Dynamic Regression Discontinuity under Treatment Effect Heterogeneity

Yu-Chin Hsu[†] **Shu Shen**[‡]

This version: May 20, 2024

Abstract

Regression discontinuity is a popular tool for analyzing economic policies or treatment interventions. This research extends the classic static RD model to a dynamic framework, where observations are eligible for repeated RD events and, therefore, treatments. Such dynamics often complicate the identification and estimation of long-term average treatment effects. Empirical papers with such designs have so far ignored the dynamics or adopted restrictive identifying assumptions. This paper presents identification strategies under various sets of weaker identifying assumptions and proposes associated estimation and inference methods. The proposed methods are applied to revisit the seminal study of Cellini et al. (2010) on long-term effects of California local school bonds.

Keywords: long-term treatment effects, dynamic regression discontinuity, semi-parametric, varying coefficient Logit

[†] Yu-Chin Hsu, Institute of Economics, Academia Sinica; Department of Finance, National Central University; Department of Economics, National Chengchi University; CRETA, National Taiwan University. E-mail: ychsu@econ.sinica.edu.tw.

[‡] Shu Shen, Department of Economics, University of California, Davis. E-mail: shushen@ucdavis.edu.

[*] This paper was first presented at a University of California, Berkeley causal inference group seminar in March 2020 under the title of "Dynamic Regression Discontinuity". The authors are grateful to the Editor, Stephane Bonhomme, and three anonymous referees for valuable comments and suggestions on previous versions of the paper. The authors also thank seminar participants from University of California, Berkeley, University of California, Irvine, University of Nevada, Reno, University of Chicago, and

1 Introduction

Regression discontinuity (RD) models, which can be traced back to Thistlethwaite and Campbell (1960), are popular in policy evaluations or other settings of treatment effect analysis. The set-up exploits discontinuity in the design of many policies to nonparametrically identify treatment effects for observations near the eligibility cutoff. Classic studies of RD identification, estimation, and inference include Hahn et al. (2001), Porter (2003), Lee (2008), Imbens and Lemieux (2008), Calonico et al. (2014), and Calonico et al. (2019), among many others. Cattaneo and Titiunik (2022) provides a comprehensive literature review.

While the classic RD set-up, either sharp or fuzzy, is static, in empirical applications we often see situations where each individual faces multiple rounds of RD and, therefore, could potentially receive repeated treatments. For example, voter-approved measures such as unionization (e.g., DiNardo and Lee, 2004; Lee and Mas, 2012) or local school bonds (e.g., Cellini et al., 2010) could be put in front of voters recurrently. A large body of literature in political science and economics (e.g., Lee, 2008; Pettersson-Lidbom, 2008; Ferreira and Gyourko, 2009; Colonnelli et al., 2020) uses RD to study the effect of political races that happen on a regular basis. Dube et al. (2019) and Johnson (2020) examine the peer effect of RD treatments. Their setting is repeated as well because the same individual can be exposed to different peer treatments at different points of time.

The repeated treatment design brings complications to identifying long-term effects in the RD setting just as in other non-RD settings. As Heckman et al. (2016) discuss, in repeated/sequential treatment settings reduced-form analysis can only identify a mixed bag of effects, called the long-term *total* effect. For policy analysis, researchers often want to distinguish a long-term *direct* effect, or the “clean” long-term effect of an earlier treatment with no subsequent treatments (Heckman et al., 2016), from the mixed bag.

In the California education bond example studied by Cellini et al. (2010) (CFR), for example, policy-makers are interested in identifying the long-run effect of passing a local education bond on education expenditure, test scores, and local house prices. For the

Chamberlain Seminar for helpful comments. Yu-Chin Hsu gratefully acknowledges the research support from Ministry of Science and Technology of Taiwan (MOST107-2410-H-001-034-MY3, MOST110-2634-F-002-045), National Science and Technology Council of Taiwan (NSTC 112-2628-H-001-001), Academia Sinica Investigator Award of Academia Sinica (AS-IA-110-H01), and Center for Research in Econometric Theory and Applications (113L8601) from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education of Taiwan. Shu Shen gratefully acknowledges the research support from UC Davis Small Grants in Aid of Research, Creative Activities, and Scholarship.

purpose of identification, one wishes that the education bond voting only takes place once. Then, comparing long-run outcomes between school districts that barely pass and barely miss the vote share cutoff would give average long-term direct effects of interests (for those school districts at the vote share cutoff). In reality, education bond measures could be put forward repeatedly. As is discussed in CFR, the difference in observed long-run outcomes at the RD cutoff only captures an average total effect that is also influenced by subsequent bond authorizations. For example, school districts failing to pass the initial bond measure are more likely to put forward another bond measure in later periods (Cellini et al., 2010).

Among the empirical literature on repeated RD settings, CFR and subsequent studies adopting their methods (e.g., Darolia, 2013; Abott et al., 2020) are the only ones, as far as the authors know, that seek proper identification of average long-term direct effects. Other empirical studies either ignore the repeated design or opt to only study immediate effects and/or long-term total effects. CFR proposes to identify average long-term direct effects from total effects with a recursive RD strategy and a more parametric event study strategy. Recently, Gallen et al. (2023) extend CFR's recursive identification strategy to a non-RD IV setting.

This paper formalizes the repeated RD design under the potential outcome framework. We find that, under treatment effect heterogeneity, preserving the recursive CFR strategy to identify long-term average direct effects requires strong assumptions. In the California education bond application, for example, using the recursive CFR strategy would require exogeneity in the decision of putting forward other bond measures after the focal round, among other identifying conditions such as path-independency and homogeneity in bond effects. As Dong (2019) points out in a study of RD design with sample selection, although RD models could be regarded as local random experiments, such randomness around running variable cutoffs could not be used to argue for the ignorance of endogenous selection into RD events. In the dynamic RD design, random participation in subsequent rounds of RD is rarely plausible. Homogeneity in average effects across different rounds and regardless of past treatment paths is also very strong.

Given the aforementioned considerations, the main contribution of this paper is to propose a new identification strategy for long-term average direct effects. In contrast to recursive CFR, our proposed identification strategy does not impose any restriction on potentially endogenous RD participation decisions. Our strategy also allows treatment effects to depend on last period's treatment take-up. Our key identification restriction is a conditional mean independence assumption (CIA) that requires mean independence between the second-round running variable and the second-round potential outcomes *without* the second-round treatment, conditional on RD participation. For tractability in longer-term effect identification, we also impose a Markovian-type assumption to simplify the path-dependency structure of average treatment effects. Besides point identification, we provide partial identification results in the online appendix under Manski-type (e.g., Manski and Pepper, 2000) monotonicity conditions.

Our paper relates to the dynamic treatment effect literature outside the RD setting. As Han (2021) discusses, the biostatistics literature (e.g., Robins, 1986, 1987; Murphy et al., 2001; Murphy, 2003; Chakraborty and Murphy, 2014) has a long history of studying dynamic causal effects under the assumption of sequential randomization. As we shall explain in the paper, although sharp RD designs are often understood as local random experiments, the dynamic RD set-up does not enjoy sequential randomization by design. In addition, imposing sequential randomization or ignorability (e.g., Blackwell, 2013; Imai and Ratkovic, 2015; Imbens and Lemieux, 2008; Bojinov et al., 2021) on treatments as an identification condition can be undesirable in empirical RD studies.

Our paper relates to Heckman et al. (2016) who evaluate treatment effects in ordered and unordered multi-stage decision problems with an instrumental variable approach, to Sun and Abraham (2021), Callaway and Sant'Anna (2021), and Athey and Imbens (2022) who examine treatment effects in panel event studies with one single irreversible treatment, to De Chaisemartin and d'Haultfoeuille (2020) who study linear two-way fixed-effect regressions for panel data models with treatment effect heterogeneity across groups or over time, and to De Chaisemartin and d'Haultfoeuille (forthcoming) who investigate path-specific long-term average treatment effects using the parallel trend condition. All aforementioned papers have different model set-ups and identifying assumptions than ours. Han (2021) proposes to identify path-specific ATEs using a sequence of dynamic treatment selection equations and excluded instruments. We do not have natural instruments in the repeated RD setting.

Our dynamic RD model also relates to RD models with repeated designs or multiple scores/cutoffs, including Grembi et al. (2016) who propose a difference-in-discontinuities strategy to partial out the effect of a confounding policy, Lv et al. (2019) who consider RD survival analysis where individual treatments are allowed to be allocated at different pre-treatment duration, multi-score models studied by Papay and Murnane (2011), Reardon and Robinson (2012), and Wong et al. (2013), and the multi-cutoff RD model of Cattaneo et al. (2016). Given our focus on long-term direct effects, this paper is significantly different from the others.

Aside from identification, our paper contributes to the literature by designing a new two-step semi-parametric boundary estimation procedure. Specifically, our identified long-term effects has a form of inverse propensity score weighting (IPW), and so the estimation follows two steps. In the first step, we model the propensity score function semi-parametrically and estimate it with the local MLE approach in Cai et al. (2000). In the second step, we plug in estimated propensity scores to local-linear regressions. Our proposed first-step local MLE algorithm is particularly suitable for the RD setting, because it allows the propensity score estimator to stay local to the RD cutoff along the dimension of the running variable (cf. Gelman and Imbens, 2019) while not overburdening the final two-step estimator with the "curse of dimensionality". In terms of inference, our paper is the first to apply the weighted bootstrap designed in Ma and Kosorok (2005) to kernel-based boundary estimation which is the main workhorse of the RD literature. The weighted bootstrap method has also been adopted by Chen and Pouzo (2009), Chernozhukov et al. (2015a,b), and Fernàndez-Val et al. (2021), among others, in other estimation settings.

The rest of the paper is organized as follows. Section 2 starts with a simple two-period dynamic RD model, explaining why long-term direct effects are important policy parameters and why their identification cannot be obtained directly from the RD design. Section 2.2 formalizes the recursive CFR identification strategy under treatment effect heterogeneity and discuss its limitations. Sections 2.3 and 2.4 present a new identification and estimation strategy for the one-period-after average direct effect in the benchmark two-period model. Section 3 presents the general multi-period dynamic RD model and the identification of longer-term average direct effects. Section 4 studies estimation and inference of the general model. Section 5 revisits the empirical study of California education bonds using CFR's published dataset. Section 6 concludes. Monte Carlo simulations and proofs are provided in the online appendix, which also includes partial identification results and several empirical-relevant special cases.

2 A Benchmark Two-period Model

In this section, we use a simple two-period model to define the dynamic RD design under the potential outcome framework. We take the simple model to point out complications brought by having a repetition in the RD design, as well as the differences between dynamic RD models and other non-RD dynamic models previously studied in the literature. We propose a novel identification strategy in this two-period model and discuss its advantages compared to previous methods.

Before we begin, it is important to point out that the simple two-period model defined in this section is only a starting point. Some important identification results provided in this paper are only relevant for the general case presented in Section 3.

2.1 The Classic Potential Outcome Framework

Consider a repeated RD setting, where the RD event (e.g., election, testing, etc.) takes place *at the beginning* of each period, and the treatment is administrated immediately following the event for participants who pass a running variable threshold. An outcome is observed *at the end* of each period.

In period one, we assume for now that everyone takes part in the RD event, so the *observed* treatment D_{i1} and outcome Y_{i1} of individual i satisfy:

$$D_{i1} = 1(Z_{i1} \geq 0), \quad Y_{i1} = Y_{i1}(0) \cdot (1 - D_{i1}) + Y_{i1}(1) \cdot D_{i1},$$

where Z_{i1} is the first-round running variable and $Y_{i1}(d_1)$ is the *potential* first-period outcome with $D_{i1} = d_1$, for $d_1 = 0, 1$. Without loss of generality, all RD cutoffs in this paper are normalized to zero.

In period two, the potential outcome framework gives:

$$\begin{aligned}
D_{i2} =& D_{i2}(0) \cdot (1 - D_{i1}) + D_{i2}(1) \cdot D_{i1}, \text{ and} \\
Y_{i2} =& Y_{i2}(0,0) \cdot (1 - D_{i1}) \cdot (1 - D_{i2}(0)) + Y_{i2}(0,1) \cdot (1 - D_{i1}) \cdot D_{i2}(0) \\
&+ Y_{i2}(1,0) \cdot D_{i1} \cdot (1 - D_{i2}(1)) + Y_{i2}(1,1) \cdot D_{i1} \cdot D_{i2}(1) \\
\equiv& \sum_{\ell^2 \in \mathcal{L}^2} Y_{i2}(\ell^2) \cdot \mathfrak{D}_i(\ell^2), \quad \mathcal{L}^2 = \{(0,0), (0,1), (1,0), (1,1)\},
\end{aligned}$$

where D_{i2} and Y_{i2} are *observed* treatment decisions and outcomes, while $D_{i2}(d_1)$ and $Y_{i2}(d_1, d_2)$, $d_1, d_2 = 0, 1$, are their *potential* counterparts. $\mathfrak{D}_i(.)$ is the path indicator of individual i and $\mathcal{L}^2$ is the set of all possible treatment paths in two periods. This dynamic potential outcome framework is common in dynamic causal effect settings outside RD. See, for example, Section III of the biostatistics textbook Hernán and Robins (2023) and previous work in econometrics, including Hahn et al. (2001) and Bojinov et al. (2021).

The RD setting also brings special features to the dynamic model. Specifically, the second-period treatment decision $D_{i2}(d_1)$ is determined by a potentially endogenous RD participation indicator $S_{i2}(d_1)$ and a potentially endogenous second-round running variable $Z_{i2}(d_1)$ such that $D_{i2}(d_1) = S_{i2}(d_1) \cdot 1(Z_{i2}(d_1) \geq 0)$. The second-period *observed* participation decision and treatment decision satisfy:

$$
\begin{aligned}
S_{i2} =& S_{i2}(0) \cdot (1 - D_{i1}) + S_{i2}(1) \cdot D_{i1}, \\
D_{i2} =& S_{i2}(0) \cdot 1(Z_{i2}(0) \geq 0) \cdot (1 - D_{i1}) + S_{i2}(1) \cdot 1(Z_{i2}(1) \geq 0) \cdot D_{i1}.
\end{aligned}
$$

The second-round running variable $Z_{i2} = Z_{i2}(0) \cdot (1 - D_{i1}) + Z_{i2}(1) \cdot D_{i1}$ is only observed given RD participation, i.e., $S_{i2} = 1$.

Example - California Education Bonds: In the California education bond example studied in CFR, outcome measures include local education expenditures, house prices, student achievements, etc. A bond measure is approved if its vote share exceeds the legislative cutoff (normalized to 0). No matter whether a school district gets its bond approved in the first round (D_{i1}) or not, it could choose to initiate a new bond measure in the next election year ($S_{i2} = 1$) and choose how much campaign efforts to put into the new measure to improve its vote share result ($Z_{i2}(d_1)$, for $d_1 = 0, 1$).

We formally define the following *individual* treatment effects for the two-period model:

$$
\begin{aligned}
&\text{immediate effect of } D_{i1}\text{: } \theta_{i;0,1} = Y_{i1}(1) - Y_{i1}(0); \\
&\text{immediate effect of } D_{i2}\text{: } \theta^{d_1}_{i;0,2} = Y_{i2}(d_1, 1) - Y_{i2}(d_1, 0), \ \textit{for } d_1 = 0, 1; \\
&\text{one-period-after direct effect of } D_{i1}\text{: } \theta_{i;1,1} = Y_{i2}(1, 0) - Y_{i2}(0, 0); \\
&\text{one-period-after total effect of } D_{i1}\text{: } \tilde{\theta}_{i;1,1} = \tilde{Y}_{i2}(1) - \tilde{Y}_{i2}(0), \\
&\qquad \textit{where } \tilde{Y}_{i2}(d_1) \equiv Y_{i2}(d_1, D_{i2}(d_1)) = Y_{i2}(d_1, 0)(1 - D_{i2}(d_1)) + Y_{i2}(d_1, 1) D_{i2}(d_1).
\end{aligned}
$$

Note that the immediate effect of the second-round treatment is path-dependent. Following Heckman et al. (2016) and CFR, we define long-term direct effects by prohibiting treatment take-ups after the focal round. The total effect, on the other hand, does not limit treatment decisions after the focal round. Direct and total effects are also commonly seen in the mediation literature.[1] See, Huber (2020) for a literature review.

Assumption 2.1 *There exists an $\epsilon > 0$, such that:*

1. Z_{i1} *is continuous in* $z_1 \in (-\epsilon, \epsilon) \equiv \mathcal{N}_\epsilon$ *with* $P[Z_{i1} \geq 0] \in (0, 1)$*;*

2. $E[Y_{i1}(d_1)|Z_{i1} = z_1]$*,* $E[D_{i2}(d_1)|Z_{i1} = z_1]$*, and* $E[\tilde{Y}_{i2}(d_1)|Z_{i1} = z_1]$ *are all continuous in* $z_1 \in \mathcal{N}_\epsilon$*, for both* $d_1 = 0, 1$*.*

[1] In the mediation literature, $\theta_{i;1,1} = Y_{i2}(1, 0) - Y_{i2}(0, 0)$ is the controlled direct effect, in contrast to the pure direct effect, which would be $Y_{i2}(1, D_{i2}(0)) - Y_{i2}(0, D_{i2}(0))$ in our notation. Flores and Flores-Lagunes (2009, 2010) study the pure direct effect in the econometrics literature. As Huber (2020) discusses, the pure direct effect is not interesting in the dynamic setting.

Assumption 2.1 imposes traditional RD smoothness conditions that provide identification for the average immediate effect $E[\theta_{i;0,1}|Z_{i1} = 0]$ and the average first-stage effect $E[D_{i2}(1) - D_{i2}(0)|Z_{i1} = 0]$. Since the random variable $\tilde{Y}_{i2}(d_1)$, which takes place in the second period but only specifies the first-round treatment status, can be viewed as a potential outcome of the first-round treatment, smoothness conditions in Assumption 2.1 also identifies the average one-period-after total effect:

$$E[\tilde{\theta}_{i;1,1}|Z_{i1} = z_1] = E[\tilde{Y}_{i2}(1) - \tilde{Y}_{i2}(0)|Z_{i1} = 0] \\ = \lim_{z_1 \searrow 0} E[Y_{i2}|Z_{i1} = z_1] - \lim_{z_1 \nearrow 0} E[Y_{i2}|Z_{i1} = z_1]. \tag{2.1}$$

In the rest of Section 2, we explore conditions that identify the average direct effect $E[\theta_{i;1,1}|Z_{i1} = 0]$, which has important policy implications as argued in Heckman et al. (1997) and CFR. For conciseness, we call $E[\theta_{i;1,1}|Z_{i1} = 0]$ the one-period-after ATE. It has the following relationship with the one-period-after average total effect:

$$E[\tilde{\theta}_{i;1,1}|Z_{i1} = 0] = E[\theta_{i;1,1}|Z_{i1} = 0] + E[\theta^1_{i;0,2}D_{i2}(1)|Z_{i1} = 0] - E[\theta^0_{i;0,2}D_{i2}(0)|Z_{i1} = 0]. \tag{2.2}$$

Example - continued: In the California education bond example, the one-period-after ATE ($E[\theta_{i;1,1}|Z_{i1} = 0]$) is the average effect of passing an education bond in the first period on the second-period outcome with no bond authorization after the first period, among all school districts at the first-period vote share cutoff. The one-period-after ATE influences the one-period-after average total effect ($E[\tilde{\theta}_{i;1,1}|Z_{i1} = 0]$), which is identified by the period two outcome discontinuity observed at the period one vote share cutoff. However, the one-period-after total effect is also influenced by the fact that school districts can pass new bond measures in the second period and hence receive immediate effects from those additional treatments.

2.2 Recursive Identification Strategy in CFR

CFR propose a recursive identification strategy for long-term direct effects based on an implicit treatment effect homogeneity assumpiton, which requires individual treatment effects to vary only by the number of periods between potential outcomes and the focal round of treatment.[2] For the two-period model, that is to require

$$\theta_0 \equiv \theta_{i,0,1} = \theta^0_{i,0,2} = \theta^1_{i,0,2} \quad \text{and} \quad \theta_1 \equiv \theta_{i,1,1} \quad \text{for all } i, \tag{2.3}$$

[2]See, for example, Section IV.B of CFR. In CFR, our direct effect is called the "treatment-on-the-treated" effect and our total effect is called the "intent-to-treat" effect.

where θ_0 and θ_1 are unknown fixed parameters.

The strong assumption reduces the relationship in (2.2) to:

$$E[\tilde{\theta}_{i;1,1}|Z_{i1}=0]=\theta_1+\theta_0 E[D_{i2}(1)-D_{i2}(0)|Z_{i1}=0].$$

The simplification implies recursive identification of the one-period-after ATE θ_1, since all other components of the equation are directly identifiable through smoothness conditions.

CFR's recursive strategy can be extended to allow for individual treatment effect heterogeneity under the potential outcome framework outlined in the previous section.

Lemma 2.1 *Under Assumption 2.1, the homogeneous ATE requirement:*

$$ATE_0 \equiv E[\theta_{i;0,1}|Z_{i1}=0]=E[\theta^1_{i;0,2}|Z_{i1}=0]=E[\theta^0_{i;0,2}|Z_{i1}=0], \tag{2.4}$$

and the random second-round treatment selection requirement:

$$E[\theta^{d_1}_{i;0,2}|D_{i2}(d_1),Z_{i1}=0]=E[\theta^{d_1}_{i;0,2}|Z_{i1}=0], \quad d_1=0,1, \tag{2.5}$$

the recursive identification strategy in CFR can be preserved:

$$\begin{aligned}
ATE_0 &= \lim_{z_1\searrow 0}E[Y_{i1}|Z_{i1}=z_1]-\lim_{z_1\nearrow 0}E[Y_{i1}|Z_{i1}=z_1],\\
ATE_1 \equiv& E[\theta_{i;1,1}|Z_{i1}=0]=\lim_{z_1\searrow 0}E[Y_{i2}|Z_{i1}=z_1]-\lim_{z_1\nearrow 0}E[Y_{i2}|Z_{i1}=z_1]\\
&-ATE_0\cdot\left(\lim_{z_1\searrow 0}E[D_{i2}|Z_{i1}=z_1]-\lim_{z_1\nearrow 0}E[D_{i2}|Z_{i1}=z_1]\right).
\end{aligned}$$

To preserve CFR's recursive identification strategy under individual treatment effect heterogeneity, conditions (2.4) and (2.5) are required. The conditions include the strong homogeneity assumption in (2.3) as a special case.[3] If the homogeneity ATE condition in (2.4) is violated due to diminishing marginal returns to repeated treatments (i.e., $E[\theta^1_{i;0,2}|Z_{i1}=0]<E[\theta^0_{i;0,2}|Z_{i1}=0]=E[\theta_{i;0,1}|Z_{i1}=0]$), then ATE_1 identified in Lemma 2.1 is smaller than the true value of $E[\theta_{i;1,1}|Z_{i1}=0]$.

The random treatment selection condition in (2.5) should not to be confused with the local randomness intuition of the RD design. A treatment intervention satisfying

[3] A more general multi-period version of the lemma is discussed in the online appendix, where we also extend Lemma 2.1 to include conditioning covariates $X_i\in\mathcal{X}$. The extension replaces the homogeneous ATE assumption in (2.4) and the random treatment selection assumption in (2.5) with:

$$E[\theta^{d_1}_{i;0,2}|D_{i2}(d_1),X_i=x,Z_{i1}=0]=E[\theta^{d_1}_{i;0,2}|X_i=x,Z_{i1}=0], \text{ and} \tag{2.6}$$

$$CATE_0(x)=E[\theta_{i;0,1}|X_i=x,Z_{i1}=0]=E[\theta^{d_1}_{i;0,2}|X_i=x,Z_{i1}=0], \ \forall d_1=0,1, \ x\in\mathcal{X} \tag{2.7}$$

the sharp RD design is only *locally* random at its *own* running variable cutoff. In other words, the RD design associated with $D_{i2}(d_1)$ only indicates local randomness among individuals at $Z_{i2}(d_1) = 0$, for $d_1 = 0, 1$. Equation (2.5) is different. It is not implied by the RD design and is a strong condition used to obtain the recursive identification result.

Example - continued: In the California education bond example, data reveal that school districts barely passing their education bond in the first round do not put forward another bond in the second period (i.e., $E[S_{i2}(1)|Z_{i1} = 0] = \lim_{z_1 \searrow 0} E[S_{i2}|Z_{i1} = z_1] = 0$). Therefore, equation (2.5) essentially restricts that, for a school district i that marginally failed the first-round voting, the decision of putting forward another measure in the next election cycle ($S_{i2}(0)$) is not related to the immediate treatment effect of the new measure ($\theta^0_{i;0,2}$). The equation also implies that, conditional on putting forward another measure ($S_{i2}(0) = 1$), the amount of resources devoted toward campaigning to improve the second-round vote share ($Z_{i2}(0)$) is not related to the immediate treatment effect ($\theta^0_{i,0,2}$), either. If school districts with higher $\theta^0_{i,0,2}$'s are more likely to pass their second-round measures, ATE_1 identified by Lemma 2.1 is smaller than the true value of $E[\theta_{i;1,1}|Z_{i1} = 0]$.

The recursive identification structure of Lemma 2.1 is to be distinguished from the one-step approach commonly seen in the biostatistics literature (e.g., Hernán and Robins, 2023). To see this, we note that an alternative way of imposing randomization on D_{i2} is to rule out the possibility that individuals endogenously choose D_{i2} based on how quickly the effect of D_{i1} dies out over time. In such a case, ATE_1 could be identified in one step as $\lim_{z_1 \searrow 0} E[Y_{i2}|D_{i2} = 0, Z_{i1} = z_1] - \lim_{z_1 \nearrow 0} E[Y_{i2}|D_{i2} = 0, Z_{i1} = z_1]$ (or a more robust version allowing for conditioning covariates). In many empirical applications including the California education bond example, it is likely more plausible to assume that D_{i2} is exogenous to its own immediate effect as in (2.5) than to assume that D_{i2} is irrelevant to the long-run effect of D_{i1}, hence favoring the recursive identification strategy proposed by CFR. Having said that, condition (2.5) is still potentially strong. In the following sections, we seek to relax both identifying conditions (2.4) and (2.5) required in Lemma 2.1.

2.3 Proposed Identification Strategy

In this section, we propose a new identification strategy for the one-period-after ATE. Compared to Lemma 2.1, the new method imposes no restrictions on the potentially endogenous second-round RD participation decision. In addition, the new strategy allows for arbitrary path-dependency in the second-round immediate effect.

Assumption 2.2 *There exists $\epsilon > 0$ such that for both $d_1, d_2 = 0, 1$:*

1. *(CIA)* $E[Y_{i2}(d_1,0)|Z_{i2}(d_1) = z_2, S_{i2}(d_1) = 1, Z_{i1} = z_1] = E[Y_{i2}(d_1,0)|S_{i2}(d_1) = 1, Z_{i1} = z_1]$ *for all* $z_1 \in \mathcal{N}_\epsilon$;

2. *(Smoothness)* $E[Y_{i2}(d_1,d_2)|D_{i2}(d_1) = d_2, Z_{i1} = z_1]$ *and* $E[D_{i2}(d_1)|S_{i2}(d_1) = 1, Z_{i1} = z_1]$ *are both continuous in* $z_1 \in \mathcal{N}_\epsilon$;

3. *(Overlapping)* $E[D_{i2}(d_1)|S_{i2}(d_1) = 1, Z_{i1} = z_1] \in (0,1)$ *for all* $z_1 \in \mathcal{N}_\epsilon$.

The key identifying condition of Assumption 2.2 is the first CIA condition. It imposes a mean independence condition on marginal individuals at the first-round RD cutoff who choose to participate in the second-round RD. The CIA condition could be weakened using covariates and observationally equivalent subpopulations. The second and third parts of Assumption 2.2 are standard RD smoothness and overlapping conditions. Section A.4.1 in the online appendix discusses an important departure from Assumption 2.2.3 for having an absorbing treatment in the dynamic RD model.

Example - continued: In the California education bond example, the CIA condition in Assumption 2.2 requires that, should the second-round education bond measure fail, the potential second-period outcome (of a school district at the first-round vote share cutoff) is mean independent of the second-round vote share. Note that the CIA condition is *not* about potential outcomes *with* second-round bond authorization. This is important because, for example, the size of the proposed bond can affect both the vote share and the potential outcome *with* second-round bond approval. The CIA condition also does *not* impose any restriction on the potentially endogenous decision of putting forward a new bond measure in the second period.

By standard RD identification arguments, Assumptions 2.2.2 and 2.2.3 imply that:

$$E[Y_{i2}(1,1)|D_{i2}(1) = 1, Z_{i1} = 0] = \lim_{z_1 \searrow 0} E[Y_{i2}|S_{i2} = 1, D_{i2} = 1, Z_{i1} = z_1].$$

Assumptions 2.2.1, 2.2.2, and 2.2.3 together imply that:

$$E[Y_{i2}(1,0)|D_{i2}(1) = 1, Z_{i1} = 0] = \lim_{z_1 \searrow 0} E[Y_{i2}(1,0)|S_{i2}(1) = 1, Z_{i2}(1) \geq 0, Z_{i1} = z_1]$$
$$= \lim_{z_1 \searrow 0} E[Y_{i2}|S_{i2}(1) = 1, Z_{i2}(1) < 0, Z_{i1} = z_1] = \lim_{z_1 \searrow 0} E[Y_{i2}|S_{i2} = 1, D_{i2} = 0, Z_{i1} = z_1],$$

where the second equality holds by the CIA. Summarizing the above two equations and obtaining a similar result for $d_1 = 0$, we have:

$$E[Y_{i2}(1,d_2)|D_{i2}(1) = 1, Z_{i1} = 0] = \lim_{z_1 \searrow 0} E[Y_{i2}|S_{i2} = 1, D_{i2} = d_2, Z_{i1} = z_1], \quad (2.8)$$

$$E[Y_{i2}(0,d_2)|D_{i2}(0) = 1, Z_{i1} = 0] = \lim_{z_1 \nearrow 0} E[Y_{i2}|S_{i2} = 1, D_{i2} = d_2, Z_{i1} = z_1], \quad (2.9)$$

which further imply that:

$$
\begin{aligned}
E[\theta^1_{i;0,2} D_{i2}(1)|Z_{i1}=0] &= \lim_{z_1 \searrow 0} E\left[Y_{i2} S_{i2}(D_{i2} - \lambda^1_{D_2|S_2}) \middle| Z_{i1} = z_1\right] / (1 - \lambda^1_{D_2|S_2}), \\
E[\theta^0_{i;0,2} D_{i2}(0)|Z_{i1}=0] &= \lim_{z_1 \nearrow 0} E\left[Y_{i2} S_{i2}(D_{i2} - \lambda^0_{D_2|S_2}) \middle| Z_{i1} = z_1\right] / (1 - \lambda^0_{D_2|S_2}), \\
\lambda^{d_1}_{D_2|S_2} &= P[D_{i2}(d_1) = 1 | S_{i2}(d_1) = 1, Z_{i1} = 0].
\end{aligned}
$$

Together with the decomposition in (2.2), the one-period-after ATE is identified.

The CIA condition in Assumption 2.2 can be extended to include covariates. Specifically, let $X_i \in \mathcal{X}$ be the vector of covariates. The extended CIA condition is:

$$
\begin{aligned}
&E[Y_{i2}(d_1,0)|X_i = x, Z_{i2}(d_1) = z_2, S_{i2}(d_1) = 1, Z_{i1} = z_1] \\
=&E[Y_{i2}(d_1,0)|X_i = x, S_{i2}(d_1) = 1, Z_{i1} = z_1], \quad \forall x \in \mathcal{X}. \qquad (2.10)
\end{aligned}
$$

The vector of covariates X_i could include both pre-treatment controls and first-period outcomes. The role of X_i here is different from that in static RD, where covariates are used to improve estimation efficiency (e.g., Calonico et al., 2019).

Example - continued: Suppose school expenditure is the outcome of interest. The vanilla CIA condition without covariates might not be appropriate if school districts with low baseline funding levels are more incentivized to carry out an active campaign, resulting in a higher vote share ($Z_{i2}(d_1)$). A useful conditioning covariate for the extended CIA condition in (2.10) hence is the first-period school expenditure ($Y_{i1}(d_1)$). If a district's expenditure without new bond authorization follows an AR(1) process with the random shock in each period being mean independent of the bond vote share, then the extended CIA condition in (2.10) holds.

Identification results in (2.8) and (2.9) could be easily extended to incorporate covariates. Let

$$
\lambda^{d_1}(x) \equiv \lambda^{d_1}_{D_2|S_2}(x) = P[D_{i2}(d_1) = 1 | X_i = x, S_{i2}(d_1) = 1, Z_{i1} = 0], \quad d_1 = 0, 1,
$$

be the propensity score function conditional on second-round RD participation. Strengthen the smoothness and overlapping conditions in Assumption 2.2 to require that for all $x \in \mathcal{X}$

$$
\begin{aligned}
&E[Y_{i2}(d_1,d_2)|X_i = x, D_{i2}(d_1) = d_2, Z_{i1} = z_1] \text{ and } E[D_{i2}(d_1)|X_i = x, S_{i2}(d_1) = 1, Z_{i1} = z_1] \\
&\qquad \text{are continuous in } z_1 \in \mathcal{N}_\epsilon, \text{ and} \\
&E[D_{i2}(d_1)|X_i = x, S_{i2}(d_1) = 1, Z_{i1} = z_1] \in (0,1) \text{ for all } z_1 \in \mathcal{N}_\epsilon. \qquad (2.11)
\end{aligned}
$$

The following lemma summarizes identification of the one-period-after ATE using IPW.[4]

Lemma 2.2 *Under Assumptions 2.1, and the extended CIA, smoothness, and overlapping conditions in* (2.10) *and* (2.11)*, the one-period-after ATE is identified:*

$$E[\theta_{i;1,1}|Z_{i1}=0] = \alpha^1 - \alpha^0, \text{ where}$$
$$\alpha^1 = \lim_{z_1 \searrow 0} E\left[Y_{i2} - \frac{Y_{i2}S_{i2}(D_{i2}-\lambda^1(X_i))}{1-\lambda^1(X_i)}\Big|Z_{i1}=z_1\right],$$
$$\alpha^0 = \lim_{z_1 \nearrow 0} E\left[Y_{i2} - \frac{Y_{i2}S_{i2}(D_{i2}-\lambda^0(X_i))}{1-\lambda^0(X_i)}\Big|Z_{i1}=z_1\right].$$

The proof is provided in the appendix. In the next section, we propose a semi-parametric local MLE method (cf. Cai et al., 2000) for propensity score estimation. The method will require additional functional form assumptions on the propensity score functions. The semi-parametric approach is particularly suitable for the RD setting as it will allow the propensity score estimator to stay local to the RD cutoff along the dimension of the running variable (Gelman and Imbens, 2019) while not overburdening the final ATE estimator with the "curse of dimensionality".

2.4 Proposed Estimation Strategy

Assume that the conditional probability function $P[D_{i2}=1|X_i=x, S_{i2}=1, Z_{i1}=z_1]$ follows a class of semi-parametric models $p(x,\gamma(z_1))$, where $p(.,.):\mathcal{X}\times\Gamma\to\mathbb{R}$ is known but $\gamma(.):\mathbb{R}\to\Gamma$ is unknown. For example, if $p(x,\gamma(z_1)) = L\left((1\ x')\gamma(z_1)\right)$ with $L(a) = \exp(a)/(1+\exp(a))$, then $P[D_{i2}=1|X_i=x,S_{i2}=1,Z_{i1}=z_1] = \frac{\exp((1\ x')\gamma(z_1))}{1+\exp((1\ x')\gamma(z_1))}$ follows a varying coefficient Logit model, nesting both parametric Logit and semi-parametric partial-linear Logit as special cases. Let $\gamma^0 = \lim_{z_1\nearrow 0}\gamma(z_1)$ and $\gamma^1 = \lim_{z_1\searrow 0}\gamma(z_1)$. Propensity score functions $\lambda^0(x)$ and $\lambda^1(x)$ could be written as $p(x,\gamma^0)$ and $p(x,\gamma^1)$.

Let $\beta^0_{FS} = \lim_{z_1\nearrow 0}\nabla\gamma(z_1)$ and $\beta^1_{FS} = \lim_{z_1\searrow 0}\nabla\gamma(z_1)$ be the left and the right limits of the gradient of $\gamma(.)$ at the RD cutoff. Let $K(\cdot)$ be the kernel function and h the bandwidth. Let $\hat{\gamma}^0$, $\hat{\gamma}^1$, $\hat{\beta}^0_{FS}$ and $\hat{\beta}^1_{FS}$ denote estimators of γ^0, γ^1, β^0, and β^1. They solve

[4] Angrist et al. (2018) adopted IPW in the static RD model for extrapolating treatment effects away from the running variable cutoff. IPW is also commonly used outside RD, for both cross-sectional and dynamic treatment models (e.g., Blackwell, 2013; Huber, 2020; Bojinov et al., 2021; Imai et al., 2023; Hernán and Robins, 2023, among many others).

the following maximization problems:

$$(\hat{\gamma}^1, \hat{\beta}^1_{FS}) = \arg\max_{\gamma,\beta} \sum_{i=1}^n S_{2i} 1(Z_{1i} \geq 0) K\left(\frac{Z_{1i}}{h}\right) \cdot \Big[D_{2i} \log p(X_i, \gamma + \beta Z_{1i}) + (1 - D_{2i}) \log(1 - p(X_i, \gamma + \beta Z_{1i}))\Big],$$

$$(\hat{\gamma}^0, \hat{\beta}^0_{FS}) = \arg\max_{\gamma,\beta} \sum_{i=1}^n S_{2i} 1(Z_{1i} < 0) K\left(\frac{Z_{1i}}{h}\right) \cdot \Big[D_{2i} \log p(X_i, \gamma + \beta Z_{1i}) + (1 - D_{2i}) \log(1 - p(X_i, \gamma + \beta Z_{1i}))\Big].$$

When $p(x, \gamma(z_1)) = L\left((1\ x')\gamma(z_1)\right)$ with $L(a) = \exp(a)/(1 + \exp(a))$, the optimization problem described above are weighted Logit regressions of D_{2i} on X_i, Z_i, and $X_i Z_i$, with weights determined by $K(Z_{1i}/h)$. If the CIA condition in the previous section holds without any conditioning covariates, the first-step estimation is fully non-parametric.

Denote $E[\theta_{i;1,1}|Z_1 = 0]$ by $\bar{\theta}_{1,1}$ and its estimator by $\hat{\bar{\theta}}_{1,1}$. Given the above-defined first-step propensity score estimators, $\bar{\theta}_{1,1}$ could be estimated by boundary local linear regressions of generated outcomes on the RD running variable. Specifically,

$$\hat{\bar{\theta}}_{1,1} = \hat{\alpha}^1 - \hat{\alpha}^0,$$

$$(\hat{\alpha}^1, \hat{\beta}^1) = \arg\min_{\alpha,\beta} \sum_{i=1}^n 1(Z_{1i} \geq 0) K\left(\frac{Z_{1i}}{h}\right) \left[A_i(\alpha, \beta; \hat{\gamma}^1)\right]^2,$$

$$(\hat{\alpha}^0, \hat{\beta}^0) = \arg\min_{\alpha,\beta} \sum_{i=1}^n 1(Z_{1i} < 0) K\left(\frac{Z_{1i}}{h}\right) \left[A_i(\alpha, \beta; \hat{\gamma}^0)\right]^2,$$

where $A_i(\alpha, \beta; \gamma) = Y_{2i} - \frac{Y_{2i} S_{2i}(D_{2i} - p(X_i, \gamma))}{1 - p(X_i, \gamma)} - \alpha - \beta Z_{1i}$.

Estimators $\hat{\lambda}^0$, $\hat{\lambda}^1$, and $\hat{\bar{\theta}}_{1,1}$ are all consistent and asymptotically normally distributed under proper assumptions. In the above estimation procedure, the kernel bandwidth is kept the same in both steps. If the bandwidth in the first-step propensity score estimation is chosen to shrink at a slower rate, the first-step estimation error would vanish asymptotically for the two-step estimator, simplifying the asymptotic variance of $\hat{\bar{\theta}}_{1,1}$. We propose to use the same bandwidth in both steps for easier interpretations of empirical results. More details about estimation, as well as a weighted bootstrap inference strategy, will be discussed in Section 4.

When the conditioning covariate X_i is non-empty, the proposed estimation procedure is semi-parametric. An interesting extension for future research is to extend the first-step propensity score estimation procedure along the lines of high-dimensional covariate selection (c.f. Chernozhukov et al., 2018), doubly robust estimation (c.f. Kennedy et al., 2017), or Fan et al. (2022) who combine doubly robust estimation with high-dimensional covariate selection in treatment evaluation.

3 Identification in the General Multi-period Setting

3.1 Set-up

Let $k = 1, 2, ..., K$ denote the round of treatment intervention whose eligibility is determined by an RD design. Let $t = 1, 2, ..., T$ denote the period of observed outcomes, and assume without loss of generality that $T = K$.[5] Let D_{ik} be the *observed* treatment status of an individual in round k, S_{ik} the *observed* participation indicator, and Z_{ik} the *observed* running variable which is observed only when $S_{ik} = 1$. Recall from Section 2 that $S_{i1} = 1$ and $D_{i1} = 1(Z_{i1} \geq 0)$. For any $k \geq 2$,

$$D_{ik} = \sum_{\ell^{k-1} \in \mathcal{L}^{k-1}} D_{ik}(\ell^{k-1}) \cdot \mathfrak{D}_i(\ell^{k-1}) \equiv \sum_{\ell^{k-1} \in \mathcal{L}^{k-1}} S_{ik}(\ell^{k-1}) 1(Z_{ik}(\ell^{k-1}) \geq 0) \cdot \mathfrak{D}_i(\ell^{k-1}),$$

$$S_{ik} = \sum_{\ell^{k-1} \in \mathcal{L}^{k-1}} S_{ik}(\ell^{k-1}) \cdot \mathfrak{D}_i(\ell^{k-1}), \quad Z_{ik} = \sum_{\ell^{k-1} \in \mathcal{L}^{k-1}} Z_{ik}(\ell^{k-1}) \cdot \mathfrak{D}_i(\ell^{k-1}) \text{ if } S_{ik} = 1,$$

where $D_{ik}(\ell^{k-1})$, $S_{ik}(\ell^{k-1})$, and $Z_{ik}(\ell^{k-1})$ are *potential* counterparts (with past treatment path ℓ^{k-1}) of D_{ik}, S_{ik}, and Z_{ik}, respectively. The path indicator $\mathfrak{D}_i(.)$ and set $\mathcal{L}^{k-1}$ extend definitions in Section 2. Specifically, the event $\{\mathfrak{D}_i(\ell^k) = 1\}$ is equivalent to $\{D_{i1} = \ell^k_1, D_{i2}(\ell^k_1) = \ell^k_2, ..., D_{ik}(\ell^k_{1:(k-1)}) = \ell^k_k\} = \{D_{i1} = \ell^k_1, D_{i2} = \ell^k_2, ..., D_{ik} = \ell^k_k\}$, where ℓ^k_j ($\ell^k_{j:j'}$) represent the j-th (j-th to j'-th) element(s) of path ℓ^k.

Let Y_{it} be the *observed* outcome in period t and $Y_{it}(\ell^t)$ the potential outcome with past treatment path $\ell^t \in \mathcal{L}^t$ such that:

$$Y_{it} = \sum_{\ell^t \in \mathcal{L}^t} Y_{it}(\ell^t) \cdot \mathfrak{D}_i(\ell^t).$$

Let $\mathbf{0}_\tau$ be a τ-dimensional vector of zeros. Let $\tilde{Y}_{i(k+\tau)}(\ell^k)$ be a $(k+\tau)$-th period *quasi-potential* outcome, where only the first k rounds of the total $(k+\tau)$ rounds of past treatment status are specified. Specifically, $\tilde{Y}_{i(k+\tau)}(\ell^k) = \sum_{\eta \in \mathcal{L}^\tau} Y_{i(k+\tau)}(\ell^k, \eta) \mathfrak{D}_{i;(k+1):(k+\tau)}(\ell^k, \eta)$, where partial path indicator $\{\mathfrak{D}_{i;l:l'}(\ell^k) = 1\} \Leftrightarrow \{D_{il}(\ell^k_{1:(l-1)}) = \ell^k_l, D_{i(l+1)}(\ell^k_{1:l}) = \ell^k_{l+1}, ..., D_{il'}(\ell^k_{1:(l'-1)}) = \ell^k_{l'}\}$. Extending definitions in Section 2, we have long-term effects of the first-round treatment for all $\tau \geq 1$:

τ-period-after direct effect of D_{i1}: $\theta_{i;\tau,1} = Y_{i(1+\tau)}(1, \mathbf{0}_\tau) - Y_{i(1+\tau)}(0, \mathbf{0}_\tau)$,

τ-period-after total effect of D_{i1}: $\tilde{\theta}_{i;\tau,1} = \tilde{Y}_{i(1+\tau)}(1) - \tilde{Y}_{i(1+\tau)}(0)$.

[5] Although in many settings outcomes can be observed in periods after the last round of treatment intervention, for identification purposes it is not necessary to distinguish such longer-term outcomes from outcomes that occur right after the last round of treatment.

In addition, define the following treatment effects of the k-th round treatment:

immediate effect of D_{ik}: $\theta_{i;0,k}^{\ell^{k-1}} = Y_k(\ell^{k-1}, 1) - Y_k(\ell^{k-1}, 0)$,
τ-period-after direct effect of D_{ik}: $\theta_{i;\tau,k}^{\ell^{k-1}} = Y_{i(k+\tau)}(\ell^{k-1}, 1, \mathbf{0}_\tau) - Y_{i(k+\tau)}(\ell^{k-1}, 0, \mathbf{0}_\tau)$,
τ-period-after total effect of D_{ik}: $\tilde{\theta}_{i;\tau,k}^{\ell^{k-1}} = \tilde{Y}_{i(k+\tau)}(\ell^{k-1}, 1) - \tilde{Y}_{i(k+\tau)}(\ell^{k-1}, 0)$,

for all $\tau \geq 1$, $k \geq 2$, and past treatment path $\ell^{k-1} \in \mathcal{L}^{k-1}$.

Before concluding this set-up section, we present a lemma that motivates the study of long-term direct effects. The following lemma shows that treatment effects not falling into the categories discussed above, such as $Y_{i2}(1,1) - Y_{i2}(0,1)$ or $Y_{i2}(1,0) - Y_{i2}(0,1)$, can be obtained from the immediate effects and long-term direct effects defined above.

Lemma 3.1 *The difference between any pair of potential outcomes can be represented by a linear combination of above-defined immediate effects and long-term direct effects.*

3.2 Identification

In this section, we seek identification of the τ-period-after ATE, or $E[\theta_{i;\tau,1}|Z_1 = 0]$, for all $\tau \geq 2$. Following discussions in Section 2, we continue to impose no assumptions on subsequent rounds of RD participation decisions, which could be potentially endogenous.

Just as in the simple two-period model, our proposed identification strategy is built upon a decomposition of long-term average total effects. The decomposition shows that any long-term total effect is equal to its corresponding direct effect plus a sum of various shorter-term total effects adjusted by subsequent round first-stage treatment decisions.

$$\begin{aligned}
\tilde{\theta}_{i;\tau,1} &= \theta_{i;\tau,1} + \Big(\tilde{\theta}^{1}_{i;\tau-1,2} \cdot D_{i2}(1) - \tilde{\theta}^{0}_{i;\tau-1,2} \cdot D_{i2}(0)\Big) \\
&+ \sum_{s=1}^{\tau-2} \Big(\tilde{\theta}^{(1,\mathbf{0}_{\tau-1-s})}_{i;s,\tau+1-s} \cdot D_{i(\tau+1-s)}(1, \mathbf{0}_{\tau-1-s}) - \tilde{\theta}^{(0,\mathbf{0}_{\tau-1-s})}_{i;s,\tau+1-s} \cdot D_{i(\tau+1-s)}(0, \mathbf{0}_{\tau-1-s})\Big) \\
&+ \theta^{(1,\mathbf{0}_{\tau-1})}_{i;0,\tau+1} \cdot D_{i(\tau+1)}(1, \mathbf{0}_{\tau-1}) - \theta^{(0,\mathbf{0}_{\tau-1})}_{i;0,\tau+1} \cdot D_{i(\tau+1)}(0, \mathbf{0}_{\tau-1}). \qquad (3.1)
\end{aligned}$$

When $\tau = 2$, the decomposition reduces to:

$$\tilde{\theta}_{i;2,1} = \theta_{i;2,1} + \tilde{\theta}^{1}_{i;1,2} D_{i2}(1) - \tilde{\theta}^{0}_{i;1,2} D_{i2}(0) + \theta^{(1,0)}_{i;0,3} D_{i3}(1,0) - \theta^{(0,0)}_{i;0,3} D_{i3}(0,0). \qquad (3.2)$$

The decomposition is new to the literature, as far as the authors know. It is different from the decomposition used behind the recursive CFR identification strategy, where the long-term total effect is decomposed to its corresponding direct effect plus a sum of various shorter-term *direct* effects adjusted by first-stage treatment decisions.[6] In the rest of the section, we use the first decomposition, or equation (3.1), to develop a new identification strategy, because the decomposition involves a much smaller number of path-dependent treatment effects.

Aside from utilizing the new decomposition result, we impose a Markovian-type condition to further reduce the number of path-dependent treatment effect parameters involved in identification. Let $\eta_{i;0,1}$, $\eta^{\ell^{k-1}}_{i;0,k}$, $\eta_{i;\tau,1}$, $\eta^{\ell^{k-1}}_{i;\tau,k}$, $\tilde{\eta}_{i;\tau,1}$, and $\tilde{\eta}^{\ell^{k-1}}_{i;\tau,k}$ be first-stage counterparts of immediate effects $\theta_{i;0,1}$ and $\theta^{\ell^{k-1}}_{i;0,k}$, direct effects $\theta_{i;\tau,1}$ and $\theta^{\ell^{k-1}}_{i;\tau,k}$, and total effects $\tilde{\theta}_{i;\tau,1}$ and $\tilde{\theta}^{\ell^{k-1}}_{i;\tau,k}$, respectively.[7] The following assumption summarizes the identifying restrictions used in the general multi-period dynamic RD model.

Assumption 3.1 (Longer-term ATEs) *There exists $\epsilon > 0$ such that for all $z_1 \in \mathcal{N}_\epsilon$, we have:*

1. *(Markovian) for any $k = 3, 4, ..., K$, $\ell^{k-2} \in \mathcal{L}^{k-2}$, and $d = 0, 1$, immediate effects and τ-period-after total effects satisfy that $E\left[\theta^{(\ell^{k-2},d)}_{i;0,k} | D_{ik}(\ell^{k-2}, d) = 1, Z_{i1} = z_1\right] = E\left[\theta^{d}_{i;0,2} | D_{i2}(d) = 1, Z_{i1} = z_1\right] \equiv \mu^d_0$ and $E\left[\tilde{\theta}^{(\ell^{k-2},d)}_{i;\tau,k} | D_{ik}(\ell^{k-2}, d) = 1, Z_{i1} = z_1\right] = E\left[\tilde{\theta}^{d}_{i;\tau,2} | D_{i2}(d) = 1, Z_{i1} = z_1\right] \equiv \tilde{\mu}^d_\tau$, for all $\tau = 1, ..., K-k$; similar conditions also hold for immediate and long-term first-stage effects;*

2. *(CIA: multi-period) for any $d_1 = 0, 1$, $z_2 \in \mathbb{R}$, and $x \in \mathcal{X}$, $E[M_i(d_1, 0)|X_i = x, Z_{i2}(d_1) = z_2, S_{i2}(d_1) = 1, Z_{i1} = z_1] = E[M_i(d_1, 0)|X_i = x, S_{i2}(d_1) = 1, Z_{i1} = z_1]$, where the random variable $M_i(d_1, 0)$ can be $D_{i3}(d_1, 0)$, $\tilde{Y}_{i(2+\tau)}(d_1, 0)$ for all $\tau = 1, \cdots, K-2$, or $\tilde{D}_{i(3+\tau)}(d_1, 0)$ for all $\tau = 1, \cdots, K-3$;*

3. *(Smoothness: multi-period) for all $d_1, d_2 = 0, 1$ and $x \in \mathcal{X}$, $E[M_i(d_1, d_2)|X_i = x, D_{i2}(d_1) = d_2, Z_{i1} = z_1]$ is continuous in z_1, where the random variable $M_i(d_1, d_2)$ can be $D_{i3}(d_1, d_2)$, $\tilde{Y}_{i(2+\tau)}(d_1, d_2)$ for all $\tau = 1, \cdots, K-2$, or $\tilde{D}_{i(3+\tau)}(d_1, d_2)$ for all $\tau = 1, \cdots, K-3$;*

[6] When $\tau = 2$, for example, the decomposition underlying the recursive CFR identification strategy is $\tilde{\theta}_{i;2,1} = \theta_{i;2,1} + \theta^1_{i;1,2} D_{i2}(1) + \theta^{(1,0)}_{i;0,3}(1 - D_{i2}(1)) D_{i3}(1, 0) + \theta^{(1,1)}_{i;0,3} D_{i2}(1) D_{i3}(1, 1) - \theta^0_{i;1,2} D_{i2}(0) - \theta^{(0,0)}_{i;0,3}(1 - D_{i2}(0)) D_{i3}(0, 0) + \theta^{(0,1)}_{i;0,3} D_{i2}(0) D_{i3}(0, 1)$.

[7] Let $\tilde{D}_{i(k+1+\tau)}(\ell^k)$ be the $(k+1+\tau)$-th period quasi-potential treatment decision with only the first k rounds of previous treatment status specified, for any $\tau \geq 1$ and $k \geq 1$. Then, $\eta_{i;0,1} = D_{i2}(1) - D_{i2}(0)$, $\eta_{i;\tau,1} = D_{i(2+\tau)}(1, \mathbf{0}_\tau) - D_{i(2+\tau)}(0, \mathbf{0}_\tau)$, and $\tilde{\eta}_{i;\tau,1} = \tilde{D}_{i(2+\tau)}(1) - \tilde{D}_{i(2+\tau)}(0)$ for all $\tau \geq 1$. For all $k \geq 2$, $\eta^{\ell^{k-1}}_{i;0,k} = D_{i(k+1)}(\ell^{k-1}, 1) - D_{i(k+1)}(\ell^{k-1}, 0)$, $\eta^{\ell^{k-1}}_{i;\tau,k} = D_{i(k+1+\tau)}(\ell^{k-1}, 1, \mathbf{0}_\tau) - D_{i(k+1+\tau)}(\ell^{k-1}, 0, \mathbf{0}_\tau)$, and $\tilde{\eta}^{\ell^{k-1}}_{i;\tau,k} = \tilde{D}_{i(k+1+\tau)}(\ell^{k-1}, 1) - \tilde{D}_{i(k+1+\tau)}(\ell^{k-1}, 0)$ for all $\tau \geq 1$.

The second and third parts of Assumption 3.1 are vanilla extensions of Assumption 2.2 viewing $D_{i3}(d_1, d_2)$, $\tilde{Y}_{i(2+\tau)}(d_1, d_2)$ and $\tilde{D}_{i(2+\tau)}(d_1, d_2)$ as potential outcomes associated with the first two treatments but taking place in a later period. The key new condition in Assumption 3.1 is the Markovian restriction. In the California education bond application, the Markovian restriction allows an education bond's various immediate and long-term average effects to depend arbitrarily on last election cycle's bond authorization but not any other bond authorizations further in the past. Although non-trivial, the Markovian condition is much less restrictive than the homogeneous ATE condition used in Lemma 2.1 (or Lemma A.1) for the recursive CFR identification strategy.

Similar Markovian-type restrictions are also used in De Chaisemartin and d'Haultfoeuille (forthcoming) and Imai et al. (2023) for non-RD dynamic treatment effect settings. When $\tau = 2$, the Markovian assumption together with the decomposition in (3.2) imply that

$$\begin{aligned} E[\tilde{\theta}_{i;2,1}|Z_{i1} = 0] =& E[\theta_{i;2,1}|Z_{i1} = 0] + \tilde{\mu}_1^1 \cdot E[D_{i2}(1)|Z_{i1} = 0] - \tilde{\mu}_1^0 \cdot E[D_{i2}(0)|Z_{i1} = 0] \\ &+ \mu_0^0 \cdot E[\eta_{i;1,1}|Z_{i1} = 0]. \end{aligned}$$

Lemma 3.2 *Under Assumptions used in Lemma 2.2, Assumption A.1 for identifying longer-term average total effects, and Assumption 3.1, we have that for* $\tau = 2, ..., K-1$*:*

$$\begin{aligned} E[\theta_{i;\tau,1}|Z_{i1} = 0] = & \lim_{z_1 \searrow 0} E[Y_{i(\tau+1)}|Z_{i1} = z_1] - \lim_{z_1 \nearrow 0} E[Y_{i(\tau+1)}|Z_{i1} = z_1] \\ & - \tilde{\mu}_{\tau-1}^1 \cdot \lim_{z_1 \searrow 0} E[D_{i2}|Z_{i1} = z_1] + \tilde{\mu}_{\tau-1}^0 \cdot \lim_{z_1 \nearrow 0} E[D_{i2}|Z_{i1} = z_1] \\ & - \sum_{s=1}^{\tau-2} \tilde{\mu}_s^0 \cdot E[\eta_{i;\tau-1-s,1}|Z_{i1} = 0] - \mu_0^0 \cdot E[\eta_{i;\tau-1,1}|Z_{i1} = 0], \end{aligned} \tag{3.3}$$

where $\mu_0^0 = \lim_{z_1 \nearrow 0} E\left[\frac{Y_{i2} S_{i2}(D_{i2} - \lambda^0(X_i))}{(1-\lambda^0(X_i))E[D_{i2}|Z_{i1}=z_1]}|Z_{i1} = z_1\right]$, $\tilde{\mu}_s^0 = \lim_{z_1 \nearrow 0} E\left[\frac{Y_{i(2+s)} S_{i2}(D_{i2} - \lambda^0(X_i))}{(1-\lambda^0(X_i))E[D_{i2}|Z_{i1}=z_1]}|Z_{i1} = z_1\right]$, *and* $\tilde{\mu}_s^1 = \lim_{z_1 \searrow 0} E\left[\frac{Y_{i(2+s)} S_{i2}(D_{i2} - \lambda^1(X_i))}{(1-\lambda^1(X_i))E[D_{i2}|Z_{i1}=z_1]}|Z_{i1} = z_1\right]$*, for all* $s \geq 1$*. In addition, treating first-stage decisions as outcomes,* $E[\eta_{i;1,1}|Z_1 = 0]$ *is identified by Lemma 2.2 and* $E[\eta_{i;k,1}|Z_1 = 0]$ *for all* $k = 2, ..., \tau - 1$ *is identified by* (3.3) *recursively.*

The proof is given in the online appendix. In Section A.4 of the online appendix, we also discuss several important special cases of the dynamic RD model, including having an absorbing state treatment and not observing some initial rounds of RD data.

4 Estimation and Inference

Estimation of the one-period-after ATE is discussed in Section 2.4 for the benchmark two-period model. In the next, we first study inference of the proposed one-period-after

ATE estimator. Then we extend the two-period estimation and inference strategy to the general multi-period setting introduced in Section 3.

4.1 Inference in the Benchmark Two-period Setting

The inference procedure proposed in this section for the one-period-after ATE estimator defined in Section 2.4 adapts the weighted bootstrap procedure following Ma and Kosorok (2005). The procedure is tractable in empirical applications as it keeps estimation and inference of different long-term ATEs within a uniform format. The procedure does not pursue bandwidth choice optimality in the sense of asymptotic mean squared errors (AMSE), however. In the online appendix, we discuss an alternative AMSE-optimal estimation and inference procedure for the one-period-after ATE based on Calonico et al. (2014, 2018, 2020, 2022).[8] The weighted bootstrap procedure discussed in this section could also be adapted for recursive CFR estimators.

4.1.1 Assumptions and Asymptotic Properties

Let $\phi_{\gamma^1,ni}(D_{2i}, S_{2i}, Z_{1i}, X_i)$ and $\phi_{\gamma^0,ni}(D_{2i}, S_{2i}, Z_{1i}, X_i)$ be influence functions of estimators $\hat{\gamma}^1$ and $\hat{\gamma}^0$ defined in Section 2.4, respectively, such that:

$$\sqrt{nh}(\hat{\gamma}^1 - \gamma^1) = \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \phi_{\gamma^1,ni}(D_{2i}, S_{2i}, Z_{1i}, X_i) + o_p(1),$$

$$\sqrt{nh}(\hat{\gamma}^0 - \gamma^0) = \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \phi_{\gamma^0,ni}(D_{2i}, S_{2i}, Z_{1i}, X_i) + o_p(1).$$

Let $\tilde{\phi}_{\alpha^1,ni}(Y_{2i}, D_{2i}, S_{2i}, Z_{1i}, X_i)$ and $\tilde{\phi}_{\alpha^0,ni}(Y_{2i}, D_{2i}, S_{2i}, Z_{1i}, X_i)$ be influence functions of infeasible estimators $\tilde{\alpha}^1$ and $\tilde{\alpha}^0$ defined by the following:

$$(\tilde{\alpha}^1, \tilde{\beta}^1) = \arg\min_{\alpha,\beta} \sum_{i=1}^{n} 1(Z_{1i} \geq 0) K\left(\frac{Z_{1i}}{h}\right) \left[A_i(\alpha, \beta; \gamma^1)\right]^2,$$

$$(\tilde{\alpha}^0, \tilde{\beta}^0) = \arg\min_{\alpha,\beta} \sum_{i=1}^{n} 1(Z_{1i} < 0) K\left(\frac{Z_{1i}}{h}\right) \left[A_i(\alpha, \beta; \gamma^0)\right]^2.$$

The infeasible estimators are generated by the same local linear regressions as used in Section 2.4 for estimators $\hat{\alpha}^1$ and $\hat{\alpha}^0$, but assuming known first-step propensity scores.

[8] Extending the robust inference procedure to longer-term ATE estimators could be realized through calculating first-order linear approximations of each τ-period-after estimator (e.g., Sections 4.1 and 4.2 of Calonico et al., 2014) for bias correction and optimal bandwidth choice calculation. It would be an interesting topic for future research.

Define gradient terms $\nabla_\gamma^1 = \lim_{z_1 \searrow 0} E\left[\nabla_\gamma \left[\frac{Y_2 S_2 (D_2 - p(X,\gamma))}{1 - p(X,\gamma)}\right] |_{\gamma=\gamma^1} \middle| Z_1 = z_1\right]$ and $\nabla_\gamma^0 = \lim_{z_1 \nearrow 0} E\left[\nabla_\gamma \left[\frac{Y_2 S_2 (D_2 - p(X,\gamma))}{1 - p(X,\gamma)}\right] |_{\gamma=\gamma^0} \middle| Z_1 = z_1\right]$. By the delta method, we obtain the following influence function representation of $\hat{\alpha}^0$ and $\hat{\alpha}^1$ such that for $d_1 = 0, 1$, we have:

$$
\begin{aligned}
&\sqrt{nh}(\hat{\alpha}^{d_1} - \alpha^{d_1}) \\
=&\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \left(\tilde{\phi}_{\alpha^{d_1},ni}(Y_{2i}, S_{2i}, D_{2i}, Z_{1i}, X_i) - \nabla_\gamma^{d_1} \cdot \phi_{\gamma^{d_1},ni}(D_{2i}, S_{2i}, Z_{1i}, X_i)\right) + o_p(1) \\
\equiv&\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \phi_{\alpha^{d_1},ni}(Y_{2i}, S_{2i}, D_{2i}, Z_{1i}, X_i) + o_p(1). \qquad (4.1)
\end{aligned}
$$

The representation implies that the asymptotic variance of $\hat{\hat{\theta}}_{1,1}$ could be estimated by:

$$
\hat{V}_{11} = \frac{1}{nh} \sum_{i=1}^{n} \left(\hat{\phi}_{\alpha^1,ni}(Y_{2i}, D_{2i}, S_{2i}, Z_{1i}, X_i)\right)^2 + \left(\hat{\phi}_{\alpha^0,ni}(Y_{2i}, D_{2i}, S_{2i}, Z_{1i}, X_i)\right)^2,
$$

where $\hat{\phi}_{\alpha^{d_1},ni}(Y_{2i}, D_{2i}, S_{2i}, Z_{1i}, X_i)$ is the estimated version of $\phi_{\alpha^{d_1},ni}(Y_{2i}, D_{2i}, S_{2i}, Z_{1i}, X_i)$ with all unknown parameters replaced with corresponding estimators; $d_1 = 0, 1$.

We next provide detailed assumptions and asymptotic properties of the proposed two-step estimator with respect to a varying coefficient Logit first stage where $p(x, \gamma) = L((1\ x')\gamma)$ with $L(a) = \exp(a)/(1 + \exp(a))$.

Assumption 4.1 *$\lambda(x; z_1) = L((1\ x')\gamma(z_1))$ is the correct specification on $z_1 \in \mathcal{N}_\epsilon$ for some $\epsilon > 0$.*

Assumption 4.2 *Density $f_{z_1}(z_1)$ is twice continuously differentiable in z_1 on $\mathcal{N}_\epsilon$, and $f_{z_1}(z_1)$ is bounded away from zero on $\mathcal{N}_\epsilon$ for some $\epsilon > 0$.*

Assumption 4.3 *Assume that:*

1. *The kernel function $K(\cdot)$ is a non-negative symmetric bounded kernel with support $[-1, 1]$; $\int K(u)du = 1$.*

2. *The bandwidth satisfies that $h \to 0$, $nh^3 \to \infty$, and $nh^5 \to 0$ as $n \to \infty$.*

Assumption 4.1 requires that the varying coefficient Logit model is correctly specified. Assumption 4.2 imposes standard smoothness conditions on the density of the running variable. Assumption 4.3 imposes standard conditions on the kernel function and undersmoothed bandwidth. Undersmoothing is required such that the bias of kernel estimators becomes asymptotically negligible. In practice, we recommend using the triangular kernel (i.e., $K(x) = |x| \cdot 1(|x| < 1)$) and under-smoothing the robust RD bandwidth introduced in Calonico et al. (2014) (CCT), which is of order $n^{1/5}$. As discussed earlier, under-smoothing is not AMSE optimal. In the online appendix, we propose an alternative procedure that uses bias correction and AMSE-optimal bandwidth in the second step. The alternative procedure also requires a higher-order local polynomial and a larger bandwidth in the first step such that estimation errors from the first step do not affect AMSE of the final estimator.

The following lemma provides asymptotic properties of the first-step varying coefficient Logit estimators. Similar results could be derived if the propensity score function $\lambda(.;.)$ follows other semi-parametric models such as varying coefficient Probit.

Lemma 4.1 *Suppose that Assumptions 4.1-4.3 and B.1-B.2 hold, then for $d = 0, 1$, we have:*

$$\sqrt{nh}\begin{pmatrix} \hat{\gamma}^d - \gamma^d \\ h\hat{\beta}^d_{FS} - h\beta^d \end{pmatrix} = \frac{1}{\sqrt{nh}}\sum_{i=1}^{n}(\Delta^d)^{-1}S_{2i} \cdot 1(Z_{1i} \geq 0)^d \cdot 1(Z_{1i} < 0)^{1-d}$$

$$\cdot K\left(Z_{1i}/h\right)\left(D_{2i} - L\big(X_i'(\gamma^d + \beta^d Z_{1i})\big)\right)\begin{pmatrix} X_i \\ Z_{1i}X_i/h \end{pmatrix} + o_p(1),$$

where Δ^d is given in equation (D.2) *in the online appendix. In addition, for $d = 0, 1$, we have:*

$$\sqrt{nh}\begin{pmatrix} \hat{\gamma}^d - \gamma^d \\ h\hat{\beta}^d_{FS} - h\beta^d \end{pmatrix} \Rightarrow N\left(0, \left(\Delta^d\right)^{-1}\Omega^d\left(\Delta^d\right)^{-1}\right),$$

where Ω^d is given in equation (D.3) *in the online appendix.*

Aside from Assumptions stated earlier, the lemma also requires Assumptions B.1 and B.2 in the online appendix. The former imposes smoothness conditions on the varying coefficient in neighborhoods right above and below the RD cutoff. The latter imposes moment conditions on the conditioning covariates.

For asymptotic properties of $\hat{\alpha}^0$ and $\hat{\alpha}^1$, we impose additional Assumptions B.3 and B.4, stated in the online appendix. The former includes smoothness conditions for the infeasible estimators $\tilde{\alpha}^0$ and $\tilde{\alpha}^1$ using true values of first-step propensity score functions. The latter imposes conditions that control the impact of first-step estimation errors on the asymptotic properties of feasible two-step estimators $\hat{\alpha}^0$ and $\hat{\alpha}^1$. Given the assumptions, $\sqrt{nh}(\hat{\alpha}^{d_1} - \alpha^{d_1})$, for $d_1 = 0, 1$, has linear representations as in (4.1) with:

$$\tilde{\phi}_{\alpha^{d_1},ni}(Y_{2i}, D_{2i}, S_{2i}, Z_{1i}, X_i) = (1 \;\; 0) \cdot \Delta_z^{-1} \cdot 1(Z_{1i} \geq 0)^{d_1} \cdot 1(Z_{1i} < 0)^{1-d} \cdot K\left(Z_{1i}/h\right)$$
$$\cdot \Big(Y_{2i} - E[Y_{2i}|Z_{1i}] - \frac{Y_{2i}S_{2i}(D_{2i} - L(X_i'\gamma^{d_1}))}{1 - L(X_i'\gamma^{d_1})} + E\Big[\frac{Y_{2i}S_{2i}(D_{2i} - L(X_i'\gamma^{d_1}))}{1 - L(X_i'\gamma^{d_1})}\Big|Z_{1i}\Big]\Big) \begin{pmatrix} 1 \\ Z_{1i}/h \end{pmatrix},$$

where $\Delta_z = f_{z_1}(0) \cdot \begin{pmatrix} \mu_{z,0} & \mu_{z,1} \\ \mu_{z,1} & \mu_{z,2} \end{pmatrix}$, and $\mu_{z,j} = \int_{u \geq 0} u^j K(u) du$ for $j = 1, 2, \ldots$. The influence function representation then implies the following asympototic results.

Theorem 4.1 *Suppose that Assumptions 4.1-4.3 and B.1-B.4 hold. For* $d_1 = 0, 1$*, we then have:*

$$\sqrt{nh}\left(\hat{\alpha}^{d_1} - \alpha^{d_1}\right) \xrightarrow{d} N(0, V_{\alpha^{d_1}}), \quad d_1 = 0, 1;$$
$$\sqrt{nh}(\hat{\bar{\theta}}_{1,1} - \bar{\theta}_{1,1}) \xrightarrow{d} N(0, V_{\alpha^1} + V_{\alpha^0}),$$

where $V_{\alpha^0} = \lim_{n\to\infty} \lim_{z_1 \nearrow 0} h^{-1} E[\phi^2_{\alpha^0,ni}(Y_{2i}, D_{2i}, S_{2i}, Z_{1i}, X_i)|Z_{1i} = z_1]$ *and* $V_{\alpha^1} = \lim_{n\to\infty} \lim_{z_1 \searrow 0} h^{-1} E[\phi^2_{\alpha^1,ni}(Y_{2i}, D_{2i}, S_{2i}, Z_{1i}, X_i)|Z_{1i} = z_1]$.

The exact expressions of V_{α^0} and V_{α^1} are tedious to calculate in general. In the next section, we adapt the weighted bootstrap procedure first introduced in Ma and Kosorok (2005) to simulate the limiting distribution of the proposed estimator. Studies adopting the procedure in other settings include Chen and Pouzo (2009), Chernozhukov et al. (2015a,b) and Fernàndez-Val et al. (2021), among many others. Our paper is the first to apply weighted bootstrap to kernel-based boundary estimation, which is the main workhorse of the RD literature.

4.1.2 Weighted Bootstrap Inference

Let $\{W_i\}_{i=1}^n$ be a sequence of pseudo random variables independent of the sample path with unit mean and variance. Define the weighted bootstrap estimator for $\bar{\theta}_{1,1}$ as:

$$\hat{\bar{\theta}}^w_{1,1} = \hat{\alpha}^{1,w} - \hat{\alpha}^{0,w},$$
$$(\hat{\alpha}^{1,w}, \hat{\beta}^{1,w}) = \arg\min_{\alpha,\beta} \sum_{i=1}^n W_i \cdot 1(Z_{1i} \geq 0) K\left(\frac{Z_{1i}}{h}\right)\left[A_i(\alpha, \beta; \hat{\gamma}^{1,w})\right]^2,$$
$$(\hat{\alpha}^{0,w}, \hat{\beta}^{0,w}) = \arg\min_{\alpha,\beta} \sum_{i=1}^n W_i \cdot 1(Z_{1i} < 0) K\left(\frac{Z_{1i}}{h}\right)\left[A_i(\alpha, \beta; \hat{\gamma}^{0,w})\right]^2,$$

where:

$$(\hat{\gamma}^{1,w}, \hat{\beta}_{FS}^{1,w}) = \arg\max_{\gamma,\beta} \sum_{i=1}^{n} W_i \cdot S_{2i} 1(Z_{1i} \geq 0) K\left(\frac{Z_{1i}}{h}\right) \cdot \Big[D_{2i} \log p(X_i, \gamma + \beta Z_{1i}) + (1 - D_{2i}) \cdot \log(1 - p(X_i, \gamma + \beta Z_{1i}))\Big],$$

$$(\hat{\gamma}^{0,w}, \hat{\beta}_{FS}^{0,w}) = \arg\max_{\gamma,\beta} \sum_{i=1}^{n} W_i \cdot S_{2i} 1(Z_{1i} < 0) K\left(\frac{Z_{1i}}{h}\right) \cdot \Big[D_{2i} \log p(X_i, \gamma + \beta Z_{1i}) + (1 - D_{2i}) \cdot \log(1 - p(X_i, \gamma + \beta Z_{1i}))\Big].$$

The weighted bootstrap procedure is simple to carry out. Given a simulated copy of $\{W_i\}_{i=1}^n$, weighted bootstrap repeats the original two-step kernel-based estimation procedure (i.e., first-step local MLE propensity score estimation and second-stage local linear regressions), replacing the original kernel weights $K(Z_{1i}/h)$ with new weights $W_i \cdot K(Z_{1i}/h)$. The procedure could also be easily extended to adjust for within-cluster correlations by assigning random W_i at the cluster level.

Following Ma and Kosorok (2005), $\sqrt{nh}(\hat{\hat{\theta}}_{1,1}^w - \hat{\hat{\theta}}_{1,1})$ and $\sqrt{nh}(\hat{\hat{\theta}}_{1,1} - \bar{\theta}_{1,1})$ have the same limiting distribution under suitable conditions. The next theorem formalizes the validity of the weighted bootstrap estimator given a varying coefficient Logit first stage.

Theorem 4.2 *Suppose that Assumptions 4.1-4.3 and B.1-B.4 hold and that $\{W_i\}_{i=1}^n$ is a sequence of i.i.d. pseudo random variables independent of the sample path with $E[W_i] = Var[W_i] = 1$ for all i. We then have:*

$$\sqrt{nh}(\hat{\hat{\theta}}_{1,1}^w - \hat{\hat{\theta}}_{1,1}) \xrightarrow{d} N(0, V_{\alpha^1} + V_{\alpha^0})$$

conditional on sample path with probability approaching one.

Although W_i can follow any distribution with unit mean and variance, in the simulation and empirical sessions where the first-step propensity score function is modeled with varying-coefficient Logit, we use a discrete distribution where $W_i = 0.5$ or 3 with probabilities 0.8 and 0.2, respectively. The binary random variable with positive support ensures that the weighted Logit objective functions remain globally concave.

4.2 Estimation and Inference in the General Multi-period Setting

Long-term ATEs in the multi-period setting, or $E[\theta_{i;\tau,1}|Z_{i1} = 0]$ for $\tau \geq 2$, are identified in Lemma 3.2. This section elaborates their estimation and inference.

Let us start with case of $\tau = 2$. To estimate the two-period-after ATE, or $\bar{\theta}_{2,1} \equiv E[\theta_{i;2,1}|Z_{i1} = 0]$, one can first estimate the one-period-after average first-stage effect

$\bar{\eta}_{1,1} \equiv E[\eta_{i;1,1}|Z_{i1} = 0]$ following the estimation procedure for one-period-after ATE, or $\bar{\theta}_{1,1}$, described in Section 2.4, treating first-stage treatment decisions as outcomes. Then, one can estimate other components of equation (3.3). Similarly, the estimation of any τ-period-after ATE for $\tau \geq 3$ involves estimating the average first-stage effect $\bar{\eta}_{k,1} \equiv E[\eta_{i;k,1}|Z_{i1} = 0]$ for all $k = 1, 2, \cdots, \tau - 1$, which can be done recursively following Lemma 3.2, and separate estimation of other components in equation (3.3).

We formally define the estimator of $\bar{\theta}_{2,1}$. Rewriting Lemma 3.2 with $\tau = 2$ gives

$$\bar{\theta}_{2,1} = \alpha_1^1 - \alpha_1^0 - (\tilde{\mu}_{nu}^0/\tilde{\mu}_{de})\bar{\eta}_{1,1}, \text{ where } \tilde{\mu}_{nu}^0 = \lim_{z_1 \nearrow 0} E\left[\frac{Y_{i2}S_{i2}(D_{i2} - \lambda^0(X_i))}{1 - \lambda^0(X_i)}|Z_{i1} = z_1\right],$$
$$\tilde{\mu}_{de} = E[D_{i2}|Z_{i1} = z_1], \ \alpha_1^1 = \lim_{z_1 \searrow 0} E\left[Y_{i3} - \frac{Y_{i3}S_{i2}(D_{i2} - \lambda^1(X_i))}{1 - \lambda^1(X_i)}|Z_{i1} = z_1\right],$$
$$\alpha_1^0 = \lim_{z_1 \nearrow 0} E\left[Y_{i3} - \frac{Y_{i3}S_{i2}(D_{i2} - \lambda^0(X_i))}{1 - \lambda^0(X_i)}|Z_{i1} = z_1\right].$$

Let $\hat{\alpha}_1^1$, $\hat{\alpha}_1^0$, $\hat{\tilde{\mu}}_{nu}^0$, $\hat{\tilde{\mu}}_{de}$, $\hat{\alpha}_{fs}^1$, and $\hat{\alpha}_{fs}^0$ be estimators of α_1^1, α_1^0, $\tilde{\mu}_{nu}^0$, $\tilde{\mu}_{de}$, α_{fs}^1, and α_{fs}^0, respectively. All of them can be defined using the same two-step semi-parametric procedure described in Section 2.4 for the estimation of $\hat{\alpha}^0$ and $\hat{\alpha}^1$, where the first-step propensity estimation uses kernel-based local MLE. Let $\phi_{\alpha_1^1,ni}$, $\phi_{\alpha_1^0,ni}$, $\phi_{\tilde{\mu}_{nu}^0,ni}$, $\phi_{\tilde{\mu}_{de},ni}$, $\phi_{\alpha_{fs}^1,ni}$, and $\phi_{\alpha_{fs}^0,ni}$ denote the influence functions of the estimators, respectively. Definitions of $\phi_{\alpha_1^1,ni}$, $\phi_{\alpha_1^0,ni}$, $\phi_{\tilde{\mu}_{nu}^0,ni}$, $\phi_{\alpha_{fs}^1,ni}$, and $\phi_{\alpha_{fs}^0,ni}$ are similar to influence functions given in equation (4.1) for $\hat{\alpha}^0$ and $\hat{\alpha}^1$. The influence function for $\hat{\tilde{\mu}}_{de}$ is defined as $\phi_{\tilde{\mu}_{de},ni} = \frac{1}{f_{z_1}(0)} K\left(\frac{Z_{1i}}{h}\right)(D_{2i} - E[D_{2i}|Z_{1i}])$.

Let $\hat{\bar{\theta}}_{2,1} = \hat{\alpha}_1^1 - \hat{\alpha}_1^0 - \hat{\tilde{\mu}}_{nu}^0/\hat{\tilde{\mu}}_{de}(\hat{\alpha}_{fs}^1 - \hat{\alpha}_{fs}^0)$ be the estimator of $\bar{\theta}_{2,1}$. By the delta method, we have $\sqrt{nh}(\hat{\bar{\theta}}_{2,1} - \bar{\theta}_{2,1}) = \frac{1}{\sqrt{nh}} \sum_{i=1}^n \phi_{\bar{\theta}_{2,1},ni}(Y_{3i}, Y_{2i}, D_{3i}, D_{2i}, S_{2i}, Z_{1i}, X_i) + o_p(1)$, where $\phi_{\bar{\theta}_{2,1},ni}(Y_{3i}, Y_{2i}, D_{3i}, D_{2i}, S_{2i}, Z_{1i}, X_i) = \phi_{\alpha_1^1,ni} - \phi_{\alpha_1^0,ni} - \frac{\alpha_{fs}^1 - \alpha_{fs}^0}{\tilde{\mu}_{de}}\phi_{\tilde{\mu}_{nu}^0,ni} + \frac{\tilde{\mu}_{nu}^0(\alpha_{fs}^1 - \alpha_{fs}^0)}{(\tilde{\mu}_{de})^2}\phi_{\tilde{\mu}_{de},ni} - \frac{\tilde{\mu}_{nu}^0}{\tilde{\mu}_{de}}(\phi_{\alpha_{fs}^1,ni} - \phi_{\alpha_{fs}^0,ni})$. The asymptotic normality of $\hat{\bar{\theta}}_{2,1}$ follows from the influence function representation. Given the influence function representation, it is easy to see that the weighted bootstrap procedure could be applied here too, and, in general, to the inference problem of any τ-period-after ATE with $\tau \geq 2$.

Monte Carlo simulations for the proposed estimation and inference procedure are summarized in the online appendix.

5 Empirical Example: The Effect of CA School Bonds

This section revisits the study of local education bonds using the dataset published by CFR. As described in CFR, school districts in California became eligible for issuing general obligation bonds through Proposition 46 in 1984. CFR study the effects of bond authorization on local house prices, student achievements, and other outcomes using California data from 1987 to 2006. Due to data limitations, we study two outcome variables: total expenditure per pupil and capital loading per pupil in a school district.

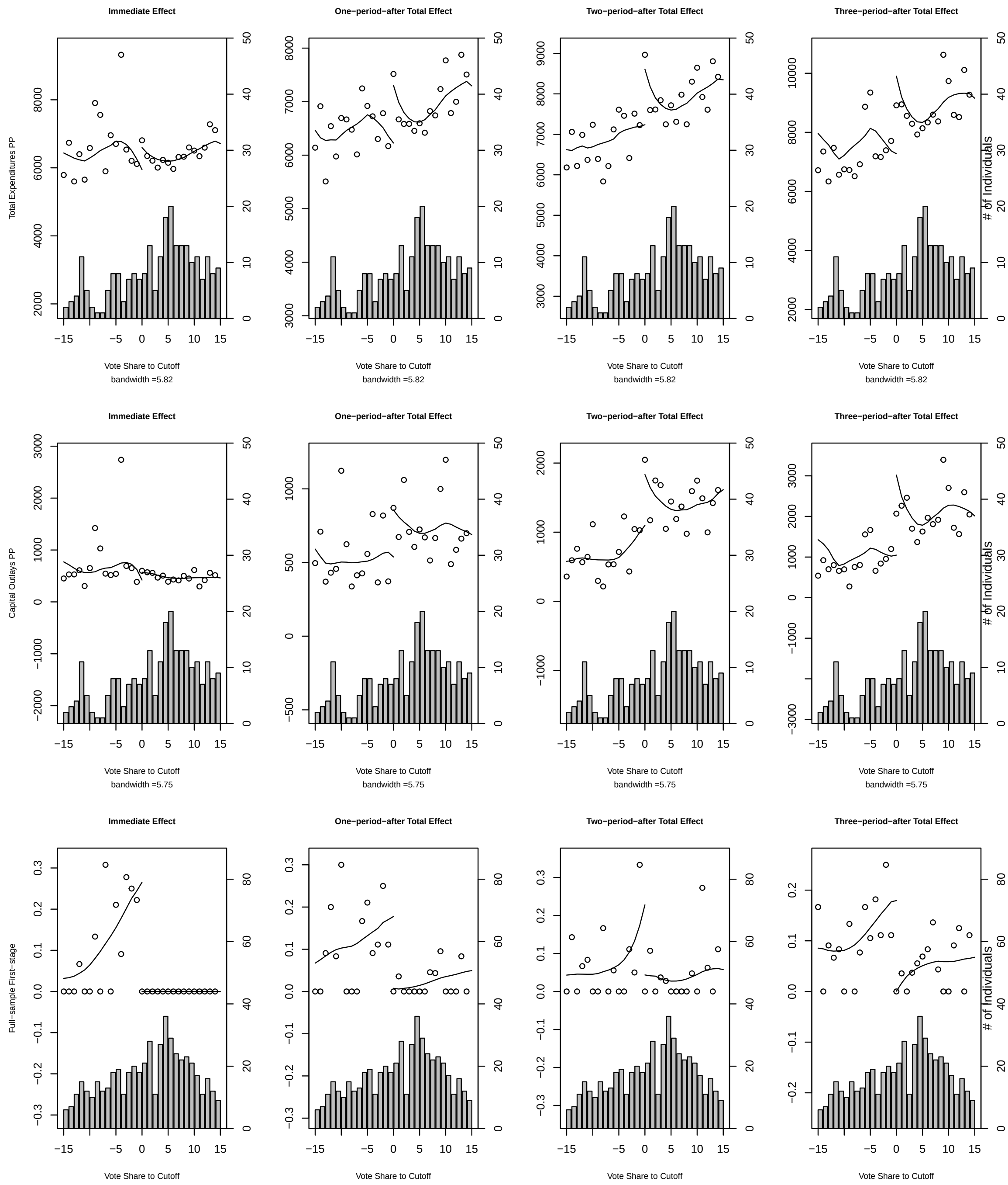
There are 1551 bond measures from 655 school districts with non-missing vote share information. Although the dataset starts from the first years of education bonds, the expenditure outcomes we use are not observed until 1995. Among all the bond measures, 282 are proposed after five or more years of inaction (no measure) and have non-missing expenditure outcome data up to four periods after the measure year. We focus on long-term effects of these bond measures.

Figure 1 gives a visual illustration of the data. The bottom row of the figure shows immediate and long-term total first-stage effects of passing an education bond measure. No school districts that barely passed the vote share cutoff in the first trial authorized another bond the following year. The probabilities increase slightly in years three and four. Around 27% of school districts that barely missed the vote share cutoff in the first trial successfully authorized their first education bond in the following year. The negative first-stage effects imply that long-term average total effects observed directly from outcome discontinuity are smaller than the long-term ATEs of policy interests.

Table 1 reports non-parametric estimation results of immediate and long-term average direct effects following the extended recursive CFR strategy formalized in Lemmas 2.1 and A.1. The estimated average effects for the capital outlays outcome are highly statistically significant. Effects of the total expenditures outcome are only marginally significant. Estimates in Table 1 are larger than those reported in Table 4 of CFR. The difference comes from two sources. First, Table 4 of CFR is estimated by pooling different rounds of RDs. Under treatment effect heterogeneity, such a pooling strategy is not appropriate, because marginal individuals at different rounds of RD cutoffs have different treatment effects. Second, Table 4 of CFR is estimated parametrically with global polynomials while Table 1 is estimated non-parametrically using the local linear method.

Table 2 reports estimation results following the proposed method. Varying-coefficient Logit is used in first-step propensity score estimation. Panel A of the table reports estimation and inference results when the CIA condition in Section 2 holds without any conditioning covariate. Panel B of the table reports results when the CIA condition holds when the first-period outcome is used as the conditioning covariate. The estimates are semi-parametric and rely on the first-step varying-coefficient Logit functional form of propensity scores. Estimates of the two panels share the same patterns. Compared

Figure 1: Average Total Effects and Histograms



Notes: Dataset is from Cellini et al. (2010). The kernel bandwidth of each row is set to the same value, which is the average of CCT bandwidths among all four RD regressions of the row. The data sample of the first two rows is a subset of that of the last row, because of missing values in the expense outcomes.

Table 1: Average Direct Effects: Non-parametric CFR based on Lemmas 2.1 and A.1

Immediate		One-period-after		Two-period-after		Three-period-after	
k=4.25	4.5	4.25	4.5	4.25	4.5	4.25	4.5
Total expenditures per pupil							
897	767	$1,502^{*}$	$1,385^{*}$	$2,101$	$2,025$	$3,581^{*}$	$3,626^{**}$
(664)	(601)	(885)	(826)	$(1,384)$	$(1,294)$	$(1,888)$	$(1,775)$
Capital outlays per pupil							
361	275	497^{**}	455^{**}	$1,002^{*}$	977^{**}	$2,379^{**}$	$2,399^{***}$
(248)	(216)	(203)	(186)	(569)	(497)	$(1,026)$	(898)

Notes: Dataset is from Cellini et al. (2010). Standard errors are calculated using weighted bootstrap with 1,000 bootstrap repetitions. Undersmoothed CCT bandwidth is calculated following suggestions in Section E of the online appendix, with the CCT bandwidth reported in Figure 1.

to the results of Table 1, long-term ATE estimates for the total expenditure outcome become smaller, although the estimation is not very precise. Long-term ATE estimates for the capital outlays outcome, on the other hand, are larger than those reported in Table 1 and remain highly statistically significant.

6 Conclusion

Static RD models with a single eligibility test have been very popular in the last two decades. Recently, more empirical studies have targeted situations where individuals are eligible for repeated RD treatments. Most studies under such a setting have either ignored dynamics in repeated RD models or employed restrictive identifying assumptions. This paper is the first to employ the conventional potential outcome framework to identify long-term average direct treatment effects. Our proposed identification strategy allow explicitly for treatment effect heterogeneity under a conditional independence assumption and does not impose any assumption on endogenous RD participation decisions of later rounds. For estimation and inference, we propose a novel multi-step semi-parametric estimation procedure. Employing the proposed method, we revisit the study of local education bonds following CFR. We find much larger long-term average direct effects of education bonds on the capital outlays per pupil outcome.

Table 2: Average Direct Effects: the Proposed Procedure

Immediate		One-period-after		Two-period-after		Three-period-after	
k=4.25	4.5	k=4.25	4.5	4.25	4.5	4.25	4.5
Panel A:							
Total expenditures per pupil							
897	767	1,070	959	1,440	1,266	2,532	2,331
(664)	(601)	(776)	(716)	(1,453)	(1,244)	(2,874)	(1,992)
Capital outlays per pupil							
361	275	405**	398***	1,245***	1,240***	2,850***	2,884***
(248)	(216)	(158)	(149)	(448)	(401)	(882)	(760)
Panel B:							
Total expenditures per pupil							
897	767	903	808	1,016	863	1,682	1,511
(664)	(601)	(964)	(882)	(2,073)	(1,863)	(4,684)	(3,629)
Capital outlays per pupil							
361	275	433**	416**	1,387***	1,364***	3,176***	3,195***
(248)	(216)	(176)	(164)	(495)	(472)	(1,156)	(1,146)

Notes: Dataset is from Cellini et al. (2010). Standard errors are calculated using weighted bootstrap with 1,000 bootstrap repetitions. Undersmoothed CCT bandwidth is calculated following suggestions in Section E of the online appendix, with the CCT bandwidth reported in Figure 1. Panel A reports estimation results with nonparametric first-step propensity score estimators with no conditioning covariates. Panel B uses the first-period outcome as the conditioning covariate in the first-step propensity score estimation.

References

Abott, C., V. Kogan, S. Lavertu, and Z. Peskowitz (2020): "School district operational spending and student outcomes: Evidence from tax elections in seven states," *Journal of Public Economics*, 183, 104–142.

Angrist, J. D., O. Jordà, and G. M. Kuersteiner. (2018): "Semiparametric estimates of monetary policy effects: string theory revisited," *Journal of Business & Economic Statistics*, 36(3), 371–387.

Athey, S. and G. W. Imbens (2022): "Design-based Analysis in Difference-in-Differences Settings with Staggered Adoption," *Journal of Econometrics*.

Blackwell, M. (2013): "A framework for dynamic causal inference in political science," *American Journal of Political Science*, 57(2), 504–520.

Blackwell, M. and A. N. Glynn (2018): "How to make causal inferences with time-series cross-sectional data under selection on observables," *American Political Science Review*, 112.(4), 1067–1082.

Bojinov, I., A. Rambachan, and N. Shephard (2021): "Panel experiments and dynamic causal effects: A finite population perspective," *Quantitative Economics*, 12(4), 1171–1196.

Cai, Z., J. Fan, and R. Li (2000): "Efficient Estimation and Inferences for Varying-Coefficient Models," *Journal of the American Statistical Association*, 95, 888–902.

Callaway, B. and P. H. Sant'Anna (2021): "Difference-in-differences with multiple time periods," *Journal of Econometrics*.

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

——— (2020): "Optimal Bandwidth Choice for Robust Bias-corrected Inference in Regression Discontinuity Designs," *Econometric Journal*, 23, 192–210.

——— (2022): "Coverage error optimal confidence intervals for local polynomial regression," *Bernoulli*, 28, 2998–3022.

Calonico, S., M. D. Cattaneo, M. H. Farrell, and R. Titiunik (2019): "Regression discontinuity designs using covariates," *Review of Economics and Statistics*, 101(3), 442–451.

Calonico, S., M. D. Cattaneo, and R. Titiunik (2014): "Robust Nonparametric Confidence Intervals for Regression Discontinuity Designs," *Econometrica*, 82(6), 2295–2326.

——— (2015): "rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs," *R Journal*, 7(1), 38.

Cattaneo, M. D., L. Keele, R. Titiunik, and G. Vazquez-Bare (2016): "Interpreting Regression Discontinuity Designs with Multiple Cutoffs," *The Journal of Politics*, 78(4), 1229–1248.

Cattaneo, M. D. and R. Titiunik (2022): "Regression Discontinuity Designs," *Annual Review of Economics, forthcoming*, 14.

Cellini, S. R., F. Ferreira, and J. Rothstein (2010): "The value of school facility investments: Evidence from a dynamic regression discontinuity design," *The Quarterly Journal of Economics*, 25.1, 215–261.

Chakraborty, B. and S. Murphy (2014): "Dynamic treatment regimes," *Annual review of statistics and its application*, 1, 447–464.

Chen, X. and D. Pouzo (2009): "Efficient estimation of semiparametric conditional moment models with possibly nonsmooth residuals," *Journal of Econometrics*, 152(1), 46–60.

Chernozhukov, V., D. Chetverikov, M. Demirer, E. Duflo, C. Hansen, W. Newey, and J. Robins (2018): "Double/debiased machine learning for treatment and structural parameters," *The Econometrics Journal*, 21(1), C1–C68.

Chernozhukov, V., I. Fern'andez-Val, S. Hoderlein, H. Holzmann, and W. Newey (2015a): "Nonparametric identification in panels using quantiles," *Journal of Econometrics*, 188(2), 378–392.

Chernozhukov, V., I. Fern'andez-Val, and A. Kowalski (2015b): "Quantile regression with censoring and endogeneity," *Journal of Econometrics*, 186(1), 201–221.

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

Clark, D. and P. Martorell (2014): "The signaling value of a high school diploma," *Journal of Political Economy*, 122, 282–318.

Colonnelli, E., M. Prem, and E. Teso. (2020): "Patronage and selection in public sector organizations," *American Economic Review*, 110, 3071–99.

Darolia, R. (2013): "Integrity versus access? The effect of federal financial aid availability on postsecondary enrollment," *Journal of Public Economics*, 106, 101–114.

De Chaisemartin, C. and X. d'Haultfoeuille (2020): "Two-way fixed effects estimators with heterogeneous treatment effects," *American Economic Review*, 110, 2964–96.

——— (forthcoming): "Difference-in-differences estimators of intertemporal treatment effects," *Review of Economics and Statistics*, No. w29873.

DiNardo, J. and D. S. Lee (2004): "Economic impacts of new unionization on private sector employers: 1984-2001," *The Quarterly Journal of Economics*, 119, 1383–1441.

Dong, Y. (2019): "Regression discontinuity designs with sample selection," *Journal of Business & Economic Statistics*, 37(1), 171–186.

Dube, A., L. Giuliano, and J. Leonard (2019): "Fairness and Frictions: The Impact of Unequal Raiseson Quit Behavior," *The American Economic Review*.

Fan, J. and I. Gijbels. (1996): *Local polynomial modelling and its applications: monographs on statistics and applied probability 66*, Chapman and Hall/CR.

Fan, Q., Y. Hsu, R. Lieli, and Y. Zhang (2022): "Estimation of conditional average treatment effects with high-dimensional data," *Journal of Business & Economic Statistics*, 40(1), 313–327.

Fernàndez-Val, I., A. van Vuuren, and F. Vella (2021): "Nonseparable sample selection models with censored selection rules," *Journal of Econometrics*.

Ferreira, F. and J. Gyourko (2009): "Do political parties matter? Evidence from US cities," *The Quarterly Journal of Economics*, 124, 399–422.

Flores, C. and Flores-Lagunes (2009): "Identification and estimation of causal mechanisms and net effects of a treatment under unconfoundedness," *working paper*.

——— (2010): "Nonparametric partial identification of causal net and mechanism average treatment effects," *working paper*.

Gallen, Y., J. S. Joensen, E. R. Johansen, and G. F. Veramendi (2023): "The Labor Market Returns to Delaying Pregnancy," *SSRN working paper*.

Gelman, A. and G. Imbens (2019): "Why high-order polynomials should not be used in regression discontinuity designs," *Journal of Business & Economic Statistics*, 37, 447–456.

Grembi, V., T. Nannicini, and U. Troiano (2016): "Do Fiscal Rules Matter?" *American Economic Journal: Applied Economics*, 8, 1–30.

Hahn, J., P. Todd, and W. van der Klaauw (2001): "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design," *Econometrica*, 69, 201–209.

Han, S. (2021): "Identification in nonparametric models for dynamic treatment effects," *Journal of Econometrics*, 225, 132–147.

Heckman, J., J. Smith, and N. Clements (1997): "Making the Most Out of Programme Evaluations and Social Experiments: Accounting for Heterogeneity in Programme Impacts," *Review of Economic Studies*, 64, 487–535.

Heckman, J. J., J. E. Humphries, and G. Veramendi (2016): "Dynamic Treatment Effects," *Journal of econometrics*, 191, 276–292.

Hernán, M. A. and J. M. Robins (2023): *Causal inference: What If*, CRC Press.

Huber, M. (2020): "Mediation analysis," in *Handbook of labor, human resources and population economics*, Springer.

Imai, K., I. Kim, and E. Wang (2023): "Matching methods for causal inference with time-series cross-sectional data," *American Journal of Political Science*, 67(3), 587–605.

Imai, K. and M. Ratkovic (2015): "Robust estimation of inverse probability weights for marginal structural models," *Journal of the American Statistical Association*, 110(511), 1013–1023.

Imbens, G. W. and T. Lemieux (2008): "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics*, 142, 615 – 635.

Johnson, M. S. (2020): "Regulation by Shaming: Deterrence Effects of PublicizingViolations of Workplace Safety and Health Laws," *The American Economic Review*, 110.6, 1866–1904.

Kennedy, E., Z. Ma, M. McHugh, and D. Small (2017): "Non-parametric methods for doubly robust estimation of continuous treatment effects," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(4), 1229–1245.

Lee, D. S. (2008): "Randomized Experiments from Non-random Selection in U.S. House Elections," *Journal of Econometrics*, 142, 675–697.

Lee, D. S. and A. Mas (2012): "Long-run impacts of unions on firms: New evidence from financial markets, 1961-1999." *The Quarterly Journal of Economics*, 127, 333–378.

Lv, X., X.-R. Sun, Y. Lu, and R. Li (2019): "Nonparametric identification and estimation of dynamic treatment effects for survival data in a regression discontinuity design," *American Economic Journal: Applied Economics*, 184, 108665.

Ma, S. and M. R. Kosorok (2005): "Robust semiparametric M-estimation and the weighted bootstrap," *Journal of Multivariate Analysis*, 96, 190–217.

Manski, C. F. and J. V. Pepper (2000): "Monotone instrumental variables: With an application to the returns to schooling." *Econometrica*, 68, 997–1010.

Murphy, S. (2003): "Optimal dynamic treatment regimes," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 65(2), 331–355.

Murphy, S., van der Laan, J. M.J., Robins, and C. P. P. R. Group (2001): "Marginal mean models for dynamic regimes," *Journal of the American Statistical Association*, 96(456), 1410–1423.

Papay, John P.and Willett, J. B. and R. J. Murnane (2011): "Extending the regressiondiscontinuity approach to multiple assignment variables," *Journal of Econometrics*, 161, 203–207.

Pettersson-Lidbom, P. (2008): "Do Parties Matter for Economic Outcomes? A Regression-discontinuity Approach," *Journal of the European Economic Association*, 6, 1037–1056.

PORTER, J. (2003): "Estimation in the Regression Discontinuity Model," *working paper*.

REARDON, S. F. AND J. P. ROBINSON (2012): "Regression Discontinuity Designs with Multiple Rating-Score Variables," *Journal of Research on Educational Effectiveness*, 5, 83–104.

ROBINS, J. (1986): "A new approach to causal inference in mortality studies with a sustained exposure period—application to control of the healthy worker survivor effect," *Mathematical modelling*, 7(9-12), 1393–1512.

——— (1987): "A graphical approach to the identification and estimation of causal parameters in mortality studies with sustained exposure periods," *Journal of chronic diseases*, 40, 139S–161S.

SUN, L. AND S. ABRAHAM (2021): "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*.

THISTLETHWAITE, D. L. AND D. T. CAMPBELL (1960): "Regression-discontinuity analysis: An alternative to the ex post facto experiment," *Journal of Educational Psychology*, 51, 309–317.

WONG, V. C., P. M. STEINER, AND T. D. COOK (2013): "Analyzing Regression-Discontinuity Designs With Multiple Assignment Variables: A Comparative Study of Four Estimation Methods," *Journal of Educational and Behavioral Statistics*, 38, 107–141.