Dynamic Regression Discontinuity:
An Event-Study Approach[*]

Francesco Ruggieri[†]

March 20, 2025[‡]

Abstract

I propose a novel argument to identify economically interpretable intertemporal treatment effects in dynamic regression discontinuity designs (RDDs). Specifically, I develop a dynamic potential outcomes model and reformulate two assumptions from the difference-in-differences literature—no anticipation and common trends—to attain point identification of cutoff-specific impulse responses. The estimand of each target parameter can be expressed as the sum of two static RDD contrasts, thereby allowing for nonparametric estimation and inference with standard local polynomial methods. I also propose a nonparametric approach to aggregate treatment effects across calendar time and treatment paths, leveraging a limited path independence restriction to reduce the dimensionality of the parameter space. I apply this method to estimate the dynamic effects of school district expenditure authorizations on housing prices in Wisconsin.

1   Introduction

Regression discontinuity designs (RDDs) are commonly used to identify causal parameters when the probability of exposure to a treatment changes discontinuously at a known deterministic threshold. In a seminal contribution, Cellini, Ferreira and Rothstein (2010) extended the canonical design to multi-period settings in which treatment assignment is sequential and exhibits path dependence. Dynamic RDDs have proven popular in the empirical public finance literature that exploits local referenda to estimate the effect of increased government expenditure on various outcomes (Darolia 2013, Hong and Zimmer 2016, Martorell, Stange and McFarlin 2016, Abott et al. 2020, Baron 2022, Enami, Reynolds and Rohlin 2023, Baron, Hyman and Vasquez 2024, Biasi, Lafortune and Schönholzer 2025).

Despite their growing relevance in empirical research, dynamic RDDs are methodologically understudied. This is especially true in relation to the challenge posed by the identification of economically interpretable long-term causal parameters. Because units are repeatedly exposed to treatment assignment and the outcome at any point in time may reflect the contribution of past and future treatment states, disentangling effects associated with sequential RD settings is inherently difficult. In recent work, Hsu and Shen (2024) addresses this challenge in a potential outcomes framework that allows for flexible patterns of path dependence and heterogeneity in treatment effects. To identify the intertemporal effects of treatment exposure in a given period, they impose a selection-on-observables assumption that exploits future realizations of the running variable and, potentially, additional covariates.

In this paper, I consider a dynamic model in which potential outcomes at any point in time depend on past, contemporaneous, and future treatment states (Robins 1986). I subsequently propose an identification argument that does not incorporate external observables and formalizes a connection between dynamic RDDs and event-study designs—one that has been implicitly recognized in recent empirical work (Biasi, Lafortune and Schönholzer 2025). Specifically, I reformulate two assumptions from the difference-in-differences literature—no anticipation and common trends—and show that they are sufficient to point identify interpretable causal parameters. The resulting, time-specific estimands can be expressed as sums of two standard RD outcome contrasts and can be aggregated, for interpretability and exposition, using a set of nonnegative weights that add to one. Finally, recognizing that the

number of estimands grows exponentially with the number of time periods in the model, I reduce the dimensionality of the parameter space with a limited path independence assumption. This restriction takes advantage of the sparse and cyclical nature of treatment assignment that is typically true of empirical applications of dynamic RDDs. A more detailed comparison between the approach developed in this paper and that of Hsu and Shen (2024) is provided in Section 3.6.

For nonparametric estimation and statistical inference, I use local polynomial methods that are widely applied in empirical research. Specifically, I estimate time-specific parameters using local linear regression and aggregate them with convex sample weights (as in Callaway and Sant'Anna 2021). In the process, I incorporate standard bias correction techniques to address the distortion that arises when estimating cutoff-specific parameters with observations away from the cutoff. To quantify statistical uncertainty, I follow the method developed by Calonico, Cattaneo and Titiunik (2014a) and construct nonparametric confidence intervals that account for the noise introduced by first-stage bias correction. Additionally, I derive bandwidth formulas that minimize the estimator's mean squared error, following standard optimality criteria (Imbens and Kalyanaraman 2012).

I implement the proposed method in the most common empirical application of dynamic RDDs: referenda through which school districts and other local jurisdictions authorize additional expenditures and bypass state-imposed fiscal policy constraints. I focus on Wisconsin, where such referenda are routinely held to finance both operational and capital spending (Baron 2022). The outcome of interest is housing prices in the years following referendum approval, a margin that has long been studied in public finance as a measure of how potential homebuyers value the tradeoff between improved public services and higher property taxes (see Ross and Yinger 1999 for a review of the earlier literature).

The approach developed in this paper is similar in spirit to recently proposed solutions to extrapolate effects away from the threshold in static designs only leveraging features of the joint distribution of the outcome, the running variable, and one or more cutoffs (Dong and Lewbel 2015, Bertanha and Imbens 2020, Cattaneo et al. 2021). I view this approach as offering several advantages. First, by construction, it is applicable to settings in which external information may not be available or fail to satisfy prerequisites for validity, such as observed characteristics co-determined by the treatment of interest. This concern may

2

be especially salient in dynamic settings in which units are repeatedly subject to treatment assignment[1]. Second, my proposed approach is based on well-understood identification assumptions with testable suggestive implications. Third, the method is easy to implement because the resulting estimators can be expressed as sums of standard regression discontinuity contrasts.

This paper contributes to three related strands of the econometric literature. First, it proposes a novel approach to point identify intertemporal treatment effects in dynamic RDDs. Second, it adds to the broad literature on the identification of economically interpretable dynamic effects under treatment effect heterogeneity (de Chaisemartin and D'Haultfœuille 2020, Goodman-Bacon 2021, Callaway and Sant'Anna 2021, Sun and Abraham 2021, de Chaisemartin and D'Haultfœuille 2024, Callaway, Goodman-Bacon and Sant'Anna 2024, de Chaisemartin, D'Haultfœuille and Vazquez-Bare 2024) and path dependence in treatment assignment (Heckman, Humphries and Veramendi 2016). Third, it contributes to the growing literature on the causal interpretation of standard estimands in the presence of multiple treatments (de Chaisemartin and D'Haultfœuille 2023, Goldsmith-Pinkham, Hull and Kolesár 2024) or multiple instrumental variables (Mogstad, Torgovitsky and Walters 2021, Mogstad and Torgovitsky 2024) when treatment effects are allowed to be stochastic.

The remainder of the paper is organized as follows. In Section 2, I introduce the main identification challenge and discuss previous approaches to address it. In Section 3, I develop a simple two-period regression discontinuity design and illustrate the core of my identification strategy. In Section 4, I generalize the model to a setting with multiple time periods. In Section 5, I propose an approach to aggregate time-specific estimands and reduce the dimensionality of the parameter space. In Section 6, I illustrate how to use standard local polynomial methods to estimate the target parameters and construct nonparametric confidence intervals around them. In Section 7, I conduct a Monte Carlo simulation to confirm that my proposed approach correctly recovers the parameters of interest. In Section 8, I apply my method to estimate the average effects of school district expenditure authorizations on housing prices in Wisconsin. Section 9 concludes.

---

[1]Statisticians and econometricians have long cautioned against adjusting for post-treatment observed confounders. See, for example, Rosenbaum (1984), Pearl (1995), Rosenbaum (2002), Wooldridge (2005), Angrist and Pischke (2008), Rubin (2009), Wooldridge (2010), Angrist and Pischke (2014), Imbens and Rubin (2015), Gelman, Hill and Vehtari (2021).

2 Background

In standard regression discontinuity designs, a treatment is assigned based on whether an observed running variable exceeds a known deterministic threshold. For example, Lee (2008) exploits U.S. House elections to estimate the average effect of a party's candidate narrowly winning an election on the party's subsequent victory in future elections for the same seat. Dynamic RDDs extend this framework by distinguishing between current and future treatment states, which depend on whether the running variable crosses the cutoff in each period.

In the context of Lee (2008), the outcome of a House election in year t determines political representation in both t and $t+1$, allowing for the estimation of short-term treatment effects[2]. Estimating the average effect of winning an election in year t on outcomes in years $t+2$ and beyond introduces an additional challenge. Because House elections occur every two years, a party that wins a seat in year t may lose it in year $t+2$, and a party that loses in year t may regain control. This dynamic turnover complicates the identification of interpretable long-term causal parameters. The key challenge lies in disentangling the long-term effects of the year-t election outcome from the short-term effects of the year-$t+2$ election outcome. More broadly, the recurrence of elections introduces dynamic fuzziness, as future treatment states may not align with initial treatment assignments. Figure 1 below provides a stylized representation of the main identification challenge in dynamic RDDs.

Figure 1: Stylized Representation of the Identification Challenge in Dynamic RDDs
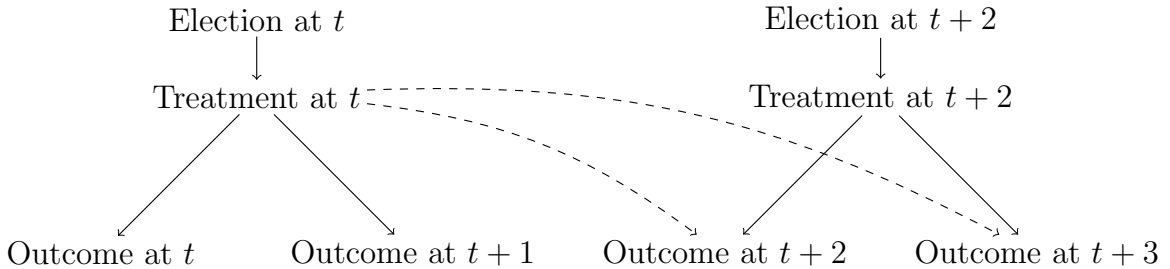
NOTES: This figure offers a stylized depiction of the identification challenge in dynamic regression discontinuity designs, drawing on the U.S. House election setting from Lee (2008). The election result in year t affects both short-term outcomes (years t and $t+1$) and longer-term outcomes (years $t+2$ and $t+3$). However, if the election result in year $t+2$ differs from that in year t, the treatment state changes, making it difficult to disentangle the long-term effects of the year-t election from the effects of the subsequent election.

---

[2]This example assumes that House elections occur at the beginning of a calendar year and that outcomes are observed at the end of it.

Since the seminal contribution by Cellini, Ferreira and Rothstein (2010), dynamic regression discontinuity designs have been employed almost exclusively in the field of local public finance. Numerous empirical studies leverage school district referenda on property tax hikes to estimate the effects of increased school expenditures on outcomes such as housing prices and student test scores. Similar to the stylized example above, dynamic fuzziness arises because school districts hold repeated referenda to authorize additional spending. A ballot measure that is initially rejected may be resubmitted and later approved. As a result, a researcher might observe that a referendum was rejected at time t but approved at time $t + 1$. If a study uses the year-t referendum outcome to estimate the long-term effects of extra spending at time t, labeling the district as untreated at time t overlooks its transition into a treated state at time $t + 1$, potentially conflating the effects of treatment received in successive periods.

Most empirical studies employing a dynamic regression discontinuity design adopt the so called "one-step" approach proposed by Cellini, Ferreira and Rothstein (2010). This method relies on the following dynamic two-way fixed effects specification:

$$Y_{it} = \alpha_i + \beta_t + \sum_{s=0}^{\overline{s}} \left(\theta_s D_{i,t-s} + \gamma_s Q_{i,t-s} + p_g \left(\delta_s, R_{i,t-s} \right) \right) + U_{it} \tag{1}$$

where D is a binary treatment indicator, Q is an indicator for whether a referendum is held, and $p_g (\cdot)$ represents a gth-degree polynomial of the approval vote share R. This specification is typically estimated without bandwidth restrictions, allowing researchers to flexibly control for both current and past treatment states, as well as for current and past realizations of the running variable (Darolia 2013, Hong and Zimmer 2016, Martorell, Stange and McFarlin 2016, Baron, Hyman and Vasquez 2024).

More recently, empirical researchers have proposed modifications to the "one-step" approach introduced by Cellini, Ferreira and Rothstein (2010). Abott et al. (2020) and Baron (2022) extend the estimating equation in (1) by including lead treatment indicators, allowing s to take negative values. Similarly, Biasi, Lafortune and Schönholzer (2025) incorporates leads but estimates the resulting linear regression in a "stacked" sample, where each component corresponds to a cohort-specific subsample. Here, a cohort is defined as a calendar year in which referenda occur[3]. The authors interact each regressor in specification (1) with co-

---

[3]Biasi, Lafortune and Schönholzer (2025) restricts each cohort to include only treated jurisdictions and

hort indicators, effectively estimating the model for each cohort separately. Departing from Cellini, Ferreira and Rothstein (2010), Hsu and Shen (2024) develops a dynamic potential outcomes model that accounts for heterogeneity in treatment effects and path dependence in treatment assignments. To identify interpretable long-term effects, the authors impose a selection-on-observables assumption that exploits future approval vote shares and, possibly, additional covariates. More details on their framework are provided in the following section.

3 A Two-Period Regression Discontinuity Design

In this section, I present the core of my identification strategy within a simple two-period model. Sections 4 and 5 extend this framework to accommodate an arbitrary number of time periods.

3.1 Model Setup

Consider a dynamic setting in which time periods are indexed by $t \in \{1, 2\}$. Let $D_t \in \{0, 1\}$ denote a binary treatment that is assigned at the beginning of period t. The treatment is determined by whether the realization of an observed running variable $R_t \in \mathbb{R}$ exceeds a known deterministic constant $c \in \mathbb{R}$:

$$D_t \equiv \mathbb{I}[R_t \geq c] \tag{2}$$

Let $Y_t \in \mathbb{R}$ be an outcome that is observed at the end of period t. In the canonical static setting, potential outcomes depend on a single treatment state, and the observed outcome satisfies the switching equation $Y_t = \sum_{d \in \{0,1\}} \mathbb{I}[D_t = d] Y_t(d)$. In a two-period model with repeated treatment assignment, the potential outcome in period 1 may reflect the treatment state in period 2 due to anticipation and, symmetrically, the potential outcome in period 2 may be affected by the treatment state in period 1 due to path dependence. Thus, following Robins (1986), it is natural to model potential outcomes as depending on the path of past, contemporaneous, and future treatment states at any point in time. Specifically, let the period-t potential outcome be $Y_t(d_1, d_2)$, where $d_1, d_2 \in \{0, 1\}$. Without further assumptions, the evolution of treatment states is unrestricted. In fact, a unit may be repeatedly

---

untreated jurisdictions that held a referendum in the same year but did not approve any ballot measure in the subsequent ten years.

assigned the treatment, never receive it, or be exposed to it in only one period. As in Hsu and Shen (2024), I allow for path dependence in treatment assignment by modeling potential treatments as functions of the history of treatment states. Specifically, let D_1 and $D_2(d_1)$ denote the first- and second-period potential treatments, respectively[4]. In the canonical application of dynamic RDDs to local referenda, it is useful to interpret potential treatments as characterizing jurisdiction "types". For example, a school district in which voters repeatedly reject a bond for school construction ($D_1 = 0$ and $D_2(0) = D_2(1) = 0$) is likely more fiscally conservative than one in which increased public expenditure is approved at the first round of elections ($D_1 = 1$) or at the second round ($D_1 = 0$ and $D_2(0) = 1$). In this two-period model, the observed outcome in period t is related to potential outcomes as

$$Y_t = \sum_{(d_1,d_2)\in\{0,1\}^2} \mathbb{I}[D_1 = d_1, D_2(d_1) = d_2] Y_t(d_1, d_2) \tag{3}$$

3.2   Target Parameters

Having outlined features of the dynamic potential outcomes model, it is natural to define target parameters. As in static RDDs, all causal parameters of interest are local to the nonstochastic threshold above which the treatment is assigned (Hahn, Todd and Van der Klaauw 2001). Without loss, the analysis will focus on target parameters implied by the first-period discontinuity. Letting $\tau \in \{0,1\}$ denote the number of leads, consider a class of threshold-specific Average Treatment Effects (ATEs),

$$\text{ATE}_{1,\tau}(c, d_2, d_2') \equiv \mathbb{E}[Y_{1+\tau}(1, d_2) - Y_{1+\tau}(0, d_2') | R_1 = c] \tag{4}$$

where $d_2, d_2' \in \{0,1\}$. Each of these average effects is constructed as the average contrast between a potential outcome that is treated in the first period and a potential outcome that is not. The choice of a second-period treatment state naturally gives rise to causal parameters with a different interpretation.

Since Cellini, Ferreira and Rothstein (2010), the empirical public finance literature that studies the effects of increased public expenditure authorized by local referenda has focused

---

[4]A more flexible setting could allow for anticipation in treatment assignment by letting the potential treatment in period 1 depend on the treatment state in period 2, i.e., $D_1(d_2)$. For simplicity, I abstract from this scenario and allow for anticipation only in potential outcomes.

on a class of Average Direct Treatment Effects that sets $d_2 = d'_2 = 0$,

$$\text{ADTE}_{1,\tau}(c) \equiv \mathbb{E}\left[Y_{1+\tau}(1,0) - Y_{1+\tau}(0,0) \mid R_1 = c\right] \tag{5}$$

Hereafter, I will refer to $\text{ADTE}_{1,0}(c)$ and $\text{ADTE}_{1,1}(c)$ as the instantaneous and cumulative Average Direct Treatment Effects, respectively. These target parameters can be interpreted as tracing out the impulse response of the outcome to the first-period discontinuity. Economically, they correspond to the thought experiment of measuring the cutoff-specific average effect of first-period treatment assignment in the counterfactual scenario in which exposure to the treatment is set to zero thereafter.

3.3 What Do Standard RD Estimands Identify?

In what follows, I will show that standard RD estimands do not, in general, identify ADTEs and will subsequently specify assumptions to achieve this goal. First, the (sharp) RD estimand implied by the first-period discontinuity is defined as

$$\text{RD}_t(c) \equiv \lim_{r\downarrow c} \mathbb{E}\left[Y_t \mid R_1 = r\right] - \lim_{r\uparrow c} \mathbb{E}\left[Y_t \mid R_1 = r\right] \tag{6}$$

Hahn, Todd and Van der Klaauw (2001) shows that continuity in average potential outcomes at the cutoff is sufficient to point identify the Average Treatment Effect at the threshold. It is natural to generalize this assumption to a dynamic setting.

Assumption 1 (Continuity at the Cutoff) *For any $t \in \{1,2\}$ and $d_1, d_2 \in \{0,1\}$,*

$$\mathbb{E}\left[D_2(d_1) \mid R_1 = r\right] \quad \text{and} \quad \mathbb{E}\left[Y_t(d_1, d_2) \mid R_1 = r, D_2(d_1)\right]$$

are continuous at $r = c$.

Assumption 1 generalizes the restriction that agents do not sort nonrandomly around the threshold. First, the probability of potential treatment assignment in the second period evolves smoothly at the first-period cutoff. Importantly, this statement does *not* imply that the probability of observing $D_2 = 1$ is unaffected by the first-period treatment state. Instead, it restricts the probability of any jurisdiction "type" – as pinned down by the potential treatment $D_2(d_1)$ – not to vary abruptly at the first-period cutoff. Analogously, Assumption 1 rules out a scenario in which agents systematically choose either side of the first-period

8

cutoff based on observed and unobserved determinants of the outcome, within each juris-
diction "type" and for each possible path of treatment states. Under this assumption, it is
immediate to derive the following identification argument.

Proposition 1 (Standard RD Estimand) *Suppose that Assumption 1 holds. Then*

$$\text{RD}_t(c) = \sum_{d_2 \in \{0,1\}} \mathbb{E}\left[Y_t(1, d_2) | R_1 = c, D_2(1) = d_2\right] \times \mathbb{P}\left(D_2(1) = d_2 | R_1 = c\right)$$
$$- \sum_{d_2 \in \{0,1\}} \mathbb{E}\left[Y_t(0, d_2) | R_1 = c, D_2(0) = d_2\right] \times \mathbb{P}\left(D_2(0) = d_2 | R_1 = c\right)$$

Without further assumptions, the sharp regression discontinuity estimand identifies the con-
trast between two weighted sums of potential outcomes over second-period treatment states,
where weights are nonnegative and add up to one. As discussed in Cellini, Ferreira and
Rothstein (2010) and similarly in Heckman, Humphries and Veramendi (2016), this target
parameter may be interpreted as the "intent-to-treat" implied by the first-period discontinu-
ity. Because treatment assignment exhibits path dependence and outcomes reflect past and
future treatment states, the ITT is generally hard to interpret. By way of exposition, con-
sider the extreme scenario in which $\mathbb{P}(D_2(1) = 1 | R_1 = c) = 1$ and $\mathbb{P}(D_2(0) = 0 | R_1 = c) = 1$.
In this case, $\text{RD}_2(c)$ identifies $\mathbb{E}\left[Y_2(1,1) - Y_2(0,0) | R_1 = c\right]$. However, the treatment state
in period 1 perfectly predicts the treatment state in period 2, making it infeasible to disen-
tangle the cumulative effect of the first-period discontinuity, namely $\text{ADTE}_{1,1}(c)$, from the
instantaneous effect of the second-period discontinuity. Formally,

$$\text{RD}_2(c) = \underbrace{\mathbb{E}\left[Y_2(1,1) - Y_2(1,0) | R_1 = c\right]}_{\text{instantaneous effect of } D_2} + \underbrace{\mathbb{E}\left[Y_2(1,0) - Y_2(0,0) | R_1 = c\right]}_{\text{cumulative effect of } D_1} \qquad (7)$$

Estimating the "intent-to-treat" from Proposition 1 may suffice should a researcher be
interested in measuring the reduced-form (or "total") effect of a policy initiated by the first-
period discontinuity, regardless of the subsequent path of treatment states. However, the
composition of a long-term causal parameter is intrinsically of value in several empirical
settings. For instance, does increased funding for local schools lead to sustained and perma-
nent gains in student achievement or the effect of extra expenditure decays over time? To
answer this and similar policy-relevant questions, a researcher may prefer to impose addi-
tional restrictions and estimate Average Direct Treatment Effects that trace out the impulse
response of an outcome to the first-period round of treatment assignment.

Motivated by this negative result, in the next two sections I specify additional sufficient restrictions to identify the instantaneous and cumulative Average Direct Treatment Effects implied by the first-period discontinuity.

3.4 Identification of the Instantaneous ADTE

The goal of this section is to provide an identification argument for the instantaneous Average Direct Treatment Effect, $\mathbb{E}\left[Y_1\left(1,0\right)|R_1=c\right] - \mathbb{E}\left[Y_1\left(0,0\right)|R_1=c\right]$. By an application of the Law of Iterated Expectations, the first term can be expressed as

$$\mathbb{E}\left[Y_1\left(1,0\right)|R_1=c\right] = \overbrace{\mathbb{E}\left[Y_1\left(1,0\right)|R_1=c,D_2\left(1\right)=0\right]}^{\text{identified}} \times \overbrace{\mathbb{P}\left(D_2\left(1\right)=0|R_1=c\right)}^{\text{identified}}$$
$$+ \underbrace{\mathbb{E}\left[Y_1\left(1,0\right)|R_1=c,D_2\left(1\right)=1\right]}_{\text{counterfactual}} \times \underbrace{\mathbb{P}\left(D_2\left(1\right)=1|R_1=c\right)}_{\text{identified}} \quad (8)$$

Notice that both probabilities are identified under Assumption 1. In fact, for $d_2 \in \{0,1\}$,

$$\mathbb{P}\left(D_2\left(1\right)=d_2|R_1=c\right) = \lim_{r\downarrow c}\mathbb{P}\left(D_2=d_2|R_1=r\right) \quad (9)$$

Similarly, the first conditional expectation in equation (8) is identified as

$$\mathbb{E}\left[Y_1\left(1,0\right)|R_1=c,D_2\left(1\right)=0\right] = \lim_{r\downarrow c}\mathbb{E}\left[Y_1|R_1=r,D_2=0\right] \quad (10)$$

On the other hand, $\mathbb{E}\left[Y_1\left(1,0\right)|R_1=c,D_2\left(1\right)=1\right]$ is counterfactual. Intuitively, this term captures the average treated-untreated potential outcome among units who would receive the treatment in $t=2$ after receiving it in $t=1$. To identify this counterfactual expectation, it is sufficient to assume that, on average and within bins implied by potential treatment states, the first-period outcome is independent of the second-period treatment state.

Assumption 2 (No Anticipation) *For any $d_1, d_2 \in \{0,1\}$,*

$$\mathbb{E}\left[Y_1\left(d_1,0\right)|R_1=c,D_2\left(d_1\right)=d_2\right] = \mathbb{E}\left[Y_1\left(d_1,1\right)|R_1=c,D_2\left(d_1\right)=d_2\right]$$

Assumption 2 is a cutoff-specific version of the canonical no anticipation assumption required to point identify the Average Treatment Effect on the Treated in difference-in-differences designs (Abbring and Van Der Berg 2003). It restricts current potential outcomes not to reflect (expectations over) future treatment states. While this statement may appear to rule

out the applicability of this argument to forward-looking outcome variables, it should be noted that the lack of anticipatory behavior must hold only at the first-period cutoff. In the application of dynamic RDDs to the effect of increased school spending on student outcomes, Assumption 2 entails that average test scores after a narrowly rejected bond do not reflect the possibility that a bond proposition may be approved or rejected in a subsequent referendum. Under Assumptions 1 and 2, the counterfactual conditional mean is identified as

$$\mathbb{E}\left[Y_1\left(1,0\right)|R_1=c,D_2\left(1\right)=1\right]=\lim_{r\downarrow c}\mathbb{E}\left[Y_1|R_1=r,D_2=1\right] \tag{11}$$

Thus, the left-hand side of $\text{ADTE}_{1,0}(c)$ can be expressed as a function of the following observables:

$$\begin{aligned}\mathbb{E}\left[Y_1\left(1,0\right)|R_1=c\right]&=\lim_{r\downarrow c}\mathbb{E}\left[Y_1|R_1=r,D_2=0\right]\times\lim_{r\downarrow c}\mathbb{P}\left(D_2=0|R_1=r\right)\\&+\lim_{r\downarrow c}\mathbb{E}\left[Y_1|R_1=r,D_2=1\right]\times\lim_{r\downarrow c}\mathbb{P}\left(D_2=1|R_1=r\right)\\&=\lim_{r\downarrow c}\mathbb{E}\left[Y_1|R_1=r\right]\end{aligned} \tag{12}$$

where the last equality follows from an application of the Law of Iterated Expectations. By a symmetric argument, the right-hand side of $\text{ADTE}_{1,0}(c)$ is point identified, too. Combining results yields the following proposition.

Proposition 2 (Identification of the Instantaneous ADTE) *Suppose that Assumptions 1 and 2 hold. Then*

$$\text{ADTE}_{1,0}(c)=\lim_{r\downarrow c}\mathbb{E}\left[Y_1|R_1=r\right]-\lim_{r\uparrow c}\mathbb{E}\left[Y_1|R_1=r\right]\equiv\text{RD}_1\left(c\right)$$

This result follows from considering Proposition 1 in the case in which Assumption 2 holds. Intuitively, the lack of anticipation implies that the second-period potential treatment state is irrelevant for the first-period outcome. Setting $d_2=0$ yields the identification result in Proposition 2. To summarize, continuity and no anticipation are sufficient for the static regression discontinuity estimand to identify an interpretable causal parameter in this model.

3.5 Identification of the Cumulative ADTE

Consider, instead, the more interesting goal of identifying the cumulative Average Direct Treatment Effect, $\mathbb{E}\left[Y_2\left(1,0\right)|R_1=c\right]-\mathbb{E}\left[Y_2\left(0,0\right)|R_1=c\right]$. Mirroring the argument in the

previous section, an application of the Law of Iterated Expectations implies that the first term can be expressed as

$$\mathbb{E}\left[Y_2\left(1,0\right)|R_1=c\right]=\overbrace{\mathbb{E}\left[Y_2\left(1,0\right)|R_1=c,D_2\left(1\right)=0\right]}^{\text{identified}}\times\overbrace{\mathbb{P}\left(D_2\left(1\right)=0|R_1=c\right)}^{\text{identified}}$$
$$+\underbrace{\mathbb{E}\left[Y_2\left(1,0\right)|R_1=c,D_2\left(1\right)=1\right]}_{\text{counterfactual}}\times\underbrace{\mathbb{P}\left(D_2\left(1\right)=1|R_1=c\right)}_{\text{identified}}\qquad(13)$$

As shown in equation (9), both conditional probabilities are identified under Assumption 1. In addition, the first conditional expectation in equation (13) is identified as

$$\mathbb{E}\left[Y_2\left(1,0\right)|R_1=c,D_2\left(1\right)=0\right]=\lim_{r\downarrow c}\mathbb{E}\left[Y_2|R_1=r,D_2=0\right]\qquad(14)$$

On the other hand, $\mathbb{E}\left[Y_2\left(1,0\right)|R_1=c,D_2\left(1\right)=1\right]$ is counterfactual because $Y_2\left(1,0\right)$ is unobserved among units who would sort into the treated arm in $t=2$ after being treated in $t=1$. Differently from the previous section, Assumption 2 alone is not sufficient to identify the counterfactual mean because the outcome is measured in $t=2$. However, point identification of the target parameter may be achieved by also assuming that the expected growth in untreated-at-2 potential outcomes is independent of the potential second-period treatment state at the cutoff.

Assumption 3 (Common Trends) *For $d_1\in\{0,1\}$,*

$$\mathbb{E}\left[\Delta_{1,2}\left(d_1,0\right)|R_1=c,D_2\left(d_1\right)=0\right]=\mathbb{E}\left[\Delta_{1,2}\left(d_1,0\right)|R_1=c,D_2\left(d_1\right)=1\right]$$

where $\Delta_{1,2}\left(d_1,0\right)\equiv Y_2\left(d_1,0\right)-Y_1\left(d_1,0\right)$.

Assumption 3 is a cutoff-specific version of the common trends assumption required to point identify the Average Treatment Effect on the Treated in difference-in-differences designs (Abadie 2005). As in that setting, groups ("treatment" and "control") are implied by paths of potential treatment states. Differently from that setting, the treatment is available also in the first period, implying that there are four, not two, groups or "types". As a consequence, the parallel trends restriction also applies to units who are potentially treated in the first period. In the classic application of dynamic RDDs to local public finance, Assumption 3 entails that, at the cutoff implied by the first-period referendum, the expected growth in potential test scores absent a second referendum approval is independent of a jurisdiction's

"type", i.e., it is independent of whether a jurisdiction would approve or reject a school bond referendum in the second period. Under Assumptions 1, 2, and 3, the counterfactual conditional mean is identified as

$$\mathbb{E}\left[Y_2\left(1,0\right)|R_1 = c, D_2\left(1\right) = 1\right] = \lim_{r\downarrow c}\mathbb{E}\left[Y_1|R_1 = r, D_2 = 1\right] + \lim_{r\downarrow c}\mathbb{E}\left[Y_2 - Y_1|R_1 = r, D_2 = 0\right] \quad (15)$$

As in the canonical difference-in-differences design, the counterfactual term is identified as the sum of a "level" and a "growth" term. In this case, the former measures the average observed first-period outcome among units who later sort into the treated arm. The latter captures the average observed outcome growth among units who sort into the untreated arm in the second period. Thus, the left-hand side of $\text{ADTE}_{1,1}\left(c\right)$ can be expressed as a function of the following observables:

$$\begin{aligned}
\mathbb{E}\left[Y_2\left(1,0\right)|R_1 = c\right] = {} & \lim_{r\downarrow c}\mathbb{E}\left[Y_2|R_1 = r, D_2 = 0\right] \times \lim_{r\downarrow c}\mathbb{P}\left(D_2 = 0|R_1 = r\right) \\
& + \left(\lim_{r\downarrow c}\mathbb{E}\left[Y_1|R_1 = r, D_2 = 1\right] + \lim_{r\downarrow c}\mathbb{E}\left[Y_2 - Y_1|R_1 = r, D_2 = 0\right]\right) \\
& \times \lim_{r\downarrow c}\mathbb{P}\left(D_2 = 1|R_1 = r\right) \quad (16)
\end{aligned}$$

By a symmetric argument, the right-hand side of $\text{ADTE}_{1,1}(c)$ is point identified, too. Combining results yields the following proposition.

Proposition 3 (Identification of the Cumulative ADTE) *Suppose that Assumptions 1, 2, and 3 hold. Then*

$$\begin{aligned}
\text{ADTE}_{1,1}(c) = {} & \lim_{r\downarrow c}\mathbb{E}\left[Y_1|R_1 = r\right] - \lim_{r\uparrow c}\mathbb{E}\left[Y_1|R_1 = r\right] \\
& + \lim_{r\downarrow c}\mathbb{E}\left[Y_2 - Y_1|R_1 = r, D_2 = 0\right] - \lim_{r\uparrow c}\mathbb{E}\left[Y_2 - Y_1|R_1 = r, D_2 = 0\right]
\end{aligned}$$

This proposition entails that an interpretable long-term causal parameter determined by the first-period discontinuity is identified by the sum of two standard sharp Regression Discontinuity estimands, namely the RD estimand that uses Y_1 as an outcome and the RD estimand that uses $Y_2 - Y_1$ as an outcome in the subpopulation implied by $D_2 = 0$.

3.6   Relation to Hsu and Shen (2024)

The fundamental difference between this paper and Hsu and Shen (2024) lies in the assumption used to identify the counterfactual expectation $\mathbb{E}\left[Y_2\left(1,0\right)|R_1 = c, D_2\left(1\right) = 1\right]$. Hsu and

Shen (2024) introduces a potential indicator $Q_2(d_1)$, which denotes whether a round of treatment assignment would occur in the second period if the first-period treatment state were $d_1 \in \{0, 1\}$. When $Q_2(d_1) = 1$, a corresponding potential second-period running variable $R_2(d_1)$ is defined. Their key identifying assumption states that, if a second-period treatment assignment occurs, the potential outcome $Y_2(1, 0)$ is mean independent of $R_2(d_1)$ in a neighborhood of the first-period cutoff. This condition is further generalized to incorporate covariates X, such that, for any $r \in \text{supp}(R_2)$,

$$\mathbb{E}\left[Y_2(d_1, 0) \middle| R_1 = c, Q_2(d_1) = 1, X = x, R_2(d_1) = r\right]$$
$$= \mathbb{E}\left[Y_2(d_1, 0) \middle| R_1 = c, Q_2(d_1) = 1, X = x\right] \tag{17}$$

This assumption has the appealing feature of restricting comparisons to units for which treatment assignment occurs in the second period as well. However, it may be vulnerable to violations if unobserved determinants of the second-period outcome also influence the second-period vote share. For instance, in the canonical application examining the effect of school expenditure referenda on student test scores, unobserved parental valuation of education may drive both support for a second-round referendum and student performance—e.g., because more involved parents are more likely to vote in favor of school funding (Poterba 1997) and also assist their children academically (Guryan, Hurst and Kearney 2008). Notably, such a violation can occur even if the unobservable is time-invariant and the first-round approval margin is idiosyncratically close to the cutoff.

Another distinction between the identification strategy developed in this paper and that in Hsu and Shen (2024) pertains to the role of anticipation in the potential outcomes framework. The dynamic model presented here allows potential outcomes at any given time to depend on past, contemporaneous, and future treatment states. In a setting with two periods, this implies that potential outcomes in the first period are denoted as $Y_1(d_1, d_2)$, while those in the second period are given by $Y_2(d_1, d_2)$. By contrast, the setup considered by Hsu and Shen (2024) restricts the sequence of treatment states to end in the period in which the outcome is realized, with first-period outcomes modeled as $Y_1(d_1)$ and second-period outcomes as $Y_2(d_1, d_2)$. In empirical applications where the outcome of interest may respond to anticipated future treatments—such as settings involving housing prices or other forward-looking variables commonly analyzed in public finance (e.g., Cellini, Ferreira and Rothstein

14

2010 and Biasi, Lafortune and Schönholzer 2025)—it may be useful to explicitly state the role of anticipation in the underlying model.

Finally, the approach developed in this paper is related in spirit to event-study-type specifications commonly employed by empirical researchers in dynamic RDD applications. The inclusion of leads in the estimating equation (1) by Abott et al. (2020) and Baron (2022), together with the cohort-by-cohort identification strategy based on "clean controls" proposed by Biasi, Lafortune and Schönholzer (2025), suggests that a conceptual link between dynamic RDDs and difference-in-differences designs has already been implicitly recognized in the literature. This paper formalizes that connection.

4   Generalization to Multiple Leads

It is natural to generalize previous arguments to identify long-term causal parameters in a dynamic potential outcomes model with more than two periods. In this section, I extend the setup and identifying assumptions of the two-period model to a setting with an arbitrary number of time periods indexed by $t \in \left\{ 1, 2, \ldots, \bar{t} \right\}$.

First, potential outcomes at any point in time reflect the path of past, contemporaneous, and future treatment states. Specifically, for $\{d_j\}_{j=1}^{\bar{t}}$ with $d_j \in \{0,1\}$ for all j, the period-t potential outcome is denoted as $Y_t\left(d_1, \ldots, d_{\bar{t}} \right)$. Second, to allow for path dependence in treatment assignment, the period-t potential treatment is modeled as $D_t\left(d_1, \ldots, d_{t-1} \right)$. As a generalization of equation (3), the relationship between potential and observed outcomes can thus be characterized as

$$Y_t = \sum_{(d_1,\ldots,d_{\bar{t}})\in\{0,1\}^{\bar{t}}} \mathbb{I}\left[D_1 = d_1, D_2\left(d_1\right) = d_2, \ldots, D_{\bar{t}}\left(d_1,\ldots,d_{\bar{t}-1}\right) = d_{\bar{t}} \right] Y_t\left(d_1,\ldots,d_{\bar{t}}\right) \quad (18)$$

To keep notation concise, let $P : \{0,1\}^{\bar{t}} \to \{0,1\}$ be a function mapping a path of potential treatment states to an indicator that takes the value one if a unit's treatment-taking behavior for a given $d_1 \in \{0,1\}$ is described by that path, i.e.,

$$P\left(d_1,\ldots,d_{\bar{t}}\right) = \mathbb{I}\left[D_2\left(d_1\right) = d_2, D_3\left(d_1,d_2\right) = d_3, \ldots, D_{\bar{t}}\left(d_1,\ldots,d_{\bar{t}-1}\right) = d_{\bar{t}} \right] \quad (19)$$

Therefore, a compact formulation of equation (18) is

$$Y_t = \sum_{(d_1,\ldots,d_{\bar{t}})\in\{0,1\}^{\bar{t}}} \mathbb{I}\left[D_1 = d_1 \right] P\left(d_1,\ldots,d_{\bar{t}}\right) Y_t\left(d_1,\ldots,d_{\bar{t}}\right) \quad (20)$$

15

The target parameter of interest is a class of Average Direct Treatment Effects implied by the first-period discontinuity. Letting $0_{\bar{t}-1}$ denote a vector of $\bar{t}-1$ zeros, the τ-period-ahead ADTE at the first-period cutoff is defined as

$$\text{ADTE}_{1,\tau}(c) \equiv \mathbb{E}\left[Y_{1+\tau}(1, 0_{\bar{t}-1}) - Y_{1+\tau}(0, 0_{\bar{t}-1}) \,|\, R_1 = c\right] \tag{21}$$

It is convenient to interpret this class of target parameters as tracing out the impulse response function of the outcome Y in a dynamic model in which the initial "shock" is determined by the threshold-crossing rule $D_1 \equiv \mathbb{I}[R_1 \geq c]$. Because treatment assignment exhibits path dependence, the sequence of states $0_{\bar{t}-1}$ may not be observed. As a matter of fact, it is counterfactual in most empirical applications. For instance, local jurisdictions in which voters reject a referendum for increased public spending may submit the same proposition again and have it approved in a subsequent round of elections. In this scenario, a standard regression discontinuity estimand will fail to identify an ADTE or a causal parameter with a clear economic interpretation. To formalize this argument, the following assumption generalizes the continuity restriction naturally embedded in the standard RDD model.

Assumption 4 (Continuity at the Cutoff) *For any* $t \in \{1, \ldots, \bar{t}\}$, *any* $\{d_j\}_{j=1}^{\bar{t}}$ *with* $d_j \in \{0, 1\}$ *for all* j, *and any* $\{d'_j\}_{j=2}^{\bar{t}}$ *with* $d'_j \in \{0, 1\}$ *for all* j,

$$\mathbb{E}\left[P\left(d_1, \ldots, d_{\bar{t}}\right) | R_1 = r\right] \quad and \quad \mathbb{E}\left[Y_t\left(d_1, d_2, \ldots, d_{\bar{t}}\right) | R_1 = r, P\left(d_1, d'_2, \ldots, d'_{\bar{t}}\right)\right]$$

are continuous at $r = c$.

Assumption 4 generalizes Assumption 1 to a model with an arbitrary number of time periods. Intuitively, restricting $\mathbb{E}\left[P\left(d_1, \ldots, d_{\bar{t}}\right) | R_1 = r\right]$ to be continuous at the first-period cutoff entails that agents must not systematically sort around the threshold to make one treatment path more likely than another. Similarly, the continuity of $\mathbb{E}\left[Y_t\left(d_1, d_2, \ldots, d_{\bar{t}}\right) | R_1 = r, P\left(d_1, d'_2, \ldots, d'_{\bar{t}}\right)\right]$ implies that observed and unobserved determinants of the outcome evolve smoothly at the cutoff, within each possible "type" defined by a path of potential treatment states. Under Assumption 4 and without further restrictions, it is immediate to show that a naive regression discontinuity estimand identifies a causal parameter with an undesirable interpretation.

Proposition 4 (Standard RD Estimand) *Suppose that Assumption 4 holds. Then the estimand* $\mathrm{RD}_t(c) \equiv \lim_{r \downarrow c} \mathbb{E}[Y_t | R_1 = r] - \lim_{r \uparrow c} \mathbb{E}[Y_t | R_1 = r]$ *identifies*

$$\sum_{(d_2,\dots,d_{\bar{t}}) \in \{0,1\}^{\bar{t}-1}} \mathbb{E}[Y_t(1, d_2, \dots, d_{\bar{t}}) | R_1 = c, P(1, d_2, \dots, d_{\bar{t}}) = 1] \times \mathbb{E}[P(1, d_2, \dots, d_{\bar{t}}) | R_1 = c]$$

$$- \sum_{(d_2,\dots,d_{\bar{t}}) \in \{0,1\}^{\bar{t}-1}} \mathbb{E}[Y_t(0, d_2, \dots, d_{\bar{t}}) | R_1 = c, P(0, d_2, \dots, d_{\bar{t}}) = 1] \times \mathbb{E}[P(0, d_2, \dots, d_{\bar{t}}) | R_1 = c]$$

Since $P(d_1, \dots, d_{\bar{t}})$ is a Bernoulli random variable, each $\mathbb{E}[P(d_1, \dots, d_{\bar{t}}) | R_1 = c]$ is a conditional probability and $\sum_{(d_2,\dots,d_{\bar{t}}) \in \{0,1\}^{\bar{t}-1}} \mathbb{E}[P(d_1, d_2, \dots, d_{\bar{t}}) | R_1 = c]$ is equal to one for $d_1 \in \{0, 1\}$. Thus, the causal parameter identified by a standard RD estimand under Assumption 4 can be viewed as the contrast between two weighted sums of potential outcome means, where only the first-period treatment state is held fixed and each subsequent treatment path is weighted according to its probability at the focal threshold.

In the remainder of this section, I first provide an identification argument for the instantaneous ADTE and then consider long-term parameters. To identify the cutoff-specific effect of the first-period discontinuity on the outcome at the end of the same period, it is sufficient to assume that the outcome of interest does not reflect agents' knowledge of or expectations over future treatment states.

Assumption 5 (No Anticipation) *For any* $\{d_j\}_{j=1}^{\bar{t}}$ *with* $d_j \in \{0, 1\}$ *for all* j, *any* $\{d_j'\}_{j=2}^{\bar{t}}$ *with* $d_j' \in \{0, 1\}$ *for all* j, *and any* $\{d_j''\}_{j=2}^{\bar{t}}$ *with* $d_j'' \in \{0, 1\}$ *for all* j,

$$\mathbb{E}\left[Y_1(d_1, d_2, \dots, d_{\bar{t}}) - Y_1(d_1, d_2', \dots, d_{\bar{t}}') \,\middle|\, R_1 = c, P(d_1, d_2'', \dots, d_{\bar{t}}'')\right] = 0$$

Starting from Proposition 4, it is easy to show that that the instantaneous ADTE is point identified under Assumption 5. In fact, letting $t = 1$, the no anticipation restriction entails that the treatment path subsequent to the first period can be equivalently replaced with a sequence of $t - 1$ zeros. Then, repeated applications of the Law of Iterated Expectations and Assumption 4 yield the following proposition.

Proposition 5 (Identification of the Instantaneous ADTE) *Suppose that Assumptions* (4) *and* (5) *hold. Then*

$$\mathrm{ADTE}_{1,0}(c) = \lim_{r \downarrow c} \mathbb{E}[Y_1 | R_1 = r] - \lim_{r \uparrow c} \mathbb{E}[Y_1 | R_1 = r] \equiv \mathrm{RD}_1(c)$$

This result generalizes Proposition 2 to a dynamic potential outcomes model with an arbitrary number of periods. In words, the instantaneous regression discontinuity estimand identifies an impulse response parameter provided that the outcome is not forward-looking at the first-period cutoff. However, by placing restrictions only on the conditional expectation of the first-period outcome, the no anticipation assumption is not sufficient to trace out the full impulse response function. To identify long-term Average Direct Treatment Effects, it is sufficient to generalize the common trends assumption introduced in the previous section.

Assumption 6 (Common Trends) *For all* $\{d_j\}_{j=1}^{\bar{t}}$ *with* $d_j \in \{0,1\}$,

$$\mathbb{E}\left[\Delta_{1,t}\left(d_1, 0_{\bar{t}-1}\right) | R_1 = c, P\left(d_1, 0_{\bar{t}-1}\right)\right] = \mathbb{E}\left[\Delta_{1,t}\left(d_1, 0_{\bar{t}-1}\right) | R_1 = c, P\left(d_1, d_2, \ldots, d_{\bar{t}}\right)\right]$$

with $\Delta_{1,t}\left(d_1, 0_{\bar{t}-1}\right) \equiv Y_t\left(d_1, 0_{\bar{t}-1}\right) - Y_1\left(d_1, 0_{\bar{t}-1}\right)$.

For any initial potential treatment state d_1, $\Delta_{1,t}\left(d_1, 0_{\bar{t}-1}\right)$ is a random variable that measures the level shift experienced by the never-treated potential outcome between the first and a subsequent period. Assumption 6 restricts the expectation of this random variable, conditional on $R_1 = c$, not to depend on the potential treatment path after d_1. In the canonical application of dynamic RDDs to the effect of local school bonds on student test scores, the path indicator $P\left(d_1, d_2, \ldots, d_{\bar{t}}\right)$ collapses the heterogeneity in jurisdiction "types" into a sequence of potential treatment states. Thus, Assumption 6 restricts the expected change in test scores after the first period, absent any subsequent referendum approval, not to depend on whether a jurisdiction is more or less likely to approve said referenda. This local "parallel trends" assumption, jointly with continuity and no anticipation, are sufficient to point identify all Average Direct Treatment Effects.

Proposition 6 (Identification of the Cumulative ADTE) *Suppose that Assumptions 4, 5, and 6 hold. Then, for* $\tau \in \{1, \ldots, \bar{t} - 1\}$,

$$\text{ADTE}_{1,\tau}(c) = \lim_{r \downarrow c} \mathbb{E}\left[Y_1 | R_1 = r\right] - \lim_{r \uparrow c} \mathbb{E}\left[Y_1 | R_1 = r\right]$$

$$+ \lim_{r \downarrow c} \mathbb{E}\left[Y_{1+\tau} - Y_1 \,\middle|\, R_1 = r, \bigcap_{s=2}^{1+\tau} \{D_s = 0\}\right] - \lim_{r \uparrow c} \mathbb{E}\left[Y_{1+\tau} - Y_1 \,\middle|\, R_1 = r, \bigcap_{s=2}^{1+\tau} \{D_s = 0\}\right]$$

Proposition 6 states that continuity, no anticipation, and common trends are sufficient for any long-term Average Direct Treatment Effect to be point identified as the sum of two

regression discontinuity estimands implied by the first-period cutoff. The first estimand measures the average outcome contrast at the cutoff immediately after the first round of treatment assignment. On the other hand, the second term compares the average outcome growth on the two sides of the cutoff among agents who are not exposed to the treatment forever after.

5 Aggregation of Time-Specific Parameters

The identification arguments presented so far have focused on causal parameters implied by the first-period discontinuity. In practice, the focal round of treatment assignment may not occur at $t = 1$. Moreover, agents may be exposed to multiple rounds of treatment assignment at different points in time.

To allow for the focal discontinuity to occur in an arbitrary period t, it is convenient to introduce a random variable $H_{t-1} \in \{0,1\}^{t-1}$ that summarizes the history of treatment states up to and excluding period t. Without further restrictions, past treatment states may affect the probability distribution of potential outcomes when the treatment is assigned. Thus, for a causal parameter to be interpretable, it must contrast agents who have been exposed to an identical path of treatment states. Letting t denote the focal discontinuity period, the Average Direct Treatment Effect at the period-t cutoff subsequent to the treatment history $h_{t-1} \in \text{supp}(H_{t-1})$ is defined as

$$\text{ADTE}_{t,\tau}(h_{t-1}, c) \equiv \mathbb{E}\left[Y_{t+\tau}(h_{t-1}, 1, 0_{\bar{t}-t}) - Y_{t+\tau}(h_{t-1}, 0, 0_{\bar{t}-t}) \,|\, H_{t-1} = h_{t-1}, R_t = c\right] \quad (22)$$

for any $\tau \in \{0, 1, \ldots, \bar{t} - t\}$[5]. Importantly, potential outcomes differ solely by the period-t treatment state, implying that this class of impulse response parameters generalizes equation (21).

Because each $\text{ADTE}_{t,\tau}(h_{t-1}, c)$ is specific to a history of treatment states, focal period of treatment assignment, and number of leads, it is natural to think of $\text{ADTE}_{t,\tau}(h_{t-1}, c)$ as a building block to construct aggregate policy-relevant parameters. This interpretation

---

[5]It is not necessary for the conditioning set to include the event $H_{t-1} = h_{t-1}$. However, an empirical researcher is unlikely to be interested in the average effect of a treatment subsequent to a treatment history h_{t-1} in the subpopulation implied by a *different* treatment history h'_{t-1}. For practical purposes, it is natural for *one* treatment history to both be an argument of potential outcomes and belong to the conditioning set.

is similar in spirit to the one put forward by Callaway and Sant'Anna (2021), who proposes several approaches to aggregate cohort- and time-specific causal parameters in the context of difference-in-differences designs with staggered adoption of an absorbing treatment. Analogously, Bojinov, Rambachan and Shephard (2021) aggregates unit-time-specific causal parameters in sequentially randomized experiments. In dynamic RDDs, a natural first step is to aggregate target parameters over histories of treatment states, for given focal period of treatment assignment and number of leads. The resulting parameter is defined as

$$\text{ADTE}_{t,\tau}\left(c\right) \equiv \sum_{h_{t-1}\in\text{supp}(H_{t-1})} \text{ADTE}_{t,\tau}\left(h_{t-1},c\right) \times \mathbb{P}\left(H_{t-1} = h_{t-1}\right) \tag{23}$$

$\text{ADTE}_{t,\tau}\left(c\right)$ integrates Average Direct Treatment Effects over the probability distribution of histories up to the focal round of treatment assignment. It can thus be interpreted as the average impulse response of the outcome to the period-t discontinuity, τ periods after the initial shock.

In addition, an empirical researcher may be interested in aggregating time-specific parameters into a single summary measure of the τ-period-ahead effect of the treatment. In this case, assuming a balanced panel, each term of the aggregate impulse response function is defined as

$$\text{ADTE}_\tau\left(c\right) \equiv \frac{1}{\bar{t}-\tau} \sum_{t=1}^{\bar{t}-\tau} \text{ADTE}_{t,\tau}\left(c\right) \tag{24}$$

5.1 Dimension Reduction via Limited Path Independence

Assumptions 4, 5, and 6 are easily generalized to achieve the point identification of each possible $\text{ADTE}_{t,\tau}\left(h_{t-1},c\right)$. In observational settings, however, it is generally infeasible to estimate one target parameter for each possible treatment history, focal calendar time, and number of leads. For instance, in the canonical application of dynamic RDDs to the effects of local referenda, units are infrequently subject to treatment assignment because jurisdictions typically hold new elections only once funds approved by a previously approved proposition are close to expiration. As a consequence, only a limited number of local governments participates in treatment assignment every year and any given jurisdiction does so in a cyclical fashion. In addition, a narrowly rejected referendum is generally submitted to voters again only once or twice briefly after the initial attempt, implying that the identification

challenge posed by repeated treatment assignment is practically salient only in a narrow time window. To summarize, in the local public finance application of dynamic RDDs, treatment assignment is sparse and cyclical.

Motivated by this fact, it is natural to impose additional restrictions to limit the number of target parameters at stake and reduce the dimensionality of the problem. In a similar fashion to Hsu and Shen (2024), I assume that all causal parameters of interest share a weak form of limited path independence. Specifically, for any given focal round of treatment assignment, I restrict the Average Direct Treatment Effect not to depend on the history of treatment states provided that units are not exposed to the treatment for a specific number of periods prior to the focal round of treatment assignment.

Assumption 7 (Limited Path Independence) *Let $k \in \mathbb{N}_+$ be a deterministic constant. For any $h_{t-k-1}, h'_{t-k-1} \in \mathrm{supp}\,(H_{t-k-1})$ such that $h_{t-1} = (h_{t-k-1}, 0_k)$ and $h'_{t-1} = \left(h'_{t-k-1}, 0_k\right)$,*

$$\mathbb{E}\left[Y_t\left(h_{t-1}, 1, 0_{\bar{t}-t}\right) - Y_t\left(h_{t-1}, 0, 0_{\bar{t}-t}\right) | H_{t-1} = h_{t-1}, R_t = c\right]$$
$$= \mathbb{E}\left[Y_t\left(h'_{t-1}, 1, 0_{\bar{t}-t}\right) - Y_t\left(h'_{t-1}, 0, 0_{\bar{t}-t}\right) | H_{t-1} = h'_{t-1}, R_t = c\right]$$

for all $t > k$. Thus, $\mathrm{ADTE}_{t,\tau}\left(h_{t-1}, c\right) = \mathrm{ADTE}_{t,\tau}\left(h'_{t-1}, c\right)$ for all $\tau \in \{0, 1, \ldots, \bar{t} - t\}$.

The constant k is to be determined by the researcher depending on the empirical application[6]. Clearly, the choice of a larger k entails a more stringent requirement for the limited path independence assumption to be satisfied. On the other hand, a smaller k restricts ADTEs not to be determined in time periods further antecedent to the focal round of treatment assignment.

The practical implication of this Markov-type assumption is that, for the purpose of estimating a target parameter, units that share a common "recent" past can be pooled, thereby improving statistical inference. Another consequence of Assumption 7 is that the plausibility of the local common trends restriction, namely Assumption 6, may be tested as in difference-in-differences and event-study designs. Because all units are untreated for k periods before the focal calendar time period, a researcher may test whether the conditional mean of the outcome exhibits pre-trends at the focal cutoff.

---

[6]In a similar spirit, Callaway and Sant'Anna (2021) considers a limited treatment anticipation assumption, based on which the researcher must specify the number of time periods the outcome variable may reflect agents' knowledge of (or expectations over) future exposure to the treatment.

For ease of exposition, consider a four-period model $(\bar{t} = 4)$ in which all units are un-treated until $t = 3$, when a round of treatment assignment occurs. For $(t, \tau) = (3, 1)$, let the target parameter be

$$\text{ADTE}_{t,\tau}(0_2, c) \equiv \mathbb{E}\left[Y_{t+\tau}(0_2, 1, 0) - Y_{t+\tau}(0_2, 0, 0) \,|\, H_2 = 0_2, R_3 = c\right] \tag{25}$$

This is the cutoff-specific, one-period-ahead average effect of the treatment assigned in $t = 3$, conditional on a never-treated path in the first two periods. Because, by construction, $D_1 = D_2 = 0$, the history of treatment states $H_2 = 0_2$ is observed. To identify $\text{ADTE}_{3,1}(0_2, c)$, Assumption 6 restricts $\mathbb{E}\left[Y_4(0_2, d_1, 0) - Y_3(0_2, d_1, 0) \,|\, H_2 = 0_2, R_3 = c, P(0_2, d_1, 0) = 1\right]$ to be equal to $\mathbb{E}\left[Y_4(0_2, d_1, 0) - Y_3(0_2, d_1, 0) \,|\, H_2 = 0_2, R_3 = c, P(0_2, d_1, 1) = 1\right]$ for both $d_1 = 0$ and $d_1 = 1$. Leveraging the intuition that the path indicator random variable P describes a unit's treatment-taking behavior and thus their "type", it is natural to expect that a similar common trends assumption hold in time periods preceding the focal round of treatment assignment. In other words, if a unit's "type" is time-invariant, an empirical researcher may plausibly conjecture that

$$\mathbb{E}\left[Y_2(0_2, d_1, 0) - Y_1(0_2, d_1, 0) \,|\, H_2 = 0_2, R_3 = c, P(0_2, d_1, 0) = 1\right]$$
$$= \mathbb{E}\left[Y_2(0_2, d_1, 0) - Y_1(0_2, d_1, 0) \,|\, H_2 = 0_2, R_3 = c, P(0_2, d_1, 1) = 1\right] \tag{26}$$

for $d_1 \in \{0, 1\}$. If the outcome variable is non-anticipatory and Assumption 4 holds, these conditional expectations are identified for both the upper and lower limit of the running variable in $t = 3$. Specifically,

$$\lim_{r \downarrow c} \mathbb{E}\left[Y_2 - Y_1 | H_2 = 0_2, R_3 = r, D_4 = 0\right] = \lim_{r \downarrow c} \mathbb{E}\left[Y_2 - Y_1 | H_2 = 0_2, R_3 = r, D_4 = 1\right] \tag{27}$$

and symmetrically for the lower limit.

These equalities naturally give rise to testable restrictions aimed at assessing the plausibility of Assumption 6. Specifically, an empirical researcher may provide a more compelling argument in favor of the local common trends restriction by failing to reject the equality of average differences in outcome trends at the cutoff in all available "pre-periods". In the canonical difference-in-differences design, it is common practice to test the equality of pre-period outcome trends across groups implied by whether units are assigned to the "treatment" or "control" arm. In a dynamic RDD, groups are implicitly determined by the

treatment path subsequent to the focal time period. Unless more restrictions are placed on the evolution of treatment states, the number of groups or "types" will in general exceed two, thereby increasing the number of testable equalities.

Under Assumption 7, the focal time period is preceded by k periods for which $D_{t-1} = \cdots = D_{t-k} = 0$. Then, based on Assumption 6, the null hypothesis in equation (26) may be generalized to any sequence of future treatment states $d_{t+1}, \ldots, d_{t+\tau} \in \{0, 1\}$ and any pair of periods u and v such that $t - k \leq u < v \leq t$. Specifically,

$$\mathbb{E}\left[\Delta_{u,v}\left(h_{t-k-1}, 0_k, d_t, 0_\tau\right) | H_{t-1} = h_{t-1}, R_t = c, P\left(h_{t-k-1}, 0_k, d_t, 0_\tau\right) = 1\right]$$
$$= \mathbb{E}\left[\Delta_{u,v}\left(h_{t-k-1}, 0_k, d_t, 0_\tau\right) | H_{t-1} = h_{t-1}, R_t = c, P\left(h_{t-k-1}, 0_k, d_t, d_{t+1}, \ldots, d_{t+\tau}\right) = 1\right] \quad (28)$$

with $\Delta_{u,v}\left(h_{t-k-1}, 0_k, d_t, 0_\tau\right) \equiv Y_v\left(h_{t-k-1}, 0_k, d_t, 0_\tau\right) - Y_u\left(h_{t-k-1}, 0_k, d_t, 0_\tau\right)$. Because units are untreated in the k periods preceding t, both conditional expectations are identified using the upper or lower limit of R_t:

$$\lim_{r \downarrow c} \mathbb{E}\left[Y_v - Y_u \Big| H_{t-1} = h_{t-1}, R_t = r, \bigcap_{s=1}^\tau \{D_{t+s} = 0\}\right]$$
$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_v - Y_u \Big| H_{t-1} = h_{t-1}, R_t = r, \bigcup_{s=1}^\tau \{D_{t+s} \neq 0\}\right] \quad (29)$$

and the equality of lower limits follows from a symmetric argument. Appendix B.5 develops statistical theory to conduct tests of hypotheses pertaining to the plausibility of the common trends assumption.

6   Estimation and Statistical Inference

In this section, I illustrate that standard local polynomial methods are applicable to the estimation of Average Direct Treatment Effects in a dynamic RDD. Hereafter, I do not explicitly condition on the history of treatment states $H_{t-1} = h_{t-1}$ to keep notation concise. Then, by an application of the Law Iterated Expectations and the properties of limits, the target estimand from Proposition 6 can be equivalently expressed as

$$\text{ADTE}_{t,\tau}(c) = \lim_{r \downarrow c} \mathbb{E}\left[Y_t | R_t = r\right] - \lim_{r \uparrow c} \mathbb{E}\left[Y_t | R_t = r\right]$$
$$+ \frac{\lim_{r \downarrow c} \mathbb{E}\left[(Y_{t+\tau} - Y_t) \prod_{s=1}^\tau (1 - D_{t+s}) | R_t = r\right]}{\lim_{r \downarrow c} \mathbb{E}\left[\prod_{s=1}^\tau (1 - D_{t+s}) | R_t = r\right]}$$

$$- \frac{\lim_{r \uparrow c} \mathbb{E}\left[(Y_{t+\tau} - Y_t) \prod_{s=1}^{\tau} (1 - D_{t+s}) \,|\, R_t = r\right]}{\lim_{r \uparrow c} \mathbb{E}\left[\prod_{s=1}^{\tau} (1 - D_{t+s}) \,|\, R_t = r\right]} \tag{30}$$

In this expression, the parameter of interest is a nonlinear function of cutoff-specific conditional means, all of which can be nonparametrically estimated in the full sample, rather than in subsamples implied by realizations of $\{D_{t+s}\}_{s=1}^{\tau}$. This feature will prove convenient for the purpose of constructing robust nonparametric confidence intervals. To keep notation compact, I also define the following random variables:

$$R \equiv R_t \qquad Y \equiv Y_t \qquad W \equiv (Y_{t+\tau} - Y_t) \prod_{s=1}^{\tau} (1 - D_{t+s}) \qquad D \equiv \prod_{s=1}^{\tau} (1 - D_{t+s}) \tag{31}$$

For any random variable A, let $\mu_A(r) \equiv \mathbb{E}[A|R = r]$ and $\sigma_{AA}^2(r) \equiv \mathbb{V}[A|R = r]$. Furthermore, let $\mu_{A+} \equiv \lim_{r \downarrow c} \mu_A(r)$ and $\mu_{A-} \equiv \lim_{r \uparrow c} \mu_A(r)$. Then the parameter of interest can be written as

$$\theta \equiv \mu_{Y+} - \mu_{Y-} + \frac{\mu_{W+}}{\mu_{D+}} - \frac{\mu_{W-}}{\mu_{D-}} \tag{32}$$

Prior to discussing statistical inference on θ, Assumption 8 outlines a number of necessary conditions pertaining to moments and features of the running variable. These requirements match those listed in Assumptions 1 and 3 of Calonico, Cattaneo and Titiunik (2014a).

Assumption 8 *Let $\delta \in \mathbb{N}$. For $\kappa_0 \in \mathbb{R}_{++}$, the following conditions hold in a neighborhood of the cutoff $(c - \kappa_0, c + \kappa_0)$.*

(a) The density $f_R(r)$ is continuous and bounded away from zero.

(b) $\mathbb{E}[Y^4|R = r]$ is bounded.

(c) $\mu_Y(r)$, $\mu_W(r)$, $\mu_D(r)$ are δ times continuously differentiable.

(d) $\sigma_{YY}^2(r)$, $\sigma_{WW}^2(r)$, $\sigma_{DD}^2(r)$ are continuous and bounded away from zero.

Consider a random sample $\left\{[R_i, Y_i, W_i, D_i]'\right\}_{i=1}^{n}$. Given a positive bandwidth $h_n > 0$, let $\mathcal{R}(h_n) = [c - h_n, c + h_n]$ be a discontinuity window implied by realizations of the running variable around the first-period cutoff. Moreover, let $\mathcal{R}^-(h_n) = [c - h_n, c)$ and $\mathcal{R}^+(h_n) = [c, c + h_n]$ indicate the left and right discontinuity windows, respectively. Throughout this section, I employ standard local polynomial estimators to approximate the value of several

24

conditional means on either side of the cutoff. Specifically, for any random variable A, I consider local linear estimators defined as

$$\left[\widehat{\mu}_{A+,1}^{(0)}\left(h_n\right), \widehat{\mu}_{A+,1}^{(1)}\left(h_n\right)\right]' \equiv \arg\min_{b_0,b_1\in\mathbb{R}} \sum_{i=1}^{n} \mathbb{I}\left[R_i \in \mathcal{R}^+\left(h_n\right)\right]\left(A_i - b_0 - b_1 R_i\right)^2 k_{h_n}\left(R_i\right) \quad (33)$$

$$\left[\widehat{\mu}_{A-,1}^{(0)}\left(h_n\right), \widehat{\mu}_{A-,1}^{(1)}\left(h_n\right)\right]' \equiv \arg\min_{b_0,b_1\in\mathbb{R}} \sum_{i=1}^{n} \mathbb{I}\left[R_i \in \mathcal{R}^-\left(h_n\right)\right]\left(A_i - b_0 - b_1 R_i\right)^2 k_{h_n}\left(R_i\right) \quad (34)$$

where $k_h\left(u\right) \equiv k\left(\frac{u}{h}\right)/h$, with $k\left(\cdot\right)$ denoting a kernel function. Mirroring Assumption 2 in Calonico, Cattaneo and Titiunik (2014a), Assumption 9 below reports minimal conditions for the validity of $k\left(\cdot\right)$.

Assumption 9 *For $\kappa \in \mathbb{R}_{++}$, the kernel function $k\left(\cdot\right) : [0,\kappa] \to \mathbb{R}$ is bounded and nonnegative, zero outside its support, and positive and continuous on $(0,\kappa)$.*

In the notation of equations (33) and (34), the subscript p in $\widehat{\mu}_{+,p}^{(\nu)}$ indicates the chosen polynomial degree in the regression specification, while the superscript ν denotes the order of the derivative of interest. For example, in standard sharp designs, the estimated coefficient of interest is $\widehat{\mu}_{+,1}^{(0)}\left(h_n\right) - \widehat{\mu}_{-,1}^{(0)}\left(h_n\right)$, namely the difference between two intercepts in a local linear regression specification. In this paper's dynamic design, the estimator of interest replaces the conditional means in equation (30) with their sample counterparts. As a matter of fact, a natural estimator for θ is

$$\widehat{\theta}\left(h_n\right) \equiv \widehat{\mu}_{Y+,1}^{(0)}\left(h_n\right) - \widehat{\mu}_{Y-,1}^{(0)}\left(h_n\right) + \frac{\widehat{\mu}_{W+,1}^{(0)}\left(h_n\right)}{\widehat{\mu}_{D+,1}^{(0)}\left(h_n\right)} - \frac{\widehat{\mu}_{W-,1}^{(0)}\left(h_n\right)}{\widehat{\mu}_{D-,1}^{(0)}\left(h_n\right)} \quad (35)$$

As in any regression discontinuity design, the estimator in equation (35) is a function of the width of the discontinuity window h_n. The choice of a larger bandwidth may lead to more precise estimates, but also increase their bias relative to the true intercept at the cutoff. Concerns about bias are particularly salient when adopting the prevalent data-driven approach to bandwidth selection, namely choosing one that minimizes the mean squared error of the estimator (Imbens and Kalyanaraman 2012). A natural solution is to first estimate the bias using a pilot bandwidth b_n and then subtract the estimated bias from the point estimate of interest. Calonico, Cattaneo and Titiunik (2014a) shows how to account for the statistical uncertainty stemming from the first step and construct robust nonparametric

confidence intervals for canonical estimands in sharp and fuzzy regression discontinuity and kink designs.

In this section, I leverage the same approach to construct robust nonparametric confidence intervals for the target estimand in equation (30). The choice of notation and the structure of proofs largely mirror those in Calonico, Cattaneo and Titiunik (2014b). Similarly to standard estimators in fuzzy regression discontinuity and kink designs, the estimator in equation (35) is nonlinear. Thus, to obtain the bias expression in equation (39), I proceed as in Calonico, Cattaneo and Titiunik (2014a) and linearize $\widehat{\theta}(h_n)$ with a second-order Taylor expansion. The resulting estimator is defined as $\widetilde{\theta}(h_n) \equiv \widetilde{\eta}_+(h_n) - \widetilde{\eta}_-(h_n)$, with

$$\widetilde{\eta}_+(h_n) \equiv \frac{1}{\mu_{D+}}\left(\widehat{\mu}_{W+,1}^{(0)}(h_n) - \mu_{W+}\right) - \frac{\mu_{W+}}{\mu_{D+}^2}\left(\widehat{\mu}_{D+,1}^{(0)}(h_n) - \mu_{D+}\right) \tag{36}$$

$$\widetilde{\eta}_-(h_n) \equiv \frac{1}{\mu_{D-}}\left(\widehat{\mu}_{W-,1}^{(0)}(h_n) - \mu_{W-}\right) - \frac{\mu_{W-}}{\mu_{D-}^2}\left(\widehat{\mu}_{D-,1}^{(0)}(h_n) - \mu_{D-}\right) \tag{37}$$

Letting $\widehat{\mathrm{B}}(h_n, b_n)$ denote the estimated bias of $\widetilde{\theta}(h_n)$, the bias-corrected estimator of θ is

$$\widehat{\theta}^{\mathrm{BC}}(h_n, b_n) \equiv \widehat{\theta}(h_n) - h_n^2 \widehat{\mathrm{B}}(h_n, b_n) \tag{38}$$

with

$$
\begin{aligned}
\widehat{\mathrm{B}}(h_n, b_n) \equiv &\left[\frac{\widehat{\mu}_{Y+,2}^{(2)}(b_n)}{2} + \frac{1}{\widehat{\mu}_{D+,1}^{(0)}(h_n)}\frac{\widehat{\mu}_{W+,2}^{(2)}(b_n)}{2} - \frac{\widehat{\mu}_{W+,1}^{(0)}(h_n)}{\left(\widehat{\mu}_{D+,1}^{(0)}(h_n)\right)^2}\frac{\widehat{\mu}_{D+,2}^{(2)}(b_n)}{2}\right]\mathcal{B}_+(h_n) \\
&- \left[\frac{\widehat{\mu}_{Y-,2}^{(2)}(b_n)}{2} + \frac{1}{\widehat{\mu}_{D-,1}^{(0)}(h_n)}\frac{\widehat{\mu}_{W-,2}^{(2)}(b_n)}{2} - \frac{\widehat{\mu}_{W-,1}^{(0)}(h_n)}{\left(\widehat{\mu}_{D-,1}^{(0)}(h_n)\right)^2}\frac{\widehat{\mu}_{D-,2}^{(2)}(b_n)}{2}\right]\mathcal{B}_-(h_n)
\end{aligned}
\tag{39}
$$

where $\mathcal{B}_+(h_n)$ and $\mathcal{B}_-(h_n)$ are observed and asymptotically bounded quantities defined in Appendix B.2.1. Proposition 7 below establishes the asymptotic distribution of the bias-corrected estimator $\widehat{\theta}^{\mathrm{BC}}(h_n, b_n)$.

Proposition 7 (Asymptotic Distribution) *Suppose that Assumptions 8 and 9 hold. Let* $\mu_{D+} \neq 0$ *and* $\mu_{D-} \neq 0$. *If* $n\min\{h_n^5, b_n^5\}\max\{h_n^2, b_n^2\} \to 0$ *and* $n\min\{h_n, b_n\} \to \infty$,

$$T^{\mathrm{RBC}}(h_n, b_n) \equiv \frac{\widehat{\theta}^{\mathrm{BC}}(h_n, b_n) - \theta}{\sqrt{\mathrm{V}^{\mathrm{BC}}(h_n, b_n)}} \xrightarrow{d} \mathcal{N}(0,1)$$

provided that $h_n \to 0$ *and* $\kappa b_n < \kappa_0$. *The exact expression for* $\mathrm{V}^{\mathrm{BC}}(h_n, b_n)$ *is provided in Appendix B.3.6.*

6.1 Bandwidth Selection

I adopt the solution proposed by Calonico, Cattaneo and Titiunik (2014a) by computing both an optimal main bandwidth h and an optimal pilot bandwidth b. The main bandwidth is used to estimate the parameter of interest, while the pilot bandwidth is employed to estimate its bias.

Following Imbens and Kalyanaraman (2012), I define optimality in terms of minimizing the mean squared error (MSE) of the estimator. Two distinct MSE criteria are considered depending on the bandwidth. First, the main bandwidth minimizes the MSE of the linearized estimator $\widetilde{\theta}(h_n)$, which consists of a linear combination of intercepts in local linear regression specifications. Second, the pilot bandwidth minimizes the MSE of an alternative linearized estimator $\widetilde{\theta}_{2,2}(h_n)$, in which the estimated intercepts are replaced by second-order derivatives in local quadratic specifications and, analogously, population means are replaced by the corresponding second-order derivatives[7]. The following proposition provides expression for both MSEs and derives the corresponding optimal bandwidths.

Proposition 8 *Suppose Assumptions 8 and 9 hold with $\delta \geq 3$. Let $h_n \to 0$ and $nh_n \to \infty$. The mean squared error of the linearized estimator $\widetilde{\theta}(h_n)$ is*

$$MSE\left(\widetilde{\theta}(h_n)\right) = h_n^4 \left(\dot{\mathrm{B}}_h^2 + o_p(1)\right) + \frac{1}{nh_n}\left(\dot{\mathrm{V}}_h + o_p(1)\right)$$

and the MSE-optimal main bandwidth is

$$h_{\mathrm{MSE}} = \left(\frac{\dot{\mathrm{V}}_h}{4\dot{\mathrm{B}}_h^2}\right)^{1/5} n^{-1/5}$$

The mean squared error of the linearized estimator $\widetilde{\theta}_{2,2}(h_n)$ is

$$MSE\left(\widetilde{\theta}_{2,2}(h_n)\right) = h_n^2 \left(\dot{\mathrm{B}}_b^2 + o_p(1)\right) + \frac{1}{nh_n^5}\left(\dot{\mathrm{V}}_b + o_p(1)\right)$$

and the MSE-optimal pilot bandwidth is

$$b_{\mathrm{MSE}} = \left(\frac{5\dot{\mathrm{V}}_b}{2\dot{\mathrm{B}}_b^2}\right)^{1/7} n^{-1/7}$$

Exact expressions for $\dot{\mathrm{B}}_h$, $\dot{\mathrm{V}}_h$, $\dot{\mathrm{B}}_b$, and $\dot{\mathrm{V}}_b$ are provided in Appendix B.4.5.

---

[7]In general, $\widetilde{\theta}_{\nu,p}(h_n)$ denotes the linearized estimator involving νth-order derivatives in local polynomial specifications of order p. Thus, the main linearized estimator $\widetilde{\theta}(h_n)$ corresponds to $\widetilde{\theta}_{0,1}(h_n)$.

It is important to note that the MSE-optimal bandwidths specified in Proposition 8 are asymptotic and, hence, not directly implementable. In practice, I follow the steps outlined in Section S.2.6 of Calonico, Cattaneo and Titiunik (2014b) and employ analogous direct plug-in bandwidth selectors.

6.2   Standard Errors

The variance of the bias-corrected estimator $\mathrm{V}^{\mathrm{BC}}(h_n, b_n)$ in Proposition 7 is asymptotic. To compute standard errors and conduct statistical inference on the parameter of interest, it is natural to replace unknown population moments with their sample analogues. First, note that the linearized estimator in equations (36) and (37) is defined as a linear combination of estimated intercepts, where the constant coefficients are unknown population means. I replace these means with their corresponding estimates computed within each discontinuity half-window. Second, population variances and covariances are unknown. For any pair of random variables A and B, define these second-order moments as $\sigma_{AB+}^2 \equiv \lim_{r \downarrow c} \mathbb{C}[A, B|R = r]$ and $\sigma_{AB-}^2 \equiv \lim_{r \uparrow c} \mathbb{C}[A, B|R = r]$. To estimate these moments, I follow the approach in Abadie and Imbens (2006) and Calonico, Cattaneo and Titiunik (2014a), employing a nearest-neighbor method that computes each sample covariance using a fixed number of observations. Specifically, letting j^* denote this tuning parameter,

$$\widehat{\sigma}_{AB+}^2 (R_i) \equiv \mathbb{I}[R_i \geq c] \frac{j^*}{j^* + 1} \left(A_i - \sum_{j \in \mathcal{J}_i^+} \frac{A_j}{j^*} \right) \left(B_i - \sum_{j \in \mathcal{J}_i^+} \frac{B_j}{j^*} \right) \tag{40}$$

$$\widehat{\sigma}_{AB-}^2 (R_i) \equiv \mathbb{I}[R_i < c] \frac{j^*}{j^* + 1} \left(A_i - \sum_{j \in \mathcal{J}_i^-} \frac{A_j}{j^*} \right) \left(B_i - \sum_{j \in \mathcal{J}_i^-} \frac{B_j}{j^*} \right) \tag{41}$$

where $\mathcal{J}_i^+$ and $\mathcal{J}_i^-$ denote the sets of the j^* closest units to unit i in the right and left discontinuity half-windows, respectively. Once all of the first- and second-order moments are estimated, the resulting estimated variance $\widehat{\mathrm{V}}^{\mathrm{BC}}(h_n, b_n)$ can be used to compute the standard error of $\widehat{\theta}^{\mathrm{BC}}(h_n, b_n)$ and construct confidence intervals for the target parameter.

7   Simulation

In this section, I apply the methods developed in the preceding sections to simulated data. I employ a data generating process similar to that presented by Biasi, Lafortune

and Schönholzer (2023)[8], which models both school districts' decisions to hold local referenda and the subsequent evolution of an outcome of interest based on referendum results. The primary objective of this exercise is to demonstrate that the estimator developed in this paper effectively recovers the parameter of interest.

7.1 Data Generating Process

Let $i \in \{1, \ldots, n\}$ and $t \in \{1, \ldots, \bar{t}\}$ index schools districts and calendar years, respectively. Let Q_{it} be a Bernoulli random variable equal to one if district i holds a referendum in year t. Denote by R_{it} the approval vote margin and define $D_{it} \equiv \mathbb{I}[R_{it} \geq 0]$ as the indicator for referendum approval. In each period, a district's decision to hold a referendum is influenced, among other factors, by whether referenda were recently approved or rejected. Specifically,

$$Q_{it} = \mathbb{I}\left[A_i^q + B_t^q + U_{it}^q + \sum_{s=1}^{\bar{s}} \rho_s^q D_{i,t-s} + \sum_{s=1}^{\bar{s}} \delta_s^q Q_{i,t-s}\left(1 - D_{i,t-s}\right) > \bar{q}\right] \tag{42}$$

where A_i^q is a district-specific intercept, B_t^q is a time-specific intercept, U_{it}^q is an idiosyncratic shock, and $\bar{q}$ is a deterministic threshold that must be exceeded for a referendum to be proposed. If a referendum is held, the approval vote margin is determined according to a similar statistical model:

$$R_{it} = A_i^r + B_t^r + U_{it}^r + \sum_{s=1}^{\bar{s}} \rho_s^r D_{i,t-s} + \sum_{s=1}^{\bar{s}} \delta_s^r Q_{i,t-s}\left(1 - D_{i,t-s}\right) + \sum_{s=1}^{\bar{s}} \gamma_s\left(1 - Q_{i,t-s}\right) \tag{43}$$

where A_i^r and B_t^r are district- and time-specific intercepts, respectively, and U_{it}^r is an idiosyncratic shock. This specification allows the vote margin to flexibly depend on whether similar ballot measures were recently proposed and, if so, whether they were approved or rejected. The outcome of interest is generated according to

$$Y_{it} = \sum_{g=1}^{t} \theta_{igt} D_{ig} + \eta \overline{R}_i + V_{it} \tag{44}$$

where V_{it} is an idiosyncratic shock and $\overline{R}_i \equiv \left(\sum_{g=1}^{\bar{t}} Q_{ig}\right)^{-1} \sum_{g=1}^{\bar{t}} Q_{ig} R_{ig}$ represents the average approval vote margin for district i over the period of interest. Importantly, the stochastic treatment effect θ_{igt} is determined by

$$\theta_{igt} = \theta_{0g} + \theta_{1g}\left(t - g\right) + \phi R_{ig} \tag{45}$$

---

[8]Specifically, the authors conduct a Monte Carlo simulation in Appendix B.

This formulation allows the effect of referendum approval to vary across districts, calendar times, and cohorts (with a cohort defined as the calendar time at which a referendum is approved). Here, θ_{0g} is a cohort-specific random variable capturing the permanent effect of a referendum approved in calendar time g, and θ_{1g} is a cohort-specific random variable that measures the incremental effect per unit of time elapsed since approval, meaning that this additional effect is proportional to the relative time $t - g$. Finally, ϕ is a deterministic coefficient that quantifies the extent to which the treatment effect varies with the magnitude of the approval vote margin.

7.2 Parameterization

The distributional assumptions and parameter values largely mirror those in Biasi, Lafortune and Schönholzer (2023). In the referendum equation (42), the unit intercept is assumed to be $A_i^q \sim \mathcal{N}\left(0, \sigma_q^2\right)$, the time intercept is $B_t^q \sim \mathcal{N}\left(0, \sigma_q^2\right)$, and the idiosyncratic shock is $U_{it}^q \sim \mathcal{N}\left(0, \sigma_{uq}^2\right)$. Similarly, in the vote margin equation (43), $A_i^r \sim \mathcal{N}\left(0, \sigma_r^2\right)$, $B_t^r \sim \mathcal{N}\left(0, \sigma_r^2\right)$, and $U_{it}^r \sim \mathcal{N}\left(0, \sigma_{ur}^2\right)$. In both cases, draws are independently distributed. In the outcome equation (44), the idiosyncratic shock is given by $V_{it} \sim \mathcal{N}\left(0, \sigma_v^2\right)$. In the treatment effect equation (45), the permanent effect is modeled as $\theta_{0g} \sim \mathcal{N}\left(\overline{\theta}_0, \sigma_0^2\right)$ and the additional, time-proportional effect is $\theta_{1g} \sim \mathcal{N}\left(\overline{\theta}_1, \sigma_1^2\right)$. I simulate data for $n = 30,000$ school districts over a period of $\overline{t} = 30$ years. To fix initial conditions, I set $Q_{i0} \sim \text{Binomial}\left(1, 0.1\right)$ and, conditional on $Q_{i0} = 1$, $R_{i0} \sim \mathcal{N}\left(\mu_r, \sigma_r^2\right)$. The chosen parameter values are as follows: $\overline{q} = 1.7$; $\sigma_q^2 = 0.5$; $\sigma_{uq}^2 = 1$; $\sigma_r^2 = 0.01$; $\sigma_{ur}^2 = 0.02$; $\mu_r = 0.05$; $\sigma_v^2 = 0.2$; $\eta = 0$; $\phi = 0.1$; $\overline{\theta}_0 = 0.4$; $\overline{\theta}_1 = -0.1$; $\sigma_0^2 = 0.08^2$; $\sigma_1^2 = 0.02^2$. In addition, I cap the incremental effect per unit of time at $t - g = 4$, so that $\theta_1\left(t - g\right) = 4 \times \theta_1$ for any $t - g > 4$. Finally, I set $\overline{s} = 3$ and $(\rho_1^q, \rho_2^q, \rho_3^q) = (-0.9, -0.5, -0.1)$, $(\delta_1^q, \delta_2^q, \delta_3^q) = (0.5, 0.3, 0.1)$, $(\rho_1^r, \rho_2^r, \rho_3^r) = (-0.05, -0.03, -0.01)$, $(\delta_1^r, \delta_2^r, \delta_3^r) = (0.05, 0.03, 0.01)$, and $(\gamma_1, \gamma_2, \gamma_3) = (0.05, 0.03, 0.01)$.

7.3 Target Parameters

By definition, θ_{igt} represents the unit-specific effect of a referendum approved in period g on the outcome observed in period t. Equivalently, letting $\tau = t - g$ denote the number of periods since approval, θ_{igt} can be written as $\theta_{ig\tau}$. The average effect of referendum approval

for cohort g at relative time τ is then

$$\theta_{g\tau} \equiv \mathbb{E}[\theta_{ig\tau}] = \theta_{0g} + \tau\theta_{1g} + \phi\mathbb{E}[R_{ig}] \tag{46}$$

where the expectation is taken over the distribution of R_{ig} across units in cohort g. As in any regression discontinuity design, the running variable must be evaluated at the cutoff, implying that the target parameter of interest is $\theta_{0g} + \tau\theta_{1g}$. Using the notation introduced earlier, this corresponds to $\text{ADTE}_{g,\tau}(0)$, i.e., the τ-period-ahead Average Direct Treatment Effect when the focal discontinuity occurs in period g. To summarize effects across cohorts, I aggregate the cohort-specific parameters for each relative time τ following equation (24):

$$\text{ADTE}_\tau(0) \equiv \sum_{g=1}^{\bar{t}-\tau} \omega_{g\tau}\text{ADTE}_{g,\tau}(0) \tag{47}$$

where the weight $\omega_{g\tau} \equiv \mathbb{P}(Q_{ig} = 1)/\sum_{\ell=1}^{\bar{t}-\tau}\mathbb{P}(Q_{i\ell} = 1)$ is the probability that a referendum occurs in period g, conditional on a referendum occurring at any time within the window $g \leq \bar{t}-\tau$. This weighting ensures that the aggregate effect appropriately reflects the distribution of referenda over time (Callaway and Sant'Anna 2021).

The remainder of this section presents a simulation exercise that evaluates how accurately the estimated $\text{ADTE}_\tau(0)$ recovers the true average effect at relative time τ, defined as

$$\theta_\tau \equiv \mathbb{E}[\theta_{g\tau}] = \mathbb{E}[\theta_{0g}] + \tau\mathbb{E}[\theta_{1g}] \tag{48}$$

where the expectation is taken over the distribution of θ_{0g} and θ_{1g} across cohorts.

7.4 Results

To reduce the number of estimands, I leverage Assumption 7 by defining a cohort g as the set of jurisdictions that held a referendum in period g and did not approve a ballot measure in any of the preceding $k = 3$ periods, i.e., $Q_{ig} = 1$ and $D_{i,g-1} = D_{i,g-2} = D_{i,g-3} = 0$. Then, for each cohort and for relative times $\tau \in \{1, 2, 3, 4, 5\}$, I estimate $\text{ADTE}_{g,\tau}(0)$ using a local linear specification within the discontinuity window implied by the MSE-optimal bandwidth, as detailed in Section 6.1. I then aggregate cohort-specific point estimates by replacing the weights in equation (47) with their sample counterparts in the simulated data. To estimate the instantaneous causal parameter $\text{ADTE}_{g,0}(0)$ and to conduct a placebo analysis on pre-period outcomes ($\tau \in \{-3, -2, -1\}$), I use standard (i.e., static) sharp RD estimators.

Figure 2 below reports the true DGP-implied treatment effects alongside the corresponding estimates, all of which are averaged over 250 replications.

Figure 2: Comparison of True and Estimated $\{\theta_\tau\}_{\tau=-3}^{5}$
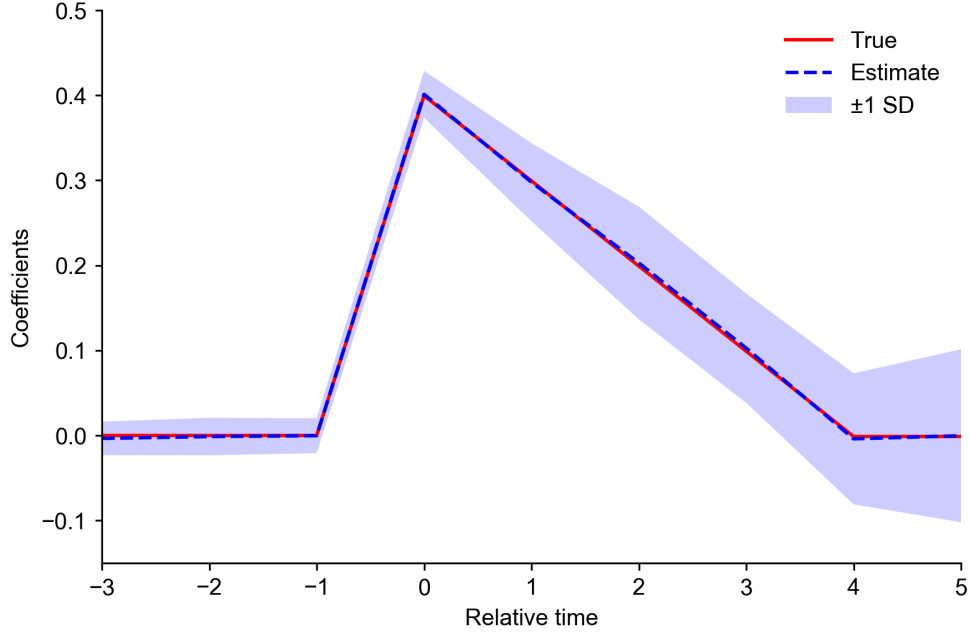
NOTES: This figure presents the results of a simulation whose data generating process is described in Section 7.1 and whose parameterization is detailed in Section 7.2. The continuous red line shows the true target parameter θ_τ at each relative time τ, as introduced in Section 7.3, while the dashed blue line represents the corresponding estimates derived from equation (47). Both lines are averaged over 250 replications. The shaded region around the dashed blue line shows the pointwise Monte Carlo variability of the estimates of θ_τ, measured by the sample standard deviation across 250 replications.

Figure 2 confirms that the estimated ADTEs accurately capture the dynamic effects of referendum approval on the outcome. As expected, outcomes measured in negative relative time periods remain unaffected by the approval of referenda in period $t = g$. In Section 5.1, I suggest that pre-period outcomes can serve as a useful check on the plausibility of the local common trends assumption. For example, for any cohort g, consider the first difference $\Delta_{g-3,g-1} \equiv Y_{g-1} - Y_{g-3}$. By construction, this difference reflects untreated potential outcomes. It is therefore natural to test whether, near the cutoff at $\tau = 0$, the difference $\Delta_{g-3,g-1}$ is mean independent of the treatment path following the focal referendum. For instance, one might compare the $D_{g+1} = D_{g+2} = D_{g+3} = D_{g+4} = D_{g+5} = 0$ path with any

alternative path involving at least one treated period:

$$\lim_{r\downarrow 0}\mathbb{E}\left[\Delta_{g-3,g-1}\middle|R_g=r,\bigcap_{s=1}^{5}\{D_{g+s}=0\}\right]=\lim_{r\downarrow 0}\mathbb{E}\left[\Delta_{g-3,g-1}\middle|R_g=r,\bigcup_{s=1}^{5}\{D_{g+s}\neq 0\}\right] \quad (49)$$

and similarly for the lower limit. I conduct a test of the null hypothesis implied by the equality in (49) and its symmetric counterpart for $r\uparrow 0$. Following the procedure described in Appendix B.5, I construct a test statistic for these two equalities and compare its value in each simulated sample to the appropriate quantile of a $\chi^2(2)$-distributed random variable. Across 250 replications, the null hypothesis is not rejected at the conventional 0.05 significance level in approximately 92 percent of cases.

8 Empirical Application: School District Expenditure Authorization and Housing Prices in Wisconsin

In this section, I apply the proposed method to estimate the intertemporal effects of changes in school district expenditures on housing prices in Wisconsin. Due to property tax limits imposed by state governments, local jurisdictions often face constraints in their fiscal policy decisions (Lincoln Institute of Land Policy and George Washington Institute of Public Policy, 2025). To secure additional funding, school districts may hold referenda seeking voter authorization for property tax increases designated to finance specific expenditure initiatives.

Housing prices are chosen as the outcome variable for two primary reasons. First, research in local public finance has long recognized that the extent to which property tax changes are capitalized into housing prices is an informative measure of the efficiency of local public goods provision (Bickerdike 1902, Marshall 1948, Oates 1969, Brueckner 1982, Cushing 1984, Barrow and Rouse 2004, Figlio and Lucas 2004, Cellini, Ferreira and Rothstein 2010, Biasi, Lafortune and Schönholzer 2025). Evidence of positive capitalization following an expenditure increase is typically interpreted as an indication that prospective homebuyers value the associated improvement in public service provision more than the change in the tax burden required to fund it. Second, many capital investment projects approved through school referenda unfold over multiple years, and their long-term implications may not be fully internalized by potential buyers at the time of the vote. This feature underscores the importance of identifying interpretable dynamic treatment effects over extended horizons.

8.1 Data

I focus my application on the state of Wisconsin, where school districts routinely hold referenda to authorize both operational and capital expenditures (Baron 2022). The Wisconsin Department of Public Instruction collects and publishes comprehensive data on all such local referenda held in the state since 1990 (Wisconsin Department of Public Instruction 2025). This dataset includes, among other variables, information on the approval vote share, which I use as the running variable in the dynamic RDD.

To construct the outcome of interest, I follow an approach similar to that of Biasi, Lafortune and Schönholzer (2025). Specifically, I rely on a repeat-sales house price index developed by Contat and Larson (2024), which covers all Census tracts located within Core-Based Statistical Areas[9] in the United States from 1989 to 2021. The index is normalized to 100 in 1989 for all tracts, allowing for within-tract temporal comparisons but not cross-sectional ones. To allow for level comparisons across school districts, I incorporate data on the average value of owner-occupied single-family homes at the Census tract level, as reported in the U.S. Census Bureau's 2000 Decennial Census[10]. For each tract, I compute a calibration factor as the ratio of the 2000 Census home value to the 2000 value of the house price index from Contat and Larson (2024), and apply this factor to the full time series of the index. The resulting measure of housing prices allows for both cross-sectional and intertemporal comparisons. Next, I compute the centroid of each Census tract and assign it to the corresponding elementary, secondary, or unified school district based on the 2010 TIGER/Line shapefiles provided by the U.S. Census Bureau[11] (U.S. Census Bureau 2025b). Finally, for each district, I calculate a population-weighted average of housing prices across its constituent Census tracts[12]. This yields the outcome variable used in the dynamic RDD.

The matched sample includes 1,001 referenda[13], of which 57.6 percent were approved. The average approval vote share margin is 2.05 percentage points, with an average of 9.86

---

[9]The term "Core-Based Statistical Area" refers collectively to both Metropolitan Statistical Areas and Micropolitan Statistical Areas (U.S. Census Bureau 2025a).

[10]The collection of this variable was discontinued beginning with the 2010 Decennial Census.

[11]I use 2010 tract and school district boundaries because the house price index constructed by Contat and Larson (2024) is based on 2010 Census tracts.

[12]Although I compute population-weighted averages, this choice is not consequential, as Census tracts are designed to contain approximately 4,000 inhabitants (U.S. Census Bureau 2025a).

[13]To construct a balanced panel in event time, I restrict the sample to referenda with outcomes observed in all relative years from –5 to 5, where relative year is defined with respect to the year of the referendum.

percentage points among approved referenda and –8.57 percentage points among those that were rejected. Figure 3 displays a histogram of the approval vote share margin. To assess the validity of the design, I test for discontinuities in the density of the running variable at the cutoff using the local polynomial density estimators developed by Cattaneo, Jansson and Ma (2020). The null hypothesis of equal densities on either side of the cutoff is not rejected (p-value $= 0.80$), suggesting that manipulation of the running variable around the threshold is unlikely to be a concern in this setting.

Figure 3: Density of the Approval Vote Share Margin
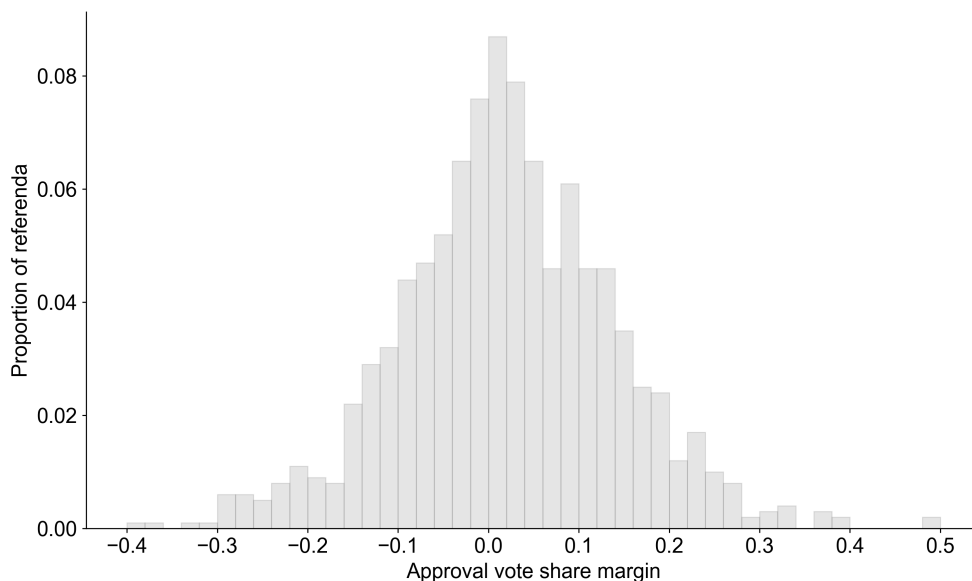
NOTES: This figure displays a histogram of the approval vote share margin, defined as the difference between the share of votes in favor of the proposed expenditure measure and the 50 percent approval threshold, for 1,001 referenda held by Wisconsin school districts between 1995 and 2015. The sample is restricted to districts located within Census-defined Core-Based Statistical Areas.

8.2 Results

As in Section 7, I leverage Assumption 7 by defining a cohort as the set of jurisdictions that held a referendum in a given year and did not approve a ballot measure in any of the preceding three years. Given the limited sample size, estimating a separate parameter for each cohort is not feasible. To address this constraint, I partition the sample into groups of adjacent years, ensuring that each group contains a roughly similar number of observations. The results presented in this section are based on a specification with six groups, each comprising

approximately 150 observations. The findings are broadly robust to alternative groupings[14].
To quantify statistical uncertainty, I compute standard errors using the nearest-neighbor
method described in Section 6.2, setting the tuning parameter to $j^* = 3$[15]. Cohort-specific
standard errors are then aggregated using the Delta method.

Table 1: Estimated ADTEs of Referendum Approval on Housing Prices

| | Year Relative to Referendum | | | | |
	1	2	3	4	5
Panel A: Cohort-by-Cohort Estimates					
1995-1996	-0.033	0.302	0.584	0.547	-0.025
	(0.198)	(0.267)	(0.456)	(0.382)	(0.165)
1997-1998	-0.099	-0.069	-0.103	-0.177	
	(0.174)	(0.177)	(0.191)	(0.196)	
1999-2001	0.298	0.304	0.325	0.355	0.286
	(0.200)	(0.214)	(0.218)	(0.234)	(0.236)
2002-2005	0.168	0.189	0.211	0.266	0.199
	(0.136)	(0.127)	(0.133)	(0.140)	(0.144)
2006-2008	-0.009	-0.042	-0.032	-0.085	-0.072
	(0.119)	(0.127)	(0.122)	(0.120)	(0.129)
2009-2015	0.098	0.107	0.109	0.156	0.182
	(0.106)	(0.111)	(0.125)	(0.151)	(0.139)
Panel B: Aggregated Estimates					
	0.080	0.127	0.167	0.168	0.124
	(0.062)	(0.068)	(0.085)	(0.083)	(0.075)

NOTES: This table reports estimates of average direct treatment effects (ADTEs) of the approval of school district expenditure referenda on housing prices in Wisconsin between 1995 and 2015. Panel A presents cohort-specific estimates of $\text{ADTE}_{g,\tau}(0)$, whose estimand is stated in equation (30). Cohorts (indexed by g) correspond to groups of adjacent years shown in the leftmost column, and relative time (indexed by τ) ranges from 1 to 5 years. Panel B reports aggregated ADTEs, calculated as weighted averages of cohort-specific estimates according to equation (47), with weights proportional to each cohort's share of the sample. Standard errors are computed using the nearest-neighbor method described in Section 6.2, with tuning parameter $j^* = 3$, and aggregated using the Delta method.

Table 1 presents the results. Panel A reports cohort-specific estimates, while Panel B

---

[14]Pooling observations across adjacent cohorts implicitly imposes the restriction $\text{ADTE}_{t,\tau} = \text{ADTE}_{t',\tau}$ for all periods t and t' within a group.

[15]This choice corresponds to the default setting in the `rdrobust` package (Calonico, Cattaneo and Titiunik 2014c, Calonico, Cattaneo and Titiunik 2015, Calonico et al. 2017).

shows aggregated parameters. The latter indicate that, on average, the approval of school expenditure measures leads to cumulative increases in housing prices of approximately 12 percent after five years. These estimates are consistent with recent findings in the literature (Biasi, Lafortune and Schönholzer 2025), although they also highlight substantial heterogeneity across cohorts. In particular, the aggregate effects appear to be primarily driven by the 1999–2001 and 2002-2005 year groups.

Figure 4: Estimated ADTEs of Referendum Approval on Housing Prices

(a) Aggregated Cohort-Specific Estimates   (b) Pooled Estimates
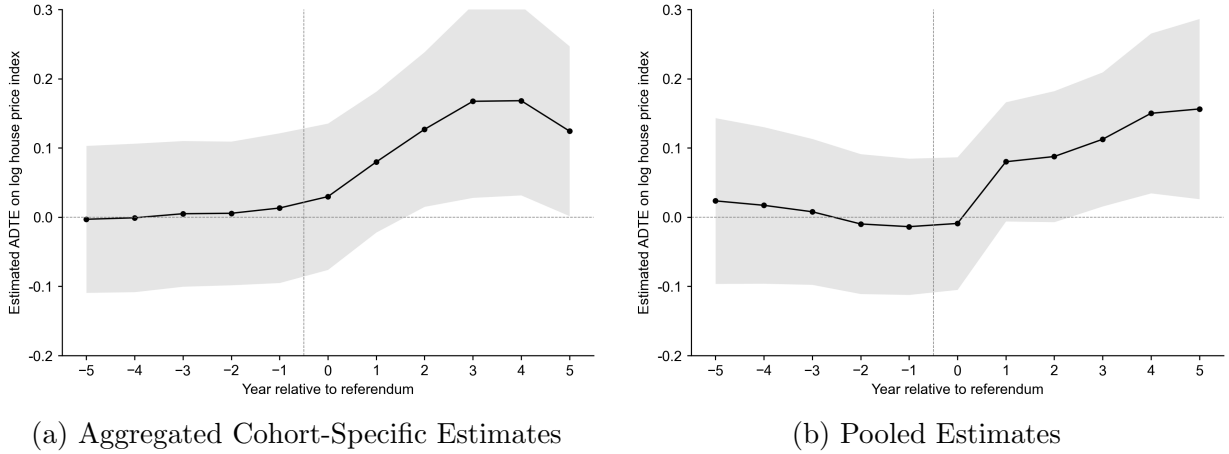
NOTES: This figure displays estimates of average direct treatment effects (ADTEs) of the approval of school district expenditure referenda on housing prices in Wisconsin between 1995 and 2015. Estimates for relative years 1 through 5 correspond to the estimand stated in equation (30). Estimates for relative years –5 through 0 are based on standard (i.e., static) local linear regression discontinuity estimators; those for years -5 to -1 serve as placebo tests. Panel (a) reports aggregated estimates of cohort-specific effects, where cohorts are defined as groups of adjacent years, as in Table 1. Aggregated ADTEs are computed as weighted averages of cohort-specific ADTEs, with weights proportional to each cohort's share of the sample. Panel (b) displays estimates from a pooled specification in which a single parameter is estimated for each relative year, using data from all cohorts combined. Shaded gray areas denote 90 percent confidence intervals. Standard errors are computed using the nearest-neighbor method described in Section 6.2, with tuning parameter $j^* = 3$. In Panel (a), standard errors are aggregated using the Delta method.

To assess the validity of the identification strategy, I conduct a placebo analysis by estimating treatment effects on housing prices in periods prior to the referenda. These estimates rely on standard (i.e., static) local linear regression discontinuity estimators. Figure 4 presents event-study-style plots that show average direct treatment effects for each relative year, based on two specifications: one that aggregates cohort-specific estimates, and another that pools observations across all cohorts. The two approaches yield comparable results in both magnitude and precision. Reassuringly, the estimated effects in pre-treatment periods are close to zero and not statistically significant, suggesting no systematic differences in

housing prices between jurisdictions that later approved versus rejected referenda.

Table 2: Checks on the Plausibility of the Common Trends Assumption

	$\Delta_{g-u,g-v} \equiv Y_{g-v} - Y_{g-u}$		
	$u = 3, v = 2$	$u = 2, v = 1$	$u = 3, v = 1$
Panel A: Aggregated Cohort-Specific Estimates			
$R_g \in \mathcal{R}^+$	-0.004	-0.020	-0.036
	(0.018)	(0.011)	(0.026)
$R_g \in \mathcal{R}^-$	-0.013	-0.021	0.055
	(0.029)	(0.015)	(0.365)
p-value	0.873	0.068	0.381
Panel B: Pooled Estimates			
$R_g \in \mathcal{R}^+$	0.000	0.010	0.010
	(0.026)	(0.022)	(0.047)
$R_g \in \mathcal{R}^-$	0.008	0.000	0.008
	(0.010)	(0.013)	(0.024)
p-value	0.747	0.909	0.923

NOTES: This table reports estimates used to assess the plausibility of the local common trends assumption, as described in Section 5.1. For each pair of relative years $(u, v) \in \{(3, 2), (2, 1), (3, 1)\}$, the table displays estimated differences in average pre-referendum outcome trends, defined as $\Delta_{g-u,g-v} \equiv Y_{g-v} - Y_{g-u}$, between the subpopulation implied by $D_{g+1} = D_{g+2} = D_{g+3} = D_{g+4} = D_{g+5} = 0$ and its complement. These differences are evaluated separately on either side of the cutoff, with $\mathcal{R}^+$ and $\mathcal{R}^-$ denoting the right and left discontinuity half-windows, respectively. Panel A presents estimates aggregated across cohorts, while Panel B reports estimates from a pooled specification in which all cohorts are combined. Standard errors, shown in parentheses, are computed using the nearest-neighbor method described in Section 6.2, with tuning parameter $j^* = 3$. The final row in each panel reports the p-value from an F-test of the null hypothesis that the upper and lower limits of the estimated difference are jointly equal to zero.

As a final validation exercise, I assess the plausibility of the local common trends assumption, as discussed in Section 5.1. By construction, outcomes are untreated in the three periods preceding each referendum. If the common trends assumption holds, an empirical researcher may be willing to conjecture that, at the cutoff of the running variable, trends in untreated outcomes prior to the referendum are similar across jurisdictions that are subsequently treated and those that remain untreated after the focal period. To test this suggestive implication, I examine all possible pairs of relative years $(u, v) \in \{(3, 2), (2, 1), (3, 1)\}$ and nonparametrically estimate the difference in conditional means of $Y_{g-v} - Y_{g-u}$ between two subpopulations: jurisdictions for which $D_{g+1} = D_{g+2} = D_{g+3} = D_{g+4} = D_{g+5} = 0$, and its complement. These differences correspond to contrasts in pre-referendum outcome trends,

evaluated at the cutoff. Table 2 reports the estimated differences for each pair of years, both using aggregated cohort-specific estimates and a specification that pools observations across all cohorts. In all cases, the estimated differences are not statistically distinguishable from zero at the 0.05 significance level. For each pair (u, v), I also conduct an F-test of the null hypothesis that the upper and lower limits of the conditional mean differences are jointly equal to zero. The null is not rejected in any case, suggesting that the local common trends assumption might be plausible in this setting.

9 Conclusion

This paper proposed a novel argument to point identify economically interpretable intertemporal treatment effects in dynamic regression discontinuity designs. I developed a dynamic potential outcomes model and reformulated two assumptions from the difference-in-differences literature—no anticipation and common trends—to point identify impulse response-like treatment effects at the cutoff. The estimand of each target parameter can be expressed as the sum of two static RD outcome contrasts, thereby allowing for nonparametric estimation and inference with standard local polynomial methods. I also proposed a nonparametric approach to aggregate treatment effects across calendar time and treatment paths, relying on a limited path independence restriction to reduce the dimensionality of the parameter space. I applied the proposed method to study the dynamic effects of school district expenditure referenda on housing prices in Wisconsin. The results indicate that, on average, referendum approval leads to a cumulative increase in housing prices of approximately 12 percent within five years.

References

Abadie, Alberto. 2005. "Semiparametric Difference-in-Differences Estimators." *Review of Economic Studies*, 72: 1–19.

Abadie, Alberto, and Guido Imbens. 2006. "Large Sample Properties of Matching Estimators for Average Treatment Effects." *Econometrica*, 74(1): 235–267.

Abbring, Jaap, and Gerard Van Der Berg. 2003. "The Nonparametric Identification of Treatment Effects in Duration Models." *Econometrica*, 71(5): 1491–1517.

Abott, Carolyn, Vladimir Kogan, Stéphane Lavertu, and Zachary Peskowitz. 2020. "School District Operational Spending and Student Outcomes: Evidence from Tax Elections in Seven States." *Journal of Public Economics*, 183: 104142.

Angrist, Joshua, and Jörn-Steffen Pischke. 2008. *Mostly Harmless Econometrics: An Empiricist's Companion.* Princeton University Press.

Angrist, Joshua, and Jörn-Steffen Pischke. 2014. *Mastering 'Metrics: The Path from Cause to Effect.* Princeton University Press.

Baron, Jason. 2022. "School Spending and Student Outcomes: Evidence from Revenue Limit Elections in Wisconsin." *American Economic Journal: Economic Policy*, 14(1): 1–39.

Baron, Jason, Joshua Hyman, and Brittany Vasquez. 2024. "Public School Funding, School Quality, and Adult Crime." *Review of Economics and Statistics*, 1–46.

Barrow, Lisa, and Cecilia Elena Rouse. 2004. "Using Market Valuation to Assess Public School Spending." *Journal of Public Economics*, 88(9): 1747–1769.

Bertanha, Marinho, and Guido Imbens. 2020. "External Validity in Fuzzy Regression Discontinuity Designs." *Journal of Business and Economic Statistics*, 38(3): 593–612.

Biasi, Barbara, Julien Lafortune, and David Schönholzer. 2023. "Effectiveness and Efficiency of School Capital Investments Across the U.S." *Working Paper.*

Biasi, Barbara, Julien Lafortune, and David Schönholzer. 2025. "What Works and for Whom? Effectiveness and Efficiency of School Capital Investments Across the U.S." *Quarterly Journal of Economics*, qjaf013.

Bickerdike, Charles Frederick. 1902. "Taxation of Site Values." *The Economic Journal*, 12(48): 472–484.

Bojinov, Iavor, Ashesh Rambachan, and Neil Shephard. 2021. "Panel Experiments and Dynamic Causal Effects: A Finite Population Perspective." *Quantitative Economics*, 12: 1171–1196.

Brueckner, Jan. 1982. "A Test for Allocative Efficiency in the Local Public Sector." *Journal of Public Economics*, 19(3): 311–331.

Callaway, Brantly, and Pedro Sant'Anna. 2021. "Difference-in-Differences with Multiple Time Periods." *Journal of Econometrics*, 225(2): 200–230.

Callaway, Brantly, Andrew Goodman-Bacon, and Pedro Sant'Anna. 2024. "Event Studies with a Continuous Treatment." *AEA Papers and Proceedings*, 114: 601–605.

Calonico, Sebastian, Matias Cattaneo, and Rocío Titiunik. 2014*a*. "Robust Nonparametric Confidence Intervals for Regression Discontinuity Designs." *Econometrica*, 82(6): 2295–2326.

Calonico, Sebastian, Matias Cattaneo, and Rocío Titiunik. 2014*b*. "Supplement to 'Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs'." *Econometrica Supplementary Material*, 82(6): 2295–2326.

Calonico, Sebastian, Matias Cattaneo, and Rocío Titiunik. 2014*c*. "Robust Data-Driven Inference in the Regression-Discontinuity Design." *Stata Journal*, 14(4): 909–946.

Calonico, Sebastian, Matias Cattaneo, and Rocío Titiunik. 2015. "rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs." *R Journal*, 7(1): 38–51.

Calonico, Sebastian, Matias Cattaneo, Max Farrell, and Rocío Titiunik. 2017. "rdrobust: Software for Regression-Discontinuity Designs." *Stata Journal*, 17(2): 372–404.

Cattaneo, Matias, Luke Keele, Rocío Titiunik, and Gonzalo Vazquez-Bare. 2021. "Extrapolating Treatment Effects in Multi-Cutoff Regression Discontinuity Designs." *Journal of the American Statistical Association*, 116(536): 1941–1952.

Cattaneo, Matias, Michael Jansson, and Xinwei Ma. 2020. "Simple Local Polynomial Density Estimators." *Journal of the American Statistical Association*, 115(531): 1449–1455.

Cellini, Stephanie, Fernando Ferreira, and Jesse Rothstein. 2010. "The Value of School Facility Investments: Evidence from a Dynamic Regression Discontinuity Design." *Quarterly Journal of Economics*, 125(1): 215–261.

Contat, Justin, and William Larson. 2024. "A Flexible Method of Housing Price Index Construction using Repeat-Sales Aggregates." *Real Estate Economics*, 52(6): 1551–1583.

Cushing, Brian. 1984. "Capitalization of Interjurisdictional Fiscal Differentials: An Alternative Approach." *Journal of Urban Economics*, 15(3): 317–326.

Darolia, Rajeev. 2013. "Integrity versus Access? The Effect of Federal Financial Aid Availability on Postsecondary Enrollment." *Journal of Public Economics*, 106: 101–114.

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2020. "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review*, 110(9): 2964–2996.

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2024. "Difference-in-Differences Estimators of Intertemporal Treatment Effects." *Review of Economics and Statistics*, 1–45.

de Chaisemartin, Clément, and Xavier D'Haultfœuille. 2023. "Two-Way Fixed Effects and Differences-in-Differences Estimators with Several Treatments." *Journal of Econometrics*, 236(2): 105480.

de Chaisemartin, Clément, Xavier D'Haultfœuille, and Gonzalo Vazquez-Bare. 2024. "Difference-in-Difference Estimators with Continuous Treatments and No Stayers." *AEA Papers and Proceedings*, 114: 610–613.

Dong, Yingying, and Arthur Lewbel. 2015. "Identifying the Effect of Changing the Policy Threshold in Regression Discontinuity Models." *Review of Economics and Statistics*, 97(5): 1081–1092.

Enami, Ali, Lockwood Reynolds, and Shawn Rohlin. 2023. "The Effect of Property Taxes on Businesses: Evidence from a Dynamic Regression Discontinuity Approach." *Regional Science and Urban Economics*, 100: 103895.

Figlio, David, and Maurice Lucas. 2004. "What's in a Grade? School Report Cards and the Housing Market." *American Economic Review*, 94(3): 591–604.

Gelman, Andrew, Jennifer Hill, and Aki Vehtari. 2021. *Regression and Other Stories. Analytical Methods for Social Research*, Cambridge University Press.

Goldsmith-Pinkham, Paul, Peter Hull, and Michal Kolesár. 2024. "Contamination Bias in Linear Regressions." *American Economic Review*, 114(12): 4015–4051.

Goodman-Bacon, Andrew. 2021. "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics*, 225(2): 254–277.

Guryan, Jonathan, Erik Hurst, and Melissa Kearney. 2008. "Parental Education and Parental Time with Children." *Journal of Economic Perspectives*, 22(3): 23–46.

Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw. 2001. "Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design." *Econometrica*, 69(1): 201–209.

Heckman, James, John Eric Humphries, and Gregory Veramendi. 2016. "Dynamic Treatment Effects." *Journal of Econometrics*, 191: 276–292.

Hong, Kai, and Ron Zimmer. 2016. "Does Investing in School Capital Infrastructure Improve Student Achievement?" *Economics of Education Review*, 53: 143–158.

Hsu, Yu-Chin, and Shu Shen. 2024. "Dynamic Regression Discontinuity under Treatment Effect Heterogeneity." *Quantitative Economics*, 15(4): 1035–1064.

Imbens, Guido, and Donald Rubin. 2015. *Causal Inference for Statistics, Social, and Biomedical Sciences: An Introduction.* Cambridge University Press.

Imbens, Guido, and Karthik Kalyanaraman. 2012. "Optimal Bandwidth Choice for the Regression Discontinuity Estimator." *Review of Economic Studies*, 79(3): 933–959.

Lee, David. 2008. "Randomized Experiments from Non-Random Selection in U.S. House Elections." *Journal of Econometrics*, 142(2): 675–697.

Lincoln Institute of Land Policy, and George Washington Institute of Public Policy. 2025. "Significant Features of the Property Tax: Tax Limits and Truth in Taxation." *https: // www. lincolninst. edu/ data/ significant-features-property-tax/ access-property-tax-database/ tax-limits-truth-taxation/*, Accessed on March 12, 2025.

Marshall, Alfred. 1948. *Principles of Economics.* . Eighth ed., MacMillan.

Martorell, Paco, Kevin Stange, and Isaac McFarlin. 2016. "Investing in Schools: Capital Spending, Facility Conditions, and Student Achievement." *Journal of Public Economics*, 140: 13–29.

Mogstad, Magne, Alexander Torgovitsky, and Christopher Walters. 2021. "The Causal Interpretation of Two-Stage Least Squares with Multiple Instrumental Variables." *American Economic Review*, 111(11): 3663–3698.

Mogstad, Magne, and Alexander Torgovitsky. 2024. "Chapter 1 – Instrumental Variables with Unobserved Heterogeneity in Treatment Effects." In . Vol. 5 of *Handbook of Labor Economics*, , ed. Christian Dustmann and Thomas Lemieux, 1–114. Elsevier.

Oates, Wallace. 1969. "The Effects of Property Taxes and Local Public Spending on Property Values: An Empirical Study of Tax Capitalization and the Tiebout Hypothesis." *Journal of Political Economy*, 77(6): 957–971.

Pearl, Judea. 1995. "Causal Diagrams for Empirical Research." *Biometrika*, 82(4): 669–688.

Poterba, James. 1997. "Demographic Structure and the Political Economy of Public Education." *Journal of Policy Analysis and Management*, 16(1): 48–66.

Robins, James. 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: 1393–1512.

Rosenbaum, Paul. 1984. "The Consequences of Adjustment for a Concomitant Variable that Has Been Affected by the Treatment." *Journal of the Royal Statistical Society. Series A (General)*, 147(5): 656–666.

Rosenbaum, Paul. 2002. *Observational Studies. Springer Series in Statistics*, Springer.

Ross, Stephen, and John Yinger. 1999. "Sorting and Voting: A Review of the Literature on Urban Public Finance." In *Applied Urban Economics*. Vol. 3 of *Handbook of Regional and Urban Economics*, 2001–2060. Elsevier.

Rubin, Donald. 2009. "Should Observational Studies Be Designed to Allow Lack of Balance in Covariate Distributions Across Treatment Groups?" *Statistics in Medicine*, 28: 1420–1423.

Sun, Liyang, and Sarah Abraham. 2021. "Estimating Dynamic Treatment Effects in Event Studies with Heterogeneous Treatment Effects." *Journal of Econometrics*, 225(2): 175–199.

U.S. Census Bureau. 2025*a*. "Glossary." *https://www.census.gov/glossary/*, Accessed on March 12, 2025.

U.S. Census Bureau. 2025*b*. "TIGER/Line Shapefiles." *https://www.census.gov/geographies/mapping-files/time-series/geo/tiger-line-file.html*, Accessed on March 12, 2025.

Wisconsin Department of Public Instruction. 2025. "Referenda Reports." *https://sfs.dpi.wi.gov/Referenda/CustomReporting.aspx*, Accessed on March 12, 2025.

Wooldridge, Jeffrey. 2005. "Violating Ignorability of Treatment by Controlling for Too Many Factors." *Econometric Theory*, 21(5): 1026–1028.

Wooldridge, Jeffrey. 2010. *Econometric Analysis of Cross Section and Panel Data, Second Edition.* MIT Press.

Appendix

A   Identification

A.1   Proof of Proposition 1

The sharp RD estimand implied by the first-period discontinuity is defined as

$$\mathrm{RD}_t\left(c\right) \equiv \lim_{r \downarrow c} \mathbb{E}\left[Y_t | R_1 = r\right] - \lim_{r \uparrow c} \mathbb{E}\left[Y_t | R_1 = r\right]$$

Consider the first term:

$$\lim_{r \downarrow c} \mathbb{E}\left[Y_t | R_1 = r\right] = \lim_{r \downarrow c} \mathbb{E}\left[Y_t | R_1 = r, D_1 = 1\right] \tag{A.1}$$
$$= \lim_{r \downarrow c} \mathbb{E}\left[\mathbb{I}\left[D_1 = 1, D_2\left(1\right) = 0\right] Y_t\left(1, 0\right) | R_1 = r, D_1 = 1\right]$$
$$+ \lim_{r \downarrow c} \mathbb{E}\left[\mathbb{I}\left[D_1 = 1, D_2\left(1\right) = 1\right] Y_t\left(1, 1\right) | R_1 = r, D_1 = 1\right]$$
$$+ \lim_{r \downarrow c} \mathbb{E}\left[\mathbb{I}\left[D_1 = 0, D_2\left(1\right) = 0\right] Y_t\left(0, 0\right) | R_1 = r, D_1 = 1\right]$$
$$+ \lim_{r \downarrow c} \mathbb{E}\left[\mathbb{I}\left[D_1 = 0, D_2\left(1\right) = 1\right] Y_t\left(0, 1\right) | R_1 = r, D_1 = 1\right] \tag{A.2}$$

The first equality exploits the assumption that $D_1 \equiv \mathbb{I}\left[R_1 \geq c\right]$, which implies that the conditioning set can be expanded to include the event $D_1 = 1$. The second equality replaces the observed outcome Y_t with its stochastic linear combination of potential outcomes, according to equation (3). Furthermore, because each of the four expectations is conditional on the event $D_1 = 1$, any indicator containing the event $D_1 = 0$ will be false. For the same reason, the event $D_1 = 1$ can be subsumed from any indicator that contains it. Thus,

$$\lim_{r \downarrow c} \mathbb{E}\left[Y_t | R_1 = r\right] = \lim_{r \downarrow c} \mathbb{E}\left[\mathbb{I}\left[D_2\left(1\right) = 0\right] Y_t\left(1, 0\right) | R_1 = r, D_1 = 1\right]$$
$$+ \lim_{r \downarrow c} \mathbb{E}\left[\mathbb{I}\left[D_2\left(1\right) = 1\right] Y_t\left(1, 1\right) | R_1 = r, D_1 = 1\right] \tag{A.3}$$
$$= \lim_{r \downarrow c} \mathbb{E}\left[\mathbb{I}\left[D_2\left(1\right) = 0\right] Y_t\left(1, 0\right) | R_1 = r\right]$$
$$+ \lim_{r \downarrow c} \mathbb{E}\left[\mathbb{I}\left[D_2\left(1\right) = 1\right] Y_t\left(1, 1\right) | R_1 = r\right] \tag{A.4}$$

$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_t\left(1,0\right)|R_1 = r, D_2\left(1\right) = 0\right] \times \mathbb{P}\left(D_2\left(1\right) = 0|R_1 = r\right)$$

$$+ \lim_{r \downarrow c} \mathbb{E}\left[Y_t\left(1,1\right)|R_1 = r, D_2\left(1\right) = 1\right] \times \mathbb{P}\left(D_2\left(1\right) = 1|R_1 = r\right) \quad \text{(A.5)}$$

$$= \mathbb{E}\left[Y_t\left(1,0\right)|R_1 = c, D_2\left(1\right) = 0\right] \times \mathbb{P}\left(D_2\left(1\right) = 0|R_1 = c\right)$$

$$+ \mathbb{E}\left[Y_t\left(1,1\right)|R_1 = c, D_2\left(1\right) = 1\right] \times \mathbb{P}\left(D_2\left(1\right) = 1|R_1 = c\right) \quad \text{(A.6)}$$

As above, the second equality exploits the assumption that $D_1 \equiv \mathbb{I}\left[R_1 \geq c\right]$. The third equality follows from an application of the Law of Iterated Expectations. The fourth equality leverages Assumption 1 and the algebraic properties of limits. By a symmetric argument, the right-hand side of $\mathrm{RD}_t\left(c\right)$ identifies

$$\lim_{r \uparrow c} \mathbb{E}\left[Y_t|R_1 = r\right] = \mathbb{E}\left[Y_t\left(0,0\right)|R_1 = c, D_2\left(0\right) = 0\right] \times \mathbb{P}\left(D_2\left(0\right) = 0|R_1 = c\right)$$

$$+ \mathbb{E}\left[Y_t\left(0,1\right)|R_1 = c, D_2\left(0\right) = 1\right] \times \mathbb{P}\left(D_2\left(0\right) = 1|R_1 = c\right) \quad \text{(A.7)}$$

which completes the proof.

A.2   Proof of Proposition 2

The instantaneous Average Direct Treatment Effect (ADTE) implied by the first-period discontinuity is defined as

$$\mathrm{ADTE}_{1,0}\left(c\right) \equiv \mathbb{E}\left[Y_1\left(1,0\right)|R_1 = c\right] - \mathbb{E}\left[Y_1\left(0,0\right)|R_1 = c\right]$$

By an application of the Law of Iterated Expectations, the first term can be expressed as

$$\mathbb{E}\left[Y_1\left(1,0\right)|R_1 = c\right] = \mathbb{E}\left[Y_1\left(1,0\right)|R_1 = c, D_2\left(1\right) = 0\right] \times \mathbb{P}\left(D_2\left(1\right) = 0|R_1 = c\right)$$

$$+ \mathbb{E}\left[Y_1\left(1,0\right)|R_1 = c, D_2\left(1\right) = 1\right] \times \mathbb{P}\left(D_2\left(1\right) = 1|R_1 = c\right) \quad \text{(A.8)}$$

Both conditional probabilities are identified under Assumption 1. In fact, for $d_2 \in \{0,1\}$,

$$\mathbb{P}\left(D_2\left(1\right) = d_2|R_1 = c\right) = \lim_{r \downarrow c} \mathbb{P}\left(D_2\left(1\right) = d_2|R_1 = r\right) \quad \text{(A.9)}$$

$$= \lim_{r \downarrow c} \mathbb{P}\left(D_2\left(1\right) = d_2|R_1 = r, D_1 = 1\right) \quad \text{(A.10)}$$

$$= \lim_{r \downarrow c} \mathbb{P}\left(D_2 = d_2|R_1 = r, D_1 = 1\right) \quad \text{(A.11)}$$

$$= \lim_{r \downarrow c} \mathbb{P}\left(D_2 = d_2|R_1 = r\right) \quad \text{(A.12)}$$

The first equality leverages Assumption 1, which states that $\mathbb{E}\left[D_2\left(1\right)|R_1 = r\right]$ is a continuous function of r at $r = c$. The second and fourth equalities follow from the threshold-crossing

47

definition of the treatment, $D_1 \equiv \mathbb{I}[R_1 \geq c]$. The third equality stems from the fact that, conditional on $D_1 = 1$, the second-period potential treatment $D_2(1)$ is observed. By a similar argument, the first conditional expectation in equation (A.52) is identified as

$$\mathbb{E}[Y_1(1,0)|R_1 = c, D_2(1) = 0] = \lim_{r \downarrow c} \mathbb{E}[Y_1(1,0)|R_1 = r, D_2(1) = 0] \tag{A.13}$$

$$= \lim_{r \downarrow c} \mathbb{E}[Y_1(1,0)|R_1 = r, D_1 = 1, D_2(1) = 0] \tag{A.14}$$

$$= \lim_{r \downarrow c} \mathbb{E}[Y_1(1,0)|R_1 = r, D_1 = 1, D_2 = 0] \tag{A.15}$$

$$= \lim_{r \downarrow c} \mathbb{E}[Y_1|R_1 = r, D_1 = 1, D_2 = 0] \tag{A.16}$$

$$= \lim_{r \downarrow c} \mathbb{E}[Y_1|R_1 = r, D_2 = 0] \tag{A.17}$$

The first equality leverages Assumption 1, which states that $\mathbb{E}[Y_1(1,0)|R_1 = r, D_2(1) = 0]$ is a continuous function of r at $r = c$. The second and fifth equalities follow from the threshold-crossing definition of the treatment, $D_1 \equiv \mathbb{I}[R_1 \geq c]$. The third equality stems from the fact that, conditional on $D_1 = 1$, the second-period potential treatment $D_2(1)$ is observed. Analogously, the potential outcome $Y_1(1,0)$ is observed conditional on $D_1 = 1$ and $D_2 = 0$, thus justifying the fourth equality. Finally, under Assumptions 1 Assumption 2, the counterfactual conditional mean in equation (A.52) is identified as

$$\mathbb{E}[Y_1(1,0)|R_1 = c, D_2(1) = 1] = \mathbb{E}[Y_1(1,1)|R_1 = c, D_2(1) = 1] \tag{A.18}$$

$$= \lim_{r \downarrow c} \mathbb{E}[Y_1(1,1)|R_1 = r, D_2(1) = 1] \tag{A.19}$$

$$= \lim_{r \downarrow c} \mathbb{E}[Y_1(1,1)|R_1 = r, D_1 = 1, D_2(1) = 1] \tag{A.20}$$

$$= \lim_{r \downarrow c} \mathbb{E}[Y_1(1,1)|R_1 = r, D_1 = 1, D_2 = 1] \tag{A.21}$$

$$= \lim_{r \downarrow c} \mathbb{E}[Y_1|R_1 = r, D_1 = 1, D_2 = 1] \tag{A.22}$$

$$= \lim_{r \downarrow c} \mathbb{E}[Y_1|R_1 = r, D_2 = 1] \tag{A.23}$$

The first equality applies the no anticipation restriction embedded in Assumption 2. The second equality leverages Assumption 1, which states that $\mathbb{E}[Y_1(1,1)|R_1 = r, D_2(1) = 1]$ is a continuous function of r at $r = c$. The third and sixth equalities follow from the threshold-crossing definition of the treatment, $D_1 \equiv \mathbb{I}[R_1 \geq c]$. The fourth equality stems from the fact that, conditional on $D_1 = 1$, the second-period potential treatment $D_2(1)$ is observed. Analogously, the potential outcome $Y_1(1,1)$ is observed conditional on $D_1 = 1$ and $D_2 = 1$, thus justifying the fifth equality. Combining these identification results, the first term in

$\text{ADTE}_{1,0}(c)$ can be expressed as

$$\mathbb{E}\left[Y_1\left(1,0\right)|R_1 = c\right] = \lim_{r \downarrow c} \mathbb{E}\left[Y_1|R_1 = r, D_2 = 0\right] \times \lim_{r \downarrow c} \mathbb{P}\left(D_2 = 0|R_1 = r\right)$$

$$+ \lim_{r \downarrow c} \mathbb{E}\left[Y_1|R_1 = r, D_2 = 1\right] \times \lim_{r \downarrow c} \mathbb{P}\left(D_2 = 1|R_1 = r\right) \quad \text{(A.24)}$$

$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_1|R_1 = r\right] \quad \text{(A.25)}$$

where the second equality follows from another application of the Law of Iterated Expectations. By a symmetric argument, $\lim_{r \uparrow c} \mathbb{E}\left[Y_1|R_1 = r\right]$ identifies $\mathbb{E}\left[Y_1\left(0,0\right)|R_1 = c\right]$. Thus,

$$\text{ADTE}_{1,0}\left(c\right) = \lim_{r \downarrow c} \mathbb{E}\left[Y_1|R_1 = r\right] - \lim_{r \uparrow c} \mathbb{E}\left[Y_1|R_1 = r\right] \quad \text{(A.26)}$$

A.3 Proof of Proposition 3

The cumulative Average Direct Treatment Effect (ADTE) implied by the first-period discontinuity is defined as

$$\text{ADTE}_{1,1}\left(c\right) \equiv \mathbb{E}\left[Y_2\left(1,0\right)|R_1 = c\right] - \mathbb{E}\left[Y_2\left(0,0\right)|R_1 = c\right]$$

By an application of the Law of Iterated Expectations, the first term can be expressed as

$$\mathbb{E}\left[Y_2\left(1,0\right)|R_1 = c\right] = \mathbb{E}\left[Y_2\left(1,0\right)|R_1 = c, D_2\left(1\right) = 0\right] \times \mathbb{P}\left(D_2\left(1\right) = 0|R_1 = c\right)$$

$$+ \mathbb{E}\left[Y_2\left(1,0\right)|R_1 = c, D_2\left(1\right) = 1\right] \times \mathbb{P}\left(D_2\left(1\right) = 1|R_1 = c\right) \quad \text{(A.27)}$$

As shown in the previous section, both conditional probabilities are identified under Assumption 1. Furthermore, the first conditional expectation is identified as

$$\mathbb{E}\left[Y_2\left(1,0\right)|R_1 = c, D_2\left(1\right) = 0\right] = \lim_{r \downarrow c} \mathbb{E}\left[Y_2\left(1,0\right)|R_1 = r, D_2\left(1\right) = 0\right] \quad \text{(A.28)}$$

$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_2\left(1,0\right)|R_1 = r, D_1 = 1, D_2\left(1\right) = 0\right] \quad \text{(A.29)}$$

$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_2\left(1,0\right)|R_1 = r, D_1 = 1, D_2 = 0\right] \quad \text{(A.30)}$$

$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_2|R_1 = r, D_1 = 1, D_2 = 0\right] \quad \text{(A.31)}$$

$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_2|R_1 = r, D_2 = 0\right] \quad \text{(A.32)}$$

The first equality leverages Assumption 1, which states that $\mathbb{E}\left[Y_2\left(1,0\right)|R_1 = r, D_2\left(1\right) = 0\right]$ is a continuous function of r at $r = c$. The second and fifth equalities follow from the threshold-crossing definition of the treatment, $D_1 \equiv \mathbb{I}\left[R_1 \geq c\right]$. The third equality stems

from the fact that, conditional on $D_1 = 1$, the second-period potential treatment $D_2(1)$ is observed. Analogously, the potential outcome $Y_2(1,0)$ is observed conditional on $D_1 = 1$ and $D_2 = 0$, thus justifying the fourth equality. Finally, under Assumptions 1, 2, and 3, the counterfactual conditional mean in equation (A.71) is identified as

$$\mathbb{E}\left[Y_2(1,0)\,|\,R_1 = c, D_2(1) = 1\right] \tag{A.33}$$

$$= \mathbb{E}\left[Y_1(1,0)\,|\,R_1 = c, D_2(1) = 1\right]$$
$$+ \mathbb{E}\left[Y_2(1,0) - Y_1(1,0)\,|\,R_1 = c, D_2(1) = 0\right] \tag{A.34}$$

$$= \mathbb{E}\left[Y_1(1,1)\,|\,R_1 = c, D_2(1) = 1\right]$$
$$+ \mathbb{E}\left[Y_2(1,0) - Y_1(1,0)\,|\,R_1 = c, D_2(1) = 0\right] \tag{A.35}$$

$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_1(1,1)\,|\,R_1 = r, D_2(1) = 1\right]$$
$$+ \lim_{r \downarrow c} \mathbb{E}\left[Y_2(1,0) - Y_1(1,0)\,|\,R_1 = r, D_2(1) = 0\right] \tag{A.36}$$

$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_1(1,1)\,|\,R_1 = r, D_1 = 1, D_2(1) = 1\right]$$
$$+ \lim_{r \downarrow c} \mathbb{E}\left[Y_2(1,0) - Y_1(1,0)\,|\,R_1 = r, D_1 = 1, D_2(1) = 0\right] \tag{A.37}$$

$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_1(1,1)\,|\,R_1 = r, D_1 = 1, D_2 = 1\right]$$
$$+ \lim_{r \downarrow c} \mathbb{E}\left[Y_2(1,0) - Y_1(1,0)\,|\,R_1 = r, D_1 = 1, D_2 = 0\right] \tag{A.38}$$

$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_1\,|\,R_1 = r, D_1 = 1, D_2 = 1\right]$$
$$+ \lim_{r \downarrow c} \mathbb{E}\left[Y_2 - Y_1\,|\,R_1 = r, D_1 = 1, D_2 = 0\right] \tag{A.39}$$

$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_1\,|\,R_1 = r, D_2 = 1\right]$$
$$+ \lim_{r \downarrow c} \mathbb{E}\left[Y_2 - Y_1\,|\,R_1 = r, D_2 = 0\right] \tag{A.40}$$

The first equality uses Assumption 3, which states that $Y_2(1,0) - Y_1(1,0)$ is mean independent of $D_2(1)$ at the first-period cutoff. The second equality applies the no anticipation restriction embedded in Assumption 2. The third equality leverages Assumption 1, which states that $\mathbb{E}\left[Y_1(1,1)\,|\,R_1 = r, D_2(1) = 1\right]$, $\mathbb{E}\left[Y_2(1,0)\,|\,R_1 = r, D_2(1) = 0\right]$, and $\mathbb{E}\left[Y_1(1,0)\,|\,R_1 = r, D_2(1) = 0\right]$ are continuous functions of r at $r = c$. The fourth and seventh equalities follow from the threshold-crossing definition of the treatment, $D_1 \equiv \mathbb{I}\left[R_1 \geq c\right]$. The fifth equality stems from the fact that, conditional on $D_1 = 1$, the second-period potential treatment $D_2(1)$ is observed. Analogously, the potential outcome $Y_1(1,1)$ is observed conditional on $D_1 = D_2 = 1$ and the potential outcomes $Y_2(1,0)$ and $Y_1(1,0)$ are observed

conditional on $D_1 = 1$ and $D_2 = 0$, thus justifying the sixth equality. Combining these results, the first term in $\text{ADTE}_{1,1}(c)$ can be expressed as

$$\mathbb{E}\left[Y_2\left(1,0\right)|R_1 = c\right] = \lim_{r \downarrow c} \mathbb{E}\left[Y_2|R_1 = r, D_2 = 0\right] \times \lim_{r \downarrow c} \mathbb{P}\left(D_2 = 0|R_1 = r\right)$$
$$+ \left(\lim_{r \downarrow c} \mathbb{E}\left[Y_1|R_1 = r, D_2 = 1\right] + \lim_{r \downarrow c} \mathbb{E}\left[Y_2 - Y_1|R_1 = r, D_2 = 0\right]\right)$$
$$\times \lim_{r \downarrow c} \mathbb{P}\left(D_2 = 1|R_1 = r\right) \tag{A.41}$$

$\mathbb{E}\left[Y_2\left(0,0\right)|R_1 = c\right]$ is point identified with a symmetric argument. Thus,

$$\text{ADTE}_{1,1}\left(c\right) = \lim_{r \downarrow c} \mathbb{E}\left[Y_2|R_1 = r, D_2 = 0\right] \times \lim_{r \downarrow c} \mathbb{P}\left(D_2 = 0|R_1 = r\right)$$
$$+ \left(\lim_{r \downarrow c} \mathbb{E}\left[Y_1|R_1 = r, D_2 = 1\right] + \lim_{r \downarrow c} \mathbb{E}\left[Y_2 - Y_1|R_1 = r, D_2 = 0\right]\right)$$
$$\times \lim_{r \downarrow c} \mathbb{P}\left(D_2 = 1|R_1 = r\right)$$
$$- \lim_{r \uparrow c} \mathbb{E}\left[Y_2|R_1 = r, D_2 = 0\right] \times \lim_{r \uparrow c} \mathbb{P}\left(D_2 = 0|R_1 = r\right)$$
$$- \left(\lim_{r \uparrow c} \mathbb{E}\left[Y_1|R_1 = r, D_2 = 1\right] + \lim_{r \uparrow c} \mathbb{E}\left[Y_2 - Y_1|R_1 = r, D_2 = 0\right]\right)$$
$$\times \lim_{r \uparrow c} \mathbb{P}\left(D_2 = 1|R_1 = r\right) \tag{A.42}$$
$$= \lim_{r \downarrow c} \mathbb{E}\left[(1 - D_2) Y_2|R_1 = r\right] - \lim_{r \uparrow c} \mathbb{E}\left[(1 - D_2) Y_2|R_1 = r\right]$$
$$+ \lim_{r \downarrow c} \mathbb{E}\left[D_2 Y_1|R_1 = r\right] - \lim_{r \uparrow c} \mathbb{E}\left[D_2 Y_1|R_1 = r\right]$$
$$- \lim_{r \downarrow c} \mathbb{E}\left[(1 - D_2)\left(Y_2 - Y_1\right)|R_1 = r\right] - \lim_{r \uparrow c} \mathbb{E}\left[(1 - D_2)\left(Y_2 - Y_1\right)|R_1 = r\right]$$
$$+ \lim_{r \downarrow c} \mathbb{E}\left[Y_2 - Y_1|R_1 = r, D_2 = 0\right] - \lim_{r \uparrow c} \mathbb{E}\left[Y_2 - Y_1|R_1 = r, D_2 = 0\right] \tag{A.43}$$
$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_1|R_1 = r\right] - \lim_{r \uparrow c} \mathbb{E}\left[Y_1|R_1 = r\right]$$
$$+ \lim_{r \downarrow c} \mathbb{E}\left[Y_2 - Y_1|R_1 = r, D_2 = 0\right] - \lim_{r \uparrow c} \mathbb{E}\left[Y_2 - Y_1|R_1 = r, D_2 = 0\right] \tag{A.44}$$

where the second equality follows from an application of the Law of Iterated Expectations.

A.4   Proof of Proposition 4

The sharp RD estimand implied by the first-period discontinuity is defined as

$$\text{RD}_t\left(c\right) \equiv \lim_{r \downarrow c} \mathbb{E}\left[Y_t|R_1 = r\right] - \lim_{r \uparrow c} \mathbb{E}\left[Y_t|R_1 = r\right]$$

Consider the first term:

$$\lim_{r \downarrow c} \mathbb{E}\left[Y_t|R_1 = r\right]$$

51

$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_t | R_1 = r, D_1 = 1\right] \tag{A.45}$$

$$= \lim_{r \downarrow c} \sum_{(d_2,\dots,d_{\bar{t}})} \mathbb{E}\left[\mathbb{I}\left[D_1 = 1\right] P\left(1, d_2, \dots, d_{\bar{t}}\right) Y_t\left(1, d_2, \dots, d_{\bar{t}}\right) | R_1 = r, D_1 = 1\right]$$

$$+ \lim_{r \downarrow c} \sum_{(d_2,\dots,d_{\bar{t}})} \mathbb{E}\left[\mathbb{I}\left[D_1 = 0\right] P\left(0, d_2, \dots, d_{\bar{t}}\right) Y_t\left(0, d_2, \dots, d_{\bar{t}}\right) | R_1 = r, D_1 = 1\right] \tag{A.46}$$

The first equality exploits the assumption that $D_1 \equiv \mathbb{I}\left[R_1 \geq c\right]$, which implies that the conditioning set can be expanded to include the event $D_1 = 1$. The second equality replaces the observed outcome Y_t with its stochastic linear combination of potential outcomes, according to equation (20). Furthermore, because each of the four expectations is conditional on the event $D_1 = 1$, any indicator containing the event $D_1 = 0$ will be false. For the same reason, the event $D_1 = 1$ can be subsumed from any indicator that contains it. Thus,

$$\lim_{r \downarrow c} \mathbb{E}\left[Y_t | R_1 = r\right] = \lim_{r \downarrow c} \sum_{(d_2,\dots,d_{\bar{t}})} \mathbb{E}\left[P\left(1, d_2, \dots, d_{\bar{t}}\right) Y_t\left(1, d_2, \dots, d_{\bar{t}}\right) | R_1 = r, D_1 = 1\right] \tag{A.47}$$

$$= \lim_{r \downarrow c} \sum_{(d_2,\dots,d_{\bar{t}})} \mathbb{E}\left[P\left(1, d_2, \dots, d_{\bar{t}}\right) Y_t\left(1, d_2, \dots, d_{\bar{t}}\right) | R_1 = r\right] \tag{A.48}$$

$$= \lim_{r \downarrow c} \sum_{(d_2,\dots,d_{\bar{t}})} \mathbb{E}\left[Y_t\left(1, d_2, \dots, d_{\bar{t}}\right) | R_1 = r, P\left(1, d_2, \dots, d_{\bar{t}}\right) = 1\right]$$

$$\times \mathbb{E}\left[P\left(1, d_2, \dots, d_{\bar{t}}\right) | R_1 = r\right] \tag{A.49}$$

$$= \sum_{(d_2,\dots,d_{\bar{t}})} \mathbb{E}\left[Y_t\left(1, d_2, \dots, d_{\bar{t}}\right) | R_1 = c, P\left(1, d_2, \dots, d_{\bar{t}}\right) = 1\right]$$

$$\times \mathbb{E}\left[P\left(1, d_2, \dots, d_{\bar{t}}\right) | R_1 = c\right] \tag{A.50}$$

As above, the second equality exploits the assumption that $D_1 \equiv \mathbb{I}\left[R_1 \geq c\right]$. The third equality follows from an application of the Law of Iterated Expectations. The fourth equality leverages Assumption 4 and the algebraic properties of limits. By a symmetric argument, the right-hand side of $\mathrm{RD}_t\left(c\right)$ identifies

$$\lim_{r \uparrow c} \mathbb{E}\left[Y_t | R_1 = r\right] = \sum_{(d_2,\dots,d_{\bar{t}})} \mathbb{E}\left[Y_t\left(0, d_2, \dots, d_{\bar{t}}\right) | R_1 = c, P\left(0, d_2, \dots, d_{\bar{t}}\right) = 1\right]$$

$$\times \mathbb{E}\left[P\left(0, d_2, \dots, d_{\bar{t}}\right) | R_1 = c\right] \tag{A.51}$$

which completes the proof.

A.5 Proof of Proposition 5

The instantaneous Average Direct Treatment Effect (ADTE) implied by the first-period discontinuity is defined as

$$\text{ADTE}_{1,0}(c) \equiv \mathbb{E}\left[Y_1\left(1, 0_{\bar{t}-1}\right) | R_1 = c\right] - \mathbb{E}\left[Y_1\left(0, 0_{\bar{t}-1}\right) | R_1 = c\right]$$

By an application of the Law of Iterated Expectations, the first term can be expressed as

$$\mathbb{E}\left[Y_1\left(1, 0_{\bar{t}-1}\right) | R_1 = c\right] = \sum_{(d_2,\ldots,d_{\bar{t}}) \in \{0,1\}^{\bar{t}-1}} \mathbb{E}\left[Y_t\left(1, 0_{\bar{t}-1}\right) | R_1 = c, P\left(1, d_2, \ldots, d_{\bar{t}}\right) = 1\right]$$
$$\times \mathbb{E}\left[P\left(1, d_2, \ldots, d_{\bar{t}}\right) | R_1 = c\right] \tag{A.52}$$

Each conditional expectation of P is identified under Assumption 4. In fact, for any sequence of treatment states $(d_2, \ldots, d_{\bar{t}})$,

$$\mathbb{E}\left[P\left(1, d_2, \ldots, d_{\bar{t}}\right) | R_1 = c\right] = \lim_{r \downarrow c} \mathbb{E}\left[P\left(1, d_2, \ldots, d_{\bar{t}}\right) | R_1 = r\right] \tag{A.53}$$
$$= \lim_{r \downarrow c} \mathbb{E}\left[P\left(1, d_2, \ldots, d_{\bar{t}}\right) | R_1 = r, D_1 = 1\right] \tag{A.54}$$
$$= \lim_{r \downarrow c} \mathbb{P}\left(D_2 = d_2, \ldots, D_{\bar{t}} = d_{\bar{t}} | R_1 = r, D_1 = 1\right) \tag{A.55}$$
$$= \lim_{r \downarrow c} \mathbb{P}\left(D_2 = d_2, \ldots, D_{\bar{t}} = d_{\bar{t}} | R_1 = r\right) \tag{A.56}$$

The first equality leverages Assumption 4, which states that $\mathbb{E}\left[P\left(1, d_2, \ldots, d_{\bar{t}}\right) | R_1 = r\right]$ is a continuous function of r at $r = c$. The second and fourth equalities follow from the threshold-crossing definition of the treatment, $D_1 \equiv \mathbb{I}\left[R_1 \geq c\right]$. The third equality stems from the fact that, conditional on $D_1 = 1$, the path of potential treatment states $P\left(1, d_2, \ldots, d_{\bar{t}}\right)$ is observed. By a similar argument, one conditional expectation in equation (A.52) is identified without further restrictions. Specifically,

$$\mathbb{E}\left[Y_1\left(1, 0_{\bar{t}-1}\right) | R_1 = c, P\left(1, 0_{\bar{t}-1}\right) = 1\right]$$
$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_1\left(1, 0_{\bar{t}-1}\right) | R_1 = r, P\left(1, 0_{\bar{t}-1}\right) = 1\right] \tag{A.57}$$
$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_1\left(1, 0_{\bar{t}-1}\right) | R_1 = r, D_1 = 1, P\left(1, 0_{\bar{t}-1}\right) = 1\right] \tag{A.58}$$
$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_1\left(1, 0_{\bar{t}-1}\right) | R_1 = r, D_1 = 1, D_2 = 0, \ldots, D_{\bar{t}} = 0\right] \tag{A.59}$$
$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_1 | R_1 = r, D_1 = 1, D_2 = 0, \ldots, D_{\bar{t}} = 0\right] \tag{A.60}$$
$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_1 | R_1 = r, D_2 = 0, \ldots, D_{\bar{t}} = 0\right] \tag{A.61}$$

The first equality leverages Assumption 4, which states that $\mathbb{E}\left[Y_1\left(1, 0_{\bar{t}-1}\right) | R_1 = r, P\left(1, 0_{\bar{t}-1}\right) = 1\right]$ is a continuous function of r at $r = c$. The second and fifth equalities follow from the threshold-crossing definition of the treatment, $D_1 \equiv \mathbb{I}\left[R_1 \geq c\right]$. The third equality stems from the fact that, conditional on $D_1 = 1$, the path of potential treatment states $P\left(1, d_2, \ldots, d_{\bar{t}}\right)$ is observed. Analogously, the potential outcome $Y_1\left(1, 0_{\bar{t}-1}\right)$ is observed conditional on $D_1 = 1$ and $D_2 = \cdots = D_{\bar{t}} = 0$, thus justifying the fourth equality. Finally, under Assumptions 4 and 5, any counterfactual conditional mean in equation (A.52) is identified as

$$\mathbb{E}\left[Y_1\left(1, 0_{\bar{t}-1}\right) | R_1 = c, P\left(1, d_2, \ldots, d_{\bar{t}}\right) = 1\right]$$

$$= \mathbb{E}\left[Y_1\left(1, d_2, \ldots, d_{\bar{t}}\right) | R_1 = c, P\left(1, d_2, \ldots, d_{\bar{t}}\right) = 1\right] \tag{A.62}$$

$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_1\left(1, d_2, \ldots, d_{\bar{t}}\right) | R_1 = r, P\left(1, d_2, \ldots, d_{\bar{t}}\right) = 1\right] \tag{A.63}$$

$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_1\left(1, d_2, \ldots, d_{\bar{t}}\right) | R_1 = r, D_1 = 1, P\left(1, d_2, \ldots, d_{\bar{t}}\right) = 1\right] \tag{A.64}$$

$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_1\left(1, d_2, \ldots, d_{\bar{t}}\right) | R_1 = r, D_1 = 1, D_2 = d_2, \ldots, D_{\bar{t}} = d_{\bar{t}}\right] \tag{A.65}$$

$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_1 | R_1 = r, D_1 = 1, D_2 = d_2, \ldots, D_{\bar{t}} = d_{\bar{t}}\right] \tag{A.66}$$

$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_1 | R_1 = r, D_2 = d_2, \ldots, D_{\bar{t}} = d_{\bar{t}}\right] \tag{A.67}$$

The first equality applies the no anticipation restriction embedded in Assumption 5. The second equality leverages Assumption 4, which states that $\mathbb{E}\left[Y_1\left(1,, 0_{\bar{t}-1}\right) | R_1 = r, P(1, d_2, \ldots, d_{\bar{t}}) = 1\right]$ is a continuous function of r at $r = c$. The third and sixth equalities follow from the threshold-crossing definition of the treatment, $D_1 \equiv \mathbb{I}\left[R_1 \geq c\right]$. The fourth equality stems from the fact that, conditional on $D_1 = 1$, the path of potential treatment states $P\left(1, d_2, \ldots, d_{\bar{t}}\right)$ is observed. Analogously, the potential outcome $Y_1\left(1, d_2, \ldots, d_{\bar{t}}\right)$ is observed conditional on $D_1 = 1$ and $D_2 = d_2, \ldots, D_{\bar{t}} = d_{\bar{t}}$, thus justifying the fifth equality. Combining these identification results, the first term in $\mathrm{ADTE}_{1,0}(c)$ can be expressed as

$$\mathbb{E}\left[Y_1\left(1, 0_{\bar{t}-1}\right) | R_1 = c\right] = \sum_{(d_2, \ldots, d_{\bar{t}}) \in \{0,1\}^{\bar{t}-1}} \lim_{r \downarrow c} \mathbb{E}\left[Y_1 | R_1 = r, D_2 = 0, \ldots, D_{\bar{t}} = d_{\bar{t}}\right]$$

$$\times \lim_{r \downarrow c} \mathbb{P}\left(D_2 = 0, \ldots, D_{\bar{t}} = d_{\bar{t}} | R_1 = r\right) \tag{A.68}$$

$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_1 | R_1 = r\right] \tag{A.69}$$

where the second equality follows from another application of the Law of Iterated Expectations and the algebraic properties of limits. By a symmetric argument, $\lim_{r \uparrow c} \mathbb{E}\left[Y_1 | R_1 = r\right]$

identifies $\mathbb{E}\left[Y_1\left(0, 0_{\bar{t}-1}\right) | R_1 = c\right]$. Thus,

$$\text{ADTE}_{1,0}\left(c\right) = \lim_{r\downarrow c}\mathbb{E}\left[Y_1 | R_1 = r\right] - \lim_{r\uparrow c}\mathbb{E}\left[Y_1 | R_1 = r\right] \tag{A.70}$$

A.6 Proof of Proposition 6

For $\tau \in \left\{1, \ldots, \bar{t} - 1\right\}$, the cumulative Average Direct Treatment Effect (ADTE) implied by the first-period discontinuity is defined as

$$\text{ADTE}_{1,\tau}\left(c\right) \equiv \mathbb{E}\left[Y_{1+\tau}\left(1, 0_{\bar{t}-1}\right) | R_1 = c\right] - \mathbb{E}\left[Y_{1+\tau}\left(0, 0_{\bar{t}-1}\right) | R_1 = c\right]$$

To keep notation concise, let Y_t denote $Y_{1+\tau}$ for $t \in \left\{2, \ldots, \bar{t}\right\}$. By an application of the Law of Iterated Expectations, the first term can be expressed as

$$\mathbb{E}\left[Y_t\left(1, 0_{\bar{t}-1}\right) | R_1 = c\right] = \sum_{\left(d_2, \ldots, d_{\bar{t}}\right) \in \{0,1\}^{\bar{t}-1}} \mathbb{E}\left[Y_t\left(1, 0_{\bar{t}-1}\right) | R_1 = c, P\left(1, d_2, \ldots, d_{\bar{t}}\right) = 1\right]$$
$$\times \mathbb{E}\left[P\left(1, d_2, \ldots, d_{\bar{t}}\right) | R_1 = c\right] \tag{A.71}$$

As shown in the previous section, each $\mathbb{E}\left[P\left(1, d_2, \ldots, d_{\bar{t}}\right) | R_1 = c\right]$ is identified under Assumption 4. Furthermore, the expected potential outcome conditional on a never-treated path indicator is identified as

$$\mathbb{E}\left[Y_t\left(1, 0_{\bar{t}-1}\right) | R_1 = c, P\left(1, 0_{\bar{t}-1}\right) = 1\right]$$
$$= \lim_{r\downarrow c}\mathbb{E}\left[Y_t\left(1, 0_{\bar{t}-1}\right) | R_1 = r, P\left(1, 0_{\bar{t}-1}\right) = 1\right] \tag{A.72}$$
$$= \lim_{r\downarrow c}\mathbb{E}\left[Y_t\left(1, 0_{\bar{t}-1}\right) | R_1 = r, D_1 = 1, P\left(1, 0_{\bar{t}-1}\right) = 1\right] \tag{A.73}$$
$$= \lim_{r\downarrow c}\mathbb{E}\left[Y_t\left(1, 0_{\bar{t}-1}\right) | R_1 = r, D_1 = 1, D_2 = 0, \ldots, D_{\bar{t}} = 0\right] \tag{A.74}$$
$$= \lim_{r\downarrow c}\mathbb{E}\left[Y_t | R_1 = r, D_1 = 1, D_2 = 0, \ldots, D_{\bar{t}} = 0\right] \tag{A.75}$$
$$= \lim_{r\downarrow c}\mathbb{E}\left[Y_t | R_1 = r, D_2 = 0, \ldots, D_{\bar{t}} = 0\right] \tag{A.76}$$

The first equality leverages Assumption 4, which states that $\mathbb{E}\left[Y_t\left(1, 0_{\bar{t}-1}\right) | R_1 = r, P\left(1, 0_{\bar{t}-1}\right) = 1\right]$ is a continuous function of r at $r = c$. The second and fifth equalities follow from the threshold-crossing definition of the treatment, $D_1 \equiv \mathbb{I}\left[R_1 \geq c\right]$. The third equality stems from the fact that, conditional on $D_1 = 1$, the path of potential treatment states $P\left(1, d_2, \ldots, d_{\bar{t}}\right)$ is observed. Analogously, the potential outcome $Y_t\left(1, d_2, \ldots, d_{\bar{t}}\right)$ is observed conditional on $D_1 = 1$ and $D_2 = d_2, \ldots, D_{\bar{t}} = d_{\bar{t}}$, thus justifying the fourth equality. Finally,

under Assumptions 4, 5, and 6, any counterfactual conditional mean in equation (A.71) is identified as

$$\mathbb{E}\left[Y_t\left(1, 0_{\bar{t}-1}\right)|R_1 = c, P\left(1, d_2, \ldots, d_{\bar{t}}\right) = 1\right]$$

$$= \mathbb{E}\left[Y_1\left(1, 0_{\bar{t}-1}\right)|R_1 = c, P\left(1, d_2, \ldots, d_{\bar{t}}\right) = 1\right]$$

$$+ \mathbb{E}\left[Y_t\left(1, 0_{\bar{t}-1}\right) - Y_1\left(1, 0_{\bar{t}-1}\right)|R_1 = c, P\left(1, 0_{\bar{t}-1}\right) = 1\right] \tag{A.77}$$

$$= \mathbb{E}\left[Y_1\left(1, d_2, \ldots, d_{\bar{t}}\right)|R_1 = c, P\left(1, d_2, \ldots, d_{\bar{t}}\right) = 1\right]$$

$$+ \mathbb{E}\left[Y_t\left(1, 0_{\bar{t}-1}\right) - Y_1\left(1, 0_{\bar{t}-1}\right)|R_1 = c, P\left(1, 0_{\bar{t}-1}\right) = 1\right] \tag{A.78}$$

$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_1\left(1, d_2, \ldots, d_{\bar{t}}\right)|R_1 = r, P\left(1, d_2, \ldots, d_{\bar{t}}\right) = 1\right]$$

$$+ \lim_{r \downarrow c} \mathbb{E}\left[Y_t\left(1, 0_{\bar{t}-1}\right) - Y_1\left(1, 0_{\bar{t}-1}\right)|R_1 = r, P\left(1, 0_{\bar{t}-1}\right) = 1\right] \tag{A.79}$$

$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_1\left(1, d_2, \ldots, d_{\bar{t}}\right)|R_1 = r, D_1 = 1, P\left(1, d_2, \ldots, d_{\bar{t}}\right) = 1\right]$$

$$+ \lim_{r \downarrow c} \mathbb{E}\left[Y_t\left(1, 0_{\bar{t}-1}\right) - Y_1\left(1, 0_{\bar{t}-1}\right)|R_1 = r, D_1 = 1, P\left(1, 0_{\bar{t}-1}\right) = 1\right] \tag{A.80}$$

$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_1\left(1, d_2, \ldots, d_{\bar{t}}\right)|R_1 = r, D_1 = 1, D_2 = d_2, \ldots, D_{\bar{t}} = d_{\bar{t}}\right]$$

$$+ \lim_{r \downarrow c} \mathbb{E}\left[Y_t\left(1, 0_{\bar{t}-1}\right) - Y_1\left(1, 0_{\bar{t}-1}\right)|R_1 = r, D_1 = 1, D_2 = 0, \ldots, D_{\bar{t}} = 0\right] \tag{A.81}$$

$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_1|R_1 = r, D_1 = 1, D_2 = d_2, D_{\bar{t}} = d_{\bar{t}}\right]$$

$$+ \lim_{r \downarrow c} \mathbb{E}\left[Y_t - Y_1|R_1 = r, D_1 = 1, D_2 = 0, \ldots, D_{\bar{t}} = 0\right] \tag{A.82}$$

$$= \lim_{r \downarrow c} \mathbb{E}\left[Y_1|R_1 = r, D_2 = d_2, \ldots, D_{\bar{t}} = d_{\bar{t}}\right]$$

$$+ \lim_{r \downarrow c} \mathbb{E}\left[Y_t - Y_1|R_1 = r, D_2 = 0, \ldots, D_{\bar{t}} = 0\right] \tag{A.83}$$

The first equality uses Assumption 6, which states that $Y_t\left(1, 0_{\bar{t}-1}\right) - Y_1\left(1, 0_{\bar{t}-1}\right)$ is mean independent of $P\left(1, d_2, \ldots, d_{\bar{t}}\right)$ at the first-period cutoff. The second equality applies the no anticipation restriction embedded in Assumption 5. The third equality leverages Assumption 4, which states that $\mathbb{E}\left[Y_1\left(1, d_2, \ldots, d_{\bar{t}}\right)|R_1 = c, P\left(1, d_2, \ldots, d_{\bar{t}}\right) = 1\right]$, $\mathbb{E}\left[Y_t\left(1, 0_{\bar{t}-1}\right)|R_1 = c, P\left(1, 0_{\bar{t}-1}\right) = 1\right]$, and $\mathbb{E}\left[Y_1\left(1, 0_{\bar{t}-1}\right)|R_1 = c, P\left(1, 0_{\bar{t}-1}\right) = 1\right]$ are continuous functions of r at $r = c$. The fourth and seventh equalities follow from the threshold-crossing definition of the treatment, $D_1 \equiv \mathbb{I}\left[R_1 \geq c\right]$. The fifth equality stems from the fact that, conditional on $D_1 = 1$, the paths of potential treatment states $P\left(1, d_2, \ldots, d_{\bar{t}}\right)$ and $P\left(1, 0_{\bar{t}-1}\right)$ are observed. Analogously, the potential outcome $Y_1\left(1, d_2, \ldots, d_{\bar{t}}\right)$ is observed conditional on $D_1 = d_1, \ldots, D_{\bar{t}} = d_{\bar{t}}$ and the potential outcomes $Y_t\left(1, 0_{\bar{t}-1}\right)$ and $Y_1\left(1, 0_{\bar{t}-1}\right)$ are observed conditional on $D_1 = 0, D_2 = 0, \ldots, D_{\bar{t}} = 0$, thus justifying the sixth equality. Combining

these results, the first term in $\text{ADTE}_{1,\tau}(c)$ can be expressed as

$$\mathbb{E}\left[Y_t\left(1, 0_{\bar{t}-1}\right)|R_1 = c\right] = \lim_{r\downarrow c} \mathbb{E}\left[Y_t|R_1 = r, D_2 = 0, \ldots, D_{\bar{t}} = 0\right]$$

$$\times \lim_{r\downarrow c} \mathbb{P}\left(D_2 = 0, \ldots, D_{\bar{t}} = 0|R_1 = r\right)$$

$$+ \sum_{(d_2,\ldots,d_{\bar{t}})\neq 0^{\bar{t}-1}} \left(\lim_{r\downarrow c} \mathbb{E}\left[Y_1|R_1 = r, D_2 = d_2, \ldots, D_{\bar{t}} = d_{\bar{t}}\right] \right.$$

$$\left. + \lim_{r\downarrow c} \mathbb{E}\left[Y_t - Y_1|R_1 = r, D_2 = 0, \ldots, D_{\bar{t}} = 0\right] \right)$$

$$\times \lim_{r\downarrow c} \mathbb{P}\left(D_2 = d_2, \ldots, D_{\bar{t}} = d_{\bar{t}}|R_1 = r\right) \tag{A.84}$$

$\mathbb{E}\left[Y_t\left(0, 0_{\bar{t}-1}\right)|R_1 = c\right]$ is point identified with a symmetric argument. Thus,

$$\text{ADTE}_{1,1}(c)$$

$$= \lim_{r\downarrow c} \mathbb{E}\left[Y_t|R_1 = r, D_2 = 0, \ldots, D_{\bar{t}} = 0\right] \times \lim_{r\downarrow c} \mathbb{P}\left(D_2 = 0, \ldots, D_{\bar{t}} = 0|R_1 = r\right)$$

$$+ \sum_{(d_2,\ldots,d_{\bar{t}})\neq 0_{\bar{t}-1}} \left(\lim_{r\downarrow c} \mathbb{E}\left[Y_1|R_1 = r, D_2 = d_2, \ldots, D_{\bar{t}} = d_{\bar{t}}\right] \right.$$

$$\left. + \lim_{r\downarrow c} \mathbb{E}\left[Y_t - Y_1|R_1 = r, D_2 = 0, \ldots, D_{\bar{t}} = 0\right] \right)$$

$$\times \lim_{r\downarrow c} \mathbb{P}\left(D_2 = d_2, \ldots, D_{\bar{t}} = d_{\bar{t}}|R_1 = r\right)$$

$$- \lim_{r\uparrow c} \mathbb{E}\left[Y_t|R_1 = r, D_2 = 0, \ldots, D_{\bar{t}} = 0\right] \times \lim_{r\uparrow c} \mathbb{P}\left(D_2 = 0, \ldots, D_{\bar{t}} = 0|R_1 = r\right)$$

$$- \sum_{(d_2,\ldots,d_{\bar{t}})\neq 0_{\bar{t}-1}} \left(\lim_{r\uparrow c} \mathbb{E}\left[Y_1|R_1 = r, D_2 = d_2, \ldots, D_{\bar{t}} = d_{\bar{t}}\right] \right.$$

$$\left. + \lim_{r\uparrow c} \mathbb{E}\left[Y_t - Y_1|R_1 = r, D_2 = 0, \ldots, D_{\bar{t}} = 0\right] \right)$$

$$\times \lim_{r\uparrow c} \mathbb{P}\left(D_2 = d_2, \ldots, D_{\bar{t}} = d_{\bar{t}}|R_1 = r\right) \tag{A.85}$$

By an application of the Law of Iterated Expectations, the target parameter can be compactly expressed as

$$\text{ADTE}_{1,1}(c)$$

$$= \lim_{r\downarrow c} \mathbb{E}\left[\prod_{s=2}^{\bar{t}} \mathbb{I}\left[D_s = 0\right] Y_t \Big| R_1 = r \right] - \lim_{r\uparrow c} \mathbb{E}\left[\prod_{s=2}^{\bar{t}} \mathbb{I}\left[D_s = 0\right] Y_t \Big| R_1 = r \right]$$

$$+ \sum_{(d_2,\ldots,d_{\bar{t}})\neq 0_{\bar{t}-1}} \lim_{r\downarrow c} \mathbb{E}\left[\prod_{s=2}^{\bar{t}} \mathbb{I}\left[D_s = d_s\right] Y_1 \Big| R_1 = r \right]$$

57

$$
-\sum_{(d_2,\ldots,d_{\bar{t}})\neq 0_{\bar{t}-1}} \lim_{r\uparrow c} \mathbb{E}\left[\prod_{s=2}^{\bar{t}} \mathbb{I}\left[D_s = d_s\right] Y_1 | R_1 = r\right]
$$

$$
-\lim_{r\downarrow c} \mathbb{E}\left[\prod_{s=2}^{\bar{t}} \mathbb{I}\left[D_s = 0\right](Y_t - Y_1)\,|R_1 = r\right]
$$

$$
+\lim_{r\uparrow c} \mathbb{E}\left[\prod_{s=2}^{\bar{t}} \mathbb{I}\left[D_s = 0\right](Y_t - Y_1)\,|R_1 = r\right]
$$

$$
+\lim_{r\downarrow c} \mathbb{E}\left[Y_t - Y_1 | R_1 = r, D_2 = 0, \ldots, D_{\bar{t}} = 0\right]
$$

$$
-\lim_{r\uparrow c} \mathbb{E}\left[Y_t - Y_1 | R_1 = r, D_2 = 0, \ldots, D_{\bar{t}} = 0\right] \tag{A.86}
$$

To conclude,

$$
\begin{aligned}
\mathrm{ADTE}_{1,\tau}(c) = {}& \lim_{r\downarrow c} \mathbb{E}\left[Y_1 | R_1 = r\right] - \lim_{r\uparrow c} \mathbb{E}\left[Y_1 | R_1 = r\right] \\
& + \lim_{r\downarrow c} \mathbb{E}\left[Y_{1+\tau} - Y_1 | R_1 = r, D_2 = 0, \ldots, D_{\bar{t}} = 0\right] \\
& - \lim_{r\uparrow c} \mathbb{E}\left[Y_{1+\tau} - Y_1 | R_1 = r, D_2 = 0, \ldots, D_{\bar{t}} = 0\right]
\end{aligned} \tag{A.87}
$$

which completes the proof.

B   Statistical Inference

In this section, I state and prove more general versions of Propositions 7 and 8. These theorems apply to derivatives of arbitrary order of the conditional means in the estimator of interest $\widehat{\theta}(h_n)$. In a similar fashion, I also develop statistical theory to assess the plausibility of Assumption 6. The notation, intermediate propositions, and structure of proofs largely mirror those in Calonico, Cattaneo and Titiunik (2014b).

B.1   Notation

For any pair of random variables A and B, consider the following conditional moments:

$$
\mu_A(r) \equiv \mathbb{E}\left[A | R = r\right] \qquad \sigma_{AA}^2(r) \equiv \mathbb{V}\left[A | R = r\right] \qquad \sigma_{AB}^2(r) \equiv \mathbb{C}\left[A, B | R = r\right] \tag{B.1}
$$

In addition, for $s \in \mathbb{N}$, the sth order derivative of the conditional mean with respect to the running variable is

$$
\mu_A^{(s)} \equiv \frac{\partial^s \mu_A(r)}{\partial r^s} \tag{B.2}
$$

Clearly, $\mu_A^{(0)}(r) = \mu_A(r)$. Then define the following limits for the running variable to the deterministic cutoff c:

$$\mu_{A+}^{(s)} \equiv \lim_{r \downarrow c} \mu_A^{(s)}(r) \qquad \mu_{A-}^{(s)} \equiv \lim_{r \uparrow c} \mu_A^{(s)}(r) \tag{B.3}$$

$$\sigma_{AA+}^2 \equiv \lim_{r \downarrow c} \sigma_{AA}^2(r) \qquad \sigma_{AA-}^2 \equiv \lim_{r \uparrow c} \sigma_{AA}^2(r) \tag{B.4}$$

$$\sigma_{AB+}^2 \equiv \lim_{r \downarrow c} \sigma_{AB}^2(r) \qquad \sigma_{AB-}^2 \equiv \lim_{r \uparrow c} \sigma_{AB}^2(r) \tag{B.5}$$

Focusing on the conditional mean, define

$$\beta_{A+,p} \equiv \left[\mu_{A+}, \mu_{A+}^{(1)}, \frac{\mu_{A+}^{(2)}}{2}, \dots, \frac{\mu_{A+}^{(p)}}{p!}\right]' \qquad \beta_{A-,p} \equiv \left[\mu_{A-}, \mu_{A-}^{(1)}, \frac{\mu_{A-}^{(2)}}{2}, \dots, \frac{\mu_{A-}^{(p)}}{p!}\right]' \tag{B.6}$$

so that

$$\mu_{A+}^{(s)} = s! e_s' \beta_{A+,p} \qquad \mu_{A-}^{(s)} = s! e_s' \beta_{A-,p} \tag{B.7}$$

where e_s is the conformable $(s+1)$th unit vector. For instance, if $p = 2$, then $e_1 = [0, 1, 0]'$. Recall that the estimand of interest is

$$\theta \equiv \mu_{Y+}^{(0)} - \mu_{Y-}^{(0)} + \frac{\mu_{W+}^{(0)}}{\mu_{D+}^{(0)}} - \frac{\mu_{W-}^{(0)}}{\mu_{D-}^{(0)}} \tag{B.8}$$

For consistency with the original paper, I consider the general target parameter

$$\theta_\nu \equiv \mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)} + \frac{\mu_{W+}^{(\nu)}}{\mu_{D+}^{(\nu)}} - \frac{\mu_{W-}^{(\nu)}}{\mu_{D-}^{(\nu)}} \tag{B.9}$$

where $\nu \in \mathbb{N}$. Let h be a positive bandwidth sequence and let

$$k_h(u) = k\left(\frac{u}{h}\right) \bigg/ h \tag{B.10}$$

where $k(\cdot)$ is the kernel function. Then, given the polynomial basis $x_p(r) = [1, r, r^2, \dots, r^p]'$, the local polynomial estimators of the conditional mean of A are

$$\widehat{\beta}_{A+,p}(h_n) \equiv \arg\min_{b \in \mathbb{R}^p} \sum_{i=1}^n \mathbb{I}[R_i \geq c] \left(A_i - x_p(R_i)' b\right)^2 k_{h_n}(R_i) \tag{B.11}$$

$$\widehat{\beta}_{A-,p}(h_n) \equiv \arg\min_{b \in \mathbb{R}^p} \sum_{i=1}^n \mathbb{I}[R_i < c] \left(A_i - x_p(R_i)' b\right)^2 k_{h_n}(R_i) \tag{B.12}$$

It follows that one can estimate the conditional νth derivative of a random variable A with

$$\widehat{\mu}_{A+,p}^{(\nu)}(h_n) = \nu! e_\nu' \widehat{\beta}_{A+,p}(h_n) \qquad \widehat{\mu}_{A-,p}^{(\nu)}(h_n) = \nu! e_\nu' \widehat{\beta}_{A-,p}(h_n) \tag{B.13}$$

As a consequence, a local polynomial estimator for the target parameter θ_ν is

$$\widehat{\theta}_{\nu,p}(h_n) \equiv \widehat{\mu}^{(\nu)}_{Y+,p}(h_n) - \widehat{\mu}^{(\nu)}_{Y-,p}(h_n) + \frac{\widehat{\mu}^{(\nu)}_{W+,p}(h_n)}{\widehat{\mu}^{(\nu)}_{D+,p}(h_n)} - \frac{\widehat{\mu}^{(\nu)}_{W-,p}(h_n)}{\widehat{\mu}^{(\nu)}_{D-,p}(h_n)} \tag{B.14}$$

Define the following data vectors and matrices:

$$\boldsymbol{Y} \equiv [Y_1,\ldots,Y_n]' \quad \boldsymbol{W} \equiv [W_1,\ldots,W_n]' \quad \boldsymbol{D} \equiv [D_1,\ldots,D_n]' \quad \boldsymbol{R} \equiv [R_1,\ldots,R_n]' \tag{B.15}$$

Also, for any $i \in \{1,\ldots,n\}$, define the residual

$$\varepsilon_{A,i} \equiv A_i - \mu_A(R_i) \tag{B.16}$$

and

$$\boldsymbol{\varepsilon_A} \equiv [\varepsilon_{A,1},\ldots,\varepsilon_{A,n}]' \qquad \boldsymbol{\Sigma_{AB}} \equiv \mathbb{E}[\boldsymbol{\varepsilon_A}\boldsymbol{\varepsilon'_B}|\boldsymbol{R}] = \mathrm{diag}\left(\sigma^2_{AB}(R_1),\ldots,\sigma^2_{AB}(R_n)\right) \tag{B.17}$$

In addition, define the following intermediate data vectors and matrices:

$$\boldsymbol{X_p}(h_n) \equiv [x_p(R_1/h_n),\ldots,x_p(R_n/h_n)]' \tag{B.18}$$

$$\boldsymbol{S_q}(h_n) \equiv [(R_1/h_n)^q,\ldots,(R_n/h_n)^q]' \tag{B.19}$$

$$\boldsymbol{Z_+}(h_n) \equiv \mathrm{diag}\left(\mathbb{I}[R_1 \geq c]\,k_{h_n}(R_1),\ldots,\mathbb{I}[R_n \geq c]\,k_{h_n}(R_n)\right) \tag{B.20}$$

$$\boldsymbol{Z_-}(h_n) \equiv \mathrm{diag}\left(\mathbb{I}[R_1 < c]\,k_{h_n}(R_1),\ldots,\mathbb{I}[R_n < c]\,k_{h_n}(R_n)\right) \tag{B.21}$$

Finally, given another positive bandwidth sequence b_n, define the scaled matrices

$$\Gamma_{+,p}(h_n) \equiv \boldsymbol{X_p}(h_n)'\,\boldsymbol{Z_+}(h_n)\,\boldsymbol{X_p}(h_n)/n \tag{B.22}$$

$$\Gamma_{-,p}(h_n) \equiv \boldsymbol{X_p}(h_n)'\,\boldsymbol{Z_-}(h_n)\,\boldsymbol{X_p}(h_n)/n \tag{B.23}$$

$$\vartheta_{+,p,q}(h_n) \equiv \boldsymbol{X_p}(h_n)'\,\boldsymbol{Z_+}(h_n)\,\boldsymbol{S_q}(h_n)/n \tag{B.24}$$

$$\vartheta_{-,p,q}(h_n) \equiv \boldsymbol{X_p}(h_n)'\,\boldsymbol{Z_-}(h_n)\,\boldsymbol{S_q}(h_n)/n \tag{B.25}$$

$$\Psi_{AB+,p,q}(h_n,b_n) \equiv \boldsymbol{X_p}(h_n)'\,\boldsymbol{Z_+}(h_n)\,\boldsymbol{\Sigma_{AB}}\boldsymbol{Z_+}(b_n)\,\boldsymbol{X_q}(b_n)/n \tag{B.26}$$

$$\Psi_{AB-,p,q}(h_n,b_n) \equiv \boldsymbol{X_p}(h_n)'\,\boldsymbol{Z_-}(h_n)\,\boldsymbol{\Sigma_{AB}}\boldsymbol{Z_-}(b_n)\,\boldsymbol{X_q}(b_n)/n \tag{B.27}$$

$$\Psi_{AB+,p}(h_n) \equiv \Psi_{AB+,p,p}(h_n,h_n) \tag{B.28}$$

$$\Psi_{AB-,p}(h_n) \equiv \Psi_{AB-,p,p}(h_n,h_n) \tag{B.29}$$

It follows that the vector of polynomial estimators in (B.11) and (B.12) can be expressed as

$$\widehat{\beta}_{A+,p}(h_n) = H_p(h_n)\,\Gamma^{-1}_{+,p}(h_n)\,\boldsymbol{X_p}(h_n)'\,\boldsymbol{Z_+}(h_n)\,\boldsymbol{A}/n \tag{B.30}$$

$$\widehat{\beta}_{A-,p}(h_n) = H_p(h_n)\,\Gamma_{-,p}^{-1}(h_n)\,\boldsymbol{X_p}(\boldsymbol{h_n})'\,\boldsymbol{Z_-}(\boldsymbol{h_n})\,\boldsymbol{A}/n \tag{B.31}$$

with $H_p(h_n) \equiv \mathrm{diag}\,(1, h_n^{-1}, h_n^{-2}, \ldots, h_n^{-p})$. To conclude,

$$\Gamma_p \equiv \int_0^\infty k(u)\,x_p(u)\,x_p(u)'\,du \tag{B.32}$$

$$\vartheta_{p,q} \equiv \int_0^\infty k(u)\,u^q x_p(u)\,du \tag{B.33}$$

$$\Psi_p \equiv \int_0^\infty k(u)^2\,x_p(u)\,x_p(u)'\,du \tag{B.34}$$

B.2   Bias-Corrected Confidence Intervals

In this section, I derive the asymptotic distribution of the bias-corrected estimator for the target parameter θ_ν. Following a similar approach to Calonico, Cattaneo and Titiunik (2014a), I first linearize the estimator $\widehat{\theta}_{\nu,p}(h_n)$ and compute the asymptotic bias and variance of the linearized estimator $\widetilde{\theta}_{\nu,p}(h_n)$. Then, I derive the asymptotic distribution of an estimator that subtracts the true (rather than estimated) bias from $\widehat{\theta}_{\nu,p}(h_n)$. Before stating the main proposition, it is convenient to define a few recurring bias- and variance-related terms.

B.2.1   Definitions of Bias Terms

First, let us define two widely used bias terms:

$$\mathcal{B}_{+,\nu,p,q}(h_n) \equiv \nu!\,e_\nu'\,\Gamma_{+,p}^{-1}(h_n)\,\vartheta_{+,p,q}(h_n) \qquad \mathcal{B}_{-,\nu,p,q}(h_n) \equiv \nu!\,e_\nu'\,\Gamma_{-,p}^{-1}(h_n)\,\vartheta_{-,p,q}(h_n) \tag{B.35}$$

As outlined in the proof of part (B) of Lemma S.A.3 in Calonico, Cattaneo and Titiunik (2014b), Lemma S.A.1 in the same paper implies that, if $nh_n \to \infty$ and $h_n \to 0$,

$$\mathcal{B}_{+,\nu,p,q}(h_n) = \nu!\,e_\nu'\,\Gamma_p^{-1}\vartheta_{p,q} + o_p(1) \qquad \mathcal{B}_{-,\nu,p,q}(h_n) = (-1)^{\nu+q}\,\nu!\,e_\nu'\,\Gamma_p^{-1}\vartheta_{p,q} + o_p(1) \tag{B.36}$$

where $e_\nu'\,\Gamma_p^{-1}\vartheta_{p,q}$ is observed and bounded. In addition, for any random variable A, define the following recurring bias terms:

$$\mathrm{B}_{A+,\nu,p,q}(h_n) \equiv \frac{\mu_{A+}^{(q)}}{q!}\mathcal{B}_{-,\nu,p,q}(h_n) \qquad \mathrm{B}_{A-,\nu,p,q}(h_n) \equiv \frac{\mu_{A-}^{(q)}}{q!}\mathcal{B}_{-,\nu,p,q}(h_n) \tag{B.37}$$

61

B.2.2 Definitions of Variance Terms

Let us define several variance and covariance terms. First, for any random variable A, the conditional variances of $\mu_{A+,p}^{(\nu)}(h_n)$ and $\mu_{A-,p}^{(\nu)}(h_n)$ are, respectively,

$$\mathcal{V}_{AA+,\nu,p}(h_n) \equiv \mathbb{V}\left[\mu_{A+,p}^{(\nu)}(h_n)\,\Big|\,\boldsymbol{R}\right] = \frac{\nu!^2}{nh_n^{2\nu}}e_\nu'\Gamma_{+,p}^{-1}(h_n)\,\Psi_{AA+,p}(h_n)\,\Gamma_{+,p}^{-1}(h_n)\,e_\nu \tag{B.38}$$

$$\mathcal{V}_{AA-,\nu,p}(h_n) \equiv \mathbb{V}\left[\mu_{A-,p}^{(\nu)}(h_n)\,\Big|\,\boldsymbol{R}\right] = \frac{\nu!^2}{nh_n^{2\nu}}e_\nu'\Gamma_{-,p}^{-1}(h_n)\,\Psi_{AA-,p}(h_n)\,\Gamma_{-,p}^{-1}(h_n)\,e_\nu \tag{B.39}$$

Moreover, by part (V) of Lemma S.A.3 in Calonico, Cattaneo and Titiunik (2014b), if $nh_n \to \infty$ and $h_n \to 0$,

$$\mathcal{V}_{AA+,\nu,p}(h_n) = \frac{1}{nh_n^{1+2\nu}}\frac{\sigma_{AA+}^2}{f_R(c)}\nu!^2 e_\nu'\Gamma_p^{-1}\Psi_p\Gamma_p^{-1}e_\nu\{1+o_p(1)\} \tag{B.40}$$

$$\mathcal{V}_{AA-,\nu,p}(h_n) = \frac{1}{nh_n^{1+2\nu}}\frac{\sigma_{AA-}^2}{f_R(c)}\nu!^2 e_\nu'\Gamma_p^{-1}\Psi_p\Gamma_p^{-1}e_\nu\{1+o_p(1)\} \tag{B.41}$$

where $e_\nu'\Gamma_p^{-1}\Psi_p\Gamma_p^{-1}e_\nu$ is observed and bounded. Analogously, for any pair of random variables A and B, the conditional covariance between $\mu_{A+,p}^{(\nu)}(h_n)$ and $\mu_{B+,p}^{(\nu)}(h_n)$ is

$$\mathcal{V}_{AB+,\nu,p}(h_n)$$
$$\equiv \mathbb{C}\left[\mu_{A+,p}^{(\nu)}(h_n),\mu_{B+,p}^{(\nu)}(h_n)\,\Big|\,\boldsymbol{R}\right]$$
$$= n^{-2}\nu!^2 e_\nu' H_p(h_n)\Gamma_{+,p}^{-1}(h_n)\,\boldsymbol{X_p}(\boldsymbol{h_n})'\,\boldsymbol{Z_+}(\boldsymbol{h_n})\,\boldsymbol{\Sigma_{AB}}\boldsymbol{Z_+}(\boldsymbol{h_n})\,\boldsymbol{X_p}(\boldsymbol{h_n})\,\Gamma_{+,p}^{-1}H_p(h_n)\,e_\nu$$
$$= n^{-2}\nu!^2 h_n^{-2\nu}e_\nu'\Gamma_{+,p}^{-1}(h_n)\,\boldsymbol{X_p}(\boldsymbol{h_n})'\,\boldsymbol{Z_+}(\boldsymbol{h_n})\,\boldsymbol{\Sigma_{AB}}\boldsymbol{Z_+}(\boldsymbol{h_n})\,\boldsymbol{X_p}(\boldsymbol{h_n})\,\Gamma_{+,p}^{-1}e_\nu$$
$$= n^{-1}\nu!^2 h_n^{-2\nu}e_\nu'\Gamma_{+,p}^{-1}(h_n)\,\Psi_{AB+,p}(h_n)\,\Gamma_{+,p}^{-1}e_\nu$$
$$= \frac{\nu!^2}{nh_n^{2\nu}}e_\nu'\Gamma_{+,p}^{-1}(h_n)\,\Psi_{AB+,p}(h_n)\,\Gamma_{+,p}^{-1}(h_n)\,e_\nu \tag{B.42}$$

Symmetrically,

$$\mathcal{V}_{AB-,\nu,p}(h_n) \equiv \mathbb{C}\left[\mu_{A-,p}^{(\nu)}(h_n),\mu_{B-,p}^{(\nu)}(h_n)\,\Big|\,\boldsymbol{R}\right]$$
$$= \frac{\nu!^2}{nh_n^{2\nu}}e_\nu'\Gamma_{-,p}^{-1}(h_n)\,\Psi_{AB-,p}(h_n)\,\Gamma_{-,p}^{-1}(h_n)\,e_\nu \tag{B.43}$$

As above, Lemma S.A.1 in Calonico, Cattaneo and Titiunik (2014b) implies that, if $nh_n \to \infty$ and $h_n \to 0$,

$$\mathcal{V}_{AB+,\nu,p}(h_n) = \frac{1}{nh_n^{1+2\nu}}\frac{\sigma_{AB+}^2}{f_R(c)}\nu!^2 e_\nu'\Gamma_p^{-1}\Psi_p\Gamma_p^{-1}e_\nu\{1+o_p(1)\} \tag{B.44}$$

$$\mathcal{V}_{AB-,\nu,p}(h_n) = \frac{1}{nh_n^{1+2\nu}}\frac{\sigma_{AB-}^2}{f_R(c)}\nu!^2 e_\nu'\Gamma_p^{-1}\Psi_p\Gamma_p^{-1}e_\nu\{1+o_p(1)\} \tag{B.45}$$

where $e_\nu'\Gamma_p^{-1}\Psi_p\Gamma_p^{-1}e_\nu$ is observed and bounded.

B.2.3 Linearization of the Estimator

Recall that the target estimand is

$$\theta_\nu \equiv \mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)} + \frac{\mu_{W+}^{(\nu)}}{\mu_{D+}^{(\nu)}} - \frac{\mu_{W-}^{(\nu)}}{\mu_{D-}^{(\nu)}} \tag{B.46}$$

and the estimator of interest is

$$\widehat{\theta}_{\nu,p}(h_n) \equiv \widehat{\mu}_{Y+,p}^{(\nu)}(h_n) - \widehat{\mu}_{Y-,p}^{(\nu)}(h_n) + \frac{\widehat{\mu}_{W+,p}^{(\nu)}(h_n)}{\widehat{\mu}_{D+,p}^{(\nu)}(h_n)} - \frac{\widehat{\mu}_{W-,p}^{(\nu)}(h_n)}{\widehat{\mu}_{D-,p}^{(\nu)}(h_n)} \tag{B.47}$$

Let us define the following ratios of polynomial estimators on either side of the cutoff:

$$\widehat{\eta}_{+,\nu,p}(h_n) \equiv \frac{\widehat{\mu}_{W+,p}^{(\nu)}(h_n)}{\widehat{\mu}_{D+,p}^{(\nu)}(h_n)} \qquad \widehat{\eta}_{-,\nu,p}(h_n) \equiv \frac{\widehat{\mu}_{W-,p}^{(\nu)}(h_n)}{\widehat{\mu}_{D-,p}^{(\nu)}(h_n)} \tag{B.48}$$

Their population counterparts are

$$\eta_{+,\nu} \equiv \frac{\mu_{W+}^{(\nu)}}{\mu_{D+}^{(\nu)}} \qquad \eta_{-,\nu} \equiv \frac{\mu_{W-}^{(\nu)}}{\mu_{D-}^{(\nu)}} \tag{B.49}$$

To determine the asymptotic behavior of $\widehat{\eta}_{+,\nu,p}(h_n)$ and $\widehat{\eta}_{-,\nu,p}(h_n)$, I proceed as in Calonico, Cattaneo and Titiunik (2014a) and perform a second-order Taylor expansion:

$$\begin{aligned}
\widehat{\eta}_{+,\nu,p}(h_n) - \eta_{+,\nu} &\approx \frac{1}{\mu_{D+}^{(\nu)}}\left(\widehat{\mu}_{W+,p}^{(\nu)}(h_n) - \mu_{W+}^{(\nu)}\right) - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}\left(\widehat{\mu}_{D+,p}^{(\nu)}(h_n) - \mu_{D+}^{(\nu)}\right) \\
&\quad - \frac{1}{\left(\mu_{D+}^{(\nu)}\right)^2}\left(\widehat{\mu}_{W+,p}^{(\nu)}(h_n) - \mu_{W+}^{(\nu)}\right)\left(\widehat{\mu}_{D+,p}^{(\nu)}(h_n) - \mu_{D+}^{(\nu)}\right) \\
&\quad + \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^3}\left(\widehat{\mu}_{D+,p}^{(\nu)}(h_n) - \mu_{D+}^{(\nu)}\right)^2
\end{aligned} \tag{B.50}$$

Analogously,

$$\begin{aligned}
\widehat{\eta}_{-,\nu,p}(h_n) - \eta_{-,\nu} &\approx \frac{1}{\mu_{D-}^{(\nu)}}\left(\widehat{\mu}_{W-,p}^{(\nu)}(h_n) - \mu_{W-}^{(\nu)}\right) - \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^2}\left(\widehat{\mu}_{D-,p}^{(\nu)}(h_n) - \mu_{D-}^{(\nu)}\right) \\
&\quad - \frac{1}{\left(\mu_{D-}^{(\nu)}\right)^2}\left(\widehat{\mu}_{W-,p}^{(\nu)}(h_n) - \mu_{W-}^{(\nu)}\right)\left(\widehat{\mu}_{D-,p}^{(\nu)}(h_n) - \mu_{D-}^{(\nu)}\right) \\
&\quad + \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^3}\left(\widehat{\mu}_{D-,p}^{(\nu)}(h_n) - \mu_{D-}^{(\nu)}\right)^2
\end{aligned} \tag{B.51}$$

Let us now define a linear term and a remainder term:

$$\widetilde{\eta}_{+,\nu,p}\left(h_n\right) \equiv \frac{1}{\mu_{D+}^{(\nu)}}\left(\widehat{\mu}_{W+,p}^{(\nu)}\left(h_n\right) - \mu_{W+}^{(\nu)}\right) - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}\left(\widehat{\mu}_{D+,p}^{(\nu)}\left(h_n\right) - \mu_{D+}^{(\nu)}\right) \tag{B.52}$$

$$R_{+,n} \equiv \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^3}\left(\widehat{\mu}_{D+,p}^{(\nu)}\left(h_n\right) - \mu_{D+}^{(\nu)}\right)^2$$

$$- \frac{1}{\left(\mu_{D+}^{(\nu)}\right)^2}\left(\widehat{\mu}_{W+,p}^{(\nu)}\left(h_n\right) - \mu_{W+}^{(\nu)}\right)\left(\widehat{\mu}_{D+,p}^{(\nu)}\left(h_n\right) - \mu_{D+}^{(\nu)}\right) \tag{B.53}$$

Symmetrically,

$$\widetilde{\eta}_{-,\nu,p}\left(h_n\right) \equiv \frac{1}{\mu_{D-}^{(\nu)}}\left(\widehat{\mu}_{W-,p}^{(\nu)}\left(h_n\right) - \mu_{W-}^{(\nu)}\right) - \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^2}\left(\widehat{\mu}_{D-,p}^{(\nu)}\left(h_n\right) - \mu_{D-}^{(\nu)}\right) \tag{B.54}$$

$$R_{-,n} \equiv \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^3}\left(\widehat{\mu}_{D-,p}^{(\nu)}\left(h_n\right) - \mu_{D-}^{(\nu)}\right)^2$$

$$- \frac{1}{\left(\mu_{D-}^{(\nu)}\right)^2}\left(\widehat{\mu}_{W-,p}^{(\nu)}\left(h_n\right) - \mu_{W-}^{(\nu)}\right)\left(\widehat{\mu}_{D-,p}^{(\nu)}\left(h_n\right) - \mu_{D-}^{(\nu)}\right) \tag{B.55}$$

Clearly,

$$\widehat{\eta}_{+,\nu,p}\left(h_n\right) - \eta_{+,\nu} = \widetilde{\eta}_{+,\nu,p}\left(h_n\right) + R_{+,n} \tag{B.56}$$

$$\widehat{\eta}_{-,\nu,p}\left(h_n\right) - \eta_{-,\nu} = \widetilde{\eta}_{-,\nu,p}\left(h_n\right) + R_{-,n} \tag{B.57}$$

Finally, define

$$\widetilde{\eta}_{\nu,p}\left(h_n\right) \equiv \widetilde{\eta}_{+,\nu,p}\left(h_n\right) - \widetilde{\eta}_{-,\nu,p}\left(h_n\right) \qquad R_n \equiv R_{+,n} - R_{-,n} \tag{B.58}$$

which implies that

$$\widehat{\eta}_{\nu,p}\left(h_n\right) = \widetilde{\eta}_{\nu,p}\left(h_n\right) + R_n \tag{B.59}$$

To conclude, the linearized estimator for θ_ν is

$$\widetilde{\theta}_{\nu,p}\left(h_n\right) \equiv \widehat{\mu}_{Y+,p}^{(\nu)}\left(h_n\right) - \widehat{\mu}_{Y-,p}^{(\nu)}\left(h_n\right) + \widetilde{\eta}_{\nu,p}\left(h_n\right) \tag{B.60}$$

B.2.4   General Theorem: Statement

Theorem B.1 *Suppose Assumptions 8 and 9 hold with $\delta \geq p + 2$ and $nh_n \to \infty$.*

(R) If $h_n \to 0$ and $n h_n^{1+2\nu} \to 0$, then

$$R_n = O_p \left(h_n^{2(p+1-\nu)} + \frac{1}{n h_n^{1+2\nu}} \right)$$

(B) If $h_n \to 0$, then

$$\mathbb{E} \left[\widetilde{\theta}_{\nu,p} (h_n) \,\middle|\, \mathbf{R} \right] = \mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)} + h_n^{p+1-\nu} \mathrm{B}_{\nu,p,p+1} (h_n)$$
$$+ h_n^{p+2-\nu} \mathrm{B}_{\nu,p,p+2} (h_n) + o_p \left(h_n^{p+2-\nu} \right)$$

where

$$\mathrm{B}_{\nu,p,q} (h_n) \equiv \mathrm{B}_{Y,\nu,p,q} (h_n) + \widetilde{\mathrm{B}}_{\nu,p,q} (h_n)$$

with

$$\mathrm{B}_{Y,\nu,p,q} (h_n) \equiv \mathrm{B}_{Y+,\nu,p,q} (h_n) - \mathrm{B}_{Y-,\nu,p,q} (h_n)$$
$$\widetilde{\mathrm{B}}_{\nu,p,q} (h_n) \equiv \widetilde{\mathrm{B}}_{+,\nu,p,q} (h_n) - \widetilde{\mathrm{B}}_{-,\nu,p,q} (h_n)$$

and

$$\widetilde{\mathrm{B}}_{+,\nu,p,q} (h_n) \equiv \frac{1}{\mu_{D+}^{(\nu)}} \mathrm{B}_{W+,\nu,p,q} (h_n) - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)} \right)^2} \mathrm{B}_{D+,\nu,p,q} (h_n)$$

$$\widetilde{\mathrm{B}}_{-,\nu,p,q} (h_n) \equiv \frac{1}{\mu_{D-}^{(\nu)}} \mathrm{B}_{W-,\nu,p,q} (h_n) - \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)} \right)^2} \mathrm{B}_{D-,\nu,p,q} (h_n)$$

(V) If $h_n \to 0$, then

$$\mathrm{V}_{\nu,p} (h_n) \equiv \mathbb{V} \left[\widetilde{\theta}_{\nu,p} (h_n) \,\middle|\, \mathbf{R} \right] = \mathrm{V}_{Y,\nu,p} (h_n) + \widetilde{\mathrm{V}}_{\nu,p} (h_n) + 2 \widetilde{\mathrm{V}}_{Y,\nu,p} (h_n)$$

where

$$\mathrm{V}_{Y,\nu,p} (h_n) \equiv \mathcal{V}_{YY+,\nu,p} (h_n) + \mathcal{V}_{YY-,\nu,p} (h_n)$$
$$\widetilde{\mathrm{V}}_{\nu,p} (h_n) \equiv \widetilde{\mathrm{V}}_{+,\nu,p} (h_n) + \widetilde{\mathrm{V}}_{-,\nu,p} (h_n)$$
$$\widetilde{\mathrm{V}}_{Y,\nu,p} (h_n) \equiv \widetilde{\mathrm{V}}_{Y+,\nu,p} (h_n) + \widetilde{\mathrm{V}}_{Y-,\nu,p} (h_n)$$

with

$$\widetilde{\mathrm{V}}_{+,\nu,p} (h_n) \equiv \frac{1}{\left(\mu_{D+}^{(\nu)} \right)^2} \mathcal{V}_{WW+,\nu,p} (h_n) + \frac{\left(\mu_{W+}^{(\nu)} \right)^2}{\left(\mu_{D+}^{(\nu)} \right)^4} \mathcal{V}_{DD+,\nu,p} (h_n) - \frac{2 \mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)} \right)^3} \mathcal{V}_{WD+,\nu,p} (h_n)$$

65

$$\widetilde{V}_{-,\nu,p}(h_n) \equiv \frac{1}{\left(\mu_{D-}^{(\nu)}\right)^2}\mathcal{V}_{WW-,\nu,p}(h_n) + \frac{\left(\mu_{W-}^{(\nu)}\right)^2}{\left(\mu_{D-}^{(\nu)}\right)^4}\mathcal{V}_{DD-,\nu,p}(h_n) - \frac{2\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^3}\mathcal{V}_{WD-,\nu,p}(h_n)$$

and

$$\widetilde{V}_{Y+,\nu,p}(h_n) \equiv \frac{1}{\mu_{D+}^{(\nu)}}\mathcal{V}_{YW+,\nu,p}(h_n) - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}\mathcal{V}_{YD+,\nu,p}(h_n)$$

$$\widetilde{V}_{Y-,\nu,p}(h_n) \equiv \frac{1}{\mu_{D-}^{(\nu)}}\mathcal{V}_{YW-,\nu,p}(h_n) - \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^2}\mathcal{V}_{YD-,\nu,p}(h_n)$$

(D) If $nh_n^{2p+5} \to 0$ and $nh_n^{1+2\nu} \to 0$, then

$$T^{\text{BC}}(h_n) \equiv \frac{\widetilde{\theta}_{\nu,p}(h_n) - \left(\mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)}\right) - h_n^{p+1-\nu}\text{B}_{\nu,p,p+1}(h_n)}{\sqrt{\text{V}_{\nu,p}(h_n)}} \xrightarrow{d} \mathcal{N}(0,1)$$

B.2.5 General Theorem: Proof

(R) Consider $R_{+,n}$. If $h_n \to 0$ and $nh_n^{2p+5} \to 0$, part (D) of Lemma S.A.3 in Calonico, Cattaneo and Titiunik (2014b) implies that

$$T_{W+,\nu,p}(h_n) \equiv \frac{\widehat{\mu}_{W+,p}^{(\nu)}(h_n) - \mu_{W+}^{(\nu)} - h_n^{p+1-\nu}\frac{\mu_+^{(p+1)}}{(p+1)!}\mathcal{B}_{+,\nu,p,p+1}(h_n)}{\sqrt{\mathcal{V}_{WW+,\nu,p}(h_n)}} \xrightarrow{d} \mathcal{N}(0,1)$$

with

$$\mathcal{B}_{+,\nu,p,r}(h_n) = \nu! e_\nu' \Gamma_p^{-1} \vartheta_{p,r} + o_p(1)$$

and

$$\mathcal{V}_{WW+,\nu,p}(h_n) = \frac{1}{nh_n^{1+2\nu}}\frac{\sigma_{WW+}^2}{f_R(c)}\nu!^2 e_\nu' \Gamma_p^{-1}\Psi_p\Gamma_p^{-1}e_\nu\{1+o_p(1)\}$$

Thus,

$$T_{W+,\nu,p}(h_n) = O_p(1) \iff \widehat{\mu}_{W+,p}^{(\nu)}(h_n) - \mu_{W+}^{(\nu)} = O_p\left(h_n^{p+1-\nu}\right) + O_p\left(\sqrt{\frac{1}{nh_n^{1+2\nu}}}\right)$$

$$\iff \widehat{\mu}_{W+,p}^{(\nu)}(h_n) - \mu_{W+}^{(\nu)} = O_p\left(h_n^{p+1-\nu} + \sqrt{\frac{1}{nh_n^{1+2\nu}}}\right)$$

Following similar steps,

$$\widehat{\mu}_{D+,p}^{(\nu)}(h_n) - \mu_{D+}^{(\nu)} = O_p\left(h_n^{p+1-\nu} + \sqrt{\frac{1}{nh_n^{1+2\nu}}}\right)$$

As a consequence, their product is

$$\left(\widehat{\mu}_{W+,p}^{(\nu)}(h_n) - \mu_{W+}^{(\nu)}\right)\left(\widehat{\mu}_{D+,p}^{(\nu)}(h_n) - \mu_{D+}^{(\nu)}\right) = O_p\left(\left(h_n^{p+1-\nu} + \sqrt{\frac{1}{nh_n^{1+2\nu}}}\right)^2\right)$$

$$= O_p\left(h_n^{2(p+1-\nu)} + \frac{1}{nh_n^{1+2\nu}}\right)$$

where the second equality follows from the fact that, for any $a, b \in \mathbb{R}_+$, $a + b \leq \left(\sqrt{a} + \sqrt{b}\right)^2 \leq 2(a+b)$ and thus $\left(\sqrt{a} + \sqrt{b}\right)^2 = O(a+b)$. Analogously,

$$\left(\widehat{\mu}_{D+,p}^{(\nu)}(h_n) - \mu_{D+}^{(\nu)}\right)^2 = O_p\left(h_n^{2(p+1-\nu)} + \frac{1}{nh_n^{1+2\nu}}\right)$$

Finally, because $\frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^3}$ and $\frac{1}{\left(\mu_{D+}^{(\nu)}\right)^2}$ are deterministic constants,

$$R_{+,n} = \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^3}O_p\left(h_n^{2(p+1-\nu)} + \frac{1}{nh_n^{1+2\nu}}\right) + \frac{1}{\left(\mu_{D+}^{(\nu)}\right)^2}O_p\left(h_n^{2(p+1-\nu)} + \frac{1}{nh_n^{1+2\nu}}\right)$$

$$= O_p\left(h_n^{2(p+1-\nu)} + \frac{1}{nh_n^{1+2\nu}}\right) + O_p\left(h_n^{2(p+1-\nu)} + \frac{1}{nh_n^{1+2\nu}}\right)$$

$$= O_p\left(h_n^{2(p+1-\nu)} + \frac{1}{nh_n^{1+2\nu}}\right)$$

which is the desired result. The proof for $R_{-,n}$ is symmetric. Combining results,

$$R_n \equiv R_{+,n} - R_{-,n}$$

$$= O_p\left(h_n^{2(p+1-\nu)} + \frac{1}{nh_n^{1+2\nu}}\right) - O_p\left(h_n^{2(p+1-\nu)} + \frac{1}{nh_n^{1+2\nu}}\right)$$

$$= O_p\left(h_n^{2(p+1-\nu)} + \frac{1}{nh_n^{1+2\nu}}\right)$$

(B) First, consider $\widehat{\mu}_{Y+,p}^{(\nu)}(h_n)$ and $\widehat{\mu}_{Y-,p}^{(\nu)}(h_n)$. Part (B) of Lemma S.A.3 in Calonico, Cattaneo and Titiunik (2014b) implies that, as $h_n \to 0$, their conditional expectations are, respectively,

$$\mathbb{E}\left[\widehat{\mu}_{Y+,p}^{(\nu)}(h_n)\middle|\mathbf{R}\right] = \mu_{Y+}^{(\nu)} + h_n^{p+1-\nu}\mathrm{B}_{Y+,\nu,p,p+1}(h_n)$$

$$+ h_n^{p+2-\nu} \mathrm{B}_{Y+,\nu,p,p+2} (h_n) + o_p \left(h_n^{p+2-\nu} \right)$$

$$\mathbb{E} \left[\widehat{\mu}_{Y-,p}^{(\nu)} (h_n) \,\middle|\, \boldsymbol{R} \right] = \mu_{Y-}^{(\nu)} + h_n^{p+1-\nu} \mathrm{B}_{Y-,\nu,p,p+1} (h_n)$$

$$+ h_n^{p+2-\nu} \mathrm{B}_{Y-,\nu,p,p+2} (h_n) + o_p \left(h_n^{p+2-\nu} \right)$$

Define

$$\mathrm{B}_{Y,\nu,p,q} (h_n) \equiv \mathrm{B}_{Y+,\nu,p,q} (h_n) - \mathrm{B}_{Y-,\nu,p,q} (h_n)$$

Then, by the linearity of the expectation operator,

$$\mathbb{E} \left[\widehat{\mu}_{Y+,p}^{(\nu)} (h_n) - \widehat{\mu}_{Y-,p}^{(\nu)} (h_n) \,\middle|\, \boldsymbol{R} \right] = \mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)} + h_n^{p+1-\nu} \mathrm{B}_{Y,\nu,p,p+1} (h_n)$$

$$+ h_n^{p+2-\nu} \mathrm{B}_{Y,\nu,p,p+2} (h_n) + o_p \left(h_n^{p+2-\nu} \right)$$

Second, consider $\widetilde{\eta}_{+,\nu,p} (h_n)$. Part (B) of Lemma S.A.3 in Calonico, Cattaneo and Titiunik (2014b) implies that, as $h_n \to 0$,

$$\mathbb{E} \left[\widehat{\mu}_{W+,p}^{(\nu)} (h_n) - \mu_{W+}^{(\nu)} \,\middle|\, \boldsymbol{R} \right] = h_n^{p+1-\nu} \mathrm{B}_{W+,\nu,p,p+1} (h_n)$$

$$+ h_n^{p+2-\nu} \mathrm{B}_{W+,\nu,p,p+2} (h_n) + o_p \left(h_n^{p+2-\nu} \right)$$

$$\mathbb{E} \left[\widehat{\mu}_{D+,p}^{(\nu)} (h_n) - \mu_{D+}^{(\nu)} \,\middle|\, \boldsymbol{R} \right] = h_n^{p+1-\nu} \mathrm{B}_{D+,\nu,p,p+1} (h_n)$$

$$+ h_n^{p+2-\nu} \mathrm{B}_{D+,\nu,p,p+2} (h_n) + o_p \left(h_n^{p+2-\nu} \right)$$

Then, by the linearity of the expectation operator,

$$\mathbb{E} \left[\widetilde{\eta}_{+,\nu,p} (h_n) \,\middle|\, \boldsymbol{R} \right] = \frac{1}{\mu_{D+}^{(\nu)}} \left(h_n^{p+1-\nu} \mathrm{B}_{W+,\nu,p,p+1} (h_n) + h_n^{p+2-\nu} \mathrm{B}_{W+,\nu,p,p+2} (h_n) \right)$$

$$- \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)} \right)^2} \left(h_n^{p+1-\nu} \mathrm{B}_{D+,\nu,p,p+1} (h_n) + h_n^{p+2-\nu} \mathrm{B}_{D+,\nu,p,p+2} (h_n) \right)$$

$$+ o_p \left(h_n^{p+2-\nu} \right)$$

Further define

$$\widetilde{\mathrm{B}}_{+,\nu,p,r} (h_n) \equiv \frac{1}{\mu_{D+}^{(\nu)}} \mathrm{B}_{W+,\nu,p,r} (h_n) - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)} \right)^2} \mathrm{B}_{D+,\nu,p,r} (h_n)$$

Thus,

$$\mathbb{E} \left[\widetilde{\eta}_{+,\nu,p} (h_n) \,\middle|\, \boldsymbol{R} \right] = h_n^{p+1-\nu} \widetilde{\mathrm{B}}_{+,\nu,p,p+1} (h_n) + h_n^{p+2-\nu} \widetilde{\mathrm{B}}_{+,\nu,p,p+2} (h_n) + o_p \left(h_n^{p+2-\nu} \right)$$

A similar set of arguments yields

$$\mathbb{E}\left[\widetilde{\eta}_{-,\nu,p}\left(h_n\right)|\boldsymbol{R}\right] = h_n^{p+1-\nu}\widetilde{\mathrm{B}}_{-,\nu,p,p+1}\left(h_n\right) + h_n^{p+2-\nu}\widetilde{\mathrm{B}}_{-,\nu,p,p+2}\left(h_n\right) + o_p\left(h_n^{p+2-\nu}\right)$$

where

$$\widetilde{\mathrm{B}}_{-,\nu,p,r}\left(h_n\right) \equiv \frac{1}{\mu_{D-}^{(\nu)}}\mathrm{B}_{W-,\nu,p,r}\left(h_n\right) - \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^2}\mathrm{B}_{D-,\nu,p,r}\left(h_n\right)$$

Also define

$$\widetilde{\mathrm{B}}_{\nu,p,r}\left(h_n\right) \equiv \widetilde{\mathrm{B}}_{+,\nu,p,r}\left(h_n\right) - \widetilde{\mathrm{B}}_{-,\nu,p,r}\left(h_n\right)$$

Leveraging the linearity of the expectation operator,

$$\mathbb{E}\left[\widetilde{\eta}_{\nu,p}\left(h_n\right)|\boldsymbol{R}\right] = h_n^{p+1-\nu}\widetilde{\mathrm{B}}_{\nu,p,p+1}\left(h_n\right) + h_n^{p+2-\nu}\widetilde{\mathrm{B}}_{\nu,p,p+2}\left(h_n\right) + o_p\left(h_n^{p+2-\nu}\right)$$

Finally, define

$$\mathrm{B}_{\nu,p,r}\left(h_n\right) \equiv \mathrm{B}_{Y,\nu,p,r}\left(h_n\right) + \widetilde{\mathrm{B}}_{\nu,p,r}\left(h_n\right)$$

Since $\widetilde{\theta}_{\nu,p}\left(h_n\right) \equiv \widehat{\mu}_{Y+,p}^{(\nu)}\left(h_n\right) - \widehat{\mu}_{Y-,p}^{(\nu)}\left(h_n\right) + \widetilde{\eta}_{\nu,p}\left(h_n\right)$, the linearity of the expectation implies that

$$\mathbb{E}\left[\widetilde{\theta}_{\nu,p}\left(h_n\right)\Big|\boldsymbol{R}\right] = \mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)} + h_n^{p+1-\nu}\mathrm{B}_{\nu,p,p+1}\left(h_n\right)$$
$$+ h_n^{p+2-\nu}\mathrm{B}_{\nu,p,p+2}\left(h_n\right) + o_p\left(h_n^{p+2-\nu}\right)$$

which completes the proof.

(V) The conditional variance of the linearized estimator $\widetilde{\theta}_{\nu,p}\left(h_n\right)$ can be written as

$$\mathbb{V}\left[\widetilde{\theta}_{\nu,p}\left(h_n\right)\Big|\boldsymbol{R}\right] = \mathbb{V}\left[\widehat{\mu}_{Y+,p}^{(\nu)}\left(h_n\right) - \widehat{\mu}_{Y-,p}^{(\nu)}\left(h_n\right)\Big|\boldsymbol{R}\right] + \mathbb{V}\left[\widetilde{\eta}_{\nu,p}\left(h_n\right)\Big|\boldsymbol{R}\right]$$
$$+ 2 \times \mathbb{C}\left[\widehat{\mu}_{Y+,p}^{(\nu)}\left(h_n\right) - \widehat{\mu}_{Y-,p}^{(\nu)}\left(h_n\right), \widetilde{\eta}_{\nu,p}\left(h_n\right)\Big|\boldsymbol{R}\right]$$

First, consider the conditional variance of $\widehat{\mu}_{Y+,p}^{(\nu)}\left(h_n\right) - \widehat{\mu}_{Y-,p}^{(\nu)}\left(h_n\right)$. Notice that

$$\mathrm{V}_{Y,\nu,p}\left(h_n\right) \equiv \mathbb{V}\left[\widehat{\mu}_{Y+,p}^{(\nu)}\left(h_n\right) - \widehat{\mu}_{Y-,p}^{(\nu)}\left(h_n\right)\Big|\boldsymbol{R}\right]$$
$$= \mathbb{V}\left[\widehat{\mu}_{Y+,p}^{(\nu)}\left(h_n\right)\Big|\boldsymbol{R}\right] + \mathbb{V}\left[\widehat{\mu}_{Y-,p}^{(\nu)}\left(h_n\right)\Big|\boldsymbol{R}\right]$$

69

where the second equality follows from the bilinearity of the covariance and the conditional independence of observations on either side of the cutoff c. Furthermore, part (V) of Lemma S.A.3 in Calonico, Cattaneo and Titiunik (2014b) implies that, as $h_n \to 0$,

$$\mathrm{V}_{Y,\nu,p}(h_n) = \mathcal{V}_{YY+,\nu,p}(h_n) + \mathcal{V}_{YY-,\nu,p}(h_n)$$

Second, consider the conditional variance of $\widetilde{\eta}_{\nu,p}(h_n)$. By its definition,

$$\mathbb{V}\left[\widetilde{\eta}_{\nu,p}(h_n)\,\Big|\,\boldsymbol{R}\right] = \mathbb{V}\left[\widetilde{\eta}_{+,\nu,p}(h_n) - \widetilde{\eta}_{-,\nu,p}(h_n)\,\Big|\,\boldsymbol{R}\right]$$
$$= \mathbb{V}\left[\widetilde{\eta}_{+,\nu,p}(h_n)\,\Big|\,\boldsymbol{R}\right] + \mathbb{V}\left[\widetilde{\eta}_{-,\nu,p}(h_n)\,\Big|\,\boldsymbol{R}\right]$$

where, as above, the second equality follows from the bilinearity of the covariance and the conditional independence of observations on either side of the cutoff c. Furthermore, by the bilinearity of the covariance,

$$\mathbb{V}\left[\widetilde{\eta}_{+,\nu,p}(h_n)\,\Big|\,\boldsymbol{R}\right] = \frac{1}{\left(\mu_{D+}^{(\nu)}\right)^2}\mathbb{V}\left[\widehat{\mu}_{W+,p}^{(\nu)}(h_n)\,\Big|\,\boldsymbol{R}\right] + \frac{\left(\mu_{W+}^{(\nu)}\right)^2}{\left(\mu_{D+}^{(\nu)}\right)^4}\mathbb{V}\left[\widehat{\mu}_{D+,p}^{(\nu)}(h_n)\,\Big|\,\boldsymbol{R}\right]$$
$$- \frac{2\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^3}\mathbb{C}\left[\widehat{\mu}_{W+,p}^{(\nu)}(h_n),\widehat{\mu}_{D+,p}^{(\nu)}(h_n)\,\Big|\,\boldsymbol{R}\right]$$

$$\mathbb{V}\left[\widetilde{\eta}_{-,\nu,p}(h_n)\,\Big|\,\boldsymbol{R}\right] = \frac{1}{\left(\mu_{D-}^{(\nu)}\right)^2}\mathbb{V}\left[\widehat{\mu}_{W-,p}^{(\nu)}(h_n)\,\Big|\,\boldsymbol{R}\right] + \frac{\left(\mu_{W-}^{(\nu)}\right)^2}{\left(\mu_{D-}^{(\nu)}\right)^4}\mathbb{V}\left[\widehat{\mu}_{D-,p}^{(\nu)}(h_n)\,\Big|\,\boldsymbol{R}\right]$$
$$- \frac{2\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^3}\mathbb{C}\left[\widehat{\mu}_{W-,p}^{(\nu)}(h_n),\widehat{\mu}_{D-,p}^{(\nu)}(h_n)\,\Big|\,\boldsymbol{R}\right]$$

Leveraging part (V) of Lemma S.A.3 in Calonico, Cattaneo and Titiunik (2014b), define the following variance terms:

$$\widetilde{\mathrm{V}}_{+,\nu,p}(h_n) \equiv \mathbb{V}\left[\widetilde{\eta}_{+,\nu,p}(h_n)\,\Big|\,\boldsymbol{R}\right]$$
$$= \frac{1}{\left(\mu_{D+}^{(\nu)}\right)^2}\mathcal{V}_{WW+,\nu,p}(h_n) + \frac{\left(\mu_{W+}^{(\nu)}\right)^2}{\left(\mu_{D+}^{(\nu)}\right)^4}\mathcal{V}_{DD+,\nu,p}(h_n) - \frac{2\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^3}\mathcal{V}_{WD+,\nu,p}(h_n)$$
$$\widetilde{\mathrm{V}}_{-,\nu,p}(h_n) \equiv \mathbb{V}\left[\widetilde{\eta}_{-,\nu,p}(h_n)\,\Big|\,\boldsymbol{R}\right]$$
$$= \frac{1}{\left(\mu_{D-}^{(\nu)}\right)^2}\mathcal{V}_{WW-,\nu,p}(h_n) + \frac{\left(\mu_{W-}^{(\nu)}\right)^2}{\left(\mu_{D-}^{(\nu)}\right)^4}\mathcal{V}_{DD-,\nu,p}(h_n) - \frac{2\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^3}\mathcal{V}_{WD-,\nu,p}(h_n)$$

Thus, the conditional variance of $\widetilde{\eta}_{\nu,p}(h_n)$ is

$$\widetilde{V}_{\nu,p}(h_n) \equiv \mathbb{V}\left[\widetilde{\eta}_{\nu,p}(h_n)\,\middle|\,\boldsymbol{R}\right] = \widetilde{V}_{+,\nu,p}(h_n) + \widetilde{V}_{-,\nu,p}(h_n)$$

Finally, consider the covariance of $\widehat{\mu}_{Y+,p}^{(\nu)}(h_n) - \widehat{\mu}_{Y-,p}^{(\nu)}(h_n)$ and $\widetilde{\eta}_{\nu,p}(h_n)$:

$$\mathbb{C}\left[\widehat{\mu}_{Y+,p}^{(\nu)}(h_n) - \widehat{\mu}_{Y-,p}^{(\nu)}(h_n),\widetilde{\eta}_{\nu,p}(h_n)\,\middle|\,\boldsymbol{R}\right] = \mathbb{C}\left[\widehat{\mu}_{Y+,p}^{(\nu)}(h_n),\widetilde{\eta}_{+,\nu,p}(h_n)\,\middle|\,\boldsymbol{R}\right]$$
$$+ \mathbb{C}\left[\widehat{\mu}_{Y-,p}^{(\nu)}(h_n),\widetilde{\eta}_{-,\nu,p}(h_n)\,\middle|\,\boldsymbol{R}\right]$$

where the second equality follows from the bilinearity of the covariance and the conditional independence of observations on either side of the cutoff c. Expanding the covariances,

$$\mathbb{C}\left[\widehat{\mu}_{Y+,p}^{(\nu)}(h_n),\widetilde{\eta}_{+,\nu,p}(h_n)\,\middle|\,\boldsymbol{R}\right] = \frac{1}{\mu_{D+}^{(\nu)}}\mathbb{C}\left[\widehat{\mu}_{Y+,p}^{(\nu)}(h_n),\widehat{\mu}_{W+,p}^{(\nu)}(h_n)\,\middle|\,\boldsymbol{R}\right]$$
$$- \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}\mathbb{C}\left[\widehat{\mu}_{Y+,p}^{(\nu)}(h_n),\widehat{\mu}_{D+,p}^{(\nu)}(h_n)\,\middle|\,\boldsymbol{R}\right]$$
$$= \frac{1}{\mu_{D+}^{(\nu)}}\mathcal{V}_{YW+,\nu,p}(h_n) - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}\mathcal{V}_{YD+,\nu,p}(h_n)$$

$$\mathbb{C}\left[\widehat{\mu}_{Y-,p}^{(\nu)}(h_n),\widetilde{\eta}_{-,\nu,p}(h_n)\,\middle|\,\boldsymbol{R}\right] = \frac{1}{\mu_{D-}^{(\nu)}}\mathbb{C}\left[\widehat{\mu}_{Y-,p}^{(\nu)}(h_n),\widehat{\mu}_{W-,p}^{(\nu)}(h_n)\,\middle|\,\boldsymbol{R}\right]$$
$$- \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^2}\mathbb{C}\left[\widehat{\mu}_{Y-,p}^{(\nu)}(h_n),\widehat{\mu}_{D-,p}^{(\nu)}(h_n)\,\middle|\,\boldsymbol{R}\right]$$
$$= \frac{1}{\mu_{D-}^{(\nu)}}\mathcal{V}_{YW-,\nu,p}(h_n) - \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^2}\mathcal{V}_{YD-,\nu,p}(h_n)$$

For compactness, define

$$\widetilde{V}_{Y+,\nu,p}(h_n) \equiv \mathbb{C}\left[\widehat{\mu}_{Y+,p}^{(\nu)}(h_n),\widetilde{\eta}_{+,\nu,p}(h_n)\,\middle|\,\boldsymbol{R}\right]$$
$$= \frac{1}{\mu_{D+}^{(\nu)}}\mathcal{V}_{YW+,\nu,p}(h_n) - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}\mathcal{V}_{YD+,\nu,p}(h_n)$$

$$\widetilde{V}_{Y-,\nu,p}(h_n) \equiv \mathbb{C}\left[\widehat{\mu}_{Y-,p}^{(\nu)}(h_n),\widetilde{\eta}_{-,\nu,p}(h_n)\,\middle|\,\boldsymbol{R}\right]$$
$$= \frac{1}{\mu_{D-}^{(\nu)}}\mathcal{V}_{YW-,\nu,p}(h_n) - \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^2}\mathcal{V}_{YD-,\nu,p}(h_n)$$

71

Then the conditional covariance of of $\widehat{\mu}_{Y+,p}^{(\nu)}(h_n) - \widehat{\mu}_{Y-,p}^{(\nu)}(h_n)$ and $\widetilde{\eta}_{\nu,p}(h_n)$ can be expressed as

$$\widetilde{V}_{Y,\nu,p}(h_n) \equiv \mathbb{C}\left[\widehat{\mu}_{Y+,p}^{(\nu)}(h_n) - \widehat{\mu}_{Y-,p}^{(\nu)}(h_n), \widetilde{\eta}_{\nu,p}(h_n)\Big|\boldsymbol{R}\right] = \widetilde{V}_{Y+,\nu,p}(h_n) + \widetilde{V}_{Y-,\nu,p}(h_n)$$

To conclude, the conditional variance of the linearized estimator $\widetilde{\theta}_{\nu,p}(h_n)$ is

$$V_{\nu,p}(h_n) \equiv \mathbb{V}\left[\widetilde{\theta}_{\nu,p}(h_n)\Big|\boldsymbol{R}\right] = V_{Y,\nu,p}(h_n) + \widetilde{V}_{\nu,p}(h_n) + 2\widetilde{V}_{Y,\nu,p}(h_n)$$

which completes the proof.

(D) Recall that, for any random variable A,

$$\widehat{\mu}_{A+,p}^{(\nu)}(h_n) = \nu! e_\nu' \widehat{\beta}_{A+,p}(h_n) \qquad \widehat{\mu}_{A-,p}^{(\nu)}(h_n) = \nu! e_\nu' \widehat{\beta}_{A-,p}(h_n)$$

with

$$\widehat{\beta}_{A+,p}(h_n) = H_p(h_n)\Gamma_{+,p}^{-1}(h_n)\boldsymbol{X_p}(\boldsymbol{h_n})'\boldsymbol{Z_+}(\boldsymbol{h_n})\boldsymbol{A}/n$$
$$\widehat{\beta}_{A-,p}(h_n) = H_p(h_n)\Gamma_{-,p}^{-1}(h_n)\boldsymbol{X_p}(\boldsymbol{h_n})'\boldsymbol{Z_-}(\boldsymbol{h_n})\boldsymbol{A}/n$$

Then, define

$$\widetilde{\beta}_{+,p}(h_n) \equiv \widehat{\beta}_{Y+,p}(h_n) + \frac{1}{\mu_{D+}^{(\nu)}}\widehat{\beta}_{W+,p}(h_n) - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}\widehat{\beta}_{D+,p}(h_n)$$

$$= H_p(h_n)\Gamma_{+,p}^{-1}(h_n)\boldsymbol{X_p}(\boldsymbol{h_n})'\boldsymbol{Z_+}(\boldsymbol{h_n})\left(\boldsymbol{Y} + \frac{1}{\mu_{D+}^{(\nu)}}\boldsymbol{W} - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}\boldsymbol{D}\right)/n$$

$$\widetilde{\beta}_{-,p}(h_n) \equiv \widehat{\beta}_{Y-,p}(h_n) + \frac{1}{\mu_{D-}^{(\nu)}}\widehat{\beta}_{W-,p}(h_n) - \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^2}\widehat{\beta}_{D-,p}(h_n)$$

$$= H_p(h_n)\Gamma_{-,p}^{-1}(h_n)\boldsymbol{X_p}(\boldsymbol{h_n})'\boldsymbol{Z_-}(\boldsymbol{h_n})\left(\boldsymbol{Y} + \frac{1}{\mu_{D-}^{(\nu)}}\boldsymbol{W} - \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^2}\boldsymbol{D}\right)/n$$

Further define

$$\widetilde{\mu}_{+,p}^{(\nu)}(h_n) \equiv \nu! e_\nu' \widetilde{\beta}_{+,p}(h_n) \qquad \widetilde{\mu}_{-,p}^{(\nu)}(h_n) \equiv \nu! e_\nu' \widetilde{\beta}_{-,p}(h_n)$$

In words, the linearized estimator of interest $\widetilde{\theta}_{\nu,p}(h_n)$ is numerically equivalent to $\widetilde{\mu}_{+,p}^{(\nu)}(h_n) - \widetilde{\mu}_{-,p}^{(\nu)}(h_n)$, the sharp regression discontinuity estimator that uses $Y + \frac{1}{\mu_D^{(\nu)}}W - $

$\frac{\mu_W^{(\nu)}}{\left(\mu_D^{(\nu)}\right)^2} D$ as the outcome. Thus, by part (D) of Lemma A.1 in Calonico, Cattaneo and Titiunik (2014*b*), if $nh_n^{2p+5} \to 0$,

$$\frac{\widetilde{\theta}_{\nu,p}(h_n) - \left(\mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)}\right) - h_n^{p+1-\nu} \mathrm{B}_{\nu,p,p+1}(h_n)}{\sqrt{\mathrm{V}_{\nu,p}(h_n)}} \xrightarrow{d} \mathcal{N}(0,1)$$

The bias-corrected test statistic for the non-linearized estimator $\widehat{\theta}_{\nu,p}(h_n)$ is

$$\begin{aligned}
T^{\mathrm{BC}}(h_n) &\equiv \frac{\widehat{\theta}_{\nu,p}(h_n) - \theta_\nu - h_n^{p+1-\nu} \mathrm{B}_{\nu,p,p+1}(h_n)}{\sqrt{\mathrm{V}_{\nu,p}(h_n)}} \\
&= \frac{\widetilde{\theta}_{\nu,p}(h_n) - \left(\mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)}\right) - h_n^{p+1-\nu} \mathrm{B}_{\nu,p,p+1}(h_n)}{\sqrt{\mathrm{V}_{\nu,p}(h_n)}} + \frac{R_n}{\sqrt{\mathrm{V}_{\nu,p}(h_n)}}
\end{aligned}$$

Recall that part (R) of this theorem states that

$$R_n = O_p\left(h_n^{2(p+1-\nu)} + \frac{1}{nh_n^{1+2\nu}}\right)$$

Leveraging part (V) of this theorem,

$$\begin{aligned}
\frac{R_n}{\sqrt{\mathrm{V}_{\nu,p}(h_n)}} &= O_p\left(\sqrt{nh_n^{1+2\nu}} h_n^{2(p+1-\nu)} + \frac{\sqrt{nh_n^{1+2\nu}}}{nh_n^{1+2\nu}}\right) \\
&= O_p\left(\sqrt{nh_n^{1+2\nu} h_n^{4(p+1-\nu)}} + \frac{1}{\sqrt{nh_n^{1+2\nu}}}\right) \\
&= O_p\left(\sqrt{nh_n^{2p+5} h_n^{2(p-\nu)}} + \frac{1}{\sqrt{nh_n^{1+2\nu}}}\right) \\
&= o_p(1)
\end{aligned}$$

because $p \geq \nu$ and, by assumption, $nh_n^{1+2\nu} \to \infty$ and $nh_n^{2p+5} \to 0$. Thus,

$$T^{\mathrm{BC}}(h_n) \equiv \frac{\widehat{\theta}_{\nu,p}(h_n) - \theta_\nu - h_n^{p+1-\nu} \mathrm{B}_{\nu,p,p+1}(h_n)}{\sqrt{\mathrm{V}_{\nu,p}(h_n)}} \xrightarrow{d} \mathcal{N}(0,1)$$

which completes the proof.

B.3   Robust Bias-Corrected Confidence Intervals

In this section, I derive the asymptotic distribution of the robust bias-corrected estimator for the target parameter θ_ν. This estimator is defined as the difference between $\widehat{\theta}_{\nu,p}(h_n)$ and its estimated (rather than true) bias. Before stating the main proposition, it is convenient to define a few additional bias- and variance-related terms.

B.3.1 Definitions of Bias Terms

To begin with, recall that the bias estimand of interest is

$$\mathrm{B}_{\nu,p,q}\left(h_n\right) \equiv \mathrm{B}_{Y,\nu,p,q}\left(h_n\right) + \widetilde{\mathrm{B}}_{\nu,p,q}\left(h_n\right) \tag{B.61}$$

where

$$\mathrm{B}_{Y,\nu,p,q}\left(h_n\right) \equiv \mathrm{B}_{Y+,\nu,p,q}\left(h_n\right) - \mathrm{B}_{Y-,\nu,p,q}\left(h_n\right) \tag{B.62}$$

$$\widetilde{\mathrm{B}}_{\nu,p,q}\left(h_n\right) \equiv \widetilde{\mathrm{B}}_{+,\nu,p,q}\left(h_n\right) - \widetilde{\mathrm{B}}_{-,\nu,p,q}\left(h_n\right) \tag{B.63}$$

with

$$\widetilde{\mathrm{B}}_{+,\nu,p,q}\left(h_n\right) \equiv \frac{1}{\mu_{D+}^{(\nu)}}\mathrm{B}_{W+,\nu,p,q}\left(h_n\right) - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}\mathrm{B}_{D+,\nu,p,q}\left(h_n\right) \tag{B.64}$$

$$\widetilde{\mathrm{B}}_{-,\nu,p,q}\left(h_n\right) \equiv \frac{1}{\mu_{D-}^{(\nu)}}\mathrm{B}_{W-,\nu,p,q}\left(h_n\right) - \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^2}\mathrm{B}_{D-,\nu,p,q}\left(h_n\right) \tag{B.65}$$

Also recall that, for any random variable A,

$$\mathrm{B}_{A+,\nu,p,q}\left(h_n\right) \equiv \frac{\mu_{A+}^{(q)}}{q!}\mathcal{B}_{-,\nu,p,q}\left(h_n\right) \qquad \mathrm{B}_{A-,\nu,p,q}\left(h_n\right) \equiv \frac{\mu_{A-}^{(q)}}{q!}\mathcal{B}_{-,\nu,p,q}\left(h_n\right) \tag{B.66}$$

where

$$\mathcal{B}_{+,\nu,p,q}\left(h_n\right) \equiv \nu! e_\nu' \Gamma_{+,p}^{-1}\left(h_n\right) \vartheta_{+,p,q}\left(h_n\right) = \nu! e_\nu' \Gamma_p^{-1}\vartheta_{p,q} + o_p\left(1\right) \tag{B.67}$$

$$\mathcal{B}_{-,\nu,p,q}\left(h_n\right) \equiv \nu! e_\nu' \Gamma_{-,p}^{-1}\left(h_n\right) \vartheta_{-,p,q}\left(h_n\right) = \left(-1\right)^{\nu+q}\nu! e_\nu' \Gamma_p^{-1}\vartheta_{p,q} + o_p\left(1\right) \tag{B.68}$$

with the second equalities following from applications of Lemma S.A.1 in Calonico, Cattaneo and Titiunik (2014b). Finally, for any random variable A, define the following recurring bias-corrected bias terms:

$$\mathrm{B}_{A+,\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right) \equiv \frac{\mu_{A+}^{(q+1)}}{(q+1)!}\mathcal{B}_{+,p+1,q,q+1}\left(b_n\right)\frac{\mathcal{B}_{+,\nu,p,p+1}\left(h_n\right)}{(p+1)!} \tag{B.69}$$

$$\mathrm{B}_{A-,\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right) \equiv \frac{\mu_{A-}^{(q+1)}}{(q+1)!}\mathcal{B}_{-,p+1,q,q+1}\left(b_n\right)\frac{\mathcal{B}_{-,\nu,p,p+1}\left(h_n\right)}{(p+1)!} \tag{B.70}$$

Natural estimators for $\mathrm{B}_{A+,\nu,p,q}\left(h_n\right)$ and $\mathrm{B}_{A-,\nu,p,q}\left(h_n\right)$ are the following:

$$\widehat{\mathrm{B}}_{A+,\nu,p,q}\left(h_n, b_n\right) \equiv \frac{\widehat{\mu}_{A+,q}^{(p+1)}\left(b_n\right)}{(p+1)!}\mathcal{B}_{+,\nu,p,p+1}\left(h_n\right) \tag{B.71}$$

$$\widehat{\mathrm{B}}_{A-,\nu,p,q}\left(h_n, b_n\right) \equiv \frac{\widehat{\mu}_{A+,q}^{(p+1)}\left(b_n\right)}{\left(p+1\right)!} \mathcal{B}_{-,\nu,p,p+1}\left(h_n\right) \tag{B.72}$$

with h_n and b_n denoting the main and pilot bandwidths, respectively. Thus, an estimator for $\mathrm{B}_{Y,\nu,p,q}\left(h_n\right)$ is

$$\widehat{\mathrm{B}}_{Y,\nu,p,q}\left(h_n\right) \equiv \widehat{\mathrm{B}}_{Y+,\nu,p,q}\left(h_n\right) - \widehat{\mathrm{B}}_{Y-,\nu,p,q}\left(h_n\right) \tag{B.73}$$

In addition, define two distinct estimators for the bias of the linearized estimator $\widetilde{\eta}\left(h_n\right)$. First, consider

$$\breve{\widetilde{\mathrm{B}}}_{\nu,p,q}\left(h_n, b_n\right) \equiv \breve{\widetilde{\mathrm{B}}}_{+,\nu,p,q}\left(h_n, b_n\right) - \breve{\widetilde{\mathrm{B}}}_{-,\nu,p,q}\left(h_n, b_n\right) \tag{B.74}$$

with

$$\breve{\widetilde{\mathrm{B}}}_{+,\nu,p,q}\left(h_n, b_n\right) \equiv \frac{1}{\mu_{D+}^{(\nu)}} \widehat{\mathrm{B}}_{W+,\nu,p,q}\left(h_n, b_n\right) - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2} \widehat{\mathrm{B}}_{D+,\nu,p,q}\left(h_n, b_n\right) \tag{B.75}$$

$$\breve{\widetilde{\mathrm{B}}}_{-,\nu,p,q}\left(h_n, b_n\right) \equiv \frac{1}{\mu_{D-}^{(\nu)}} \widehat{\mathrm{B}}_{W-,\nu,p,q}\left(h_n, b_n\right) - \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^2} \widehat{\mathrm{B}}_{D-,\nu,p,q}\left(h_n, b_n\right) \tag{B.76}$$

Second, define an estimator that replaces population conditional means with their sample counterparts:

$$\widehat{\widetilde{\mathrm{B}}}_{\nu,p,q}\left(h_n, b_n\right) \equiv \widehat{\widetilde{\mathrm{B}}}_{+,\nu,p,q}\left(h_n, b_n\right) - \widehat{\widetilde{\mathrm{B}}}_{-,\nu,p,q}\left(h_n, b_n\right) \tag{B.77}$$

with

$$\widehat{\widetilde{\mathrm{B}}}_{+,\nu,p,q}\left(h_n, b_n\right) \equiv \frac{1}{\widehat{\mu}_{D+,p}^{(\nu)}\left(h_n\right)} \widehat{\mathrm{B}}_{W+,\nu,p,q}\left(h_n, b_n\right) - \frac{\widehat{\mu}_{W+,p}^{(\nu)}\left(h_n\right)}{\left(\widehat{\mu}_{D+,p}^{(\nu)}\left(h_n\right)\right)^2} \widehat{\mathrm{B}}_{D+,\nu,p,q}\left(h_n, b_n\right) \tag{B.78}$$

$$\widehat{\widetilde{\mathrm{B}}}_{-,\nu,p,q}\left(h_n, b_n\right) \equiv \frac{1}{\widehat{\mu}_{D-,p}^{(\nu)}\left(h_n\right)} \widehat{\mathrm{B}}_{W-,\nu,p,q}\left(h_n, b_n\right) - \frac{\widehat{\mu}_{W-,p}^{(\nu)}\left(h_n\right)}{\left(\widehat{\mu}_{D-,p}^{(\nu)}\left(h_n\right)\right)^2} \widehat{\mathrm{B}}_{D-,\nu,p,q}\left(h_n, b_n\right) \tag{B.79}$$

Thus, two candidate estimands for the bias of the linearized estimator $\widetilde{\theta}_{\nu,p}\left(h_n\right)$ are the following:

$$\breve{\mathrm{B}}_{\nu,p,q}\left(h_n, b_n\right) \equiv \widehat{\mathrm{B}}_{Y,\nu,p,q}\left(h_n, b_n\right) + \breve{\widetilde{\mathrm{B}}}_{\nu,p,q}\left(h_n, b_n\right) \tag{B.80}$$

$$\widehat{\mathrm{B}}_{\nu,p,q}\left(h_n, b_n\right) \equiv \widehat{\mathrm{B}}_{Y,\nu,p,q}\left(h_n, b_n\right) + \widehat{\widetilde{\mathrm{B}}}_{\nu,p,q}\left(h_n, b_n\right) \tag{B.81}$$

B.3.2 Definitions of Bias-Corrected Estimators

The bias-corrected estimator for the target parameter $\widehat{\theta}_{\nu,p}(h_n)$ is

$$\widehat{\theta}^{\mathrm{BC}}_{\nu,p,q}(h_n, b_n) \equiv \widehat{\theta}_{\nu,p}(h_n) - h_n^{p+1-\nu}\widehat{\mathrm{B}}_{\nu,p,q}(h_n, b_n) \tag{B.82}$$

Analogously, define the bias-corrected linearized estimator $\widetilde{\theta}_{\nu,p}(h_n)$ as

$$\widetilde{\theta}^{\mathrm{BC}}_{\nu,p,q}(h_n, b_n) \equiv \widetilde{\theta}_{\nu,p}(h_n) - h_n^{p+1-\nu}\check{\mathrm{B}}_{\nu,p,q}(h_n, b_n) \tag{B.83}$$

Recall that

$$R_n \equiv \widehat{\eta}_{\nu,p}(h_n) - \widetilde{\eta}_{\nu,p}(h_n) = \widehat{\theta}_{\nu,p}(h_n) - \widetilde{\theta}_{\nu,p}(h_n) \tag{B.84}$$

Similarly, the difference between bias-corrected estimators is

$$\begin{aligned}
\widehat{\theta}^{\mathrm{BC}}_{\nu,p,q}(h_n, b_n) - \check{\theta}^{\mathrm{BC}}_{\nu,p,q}(h_n, b_n) &= \widehat{\theta}_{\nu,p}(h_n) - \widetilde{\theta}_{\nu,p}(h_n) \\
&\quad - h_n^{p+1-\nu}\left(\widehat{\mathrm{B}}_{\nu,p,q}(h_n, b_n) - \check{\mathrm{B}}_{\nu,p,q}(h_n, b_n)\right) \tag{B.85} \\
&= R_n - h_n^{p+1-\nu}\left(\widehat{\widetilde{\mathrm{B}}}_{\nu,p,q}(h_n, b_n) - \check{\widetilde{\mathrm{B}}}_{\nu,p,q}(h_n, b_n)\right) \tag{B.86}
\end{aligned}$$

where the last equality follows from the fact that $\widehat{\mathrm{B}}_{Y,\nu,p,q}(h_n, b_n)$ enters additively in both definitions of estimated bias. Finally, define

$$R_n^{\mathrm{BC}} \equiv h_n^{p+1-\nu}\left(\widehat{\widetilde{\mathrm{B}}}_{\nu,p,q}(h_n, b_n) - \check{\widetilde{\mathrm{B}}}_{\nu,p,q}(h_n, b_n)\right) \tag{B.87}$$

B.3.3 Definitions of Variance Terms

In this section, I define general covariance terms involving estimators of conditional derivatives of random variables, one being estimated using the main bandwidth h_n and the other using the pilot bandwidth b_n. Recall that, for any pair of random variables A and B,

$$\mathbb{C}\left[\widehat{\mu}^{(\nu)}_{A+,p}(h_n), \widehat{\mu}^{(p+1)}_{B+,q}(b_n) \,\Big|\, \boldsymbol{R}\right] = \mathbb{C}\left[\nu! e'_\nu \widehat{\beta}_{A+,p}(h_n), (p+1)! e'_{p+1}\widehat{\beta}_{B+,q}(b_n) \,\Big|\, \boldsymbol{R}\right] \tag{B.88}$$

with

$$\widehat{\beta}_{A+,p}(h_n) = H_p(h_n)\,\Gamma^{-1}_{+,p}(h_n)\,\boldsymbol{X_p}(\boldsymbol{h_n})'\,\boldsymbol{Z}_+(\boldsymbol{h_n})\,\boldsymbol{A}/n \tag{B.89}$$

$$\widehat{\beta}_{B+,q}(b_n) = H_q(h_n)\,\Gamma^{-1}_{+,q}(b_n)\,\boldsymbol{X_q}(\boldsymbol{b_n})'\,\boldsymbol{Z}_+(\boldsymbol{b_n})\,\boldsymbol{B}/n \tag{B.90}$$

Then

$$
\mathcal{C}_{AB+,\nu,p,q}\left(h_n, b_n\right) \equiv \mathbb{C}\left[\widehat{\mu}_{A+,p}^{(\nu)}\left(h_n\right), \widehat{\mu}_{B+,q}^{(p+1)}\left(b_n\right) \Big| \boldsymbol{R}\right]
$$

$$
= n^{-2}\nu!\,(p+1)!e_\nu' H_p\left(h_n\right)\Gamma_{+,p}^{-1}\left(h_n\right)\boldsymbol{X_p}\left(\boldsymbol{h_n}\right)'\boldsymbol{Z_+}\left(\boldsymbol{h_n}\right)\boldsymbol{\Sigma_{AB}}
$$

$$
\times\,\boldsymbol{Z_+}\left(\boldsymbol{b_n}\right)\boldsymbol{X_q}\left(\boldsymbol{b_n}\right)\Gamma_{+,q}^{-1}\left(b_n\right)H_q\left(b_n\right)e_{p+1} \tag{B.91}
$$

$$
= n^{-2}\nu!\,(p+1)!h_n^{-\nu}b_n^{-(p+1)}e_\nu'\Gamma_{+,p}^{-1}\left(h_n\right)\boldsymbol{X_p}\left(\boldsymbol{h_n}\right)'\boldsymbol{Z_+}\left(\boldsymbol{h_n}\right)\boldsymbol{\Sigma_{AB}}
$$

$$
\times\,\boldsymbol{Z_+}\left(\boldsymbol{b_n}\right)\boldsymbol{X_q}\left(\boldsymbol{b_n}\right)\Gamma_{+,q}^{-1}\left(b_n\right)H_q\left(b_n\right)e_{p+1} \tag{B.92}
$$

$$
= n^{-1}\nu!\,(p+1)!h_n^{-\nu}b_n^{-(p+1)}e_\nu'\Gamma_{+,p}^{-1}\left(h_n\right)\Psi_{AB+,p,q}\left(h_n, b_n\right)\Gamma_{+,q}^{-1}\left(b_n\right)e_{p+1} \tag{B.93}
$$

$$
= \frac{\nu!\,(p+1)!}{nh_n^\nu b_n^{p+1}}e_\nu'\Gamma_{+,p}^{-1}\left(h_n\right)\Psi_{AB+,p,q}\left(h_n, b_n\right)\Gamma_{+,q}^{-1}\left(b_n\right)e_{p+1} \tag{B.94}
$$

$$
= \frac{1}{nh_n^\nu b_n^{p+1}}\frac{\sigma_{AB+}^2}{f_R\left(c\right)}\nu!\,(p+1)!e_\nu'\Gamma_p^{-1}\Psi_p\Gamma_q^{-1}e_{p+1}\left\{1+o_p\left(1\right)\right\} \tag{B.95}
$$

where the last equality follows from an application of Lemma S.A.1 in Calonico, Cattaneo and Titiunik (2014b). Symmetrically,

$$
\mathcal{C}_{AB-,\nu,p,q}\left(h_n, b_n\right) \equiv \mathbb{C}\left[\widehat{\mu}_{A-,p}^{(\nu)}\left(h_n\right), \widehat{\mu}_{B-,q}^{(p+1)}\left(b_n\right) \Big| \boldsymbol{R}\right]
$$

$$
= \frac{\nu!\,(p+1)!}{nh_n^\nu b_n^{p+1}}e_\nu'\Gamma_{-,p}^{-1}\left(h_n\right)\Psi_{AB-,p,q}\left(h_n, b_n\right)\Gamma_{-,q}^{-1}\left(b_n\right)e_{p+1} \tag{B.96}
$$

$$
= \frac{1}{nh_n^\nu b_n^{p+1}}\frac{\sigma_{AB-}^2}{f_R\left(c\right)}\nu!\,(p+1)!e_\nu'\Gamma_p^{-1}\Psi_p\Gamma_q^{-1}e_{p+1}\left\{1+o_p\left(1\right)\right\} \tag{B.97}
$$

B.3.4 General Theorem: Statement

Theorem B.2 *Suppose Assumptions 8 and 9 hold with $\delta \geq p+2$, and $n\min\{h_n, b_n\} \to \infty$.*

(R) If $h_n \to 0$ and $nh_n^{1+2\nu} \to \infty$, and provided that $\kappa b_n < \kappa_0$, then

$$
R_n^{\mathrm{BC}} = O_p\left(h_n^{2(p+1-\nu)} + \frac{h_n^{p+1-\nu}}{\sqrt{nh_n^{1+2\nu}}}\right)O_p\left(1+\sqrt{\frac{1}{nb_n^{2p+3}}}\right)
$$

(B) If $\max\{h_n, b_n\} \to 0$, then

$$
\mathbb{E}\left[\widetilde{\theta}_{\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right) \Big| \boldsymbol{R}\right] = h_n^{p+2-\nu}\mathrm{B}_{\nu,p,p+2}\left(h_n\right)\left\{1+o_p\left(1\right)\right\}
$$

$$
+ h_n^{p+1-\nu}b_n^{q-p}\mathrm{B}_{\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right)\left\{1+o_p\left(1\right)\right\}
$$

where

$$\mathrm{B}_{\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right) \equiv \mathrm{B}_{Y,\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right) + \widetilde{\mathrm{B}}_{\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right)$$

with

$$\mathrm{B}_{Y,\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right) \equiv \mathrm{B}_{Y+,\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right) - \mathrm{B}_{Y-,\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right)$$

$$\widetilde{\mathrm{B}}_{\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right) \equiv \widetilde{\mathrm{B}}_{+,\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right) - \widetilde{\mathrm{B}}_{-,\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right)$$

and

$$\widetilde{\mathrm{B}}_{+,\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right) \equiv \frac{1}{\mu_{D+}^{(\nu)}}\mathrm{B}_{W+,\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right) - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}\mathrm{B}_{D+,\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right)$$

$$\widetilde{\mathrm{B}}_{-,\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right) \equiv \frac{1}{\mu_{D-}^{(\nu)}}\mathrm{B}_{W-,\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right) - \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^2}\mathrm{B}_{D-,\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right)$$

(V) If $\max\{h_n, b_n\} \to 0$, *then*

$$\mathrm{V}_{\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right) \equiv \mathbb{V}\left[\widetilde{\theta}_{\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right)\bigg|\boldsymbol{R}\right] = \mathrm{V}_{+,\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right) + \mathrm{V}_{-,\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right)$$

where

$$\mathrm{V}_{+,\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right) \equiv \mathrm{V}_{+,\nu,p}\left(h_n\right) + h_n^{2(p+1-\nu)}\mathrm{V}_{+,p+1,q}\left(b_n\right)\frac{\mathcal{B}_{+,\nu,p,p+1}^2\left(h_n\right)}{(p+1)!^2}$$

$$- 2h_n^{p+1-\nu}\mathrm{C}_{+,\nu,p,q}\left(h_n, b_n\right)\frac{\mathcal{B}_{+,\nu,p,p+1}\left(h_n\right)}{(p+1)!}$$

$$\mathrm{V}_{-,\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right) \equiv \mathrm{V}_{-,\nu,p}\left(h_n\right) + h_n^{2(p+1-\nu)}\mathrm{V}_{-,p+1,q}\left(b_n\right)\frac{\mathcal{B}_{-,\nu,p,p+1}^2\left(h_n\right)}{(p+1)!^2}$$

$$- 2h_n^{p+1-\nu}\mathrm{C}_{-,\nu,p,q}\left(h_n, b_n\right)\frac{\mathcal{B}_{-,\nu,p,p+1}\left(h_n\right)}{(p+1)!}$$

with

$$\mathrm{V}_{+,u,v}\left(x_n\right) \equiv \mathcal{V}_{YY+,u,v}\left(x_n\right) + \widetilde{\mathrm{V}}_{+,u,v}\left(x_n\right) + 2\widetilde{\mathrm{V}}_{Y+,u,v}\left(x_n\right)$$

$$\mathrm{V}_{-,u,v}\left(x_n\right) \equiv \mathcal{V}_{YY-,u,v}\left(x_n\right) + \widetilde{\mathrm{V}}_{-,u,v}\left(x_n\right) + 2\widetilde{\mathrm{V}}_{Y-,u,v}\left(x_n\right)$$

$$\mathrm{C}_{+,\nu,p,q}\left(h_n, b_n\right) \equiv \widetilde{\mathrm{C}}_{Y+,\nu,p,q}\left(h_n, b_n\right) + \widetilde{\mathrm{C}}_{+,\nu,p,q}\left(h_n, b_n\right)$$

$$\mathrm{C}_{-,\nu,p,q}\left(h_n, b_n\right) \equiv \widetilde{\mathrm{C}}_{Y-,\nu,p,q}\left(h_n, b_n\right) + \widetilde{\mathrm{C}}_{-,\nu,p,q}\left(h_n, b_n\right)$$

78

and

$$\widetilde{C}_{Y+,\nu,p,q}\left(h_n, b_n\right) \equiv \mathcal{C}_{YY+,\nu,p,q}\left(h_n, b_n\right)$$

$$+ \frac{1}{\mu_{D+}^{(\nu)}}\mathcal{C}_{YW+,\nu,p,q}\left(h_n, b_n\right) - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}\mathcal{C}_{YD+,\nu,p,q}\left(h_n, b_n\right)$$

$$\widetilde{C}_{Y-,\nu,p,q}\left(h_n, b_n\right) \equiv \mathcal{C}_{YY-,\nu,p,q}\left(h_n, b_n\right)$$

$$+ \frac{1}{\mu_{D-}^{(\nu)}}\mathcal{C}_{YW-,\nu,p,q}\left(h_n, b_n\right) - \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^2}\mathcal{C}_{YD-,\nu,p,q}\left(h_n, b_n\right)$$

and

$$\widetilde{C}_{+,\nu,p,q}\left(h_n, b_n\right) \equiv \frac{1}{\mu_{D+}^{(\nu)}}\mathcal{C}_{WY+,\nu,p,q}\left(h_n, b_n\right) - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}\mathcal{C}_{DY+,\nu,p,q}\left(h_n, b_n\right)$$

$$+ \frac{1}{\left(\mu_{D+}^{(\nu)}\right)^2}\mathcal{C}_{WW+,\nu,p,q}\left(h_n, b_n\right) - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^3}\mathcal{C}_{DW+,\nu,p,q}\left(h_n, b_n\right)$$

$$- \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^3}\mathcal{C}_{WD+,\nu,p,q}\left(h_n, b_n\right) + \frac{\left(\mu_{W+}^{(\nu)}\right)^2}{\left(\mu_{D+}^{(\nu)}\right)^4}\mathcal{C}_{DD+,\nu,p,q}\left(h_n, b_n\right)$$

$$\widetilde{C}_{-,\nu,p,q}\left(h_n, b_n\right) \equiv \frac{1}{\mu_{D-}^{(\nu)}}\mathcal{C}_{WY-,\nu,p,q}\left(h_n, b_n\right) - \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^2}\mathcal{C}_{DY-,\nu,p,q}\left(h_n, b_n\right)$$

$$+ \frac{1}{\left(\mu_{D-}^{(\nu)}\right)^2}\mathcal{C}_{WW-,\nu,p,q}\left(h_n, b_n\right) - \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^3}\mathcal{C}_{DW-,\nu,p,q}\left(h_n, b_n\right)$$

$$- \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^3}\mathcal{C}_{WD-,\nu,p,q}\left(h_n, b_n\right) + \frac{\left(\mu_{W-}^{(\nu)}\right)^2}{\left(\mu_{D-}^{(\nu)}\right)^4}\mathcal{C}_{DD-,\nu,p,q}\left(h_n, b_n\right)$$

(D) If $n \min\left\{h_n^{2p+3}, b_n^{2p+3}\right\} \max\left\{h_n^2, b_n^{2(q-p)}\right\} \to 0$ and $\kappa \max\left\{h_n, b_n\right\} < \kappa_0$, then

$$T^{\mathrm{RBC}}\left(h_n, b_n\right) \equiv \frac{\widehat{\theta}_{\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right) - \theta_\nu}{\sqrt{\mathrm{V}_{\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right)}} \xrightarrow{d} \mathcal{N}\left(0, 1\right)$$

B.3.5   General Theorem: Proof

(R) Recall that

$$R_n^{\mathrm{BC}} \equiv h_n^{p+1-\nu}\left(\widehat{\widetilde{\mathrm{B}}}_{\nu,p,q}\left(h_n, b_n\right) - \widecheck{\widetilde{\mathrm{B}}}_{\nu,p,q}\left(h_n, b_n\right)\right)$$

$$= h_n^{p+1-\nu} \left(\widehat{\widehat{\mathrm{B}}}_{+,\nu,p,q} \left(h_n, b_n \right) - \widecheck{\widehat{\mathrm{B}}}_{+,\nu,p,q} \left(h_n, b_n \right) \right)$$

$$- h_n^{p+1-\nu} \left(\widehat{\widehat{\mathrm{B}}}_{-,\nu,p,q} \left(h_n, b_n \right) - \widecheck{\widehat{\mathrm{B}}}_{-,\nu,p,q} \left(h_n, b_n \right) \right)$$

with

$$\widehat{\widehat{\mathrm{B}}}_{+,\nu,p,q} \left(h_n, b_n \right) \equiv \frac{1}{\widehat{\mu}_{D+,p}^{(\nu)} \left(h_n \right)} \widehat{\mathrm{B}}_{W+,\nu,p,q} \left(h_n, b_n \right) - \frac{\widehat{\mu}_{W+,p}^{(\nu)} \left(h_n \right)}{\left(\widehat{\mu}_{D+,p}^{(\nu)} \left(h_n \right) \right)^2} \widehat{\mathrm{B}}_{D+,\nu,p,q} \left(h_n, b_n \right)$$

$$\widehat{\widehat{\mathrm{B}}}_{-,\nu,p,q} \left(h_n, b_n \right) \equiv \frac{1}{\widehat{\mu}_{D-,p}^{(\nu)} \left(h_n \right)} \widehat{\mathrm{B}}_{W-,\nu,p,q} \left(h_n, b_n \right) - \frac{\widehat{\mu}_{W-,p}^{(\nu)} \left(h_n \right)}{\left(\widehat{\mu}_{D-,p}^{(\nu)} \left(h_n \right) \right)^2} \widehat{\mathrm{B}}_{D-,\nu,p,q} \left(h_n, b_n \right)$$

$$\widecheck{\widehat{\mathrm{B}}}_{+,\nu,p,q} \left(h_n, b_n \right) \equiv \frac{1}{\mu_{D+}^{(\nu)}} \widehat{\mathrm{B}}_{W+,\nu,p,q} \left(h_n, b_n \right) - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)} \right)^2} \widehat{\mathrm{B}}_{D+,\nu,p,q} \left(h_n, b_n \right)$$

$$\widecheck{\widehat{\mathrm{B}}}_{-,\nu,p,q} \left(h_n, b_n \right) \equiv \frac{1}{\mu_{D-}^{(\nu)}} \widehat{\mathrm{B}}_{W-,\nu,p,q} \left(h_n, b_n \right) - \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)} \right)^2} \widehat{\mathrm{B}}_{D-,\nu,p,q} \left(h_n, b_n \right)$$

Thus,

$$R_n^{\mathrm{BC}} = h_n^{p+1-\nu} \left(\frac{1}{\widehat{\mu}_{D+,p}^{(\nu)} \left(h_n \right)} - \frac{1}{\mu_{D+}^{(\nu)}} \right) \widehat{\mathrm{B}}_{W+,\nu,p,q} \left(h_n, b_n \right)$$

$$- h_n^{p+1-\nu} \left(\frac{\widehat{\mu}_{W+,p}^{(\nu)} \left(h_n \right)}{\left(\widehat{\mu}_{D+,p}^{(\nu)} \left(h_n \right) \right)^2} - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)} \right)^2} \right) \widehat{\mathrm{B}}_{D+,\nu,p,q} \left(h_n, b_n \right)$$

$$- h_n^{p+1-\nu} \left(\frac{1}{\widehat{\mu}_{D-,p}^{(\nu)} \left(h_n \right)} - \frac{1}{\mu_{D-}^{(\nu)}} \right) \widehat{\mathrm{B}}_{W-,\nu,p,q} \left(h_n, b_n \right)$$

$$+ h_n^{p+1-\nu} \left(\frac{\widehat{\mu}_{W-,p}^{(\nu)} \left(h_n \right)}{\left(\widehat{\mu}_{D-,p}^{(\nu)} \left(h_n \right) \right)^2} - \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)} \right)^2} \right) \widehat{\mathrm{B}}_{D-,\nu,p,q} \left(h_n, b_n \right)$$

Expanding further,

$$R_n^{\mathrm{BC}} = h_n^{p+1-\nu} \left(\frac{1}{\widehat{\mu}_{D+,p}^{(\nu)} \left(h_n \right)} - \frac{1}{\mu_{D+}^{(\nu)}} \right) e_{p+1}' \widehat{\beta}_{W+,q} \left(b_n \right) \mathcal{B}_{+,\nu,p,p+1} \left(h_n \right)$$

$$- h_n^{p+1-\nu} \left(\frac{\widehat{\mu}_{W+,p}^{(\nu)} \left(h_n \right)}{\left(\widehat{\mu}_{D+,p}^{(\nu)} \left(h_n \right) \right)^2} - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)} \right)^2} \right) e_{p+1}' \widehat{\beta}_{D+,q} \left(b_n \right) \mathcal{B}_{+,\nu,p,p+1} \left(h_n \right)$$

$$- h_n^{p+1-\nu} \left(\frac{1}{\widehat{\mu}_{D-,p}^{(\nu)} \left(h_n \right)} - \frac{1}{\mu_{D-}^{(\nu)}} \right) e_{p+1}' \widehat{\beta}_{W-,q} \left(b_n \right) \mathcal{B}_{-,\nu,p,p+1} \left(h_n \right)$$

$$+ h_n^{p+1-\nu} \left(\frac{\widehat{\mu}_{W-,p}^{(\nu)}(h_n)}{\left(\widehat{\mu}_{D-,p}^{(\nu)}(h_n) \right)^2} - \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)} \right)^2} \right) e_{p+1}' \widehat{\beta}_{D-,q}(b_n) \mathcal{B}_{-,\nu,p,p+1}(h_n)$$

Recall that, by Lemma S.A.1 in Calonico, Cattaneo and Titiunik (2014b),

$$\mathcal{B}_{+,\nu,p,q}(h_n) = \nu! e_\nu' \Gamma_p^{-1} \vartheta_{p,q} + o_p(1) \qquad \mathcal{B}_{-,\nu,p,q}(h_n) = (-1)^{\nu+q} \nu! e_\nu' \Gamma_p^{-1} \vartheta_{p,q} + o_p(1)$$

In addition, by a similar set of arguments as those in the proof of part (R) of Theorem B.1,

$$e_{p+1}' \widehat{\beta}_{W+,q}(b_n) = \frac{\widehat{\mu}_{W+,q}^{(p+1)}(b_n)}{(p+1)!} = O_p \left(b_n^{p+1-(p+1)} + \sqrt{\frac{1}{nb_n^{1+2(p+1)}}} \right) = O_p \left(1 + \sqrt{\frac{1}{nb_n^{2p+3}}} \right)$$

$$e_{p+1}' \widehat{\beta}_{D+,q}(b_n) = \frac{\widehat{\mu}_{D+,q}^{(p+1)}(b_n)}{(p+1)!} = O_p \left(b_n^{p+1-(p+1)} + \sqrt{\frac{1}{nb_n^{1+2(p+1)}}} \right) = O_p \left(1 + \sqrt{\frac{1}{nb_n^{2p+3}}} \right)$$

$$e_{p+1}' \widehat{\beta}_{W-,q}(b_n) = \frac{\widehat{\mu}_{W-,q}^{(p+1)}(b_n)}{(p+1)!} = O_p \left(b_n^{p+1-(p+1)} + \sqrt{\frac{1}{nb_n^{1+2(p+1)}}} \right) = O_p \left(1 + \sqrt{\frac{1}{nb_n^{2p+3}}} \right)$$

$$e_{p+1}' \widehat{\beta}_{D-,q}(b_n) = \frac{\widehat{\mu}_{D-,q}^{(p+1)}(b_n)}{(p+1)!} = O_p \left(b_n^{p+1-(p+1)} + \sqrt{\frac{1}{nb_n^{1+2(p+1)}}} \right) = O_p \left(1 + \sqrt{\frac{1}{nb_n^{2p+3}}} \right)$$

Following again the arguments proposed in the proof of part (R) of Theorem B.1 and with repeated applications of the Continuous Mapping Theorem and the Delta Method,

$$\frac{1}{\widehat{\mu}_{D+,p}^{(\nu)}(h_n)} - \frac{1}{\mu_{D+}^{(\nu)}} = O_p \left(h_n^{p+1-\nu} + \sqrt{\frac{1}{nh_n^{1+2\nu}}} \right)$$

$$\frac{\widehat{\mu}_{W+,p}^{(\nu)}(h_n)}{\left(\widehat{\mu}_{D+,p}^{(\nu)}(h_n) \right)^2} - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)} \right)^2} = O_p \left(h_n^{p+1-\nu} + \sqrt{\frac{1}{nh_n^{1+2\nu}}} \right)$$

$$\frac{1}{\widehat{\mu}_{D-,p}^{(\nu)}(h_n)} - \frac{1}{\mu_{D-}^{(\nu)}} = O_p \left(h_n^{p+1-\nu} + \sqrt{\frac{1}{nh_n^{1+2\nu}}} \right)$$

$$\frac{\widehat{\mu}_{W-,p}^{(\nu)}(h_n)}{\left(\widehat{\mu}_{D-,p}^{(\nu)}(h_n) \right)^2} - \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)} \right)^2} = O_p \left(h_n^{p+1-\nu} + \sqrt{\frac{1}{nh_n^{1+2\nu}}} \right)$$

because the variance of any polynomial estimator for $\mu_{A+}^{(\nu)}$ and $\mu_{A-}^{(\nu)}$ is multiplicative in $\frac{1}{nh_n^{1+2\nu}}$, and this feature is preserved by the variance of continuously differentiable functions of these estimators. Thus, if $nh_n^{1+2\nu} \to \infty$,

$$R_n^{\mathrm{BC}} = h_n^{p+1-\nu} O_p \left(h_n^{p+1-\nu} + \sqrt{\frac{1}{nh_n^{1+2\nu}}} \right) O_p \left(1 + \sqrt{\frac{1}{nb_n^{2p+3}}} \right)$$

$$= O_p \left(h_n^{2(p+1-\nu)} + \frac{h_n^{p+1-\nu}}{\sqrt{nh_n^{1+2\nu}}} \right) O_p \left(1 + \sqrt{\frac{1}{nb_n^{2p+3}}} \right)$$

which completes the proof.

(B) The conditional expectation of the bias-corrected linearized estimator $\widetilde{\theta}_{\nu,p,q}^{\mathrm{BC}}(h_n, b_n)$ is

$$\mathbb{E}\left[\widetilde{\theta}_{\nu,p,q}^{\mathrm{BC}}(h_n, b_n) \,\Big|\, \boldsymbol{R} \right] = \mathbb{E}\left[\widetilde{\theta}_{\nu,p}(h_n) - h_n^{p+1-\nu}\check{\mathrm{B}}_{\nu,p,q}(h_n, b_n) \,|\, \boldsymbol{R} \right]$$
$$= \mathrm{B}_1 - h^{p+1-\nu}\mathrm{B}_2$$

First, define

$$\mathrm{B}_1 \equiv \mathbb{E}\left[\widetilde{\theta}_{\nu,p}(h_n) - h_n^{p+1-\nu}\mathrm{B}_{\nu,p,p+1}(h_n) \,\Big|\, \boldsymbol{R} \right]$$
$$= \mathbb{E}\left[\widetilde{\theta}_{\nu,p}(h_n) \,\Big|\, \boldsymbol{R} \right] - h_n^{p+1-\nu}\mathrm{B}_{\nu,p,p+1}(h_n)$$
$$= \mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)} + h_n^{p+1-\nu}\mathrm{B}_{\nu,p,p+1}(h_n) + h_n^{p+2-\nu}\mathrm{B}_{\nu,p,p+2}(h_n) + o_p\left(h_n^{p+2-\nu} \right)$$
$$- h_n^{p+1-\nu}\mathrm{B}_{\nu,p,p+1}(h_n)$$
$$= \mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)} + h_n^{p+2-\nu}\mathrm{B}_{\nu,p,p+2}(h_n) + o_p\left(h_n^{p+2-\nu} \right)$$
$$= \mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)} + h_n^{p+2-\nu}\mathrm{B}_{\nu,p,p+2}(h_n) + h_n^{p+2-\nu}o_p(1)$$
$$= \mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)} + h_n^{p+2-\nu}\mathrm{B}_{\nu,p,p+2}(h_n)\{1 + o_p(1)\}$$

where the third equality follows from part (B) of Theorem B.1. Second, define

$$\mathrm{B}_2 \equiv \mathbb{E}\left[\check{\mathrm{B}}_{\nu,p,q}(h_n, b_n) - \mathrm{B}_{\nu,p,p+1}(h_n) \,\Big|\, \boldsymbol{R} \right]$$
$$= \mathbb{E}\left[\check{\mathrm{B}}_{\nu,p,q}(h_n, b_n) \,\Big|\, \boldsymbol{R} \right] - \mathrm{B}_{\nu,p,p+1}(h_n)$$
$$= \mathbb{E}\left[\widehat{\mathrm{B}}_{Y,\nu,p,q}(h_n, b_n) + \check{\widetilde{\mathrm{B}}}_{\nu,p,q}(h_n, b_n) \,\Big|\, \boldsymbol{R} \right] - \left(\mathrm{B}_{Y,\nu,p,p+1}(h_n) + \widetilde{\mathrm{B}}_{\nu,p,p+1}(h_n) \right)$$
$$= \mathbb{E}\left[\widehat{\mathrm{B}}_{Y,\nu,p,q}(h_n, b_n) + \check{\widetilde{\mathrm{B}}}_{+,\nu,p,q}(h_n, b_n) - \check{\widetilde{\mathrm{B}}}_{-,\nu,p,q}(h_n, b_n) \,\Big|\, \boldsymbol{R} \right]$$
$$- \left(\mathrm{B}_{Y,\nu,p,p+1}(h_n) + \widetilde{\mathrm{B}}_{+,\nu,p,p+1}(h_n) - \widetilde{\mathrm{B}}_{-,\nu,p,p+1}(h_n) \right)$$

Recall that

$$\mathbb{E}\left[\widehat{\mathrm{B}}_{Y,\nu,p,q}(h_n, b_n) \,\Big|\, \boldsymbol{R} \right] = \mathbb{E}\left[e_{p+1}'\widehat{\beta}_{Y+,q}(b_n)\mathcal{B}_{+,\nu,p,p+1}(h_n) \right.$$
$$\left. - e_{p+1}'\widehat{\beta}_{Y-,q}(b_n)\mathcal{B}_{-,\nu,p,p+1}(h_n) \,\Big|\, \boldsymbol{R} \right]$$
$$\mathrm{B}_{Y,\nu,p,p+1}(h_n) \equiv e_{p+1}'\beta_{Y+,q}\mathcal{B}_{+,\nu,p,p+1}(h_n) - e_{p+1}'\beta_{Y-,q}\mathcal{B}_{-,\nu,p,p+1}(h_n)$$

which implies that

$$\mathbb{E}\left[\widehat{\mathrm{B}}_{Y,\nu,p,q}\left(h_n,b_n\right)\middle|\boldsymbol{R}\right]-\mathrm{B}_{Y,\nu,p,p+1}\left(h_n\right)$$

$$=e'_{p+1}\left(\mathbb{E}\left[\widehat{\beta}_{Y+,q}\left(b_n\right)\middle|\boldsymbol{R}\right]-\beta_{Y+,q}\right)\mathcal{B}_{+,\nu,p,p+1}\left(h_n\right)$$

$$-e'_{p+1}\left(\mathbb{E}\left[\widehat{\beta}_{Y-,q}\left(b_n\right)\middle|\boldsymbol{R}\right]-\beta_{Y-,q}\right)\mathcal{B}_{-,\nu,p,p+1}\left(h_n\right)$$

Similarly,

$$\mathbb{E}\left[\widecheck{\widetilde{\mathrm{B}}}_{+,\nu,p,q}\left(h_n,b_n\right)\middle|\boldsymbol{R}\right]=\mathbb{E}\left[\frac{1}{\mu_{D+}^{(\nu)}}\widehat{\mathrm{B}}_{W+,\nu,p,q}\left(h_n,b_n\right)-\frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}\widehat{\mathrm{B}}_{D+,\nu,p,q}\left(h_n,b_n\right)\middle|\boldsymbol{R}\right]$$

$$=\frac{1}{\mu_{D+}^{(\nu)}}\mathbb{E}\left[\widehat{\mathrm{B}}_{W+,\nu,p,q}\left(h_n,b_n\right)\middle|\boldsymbol{R}\right]$$

$$-\frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}\mathbb{E}\left[\widehat{\mathrm{B}}_{D+,\nu,p,q}\left(h_n,b_n\right)\middle|\boldsymbol{R}\right]$$

$$=\frac{1}{\mu_{D+}^{(\nu)}}\mathbb{E}\left[e'_{p+1}\widehat{\beta}_{W+,q}\left(b_n\right)\mathcal{B}_{+,\nu,p,p+1}\left(h_n\right)\middle|\boldsymbol{R}\right]$$

$$-\frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}\mathbb{E}\left[e'_{p+1}\widehat{\beta}_{D+,q}\left(b_n\right)\mathcal{B}_{+,\nu,p,p+1}\left(h_n\right)\middle|\boldsymbol{R}\right]$$

$$\widetilde{\mathrm{B}}_{+,\nu,p,p+1}\left(h_n\right)\equiv\frac{1}{\mu_{D+}^{(\nu)}}\mathrm{B}_{W+,\nu,p,p+1}\left(h_n\right)-\frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}\mathrm{B}_{D+,\nu,p,p+1}\left(h_n\right)$$

$$=\frac{1}{\mu_{D+}^{(\nu)}}e'_{p+1}\beta_{W+,q}\mathcal{B}_{+,\nu,p,p+1}\left(h_n\right)$$

$$-\frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}e'_{p+1}\beta_{D+,q}\mathcal{B}_{+,\nu,p,p+1}\left(h_n\right)$$

which implies that

$$\mathbb{E}\left[\widecheck{\widetilde{\mathrm{B}}}_{+,\nu,p,q}\left(h_n,b_n\right)\middle|\boldsymbol{R}\right]-\widetilde{\mathrm{B}}_{+,\nu,p,p+1}\left(h_n\right)$$

$$=\frac{1}{\mu_{D+}^{(\nu)}}e'_{p+1}\left(\mathbb{E}\left[\widehat{\beta}_{W+,q}\left(b_n\right)\middle|\boldsymbol{R}\right]-\beta_{W+,q}\right)\mathcal{B}_{+,\nu,p,p+1}\left(h_n\right)$$

$$-\frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}e'_{p+1}\left(\mathbb{E}\left[\widehat{\beta}_{D+,q}\left(b_n\right)\middle|\boldsymbol{R}\right]-\beta_{D+,q}\right)\mathcal{B}_{+,\nu,p,p+1}\left(h_n\right)$$

Symmetrically,

$$\mathbb{E}\left[\widecheck{\widetilde{\mathrm{B}}}_{-,\nu,p,q}\left(h_n,b_n\right)\middle|\boldsymbol{R}\right]-\widetilde{\mathrm{B}}_{-,\nu,p,p+1}\left(h_n\right)$$

$$= \frac{1}{\mu_{D-}^{(\nu)}} e'_{p+1} \left(\mathbb{E}\left[\widehat{\beta}_{W-,q}(b_n) \,\middle|\, \boldsymbol{R} \right] - \beta_{W-,q} \right) \mathcal{B}_{-,\nu,p,p+1}(h_n)$$

$$- \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)} \right)^2} e'_{p+1} \left(\mathbb{E}\left[\widehat{\beta}_{D-,q}(b_n) \,\middle|\, \boldsymbol{R} \right] - \beta_{D-,q} \right) \mathcal{B}_{-,\nu,p,p+1}(h_n)$$

Combining these derivations,

$$
\begin{aligned}
\mathrm{B}_2 &\equiv e'_{p+1} \left(\mathbb{E}\left[\widehat{\beta}_{Y+,q}(b_n) \,\middle|\, \boldsymbol{R} \right] - \beta_{Y+,q} \right) \mathcal{B}_{+,\nu,p,p+1}(h_n) \\
&\quad - e'_{p+1} \left(\mathbb{E}\left[\widehat{\beta}_{Y-,q}(b_n) \,\middle|\, \boldsymbol{R} \right] - \beta_{Y-,q} \right) \mathcal{B}_{-,\nu,p,p+1}(h_n) \\
&\quad + \frac{1}{\mu_{D+}^{(\nu)}} e'_{p+1} \left(\mathbb{E}\left[\widehat{\beta}_{W+,q}(b_n) \,\middle|\, \boldsymbol{R} \right] - \beta_{W+,q} \right) \mathcal{B}_{+,\nu,p,p+1}(h_n) \\
&\quad - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)} \right)^2} e'_{p+1} \left(\mathbb{E}\left[\widehat{\beta}_{D+,q}(b_n) \,\middle|\, \boldsymbol{R} \right] - \beta_{D+,q} \right) \mathcal{B}_{+,\nu,p,p+1}(h_n) \\
&\quad - \frac{1}{\mu_{D-}^{(\nu)}} e'_{p+1} \left(\mathbb{E}\left[\widehat{\beta}_{W-,q}(b_n) \,\middle|\, \boldsymbol{R} \right] - \beta_{W-,q} \right) \mathcal{B}_{-,\nu,p,p+1}(h_n) \\
&\quad + \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)} \right)^2} e'_{p+1} \left(\mathbb{E}\left[\widehat{\beta}_{D-,q}(b_n) \,\middle|\, \boldsymbol{R} \right] - \beta_{D-,q} \right) \mathcal{B}_{-,\nu,p,p+1}(h_n)
\end{aligned}
$$

Furthermore, by several applications of Lemma S.A.3 in Calonico, Cattaneo and Titiunik (2014*b*), for any random variable A,

$$\mathbb{E}\left[(p+1)! \, e'_{p+1} \widehat{\beta}_{A+,q}(b_n) \,\middle|\, \boldsymbol{R} \right] = e'_{p+1}\beta_{A+,q} + b_n^{q-p} \frac{\mu_{A+}^{(q+1)}}{(q+1)!} \mathcal{B}_{+,p+1,q,q+1}(b_n) + o_p\left(b_n^{q-p} \right)$$

$$\mathbb{E}\left[(p+1)! \, e'_{p+1} \widehat{\beta}_{A-,q}(b_n) \,\middle|\, \boldsymbol{R} \right] = e'_{p+1}\beta_{A-,q} + b_n^{q-p} \frac{\mu_{A-}^{(q+1)}}{(q+1)!} \mathcal{B}_{-,p+1,q,q+1}(b_n) + o_p\left(b_n^{q-p} \right)$$

Thus,

$$
\begin{aligned}
\mathrm{B}_2 &= \left(b_n^{q-p} \frac{\mu_{Y+}^{(q+1)}}{(q+1)!} \mathcal{B}_{+,p+1,q,q+1}(b_n) + o_p\left(b_n^{q-p} \right) \right) \frac{\mathcal{B}_{+,\nu,p,p+1}(h_n)}{(p+1)!} \\
&\quad - \left(b_n^{q-p} \frac{\mu_{Y-}^{(q+1)}}{(q+1)!} \mathcal{B}_{-,p+1,q,q+1}(b_n) + o_p\left(b_n^{q-p} \right) \right) \frac{\mathcal{B}_{-,\nu,p,p+1}(h_n)}{(p+1)!} \\
&\quad + \frac{1}{\mu_{D+}^{(\nu)}} \left(b_n^{q-p} \frac{\mu_{W+}^{(q+1)}}{(q+1)!} \mathcal{B}_{+,p+1,q,q+1}(b_n) + o_p\left(b_n^{q-p} \right) \right) \frac{\mathcal{B}_{+,\nu,p,p+1}(h_n)}{(p+1)!} \\
&\quad - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)} \right)^2} \left(b_n^{q-p} \frac{\mu_{D+}^{(q+1)}}{(q+1)!} \mathcal{B}_{+,p+1,q,q+1}(b_n) + o_p\left(b_n^{q-p} \right) \right) \frac{\mathcal{B}_{+,\nu,p,p+1}(h_n)}{(p+1)!}
\end{aligned}
$$

$$- \frac{1}{\mu_{D-}^{(\nu)}} \left(b_n^{q-p} \frac{\mu_{W-}^{(q+1)}}{(q+1)!} \mathcal{B}_{-,p+1,q,q+1} (b_n) + o_p \left(b_n^{q-p} \right) \right) \frac{\mathcal{B}_{-,\nu,p,p+1} (h_n)}{(p+1)!}$$

$$+ \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)} \right)^2} \left(b_n^{q-p} \frac{\mu_{D-}^{(q+1)}}{(q+1)!} \mathcal{B}_{-,p+1,q,q+1} (b_n) + o_p \left(b_n^{q-p} \right) \right) \frac{\mathcal{B}_{-,\nu,p,p+1} (h_n)}{(p+1)!}$$

Define the following bias terms:

$$\mathrm{B}_{Y+,\nu,p,q}^{\mathrm{BC}} (h_n, b_n) \equiv \frac{\mu_{Y+}^{(q+1)}}{(q+1)!} \mathcal{B}_{+,p+1,q,q+1} (b_n) \frac{\mathcal{B}_{+,\nu,p,p+1} (h_n)}{(p+1)!}$$

$$\mathrm{B}_{Y-,\nu,p,q}^{\mathrm{BC}} (h_n, b_n) \equiv \frac{\mu_{Y-}^{(q+1)}}{(q+1)!} \mathcal{B}_{-,p+1,q,q+1} (b_n) \frac{\mathcal{B}_{-,\nu,p,p+1} (h_n)}{(p+1)!}$$

$$\mathrm{B}_{W+,\nu,p,q}^{\mathrm{BC}} (h_n, b_n) \equiv \frac{\mu_{W+}^{(q+1)}}{(q+1)!} \mathcal{B}_{+,p+1,q,q+1} (b_n) \frac{\mathcal{B}_{+,\nu,p,p+1} (h_n)}{(p+1)!}$$

$$\mathrm{B}_{D+,\nu,p,q}^{\mathrm{BC}} (h_n, b_n) \equiv \frac{\mu_{D+}^{(q+1)}}{(q+1)!} \mathcal{B}_{+,p+1,q,q+1} (b_n) \frac{\mathcal{B}_{+,\nu,p,p+1} (h_n)}{(p+1)!}$$

$$\mathrm{B}_{W-,\nu,p,q}^{\mathrm{BC}} (h_n, b_n) \equiv \frac{\mu_{W-}^{(q+1)}}{(q+1)!} \mathcal{B}_{-,p+1,q,q+1} (b_n) \frac{\mathcal{B}_{-,\nu,p,p+1} (h_n)}{(p+1)!}$$

$$\mathrm{B}_{D-,\nu,p,q}^{\mathrm{BC}} (h_n, b_n) \equiv \frac{\mu_{D-}^{(q+1)}}{(q+1)!} \mathcal{B}_{-,p+1,q,q+1} (b_n) \frac{\mathcal{B}_{-,\nu,p,p+1} (h_n)}{(p+1)!}$$

Furthermore,

$$\mathrm{B}_{Y,\nu,p,q}^{\mathrm{BC}} (h_n, b_n) \equiv \mathrm{B}_{Y+,\nu,p,q}^{\mathrm{BC}} (h_n, b_n) - \mathrm{B}_{Y-,\nu,p,q}^{\mathrm{BC}} (h_n, b_n)$$

$$\widetilde{\mathrm{B}}_{+,\nu,p,q}^{\mathrm{BC}} (h_n, b_n) \equiv \frac{1}{\mu_{D+}^{(\nu)}} \mathrm{B}_{W+,\nu,p,q}^{\mathrm{BC}} (h_n, b_n) - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)} \right)^2} \mathrm{B}_{D+,\nu,p,q}^{\mathrm{BC}} (h_n, b_n)$$

$$\widetilde{\mathrm{B}}_{-,\nu,p,q}^{\mathrm{BC}} (h_n, b_n) \equiv \frac{1}{\mu_{D-}^{(\nu)}} \mathrm{B}_{W-,\nu,p,q}^{\mathrm{BC}} (h_n, b_n) - \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)} \right)^2} \mathrm{B}_{D-,\nu,p,q}^{\mathrm{BC}} (h_n, b_n)$$

And

$$\widetilde{\mathrm{B}}_{\nu,p,q}^{\mathrm{BC}} (h_n, b_n) \equiv \widetilde{\mathrm{B}}_{+,\nu,p,q}^{\mathrm{BC}} (h_n, b_n) - \widetilde{\mathrm{B}}_{-,\nu,p,q}^{\mathrm{BC}} (h_n, b_n)$$

Finally,

$$\mathrm{B}_{\nu,p,q}^{\mathrm{BC}} (h_n, b_n) \equiv \mathrm{B}_{Y,\nu,p,q}^{\mathrm{BC}} (h_n, b_n) + \widetilde{\mathrm{B}}_{\nu,p,q}^{\mathrm{BC}} (h_n, b_n)$$

As a consequence,

$$\mathrm{B}_2 = b_n^{q-p} \mathrm{B}_{\nu,p,q}^{\mathrm{BC}} (h_n, b_n) + o_p \left(b_n^{q-p} \right)$$

85

$$= b_n^{q-p} \mathrm{B}_{\nu,p,q}^{\mathrm{BC}} \left(h_n, b_n \right) + b_n^{q-p} o_p \left(1 \right)$$

$$= b_n^{q-p} \mathrm{B}_{\nu,p,q}^{\mathrm{BC}} \left(h_n, b_n \right) \left\{ 1 + o_p \left(1 \right) \right\}$$

Taken together, the conditional expectation of the linearized bias-corrected estimator $\widetilde{\theta}_{\nu,p}^{\mathrm{BC}} \left(h_n, b_n \right)$ is

$$\mathbb{E} \left[\widetilde{\theta}_{\nu,p}^{\mathrm{BC}} \left(h_n, b_n \right) \Big| \boldsymbol{R} \right] = \mathrm{B}_1 - h^{p+1-\nu} \mathrm{B}_2$$

$$= \mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)} + h_n^{p+2-\nu} \mathrm{B}_{\nu,p,p+2} \left(h_n \right) \left\{ 1 + o_p \left(1 \right) \right\}$$

$$- h^{p+1-\nu} b_n^{q-p} \mathrm{B}_{\nu,p,q}^{\mathrm{BC}} \left(h_n, b_n \right) \left\{ 1 + o_p \left(1 \right) \right\}$$

which completes the proof.

(V) The conditional variance of the linearized bias-corrected estimator $\widetilde{\theta}_{\nu,p}^{\mathrm{BC}} \left(h_n, b_n \right)$ is

$$\mathbb{V} \left[\widetilde{\theta}_{\nu,p}^{\mathrm{BC}} \left(h_n, b_n \right) \Big| \boldsymbol{R} \right]$$

$$= \mathbb{V} \left[\widehat{\mu}_{Y+,\nu,p}^{(\nu)} \left(h_n \right) + \widetilde{\eta}_{+,\nu,p} \left(h_n \right) - h_n^{p+1-\nu} \left(\widehat{\mathrm{B}}_{Y+,\nu,p,q} \left(h_n, b_n \right) + \breve{\widetilde{\mathrm{B}}}_{+,\nu,p,q} \left(h_n, b_n \right) \right) \Big| \boldsymbol{R} \right]$$

$$+ \mathbb{V} \left[\widehat{\mu}_{Y-,\nu,p}^{(\nu)} \left(h_n \right) + \widetilde{\eta}_{-,\nu,p} \left(h_n \right) - h_n^{p+1-\nu} \left(\widehat{\mathrm{B}}_{Y-,\nu,p,q} \left(h_n, b_n \right) + \breve{\widetilde{\mathrm{B}}}_{-,\nu,p,q} \left(h_n, b_n \right) \right) \Big| \boldsymbol{R} \right]$$

where the equality follows from the bilinearity of the covariance and the conditional independence of observations on either side of the cutoff c. For compactness, define

$$\mathrm{V}_{+,\nu,p,q}^{\mathrm{BC}} \left(h_n, b_n \right) \equiv \mathbb{V} \left[\widehat{\mu}_{Y+,\nu,p}^{(\nu)} \left(h_n \right) + \widetilde{\eta}_{+,\nu,p} \left(h_n \right) \right.$$

$$\left. - h_n^{p+1-\nu} \left(\widehat{\mathrm{B}}_{Y+,\nu,p,q} \left(h_n, b_n \right) + \breve{\widetilde{\mathrm{B}}}_{+,\nu,p,q} \left(h_n, b_n \right) \right) \Big| \boldsymbol{R} \right]$$

$$= \mathbb{V} \left[\widehat{\mu}_{Y+,\nu,p}^{(\nu)} \left(h_n \right) + \widetilde{\eta}_{+,\nu,p} \left(h_n \right) \Big| \boldsymbol{R} \right]$$

$$+ h_n^{2(p+1-\nu)} \mathbb{V} \left[\widehat{\mathrm{B}}_{Y+,\nu,p,q} \left(h_n, b_n \right) + \breve{\widetilde{\mathrm{B}}}_{+,\nu,p,q} \left(h_n, b_n \right) \Big| \boldsymbol{R} \right]$$

$$- 2 h_n^{p+1-\nu} \mathbb{C} \left[\widehat{\mu}_{Y+,\nu,p}^{(\nu)} \left(h_n \right), \widehat{\mathrm{B}}_{Y+,\nu,p,q} \left(h_n, b_n \right) + \breve{\widetilde{\mathrm{B}}}_{+,\nu,p,q} \left(h_n, b_n \right) \Big| \boldsymbol{R} \right]$$

$$- 2 h_n^{p+1-\nu} \mathbb{C} \left[\widetilde{\eta}_{+,\nu,p} \left(h_n \right), \widehat{\mathrm{B}}_{Y+,\nu,p,q} \left(h_n, b_n \right) + \breve{\widetilde{\mathrm{B}}}_{+,\nu,p,q} \left(h_n, b_n \right) \Big| \boldsymbol{R} \right]$$

Symmetrically,

$$\mathrm{V}_{-,\nu,p,q}^{\mathrm{BC}} \left(h_n, b_n \right) \equiv \mathbb{V} \left[\widehat{\mu}_{Y-,\nu,p}^{(\nu)} \left(h_n \right) + \widetilde{\eta}_{-,\nu,p} \left(h_n \right) \right.$$

$$- h_n^{p+1-\nu} \left(\widehat{\mathrm{B}}_{Y-,\nu,p,q}(h_n, b_n) + \widecheck{\widetilde{\mathrm{B}}}_{-,\nu,p,q}(h_n, b_n) \right) \Big| \boldsymbol{R} \Big]$$

$$= \mathbb{V}\left[\widehat{\mu}_{Y-,\nu,p}^{(\nu)}(h_n) + \widetilde{\eta}_{-,\nu,p}(h_n) \,\Big|\, \boldsymbol{R} \right]$$

$$+ h_n^{2(p+1-\nu)} \mathbb{V}\left[\widehat{\mathrm{B}}_{Y-,\nu,p,q}(h_n, b_n) + \widecheck{\widetilde{\mathrm{B}}}_{-,\nu,p,q}(h_n, b_n) \,\Big|\, \boldsymbol{R} \right]$$

$$- 2 h_n^{p+1-\nu} \mathbb{C}\left[\widehat{\mu}_{Y-,\nu,p}^{(\nu)}(h_n), \widehat{\mathrm{B}}_{Y-,\nu,p,q}(h_n, b_n) + \widecheck{\widetilde{\mathrm{B}}}_{-,\nu,p,q}(h_n, b_n) \,\Big|\, \boldsymbol{R} \right]$$

$$- 2 h_n^{p+1-\nu} \mathbb{C}\left[\widetilde{\eta}_{-,\nu,p}(h_n), \widehat{\mathrm{B}}_{Y-,\nu,p,q}(h_n, b_n) + \widecheck{\widetilde{\mathrm{B}}}_{-,\nu,p,q}(h_n, b_n) \,\Big|\, \boldsymbol{R} \right]$$

First, leveraging the bilinearity of the covariance and the definitions from part (V) of Theorem B.1,

$$\mathrm{V}_{+,\nu,p}(h_n) \equiv \mathbb{V}\left[\widehat{\mu}_{Y+,\nu,p}^{(\nu)}(h_n) + \widetilde{\eta}_{+,\nu,p}(h_n) \,\Big|\, \boldsymbol{R} \right]$$

$$= \mathbb{V}\left[\widehat{\mu}_{Y+,\nu,p}^{(\nu)}(h_n) \,\Big|\, \boldsymbol{R} \right] + \mathbb{V}\left[\widetilde{\eta}_{+,\nu,p}(h_n) \,\Big|\, \boldsymbol{R} \right]$$

$$+ 2 \times \mathbb{C}\left[\widehat{\mu}_{Y+,\nu,p}^{(\nu)}(h_n), \widetilde{\eta}_{+,\nu,p}(h_n) \,\Big|\, \boldsymbol{R} \right]$$

$$= \mathcal{V}_{YY+,\nu,p}(h_n) + \widetilde{\mathrm{V}}_{+,\nu,p}(h_n) + 2 \widetilde{\mathrm{V}}_{Y+,\nu,p}(h_n)$$

Symmetrically,

$$\mathrm{V}_{-,\nu,p}(h_n) \equiv \mathbb{V}\left[\widehat{\mu}_{Y-,\nu,p}^{(\nu)}(h_n) + \widetilde{\eta}_{-,\nu,p}(h_n) \,\Big|\, \boldsymbol{R} \right]$$

$$= \mathbb{V}\left[\widehat{\mu}_{Y-,\nu,p}^{(\nu)}(h_n) \,\Big|\, \boldsymbol{R} \right] + \mathbb{V}\left[\widetilde{\eta}_{-,\nu,p}(h_n) \,\Big|\, \boldsymbol{R} \right]$$

$$+ 2 \times \mathbb{C}\left[\widehat{\mu}_{Y-,\nu,p}^{(\nu)}(h_n), \widetilde{\eta}_{-,\nu,p}(h_n) \,\Big|\, \boldsymbol{R} \right]$$

$$= \mathcal{V}_{YY-,\nu,p}(h_n) + \widetilde{\mathrm{V}}_{-,\nu,p}(h_n) + 2 \widetilde{\mathrm{V}}_{Y-,\nu,p}(h_n)$$

Second,

$$\mathbb{V}\left[\widehat{\mathrm{B}}_{Y+,\nu,p,q}(h_n, b_n) + \widecheck{\widetilde{\mathrm{B}}}_{+,\nu,p,q}(h_n, b_n) \,\Big|\, \boldsymbol{R} \right]$$

$$= \mathbb{V}\left[\widehat{\mathrm{B}}_{Y+,\nu,p,q}(h_n, b_n) \,\Big|\, \boldsymbol{R} \right] + \mathbb{V}\left[\widecheck{\widetilde{\mathrm{B}}}_{+,\nu,p,q}(h_n, b_n) \,\Big|\, \boldsymbol{R} \right]$$

$$+ 2 \times \mathbb{C}\left[\widehat{\mathrm{B}}_{Y+,\nu,p,q}(h_n, b_n), \widecheck{\widetilde{\mathrm{B}}}_{+,\nu,p,q}(h_n, b_n) \,\Big|\, \boldsymbol{R} \right]$$

Expanding the variance terms,

$$\mathbb{V}\left[\widehat{\mathrm{B}}_{Y+,\nu,p,q}(h_n, b_n) \,\Big|\, \boldsymbol{R} \right] = \mathbb{V}\left[\widehat{\mu}_{Y+,q}^{(p+1)}(b_n) \,\Big|\, \boldsymbol{R} \right] \frac{\mathcal{B}_{+,\nu,p,p+1}^2(h_n)}{(p+1)!^2}$$

87

$$= \mathcal{V}_{YY+,p+1,q}\left(b_n\right) \frac{\mathcal{B}^2_{+,\nu,p,p+1}\left(h_n\right)}{(p+1)!^2}$$

$$\mathbb{V}\left[\breve{\widehat{\mathrm{B}}}_{+,\nu,p,q}\left(h_n,b_n\right)\Big|\boldsymbol{R}\right] = \frac{1}{\left(\mu^{(\nu)}_{D+}\right)^2}\mathbb{V}\left[\widehat{\mathrm{B}}_{W+,\nu,p,q}\left(h_n,b_n\right)\Big|\boldsymbol{R}\right]$$

$$+ \frac{\left(\mu^{(\nu)}_{W+}\right)^2}{\left(\mu^{(\nu)}_{D+}\right)^4}\mathbb{V}\left[\widehat{\mathrm{B}}_{D+,\nu,p,q}\left(h_n,b_n\right)\Big|\boldsymbol{R}\right]$$

$$- 2\frac{\mu^{(\nu)}_{W+}}{\left(\mu^{(\nu)}_{D+}\right)^3}\mathbb{C}\left[\widehat{\mathrm{B}}_{W+,\nu,p,q}\left(h_n,b_n\right),\widehat{\mathrm{B}}_{D+,\nu,p,q}\left(h_n,b_n\right)\Big|\boldsymbol{R}\right]$$

$$= \frac{1}{\left(\mu^{(\nu)}_{D+}\right)^2}\mathbb{V}\left[\widehat{\mu}^{(p+1)}_{W+,q}\left(b_n\right)\Big|\boldsymbol{R}\right]\frac{\mathcal{B}^2_{+,\nu,p,p+1}\left(h_n\right)}{(p+1)!^2}$$

$$+ \frac{\left(\mu^{(\nu)}_{W+}\right)^2}{\left(\mu^{(\nu)}_{D+}\right)^4}\mathbb{V}\left[\widehat{\mu}^{(p+1)}_{D+,q}\left(b_n\right)\Big|\boldsymbol{R}\right]\frac{\mathcal{B}^2_{+,\nu,p,p+1}\left(h_n\right)}{(p+1)!^2}$$

$$- 2\frac{\mu^{(\nu)}_{W+}}{\left(\mu^{(\nu)}_{D+}\right)^3}\mathbb{C}\left[\widehat{\mu}^{(p+1)}_{W+,q}\left(b_n\right),\widehat{\mu}^{(p+1)}_{D+,q}\left(b_n\right)\Big|\boldsymbol{R}\right]\frac{\mathcal{B}^2_{+,\nu,p,p+1}\left(h_n\right)}{(p+1)!^2}$$

$$= \frac{1}{\left(\mu^{(\nu)}_{D+}\right)^2}\mathcal{V}_{WW+,p+1,q}\left(b_n\right)\frac{\mathcal{B}^2_{+,\nu,p,p+1}\left(h_n\right)}{(p+1)!^2}$$

$$+ \frac{\left(\mu^{(\nu)}_{W+}\right)^2}{\left(\mu^{(\nu)}_{D+}\right)^4}\mathcal{V}_{DD+,p+1,q}\left(b_n\right)\frac{\mathcal{B}^2_{+,\nu,p,p+1}\left(h_n\right)}{(p+1)!^2}$$

$$- 2\frac{\mu^{(\nu)}_{W+}}{\left(\mu^{(\nu)}_{D+}\right)^3}\mathcal{V}_{WD+,p+1,q}\left(b_n\right)\frac{\mathcal{B}^2_{+,\nu,p,p+1}\left(h_n\right)}{(p+1)!^2}$$

$$= \widetilde{\mathrm{V}}_{+,p+1,q}\left(b_n\right)\frac{\mathcal{B}^2_{+,\nu,p,p+1}\left(h_n\right)}{(p+1)!^2}$$

Expanding the covariance term,

$$\mathbb{C}\left[\widehat{\mathrm{B}}_{Y+,\nu,p,q}\left(h_n,b_n\right),\breve{\widehat{\mathrm{B}}}_{+,\nu,p,q}\left(h_n,b_n\right)\Big|\boldsymbol{R}\right]$$

$$= \frac{1}{\mu^{(\nu)}_{D+}}\mathbb{C}\left[\widehat{\mathrm{B}}_{Y+,\nu,p,q}\left(h_n,b_n\right),\widehat{\mathrm{B}}_{W+,\nu,p,q}\left(h_n,b_n\right)\Big|\boldsymbol{R}\right]$$

$$- \frac{\mu^{(\nu)}_{W+}}{\left(\mu^{(\nu)}_{D+}\right)^2}\mathbb{C}\left[\widehat{\mathrm{B}}_{Y+,\nu,p,q}\left(h_n,b_n\right),\widehat{\mathrm{B}}_{D+,\nu,p,q}\left(h_n,b_n\right)\Big|\boldsymbol{R}\right]$$

$$= \frac{1}{\mu_{D+}^{(\nu)}} \mathbb{C} \left[\widehat{\mu}_{Y+,q}^{(p+1)} (b_n) , \widehat{\mu}_{W+,q}^{(p+1)} (b_n) \,\Big|\, \boldsymbol{R} \right] \frac{\mathcal{B}_{+,\nu,p,p+1}^2 (h_n)}{(p+1)!^2}$$

$$- \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)} \right)^2} \mathbb{C} \left[\widehat{\mu}_{Y+,q}^{(p+1)} (b_n) , \widehat{\mu}_{D+,q}^{(p+1)} (b_n) \,\Big|\, \boldsymbol{R} \right] \frac{\mathcal{B}_{+,\nu,p,p+1}^2 (h_n)}{(p+1)!^2}$$

$$= \widetilde{\mathrm{V}}_{Y+,p+1,q} (b_n) \frac{\mathcal{B}_{+,\nu,p,p+1}^2 (h_n)}{(p+1)!^2}$$

Combining these terms,

$$\mathbb{V} \left[\widehat{\mathrm{B}}_{Y+,\nu,p,q} (h_n, b_n) + \widecheck{\widetilde{\mathrm{B}}}_{+,\nu,p,q} (h_n, b_n) \,\Big|\, \boldsymbol{R} \right]$$

$$= \left(\mathcal{V}_{YY+,p+1,q} (b_n) + \widetilde{\mathrm{V}}_{+,p+1,q} (b_n) + 2\widetilde{\mathrm{V}}_{Y+,p+1,q} (b_n) \right) \frac{\mathcal{B}_{+,\nu,p,p+1}^2 (h_n)}{(p+1)!^2}$$

$$= \mathrm{V}_{+,p+1,q} (b_n) \frac{\mathcal{B}_{+,\nu,p,p+1}^2 (h_n)}{(p+1)!^2}$$

Symmetrically,

$$\mathbb{V} \left[\widehat{\mathrm{B}}_{Y-,\nu,p,q} (h_n, b_n) + \widecheck{\widetilde{\mathrm{B}}}_{-,\nu,p,q} (h_n, b_n) \,\Big|\, \boldsymbol{R} \right]$$

$$= \left(\mathcal{V}_{YY-,p+1,q} (b_n) + \widetilde{\mathrm{V}}_{-,p+1,q} (b_n) + 2\widetilde{\mathrm{V}}_{Y-,p+1,q} (b_n) \right) \frac{\mathcal{B}_{-,\nu,p,p+1}^2 (h_n)}{(p+1)!^2}$$

$$= \mathrm{V}_{-,p+1,q} (b_n) \frac{\mathcal{B}_{-,\nu,p,p+1}^2 (h_n)}{(p+1)!^2}$$

Third,

$$\mathbb{C} \left[\widehat{\mu}_{Y+,\nu,p}^{(\nu)} (h_n) , \widehat{\mathrm{B}}_{Y+,\nu,p,q} (h_n, b_n) + \widecheck{\widetilde{\mathrm{B}}}_{+,\nu,p,q} (h_n, b_n) \,\Big|\, \boldsymbol{R} \right]$$

$$= \mathbb{C} \left[\widehat{\mu}_{Y+,\nu,p}^{(\nu)} (h_n) , \widehat{\mathrm{B}}_{Y+,\nu,p,q} (h_n, b_n) \,\Big|\, \boldsymbol{R} \right]$$

$$+ \frac{1}{\mu_{D+}^{(\nu)}} \mathbb{C} \left[\widehat{\mu}_{Y+,\nu,p}^{(\nu)} (h_n) , \widehat{\mathrm{B}}_{W+,\nu,p,q} (h_n, b_n) \,\Big|\, \boldsymbol{R} \right]$$

$$- \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)} \right)^2} \mathbb{C} \left[\widehat{\mu}_{Y+,\nu,p}^{(\nu)} (h_n) , \widehat{\mathrm{B}}_{D+,\nu,p,q} (h_n, b_n) \,\Big|\, \boldsymbol{R} \right]$$

$$= \mathbb{C} \left[\widehat{\mu}_{Y+,\nu,p}^{(\nu)} (h_n) , \widehat{\mu}_{Y+,q}^{(p+1)} (b_n) \,\Big|\, \boldsymbol{R} \right] \frac{\mathcal{B}_{+,\nu,p,p+1} (h_n)}{(p+1)!}$$

$$+ \frac{1}{\mu_{D+}^{(\nu)}} \mathbb{C} \left[\widehat{\mu}_{Y+,\nu,p}^{(\nu)} (h_n) , \widehat{\mu}_{W+,q}^{(p+1)} (b_n) \,\Big|\, \boldsymbol{R} \right] \frac{\mathcal{B}_{+,\nu,p,p+1} (h_n)}{(p+1)!}$$

$$- \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)} \right)^2} \mathbb{C} \left[\widehat{\mu}_{Y+,\nu,p}^{(\nu)} (h_n) , \widehat{\mu}_{D+,q}^{(p+1)} (b_n) \,\Big|\, \boldsymbol{R} \right] \frac{\mathcal{B}_{+,\nu,p,p+1} (h_n)}{(p+1)!}$$

89

Define

$$\widetilde{\mathrm{C}}_{Y+,\nu,p,q}\left(h_n,b_n\right)\equiv \mathcal{C}_{YY+,\nu,p,q}\left(h_n,b_n\right)$$

$$+\frac{1}{\mu_{D+}^{(\nu)}}\mathcal{C}_{YW+,\nu,p,q}\left(h_n,b_n\right)-\frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}\mathcal{C}_{YD+,\nu,p,q}\left(h_n,b_n\right)$$

and $\widetilde{\mathrm{C}}_{Y-,\nu,p,q}\left(h_n,b_n\right)$ analogously. Thus,

$$\mathbb{C}\left[\widehat{\mu}_{Y+,\nu,p}^{(\nu)}\left(h_n\right),\widehat{\mathrm{B}}_{Y+,\nu,p,q}\left(h_n,b_n\right)+\overset{\smallsmile}{\mathrm{B}}_{+,\nu,p,q}\left(h_n,b_n\right)\middle|\boldsymbol{R}\right]$$

$$=\widetilde{\mathrm{C}}_{Y+,\nu,p,q}\left(h_n,b_n\right)\frac{\mathcal{B}_{+,\nu,p,p+1}\left(h_n\right)}{(p+1)!}$$

Symmetrically,

$$\mathbb{C}\left[\widehat{\mu}_{Y-,\nu,p}^{(\nu)}\left(h_n\right),\widehat{\mathrm{B}}_{Y-,\nu,p,q}\left(h_n,b_n\right)+\overset{\smallsmile}{\mathrm{B}}_{-,\nu,p,q}\left(h_n,b_n\right)\middle|\boldsymbol{R}\right]$$

$$=\widetilde{\mathrm{C}}_{Y-,\nu,p,q}\left(h_n,b_n\right)\frac{\mathcal{B}_{-,\nu,p,p+1}\left(h_n\right)}{(p+1)!}$$

Fourth, recalling that $\widetilde{\eta}_{+,\nu,p}\left(h_n\right)\equiv\frac{1}{\mu_{D+}^{(\nu)}}\widehat{\mu}_{W+,p}^{(\nu)}\left(h_n\right)-\frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}\widehat{\mu}_{D+,p}^{(\nu)}\left(h_n\right)$,

$$\mathbb{C}\left[\widetilde{\eta}_{+,\nu,p}\left(h_n\right),\widehat{\mathrm{B}}_{Y+,\nu,p,q}\left(h_n,b_n\right)+\overset{\smallsmile}{\mathrm{B}}_{+,\nu,p,q}\left(h_n,b_n\right)\middle|\boldsymbol{R}\right]$$

$$=\mathbb{C}\left[\widetilde{\eta}_{+,\nu,p}\left(h_n\right),\widehat{\mathrm{B}}_{Y+,\nu,p,q}\left(h_n,b_n\right)\middle|\boldsymbol{R}\right]$$

$$+\frac{1}{\mu_{D+}^{(\nu)}}\mathbb{C}\left[\widetilde{\eta}_{+,\nu,p}\left(h_n\right),\widehat{\mathrm{B}}_{W+,\nu,p,q}\left(h_n,b_n\right)\middle|\boldsymbol{R}\right]$$

$$-\frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}\mathbb{C}\left[\widetilde{\eta}_{+,\nu,p}\left(h_n\right),\widehat{\mathrm{B}}_{D+,\nu,p,q}\left(h_n,b_n\right)\middle|\boldsymbol{R}\right]$$

$$=\mathbb{C}\left[\widetilde{\eta}_{+,\nu,p}\left(h_n\right),\widehat{\mu}_{Y+,q}^{(p+1)}\left(b_n\right)\middle|\boldsymbol{R}\right]\frac{\mathcal{B}_{+,\nu,p,p+1}\left(h_n\right)}{(p+1)!}$$

$$+\frac{1}{\mu_{D+}^{(\nu)}}\mathbb{C}\left[\widetilde{\eta}_{+,\nu,p}\left(h_n\right),\widehat{\mu}_{W+,q}^{(p+1)}\left(b_n\right)\middle|\boldsymbol{R}\right]\frac{\mathcal{B}_{+,\nu,p,p+1}\left(h_n\right)}{(p+1)!}$$

$$-\frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}\mathbb{C}\left[\widetilde{\eta}_{+,\nu,p}\left(h_n\right),\widehat{\mu}_{D+,q}^{(p+1)}\left(b_n\right)\middle|\boldsymbol{R}\right]\frac{\mathcal{B}_{+,\nu,p,p+1}\left(h_n\right)}{(p+1)!}$$

$$=\frac{1}{\mu_{D+}^{(\nu)}}\mathbb{C}\left[\widehat{\mu}_{W+,p}^{(\nu)}\left(h_n\right),\widehat{\mu}_{Y+,q}^{(p+1)}\left(b_n\right)\middle|\boldsymbol{R}\right]\frac{\mathcal{B}_{+,\nu,p,p+1}\left(h_n\right)}{(p+1)!}$$

$$-\frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}\mathbb{C}\left[\widehat{\mu}_{D+,p}^{(\nu)}\left(h_n\right),\widehat{\mu}_{Y+,q}^{(p+1)}\left(b_n\right)\middle|\boldsymbol{R}\right]\frac{\mathcal{B}_{+,\nu,p,p+1}\left(h_n\right)}{(p+1)!}$$

$$+ \frac{1}{\left(\mu_{D+}^{(\nu)}\right)^2} \mathbb{C}\left[\widehat{\mu}_{W+,p}^{(\nu)}(h_n), \widehat{\mu}_{W+,q}^{(p+1)}(b_n) \,\middle|\, \boldsymbol{R}\right] \frac{\mathcal{B}_{+,\nu,p,p+1}(h_n)}{(p+1)!}$$

$$- \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^3} \mathbb{C}\left[\widehat{\mu}_{D+,p}^{(\nu)}(h_n), \widehat{\mu}_{W+,q}^{(p+1)}(b_n) \,\middle|\, \boldsymbol{R}\right] \frac{\mathcal{B}_{+,\nu,p,p+1}(h_n)}{(p+1)!}$$

$$- \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^3} \mathbb{C}\left[\widehat{\mu}_{W+,p}^{(\nu)}(h_n), \widehat{\mu}_{D+,q}^{(p+1)}(b_n) \,\middle|\, \boldsymbol{R}\right] \frac{\mathcal{B}_{+,\nu,p,p+1}(h_n)}{(p+1)!}$$

$$+ \frac{\left(\mu_{W+}^{(\nu)}\right)^2}{\left(\mu_{D+}^{(\nu)}\right)^4} \mathbb{C}\left[\widehat{\mu}_{D+,p}^{(\nu)}(h_n), \widehat{\mu}_{D+,q}^{(p+1)}(b_n) \,\middle|\, \boldsymbol{R}\right] \frac{\mathcal{B}_{+,\nu,p,p+1}(h_n)}{(p+1)!}$$

Define

$$\widetilde{\mathrm{C}}_{+,\nu,p,q}(h_n, b_n) \equiv \frac{1}{\mu_{D+}^{(\nu)}} \mathcal{C}_{WY+,\nu,p,q}(h_n, b_n) - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2} \mathcal{C}_{DY+,\nu,p,q}(h_n, b_n)$$

$$+ \frac{1}{\left(\mu_{D+}^{(\nu)}\right)^2} \mathcal{C}_{WW+,\nu,p,q}(h_n, b_n) - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^3} \mathcal{C}_{DW+,\nu,p,q}(h_n, b_n)$$

$$- \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^3} \mathcal{C}_{WD+,\nu,p,q}(h_n, b_n) + \frac{\left(\mu_{W+}^{(\nu)}\right)^2}{\left(\mu_{D+}^{(\nu)}\right)^4} \mathcal{C}_{DD+,\nu,p,q}(h_n, b_n)$$

and $\widetilde{\mathrm{C}}_{-,\nu,p,q}(h_n, b_n)$ analogously. Thus,

$$\mathbb{C}\left[\widetilde{\eta}_{+,\nu,p}(h_n), \widehat{\mathrm{B}}_{Y+,\nu,p,q}(h_n, b_n) + \widecheck{\mathrm{B}}_{+,\nu,p,q}(h_n, b_n) \,\middle|\, \boldsymbol{R}\right]$$

$$= \widetilde{\mathrm{C}}_{+,\nu,p,q}(h_n, b_n) \frac{\mathcal{B}_{+,\nu,p,p+1}(h_n)}{(p+1)!}$$

Symmetrically,

$$\mathbb{C}\left[\widetilde{\eta}_{-,\nu,p}(h_n), \widehat{\mathrm{B}}_{Y-,\nu,p,q}(h_n, b_n) + \widecheck{\mathrm{B}}_{-,\nu,p,q}(h_n, b_n) \,\middle|\, \boldsymbol{R}\right]$$

$$= \widetilde{\mathrm{C}}_{-,\nu,p,q}(h_n, b_n) \frac{\mathcal{B}_{-,\nu,p,p+1}(h_n)}{(p+1)!}$$

Also define

$$\mathrm{C}_{+,\nu,p,q}(h_n, b_n) \equiv \widetilde{\mathrm{C}}_{Y+,\nu,p,q}(h_n, b_n) + \widetilde{\mathrm{C}}_{+,\nu,p,q}(h_n, b_n)$$

$$\mathrm{C}_{-,\nu,p,q}(h_n, b_n) \equiv \widetilde{\mathrm{C}}_{Y-,\nu,p,q}(h_n, b_n) + \widetilde{\mathrm{C}}_{-,\nu,p,q}(h_n, b_n)$$

Combining these derivations,

$$
\begin{aligned}
\mathrm{V}^{\mathrm{BC}}_{+,\nu,p,q}\left(h_n, b_n\right) &\equiv \mathbb{V}\left[\widehat{\mu}^{(\nu)}_{Y+,\nu,p}\left(h_n\right) + \widetilde{\eta}_{+,\nu,p}\left(h_n\right)\right.\\
&\quad\left. - h_n^{p+1-\nu}\left(\widehat{\mathrm{B}}_{Y+,\nu,p,q}\left(h_n, b_n\right) + \widecheck{\mathrm{B}}_{+,\nu,p,q}\left(h_n, b_n\right)\right)\,\middle|\,\boldsymbol{R}\right]\\
&= \mathrm{V}_{+,\nu,p}\left(h_n\right) + h_n^{2(p+1-\nu)}\mathrm{V}_{+,p+1,q}\left(b_n\right)\frac{\mathcal{B}^2_{+,\nu,p,p+1}\left(h_n\right)}{(p+1)!^2}\\
&\quad - 2h_n^{p+1-\nu}\mathrm{C}_{+,\nu,p,q}\left(h_n, b_n\right)\frac{\mathcal{B}_{+,\nu,p,p+1}\left(h_n\right)}{(p+1)!}
\end{aligned}
$$

Symmetrically,

$$
\begin{aligned}
\mathrm{V}^{\mathrm{BC}}_{-,\nu,p,q}\left(h_n, b_n\right) &\equiv \mathbb{V}\left[\widehat{\mu}^{(\nu)}_{Y-,\nu,p}\left(h_n\right) + \widetilde{\eta}_{-,\nu,p}\left(h_n\right)\right.\\
&\quad\left. - h_n^{p+1-\nu}\left(\widehat{\mathrm{B}}_{Y-,\nu,p,q}\left(h_n, b_n\right) + \widecheck{\mathrm{B}}_{-,\nu,p,q}\left(h_n, b_n\right)\right)\,\middle|\,\boldsymbol{R}\right]\\
&= \mathrm{V}_{-,\nu,p}\left(h_n\right) + h_n^{2(p+1-\nu)}\mathrm{V}_{-,p+1,q}\left(b_n\right)\frac{\mathcal{B}^2_{-,\nu,p,p+1}\left(h_n\right)}{(p+1)!^2}\\
&\quad - 2h_n^{p+1-\nu}\mathrm{C}_{-,\nu,p,q}\left(h_n, b_n\right)\frac{\mathcal{B}_{-,\nu,p,p+1}\left(h_n\right)}{(p+1)!}
\end{aligned}
$$

To conclude, the conditional variance of the linearized bias-corrected estimator $\widetilde{\theta}^{\mathrm{BC}}_{\nu,p}\left(h_n, b_n\right)$ is

$$
\mathrm{V}^{\mathrm{BC}}_{\nu,p,q}\left(h_n, b_n\right) \equiv \mathbb{V}\left[\widetilde{\theta}^{\mathrm{BC}}_{\nu,p}\left(h_n, b_n\right)\,\middle|\,\boldsymbol{R}\right] = \mathrm{V}^{\mathrm{BC}}_{+,\nu,p,q}\left(h_n, b_n\right) + \mathrm{V}^{\mathrm{BC}}_{-,\nu,p,q}\left(h_n, b_n\right)
$$

(D) Part (V) of this paper implies that

$$
\mathrm{V}^{\mathrm{BC}}_{\nu,p,q}\left(h_n, b_n\right) = O_p\left(\frac{1}{nh_n^{1+2\nu}} + \frac{h_n^{2(p+1-\nu)}}{nb_n^{2p+3}}\right)
$$

In addition, by part (R) of Theorem B.1,

$$
R_n = O_p\left(h_n^{2(p+1-\nu)} + \frac{1}{nh_n^{1+2\nu}}\right)
$$

which implies that

$$
R_n^2 = O_p\left(h_n^{4(p+1-\nu)} + \frac{1}{n^2h_n^{2+4\nu}}\right)
$$

As a consequence,

$$
\frac{R_n^2}{\mathrm{V}^{\mathrm{BC}}_{\nu,p,q}\left(h_n, b_n\right)} = O_p\left(\min\left\{nh_n^{1+2\nu}, \frac{nb_n^{2p+3}}{h_n^{2(p+1-\nu)}}\right\}\right)O_p\left(h_n^{4(p+1-\nu)} + \frac{1}{n^2h_n^{2+4\nu}}\right)
$$

$$= O_p \left(\min \left\{ h_n^{2p+3}, b_n^{2p+3} \right\} \right) O_p \left(nh_n^{2(p+1-\nu)} + \frac{1}{nh_n^{2p+4+2\nu}} \right)$$

$$= O_p \left(\min \left\{ \frac{1}{nh_n^{1+2\nu}}, \frac{b_n^{2p+3}}{nh_n^{2p+4+2\nu}} \right\} \right) O_p \left(n^2 h_n^{2(2p+3)} + 1 \right)$$

$$= O_p \left(\min \left\{ \frac{1}{nh_n^{1+2\nu}}, \frac{b_n^{2p+3}}{nh_n^{2p+4+2\nu}} \right\} \right)$$
$$+ O_p \left(\min \left\{ nh_n^{4p+5-2\nu}, nb_n^{2p+3} h_n^{2(p+1-\nu)} \right\} \right)$$

$$= O_p \left(\frac{1}{nh_n^{1+2\nu}} \min \left\{ 1, \frac{b_n^{2p+3}}{h_n^{2p+3}} \right\} \right)$$
$$+ O_p \left(nh_n^{2(p+1-\nu)} \min \left\{ h_n^{2p+3}, b_n^{2p+3} \right\} \right)$$

$$= O_p \left(\frac{1}{nh_n^{1+2\nu}} \min \left\{ 1, \frac{1}{\rho_n^{2p+3}} \right\} \right) + O_p \left(nh_n^{2(p+1-\nu)} \min \left\{ h_n^{2p+3}, b_n^{2p+3} \right\} \right)$$

where $\rho_n \equiv \frac{h_n}{b_n}$. By assumption, $nh_n^{1+2\nu} \to \infty$, implying that

$$O_p \left(\frac{1}{nh_n^{1+2\nu}} \min \left\{ 1, \frac{1}{\rho_n^{2p+3}} \right\} \right) = o_p (1)$$

Again by assumption, $nh_n^2 \min \left\{ h_n^{2p+3}, b_n^{2p+3} \right\} \to 0$, implying that

$$O_p \left(nh_n^{2(p+1-\nu)} \min \left\{ h_n^{2p+3}, b_n^{2p+3} \right\} \right) = o_p (1)$$

for any $p \geq \nu$. Taken together,

$$\frac{R_n^2}{V_{\nu,p,q}^{\text{BC}} (h_n, b_n)} = o_p (1)$$

Furthermore, by part (R) of this theorem,

$$R_n^{\text{BC}} = O_p \left(h_n^{2(p+1-\nu)} + \frac{h_n^{p+1-\nu}}{\sqrt{nh_n^{1+2\nu}}} \right) O_p \left(1 + \sqrt{\frac{1}{nb_n^{2p+3}}} \right)$$

which implies that

$$(R_n^{\text{BC}})^2 = O_p \left(h_n^{4(p+1-\nu)} + \frac{h_n^{2(p+1-\nu)}}{nh_n^{1+2\nu}} \right) O_p \left(1 + \frac{1}{nb_n^{2p+3}} \right)$$

$$= O_p \left(h_n^{4(p+1-\nu)} + \frac{h_n^{2p+3}}{nh_n^{2+4\nu}} \right) O_p \left(1 + \frac{1}{nb_n^{2p+3}} \right)$$

As a consequence,

$$\frac{R_n^{\text{BC}}}{V_{\nu,p,q}^{\text{BC}} (h_n, b_n)} = O_p \left(\min \left\{ nh_n^{1+2\nu}, \frac{nb_n^{2p+3}}{h_n^{2(p+1-\nu)}} \right\} \right)$$

$$\times O_p\left(h_n^{4(p+1-\nu)} + \frac{h_n^{2p+3}}{nh_n^{2+4\nu}}\right) O_p\left(1 + \frac{1}{nb_n^{2p+3}}\right)$$

First, with similar derivations as in previous steps,

$$O_p\left(\min\left\{nh_n^{1+2\nu}, \frac{nb_n^{2p+3}}{h_n^{2(p+1-\nu)}}\right\}\right) O_p\left(h_n^{4(p+1-\nu)} + \frac{h_n^{2p+3}}{nh_n^{2+4\nu}}\right)$$

$$= O_p\left(\min\left\{h_n^{2p+3}, b_n^{2p+3}\right\}\right) O_p\left(nh_n^{2(p+1-\nu)} + \frac{h_n^{2p+3}}{nh_n^{2p+4+2\nu}}\right)$$

$$= O_p\left(\min\left\{\frac{1}{nh_n^{1+2\nu}}, \frac{b_n^{2p+3}}{nh_n^{2p+4+2\nu}}\right\}\right) O_p\left(n^2 h_n^{2(2p+3)} + h_n^{2p+3}\right)$$

$$= O_p\left(h_n^{2p+3} \min\left\{\frac{1}{nh_n^{1+2\nu}}, \frac{b_n^{2p+3}}{nh_n^{2p+4+2\nu}}\right\}\right)$$

$$+ O_p\left(\min\left\{nh_n^{4p+5-2\nu}, nb_n^{2p+3}h_n^{2(p+1-\nu)}\right\}\right)$$

$$= O_p\left(\frac{h_n^{2p+3}}{nh_n^{1+2\nu}} \min\left\{1, \frac{b_n^{2p+3}}{h_n^{2p+3}}\right\}\right)$$

$$+ O_p\left(nh_n^{2(p+1-\nu)} \min\left\{h_n^{2p+3}, b_n^{2p+3}\right\}\right)$$

$$= O_p\left(\frac{h_n^{2p+3}}{nh_n^{1+2\nu}} \min\left\{1, \frac{1}{\rho_n^{2p+3}}\right\}\right) + O_p\left(nh_n^{2(p+1-\nu)} \min\left\{h_n^{2p+3}, b_n^{2p+3}\right\}\right)$$

$$= O_p\left(\frac{1}{nh_n^{1+2\nu}} \min\left\{h_n^{2p+3}, b_n^{2p+3}\right\}\right) + o_p(1)$$

$$= o_p(1)$$

where the last equality follows from the assumption that $nh_n^{1+2\nu} \to \infty$. Thus,

$$\frac{R_n^{\mathrm{BC}}}{\mathrm{V}_{\nu,p,q}^{\mathrm{BC}}(h_n, b_n)} = o_p(1)$$

$$+ O_p\left(\frac{h_n^{2p+3}}{nh_n^{1+2\nu}} \min\left\{1, \frac{1}{\rho_n^{2p+3}}\right\}\right) O_p\left(\frac{1}{nb_n^{2p+3}}\right)$$

$$+ O_p\left(nh_n^{2(p+1-\nu)} \min\left\{h_n^{2p+3}, b_n^{2p+3}\right\}\right) O_p\left(\frac{1}{nb_n^{2p+3}}\right)$$

$$= o_p(1)$$

$$+ O_p\left(\frac{1}{nh_n^{1+2\nu}} \frac{1}{n} \min\left\{\rho_n^{2p+3}, 1\right\}\right) + O_p\left(h_n^{2(p+1-\nu)} \min\left\{\rho_n^{2p+3}, 1\right\}\right)$$

$$= o_p(1)$$

where the last equality follows from the assumptions that $nh_n^{1+2\nu} \to \infty$ and $h_n \to 0$. The robust bias-corrected test statistic for the estimator $\widehat{\theta}_{\nu,p}^{\mathrm{BC}}(h_n, b_n)$ is

$$T^{\mathrm{RBC}}(h_n, b_n) \equiv \frac{\widehat{\theta}_{\nu,p}^{\mathrm{BC}}(h_n, b_n) - \theta_\nu}{\sqrt{\mathrm{V}_{\nu,p,q}^{\mathrm{BC}}(h_n, b_n)}}$$

$$\begin{aligned}
&= \frac{\widetilde{\theta}_{\nu,p}^{\text{BC}}(h_n, b_n) - \left(\mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)}\right)}{\sqrt{\mathrm{V}_{\nu,p,q}^{\text{BC}}(h_n, b_n)}} + \frac{R_n + R_n^{\text{BC}}}{\sqrt{\mathrm{V}_{\nu,p,q}^{\text{BC}}(h_n, b_n)}} \\
&= \frac{\widetilde{\theta}_{\nu,p}^{\text{BC}}(h_n, b_n) - \left(\mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)}\right)}{\sqrt{\mathrm{V}_{\nu,p,q}^{\text{BC}}(h_n, b_n)}} + o_p(1) \\
&\xrightarrow{d} \mathcal{N}(0, 1)
\end{aligned}$$

where the last line is proved by recognizing that the bias-corrected linearized estimator $\widetilde{\theta}_{\nu,p}^{\text{BC}}(h_n, b_n)$ is numerically equivalent to the sharp regression discontinuity estimator that uses the sum of $Y + \frac{1}{\mu_D^{(\nu)}}W - \frac{\mu_W^{(\nu)}}{\left(\mu_D^{(\nu)}\right)^2}D$ and a bias correction term as the outcome. Thus, provided that $n\min\{h_n^{2p+3}, b_n^{2p+3}\}\max\left\{h_n^2, b_n^{2(q-p)}\right\} \to 0$ and $\kappa\max\{h_n, b_n\} < \kappa_0$, the result follows from part (D) of Theorem A.1 in Calonico, Cattaneo and Titiunik (2014b).

B.3.6 Proof of Proposition 7

The statement follows directly from specializing Theorem B.2 to $\nu = 0$, $p = 1$, and $q = 2$. As a matter of fact, the asymptotic variance of the bias-corrected estimator $\widetilde{\theta}(h_n, b_n)$ is

$$\mathrm{V}^{\text{BC}}(h_n, b_n) \equiv \mathbb{V}\left[\widetilde{\theta}^{\text{BC}}(h_n, b_n)\,\Big|\,\boldsymbol{R}\right] = \mathrm{V}_{+,0,1,2}^{\text{BC}}(h_n, b_n) + \mathrm{V}_{-,0,1,2}^{\text{BC}}(h_n, b_n)$$

where $\mathrm{V}_{+,\nu,p,q}^{\text{BC}}(h_n, b_n)$ and $\mathrm{V}_{+,\nu,p,q}^{\text{BC}}(h_n, b_n)$ are defined in the statement of Theorem B.2.

B.4 MSE-Optimal Bandwidths

In this section, I derive the asymptotic (infeasible) bandwidth $h_{\text{MSE},\nu,p}$ that minimizes the Mean Squared Error (MSE) of the linearized estimator $\widetilde{\theta}_{\nu,p}(h_n)$. I subsequently specialize this expression to compute MSE-optimal main and pilot bandwidths for the dynamic RD estimand in the paper. Before stating the general theorem, it is convenient to define additional recurring bias- and variance-related terms.

B.4.1 Definitions of Bias Terms

Recall that, by part (B) of Lemma S.A.3 in Calonico, Cattaneo and Titiunik (2014b),

$$\mathcal{B}_{+,\nu,p,q}(h_n) = \nu!e_\nu'\Gamma_p^{-1}\vartheta_{p,q} + o_p(1) \qquad \mathcal{B}_{-,\nu,p,q}(h_n) = (-1)^{\nu+q}\nu!e_\nu'\Gamma_p^{-1}\vartheta_{p,q} + o_p(1) \quad \text{(B.98)}$$

which implies that the h_n argument can be suppressed. To eliminate the dependence on h_n, define the following bias terms:

$$\dot{\mathcal{B}}_{+,\nu,p,q} \equiv \nu! e_\nu' \Gamma_p^{-1} \vartheta_{p,q} \qquad \dot{\mathcal{B}}_{-,\nu,p,q} \equiv (-1)^{\nu+q} \nu! e_\nu' \Gamma_p^{-1} \vartheta_{p,q} \tag{B.99}$$

Clearly,

$$\mathcal{B}_{+,\nu,p,q}(h_n) = \dot{\mathcal{B}}_{+,\nu,p,q} + o_p(1) \qquad \mathcal{B}_{-,\nu,p,q}(h_n) = \dot{\mathcal{B}}_{-,\nu,p,q} + o_p(1) \tag{B.100}$$

In turn, for any random variable A, define

$$\dot{\mathrm{B}}_{A+,\nu,p,q} \equiv \frac{\mu_{A+}^{(q)}}{q!} \dot{\mathcal{B}}_{+,\nu,p,q} \qquad \dot{\mathrm{B}}_{A-,\nu,p,q} \equiv \frac{\mu_{A-}^{(q)}}{q!} \dot{\mathcal{B}}_{-,\nu,p,q} \tag{B.101}$$

B.4.2 Definitions of Variance Terms

Recall that, by part (V) of Lemma S.A.3 in Calonico, Cattaneo and Titiunik (2014b), for any pair of random variables A and B,

$$\mathcal{V}_{AB+,\nu,p}(h_n) = \frac{1}{nh_n^{1+2\nu}} \frac{\sigma_{AB+}^2}{f_R(c)} \nu!^2 e_\nu' \Gamma_p^{-1} \Psi_p \Gamma_p^{-1} e_\nu (1 + o_p(1)) \tag{B.102}$$

$$\mathcal{V}_{AB-,\nu,p}(h_n) = \frac{1}{nh_n^{1+2\nu}} \frac{\sigma_{AB-}^2}{f_R(c)} \nu!^2 e_\nu' \Gamma_p^{-1} \Psi_p \Gamma_p^{-1} e_\nu (1 + o_p(1)) \tag{B.103}$$

To eliminate the dependence on h_n, define the following variance terms:

$$\dot{\mathcal{V}}_{AB+,\nu,p} \equiv \frac{\sigma_{AB+}^2}{f_R(c)} \nu!^2 e_\nu' \Gamma_p^{-1} \Psi_p \Gamma_p^{-1} e_\nu \tag{B.104}$$

$$\dot{\mathcal{V}}_{AB-,\nu,p} \equiv \frac{\sigma_{AB-}^2}{f_R(c)} \nu!^2 e_\nu' \Gamma_p^{-1} \Psi_p \Gamma_p^{-1} e_\nu \tag{B.105}$$

Clearly,

$$\mathcal{V}_{AB+,\nu,p}(h_n) = \frac{1}{nh_n^{1+2\nu}} \dot{\mathcal{V}}_{AB+,\nu,p} (1 + o_p(1)) \tag{B.106}$$

$$\mathcal{V}_{AB-,\nu,p}(h_n) = \frac{1}{nh_n^{1+2\nu}} \dot{\mathcal{V}}_{AB-,\nu,p} (1 + o_p(1)) \tag{B.107}$$

B.4.3 General Theorem: Statement

Theorem B.3 *Suppose Assumptions 8 and 9 hold with $\delta \geq p+1$ and $\nu \leq p$. If $h_n \to 0$ and $nh_n \to \infty$, then the Mean Squared Error of the linearized estimator $\widetilde{\theta}_{\nu,p}(h_n)$ is*

$$MSE\left(\widetilde{\theta}_{\nu,p}(h_n)\right) = h_n^{2(p+1-\nu)} \left(\dot{\mathrm{B}}_{\nu,p,p+1}^2 + o_p(1)\right) + \frac{1}{nh_n^{1+2\nu}} \left(\dot{\mathrm{V}}_{\nu,p} + o_p(1)\right)$$

First,

$$\dot{B}_{\nu,p,q} \equiv \dot{B}_{Y,\nu,p,q} + \dot{\tilde{B}}_{\nu,p,q}$$

with

$$\dot{B}_{Y,\nu,p,q} \equiv \dot{B}_{Y+,\nu,p,q} - \dot{B}_{Y-,\nu,p,q}$$
$$\dot{\tilde{B}}_{\nu,p,q} \equiv \dot{\tilde{B}}_{+,\nu,p,q} - \dot{\tilde{B}}_{-,\nu,p,q}$$

and

$$\dot{\tilde{B}}_{+,\nu,p,q} \equiv \frac{1}{\mu_{D+}^{(\nu)}} \dot{B}_{W+,\nu,p,q} - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2} \dot{B}_{D+,\nu,p,q}$$

$$\dot{\tilde{B}}_{-,\nu,p,q} \equiv \frac{1}{\mu_{D-}^{(\nu)}} \dot{B}_{W-,\nu,p,q} - \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^2} \dot{B}_{D-,\nu,p,q}$$

Second,

$$\dot{V}_{\nu,p} \equiv \dot{V}_{Y,\nu,p} + \dot{\tilde{V}}_{\nu,p} + 2\dot{\tilde{V}}_{Y,\nu,p}$$

where

$$\dot{V}_{Y,\nu,p} \equiv \dot{\mathcal{V}}_{YY+,\nu,p} + \dot{\mathcal{V}}_{YY-,\nu,p}$$
$$\dot{\tilde{V}}_{\nu,p} \equiv \dot{\tilde{V}}_{+,\nu,p} + \dot{\tilde{V}}_{-,\nu,p}$$
$$\dot{\tilde{V}}_{Y,\nu,p} \equiv \dot{\tilde{V}}_{Y+,\nu,p} + \dot{\tilde{V}}_{Y-,\nu,p}$$

with

$$\dot{\tilde{V}}_{+,\nu,p} \equiv \frac{1}{\left(\mu_{D+}^{(\nu)}\right)^2} \dot{\mathcal{V}}_{WW+,\nu,p} + \frac{\left(\mu_{W+}^{(\nu)}\right)^2}{\left(\mu_{D+}^{(\nu)}\right)^4} \dot{\mathcal{V}}_{DD+,\nu,p} - \frac{2\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^3} \dot{\mathcal{V}}_{WD+,\nu,p}$$

$$\dot{\tilde{V}}_{-,\nu,p} \equiv \frac{1}{\left(\mu_{D-}^{(\nu)}\right)^2} \dot{\mathcal{V}}_{WW-,\nu,p} + \frac{\left(\mu_{W-}^{(\nu)}\right)^2}{\left(\mu_{D-}^{(\nu)}\right)^4} \dot{\mathcal{V}}_{DD-,\nu,p} - \frac{2\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^3} \dot{\mathcal{V}}_{WD-,\nu,p}$$

and

$$\dot{\tilde{V}}_{Y+,\nu,p} \equiv \frac{1}{\mu_{D+}^{(\nu)}} \dot{\mathcal{V}}_{YW+,\nu,p} - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2} \dot{\mathcal{V}}_{YD+,\nu,p}$$

$$\dot{\tilde{\mathbf{V}}}_{Y-,\nu,p} \equiv \frac{1}{\mu_{D-}^{(\nu)}} \dot{\mathcal{V}}_{YW-,\nu,p} - \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^2} \dot{\mathcal{V}}_{YD-,\nu,p}$$

Finally, the MSE-minimizing bandwidth is

$$h_{\mathrm{MSE},\nu,p} = \left(\frac{(1+2\nu)\,\dot{\mathbf{V}}_{\nu,p}}{2(p+1-\nu)\dot{\mathbf{B}}_{\nu,p,p+1}^2}\right)^{\frac{1}{2p+3}} n^{-\frac{1}{2p+3}}$$

B.4.4   General Theorem: Proof

The conditional Mean Squared Error of $\widetilde{\theta}_{\nu,p}(h_n)$ is

$$\mathrm{MSE}\left(\widetilde{\theta}_{\nu,p}(h_n)\right) \equiv \mathbb{E}\left[\left(\widetilde{\theta}_{\nu,p}(h_n) - \left(\mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)}\right)\right)^2 \middle| \mathbf{R}\right]$$

$$= \left(\mathbb{E}\left[\widetilde{\theta}_{\nu,p}(h_n)\middle|\mathbf{R}\right] - \left(\mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)}\right)\right)^2 + \mathbb{V}\left[\widetilde{\theta}_{\nu,p}(h_n)\middle|\mathbf{R}\right]$$

First, part (B) of Theorem B.1 states that, as $h_n \to 0$,

$$\mathbb{E}\left[\widetilde{\theta}_{\nu,p}(h_n)\middle|\mathbf{R}\right] = \mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)} + h_n^{p+1-\nu}\mathrm{B}_{\nu,p,p+1}(h_n)$$

$$+ h_n^{p+2-\nu}\mathrm{B}_{\nu,p,p+2}(h_n) + o_p\left(h_n^{p+2-\nu}\right)$$

$$= \mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)} + h_n^{p+1-\nu}\mathrm{B}_{\nu,p,p+1}(h_n)$$

$$+ h_n^{p+1-\nu}\mathrm{B}_{\nu,p,p+1}(h_n)\left(\frac{h_n\mathrm{B}_{\nu,p,p+2}(h_n)}{\mathrm{B}_{\nu,p,p+1}(h_n)} + \frac{o_p(h_n)}{\mathrm{B}_{\nu,p,p+1}(h_n)}\right)$$

$$= \mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)} + h_n^{p+1-\nu}\mathrm{B}_{\nu,p,p+1}(h_n) + h_n^{p+1-\nu}\mathrm{B}_{\nu,p,p+1}(h_n)\,o_p(1)$$

$$= \mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)} + h_n^{p+1-\nu}\mathrm{B}_{\nu,p,p+1}(h_n)(1 + o_p(1))$$

Thus,

$$\left(\mathbb{E}\left[\widetilde{\theta}_{\nu,p}(h_n)\middle|\mathbf{R}\right] - \left(\mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)}\right)\right)^2 = \left(h_n^{p+1-\nu}\mathrm{B}_{\nu,p,p+1}(h_n)(1+o_p(1))\right)^2$$

$$= h_n^{2(p+1-\nu)}\mathrm{B}_{\nu,p,p+1}^2(h_n)(1+o_p(1))^2$$

$$= h_n^{2(p+1-\nu)}\mathrm{B}_{\nu,p,p+1}^2(h_n)(1+o_p(1))$$

$$= h_n^{2(p+1-\nu)}\left(\mathrm{B}_{\nu,p,p+1}^2(h_n) + o_p(1)\right)$$

Then, for $h_n \to 0$, the bias terms can be redefined as follows:

$$\dot{\mathrm{B}}_{Y,\nu,p,q} \equiv \dot{\mathrm{B}}_{Y+,\nu,p,q} - \dot{\mathrm{B}}_{Y-,\nu,p,q}$$

$$\dot{\widetilde{B}}_{+,\nu,p,q} \equiv \frac{1}{\mu_{D+}^{(\nu)}} \dot{B}_{W+,\nu,p,q} - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2} \dot{B}_{D+,\nu,p,q}$$

$$\dot{\widetilde{B}}_{-,\nu,p,q} \equiv \frac{1}{\mu_{D-}^{(\nu)}} \dot{B}_{W-,\nu,p,q} - \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^2} \dot{B}_{D-,\nu,p,q}$$

Furthermore,

$$\dot{\widetilde{B}}_{\nu,p,q} \equiv \dot{\widetilde{B}}_{+,\nu,p,q} - \dot{\widetilde{B}}_{-,\nu,p,q}$$

Finally,

$$\dot{B}_{\nu,p,q} \equiv \dot{B}_{Y,\nu,p,q} + \dot{\widetilde{B}}_{\nu,p,q}$$

To conclude,

$$\left(\mathbb{E}\left[\widetilde{\theta}_{\nu,p}\left(h_n\right)\Big|\boldsymbol{R}\right] - \left(\mu_{Y+}^{(\nu)} - \mu_{Y-}^{(\nu)}\right)\right)^2 = h_n^{2(p+1-\nu)}\left(\dot{B}_{\nu,p,p+1}^2 + o_p\left(1\right)\right)$$

Second, part (V) of Theorem B.1 states that, as $h_n \to 0$,

$$\mathbb{V}\left[\widetilde{\theta}_{\nu,p}\left(h_n\right)\Big|\boldsymbol{R}\right] = V_{Y,\nu,p}\left(h_n\right) + \widetilde{V}_{\nu,p}\left(h_n\right) + 2\widetilde{V}_{Y,\nu,p}\left(h_n\right)$$

Then, for $h_n \to 0$, the variance terms can be redefined as follows:

$$\dot{\widetilde{V}}_{Y+,\nu,p} \equiv \frac{1}{\mu_{D+}^{(\nu)}} \dot{V}_{YW+,\nu,p} - \frac{\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2} \dot{V}_{YD+,\nu,p}$$

$$\dot{\widetilde{V}}_{Y-,\nu,p} \equiv \frac{1}{\mu_{D-}^{(\nu)}} \dot{V}_{YW-,\nu,p} - \frac{\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^2} \dot{V}_{YD-,\nu,p}$$

Furthermore,

$$\dot{\widetilde{V}}_{+,\nu,p} \equiv \frac{1}{\left(\mu_{D+}^{(\nu)}\right)^2} \dot{V}_{WW+,\nu,p} + \frac{\left(\mu_{W+}^{(\nu)}\right)^2}{\left(\mu_{D+}^{(\nu)}\right)^4} \dot{V}_{DD+,\nu,p} - \frac{2\mu_{W+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^3} \dot{V}_{WD+,\nu,p}$$

$$\dot{\widetilde{V}}_{-,\nu,p} \equiv \frac{1}{\left(\mu_{D-}^{(\nu)}\right)^2} \dot{V}_{WW-,\nu,p} + \frac{\left(\mu_{W-}^{(\nu)}\right)^2}{\left(\mu_{D-}^{(\nu)}\right)^4} \dot{V}_{DD-,\nu,p} - \frac{2\mu_{W-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^3} \dot{V}_{WD-,\nu,p}$$

And

$$\dot{V}_{Y,\nu,p} \equiv \dot{V}_{YY+,\nu,p} + \dot{V}_{YY-,\nu,p}$$

$$\dot{\tilde{\mathrm{V}}}_{\nu,p} \equiv \dot{\tilde{\mathrm{V}}}_{+,\nu,p} + \dot{\tilde{\mathrm{V}}}_{-,\nu,p}$$

$$\dot{\tilde{\mathrm{V}}}_{Y,\nu,p} \equiv \dot{\tilde{\mathrm{V}}}_{Y+,\nu,p} + \dot{\tilde{\mathrm{V}}}_{Y-,\nu,p}$$

Finally,

$$\dot{\mathrm{V}}_{\nu,p} \equiv \dot{\mathrm{V}}_{Y,\nu,p} + \dot{\tilde{\mathrm{V}}}_{\nu,p} + 2\dot{\tilde{\mathrm{V}}}_{Y,\nu,p}$$

To conclude,

$$\mathbb{V}\left[\widetilde{\theta}_{\nu,p}\left(h_n\right)\bigg|\boldsymbol{R}\right] = \frac{1}{nh_n^{1+2\nu}}\dot{\mathrm{V}}_{\nu,p}\left(1 + o_p\left(1\right)\right) = \frac{1}{nh_n^{1+2\nu}}\left(\dot{\mathrm{V}}_{\nu,p} + o_p\left(1\right)\right)$$

Taken together, as $h_n \to 0$, the Mean Squared Error of the linearized estimator $\widetilde{\theta}_{\nu,p}\left(h_n\right)$ is

$$\mathrm{MSE}\left(\widetilde{\theta}_{\nu,p}\left(h_n\right)\right) = h_n^{2(p+1-\nu)}\left(\dot{\mathrm{B}}_{\nu,p,p+1}^2 + o_p\left(1\right)\right) + \frac{1}{nh_n^{1+2\nu}}\left(\dot{\mathrm{V}}_{\nu,p} + o_p\left(1\right)\right)$$

The MSE-optimal bandwidth for the linearized estimator $h_{\mathrm{MSE},\nu,p}$ can then be computed by minimizing $\mathrm{MSE}\left(\widetilde{\theta}_{\nu,p}\left(h_n\right)\right)$ with respect to h_n:

$$h_{\mathrm{MSE},\nu,p} \equiv \arg\min_{h}\left\{h^{2(p+1-\nu)}\dot{\mathrm{B}}_{\nu,p,p+1}^2 + \frac{1}{nh^{1+2\nu}}\dot{\mathrm{V}}_{\nu,p}\right\}$$

Taking the first-order condition and solving for the MSE-minimizing bandwidth yields

$$2(p+1-\nu)h_{\mathrm{MSE},\nu,p}^{2p+1-2\nu}\dot{\mathrm{B}}_{\nu,p,p+1}^2 = (1+2\nu)\,h_{\mathrm{MSE},\nu,p}^{-2-2\nu}n^{-1}\dot{\mathrm{V}}_{\nu,p}$$

$$\Longleftrightarrow h_{\mathrm{MSE},\nu,p}^{2p+3} = \frac{(1+2\nu)\,\dot{\mathrm{V}}_{\nu,p}}{2(p+1-\nu)\dot{\mathrm{B}}_{\nu,p,p+1}^2} \times n^{-1}$$

$$\Longleftrightarrow h_{\mathrm{MSE},\nu,p} = \left(\frac{(1+2\nu)\,\dot{\mathrm{V}}_{\nu,p}}{2(p+1-\nu)\dot{\mathrm{B}}_{\nu,p,p+1}^2}\right)^{\frac{1}{2p+3}}n^{-\frac{1}{2p+3}}$$

which completes the proof.

B.4.5   Proof of Proposition 8

The first half of the statement follows from specializing Theorem B.3 to $\nu = 0$ and $p = 1$. As a matter of fact, the Mean Squared Error of the linearized estimator $\widetilde{\theta}\left(h_n\right)$ is

$$\mathrm{MSE}\left(\widetilde{\theta}\left(h_n\right)\right) = \mathrm{MSE}\left(\widetilde{\theta}_{0,1}\left(h_n\right)\right) = h_n^4\left(\dot{\mathrm{B}}_{0,1,2}^2 + o_p\left(1\right)\right) + \frac{1}{nh_n}\left(\dot{\mathrm{V}}_{0,1} + o_p\left(1\right)\right)$$

Thus, the MSE-optimal main bandwidth is

$$
h_{\text{MSE}} = \left(\frac{\dot{V}_{0,1}}{4\dot{B}_{0,1,2}^2} \right)^{1/5} n^{-1/5}
$$

To complete the proof, define $\dot{B}_h \equiv \dot{B}_{0,1,2}$ and $\dot{V}_h \equiv \dot{V}_{0,1}$. Analogously, the second half of the statement follows from specializing Theorem B.3 to $\nu = 2$ and $p = 2$. As a matter of fact, the Mean Squared Error of the linearized estimator $\widetilde{\theta}_{2,2}(b_n)$ is

$$
\text{MSE}\left(\widetilde{\theta}_{2,2}(b_n) \right) = b_n^2 \left(\dot{B}_{2,2,3}^2 + o_p(1) \right) + \frac{1}{nb_n^5} \left(\dot{V}_{2,2} + o_p(1) \right)
$$

Thus, the MSE-optimal pilot bandwidth is

$$
b_{\text{MSE}} = \left(\frac{5\dot{V}_{2,2}}{2\dot{B}_{2,2,3}^2} \right)^{1/7} n^{-1/7}
$$

To complete the proof, define $\dot{B}_b \equiv \dot{B}_{2,2,3}$ and $\dot{V}_b \equiv \dot{V}_{2,2}$.

B.5 Assessing the Plausibility of the Common Trends Assumption

The goal of this section is to derive a statistic to test hypotheses pertaining to the plausibility of the common trends assumption. As in standard difference-in-differences designs, this test hinges on the availability of "pre-periods", namely periods prior to the focal round of treatment assignment. Under Assumption 7, the focal time period is preceded by k periods for which $D_{t-1} = \cdots = D_{t-k} = 0$. Then, based on Assumption 6, an empirical researcher may expect that, for $t - k \leq u < v \leq t$ and any $d_{t+1}, \ldots, d_{t+\tau} \in \{0,1\}$,

$$
\mathbb{E}\left[\Delta_{u,v}(h_{t-k-1}, 0_k, d_t, 0_\tau) | H_{t-1} = h_{t-1}, R_t = c, P(h_{t-k-1}, 0_k, d_t, 0_\tau) = 1 \right]
$$
$$
= \mathbb{E}\left[\Delta_{u,v}(h_{t-k-1}, 0_k, d_t, 0_\tau) | H_{t-1} = h_{t-1}, R_t = c, P(h_{t-k-1}, 0_k, d_t, d_{t+1}, \ldots, d_{t+\tau}) = 1 \right]
$$
$$
\text{(B.108)}
$$

with $\Delta_{u,v}(h_{t-k-1}, 0_k, d_t, 0_\tau) \equiv Y_v(h_{t-k-1}, 0_k, d_t, 0_\tau) - Y_u(h_{t-k-1}, 0_k, d_t, 0_\tau)$. Because units are untreated in the k periods preceding t, both conditional expectations are identified using the upper or lower limit of R_t. Specifically,

$$
\lim_{r \downarrow c} \mathbb{E}\left[Y_v - Y_u \middle| H_{t-1} = h_{t-1}, R_t = r, \bigcap_{s=1}^{\tau} \{D_{t+s} = 0\} \right]
$$
$$
= \lim_{r \downarrow c} \mathbb{E}\left[Y_v - Y_u \middle| H_{t-1} = h_{t-1}, R_t = r, \bigcup_{s=1}^{\tau} \{D_{t+s} \neq 0\} \right] \qquad \text{(B.109)}
$$

and the equality of lower limits follows from a symmetric argument. In addition, by an application of the Law Iterated Expectations and the properties of limits, equation (B.109) can be written as

$$\frac{\lim_{r\downarrow c} \mathbb{E}\left[(Y_v - Y_u)\prod_{s=1}^{\tau}(1 - D_{t+s}) \,|H_{t-1} = h_{t-1}, R_t = r\right]}{\lim_{r\downarrow c} \mathbb{E}\left[\prod_{s=1}^{\tau}(1 - D_{t+s}) \,|H_{t-1} = h_{t-1}, R_t = r\right]}$$
$$= \frac{\lim_{r\downarrow c} \mathbb{E}\left[(Y_v - Y_u)\left(1 - \prod_{s=1}^{\tau}(1 - D_{t+s})\right) \,|H_{t-1} = h_{t-1}, R_t = r\right]}{\lim_{r\downarrow c} \mathbb{E}\left[1 - \prod_{s=1}^{\tau}(1 - D_{t+s}) \,|H_{t-1} = h_{t-1}, R_t = r\right]} \tag{B.110}$$

Hereafter, to keep notation compact, I omit the event $H_{t-1} = h_{t-1}$ from the conditioning set and define the following random variables:

$$J \equiv (Y_v - Y_u)\prod_{s=1}^{\tau}(1 - D_{t+s}) \qquad\qquad D \equiv \prod_{s=1}^{\tau}(1 - D_{t+s}) \tag{B.111}$$

$$K \equiv (Y_v - Y_u)\left(1 - \prod_{s=1}^{\tau}(1 - D_{t+s})\right) \qquad G \equiv 1 - \prod_{s=1}^{\tau}(1 - D_{t+s}) \tag{B.112}$$

so that the two equalities of interest can be concisely expressed as

$$\pi_+ \equiv \frac{\mu_{J+}}{\mu_{D+}} - \frac{\mu_{K+}}{\mu_{G+}} = 0 \qquad\qquad \pi_- \equiv \frac{\mu_{J-}}{\mu_{D-}} - \frac{\mu_{K-}}{\mu_{G-}} = 0 \tag{B.113}$$

For consistency with previous sections, I consider more general parameters that involve derivatives of arbitrary order ν:

$$\pi_{+,\nu} \equiv \frac{\mu_{J+}^{(\nu)}}{\mu_{D+}^{(\nu)}} - \frac{\mu_{K+}^{(\nu)}}{\mu_{G+}^{(\nu)}} \qquad\qquad \pi_{-,\nu} \equiv \frac{\mu_{J-}^{(\nu)}}{\mu_{D-}^{(\nu)}} - \frac{\mu_{K-}^{(\nu)}}{\mu_{G-}^{(\nu)}} \tag{B.114}$$

In the remainder of this section, I construct a statistic to test the null hypothesis implied by $\pi_{+,\nu} = 0$ and $\pi_{-,\nu} = 0$ jointly. The argument for constructing bias-corrected estimators for these parameters and computing their variances is analogous to the one required to derive the asymptotic distribution of $\widehat{\theta}_{\nu,p,q}^{\text{BC}}(h_n, b_n)$.

B.5.1 Definitions of Estimators

Given a bandwidth sequence h_n, natural estimators for the target parameters are

$$\widehat{\pi}_{+,\nu,p}(h_n) \equiv \frac{\widehat{\mu}_{J+,p}^{(\nu)}(h_n)}{\widehat{\mu}_{D+,p}^{(\nu)}(h_n)} - \frac{\widehat{\mu}_{K+,p}^{(\nu)}(h_n)}{\widehat{\mu}_{G+,p}^{(\nu)}(h_n)} \qquad \widehat{\pi}_{-,\nu,p}(h_n) \equiv \frac{\widehat{\mu}_{J-,p}^{(\nu)}(h_n)}{\widehat{\mu}_{D-,p}^{(\nu)}(h_n)} - \frac{\widehat{\mu}_{K-,p}^{(\nu)}(h_n)}{\widehat{\mu}_{G-,p}^{(\nu)}(h_n)} \tag{B.115}$$

Proceeding as in previous sections, a second-order Taylor expansion yields the following linearized estimators:

$$\widetilde{\pi}_{+,\nu,p}(h_n) \equiv \frac{1}{\mu_{D+}^{(\nu)}}\left(\widehat{\mu}_{J+,p}^{(\nu)}(h_n) - \mu_{J+}^{(\nu)}\right) - \frac{\mu_{J+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}\left(\widehat{\mu}_{D+,p}^{(\nu)}(h_n) - \mu_{D+}^{(\nu)}\right)$$

$$- \frac{1}{\mu_{G+}^{(\nu)}} \left(\widehat{\mu}_{K+,p}^{(\nu)} (h_n) - \mu_{K+}^{(\nu)} \right) + \frac{\mu_{K+}^{(\nu)}}{\left(\mu_{G+}^{(\nu)} \right)^2} \left(\widehat{\mu}_{G+,p}^{(\nu)} (h_n) - \mu_{G+}^{(\nu)} \right) \tag{B.116}$$

$$\widetilde{\pi}_{-,\nu,p} (h_n) \equiv \frac{1}{\mu_{D-}^{(\nu)}} \left(\widehat{\mu}_{J-,p}^{(\nu)} (h_n) - \mu_{J-}^{(\nu)} \right) - \frac{\mu_{J-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)} \right)^2} \left(\widehat{\mu}_{D-,p}^{(\nu)} (h_n) - \mu_{D-}^{(\nu)} \right)$$

$$- \frac{1}{\mu_{G-}^{(\nu)}} \left(\widehat{\mu}_{K-,p}^{(\nu)} (h_n) - \mu_{K-}^{(\nu)} \right) + \frac{\mu_{K-}^{(\nu)}}{\left(\mu_{G-}^{(\nu)} \right)^2} \left(\widehat{\mu}_{G-,p}^{(\nu)} (h_n) - \mu_{G-}^{(\nu)} \right) \tag{B.117}$$

B.5.2   Bias of Linearized Estimators

As in Theorem B.1, suppose that Assumptions 8 and 9 hold with $\delta \geq p + 2$, and $nh_n \to \infty$. Then, if $h_n \to 0$, the conditional expectation of the linearized estimator $\widetilde{\pi}_{+,\nu,p} (h_n)$ is

$$\mathbb{E} \left[\widetilde{\pi}_{+,\nu,p} (h_n) \,\middle|\, \boldsymbol{R} \right] = h_n^{p+1-\nu} \mathrm{B}_{+,\nu,p,p+1} (h_n) + h_n^{p+2-\nu} \mathrm{B}_{+,\nu,p,p+2} (h_n) + o_p \left(h_n^{p+2-\nu} \right) \tag{B.118}$$

where

$$\mathrm{B}_{+,\nu,p,q} (h_n) \equiv \mathrm{B}_{1+,\nu,p,q} (h_n) - \mathrm{B}_{2+,\nu,p,q} (h_n) \tag{B.119}$$

with

$$\mathrm{B}_{1+,\nu,p,q} (h_n) \equiv \frac{1}{\mu_{D+}^{(\nu)}} \mathrm{B}_{J+,\nu,p,q} (h_n) - \frac{\mu_{J+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)} \right)^2} \mathrm{B}_{D+,\nu,p,q} (h_n) \tag{B.120}$$

$$\mathrm{B}_{2+,\nu,p,q} (h_n) \equiv \frac{1}{\mu_{G+}^{(\nu)}} \mathrm{B}_{K+,\nu,p,q} (h_n) - \frac{\mu_{K+}^{(\nu)}}{\left(\mu_{G+}^{(\nu)} \right)^2} \mathrm{B}_{G+,\nu,p,q} (h_n) \tag{B.121}$$

Symmetrically, the conditional expectation of the linearized estimator $\widetilde{\pi}_{-,\nu,p} (h_n)$ is

$$\mathbb{E} \left[\widetilde{\pi}_{-,\nu,p} (h_n) \,\middle|\, \boldsymbol{R} \right] = h_n^{p+1-\nu} \mathrm{B}_{-,\nu,p,p+1} (h_n) + h_n^{p+2-\nu} \mathrm{B}_{-,\nu,p,p+2} (h_n) + o_p \left(h_n^{p+2-\nu} \right) \tag{B.122}$$

where

$$\mathrm{B}_{-,\nu,p,q} (h_n) \equiv \mathrm{B}_{1-,\nu,p,q} (h_n) - \mathrm{B}_{2-,\nu,p,q} (h_n) \tag{B.123}$$

with

$$\mathrm{B}_{1-,\nu,p,q} (h_n) \equiv \frac{1}{\mu_{D-}^{(\nu)}} \mathrm{B}_{J-,\nu,p,q} (h_n) - \frac{\mu_{J-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)} \right)^2} \mathrm{B}_{D-,\nu,p,q} (h_n) \tag{B.124}$$

$$\mathrm{B}_{2-,\nu,p,q} (h_n) \equiv \frac{1}{\mu_{G-}^{(\nu)}} \mathrm{B}_{K-,\nu,p,q} (h_n) - \frac{\mu_{K-}^{(\nu)}}{\left(\mu_{G-}^{(\nu)} \right)^2} \mathrm{B}_{G-,\nu,p,q} (h_n) \tag{B.125}$$

B.5.3 Variance of Linearized Estimators

Maintaining the assumptions of Theorem B.1, if $h_n \to 0$, the conditional variance of the linearized estimator $\widetilde{\pi}_{+,\nu,p}(h_n)$ is

$$\mathrm{V}_{+,\nu,p}(h_n) \equiv \mathbb{V}\left[\widetilde{\pi}_{+,\nu,p}(h_n)\,\middle|\,\boldsymbol{R}\right] = \mathrm{V}_{11+,\nu,p}(h_n) + \mathrm{V}_{22+,\nu,p}(h_n) - 2\mathrm{V}_{12+,\nu,p}(h_n) \quad \text{(B.126)}$$

where

$$\mathrm{V}_{11+,\nu,p}(h_n) \equiv \frac{1}{\left(\mu_{D+}^{(\nu)}\right)^2}\mathcal{V}_{JJ+,\nu,p}(h_n) + \frac{\left(\mu_{J+}^{(\nu)}\right)^2}{\left(\mu_{D+}^{(\nu)}\right)^4}\mathcal{V}_{DD+,\nu,p}(h_n) - \frac{2\mu_{J+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^3}\mathcal{V}_{JD+,\nu,p}(h_n)$$

$$\text{(B.127)}$$

$$\mathrm{V}_{22+,\nu,p}(h_n) \equiv \frac{1}{\left(\mu_{G+}^{(\nu)}\right)^2}\mathcal{V}_{KK+,\nu,p}(h_n) + \frac{\left(\mu_{K+}^{(\nu)}\right)^2}{\left(\mu_{G+}^{(\nu)}\right)^4}\mathcal{V}_{GG+,\nu,p}(h_n) - \frac{2\mu_{K+}^{(\nu)}}{\left(\mu_{G+}^{(\nu)}\right)^3}\mathcal{V}_{KG+,\nu,p}(h_n)$$

$$\text{(B.128)}$$

and

$$\mathrm{V}_{12+,\nu,p}(h_n) \equiv \frac{1}{\mu_{D+}^{(\nu)}\mu_{G+}^{(\nu)}}\mathcal{V}_{JK+,\nu,p}(h_n) - \frac{\mu_{K+}^{(\nu)}}{\mu_{D+}^{(\nu)}\left(\mu_{G+}^{(\nu)}\right)^2}\mathcal{V}_{JG+,\nu,p}(h_n)$$

$$- \frac{\mu_{J+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2\mu_{G+}^{(\nu)}}\mathcal{V}_{DK+,\nu,p}(h_n) + \frac{\mu_{J+}^{(\nu)}\mu_{K+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\mu_{G+}^{(\nu)}\right)^2}\mathcal{V}_{DG+,\nu,p}(h_n) \quad \text{(B.129)}$$

Symmetrically, the conditional variance of the linearized estimator $\widetilde{\pi}_{-,\nu,p}(h_n)$ is

$$\mathrm{V}_{-,\nu,p}(h_n) \equiv \mathbb{V}\left[\widetilde{\pi}_{-,\nu,p}(h_n)\,\middle|\,\boldsymbol{R}\right] = \mathrm{V}_{11-,\nu,p}(h_n) + \mathrm{V}_{22-,\nu,p}(h_n) - 2\mathrm{V}_{12-,\nu,p}(h_n) \quad \text{(B.130)}$$

where

$$\mathrm{V}_{11-,\nu,p}(h_n) \equiv \frac{1}{\left(\mu_{D-}^{(\nu)}\right)^2}\mathcal{V}_{JJ-,\nu,p}(h_n) + \frac{\left(\mu_{J-}^{(\nu)}\right)^2}{\left(\mu_{D-}^{(\nu)}\right)^4}\mathcal{V}_{DD-,\nu,p}(h_n) - \frac{2\mu_{J-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^3}\mathcal{V}_{JD-,\nu,p}(h_n)$$

$$\text{(B.131)}$$

$$\mathrm{V}_{22-,\nu,p}(h_n) \equiv \frac{1}{\left(\mu_{G-}^{(\nu)}\right)^2}\mathcal{V}_{KK-,\nu,p}(h_n) + \frac{\left(\mu_{K-}^{(\nu)}\right)^2}{\left(\mu_{G-}^{(\nu)}\right)^4}\mathcal{V}_{GG-,\nu,p}(h_n) - \frac{2\mu_{K-}^{(\nu)}}{\left(\mu_{G-}^{(\nu)}\right)^3}\mathcal{V}_{KG-,\nu,p}(h_n)$$

$$\text{(B.132)}$$

and

$$V_{12-,\nu,p}(h_n) \equiv \frac{1}{\mu_{D-}^{(\nu)}\mu_{G+}^{(\nu)}}\mathcal{V}_{JK-,\nu,p}(h_n) - \frac{\mu_{K-}^{(\nu)}}{\mu_{D-}^{(\nu)}\left(\mu_{G-}^{(\nu)}\right)^2}\mathcal{V}_{JG-,\nu,p}(h_n)$$

$$- \frac{\mu_{J-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^2\mu_{G-}^{(\nu)}}\mathcal{V}_{DK-,\nu,p}(h_n) + \frac{\mu_{J-}^{(\nu)}\mu_{K-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\mu_{G+}^{(\nu)}\right)^2}\mathcal{V}_{DG-,\nu,p}(h_n) \qquad \text{(B.133)}$$

B.5.4 Definitions of Bias-Corrected Estimators

The bias-corrected estimators for $\pi_{+,\nu}$ and $\pi_{-,\nu}$ are defined as

$$\widehat{\pi}_{+,\nu,p,q}^{\text{BC}}(h_n,b_n) \equiv \widehat{\pi}_{+,\nu,p}(h_n) - h^{p+1-\nu}\widehat{B}_{+,\nu,p,q}(h_n,b_n) \qquad \text{(B.134)}$$

$$\widehat{\pi}_{-,\nu,p,q}^{\text{BC}}(h_n,b_n) \equiv \widehat{\pi}_{-,\nu,p}(h_n) - h^{p+1-\nu}\widehat{B}_{-,\nu,p,q}(h_n,b_n) \qquad \text{(B.135)}$$

where $\widehat{B}_{+,\nu,p,q}(h_n,b_n)$ and $\widehat{B}_{-,\nu,p,q}(h_n,b_n)$ are estimators for the bias of $\widetilde{\pi}_{+,\nu,p}(h_n)$ and $\widetilde{\pi}_{-,\nu,p}(h_n)$, respectively. Specifically,

$$\widehat{B}_{+,\nu,p,q}(h_n,b_n) \equiv \widehat{B}_{1+,\nu,p,q}(h_n,b_n) - \widehat{B}_{2+,\nu,p,q}(h_n,b_n) \qquad \text{(B.136)}$$

$$\widehat{B}_{-,\nu,p,q}(h_n,b_n) \equiv \widehat{B}_{1-,\nu,p,q}(h_n,b_n) - \widehat{B}_{2-,\nu,p,q}(h_n,b_n) \qquad \text{(B.137)}$$

with

$$\widehat{B}_{1+,\nu,p,q}(h_n,b_n) \equiv \frac{1}{\widehat{\mu}_{D+,p}^{(\nu)}(h_n)}\widehat{B}_{J+,\nu,p,q}(h_n,b_n) - \frac{\widehat{\mu}_{J+,p}^{(\nu)}(h_n)}{\left(\widehat{\mu}_{D+,p}^{(\nu)}(h_n)\right)^2}\widehat{B}_{D+,\nu,p,q}(h_n,b_n) \quad \text{(B.138)}$$

$$\widehat{B}_{2+,\nu,p,q}(h_n,b_n) \equiv \frac{1}{\widehat{\mu}_{G+,p}^{(\nu)}(h_n)}\widehat{B}_{K+,\nu,p,q}(h_n,b_n) - \frac{\widehat{\mu}_{K+,p}^{(\nu)}(h_n)}{\left(\widehat{\mu}_{G+,p}^{(\nu)}(h_n)\right)^2}\widehat{B}_{G+,\nu,p,q}(h_n,b_n) \quad \text{(B.139)}$$

$$\widehat{B}_{1-,\nu,p,q}(h_n,b_n) \equiv \frac{1}{\widehat{\mu}_{D-,p}^{(\nu)}(h_n)}\widehat{B}_{J-,\nu,p,q}(h_n,b_n) - \frac{\widehat{\mu}_{J-,p}^{(\nu)}(h_n)}{\left(\widehat{\mu}_{D-,p}^{(\nu)}(h_n)\right)^2}\widehat{B}_{D-,\nu,p,q}(h_n,b_n) \quad \text{(B.140)}$$

$$\widehat{B}_{2-,\nu,p,q}(h_n,b_n) \equiv \frac{1}{\widehat{\mu}_{G-,p}^{(\nu)}(h_n)}\widehat{B}_{K-,\nu,p,q}(h_n,b_n) - \frac{\widehat{\mu}_{K-,p}^{(\nu)}(h_n)}{\left(\widehat{\mu}_{G-,p}^{(\nu)}(h_n)\right)^2}\widehat{B}_{G-,\nu,p,q}(h_n,b_n) \quad \text{(B.141)}$$

In addition, the bias-corrected linearized estimators for $\pi_{+,\nu}$ and $\pi_{-,\nu}$ are defined as

$$\widetilde{\pi}_{+,\nu,p,q}^{\text{BC}}(h_n,b_n) \equiv \widetilde{\pi}_{+,\nu,p}(h_n) - h^{p+1-\nu}\breve{B}_{+,\nu,p,q}(h_n,b_n) \qquad \text{(B.142)}$$

$$\widetilde{\pi}_{-,\nu,p,q}^{\text{BC}}(h_n,b_n) \equiv \widetilde{\pi}_{-,\nu,p}(h_n) - h^{p+1-\nu}\breve{B}_{-,\nu,p,q}(h_n,b_n) \qquad \text{(B.143)}$$

where

$$\check{B}_{+,\nu,p,q}\left(h_n, b_n\right) \equiv \check{B}_{1+,\nu,p,q}\left(h_n, b_n\right) - \check{B}_{2+,\nu,p,q}\left(h_n, b_n\right) \tag{B.144}$$

$$\check{B}_{-,\nu,p,q}\left(h_n, b_n\right) \equiv \check{B}_{1-,\nu,p,q}\left(h_n, b_n\right) - \check{B}_{2-,\nu,p,q}\left(h_n, b_n\right) \tag{B.145}$$

with

$$\check{B}_{1+,\nu,p,q}\left(h_n, b_n\right) \equiv \frac{1}{\mu_{D+}^{(\nu)}}\widehat{B}_{J+,\nu,p,q}\left(h_n, b_n\right) - \frac{\mu_{J+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}\widehat{B}_{D+,\nu,p,q}\left(h_n, b_n\right) \tag{B.146}$$

$$\check{B}_{2+,\nu,p,q}\left(h_n, b_n\right) \equiv \frac{1}{\mu_{G+}^{(\nu)}}\widehat{B}_{K+,\nu,p,q}\left(h_n, b_n\right) - \frac{\mu_{K+}^{(\nu)}}{\left(\mu_{G+}^{(\nu)}\right)^2}\widehat{B}_{G+,\nu,p,q}\left(h_n, b_n\right) \tag{B.147}$$

$$\check{B}_{1-,\nu,p,q}\left(h_n, b_n\right) \equiv \frac{1}{\mu_{D-}^{(\nu)}}\widehat{B}_{J-,\nu,p,q}\left(h_n, b_n\right) - \frac{\mu_{J-}^{(\nu)}}{\left(\mu_{D-}^{(\nu)}\right)^2}\widehat{B}_{D-,\nu,p,q}\left(h_n, b_n\right) \tag{B.148}$$

$$\check{B}_{2-,\nu,p,q}\left(h_n, b_n\right) \equiv \frac{1}{\mu_{G-}^{(\nu)}}\widehat{B}_{K-,\nu,p,q}\left(h_n, b_n\right) - \frac{\mu_{K-}^{(\nu)}}{\left(\mu_{G-}^{(\nu)}\right)^2}\widehat{B}_{G-,\nu,p,q}\left(h_n, b_n\right) \tag{B.149}$$

B.5.5   Variance of Bias-Corrected Linearized Estimators

As in Theorem B.2, suppose that Assumptions 8 and 9 hold with $\delta \geq p+2$, and $n \min\{h_n, b_n\} \to \infty$. Then, if $\max\{h_n, b_n\} \to 0$, the conditional variance of the bias-corrected linearized estimator $\widetilde{\pi}_{+,\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right)$ is

$$\begin{aligned}
V_{+,\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right) &\equiv \mathbb{V}\left[\widetilde{\pi}_{+,\nu,p,q}^{\mathrm{BC}}\left(h_n, b_n\right) \middle| \boldsymbol{R}\right] \\
&= V_{+,\nu,p}\left(h_n\right) + h_n^{2(p+1-\nu)}V_{+,p+1,q}\left(b_n\right)\frac{\mathcal{B}_{+,\nu,p,p+1}^2\left(h_n\right)}{(p+1)!^2} \\
&\quad - 2h_n^{p+1-\nu}C_{+,\nu,p,q}\left(h_n, b_n\right)\frac{\mathcal{B}_{+,\nu,p,p+1}\left(h_n\right)}{(p+1)!}
\end{aligned} \tag{B.150}$$

where

$$\begin{aligned}
&C_{+,\nu,p,q}\left(h_n, b_n\right) \\
&\equiv C_{11+,\nu,p,q}\left(h_n, b_n\right) - C_{12+,\nu,p,q}\left(h_n, b_n\right) - C_{21+,\nu,p,q}\left(h_n, b_n\right) + C_{22+,\nu,p,q}\left(h_n, b_n\right)
\end{aligned} \tag{B.151}$$

with

$$C_{11+,\nu,p,q}\left(h_n, b_n\right) \equiv \mathbb{C}\left[\frac{1}{\mu_{D+}^{(\nu)}}\widehat{\mu}_{J+,p}^{(\nu)}\left(h_n\right) - \frac{\mu_{J+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}\widehat{\mu}_{D+,p}^{(\nu)}\left(h_n\right), \check{B}_{1+,\nu,p,q}\left(h_n, b_n\right)\right] \tag{B.152}$$

106

$$C_{12+,\nu,p,q}(h_n, b_n) \equiv \mathbb{C}\left[\frac{1}{\mu_{D+}^{(\nu)}}\widehat{\mu}_{J+,p}^{(\nu)}(h_n) - \frac{\mu_{J+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2}\widehat{\mu}_{D+,p}^{(\nu)}(h_n), \check{B}_{2+,\nu,p,q}(h_n, b_n)\right] \quad \text{(B.153)}$$

$$C_{21+,\nu,p,q}(h_n, b_n) \equiv \mathbb{C}\left[\frac{1}{\mu_{G+}^{(\nu)}}\widehat{\mu}_{K+,p}^{(\nu)}(h_n) - \frac{\mu_{K+}^{(\nu)}}{\left(\mu_{G+}^{(\nu)}\right)^2}\widehat{\mu}_{G+,p}^{(\nu)}(h_n), \check{B}_{1+,\nu,p,q}(h_n, b_n)\right] \quad \text{(B.154)}$$

$$C_{22+,\nu,p,q}(h_n, b_n) \equiv \mathbb{C}\left[\frac{1}{\mu_{G+}^{(\nu)}}\widehat{\mu}_{K+,p}^{(\nu)}(h_n) - \frac{\mu_{K+}^{(\nu)}}{\left(\mu_{G+}^{(\nu)}\right)^2}\widehat{\mu}_{G+,p}^{(\nu)}(h_n), \check{B}_{2+,\nu,p,q}(h_n, b_n)\right] \quad \text{(B.155)}$$

or, equivalently,

$$C_{11+,\nu,p,q}(h_n, b_n) = \frac{1}{\left(\mu_{D+}^{(\nu)}\right)^2}\mathcal{C}_{JJ+,\nu,p,q}(h_n, b_n) - \frac{\mu_{J+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^3}\mathcal{C}_{JD+,\nu,p,q}(h_n, b_n)$$

$$- \frac{\mu_{J+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^3}\mathcal{C}_{DJ+,\nu,p,q}(h_n, b_n) + \frac{\left(\mu_{J+}^{(\nu)}\right)^2}{\left(\mu_{D+}^{(\nu)}\right)^4}\mathcal{C}_{DD+,\nu,p,q}(h_n, b_n) \quad \text{(B.156)}$$

$$C_{12+,\nu,p,q}(h_n, b_n) = \frac{1}{\mu_{D+}^{(\nu)}\mu_{G+}^{(\nu)}}\mathcal{C}_{JK+,\nu,p,q}(h_n, b_n) - \frac{\mu_{K+}^{(\nu)}}{\mu_{D+}^{(\nu)}\left(\mu_{G+}^{(\nu)}\right)^2}\mathcal{C}_{JG+,\nu,p,q}(h_n, b_n)$$

$$- \frac{\mu_{J+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2\mu_{G+}^{(\nu)}}\mathcal{C}_{DK+,\nu,p,q}(h_n, b_n) + \frac{\mu_{J+}^{(\nu)}\mu_{K+}^{(\nu)}}{\left(\mu_{D+}^{(\nu)}\right)^2\left(\mu_{G+}^{(\nu)}\right)^2}\mathcal{C}_{DG+,\nu,p,q}(h_n, b_n)$$

$$\text{(B.157)}$$

$$C_{21+,\nu,p,q}(h_n, b_n) = \frac{1}{\mu_{G+}^{(\nu)}\mu_{D+}^{(\nu)}}\mathcal{C}_{KJ+,\nu,p,q}(h_n, b_n) - \frac{\mu_{K+}^{(\nu)}}{\mu_{G+}^{(\nu)}\left(\mu_{D+}^{(\nu)}\right)^2}\mathcal{C}_{KD+,\nu,p,q}(h_n, b_n)$$

$$- \frac{\mu_{J+}^{(\nu)}}{\left(\mu_{G+}^{(\nu)}\right)^2\mu_{D+}^{(\nu)}}\mathcal{C}_{GJ+,\nu,p,q}(h_n, b_n) + \frac{\mu_{K+}^{(\nu)}\mu_{J+}^{(\nu)}}{\left(\mu_{G+}^{(\nu)}\right)^2\left(\mu_{D+}^{(\nu)}\right)^2}\mathcal{C}_{GD+,\nu,p,q}(h_n, b_n)$$

$$\text{(B.158)}$$

$$C_{22+,\nu,p,q}(h_n, b_n) = \frac{1}{\left(\mu_{G+}^{(\nu)}\right)^2}\mathcal{C}_{KK+,\nu,p,q}(h_n, b_n) - \frac{\mu_{K+}^{(\nu)}}{\left(\mu_{G+}^{(\nu)}\right)^3}\mathcal{C}_{KG+,\nu,p,q}(h_n, b_n)$$

$$- \frac{\mu_{K+}^{(\nu)}}{\left(\mu_{G+}^{(\nu)}\right)^3}\mathcal{C}_{GK+,\nu,p,q}(h_n, b_n) + \frac{\left(\mu_{K+}^{(\nu)}\right)^2}{\left(\mu_{G+}^{(\nu)}\right)^4}\mathcal{C}_{GG+,\nu,p,q}(h_n, b_n) \quad \text{(B.159)}$$

Symmetrically, the conditional variance of the bias-corrected linearized estimator $\widetilde{\pi}^{\text{BC}}_{-,\nu,p,q}(h_n, b_n)$ is

$$
\begin{aligned}
\text{V}^{\text{BC}}_{-,\nu,p,q}(h_n, b_n) &\equiv \mathbb{V}\left[\widetilde{\pi}^{\text{BC}}_{-,\nu,p,q}(h_n, b_n)\,\Big|\,\boldsymbol{R}\right] \\
&= \text{V}_{-,\nu,p}(h_n) + h_n^{2(p+1-\nu)}\text{V}_{-,p+1,q}(b_n)\frac{\mathcal{B}^2_{-,\nu,p,p+1}(h_n)}{(p+1)!^2} \\
&\quad - 2h_n^{p+1-\nu}\text{C}_{-,\nu,p,q}(h_n, b_n)\frac{\mathcal{B}_{-,\nu,p,p+1}(h_n)}{(p+1)!}
\end{aligned}
\tag{B.160}
$$

where

$$
\begin{aligned}
&\text{C}_{-,\nu,p,q}(h_n, b_n) \\
&\equiv \text{C}_{11-,\nu,p,q}(h_n, b_n) - \text{C}_{12-,\nu,p,q}(h_n, b_n) - \text{C}_{21-,\nu,p,q}(h_n, b_n) + \text{C}_{22-,\nu,p,q}(h_n, b_n)
\end{aligned}
\tag{B.161}
$$

with the definitions of $\text{C}_{11-,\nu,p,q}(h_n, b_n)$, $\text{C}_{12-,\nu,p,q}(h_n, b_n)$, $\text{C}_{21-,\nu,p,q}(h_n, b_n)$, and $\text{C}_{22-,\nu,p,q}(h_n, b_n)$ mirroring the positive case.

B.5.6 Asymptotic Distribution of Bias-Corrected Estimators

As in part (D) of Theorem B.2, provided that $n \min\{h_n^{2p+3}, b_n^{2p+3}\}\max\left\{h_n^2, b_n^{2(q-p)}\right\} \to 0$ and $\kappa \max\{h_n, b_n\} < \kappa_0$, then

$$
T^{\text{RBC}}_+(h_n, b_n) \equiv \frac{\widehat{\pi}^{\text{BC}}_{+,\nu,p,q}(h_n, b_n) - \pi_{+,\nu}}{\sqrt{\text{V}^{\text{BC}}_{+,\nu,p,q}(h_n, b_n)}} \xrightarrow{d} \mathcal{N}(0, 1)
\tag{B.162}
$$

$$
T^{\text{RBC}}_-(h_n, b_n) \equiv \frac{\widehat{\pi}^{\text{BC}}_{-,\nu,p,q}(h_n, b_n) - \pi_{-,\nu}}{\sqrt{\text{V}^{\text{BC}}_{-,\nu,p,q}(h_n, b_n)}} \xrightarrow{d} \mathcal{N}(0, 1)
\tag{B.163}
$$

These statistics can be used to test whether $\pi_{+,\nu} = 0$ and $\pi_{-,\nu} = 0$, respectively. Failure to reject any of these null hypotheses may provide suggestive evidence in favor of the plausibility of the common trends assumption.