in these settings by using the semiparametric estimator developed in the current study.

The theoretical literature on RD designs mostly focuses on the standard case of a binary treatment variable. [Caetano et al. \(2020\)](#) study RD designs with a multi-valued discrete treatment variable. The most relevant paper to ours is the recent work by [Dong et al. \(2021\)](#), which studies RD designs specifically with a continuous treatment variable. Under simple conditions, they propose a way to identify and estimate the *Quantile specific LATE* (Q-LATE) defined as:

$$\frac{\lim_{r \uparrow \bar{r}} \mathbb{E}[Y|U = u, R = r] - \lim_{r \downarrow \bar{r}} \mathbb{E}[Y|U = u, R = r]}{\lim_{r \uparrow \bar{r}} \mathbb{E}[T|U = u, R = r] - \lim_{r \downarrow \bar{r}} \mathbb{E}[T|U = u, R = r]},$$

where $U = F_{T|R}(T|R)$ is the conditional rank of T given R with $F_{T|R}$ denoting the conditional distribution of T given R . The Q-LATE parameter is the Wald ratio given a particu-

Context	Outcome Y	Treatment T	Running R	Studies
US time zone system	Health status	Sleep time	Distance to time zone boundary	Giuntella and Mazzonna (2019)
Heating policy in China	Life expectancy	Air pollution	Distance to Huai River	Chen et al. (2013) Ebenstein et al. (2017)
Minimum capital requirement	Bank failue	Capital	Town size	Dong et al. (2021)
Very low birth weight policy	Infant mortality	Medical spending / Hospitalization	Birth weight	Almond et al. (2010) Barreca et al. (2011) Bharadwaj et al. (2013)
Tax distribution in Brazil	Electoral chance of the incumbent	Local government spending	Municipality population	Litschig and Morrison (2010)
Child-related tax benefits	Personal achievements	Family income	Birthdate	Barr et al. (2021) Cole (2021)

Table 1: Selected empirical RD studies with a continuous treatment variable.

lar quantile level of the treatment. This parameter is a weighted average of the derivative of the structural function (Dong et al., 2021, Section 2.2). In certain situations, however, the policy design process may require information beyond the weighted average of the structural function. The contribution of the current paper is to take a different approach and directly tackle the identification and estimation issues of the structural function.

It has become common in the literature to identify a certain weighted average of the structural function as the causal estimand.⁴ The reason for this trend is twofold: the direct identification of the structural function is difficult, and the researchers want their assumptions to be minimal. However, the weighted average only provides summary information on the structural function, which is insufficient in optimal policy designs. In this paper, we make an effort to identify the structural function itself at the expense of making stronger assumptions. In the empirical study in Section 4, we show that the estimated nonlinear structural function can help determine the optimal sleep time while the TSLS estimates cannot.

Since RD design can be interpreted as a local *instrumental variable* (IV) approach, our paper is naturally connected to the large body of nonparametric IV literature. In partic-

⁴One of such examples is the TSLS estimation with a multivalued treatment (Angrist and Imbens, 1995).

ular, the identification result in this paper is related to the literature on instruments with small support: [Torgovitsky \(2015, 2017\)](#), [D’Haultfœuille and Février \(2015\)](#), [Masten and Torgovitsky \(2016\)](#) and [Ishihara \(2021\)](#). These papers examine a model with a discrete IV and a continuous treatment variable. To directly apply the IV identification result to our setting, we would require the running variable R to be a valid instrument: R is (conditionally) independent with ϵ and is excluded from the structural function g^* .

The challenge in the RD setting is that the independence and the exclusion restriction for the running variable R are typically violated. In the time zone example, the location one lives is correlated with the unobserved eating habits (contained in ϵ) and may directly affect the health status Y .⁵ The contribution of our paper is to formally establish the identification of the structural function in the RD setting by utilizing only the discontinuity at the cutoff. In particular, the running variable R is allowed to be correlated with ϵ and included in the structural function as presented in (1). The estimation procedure in our RD design is also more challenging than IV estimation because we focus on a local neighborhood of the cutoff. Unlike in the IV setup, where \sqrt{n} -consistent estimators exist ([Torgovitsky, 2017](#)), the proposed semiparametric estimator in our RD design is shown to be $n^{-2/5}$ -consistent.

The problem studied by this paper is also related to the broad literature on the nonparametric identification of nonseparable structural functions. Relevant papers include [Matzkin \(2003\)](#), [Hoderlein and Mammen \(2007, 2009\)](#), [Sasaki \(2015\)](#), and [Su et al. \(2019\)](#). The identification there relies on the exogeneity of the treatment, which is not assumed in RD designs.

2 RD Design with a continuous treatment

This section describes the RD model with a continuous treatment, explains the assumptions of the model, and discusses the nonparametric identification of the structural function local to the cutoff.

⁵As another example, consider the very low birth weight policy in Table 1. The running variable birth weight is a strong indicator of infants’ health even after fixing a treatment level.

2.1 The model

We study the following causal equation:

$$Y = g^*(T, R, \varepsilon), \quad (2)$$

where Y is the outcome of interest, T is the treatment, and R is the running variable. The scalar variable ε represents unobserved causal factors in the outcome equation. We assume all the random variables are absolutely continuous. The function g^* is the unknown true structural function.

The running variable R partly determines the treatment T by the following treatment choice function:

$$T = \begin{cases} m_0(R, U_0), & R < \bar{r}, \\ m_1(R, U_1), & R \geq \bar{r}, \end{cases} \quad (3)$$

where \bar{r} is the cutoff value, and U_0 and U_1 are scalar variables, representing other factors that are not observable to an econometrician. For easy reference, they will be referred to as the error terms hereafter. The important feature of the RD design is that the treatment varies discontinuously when the running variable crosses the cutoff \bar{r} . The functions m_0 and m_1 represent respectively the treatment choice mechanism when R is below and above the cutoff.

It is important to point out that the variables U_0, U_1, T and R are allowed to be correlated with the error term ε . If we assume ε to be independent of (T, R) , then we can follow [Matzkin \(2003\)](#) or [Hoderlein and Mammen \(2007\)](#) to identify the structural function. If we assume $\varepsilon \perp R$ and R is excluded from m_0, m_1 and g , then we can follow [Torgovitsky \(2015\)](#) to identify the structural function by treating the binary variable $\mathbf{1}\{R \geq \bar{r}\}$ as the instrument.

We make the following assumptions on the model imposed by (2) - (3). Let \mathcal{G} be the set of candidate structural functions such that the true g^* is contained in \mathcal{G} . That is, \mathcal{G} is the infinite-dimensional parameter space where the structural function belongs to. Denote the conditional distribution function by $F_{\cdot|\cdot}(\cdot|\cdot)$, the conditional density function

by $f_{\cdot|\cdot}(\cdot|\cdot)$, and the conditional quantile function by $F_{\cdot|\cdot}^{-1}(\cdot|\cdot)$.

Assumption 1 (Dual Monotonicity).

- (i) Every $g \in \mathcal{G}$ satisfies that for each given $T = t$ and $R = r$, g is strictly increasing in ε .
- (ii) For each given $R = r$, m_0 is strictly increasing in U_0 and m_1 is strictly increasing in U_1 .

Assumption 2 (Smoothness).

- (i) The functions m_0, m_1 and every $g \in \mathcal{G}$ are continuous on their respective domains.
- (ii) The conditional quantile functions $F_{U_0|R}^{-1}(u|r)$ and $F_{U_1|R}^{-1}(u|r)$ are strictly increasing in u and continuous in (u, r) .
- (iii) The conditional distribution functions $F_{\varepsilon|U_0,R}(e|u, r)$ and $F_{\varepsilon|U_1,R}(e|u, r)$ are strictly increasing in e and are continuous in r at \bar{r} .
- (iv) The running variable R is absolutely continuous, and its density is strictly positive around the cutoff \bar{r} .

Assumption 3 (Rank Similarity). $U_0|(\varepsilon, R = \bar{r}^-)$ has the same distribution as $U_1|(\varepsilon, R = \bar{r}^+)$.

That is, for every u in the common support of U_0 and U_1 and every e in the support of ε ,

$$\lim_{r \uparrow \bar{r}} f_{U_0|\varepsilon,R}(u|e, r) = \lim_{r \downarrow \bar{r}} f_{U_1|\varepsilon,R}(u|e, r).$$

Assumption 1 restricts the shape of the structural function and the treatment choice function. Assumption 1(i) has implications on the treatment effect heterogeneity. In particular, it requires the error term ε to be one-dimensional. Such a condition is imposed in [Torgovitsky \(2015\)](#) and [D'Haultfœuille and Février \(2015\)](#) for the IV setting and [Matzkin \(2003\)](#) for the setting with an exogenous treatment. This condition is considered a reasonable sacrifice for deriving a strong result of directly identifying the structural function. Assumption 1(ii) imposes monotonicity on the unobserved heterogeneity in the treatment choice. This condition is similar to Assumption 1 in [Dong et al. \(2021\)](#). In fact, the monotonicity assumption in the treatment choice model is common in the IV literature (e.g., [Imbens and Newey, 2009](#)). From the theoretical perspective, the entirety of Assumption

1 suggests that there is a one-to-one mapping between (Y, T) and (ε, U_0, U_1) for a given value of the running variable R .

Assumption 2 states that except for the discontinuity introduced in (3), everything else is reasonably smooth. In particular, the conditional distribution of ε given (U, R) needs to be smooth with respect to the running variable R at the cutoff \bar{r} . This condition indicates that the discontinuity in the outcome Y is purely generated by the discontinuity in the treatment T .

Assumption 3 is similar to Assumption 3 in Dong et al. (2021). It imposes the rank similarity condition on the potential treatments. In the time zone example, this assumption requires that the probability for sleep time to stay at a certain rank remains the same regardless of whether the individual lives just east or west of the time zone boundary.⁶

The treatment choice functions (m_0, m_1) are not identified. Rather than trying to identify them, it is more convenient to consider a normalization to a quantile representation. By using the monotonicity of m_0 and m_1 in Assumption 1(ii), we define

$$U = \mathbf{1}\{R < \bar{r}\}F_{U_0|R}(U_0|R) + \mathbf{1}\{R \geq \bar{r}\}F_{U_1|R}(U_1|R) = F_{T|R}(T|R), \quad (4)$$

as the conditional rank of T given R .⁷ Then the treatment choice model in (3) can be written as

$$T = h(R, U) = \begin{cases} h_0(R, U), & R < \bar{r}, \\ h_1(R, U), & R \geq \bar{r}, \end{cases}$$

where

$$h_0(r, u) = m_0(r, F_{U_0|R}^{-1}(u|r)) \text{ and } h_1(r, u) = m_1(r, F_{U_1|R}^{-1}(u|r)).$$

By using $[r_0, r_1]$ to denote the support of R , we can write the domains of h_0 and h_1 respectively as $[r_0, \bar{r}] \times [0, 1]$ and $[\bar{r}, r_1] \times [0, 1]$.

⁶Notice that Assumption 3 is different from the rank similarity condition in the IV quantile regression model Chernozhukov and Hansen (2005), where the similarity is imposed on the rank of potential outcomes.

⁷The second equality in Equation (4) is proved in Lemma 1

The following lemma shows that the function h defined above is the conditional quantile function of T given R , and the quantile representation is a valid normalization in the sense that it preserves the monotonicity and smoothness conditions. Consequently, the function h (including both h_0 and h_1) and the conditional rank $U = h^{-1}(R, T)$ are identified from the data, where h^{-1} denotes the inverse of h with respect to the second argument U . In fact, the two functions $h_0(\bar{r}, \cdot)$ and $h_1(\bar{r}, \cdot)$ are the curves shown in Figure 2(b), in which the horizontal axis represents the conditional rank U .

Lemma 1 (Quantile Representation). *The following statements hold under Assumptions 1 - 3:*

- (i) $U \perp R$, $U|R \sim \text{Unif}[0, 1]$, and $\mathbb{P}(T \leq h(R, u)|R) = u, u \in [0, 1]$.
- (ii) For each $R = r$, h_0 and h_1 are strictly increasing in U .
- (iii) The functions h_0 and h_1 are continuous.
- (iv) The conditional distribution function $F_{\varepsilon|U,R}(e|u, r)$ is strictly increasing in e and is continuous in r at \bar{r} , that is,

$$\lim_{r \uparrow \bar{r}} F_{\varepsilon|U,R}(e|u, r) = \lim_{r \downarrow \bar{r}} F_{\varepsilon|U,R}(e|u, r), \text{ for every } (e, u). \quad (5)$$

By construction, U is independent of R , but U and ε are possibly correlated even after conditioning on R . The following assumption states that the support of the unobserved ε does not vary with U or R . This invariance of the support is not strong since it still allows ε to be correlated with U or R in any way.

Assumption 4 (Support Invariance). *Supp($\varepsilon|U = u, R = r$) does not depend on u or r in the neighborhood of \bar{r} . This common support is denoted by \mathcal{E} .*

We use the following notation to denote the left and right limits. Let $F_{Y|T,R}$ be the

conditional distribution function of Y given T and R . We define

$$F_{Y|T,R}^-(y|t,r) = \begin{cases} F_{Y|T,R}(y|t,r), & \text{if } r < \bar{r}, \\ \lim_{r \uparrow \bar{r}} F_{Y|T,R}(y|t,r), & \text{if } r = \bar{r}. \end{cases}$$

$$F_{Y|T,R}^+(y|t,r) = \begin{cases} F_{Y|T,R}(y|t,r), & \text{if } r > \bar{r}, \\ \lim_{r \downarrow \bar{r}} F_{Y|T,R}(y|t,r), & \text{if } r = \bar{r}. \end{cases}$$

The conditional density functions $f_{T|R}^-$ and $f_{T|R}^+$ are analogously defined. These left and right limits exist in view of Assumptions 1 and 2.

2.2 Nonparametric identification

We derive the key identification equation starting from (5), which is presented below for reference:

$$\lim_{r \uparrow \bar{r}} F_{\varepsilon|U,R}(e|u,r) = \lim_{r \downarrow \bar{r}} F_{\varepsilon|U,R}(e|u,r).$$

This equation means that the conditional distribution of ε is smooth at the cutoff of the running variable. Then we recall that Assumption 1 (the dual monotonicity condition), together with Lemma 1, establishes a one-to-one mapping between (ε, U) and (Y, T) . Consequently, we can relate the unobserved distribution of ε to the observed distribution of (Y, T) as follows:

$$\lim_{r \uparrow \bar{r}} F_{\varepsilon|U,R}(e|u,r) = F_{Y|T,R}^-(g^*(h_0(\bar{r}, u), \bar{r}, e)|h_0(\bar{r}, u), \bar{r}),$$

$$\lim_{r \downarrow \bar{r}} F_{\varepsilon|U,R}(e|u,r) = F_{Y|T,R}^+(g^*(h_1(\bar{r}, u), \bar{r}, e)|h_1(\bar{r}, u), \bar{r}).$$

Combing the above equations, we obtain that

$$F_{Y|T,R}^-(g^*(h_0(\bar{r}, u), \bar{r}, e)|h_0(\bar{r}, u), \bar{r}) = F_{Y|T,R}^+(g^*(h_1(\bar{r}, u), \bar{r}, e)|h_1(\bar{r}, u), \bar{r}), \text{ for every } (e, u).$$

Given that we can identify $F_{Y|T,R}^\pm$ and (h_0, h_1) , the above condition imposes a constraint on the structural function $g^*(\cdot, \bar{r}, \cdot)$. This constraint helps us identify $g^*(\cdot, \bar{r}, \cdot)$.

In order to precisely refer to the above condition, we introduce the following definition: A function $g \in \mathcal{G}$ is said to satisfy Condition (6) if for every $e \in \mathcal{E}$ and $u \in [0, 1]$,

$$F_{Y|T,R}^-(g(h_0(\bar{r}, u), \bar{r}, e)|h_0(\bar{r}, u), \bar{r}) = F_{Y|T,R}^+(g(h_1(\bar{r}, u), \bar{r}, e)|h_1(\bar{r}, u), \bar{r}). \quad (6)$$

The previous analysis shows that the true structural function g^* satisfies Condition (6), which is formally presented in the following lemma.

Lemma 2 (Local Control Function). *Under Assumptions 1 - 3, g^* satisfies Condition (6) with both sides of the equation equal to $F_{\varepsilon|U,R}(e|u, \bar{r})$.*

For the true structural function g^* , Condition (6) can also be written as

$$\lim_{r \uparrow \bar{r}} \mathbb{P}(\varepsilon|T = h_0(r, u), R = r) = \lim_{r \downarrow \bar{r}} \mathbb{P}(\varepsilon|T = h_1(r, u), R = r).$$

The above equation leads to another interpretation of Lemma 2: U can serve as a control function local to the cutoff. After fixing the value of U , the variation in the treatment T becomes locally exogenous. This is because given U and R , the treatment T becomes deterministic. The only variation left in T around the cutoff is due to the discontinuity in the treatment choice function. Lemma 2 is related to Lemma 1 in Dong et al. (2021), but unlike in that paper, here we explicitly express the distribution of ε in terms of the observed distribution of (Y, T) and the structural function g^* .

From the IV perspective, we can relate Lemma 2 to two results in the literature. First, Lemma 2 corresponds to Theorem 1 in Imbens and Newey (2009), where they show that the normalized unobserved heterogeneity in the first stage is a valid control variable. Second, Lemma 2 is similar to Theorem 1 in Torgovitsky (2015) in the sense that it provides a (necessary) characterization of the identified set of the structural function.⁸

Lemma 2 can be used to invalid a candidate structural function g in the parameter space \mathcal{G} . Particularly, if

$$F_{Y|U,R}^-(g(h_0(\bar{r}, u), \bar{r}, e)|u, \bar{r}) \neq F_{Y|U,R}^+(g(h_1(\bar{r}, u), \bar{r}, e)|u, \bar{r}), \text{ for some } e \text{ and } u,$$

⁸The identified set can be defined as the subset of \mathcal{G} that contains the functions g that can generate the observed distribution of (Y, T, R) . However, it is rather a detour to formally define such a set because in Section 3 we directly use Condition (6) for estimation.

then g can not be the true structural function. This means that Condition (6) is a necessary condition for the identification of g^* . To make Condition (6) also a sufficient condition that uniquely specifies the true g^* , we further introduce some regularity conditions below.

Assumption 5.

- (i) (Fuzzy RD). The support of $T|R$ from just below and above the cutoff are intervals denoted respectively by $\text{Supp}(h_0(\bar{r}, U)) = [t'_0, t''_0]$ and $\text{Supp}(h_1(\bar{r}, U)) = [t'_1, t''_1]$. The two supports are overlapping: $[t'_0, t''_0] \cap [t'_1, t''_1] \neq \emptyset$.⁹
- (ii) (Strong Discontinuity). Local to the cutoff, the functions h_0 and h_1 intersects and only intersects finitely many times. That is, the following set is nonempty and finite:

$$\{h_0(\bar{r}, u) : h_0(\bar{r}, u) = h_1(\bar{r}, u) \in [t'_0, t''_0] \cap [t'_1, t''_1], u \in [0, 1]\}.$$

Assumption 5 imposes restrictions on the nature of the discontinuity. Assumption 5(i) requires that the RD design is fuzzy in that there are treatment levels that are taken both below and above the cutoff. Assumption 5(ii) imposes restrictions on the strength of the discontinuity. It requires that the conditional quantile functions $h_0(\bar{r}, \cdot)$ and $h_1(\bar{r}, \cdot)$ only intersects finitely many times. The two curves should intersect but not overlap. If the two functions $h_0(\bar{r}, \cdot)$ and $h_1(\bar{r}, \cdot)$ overlaps on some interval, then the structural function is not identified on that interval because there is no exogenous variation in the treatment inside that interval.¹⁰ In the extreme case where the two curves completely overlap, there is no discontinuity. Figure 4 provides two examples and one counterexample of the strong discontinuity condition.

With the above assumptions, we present the main identification result of the paper. The following theorem shows that Condition (6) identifies the true structural function up to a monotone transformation of the error term ε .

Theorem 1 (Nonparametric Identification). *Let Assumptions 1 - 5 hold. If $g \in \mathcal{G}$ satisfies Condition (6), then there exists a continuous and strictly increasing function λ^g such that for*

⁹Infinite intervals are also allowed. For example, $\text{Supp}(h_0(\bar{r}, U))$ can be $(-\infty, t''_0]$, $[t'_0, \infty)$, or \mathbb{R} . We use the notation $[t'_0, t''_0]$ to represent all these cases.

¹⁰In that case, it is possible to partially identify the structural function.

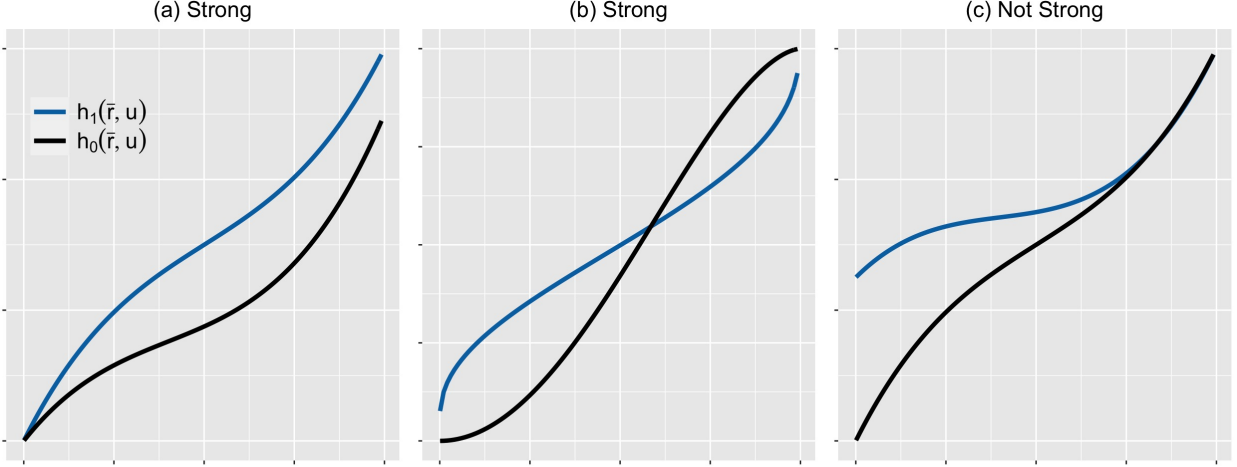


Figure 4: Graphical illustration of the strong discontinuity condition.

Under the strong discontinuity condition stated in Assumption 5(ii), the two curves $h_0(\bar{r}, \cdot)$ and $h_1(\bar{r}, \cdot)$ can either follow the stochastic dominance relationship as in (a) or cross each other for multiple times as in (b). Graph (c) is a violation of the strong discontinuity condition, where the two curves completely overlap on an interval.

every $t \in [t'_0, t''_0] \cup [t'_1, t''_1], e \in \mathcal{E}$, and $u \in [0, 1]$, $g^*(t, \bar{r}, e) = g(t, \bar{r}, \lambda^g(e))$. The specific form of λ^g is given in the proof.

Remark. To prove Theorem 1, we take any $g \in \mathcal{G}$ that satisfies Condition (6). Then we apply the sequencing approach developed in the proof of Theorem 2 in [Torgovitsky \(2015\)](#) to show that g and g^* are related by the transformation λ^g .

Theorem 1 is the best one can achieve in terms of identifying the nonseparable structural function because the error term is unobserved. Any $g \in \mathcal{G}$ that satisfies Condition (6) is equally good as the true g^* . The only difference is that the error term is rescaled by the monotone transformation λ^g . Therefore, such a function g can also be seen as a “version” of g^* .

Inspecting the conditions of Theorem 1, we can see that no independence assumption is needed. However, in some theoretical studies of RD designs, a local independence assumption is imposed on the running variable around the cutoff. For example, Assumption A3(i) in [Hahn et al. \(2001\)](#) requires (ε, U) to be jointly independent of R conditioning on R near \bar{r} . In the binary treatment case, [Dong \(2018a\)](#) shows that this local independence condition is not needed to achieve identification.

Based on the observation made in Lemma 2, we can recover the conditional distribution of ε given U and $R = \bar{r}$. For any $g \in \mathcal{G}$, if g is the true structural function, then the

corresponding conditional distribution of ε is

$$F_{\varepsilon|U,R}^g(e|u, \bar{r}) = F_{Y|T,R}^-(g(h_0(\bar{r}, u), \bar{r}, e)|h_0(\bar{r}, u), \bar{r}) = F_{Y|T,R}^+(g(h_1(\bar{r}, u), \bar{r}, e)|h_1(\bar{r}, u), \bar{r}). \quad (7)$$

In fact, the above conditional distribution $F_{\varepsilon|U,R}^g$ is a transformed version of the true conditional distribution $F_{\varepsilon|U,R}$, where the transformation is the λ^g defined in Theorem 1. This means that the conditional distribution of $\varepsilon|U, R = \bar{r}$ is identified up to the same monotone transformation as the structural function. To eliminate such inconvenience caused by the error term, we can integrate out ε and obtain a unique *conditional average structural function* (CASF):

$$\beta^*(t) = \mathbb{E}[g^*(t, r, \varepsilon)|R = \bar{r}],$$

where the expectation is taken with respect to the true conditional distribution of ε given $R = \bar{r}$. The following corollary summarizes the above analysis.

Corollary 1 (CASF). *Let Assumptions 1 - 5 hold. For any $g^\circ \in \mathcal{G}$ that satisfies Condition (6), let λ^g be the transformation defined in Theorem 1. The following two statements hold true.¹¹*

(i) *For every $e \in \mathcal{E}$ and $u \in [0, 1]$, $F_{\varepsilon|U,R}^{g^\circ}(\lambda^{g^\circ}(e)|u, \bar{r}) = F_{\varepsilon|U,R}(e|u, \bar{r})$.*

(ii) *For any $t \in [t'_0, t''_0] \cup [t'_1, t''_1]$, the CASF $\beta^*(t, \bar{r})$ is uniquely identified as*

$$\beta^*(t) = \int g^\circ(t, \bar{r}, e) dF_{\varepsilon|U,R}^{g^\circ}(e|u, \bar{r}) du.$$

The CASF $\beta^*(t)$ gives the average outcome the policy-maker can achieve when the treatment level for individuals with characteristic $R = \bar{r}$ is set to t . It is worth noting the difference between the CASF and the *local average structural function* (LASF) commonly seen in the LATE literature. The LASF represents the average outcome for the so-called compliers, an unobservable subpopulation. Therefore, the policy-maker cannot assign treatment to the compliers even when the LASF is identified. On the other hand, the identified CASF can directly guide the treatment assignment to the subpopulation with

¹¹Since the conditional distribution of ε given $U, R = \bar{r}$ is identified. We can also identify other structural parameters, including the conditional quantile structural function.

$R = \bar{r}$. The derivative of the CASF is not the causal effect specific to any subpopulation. Following the spirit of, for example, Heckman and Vytlačil (2001), we may call CASF a policy-relevant parameter.

3 Semiparametric estimation

In this section, we consider parametrizations of the structural function that maintain non-linearity and nonseparability. We propose a semiparametric estimation procedure and derive its large-sample properties. The estimator is semiparametric because the structural function is parametrically specified, while the treatment choice model is left nonparametrically specified.

We do not consider a fully nonparametric estimator since such a procedure can be too data-demanding for practical use, which is especially true for the RD design since the estimation is in the local neighborhood of the cutoff.¹²

3.1 Construction of the estimator

Consider the following parametrization of \mathcal{G} local to the cutoff.

Assumption 6 (Semiparametric Specification). *There is a one-to-one mapping from the class $\{(t, e) \mapsto g(t, \bar{r}, e) : g \in \mathcal{G}\}$ of functions to a finite-dimensional parameter space $\Gamma \subset \mathbb{R}^{d_\Gamma}$. We write such parametrization as $\{g_\gamma(\cdot, \bar{r}, \cdot) : \gamma \in \Gamma\}$. Assume this parametric model is correctly specified, that is, there exists $\gamma^* \in \Gamma$ such that $g_{\gamma^*}(\cdot, \bar{r}, \cdot) = g^*(\cdot, \bar{r}, \cdot)$.*

Assumption 7 (Normalization of Γ). *For any $\gamma, \gamma' \in \Gamma$, if there exists a transformation λ such that*

$$g_\gamma(\cdot, \bar{r}, \cdot) = g_{\gamma'}(\cdot, \bar{r}, \lambda(\cdot)),$$

then $\gamma = \gamma'$ and λ is the identity transformation.

¹²From the theoretical perspective, it can be challenging to construct a fully nonparametric estimator. If we follow the sieve approach, for example, we would need to consider a basis of functions that are strictly increasing in one of the arguments to accommodate the monotonicity of the structural function, which is a non-trivial task.

Assumption 7 is a normalization condition that fixes the scale of the error term ε . An example is provided below to illustrate the parametrization of the structural function. One way to achieve such normalization is to have some treatment value \tilde{t} such that $g_\gamma(\tilde{t}, \bar{r}, e) = e$, for all $\gamma \in \Gamma$.

Example 1. Let $\tilde{t} \in [t'_0, t''_0] \cap [t'_1, t''_1]$. We can specify the structural function by

$$g_\gamma(T, \bar{r}, \varepsilon) = \gamma_1(T - \tilde{t}) + \gamma_2(T - \tilde{t})^2 + \gamma_3(T - \tilde{t})\varepsilon + \varepsilon.$$

The parameter $\gamma = (\gamma_1, \gamma_2, \gamma_3)$ is three-dimensional. The function $g_\gamma(T, \bar{r}, \varepsilon)$ is strictly increasing in ε for all γ satisfying $\mathbb{P}(\gamma_3(T - \tilde{t}) + 1 > 0) = 1$. The parametrization satisfies Assumption 7 because, by construction, $g_\gamma(\tilde{t}, \bar{r}, e) = e$ for all values of γ and e .¹³ The model is quadratic in T and nonseparable between T and ε . The effect of T on Y is allowed to be nonlinear and contain unobserved heterogeneity. The distribution of ε is not parametrized and thus can be very general.

The true parameter γ^* in the normalized semiparametric model can be identified as follows. Hereafter, we use $h^* = (h_0^*, h_1^*)$ to signify the true conditional quantile functions and $h = (h_0, h_1)$ a generic pair of conditional quantile functions. Let $w(e, u)$ be a weighting function defined on $\mathbb{R} \times [0, 1]$. Define the criterion function as

$$\|D_{\gamma, h}\|_w = \left(\int_0^1 \int_{\mathbb{R}} |D_{\gamma, h}(e, u)|^2 w(e, u) de du \right)^{1/2}, \quad (8)$$

where $D_{\gamma, h}(e, u)$ is defined to be

$$\int_0^u \left(F_{Y|T, R}^-(g_\gamma(h_0(\bar{r}, v), \bar{r}, e) | h_0(\bar{r}, v), \bar{r}) - F_{Y|T, R}^+(g_\gamma(h_1(\bar{r}, v), \bar{r}, e) | h_1(\bar{r}, v), \bar{r}) \right) dv. \quad (9)$$

This criterion function is based on Equation (6), which by Lemma 2 is a necessary characterization of the identified set. We take an integral form of Condition (6) because it gives a faster convergence rate of the resulting estimator.

Assumption 8 (Weighting function w). *The function w is nonnegative, integrates to one, and is bounded on $\mathbb{R} \times [0, 1]$. The support of w contains $\mathcal{E} \times [0, 1]$. The criterion function $\|D_{\gamma, h}\|_w$ defined in (8) is finite.*

¹³This normalization strategy is presented in Equation (2.5) of Matzkin (2003).

In practice, we may use a weighting function w that is supported on the entire domain $\mathbb{R} \times [0, 1]$ since \mathcal{E} is unknown. For the numerical results in Section 4, we use the normal density function as the weighting w . The following corollary provides the semiparametric identification result, which is based on the nonparametric identification result in Section 2. It shows that the criterion function, when evaluated at the true nuisance parameter value h^* , is uniquely minimized by the true γ^* .

Corollary 2 (Semiparametric Identification). *Let Assumptions 1 - 8 hold. For any $\gamma \in \Gamma$ such that $\gamma \neq \gamma^*$, we have $\|D_{\gamma, h^*}\|_w > \|D_{\gamma^*, h^*}\|_w = 0$. In other words, γ^* is the unique minimizer of $\|D_{\gamma, h^*}\|_w$.*

Assume there is an *independent and identically distributed* (iid) sample $(Y_i, T_i, R_i)_{i=1}^n$ available. We propose an estimation procedure based on the above semiparametric identification result. The idea is that we first estimate the nonparametric components (h_0, h_1) and $(F_{Y|T,R}^-, F_{Y|T,R}^+)$ that appear in the criterion function. Then we construct an empirical version of the criterion function and take its minimizer to be the estimator.

The estimation procedure of γ is more specifically divided into three steps. The first step is to estimate the conditional quantile functions h_0 and h_1 . The second step uses *local linear regression* (LLR) to estimate the conditional distributions $F_{Y|T,R}^-$ and $F_{Y|T,R}^+$. It is standard to use local polynomials in the estimation of RD designs (Porter, 2003; Sun, 2005). The challenge here is that, unlike in the classical RD methods where local polynomial is used to estimate the conditional expectation function of Y given R , here we estimate the conditional distribution of Y .¹⁴ The third step constructs an estimate of the criterion function by replacing the nonparametric nuisance parameters in (9) by their estimated counterparts and then finds the estimate of γ^* by minimizing the estimated criterion function. We describe the detail of the estimation procedure as follows. Denote \mathcal{Y} as the range of the outcome Y .

- **STEP 1.** Choose estimators $\hat{h}_0(\bar{r}, \cdot)$ and $\hat{h}_1(\bar{r}, \cdot)$ of the corresponding conditional quantile processes, $h_0(\bar{r}, \cdot)$ and $h_1(\bar{r}, \cdot)$. Specific constructions are provided in Section 3.3.

¹⁴Local linear estimation of the conditional distribution function can be found in Hansen (2004); Xie (2021).

- **STEP 2.** Choose two bandwidth sequences $b_1 = b_{1n}$ and $b_2 = b_{2n}$ and three kernel functions k_Y , k_T , and k_R . Define $K_Y(y) = \int_{-\infty}^y k_Y(\tilde{y})d\tilde{y}$. For each $y \in \mathcal{Y}$ and $t \in [t'_0, t''_0]$, solve the following minimization problem:

$$\min_{a^-, a_T^-, a_R^-} \sum_{i: R_i < \bar{r}} \left(K_Y \left(\frac{y - Y_i}{b_2} \right) - a^- - a_T^-(T_i - t) - a_R^-(R_i - \bar{r}) \right)^2 \\ \times k_T \left(\frac{T_i - t}{b_1} \right) k_R \left(\frac{R_i - \bar{r}}{b_1} \right).$$

The minimizer \hat{a}^- is the estimate $\hat{F}_{Y|T,R}^-(y|t, \bar{r})$. For each $y \in \mathcal{Y}$ and $t \in [t'_1, t''_1]$, solve the following minimization problem:

$$\min_{a^+, a_T^+, a_R^+} \sum_{i: R_i \geq \bar{r}} \left(K_Y \left(\frac{y - Y_i}{b_2} \right) - a^+ - a_T^+(T_i - t) - a_R^+(R_i - \bar{r}) \right)^2 \\ \times k_T \left(\frac{T_i - t}{b_1} \right) k_R \left(\frac{R_i - \bar{r}}{b_1} \right).$$

The minimizer \hat{a}^+ is the estimate $\hat{F}_{Y|T,R}^+(y|t, \bar{r})$.

- **STEP 3.** Construct the empirical version of the criterion function:

$$\left\| \hat{D}_{\gamma, \hat{h}} \right\|_w = \left(\int_0^1 \int_{\mathbb{R}} |\hat{D}_{\gamma, \hat{h}}(e, u)|^2 w(e, u) de du \right)^{1/2},$$

where $\hat{D}_{\gamma, \hat{h}}(e, u)$ is defined to be

$$\int_0^u \left(\hat{F}_{Y|T,R}^-(g_\gamma(\hat{h}_0(\bar{r}, v), \bar{r}, e) | \hat{h}_0(\bar{r}, v), \bar{r}) - \hat{F}_{Y|T,R}^+(g_\gamma(\hat{h}_1(\bar{r}, v), \bar{r}, e) | \hat{h}_1(\bar{r}, v), \bar{r}) \right) dv.$$

The estimator $\hat{\gamma}$ is any parameter value in Γ that satisfies

$$\left\| \hat{D}_{\hat{\gamma}, h} \right\|_w \leq \inf_{\gamma \in \Gamma} \left\| \hat{D}_{\gamma, h} \right\|_w + O_p(\alpha_n), \quad (10)$$

where $\alpha_n \rightarrow 0$ is a deterministic sequence specified by Equation (B.1) in Appendix B.

3.2 Asymptotic normality

In order to derive the asymptotic distribution of the semiparametric estimator, we consider the regularity assumptions listed below.

Assumption 9 (Distributions of Y , T , and R).

- (i) *The support of T does not vary with R except when crossing the cutoff \bar{r} , i.e., $\text{Supp}(T|R = r) = [t'_0, t''_0]$ for $r < \bar{r}$ and $\text{Supp}(T|R = r) = [t'_1, t''_1]$ for $r > \bar{r}$. The density functions $f_{T,R}^-$ and $f_{T,R}^+$ are bound away from zero.*
- (ii) *The density functions $f_{T,R}^-$ and $f_{T,R}^+$ are twice continuously differentiable, and $\frac{\partial^2}{\partial t^2} f_{T,R}^-(t, \bar{r})$ and $\frac{\partial^2}{\partial t^2} f_{T,R}^+(t, \bar{r})$ are Lipschitz continuous with respect to t .*
- (iii) *The support of Y , \mathcal{Y} , is compact. The conditional distribution functions $F_{Y|T,R}^-$ and $F_{Y|T,R}^+$ are three-times continuously differentiable over $\mathcal{Y} \times [t'_0, t''_0] \times [r_0, \bar{r}]$ and $\mathcal{Y} \times [t'_1, t''_1] \times [\bar{r}, r_1]$, respectively.*

Assumption 10 (Complexity of the Parametric Model). *The parametrization $\{g_\gamma(\cdot, \bar{r}, \cdot) : \gamma \in \Gamma\}$ satisfies the following conditions.*

- (i) *The parameter space Γ is compact.*
- (ii) *The class of functions $\{T \mapsto g_\gamma(T + v, \bar{r}, e) : \gamma \in \Gamma, v \in (-1, 1), e \in \mathcal{E}\}$ is finite-dimensional.*
- (iii) *The function $g_\gamma(t, \bar{r}, e)$ is twice continuously differentiable over $\gamma \in \Gamma, t \in [t'_0, t''_0] \cup [t'_1, t''_1]$, and $e \in \mathcal{E}$.*
- (iv) *The gradient $\nabla_\gamma D_{\gamma^*, h^*}(e, u)$ is a vector of linearly independent functions of (e, u) .*

Assumption 11 (Kernels).

- (i) *The kernel functions k_T and k_R are (1) supported on $[-1, 1]$, (2) strictly greater than zero in the interior of the support, (3) of bounded variation, (4) continuously differentiable on \mathbb{R} .*
- (ii) *The kernel function k_Y is (1) nonnegative and (2) integrable on \mathbb{R} with $\int k_Y(y) dy = 1$ and satisfies (3) $\int y k_Y(y) dy = 0$.*

Assumption 12 (Bandwidth). *The bandwidth b_1 and b_2 satisfy the following conditions:*

- (i) $b_1 \asymp b_2$.¹⁵
- (ii) $(n \log n)b_1^6 = o(1)$.
- (iii) $nb_1^{\frac{13}{3}+\epsilon} \rightarrow \infty$, for some sufficiently small $\epsilon > 0$.

Assumption 13 (First-step Conditional Quantile Estimators). *The estimators \hat{h}_0 and \hat{h}_1 satisfy the following conditions.*

- (i) *Monotonicity and smoothness: for every n sufficiently large, there exist $C > 0$ and deterministic and finite partitions \mathcal{P}_0^n and \mathcal{P}_1^n on $(0, 1)$ such that*

$$\mathbb{P} \left(\hat{h}_0(\bar{r}, \cdot) \notin \mathcal{H}_0(\mathcal{P}_0^n) \right), \mathbb{P} \left(\hat{h}_1(\bar{r}, \cdot) \notin \mathcal{H}_1(\mathcal{P}_1^n) \right) = O(\sqrt{b_1}),$$

where

$\mathcal{H}_0(\mathcal{P}_0^n) = \{ \text{function } h \text{ from } [0, 1] \text{ into } [t'_0, t''_0] : \text{on each element of } \mathcal{P}_0^n, h \text{ is strictly increasing, its inverse } h^{-1} \text{ is three-times continuously differentiable, and } (h^{-1})^{(3)} \text{ is Lipschitz continuous} \},$

and $\mathcal{H}_1(\mathcal{P}_1^n)$ is defined analogously by replacing \mathcal{P}_0^n with \mathcal{P}_1^n .

- (ii) *Uniform Bahadur representation:*

$$\begin{aligned} \hat{h}_0(\bar{r}, u) - h_0^*(\bar{r}, u) &= b_1^2 v_0(u) + O_p(b_1^3) \\ &\quad + \frac{1}{nb_1} \sum_{i=1}^n q_0(T_i, R_i; u) k_{Q,0} \left(\frac{R_i - \bar{r}}{b_1} \right) \mathbf{1}\{R_i < \bar{r}\} + o_p(1/\sqrt{nb_1}), \\ \hat{h}_1(\bar{r}, u) - h_1^*(\bar{r}, u) &= b_1^2 v_1(u) + O_p(b_1^3) \\ &\quad + \frac{1}{nb_1} \sum_{i=1}^n q_1(T_i, R_i; u) k_{Q,1} \left(\frac{R_i - \bar{r}}{b_1} \right) \mathbf{1}\{R_i \geq \bar{r}\} + o_p(1/\sqrt{nb_1}), \end{aligned}$$

uniformly over $u \in (0, 1)$. The functions v_0 and v_1 are bounded. The functions q_0 and q_1 are (1) bounded, (2) centered, that is, $\mathbb{E}[q_0(T, R; u)|T, R] = \mathbb{E}[q_1(T, R; u)|T, R] = 0$, and

¹⁵The notation $b_1 \asymp b_2$ means that there exists $C > 1$ such that $b_1/b_2 \in [1/C, C]$.

(3) does not vary with n . The functions $k_{Q,0}$ and $k_{Q,1}$ are bounded.

(iii) Uniform convergence rate:

$$\begin{aligned}\|\hat{h} - h^*\|_\infty &= \sup_{u \in (0,1)} |\hat{h}_0(\bar{r}, u) - h_0^*(\bar{r}, u)| \vee |\hat{h}_1(\bar{r}, u) - h_1^*(\bar{r}, u)| \\ &= O_p \left(\sqrt{\log n / (nb_1)} + b_1^2 \right).\end{aligned}$$

A brief discussion of the assumptions is in order. Assumption 9 imposes smoothness restrictions on the joint distribution of (Y, T, R) . In the previous section, the identification result only requires continuity of the relevant functions. For estimation, we need higher-order smoothness regarding the distribution functions. Assumption 10 imposes restrictions on the parametric model of the structural function. Part (ii) restricts the complexity of the model. Part (iii) imposes high-order smoothness on the structural function. Part (iv) is similar to Assumption D4 in Torgovitsky (2017) and requires that $\nabla_\gamma D_{\gamma^*, h^*}$ to carry information about each component of the parameter.

Assumption 11 imposes restrictions on the kernel functions k_T , k_R , and k_Y . The differentiability is needed to prove a stochastic equicontinuity condition. Assumption 12 restricts that b_1 and b_2 are of the same asymptotic order, which is slightly faster than $n^{-1/6}$ and slightly slower than $n^{-3/13}$. This assumption is not restrictive and allows for the asymptotic mean squared error (AMSE) optimal bandwidth as well as undersmoothing.

Assumption 13 imposes high-level restrictions on the first-stage nonparametric conditional quantile estimators. Part (i) assumes that the quantile estimators are piece-wise monotonic and smooth with a high probability. Part (ii) and (iii) give the uniform Bahadur representation and the uniform convergence rate, which are fairly standard in the quantile estimation literature. In Section 3.3, we discuss a specific nonparametric quantile estimator that satisfies Assumption 13.

Theorem 2 (Asymptotic Distribution of the Semiparametric Estimator). *Let Assumptions 1 - 13 hold. Then $\|\hat{\gamma} - \gamma^*\|_2 = O_p(b_1^2 + 1/\sqrt{nb_1})$ and*

$$(\sqrt{nb_1}(\Sigma_- + \Sigma_+)^{-1/2})(\Delta(\hat{\gamma} - \gamma^*) - b_1^2(B_- - B_+)) \xrightarrow{d} N(0, \mathbf{I}_{d_\Gamma}),$$

where I_{d_Γ} is the d_Γ -dimensional identity matrix. The exact forms of Δ , B_- , B_+ , Σ_- and Σ_+ are given in Equations (B.9), (B.10), (B.11), (B.12) and (B.13) in Appendix B, respectively.

Remark. The convergence rate of $\hat{\gamma}$ is $b_1^2 + 1 / \sqrt{nb_1}$, which is equal to $n^{-2/5}$ when $b_1 \asymp n^{-1/5}$. This rate is the same as the one obtained in the classical RD design with a binary treatment variable (Hahn et al., 2001). Having a continuous treatment does not slow down the convergence rate of the estimator in this case. This is due to the integral smoothing in the definition of the criterion function $D_{\gamma,h}$ in (9).

Remark. The proof of Theorem 2 follows the general steps of proving asymptotic normality of semiparametric estimators as in, for example, Torgovitsky (2017) and Chen et al. (2003). The main difficulty is that the usual stochastic equicontinuity condition is not sharp because the criterion function is nonparametrically estimated. More specifically, the consistency of $\hat{\gamma}$ by itself is not enough to eliminate the estimation errors asymptotically. To overcome this issue, we first use the empirical process theory to derive a uniform convergence rate for the estimated criterion function, which gives an initial bound on the convergence rate of $\hat{\gamma}$. A sharper stochastic equicontinuity result is then derived based on this initial bound together with more applications of the empirical process theory. This sharper stochastic equicontinuity result helps demonstrate that the usual linearization of the criterion function is valid. See the proof in Appendix B for details.

Once the asymptotic normal distribution of the estimator $\hat{\gamma}$ is established, we can conduct inference for γ^* . Theorem 2 together with the undersmoothing condition that $nb_1^5 = o(1)$ gives that

$$\sqrt{nb_1}(\hat{\gamma} - \gamma^*) \xrightarrow{d} N(0, \Delta^{-1}(\Sigma_- + \Sigma_+)\Delta^{-1}).$$

A linear null hypothesis regarding γ can be written as $H\gamma = \eta$, where $\eta \in \mathbb{R}^{d_\eta}$ and H is a $d_\eta \times d_\gamma$ full-rank matrix. Consider the test statistic

$$nb_1(\hat{\gamma} - \gamma^*)' \left(H\hat{\Delta}^{-1}(\hat{\Sigma}_- + \hat{\Sigma}_+)\hat{\Delta}^{-1}H' \right)^{-1} (\hat{\gamma} - \gamma^*),$$

where $\hat{\Delta}$, $\hat{\Sigma}_-$, and $\hat{\Sigma}_+$ are consistent estimators of Δ , Σ_- , and Σ_+ , respectively. By Slutsky's theorem, the above test statistic converges in distribution to the χ^2 distribution with d_η

degrees of freedom. In Appendix B, we discuss how to construct consistent estimators for Δ , Σ_- , and Σ_+ .

3.3 First-step nonparametric quantile estimators

This section discusses how to construct nonparametric conditional quantile estimators that satisfy Assumption 13. Consider the following two-step estimation procedure introduced by Qu and Yoon (2015). Define $\rho_u(t) = t(u - \mathbf{1}\{t < 0\})$.

- **STEP 1.** Choose a bandwidth sequence $b_3 = b_{3n} = o(1)$ and a kernel function k_{FS} . Partition the unit interval $(0, 1)$ into a grid of equally spaced points $\{u_1, \dots, u_{J_n}\}$, where $J_n / (nb_3)^{1/4} \rightarrow \infty$. Solve the following optimization problem:

$$\min_{\{h_j, h'_j\}_{j=1}^{J_n}} \sum_{j=1}^{J_n} \sum_{i=1}^n \rho_{u_j} \left(T_i - h_j - h'_j(R_i - \bar{r}) \right) k_{FS} \left(\frac{R_i - \bar{r}}{b_3} \right) \mathbf{1}\{R_i < \bar{r}\}. \quad (11)$$

Denote the minimizers by $(\hat{h}_0(\bar{r}, u_1), \dots, \hat{h}_0(\bar{r}, u_{J_n}))$.

- **STEP 2.** Let $u_0 = 0$ and $u_{J_n+1} = 1$. Let $\hat{h}_0(\bar{r}, u_0) = \min_{i: R_i < \bar{r}} T_i$ and $\hat{h}_0(\bar{r}, u_{J_n+1}) = \max_{i: R_i < \bar{r}} T_i$. Linearly interpolate between the estimates to obtain an estimate for the entire quantile process. That is, for any $u \in (u_j, u_{j+1})$, define

$$\hat{h}_0(\bar{r}, u) = \frac{u_{j+1} - u}{u_{j+1} - u_j} \hat{h}_0(\bar{r}, u_j) + \frac{u - u_j}{u_{j+1} - u_j} \hat{h}_0(\bar{r}, u_{j+1}).$$

The estimator $\hat{h}_1(\bar{r}, \cdot)$ can be analogously defined by using the data with $R_i \geq \bar{r}$.¹⁶ We can verify Assumption 13 for the estimator constructed above. Denote

$$\begin{aligned} \Omega_{Q,0} &= \int (1, x)(1, x)' \mathbf{1}\{x < 0\} k_{FS}(x) dx, \\ \Omega_{Q,1} &= \int (1, x)(1, x)' \mathbf{1}\{x \geq 0\} k_{FS}(x) dx. \end{aligned}$$

¹⁶There are three estimators of conditional quantile process in Qu and Yoon (2015). The estimator explained here is their second one, denoted by \hat{a}^* in that paper. Their third estimator imposes a monotonicity constrain to the minimization problem (11).

Proposition 1. *Let Assumptions 2(iv) and 9(i)-(ii) hold. Assume that the third-order derivatives $\frac{\partial^3}{\partial R^3} h_0^*(r, u)$ and $\frac{\partial^3}{\partial R^3} h_1^*(r, u)$ are Lipschitz continuous respectively on $[r_0, \bar{r}] \times [0, 1]$ and $[\bar{r}, r_1] \times [0, 1]$. Assume that the bandwidth $b_3 = cb_1$ for some constant $c > 0$. Assume that the kernel k_{FS} is nonnegative, of bounded variation, compactly supported, having finite first-order derivatives and satisfying*

$$\int k_{FS}(x)dx = 1, \int xk_{FS}(x)dx = 0, \int x^2k_{FS}(x)dx < \infty.$$

Then the estimators $\hat{h}_0(\bar{r}, \cdot)$ and $\hat{h}_1(\bar{r}, \cdot)$ described above satisfy Assumption 13. The specific forms of v_0 and v_1 are

$$\begin{aligned} v_0(u) &= \frac{c^2}{2} \frac{\partial^2}{\partial r^2} h_0^*(\bar{r}, u) \iota' \Omega_{Q,0}^{-1} \int x^2 (1, x)' k_{FS}(x) \mathbf{1}\{x < 0\} dx, \\ v_1(u) &= \frac{c^2}{2} \frac{\partial^2}{\partial r^2} h_1^*(\bar{r}, u) \iota' \Omega_{Q,1}^{-1} \int x^2 (1, x)' k_{FS}(x) \mathbf{1}\{x \geq 0\} dx. \end{aligned}$$

The specific forms of $k_{Q,0}$ and $k_{Q,1}$ are

$$k_{Q,j}(x) = \iota' \Omega_{Q,j}^{-1} (1, x/c)' k_{FS}(x/c) / c, j = 0, 1.$$

The functions q_0 and q_1 are

$$\begin{aligned} q_0(T, R; u) &= (u - \mathbf{1}\{T \leq h_0^*(\bar{r}, u)\}) / (f_R(\bar{r}) f_{T|R}^-(h_0^*(\bar{r}, u) | \bar{r})), \\ q_1(T, R; u) &= (u - \mathbf{1}\{T \leq h_1^*(\bar{r}, u)\}) / (f_R(\bar{r}) f_{T|R}^+(h_1^*(\bar{r}, u) | \bar{r})), \end{aligned}$$

which take the form of an influence function for quantiles.

Other quantile estimation methods are also available. For example, one can consider the generic framework proposed by Chernozhukov et al. (2010) for rearrangement. In particular, they show that the rearrangement of a preliminary estimated quantile process delivers a monotonic estimator that preserves the asymptotic properties. This result gives a different way to generate estimators that satisfy Assumption 13. We can start with an estimator with desired asymptotic properties that give rise to Assumption 13(ii) and (iii), and then apply the rearrangement procedure. The resulting estimator would be

monotonic on the entire domain, and partitioning is unnecessary.

4 Numerical results

This section presents the empirical study and the simulation results. The empirical study shows that the semiparametric estimator is considerably better than the simple TSLS estimator in discovering quantitative information regarding the structural function. The simulation studies show that the semiparametric procedure can accurately estimate the parameters with moderate sample size.¹⁷

4.1 Empirical study

In this empirical study, we examine the causal effect of sleep time on health status by exploiting the discontinuity in the timing of natural light at time zone boundaries. The unit of observation is the individual in the American Time Use Survey (ATUS). The outcome Y is the individual's health status measured by the body-mass index (BMI).¹⁸ The treatment T is the sleep time. The running variable R is the longitudinal distance to the nearest time zone boundary, with cutoff $\bar{r} = 0$ denoting the time zone border. As explained in the introduction, the identification is based on the exogenous variation in the sleep time around the time zone boundary. This exogenous variation is due to the difference in the timing of natural light on each side of the time zone boundary.

Many studies in the medical literature examine the effect of sleep time on overweight issues. See [Beccuti and Pannain \(2011\)](#) and the references therein. These studies typically use survey or laboratory data. This problem is first studied by using the RD design in [Giuntella and Mazzonna \(2019\)](#).¹⁹ The relevant outcome variable they use is a binary indicator of the obesity (or overweight) status indicating whether the BMI is above some threshold. They use the TSLS procedure to estimate the Wald ratio across the time zone boundary. We consider two improvements based on their work. First, we directly use

¹⁷Replication files for the empirical and simulation studies are available from the author upon request.

¹⁸BMI is a person's weight in kilograms divided by the square of height in meters. The Centers for Disease Control and Prevention define overweight as BMI > 25 and obesity as BMI > 30.

¹⁹[Giuntella and Mazzonna \(2019\)](#) study many health and economics-related issues. Here we only mention the relevant ones.

BMI as the outcome variable, providing a more quantitative measure of the health status. Second, we use the proposed semiparametric estimator to estimate a nonlinear structural function. Previous medical studies have provided evidence of the nonlinearity of the structural function. For example, [Hairston et al. \(2010\)](#) show that both undersleeping and oversleeping lead to an increase in BMI while sleeping around 8 hours leads to a more healthy BMI level.

The data for this empirical study is collected from IPUMS CPS ([Flood et al., 2020](#)) and IPUMS ATUS ([Hofferth et al., 2020](#)) during the periods 2006 - 2008 and 2014 - 2016. By linking these datasets, we can locate the county where the individual lives and then use the county's centroid as the location of the individual. We focus on counties near the time zone boundary between the Eastern and Central time zone. The counties are divided into two regions based on their latitude. We estimate the model separately for each region.

The estimated marginal effects of sleep on BMI from the semiparametric estimator and the TSLS estimator are shown in Figure 5. Several interesting findings are observed based on the semiparametric estimates. First, the marginal effects are increasing and increase from negative to positive. This lends some support to the previous argument that the structural function is nonlinear and neither sleeping too little nor too much is preferable. Second, we can determine the optimal (in terms of BMI) sleep time by finding the zero of the marginal effect curve. In both cases, the optimal sleep time is estimated to be between 7 and 8 hours, which also aligns with the findings in previous medical studies. Third, the results from the two regions are similar, meaning that the variation across different latitudes is small.

From Figure 5, we can also see that the TSLS estimates are not capable of demonstrating the above results. First, the TSLS procedure only provides a constant estimate of the marginal effect across all levels of sleep time. This means an extra hour of sleep would lead to the same effect on health regardless of the person's current sleep time, which is inappropriate in this setting. Moreover, we cannot estimate the optimal sleep time based on the linear structural function. Second, the magnitude of the TSLS estimates is small. This is because the TSLS provides a weighted average of the marginal effects across the entire range of sleep time. By averaging the negative and positive effects, the TSLS delivers an

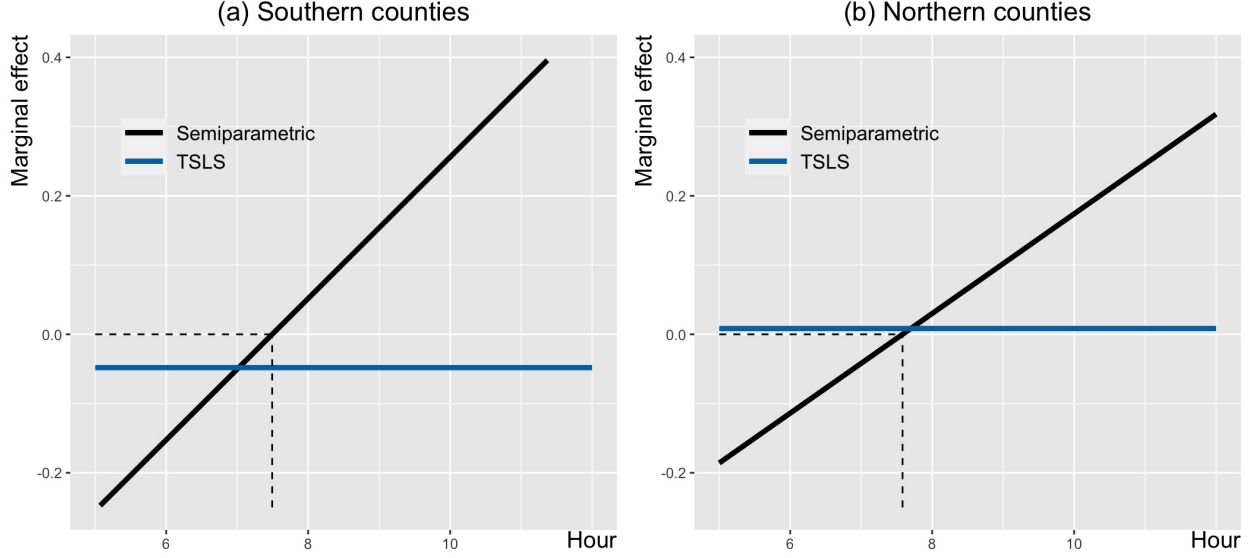


Figure 5: Estimated marginal effects of sleep time on BMI (kg/m^2).

The plots show the estimated marginal effects based on the semiparametric and the TSLS estimator. Graph (a) is computed based on counties with latitude < 37 . Graph (b) is computed based on counties with latitude > 37 . We can see that the marginal effects are increasing, indicating a nonlinear (U-shaped) structural function. The optimal sleep time computed as the zero of the marginal effect curve is between 7 and 8 hours. However, the marginal effect estimated from the TSLS procedure is constant across different sleep times. These estimates are small in magnitude and less informative for the researcher.

estimate attenuated toward zero, which is not informative for the researcher.²⁰

4.2 Simulations

We use simulation studies to investigate the performance of the proposed semiparametric method and compare it with the performance of the TSLS estimator. The data generating process (DGP) for these simulations was chosen to roughly approximate the ATUS data used in the empirical study. Let marginal distributions of U and ε be given by $F_U = \text{Unif}(0, 1)$ and $F_\varepsilon = \text{Beta}(2, 2)$, respectively. Two marginal distributions for R are considered, $\text{Unif}(0, 1)$ and $N(0, 1)$. The joint distribution of (R, ε, U) be characterized by the Gaussian copula with correlation structure $\text{corr}(R, U) = 0$, $\text{corr}(\varepsilon, R) = \rho_R$, and $\text{corr}(\varepsilon, U) = \rho_U$. The treatment choice model is given by $h_0^*(r, u) = r + 2 \sin(\pi u / 2)$ and

²⁰Giuntella and Mazzonna (2019) find a more significant effect of sleep time on obesity. There are two possible reasons: they consider the binary indicator of obesity, and they include more control variables in the regression.

$h_1^*(r, u) = r + 2u^3$. The structural function is given by

$$g_\gamma(T, R, \varepsilon) = \gamma_1(T - 0.5) + \gamma_2(T^2 - 0.5^2) + \gamma_3(T - 0.5)\varepsilon + \varepsilon + R.$$

The true γ^* is taken to be $(1, 1, 1)$.

Further implementation details are described below. Construct a kernel function k that is an even function given by

$$k(x) = \begin{cases} 2x^3 - 3x^2 + 1, & \text{if } x \in [0, 1], \\ 0, & \text{if } x > 1. \end{cases}$$

We can verify that k is continuously differentiable on the real line and compactly supported on $[0, 1]$. Within the interior of its support, k is strictly positive. We use this function k to be the kernels k_T , k_R , k_Y , and k_{FS} in the estimation. The bandwidth is chosen to be $b_1 = b_2 = b_3 = 2n^{-1/5}$. The weighting function $w(e, u)$ is chosen to be constant in u and equal to the standard normal density function with respect to e .

Table 2 contains the simulation results of the performance of the semiparametric estimator for different choices of the marginal distribution of R , the correlation parameters (ρ_U, ρ_R) and the sample size. In each case, the number of replications is set at 500. We can see that the estimator performs well with a moderate sample size ($n = 1000$). The marginal distribution of R does not have a large impact on the performance. When the sample size is small, larger values of ρ_R or ρ_U can lead to poorer performance of the estimator. The plausible reason is that larger values of the correlation parameters would lead to more severe endogeneity issues in finite samples. When the sample size becomes large, the performance of the estimator does not vary significantly with the choices of (ρ_R, ρ_U) .

It is also of interest to compare the semiparametric estimator with the TSLS estimator. Directly comparing the two estimators can be difficult since they are of different dimensions and converge to different limits. Instead, we can compare their performance in estimating the marginal effect. For the structural function $g_\gamma(t, \bar{r}, e) = \gamma_1 t + \gamma_2 t^2 + \gamma_3 t e + e$, the marginal effect of the treatment on the outcome is $\frac{\partial}{\partial t} g_\gamma(t, \bar{r}, e) = \gamma_1 + 2\gamma_2 t + \gamma_3 e$, which takes on different values for different treatment and outcome levels. For a given

Dist. R	ρ_U	ρ_R	Param	$n = 500$			$n = 1000$			$n = 1500$		
				bias	sd	mse	bias	sd	mse	bias	sd	mse
$U(0, 1)$.3	.3	γ_1	-.257	.206	.108	-.233	.134	.072	-.219	.119	.062
			γ_2	.038	.153	.025	.062	.098	.013	.068	.081	.011
			γ_3	.172	.139	.049	.158	.091	.033	.154	.073	.029
		.5	γ_1	-.260	.464	.283	-.231	.134	.071	-.215	.116	.060
			γ_2	.027	.514	.265	.053	.097	.012	.060	.080	.010
			γ_3	.178	.168	.060	.159	.097	.035	.155	.076	.030
	.5	.3	γ_1	-.256	.199	.105	-.225	.129	.067	-.210	.111	.056
			γ_2	.039	.157	.026	.061	.098	.013	.068	.078	.011
			γ_3	.189	.140	.055	.171	.086	.037	.163	.070	.031
		.5	γ_1	-.248	.456	.269	-.217	.125	.063	-.202	.104	.052
			γ_2	.015	.512	.262	.044	.095	.011	.052	.075	.008
			γ_3	.222	.419	.225	.178	.100	.042	.171	.076	.035
$N(0, 1)$.3	.3	γ_1	-.224	.479	.280	-.228	.156	.076	-.216	.136	.065
			γ_2	.006	.527	.278	.059	.114	.016	.067	.093	.013
			γ_3	.152	.170	.052	.152	.105	.034	.148	.085	.029
		.5	γ_1	-.254	.362	.196	-.224	.153	.074	-.212	.124	.060
			γ_2	.030	.268	.073	.048	.112	.015	.058	.091	.012
			γ_3	.167	.178	.060	.152	.113	.036	.149	.086	.030
	.5	.3	γ_1	-.243	.233	.113	-.221	.149	.071	-.209	.127	.060
			γ_2	.027	.185	.035	.059	.113	.016	.069	.090	.013
			γ_3	.176	.172	.061	.163	.102	.037	.157	.082	.031
		.5	γ_1	-.230	.500	.303	-.210	.144	.065	-.198	.118	.053
			γ_2	-.012	.619	.383	.037	.113	.014	.049	.086	.010
			γ_3	.193	.248	.099	.172	.116	.043	.167	.088	.036

Table 2: Performance of the semiparametric estimator.

The structural function follows the three-parameter specification in Example 1. The number of replications is 500. The marginal distribution of R is chosen to be the uniform distribution on $[0, 1]$ or the standard normal distribution. The two correlation parameters $\rho_R = \text{corr}(\varepsilon, R)$ and $\rho_U = \text{corr}(\varepsilon, U)$ are chosen from $\{0.3, 0.5\}$. The results demonstrate the following points. First, the semiparametric estimator performs well with a moderate sample size of 1000. Second, the performance of the estimator is not affected by The marginal distribution of R . Third, when the sample size is as large as 1000, the performance of the estimator does not vary significantly with the choices of (ρ_R, ρ_U) .

treatment level t , we can use the semiparametric estimator to obtain an estimate $\hat{\gamma}_1 + 2\hat{\gamma}_2t + \hat{\gamma}_3e$ of the marginal effect. However, a TSLS procedure would deliver a scalar estimate that is a mixture of marginal effects across different treatment and outcome levels. Figure 6 shows that with a nonlinear specification, the semiparametric estimator outperforms the TSLS estimator. We compare the two estimates of the marginal effect at four quantile levels of the treatment: 20%, 40%, 60%, and 80%. The simulated distribution of the semiparametric estimator is correctly centered, while the TSLS estimator incurs a large bias. Figure 7 shows similar findings with a fully nonlinear and nonseparable specification, where we compare the two estimates at four levels of e : 0.2, 0.4, 0.6, 0.8.

Next, we compare the semiparametric estimator with the TSLS estimator when the structural function g is linear. This is achieved by imposing $\gamma_2 = \gamma_3 = 0$. The remaining slope coefficient γ_1 is equal to the marginal effect. In this case, the TSLS estimator is consistent for the coefficient γ_1 . However, the identification of the TSLS estimator is based solely on the difference between the two means. If the two distributions corresponding to h_0 and h_1 have the same mean, then the TSLS procedure suffers from weak identification issues. In contrast, the identification of the semiparametric estimator $\hat{\gamma}$ is based on the entire difference between h_0 and h_1 . As a result, the semiparametric estimator continues to work under the linear specification even if the estimand of the TSLS estimator is weakly identified. For the simulation, we let h_0 be the quantile function of $Beta(0.1, 0.1)$, and h_1 be the quantile function of $Beta(10, 10)$. These two distributions are significantly different, but they have the same mean, which is equal to 0.5. Figure 8 shows that, in this case, the semiparametric estimator outperforms the TSLS estimator in estimating the marginal effect.

5 Conclusion

In this study, we have examined the identification and estimation of the structural function in an RD design with a continuous treatment variable. We have established the nonparametric identification result and proposed a semiparametric estimator for the possibly nonlinear and nonseparable structural function. The estimator is proven to be consistent and asymptotically normal. The empirical study and simulation results demonstrate the

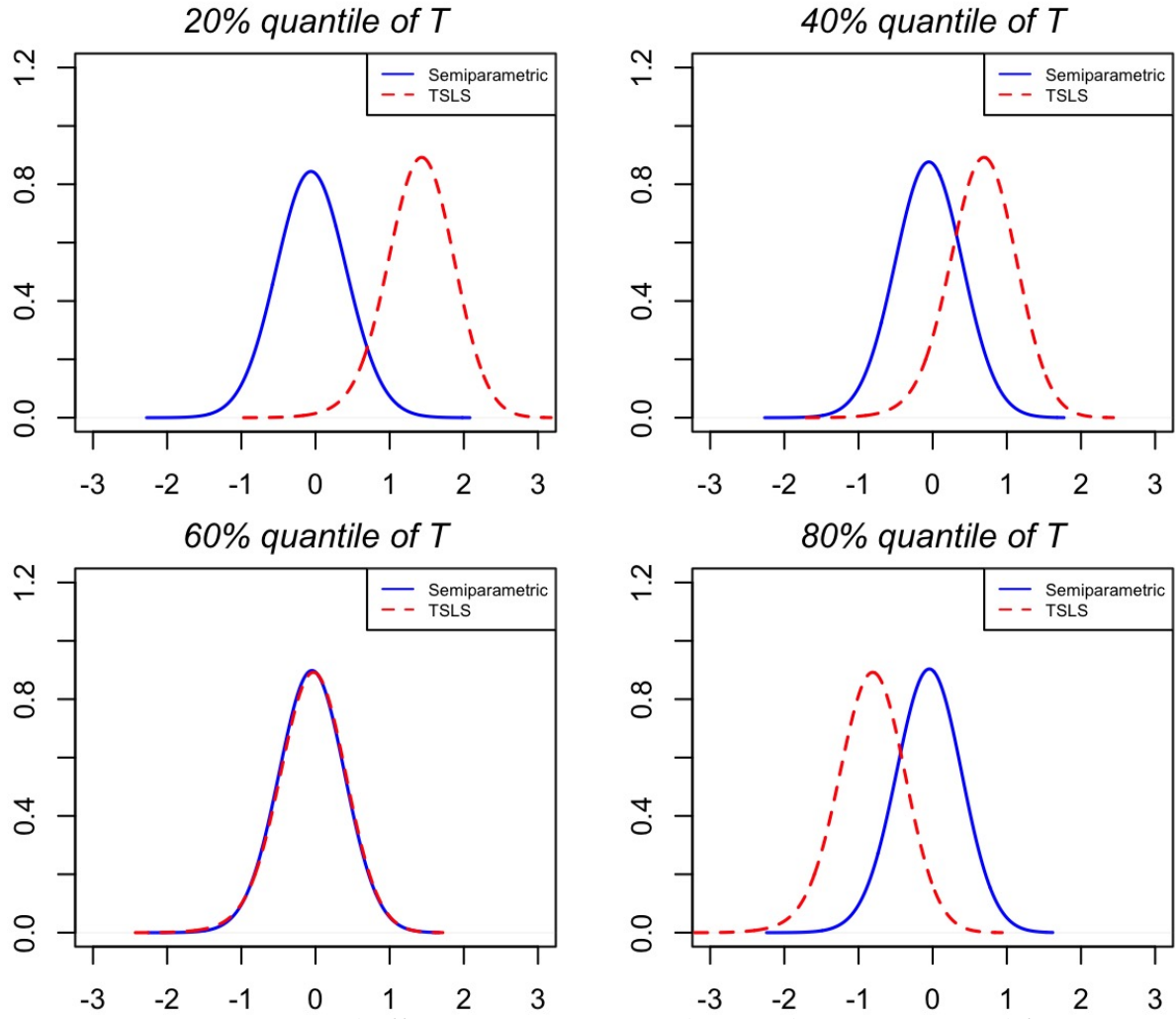


Figure 6: Marginal effects comparison with a nonlinear structural function.

Kernel density estimates of estimated marginal effect by the semiparametric and TSLS estimators (minus the true marginal effect) based on 500 replications. The sample size is 1000. The true structural function is specified to be $g(t, \bar{r}, e) = t/2 + t^2 + e$, where the marginal effect is $1/2 + 2t$. The graphs show the estimation results of four quantile levels of the treatment: 20%, 40%, 60%, and 80%. The distribution of the semiparametric estimator is correctly centered while the TSLS estimator incurs a large bias. The TSLS estimator gives an approximately unbiased estimate of the marginal effect only around the 60% quantile level.

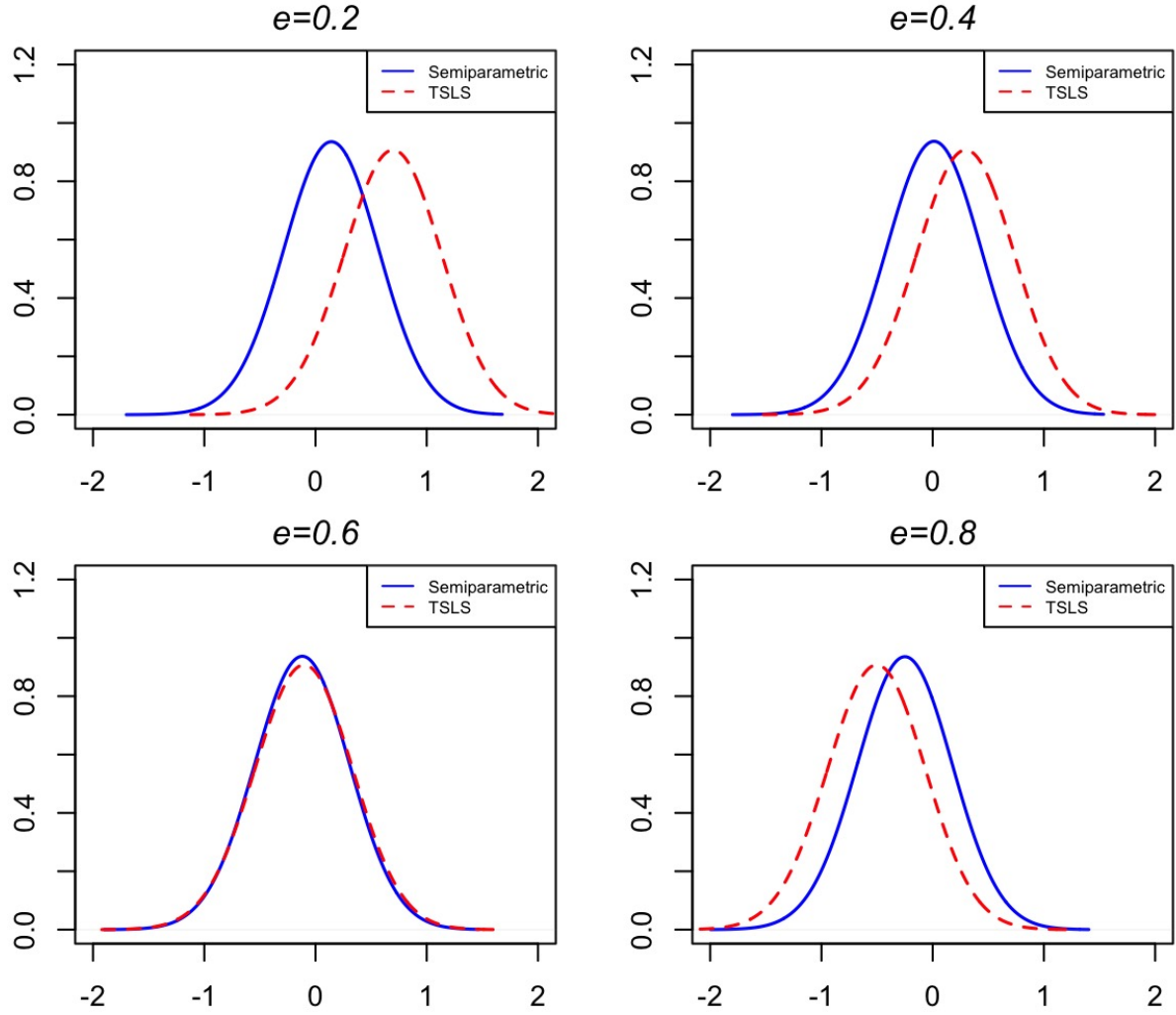


Figure 7: Marginal effects comparison with a nonseparable structural function.

Kernel density estimates of estimated marginal effect by the semiparametric and TSLS estimators (minus the true marginal effect) based on 500 replications. The sample size is 1000. The true structural function is specified to be $g(t, \bar{r}, e) = t/2 + t^2 + 2te + e$, where the marginal effect is $1/2 + 2t + 2e$. The treatment level is specified to be the median. The graphs show the estimation results of four levels of e : 0.2, 0.4, 0.6, 0.8. The distribution of the semiparametric estimator is correctly centered while the TSLS estimator incurs a large bias. The TSLS estimator gives an unbiased estimate of the marginal effect only around $e = 0.6$.

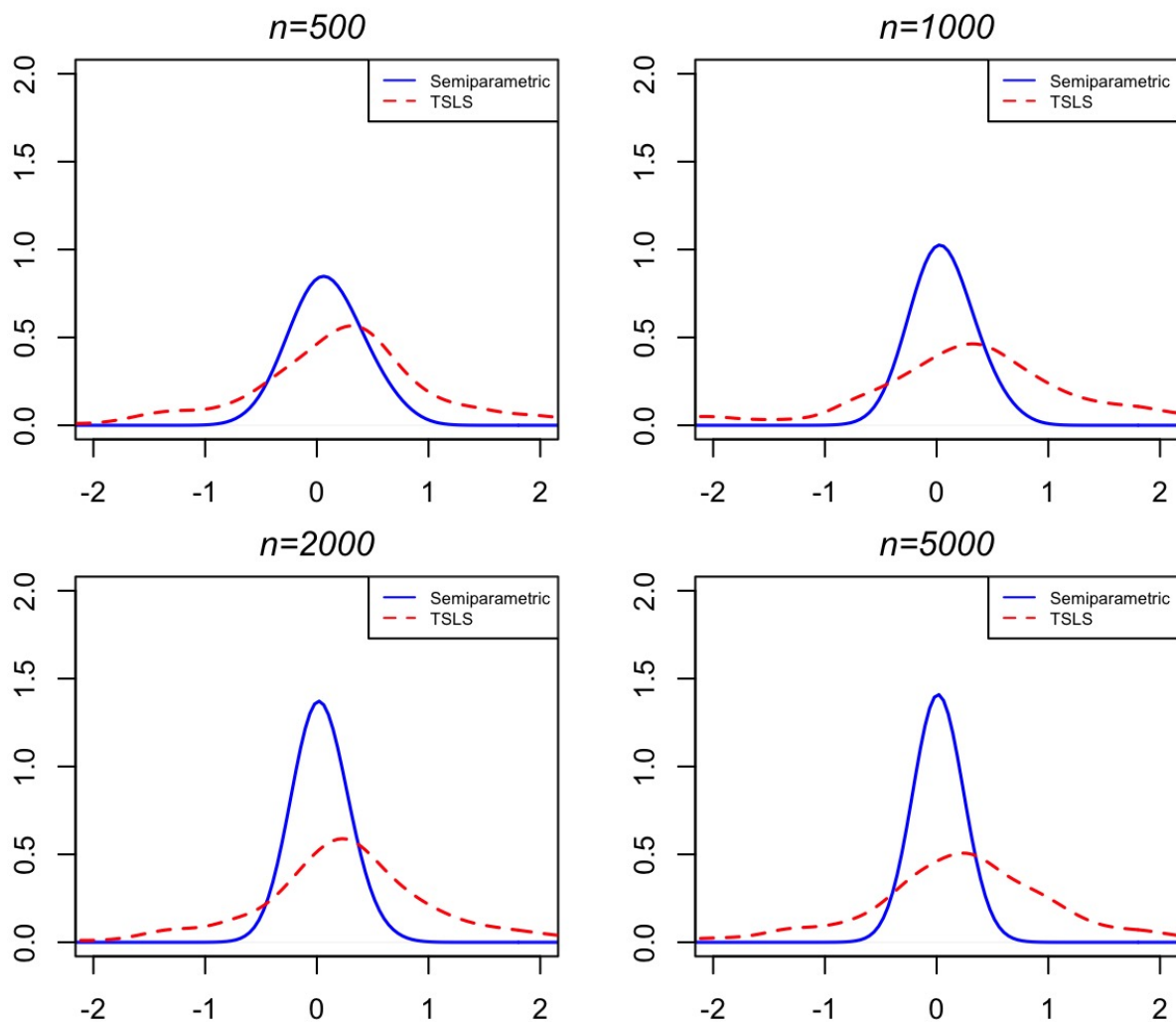


Figure 8: Semiparametric and TSLS estimators when the structural function is linear.

Kernel density estimates of the semiparametric and TSLS estimators (minus the true $\gamma^* = 1$) based on 500 replications. In the DGP, h_0 is equal to the quantile function of $Beta(0.1, 0.1)$, and h_1 is equal to the quantile function of $Beta(10, 10)$. These two distributions are significantly different, but they have the same mean. In this case, the semiparametric estimator outperforms the TSLS estimator because the latter is weakly identified.

advantage of the semiparametric estimator compared to the TSLS estimator.

There are two promising ways to extend the results in this paper in the future. First, we can consider extrapolating the identification result away from the cutoff. The extrapolation can be done by identifying the derivative of the structural function with respect to the running variable at the cutoff as in [Dong and Lewbel \(2015\)](#). Second, we can apply the methodology developed in this paper to the regression kink design model studied by [Card et al. \(2015\)](#); [Dong \(2018b\)](#), where the treatment choice function exhibits a kink instead of a discontinuity at the cutoff.

A Proof of identification results

In this section we prove the identification results including Lemma 1, Lemma 2, Theorem 1, Corollary 1, and Corollary 2.

Proof of Lemma 1. We first prove the second equality in Equation (4). The conditional distribution function of T given R is

$$\begin{aligned} F_{T|R}(t|r) &= \mathbb{P}(T \leq t | R = r) \mathbf{1}\{r < \bar{r}\} + \mathbb{P}(T \leq t | R = r) \mathbf{1}\{r \geq \bar{r}\} \\ &= \mathbb{P}(U_0 \leq m_0^{-1}(r, t) | R = r) \mathbf{1}\{r < \bar{r}\} + \mathbb{P}(U_1 \leq m_1^{-1}(r, t) | R = r) \mathbf{1}\{r \geq \bar{r}\} \\ &= F_{U_0|R}(m_0^{-1}(r, t) | r) \mathbf{1}\{r < \bar{r}\} + F_{U_1|R}(m_1^{-1}(r, t) | r) \mathbf{1}\{r \geq \bar{r}\}, \end{aligned}$$

where the last line follows from the monotonicity of m_0 and m_1 . Therefore, we have

$$\begin{aligned} F_{T|R}(T|R) &= F_{U_0|R}(m_0^{-1}(R, T) | R) \mathbf{1}\{R < \bar{r}\} + F_{U_1|R}(m_1^{-1}(R, T) | R) \mathbf{1}\{R \geq \bar{r}\} \\ &= F_{U_0|R}(U_0 | R) \mathbf{1}\{R < \bar{r}\} + F_{U_1|R}(U_1 | R) \mathbf{1}\{R \geq \bar{r}\} = U. \end{aligned}$$

For (i) of Lemma 1, take any $u \in [0, 1]$ and $r < \bar{r}$, we have

$$\mathbb{P}(U \leq u | R = r) = \mathbb{P}(F_{U_0|R}(U_0 | r) \leq u | R = r) = u.$$

Similarly, we can show that $\mathbb{P}(U \leq u | R = r) = u$ when $r \geq \bar{r}$. Therefore, $U | R$ follows the uniform distribution. Then the second argument follows from the monotonicity of h with

respect to the second argument. The statements (ii) and (iii) are straightforward from the definition of h_0 and h_1 and the monotonicity and continuity assumptions. For (iv), notice that for $r < \bar{r}$,

$$\begin{aligned} F_{\varepsilon|U,R}(e|u,r) &= \mathbb{P}(\varepsilon \leq e|U = u, R = r) \\ &= \mathbb{P}(\varepsilon \leq e|F_{U_0|R}(U_0|r) = u, R = r) \\ &= \mathbb{P}(\varepsilon \leq e|U_0 = F_{U_0|R}^{-1}(u|r), R = r) \\ &= F_{\varepsilon|U_0,R}(e|F_{U_0|R}^{-1}(u|r), r), \end{aligned}$$

where the third equality follows from the strict monotonicity of $F_{U_0|R}^{-1}(u|r)$ in u imposed in Assumption 2(ii). Similarly, we can show that for $r \geq \bar{r}$,

$$F_{\varepsilon|U,R}(e|u,r) = \mathbb{P}(\varepsilon \leq e|U = u, R = r) = F_{\varepsilon|U_1,R}(e|F_{U_1|R}^{-1}(u|r), r).$$

Combining the two equations together, we have

$$F_{\varepsilon|U,R}(e|u,r) = \begin{cases} F_{\varepsilon|U_0,R}(e|F_{U_0|R}^{-1}(u|r), r), & r < \bar{r}, \\ F_{\varepsilon|U_1,R}(e|F_{U_1|R}^{-1}(u|r), r), & r \geq \bar{r}. \end{cases}$$

By Assumption 2(iii), we know that $F_{\varepsilon|U,R}(e|u,r)$ is strictly increasing in the first argument e . By Bayes rule, the rank similarity condition in Assumption 3 implies that $U_0|R = \bar{r}^-$ has the same distribution as $U_1|R = \bar{r}^+$, and $\varepsilon|U_0, R = \bar{r}^-$ has the same distribution as $\varepsilon|U_1, R = \bar{r}^+$. Then we have

$$\lim_{r \uparrow \bar{r}} F_{\varepsilon|U,R}(e|u,r) = F_{\varepsilon|U_0,R}(e|F_{U_0|R}^{-1}(u|\bar{r}), \bar{r}) = F_{\varepsilon|U_1,R}(e|F_{U_1|R}^{-1}(u|\bar{r}), \bar{r}) = \lim_{r \downarrow \bar{r}} F_{\varepsilon|U,R}(e|u,r).$$

□

Lemma A.1. *Let $f : (x, y) \mapsto z$ be a real-valued bivariate function defined on a compact set in $\mathbb{R} \times \mathbb{R}^d$. Assume that f is continuous on its entire domain and strictly increasing in the first argument x . Let f^{-1} denote the inverse of f with respect to the first argument. Then f^{-1} is continuous on its domain and strictly increasing in the first argument.*

Therefore, under Assumptions 1 and 2, the inverse of g, h_0 and h_1 (with respect to the last argument), which are respectively $g^{-1}, (h_0)^{-}$ and $(h_1)^{-}$, are all continuous and strictly increasing with respect to the last argument.

Proof of Lemma A.1. Fix any (z_0, y_0) , we want to show that $f^{-1}(z_0, \cdot)$ is continuous at (z_0, y_0) . If not, then there exists $\delta > 0$ and a sequence $\{(z_k, y_k)\}$ such that $\|(z_k, y_k) - (z_0, y_0)\| < 1/k$ but

$$|f^{-1}(z_0, y_k) - f^{-1}(z_0, y_0)| > \delta.$$

Denote $x_k = f^{-1}(z_k, y_k)$ and $x_0 = f^{-1}(z_0, y_0)$. Because the sequence $\{x_k\}$ lies in a compact set, it has a convergent subsequence. Without loss of generality, we assume $\{x_k\}$ itself is converging. Then $\lim x_k \neq x_0$. However, by the continuity of f ,

$$f(\lim x_k, y_0) = f(\lim x_k, \lim y_k) = \lim f(x_k, y_k) = z_0 = f(x_0, y_0).$$

This leads to a contradiction since $f(\cdot, y_0)$ is strictly increasing.

To show that f^{-1} is strictly increasing with respect to the first argument, take any y and $z_1 > z_0$. If $f^{-1}(z_1, y) \leq f^{-1}(z_0, y)$, then $z_1 = f(f^{-1}(z_1, y), y) \leq f(f^{-1}(z_0, y), y) = z_0$, which leads to a contradiction. \square

Proof of Lemma 2. By the definition of $F_{Y|T,R}^-$ and the monotonicity and continuity of g^* and h_0 ,

$$\begin{aligned} \text{LHS of (6) with } g = g^* &= \lim_{r \uparrow \bar{r}} F_{Y|T,R}(g^*(h_0(r, u), r, e) | h_0(r, u), r) \\ &= \lim_{r \uparrow \bar{r}} \mathbb{P}(Y \leq g^*(h_0(r, u), r, e) | T = h_0(r, u), R = r) \\ &= \lim_{r \uparrow \bar{r}} \mathbb{P}(g^*(h_0(r, u), r, \varepsilon) \leq g^*(h_0(r, u), r, e) | T = h_0(r, u), R = r) \\ &= \lim_{r \uparrow \bar{r}} \mathbb{P}(\varepsilon \leq e | U = u, R = r) \\ &= F_{\varepsilon|U,R}(e | u, r) \end{aligned}$$

where the last line follows from the continuity of $F_{\varepsilon|U,R}(e | u, r)$ with respect to the last argument r (Lemma 1). Similarly, we can show that the RHS of (6) is equal to $F_{\varepsilon|U,R}(e | u, \bar{r})$.

Then the result follows. □

Proof of Theorem 1. Denote

$$\mathcal{T}^\times = \{h_0(\bar{r}, u) : h_0(\bar{r}, u) = h_1(\bar{r}, u) \in [t'_0, t''_0] \cap [t'_1, t''_1], u \in [0, 1]\}.$$

By Assumption 5(ii), \mathcal{T}^\times is nonempty and finite. Then $[t'_0, t''_0] \cap [t'_1, t''_1]$ is a closed interval with nonempty interior.²¹ Let

$$\inf([t'_0, t''_0] \cap [t'_1, t''_1]) = t_1 \leq t_2 \leq \dots \leq t_L = \sup([t'_0, t''_0] \cap [t'_1, t''_1])$$

denote the unique elements of $\mathcal{T}^\times \cup \{\inf([t'_0, t''_0] \cap [t'_1, t''_1]), \sup([t'_0, t''_0] \cap [t'_1, t''_1])\}$.

Here is the strategy of the proof. For each $g \in \mathcal{G}$ that satisfies Equation (6), define

$$\tilde{\lambda}^g(t, e) = g^{-1}(t, \bar{r}, g^*(t, \bar{r}, e)). \quad (\text{A.1})$$

The goal is to show that $\tilde{\lambda}^g$ is constant as a function of t for every $e \in \mathcal{E}$. We proceed in five steps. Step 1 derives some useful properties of the function $\tilde{\lambda}^g$, including an important identity, Equation (A.2). Step 2 shows that $\tilde{\lambda}^g$ is constant in t on the interval (t_1, t_2) . Step 3 shows that $\tilde{\lambda}^g$ is constant in t on the entire region $[t_1, t_L] = [t'_0, t''_0] \cap [t'_1, t''_1]$. Step 4 further expands this constancy to $[t'_0, t''_0] \cup [t'_1, t''_1]$. Step 5 concludes.

Step 1. Since g^* and g^{-1} are continuous and are strictly increasing in the last argument, $\tilde{\lambda}^g$ is also continuous and strictly increasing in the last argument. Also, $F_{Y|T,R}^-(\cdot | t, \bar{r})$ is strictly increasing since

$$F_{Y|T,R}^-(y | t, \bar{r}) = F_{\varepsilon|U,R}((g^*)^{-1}(t, \bar{r}, y) | h_0^{-1}(\bar{r}, t), \bar{r})$$

is strictly increasing in y (Assumption 2(ii)).

²¹Notice that if $[t'_0, t''_0] \cap [t'_1, t''_1]$ is a singleton, then \mathcal{T}^\times is empty.

Notice that $g(t, \bar{r}, \tilde{\lambda}^g(t, e)) = g^*(t, \bar{r}, e)$. Then for any $e \in \mathcal{E}$ and $u \in [0, 1]$,

$$\begin{aligned}
& F_{Y|T,R}^-(g(h_0(\bar{r}, u), \bar{r}, \tilde{\lambda}^g(h_0(\bar{r}, u), e)) | h_0(\bar{r}, u), \bar{r}) \\
&= F_{Y|T,R}^-(g^*(h_0(\bar{r}, u), \bar{r}, e) | h_0(\bar{r}, u), \bar{r}) \\
&= F_{Y|T,R}^+(g^*(h_1(\bar{r}, u), \bar{r}, e) | h_1(\bar{r}, u), \bar{r}) \\
&= F_{Y|T,R}^+(g(h_1(\bar{r}, u), \bar{r}, \tilde{\lambda}^g(h_1(\bar{r}, u), e)) | h_1(\bar{r}, u), \bar{r}),
\end{aligned}$$

where the second inequality follows from Lemma 2. The above equality implies

$$\tilde{\lambda}^g(h_0(\bar{r}, u), e) = \tilde{\lambda}^g(h_1(\bar{r}, u), e). \quad (\text{A.2})$$

To see that, suppose there exists e and u such that $\tilde{\lambda}^g(h_0(\bar{r}, u), e) \neq \tilde{\lambda}^g(h_1(\bar{r}, u), e)$. Since g satisfies Condition (6),

$$\begin{aligned}
& F_{Y|T,R}^-(g(h_0(\bar{r}, u), \bar{r}, \tilde{\lambda}^g(h_1(\bar{r}, u), e)) | h_0(\bar{r}, u), \bar{r}) \\
&= F_{Y|T,R}^+(g(h_1(\bar{r}, u), \bar{r}, \tilde{\lambda}^g(h_1(\bar{r}, u), e)) | h_1(\bar{r}, u), \bar{r}) \\
&= F_{Y|T,R}^-(g(h_0(\bar{r}, u), \bar{r}, \tilde{\lambda}^g(h_0(\bar{r}, u), e)) | h_0(\bar{r}, u), \bar{r}),
\end{aligned}$$

which violates the fact that $F_{Y|T,R}^-(g(h_0(\bar{r}, u), \bar{r}, \cdot) | h_0(\bar{r}, u), \bar{r})$ is strictly increasing.

Step 2. Consider the interval (t_1, t_2) . By construction, $\{t_1, t_2\} \cap \mathcal{T}^\times \neq \emptyset$. Notice that over (t_1, t_2) , $h_0^{-1}(\bar{r}, \cdot)$ and $h_1^{-1}(\bar{r}, \cdot)$ do not intersect. Then by continuity, one of them is always strictly greater than the other. The goal is to show that $\tilde{\lambda}^g(t, e)$ is constant as a function of t over (t_1, t_2) . There are four cases to consider, depending on whether $t_1 \in \mathcal{T}^\times$ or $t_2 \in \mathcal{T}^\times$ and whether $h_0^{-1}(\bar{r}, \cdot)$ is strictly greater or smaller than $h_1^{-1}(\bar{r}, \cdot)$ over (t_1, t_2) .

We first focus on the case of $t_1 \in \mathcal{T}^\times$ and $h_0^{-1}(\bar{r}, \cdot) < h_1^{-1}(\bar{r}, \cdot)$ over (t_1, t_2) . The other cases are essentially the same. Define a mapping $\pi(t) = h_1(\bar{r}, h_0^{-1}(\bar{r}, t))$. Such a mapping π maps the interval (t_1, t_2) back to itself. To see that, we first notice that $\pi(t)$ is less than t for any $t \in (t_1, t_2)$ since

$$\pi(t) = h_1(\bar{r}, h_0^{-1}(\bar{r}, t)) \leq h_1(\bar{r}, h_1^{-1}(\bar{r}, t)) = t < t_2.$$

Suppose $\pi(t) \leq t_1$, then

$$\begin{aligned} h_0^{-1}(\bar{r}, t_1) &= h_1^{-1}(\bar{r}, t_1) \\ &\leq h_1^{-1}(\bar{r}, \pi(t)) \\ &= h_0^{-1}(\bar{r}, t), \end{aligned}$$

where the first line follows from $t_1 \in \mathcal{T}^\times$, the second line follows from the monotonicity of h_1^{-1} and $\pi(t) \leq t_1$, and the last line follows from the definition of π . This contradicts the strict monotonicity of $h_0^{-1}(\bar{r}, \cdot)$ since $t_1 < t$.

Now pick any $\tilde{t}_0 \in (t_1, t_2)$, the recursive sequence $\tilde{t}_{k+1} = \pi(\tilde{t}_k)$ is well-defined. This sequence is non-increasing and bounded below by t_1 . Therefore, $\lim_{k \rightarrow \infty} \tilde{t}_k$ exists and lies in the interval $[t_1, t_2)$. By the continuity of h_0^{-1} and h_1^{-1} ,

$$\begin{aligned} h_0^{-1}(\bar{r}, \lim \tilde{t}_k) &= \lim h_0^{-1}(\bar{r}, \tilde{t}_k) \\ &= \lim h_1^{-1}(\bar{r}, \pi(\tilde{t}_k)) \\ &= \lim h_1^{-1}(\bar{r}, \tilde{t}_{k+1}) \\ &= h_1^{-1}(\bar{r}, \lim \tilde{t}_{k+1}), \end{aligned}$$

where the second line follows from the definition of π and the third line follows from the construction of the sequence $\{\tilde{t}_k\}$. Then it must be true that $\lim \tilde{t}_k = t_1$ since we are studying the case where $h_0^{-1}(\bar{r}, \cdot) < h_1^{-1}(\bar{r}, \cdot)$ over (t_1, t_2) .

Equation (A.2) implies that $\tilde{\lambda}^g(\cdot, e)$ is invariant with respect to the transformation π :

$$\tilde{\lambda}^g(\pi(t), \bar{r}, e) = \tilde{\lambda}^g(t, e),$$

for every $t \in [t'_0, t''_0]$ and $e \in \mathcal{E}$. Then $\tilde{\lambda}^g$ is invariant along the sequence $\{\tilde{t}_k\}$. By the continuity of $\tilde{\lambda}^g$,

$$\tilde{\lambda}^g(\tilde{t}_0, e) = \lim_{k \rightarrow \infty} \tilde{\lambda}^g(\tilde{t}_0, e) = \lim_{k \rightarrow \infty} \tilde{\lambda}^g(\tilde{t}_k, e) = \tilde{\lambda}^g(\lim_{k \rightarrow \infty} \tilde{t}_k, e) = \tilde{\lambda}^g(t_1, e) = \lambda^g(e),$$

where the first equality holds since $\tilde{\lambda}^g(\tilde{t}_0, \bar{r}, e)$ is constant with respect to k , the second equality holds since $\tilde{\lambda}^g$ is invariant along the sequence $\{\tilde{t}_k\}$, the third equality follows

from the continuity of $\tilde{\lambda}^g$, and the last equality is the definition of the function λ on \mathcal{E} .

Since the initial point \tilde{t}_0 is chosen arbitrarily from the interval (t_1, t_2) , the above analysis shows that $\tilde{\lambda}^g(t, e) = \lambda^g(e)$ for $t \in (t_1, t_2)$ (hence for $t \in [t_1, t_2]$, by continuity) and $e \in \mathcal{E}$.

Recall that this analysis is conducted for the case where $t_1 \in \mathcal{T}^\times$ and $h_0^{-1}(\bar{r}, \cdot)$ is strictly smaller than $h_1^{-1}(\bar{r}, \cdot)$ over (t_1, t_2) . The other three cases reach the same conclusion that $\tilde{\lambda}^g(t, e)$ is equal to $\lambda^g(e)$ over $[t_1, t_2] \times \mathcal{E}$ through symmetric arguments. More specifically, we can switch h_1 and h_0 in defining π so that the sequence $\{\tilde{t}_k\}$ tends to a point in \mathcal{T}^\times .

Step 3. Repeat step 2 on each interval (t_l, t_{l+1}) , $l = 2, \dots, L$. It follows that $\tilde{\lambda}^g(t, e) = \lambda^g(e)$ over $([t'_0, t''_0] \cap [t'_1, t''_1]) \times [0, 1]$.

Step 4. Pick any $t' \in \text{int}([t'_0, t''_0] \cup [t'_1, t''_1]) \setminus [t_1, t_L]$ (if this set is nonempty). There are four cases to consider, depending on whether $t' \in [t'_0, t''_0]$ or $t' \in [t'_1, t''_1]$ and whether $t''_0 < t''_1$ or $t''_0 > t''_1$. Without loss of generality, assume that $t' \in [t'_0, t''_0]$ and $t''_0 < t''_1$. The other three cases can be dealt with symmetric arguments. In this case, $t'_1 < t''_0$ because $[t'_0, t''_0] \cap [t'_1, t''_1]$ is a non-degenerate interval. Denote $t'' \in [t'_1, t''_1]$ such that $h_0^{-1}(\bar{r}, t) = h_1^{-1}(\bar{r}, t'')$. It must be the case that $t'' \in [t'_1, t''_0]$. If that is not the case, then $t'' > t''_0$. Then for any $t \in [t'_1, t''_0]$,

$$h_0^{-1}(\bar{r}, t) > h_0^{-1}(\bar{r}, t') = h_1^{-1}(\bar{r}, t'') > h_1^{-1}(\bar{r}, t), t \in [t'_0, t''_0] \cap [t'_1, t''_1],$$

by the strict monotonicity of $h_0^{-1}(\bar{r}, \cdot)$ and $h_1^{-1}(\bar{r}, \cdot)$. However, this contradicts the assumption that \mathcal{T}^\times is nonempty. Then by (A.2),

$$\tilde{\lambda}^g(t', e) = \tilde{\lambda}^g(t'', e) = \lambda^g(e).$$

Step 5. By the definition of $\tilde{\lambda}^g$ in (A.1), we now have

$$g^*(t, \bar{r}, e) = g(t, \bar{r}, \lambda^g(e)), \text{ for } t \in [t'_0, t''_0] \cup [t'_1, t''_1], e \in \mathcal{E}.$$

By the properties of $\tilde{\lambda}^g$, we know λ^g is continuous and strictly increasing. The above statement holds for any $g \in \mathcal{G}$ that satisfies Equation (6).

□

Proof of Corollary 1. Based on the definition of $F_{\varepsilon|U,R}^{g^\circ}$ in (7), Theorem 1 and Lemma 2, we have

$$\begin{aligned} F_{\varepsilon|U,R}^{g^\circ}(\lambda^g(e)|u, \bar{r}) &= F_{Y|T,R}^-(g^\circ(h_0(\bar{r}, u), \bar{r}, \lambda^{g^\circ}(e))|h_0(\bar{r}, u), \bar{r}) \\ &= F_{Y|T,R}^-(g^\circ(h_0(\bar{r}, u), \bar{r}, e)|h_0(\bar{r}, u), \bar{r}) \\ &= F_{\varepsilon|U,R}(e|u, \bar{r}). \end{aligned}$$

The second claim follows from a change of variable. □

Proof of Corollary 2. By construction, $\|D_{\gamma, h^*}\|_w \geq 0$ for any $\gamma \in \Gamma$. By Lemma 2, we have $\|D_{\gamma^*, h^*}\|_w = 0$. We want to show that γ^* is the unique zero. Since $w > 0$, we have

$$\begin{aligned} \|D_{\gamma, h^*}\|_w = 0 &\implies D_{\gamma, h^*}(e, u) = 0, \text{ for all } e \in \mathcal{E}, u \in [0, 1] \\ &\implies \text{Condition (6) is satisfied by } g_\gamma(\cdot, \bar{r}, \cdot), \end{aligned}$$

where the second line follows by taking the partial derivative with respect to u on both sides. By Theorem 1, this implies that $g_\gamma(\cdot, \bar{r}, \cdot) = g_{\gamma'}(\cdot, \bar{r}, \lambda(\cdot))$. By Assumption 7, it must be that $\gamma = \gamma^*$. Therefore, γ^* is the unique minimizer of $\|D_{\gamma, h^*}\|_w$. □

B Proof of estimation results

This section proceeds as follows. Section B.1 provides the proofs of Theorem 2 and Proposition 1. Section B.2 introduces the empirical process theory and presents the lemmas on the uniform convergence results used in Section B.1. Section B.3 discusses the consistent estimation of the asymptotic covariance matrix.

B.1 Proofs of Theorem 2 and Proposition 1

Proof of Theorem 2. In this proof, the functions $h_1(r, u)$, $h_0(r, u)$, and $g(t, r, e)$ are only evaluated at $r = \bar{r}$. For simplicity, we omit this argument \bar{r} throughout. The proof proceeds with seven steps:

- [Step 1](#) contains preliminary results on the LLR estimator of the condition distribution $Y|T, R$.
- [Step 2](#) derives the consistency of $\hat{\gamma}$.
- [Step 3](#) derives an initial estimate of the convergence rate of $\hat{\gamma}$.
- [Step 4](#) proves a stochastic equicontinuity condition on the criterion function.
- [Step 5](#) presents a linear approximation of the criterion function.
- [Step 6](#) shows the asymptotic normality of the minimizer of the linearized criterion function.
- [Step 7](#) derives the asymptotic normal distribution of $\hat{\gamma}$.

Step 1. (Preliminary results on LLR.) Let $X_i(t) = (1, (T_i - t)/b_1, (R_i - \bar{r})/b_1)'$ denote the vector containing the regressors in the LLR. For $\mathbf{x} = (1, x_1, x_2)$, let

$$\begin{aligned} k_0(\mathbf{x}) &= k_T(x_1)k_R(x_2)\mathbf{1}\{x_2 < 0\}, \\ k_1(\mathbf{x}) &= k_T(x_1)k_R(x_2)\mathbf{1}\{x_2 \geq 0\}. \end{aligned}$$

The kernel weights in the LLR can be written as

$$\begin{aligned} k_0(X_i(t)) &= k_T((T_i - t)/b_1)k_R((R_i - \bar{r})/b_1)\mathbf{1}\{R_i < \bar{r}\}, \\ k_1(X_i(t)) &= k_T((T_i - t)/b_1)k_R((R_i - \bar{r})/b_1)\mathbf{1}\{R_i \geq \bar{r}\}. \end{aligned}$$

From [Xie \(2021\)](#), we have the following uniform asymptotic linear representation for the LLR estimator of the conditional distribution functions:

$$\begin{aligned} \hat{F}_{Y|T,R}^-(y|t, \bar{r}) - F_{Y|T,R}^-(y|t, \bar{r}) &= b_1^2 \mu_0(y, t) + \iota' \Xi_0(t)^{-1} \frac{1}{nb_1^2} \sum_{i=1}^n s_0(Y_i, T_i, R_i, y, t) \\ &\quad + O_p \left(b_1^3 + \frac{|\log b_1|}{nb_1^2} \right), \end{aligned}$$

uniformly over $y \in \mathbb{R}, t \in [t'_0, t''_0]$, and

$$\begin{aligned} \hat{F}_{Y|T,R}^+(y|t, \bar{r}) - F_{Y|T,R}^+(y|t, \bar{r}) &= b_1^2 \mu_1(y, t) + \iota' \Xi_1(t)^{-1} \frac{1}{nb_1^2} \sum_{i=1}^n s_1(Y_i, T_i, R_i, y, t) \\ &\quad + O_p \left(b_1^3 + \frac{|\log b_1|}{nb_1^2} \right), \end{aligned}$$

uniformly over $y \in \mathbb{R}, t \in [t'_1, t''_1]$. In the above expressions, $\iota = (1, 0, \dots, 0)$. The functions $\mu_0(y, t)$ and $\mu_1(y, t)$ are defined by

$$\begin{aligned} \mu_0(y, t) &= \frac{b_1^2}{2} \iota' \Omega_0(t)^{-1} \frac{\partial^2}{\partial t^2} F_{Y|T,R}^-(y|t, \bar{r}) \int \mathbf{x} x_1^2 k_0(\mathbf{x}) \mathbf{1}\{t + b_1 x_1 \in [t'_0, t''_0]\} dx_1 dx_2 \\ &\quad + \frac{b_1^2}{2} \iota' \Omega_0(t)^{-1} \frac{\partial^2}{\partial r^2} F_{Y|T,R}^-(y|t, \bar{r}) \int \mathbf{x} x_2^2 k_0(\mathbf{x}) \mathbf{1}\{t + b_1 x_1 \in [t'_0, t''_0]\} dx_1 dx_2 \\ &\quad + b_1^2 \iota' \Omega_0(t)^{-1} \frac{\partial^2}{\partial t \partial r} F_{Y|T,R}^-(y|t, \bar{r}) \int \mathbf{x} x_1 x_2 k_0(\mathbf{x}) \mathbf{1}\{t + b_1 x_1 \in [t'_0, t''_0]\} dx_1 dx_2 \\ &\quad + \frac{b_1^2}{2} \Omega_0(t)^{-1} \frac{\partial^2}{\partial y^2} F_{Y|T,R}^-(y|t, \bar{r}) \int \mathbf{x} k_0(\mathbf{x}) \mathbf{1}\{t + b_1 x_1 \in [t'_0, t''_0]\} dx_1 dx_2, \end{aligned}$$

and

$$\begin{aligned} \mu_1(y, t) &= \frac{b_1^2}{2} \iota' \Omega_1(t)^{-1} \frac{\partial^2}{\partial t^2} F_{Y|T,R}^+(y|t, \bar{r}) \int \mathbf{x} x_1^2 k_1(\mathbf{x}) \mathbf{1}\{t + b_1 x_1 \in [t'_1, t''_1]\} dx_1 dx_2 \\ &\quad + \frac{b_1^2}{2} \iota' \Omega_1(t)^{-1} \frac{\partial^2}{\partial r^2} F_{Y|T,R}^+(y|t, \bar{r}) \int \mathbf{x} x_2^2 k_1(\mathbf{x}) \mathbf{1}\{t + b_1 x_1 \in [t'_1, t''_1]\} dx_1 dx_2 \\ &\quad + b_1^2 \iota' \Omega_1(t)^{-1} \frac{\partial^2}{\partial t \partial r} F_{Y|T,R}^+(y|t, \bar{r}) \int \mathbf{x} x_1 x_2 k_1(\mathbf{x}) \mathbf{1}\{t + b_1 x_1 \in [t'_1, t''_1]\} dx_1 dx_2 \\ &\quad + \frac{b_1^2}{2} \Omega_1(t)^{-1} \frac{\partial^2}{\partial y^2} F_{Y|T,R}^+(y|t, \bar{r}) \int \mathbf{x} k_1(\mathbf{x}) \mathbf{1}\{t + b_1 x_1 \in [t'_1, t''_1]\} dx_1 dx_2. \end{aligned}$$

The matrices $\Omega_0(t)$, $\Omega_1(t)$, $\Xi_0(t)$, and $\Xi_1(t)$ are defined by

$$\begin{aligned} \Omega_0(t) &= \int \mathbf{x} \mathbf{x}' k_0(\mathbf{x}) \mathbf{1}\{t + b_1 x_1 \in [t'_0, t''_0]\} dx_1 dx_2, \\ \Omega_1(t) &= \int \mathbf{x} \mathbf{x}' k_1(\mathbf{x}) \mathbf{1}\{t + b_1 x_1 \in [t'_1, t''_1]\} dx_1 dx_2, \end{aligned}$$

and

$$\begin{aligned}\Xi_0(t) &= \int \mathbf{x} \mathbf{x}' k_0(\mathbf{x}) f_{T,R}^-(t + b_1 x_1, \bar{r} + b_1 x_2) dx_1 dx_2, \\ \Xi_1(t) &= \int \mathbf{x} \mathbf{x}' k_1(\mathbf{x}) f_{T,R}^-(t + b_1 x_1, \bar{r} + b_1 x_2) dx_1 dx_2.\end{aligned}$$

The terms s_0 and s_1 are defined by

$$\begin{aligned}s_0(Y_i, T_i, R_i, y, t) &= X_i(t) \tilde{K}_Y(Y_i, T_i, R_i; y, t) k_0(X_i(t)), \\ s_1(Y_i, T_i, R_i, y, t) &= X_i(t) \tilde{K}_Y(Y_i, T_i, R_i; y, t) k_1(X_i(t)), \\ \tilde{K}_Y(Y_i, T_i, R_i, y) &= K_Y((y - Y_i)/b_1) - \mathbb{E}[K_Y((y - Y_i)/b_1) | T_i, R_i].\end{aligned}$$

Under Assumption 9(i) and (ii) and Assumption 11(i), we can apply Lemma 1 in Xie (2021), which is a modification of Lemma 11 in Fan and Guerre (2016), and obtain that the eigenvalues of $\Omega_0(t)$, $\Omega_1(t)$, $\Xi_0(t)$ and $\Xi_1(t)$ are bounded and bounded away from zero for all values of t and b_1 . Consequently, the norm of these matrices and their inverses are bounded.

Notice that in Xie (2021), the remainder term from the bias expansion is $o(b_1^2)$ while in the above asymptotic linear representation, the corresponding term is $O(b_1^3)$. This is because we assume that $F_{Y|T,R}^-$ and $F_{Y|T,R}^+$ are three-times continuously differentiable (Assumption 9). Under this assumption, we can go through the same steps as in the proof of Theorem 1 in Xie (2021) and show that the remainder from the bias expansion is $O(b_1^3)$. The details are omitted for brevity. The bias terms $\mu_0(y, t)$ and $\mu_1(y, t)$ are continuously differentiable under Assumption 9. The indicator functions inside the integral, for example, $\mathbf{1}\{t + b_1 x_1 \in [t'_0, t''_0]\}$, can be eliminated by explicitly indicating the lower and upper limits of the corresponding integral. The derivative can then be taken by using the Leibniz rule.

Step 2. (Consistency of $\hat{\gamma}$.) Because w is positive and integrates to 1, we have $\|\hat{D}_{\gamma, \hat{h}} - D_{\gamma, \hat{h}}\|_w \leq \|\hat{D}_{\gamma, \hat{h}} - D_{\gamma, \hat{h}}\|_\infty$. For any $\gamma \in \Gamma$, we can apply Fubini's theorem to the uniform

asymptotic linear representation and obtain that

$$\hat{D}_{\gamma, \hat{h}}(e, u) - D_{\gamma, \hat{h}}(e, u) = \text{I} + \text{II} + O_p\left(b_1^2 + \log n / (nb_1^2)\right)$$

uniformly over $\gamma \in \Gamma, e \in \mathcal{E}$, and $u \in (0, 1)$, where

$$\begin{aligned} \text{I} &= \frac{1}{nb_1^2} \sum_{i=1}^n \int_0^u l' \Xi_0(\hat{h}_0(v))^{-1} s_0(Y_i, T_i, R_i; g_\gamma(\hat{h}_0(v), e), \hat{h}_0(v)) dv, \\ \text{II} &= \frac{1}{nb_1^2} \sum_{i=1}^n \int_0^u l' \Xi_1(\hat{h}_1(v))^{-1} s_1(Y_i, T_i, R_i; g_\gamma(\hat{h}_1(v), e), \hat{h}_1(v)) dv. \end{aligned}$$

By symmetry, we only need to study the term I. Denote $\mathbf{1}_{\hat{h}_0} = \mathbf{1}\{\hat{h}_0 \in \mathcal{H}_0(\mathcal{P}_0^n)\}$ as the indicator of whether $\hat{h}_0 \in \mathcal{H}_0(\mathcal{P}_0^n)$. Then $\text{I} \leq \text{I.1} + \text{I.2}$, where

$$\begin{aligned} \text{I.1} &= \sup_{\gamma \in \Gamma, h_0 \in \mathcal{H}_0(\mathcal{P}_0^n), e \in \mathcal{E}, u \in (0, 1)} \left| \frac{1}{nb_1^2} \sum_{i=1}^n \int_0^u l' \Xi_0(h_0(v))^{-1} s_0(Y_i, T_i, R_i; g_\gamma(h_0(v), e), h_0(v)) dv \right|, \\ \text{I.2} &= (1 - \mathbf{1}_{\hat{h}_0}) \sup_{y \in \mathcal{Y}, t \in [t'_0, t''_0]} \left| \frac{1}{nb_1^2} \sum_{i=1}^n l' \Xi_0(t)^{-1} s_0(Y_i, T_i, R_i; y, t) \right| \end{aligned}$$

Define

$$\tilde{\alpha}_n = n^{-1/2} b_1^{-7/12 - \bar{\epsilon}/5}, \text{ and } \alpha_n = \left(b_1^2 + \sqrt{\log n / (nb_1^4)} \right) (b_1^2 + \tilde{\alpha}_n), \quad (\text{B.1})$$

where $\bar{\epsilon}$ is defined in Assumption 12. In Lemma B.5, we show that $\text{I.1} = O_p(\tilde{\alpha}_n)$. Combining Assumption 13(i) and Lemma B.7, we have $\text{I.2} = O_p(\tilde{\alpha}_n)$. Therefore,

$$\sup_{\gamma \in \Gamma} \left\| \hat{D}_{\gamma, \hat{h}} - D_{\gamma, \hat{h}} \right\|_w O_p(b_1^2 + \tilde{\alpha}_n). \quad (\text{B.2})$$

By the smoothness of $F_{Y, T, R}^-$ and g_γ , the following term is $O(\|\hat{h} - h^*\|_\infty)$:

$$\sup_{\gamma \in \Gamma} \left| \int_0^u \left(F_{Y|T, R}^-(g_\gamma(\hat{h}_0(v), e) | h_0(v), \bar{r}) - F_{Y|T, R}^-(g_\gamma(h_0^*(v), e) | h_0(v), \bar{r}) \right) dv \right|$$

Therefore, we obtain that uniformly over $\gamma \in \Gamma$,

$$\left\| D_{\gamma, \hat{h}} - D_{\gamma, h^*} \right\|_w = O(\|\hat{h} - h^*\|_\infty) = O_p \left(b_1^2 + \sqrt{\log n / (nb_1)} \right).$$

By the triangle inequality, we have

$$\left\| D_{\hat{\gamma}, h^*} \right\|_w \leq \left\| D_{\hat{\gamma}, h^*} - D_{\hat{\gamma}, \hat{h}} \right\|_w + \left\| \hat{D}_{\hat{\gamma}, \hat{h}} - D_{\hat{\gamma}, \hat{h}} \right\|_w + \left\| \hat{D}_{\hat{\gamma}, \hat{h}} \right\|_w.$$

By the definition of $\hat{\gamma}$ in (10), we have

$$\left\| \hat{D}_{\hat{\gamma}, \hat{h}} \right\|_w \leq \left\| \hat{D}_{\gamma^*, \hat{h}} \right\|_w + o_p(\alpha_n) \leq \left\| \hat{D}_{\gamma^*, \hat{h}} - D_{\gamma^*, \hat{h}} \right\|_w + \left\| D_{\gamma^*, \hat{h}} - D_{\gamma^*, h^*} \right\|_w + o_p(\alpha_n).$$

Combining the above two inequalities together, we obtain that

$$\left\| D_{\hat{\gamma}, h^*} \right\|_w \leq 2 \sup_{\gamma \in \Gamma} \left\| D_{\gamma, \hat{h}} - D_{\gamma, h^*} \right\|_w + 2 \sup_{\gamma \in \Gamma} \left\| \hat{D}_{\gamma, \hat{h}} - D_{\gamma, \hat{h}} \right\|_w + o_p(\alpha_n) = O_p \left(b_1^2 + \tilde{\alpha}_n \right). \quad (\text{B.3})$$

In particular, the above quantity is $o_p(1)$. Because Γ is compact, and $\|D_{\cdot, h^*}\|_w$ is continuous and has a unique minimizer γ^* (Corollary 2), for any $\epsilon > 0$ there exists $\delta > 0$ such that $\|\gamma - \gamma^*\|_2 > \epsilon \implies \|D_{\gamma, h^*}\|_w > \delta$. Therefore, $\mathbb{P}(\|\hat{\gamma} - \gamma^*\|_2 > \epsilon) \leq \mathbb{P}(\|D_{\hat{\gamma}, h^*}\|_w > \delta) = o(1)$. This proves that $\hat{\gamma}$ is a consistent estimator.

Step 3. (Convergence rate of $\hat{\gamma}$.) Since $\hat{\gamma}$ is consistent, we can Taylor expand $D_{\hat{\gamma}, h^*}$ around γ^* . Together with the reverse triangle inequality and the fact that $D_{\gamma^*, h^*} = 0$, the expansion gives that

$$\begin{aligned} \left\| D_{\hat{\gamma}, h^*} \right\|_w &= \left\| \nabla_{\gamma} D_{\gamma^*, h^*} (\hat{\gamma} - \gamma^*) + (\hat{\gamma} - \gamma^*)' \nabla_{\gamma}^2 D_{\tilde{\gamma}, h^*} (\hat{\gamma} - \gamma^*) \right\|_w \\ &\geq \left\| \nabla_{\gamma} D_{\gamma^*, h^*} (\hat{\gamma} - \gamma^*) \right\|_w - \|\hat{\gamma} - \gamma^*\|_2^2 \int_0^1 \int_{\mathcal{E}} \left\| \nabla_{\gamma}^2 D_{\tilde{\gamma}, h^*} (u, e) \right\|_2 w(e, u) de du \\ &\geq \left\| \nabla_{\gamma} D_{\gamma^*, h^*} (\hat{\gamma} - \gamma^*) \right\|_w + O(\|\hat{\gamma} - \gamma^*\|_2^2), \end{aligned}$$

where $\tilde{\gamma}$ is some point on the line segment connecting $\hat{\gamma}$ and γ^* and the last line follows from Assumption 10(iii) that $\left\| \nabla_{\gamma}^2 D_{\tilde{\gamma}, h^*} (u, e) \right\|_2$ is bounded. We claim that there exists a

universal constant $C > 0$ such that

$$\|\nabla_\gamma D_{\gamma^*, h^*} \zeta\|_w \geq C \|\zeta\|, \text{ for all } \zeta \in \mathbb{R}^{d_\Gamma}. \quad (\text{B.4})$$

If this claim is true, then by Equation (B.3), we obtain a bound on the convergence rate of $\hat{\gamma}$:

$$\|\hat{\gamma} - \gamma^*\|_2 = O_p(b_1^2 + \tilde{\alpha}_n).$$

The remaining part of this step is devoted to the proof of (B.4). Suppose that claim is false, then for each integer $k \geq 1$, there exists $\zeta_k \in \mathbb{R}^{d_\Gamma}$ such that $\|\nabla_\gamma D_{\gamma^*, h^*} \zeta_k\|_w < 1/k \|\zeta_k\|$. Without loss of generality, we can assume $\|\zeta_k\| = 1$ (or simply redefine the sequence as $\zeta_k / \|\zeta_k\|$). By the Bolzano–Weierstrass theorem, the sequence $\{\zeta_k\}$ has a convergent subsequence. Without loss of generality, we assume $\{\zeta_k\}$ itself is convergent with the limit denoted by ζ_∞ . Then it must be the case that $\|\nabla_\gamma D_{\gamma^*, h^*} \zeta_\infty\|_w = 0$. Since $\nabla_\gamma D_{\gamma^*, h^*}$ is a continuous function, the previous equation implies that $\nabla_\gamma D_{\gamma^*, h^*} \zeta_\infty = 0$. This violates Assumption 10(iv) that $\nabla_\gamma D_{\gamma^*, h^*}$ is a vector of linearly independent functions.

Step 4. (Stochastic equicontinuity of the criterion function.) Let $\gamma_n \xrightarrow{p} \gamma^*$ be such that $\|\gamma_n - \gamma^*\| = O_p(b_1^2 + \tilde{\alpha}_n)$. We want to find the asymptotic order of the term $\|\hat{D}_{\gamma_n, \hat{h}} - D_{\gamma_n, \hat{h}} - \hat{D}_{\gamma^*, h^*}\|_w$, which is bounded by

$$\sup_{e \in \mathcal{E}, u \in (0,1)} \left| \hat{D}_{\gamma_n, \hat{h}}(e, u) - D_{\gamma_n, \hat{h}}(e, u) - (\hat{D}_{\gamma^*, h^*}(e, u) - D_{\gamma^*, h^*}(e, u)) \right| \leq \text{I} + \text{II},$$

where

$$\begin{aligned} \text{I} = & \sup_{e \in \mathcal{E}, u \in (0,1)} \left| \hat{F}_{Y|T,R}^-(g_{\gamma_n}(\hat{h}_0(v), e) | \hat{h}_0(v), \bar{r}) - F_{Y|T,R}^-(g_{\gamma_n}(\hat{h}_0(v), e) | \hat{h}_0(v), \bar{r}) \right. \\ & \left. - \left(\hat{F}_{Y|T,R}^-(g_{\gamma^*}(h_0^*(v), e) | h_0^*(v), \bar{r}) - F_{Y|T,R}^-(g_{\gamma^*}(h_0^*(v), e) | h_0^*(v), \bar{r}) \right) \right|, \end{aligned}$$

and

$$\begin{aligned} \Pi = & \sup_{e \in \mathcal{E}, u \in (0,1)} \left| \hat{F}_{Y|T,R}^+(g_{\gamma_n}(\hat{h}_1(v), e) | \hat{h}_1(v), \bar{r}) - F_{Y|T,R}^-(g_{\gamma_n}(\hat{h}_1(v), e) | \hat{h}_1(v), \bar{r}) \right. \\ & \left. - \left(\hat{F}_{Y|T,R}^+(g_{\gamma^*}(h_1^*(v), e) | h_1^*(v), \bar{r}) - F_{Y|T,R}^+(g_{\gamma^*}(h_1^*(v), e) | h_1^*(v), \bar{r}) \right) \right|. \end{aligned}$$

By symmetry, we only need to study the term I. The uniform asymptotic linear representation of the LLR estimators gives a bias-variance decomposition that

$$I \leq I.1 + I.2 + O_p \left(b_1^3 + |\log b_1| / n b_1^2 \right),$$

where

$$\begin{aligned} I.1 &= b_1^2 \left(\mu_0(g_{\gamma_n}(\hat{h}_0(v), e), \hat{h}_0(v)) - \mu_0(g_{\gamma^*}(h_0^*(v), e), h_0^*(v)) \right), \\ I.2 &= \sup_{e \in \mathcal{E}, u \in [0,1]} \left| \frac{1}{n b_1^2} \sum_{i=1}^n l' \Xi_0(\hat{h}_0(v))^{-1} s_0(Y_i, T_i, R_i; g_{\gamma_n}(\hat{h}_0(v), e), \hat{h}_0(v)) \right. \\ &\quad \left. - l' \Xi_0(h_0^*(v))^{-1} s_0(Y_i, T_i, R_i; g_{\gamma^*}(h_0^*(v), e), h_0^*(v)) \right|. \end{aligned}$$

By the smoothness of μ_0 (Step 1) and g_γ (Assumption 10), we can bound the term I.1 by

$$I.1 \leq C b_1^2 (\|\gamma_n - \gamma^*\|_2 + \|\hat{h} - h^*\|_\infty) = O_p(b_1^4 + b_1^2 \tilde{\alpha}_n).$$

For the term I.2, consider the decomposition that $I.2 \leq I.2.1 + I.2.2$, where

$$\begin{aligned} I.2.1 &= \sup_{e \in \mathcal{E}, u \in [0,1]} \left| l' \Xi_0(\hat{h}_0(v))^{-1} \frac{1}{n b_1^2} \sum_{i=1}^n s_0(Y_i, T_i, R_i; g_{\gamma_n}(\hat{h}_0(v), e), \hat{h}_0(v)) \right. \\ &\quad \left. - s_0(Y_i, T_i, R_i; g_{\gamma^*}(h_0^*(v), e), h_0^*(v)) \right|, \\ I.2.2 &= \sup_{e \in \mathcal{E}, u \in [0,1]} \left| l' \left(\Xi_0(\hat{h}_0(v))^{-1} - \Xi_0(h_0^*(v))^{-1} \right) \frac{1}{n b_1^2} \sum_{i=1}^n s_0(Y_i, T_i, R_i; g_{\gamma^*}(h_0^*(v), e), h_0^*(v)) \right|. \end{aligned}$$

As mentioned in Step 1, we know that $\|\Xi(t)^{-1}\|_2$ is bounded for $t \in [t'_0, t''_0]$ by Lemma 1

in Xie (2021). Applying the mean value theorem, we obtain that

$$\text{I.2.1} \leq C(\text{I.2.1.1} + \text{I.2.1.2} + \text{I.2.1.3})$$

where

$$\begin{aligned} \text{I.2.1.1} &= \sup_{y \in \mathbb{R}, t \in [t'_0, t''_0]} \left\| \frac{1}{nb_1^2} \sum_{i=1}^n \frac{\partial}{\partial y} s_0(Y_i, T_i, R_i; y, t) \right\|_2 \|\nabla_\gamma g_{\gamma^*}\|_\infty \|\gamma_n - \gamma^*\|_2, \\ \text{I.2.1.2} &= \sup_{y \in \mathbb{R}, t \in [t'_0, t''_0]} \left\| \frac{1}{nb_1^2} \sum_{i=1}^n \frac{\partial}{\partial y} s_0(Y_i, T_i, R_i; y, t) \right\|_2 \left\| \frac{\partial}{\partial T} g_{\gamma^*} \right\|_\infty \|\hat{h} - h^*\|_\infty, \\ \text{I.2.1.3} &= \sup_{y \in \mathbb{R}, t \in [t'_0, t''_0]} \left\| \frac{1}{nb_1^2} \sum_{i=1}^n \frac{\partial}{\partial t} s_0(Y_i, T_i, R_i; y, t) \right\|_2 \|\hat{h} - h^*\|_\infty. \end{aligned}$$

In Lemma B.6, we show that the following two terms are of order $O_p\left(\sqrt{\log n / (nb_1^4)}\right)$:

$$\sup_{y \in \mathbb{R}, t \in [t'_0, t''_0]} \left| \frac{1}{nb_1^2} \sum_{i=1}^n \frac{\partial}{\partial y} s(Y_i, T_i, R_i; y, t) \right|, \quad \sup_{y \in \mathbb{R}, t \in [t'_0, t''_0]} \left| \frac{1}{nb_1^2} \sum_{i=1}^n \frac{\partial}{\partial t} s(Y_i, T_i, R_i; y, t) \right|.$$

Because $\|\nabla_\gamma g_{\gamma^*}\|_\infty$ and $\|\partial g_{\gamma^*} / \partial T\|_\infty$ are finite, we know that

$$\text{I.2.1} = O_p\left(\sqrt{\log n / (nb_1^4)}\right) \times (\|\gamma_n - \gamma^*\|_2 + \|\hat{h} - h^*\|_\infty) = O_p\left(\sqrt{\log n / (nb_1^4)} \tilde{a}_n\right).$$

Applying the mean value theorem to I.2.2, we obtain that

$$\text{I.2.2} \leq \sup_{t \in [t'_0, t''_0]} \left| t' \frac{\partial}{\partial t} \Xi_0(t)^{-1} \right| \sup_{y \in \mathbb{R}, t \in [t'_0, t''_0]} \left\| \frac{1}{nb_1^2} \sum_{i=1}^n s_0(Y_i, T_i, R_i; y, t) \right\|_2 \|\hat{h} - h^*\|_\infty.$$

In Lemma B.7, we show that

$$\sup_{y \in \mathbb{R}, t \in [t'_0, t''_0]} \left\| \sum_{i=1}^n s_0(Y_i, T_i, R_i; y, t) / (nb_1^2) \right\|_2 = O_p\left(\sqrt{\log n / (nb_1^2)}\right).$$

Therefore, I.2.1 asymptotically dominates I.2.2. Hence, the term I is of the following order:

$$I = O_p \left(\left(b_1^2 + \sqrt{\log n / (nb_1^4)} \right) \tilde{\alpha}_n \right) = O_p(\alpha_n).$$

Based on the same argument, the above asymptotic order also applies to the term II. Thus, we have the following stochastic equicontinuity result:

$$\|\hat{D}_{\gamma_n, \hat{h}} - D_{\gamma_n, \hat{h}} - \hat{D}_{\gamma^*, h^*}\|_w = O_p(\alpha_n). \quad (\text{B.5})$$

Step 5. (Linearization of the criterion function.) Let $\partial_h^{[\hat{h}-h^*]} D_{\gamma, h^*}(e, u)$ be the Fréchet derivative of $D_{\gamma, h}(e, u)$ with respect to h at h^* , in the direction of $h - h^*$. That is,

$$\partial_h^{[\hat{h}-h^*]} D_{\gamma, h^*}(e, u) = \int_0^u (\phi_\gamma^-(e, v) - \phi_\gamma^+(e, v))(\hat{h}_0(v) - h_0^*(v)) dv,$$

where

$$\begin{aligned} \phi_\gamma^-(e, v) &= \frac{\partial}{\partial Y} F_{Y|T, R}^-(g_\gamma(h_0^*(v), e) | h_0^*(v), \bar{r}) \frac{\partial}{\partial T} g_\gamma(h_0^*(v), e) + \frac{\partial}{\partial T} F_{Y|T, R}^-(g_\gamma(h_0^*(v), e) | h_0^*(v), \bar{r}), \\ \phi_\gamma^+(e, v) &= \frac{\partial}{\partial Y} F_{Y|T, R}^+(g_\gamma(h_1^*(v), e) | h_1^*(v), \bar{r}) \frac{\partial}{\partial T} g_\gamma(h_1^*(v), e) + \frac{\partial}{\partial T} F_{Y|T, R}^+(g_\gamma(h_1^*(v), e) | h_1^*(v), \bar{r}). \end{aligned}$$

It is straightforward to see that $\|\partial_h^{[\hat{h}-h^*]} D_{\gamma, h^*}\|_\infty = O(\|\hat{h} - h^*\|_\infty)$. Following the same steps as in Lemma 4 of [Torgovitsky \(2017\)](#), we can show that $\|D_{\gamma, \hat{h}} - D_{\gamma, h^*} - \partial_h^{[\hat{h}-h^*]} D_{\gamma, h^*}\|_w = O(\|\hat{h} - h^*\|_\infty^2)$, uniformly over $\gamma \in \Gamma$, and

$$\|\partial_h^{[\hat{h}-h^*]} D_{\gamma_n, h^*} - \partial_h^{[\hat{h}-h^*]} D_{\gamma^*, h^*}\|_w = O(\|\gamma_n - \gamma^*\|_2 \|\hat{h} - h^*\|_\infty),$$

for any sequence $\gamma_n \xrightarrow{p} \gamma^*$. Define

$$\hat{L}_\gamma(e, u) = \hat{D}_{\gamma^*, h^*}(e, u) + \nabla_\gamma D_{\gamma^*, h^*}(e, u)(\gamma - \gamma^*) + \partial_h^{[\hat{h}-h^*]} D_{\gamma^*, h^*}(e, u),$$

as a linear approximation of $\hat{D}_{\gamma, \hat{h}}(e, u)$ for γ near γ^* . For any sequence $\|\gamma_n - \gamma^*\|_2 =$

$O_p(b_1^2 + \tilde{\alpha}_n)$, we have

$$\|\hat{L}_{\gamma_n}\|_w \leq \|\hat{D}_{\gamma^*, h^*} - D_{\gamma^*, h^*}\|_w + \|\nabla_{\gamma} D_{\gamma^*, h^*}\|_w \|\gamma_n - \gamma^*\|_2 + O(\|\hat{h} - h^*\|_{\infty}) = O_p(b_1^2 + \tilde{\alpha}_n), \quad (\text{B.6})$$

where the asymptotic order of the first term on the RHS is derived in (B.2). We want to bound the approximation error from the linearization of the criterion function. By adding and subtracting terms, we obtain that

$$\begin{aligned} & \|\hat{L}_{\gamma_n} - \hat{D}_{\gamma_n, \hat{h}}\|_w \\ & \leq \|\hat{D}_{\gamma^*, h^*} - \hat{D}_{\gamma_n, \hat{h}} - (D_{\gamma^*, h^*} - D_{\gamma_n, \hat{h}})\|_w + \|D_{\gamma^*, h^*} + \nabla_{\gamma} D_{\gamma^*, h^*}(\gamma - \gamma^*) - D_{\gamma_n, h^*}\|_w \\ & \quad + \|D_{\gamma_n, h^*} + \partial_h^{[\hat{h} - h^*]} D_{\gamma_n, h^*} - D_{\gamma_n, \hat{h}}\|_w + \|\partial_h^{[\hat{h} - h^*]} D_{\gamma^*, h^*} - \partial_h^{[\hat{h} - h^*]} D_{\gamma_n, h^*}\|_w. \end{aligned} \quad (\text{B.7})$$

The four terms on the RHS of the above inequality can be analyzed as the following. The order of the first term on the RHS of (B.7) is given by (B.5) in the previous step. The second term is $O(\|\gamma_n - \gamma^*\|_2)$ by the smoothness of D_{γ, h^*} . The third term is bounded by

$$\sup_{\gamma \in \Gamma} \|D_{\gamma, \hat{h}} - D_{\gamma, h^*} - \partial_h^{[\hat{h} - h^*]} D_{\gamma, h^*}\|_w = O(\|\hat{h} - h^*\|_{\infty}^2).$$

The fourth term is $O(\|\gamma_n - \gamma^*\|_2 \|\hat{h} - h^*\|_{\infty})$. Therefore, the leading term on the RHS of (B.7) is the first term, and hence the approximation error from the linearization of the criterion function is of the following order:

$$\|\hat{L}_{\gamma_n} - \hat{D}_{\gamma_n, \hat{h}}\|_w = O_p(\alpha_n). \quad (\text{B.8})$$

Step 6. (Minimizer of the linearized criterion function.) Define $\tilde{\gamma}$ as the minimizer of $\|\hat{L}_{\gamma}\|_w$. The first-order condition gives that

$$\Delta(\tilde{\gamma} - \gamma^*) = \int \nabla_{\gamma} D_{\gamma^*, h^*}(e, u) \left(\hat{D}_{\gamma^*, h^*}(e, u) + \partial_h^{[\hat{h} - h^*]} D_{\gamma^*, h^*}(e, u) \right) w(e, u) de du,$$

where

$$\Delta = \int \nabla_{\gamma} D_{\gamma^*, h^*}(e, u) \nabla_{\gamma} D_{\gamma^*, h^*}(e, u)' w(e, u) dedu. \quad (\text{B.9})$$

By the uniform asymptotic linear representation of the LLR estimators and \hat{h} , we can write

$$\begin{aligned} & \int \nabla_{\gamma} D_{\gamma^*, h^*}(e, u) \left(\hat{D}_{\gamma^*, h^*}(e, u) + \partial_h^{[\hat{h} - h^*]} D_{\gamma^*, h^*}(e, u) \right) w(e, u) dedu \\ &= \frac{1}{nb_1} \sum_{i=1}^n (\zeta_{-}^{DF}(Y_i, T_i, R_i) + \zeta_{-}^Q(Y_i, T_i, R_i)) - \frac{1}{nb_1} \sum_{i=1}^n (\zeta_{+}^{DF}(Y_i, T_i, R_i) + \zeta_{+}^Q(Y_i, T_i, R_i)) \\ &+ b_1^2(B_{-} - B_{+}) + O_p(b_1^3) + o_p(1/\sqrt{nb_1}). \end{aligned}$$

The terms B_{-} and B_{+} are deterministic bias terms defined by

$$B_{-} = \int w(e, u) \nabla_{\gamma} D_{\gamma^*, h^*}(e, u) \left(\int_0^u \mu_0(g_{\gamma^*}(h_0^*(v), e), h_0^*(v)) + \phi_{\gamma^*}^{-}(e, v) \nu_0(v) \right) dedu, \quad (\text{B.10})$$

$$B_{+} = \int w(e, u) \nabla_{\gamma} D_{\gamma^*, h^*}(e, u) \left(\int_0^u \mu_1(g_{\gamma^*}(h_1^*(v), e), h_1^*(v)) + \phi_{\gamma^*}^{+}(e, v) \nu_1(v) \right) dedu. \quad (\text{B.11})$$

The functions ζ_{-}^{DF} and ζ_{+}^{DF} represent stochastic terms from the LLR estimation of the conditional distribution $F_{Y|T, R}$. They are defined by

$$\begin{aligned} & \zeta_{-}^{DF}(Y, T, R) \\ &= \frac{1}{b_1} \int_{\mathcal{E}} \int_0^1 w(e, u) \nabla_{\gamma} D_{\gamma^*, h^*}(e, u) \int_0^u l' \Xi_0(h_0^*(v))^{-1} s_0(Y, T, R, g_{\gamma^*}(h_0^*(v), e), h_0^*(v)) dv dedu \\ &= \int_{\mathcal{E}} \int_0^1 w(e, u) \nabla_{\gamma} D_{\gamma^*, h^*}(e, u) \int_{(T-h_0^*(u))/b_1}^{(T-t_0^*)/b_1} l' \Xi_0(T + b_1 v)^{-1} (1, v, (R - \bar{r})/b_1)' \\ &\quad \times \tilde{K}_Y(Y, T, R; g_{\gamma}(T + b_1 v, e)) k_T(v) k_R^{-}((R - \bar{r})/b_1) ((h_0^*)^{-1})'(T + b_1 v) dv, \end{aligned}$$

and

$$\begin{aligned}
& \zeta_+^{DF}(Y, T, R) \\
&= \frac{1}{b_1} \int_{\mathcal{E}} \int_0^1 w(e, u) \nabla_{\gamma} D_{\gamma^*, h^*}(e, u) \int_0^u \iota' \Xi_1(h_1^*(v))^{-1} s_1(Y_i, T_i, R_i, g_{\gamma^*}(h_1^*(v), e), h_1^*(v)) dv dedu \\
&= \int_{\mathcal{E}} \int_0^1 w(e, u) \nabla_{\gamma} D_{\gamma^*, h^*}(e, u) \int_{(T-h_1(u))/b_1}^{(T-t'_1)/b_1} \iota' \Xi_1(T + b_1 v)^{-1} (1, v, (r - \bar{r})/b_1)' \\
&\quad \times \tilde{K}_Y(Y, T, R; g_{\gamma^*}(T + b_1 v, e)) k_T(v) k_R^+((R - \bar{r})/b_1) ((h_0^*)^{-1})'(T + b_1 v) dv.
\end{aligned}$$

In the above notations, $k_R^-(x) = k_R(x) \mathbb{1}\{x < 0\}$ and $k_R^+(x) = k_R(x) \mathbb{1}\{x \geq 0\}$. Similarly define $k_{Q,0}^-(x) = k_{Q,0}(x) \mathbb{1}\{x < 0\}$ and $k_{Q,1}^+(x) = k_{Q,1}(x) \mathbb{1}\{x \geq 0\}$. The functions ζ_-^Q and ζ_+^Q represent stochastic terms from the nonparametric estimation of the conditional quantile function h^* . They are defined by

$$\begin{aligned}
\zeta_-^Q(T, R) &= \int_{\mathcal{E}} \int_0^1 w(e, u) \nabla_{\gamma} D_{\gamma^*, h^*}(e, u) \int_0^u \phi_{\gamma^*}^-(e, v) q_0(T, R; v) k_{Q,0}^-((R - \bar{r})/b_1) dv dedu, \\
\zeta_+^Q(T, R) &= \int_{\mathcal{E}} \int_0^1 w(e, u) \nabla_{\gamma} D_{\gamma^*, h^*}(e, u) \int_0^u \phi_{\gamma^*}^+(e, v) q_1(T, R; v) k_{Q,1}^+((R - \bar{r})/b_1) dv dedu.
\end{aligned}$$

By Fubini's theorem, we have

$$\mathbb{E}[\zeta_{\pm}^{DF}(Y, T, R) | T, R] = \mathbb{E}[\zeta_{\pm}^Q(T, R) | T, R] = 0.$$

Notice that $((h_0^*)^{-1})'(\cdot) = f_{T|R}^-(\cdot | \bar{r})$ and $((h_1^*)^{-1})'(\cdot) = f_{T|R}^+(\cdot | \bar{r})$. The variance matrix can be computed as follows, where to save space, we use the notation of squaring a vector to mean the tensor product of that vector with itself.

$$\begin{aligned}
& \mathbb{E} \left[(\zeta_-^{DF}(Y, T, R) + \zeta_-^Q(T, R)) \otimes (\zeta_-^{DF}(Y, T, R) + \zeta_-^Q(T, R)) \right] \\
&= \int_{\mathcal{Y} \times [t'_0, t''_0] \times [r_0, \bar{r}]} \left(\int_{\mathcal{E}} \int_0^1 w(e, u) \nabla_{\gamma} D_{\gamma^*, h^*}(e, u) \int_{(t-h_0(u))/b_1}^{(t-t'_0)/b_1} \iota' \Xi_0(t + b_1 v)^{-1} (1, v, (r - \bar{r})/b_1)' \right. \\
&\quad \times \tilde{K}_Y(y, t, r; g_{\gamma^*}(t + b_1 v, e)) k_T(v) k_R^-((r - \bar{r})/b_1) f_{T|R}^-(t + b_1 v | \bar{r}) dv, \\
&\quad \left. + \int_0^u \phi_{\gamma^*}^-(e, v) q_0(t, r; v) k_{Q,0}^-((r - \bar{r})/b_1) dv dedu \right)^2 f_{Y,T,R}^-(y, t, r) dy dt dr,
\end{aligned}$$

where $f_{Y,T,R}^{\pm}$ is defined analogously as $f_{T|R}^{\pm}$ and $F_{Y,T,R}^{\pm}$. Applying the change of variables

$\tilde{r} = (r - \bar{r})/b_1$, we obtain that the above matrix is equal to b_1 times the matrix

$$\begin{aligned} & \int_{\mathcal{Y} \times [t'_0, t''_0] \times [-1, 0]} \left(\int_{\mathcal{E}} \int_0^1 w(e, u) \nabla_{\gamma} D_{\gamma^*, h^*}(e, u) \int_{(t-h_0(u))/b_1}^{(t-t'_0)/b_1} t' \Xi_0(t + b_1 v)^{-1} (1, v, \tilde{r})' \right. \\ & \times \tilde{K}_Y(y, t, \tilde{r} + b_1 \tilde{r}; g_{\gamma^*}(t + b_1 v, e)) k_T(v) k_R^-(\tilde{r}) f_{T|R}^-(t + b_1 v | \tilde{r}) dv \\ & \left. + \int_0^u \phi_{\gamma^*}^-(e, v) q_0(t, \tilde{r} + b_1 \tilde{r}; v) k_{Q,0}^-(\tilde{r}) dv dedu \right)^2 f_{Y,T,R}^-(y, t, \tilde{r} + b_1 \tilde{r}) dy dt d\tilde{r}. \end{aligned}$$

For any $t \in [t'_0 + b_1, t''_0 - b_1]$, we have $\Xi_0(t) = f_{T,R}^-(t, \bar{r}) \bar{\Omega}_0$ with $\bar{\Omega}_0 = \int x x' k_0(x) dx_1 dx_2$. By letting $n \rightarrow \infty$ (so that $b_1 \rightarrow 0$) and using the continuity of the relevant functions and the dominated convergence theorem, we know that the above matrix is asymptotically equivalent to

$$\begin{aligned} \Sigma_- = & \int \left(\int_{\mathcal{E}} \int_0^1 w(e, u) \nabla_{\gamma} D_{\gamma^*, h^*}(e, u) \left(t' \bar{\Omega}_0^{-1} (1, 0, \tilde{r})' (\mathbf{1}\{y \leq g_{\gamma^*}(t, e)\} \right. \right. \\ & - F_{Y|T,R}^-(g_{\gamma^*}(t, e) | t, \bar{r})) k_R^-(\tilde{r}) / f_R(\bar{r}) \\ & \left. \left. + \int_0^u \phi_{\gamma^*}^-(e, v) q_0(t, \bar{r}; v) k_{Q,0}^-(\tilde{r}) dv \right) dedu \right)^2 f_{Y,T,R}^-(y, t, \bar{r}) dy dt d\tilde{r}. \end{aligned} \quad (\text{B.12})$$

In particular, we have used the following convergence result in the above expression:

$$\begin{aligned} & K_Y((y - y')/b_1) \rightarrow \mathbf{1}\{y' \leq y\}, \\ & \tilde{K}_Y(y, t, \bar{r} + b_1 \tilde{r}; g_{\gamma^*}(t + b_1 v, e)) \rightarrow \mathbf{1}\{y \leq g_{\gamma^*}(t, e)\} - F_{Y|T,R}^-(g_{\gamma^*}(t, e) | t, \bar{r}). \end{aligned}$$

The above derivation shows that

$$\mathbb{E} \left[(\zeta_-^{DF}(Y, T, R) + \zeta_-^Q(T, R)) \otimes (\zeta_-^{DF}(Y, T, R) + \zeta_-^Q(T, R)) \right] \sim b_1 \Sigma_-.$$

Similarly, we can show that

$$\mathbb{E} \left[(\zeta_+^{DF}(Y, T, R) + \zeta_+^Q(T, R)) \otimes (\zeta_+^{DF}(Y, T, R) + \zeta_+^Q(T, R)) \right] \sim b_1 \Sigma_+.$$

where

$$\begin{aligned}\Sigma_+ = & \int \left(\int_{\mathcal{E}} \int_0^1 w(e, u) \nabla_{\gamma} D_{\gamma^*, h^*}(e, u) \left(\iota' \bar{\Omega}_1^{-1}(1, 0, \tilde{r})' (\mathbf{1}\{y \leq g_{\gamma^*}(t, e)\} \right. \right. \\ & - F_{Y|T, R}^+(g_{\gamma^*}(t, e) | t, \tilde{r})) k_R^+(\tilde{r}) / f_R(\tilde{r}) \\ & \left. \left. + \int_0^u \phi_{\gamma^*}^+(e, v) q_1(t, \tilde{r}; v) k_{Q, 1}^+(\tilde{r}) dv \right) dedu \right)^2 f_{Y, T, R}^+(y, t, \tilde{r}) dy dt d\tilde{r},\end{aligned}\quad (\text{B.13})$$

and $\bar{\Omega}_1 = \int \mathbf{x} \mathbf{x}' k_1(\mathbf{x}) d\mathbf{x}_1 d\mathbf{x}_2$. The terms $\zeta_-^{DF}(Y, T, R)$ and $\zeta_-^Q(T, R)$ contain the factor $\mathbf{1}\{R < 0\}$ while the terms $\zeta_+^{DF}(Y, T, R)$ and $\zeta_+^Q(T, R)$ contain the factor $\mathbf{1}\{R \geq 0\}$. Hence, we can compute the variance matrix of their sum as

$$\text{var} \left((\zeta_-^{DF}(Y, T, R) + \zeta_-^Q(T, R)) - (\zeta_+^{DF}(Y, T, R) + \zeta_+^Q(T, R)) \right) = \Sigma_- + \Sigma_+.$$

Since Σ_- and Σ_+ do not vary with n , Chebyshev's inequality implies that the following term is $O_p(1/\sqrt{nb_1})$:

$$\frac{1}{nb_1} \sum_{i=1}^n \left(\zeta_-^{DF}(Y_i, T_i, R_i) + \zeta_-^Q(T_i, R_i) \right) - \frac{1}{nb_1} \sum_{i=1}^n \left(\zeta_+^{DF}(Y_i, T_i, R_i) + \zeta_+^Q(T_i, R_i) \right).$$

Moreover, \mathcal{E} is compact and the relevant functions in the expressions of ζ_{\pm}^{DF} and ζ_{\pm}^Q are bounded (Assumptions 8, 9, 10, and 13). We can apply the Lyapnov's central limit theorem (for example, Theorem 5.11 in [White \(2001\)](#)) to obtain that

$$\begin{aligned}& (\sqrt{nb_1}(\Sigma_- + \Sigma_+)^{-1/2}) \left(\frac{1}{nb_1} \sum_{i=1}^n \left(\zeta_-^{DF}(Y_i, T_i, R_i) + \zeta_-^Q(T_i, R_i) \right) \right. \\ & \left. - \frac{1}{nb_1} \sum_{i=1}^n \left(\zeta_+^{DF}(Y_i, T_i, R_i) + \zeta_+^Q(T_i, R_i) \right) \right) \xrightarrow{d} N(0, \mathbf{I}_{d_{\Gamma}}).\end{aligned}$$

Therefore, we obtain, for $\tilde{\gamma}$, the convergence rate: $\|\tilde{\gamma} - \gamma^*\|_2 = O_p(b_1^2 + 1/\sqrt{nb_1})$, and asymptotic normality:

$$(\sqrt{nb_1}(\Sigma_- + \Sigma_+)^{-1/2})(\Delta(\tilde{\gamma} - \gamma^*) - b_1^2(B_- - B_+)) \xrightarrow{d} N(0, \mathbf{I}_{d_{\Gamma}}), \quad (\text{B.14})$$

under the condition that $nb_1^7 \rightarrow 0$ (Assumption 12).

Step 7. (Asymptotic normality of $\hat{\gamma}$.) By Equation (B.8), we can apply the triangle inequality repeatedly and obtain that

$$\|\hat{L}_{\hat{\gamma}}\|_w \leq \|\hat{Q}_{\hat{\gamma}, \hat{h}}\|_w + O_p(\alpha_n) \leq \|\hat{Q}_{\tilde{\gamma}, \hat{h}}\|_w + O_p(\alpha_n) \leq \|\hat{L}_{\tilde{\gamma}}\|_w + O_p(\alpha_n),$$

where the second inequality uses the definition of $\hat{\gamma}$ in (10). Squaring the above inequality and using (B.6) to bounded $\|\hat{L}_{\tilde{\gamma}}\|_w$, we obtain that

$$\begin{aligned} \|\hat{L}_{\hat{\gamma}}\|_w^2 &\leq \|\hat{L}_{\tilde{\gamma}}\|_w^2 + O_p(\alpha_n^2) + O_p(\alpha_n \tilde{\alpha}_n) \leq \|\hat{L}_{\tilde{\gamma}}\|_w^2 + O_p(\alpha_n \tilde{\alpha}_n) \\ &= \|\hat{L}_{\tilde{\gamma}}\|_w^2 + O_p\left(\left(b_1^2 + \sqrt{\log n / (nb_1^4)}\right) \left(b_1^4 + n^{-1}b_1^{-7/6-2\bar{\epsilon}/5}\right)\right). \end{aligned}$$

Thus, we have

$$\|\hat{L}_{\hat{\gamma}}\|_w^2 - \|\hat{L}_{\tilde{\gamma}}\|_w^2 = O_p\left(b_1^6 + n^{-1}b_1^{5/6-2\bar{\epsilon}/5} + b_1^2\sqrt{\log n / n} + \sqrt{\log nn}^{-3/2}b_1^{-3\frac{1}{6}-2\bar{\epsilon}/5}\right).$$

We want to show that the four terms inside the O_p -notation in the above equation is $o(1/(nb_1))$. Both the terms b_1^6 and $b_1^2\sqrt{\log n / n}$ are $o(1/(nb_1))$ under Assumption 12(ii). The term $n^{-1}b_1^{5/6-2\bar{\epsilon}/5}$ is $o(1/(nb_1))$ since $b_1 = o(1)$. For the fourth term, we have

$$\sqrt{\log nn}^{-3/2}b_1^{-3\frac{1}{6}-2\bar{\epsilon}/5} = o(1/(nb_1)) \iff nb_1^{4\frac{1}{3}+\bar{\epsilon}}b_1^{-\bar{\epsilon}/5} / \log n \rightarrow \infty,$$

where the statement on the RHS is true by Assumption 12(iii). The above derivations show that

$$\|\hat{L}_{\hat{\gamma}}\|_w^2 - \|\hat{L}_{\tilde{\gamma}}\|_w^2 = o_p(b_1^2 + 1/(nb_1)).$$

By adding and subtracting $(\tilde{\gamma} - \gamma^*)\nabla_{\gamma}D_{\gamma^*, h^*}$, we obtain that

$$\|\hat{L}_{\hat{\gamma}}\|_w^2 = \|\tilde{L}_{\tilde{\gamma}}\|_w^2 + \|(\hat{\gamma} - \tilde{\gamma})\nabla_{\gamma}D_{\gamma^*, h^*}\|_w^2 + 2(\hat{\gamma} - \tilde{\gamma}) \int \tilde{L}_{\tilde{\gamma}}(e, u) \nabla_{\gamma}D_{\gamma^*, h^*}(e, u) w(e, u) dedu.$$

The last term (the inner product term) above is zero because $\tilde{L}_{\tilde{\gamma}}$ is orthogonal to $\nabla_{\gamma}D_{\gamma^*, h^*}$ from the projection perspective. This can also be verified by using the definition of $\tilde{\gamma}$. Hence, we have $\|(\hat{\gamma} - \tilde{\gamma})\nabla_{\gamma}D_{\gamma^*, h^*}\|_w^2 = o_p(1/(nb_1))$. By the same argument as in Step 3,

we can show that $\|\hat{\gamma} - \tilde{\gamma}\|_2 = o_p(1/\sqrt{nb_1})$. Therefore, by (B.14) and Slutsky's theorem, we obtain the desired asymptotic distribution of $\hat{\gamma}$:

$$\begin{aligned} & (\sqrt{nb_1}(\Sigma_- + \Sigma_+)^{-1/2})(\Delta(\hat{\gamma} - \gamma^*) - b_1^2(B_- - B_+)) \\ &= (\sqrt{nb_1}(\Sigma_- + \Sigma_+)^{-1/2})(\Delta(\tilde{\gamma} - \gamma^*) - b_1^2(B_- - B_+)) + o_p(1) \xrightarrow{d} N(0, \mathbf{I}_{d_T}). \end{aligned}$$

□

Proof of Proposition 1. We only prove the results for $\hat{h}_0(\bar{r}, \cdot)$ since the results for $\hat{h}_1(\bar{r}, \cdot)$ can be proved analogously. For part (i) of Assumption 13, we can set the partition \mathcal{P}_0^n to be the class of intervals $\{[u_j, u_{j+1}] : j = 0, \dots, J_n\}$. The estimator $\hat{h}_0(\bar{r}, \cdot)$ is a linear function within each interval and hence is contained in the class $\mathcal{H}_0^n(\mathcal{P}_0^n)$.

For part (ii), notice that, under Assumption 9(i), the estimator $\hat{h}_0(\bar{r}, u_0)$ and $\hat{h}_0(\bar{r}, u_{J_n+1})$ converge to t'_0 and t''_0 , respectively, at the $1/n$ rate. Therefore, we can replace $\hat{h}_0(\bar{r}, u_0)$ by t'_0 and $\hat{h}_0(\bar{r}, u_{J_n+1})$ by t''_0 without affecting the asymptotics. Let $\tilde{h}_0(\bar{r}, u)$ denote the solution of (11) at given u . The uniform asymptotic linear representation for $\hat{h}_0(\bar{r}, u), u \in (0, 1)$ follows from Lemma 3 in the Appendix of Dong et al. (2021), which is a slight modification of Theorem 1.2 of Qu and Yoon (2015). Then we can use Step 2 in the proof of Theorem 2 in Qu and Yoon (2015) to show that the error induced by linear interpolation is asymptotically negligible.

The uniform convergence rate in Part (iii) of Assumption 13 can be shown by using the uniform asymptotic linear representation. Since ν_0 is bounded, the bias term is $O(b_1^2)$. In Lemma B.8, we show that the stochastic term is satisfies

$$\sup_{u \in [0, 1]} \left| \frac{1}{nb_1} \sum_{i=1}^n q_0(T_i, R_i; u) k_{Q,0} \left(\frac{R_i - \bar{r}}{b_1} \right) \mathbf{1}\{R_i < \bar{r}\} \right| = O_p \left(\sqrt{\log n / nb_1} \right).$$

This proves the desired result.

□

B.2 Uniform convergence rates and the empirical process theory

Below are some basic concepts and results from the empirical process theory which are used to prove several uniform convergence results.

Let \mathcal{F} be a class of uniformly bounded measurable matrix-valued functions, that is, there exists $M > 0$ such that, for all $f \in \mathcal{F}$, $\|f\|_2 \leq M$. Let $N(\mathcal{F}, P, \epsilon)$ be the ϵ -covering number of the metric space $(\mathcal{F}, L_2(P))$, that is, $N(\mathcal{F}, P, \epsilon)$ is defined as the minimal number of open $\|\cdot\|_{L_2(P)}$ -balls of radius ϵ and centers in \mathcal{F} required to cover \mathcal{F} .

We say that a uniformly bounded function class \mathcal{F} is *Euclidean* if there exists $A_1, A_2 > 0$ (that only depend on the uniform bound) such that for every probability measure P and every $\epsilon \in (0, 1]$, $N(\mathcal{F}, P, \epsilon) \leq A_1/\epsilon^{A_2}$. We say that a function class \mathcal{F} is *log-Euclidean* with coefficient $\rho \in (0, 1)$ if there exists $A > 0$ (that only depends on the uniform bound) such that for every probability measure P and every $\epsilon \in (0, 1]$, $\log N(\mathcal{F}, P, \epsilon) \leq A/\epsilon^{2\rho}$.

The above definition of Euclidean classes is introduced by [Nolan and Pollard \(1987\)](#). The same concept is also studied by [Giné and Guillou \(1999\)](#), but they refer to what we call “Euclidean” as “VC.” There is a slight difference that [Nolan and Pollard \(1987\)](#) use the L_1 -norm while [Giné and Guillou \(1999\)](#) use the L_2 -norm. We ignored the envelope in their definition since we only work with uniformly bounded \mathcal{F} . The following two lemmas demonstrates how to generate function classes that are Euclidean and log-Euclidean.

Lemma B.1. *Let \mathcal{F}_1 and \mathcal{F}_2 be uniformly bounded and Euclidean classes of functions. The following classes of functions are also uniformly bounded and Euclidean.*

- (i) $\mathcal{F}_1 \oplus \mathcal{F}_2 = \{f_1 + f_2 : f_1 \in \mathcal{F}_1, f_2 \in \mathcal{F}_2\}$.
- (ii) $\mathcal{F}_1 \mathcal{F}_2 = \{f_1 \cdot f_2 : f_1 \in \mathcal{F}_1, f_2 \in \mathcal{F}_2\}$.
- (iii) $\{\mathbb{E}[f_1(\cdot)|X] : f_1 \in \mathcal{F}_1\}$.
- (iv) $\{k((\cdot - x)/b) : x \in \mathbb{R}, b > 0\}$, where $k : \mathbb{R} \rightarrow \mathbb{R}$ is a function of bounded variation.

Proof of Lemma B.1. See Appendix B in [Xie \(2021\)](#). □

Lemma B.2. *Let \mathcal{F}_1 be a uniformly bounded and Euclidean class of functions and \mathcal{F}_2 be a uniformly bounded and log-Euclidean class of functions with coefficient ρ . Then $\mathcal{F}_1 \mathcal{F}_2$ is uniformly bounded and log-Euclidean with coefficient $\rho + \epsilon$ for any $\epsilon > 0$.*

The following two lemmas give the asymptotic order of the supremum of empirical processes generated by Euclidean and log-Euclidean classes, respectively.

Proof of Lemma B.2. This follows from the definition of Euclidean and log-Euclidean classes. \square

Lemma B.3. Let X_1, \dots, X_n be an iid sample of a random vector X in \mathbb{R}^d . Let \mathcal{G}_n be a sequence of classes of measurable real-valued functions defined on \mathbb{R}^d . Assume that there is a fixed uniformly bounded Euclidean class \mathcal{F} such that $\mathcal{F}_n \subset \mathcal{F}$ for all n . Let $\sigma_n^2 \geq \sup_{f \in \mathcal{F}_n} \mathbb{E}[f(X)^2]$. Then

$$\sup_{f \in \mathcal{F}_n} \left| \sum_{i=1}^n (f(X_i) - \mathbb{E}f(X_i)) \right| = O_p \left(\sqrt{n\sigma_n^2 |\log \sigma_n|} + |\log \sigma_n| \right).$$

In particular, if $n\sigma_n^2 / |\log \sigma_n| \rightarrow \infty$, then the above rate simplifies to $O_p \left(\sqrt{n\sigma_n^2 |\log \sigma_n|} \right)$.

Proof of Lemma B.3. This is Lemma 2 in Xie (2021). \square

Lemma B.4. Let X_1, \dots, X_n be an iid sample of a random vector X in \mathbb{R}^d . Let \mathcal{F}_n be a sequence of classes of measurable real-valued functions defined on \mathbb{R}^d . Assume that there is a fixed uniformly bounded log-Euclidean class \mathcal{F} with coefficient ρ such that $\mathcal{F}_n \subset \mathcal{F}$ for all n . Let $\sigma_n^2 = \sup_{f \in \mathcal{F}_n} \mathbb{E}[f(X)^2]$. Then

$$\sup_{f \in \mathcal{F}_n} \left| \sum_{i=1}^n (f(X_i) - \mathbb{E}f(X_i)) \right| = O_p \left(\sqrt{n}\sigma_n^{1-\rho} + n^{\rho/(1+\rho)} \right).$$

Proof of Lemma B.4. Let $M > 0$ be the uniform bound of \mathcal{F} . Since \mathcal{F} is log-Euclidean with coefficient ρ , there exists $A > 0$ such that $\log N(\mathcal{F}, P_n, \epsilon) \leq A/\epsilon^\rho$ for every $\epsilon \in (0, 1]$, where P_n is the empirical measure. Since each \mathcal{F}_n is contained in \mathcal{F} , the above result also holds when \mathcal{F} is replaced by \mathcal{F}_n . Denote $\text{Rad}_i, 1 \leq i \leq n$, as a sequence of iid Rademacher variables. By Equation (3.19) in Koltchinskii (2011) (which is a result of Theorem 3.12 in the same book), there exists a universal constant $C > 0$ such that

$$\mathbb{E} \sup_{f \in \mathcal{F}} \left| \sum_{i=1}^n \text{Rad}_i f(X_i) \right| \leq CA^\rho M^\rho \sqrt{n}\sigma_n^{1-\rho} \vee CA^{2\rho/(\rho+1)} Mn^{\rho/(1+\rho)} = O_p \left(\sqrt{n}\sigma_n^{1-\rho} + n^{\rho/(1+\rho)} \right).$$

Then the desired result follows from the usual symmetrization argument (for example, Theorem 2.1 in [Koltchinskii \(2011\)](#)) and Chebyshev's inequality. \square

The following three lemmas give uniform convergence results that are used in the proof of Theorem 2.

Lemma B.5. *Under the assumptions of Theorem 2, the following term is $O_p(\tilde{\alpha}_n)$:*

$$\sup_{e \in \mathcal{E}, u \in [0,1], \gamma \in \Gamma, h_0 \in \mathcal{H}_0(\mathcal{P}_0^n)} \left| \frac{1}{nb_1^2} \sum_{i=1}^n \int_0^u \iota' \Xi_0(h_0(v))^{-1} s_0(Y_i, T_i, R_i; g_\gamma(h_0(v), \bar{r}, e), h_0(v)) dv \right|.$$

Proof of Lemma B.5. Since \mathcal{P}_0^n is a finite partition, we can without loss of generality assume that \mathcal{P}_0^n only contains the whole interval $[t'_0, t''_0]$ so that there is effectively no partition. To simplify notation, we omit the term \mathcal{P}_0^n . By the change of variables $\tilde{v} = (T_i - h_0(v))/b_1$ and Fubini's theorem, we have

$$\begin{aligned} & \left| \frac{1}{nb_1^2} \sum_{i=1}^n \int_0^u \iota' \Xi_0(h_0(v))^{-1} s_0(Y_i, T_i, R_i; g_\gamma(h_0(v), \bar{r}, e), h_0(v)) dv \right| \\ & \leq \int \left| \frac{1}{nb_1} \sum_{i=1}^n \iota' \Xi_0(T_i + b_1 \tilde{v})^{-1} s_0(Y_i, T_i, R_i; g_\gamma(T_i + b_1 \tilde{v}, \bar{r}, e), T_i + b_1 \tilde{v}) \right. \\ & \quad \times (h_0^{-1})'(T_i + b_1 \tilde{v}) \mathbf{1}\{(T_i - h_0(u))/b_1 < \tilde{v} < (T_i - t'_0)/b_1\} d\tilde{v} \left. \right| \\ & \leq \sup_{\tilde{v} \in (-1,1)} \left| \frac{1}{nb_1} \sum_{i=1}^n \iota' \Xi_0(T_i + b_1 \tilde{v})^{-1} s_0(Y_i, T_i, R_i; g_\gamma(T_i + b_1 \tilde{v}, \bar{r}, e), T_i + b_1 \tilde{v}) \right. \\ & \quad \times (h_0^{-1})'(T_i + b_1 \tilde{v}) \mathbf{1}\{(T_i - h_0(u))/b_1 < \tilde{v} < (T_i - t'_0)/b_1\} \left. \right|, \end{aligned}$$

where, in the last inequality, the supremum is taken over $\tilde{v} \in (-1,1)$ because of the support of k_T . Define the following function of (Y, T, R) indexed by (v, u, e, γ, h_0) :

$$\begin{aligned} \psi_n(Y, T, R; v, u, e, \gamma, h_0) &= \iota' \Xi_0(T + b_1 v)^{-1} s_0(Y, T, R; g_\gamma(T + b_1 v, \bar{r}, e), T + b_1 v) \\ &\quad \times (h_0^{-1})'(T + b_1 v) \mathbf{1}\{(T - h_0(u))/b_1 < v < (T - t'_0)/b_1\}. \end{aligned}$$

Let $\Psi_n = \{\psi_n(\cdot, \cdot, \cdot; v, u, e, \gamma, h) : v \in (-1,1), u \in (0,1), e \in \mathcal{E}, \gamma \in \Gamma, h_0 \in \mathcal{H}_0\}$. Our goal

is to use empirical process theory to derive the asymptotic order of

$$\sup_{\psi_n \in \Psi_n} \left| \sum_{i=1}^n \psi_n(Y, T, R; v, u, e, \gamma, h_0) \right|.$$

Consider a larger class Ψ as the product $\Psi = \Psi_{\Xi_0} \Psi_Y \Psi_{TR} \Psi_{\mathcal{H}_0}$, where

$$\begin{aligned} \Psi_{\Xi_0} &= \{T \mapsto l' \Xi_0(T + v)^{-1} : v \in (-1, 1)\}, \\ \Psi_Y &= \{(Y, T, R) \mapsto \tilde{K}_Y(Y, T, R; g_\gamma(T + v, \bar{r}, e), T + v) : v \in (-1, 1), \gamma \in \Gamma, e \in \mathcal{E}\}, \\ \Psi_{TR} &= \{(Y, T, R) \mapsto (1, v, (R - \bar{r})/b)' k_T(v) k_R^-((R - \bar{r})/b) \\ &\quad \times \mathbf{1}\{T - h_0(u) < bv < T - t'_0\} : b, u \in (0, 1), v \in (-1, 1)\}, \\ \Psi_{\mathcal{H}_0} &= \{T \mapsto (h_0^{-1})'(T + v) : h_0 \in \mathcal{H}_0, v \in (0, 1)\}. \end{aligned}$$

Notice that in the above definition of Ψ_{Ξ_0} , Ψ_Y , and $\Psi_{\mathcal{H}_0}$, omitting the parameter b does not change the class under consideration. The class Ψ does not vary with n and $\Psi_n \subset \Psi, n \geq 1$. In the following paragraphs, we show that the classes Ψ_{Ξ_0} , Ψ_Y , and Ψ_{TR} are Euclidean while the class $\Psi_{\mathcal{H}_0}$ is log-Euclidean.

For Ψ_{Ξ_0} , we know that $\|\Xi_0\|$ and $\|\Xi_0^{-1}\|$ are uniformly bounded by Lemma 1 in [Xie \(2021\)](#). By the smoothness of $f_{T,R}^-$ in Assumption 9, the class $\{T \mapsto l' \Xi_0(T + v) : v \in (-1, 1)\}$ is Lipschitz in the parameter $v \in (-1, 1)$ and hence, by Theorem 2.7.11 in [van der Vaart and Wellner \(1996\)](#), has covering numbers bounded by that of one-dimensional intervals. This implies that $\{T \mapsto l' \Xi_0(T + v) : v \in (-1, 1)\}$ is uniformly bounded and Euclidean. Then by Theorem 3 in [Andrews \(1994\)](#), we know that Ψ_{Ξ_0} is uniformly bounded and Euclidean.

The class Ψ_Y can be written as $\Psi_Y = \Psi_{Y1} + \Psi_{Y2}$, where

$$\begin{aligned} \Psi_{Y1} &= \{(Y, T) \mapsto K_Y((g_\gamma(T + v, \bar{r}, e) - Y)/b) : b, v \in (0, 1), \gamma \in \Gamma, e \in \mathcal{E}\}, \\ \Psi_{Y2} &= \{(Y, T, R) \mapsto -\mathbb{E}[K_Y((g_\gamma(T + v, \bar{r}, e) - Y)/b) | T, R] : b, v \in (0, 1), \gamma \in \Gamma, e \in \mathcal{E}\}. \end{aligned}$$

In view of Lemma B.1(i) and (iii), we only need to show that Ψ_Y is Euclidean. The class Ψ_Y is uniformly bounded by 1. The function K_Y is increasing since k_Y is positive. The

subgraph class of Ψ_Y can be written as

$$\begin{aligned} & \{ \{ (y, t, s) : K_Y((g_\gamma(t + v, \bar{r}, e) - y)/b) \leq s \} : b, v \in (0, 1), \gamma \in \Gamma, e \in \mathcal{E} \} \\ & = \{ \{ (y, t, s) : g_\gamma(t + v, \bar{r}, e) - y - bK^{-1}(s) \leq 0 \} : b, v \in (0, 1), \gamma \in \Gamma, e \in \mathcal{E} \}. \end{aligned}$$

By Assumption 10, the following function class is finite-dimensional:

$$\{ (t, y, s) \mapsto g_\gamma(t + v, \bar{r}, e) - y - bK^{-1}(s) : b, v \in (0, 1), \gamma \in \Gamma, e \in \mathcal{E} \}.$$

By Lemma 18(ii) in [Nolan and Pollard \(1987\)](#), the subgraph class of Ψ_Y is a polynomial class, which implies (by Theorem 2.6.7 in [van der Vaart and Wellner \(1996\)](#)) that Ψ_Y is Euclidean.

For the class Ψ_{TR} , notice that the function k_T , k_R , and the indicator function are all of bounded variation. The kernel functions k_T and k_R are supported on $[-1, 1]$. Therefore, the term $(R - \bar{r})/b$ is bounded between $[-1, 1]$. By Lemma B.1(ii) and (iv), we know that Ψ_{TR} is uniformly bounded and Euclidean.

Lastly, by Assumption 13(i), the class $\Psi_{\mathcal{H}_0}$ is contained in the class of twice continuously differentiable functions whose second-order derivatives are Lipschitz continuous. By the well-known bounds on the entropy of Lipschitz classes (see, for example, Example 5.11 in Chapter 5 of [Wainwright \(2019\)](#)), we know the class $\Psi_{\mathcal{H}_0}$ is log-Euclidean with coefficient $1/2 \times 1/(2 + 1) = 1/6$. Then by Lemma B.2, we know that Ψ is log-Euclidean with coefficient $1/6 + \epsilon$ for any small $\epsilon > 0$.

Next, we want to derive a uniform variance bound for the class Ψ and appeal to Lemma B.4. By the uniform boundedness of the classes studied above and applying the usual change of variables, we obtain that

$$\mathbb{E}[\psi_n(Y, T, R; v, u, e, \gamma, h_0)^2] \leq C \mathbb{E}[k_R((R - \bar{r})/b_1)^2] = C b_1 \int k_R(\tilde{r}) f_R(\bar{r} + b_1 \tilde{r}) d\tilde{r} = O(b_1).$$

Lemma B.4 then gives that

$$\sup_{\psi \in \Psi_n} \left| \sum_{i=1}^n \psi(Y, T, R; v, u, e, \gamma, h_0) \right| = O_p \left(n^{1/2} b_1^{(1-1/6)/2-\epsilon} + n^{1/7} \right),$$

for any small $\epsilon > 0$. Notice that, in the rate specified above, the term $n^{1/7}$ is dominated in view of Assumption 12. Then the desired convergence rate follows from dividing by nb_1 on both sides. \square

Lemma B.6. *Under the assumptions of Theorem 2, we have*

$$\begin{aligned} \sup_{y \in \mathbb{R}, t \in [t'_0, t''_0]} \left\| \frac{1}{nb_1^2} \sum_{i=1}^n \frac{\partial}{\partial y} s_0(Y_i, T_i, R_i; y, t) \right\|_2 &= O_p \left(\sqrt{\log n / (nb_1^4)} \right), \\ \sup_{y \in \mathbb{R}, t \in [t'_0, t''_0]} \left\| \frac{1}{nb_1^2} \sum_{i=1}^n \frac{\partial}{\partial t} s_0(Y_i, T_i, R_i; y, t) \right\|_2 &= O_p \left(\sqrt{\log n / (nb_1^4)} \right). \end{aligned}$$

Proof of Lemma B.6. The partial derivative of s_0 with respect to y is a vector of length three whose generic element can be denoted by $\dot{s}_0(Y_i, T_i, R_i; y, t, b_1, b_2) / b_2$, where

$$\dot{s}_0(Y, T, R; y, t) = \left(\frac{T-t}{b_1} \right)^{\ell_1} \left(\frac{R-\bar{r}}{b_1} \right)^{\ell_2} (k_Y((y-Y)/b_2) - \mathbb{E}[k_Y((y-Y)/b_2) | T, R]) k_0(X(t))$$

with $(\ell_1, \ell_2) = (0, 0), (1, 0), (0, 1)$. We use the empirical process theory to derive the uniform convergence rate of the sample average of \dot{s}_0 . Recall that the kernel functions k_Y, k_T , and k_R are of bounded variation. Then by Lemma B.1, we know that the following function class is uniformly bounded and Euclidean:

$$\begin{aligned} &\{(Y, T, R) \mapsto \dot{s}_0(Y, T, R; y, t, b, b') : y \in \mathbb{R}, t \in [t'_0, t''_0], b, b' > 0\} \\ &= \{(Y, T, R) \mapsto ((T-t)/b)^{\ell_1} ((R-\bar{r})/b)^{\ell_2} (k_Y((y-Y)/b') - \mathbb{E}[k_Y((y-Y)/b') | T, R]) \\ &\quad \times k_T((T-t)/b) k_R^-(R-\bar{r})/b : y \in \mathbb{R}, t \in [t'_0, t''_0], b, b' > 0\}. \end{aligned}$$

By the law of iterated expectations and differentiation under the integral, we know that \dot{s}_0 is centered. By using the fact that k_T and k_R are supported on $[-1, 1]$ and k_Y is bounded and applying the standard change of variables, we can bound the variance of \dot{s}_0 by

$$2 \|k_Y\|_\infty^2 \mathbb{E} \left[k_T((T-t)/b_1)^2 k_R^-(R-\bar{r})/b_1^2 \right] = b_1^2 2 \|k_Y\|_\infty^2 \int k_T(x_1)^2 k_R^-(x_2)^2 = O(b_1^2),$$

uniformly over $y \in \mathbb{R}$ and $t \in [t'_0, t''_0]$. Then by Lemma B.3, we know that the uniform

convergence rate of the sample average of \dot{s}_0 is $O_p\left(\sqrt{nb_1^2 \log n}\right)$. Therefore,

$$\sup_{y \in \mathbb{R}, t \in [t'_0, t''_0]} \left\| \frac{1}{nb_1^2} \sum_{i=1}^n \frac{\partial}{\partial y} s_0(Y_i, T_i, R_i; y, t) \right\|_2 = \frac{1}{nb_1^2 b_2} O_p\left(\sqrt{nb_1^2 \log n}\right) = O_p\left(\sqrt{\log n / (nb_1^4)}\right)$$

under the condition that $b_1/b_2 \in [1/C, C]$ (Assumption 12). This proves the first claim of the lemma. For the second claim, the same argument applies. We just want to point out that k_T is differentiable on the entire real line by Assumption 11 even though its support is $[-1, 1]$. □

Lemma B.7. *Under the assumptions of Theorem 2, we have*

$$\begin{aligned} \sup_{y \in \mathcal{Y}, t \in [t'_0, t''_0]} \left| \frac{1}{nb_1^2} \sum_{i=1}^n t' \Xi_0(t)^{-1} s_0(Y_i, T_i, R_i; y, t) \right| &= O_p\left(\sqrt{\log n / (nb_1^2)}\right), \\ \sup_{y \in \mathbb{R}, t \in [t'_0, t''_0]} \left\| \frac{1}{nb_1^2} \sum_{i=1}^n s_0(Y_i, T_i, R_i; y, t) \right\|_2 &= O_p\left(\sqrt{\log n / (nb_1^2)}\right). \end{aligned}$$

Proof of Lemma B.7. Following the same steps as in the proofs of the previous two lemmas, we can show that the relevant function classes are uniformly bounded and Euclidean. By the usual change of variables, we can show that the uniform variance bound is $O(b_1^2)$ before taking into account the factor $1/(nb_1)$ in the two terms. Then the desired results follow from Lemma B.3. The details are omitted for brevity. □

Lemma B.8.

$$\sup_{u \in [0, 1]} \left| \frac{1}{nb_1} \sum_{i=1}^n q_0(T_i, R_i; u) k_{Q,0} \left(\frac{R_i - \bar{r}}{b_1} \right) \mathbf{1}\{R_i < \bar{r}\} \right| = O_p\left(\sqrt{\log n / nb_1}\right).$$

Proof of Lemma B.8. Without loss of generality, let $c = 1$. Define

$$\begin{aligned} \psi_n(T, R; u) &= q_0(T, R; u) k_{Q,0} \left(\frac{R - \bar{r}}{b_1} \right) \mathbf{1}\{R < \bar{r}\} \\ &= \frac{u - \mathbf{1}\{T \leq h_0^*(\bar{r}, u)\}}{f_R(\bar{r}) f_{T|R}(h_0^*(\bar{r}, u) | \bar{r})} t' \Omega_{Q,0}^{-1}(1, (R - \bar{r})/b_1)' K_{FS}((R - \bar{r})/b_1) \end{aligned}$$

and $\Psi_n = \{(T, R) \mapsto \psi_n(T, R; u) : u \in [0, 1]\}$. By the law of iterated expectations, ψ_n is centered. Let $M = \sup_{u \in [0, 1]} |f_R(\bar{r}) f_{T|R}^-(h_0^*(\bar{r}, u) | \bar{r})|$. Define a product class $\Psi = \Psi_T \Psi_R$ where

$$\begin{aligned}\Psi_T &= \{(T, R) \mapsto C(u - \mathbf{1}\{T \leq t\}) : u \in [0, 1], t \in [t'_0, t''_0], |C| \leq M\}, \\ \Psi_R &= \{(T, R) \mapsto \iota' \Omega_{Q,0}^{-1}(1, (R - \bar{r})/b)' K_{FS}((R - \bar{r})/b) : b > 0\}.\end{aligned}$$

The class Ψ does not vary with n , and $\Psi_n \subset \Psi, n \geq 1$. The class Ψ_T is uniformly bounded and Euclidean since the set of indicator functions $\mathbf{1}\{T \leq t\}, t \in [t'_0, t''_0]$ is Euclidean. The class Ψ_R is uniformly bounded and Euclidean since K_{FS} is of bounded variation and compactly supported. By the usual change of variables, we can show that the uniform variance bound for Ψ_n is $O(b_1)$. Then the desired convergence rate follows from Lemma B.3.

□

B.3 Covariance Matrix Estimation

In this section, we discuss the estimation of the asymptotic variance matrix of $\hat{\gamma}$, which involves the estimation of Δ , Σ_- , and Σ_+ . For concreteness, we consider the first-step nonparametric conditional quantile estimation procedure described in Section 3.3 and Proposition 1. In the expressions of Δ , Σ_- , and Σ_+ , the functions that require estimation include $\nabla_\gamma D_{\gamma^*, h^*}$, $\phi_{\gamma^*}^\pm$, $f_{Y,T,R}^\pm$, $f_{T|R}^\pm$, and f_R . By definition,

$$\begin{aligned}\nabla_\gamma D_{\gamma^*}(e, u) &= \int_0^u \left[\frac{\partial}{\partial Y} F_{Y|T,R}^-(g_{\gamma^*}(h_0^*(\bar{r}, v), \bar{r}, e) | h_0^*(\bar{r}, v), \bar{r}) \nabla_\gamma g_{\gamma^*}(h_0^*(\bar{r}, v), \bar{r}, e) \right. \\ &\quad \left. - \frac{\partial}{\partial Y} F_{Y|T,R}^+(g_{\gamma^*}(h_1^*(\bar{r}, v), \bar{r}, e) | h_1^*(\bar{r}, v), \bar{r}) \nabla_\gamma g_{\gamma^*}(h_1^*(\bar{r}, v), \bar{r}, e) \right] dv.\end{aligned}$$

In the above quantity, we only need to estimate $\frac{\partial}{\partial Y} F_{Y|T,R}^- = f_{Y|T,R}^\pm$ since we already have estimators for γ^* and h^* . By observing the definition of $\phi_{\gamma^*}^\pm$, we know that the additional term that requires estimation is $\frac{\partial}{\partial T} F_{Y|T,R}^\pm$. To summarize, we want to estimate $f_{Y,T,R}^\pm$ and $\frac{\partial}{\partial T} F_{Y|T,R}^\pm$. Once $f_{Y,T,R}^\pm$ is obtained, we can operate to get the marginal and conditional density functions.

For estimation of $f_{Y,T,R}^\pm$, we can employ the method developed by [Cattaneo et al. \(2020\)](#). They use the second-order local polynomial regression to estimate the joint density. Due to the nature of local polynomial regressions, the estimator is boundary adaptive and particularly suitable for RD designs. To estimate the partial derivative $\frac{\partial}{\partial T} F_{Y|T,R}^\pm$, we can employ a second-order local polynomial regression. The procedure is similar to STEP 2 in the construction of $\hat{\gamma}$. We add two quadratic terms into the minimization problem:

$$\sum_{i: R_i < \bar{r}} \left(K_Y \left(\frac{y - Y_i}{b_2} \right) - a^- - a_T^-(T_i - t) - a_{T,2}^-(T_i - t)^2 - a_R^-(R_i - \bar{r}) - a_{R,2}^-(R_i - \bar{r})^2 \right)^2 \\ \times k_T \left(\frac{T_i - t}{b_1} \right) k_R \left(\frac{R_i - \bar{r}}{b_1} \right).$$

The minimizer \hat{a}_T^- is the estimate of $\frac{\partial}{\partial T} F_{Y|T,R}^-(y|t, \bar{r})$. The estimate of $\frac{\partial}{\partial T} F_{Y|T,R}^+(y|t, \bar{r})$ can be analogously constructed.

We assume that the resulting estimators $\hat{f}_{Y,T,R}^\pm$ and $\frac{\partial}{\partial T} \hat{F}_{Y|T,R}^\pm$ are uniformly consistent. That is,

$$\sup_{y,t,r} |\hat{f}^\pm(y, t, r) - f^\pm(y, t, r)| = o_p(1), \\ \sup_{y,t} \left| \frac{\partial}{\partial T} \hat{F}_{Y|T,R}^\pm(y|t, \bar{r}) - \frac{\partial}{\partial T} F_{Y|T,R}^\pm(y|t, \bar{r}) \right| = o_p(1).$$

Such uniform convergence results can be proved along the lines of, for examples, [Fan and Guerre \(2016\)](#) and [Xie \(2021\)](#). The details are omitted here. We can construct the following distributional estimates:

$$\hat{f}_{T,R}^\pm(t, \bar{r}) = \int \hat{f}_{Y,T,R}^\pm(y, t, \bar{r}) dy, \\ \hat{f}_{Y|T,R}^\pm(y, t|\bar{r}) = \hat{f}_{Y,T,R}^\pm(y, t, \bar{r}) / \hat{f}_{T,R}^\pm(t, \bar{r}).$$

Under the assumption that $f_{T,R}^\pm$ is bounded away from zero, the estimator $\hat{f}_{Y|T,R}^\pm(y, t|\bar{r})$ is

uniformly consistent. Let

$$\hat{\Delta} = \int w(e, u) \left(\int_0^u [\hat{f}_{Y|T,R}^-(g_{\hat{\gamma}}(\hat{h}_0(\bar{r}, v), \bar{r}, e) | \hat{h}_0(\bar{r}, v), \bar{r}) \nabla_{\gamma} g_{\hat{\gamma}}(\hat{h}_0(\bar{r}, v), \bar{r}, e) - \hat{f}_{Y|T,R}^+(g_{\hat{\gamma}}(\hat{h}_1(\bar{r}, v), \bar{r}, e) | \hat{h}_1(\bar{r}, v), \bar{r}) \nabla_{\gamma} g_{\hat{\gamma}}(\hat{h}_1(\bar{r}, v), \bar{r}, e)] dv \right)^2 dedu.$$

Under the uniform consistency of $\hat{f}_{Y|T,R}^{\pm}$ and \hat{h} and the consistency of $\hat{\gamma}$, we can show that $\hat{\Delta}$ is a consistent estimator of Δ . For the estimation of Σ_{-} , define

$$\begin{aligned} \hat{f}_R(\bar{r}) &= \int \hat{f}_{T,R}^{\pm}(t, \bar{r}) dt, \\ \hat{f}_{T|R}^{\pm}(t | \bar{r}) &= \hat{f}_{T,R}^{\pm}(t, \bar{r}) / \hat{f}_R(\bar{r}), \\ \hat{\phi}_{\hat{\gamma}}^-(e, v) &= \hat{f}_{Y|T,R}^-(g_{\hat{\gamma}}(\hat{h}_0(v), e) | \hat{h}_0(v), \bar{r}) \frac{\partial}{\partial T} g_{\hat{\gamma}}(\hat{h}_0(v), e) + \frac{\partial}{\partial T} \hat{F}_{Y|T,R}^-(g_{\hat{\gamma}}(\hat{h}_0(v), e) | \hat{h}_0(v), \bar{r}). \end{aligned}$$

The above estimators are also uniformly consistent. In particular,

$$\sup_{e, v} \left| \hat{\phi}_{\hat{\gamma}}^-(e, v) - \phi_{\gamma^*}^-(e, v) \right| = o_p(1).$$

Let

$$\begin{aligned} \hat{\Sigma}_{-} &= \int \left(\int_{\mathcal{E}} \int_0^1 w(e, u) \left(\int_0^u [\hat{f}_{Y|T,R}^-(g_{\hat{\gamma}}(\hat{h}_0(\bar{r}, v), \bar{r}, e) | \hat{h}_0(\bar{r}, v), \bar{r}) \nabla_{\gamma} g_{\hat{\gamma}}(\hat{h}_0(\bar{r}, v), \bar{r}, e) - \hat{f}_{Y|T,R}^+(g_{\hat{\gamma}}(\hat{h}_1(\bar{r}, v), \bar{r}, e) | \hat{h}_1(\bar{r}, v), \bar{r}) \nabla_{\gamma} g_{\hat{\gamma}}(\hat{h}_1(\bar{r}, v), \bar{r}, e)] dv \right) \right. \\ &\quad \left(l' \bar{\Omega}_0^{-1}(1, 0, \bar{r})' (\mathbf{1}\{y \leq g_{\hat{\gamma}}(t, e)\} - \hat{F}_{Y|T,R}^-(g_{\hat{\gamma}}(t, e) | t, \bar{r})) k_R^-(\bar{r}) / \hat{f}_R(\bar{r}) \right. \\ &\quad \left. \left. + \frac{k_{Q,0}^-(\bar{r})}{c \hat{f}_{T,R}^-(t, \bar{r})} \int_0^u \hat{\phi}_{\hat{\gamma}}^-(e, v) (v - \mathbf{1}\{t \leq \hat{h}_0(\bar{r}, v)\}) dv \right) dedu \right)^2 \hat{f}_{Y,T,R}^-(y, t, \bar{r}) dy dt d\bar{r}. \end{aligned}$$

Under the uniform consistency of \hat{f}_R , $\hat{f}_{T,R}^{\pm}$, $\hat{f}_{Y|T,R}^{\pm}$, $\hat{F}_{Y|T,R}$, $\hat{f}_{Y,T,R}^{\pm}$, $\hat{\phi}_{\hat{\gamma}}^+$, and \hat{h} , and the consistency of $\hat{\gamma}$, $\hat{\Sigma}_{-}$ is a consistent estimator of Σ_{-} . The estimation of Σ_{+} can be performed analogously.

References

- Almond, D., J. Doyle, Joseph J., A. E. Kowalski, and H. Williams (2010, 05). Estimating Marginal Returns to Medical Care: Evidence from At-risk Newborns*. *The Quarterly Journal of Economics* 125(2), 591–634.
- Andrews, D. W. (1994). Chapter 37 empirical process methods in econometrics. Volume 4 of *Handbook of Econometrics*, pp. 2247–2294. Elsevier.
- Angrist, J. D. and G. W. Imbens (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. *Journal of the American Statistical Association* 90(430), 431–442.
- Barr, A., J. Eggleston, and A. A. Smith (2021). The effect of income during infancy: Evidence from a discontinuity in tax benefits. Working paper, Texas A&M University.
- Barreca, A. I., M. Guldi, J. M. Lindo, and G. R. Waddell (2011, 10). Saving Babies? Revisiting the effect of very low birth weight classification*. *The Quarterly Journal of Economics* 126(4), 2117–2123.
- Beccuti, G. and S. Pannain (2011). Sleep and obesity. *Current opinion in clinical nutrition and metabolic care* 14(4), 402.
- Bharadwaj, P., K. V. Løken, and C. Neilson (2013). Early life health interventions and academic achievement. *American Economic Review* 103(5), 1862–91.
- Caetano, C., G. Caetano, and J. C. Escanciano (2020). Regression discontinuity design with multivalued treatments. *arXiv preprint arXiv:2007.00185*.
- Card, D., D. S. Lee, Z. Pei, and A. Weber (2015). Inference on causal effects in a generalized regression kink design. *Econometrica* 83(6), 2453–2483.
- Cattaneo, M. D. and J. C. Escanciano (2017). *Regression discontinuity designs: Theory and applications*. Advances in Econometrics. Emerald Group Publishing.
- Cattaneo, M. D., M. Jansson, and X. Ma (2020). Simple local polynomial density estimators. *Journal of the American Statistical Association* 115(531), 1449–1455.

- Cattaneo, M. D. and R. Titiunik (2021). Regression discontinuity designs.
- Chen, X., O. Linton, and I. V. Keilegom (2003). Estimation of semiparametric models when the criterion function is not smooth. *Econometrica* 71(5), 1591–1608.
- Chen, Y., A. Ebenstein, M. Greenstone, and H. Li (2013). Evidence on the impact of sustained exposure to air pollution on life expectancy from china’s huai river policy. *Proceedings of the National Academy of Sciences* 110(32), 12936–12941.
- Chernozhukov, V., I. Fernández-Val, and A. Galichon (2010). Quantile and probability curves without crossing. *Econometrica* 78(3), 1093–1125.
- Chernozhukov, V. and C. Hansen (2005). An iv model of quantile treatment effects. *Econometrica* 73(1), 245–261.
- Cole, C. (2021). Effect of family income in infancy on child and adult outcomes: New evidence using census data and tax discontinuities. Job market paper, University of Michigan.
- D’Haultfœuille, X. and P. Février (2015). Identification of nonseparable triangular models with discrete instruments. *Econometrica* 83(3), 1199–1210.
- Dong, Y. (2018a). Alternative assumptions to identify late in fuzzy regression discontinuity designs. *Oxford Bulletin of Economics and Statistics* 80(5), 1020–1027.
- Dong, Y. (2018b). Jump or kink? regression probability jump and kink design for treatment effect evaluation. Working paper, UC Irvine.
- Dong, Y., Y.-Y. Lee, and M. Gou (2021). Regression discontinuity designs with a continuous treatment. *Journal of the American Statistical Association* 0(ja), 1–31.
- Dong, Y. and A. Lewbel (2015, 12). Identifying the Effect of Changing the Policy Threshold in Regression Discontinuity Models. *The Review of Economics and Statistics* 97(5), 1081–1092.
- Ebenstein, A., M. Fan, M. Greenstone, G. He, and M. Zhou (2017). New evidence on the impact of sustained exposure to air pollution on life expectancy from china’s huai river policy. *Proceedings of the National Academy of Sciences* 114(39), 10384–10389.

- Fan, Y. and E. Guerre (2016, 2021/10/18). *Multivariate Local Polynomial Estimators: Uniform Boundary Properties and Asymptotic Linear Representation*, Volume 36, pp. 489–537. Emerald Group Publishing Limited.
- Flood, S., M. King, R. Rodgers, S. Ruggles, and J. R. Warren (2020). Integrated public use microdata series, current population survey: Version 7.0 [dataset]. <https://doi.org/10.18128/D030.V7.0>. Minneapolis, MN: IPUMS.
- Giné, E. and A. Guillou (1999). Laws of the Iterated Logarithm for Censored Data. *The Annals of Probability* 27(4), 2042 – 2067.
- Giuntella, O. and F. Mazzonna (2019). Sunset time and the economic effects of social jetlag: evidence from us time zone borders. *Journal of Health Economics* 65, 210–226.
- Hahn, J., P. Todd, and W. Van der Klaauw (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica* 69(1), 201–209.
- Hairston, K. G., M. Bryer-Ash, J. M. Norris, S. Haffner, D. W. Bowden, and L. E. Wagenknecht (2010). Sleep duration and five-year abdominal fat accumulation in a minority cohort: the iras family study. *Sleep* 33(3), 289–295.
- Hansen, B. E. (2004). Nonparametric estimation of smooth conditional distributions. *Unpublished paper: Department of Economics, University of Wisconsin*.
- Heckman, J. J. and E. Vytlačil (2001). Policy-relevant treatment effects. *American Economic Review* 91(2), 107–111.
- Hoderlein, S. and E. Mammen (2007). Identification of marginal effects in nonseparable models without monotonicity. *Econometrica* 75(5), 1513–1518.
- Hoderlein, S. and E. Mammen (2009). Identification and estimation of local average derivatives in non-separable models without monotonicity. *The Econometrics Journal* 12(1), 1–25.
- Hofferth, S. L., S. M. Flood, M. Sobek, and D. Backman (2020). American time use survey data extract builder: Version 2.8 [dataset]. <https://doi.org/10.18128/D060.V2.8>. College Park, MD: University of Maryland and Minneapolis, MN: IPUMS.

- Imbens, G. W. and T. Lemieux (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics* 142(2), 615–635. The regression discontinuity design: Theory and applications.
- Imbens, G. W. and W. K. Newey (2009). Identification and estimation of triangular simultaneous equations models without additivity. *Econometrica* 77(5), 1481–1512.
- Ishihara, T. (2021). Partial identification of nonseparable models using binary instruments. *Econometric Theory* 37(4), 817–848.
- Koltchinskii, V. (2011). *Oracle Inequalities in Empirical Risk Minimization and Sparse Recovery Problems: Ecole d’Eté de Probabilités de Saint-Flour XXXVIII-2008*, Volume 2033. Springer Science & Business Media.
- Lee, D. S. and T. Lemieux (2010, 6). Regression discontinuity designs in economics. *Journal of Economic Literature* 48(2), 281–355.
- Litschig, S. and K. Morrison (2010). Government spending and re-election: Quasi-experimental evidence from brazilian municipalities. *UPF Discussion Paper*.
- Masten, M. A. and A. Torgovitsky (2016). Identification of instrumental variable correlated random coefficients models. *Review of Economics and Statistics* 98(5), 1001–1005.
- Matzkin, R. L. (2003). Nonparametric estimation of nonadditive random functions. *Econometrica* 71(5), 1339–1375.
- Nolan, D. and D. Pollard (1987). U-processes: Rates of convergence. *The Annals of Statistics* 15(2), 780–799.
- Porter, J. (2003). Estimation in the regression discontinuity model. Working paper, University of Wisconsin at Madison.
- Qu, Z. and J. Yoon (2015). Nonparametric estimation and inference on conditional quantile processes. *Journal of Econometrics* 185(1), 1–19.
- Sasaki, Y. (2015). What do quantile regressions identify for general structural functions? *Econometric Theory* 31(5), 1102–1116.

- Su, L., T. Ura, and Y. Zhang (2019). Non-separable models with high-dimensional data. *Journal of Econometrics* 212(2), 646–677.
- Sun, Y. (2005). Adaptive estimation of the regression discontinuity model. Working paper, UC San Diego.
- Thistlethwaite, D. L. and D. T. Campbell (1960). Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational psychology* 51(6), 309.
- Torgovitsky, A. (2015). Identification of nonseparable models using instruments with small support. *Econometrica* 83(3), 1185–1197.
- Torgovitsky, A. (2017). Minimum distance from independence estimation of nonseparable instrumental variables models. *Journal of Econometrics* 199(1), 35–48.
- van der Vaart, A. W. and J. A. Wellner (1996). *Weak Convergence and Empirical Processes*. New York, NY: Springer.
- Wainwright, M. J. (2019). *High-dimensional statistics: A non-asymptotic viewpoint*, Volume 48. Cambridge University Press.
- White, H. (2001). *Asymptotic theory for econometricians*. Academic press.
- Xie, H. (2021). Uniform convergence for local linear regression estimation of the conditional distribution. arXiv 2112.08546.