Dynamic Regression Discontinuity under Treatment Effect Heterogeneity

Yu-Chin Hsu† **Shu Shen**‡

This version: July 9, 2021

Abstract

Regression discontinuity (RD) is a popular tool for the analysis of economic policies or treatment interventions. This paper extends the classic static RD model to a dynamic framework, where observations are eligible for repeated RD experiments and, therefore, treatments. Such dynamics often complicate the identification and estimation of longer-term average treatment effects. Previous empirical research with such designs typically ignored the dynamics in the model or adopted restrictive identifying assumptions. This paper studies identification strategies under various sets of weaker identifying assumptions and proposes associated estimation and inference methods. The proposed methods are applied to revisit the effect of Californian local school bonds in the seminal study of Cellini et al. (2010).

Keywords: primary treatment effect, quasi-potential outcome variable, dynamic regression discontinuity, semiparametric, varying coefficient Logit, weighted bootstrap

† 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.

1 Introduction

Regression discontinuity (RD) models are popular in policy evaluations or other settings of treatment effect analysis. The setup exploits discontinuity in the design of many policies to nonparametrically identify treatment effects for observations near the eligibility cutoff. While the classic RD setup, either sharp or fuzzy, is static, in empirical applications we often see situations where each individual could potentially participate in multiple RD experiments and sometimes receive multiple treatments over a period of time. 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 forward in front of voters repeatedly over time. A large body of literature in political science (e.g., Ferreira and Gyourko, 2009; Caughey and Sekhon, 2011; Colonnelli et al., 2020) uses RD to study the effect of political races that happen on a regular basis. Some RD applications study the effect of having peers passing RD eligibility tests and receiving treatment interventions (e.g., Dube et al., 2019 and Johnson, 2020). In such studies, observations could again be exposed to RD experiments repeatedly.

In this paper, we study a general multi-period dynamic RD model under the potential outcome framework. We formalize the concept of *primary* treatment effects which prohibit the reception of treatments after the focal round of RD and show that primary treatment effects could be used to construct all other treatment effects defined as differences of potential outcomes. We distinguish primary treatment effects from *total* treatment effects that do not restrict treatment status after the focal round of RD. It is well-known that classic static RD models could be regarded as local random experiments under proper smoothness conditions, hence providing nonparametric identification

* This paper was first presented in the University of California, Berkeley causal inference group seminar in March 2020 under the title of "Dynamic Regression Discontinuity". The authors thank seminar participants from University of California, Berkeley, University of Nevada, Reno, University of Chicago, and 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), Academia Sinica Investigator Award of Academia Sinica (AS-IA-110-H01), and Center for Research in Econometric Theory and Applications (110L9002) from the Featured Areas Research Center Program within the framework of the Higher Education Sprout Project by the Ministry of Education of Taiwan.

of average/distributional treatment effects for marginal individuals at the RD cutoff. For dynamic RD models considered in this paper, however, smoothness conditions alone could only identify average/distributional total treatment effects and immediate primary treatment effects. If longer-term primary treatment effects are of interest, additional identifying conditions need to be imposed. We study constructive identification of longer-term average primary treatment effects under various assumptions and propose associated estimation procedures. For inference, we focus on the asymptotics of a proposed two-step semiparametric estimation procedure for longer-term average primary treatment effects, as the inference of other nonparametric estimators are standard.

Among empirical studies with dynamic RD settings, the pioneering work by Cellini et al. (2010) (CFR) and subsequent studies that adopt their method (e.g., Darolia, 2013; Abott et al., 2020) are the only ones, as far as the authors know, that take the dynamics in model seriously. CFR identifies longer-term average primary treatment effects with an innovative recursive RD strategy and an event study type strategy. Both strategies take a parametric perspective and use restrictive assumptions, including path independence and mean equivalence of treatment effects in different rounds and homogeneous individual treatment effects. We argue that the popularity of classic static RD models lies in the nonparametric nature of model identification, which only requires mild conditions, and identification in dynamic RD models shall carry the same features. In this paper, we study nonparametric identification of longer-term average primary treatment effects under identifying assumptions that are substantially weaker than those imposed in CFR.

In addition to identification, this paper contributes to the RD literature by introducing new estimation and inference procedures from the statistics literature. Our main estimators for longer-term average treatment effects follow two steps. In the first step, we model the propensity score functions semiparametrically and use the local MLE estimator in Cai et al. (2000). This kind of modeling is particularly suitable for the dynamic RD setting because the first-step propensity score functions need to condition on additional covariates as well as the running variable. The proposed local MLE estimator has the advantage of staying local to the RD cutoff along the dimension of the running variable, which is vital to RD estimation as is argued in Gelman and Imbens (2019) while remaining parametric along the dimension of other controls so as not to overburden the convergence rate of the two-step estimator. In terms of inference, this paper extends the weighted bootstrap method designed in Ma and Kosorok (2005) for semiparametric M-estimation to the proposed two-step local semiparametric estimation procedure.

Our paper is also related to the literature on dynamic treatment effects in non-RD settings. For example, Heckman et al. (2016) study treatment effects in ordered and unordered multi-stage decision problems with an instrumental variable approach. Sun and Abraham (forthcoming), Callaway and Sant'Anna (forthcoming), and Athey and Imbens (forthcoming) examine treatment effects in panel event studies with one single irreversible treatment. De Chaisemartin and d'Haultfoeuille (2020) study linear two-way fixed-effect regressions for panel data models with treatment effect heterogeneity across groups or over time. Even without considering the RD setting, models and goals considered in our paper are different. In our model, individuals have the opportunity of obtaining repeated treatments, and the goal for identification is to disentangle different average primary treatment effects that make up longer-term average total treatment effects. Our dynamic RD model allows individuals to self-select into subsequent rounds of RD experiments and allows potential RD participation decisions and running variables in later rounds to vary with treatment decisions in earlier rounds.

The rest of the paper is organized as follows. Section 2 lays out the formal framework of dynamic RD and defines various treatment effects of interests. The section also reevaluates the recursive identification strategy proposed in Cellini et al. (2010) under the potential outcome framework. Section 3 studies both point and partial identification of longer-term average primary treatment effects under different identifying assumptions. Section 4 discusses estimation and inference of various identified treatment effects. Monte Carlo simulations in Section 5 show good small sample performances of the proposed estimators. Section 6 revisits the empirical study of local school bonds in CFR using the proposed procedures. Section 7 concludes.

2 Model Setup

2.1 Potential Outcome Framework

We observe a panel of n individuals over T time periods. In each time period, an RD experiment takes place *at the beginning* of the time period, and treatment is administrated immediately following the experiment for those who pass the RD threshold. An outcome is observed *at the end* of each time period. Everyone takes part in the first RD experiment and can choose (or be chosen) to participate in one or more rounds of subsequent RDs. We use subscript k, $k = 1, 2, ..., K$, to denote the round of RD experiments or treatment interventions and t, $t = 1, 2, ..., T$, to denote the period of observed outcomes. Although in many settings, outcomes can be observed in periods after the last round of treatment intervention, it is not necessary to distinguish such longer-term outcomes from those that occur right after the last round of treatment in terms of econometrics modeling. Therefore, we assume without loss of generality that $T = K$.

Let S_k be the *observed* participation indicator for the k-th round RD experiment. If an individual participates in the k-th period, $S_k = 1$; otherwise, $S_k = 0$. Assume that everyone participates in the first period, so $S_1 = 1$ and is degenerate. Let Z_k be the running variable of the k-th RD experiment; Z_k is only observed for individuals with $S_k = 1$. The observed treatment status of an individual in round k is $D_k = 1(Z_k \geq 0) \cdot S_k$, once we normalize thresholds of all RD experiments to 0.

Let $S_k(\ell^{k-1})$ be the *potential* k-th round participation indicator and $Z_k(\ell^{k-1})$ the *potential* k-th round running variable with path of past treatment statuses summarized by $\ell^{k-1} = (d_1, d_2, ..., d_{k-1}) \in \{0, 1\}^{k-1}$; $k \geq 2$. Let $\mathcal{L}^{k-1}$ be the set of all possible paths of treatment statuses for a total of $k - 1$ periods. Without further restriction, the canonical count of the set, denoted by $|\mathcal{L}^{k-1}|$, is 2^{k-1}. For example, $\mathcal{L}^1 = \{0, 1\}$ and $\mathcal{L}^2 = \{(0, 0), (0, 1), (1, 0), (1, 1)\}$. If an individual could only receive up to one treatment, a special case that will be discussed later, $|\mathcal{L}^{k-1}| = k$. The *potential* k-th round treatment decision is then $D_k(\ell^{k-1}) = 1(Z_k(\ell^{k-1}) \geq 0) \cdot S_k(\ell^{k-1})$ for $k \geq 2$. Treatment decisions are path-dependent because treatment status in previous rounds could contribute to both RD participation decisions and values of running variables in future rounds.

The observed treatment status D_k for $k \geq 2$ could be represented by potential treatment decisions.

$$D_k = \sum_{\ell^{k-1} \in \mathcal{L}^{k-1}} D_k(\ell^{k-1}) \cdot \mathfrak{D}(\ell^{k-1}), \tag{2.1}$$

where $\mathfrak{D}(.)$ is a path indicator. For example, $\mathfrak{D}(\ell^1) = D_1^{\ell^1} \cdot (1 - D_1)^{1-\ell^1}$ with $\ell^1 \in \{0, 1\}$ and $\mathfrak{D}(\ell^2) = D_1^{\ell_1^2} \cdot (1 - D_1)^{1-\ell_1^2} \cdot D_2(\ell_1^2)^{\ell_2^2} \cdot (1 - D_2(\ell_1^2))^{1-\ell_2^2}$ with $\ell^2 \in \{(0,0), (0,1), (1,0), (1,1)\}$, where ℓ_s^{k-1} is the s-th dimension of ℓ^{k-1}. $\{\mathfrak{D}(\ell^k) = 1\} \Leftrightarrow \{D_1 = \ell_1^k, D_2(\ell_1^k) = \ell_2^k, ..., D_k(\ell_{1:(k-1)}^k) = \ell_k^k\} \Leftrightarrow \{D_1 = \ell_1^k, D_2 = \ell_2^k, ..., D_k = \ell_k^k\}$. When $k = 2$, $D_2 = D_2(0) \cdot (1 - D_1) + D_2(1) \cdot D_1$. When $k = 3$, $D_3 = D_3((0,0)) \cdot (1 - D_1) \cdot (1 - D_2(0)) + D_3((0,1)) \cdot (1 - D_1) \cdot D_2(0) + D_3((1,0)) \cdot D_1 \cdot (1 - D_2(1)) + D_3((1,1)) \cdot D_1 \cdot D_2(1)$.

Let Y_t be the observed outcome in time period t, and $Y_t(\ell^t)$ be the potential outcome with the path of past treatment statuses summarized by ℓ^t, for $t = 1, ..., T$. Then

$$Y_t = \sum_{\ell^t \in \mathcal{L}^t} Y_t(\ell^t) \cdot \mathfrak{D}(\ell^t). \tag{2.2}$$

When $t = 1$, $Y_1 = Y_1(0) \cdot (1 - D_1) + Y_1(1) \cdot D_1$. When $t = 2$, $Y_2 = Y_2((0,0)) \cdot (1 - D_1) \cdot (1 - D_2(0)) + Y_2((0,1)) \cdot (1 - D_1) \cdot D_2(0) + Y_2((1,0)) \cdot D_1 \cdot (1 - D_2(1)) + Y_2((1,1)) \cdot D_1 \cdot D_2(1)$.

The above-defined dynamic RD model is different from a setting where observations could take multiple RD eligibility tests but only receive up to one treatment after all RD eligibility tests. For example, Clark and Martorell (2014) use RD to study the effect of a high school diploma, while every student has multiple chances to take the test and qualify for the diploma. It is worthwhile to point out that the RD setting in Clark and Martorell (2014) could be treated as a classic fuzzy RD model, where those who would opt out or fail to meet later-round RD cutoffs upon failing the first round could be treated as compliers, and those who earn eligibility to treatment through later rounds of RD could be treated as always-takers. Our dynamic RD setup is different. Treatments in our model are administrated repeatedly in each time period.

2.2 Definitions of Individual Treatment Effects

First, we define the individual *primary treatment effect* of a focal k-th round treatment when the individual is prohibited to receive additional treatments after the focal round.[1]

[1]The primary individual treatment effect is referred to as the treatment-on-treated (TOT) effect in CFR.

Let $\theta_{\tau,1}$ denote the τ-period-after primary treatment effect of the first treatment and $\theta_{\tau,k}^{\ell^{k-1}}$ the τ-period-after primary treatment effect of the k-th treatment with path of past treatment statuses summarized by ℓ^{k-1}; $k = 2, ..., T$. We also call $\theta_{0,k}$ the immediate primary treatment effect of the k-th round treatment and $\theta_{\tau,k}$ the longer-term primary treatment effect when $\tau \geq 1$. Let $\mathbf{0}_\tau$ denote a τ-dimensional vector of zeros for $\tau \geq 1$. Then

$$\begin{aligned}
&\theta_{0,1} = Y_1(1) - Y_1(0),\ \theta_{\tau,1} = Y_{1+\tau}((1, \mathbf{0}_\tau)) - Y_{1+\tau}((0, \mathbf{0}_\tau)),\\
&\theta_{0,k}^{\ell^{k-1}} = Y_k((\ell^{k-1}, 1)) - Y_k((\ell^{k-1}, 0)),\\
&\theta_{\tau,k}^{\ell^{k-1}} = Y_{k+\tau}((\ell^{k-1}, 1, \mathbf{0}_\tau)) - Y_{k+\tau}((\ell^{k-1}, 0, \mathbf{0}_\tau)), \quad \text{for } k \geq 2, \tau \geq 1.
\end{aligned} \tag{2.3}$$

Notice that differences of some pairs of potential outcomes, such as $Y_2((1, 1)) - Y_2((0, 1))$ and $Y_2((1, 0)) - Y_2((0, 1))$, are not defined by equation (2.3). The following lemma shows that such differences could be represented by primary treatment effects.

Lemma 2.1 *The difference of any pair of potential outcomes could be represented by linear combinations of primary individual treatment effects defined in equation* (2.3)*.*

In the context of nonparametric RD identification, the above lemma implies that if average primary treatment effects for a set of marginal individuals are identified, any other average treatment effects for the same set of marginal individuals are identified.

It is important to distinguish primary treatment effects defined above from the total effect concept in Heckman et al. (2016), also called the intent-to-treat effect in Cellini et al. (2010). Instead of prohibiting treatment take-up after the focal k-th round, total effects do not restrict treatment status in periods after the focal k-th round. For example, $\tilde{\theta}_{1,1} = Y_2((1, 0))(1 - D_2(1)) + Y_2((1, 1))D_2(1) - [Y_2((0, 0))(1 - D_2(0)) + Y_2((0, 1))D_2(0)]$, and $\tilde{\theta}_{1,2}^{d_1} = Y_3((d_1, 1, 0))(1 - D_3((d_1, 1)) + Y_3((d_1, 1, 1))D_3((d_1, 1)) - Y_3((d_1, 0, 0))(1 - D_3((d_1, 0)) - Y_3((d_1, 0, 1))D_3((d_1, 0))$, for $d_1 = 0, 1$.

More generally, let $\tilde{\theta}_{\tau,k}^{\ell^{k-1}}$ be the τ-period-after individual *total treatment effect* of the k-th round treatment with path of past treatments summarized by ℓ^{k-1}. Then

$$\tilde{\theta}_{\tau,1} = \tilde{Y}_{1+\tau}(1) - \tilde{Y}_{1+\tau}(0), \quad \tilde{\theta}_{\tau,k}^{\ell^{k-1}} = \tilde{Y}_{k+\tau}((\ell^{k-1}, 1)) - \tilde{Y}_{k+\tau}((\ell^{k-1}, 0)), \tag{2.4}$$

where $\tilde{Y}_{k+\tau}(\ell^k)$, for $k = 1, 2, ...$ and $\tau = 0, 1, ...$, is a *quasi-potential outcome* that fixes the first k rounds of treatment status at $\ell^k \in \mathcal{L}^k$ but does not put any restrictions on treatment statuses after the k-th round. Specifically,

$$\tilde{Y}_{k+\tau}(\ell^k) = \sum_{\eta \in \mathcal{L}^\tau} Y_{k+\tau}((\ell^k, \eta))\mathfrak{D}_{(k+1):(k+\tau)}((\ell^k, \eta)),$$

where $\{\mathfrak{D}_{l:l'}(\ell^k) = 1\} \Leftrightarrow \{D_l(\ell^k_{1:(l-1)}) = \ell^k_l, D_{l+1}(\ell^k_{1:l}) = \ell^k_{l+1}, ..., D_{l'}(\ell^k_{1:(l'-1)}) = \ell^k_{l'}\}$. It is easy to see that when $\tau = 0$, quasi-potential outcomes are the same as usual potential outcomes, implying that immediate total effects defined in (2.4) are the same as immediate primary effects defined in (2.3). When $\tau \geq 1$, quasi-potential outcomes are different from classic potential outcomes. Only treatment statuses in the first k rounds are fixed in $\tilde{Y}_{k+\tau}(d)$, regardless of the value of τ. Therefore, $\tilde{Y}_{k+\tau}(\ell^k)$, although documenting the $(k+\tau)$-th period outcome could also be viewed as a classic potential k-th period outcome.

Next, we define *primary first-stage effects* and *total first-stage effects* similar to primary treatment effects and total treatment effects. Let $\eta_{\tau,1}$ and $\eta^{\ell^{k-1}}_{\tau,k}$ be the primary first-stage effects and $\tilde{\eta}_{\tau,1}$ and $\tilde{\eta}^{\ell^{k-1}}_{\tau,k}$ be the total first-stage effects. For $\tau \geq 0$,

$$\eta_{\tau,1} = D_{2+\tau}((1, \mathbf{0}_\tau)) - D_{2+\tau}((0, \mathbf{0}_\tau)),$$
$$\tilde{\eta}_{\tau,1} = \tilde{D}_{2+\tau}(1) - \tilde{D}_{2+\tau}(0),$$

where $(\ell, \mathbf{0}_\tau) = \ell$ if $\tau = 0$. For $\tau \geq 0$, $k \geq 2$, and $\ell^{k-1} \in \mathcal{L}^{k-1}$,

$$\eta^{\ell^{k-1}}_{\tau,k} = D_{k+1+\tau}((\ell^{k-1}, 1, \mathbf{0}_\tau)) - D_{k+1+\tau}((\ell^{k-1}, 0, \mathbf{0}_\tau)),$$
$$\tilde{\eta}^{\ell^{k-1}}_{\tau,k} = \tilde{D}_{k+1+\tau}((\ell^{k-1}, 1)) - \tilde{D}_{k+1+\tau}((\ell^{k-1}, 0)),$$

where $\tilde{D}_{k+1+\tau}(\ell^{k-1}) = \sum_{\eta \in \mathcal{L}^\tau} D_{k+1+\tau}((\ell^k, \eta))\mathfrak{D}_{(k+1):(k+\tau)}((\ell^k, \eta))$. These definitions will be used later in the paper to identify longer-term average primary treatment effects.

2.3 Identification under Smoothness Conditions Only

This section discusses what kinds of average treatment effect concepts can be identified with smoothness conditions alone. Let $\mathcal{N}_\epsilon = (-\epsilon, \epsilon)$ for some $\epsilon > 0$. The following assumption specifies classic RD-type smoothness conditions for our dynamic framework.

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

1. Z_1 is continuous in $z_1 \in \mathcal{N}_\epsilon$ with $P[Z_1 \geq 0] \in (0,1)$;

2. for $\tau = 0, 1, ..., T-1$, $E[Y_{\tau+1}(\ell^{\tau+1})|\mathfrak{D}(\ell^{\tau+1}) = 1, Z_1 = z_1]$ is continuous in $z_1 \in \mathcal{N}_\epsilon$ for all $\ell^{\tau+1} \in \mathcal{L}^{\tau+1}$.

3. $P[\mathfrak{D}(\ell^T) = 1|Z_1 = z_1]$ is continuous in $z_1 \in \mathcal{N}_\epsilon$ for all $\ell^T \in \mathcal{L}^T$.

Assumption 2.1 requires smoothness of average potential outcomes and average potential treatment decisions around the first-round running variable cutoff. Since $\mathfrak{D}(.)$ is the path indicator and Assumption 2.1.3 is for all paths of treatment statuses, Assumption 2.1 implies, for example, continuity of $P[\mathfrak{D}(\ell^k) = 1|Z_1 = z_1]$ and $E[D_{k+1}(\ell^k)|\mathfrak{D}(\ell^k) = 1, Z_1 = z_1]$ as well, for all $k = 1, 2, ..., K$ and $\ell^k \in \mathcal{L}^k$. The assumption also implies smoothness of average quasi-potential outcomes and quasi-potential treatment decisions.

As is discussed in the introduction, τ-period-after average total effects could be identified under smoothness conditions alone, for all $\tau = 0, 1, ..., T$. When the focal treatment is the first-round treatment, under Assumption 2.1,

$$E[\tilde{\theta}_{\tau,1}|Z_1 = 0] = \lim_{z_1 \searrow 0} E[Y_{\tau+1}|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[Y_{\tau+1}|Z_1 = z_1].$$

Assumption 2.1 could be extended to condition on the k-th round running variable. We state this extension in Assumption A.1 in the Online Appendix. When the focal treatment is the k-th-round treatment, under Assumption A.1,

$$\begin{aligned} &E[\tilde{\theta}^{\ell^{k-1}}_{\tau,k}|\mathfrak{D}(\ell^{k-1}) = 1, Z_k = 0] \\ = &\lim_{z_k \searrow 0} E[Y_{\tau+k}|\mathfrak{D}(\ell^{k-1}) = 1, Z_k = z_k] - \lim_{z_k \nearrow 0} E[Y_{\tau+k}|\mathfrak{D}(\ell^{k-1}) = 1, Z_k = z_k]. \end{aligned}$$

Since immediate primary treatment effects and immediate total treatment effects are the same, the above results imply that $E[\theta_{0,1}|Z_1 = 0]$ and $E[\theta^{\ell^{k-1}}_{0,k}|\mathfrak{D}(\ell^{k-1}) = 1, Z_k = 0]$ are identified. Assumption 2.1 alone, however, cannot identify longer-term average primary treatment effects, which would be the focus of the rest of the paper. On the other hand, in the special setting of having the only treatment administrated after all rounds of RD eligibility tests, smoothness conditions alone could identify local average treatment effects (LATE) for complies, given the fuzzy RD interpretation of the model discussed at the end of Section 2.1.

2.4 Identification under Treatment Effect Homogeneity

As is discussed in the introduction, the seminal study of CFR proposes a recursive identification strategy for longer-term primary treatment effects in dynamic RD models using parametric RD regressions. This section formalizes the recursive identification result in CFR under the potential outcome framework. The identification uses a key assumption of treatment effect homogeneity in addition to classic RD-type smoothness conditions.

CFR labels individual treatment effects only by τ, the number of periods between outcome variables and the focal treatment. Under the potential outcome framework, this implies a key assumption of treatment effect homogeneity. Specifically,

$$\begin{aligned}
\theta_0 &= Y_1(1) - Y_1(0) = Y_2((0,1)) - Y_2((0,0)) = Y_2((1,1)) - Y_2((1,0)) = \\
&= Y_3((0,0,1)) - Y_3((0,0,0)) = ... \\
\theta_1 &= Y_2((1,0)) - Y_2((0,0)) = Y_3((0,1,0)) - Y_3((0,0,0)) \\
&= Y_3((1,1,0)) - Y_3((1,0,0)) = ... \\
&...
\end{aligned} \tag{2.5}$$

The simplification is restrictive. For example, it requires that for each individual, the individual treatment effect of the first-round treatment is the same as that of the k-th-round treatment; $k \geq 2$. The simplification also implies that rankings of individual treatment effects remain the same across different rounds of treatments and, therefore, is even stronger than the condition of rank invariance (e.g., Heckman et al., 1997 and Dong and Shen, 2018) across treatment effects. The simplification also assumes that individual treatment effects are path independent.

In addition to the restrictions discussed above, the recursive identification strategy in CFR also requires individual treatment effects to be homogeneous across individuals so that θ_τ are constants for all $\tau \geq 0$. Let $\pi_k = \lim_{z_1 \searrow 0} E[D_{k+1}|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[D_{k+1}|Z_1 = z_1]$ be the jump in the $(k+1)$-th round treatment take-up rate among marginal individuals of the first-round RD; $\pi_0 = 1$, $\pi_k \in [-1,1]$ for $k \geq 1$. The following Lemma summarizes the CFR resursive identification strategy.

Lemma 2.2 *Under Assumption 2.1, the treatment effect homogeneity restriction in equation* (2.5)*, and the assumption that* $\theta_0, \theta_1, ...$ *are fixed constants, the following recursive* *identification result holds:*

$$\sum_{\tau=0}^{t-1} \theta_\tau \cdot \pi_{t-1-\tau} = \lim_{z_1 \searrow 0} E[Y_t|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[Y_t|Z_1 = z_1], \quad \textit{for } t = 1, 2, ..., T. \quad (2.6)$$

Or,

$$\theta_0 = \lim_{z_1 \searrow 0} E[Y_1|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[Y_1|Z_1 = z_1],$$

$$\theta_1 + \theta_0 \cdot \pi_1 = \lim_{z_1 \searrow 0} E[Y_2|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[Y_2|Z_1 = z_1],$$

$$\theta_2 + \theta_1 \cdot \pi_1 + \theta_0 \cdot \pi_2 = \lim_{z_1 \searrow 0} E[Y_3|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[Y_3|Z_1 = z_1],$$

....

Lemma 2.2 is proved by the method of induction, and a formal proof is given in the online appendix. Intuitively, if barely passing the first-round RD cutoff and receiving the first-round treatment does not change an individual's probability of taking treatments in later rounds, these later treatments would not contribute to the threshold discontinuity in the t-th period outcome. However, if barely passing the first-round RD cutoff changes an individual's probability of taking later round treatments, either through changed RD participation rate or running variables in subsequent rounds, observed discontinuity in the t-th period average outcome at the first-round RD cutoff would also include effects from treatments of later round.

The recursive identification strategy in Lemma 2.2 relies heavily on the treatment effect homogeneity simplification. Take the simplest case of $t = 2$ as an example:

$$\begin{aligned}
&\lim_{z_1 \searrow 0} E[Y_2|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[Y_2|Z_1 = z_1] \\
=&\lim_{z_1 \searrow 0} E[Y_2(1,0)|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[Y_2(0,0)|Z_1 = z_1] \\
&+ \lim_{z_1 \searrow 0} E[(Y_2(1,1) - Y_2(1,0))D_2(1)|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[(Y_2(0,1) - Y_2(0,0))D_2(0)|Z_1 = z_1] \\
=&E[\theta_1|Z_1 = 0] + \theta_0 \cdot \left(\lim_{z_1 \searrow 0} E[D_2|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[D_2|Z_1 = z_1] \right) \\
=&\theta_1 + \theta_0 \cdot \pi_1.
\end{aligned}$$

The second last equality holds because individual treatment effects are non-random. The recursiveness in the identification system relies on the assumption that the τ-period-after effects are the same across treatments of different rounds.

The strength of the treatment effect homogeneity restriction can also be revealed by the following over-identification result. Under updated smoothness conditions in Assumption A.1 stated in the online appendix and the treatment effect homogeneity restriction discussed above, we can show that, for $s = 1, 2, ..., T - k - 1$,

$$\sum_{\tau=0}^{s} \theta_\tau \cdot p_{s-\tau} = \lim_{z_{k+1} \searrow 0} E[Y_{k+1+s}|Z_{k+1} = z_{k+1}] - \lim_{z_{k+1} \nearrow 0} E[Y_{k+1+s}|Z_{k+1} = z_{k+1}], \tag{2.7}$$

where $p_s = \lim_{z_{k+1} \searrow 0} E[D_{k+1+s}|Z_{k+1} = z_{k+1}] - \lim_{z_{k+1} \nearrow 0} E[D_{k+1+s}|Z_{k+1} = z_{k+1}]$. Detailed derivations of (2.7) are given in the online appendix. Equation (2.7) implies that the same set of θ_0, θ_1,..., could be identified even when a first k period of data are set aside, for any $k = 1, 2, ...$ The result therefore provides many testable implications for the recursive CFR identification strategy.

Last but not least, it is worth mentioning that the recursive CFR identification strategy could be extended to the case where individual primary treatment effects are random but still path independent, mean independent of treatment decisions in subsequent rounds after conditioning on additional covariates, and have the same conditional means across different rounds of treatments. However, such assumptions could still be too strong in empirical applications. As a result, in the next section, we look into identifying longer-term average primary treatment effects under weaker identifying assumptions.

3 Identification under Treatment Effect Heterogeneity

In this section, we study the dynamic RD model under the general setting of treatment effect heterogeneity and propose a new identification strategy for average primary treatment effects. We focus on identifying the average primary treatment effects of the first round treatment under weaker assumptions than those discussed in Section 2.4. The proposed identification strategy could also be extended to distributional effects or average primary effects of the k-th round treatment conditional on past treatment statuses. In the rest of the paper, we will use ATE to refer to the average primary treatment effect and AITTE to refer to the average total treatment effect, also called the average intent-to-treat effect in CFR. Although ATEs and AITTEs, such as $E[\theta_{\tau,k}^{\ell^{k-1}}|Z_1 = 0]$ and

$E[\tilde{\theta}_{\tau,k}^{\ell^{k-1}}|Z_1 = 0]$, depend on τ, k, and ℓ^{k-1} and are for marginal individuals only, we will often refrain from iterating these details and just use ATE and AITTE for brevity.

Under smoothness conditions, all τ-period-after AITTEs are identified; $\tau \geq 0$. When $\tau = 0$, the immediate AITTE is the same as the immediate ATE. When $\tau \geq 1$, the τ-period-after AITTE of the first-round treatment could be decomposed to the τ-period-after ATE of the first-round treatment and many other shorter-term AITTEs of later treatments on the treated.[2] Mathematically, under Assumption 2.1,

$$\begin{aligned}
&E[\tilde{\theta}_{\tau,1}|Z_1 = 0] \\
=& \lim_{z_1 \searrow 0} E[Y_{\tau+1}|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[Y_{\tau+1}|Z_1 = z_1] \\
=& E[\theta_{\tau,1}|Z_1 = 0] \\
&+ \sum_{s=0}^{\tau-1} E[\tilde{\theta}_{s,\tau+1-s}^{(1,\mathbf{0}_{\tau-1-s})}|D_{\tau+1-s}(1, \mathbf{0}_{\tau-1-s}) = 1, Z_1 = 0] P[D_{\tau+1-s}(1, \mathbf{0}_{\tau-1-s}) = 1|Z_1 = 0] \\
&- \sum_{s=0}^{\tau-1} E[\tilde{\theta}_{s,\tau+1-s}^{(0,\mathbf{0}_{\tau-1-s})}|D_{\tau+1-s}(0, \mathbf{0}_{\tau-1-s}) = 1, Z_1 = 0] P[D_{\tau+1-s}(0, \mathbf{0}_{\tau-1-s}) = 1|Z_1 = 0].
\end{aligned} \tag{3.1}$$

For example, for $\tau = 1, 2$, equation (3.1) implies that

$$\begin{aligned}
&\lim_{z_1 \searrow 0} E[Y_2|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[Y_2|Z_1 = z_1] \\
=& E[\theta_{1,1}|Z_1 = 0] + E[\tilde{\theta}_{0,2}^1|D_2(1) = 1, Z_1 = 0] \cdot \lim_{z_1 \searrow 0} P[D_2 = 1|Z_1 = z_1] \\
&- E[\tilde{\theta}_{0,2}^0|D_2(0) = 1, Z_1 = 0] \cdot \lim_{z_1 \nearrow 0} P[D_2 = 1|Z_1 = z_1],
\end{aligned} \tag{3.2}$$

[2] The τ-period-after ATE of the first-round treatment could also be decomposed into the τ-period-after ATE of the first-round treatment and many other shorter-term ATEs of later treatments conditional on paths of past treatment statuses. Specifically,

$$\begin{aligned}
&E[\tilde{\theta}_{\tau,1}|Z_1 = 0] = \lim_{z_1 \searrow 0} E[Y_{\tau+1}|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[Y_{\tau+1}|Z_1 = z_1] \\
=& E[\theta_{\tau,1}|Z_1 = 0] \\
&+ \sum_{s=0}^{\tau-1} \sum_{\ell \in \mathcal{L}^{\tau-1-s}} E[\theta_{s,\tau+1-s}^{(1,\ell)}|\mathfrak{D}_{2:(\tau+1-s)}(1, \ell, 1) = 1, Z_1 = 0] P[\mathfrak{D}_{2:(\tau+1-s)}(1, \ell, 1) = 1|Z_1 = 0] \\
&- \sum_{s=0}^{\tau-1} \sum_{\ell \in \mathcal{L}^{\tau-1-s}} E[\theta_{s,\tau+1-s}^{(0,\ell)}|\mathfrak{D}_{2:(\tau+1-s)}(0, \ell, 1) = 1, Z_1 = 0] P[\mathfrak{D}_{2:(\tau+1-s)}(0, \ell, 1) = 1|Z_1 = 0]
\end{aligned}$$

where we abuse the notation and let $(d_1, \ell^0) = d_1$.

and

$$
\begin{aligned}
&\lim_{z_1 \searrow 0} E[Y_3|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[Y_3|Z_1 = z_1] \\
=&E[\theta_{1,1}|Z_1 = 0] + E[\tilde{\theta}^1_{1,1}|D_2(1) = 1, Z_1 = 0]P[D_2(1) = 1|Z_1 = 0] \\
&- E[\tilde{\theta}^0_{1,1}|D_2(0) = 1, Z_1 = 0]P[D_2(0) = 1|Z_1 = 0] \\
&+ E[\tilde{\theta}^{(1,0)}_{0,3}|D_3(1,0) = 1, Z_1 = 0]P[D_3(1,0) = 1|Z_1 = 0] \\
&- E[\tilde{\theta}^{(0,0)}_{0,3}|D_3(0,0) = 1, Z_1 = 0]P[D_3(0,0) = 1|Z_1 = 0],
\end{aligned} \tag{3.3}
$$

respectively. Note that $\tilde{\theta}^{d_1}_{0,2}$ in (3.2) and $\tilde{\theta}^{(d_1,0)}_{0,3}$ in (3.3) are equal to $\theta^{d_1}_{0,2}$ and $\theta^{(d_1,0)}_{0,3}$, respectively, for $d_1 = 0, 1$. We will use this decomposition to identify $E[\theta_{\tau,1}|Z_1 = 0]$ for $\tau = 1, 2, ...$ in the following sections.

3.1 One-period-after ATE

We first identify the one-period-after ATE, or $E[\theta_{1,1}|Z_1 = 0]$, with a conditional mean independence assumption (CIA) that requires mean independence of potential outcomes and the second-round running variable among observationally equivalent individuals who choose to participate in the second-round RD. Recall that to identify $E[\theta_{1,1}|Z_1 = 0]$, CFR assumes $\theta_0 \equiv \theta^1_{0,2} = \theta^0_{0,2} = \theta_{0,1}$ and that individual immediate treatment effect θ_0 is either fixed constant or independent of the second-round treatment decision. The following assumption that we make is substantially weaker. Let X denote confounding covariates with support $\mathcal{X}$.

Assumption 3.1 (CIA - Benchmark) *There exists an $\epsilon > 0$ such that $E[Y_2(d_1,0)|X = x, Z_2(d_1) = z_2, S_2(d_1) = 1, Z_1 = z_1] = E[Y_2(d_1,0)|X = x, S_2(d_1) = 1, Z_1 = z_1]$, for all $x \in \mathcal{X}$, $z_2 \in \mathbb{R}$, $d_1 = 0, 1$ and all $z_1 \in \mathcal{N}_\epsilon$.*

Assumption 3.1 does not restrict the relationship among $\theta^1_{0,2}$, $\theta^0_{0,2}$, and $\theta_{0,1}$. It does not impose any restriction on the endogeneity of the second-round RD participation indicator, either. In addition, Assumption 3.1 only involves two potential second-period outcomes with no treatment in the second round. This distinction could make a big difference in applied research. For example, in the empirical application in Section 6, it would be a strong assumption to require the potential expenditure associated with second-round school bond authorization to be mean independent of the potential vote-share since detailed terms of school bond measures could affect both. Assuming mean independence between the potential outcome with no second-round treatment and the running variable is much weaker.

It is worthwhile to point out that the covariate X in Assumption 3.1 could be pre-RD individual characteristics, but could also include lagged potential outcome $Y_1(d_1)$ since $Y_1(d_1)$ is realized before both $S_2(d_1)$ and $Z_2(d_1)$. If $Y_2(d_1, 0) = \rho \cdot Y_1(d_1) + u$ follows an AR(1) process with random shock u, Assumption 3.1 is clearly satisfied with $X = Y_1(d_1)$.

Assumption 3.1 follows the idea in Angrist and Rokkanen (2015) for extrapolating RD effects away from the running variable cutoff in a classic one-period RD setting. In addition to Assumption 3.1, we also need a stronger version of the smoothness conditions for each subpopulation defined by X. Similar to Assumption 3.1, the next Assumption is tailored for the identification of the one-period-after ATE.

Assumption 3.2 (Smoothness - Benchmark) *There exists an $\epsilon > 0$, such that*

1. *Z_1 is continuous in $z_1 \in \mathcal{N}_\epsilon$ with $P[Z_1 \geq 0] \in (0,1)$;*
2. *$E[D_2(d_1)|X = x, S_2(d_1) = 1, Z_1 = z_1] \in (0,1)$ is continuous in $z_1 \in \mathcal{N}_\epsilon$ for all $d_1 = 0,1$ and $x \in \mathcal{X}$;*
3. *$E[Y_2(d_1,d_2)|X = x, D_2(d_1) = d_2, Z_1 = z_1]$ is continuous in $z_1 \in \mathcal{N}_\epsilon$ for all $d_1, d_2 = 0,1$ and $x \in \mathcal{X}$.*

Under Assumption 3.2, it is easy to see that $E[Y_2(1,1)|D_2(1) = 1, Z_1 = 0]$ in equation (3.2) could be identified as $\lim_{z_1 \searrow 0} E[Y_2|D_2 = 1, Z_1 = z_1]$. Meanwhile, under Assumptions 3.1 and 3.2,

$$
\begin{aligned}
&E[Y_2(1,0)|D_2(1) = 1, Z_1 = 0] \\
=& \lim_{z_1 \searrow 0} E[E[Y_2(1,0)|X, S_2(1) = 1, Z_2(1) \geq 0, Z_1 = z_1]|S_2(1) = 1, Z_2(1) \geq 0, Z_1 = z_1] \\
=& \lim_{z_1 \searrow 0} E[E[Y_2(1,0)|X, S_2(1) = 1, Z_2(1) < 0, Z_1 = z_1]|S_2(1) = 1, Z_2(1) \geq 0, Z_1 = z_1] \\
=& \lim_{z_1 \searrow 0} E[E[Y_2|X, S_2 = 1, D_2 = 0, Z_1 = z_1]|D_2 = 1, Z_1 = z_1].
\end{aligned}
$$

Combining results above completes the identification of $E[\theta_{0,2}^1|D_2(1)=1,Z_1=0]$; similarly, $E[\theta_{0,2}^0|D_2(0)=1,Z_1=0]$ could be identified. The identification of $E[\theta_{1,1}|Z_1=0]$ then follows.

When the dimension of X is large, a propensity score version of the above identification strategy would be more practical. Let

$$\lambda^{d_1}(x)=P[D_2(d_1)=1|X=x,S_2(d_1)=1,Z_1=0],\quad d_1=0,1, \tag{3.4}$$

be the *potential* propensity scores of passing the second-round RD cutoff among marginal individuals who are at the first-round cutoff and would participate in the second-round eligibility test given the first-round treatment status d_1. The potential propensity scores are identified by

$$\lambda^0(x)=\lim_{z_1\nearrow 0}P[D_2=1|X=x,S_2=1,Z_1=z_1],$$
$$\lambda^1(x)=\lim_{z_1\searrow 0}P[D_2=1|X=x,S_2=1,Z_1=z_1].$$

Under Assumptions 3.1 and 3.2, one can show that

$$\begin{aligned}
&E[\theta_{0,2}^0|D_2(0)=1,Z_1=0]=E[\theta_{0,2}^0|S_2(0)=1,Z_2(0)\geq 0,Z_1=0]\\
&=\lim_{z_1\nearrow 0}E\left[E\left[\left(\frac{Y_2D_2}{\lambda^0(X)}-\frac{Y_2(1-D_2)}{1-\lambda^0(X)}\right)|X,S_2=1,Z_1=z_1\right]|S_2=1,Z_2\geq 0,Z_1=z_1\right]\\
&=\lim_{z_1\nearrow 0}E\left[\frac{Y_2S_2(D_2-\lambda^0(X))}{(1-\lambda^0(X))E[D_2|Z_1=z_1]}|Z_1=z_1\right],
\end{aligned} \tag{3.5}$$

and similarly,

$$E[\theta_{0,2}^1|D_2(1)=1,Z_1=0]=\lim_{z_1\searrow 0}E\left[\frac{Y_2S_2(D_2-\lambda^1(X))}{(1-\lambda^1(X))E[D_2|Z_1=z_1]}|Z_1=z_1\right]. \tag{3.6}$$

Plugging equations (3.5) and (3.6) into equation (3.2), we have the following lemma.

Lemma 3.1 *Under Assumptions 3.1 and 3.2, the one-period-after ATE of the first-round treatment is identified as*

$$\begin{aligned}
&E[\theta_{1,1}|Z_1=0]=\alpha^1-\alpha^0,\quad \textit{with}\\
&\alpha^1\equiv\lim_{z_1\searrow 0}E\left[Y_2-\frac{Y_2S_2(D_2-\lambda^1(X))}{1-\lambda^1(X)}|Z_1=z_1\right],\\
&\alpha^0\equiv\lim_{z_1\nearrow 0}E\left[Y_2-\frac{Y_2S_2(D_2-\lambda^0(X))}{1-\lambda^0(X)}|Z_1=z_1\right].
\end{aligned}$$

The proof is given in the online appendix. The lemma identifies the one-period-after ATE of the first-round treatment. As is discussed at the beginning of the section, identification results in Lemma 3.1 could be extended, given strengthened smoothness and CIA conditions, to distributional treatment effects or ATEs of the k-th round treatment conditional on the path of past treatment statuses. For example, the strategy could be applied to identify $E[\theta_{0,2}^{d_1}|S_2(d_1) = 1, Z_1 = 0]$, for $d_1 = 0, 1$, provided that a stronger version for Assumption 3.1 with all four second-round potential outcomes holds.

Note that if the CIA condition in Assumption 3.1 is relaxed to a monotonicity condition as is stated in the following, identification results in equations (3.5) and (3.6) become upper bounds of $E[\theta_{0,2}^{0}|D_2(d_1) = 1, Z_1 = 0]$ and $E[\theta_{0,2}^{1}|D_2(d_1) = 1, Z_1 = 0]$, respectively.

Assumption 3.3 (Monotone 1 - Benchmark) $E[Y_2(d_1, 0)|X = x, Z_2(d_1) = z_2, S_2(d_1) = 1, Z_1 = z_1]$ *is (weakly) monotonically increasing in* z_2 *for all* $x \in \mathcal{X}$ *and* $z_1 \in \mathcal{N}_\epsilon$.

Assumption 3.3 assumes that the potential second-round running variable has a monotonic relationship with the conditional mean of potential second-round outcomes with no second-round treatment. It nests the CIA condition in Assumption 3.1. Under Assumption 3.3, for $d_1 = 0, 1$,

$$
\begin{aligned}
&E[\theta_{0,2}^{d_1}|D_2(d_1) = 1, Z_1 = 0] \\
=&E[Y_2(d_1, 1)|D_2(d_1) = 1, Z_1 = 0] - E[Y_2(d_1, 0)|D_2(d_1) = 1, Z_1 = 0] \\
\leq&E[Y_2(d_1, 1)|D_2(d_1) = 1, Z_1 = 0] \\
&- E[E[Y_2(d_1, 0)|X, S_2(d_1) = 1, Z_2(d_1) < 0, Z_1 = 0|S_2(d_1) = 1, Z_2(d_1) \geq 0, Z_1 = 0] \\
=&\alpha^{d_1},
\end{aligned}
$$

where α^{d_1} is defined in Lemma 3.1. The upper bounds of $E[\theta_{0,2}^{d_1}|D_2(d_1) = 1, Z_1 = 0]$ alone, for $d_1 = 0, 1$, cannot be used to bound the one-period-after ATE $E[\theta_{1,1}|Z_1 = 0]$. For this purpose, we introduce a second monotonicity condition as following.

Assumption 3.4 (Monotone 2 - Benchmark) $E[\theta_{0,2}^{d_1}|X = x, Z_2(d_1) = z_2, S_2(d_1) = 1, Z_1 = z_1]$ *is (weakly) monotonically increasing in* $z_2 \in \mathbb{R}$ *for all* $x \in \mathcal{X}$ *and* $z_1 \in \mathcal{N}_\epsilon$.

Assumption 3.4 assumes that the potential second-round running variable has a monotonic relationship with the immediate second-period ATE. When the continuity conditions in Assumptions 2.1 are extended to conditional means of potential outcomes conditional on both Z_1 and Z_2, it is easy to show that under Assumption 3.4,

$$\begin{aligned}
&E[\theta^0_{0,2}|D_2(0)=1, Z_1=0] \geq E[\theta^0_{0,2}|S_2(0)=1, Z_2(0)=0, Z_1=0] \\
=&\lim_{z_1\nearrow 0, z_2\searrow 0} E[Y_2|S_2=1, Z_2=z_2, Z_1=z_1] - \lim_{z_1\nearrow 0, z_2\nearrow 0} E[Y_2|S_2=1, Z_2=z_2, Z_1=z_1] \\
\equiv&\beta^0, \\
&E[\theta^1_{0,2}|D_2(1)=1, Z_1=0] \geq E[\theta^1_{0,2}|S_2(1)=1, Z_2(1)=0, Z_1=0] \\
=&\lim_{z_1\searrow 0, z_2\searrow 0} E[Y_2|S_2=1, Z_2=z_2, Z_1=z_1] - \lim_{z_1\searrow 0, z_2\nearrow 0} E[Y_2|S_2=1, Z_2=z_2, Z_1=z_1] \\
\equiv&\beta^1.
\end{aligned}$$

Combining the inequalities above and the decomposition stated in equation (3.2), we bound the one-period-after ATE $E[\theta_{1,1}|Z_1=z_1]$ as

$$\begin{aligned}
&\alpha^1 - \left(\lim_{z_1\nearrow 0} E[Y_2|Z_1=z_1] - \beta^0 \cdot \lim_{z_1\nearrow 0} P[D_2=1|Z_1=z_1]\right) \\
\leq& E[\theta_{1,1}|Z_1=z_1] \\
\leq& \left(\lim_{z_1\searrow 0} E[Y_2|Z_1=z_1] - \beta^1 \cdot \lim_{z_1\searrow 0} P[D_2=1|Z_1=z_1]\right) - \alpha^0.
\end{aligned}$$

3.2 Longer-term ATEs

The identification strategy discussed above for one-period-after ATE could be extended to other longer-term ATEs of the first-round treatment with additional assumptions. Although the previous section does not impose any restriction on the path-dependency of potential random variables, for the identification of longer-term ATEs we introduce a new Markovian-type condition.

Assumption 3.5 (Markovian) *For all $k = 3, 4, ..., K$, $\tau = 0, ..., K-k$ and $z_1 \in \mathcal{N}_\epsilon$, we have that for all $\ell^{k-2} \in \mathcal{L}^{k-2}$ and $d = 0, 1$,*

1. $\tilde{\mu}^d_\tau \equiv E\left[\tilde{\theta}^{(\ell^{k-2},d)}_{\tau,k}\middle|D_k(\ell^{k-2},d)=1, Z_1=z_1\right] = E\left[\tilde{\theta}^d_{\tau,2}\middle|D_2(d)=1, Z_1=z_1\right]$,
2. $\tilde{\nu}^d_\tau \equiv E\left[\tilde{\eta}^{(\ell^{k-2},d)}_{\tau,k}\middle|D_k(\ell^{k-2},d)=1, Z_1=z_1\right] = E\left[\tilde{\eta}^d_{\tau,2}\middle|D_2(d)=1, Z_1=z_1\right]$.

Assumption 3.5 requires that different τ-period-after AITTEs and average total first-stage effects on the treated have the same conditional means as long as the treatment statuses in the period right before the focal treatment are the same.[3] The assumption can be restrictive, but it reduces the dimension of unknown treatment effects involved in identification since the cardinal count of individual treatment effects increases exponentially with the number of time periods if no further restriction is imposed.[4] Assumption 3.5 is still substantially weaker than the complete path-independence condition used in CFR.

Assumption 3.6 (Smoothness) *There exists an $\epsilon > 0$ such that for all $\tau \geq 0$,*

1. *Z_1 is continuous in $z_1 \in \mathcal{N}_\epsilon$ with $P[Z_1 \geq 0 | X = x, S = 1] \in (0, 1)$ for all $x \in \mathcal{X}$;*

2. *$E[D_2(d_1)|X = x, S_2(d_1) = 1, Z_1 = z_1]$ is continuous in $z_1 \in \mathcal{N}_\epsilon$ for all $d_1 = 0, 1$ and $x \in \mathcal{X}$;*

3. *$E[\tilde{Y}_{2+\tau}(d_1, d_2)|X = x, D_2(d_1) = d_2, Z_1 = z_1]$ and $E[\tilde{D}_{3+\tau}(d_1, d_2)|X = x, D_2(d_1) = d_2, Z_1 = z_1]$ are continuous in $z_1 \in \mathcal{N}_\epsilon$ for all $d_1, d_2 = 0, 1$ and $x \in \mathcal{X}$.*

Assumption 3.6 reduces to Assumption 3.2 when $\tau = 0$. For $\tau \geq 1$, recall that both the quasi-potential outcome $\tilde{Y}_{2+\tau}(d_1, d_2)$ and the quasi-potential treatment decision $\tilde{D}_{3+\tau}(d_1, d_2)$ could be viewed as a classic second-period potential outcome; $d_1, d_2 = 0, 1$. Therefore, Assumption 3.6 is essentially a reiteration of the smoothness conditions in Assumption 3.2.

Combining Assumptions 3.5 and 3.6 with the decomposition result in equation (3.1),

[3] In fact, for the identification strategy described in this section, only $d = 0$ is required. However, it might be hard to find an empirical situation where conditions in Assumption 3.5 hold with $d = 0$ but not with $d = 1$, so we do not make such a distinction in Assumption 3.5.

[4] For example, for each $\tau = 0, 1, ..., K - k$, without Assumption 3.5, the k-th round treatment would have 2^{k-1} different τ-period-after individual primary or total effects. When Assumption 3.5 is imposed, the k-th round treatment would only have 2 different τ-period-after individual primary or total effects.

we have that for $\tau \geq 1$,

$$\begin{aligned} E[\theta_{\tau,1}|Z_1 = 0] = &\lim_{z_1 \searrow 0} E[Y_{\tau+1}|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[Y_{\tau+1}|Z_1 = z_1] \\ &- \tilde{\mu}^1_{\tau-1} \cdot \lim_{z_1 \searrow 0} P[D_2 = 1|Z_1 = z_1] + \tilde{\mu}^0_{\tau-1} \cdot \lim_{z_1 \nearrow 0} P[D_2 = 1|Z_1 = z_1] \\ &- \sum_{s=0}^{\tau-2} \tilde{\mu}^0_s \cdot E[\eta_{\tau-1-s,1}|Z_1 = 0] \end{aligned} \tag{3.7}$$

and

$$\begin{aligned} E[\eta_{\tau,1}|Z_1 = 0] = &\lim_{z_1 \searrow 0} E[D_{\tau+2}|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[D_{\tau+2}|Z_1 = z_1] \\ &- \tilde{\nu}^1_{\tau-1} \cdot \lim_{z_1 \searrow 0} P[D_2 = 1|Z_1 = z_1] + \tilde{\nu}^0_{\tau-1} \cdot \lim_{z_1 \nearrow 0} P[D_2 = 1|Z_1 = z_1] \\ &- \sum_{s=0}^{\tau-2} \tilde{\nu}^0_s \cdot E[\eta_{\tau-1-s,1}|Z_1 = 0]. \end{aligned} \tag{3.8}$$

When $\tau = 1$, equation (3.7) reduces to

$$\begin{aligned} E[\theta_{1,1}|Z_1 = 0] = &\lim_{z_1 \searrow 0} E[Y_2|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[Y_2|Z_1 = z_1] \\ &- \tilde{\mu}^1_0 \cdot \lim_{z_1 \searrow 0} P[D_2 = 1|Z_1 = z_1] + \tilde{\mu}^0_0 \cdot \lim_{z_1 \nearrow 0} P[D_2 = 1|Z_1 = z_1], \end{aligned}$$

which is equivalent to equation (3.2). Equation (3.8) reduces to

$$\begin{aligned} E[\eta_{1,1}|Z_1 = 0] = &\lim_{z_1 \searrow 0} E[D_3|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[D_3|Z_1 = z_1] \\ &- \tilde{\nu}^1_0 \cdot \lim_{z_1 \searrow 0} P[D_2 = 1|Z_1 = z_1] + \tilde{\nu}^0_0 \cdot \lim_{z_1 \nearrow 0} P[D_2 = 1|Z_1 = z_1]. \end{aligned}$$

When $\tau = 2$, equations (3.7) and (3.8) reduce, correspondingly, to

$$\begin{aligned} E[\theta_{2,1}|Z_1 = 0] = &\lim_{z_1 \searrow 0} E[Y_3|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[Y_3|Z_1 = z_1] \\ &- \tilde{\mu}^1_1 \cdot \lim_{z_1 \searrow 0} P[D_2 = 1|Z_1 = z_1] + \tilde{\mu}^0_1 \cdot \lim_{z_1 \nearrow 0} P[D_2 = 1|Z_1 = z_1] \\ &- \tilde{\mu}^0_0 \cdot E[\eta_{1,1}|Z_1 = 0], \text{ and} \\ E[\eta_{2,1}|Z_1 = 0] = &\lim_{z_1 \searrow 0} E[D_4|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[D_4|Z_1 = z_1] \\ &- \tilde{\nu}^1_1 \cdot \lim_{z_1 \searrow 0} P[D_2 = 1|Z_1 = z_1] + \tilde{\nu}^0_1 \cdot \lim_{z_1 \nearrow 0} P[D_2 = 1|Z_1 = z_1] \\ &- \tilde{\nu}^0_0 \cdot E[\eta_{1,1}|Z_1 = 0]. \end{aligned}$$

To identify components in the right hand side of equations (3.7) and (3.8), we strengthen the CIA condition using quasi-potential outcomes and treatment decisions defined in Section 2.2. Again, when the quasi-potential outcome and the quasi-potential treatment decision stated in the following are viewed as classic second-period potential outcomes, the assumption below is just a reiteration of Assumption 3.1.

Assumption 3.7 (CIA) *There exists an $\epsilon > 0$ such that for all $\tau \geq 0$,*

1. $E[\tilde{Y}_{2+\tau}(d_1, 0)|X = x, Z_2(d_1) = z_2, S_2(d_1) = 1, Z_1 = z_1] = E[\tilde{Y}_{2+\tau}(d_1, 0)|X = x, S_2(d_1) = 1, Z_1 = z_0]$ *for all $x \in \mathcal{X}$, $z_2 \in \mathbb{R}$, $d_1 = 0, 1$ and all $z_1 \in \mathcal{N}_\epsilon$;*

2. $E[\tilde{D}_{3+\tau}(d_1, 0)|X = x, Z_2(d_1) = z_2, S_2(d_1) = 1, Z_1 = z_1] = E[\tilde{D}_{3+\tau}(d_1, 0)| = x, S_2(d_1) = 1, Z_1 = z_1]$ *for all $x \in \mathcal{X}$, $z_2 \in \mathbb{R}$, $d_1 = 0, 1$ and all $z_1 \in \mathcal{N}_\epsilon$.*

Under Assumptions 3.6 and 3.7,

$$\begin{aligned}
\tilde{\mu}^0_\tau &= \lim_{z_1 \nearrow 0} E\left[\frac{Y_{2+\tau} S_2(D_2 - \lambda^0(X))}{(1 - \lambda^0(X))E[D_2|Z_1 = z_1]}|Z_1 = z_1\right], \\
\tilde{\mu}^1_\tau &= \lim_{z_1 \searrow 0} E\left[\frac{Y_{2+\tau} S_2(D_2 - \lambda^1(X))}{(1 - \lambda^1(X))E[D_2|Z_1 = z_1]}|Z_1 = z_1\right], \\
\tilde{\nu}^0_\tau &= \lim_{z_1 \nearrow 0} E\left[\frac{D_{3+\tau} S_2(D_2 - \lambda^0(X))}{(1 - \lambda^0(X))E[D_2|Z_1 = z_1]}|Z_1 = z_1\right], \\
\tilde{\nu}^1_\tau &= \lim_{z_1 \searrow 0} E\left[\frac{D_{3+\tau} S_2(D_2 - \lambda^1(X))}{(1 - \lambda^1(X))E[D_2|Z_1 = z_1]}|Z_1 = z_1\right]. \qquad (3.9)
\end{aligned}$$

Plugging into the decompositions in equations (3.7) and (3.8), we obtain a recursive identification strategy of $E[\theta_{\tau,1}|Z_1 = 0]$ and $E[\eta_{\tau,1}|Z_1 = 0]$, for $\tau \geq 1$.

When $\tau = 1$, equation (3.7) simplifies to the identification in Lemma 3.1, and equation (3.8) simplifies to

$$\begin{aligned}
E[\eta_{1,1}|Z_1 = 0] = &\lim_{z_1 \searrow 0} E\left[D_3 - \frac{D_3 S_2(D_2 - \lambda^1(X))}{1 - \lambda^1(X)}|Z_1 = z_1\right] \\
&- \lim_{z_1 \nearrow 0} E\left[D_3 - \frac{D_3 S_2(D_2 - \lambda^0(X))}{1 - \lambda^0(X)}|Z_1 = z_1\right].
\end{aligned}$$

When $\tau = 2$, $E[\theta_{2,1}|Z_1 = 0]$ is identified with results in (3.9) and the identification of

$E[\eta_{1,1}|Z_1 = 0]$. Specifically,

$$
\begin{aligned}
&E[\theta_{2,1}|Z_1 = 0] \\
=& \lim_{z_1 \searrow 0} E\left[Y_3 - \frac{Y_3 S_2(D_2 - \lambda^1(X))}{1 - \lambda^1(X)}|Z_1 = z_1\right] - \lim_{z_1 \nearrow 0} E\left[Y_3 - \frac{Y_3 S_2(D_2 - \lambda^0(X))}{1 - \lambda^0(X)}|Z_1 = z_1\right] \\
&- \lim_{z_1 \nearrow 0} E\left[\frac{Y_2 S_2(D_2 - \lambda^0(X))}{1 - \lambda^0(X)}|Z_1 = z_1\right] / E[D_2|Z_1 = z_1] \\
&\times \left(\lim_{z_1 \searrow 0} E\left[D_3 - \frac{D_3 S_2(D_2 - \lambda^1(X))}{1 - \lambda^1(X)}|Z_1 = z_1\right] - \lim_{z_1 \nearrow 0} E\left[D_3 - \frac{D_3 S_2(D_2 - \lambda^0(X))}{1 - \lambda^0(X)}|Z_1 = z_1\right]\right).
\end{aligned}
$$

Similarly, $E[\eta_{2,1}|Z_1 = 0]$ is identified and

$$
\begin{aligned}
&E[\eta_{2,1}|Z_1 = 0] \\
=& \lim_{z_1 \searrow 0} E\left[D_4 - \frac{D_4 S_2(D_2 - \lambda^1(X))}{1 - \lambda^1(X)}|Z_1 = z_1\right] - \lim_{z_1 \nearrow 0} E\left[D_4 - \frac{D_4 S_2(D_2 - \lambda^0(X))}{1 - \lambda^0(X)}|Z_1 = z_1\right] \\
&- \lim_{z_1 \nearrow 0} E\left[\frac{D_3 S_2(D_2 - \lambda^0(X))}{1 - \lambda^0(X)}|Z_1 = z_1\right] / E[D_2|Z_1 = z_1] \\
&\times \left(\lim_{z_1 \searrow 0} E\left[D_3 - \frac{D_3 S_2(D_2 - \lambda^1(X))}{1 - \lambda^1(X)}|Z_1 = z_1\right] - \lim_{z_1 \nearrow 0} E\left[D_3 - \frac{D_3 S_2(D_2 - \lambda^0(X))}{1 - \lambda^0(X)}|Z_1 = z_1\right]\right).
\end{aligned}
$$

When $\tau = 3$, $E[\theta_{3,1}|Z_1 = 0]$ and $E[\eta_{3,1}|Z_1 = 0]$ are identified with the results in (3.9) and the identification of $E[\eta_{1,1}|Z_1 = 0]$ and $E[\eta_{2,1}|Z_1 = 0]$. The recursive identification pattern is then clear.

The recursive identification strategy could also be extended to partially identify $E[\tilde{\theta}_{\tau,1}|Z_1 = 0]$ for $\tau \geq 2$ if both $\tilde{\mu}_\tau^d$ and $\tilde{\nu}_\tau^d$ are bounded using monotonic conditions similar to those in Assumptions 3.3 and 3.4. We omit the details here.

4 Estimation and Inference

In this section, we propose estimation and inference procedures for the identified average effects discussed in Sections 3.1 and 3.2.[5] We propose a two-step semiparametric estimation procedure. We focus on the one-period-after ATE following the identification in Lemma 3.1. Given the recursive nature in longer-term ATE identification, the one-period-after ATE estimator that we focus on extends readily to longer-term ATEs. For inference, we propose a weighted bootstrap procedure. The procedure is especially helpful

[5]The effects identified in Sections 2.3 and 2.4 can be estimated by conventional nonparametric RD estimators. See, for example, Chiang et al. (2019). We omit the details.

for longer-term ATE estimators, which include many more conditional mean expressions when the lag between the outcome and the focal round of treatment increases.

4.1 Estimation

Assume that $\lambda(X;z_1) = P[D_2 = 1|X, S_2 = 1, Z_1 = z_1]$ follows a class of semiparametric models $p(X,\gamma(z_1))$, where $p(.,.) : \mathcal{X} \times \Gamma \to \mathbb{R}$ and $\gamma(.) : \mathbb{R} \to \Gamma$ is unknown. For example, if $p(X,\gamma) = L(X'\gamma)$ with $L(a) = \exp(a)/(1+\exp(a))$, then $\lambda(X;z_1) = \frac{\exp(X'\gamma(z_1))}{1+\exp(X'\gamma(z_1))}$ follows a varying coefficient Logit model, nesting both parametric Logit and semiparametric partial-linear Logit as special cases. In this section we describe estimation strategy with a general semi-parametric propsensity score function. In Section 4.3 we provide details with a varying coefficient Logit propsensity score model.

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 $p(X,\gamma^1)$. In addition, let $\beta^0 = \lim_{z_1 \nearrow 0} \gamma'(z_1)$ and $\beta^1 = \lim_{z_1 \searrow 0} \gamma'(z_1)$ be the left and the right limits of the first-order derivative of $\gamma(.)$ at the RD cutoff. 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, correspondingly. In order to estimate the propensity score function locally at the RD cutoff while also avoiding the "curse of dimensionality", we propose to solve the following maximization problem:

$$
\begin{aligned}
(\hat{\gamma}^1, \hat{\beta}^1_{FS}) = \arg\max_{\gamma,\beta} \sum_{i=1}^{n} & S_{2i} 1(Z_{1i} \geq 0) K\left(\frac{Z_{1i}}{h}\right) \\
& \cdot \Big[D_{2i} \log p(X_i, \gamma + \beta Z_{1i}) + (1 - D_{2i}) \log(1 - p(X_i, \gamma + \beta Z_{1i})) \Big], \\
(\hat{\gamma}^0, \hat{\beta}^0_{FS}) = \arg\max_{\gamma,\beta} \sum_{i=1}^{n} & S_{2i} 1(Z_{1i} < 0) K\left(\frac{Z_{1i}}{h}\right) \\
& \cdot \Big[D_{2i} \log p(X_i, \gamma + \beta Z_{1i}) + (1 - D_{2i}) \log(1 - p(X_i, \gamma + \beta Z_{1i})) \Big],
\end{aligned} \tag{4.1}
$$

where $K(\cdot)$ is the kernel function and h is the bandwidth.

Denote the one-period-after ATE $E[\theta_{1,1}|Z_1 = 0]$ by $\bar{\theta}_{1,1}$ and its estimator by $\hat{\bar{\theta}}_{1,1}$. Given the first-step propensity score estimators, $\bar{\theta}_{1,1}$ could be estimated by local linear regressions as the following:

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

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

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

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

Estimators $\hat{\lambda}^0$, $\hat{\lambda}^1$, and $\hat{\bar{\theta}}$ are all consistent and asymptotically normally distributed under proper assumptions, which will be elaborated in the next sections. In the above estimation procedure, the bandwidth is kept the same in both steps. If the first-step bandwidth converges to zero at a slower rate than the second-step bandwidth, asymptotic variance of $\hat{\bar{\theta}}_{1,1}$ could be simplified since the first-step estimation error would vanish asymptotically. We propose to keep bandwidths the same in both steps for easier interpretations of the empirical results. This implies that the asymptotic variance of $\hat{\bar{\theta}}_{1,1}$ needs to account for estimation error in the first step as well.

The two-step estimation procedure described above could be modified to estimate $E[\theta_{\tau,1}|Z_1 = 0]$ for $\tau \geq 2$ as well. Take the two-period-after ATE as an example. Let $\bar{\theta}_{2,1} \equiv E[\theta_{2,1}|Z_1 = 0]$ and recall that

$$\bar{\theta}_{2,1} = \alpha_1^1 - \alpha_1^0 - (\tilde{\mu}_{nu}^0 / \tilde{\mu}_{de}) \cdot (\alpha_{fs}^1 - \alpha_{fs}^0), \text{ where}$$

$$\alpha_1^1 = \lim_{z_1 \searrow 0} E\left[Y_3 - \frac{Y_3 S_2 (D_2 - \lambda^1(X))}{1 - \lambda^1(X)} | Z_1 = z_1\right],$$

$$\alpha_1^0 = \lim_{z_1 \nearrow 0} E\left[Y_3 - \frac{Y_3 S_2 (D_2 - \lambda^0(X))}{1 - \lambda^0(X)} | Z_1 = z_1\right],$$

$$\alpha_{fs}^1 = \lim_{z_1 \searrow 0} E\left[D_3 - \frac{D_3 S_2 (D_2 - \lambda^1(X))}{1 - \lambda^1(X)} | Z_1 = z_1\right],$$

$$\alpha_{fs}^0 = \lim_{z_1 \nearrow 0} E\left[D_3 - \frac{D_3 S_2 (D_2 - \lambda^0(X))}{1 - \lambda^0(X)} | Z_1 = z_1\right],$$

$$\tilde{\mu}_{nu}^0 = \lim_{z_1 \nearrow 0} E\left[\frac{Y_2 S_2 (D_2 - \lambda^0(X))}{1 - \lambda^0(X)} | Z_1 = z_1\right], \; \tilde{\mu}_{de} = E[D_2 | Z_1 = z_1].$$

All components of $\bar{\theta}_{2,1}$ are conditional means of functions of observables or propensity scores given the first-round running variable. Therefore, $\bar{\theta}_{2,1}$ could be estimated by local linear regressions once the first-step local MLE propensity score estimators are plugged in. Estimation of more longer-term ATEs would follow similarly.

Before proceeding to the next section, where we discuss asymptotic properties of the proposed estimators, we note that the local MLE method used here for propensity score estimation could be useful in other RD settings as well. For example, the varying coefficient Logit model discussed above could be applied to estimate heterogeneous first-stage effects in classic static fuzzy RD models.[6] In addition, local MLE with other link functions could be applied to estimate heterogeneous ATEs with binary, count, or duration outcome variables. We leave such RD applications for future studies.

4.2 Asymptotics

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 $\hat{\gamma}^1$ and $\hat{\gamma}^0$, 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 $\tilde{\alpha}^0$, respectively, such that

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

Define gradient terms $\nabla_\gamma^1 = \lim_{z_1 \searrow 0} E\left[\nabla_\gamma \left[\frac{Y_2 S_2 (D_2 - p(X,\gamma))}{1 - p(X,\gamma)}\right] \Big|_{\gamma=\gamma^1} \Big| Z_1 = z_1\right]$ and $\nabla_\gamma^0 = \lim_{z_1 \nearrow 0} E\left[\nabla_\gamma \left[\frac{Y_2 S_2 (D_2 - p(X,\gamma))}{1 - p(X,\gamma)}\right] \Big|_{\gamma=\gamma^0} \Big| Z_1 = z_1\right]$. By the delta method, we obtain the fol-

[6] Let Z be the running variable in a static fuzzy RD model and c be the RD cutoff. Let D be the treatment decision and X the covariate. Researchers are often interested in knowing how RD first-stage or treatment effects vary with X. Hsu and Shen (2019, 2021) propose nonparametric tests for such questions. The local MLE estimator discussed above could easily estimate conditional mean functions such as $\lim_{z \searrow c} E[D|X = x, Z = z] - \lim_{z \nearrow c} E[D|X = x, Z = z]$. A semiparametric local MLE estimator would be more restrictive than a fully nonparametric estimator but it enjoys a faster rate of convergence while still remains local at the RD cutoff.

lowing influence function representation of $\hat{\alpha}^0$ and $\hat{\alpha}^1$ such that for $d_1 = 0, 1$,

$$
\begin{aligned}
&\sqrt{nh}(\hat{\alpha}^{d_1} - \alpha^{d_1}) \\
=&\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \Big(\tilde{\phi}_{\alpha^{d_1},ni}(Y_{2i}, S_{2i}, D_{2i}, Z_{1i}, X_i) - \bigtriangledown_{\gamma}^{d_1} \cdot \phi_{\gamma^{d_1},ni}(D_{2i}, S_{2i}, Z_{1i}, X_i) \Big) + o_p(1) \\
\equiv&\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \phi_{\alpha^{d_1},ni}(Y_{2i}, S_{2i}, D_{2i}, Z_{1i}, X_i) + o_p(1). \qquad (4.2)
\end{aligned}
$$

The influence functions would give us asymptotic limit of the one-period-after ATE estimator $\hat{\bar{\theta}}_{1,1}$. The asymptotic variance of $\hat{\bar{\theta}}_{1,1}$ could also be estimated by

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

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

The analytical variance estimator of the proposed estimators could be tedious to calculate. In applications, we propose to use a weighted bootstrap procedure first introduced in Ma and Kosorok (2005) to simulate the limiting distribution of the proposed one-period-after ATE estimator.

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

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

where

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

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

Following Ma and Kosorok (2005), $\sqrt{nh}(\hat{\bar{\theta}}_{1,1}^{w} - \hat{\bar{\theta}}_{1,1})$ and $\sqrt{nh}(\hat{\bar{\theta}}_{1,1} - \bar{\theta}_{1,1})$ have the same limiting distribution under suitable conditions. We will give out detailed conditions in the next section using a varying coefficient Logit propensity score model in the first-step.

4.3 Asymptotics based on First Step Varying Coefficient Logit Model

In this section, we provide detailed assumptions and asymptotic properties of the proposed two-step estimator using a varying coefficient Logit propensity score model in the first-step. Let $p(x, \gamma) = L(x'\gamma)$ with $L(a) = \exp(a)/(1 + \exp(a))$. Then

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

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

The following set of assumptions states regularity conditions for the estimators.

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

Assumption 4.2 *For $j = 1, \ldots, k$, the j-th element of $\gamma(z_1)$, or $\gamma_j(z_1)$, is twice continuously differentiable on $(-\epsilon, 0)$ and $(0, \epsilon)$ with corresponding derivatives bounded for some $\epsilon > 0$.*

Assumption 4.3 *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 4.4 *Moment* $E[\|X\|^3|Z_1 = z_1]$ *exists and is bounded on* $\mathcal{N}_\epsilon$ *for some* $\epsilon > 0$.

Assumption 4.5 *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 smoothness conditions on the varying coefficient in neighborhoods right above and below the RD cut-off. Assumption 4.3 imposes standard smoothness conditions on the density of the running variable. Assumption 4.4 imposes a moment condition on the covariate X. Assumption 4.5 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 $n^{1/5}$.

Recall that $\phi_{\gamma^0,ni}(D_{2i}, S_{2i}, Z_{1i}, X_i)$ and $\phi_{\gamma^1,ni}(D_{2i}, S_{2i}, Z_{1i}, X_i)$ are influence functions of $\hat{\gamma}^0$ and $\hat{\gamma}^1$, respectively. Let I_k denote the $k \times k$ identity matrix and $\mathbf{0}_{k\times k}$ denote the $k \times k$ zero matrix. Under Assumptions 4.1-4.5, one can show that

$$\begin{aligned}\phi_{\gamma^d,ni}(D_{2i}, S_{2i}, Z_{1i}, X_i) =&(I_k \ \ \mathbf{0}_{k\times k})(\Delta^d)^{-1} S_{2i} \cdot 1(Z_{1i} \geq 0)^d \cdot 1(Z_{1i} < 0)^{1-d} \\ &\cdot K\left(Z_{1i}/h\right)\left(D_{2i} - L\big(X_i'(\gamma^d + \beta^d Z_{1i})\big)\right) \begin{pmatrix} X_i \\ Z_{1i}X_i/h \end{pmatrix},\end{aligned}$$

for $d = 0, 1$. The following lemma then provides asymptotic properties of the first-step local MLE estimators under a varying coefficient Logit specification. Similar results could be derived if the $\lambda(.;.)$ function follows some other semi-parametric models such as varying coefficient Probit.

Lemma 4.1 *Suppose that Assumptions 4.1-4.5 hold, then for* $d = 0, 1$,

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

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

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

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

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

Next, we discuss asymptotic properties of $\hat{\alpha}^0$ and $\hat{\alpha}^1$. We first state regularity conditions. Assumption 4.6 includes smoothness conditions for the infeasible two-step estimators with known first-step propensity score functions. Assumption 4.7 imposes conditions required to control the impact of first-stage estimation errors on the asymptotic properties of feasible two-step estimators $\hat{\alpha}^0$ and $\hat{\alpha}^1$. For notational simplicity, we use $\widetilde{Y}_2(\gamma)$ to denote $\frac{Y_2 S_2(D_2 - L(X'\gamma))}{(1-L(X'\gamma))}$. For $d_1 = 0, 1$, define $\widetilde{Y}_2^{d_1} = \frac{Y_2 S_2(D_2 - L(X'\gamma^{d_1}))}{(1-L(X'\gamma^{d_1}))}$ and $\nabla_\gamma \widetilde{Y}_2^{d_1} = \nabla_\gamma \widetilde{Y}_2(\gamma)|_{\gamma=\gamma^{d_1}}$. Let $\nabla^2_\gamma \widetilde{Y}_2(\gamma)$ be the Hessian matrix of $\widetilde{Y}_2(\gamma)$.

Assumption 4.6 *Assume that for some* $\epsilon > 0$,

1. $E[Y_2|Z = z]$ *and* $E[\widetilde{Y}_2^0|Z_1 = z_1]$ *are twice continuously differentiable on* $z \in [-\epsilon, 0)$ *with bounded corresponding derivatives;*

2. $E[Y_2|Z = z]$ *and* $E[\widetilde{Y}_2^1|Z_1 = z_1]$ *are twice continuously differentiable on* $z \in [0, \epsilon]$ *with bounded corresponding derivatives;*

3. $E[|Y_2|^3|Z_1 = z_1]$ *is bounded for* $z \in [-\epsilon, \epsilon]$, $E[|\widetilde{Y}_2^0|^3|Z_1 = z_1]$ *is bounded for* $z \in [-\epsilon, 0)$, *and* $E[|\widetilde{Y}_2^1|^3|Z_1 = z_1]$ *is bounded for* $z \in [0, \epsilon]$.

Assumption 4.7 *Assume that for some* $\epsilon > 0$,

1. *The third moment of the* j*-th element of* $\nabla_\gamma \widetilde{Y}_2^0$, *or* $E[|\nabla_\gamma \widetilde{Y}_{2j}^0|^3|Z_1 = z_1]$, *is bounded and twice continuously differentiable on* $z \in [-\epsilon, 0)$ *with bounded corresponding derivatives;*

2. *The third moment of the j-th element of* $\nabla_\gamma \widetilde{Y}_2^1$, *or* $E[|\nabla_\gamma \widetilde{Y}_{2j}^1|^3|Z_1 = z_1]$, *is bounded and twice continuously differentiable on* $z \in [0, \epsilon]$ *with bounded corresponding derivatives;*

3. $E[\sup_{\|\gamma-\gamma^0\|\leq\epsilon} \|\nabla_\gamma^2 \widetilde{Y}_2(\gamma)\|^2]$ *and* $E[\sup_{\|\gamma-\gamma^1\|\leq\epsilon} \|\nabla_\gamma^2 \widetilde{Y}_2(\gamma)\|^2]$ *are bounded.*

The following lemma summarize the inference function representations of $\hat{\alpha}^0$ and $\hat{\alpha}^1$.

Lemma 4.2 *Suppose that Assumptions 4.1-4.6 hold. Then* $\sqrt{nh}(\hat{\alpha}^0 - \alpha^d)$, *for* $d = 0, 1$, *has linear representations as in* (4.2) *with*

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

where $\Delta_z = f_{z_1}(0) \cdot \begin{pmatrix} \mu_{z,0} & \mu_{z,1} \\ \mu_{z,1} & \mu_{z,2} \end{pmatrix}$, $\bigtriangledown_\gamma^0 = \lim_{z\nearrow 0} E\left[\frac{Y_2S_2(D_2-1)L(X'\gamma^0)}{1-L(X'\gamma^0))} X' \Big| Z_1 = z_1\right]$, $\bigtriangledown_\gamma^1 = \lim_{z\searrow 0} E\left[\frac{Y_2S_2(D_2-1)L(X'\gamma^1)}{1-L(X'\gamma^1))} X' \Big| Z_1 = z_1\right]$, *and* $\mu_{z,j} = \int_{u\geq 0} u^j K(u)du$ *for* $j = 1, 2, \ldots$. *The influence function representation further implies that for* $d = 0, 1$,

$$\sqrt{nh}\left(\hat{\alpha}^d - \alpha^d\right) \xrightarrow{d} N(0, V_{\alpha^d}),$$

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 and, in general, we do not need these expressions to obtain consistent estimators for V_{α^0} and V_{α^1} or to make inferences. The next theorem summarizes asymptotic properties of $\hat{\bar{\theta}}_{1,1}$.

Theorem 4.1 *Suppose that Assumptions 4.1-4.6 hold. Then*

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

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

The result in Theorem 4.1 follows directly from Lemma 4.2 as it is straightforward to see that $\sqrt{nh}(\hat{\alpha}^0 - \alpha^0)$ and $\sqrt{nh}(\hat{\alpha}^1 - \alpha^1)$, estimated based on two mutually exclusive subsamples of observations, are asymptotically independent.

Asymptotic properties of longer-term ATE estimators could be derived in the same way. Again, take the two-period-after ATE as an example. 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 the estimators of α_1^1, α_1^0, $\tilde{\mu}_{nu}^0$, $\tilde{\mu}_{de}$, α_{fs}^1, and α_{fs}^0 defined in equation (3.2), and 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 their influence functions, correspondingly. 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 (4.2) for $\hat{\alpha}^0$ and $\hat{\alpha}^1$. The influence function $\phi_{\tilde{\mu}_{de},ni}$ 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 $\hat{\bar{\theta}}_{2,1} = \hat{\alpha}_1^1 - \hat{\alpha}_1^0 - (\hat{\tilde{\mu}}_{nu}^0/\hat{\tilde{\mu}}_{de}) \cdot (\hat{\alpha}_{fs}^1 - \hat{\alpha}_{fs}^0)$ be the estimator of $\bar{\theta}_{2,1}$. By the delta method, we have that

$$
\begin{aligned}
&\sqrt{nh}(\hat{\bar{\theta}}_{2,1} - \bar{\theta}_{2,1}) \\
=&\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \phi_{\alpha_1^1,ni} - \phi_{\alpha_1^0,ni} - \frac{\alpha_{fs}^1 - \alpha_{fs}^0}{\tilde{\mu}_{de}} \phi_{\tilde{\mu}_{nu}^0,ni} + \frac{\tilde{\mu}_{nu}^0(\alpha_{fs}^1 - \alpha_{fs}^0)}{(\tilde{\mu}_{de})^2} \phi_{\tilde{\mu}_{de},ni} \\
&\qquad - \frac{\tilde{\mu}_{nu}^0}{\tilde{\mu}_{de}} (\phi_{\alpha_{fs}^1,ni} - \phi_{\alpha_{fs}^0,ni}) + o_p(1) \\
\equiv&\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \phi_{\bar{\theta}_{2,1},ni}(Y_{3i}, Y_{2i}, D_{3i}, D_{2i}, S_{2i}, Z_{1i}, X_i) + o_p(1).
\end{aligned}
$$

In the right hand side of the first equality, we omit the arguments of influence functions for brevity. The asymptotic normality of $\hat{\bar{\theta}}_{2,1}$ follows from the influence function representation under proper smoothness conditions.

Last but not least, we provide details of the weighted bootstrap method given the varying coefficient Logit first-stage model.

Theorem 4.2 *Suppose that Assumptions 4.1-4.6 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. Then,*

$$
\begin{aligned}
&\sqrt{nh}(\hat{\bar{\theta}}_{1,1}^w - \hat{\bar{\theta}}_{1,1}) \\
=&\frac{1}{\sqrt{nh}} \sum_{i=1}^{n} (W_i - 1)\big(\phi_{\alpha^1,ni}(Y_{2i}, D_{2i}, S_{2i}, Z_{1i}, X_i) - \phi_{\alpha^0,ni}(Y_{2i}, D_{2i}, S_{2i}, Z_{1i}, X_i)\big) + o_p(1)
\end{aligned}
$$

and $\sqrt{nh}(\hat{\bar{\theta}}_{1,1}^w - \hat{\bar{\theta}}_{1,1}) \xrightarrow{d} N(0, V_{\alpha^1} + V_{\alpha^0})$ conditional on sample path with probability approaching one.

Theorem 4.2 establishes the validity of the weighted bootstrap estimator for $\bar{\theta}_{1,1}$. The proof follows Ma and Kosorok (2005) and is given in the online appendix. Although W_i

can follow any distribution with unit mean and variance, we use, in the simulation and empirical sessions, 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.

Finally, note that the weighted bootstrap procedure could be applied to local linear recursive CFR estimators following the identification in Section 2.4 and classic local linear estimation. For example, let $\hat{\bar{\theta}}^w_{2,1}$ denote the weighted bootstrap estimator. Following the same arguments as in Theorem 4.2, we can show that

$$\sqrt{nh}(\hat{\bar{\theta}}^w_{2,1} - \hat{\bar{\theta}}_{2,1}) = \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} (W_i - 1)\phi_{\bar{\theta}_{2,1},ni}(Y_{3i}, Y_{2i}, D_{3i}, D_{2i}, S_{2i}, Z_{1i}, X_i) + o_p(1),$$

so the validity of weighted bootstrap would follow and we omit the details.

5 Monte Carlo simulations

In this section, we study the small sample performance of the proposed estimator $\hat{\bar{\theta}}_{1,1}$ for the one-period-after ATE $E[\theta_{1,1}|Z_1 = 0]$ using Monte Carlo simulations. Small sample performances of two-period-after and three-period-after ATE estimators are reported in the online appendix.

We use five data-generating processes (DGPs). DGP 1 illustrates a case where the individual treatment effects are fixed and only need to be labeled by the number of periods between the outcome variable and the focal round of RD. DGP 2 illustrates a case where individual treatment effects are fixed but path-dependent. DGP 3 illustrates a case where second-round immediate treatment effects are path independent, independent of the second-round treatment decision, and share the same distribution as the first-round immediate effect. DGP 4 illustrates a case where the second-round immediate treatment effect is correlated with the second-round RD participation decision. Finally, DGP 5 illustrates a case where the potential second-period outcome is correlated with the second-round running variable even after conditioning on covariates, RD participation, and the first-stage running variable. The proposed estimator is valid under DGPs 1-4, while the recursive CFR estimator is only valid under DGPs 1 and 3. Under DGP 5, both the proposed estimator and the recursive CFR estimator are invalid.

For all DGPs,

$$X \sim U[0,10],\ Z_1 \sim X - 10 \cdot Beta(2,2),$$

$$u_{y1} \sim N(0,0.5),\ u_{s2} \sim N(0,1),\ v_{z2} \sim logis(0,1),\ u_{y2} \sim N(0,0.5),$$

$$Y_1(0) = 0.1X + 0.5Z_1 + 0.1XZ_1 + 0.1Z_1^2 + u_{y1},\ Y_1(1) = Y_1(0) + \theta_{0,1},$$

$$S_2(0) = 1(u_{s2} \geq 0),\ S_2(1) = 1(1 + u_{s2} \geq 0),$$

$$Z_2(0) = 0.3 + 0.1X + v_{z2},\ Z_2(1) = Z_2(0) + (1\ X)\gamma_0,\ \gamma_0 = (-0.4\ \ -0.2)',$$

$$Y_2(0,0) = 0.1X + 0.5Z_1 + 0.1XZ_1 + 0.1Z_1^2 + u_{y2},\ Y_2(0,1) = Y_2(0,0) + \theta_{0,2}^0,$$

$$Y_2(1,0) = Y_2(0,0) + \theta_{1,1}, \quad Y_2(1,1) = Y_2(0,0) + \theta_{1,1} + \theta_{0,2}^1.$$

Primary treatment effects and first-stage effects vary across DGPs.

DGP 1: $\theta_{0,1} = 0.5$, $\theta_{0,2}^0 = 0.5$, $\theta_{0,2}^1 = 0.5$, $\theta_{1,1} = 0.2$.

DGP 2: $\theta_{0,1} = 0.5$, $\theta_{0,2}^0 = 0.5$, $\theta_{0,2}^1 = 0.1$, $\theta_{1,1} = 0.2$.

DGP 3: $\theta_{0,1} = 0.5$, $\theta_{0,2}^0 = \theta_{0,2}^1 = 0.5 + e$, $\theta_{1,1} = 0.2 + e$, $e \sim U[-0.5, 0.5]$.

DGP 4: $\theta_{0,1} = 0.5$, $\theta_{0,2}^0 = 0.5 + 0.5u_{s2}$, $\theta_{0,2}^1 = 0.5$, $\theta_{1,1} = 0.2$.

DGP 5: $\theta_{0,1} = 0.5$, $\theta_{0,2}^0 = \theta_{0,2}^1 = 0.5 + 0.5u_{s2} + 0.5v_{z2}$, $\theta_{1,1} = 0.2 + 0.5u_{s2} + 0.5v_{z2}$.

Given the above potential random variables, observed random variables Y_1, S_2, Z_2, D_2, and Y_2 are defined following the potential outcome framework in Section 3. For each DGP, we carry out 1,000 simulations and estimate both the proposed and the recursive CFR immediate and one-period-after ATEs. Standard errors are calculated using weighted bootstrap discussed in Section 4.2. Bandwidth is chosen following $h = h_{CCT} \times n^{1/5-1/k}$, where h_{CCT} is the CCT bandwidth for classic RD estimation of $E[\tilde{\theta}_{1,1}|Z_1 = 0]$, and $k < 5$ is an under-smoothing parameter. Simulation codes are written using R. The CCT bandwidth is calculated using R package "rdrobust". In the simulation and empirical sections, we report estimation and inference results with different k choices to examine the robustness of proposed estimators with respect to bandwidth choice. Specifically, we set $k = 4.25$, 4.5, and 4.75.

Table 1 reports the mean and the mean squared error (MSE) of both the proposed and the recursive CFR one-period-after ATE estimators. As is predicted by the theory, both estimators average around the true value in DGPs 1 and 3. The proposed estimator has larger MSEs due to first-stage local likelihood estimation. Under DGPs 2 and 4, the recursive CFR estimator does not center around the true value 0.2, while the proposed estimator still performs well. Under DGP 5, neither estimators have correct centering.

Table 1: One-period-after ATE: Proposed Estimator Vs. Recursive CFR Estimator

	Proposed Estimator						Recursive CFR Estimator					
	Mean			MSE			Mean			MSE		
k	4.25	4.5	4.75	4.25	4.5	4.75	4.25	4.5	4.75	4.25	4.5	4.75
	DGP 1											
n=2000	0.198	0.199	0.200	0.144	0.140	0.138	0.210	0.210	0.210	0.110	0.107	0.106
n=4000	0.200	0.201	0.202	0.100	0.097	0.096	0.203	0.204	0.204	0.080	0.078	0.077
n=8000	0.205	0.205	0.206	0.073	0.071	0.070	0.210	0.211	0.211	0.058	0.056	0.055
	DGP 2											
n=2000	0.197	0.199	0.200	0.144	0.140	0.139	0.142	0.143	0.143	0.116	0.113	0.111
n=4000	0.203	0.205	0.206	0.101	0.098	0.098	0.136	0.138	0.138	0.082	0.080	0.080
n=8000	0.205	0.206	0.207	0.074	0.072	0.071	0.137	0.138	0.139	0.060	0.058	0.057
	DGP 3											
n=2000	0.198	0.200	0.202	0.155	0.151	0.149	0.208	0.210	0.211	0.128	0.125	0.123
n=4000	0.205	0.206	0.207	0.102	0.100	0.098	0.207	0.209	0.210	0.086	0.084	0.083
n=8000	0.205	0.205	0.206	0.072	0.070	0.070	0.208	0.209	0.209	0.062	0.060	0.060
	DGP 4											
n=2000	0.203	0.205	0.206	0.150	0.146	0.144	0.124	0.125	0.126	0.125	0.121	0.119
n=4000	0.197	0.199	0.200	0.104	0.101	0.100	0.120	0.121	0.122	0.087	0.085	0.084
n=8000	0.202	0.203	0.203	0.075	0.072	0.072	0.121	0.122	0.123	0.062	0.060	0.059
	DGP 5											
n=2000	-0.044	-0.043	-0.042	0.175	0.171	0.169	0.098	0.098	0.097	0.219	0.214	0.211
n=4000	-0.044	-0.044	-0.043	0.127	0.122	0.121	0.098	0.099	0.100	0.163	0.158	0.156
n=8000	-0.050	-0.049	-0.048	0.093	0.090	0.088	0.096	0.098	0.099	0.119	0.115	0.114

Note: All Monte Carlo experiments use 1,000 simulation repetitions and weighted bootstrap with 1,000 bootstrap repetitions.

Table 2 reports the proportion of rejections in 5% two-sided t-tests associated with proposed immediate and one-period-after ATE estimators. The left half of the table shows the size of the tests with the true value stated under the null. The right half of the table shows the power of the tests with the null set incorrectly to 0.3 for the immediate ATE and 0 for the one-period-after ATE. It is clear from the table that under DGPs 1-4 t-tests control size well under the null and have power going to one under the alternative. Under DGP 5, the tests do not control size as the DGP violates the identifying condition in Assumption 3.1. The choice of undersmoothing parameter k does not seem to affect simulation results much under the five DGPs considered in this section.

Table 2: Two-sided T-tests with Proposed Immediate and One-period-after ATE Estimators

	Size						Power					
	Immediate ATE			One-period-after ATE			Immediate ATE			One-period-after ATE		
k	4.25	4.5	4.75	4.25	4.5	4.75	4.25	4.5	4.75	4.25	4.5	4.75
	DGP 1											
n=2000	0.048	0.047	0.046	0.070	0.066	0.066	0.596	0.620	0.629	0.346	0.357	0.366
n=4000	0.050	0.043	0.040	0.054	0.052	0.054	0.837	0.867	0.879	0.542	0.563	0.574
n=8000	0.046	0.052	0.048	0.056	0.056	0.056	0.980	0.986	0.988	0.811	0.823	0.834
	DGP 2											
n=2000	0.061	0.065	0.060	0.068	0.074	0.072	0.613	0.633	0.642	0.344	0.361	0.365
n=4000	0.048	0.042	0.042	0.056	0.056	0.058	0.865	0.876	0.883	0.550	0.582	0.593
n=8000	0.051	0.057	0.059	0.064	0.064	0.068	0.978	0.984	0.990	0.814	0.835	0.847
	DGP 3											
n=2000	0.057	0.052	0.052	0.079	0.080	0.076	0.573	0.612	0.630	0.329	0.343	0.348
n=4000	0.051	0.054	0.054	0.058	0.058	0.054	0.842	0.868	0.878	0.531	0.565	0.572
n=8000	0.058	0.059	0.064	0.048	0.052	0.053	0.984	0.989	0.991	0.806	0.825	0.833
	DGP 4											
n=2000	0.049	0.049	0.048	0.074	0.082	0.076	0.572	0.594	0.613	0.366	0.377	0.381
n=4000	0.052	0.048	0.054	0.064	0.061	0.057	0.856	0.882	0.885	0.554	0.587	0.599
n=8000	0.053	0.055	0.057	0.062	0.063	0.067	0.975	0.984	0.986	0.796	0.823	0.837
	DGP 5											
n=2000	0.052	0.052	0.055	0.300	0.323	0.329	0.596	0.620	0.636	0.058	0.064	0.065
n=4000	0.059	0.054	0.053	0.513	0.539	0.548	0.879	0.897	0.911	0.072	0.065	0.066
n=8000	0.071	0.069	0.070	0.808	0.825	0.837	0.982	0.986	0.988	0.088	0.091	0.087

Note: All Monte Carlo experiments use 1,000 simulation repetitions and weighted bootstrap with 1,000 bootstrap repetitions. The true value of the estimated parameter is 0.2. All t-tests use the 5% significance level.

6 Empirical Example: The Effect of CA School Bonds

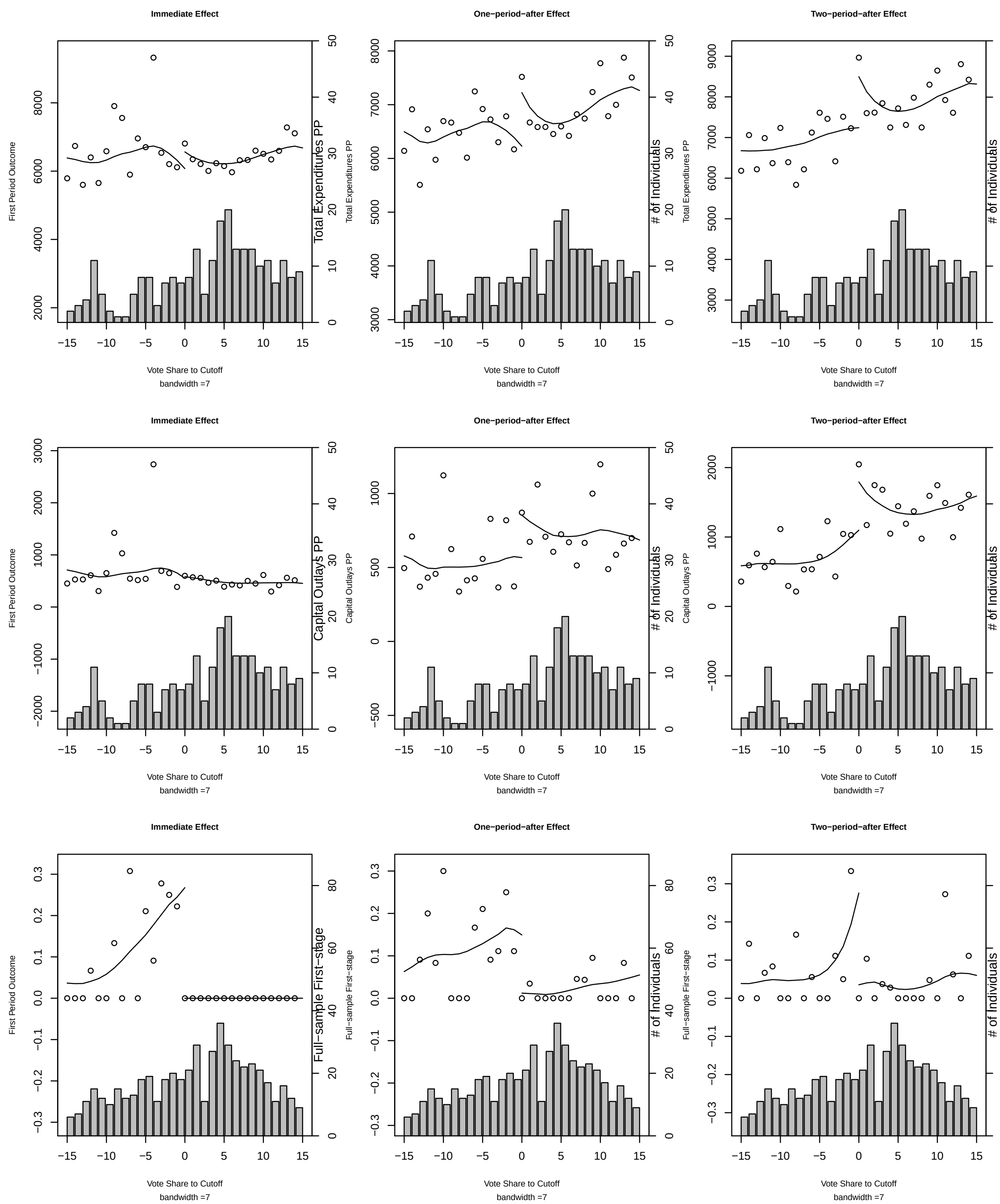
This section revisits the study of local education bonds using the dataset published by CFR. As described in CFR, school districts in California became eligible for issuing general obligation bonds through Proposition 46 in 1984. CFR studies effects of bond authorization on local house prices, student achievements, and other outcomes using Californian data from 1987 to 2005. Due to data limitations, we only study two outcome variables: total expenditure per pupil and capital loading per pupil in a school district.

The data structure we constructed for this analysis is different from that in CFR. Recall that CFR assumed path-independence of treatment effects and are free to pool various bond measures of the same school districts. In contrast, our proposed identification method allows treatment effects to be path-dependent. Therefore, we only use each school district's first education bond measure as the focal RD experiment. There are 614 school districts in the dataset. Restricting the sample to those with non-missing vote share data and the first bond measure in or before 2002, we obtain a subsample of 596 school districts. Only 282 of them seeking voter approval of the first bond measure in or after 1995 have non-missing outcome data.

Among the 596 local school districts, 74 followed up with a second bond voting in the year after the initial trial. Among these 74 districts, 73 did not succeed in the initial trial, and one succeeded by a substantial margin. This data pattern could be seen in the bottom row of Figure 1. The three first-stage graphs show the effect of authorizing an education bond in year one on the probability of authorizing another education bond in years two, three, and four, correspondingly. No school districts that barely passed the vote share cutoff at the first trial authorize another bond next year. The probabilities increase slightly in years three and four. On the other hand, around 27% of school districts that barely missed the vote share cutoff at the first trial successfully authorize their first education bond in the next year.

Figure 1 also shows the average total effects and the histogram of the first-round running variable. Consistent with discussions in CFR, education bonds have lagged effects on the total expenditure and capital loading of local school districts. Combining the information in all three rows of Figure 1, we know that the longer-term AITTEs shall be smaller than longer-term ATEs effects in this data application since receiving the first-round treatment decreases the probability of receiving treatments in future rounds.

Figure 1: Average Total Effects And Histograms of the First-round Running Variable



Note: If we calculate the CCT bandwidth for each individual figure, the average is equal to 6.23 for the three ITT effect regressions on total expenditure per pupil, 6.16 for the three ITT effect regressions on capital loading, and 8.66 for the three first-stage ITT effect regressions. To facilitate interpretation of the figures, we take an average and use bandwidth 7 across all nine graphs so that results are comparable across graphs. As stated in the paper, the sample size is 596 for the first-stage analysis and 282 for the two outcomes due to missing data.

Table 3 reports nonparametric local linear estimation results of ATEs and average primary first-stage effects following the recursive CFR identification strategy. Three-period-after average primary first-stage effects are not reported because the identification involves treatment decisions of the fifth period, which is missing for some school districts in our subsample. The estimates reported in the table are somewhat larger than the parametric recursive estimates reported in Table 4 of Cellini et al. (2010). This difference could be due to the difference between parametric vs. local linear RD estimation strategies or the fact that we only use the first school bond measure of each school district as the focal treatment. Inference of both recursive CFR and proposed two-step estimators is carried out with weighted bootstrap.

Table 3: ATEs and Average Primary First-stage Effects – Nonparametric CFR

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 expenditure per pupil, 282 school districts							
775	685	$1,391^{*}$	$1,313^{*}$	$2,027$	$1,990$	$3,616^{**}$	$3,702^{**}$
(604)	(545)	(828)	(779)	(1,301)	(1,232)	(1,792)	(1,708)
Capital loading per pupil, 282 school districts							
279	200	459^{**}	405^{**}	981^{***}	932^{***}	$2,400^{**}$	$2,407^{***}$
(215)	(186)	(190)	(175)	(538)	(472)	(967)	(854)
First-stage, 596 school districts							
-0.268^{***}	-0.265^{***}	-0.208^{***}	-0.214^{***}	-0.335^{***}	-0.323^{***}	-	-
(0.083)	(0.081)	(0.076)	(0.076)	(0.083)	(0.082)	-	-

Note: Weighted bootstrap confidence intervals are calculated with 1,000 bootstrap repetitions. Undersmoothed CCT bandwidth is calculated following suggestions in Section 5, with CCT bandwidth set to 6.23 for the total expenditure outcome, 6.16 for the capital loading outcome, and 8.66 for the full-sample first-stage analysis. These numbers are discussed in the footnote of Figure 1 as well.

Table 4 reports estimates of ATEs and average primary first-stage effects following the proposed procedure. Section (1) of the table reports estimation and inference results when the propensity score functions defined in Section 3.1 only use the constant as the conditioning covariate X. Section (2) of the table reports results when the propensity score function uses $X = (1, Y_1(d_1))$. Regression estimates of the two sections share the same pattern. The longer-term ATE estimates for the total expenditure outcome are smaller in this table, comparing to those reported in Table 3. Meanwhile, the longer-term ATE estimates for the capital loading outcome are larger in this table than those reported in Table 3. However, due to the small sample size of this data application, we cannot find any statistically significant difference between the estimates reported in the two tables.

Table 4: Average Primary Treatment Effects – Proposed CIA Estimator

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
Section (1):							
Total expenditures per pupil, 282 school districts							
775	685	963	915	1,273	1,203	2,333	2,338
(604)	(545)	(733)	(685)	(1,314)	(1,198)	(2,303)	(1,947)
Capital outlays per pupil, 282 school districts							
279	200	399***	377***	1,242***	1,218***	2,884***	2,932***
(215)	(186)	(153)	(144)	(418)	(382)	(791)	(740)
First-stage, 596 school districts							
-0.268***	-0.265***	-0.215*	-0.233**	-0.421***	-0.426***	-	-
(0.083)	(0.081)	(0.113)	(0.106)	(0.119)	(0.107)	-	-
Section (2):							
Total expenditures per pupil, 282 school districts							
775	685	811	779	867	828	1,508	1,571
(604)	(545)	(899)	(854)	(1,897)	(1,878)	(3,629)	(3,769)
Capital outlays per pupil, 282 school districts							
279	200	418**	389**	1,368***	1,327***	3,196**	3,229**
(215)	(186)	(172)	(161)	(487)	(478)	(1,282)	(1,341)
First-stage, 596 school districts							
-0.268***	-0.265***	-0.136**	-0.144**	-0.243***	-0.228***	-	-
(0.083)	(0.081)	(0.069)	(0.070)	(0.088)	(0.087)	-	-

Note: Weighted bootstrap confidence intervals are calculated with 1000 bootstrap repetitions. Undersmoothed CCT bandwidth is calculated following suggestions in Section 5, with CCT bandwidth set to 6.23 for the total expenditure outcome, 6.16 for the capital loading outcome, and 8.66 for the full-sample first-stage analysis. These numbers are discussed in the footnote of Figure 1 as well. Section (1) report estimation results with constant only covariate X in the potential propensity score functions defined in Section 3.1. Section (2) report results with $X = (1, Y_1(d_1))$ in the potential propensity score functions.

7 Conclusion

Static RD models with a single eligibility test have been very popular in the last two decades. Recently, more empirical studies are interested in situations where each individual could potentially participate in multiple RD experiments and hence receive multiple treatments over a period of time. Most papers in the literature, however, either ignore dynamics in such RD models or employ restrictive identifying assumptions. This paper is the first to employ the conventional potential outcome framework to formulate a general dynamic RD model. We propose several sets of identifying assumptions that are much weaker than those used in the literature. A novel two-step semiparametric estimation procedure and a weighted bootstrap inference method are proposed for estimation and inference of longer-term average primary treatment effects that prohibit reception of treatments after the focal round of RD. The proposed estimation and inference strategy is adopted to revisit the study of local education bonds following CFR. For future research, it will be interesting to extend the local MLE procedure adopted in the proposed two-step estimation procedure to other traditional static RD models for the analysis of first-stage and treatment effect heterogeneity.

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. and M. Rokkanen (2015): "Wanna Get Away? Regression Discontinuity Estimation of Exam School Effects Away from the Cutoff," *Journal of the American Statistical Association*, 110(512), 1331–1344.

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

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 (forthcoming): "Difference-in-differences with multiple time periods," *Journal of Econometrics*.

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

Caughey, D. and J. S. Sekhon (2011): "Elections and the regression discontinuity design: Lessons from close US house races, 1942-2008," *Political Analysis*, 19, 385–408.

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*, 125, 215–261.

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.

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. and S. Shen (2018): "Testing for rank invariance or similarity in program evaluation," *Review of Economics and Statistics*, 100, 78–85.

Dube, A., L. Giuliano, and J. Leonard (2019): “Fairness and Frictions: The Impact of Unequal Raises on 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.

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

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.

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.

Hsu, Y.-C. and S. Shen (2019): “Testing Treatment Effect Heterogeneity in Regression Discontinuity Designs,” *Journal of Econometrics*, 208, 468–486.

——— (2021): “Testing Monotonicity of Conditional Treatment Effects under Regression Discontinuity Design,” *Journal of Applied Econometrics*, 36, 346–366.

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.

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.

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

Sun, L. and S. Abraham (forthcoming): "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects," *Journal of Econometrics*.

ONLINE APPENDIX

A Proofs for Identification Results

Proof of Lemma 2.1

First consider pairs of potential outcomes with only one flipped treatment status. Denote the difference by $Y_{k+\tau}((\ell^{k-1},1,\eta))-Y_{k+\tau}((\ell^{k-1},0,\eta))$, where $\tau=1,...,T-k$, $k=1,...,T$, $\ell\in\mathcal{L}^{k-1}$ and $\eta\in\mathcal{L}^{\tau}$. If all elements of η are zero, the above difference in potential outcomes is defined in (2.3). If all but the s-th element of η are zero, then the difference

$$
\begin{aligned}
&Y_{k+\tau}((\ell^{k-1},1,\eta))-Y_{k+\tau}((\ell^{k-1},0,\eta))\\
=&Y_{k+\tau}((\ell^{k-1},1,\mathbf{0}))-Y_{k+\tau}((\ell^{k-1},0,\mathbf{0}))\\
&+Y_{k+\tau}((\ell^{k-1},1,\eta))-Y_{k+\tau}((\ell^{k-1},1,\mathbf{0}))-\Big(Y_{k+\tau}((\ell^{k-1},0,\eta))-Y_{k+\tau}((\ell^{k-1},0,\mathbf{0}))\Big)\\
=&\theta^{\ell_{k-1}}_{\tau,k}+\theta^{(\ell^{k-1},1,\mathbf{0}_{s-1})}_{\tau-s,k+s}-\theta^{(\ell^{k-1},\mathbf{0}_{s})}_{\tau-s,k+s}.
\end{aligned}
\tag{A.1}
$$

is a linear combination of the effects defined in (2.3), for all $s=1,...,T-k$. Let η' be the vector where $\eta'_s=1$ and all other elements are zero. If all but the s-th and s'-th elements of η are zero, $s<s'$, then

$$
\begin{aligned}
&Y_{k+\tau}((\ell^{k-1},1,\eta))-Y_{k+\tau}((\ell^{k-1},0,\eta))\\
=&Y_{k+\tau}((\ell^{k-1},1,\eta'))-Y_{k+\tau}((\ell^{k-1},0,\eta'))+Y_{k+\tau}((\ell^{k-1},1,\eta))-Y_{k+\tau}((\ell^{k-1},1,\eta'))\\
&-\Big(Y_{k+\tau}((\ell^{k-1},0,\eta))-Y_{k+\tau}((\ell^{k-1},0,\eta'))\Big),
\end{aligned}
$$

where the first difference is between a pair of outcomes discussed in (A.1) and the other two differences are between pairs of outcomes defined in (2.3). Similarly, the difference $Y_{k+\tau}((\ell^{k-1},1,\eta))-Y_{k+\tau}((\ell^{k-1},0,\eta))$ with three or more non-zero elements in η could all be represented by linear cominbations of the primary individual treatment effects defined in equation (2.3).

Now consider pairs of potential outcomes with two flipped treatments. It is easy to see that such differences, for example, $Y_{k+\tau}((\ell^{k-1},1,\eta,1,\rho))-Y_{k+\tau}((\ell^{k-1},0,\eta,0,\rho))$, where $\tau=2,...,T-k$, $k=1,...,T$, $\ell\in\mathcal{L}^{k-1}$, and $(1,\eta,1,\rho)\in\mathcal{L}^{\tau}$ could be represented by a linear cominbation of differences of potential outcomes with only one flipped treatment status, which has been discussed above, and therefore eventually be represented by linear cominbations of the primary individual treatment effects defined in equation (2.3). Similarly, the difference of potential outcome with three or more flipped treatment status could be defined by a linear cominbation of primary individual treatment effects defined in (2.3). This completes the proof.

Proof of Lemma 2.2

We prove the result by induction. First, it has been shown in the main paper that equation (2.6) holds for the first two time periods.

Suppose the result of the Lemma holds for any period $s = 3, ..., T$. This implies that

$$\begin{aligned}
&\lim_{z_1 \searrow 0} E[Y_s|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[Y_s|Z_1 = z_1] \\
=&E\left[\sum_{\ell^{s-1} \in \mathcal{L}^{s-1}} \left(Y_s(1, \ell^{s-1}) \cdot \mathfrak{D}_{2:s}(1, \ell^{s-1}) - Y_t(0, \ell^{s-1}) \cdot \mathfrak{D}_{2:s}(0, \ell^{s-1})\right) |Z_1 = 0\right] \\
=&\sum_{\tau=0}^{s-1} \theta_\tau \cdot \pi_{s-1-\tau},
\end{aligned} \tag{A.2}$$

where the first equality holds by the smoothness conditions in Assumption 2.1.

Now for period $s + 1$, under Assumption 2.1,

$$\begin{aligned}
&\lim_{z_1 \searrow 0} E[Y_{s+1}|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[Y_{s+1}|Z_1 = z_1] \\
=&E\left[\sum_{\ell^{s-1} \in \mathcal{L}^{s-1}} Y_{s+1}(1, \ell^{s-1}, 1)\mathfrak{D}_{2:s}(1, \ell^{s-1})D_{s+1}(1, \ell^{s-1}) + Y_{s+1}(1, \ell^{s-1}, 0)\mathfrak{D}_{2:s}(1, \ell^{s-1})(1 - D_{s+1}(1, \ell^{s-1}))|Z_1 = 0\right] \\
&- E\left[\sum_{\ell^{s-1} \in \mathcal{L}^{s-1}} Y_{s+1}(0, \ell^{s-1}, 1)\mathfrak{D}_{2:s}(0, \ell^{s-1})D_{s+1}(0, \ell^{s-1}) + Y_{s+1}(0, \ell^{s-1}, 0)\mathfrak{D}_{2:s}(0, \ell^{s-1})(1 - D_{s+1}(0, \ell^{s-1}))|Z_1 = 0\right] \\
=&E\left[\sum_{\ell^{s-1} \in \mathcal{L}^{s-1}} \left(Y_{s+1}(1, \ell^{s-1}, 0)\mathfrak{D}_{2:s}(1, \ell^{s-1}) - Y_{t+1}(0, \ell^{s-1}, 0)\mathfrak{D}_{2:s}(\mathbf{0}, \ell^{s-1})\right) |Z_1 = 0\right] \\
&+ E\left[\sum_{\ell^{s-1} \in \mathcal{L}^{s-1}} \left(Y_{s+1}(1, \ell^{s-1}, 1) - Y_{s+1}(1, \ell^{s-1}, 0)\right) \mathfrak{D}_{2:s}(1, \ell^{s-1})D_{s+1}(1, \ell^{s-1})|Z_1 = 0\right] \\
&- E\left[\sum_{\ell^{s-1} \in \mathcal{L}^{s-1}} \left(Y_{s+1}(0, \ell^{s-1}, 1) - Y_{s+1}(0, \ell^{s-1}, 0)\right) \mathfrak{D}_{2:s}(0, \ell^{s-1})D_{s+1}(0, \ell^{s-1})|Z_1 = 0\right] \\
\equiv&A + \theta_0 \cdot E\left[\sum_{\ell^{s-1} \in \mathcal{L}^{s-1}} \left(\mathfrak{D}_{2:s}(1, \ell^{s-1})D_{s+1}(1, \ell^{s-1}) - \mathfrak{D}_{2:s}(0, \ell^{s-1})D_{s+1}(0, \ell^{s-1})\right) |Z_1 = 0\right] \\
\equiv&A + \theta_0 \cdot \left(\lim_{z_1 \searrow 0} E\left[\sum_{\ell^{s-1} \in \mathcal{L}^{s-1}} D_{s+1}(1, \ell^{s-1}) \cdot \mathfrak{D}(1, \ell^{s-1})|Z_1 = z_1\right] - \lim_{z_1 \nearrow 0} E\left[\sum_{\ell^{s-1} \in \mathcal{L}^{s-1}} D_{s+1}(0, \ell^{s-1}) \cdot \mathfrak{D}(0, \ell^{s-1})|Z_1 = z_1\right]\right) \\
=&A + \theta_0 \cdot \pi_s.
\end{aligned} \tag{A.3}$$

Now note that the only difference between the A term above and the conditional mean expression

$$E\left[\sum_{\ell^{s-1}\in\mathcal{L}^{s-1}}\left(Y_s(1,\ell^{s-1})\cdot\mathfrak{D}_{2:s}(1,\ell^{s-1})-Y_t(0,\ell^{s-1})\cdot\mathfrak{D}_{2:s}(\mathbf{0},\ell^{s-1})\right)|Z_1=0\right]$$

in equation (A.2) is between potential outcomes $Y_{s+1}(1,\ell^{s-1},0)$ and $Y_s(1,\ell^{s-1})$. Since θ_τ for $\tau=0,1,2...$ are primary treatment effects that prohibit additional treatments after the focal round, it is clear that

$$A=\sum_{\tau=0}^{s-1}\theta_{\tau+1}\cdot\pi_{s-1-\tau}$$

Plugging the result into equation (A.3) therefore completes the proof.

Proof of Equation (2.7)

First, we state the updated smoothness conditions that condition on the k-th round running variable.

Assumption A.1 *For the fixed $k=1,...,K-2$ considered in the paper, there exists an $\epsilon>0$, such that*

1. *Z_{k+1} is continuous in $z_{k+1}\in\mathcal{N}_\epsilon$ with $P[Z_{k+1}\geq 0]\in(0,1)$;*
2. *for $\tau=0,1,...,(T-k-1)$, $E[Y_{\tau+k+1}(\ell^{\tau+k+1})|\mathfrak{D}(\ell^{\tau+k+1})=1,Z_{k+1}=z_{k+1}]$ is continuous in $z_{k+1}\in\mathcal{N}_\epsilon$ for all $\ell^{\tau+k+1}\in\mathcal{L}^{\tau+k+1}$.*
3. *$P[\mathfrak{D}(\ell^T)=1|Z_{k+1}=z_{k+1}]$ is continuous in $z_{k+1}\in\mathcal{N}_\epsilon$ for all $\ell^T\in\mathcal{L}^T$.*

Under Assumption A.1,

$$\begin{aligned}
&\lim_{z_{k+1}\searrow 0}E[Y_{k+1}|Z_{k+1}=z_{k+1}]-\lim_{z_{k+1}\nearrow 0}E[Y_{k+1}|Z_{k+1}=z_{k+1}]\\
=&E\left[\sum_{\ell^k\in\mathcal{L}^k}Y_{k+1}(\ell^k,1)\cdot\mathfrak{D}(\ell^k)\middle|Z_{k+1}=0\right]-E\left[\sum_{\ell^k\in\mathcal{L}^k}Y_{k+1}(\ell^k,0)\cdot\mathfrak{D}(\ell^k)\middle|Z_{k+1}=0\right]\\
=&\theta_0\cdot E\left[\sum_{\ell^k\in\mathcal{L}^k}\mathfrak{D}(\ell^k)\middle|Z_{k+1}=0\right]\\
=&\theta_0,
\end{aligned}$$

and

$$
\begin{aligned}
&\lim_{z_{k+1}\searrow 0} E[Y_{k+2}|Z_{k+1}=z_{k+1}] - \lim_{z_{k+1}\nearrow 0} E[Y_{k+2}|Z_{k+1}=z_{k+1}] \\
=&\sum_{\ell^k\in\mathcal{L}^k} E[Y_{k+2}(\ell^k,1,1)\mathfrak{D}(\ell^k)D_{k+2}(\ell^k,1)+Y_{k+2}(\ell^k,1,0)\mathfrak{D}(\ell^k)(1-D_{k+2}(\ell^k,1))|Z_{k+1}=0] \\
&-\sum_{\ell^k\in\mathcal{L}^k} E[Y_{k+2}(\ell^k,0,1)\mathfrak{D}(\ell^k)D_{k+2}(\ell^k,0)+Y_{k+2}(\ell^k,0,0)\mathfrak{D}(\ell^k)(1-D_{k+2}(\ell^k,0))|Z_{k+1}=0] \\
=&\sum_{\ell^k\in\mathcal{L}^k} E[(Y_{k+2}(\ell^k,1,0)\mathfrak{D}(\ell^k)|Z_{k+1}=0] - \sum_{\ell^k\in\mathcal{L}^k} E[Y_{k+2}(\ell^k,0,0)\mathfrak{D}(\ell^k)|Z_{k+1}=0] \\
&+\sum_{\ell^k\in\mathcal{L}^k} E\left[\Big(Y_{k+2}(\ell^k,1,1)-Y_{k+2}(\ell^k,1,0)\Big)\mathfrak{D}(\ell^k)D_{k+2}(\ell^k,1)|Z_{k+1}=0\right] \\
&-\sum_{\ell^k\in\mathcal{L}^k} E\left[\Big(Y_{k+2}(\ell^k,0,1)-Y_{k+2}(\ell^k,0,0)\Big)\mathfrak{D}(\ell^k)D_{k+2}(\ell^k,0)|Z_{k+1}=0\right] \\
=&E\left[\theta_1\cdot\sum_{\ell^k\in\mathcal{L}^k}\mathfrak{D}(\ell^k)\big|Z_{k+1}=0\right] + E\left[\theta_0\cdot\sum_{\ell^k\in\mathcal{L}^k}\mathfrak{D}(\ell^k)D_{k+2}(\ell^k,1)\big|Z_{k+1}=0\right] \\
&-E\left[\theta_0\cdot\sum_{\ell^k\in\mathcal{L}^k}\mathfrak{D}(\ell^k)D_{k+2}(\ell^k,0)\big|Z_{k+1}=0\right] \\
=&\theta_1+\theta_0\cdot\left(\lim_{z_{k+1}\searrow 0} E[D_{k+2}|Z_{k+1}=z_{k+1}] - \lim_{z_{k+1}\nearrow 0} E[D_{k+2}|Z_{k+1}=z_{k+1}]\right).
\end{aligned}
$$

The general result for decomposing $\lim_{z_{k+1}\searrow 0} E[Y_{k+s}|Z_{k+1}=z_{k+1}] - \lim_{z_{k+1}\nearrow 0} E[Y_{k+s}|Z_{k+1}=z_{k+1}]$ with any $s\geq 2$ could then be proven by induction just as in the proof of Lemma 2.2.

Proof of Equation (3.1)

When $\tau=1$, equation (3.1) is clear. When $\tau=2,3,...,T-1$, first we notice that for $d_1=0,1$, the quasi-potential outcome $\tilde{Y}_{\tau+1}(d_1)$ could be decomposed, and

$$
\begin{aligned}
\tilde{Y}_{\tau+1}(d_1)=&\tilde{Y}_{\tau+1}(d_1,0)\cdot(1-D_2(d_1))+\tilde{Y}_{\tau+1}(d_1,1)\cdot D_2(d_1) \\
=&\tilde{Y}_{\tau+1}(d_1,0)+\tilde{\theta}^{d_1}_{\tau-1,2}\cdot D_2(d_1) \\
=&\tilde{Y}_{\tau+1}(d_1,0,0)\cdot(1-D_3(d_1,0))+\tilde{Y}_{\tau+1}(d_1,0,1)\cdot D_3(d_1,0)+\tilde{\theta}^{d_1}_{\tau-1,2}\cdot D_2(d_1) \\
=&\tilde{Y}_{\tau+1}(d_1,\mathbf{0}_2)+\tilde{\theta}^{(d_1,0)}_{\tau-2,3}\cdot D_3(d_1,0)+\tilde{\theta}^{d_1}_{\tau-1,2}\cdot D_2(d_1) \\
=&\tilde{Y}_{\tau+1}(d_1,\mathbf{0}_3)\cdot(1-D_4(d_1,\mathbf{0}_2))+\tilde{Y}_{\tau+1}(d_1,\mathbf{0}_2,1)\cdot D_4(d_1,\mathbf{0}_2) \\
&+\tilde{\theta}^{(d_1,0)}_{\tau-2,3}\cdot D_3(d_1,0)+\tilde{\theta}^{d_1}_{\tau-1,2}\cdot D_2(d_1)
\end{aligned}
$$

$$
\begin{aligned}
&=\tilde{Y}_{\tau+1}(d_1,\mathbf{0}_3)+\tilde{\theta}^{(d_1,\mathbf{0}_2)}_{\tau-3,4}\cdot D_4(d_1,\mathbf{0}_2)+\tilde{\theta}^{(d_1,0)}_{\tau-2,3}\cdot D_3(d_1,0)+\tilde{\theta}^{d_1}_{\tau-1,2}\cdot D_2(d_1)\\
&=\ldots\\
&=Y_{\tau+1}(d_1,\mathbf{0}_\tau)+\tilde{\theta}^{d_1}_{\tau-1,2}\cdot D_2(d_1)+\sum_{s=0}^{\tau-2}\tilde{\theta}^{(d_1,\mathbf{0}_{\tau-1-s})}_{s,\tau+1-s}\cdot D_{\tau+1-s}(d_1,\mathbf{0}_{\tau-1-s}).
\end{aligned}
$$

Then, under smoothness conditions in Assumption 2.1,

$$
\begin{aligned}
&\lim_{z_1\searrow 0}E[Y_{\tau+1}|Z_1=z_1]-\lim_{z_1\nearrow 0}E[Y_{\tau+1}|Z_1=z_1]\\
=&E[\tilde{Y}_{\tau+1}(1)|Z_1=0]-E[\tilde{Y}_{\tau+1}(0)|Z_1=0]\\
=&E\left[Y_{\tau+1}(1,\mathbf{0}_\tau)+\tilde{\theta}^{1}_{\tau-1,2}\cdot D_2(1)+\sum_{s=0}^{\tau-2}\tilde{\theta}^{(1,\mathbf{0}_{\tau-1-s})}_{s,\tau+1-s}\cdot D_{\tau+1-s}(1,\mathbf{0}_{\tau-1-s})\big|Z_1=0\right]\\
&-E\left[Y_{\tau+1}(0,\mathbf{0}_\tau)+\tilde{\theta}^{0}_{\tau-1,2}\cdot D_2(0)+\sum_{s=0}^{\tau-2}\tilde{\theta}^{(0,\mathbf{0}_{\tau-1-s})}_{s,\tau+1-s}\cdot D_{\tau+1-s}(0,\mathbf{0}_{\tau-1-s})\big|Z_1=0\right]\\
=&E[\theta_{\tau,1}|Z_1=0]+E\left[\tilde{\theta}^{1}_{\tau-1,2}|D_2(1)=1,Z_1=0\right]\cdot P[D_2(1)=1|Z_1=0]\\
&-E\left[\tilde{\theta}^{0}_{\tau-1,2}|D_2(0)=1,Z_1=0\right]\cdot P[D_2(0)=1|Z_1=0]\\
&+\sum_{s=0}^{\tau-2}E\left[\tilde{\theta}^{(1,\mathbf{0}_{\tau-1-s})}_{s,\tau+1-s}|D_{\tau+1-s}(1,\mathbf{0}_{\tau-1-s})=1,Z_1=0\right]P\left[D_{\tau+1-s}(1,\mathbf{0}_{\tau-1-s})=1|Z_1=0\right]\\
&-\sum_{s=0}^{\tau-2}E\left[\tilde{\theta}^{(0,\mathbf{0}_{\tau-1-s})}_{s,\tau+1-s}|D_{\tau+1-s}(0,\mathbf{0}_{\tau-1-s})=1,Z_1=0\right]P\left[D_{\tau+1-s}(0,\mathbf{0}_{\tau-1-s})=1|Z_1=0\right].
\end{aligned}
$$

Proof of Lemma 3.1

By the definition of potential propensity scores, we see that under Assumption 3.2,

$$
\begin{aligned}
&E[Y_2(0,1)|X,S_2(0)=1,Z_2(0)\geq 0,Z_1=0]\\
=&\lim_{z_1\nearrow 0}E[Y_2(0,1)|X,S_2(0)=1,Z_2(0)\geq 0,Z_1=z_1]\\
=&\lim_{z_1\nearrow 0}E[Y_2(0,1)1(Z_2(0)\geq 0)|X,S_2(0)=1,Z_1=z_1]/P[Z_2(0)\geq 0|X,S_2(0)=1,Z_1=z_1]\\
=&\lim_{z_1\nearrow 0}E[Y_2D_2/\lambda^0(X)|X,S_2=1,Z_1=z_1].
\end{aligned}
$$

Together with Assumption 3.1, we also have that,

$$
\begin{aligned}
&E[Y_2(0,0)|X,S_2(0)=1,Z_2(0)\geq 0,Z_1=0]=E[Y_2(0,0)|X,S_2(0)=1,Z_2(0)<0,Z_1=0]\\
=&\lim_{z_1\nearrow 0}E[Y_2(0,0)(Z_2(0)<0)|X,S_2(0)=1,Z_1=z_1]/P[Z_2(0)<0|X,S_2(0)=1,Z_1=z_1]\\
=&\lim_{z_1\nearrow 0}E[Y_2(1-D_2)/(1-\lambda^0(X))|X,S_2=1,Z_1=z_1].
\end{aligned}
$$

The results above lead to the identification of $E[\theta^0_{0,2}|D_2(0)=1, Z_1=0]$, since

$$
\begin{aligned}
&E[\theta^0_{0,2}|D_2(0)=1, Z_1=0] = E[Y_2(0,1)-Y_2(0,0)|S_2(0)=1, Z_2(0)\geq 0, Z_1=0] \\
&= \lim_{z_1 \nearrow 0} E\left[E\left[\left(\frac{Y_2D_2}{\lambda^0(X)} - \frac{Y_2(1-D_2)}{1-\lambda^0(X)}\right)|X, S_2=1, Z_1=z_1\right]|S_2=1, Z_2\geq 0, Z_1=z_1\right] \\
&= \lim_{z_1 \nearrow 0} \frac{E\left[E\left[\frac{Y_2(D_2-\lambda^0(X))}{\lambda^0(X)(1-\lambda^0(X))}|X, S_2=1, Z_1=z_1\right]\cdot 1(Z_2\geq 0)|S_2=1, Z_1=z_1\right]}{P[Z_2\geq 0|S_2=1, Z_1=z_1]} \\
&= \lim_{z_1 \nearrow 0} E\left[\frac{Y_2(D_2-\lambda^0(X))}{\lambda^0(X)(1-\lambda^0(X))}\cdot\frac{P[Z_2\geq 0|X, S_2=1, Z_1=z_1]}{P[Z_2\geq 0|S_2=1, Z_1=z_1]}|S_2=1, Z_1=z_1\right] \\
&= \lim_{z_1 \nearrow 0} E\left[\frac{Y_2(D_2-\lambda^0(X))}{(1-\lambda^0(X))E[D_2|S_2=1, Z_1=z_1]}|S_2=1, Z_1=z_1\right] \\
&= \lim_{z_1 \nearrow 0} E\left[\frac{Y_2(D_2-\lambda^0(X))}{(1-\lambda^0(X))E[D_2|S_2=1, Z_1=z_1]}\cdot\frac{S_2}{P[S_2=1|Z_1=z_1]}|Z_1=z_1\right] \\
&= \lim_{z_1 \nearrow 0} E\left[\frac{Y_2S_2(D_2-\lambda^0(X))}{(1-\lambda^0(X))E[D_2|Z_1=z_1]}|Z_1=z_1\right].
\end{aligned}
$$

And a similar identification strategy could be used to find that

$$
E[\theta^1_{0,2}|D_2(1)=1, Z_1=0] = \lim_{z_1 \searrow 0} E\left[\frac{Y_2S_2(D_2-\lambda^1(X))}{(1-\lambda^1(X))E[D_2|Z_1=z_1]}|Z_1=z_1\right].
$$

Plugging the results to equation (3.2) proves the lemma.

B Proofs for Results in Section 4

Proof of Lemma 4.1

Recall that

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

We prove the lemma for $\hat{\gamma}^1$ following Cai et al. (2000). Results for $\hat{\gamma}^0$ could be shown similarly. To simplify notations, we will drop the superscript 1 and subscript FS in the rest of the proof. That is, we have

$$
\begin{aligned}
(\hat{\gamma}, \hat{\beta}) = \arg\max_{\gamma,\beta} \sum_{i=1}^{n} S_{2i} 1(Z_{1i} \geq 0) K\left(\frac{Z_{1i}}{h}\right) \cdot \\
\Big[D_{2i} \log L(X_i'(\gamma + \beta Z_{1i})) + (1 - D_{2i}) \log \big(1 - L(X_i'(\gamma + \beta Z_{1i}))\big)\Big] \\
\equiv \arg\max_{\gamma,\beta} \ell_n(\gamma, \beta). \qquad \text{(B.1)}
\end{aligned}
$$

Recall that $\gamma^1 = \lim_{z \searrow 0} \gamma(z)$ and $\beta^1 = \lim_{z \searrow 0} \gamma'(z)$. Define

$$
\begin{aligned}
&\gamma^* = \sqrt{nh}(\gamma - \gamma^1), \quad \beta^* = \sqrt{nh}(h\beta - h\beta^1), \\
&\hat{\gamma}^* = \sqrt{nh}(\hat{\gamma}^1 - \gamma^1), \quad \hat{\beta}^* = \sqrt{nh}(h\hat{\beta}^1 - h\beta^1), \\
&\theta = ((\gamma^*)', (\beta^*)')', \quad \hat{\theta} = ((\hat{\gamma}^*)', (\hat{\beta}^*)')', \\
&\widetilde{X}_i = (X_i' \ \frac{Z_{1i} X_i'}{h})', \quad \delta_n = \frac{1}{\sqrt{nh}}, \quad \eta(z, x) = (\gamma^1 + \beta^1 z)' x.
\end{aligned}
$$

Therefore, we have that

$$
(\gamma + \beta Z_{1i})' X_i = (\gamma^1 + \beta^1 Z_{1i})' X_i + \delta_n((\gamma^*)' X_i + (\beta^*)' \frac{Z_{1i} X_i}{h}) = \eta(Z_{1i}, X_i) + \delta_n \theta' \widetilde{X}_i,
$$

and we define $\ell_n^*(\theta)$ as

$$
\begin{aligned}
\ell_n^*(\theta) = \sum_{i=1}^{n} S_{2i} 1(Z_{1i} \geq 0) K\left(\frac{Z_{1i}}{h}\right) \cdot \\
\Big\{\Big[D_{2i} \log L(\eta(Z_{1i}, X_i) + \delta_n \theta' \widetilde{X}_i) + (1 - D_{2i}) \log \big(1 - L(\eta(Z_{1i}, X_i) + \delta_n \theta' \widetilde{X}_i)\big)\Big] - \\
\Big[D_{2i} \log L(\eta(Z_{1i}, X_i)) + (1 - D_{2i}) \log \big(1 - L(\eta(Z_{1i}, X_i))\big)\Big]\Big\}.
\end{aligned}
$$

Given that $(\hat{\gamma}', \hat{\beta}')'$ maximizes $\ell_n(\gamma, \beta)$, we have $\hat{\theta}$ maximizes $\ell_n^*(\theta)$.

Let $q_i(a) = D_{2i} \log L(a) + (1 - D_{2i}) \log(1 - L(a))$, then

$$
q_i'(a) = D_{2i} - L(a), \quad q_i''(a) = -L(a)(1 - L(a)), \quad q_i'''(a) = (2L(a) - 1)L(a)(1 - L(a)).
$$

Taking a Taylor expansion of $q_i(\eta(Z_{1i}, X_i) + \delta_n \theta' \widetilde{X}_i)$ around $\eta(Z_{1i}, X_i)$ for each i, we obtain

$$
\begin{aligned}
\ell_n^*(\theta) = \sum_{i=1}^{n} S_{2i} 1(Z_{1i} \geq 0) K\left(\frac{Z_{1i}}{h}\right) \cdot \\
\Big\{(D_{2i} - L(\eta(Z_{1i}, X_i)))\delta_n \theta' \widetilde{X}_i - \frac{1}{2} L(\eta(Z_{1i}, X_i))(1 - L(\eta(Z_{1i}, X_i)))(\delta_n \theta' \widetilde{X}_i)^2 \\
+ \frac{1}{6}(2L(\bar{\eta}_i) - 1)L(\bar{\eta}_i)(1 - L(\bar{\eta}_i))(\delta_n \theta' \widetilde{X}_i)^3\Big\},
\end{aligned}
$$

where $\bar{\eta}_i$ is between $\eta(Z_{1i}, X_i))$ and $\eta(Z_{1i}, X_i)) + \delta_n \theta' \widetilde{X}_i$ for each i. Note that for each i, the expected value of the last term, $S_{2i} 1(Z_{1i} \geq 0) K\left(\frac{Z_{1i}}{h}\right)(2L(\bar{\eta}_i) - 1)L(\bar{\eta}_i)(1 - L(\bar{\eta}_i))(\delta_n \theta' \widetilde{X}_i)^3$, is bounded by

$$O(\delta^3 E|\|X_i\| \cdot K(Z_{1i}/h)|) = O(n^{-3/2} \cdot h^{-3/2} \cdot h) = O(n^{-1}\delta_n).$$

It then follows that

$$\sum_{i=1}^{n} S_{2i} 1(Z_{1i} \geq 0) K\left(\frac{Z_{1i}}{h}\right) \cdot \frac{1}{6}(2L(\bar{\eta}_i) - 1)L(\bar{\eta}_i)(1 - L(\bar{\eta}_i))(\delta_n \theta' \widetilde{X}_i)^3 = O(\delta_n) = o(1).$$

Therefore,

$$\begin{aligned}
&\ell_n^*(\theta) = Q_n' \theta - \frac{1}{2}\theta' \Delta_n \theta + o_p(1), \text{ where} \\
&Q_n = \delta_n \sum_{i=1}^{n} S_{2i} 1(Z_{1i} \geq 0) K\left(\frac{Z_{1i}}{h}\right) \cdot (D_{2i} - L(\eta(Z_{1i}, X_i)))\widetilde{X}_i, \\
&\Delta_n = \delta_n^2 \sum_{i=1}^{n} S_{2i} 1(Z_{1i} \geq 0) K\left(\frac{Z_{1i}}{h}\right) \cdot L(\eta(Z_{1i}, X_i))(1 - L(\eta(Z_{1i}, X_i)))\widetilde{X}_i \widetilde{X}_i'.
\end{aligned}$$

For the term Δ_n, we have that

$$E[\Delta_n] = \frac{1}{h} E\left[S_2 1(Z_1 \geq 0) K\left(\frac{Z_1}{h}\right) L(\eta(Z_1, X))(1 - L(\eta(Z_1, X))) \begin{pmatrix} X \\ \frac{Z_1 X}{h} \end{pmatrix} \begin{pmatrix} X' & \frac{Z_1 X'}{h} \end{pmatrix}\right].$$

Note that for any $j = 0, 1, \ldots$ and function $g(.)$, by standard arguments,

$$\begin{aligned}
&\frac{1}{h} E\left[S_2 1(Z_1 \geq 0) K\left(\frac{Z_1}{h}\right) L(\eta(Z_1, X))(1 - L(\eta(Z_1, X)))g(X_1)\left(\frac{Z_1}{h}\right)^j\right] \\
=&E\Big[E[S_2 L(\eta(Z_1, X))(1 - L(\eta(Z_1, X)))g(X_1)|Z_1] 1(Z_1 \geq 0)\left(\frac{Z_1}{h}\right)^j K\left(\frac{Z_1}{h}\right)\Big] \\
=&f_{z_1}(0) E[S_2 L(\eta(Z_1, X))(1 - L(\eta(Z_1, X)))g(X_1)|Z_1 = 0] \int_{u \geq 0} u^j K(u) du + o(h).
\end{aligned}$$

Let

$$\Delta_z = f_z(0) \cdot \begin{pmatrix} \mu_{z,0} & \mu_{z,1} \\ \mu_{z,1} & \mu_{z,2} \end{pmatrix} \text{ with } \mu_{z,j} = \int_{u \geq 0} u^j K(u) du, \text{ for } j = 0, 1, \ldots.$$

Then, we have

$$\begin{aligned}
E[\Delta_n] =&\Delta_z \otimes E\left[S_2 L(\eta(Z_1, X))(1 - L(\eta(Z_1, X)))XX' \Big| Z_1 = 0\right] + o(1) \\
\equiv&\Delta + o(1). \qquad \text{(B.2)}
\end{aligned}$$

where $\otimes$ denotes kronecker product. Similar arguments show that for each, $(\Delta_n)_{jk}$, the (j,k)-th element of Δ_n, $Var((\Delta_n)_{jk}) = O(\delta_n) = o(1)$. Therefore, $\Delta_n \xrightarrow{p} \Delta$ and it follows that

$$\ell_n^*(\theta) = Q_n'\theta - \frac{1}{2}\theta'\Delta\theta + o_p(1).$$

Then by the quadratic approximation lemma in Fan and Gijbels. (1996), p. 210, we have that

$$\hat{\theta} = \Delta^{-1}Q_n + o_p(1).$$

For the term Q_n, we have that

$$\begin{aligned}
E[Q_n] &= n\delta_n E\left[S_2 1(Z_1 \geq 0)K\left(\frac{Z_1}{h}\right) \cdot (D_2 - L(\eta(Z_1, X)))\widetilde{X}\right] \\
&= n\delta_n E\left[E[S_2|X, Z_1]1(Z_1 \geq 0)K\left(\frac{Z_1}{h}\right) \cdot (E[D_2|S_2 = 1, X, Z_1] - L(\eta(Z_1, X)))\widetilde{X}\right] \\
&= n\delta_n E\left[E[S_2|X, Z_1]1(Z_1 \geq 0)K\left(\frac{Z_1}{h}\right) \cdot (L(\gamma(Z_1)'X) - L(\eta(Z_1, X)))\widetilde{X}\right] \\
&= O(n\delta_n h \cdot h^2) = O(\sqrt{nh^5}) = o(1).
\end{aligned}$$

To see this, note that $L(\gamma(Z_1)'X) = L(\eta(Z_1, X)) + L(\bar{\eta})(1 - L(\bar{\eta}))(\gamma(Z_1)'X - \eta(Z_1, X))$ where $\bar{\eta}$ is between $\eta(Z_1, X)$ and $\gamma(Z_1)'X$, so $L(\gamma(Z_1)'X) - L(\eta(Z_1, X)) = O_p(\gamma(Z_1)'X - \eta(Z_1, X)))$ because $L(\bar{\eta})(1 - L(\bar{\eta}))$ is bounded by $1/4$. By a mean value expansion of $\gamma(Z_1)'X$ around 0, we have $\gamma(Z_1)'X = (\gamma^1 + \beta^1 Z_1 + \gamma''(\bar{Z}_1)Z_1^2)'X$ where $\bar{Z}_1$ is between 0 and Z_1. Therefore, $\gamma(Z_1)'X - \eta(Z_1, X)' = \gamma''(\bar{Z}_1)'Z_1^2 X$. Therefore, $L(\gamma(Z_1)'X) - L(\eta(Z_1, X)) = O_p(Z_1^2)$. Given that $K(Z_1/h)$ is non-zero when $|Z_1/h| \leq 1$ or equivalently, $|Z_1| \leq h$, $K\left(\frac{Z_1}{h}\right) \cdot (L(\gamma(Z_1)'X) - L(\eta(Z_1, X))) = O_p(K(Z_1/h)h^2)$. It follows that the expectation is $O(n\delta_n h \cdot h^2) = O(\sqrt{nh^5})$ and Assumption 4.5(iii) implies that $O(\sqrt{nh^5}) = o(1)$.

In addition, the variance-covariance matrix of Q_n is given by

$$
\begin{aligned}
&V[Q_n] = \delta^2 n E[S_2 1(Z_1 \geq 0) K^2\left(\frac{Z_1}{h}\right) \cdot (D_2 - L(\eta(Z_1, X)))^2 \widetilde{X}\widetilde{X}'] \\
=&\frac{1}{h} E[E[S_2|Z_1, X] 1(Z_1 \geq 0) K^2\left(\frac{Z_1}{h}\right) \cdot L(\eta(Z_1, X))(1 - L(\eta(Z_1, X)))\widetilde{X}\widetilde{X}'] + O(h^2) \\
=&f_{z_1}(0) \begin{pmatrix} \nu_{0,+} & \nu_{1,+} \\ \nu_{1,+} & \nu_{2,+} \end{pmatrix} \otimes E\left[E[S_2|Z_1, X] L(\eta(Z_1, X))(1 - L(\eta(Z_1, X)))XX' \Big| Z_1 = 0\right] + O(h^2) \\
\equiv&\Omega + o(1), \qquad \text{(B.3)}
\end{aligned}
$$

where $\nu_{k,+} = \int_{u \geq 0} u^k K^2(u) du$ for $k = 0, 1, \ldots$.

Finally, let $\xi_i = S_{2i} 1(Z_{1i} \geq 0) K(Z_{1i}/h)(D_{2i} - L(\eta(Z_{1i}, X_i)))\widetilde{X}_i$. ξ_i satisfies the Lyapounov's condition since $n\delta_n^3 E[\|\xi_i\|^3] = O(\delta_n) \to 0$ by Assumption 4.4. It then follows that $Q_n \xrightarrow{d} (0, \Omega)$ and $\hat{\theta} \xrightarrow{d} (0, \Delta^{-1}\Omega\Delta^{-1})$.

Proof of Theorem 4.1

We derive the asymptotics of $\hat{\alpha}^1$ and $\hat{\alpha}^0$. Recall that

$$
\begin{aligned}
(\hat{\alpha}^1, \hat{\beta}^1) &= \arg\min_{\alpha,\beta} \sum_{\{i: Z_{1i} \geq 0\}} K\left(\frac{Z_{1i}}{h}\right)\left[Y_{2i} - \frac{Y_{2i} S_{2i}(D_{2i} - L(X_i'\hat{\gamma}^1))}{(1 - L(X_i'\hat{\gamma}^1))} - \alpha - \beta Z_{1i}\right]^2, \\
(\hat{\alpha}^0, \hat{\beta}^0) &= \arg\min_{\alpha,\beta} \sum_{\{i: Z_{1i} < 0\}} K\left(\frac{Z_{1i}}{h}\right)\left[Y_{2i} - \frac{Y_{2i} S_{2i}(D_{2i} - L(X_i'\hat{\gamma}^0))}{(1 - L(X_i'\hat{\gamma}^0))} - \alpha - \beta Z_{1i}\right]^2.
\end{aligned}
$$

Note that the local linear estimator is additive in the dependent variables in that if

$$
\begin{aligned}
(\hat{\alpha}_{ay+bx}, \hat{\beta}_{ay+bx}) &= \arg\min_{\alpha,\beta} \sum_{i=1}^n 1(Z_{1i} \geq 0) K\left(\frac{Z_{1i}}{h}\right)\left[(aY_i + bX_i) - \alpha - \beta Z_{1i}\right]^2, \\
(\hat{\alpha}_y, \hat{\beta}_y) &= \arg\min_{\alpha,\beta} \sum_{i=1}^n 1(Z_{1i} \geq 0) K\left(\frac{Z_{1i}}{h}\right)\left[Y_i - \alpha - \beta Z_{1i}\right]^2, \\
(\hat{\alpha}_x, \hat{\beta}_x) &= \arg\min_{\alpha,\beta} \sum_{i=1}^n 1(Z_{1i} \geq 0) K\left(\frac{Z_{1i}}{h}\right)\left[X_i - \alpha - \beta Z_{1i}\right]^2,
\end{aligned}
$$

then $(\hat{\alpha}_{ay+bx}, \hat{\beta}_{ay+bx}) = a(\hat{\alpha}_y, \hat{\beta}_y) + b(\hat{\alpha}_x, \hat{\beta}_x)$. In addition, suppose Y_i satisfies Assumption 4.6 with Y_{2i} replaced with Y_i. By Chiang et al. (2019), we have

$$
\sqrt{nh}\begin{pmatrix} \hat{\alpha}_y - \alpha_y \\ h\hat{\beta}_y - h\beta_y \end{pmatrix} = \frac{1}{\sqrt{nh}} \sum_{i=1}^n \Delta_z^{-1} 1(Z_{1i} \geq 0) K\left(\frac{Z_{1i}}{h}\right)(Y_i - E[Y_i|Z_{1i}])\begin{pmatrix} 1 \\ \frac{Z_{1i}}{h} \end{pmatrix} + o_p(1)
$$

where $\alpha_y = \lim_{z \searrow 0} E[Y|Z = z]$, $\beta_y = \lim_{z \searrow 0} dE[Y|Z = z]/dz$. For each i, we take a second order Taylor expansion of $\frac{Y_{2i}S_{2i}(D_{2i}-L(X_i'\hat{\gamma}^1))}{1-L(X_i'\hat{\gamma}^1)}$ around γ^1 and

$$\begin{aligned}\frac{Y_{2i}S_{2i}(D_{2i}-L(X_i'\hat{\gamma}^1))}{1-L(X_i'\hat{\gamma}^1)} =& \frac{Y_{2i}S_{2i}(D_{2i}-L(X_i'\gamma^1))}{1-L(X_i'\gamma^1)} \\ &+ \frac{Y_{2i}S_{2i}(D_{2i}-1)L(X_i'\gamma^1)}{1-L(X_i'\gamma^1)}X_i'(\hat{\gamma}^1-\gamma^1) + O_p(n^{-1}h^{-1}),\end{aligned}$$

where $O_p(n^{-1}h^{-1})$ holds by the fact that $(\hat{\gamma}^1 - \gamma^1)$ is $O_p(n^{-1/2}h^{-1/2})$ and its coefficient is $O_p(1)$. Therefore, it is true that

$$\begin{aligned}\hat{\alpha}^1 &= \hat{\alpha}_{y_2} - \hat{\alpha}_{\frac{Y_{2i}S_{2i}(D_{2i}-L(X_i'\gamma^1))}{1-L(X_i'\gamma^1)}} - \tilde{\alpha}_c'(\hat{\gamma}^1-\gamma^1) + o_p(\sqrt{nh}) \\ &= \tilde{\alpha}^1 - \tilde{\alpha}_c'(\hat{\gamma}^1-\gamma^1) + o_p(\sqrt{nh})\end{aligned}$$

where

$$\begin{aligned}&\tilde{\alpha}_c = (\tilde{\alpha}_{c,1}, \dots, \tilde{\alpha}_{c,1})', \\ &(\tilde{\alpha}_{c,j}, \tilde{\beta}_{c,j}) = \arg\min_{\alpha,\beta} \sum_{i=1}^n 1(Z_{1i} \geq 0) K\left(\frac{Z_{1i}}{h}\right)\Big[\frac{Y_{2i}S_{2i}(D_{2i}-1)L(X_i'\gamma^1)}{1-L(X_i'\gamma^1))}X_{ji} - \alpha - \beta Z_{1i}\Big]^2 \\ &\quad \text{for } j = 1, \dots, k.\end{aligned}$$

Then it is true that

$$\begin{aligned}\sqrt{nh}(\tilde{\alpha}^1 - \alpha^1) =& \frac{1}{\sqrt{nh}}\sum_{i=1}^n (1\ 0)\Delta_z^{-1} 1(Z_{1i} \geq 0) K\left(\frac{Z_{1i}}{h}\right)\Big(Y_{2i} - E[Y_{2i}|Z_{1i}] \\ &- \frac{Y_{2i}S_{2i}(D_{2i}-L(X_i'\gamma^1))}{1-L(X_i'\gamma^1)} + E\Big[\frac{Y_{2i}S_{2i}(D_{2i}-L(X_i'\gamma^1))}{1-L(X_i'\gamma^1)}|Z_{1i}\Big]\Big)\begin{pmatrix} 1 \\ \frac{Z_{1i}}{h}\end{pmatrix} + o_p(1).\end{aligned}$$

Given that

$$\tilde{\alpha}_{c,j} = \alpha_{c,j} + o_p(1),\ \text{with}\ \alpha_{c,j} = \lim_{z \searrow 0} E\left[\frac{Y_2S_2(D_2-1)L(X'\gamma^1)}{1-L(X'\gamma^1))}X_j\Big|Z = z\right],$$

we have

$$\begin{aligned}\sqrt{nh}\hat{\alpha}_c &= \tilde{\alpha}_c'\sqrt{nh}(\hat{\gamma}^1-\gamma^1) = \alpha_c'\sqrt{nh}(\hat{\gamma}^1-\gamma^1) + o_p(1) \\ &= \lim_{z \searrow 0} E\left[\frac{Y_2S_2(D_2-1)L(X'\gamma^1)}{1-L(X'\gamma^1))}X'\Big|Z = z\right]\sqrt{nh}(\hat{\gamma}^1-\gamma^1) + o_p(1) \\ &\equiv \frac{1}{\sqrt{nh}}\sum_{i=1}^n \nabla_\gamma^1 \cdot \phi_{\gamma^1,ni}(D_{2i}, S_{2i}, Z_{1i}, X_i) + o_p(1),\end{aligned}$$

where ∇^1_γ is the gradient and $\phi_{\gamma^1,ni}(D_{2i}, Z_{1i}, S_{2i}, X_i)$ is the inference function of $\sqrt{nh}(\hat{\gamma}^1 - \gamma^1)$. Both notations are defined in Section 4.1. Then it is true that

$$\begin{aligned}
\sqrt{nh}(\hat{\alpha}^1 - \alpha^1) =& \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} (1\ 0) \Delta_z^{-1} 1(Z_{1i} \geq 0) K\left(\frac{Z_{1i}}{h}\right) \Big(Y_{2i} - E[Y_{2i}|Z_{1i}] \\
&- \frac{Y_{2i} S_{2i}(D_{2i} - L(X_i'\gamma^1))}{1 - L(X_i'\gamma^1)} + E\Big[\frac{Y_{2i} S_{2i}(D_{2i} - L(X_i'\gamma^1))}{1 - L(X_i'\gamma^1)} |Z_{1i}\Big]\Big) \begin{pmatrix} 1 \\ \frac{Z_{1i}}{h} \end{pmatrix} \\
&- \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \nabla^1_\gamma \cdot \phi_{\gamma^1,ni}(D_{2i}, S_{2i}, Z_{1i}, X_i) + o_p(1) \\
=& \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \Big(\tilde{\phi}_{\alpha^1,ni}(Y_{2i}, D_{2i}, S_{2i}, Z_{1i}, X_i) - \nabla^1_\gamma \cdot \phi_{\gamma^1,ni}(D_{2i}, S_{2i}, Z_{1i}, X_i) \Big) + o_p(1) \\
\equiv& \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \phi_{\alpha^1}(Y_{2i}, D_{2i}, S_{2i}, Z_{1i}, X_i) + o_p(1).
\end{aligned}$$

Similarly, we have

$$\begin{aligned}
\sqrt{nh}(\hat{\alpha}^0 - \alpha^0) =& \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} (1, 0)' \Delta_{z,-}^{-1} 1(Z_{1i} < 0) K\left(\frac{Z_{1i}}{h}\right) \Big(Y_{2i} - E[Y_{2i}|Z_{1i}] \\
&- \frac{Y_{2i} S_{2i}(D_{2i} - L(X_i'\gamma^0))}{1 - L(X_i'\gamma^0)} + E\Big[\frac{Y_{2i} S_{2i}(D_{2i} - L(X_i'\gamma^0))}{1 - L(X_i'\gamma^0)} |Z_{1i}\Big]\Big) \begin{pmatrix} 1 \\ \frac{Z_{1i}}{h} \end{pmatrix} \\
&- \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \nabla^0_\gamma \phi_{\gamma^0,ni}(D_{2i}, S_{2i}, Z_{1i}, X_i) + o_p(1) \\
\equiv& \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} \phi_{\alpha^0}(Y_{2i}, D_{2i}, S_{2i}, Z_{1i}, X_i) + o_p(1).
\end{aligned}$$

These results are enough to derive the asymptotic normality of $\hat{\alpha}^1$ and $\hat{\alpha}^0$. It is straightforward to see that $\hat{\alpha}^1$ and $\hat{\alpha}^0$ are mutually independent.

Proof of Theorem 4.2

Recall that $\hat{\gamma}^{0,w}$, $\hat{\gamma}^{1,w}$, $\hat{\beta}^{0,w}_{FS}$, $\hat{\beta}^{1,w}_{FS}$ are given by

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

Again, for brevity, we focus on the $\hat{\gamma}^{1,w}$ case and drop the superscript 1 and subscript FS for notational simplicity. Therefore, by the same argument, we have

$$\ell_n^w(\theta) = (Q_n^w)'\theta - \frac{1}{2}\theta'\Delta_n^w\theta + o_p(1), \text{ where}$$
$$Q_n^w = \delta_n \sum_{i=1}^n W_i S_{2i} 1(Z_{1i} \geq 0) K\left(\frac{Z_{1i}}{h}\right) \cdot (D_{2i} - L(\eta(Z_{1i}, X_i)))\widetilde{X}_i,$$
$$\Delta_n^w = \delta_n^2 \sum_{i=1}^n W_i S_{2i} 1(Z_{1i} \geq 0) K\left(\frac{Z_{1i}}{h}\right) \cdot L(\eta(Z_{1i}, X_i))(1 - L(\eta(Z_{1i}, X_i)))\widetilde{X}_i\widetilde{X}_i'.$$

Note that

$$\begin{aligned} E[\Delta_n^w] &= \frac{1}{h}E\left[WS_2 1(Z_1 \geq 0)K\left(\frac{Z_1}{h}\right)L(\eta(Z_1, X))(1 - L(\eta(Z_1, X)))\widetilde{X}_i\widetilde{X}_i'\right] \\ &= \frac{1}{h}E\left[S_2 1(Z_1 \geq 0)K\left(\frac{Z_1}{h}\right)L(\eta(Z_1, X))(1 - L(\eta(Z_1, X)))\widetilde{X}_i\widetilde{X}_i'\right] \\ &= \Delta + o(1), \end{aligned}$$

where the second equality holds by the fact that W is independent of (S, Z_1, X) and $E[W] = 1$.

Similar arguments show that for each, $(\Delta_n^w)_{jk}$, the (j,k)-th element of Δ_n^w, $V[(\Delta_n^w)_{jk}] = O(\delta_n) = o(1)$. Therefore, $\Delta_n^w \xrightarrow{p} \Delta$ and it follows that

$$\ell_n^w(\theta) = (Q_n^w)'\theta - \frac{1}{2}\theta'\Delta\theta + o_p(1).$$

Let $\hat{\gamma}^{*,w} = \sqrt{nh}(\hat{\gamma}^{1,w} - \gamma^1)$, $\hat{\beta}^{*,w} = \sqrt{nh}(h\hat{\beta}^{1,w} - h\beta^1)$, $\hat{\theta}^w = ((\hat{\gamma}^{*,w})', (\hat{\beta}^{*,w})')'$. Then, by the quadratic approximation lemma again, we have that $\hat{\theta}^w = \Delta^{-1}Q_n^w + o_p(1)$. Therefore,

$$\begin{aligned} \hat{\theta}^w - \hat{\theta} &= \Delta^{-1}(Q_n^w - Q_n) + o_p(1) \\ &= \sum_{i=1}^n (W_i - 1)\Big[S_{2i}1(Z_{1i} \geq 0)K\left(\frac{Z_{1i}}{h}\right) \cdot (D_{2i} - L(\eta(Z_{1i}, X_i)))\widetilde{X}_i\Big] + o_p(1). \end{aligned}$$

Given that $E[W_i - 1] = 0$ and $Var(W_i - 1) = 1$ and that $\{W_i - 1\}_{i=1}^n$ is independent of the sample path, we can apply the standard multiplier bootstrap argument as in Ma and Kosorok (2005) to show that conditional on the sample path with probability one, $\hat{\theta}^w - \hat{\theta} \xrightarrow{d} (0, \Delta^{-1}\Omega\Delta^{-1})$ which shows the validity of the weighted bootstrap for the local MLE estimator.

Following the same arguments in the proof of Theorem 4.1, we can show that

$$\sqrt{nh}(\hat{\alpha}^{1,w} - \alpha^1) = \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} W_i \cdot \phi_{\alpha^1}(Y_{2i}, D_{2i}, S_{2i}, Z_{1i}, X_i) + o_p(1),$$
$$\sqrt{nh}(\hat{\alpha}^{0,w} - \alpha^0) = \frac{1}{\sqrt{nh}} \sum_{i=1}^{n} W_i \cdot \phi_{\alpha^0}(Y_{2i}, D_{2i}, S_{2i}, Z_{1i}, X_i) + o_p(1),$$

and it follows that

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

Therefore, the two left hand side expressions converge to the same distributions as $\sqrt{nh}(\hat{\alpha}^1 - \alpha^1)$ and $\sqrt{nh}(\hat{\alpha}^0 - \alpha^0)$, respectively, conditional on sample path with probability approaching one.

With all the results above, we know that $\sqrt{nh}(\hat{\bar{\theta}}^w_{1,1} - \hat{\bar{\theta}}_{1,1})$ is asymptotic normal and converges to the same limiting distribution as $\sqrt{nh}(\hat{\bar{\theta}}_{1,1} - \bar{\theta}_{1,1})$ conditional on sample path with probability approaching one.

C More Simulation Results

This section extends DGPs 1-4 in Section 5 to to examine small sample performances of proposed estimators of $E[\theta_{\tau,1}|Z_1 = 0]$ for $\tau = 0, 1, 2, 3$. Like in Section 5, the proposed estimators are valid under all four DGPs while the recursive CFR estimators are only valid under DGPs 1 and 3.

Recall that $\mathbf{0}_t$ denote a t-dimensional vector of zeros. For all DGPs, let

$$X \sim U[0, 10],\ Z_1 \sim X - 10 \cdot Beta(2, 2),\ (u_{y1}, u_{y2}, u_{y3}, u_{y4}) \sim i.i.d.\ N(0, 0.5),$$
$$Y_t(\mathbf{0}_t) = 0.1X + 0.5Z_1 + 0.1XZ_1 + 0.1Z_1^2 + u_{yt},\ \text{for } t = 1, 2, 3, 4,$$
$$(u_{s1}, u_{s2}, u_{s3}) \sim i.i.d.\ N(0, 1),\ S_t(0) = 1(u_{st} \geq 0),\ S_t(1) = 1(1 + u_{st} \geq 0),$$
$$(v_{z1}, v_{z2}, v_{z3}) \sim i.i.d.\ logis(0, 1),\ Z_t(\mathbf{0}_t) = 0.3 + 0.1X_1 + v_{zt},\ \text{for } t = 2, 3, 4.$$

Potential outcomes with nonzero treatment status are simulated with $Y_t(\mathbf{0}_t)$ defined above and individual primary effects that are labeled by the number of periods from the focal treatment and the outcome as well as the treatment status in the very last round so that the Markovian condition in Assumption 3.5 is satisfied. For example, $Y_2(1,1) = Y_2(0,0) + \theta_{1,1} + \theta_0^1$ and $Y_3(1,0,1) = Y_3(0,0,0) + \theta_{2,1} + \theta_0^0$.

DGP 1: $\theta_{0,1} = \theta_0^0 = \theta_0^1 = 0.5$, $\theta_{1,1} = \theta_1^0 = \theta_1^1 = 0.2$, $\theta_{2,1} = \theta_2^0 = \theta_2^1 = 0.3$, $\theta_{3,1} = 0$.

DGP 2: $\theta_{0,1} = \theta_0^0 = 0.5$, $\theta_0^1 = 0.1$, $\theta_{1,1} = \theta_1^0 = 0.2$, $\theta_1^1 = -0.2$, $\theta_{2,1} = \theta_2^0 = 0.3$, $\theta_2^1 = -0.3$, $\theta_{3,1} = 0$.

DGP 3: $\theta_{0,1} = \theta_0^0 = \theta_0^1 = 0.5 + e_0$, $\theta_{1,1} = \theta_1^0 = \theta_1^1 = 0.2 + e_1$, $\theta_{2,1} = \theta_2^0 = \theta_2^1 = 0.3 + e_2$, $\theta_{3,1} = e_3$, $(e_0, e_1, e_2, e_3) \sim i.i.d.\ U[-0.5, 0.5]$.

DGP 4: $\theta_{0,1} = 0.5$, $\theta_0^0 = 0.5 + 0.2(u_{s2} + u_{s3} + u_{s4})$, $\theta_0^1 = 0.5$, $\theta_{1,1} = 0.2$, $\theta_1^0 = 0.2 + 0.3(u_{s2} + u_{s3})$, $\theta_1^1 = 0.2$, $\theta_{2,1} = 0.3$, $\theta_2^0 = 0.3 + 0.5u_{s2}$, $\theta_2^1 = 0.3$, $\theta_{3,1} = 0$.

Similarly, potential running variables are simulated with $Z_t(\mathbf{0}_t)$ and individual primary effects on the potential running variables. Specifically,

$$
\begin{aligned}
&Z_2(1) = Z_2(0) + (1\ X_1)\gamma_0,\ Z_3(0,1) = Z_3(\mathbf{0}_2) + (1\ X_1)\gamma_0^0,\\
&Z_3(1,0) = Z_3(\mathbf{0}_2) + (1\ X_1)\gamma_{1,1},\ Z_3(1,1) = Z_3(\mathbf{0}_2) + (1\ X_1)(\gamma_{1,1} + \gamma_0^1),\\
&Z_4(0,0,1) = Z_4(\mathbf{0}_3) + (1\ X_1)\gamma_0^0,\ Z_4(0,1,0) = Z_4(\mathbf{0}_3) + (1\ X_1)\gamma_1^0,\\
&Z_4(0,1,1) = Z_4(\mathbf{0}_3) + (1\ X_1)(\gamma_1^0 + \gamma_0^1),\ Z_4(1,0,0) = Z_4(\mathbf{0}_3) + (1\ X_1)\gamma_{2,1},\\
&Z_4(1,1,0) = Z_4(\mathbf{0}_3) + (1\ X_1)(\gamma_{2,1} + \gamma_1^1),\ Z_4(1,0,1) = Z_4(\mathbf{0}_3) + (1\ X_1)(\gamma_{2,1} + \gamma_0^0),\\
&\gamma_{0,1} = (-0.3\ \ -0.1),\ \gamma_0^0 = (0.1\ 0.1),\ \gamma_0^1 = (-0.2\ \ -0.1),\\
&\gamma_{1,1} = \gamma_1^0 = \gamma_1^1 = \gamma_{2,1} = (-0.1\ \ -0.1).
\end{aligned}
$$

Given the above potential random variables, observed random variables are defined following the potential outcome framework discussed in Section 3.

Table 5 reports the average of the proposed and recursive CFR estimators among 1,000 simulations. The true value is 0.5, 0.2, 0.3, and 0 for the immediate, one-period-after, two-period-after, and three-period-after ATEs. As is predicted by the theory, the proposed estimators average around the true value among all four DGPs, while the recursive estimators only perform well under DGPs 1 and 3.

Table 6 reports proportions of rejections in two-sided t-tests associated with proposed ATE estimators. The first half of the table shows the size of the tests with the true value of ATEs stated under the null. The second half of the table shows the power of the tests with the null set incorrectly to 0.3 for the immediate ATE and 0 for all other longer-term ATEs. Thus, it is clear that the proposed method controls size well under the null and has power going to one under the alternative.

Table 5: Performance of Immediate, One-, Two-, and Three-period-after ATE Estimators

	Immediate			One-period-after			Two-period-after			Three-period-after		
k	4.25	4.5	4.75	4.25	4.5	4.75	4.25	4.5	4.75	4.25	4.5	4.75
							Proposed Estimation Strategy					
	DGP 1											
n=2000	0.505	0.506	0.507	0.204	0.206	0.208	0.299	0.302	0.305	-0.001	0.001	0.003
n=4000	0.507	0.507	0.508	0.202	0.204	0.205	0.296	0.298	0.299	0.0001	0.001	0.003
n=8000	0.507	0.507	0.507	0.205	0.206	0.206	0.306	0.308	0.308	0.002	0.003	0.004
	DGP 2											
n=2000	0.503	0.504	0.505	0.189	0.192	0.194	0.292	0.295	0.298	0.007	0.009	0.011
n=4000	0.506	0.507	0.507	0.205	0.206	0.207	0.307	0.308	0.310	0.010	0.012	0.013
n=8000	0.506	0.507	0.507	0.202	0.203	0.203	0.307	0.307	0.307	0.012	0.013	0.013
	DGP 3											
n=2000	0.506	0.507	0.508	0.198	0.200	0.203	0.300	0.303	0.306	-0.014	-0.011	-0.008
n=4000	0.509	0.509	0.510	0.203	0.204	0.205	0.294	0.295	0.297	0.001	0.003	0.004
n=8000	0.508	0.509	0.509	0.202	0.202	0.203	0.304	0.304	0.305	0.006	0.007	0.007
	DGP 4											
n=2000	0.509	0.509	0.509	0.188	0.191	0.193	0.300	0.302	0.304	0.003	0.005	0.007
n=4000	0.505	0.506	0.507	0.198	0.200	0.201	0.303	0.304	0.306	0.017	0.018	0.019
n=8000	0.503	0.504	0.504	0.200	0.202	0.203	0.301	0.303	0.304	0.007	0.009	0.010
							Recursive CFR Strategy					
	DGP 1											
n=2000	0.505	0.506	0.507	0.209	0.210	0.211	0.310	0.311	0.312	0.008	0.010	0.011
n=4000	0.507	0.507	0.508	0.209	0.210	0.210	0.305	0.307	0.307	0.006	0.008	0.008
n=8000	0.507	0.507	0.507	0.207	0.208	0.208	0.310	0.311	0.312	0.007	0.008	0.009
	DGP 2											
n=2000	0.503	0.504	0.505	0.100	0.101	0.102	0.244	0.246	0.247	-0.095	-0.093	-0.092
n=4000	0.506	0.507	0.507	0.105	0.106	0.106	0.248	0.249	0.249	-0.099	-0.098	-0.097
n=8000	0.506	0.507	0.507	0.101	0.102	0.103	0.245	0.246	0.246	-0.099	-0.098	-0.097
	DGP 3											
n=2000	0.506	0.507	0.508	0.216	0.217	0.218	0.317	0.319	0.320	0.011	0.012	0.013
n=4000	0.509	0.509	0.510	0.208	0.209	0.210	0.308	0.309	0.310	0.011	0.012	0.013
n=8000	0.508	0.509	0.509	0.206	0.207	0.207	0.309	0.310	0.310	0.012	0.013	0.013
	DGP 4											
n=2000	0.509	0.509	0.509	0.171	0.172	0.172	0.253	0.254	0.254	-0.078	-0.077	-0.076
n=4000	0.505	0.506	0.507	0.170	0.171	0.172	0.254	0.255	0.256	-0.066	-0.066	-0.065
n=8000	0.503	0.504	0.504	0.171	0.172	0.173	0.252	0.253	0.254	-0.075	-0.074	-0.073

Note: All Monte Carlo experiments use 1,000 simulation repetitions and weighted bootstrap with 1,000 bootstrap repetitions.

Table 6: Immediate and One-period-after Average Primary Treatment Effects: Inference of Proposed Estimators

	Immediate			One-period-after			Two-period-after			Three-period-after		
k	4.25	4.5	4.75	4.25	4.5	4.75	4.25	4.5	4.75	4.25	4.5	4.75
						Size of Two-sided T-tests						
	DGP 1											
n=2000	0.058	0.063	0.062	0.067	0.073	0.069	0.043	0.042	0.040	0.058	0.062	0.057
n=4000	0.046	0.050	0.052	0.058	0.058	0.060	0.050	0.052	0.055	0.053	0.054	0.051
n=8000	0.059	0.059	0.061	0.050	0.045	0.046	0.060	0.059	0.057	0.049	0.042	0.044
	DGP 2											
n=2000	0.056	0.055	0.058	0.063	0.060	0.057	0.053	0.047	0.052	0.050	0.060	0.056
n=4000	0.055	0.055	0.059	0.055	0.056	0.060	0.061	0.065	0.063	0.064	0.061	0.061
n=8000	0.061	0.065	0.060	0.070	0.072	0.064	0.054	0.047	0.052	0.058	0.057	0.056
	DGP 3											
n=2000	0.061	0.057	0.054	0.068	0.065	0.067	0.057	0.054	0.051	0.061	0.060	0.063
n=4000	0.064	0.065	0.061	0.057	0.061	0.063	0.070	0.070	0.068	0.052	0.053	0.052
n=8000	0.053	0.052	0.051	0.047	0.040	0.041	0.055	0.055	0.057	0.066	0.063	0.061
	DGP 4											
n=2000	0.060	0.055	0.058	0.060	0.060	0.063	0.062	0.062	0.062	0.056	0.056	0.054
n=4000	0.048	0.049	0.049	0.064	0.064	0.063	0.058	0.061	0.061	0.050	0.046	0.044
n=8000	0.063	0.054	0.053	0.060	0.061	0.059	0.049	0.046	0.046	0.058	0.058	0.062
						Power of Two-sided T-tests						
	DGP 1											
n=2000	0.581	0.621	0.632	0.328	0.348	0.362	0.517	0.550	0.559	0.476	0.495	0.498
n=4000	0.854	0.875	0.886	0.519	0.553	0.571	0.800	0.823	0.831	0.795	0.812	0.818
n=8000	0.980	0.986	0.986	0.783	0.811	0.821	0.978	0.984	0.986	0.974	0.979	0.979
	DGP 2											
n=2000	0.584	0.611	0.618	0.307	0.319	0.332	0.507	0.525	0.538	0.450	0.467	0.472
n=4000	0.846	0.872	0.882	0.530	0.554	0.562	0.817	0.843	0.853	0.721	0.742	0.746
n=8000	0.975	0.982	0.983	0.767	0.787	0.797	0.969	0.973	0.979	0.947	0.953	0.958
	DGP 3											
n=2000	0.540	0.573	0.586	0.285	0.312	0.316	0.475	0.502	0.521	0.422	0.444	0.443
n=4000	0.799	0.827	0.839	0.499	0.523	0.533	0.709	0.733	0.744	0.660	0.672	0.679
n=8000	0.973	0.978	0.982	0.749	0.775	0.780	0.936	0.952	0.954	0.896	0.903	0.915
	DGP 4											
n=2000	0.614	0.639	0.650	0.296	0.308	0.316	0.470	0.493	0.507	0.608	0.627	0.636
n=4000	0.842	0.873	0.886	0.497	0.525	0.528	0.762	0.791	0.802	0.863	0.885	0.893
n=8000	0.975	0.984	0.990	0.780	0.812	0.822	0.949	0.963	0.969	0.993	0.995	0.997

Note: All Monte Carlo experiments use 1,000 simulation repetitions and weighted bootstrap with 1,000 bootstrap repetitions. All t-tests use the 5% significance level.