ONLINE APPENDIX

A Additional Identification Results

In this section, we omit the subscript i in all random variables for notational simplicity. In addition, we suppress the subscript $D_2|S_2$ in propensity score functions throughout the online appendix

A.1 Multi-period CFR

The following lemma extends Lemma 2.1 to the general multi-period model discussed in Section 3. First we extend the smoothness condition in Assumption 2.1.2.

Assumption A.1 *There exists an* $\epsilon > 0$, *such that* $E\left[\tilde{Y}_{(1+\tau)}(d_1)|Z_1 = z_1\right]$ *is continuous in* $z_1 \in \mathcal{N}_\epsilon$ *for all* $\tau = 2, ..., K-1$ *and* $E\left[\tilde{D}_{(2+\tau)}(d_1)|Z_1 = z_1\right]$ *is continuous for all* $\tau = 1, ..., K-2$, *for both* $d_1 = 0, 1$.

Meanwhile, the mean equivalence condition in (2.4) need to be extended to

$$ATE_\tau \equiv E\left[\theta_{\tau,1}|Z_1 = 0\right] = E\left[\theta^{\ell^k}_{\tau,(k+1)}|Z_1 = 0\right], \tag{A.1}$$

for all $\tau = 0, 1, ..., K-1$, $\ell^k \in \mathcal{L}^k$, and $k = 1, ..., K-\tau-1$. The random treatment selection condition in (2.5) need to be extended to

$$E\left[\theta^{\ell^k}_{\tau,(k+1)}\mathfrak{D}(\ell^k, 1)|Z_1 = 0\right] = E\left[\theta^{\ell^k}_{\tau,(k+1)}|Z_1 = 0\right] \cdot E\left[\mathfrak{D}(\ell^k, 1)|Z_1 = 0\right], \tag{A.2}$$

for all $\tau = 0, 1, ..., K-1$, $\ell^k \in \mathcal{L}^k$, and $k = 1, ..., K-\tau-1$.

Lemma A.1 *Under Assumptions 2.1, A.1 and conditions in equations* (2.4), (2.5), (A.1), *and* (A.2), *the following CFR recursive identification result holds:*

$$\begin{aligned}
ATE_0 &= \lim_{z_1 \searrow 0} E[Y_1|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[Y_1|Z_1 = z_1], \\
ATE_\tau &= \lim_{z_1 \searrow 0} E[Y_{1+\tau}|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[Y_{1+\tau}|Z_1 = z_1] \\
&\quad - \sum_{s=0}^{\tau-1} ATE_s \cdot \left(\lim_{z_1 \searrow 0} E[D_{1+\tau-s}|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[D_{1+\tau-s}|Z_1 = z_1]\right),
\end{aligned}$$

for all $\tau = 1, 2, ..., K-1$.

Lemma A.1 reduces to Lemma 2.1 when $K = 2$. The lemma is proven in Section C.

A.2 Extended CFR with Covariates

This section extends the recursive CFR identification strategy using covariates.

Assumption A.2 *There exists an $\epsilon > 0$, such that for all $x \in \mathcal{X}$,*

1. *Z_1 is continuous in $z_1 \in \mathcal{N}_\epsilon$ with $P[Z_1 \geq 0|X = x] \in (0, 1)$;*
2. *$E[Y_1(d_1)|X = x, Z_1 = z_1]$, $E[\tilde{Y}_2(d_1)|Z_1 = z_1]$, and $E[D_2(d_1)|X = x, Z_1 = z_1]$ are all continuous in $z_1 \in \mathcal{N}_\epsilon$, for both $d_1, d_2 = 0, 1$.*

The following Lemma summarizes the extension.

Lemma A.2 *Under Assumption A.2 and conditions in equations* (2.6) *and* (2.7)*, the following recursive identification results hold:*

$$ATE_1 \equiv \lim_{z_1 \searrow 0} E[Y_2|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[Y_2|Z_1 = z_1] - E[CATE_0(X)p_2(X)|Z_1 = 0],$$

where $CATE_0(x) = \lim_{z_1 \searrow 0} E[Y_1|X = x, Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[Y_1|X = x, Z_1 = z_1]$ and $p_2(x) = \lim_{z_1 \searrow 0} E[D_2|X = x, Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[D_2|X = x, Z_1 = z_1]$, for all $x \in \mathcal{X}$.

The lemma is proven in Section C.

A.3 Partial Identification

In this section, we discuss partial identification of the one-period-after ATE by replacing the CIA condition in Assumption 2.2 to a monotonicity condition that might be more plausible in some empirical applications.

Assumption A.3 (Monotone 1 - Benchmark) *$E[Y_2(d_1, 0)|Z_2(d_1) = z_2, S_2(d_1) = 1, Z_1 = z_1]$ is (weakly) monotonically increasing in z_2 for all $z_1 \in \mathcal{N}_\epsilon$.*

Assumption A.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. The assumption nests the CIA condition in Assumption 2.2. Under Assumption A.3, for $d_1 = 0, 1$,

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

Then, following the identification results in Section 2.3, we know that

$$E[\theta^0_{0,2}D_2(0)|Z_1=0] \le \lim_{z_1 \nearrow 0} E[Y_2 S_2(D_2-\lambda^0)]/(1-\lambda^0), \text{ and}$$
$$E[\theta^1_{0,2}D_2(1)|Z_1=0] \le \lim_{z_1 \searrow 0} E[Y_2 S_2(D_2-\lambda^1)]/(1-\lambda^1).$$

Assumption A.4 (Monotone 2 - Benchmark) $E[\theta^{d_1}_{0,2}|Z_2(d_1)=z_2, S_2(d_1)=1, Z_1=z_1]$ *is (weakly) monotonically increasing in* $z_2 \in \mathbb{R}$ *for all* $x \in \mathcal{X}$ *and* $z_1 \in \mathcal{N}_\epsilon$.

Assumption A.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 2.2 are extended to conditional means of potential outcomes conditional on both Z_1 and Z_2, it is easy to show that under Assumption A.4,

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

Combining the inequalities above and the decomposition stated in equation (2.2), we bound the one-period-after ATE $E[\theta_{1,1}|Z_1=z_1]$ as

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

where α^0 and α^1 are identified in Lemma 2.2 but used here without the conditioning covariate X. The inequalities can also be extended trivially to include covariates.

A similar partial identification result of longer-term ATEs could also be obtained, given the Markovian-type condition in Assumption 2.2.1 and generalized versions of Assumptions A.3 and A.4. Details are omitted for brevity of the paper. All identified bounds can be estimated by conventional non-parametric RD estimators (see, for example, Chiang et al., 2019).

A.4 Special Cases

A.4.1 Up-to-one Treatment Case I

An important special case of the dynamic RD model is that every individual can only receive up to one treatment. For example, the effect of unionization studied in DiNardo and

Lee (2004) and Lee and Mas (2012) is a case where the treatment has an absorbing state. Then, the universe of past treatment paths, or $\mathcal{L}^k$ for all k, only includes paths with up to one treatment. For instance, $\mathcal{L}^2 = \{(0,0),(0,1)\}$, and $\mathcal{L}^3 = \{(0,0,0),(0,1,0),(0,0,1)\}$.

Consider the identification of one-period-after and two-period after ATEs. In this special case, the relationship between total and direct effects reduces to

$$
\begin{aligned}
\tilde{\theta}_{1,1} =& \theta_{1,1} - \theta^0_{0,2} D_2(0), \\
\tilde{\theta}_{2,1} =& \theta_{2,1} - \tilde{\theta}^0_{1,2} D_2(0) - \theta^{(0,0)}_{0,3} D_3(0,0) = \theta_{2,1} + \tilde{\theta}^0_{1,2}\eta_{0,1} + \theta^{(0,0)}_{0,3}\eta_{1,1}.
\end{aligned}
$$

This leads to simplification in Lemma 3.2. For one-period-after and two-period after ATEs, we have that

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

A.4.2 Up-to-one Treatment Case II

In a related but different special case, treatment is administrated after all rounds of RD have taken place, if an individual qualifies for it in *any* round. For example, Clark and Martorell (2014) use RD to study the effect of a high school diploma, whereby every student has multiple chances to take the test and qualify for the diploma. This RD setting could be regarded 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 are compliers, and those who earn eligibility for treatment through later rounds of RD are always-takers.

Use a two-round model for intuition. The potential outcome framework is

$$
\begin{aligned}
Y =& Y(1)D + Y(0)(1-D), \\
D =& D_1 + D_2 = 1(Z_1 \geq 0) + (1-D_1)\cdot S_2(0) \cdot 1(Z_2(0) \geq 0).
\end{aligned}
$$

Let $C = 1 - S_2(0)\cdot 1(Z_2(0) \geq 0)$. Under smoothness conditions,

$$
\begin{aligned}
\lim_{z_1 \searrow 0} E[Y|Z_1=z_1] &= \lim_{z_1 \searrow 0} E[Y(1)|Z_1=z_1] = E[Y(1)|Z_1=0], \\
\lim_{z_1 \nearrow 0} E[Y|Z_1=z_1] &= \lim_{z_1 \nearrow 0} E[Y(1)\cdot(1-C) + Y(0)\cdot C|Z_1=z_1] \\
&= E[Y(1)|Z_1=0] - E[(Y(1)-Y(0))\cdot C|Z_1=0].
\end{aligned}
$$

In this special case, conventional RD smoothness conditions could identify the average treatment effect among those who would opt out or fail to meet later-round RD cutoffs upon barely failing the first round.

$$E[Y(1)-Y(0)|C=1,Z_1=0]=\frac{\lim_{z_1\searrow 0}E[Y|Z_1=z_1]-\lim_{z_1\nearrow 0}E[Y|Z_1=z_1]}{\lim_{z_1\nearrow 0}P[D_2=0|Z_1=z_1]}$$

If the goal is, instead, to identify the average effect for everyone at the first-round RD cutoff, then we would like to note that

$$\begin{aligned}&\lim_{z_1\searrow 0}E[Y|Z_1=z_1]-\lim_{z_1\nearrow 0}E[Y|Z_1=z_1]\\ =&E[Y(1)-Y(0)|Z_1=0]-E[Y(1)-Y(0)|D_2(0)=1,Z_1=0]\cdot P[D_2(0)=1|Z_1=0],\end{aligned}$$

where $E[Y(1)-Y(0)|Z_1=0]$ could then be identified following the same intuition of point identifying $E[\theta_{1,1}|Z_1=0]$ in the general model.

A.4.3 Not Observing Some Initial Rounds of RD

In some empirical applications (e.g., U.S. House of Representative elections in Lee, 2008), the first observed period may not be the first round of treatment. In such a case, identification strategies described in Lemmas 2.2 and 3.2 need to be re-interpreted or modified.

Suppose the initial S rounds of RD (and treatment decisions) are unobserved. The observed outcome discontinuity at the first observed RD cutoff then identifies a *contemporaneous effect of treatment* (CET) for individuals at the cutoff, following the terminology in Blackwell and Glynn (2018). Let $L^{pre}\in\mathcal{L}^S$ be the random variable denoting the *realized* but *unobserved* treatment path of all unobserved rounds of RD. Let $Y_1(L^{pre},d_1)$ be the potential outcome of the first observed outcome, characterized by the potential treatment decision d_1 in the first observed period. It is then clear that the following average immediate treatment effect is identified under standard smoothness conditions:

$$\begin{aligned}&E[Y_1(L^{pre},1)|Z_1=0]-E[Y_1(L^{pre},0)|Z_1=0]\\ =&\lim_{z_1\searrow 0}E[Y_1|Z_1=z_1]-\lim_{z_1\nearrow 0}E[Y_1|Z_1=z_1].\end{aligned}$$

Similarly, the CET concept could be extended to capture long-term effects in the special setting with unobserved initial rounds of data.

As is explained in Blackwell and Glynn (2018), CET reflects the effect of a treatment averaged across all of the treatment histories up to the period of that treatment. In other words, CET is a weighted average of path-specific treatment effects with unknown weights. Using notations of the previous sections,

$$
\begin{aligned}
&E[Y_1(L^{pre},1)|Z_1=0]-E[Y_1(L^{pre},1)|Z_1=0]\\
=&\sum_{\ell^{pre}\in\mathcal{L}^S}E[Y_1(\ell^{pre},1)-Y_1(\ell^{pre},0)|\mathfrak{D}_{pre}(\ell^{pre})=1,Z_1=0]P[\mathfrak{D}_{pre}(\ell^{pre})=1|Z_1=0],
\end{aligned}
$$

where ℓ^{pre} is a pre-observation treatment path (e.g., $\ell^{pre}=(0,1,\mathbf{0}_{S-2})$), and $\mathfrak{D}_{pre}(.)$ is the unobserved pre-observation treatment path indicator.

If a researcher would, instead, like to identify path-specific ATEs, then an alternative strategy is to use a pre-focal-treatment condition using observed treatments, and then use a Markovian-type assumption to eliminate the dependence of treatment effects on the unobserved treatment history. The empirical section of the paper takes this approach since the expenditure outcomes are not observed in the first several years of the data.

B Additional Inference Results

B.1 Additional Assumptions the Asymptotic Results

The following two assumptions are required for the asymptotic results stated in Theorem 4.1 in Section 4.1.

Assumption B.1 *For $j=1,\dots,k$, 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 B.2 *Moment $E[\|X\|^3|Z_1=z_1]$ exists and is bounded on $\mathcal{N}_\epsilon$ for some $\epsilon>0$.*

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 4.1-4.3 and B.1-B.2, 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 two assumptions are required for the asymptotic results stated in Theorem 4.1 and Theorem 4.2.

For notational simplicity, we use $\widetilde{Y}_2(\gamma)$ to denote $\frac{Y_2S_2(D_2-L(X'\gamma))}{(1-L(X'\gamma))}$. For $d_1=0,1$, define $\widetilde{Y}_2^{d_1}=\frac{Y_2S_2(D_2-L(X'\gamma^{d_1}))}{(1-L(X'\gamma^{d_1}))}$ and $\nabla_\gamma\widetilde{Y}_2^{d_1}=\nabla_\gamma\widetilde{Y}_2(\gamma)|_{\gamma=\gamma^{d_1}}$. Let $\nabla_\gamma^2\widetilde{Y}_2(\gamma)$ be the Hessian matrix of $\widetilde{Y}_2(\gamma)$.

Assumption B.3 *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 B.4 *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.*

B.2 Alternative Inference Procedure with Robust RD Inference

In this section, we propose an alternative inference procedure that extends the robust RD inference method in CCT to the two-step one-period-after ATE estimator proposed in Section 2. The new inference procedure avoids under-smoothing in the second-step estimation. On the other hand, the first-step propensity score estimation in the new procedure needs to use a higher-order local polynomial and a larger bandwidth than the second-step ATE estimation. We detail the procedure below.

To carry out the alternative inference procedure, we redefine the first-step estimators $\hat{\gamma}_{FS}^0$ and $\hat{\gamma}_{FS}^1$ using a local quadratic method and a first-step specific bandwidth h_{FS}.

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

where h_{FS} is the bandwidth for the first-step propensity score estimator.

Given the first-step estimators, define

$$\widehat{Y}_{1,i} = Y_{2i} - \frac{Y_{2i}S_{2i}(D_{2i} - p(X_i, \hat{\gamma}^1_{FS}))}{1 - p(X_i, \hat{\gamma}^1_{FS})}, \quad \widehat{Y}_{0,i} = Y_{2i} - \frac{Y_{2i}S_{2i}(D_{2i} - p(X_i, \hat{\gamma}^0_{FS}))}{1 - p(X_i, \hat{\gamma}^0_{FS})}.$$

Let $\hat{\alpha}^1(h_n)$ and $\hat{\alpha}^0(h_n)$ be estimators of α^1 and α^0, respectively, with the second-step bandwidth h_n:

$$(\hat{\alpha}^1(h_n), \hat{\beta}^1(h_n)) = \arg\min_{\alpha,\beta} \sum_{i=1}^{n} 1(Z_{1i} \geq 0) K\left(\frac{Z_{1i}}{h_n}\right) \left[\widehat{Y}_{1,i} - \alpha - \beta Z_{1i}\right]^2,$$

$$(\hat{\alpha}^0(h_n), \hat{\beta}^0(h_n)) = \arg\min_{\alpha,\beta} \sum_{i=1}^{n} 1(Z_{1i} < 0) K\left(\frac{Z_{1i}}{h}\right) \left[\widehat{Y}_{0,i} - \alpha - \beta Z_{1i}\right]^2.$$

Following CCT, we define a bias-corrected estimator for the one-period-after ATE, $\bar{\theta}_{1,1}$, as

$$\hat{\theta}^{bc}_{1,1}(h_n, b_n) = \hat{\alpha}^1(h_n) - \hat{\alpha}^0(h_n) - h_n^2 \hat{B}_{1,1}(h_n, b_n)$$

where $h_n^2 \hat{B}_{1,1}(h_n, b_n)$ is the bias estimator of the local linear estimator $\hat{\alpha}^1(h_n) - \hat{\alpha}^0(h_n)$ using a pilot bandwidth b_n defined later. Under proper assumptions, the first-step estimation of γ^0_{FS} and γ^1_{FS} does not influence either the first-order asymptotic bias or the asymptotic variance of $\hat{\alpha}^1(h_n) - \hat{\alpha}^0(h_n)$. In other words, the bias term could be defined as

$$\hat{B}_{1,1}(h_n, b_n) = \frac{\hat{\rho}^1(b_n)}{2!} B_{+,1,1}(h_n) - \frac{\hat{\rho}^0(b_n)}{2!} B_{-,1,1}(h_n)$$

with $B_{+,1,1}(h_n)$ and $B_{-,1,1}(h_n)$ following the definitions in Lemma A.1(B) of CCT replacing outcome variables in the Lemma by $\widetilde{Y}_{1,i}$ and $\widetilde{Y}_{0,i}$. Let $V^{bc}_{1,1}(h_n, b_n)$ be the variance given in Theorem 1 of CCT with outcome variables replaced by $\widetilde{Y}_{1,i}$ and $\widetilde{Y}_{0,i}$. Under suitable conditions, we are able to show that

$$\frac{\hat{\theta}^{bc}_{1,1}(h_n, b_n) - \bar{\theta}_{1,1}}{\sqrt{V^{bc}_{1,1}(h_n, b_n)}} \xrightarrow{d} N(0, 1).$$

Next we study the asymptotic properties of the proposed two-step estimator with $p(x, \gamma) = L(x'\gamma)$ with $L(a) = \exp(a)/(1 + \exp(a))$. The following assumptions gives the new bandwidth conditions.

Assumption B.5 *Assume that*

1. *The bandwidth satisfies that* $h_{FS} \to 0$, $nh_{FS} \to \infty$ *and* $nh^7_{FS} \to 0$ *as* $n \to \infty$.

2. $h_n \to 0$, $nh_n \to \infty$ *and* $nh_n^7 \to 0$*;*

3. $b_n \to 0$, $nb_n \to \infty$ *and* $nb_n^7 \to 0$*;*

4. $h_n/h_{FS} \to 0$ *and* $b_n/h_{FS} \to 0$*.*

Redefine the influence function

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

for $d = 0, 1$. The following lemma then provides asymptotic properties of the first-step local MLE estimators under the new local quadratic regression set-up.

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

$$\begin{aligned}\sqrt{nh_{FS}}\begin{pmatrix} \hat{\gamma}_{FS}^d - \gamma_{FS}^d \\ h_{FS}\hat{\beta}_{FS}^d - h_{FS}\beta_{FS}^d \\ h_{FS}^2\hat{\rho}_{FS}^d - h_{FS}^2\rho_{FS}^d \end{pmatrix} =& \frac{1}{\sqrt{nh_{FS}}}\sum_{i=1}^n (\Delta^d)^{-1} S_{2i} \cdot 1(Z_{1i} \geq 0)^d \cdot 1(Z_{1i} < 0)^{1-d} \\ & \cdot K\left(\frac{Z_{1i}}{h_{FS}}\right)\left(D_{2i} - L\big(X_i'(\gamma_{FS}^d + \beta_{FS}^d Z_{1i} + \rho_{FS}^d Z_{1i}^2)\big)\right)\begin{pmatrix} X_i \\ \frac{X_i Z_{1i}}{h_{FS}} \\ \frac{X_i Z_{1i}^2}{h_{FS}^2}\end{pmatrix} + o_p(1),\end{aligned}$$

where Δ^d *is given in the proof of the lemma. In addition, for* $d = 0, 1$*,*

$$\sqrt{nh_{FS}}\begin{pmatrix} \hat{\gamma}_{FS}^d - \gamma_{FS}^d \\ h_{FS}\hat{\beta}_{FS}^d - h_{FS}\beta_{FS}^d \\ h_{FS}^2\hat{\rho}_{FS}^d - h_{FS}^2\rho_{FS}^d \end{pmatrix} \Rightarrow N\left(0, \left(\Delta^d\right)^{-1}\Omega^d\left(\Delta^d\right)^{-1}\right),$$

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

Let $\widetilde{Y}_{1,i}$ and $\widetilde{Y}_{0,i}$ be the infeasible versions of $\widehat{Y}_{1,i}$ and $\widehat{Y}_{0,i}$ such that

$$\widetilde{Y}_{1,i} = Y_{2i} - \frac{Y_{2i}S_{2i}(D_{2i} - p(X_i, \gamma_{FS}^1))}{1 - p(X_i, \gamma_{FS}^1)}, \quad \widetilde{Y}_{0,i} = Y_{2i} - \frac{Y_{2i}S_{2i}(D_{2i} - p(X_i, \gamma_{FS}^0))}{1 - p(X_i, \gamma_{FS}^0)}.$$

Given smoothness conditions on them, we can show asymptotic properties of the bias-corrected estimator $\hat{\theta}_{1,1}^{bc}(h_n, b_n)$.

Assumption B.6 *Assume that for some* $\epsilon > 0$*,*

1. $E[\widetilde{Y}_1^4|Z=z]$ *and* $E[\widetilde{Y}_0^4|Z=z]$ *are bounded on* $z \in [-\epsilon, \epsilon]$*;*

2. $E[\widetilde{Y}_1|Z=z]$ *and* $E[\widetilde{Y}_0|Z=z]$ *are three times continuously differentiable on* $z \in [-\epsilon, \epsilon]$*;*

3. $V[\widetilde{Y}_1|Z=z]$ *and* $V[\widetilde{Y}_0|Z=z]$ *are continuously on* $z \in [-\epsilon, \epsilon]$ *and bounded away from zero.*

Theorem B.1 *Suppose that Assumptions 4.1-B.2, 4.3, and B.6 hold. Then*

$$\frac{\hat{\theta}^{bc}_{1,1}(h_n, b_n) - \bar{\theta}_{1,1}}{\sqrt{V^{bc}_{1,1}(h_n, b_n)}} \xrightarrow{d} N(0,1).$$

Results in the Theorem follows directly from Theorem 1 of CCT. Proof is given in Section D. This is straightforward because Assumption B.5 implies that the first step estimation converges at a faster rate than the resulting $\hat{\theta}^{bc}_{1,1}(h_n, b_n)$ estimator and we can ignore its effect on the AMSE of the final estimator. Then by the same reasoning, we can show that the AMSE-optimal (infeasible) choices for h_n and b_n are the same as those in Lemma 1 of CCT after replacing outcomes with $\widetilde{Y}_{1,i}$ and $\widetilde{Y}_{0,i}$, respectively. The data-driven plug-in bandwidth selectors then follow directly from Section S.2.6 of the supplement material of CCT.

C Proofs for Identification Results

Proof of Lemma A.1

The identification result for ATE_0 is a standard result in static sharp RD. Also by standard static RD identification and the smoothness conditions in Assumption A.2, we know that $\lim_{z_1 \searrow 0} E[D_{1+\tau-s}|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[D_{1+\tau-s}|Z_1 = z_1] = E[\tilde{D}_{1+\tau-s}(1) - \tilde{D}_{1+\tau-s}(0)|Z_1 = 0] \equiv \pi_{\tau-s}$, for any $\tau \geq 1$ and $s = 0, 1, \cdots, \tau - 1$.

Set $\pi_0 = \lim_{z_1 \searrow 0} E[D_1|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[D_1|Z_1 = z_1] = 1$. To prove the lemma, we only need to prove that for all $\tau \geq 1$,

$$\lim_{z_1 \searrow 0} E[Y_{1+\tau}|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[Y_{1+\tau}|Z_1 = z_1] = \sum_{s=0}^{\tau} ATE_s \cdot E[\pi_{\tau-s}|Z_1 = 0].$$

We prove by induction. When $\tau = 1$, the equation implies that $\lim_{z_1 \searrow 0} E[Y_2|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[Y_2|Z_1 = z_1] = E[\tilde{\theta}_{1,1}|Z_1 = 0] = ATE_1 + ATE_0 E[\pi_1|Z_1 = 0]$, which is already shown in Section 2. Now, suppose that the equation above holds for some $k \geq 1$.

This implies that

$$
\begin{aligned}
&\sum_{s=0}^{k} ATE_s \cdot E[\pi_{k-s}|Z_1=0] = \lim_{z_1 \searrow 0} E[Y_{1+k}|Z_1=z_1] - \lim_{z_1 \nearrow 0} E[Y_{1+k}|Z_1=z_1] \\
&= \lim_{z_1 \searrow 0} E[\tilde{Y}_{1+k}(1)|Z_1=z_1] - \lim_{z_1 \nearrow 0} E[\tilde{Y}_{1+k}(0)|Z_1=z_1] \\
&= E[\tilde{Y}_{1+k}|Z_1=0] - E[\tilde{Y}_{1+k}|Z_1=0] \\
&= E\left[\sum_{\ell^k \in \mathcal{L}^k} \left(Y_{1+k}(1,\ell^k)\cdot \mathfrak{D}_{2:(1+k)}(1,\ell^k) - Y_t(0,\ell^k)\cdot \mathfrak{D}_{2:(1+k)}(0,\ell^k)\right)|Z_1=0\right], \qquad \text{(C.1)}
\end{aligned}
$$

where the second and fourth equalities hold by definitions and the third equality holds by smoothness conditions.

Now for period $k+1$, under smoothness conditions,

$$
\begin{aligned}
&\lim_{z_1 \searrow 0} E[Y_{(k+1)+1}|Z_1=z_1] - \lim_{z_1 \nearrow 0} E[Y_{(k+1)+1}|Z_1=z_1] \\
=&E\left[\sum_{\ell^k \in \mathcal{L}^k} Y_{(k+1)+1}(1,\ell^k,1)\mathfrak{D}_{2:(k+1)}(1,\ell^k)D_{(k+1)+1}(1,\ell^k) + Y_{(k+1)+1}(1,\ell^k,0)\mathfrak{D}_{2:(k+1)}(1,\ell^k)(1-D_{(k+1)+1}(1,\ell^k))|Z_1=0\right] \\
&-E\left[\sum_{\ell^k \in \mathcal{L}^k} Y_{(k+1)+1}(0,\ell^k,1)\mathfrak{D}_{2:(k+1)}(0,\ell^k)D_{(k+1)+1}(0,\ell^k) + Y_{(k+1)+1}(0,\ell^k,0)\mathfrak{D}_{2:(k+1)}(0,\ell^k)(1-D_{(k+1)+1}(0,\ell^k))|Z_1=0\right] \\
=&E\left[\sum_{\ell^k \in \mathcal{L}^k} \left(Y_{(k+1)+1}(1,\ell^k,0)\mathfrak{D}_{2:(k+1)}(1,\ell^k) - Y_{(k+1)+1}(0,\ell^k,0)\mathfrak{D}_{2:(k+1)}(0,\ell^k)\right)|Z_1=0\right] \\
&+E\left[\sum_{\ell^k \in \mathcal{L}^k} \theta^{(1,\ell^k)}_{0,(k+1)+1}\cdot \mathfrak{D}_{2:(k+1)+1}(1,\ell^k,1)|Z_1=0\right] - E\left[\sum_{\ell^k \in \mathcal{L}^k} \theta^{(0,\ell^k)}_{0,(k+1)+1}\cdot \mathfrak{D}_{2:(k+1)+1}(0,\ell^k,1)|Z_1=0\right] \\
=&A + ATE_0 \cdot E\left[\sum_{\ell^k \in \mathcal{L}^k} \mathfrak{D}_{2:(k+1)+1}(1,\ell^k,1)|Z_1=0\right] - ATE_0 \cdot E\left[\sum_{\ell^k \in \mathcal{L}^k} \mathfrak{D}_{2:(k+1)+1}(0,\ell^k,1)|Z_1=0\right] \\
\equiv&A + ATE_0 \cdot E\left[\tilde{D}_{(k+1)+1}(1) - \tilde{D}_{(k+1)+1}(0)|Z_1=0\right] = A + ATE_0 \cdot E[\pi_{k+1}|Z_1=0]. \qquad \text{(C.2)}
\end{aligned}
$$

The first two equalities hold by definitions. The third equality holds by the mean equalivance condition and random treatment selection condition in (A.1) and (A.2). The last two equalities again hold by definitions.

Now note that the only difference between the A term above and the conditional mean expression in the right hand side of equation (C.1) is between the quasi potential outcome $Y_{(k+1)+1}(d_1,\ell^k,0)$ in (C.2) and the potential outcome $Y_{k+1}(d_1,\ell^k)$ in (C.1). Given the definition of direct treatment effects, it is clear that

$$
A = \sum_{s=0}^{k} ATE_{s+1} \cdot E[\pi_{k-s}|Z_1=0] = \sum_{s=1}^{k+1} ATE_s \cdot E[\pi_{(k+1)-s}|Z_1=0]
$$

Plugging the result into equation (C.2) therefore completes the proof by showing that

$$
\lim_{z_1 \searrow 0} E[Y_{1+(k+1)}|Z_1=z_1] - \lim_{z_1 \nearrow 0} E[Y_{1+(k+1)}|Z_1=z_1] = \sum_{s=0}^{k+1} ATE_s \cdot E[\pi_{(k+1)-s}|Z_1=0].
$$

Proof of Lemma A.2

From Equation (2.1) and the proof of Lemma A.1, we know that

$$
\begin{aligned}
ATE_1 = & \lim_{z_1 \searrow 0} E[Y_2|Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[Y_2|Z_1 = z_1] \\
& - E\left[E[\theta^1_{0,2} D_2(1) - \theta^0_{0,2} D_2(0)|X, Z_1 = 0]|Z_1 = 0\right],
\end{aligned}
$$

where

$$
\begin{aligned}
& E[\theta^1_{0,2} D_2(1) - \theta^0_{0,2} D_2(0)|X, Z_1 = 0] \\
= & E[\theta^1_{0,2}|D_2(1) = 1, X, Z_1 = 0]E[D_2(1)|X, Z_1 = 0] \\
& - E[\theta^0_{0,2}|D_2(0) = 1, X, Z_1 = 0]E[D_2(0)|X, Z_1 = 0] \\
= & E[\theta^1_{0,2}|X, Z_1 = 0]E[D_2(1)|X, Z_1 = 0] - E[\theta^0_{0,2}|X, Z_1 = 0]E[D_2(0)|X, Z_1 = 0] \\
= & CATE_0(X)\left(E[D_2(1) - D_2(0)|X, Z_1 = 0]\right).
\end{aligned}
$$

The second equality is by the extended random treatment selection assumption in (2.6) and the third equality is by the extended homogeneous ATE assumption in (2.7). Then, by the strengthened smoothness and overlapping conditions in Assumption A.2, we know that $CATE_0(X) = \lim_{z_1 \searrow 0} E[Y_1|X, Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[Y_1|X, Z_1 = z_1]$ and $E[D_2(1) - D_2(0)|X, Z_1 = 0] = \lim_{z_1 \searrow 0} E[D_2|X, Z_1 = z_1] - \lim_{z_1 \nearrow 0} E[D_2|X, Z_1 = z_1]$. The Lemma is hence proven.

Proof of Lemma 2.2

By the definition of potential propensity scores, we have that

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

Under the CIA condition in Assumption 2.2, we also have that,

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

Then,

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

The first and third equalities holds by the law of iterated expectations. The second holds by plugging in the equations shown above. The fourth equality holds by smoothness conditions in Assumption 2.2.2.

Similarly, we have 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 (2.2) proves the lemma.

Proof of Lemma 3.1

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

$$\begin{aligned}Y_{k+\tau}(\ell^{k-1},1,\eta)-Y_{k+\tau}(\ell^{k-1},0,\eta)&=Y_{k+\tau}(\ell^{k-1},1,\mathbf{0}_\tau)-Y_{k+\tau}(\ell^{k-1},0,\mathbf{0}_\tau)\\&\quad+Y_{k+\tau}(\ell^{k-1},1,\eta)-Y_{k+\tau}(\ell^{k-1},1,\mathbf{0}_\tau)-\Big(Y_{k+\tau}(\ell^{k-1},0,\eta)-Y_{k+\tau}(\ell^{k-1},0,\mathbf{0}_\tau)\Big)\\&=\theta^{\ell_{k-1}}_{\tau,k}+\theta^{(\ell^{k-1},1,\mathbf{0}_{s-1})}_{\tau-s,k+s}-\theta^{(\ell^{k-1},0,\mathbf{0}_{s-1})}_{\tau-s,k+s}.\end{aligned}\tag{C.3}$$

is a linear combination of long-term direct effects (or immediate effects and long-term direct effects when $\tau=s$).

If all but the s-th and s'-th elements of η are zero, $s < s'$, then

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

where η' is a vector whose s-th element is one and all other elements are zero. The first difference in the right hand side is between a pair of potential outcomes discussed in (C.3). The other two differences are direct effects (or immediate effects when $\tau = s'$). 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 combinations of immediate effects and long-term direct effects.

Now consider pairs of potential outcomes with two flipped treatments. It is easy to see that such differences, for example, $Y_{k+\tau}(\ell^{k-1}, 1, \eta, 1, \rho) - Y_{k+\tau}(\ell^{k-1}, 0, \eta, 0, \rho)$, could be represented by a linear combination of differences of potential outcomes with only one flipped treatment status, which has been discussed above, and therefore could eventually be represented by a linear combination of immediate effects and long-term direct effects. Similarly, the difference of potential outcomes with three or more flipped treatment status could be defined by a linear combination of immediate effects and long-term direct effects. This completes the proof.

Proof of Equation (3.1)

Set $\tilde{\theta}_{0,k}^{\ell^{k-1}} = \theta_{0,k}^{\ell^{k-1}}$ in this proof for nation simplicity. (There is no need to differentiate direct immediate effect and total immediate effect.) First we notice that for both $d_1 =$

$0, 1$, the quasi-potential outcome $\tilde{Y}_{\tau+1}(d_1)$ could be decomposed as the following.

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

Then, it is clear that for all $\tau = 2, ..., K - 1$,

$$
\begin{aligned}
& \tilde{\theta}_{\tau,1} = \tilde{Y}_{\tau+1}(1) - \tilde{Y}_{\tau+1}(0) \\
=& \theta_{\tau,1} + \left(\tilde{\theta}^{1}_{\tau-1,2} \cdot D_2(1) - \tilde{\theta}^{0}_{\tau-1,2} \cdot D_2(0) \right) \\
& + \sum_{s=0}^{\tau-2} \left(\tilde{\theta}^{(1,\mathbf{0}_{\tau-1-s})}_{s,\tau+1-s} \cdot D_{\tau+1-s}(1, \mathbf{0}_{\tau-1-s}) - \tilde{\theta}^{(0,\mathbf{0}_{\tau-1-s})}_{s,\tau+1-s} \cdot D_{\tau+1-s}(0, \mathbf{0}_{\tau-1-s}) \right).
\end{aligned}
$$

This completes the proof of Equation (3.1).

Proof of Lemma 3.2

Combining the decomposition in Equation (3.1) and the Markovian condition in Assumption 3.1, we have that

$$
\begin{aligned}
E[\tilde{\theta}_{\tau,1}|Z_1=0] =& E[\theta_{\tau,1}|Z_1=0] + E\left[\tilde{\theta}^1_{\tau-1,2}|D_2(1)=1, Z_1=0\right]\cdot P[D_2(1)=1|Z_1=0] \\
&- E\left[\tilde{\theta}^0_{\tau-1,2}|D_2(0)=1, Z_1=0\right]\cdot P[D_2(0)=1|Z_1=0] \\
&+ \sum_{s=0}^{\tau-2} E\left[\tilde{\theta}^{(1,\mathbf{0}_{\tau-1-s})}_{s,\tau+1-s}|D_{\tau+1-s}(1,\mathbf{0}_{\tau-1-s})=1, Z_1=0\right] P\left[D_{\tau+1-s}(1,\mathbf{0}_{\tau-1-s})=1|Z_1=0\right] \\
&- \sum_{s=0}^{\tau-2} E\left[\tilde{\theta}^{(0,\mathbf{0}_{\tau-1-s})}_{s,\tau+1-s}|D_{\tau+1-s}(0,\mathbf{0}_{\tau-1-s})=1, Z_1=0\right] P\left[D_{\tau+1-s}(0,\mathbf{0}_{\tau-1-s})=1|Z_1=0\right] \\
=& E[\theta_{\tau,1}|Z_1=0] + E\left[\tilde{\theta}^1_{\tau-1,2}|D_2(1)=1, Z_1=0\right]\cdot P[D_2(1)=1|Z_1=0] \\
&- E\left[\tilde{\theta}^0_{\tau-1,2}|D_2(0)=1, Z_1=0\right]\cdot P[D_2(0)=1|Z_1=0] \\
&+ \sum_{s=0}^{\tau-2} E\left[\tilde{\theta}^0_{s,2}|D_2(0)=1, Z_1=0\right] P\left[D_{\tau+1-s}(1,\mathbf{0}_{\tau-1-s})=1|Z_1=0\right] \\
&- \sum_{s=0}^{\tau-2} E\left[\tilde{\theta}^0_{s,2}|D_2(0)=1, Z_1=0\right] P\left[D_{\tau+1-s}(0,\mathbf{0}_{\tau-1-s})=1|Z_1=0\right] \\
=& E[\theta_{\tau,1}|Z_1=0] + \tilde{\mu}^1_{\tau-1}\cdot E[D_2(1)|Z_1=0] - \tilde{\mu}^0_{\tau-1}\cdot E[D_2(0)|Z_1=0] \\
&+ \sum_{s=0}^{\tau-2} \tilde{\mu}^0_s \cdot E\left[\eta_{\tau-1-s,1}|Z_1=0\right].
\end{aligned} \tag{C.4}
$$

By smoothness conditions in Assumption 2.1, the left hand side could be identified as $\lim_{z_1\searrow 0} E[Y_{s+1}|Z_1=z_1] - \lim_{z_1\nearrow 0} E[Y_{s+1}|Z_1=z_1]$ while $E[D_2(1)|Z_1=0] = \lim_{z_1\searrow 0} E[D_2|Z_1=z_1]$, and $E[D_2(0)|Z_1=0] = \lim_{z_1\nearrow 0} E[D_2|Z_1=z_1]$ in the right hand side. Meanwhile, since

$$
E\left[\tilde{\theta}^d_{s,2}|D_2(d)=1, Z_1=0\right] = E\left[\tilde{Y}_{2+s}(d,1) - \tilde{Y}_{2+s}(d,0)|D_2(d)=1, Z_1=0\right],
$$

it can be identified following the same steps in the proof of Lemma 2.2, by treating $\tilde{Y}_{2+s}(d,0)$ and $\tilde{Y}_{2+s}(d,1)$ as potential second-period outcomes. Then, we have that $\tilde{\mu}^0_s = \lim_{z_1\nearrow 0} E\left[\frac{Y_{2+s}S_2(D_2-\lambda^0(X))}{(1-\lambda^0(X))E[D_2|Z_1=z_1]}|Z_1=z_1\right]$ and $\tilde{\mu}^1_s = \lim_{z_1\searrow 0} E\left[\frac{Y_{2+s}S_2(D_2-\lambda^1(X))}{(1-\lambda^1(X))E[D_2|Z_1=z_1]}|Z_1=z_1\right]$ for all $s=0,...,\tau-1$.

Lastly, we notice that $\eta_{\tau-1-s,1} = D_{\tau+1-s}(1,\mathbf{0}_{\tau-1-s}) - D_{\tau+1-s}(0,\mathbf{0}_{\tau-1-s})$ is a direct effect, viewing a subsequent treatment decision as an outcome. Then, applying the identification result in Lemma 2.2, we know that

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

In addition, for all $\tau = 2, ..., K-2$, following the identification in equation (C.4) and viewing a subsequent treatment decision as an outcome,

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

could be identified recursively, where $\tilde{\nu}^0_s = \lim_{z_1 \nearrow 0} E\left[\frac{D_{3+s} S_2(D_2-\lambda^0(X))}{(1-\lambda^0(X))E[D_2|Z_1=z_1]}\Big|Z_1=z_1\right]$ and $\tilde{\nu}^1_s = \lim_{z_1 \searrow 0} E\left[\frac{D_{3+s} S_2(D_2-\lambda^1(X))}{(1-\lambda^1(X))E[D_2|Z_1=z_1]}\Big|Z_1=z_1\right]$ for all $s = 0, ..., \tau-1$.

Plugging in all pieces to equation (C.4) completes the proof of Lemma 3.2.

D Proofs for Inference Results

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

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

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

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

Next, we omit subscript i when there is no confusion for notational simplicity. 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,

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

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

$$E[\Delta_n] = \Delta_z \otimes E\left[S_2 L(\eta(Z_1, X))(1 - L(\eta(Z_1, X)))XX' \middle| Z_1 = 0\right] + o(1) \equiv \Delta + o(1). \tag{D.2}$$

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 4.3(iii) implies that $O(\sqrt{nh^5}) = o(1)$.

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

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

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 B.2. 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 B.3 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 for $j = 1, \ldots, k$

$\tilde{\alpha}_c = (\tilde{\alpha}_{c,1}, \ldots, \tilde{\alpha}_{c,1})'$,

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

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 ∇_γ^1 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 4.2. Then it is true that

$$\sqrt{nh}(\hat{\alpha}^1-\alpha^1) = \frac{1}{\sqrt{nh}}\sum_{i=1}^n (1\ 0)\Delta_z^{-1}1(Z_{1i}\geq 0)K\left(\frac{Z_{1i}}{h}\right)\Big(Y_{2i} - E[Y_{2i}|Z_{1i}]$$

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

Similarly, we have

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

These results are enough to derive the asymptotic normality of $\hat{\alpha}^1$ and $\hat{\alpha}^0$ since $\hat{\alpha}^1$ and $\hat{\alpha}^0$ are mutually independent.

Proof of Theorem 4.2

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

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

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

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

Note that

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

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.

Proof of Lemma B.1

Recall that for the alternative estimation and inference procedure proposed in Section B, the first-step propensity estimation uses the local quadratic method:

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

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

$$\begin{aligned}
(\hat{\gamma}, \hat{\beta}, \hat{\rho}) = \arg\max_{\gamma,\beta,\rho} \sum_{i=1}^{n} S_{2i} 1(Z_{1i} \geq 0) K\left(\frac{Z_{1i}}{h}\right) \cdot \\
\Big[D_{2i} \log L(X_i'(\gamma + \beta Z_{1i} + \rho Z_{1i}^2)) + (1 - D_{2i}) \log\left(1 - L(X_i'(\gamma + \beta Z_{1i} + \rho Z_{1i}^2))\right)\Big] \\
\equiv \arg\max_{\gamma,\beta,\rho} \ell_n(\gamma, \beta, \rho). \qquad \text{(D.4)}
\end{aligned}$$

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

$$\begin{aligned}
&\gamma^* = \sqrt{nh}(\gamma - \gamma^1), \quad \beta^* = \sqrt{nh}(h\beta - h\beta^1), \quad \rho^* = \sqrt{nh}(h^2\rho - h^2\rho^1)\\
&\hat{\gamma}^* = \sqrt{nh}(\hat{\gamma}^1 - \gamma^1), \quad \hat{\beta}^* = \sqrt{nh}(h\hat{\beta}^1 - h\beta^1), \quad \hat{\rho}^* = \sqrt{nh}(h^2\hat{\rho}^1 - h^2\rho^1)\\
&\theta = ((\gamma^*)', (\beta^*)', (\rho^*)')', \quad \hat{\theta} = ((\hat{\gamma}^*)', (\hat{\beta}^*)', (\hat{\rho}^*)')',\\
&\widetilde{X}_i = \left(X_i' \;\; \frac{Z_{1i}X_i'}{h} \;\; \frac{Z_{1i}^2X_i'}{h^2}\right)', \quad \delta_n = \frac{1}{\sqrt{nh}}, \quad \eta(z,x) = (\gamma^1 + \beta^1 z + \rho^1 z^2)'x.
\end{aligned}$$

Therefore, we have that

$$\begin{aligned}
(\gamma + \beta Z_{1i} + \rho Z_{1i}^2)'X_i &= (\gamma^1 + \beta^1 Z_{1i} + \rho^1 Z_{1i}^2)'X_i + \delta_n((\gamma^*)'X_i + (\beta^*)'\frac{Z_{1i}X_i}{h} + (\rho^*)'\frac{Z_{1i}^2X_i}{h^2})\\
&= \eta(Z_{1i}, X_i) + \delta_n\theta'\widetilde{X}_i,
\end{aligned}$$

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 $(\hat{\gamma}', \hat{\beta}', \hat{\rho}')'$ maximizes $\ell_n(\gamma, \beta, \rho)$, 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

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

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

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

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,

$$\ell_n^*(\theta) = Q_n' \theta - \frac{1}{2}\theta' \Delta_n \theta + o_p(1), \text{ where}$$

$$Q_n = \delta_n \sum_{i=1}^{n} S_{2i} 1(Z_{1i} \geq 0) K\left(\frac{Z_{1i}}{h}\right) \cdot (D_{2i} - L(\eta(Z_{1i}, X_i)))\widetilde{X}_i,$$

$$\Delta_n = \delta_n^2 \sum_{i=1}^{n} S_{2i} 1(Z_{1i} \geq 0) K\left(\frac{Z_{1i}}{h}\right) \cdot L(\eta(Z_{1i}, X_i))(1 - L(\eta(Z_{1i}, X_i)))\widetilde{X}_i \widetilde{X}_i'.$$

For the term Δ_n, we have that

$$E[\Delta_n] = \frac{1}{h} E\left[S_2 1(Z_1 \geq 0) K\left(\frac{Z_1}{h}\right) L(\eta(Z_1, X))(1 - L(\eta(Z_1, X))) \begin{pmatrix} X \\ \frac{Z_1 X}{h} \\ \frac{Z_1^2 X}{h^2} \end{pmatrix} \begin{pmatrix} X' & \frac{Z_1 X'}{h} & \frac{Z_1^2 X'}{h^2} \end{pmatrix}\right].$$

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

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

Let

$$\Delta_z = f_z(0) \cdot \begin{pmatrix} \mu_{z,0} & \mu_{z,1} & \mu_{z,2} \\ \mu_{z,1} & \mu_{z,2} & \mu_{z,3} \\ \mu_{z,2} & \mu_{z,3} & \mu_{z,4} \end{pmatrix} \text{ with } \mu_{z,j} = \int_{u \geq 0} u^j K(u) du, \text{ for } j = 0, 1, \ldots.$$

Then, we again have

$$E[\Delta_n] = \Delta_z \otimes E\left[S_2 L(\eta(Z_1, X))(1 - L(\eta(Z_1, X)))XX' \Big| Z_1 = 0\right] + o(1) \equiv \Delta + o(1),$$

and 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

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

and

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

For the term Q_n, we have that

$$
\begin{aligned}
E[Q_n] &= n\delta_n E\left[S_2 1(Z_1 \geq 0) K\left(\frac{Z_1}{h}\right) \cdot (D_2 - L(\eta(Z_1, X)))\widetilde{X}\right] \\
&= n\delta_n E\left[E[S_2|X, Z_1] 1(Z_1 \geq 0) K\left(\frac{Z_1}{h}\right) \cdot (E[D_2|S_2 = 1, X, Z_1] - L(\eta(Z_1, X)))\widetilde{X}\right] \\
&= n\delta_n E\left[E[S_2|X, Z_1] 1(Z_1 \geq 0) K\left(\frac{Z_1}{h}\right) \cdot (L(\gamma(Z_1)'X) - L(\eta(Z_1, X)))\widetilde{X}\right] \\
&= O(n\delta_n h \cdot h^3) = O(\sqrt{nh^7}) = o(1).
\end{aligned}
$$

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 $\gamma(Z_1)'X = (\gamma^1 + \beta^1 Z_1 + \rho^2 Z_1^2 + \gamma'''(\bar{Z}_1) Z_1^3)'X$ where $\bar{Z}_1$ is between 0 and Z_1. Therefore, $\gamma(Z_1)'X - \eta(Z_1, X)' = \gamma'''(\bar{Z}_1)' Z_1^3 X$. Therefore, $L(\gamma(Z_1)'X) - L(\eta(Z_1, X)) = O_p(Z_1^3)$. Given that $K(Z_1/h)$ is non-zero when $|Z_1/h| \leq 1$ or equivalently, $|Z_1| \leq h$, $K\left(\frac{Z_1}{h}\right) \cdot (L(\gamma(Z_1)'X) - L(\eta(Z_1, X))) = O_p(K(Z_1/h)h^3)$. It follows that the expectation is $O(n\delta_n h \cdot h^3) = O(\sqrt{nh^7})$ and Assumption 4.3(iii) implies that $O(\sqrt{nh^7}) = o(1)$.

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

$$
\begin{aligned}
&V[Q_n] = \delta^2 n E[S_2 1(Z_1 \geq 0) K^2\left(\frac{Z_1}{h}\right) \cdot (D_2 - L(\eta(Z_1, X)))^2 \widetilde{X}\widetilde{X}'] \\
=&\frac{1}{h} E[E[S_2|Z_1, X] 1(Z_1 \geq 0) K^2\left(\frac{Z_1}{h}\right) \cdot L(\eta(Z_1, X))(1 - L(\eta(Z_1, X)))\widetilde{X}\widetilde{X}'] + O(h^3) \\
=&f_{z_1}(0) \begin{pmatrix} \nu_{0,+} & \nu_{1,+} & \nu_{2,+} \\ \nu_{1,+} & \nu_{2,+} & \nu_{3,+} \\ \nu_{2,+} & \nu_{3,+} & \nu_{4,+} \end{pmatrix} \otimes E\left[E[S_2|Z_1, X] L(\eta(Z_1, X))(1 - L(\eta(Z_1, X)))XX' \Big| Z_1 = 0\right] + O(h^3) \\
\equiv&\Omega + o(1),
\end{aligned}
$$

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 B.2. It then follows that $Q_n \xrightarrow{d} (0, \Omega)$ and $\hat{\theta} \xrightarrow{d} (0, \Delta^{-1}\Omega\Delta^{-1})$.

E Monte Carlo Simulations

We first use the simple case of $T = 2$ to showcase the advantages of our proposed estimator compared to the recursive CFR strategy.

We use six data-generating processes (DGPs). DGP 1 illustrates a case where individual treatment effects are non-random 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 non-random but path-dependent. DGP 3 modifies DGP 1 by simulating individual treatment effects as random variables. In DGP 3, individual treatment effects are still independent of treatment decisions of later rounds. DGP 4 modifies DGP 3 by adding a correlation between the second-round immediate treatment effect and the second-round RD participation decision. In DGP 5, potential second-period outcomes *with* second-round treatments are designed to be correlated with the second-round running variable. In DGP 6, potential second-period outcomes *without* second-round treatments are designed to be correlated with the second-round running variable. As is stated in the CIA condition of Assumption 2.2, the case of DGP 5 is compatible with the proposed estimation procedure described in Lemma 2.2, while the case of DGP 6 is not. Summing up, the proposed estimation strategy described in Lemma 2.2 is valid under DGPs 1-5. The recursive CFR estimator described in Lemma 2.1 is only valid under DGPs 1 and 3. Under DGP 6, both estimation strategies are invalid.

For all DGPs, we first simulate random variables

$$
\begin{aligned}
&X \sim U[0,10],\ Z_1 \sim X - 10 \cdot Beta(2,2),\\
&u_{y1}, u_{y2}, u_{s2}, a_s \sim N(0,0.5),\ v_{z2} \sim logis(0,1),
\end{aligned}
$$

all independent of each other. Then we simulate potential random variables following

$$
\begin{aligned}
&Y_1(0) = 0.1X + 0.5Z_1 + 0.1XZ_1 + 0.1Z_1^2 + u_{y1},\\
&S_2(0) = 1(u_{s2} + a_s \geq 0),\ S_2(1) = 1(1 + u_{s2} + a_s \geq 0),\\
&Z_2(0) = 0.3 + 0.1X + v_{z2},\ Z_2(1) = Z_2(0) + (1\ X)\gamma_0,\ \gamma_0 = (-0.4\ \ -0.2)',\\
&Y_2(0,0) = 0.1X + 0.5Z_1 + 0.1XZ_1 + 0.1Z_1^2 + u_{y2},\\
&Y_1(1) = Y_1(0) + \theta_{0,1},\ Y_2(0,1) = Y_2(0,0) + \theta_{0,2}^0,\\
&Y_2(1,0) = Y_2(0,0) + \theta_{1,1},\ Y_2(1,1) = Y_2(0,0) + \theta_{1,1} + \theta_{0,2}^1,
\end{aligned}
$$

with treatment and first-stage effect parameters $\theta_{0,1}$, $\theta_{0,2}^0$, $\theta_{0,2}^1$, and $\theta_{1,1}$ varying across different DGPs.

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

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: $\theta_{0,1} = 0.5$, $\theta_{0,2}^0 = \theta_{0,2}^1 = 0.5 + e$, $\theta_{1,1} = 0.2 + e$, $e \sim N(0, 0.5)$.

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

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

DGP 6: $\theta_{0,1} = 0.5$, $\theta_{0,2}^0 = \theta_{0,2}^1 = 0.5 + v_{z2}$, $\theta_{1,1} = 0.2 + v_{z2}$.

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 2. 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 4.1. 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" (Calonico et al., 2015). Different k choices are used to examine the robustness of proposed estimators with respect to bandwidth choice.

Table A1 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-step local likelihood estimation. Under DGPs 2, 4, and 5, the recursive CFR estimator does not center around the true value 0.2, while the proposed estimator still performs well. Under DGP 6, neither estimators have correct centering.

Table A2 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 that for all DGPs that are compatible with the proposed estimation procedure, t-tests following the proposed estimators control size well under the null and have power going to one under the alternative. Choice of the under-smoothing parameter k does not seem to affect simulation results much, either, under the DGPs considered in this section.

Table A1: One-period-after ATE: Proposed Estimator vs. Recursive CFR Estimator

	Proposed Estimation Strategy						Recursive CFR					
		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	0.208	0.209	0.210	0.021	0.020	0.020	0.214	0.215	0.215	0.014	0.013	0.013
n=4000	0.205	0.206	0.207	0.010	0.009	0.009	0.212	0.213	0.214	0.007	0.007	0.007
n=8000	0.202	0.202	0.203	0.005	0.005	0.005	0.205	0.206	0.206	0.004	0.003	0.003
	DGP 2											
n=2000	0.205	0.206	0.208	0.022	0.021	0.020	0.152	0.154	0.155	0.016	0.015	0.014
n=4000	0.196	0.197	0.198	0.010	0.010	0.009	0.145	0.146	0.147	0.011	0.010	0.010
n=8000	0.203	0.205	0.205	0.006	0.005	0.005	0.151	0.153	0.154	0.007	0.006	0.006
	DGP 3											
n=2000	0.209	0.209	0.210	0.026	0.025	0.024	0.211	0.212	0.212	0.021	0.020	0.019
n=4000	0.199	0.200	0.201	0.013	0.012	0.012	0.203	0.205	0.206	0.012	0.011	0.011
n=8000	0.203	0.203	0.204	0.006	0.006	0.006	0.205	0.206	0.206	0.005	0.005	0.005
	DGP 4											
n=2000	0.202	0.204	0.205	0.021	0.020	0.019	0.166	0.168	0.169	0.017	0.016	0.016
n=4000	0.200	0.200	0.201	0.010	0.010	0.010	0.169	0.170	0.171	0.009	0.008	0.008
n=8000	0.204	0.205	0.205	0.005	0.005	0.005	0.172	0.173	0.173	0.005	0.004	0.004
	DGP 5											
n=2000	0.203	0.204	0.205	0.021	0.019	0.019	0.076	0.077	0.077	0.035	0.034	0.033
n=4000	0.203	0.204	0.205	0.010	0.010	0.010	0.069	0.071	0.071	0.027	0.026	0.026
n=8000	0.203	0.203	0.204	0.005	0.005	0.005	0.066	0.067	0.068	0.023	0.023	0.022
	DGP 6											
n=2000	-0.004	-0.003	-0.002	0.073	0.071	0.070	0.062	0.062	0.062	0.059	0.057	0.056
n=4000	0.0004	0.002	0.003	0.055	0.053	0.053	0.068	0.069	0.070	0.038	0.036	0.036
n=8000	-0.004	-0.002	-0.002	0.049	0.048	0.047	0.065	0.066	0.067	0.028	0.027	0.027

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

Table A2: 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	0.060	0.060	0.058	0.069	0.068	0.069	0.634	0.670	0.683	0.353	0.379	0.388
n=4000	0.041	0.041	0.042	0.047	0.045	0.043	0.878	0.893	0.902	0.553	0.585	0.604
n=8000	0.049	0.050	0.055	0.044	0.044	0.045	0.982	0.987	0.990	0.786	0.803	0.817
	DGP 2											
n=2000	0.056	0.054	0.056	0.079	0.077	0.072	0.592	0.617	0.629	0.356	0.380	0.380
n=4000	0.057	0.051	0.056	0.057	0.055	0.050	0.823	0.848	0.854	0.537	0.557	0.562
n=8000	0.063	0.063	0.058	0.064	0.063	0.069	0.973	0.984	0.985	0.793	0.826	0.834
	DGP 3											
n=2000	0.052	0.048	0.050	0.072	0.075	0.076	0.471	0.491	0.499	0.314	0.331	0.339
n=4000	0.050	0.051	0.054	0.066	0.066	0.066	0.724	0.756	0.768	0.473	0.499	0.501
n=8000	0.050	0.051	0.050	0.050	0.048	0.050	0.938	0.949	0.952	0.734	0.764	0.771
	DGP 4											
n=2000	0.061	0.064	0.063	0.080	0.078	0.071	0.591	0.627	0.638	0.369	0.393	0.395
n=4000	0.056	0.052	0.054	0.063	0.065	0.063	0.846	0.872	0.886	0.535	0.557	0.575
n=8000	0.060	0.060	0.063	0.046	0.050	0.047	0.985	0.990	0.991	0.814	0.834	0.844
	DGP 5											
n=2000	0.058	0.058	0.055	0.075	0.078	0.077	0.602	0.623	0.636	0.345	0.358	0.366
n=4000	0.053	0.058	0.060	0.057	0.053	0.059	0.861	0.883	0.892	0.566	0.601	0.611
n=8000	0.048	0.043	0.039	0.063	0.059	0.061	0.985	0.992	0.992	0.815	0.842	0.849
	DGP 6											
n=2000	0.048	0.047	0.046	0.244	0.254	0.257	0.622	0.654	0.668	0.062	0.059	0.058
n=4000	0.050	0.054	0.061	0.387	0.400	0.404	0.860	0.873	0.887	0.048	0.049	0.050
n=8000	0.053	0.062	0.064	0.657	0.671	0.687	0.990	0.993	0.995	0.038	0.040	0.041

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.

Next, we extend DGPs 1-4 to examine small sample performances of proposed estimators of $E[\theta_{\tau,1}|Z_1=0]$ for $\tau=0,1,2,3$. For all DGPs, let

$$X \sim U[0,10],\ Z_1 \sim X - 10\cdot Beta(2,2),\ (u_{y1},u_{y2},u_{y3},u_{y4},a_s) \sim i.i.d.\ N(0,0.5),$$
$$Y_t(\mathbf{0}_t) = 0.1X + 0.5Z_1 + 0.1XZ_1 + 0.1Z_1^2 + u_{yt},\ \text{for } t=1,2,3,4.$$

Potential outcomes with nonzero treatment status are simulated different in each DGP, based on the longer-term direct treatment effect parameters specific to each DGP, as is stated in the following.

DGP 1-T4: $\theta_{0,1}=\theta_0^0=\theta_0^1=0.5$, $\theta_{1,1}=\theta_1^0=\theta_1^1=0.2$, $\theta_{2,1}=\theta_2^0=\theta_2^1=0.3$, $\theta_{3,1}=0$.

DGP 2-T4: $\theta_{0,1}=\theta_0^0=0.5$, $\theta_0^1=0.1$, $\theta_{1,1}=\theta_1^0=0.2$, $\theta_1^1=-0.2$, $\theta_{2,1}=\theta_2^0=0.3$, $\theta_2^1=-0.3$, $\theta_{3,1}=0$.

DGP 3-T4: $\theta_{0,1} = \theta_0^0 = \theta_0^1 = 0.5 + e_0$, $\theta_{1,1} = \theta_1^0 = \theta_1^1 = 0.2 + e_1$, $\theta_{2,1} = \theta_2^0 = \theta_2^1 = 0.3 + e_2$, $\theta_{3,1} = e_3$, $(e_0, e_1, e_2, e_3) \sim i.i.d.\ N(0, 0.5)$.

DGP 4-T4: $\theta_{0,1} = 0.5$, $\theta_0^0 = 0.5 + a_s$, $\theta_0^1 = 0.5$, $\theta_{1,1} = 0.2$, $\theta_1^0 = 0.2 + a_s$, $\theta_1^1 = 0.2$, $\theta_{2,1} = 0.3$, $\theta_2^0 = 0.3 + a_s$, $\theta_2^1 = 0.3$, $\theta_{3,1} = 0$.

Note that path-dependency in direct effects are restricted with the same Markovian assumption in Assumption 3.1. For example, $Y_2(1,1)$ is simulated using the fact that $Y_2(1,1) = Y_2(0,0) + [Y_2(1,0) - Y_2(0,0)] + [Y_2(1,1) - Y_2(1,0)] = Y_2(0,0) + \theta_{1,1} + \theta_0^1$, while $Y_3(1,0,1)$is simulated using the fact that $Y_3(1,0,1) = Y_3(0,0,0) + [Y_3(1,0,0) - Y_3(0,0,0)] + [Y_3(1,0,1) - Y_3(1,0,0)] = Y_3(0,0,0) + \theta_{2,1} + \theta_0^0$. The proposed estimators are valid under all four DGPs while the recursive CFR estimators are only valid under DGPs 1-T4 and 3-T4.

Meanwhile, potential running variables and RD participation decisions are simulated as following

$$
\begin{aligned}
&(v_{z1}, v_{z2}, v_{z3}, v_{z4}) \sim i.i.d.\ logis(0,1),\ Z_t(\mathbf{0}_t) = 0.3 + 0.1X + v_{zt},\ \text{for } t = 2,3,4,\\
&Z_2(1) = Z_2(0) + (1\ X)\gamma_0,\ Z_3(0,1) = Z_3(\mathbf{0}_2) + (1\ X)\gamma_0^0,\\
&Z_3(1,0) = Z_3(\mathbf{0}_2) + (1\ X)\gamma_{1,1},\ Z_3(1,1) = Z_3(\mathbf{0}_2) + (1\ X)(\gamma_{1,1} + \gamma_0^1),\\
&Z_4(0,0,1) = Z_4(\mathbf{0}_3) + (1\ X)\gamma_0^0,\ Z_4(0,1,0) = Z_4(\mathbf{0}_3) + (1\ X)\gamma_1^0,\\
&Z_4(0,1,1) = Z_4(\mathbf{0}_3) + (1\ X)(\gamma_1^0 + \gamma_0^1),\ Z_4(1,0,0) = Z_4(\mathbf{0}_3) + (1\ X)\gamma_{2,1},\\
&Z_4(1,1,0) = Z_4(\mathbf{0}_3) + (1\ X)(\gamma_{2,1} + \gamma_1^1),\ Z_4(1,0,1) = Z_4(\mathbf{0}_3) + (1\ X)(\gamma_{2,1} + \gamma_0^0),\\
&\gamma_{0,1} = (-0.3\ \ -0.1),\ \gamma_0^0 = (0.1\ 0.1),\ \gamma_0^1 = (-0.2\ \ -0.1),\\
&\gamma_{1,1} = \gamma_1^0 = \gamma_1^1 = \gamma_{2,1} = (-0.1\ \ -0.1),\\
&(u_{s1}, u_{s2}, u_{s3}, u_{s4}) \sim i.i.d.\ N(0, 0.5),\\
&S_t(0) = 1(u_{st} + a_s \geq 0),\ S_t(1) = 1(1 + u_{st} + a_s \geq 0),\ \text{for } t = 2,3,4.
\end{aligned}
$$

Table A3 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-T4 and 3-T4.

Table A4 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 A3: Performance of Proposed and CFR 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
True Parameter Values												
	0.5	0.5	0.5	0.2	0.2	0.2	0.3	0.3	0.3	0.0	0.0	0.0
Averages Across Simulations: the Proposed Strategy												
	DGP 1-T4											
n=2000	0.501	0.502	0.502	0.199	0.202	0.204	0.303	0.305	0.307	0.003	0.007	0.010
n=4000	0.504	0.505	0.506	0.193	0.194	0.196	0.299	0.301	0.302	-0.000	0.001	0.002
n=8000	0.504	0.504	0.505	0.203	0.204	0.205	0.307	0.308	0.309	0.003	0.004	0.005
	DGP 2-T4											
n=2000	0.505	0.505	0.505	0.200	0.202	0.204	0.296	0.298	0.300	0.008	0.010	0.012
n=4000	0.504	0.505	0.506	0.203	0.204	0.205	0.303	0.305	0.306	0.018	0.019	0.021
n=8000	0.504	0.505	0.505	0.203	0.204	0.205	0.301	0.302	0.302	0.015	0.015	0.016
	DGP 3-T4											
n=2000	0.512	0.512	0.513	0.205	0.207	0.210	0.300	0.302	0.304	0.008	0.011	0.013
n=4000	0.505	0.505	0.506	0.202	0.204	0.205	0.301	0.303	0.304	-0.000	0.002	0.004
n=8000	0.505	0.506	0.506	0.203	0.204	0.205	0.305	0.307	0.308	0.005	0.006	0.007
	DGP 4-T4											
n=2000	0.506	0.507	0.508	0.201	0.203	0.204	0.304	0.306	0.309	-0.001	0.003	0.006
n=4000	0.504	0.505	0.506	0.200	0.201	0.202	0.297	0.298	0.299	-0.001	0.0003	0.002
n=8000	0.506	0.507	0.507	0.202	0.203	0.204	0.306	0.308	0.309	0.013	0.014	0.014
Averages Across Simulations: the Recursive CFR Strategy												
	DGP 1-T4											
n=2000	0.501	0.502	0.502	0.205	0.206	0.207	0.312	0.313	0.313	0.013	0.015	0.016
n=4000	0.504	0.505	0.506	0.202	0.203	0.204	0.305	0.307	0.307	0.010	0.011	0.011
n=8000	0.504	0.504	0.505	0.204	0.205	0.206	0.306	0.308	0.309	0.008	0.009	0.010
	DGP 2-T4											
n=2000	0.505	0.505	0.505	0.097	0.098	0.098	0.238	0.239	0.240	-0.095	-0.094	-0.093
n=4000	0.504	0.505	0.506	0.103	0.104	0.105	0.247	0.248	0.249	-0.085	-0.084	-0.083
n=8000	0.504	0.505	0.505	0.099	0.100	0.101	0.243	0.244	0.245	-0.091	-0.090	-0.089
	DGP 3-T4											
n.2000.2	0.512	0.512	0.513	0.218	0.219	0.219	0.320	0.321	0.321	0.022	0.023	0.024
n.4000.2	0.505	0.505	0.506	0.207	0.208	0.208	0.307	0.308	0.309	0.006	0.007	0.007
n.8000.2	0.505	0.506	0.506	0.207	0.208	0.209	0.307	0.309	0.310	0.008	0.010	0.010
	DGP 4-T4											
n=2000	0.506	0.507	0.508	0.172	0.172	0.173	0.257	0.257	0.258	-0.033	-0.032	-0.031
n=4000	0.504	0.505	0.506	0.171	0.171	0.172	0.250	0.250	0.251	-0.036	-0.035	-0.035
n=8000	0.506	0.507	0.507	0.167	0.168	0.169	0.249	0.251	0.252	-0.033	-0.032	-0.031

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

Table A4: Performance of Proposed Estimators: Rejection Proportion of Two-sided Tests

	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-T4											
n=2000	0.056	0.057	0.056	0.063	0.069	0.074	0.048	0.052	0.052	0.055	0.058	0.056
n=4000	0.062	0.067	0.066	0.056	0.055	0.058	0.045	0.049	0.047	0.051	0.050	0.051
n=8000	0.046	0.043	0.046	0.065	0.062	0.058	0.052	0.050	0.052	0.050	0.052	0.049
	DGP 2-T4											
n=2000	0.053	0.054	0.052	0.067	0.066	0.068	0.057	0.056	0.052	0.064	0.064	0.063
n=4000	0.047	0.051	0.054	0.055	0.055	0.052	0.070	0.069	0.069	0.077	0.072	0.068
n=8000	0.047	0.048	0.050	0.066	0.060	0.062	0.050	0.051	0.053	0.057	0.059	0.056
	DGP 3-T4											
n=2000	0.046	0.045	0.046	0.075	0.072	0.074	0.062	0.060	0.057	0.052	0.051	0.052
n=4000	0.047	0.051	0.051	0.057	0.061	0.061	0.062	0.061	0.058	0.064	0.069	0.063
n=8000	0.052	0.054	0.055	0.053	0.049	0.049	0.060	0.061	0.058	0.058	0.060	0.059
	DGP 4-T4											
n=2000	0.039	0.041	0.042	0.062	0.069	0.067	0.071	0.070	0.075	0.056	0.055	0.058
n=4000	0.052	0.056	0.051	0.064	0.062	0.065	0.050	0.052	0.049	0.059	0.062	0.061
n=8000	0.053	0.051	0.058	0.052	0.048	0.049	0.051	0.050	0.049	0.063	0.059	0.060
					Power of Two-sided T-tests							
	DGP 1-T4											
n=2000	0.598	0.615	0.626	0.325	0.335	0.351	0.515	0.540	0.552	0.442	0.464	0.467
n=4000	0.824	0.850	0.862	0.460	0.479	0.485	0.783	0.806	0.816	0.757	0.770	0.784
n=8000	0.974	0.983	0.988	0.749	0.770	0.781	0.961	0.973	0.974	0.950	0.960	0.961
	DGP 2-T4											
n=2000	0.590	0.619	0.633	0.311	0.327	0.328	0.474	0.524	0.538	0.427	0.448	0.452
n=4000	0.856	0.885	0.889	0.509	0.541	0.555	0.791	0.807	0.824	0.716	0.730	0.729
n=8000	0.977	0.985	0.987	0.768	0.785	0.793	0.965	0.975	0.978	0.944	0.949	0.949
	DGP 3-T4											
n=2000	0.479	0.506	0.521	0.269	0.281	0.289	0.366	0.379	0.390	0.270	0.281	0.289
n=4000	0.739	0.756	0.775	0.454	0.470	0.488	0.594	0.616	0.630	0.481	0.497	0.507
n=8000	0.931	0.941	0.949	0.686	0.711	0.726	0.845	0.869	0.883	0.745	0.765	0.774
	DGP 4-T4											
n=2000	0.596	0.631	0.643	0.304	0.325	0.330	0.425	0.453	0.465	0.339	0.346	0.345
n=4000	0.838	0.862	0.877	0.506	0.530	0.540	0.661	0.688	0.696	0.602	0.616	0.621
n=8000	0.983	0.989	0.990	0.760	0.802	0.808	0.907	0.926	0.929	0.813	0.831	0.846

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.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

KENNEDY, E., Z. MA, M. MCHUGH, AND D. SMALL (2017): "Non-parametric methods for doubly robust estimation of continuous treatment effects," *Journal of the Royal Statistical Society Series B: Statistical Methodology*, 79(4), 1229–1245.

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

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

LV, X., X.-R. SUN, Y. LU, AND R. LI (2019): "Nonparametric identification and estimation of dynamic treatment effects for survival data in a regression discontinuity design," *American Economic Journal: Applied Economics*, 184, 108665.

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

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

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

MURPHY, S., VAN DER LAAN, J. M.J., ROBINS, AND C. P. P. R. GROUP (2001): "Marginal mean models for dynamic regimes," *Journal of the American Statistical Association*, 96(456), 1410–1423.

PAPAY, JOHN P.AND WILLETT, J. B. AND R. J. MURNANE (2011): "Extending the regressiondiscontinuity approach to multiple assignment variables," *Journal of Econometrics*, 161, 203–207.

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

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

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

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

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

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

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

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