# Regression Discontinuity Design with Spillovers*

Eric Auerbach

Department of Economics

Northwestern University

eric.auerbach@northwestern.edu

Yong Cai

Department of Economics

University of Wisconsin-Madison

yong.cai@wisc.edu

Ahnaf Rafi

Department of Economics

University of Virginia

ahnafrafi@virginia.edu

September 16, 2025

### Abstract

This paper studies regression discontinuity designs (RDD) when linear-in-means spill-overs occur between units that are close in their running variable. We show that the RDD estimand depends on the ratio of two terms: (1) the radius over which spillovers occur and (2) the choice of bandwidth used for the local linear regression. RDD estimates direct treatment effect when radius is of larger order than the bandwidth and total treatment effect when radius is of smaller order than the bandwidth. When the two are of similar order, the RDD estimand need not have a causal interpretation. To recover direct and spillover effects in the intermediate regime, we propose to incorporate estimated spillover terms into local linear regression. Our estimator is consistent and asymptotically normal and we provide bias-aware confidence intervals for direct treatment effects and spillovers. In the setting of Gonzalez (2021), we detect endogenous spillovers in voter fraud during the 2009 Afghan Presidential election. We also clarify when the donut-hole design addresses spillovers in RDD.

---

# 1    Introduction

Regression discontinuity design (RDD) is a popular method for causal inference and policy evaluation, particularly in settings where experimental manipulation is not possible. However, in the continuity-based framework of Hahn et al. (2001), identification of treatment effect at the cutoff relies on the Stable Unit Treatment Value Assumption (SUTVA). This requires each unit's potential outcomes to be unaffected by the treatment status of other units and may be unrealistic: in many economic environments, the variables of interest are equilibrium objects shaped by strategic interactions, making SUTVA violations natural. Although RDD is known to identify treatment effects local to the cutoff under SUTVA, this need not hold once SUTVA is violated.

Following Manski (1993), the literature on social interactions typically emphasizes two types of SUTVA violations: exogenous spillovers (also known as contextual effects) and endogenous spillovers. Exogenous spillovers occur when the outcome of an agent depends directly on the treatment status of their neighbors. On the other hand, endogenous spillovers arise when the outcome of an agent depends on the outcome of their neighbors. For a concrete example, consider Gonzalez (2021), which studies the effect of cell phone coverage on voter fraud during the 2009 Afghan presidential election. Their spatial RDD compares polling stations on either side of the boundary of cell phone coverage areas and finds that cell phone coverage reduced fraud. However, the author is also explicitly concerned about potential spillovers. For example, suppose coverage at a polling station reduces fraud by allowing voters to report suspicious behavior to the electoral commission. Voters at non-covered stations could potentially walk to covered stations to report fraud, giving rise to exogenous spillovers. Alternatively, corrupt politicians might strategically allocate fraud to areas with less monitoring. In this way, higher fraud in one polling station will reduce the need to commit fraud in neighboring polling stations. This is a form of endogenous spillovers. Because spillovers appear plausible in many settings, it is important to understand

their effects on the estimand of RDD.

This paper studies RDD in the presence of spillovers. To do so, we extend the framework of Hahn et al. (2001) to incorporate exogenous and endogenous spillovers that occur along the running variable. We make two important modeling assumptions. Firstly, we assume that the outcome of a given unit depends linearly on the mean treatment status and mean outcome of their neighbors. The linear-in-means specification characterizes a large literature on peer effects (see e.g. Manski 1993; Bramoullé et al. 2009; De Giorgi et al. 2010; Goldsmith-Pinkham and Imbens 2013; De Paula et al. 2024) and spatial autoregression models (e.g. Cliff and Ord 1973; Kelejian and Prucha 1998; 2010; Lee 2004; 2007). Secondly, we assume that the running variable which defines the RDD also determines the social interactions in the linear-in-means model. Specifically, units with similar values of the running variables are also assumed to be neighbors. When the running variable is geographical coordinates, this model captures interaction between units that are close in space. In the classic Thistlethwaite and Campbell (1960), which studies the effect of scholarships on academic achievements, the running variable is test scores. Our model then captures the idea that students with similar test scores exert peer effects on one another, for example by forming study groups.

We make two main contributions. Firstly, we show that the estimand of RDD depends on the ratio of two terms: (1) the radius over which spillovers occur and (2) the bandwidth used for local linear regression. Specifically, RDD estimates direct treatment effect at the boundary when the radius is of larger order than the bandwidth. This is the effect on a unit at the boundary when we switch their treatment status from control to treatment, keeping treatment assignment fixed for all other units. When radius is of smaller order than the bandwidth, RDD instead estimates total treatment effect, which is the effect on the unit at the boundary when we switch the treatment status of the entire population from control into treatment. In the regime where radius is of similar order as the bandwidth, the RDD estimand is a mixture of

the above effects and lacks a clear interpretation. We argue, however, that this is the more realistic regime wherein the asymptotic approximation captures more features of the finite sample distribution.

As such, our second contribution is a proposal for recovering direct and spillover effects in the intermediate regime. We do so by incorporating estimated spillover terms into local linear regression, in a procedure we call the local spillover regression (LSR). We show that LSR, which is essentially the local analog of peer effects regressions, leads to consistent estimators which are asymptotically normally distributed. We also adapt the method of Armstrong and Kolesár (2020) to construct bias-aware confidence intervals for direct and spillover effects.

Simulation results show that LSR performs well relative to local linear regression with naively chosen bandwidths. The adaptive bandwidth rules of Calonico et al. (2014) and Armstrong and Kolesár (2020) lead to estimators that are more robust to spillovers, but our method can achieve lower MSE and better coverage particularly in settings with intermediate amounts of spillovers. We apply LSR to the setting of Gonzalez (2021) and find evidence of positive endogenous spillovers in electoral fraud. The implied social multiplier may be relevant for cost-benefit analysis for fraud deterrence policies and may provide motivation for network-based interventions. Along the way, we also discuss the estimands of the popular RD donut and highlight settings under which it recovers either direct or total treatment effects.

This paper contributes to the vast literature on RDD (see Cattaneo and Titiunik 2022 for a recent review). Empirical researchers have long been concerned with spillovers in RDD, with papers such as Jardim et al. (2022) arguing against the use of spatial discontinuity design for policy evaluation. However, to our knowledge, prior theoretical work on this issue is limited to Aronow et al. (2017), which conducts their analysis under the local randomization framework of Cattaneo et al. (2015), finding that RDD always recovers a weighted average of direct treatment effect. We consider RDD under the continuity-based approach of Hahn et al. (2001) and find

4

that the estimand can exhibit more complex behavior, particularly when neighborhoods are determined by the running variable. Since the first version of this paper was posted, Dal Torrione et al. (2024) and Borusyak and Kolerman-Shemer (2024) have also studied RDD in the presence of spillovers. Dal Torrione et al. (2024) takes the continuity-based approach, assuming that the running variable and the variable determining neighborhood structure have a continuous joint density. They find that RDD recovers a weighted average of direct treatment effect, drawing a connection to multi-score RDD. Our paper focuses on the case where neighborhoods are completely determined by the running variable and is unique in having spillovers that appear in the asymptotic approximation, provided that radius and bandwidth are scaled appropriately. This paper also differs from Aronow et al. (2017) and Dal Torrione et al. (2024) in considering endogenous spillovers, which arises in many settings of interest to economists. Finally, we note that the aforementioned papers, as well as ours, focus on spillovers between units within an RD. Borusyak and Kolerman-Shemer (2024) studies the aggregation of multiple RDDs and considers effects of spillovers across designs.

This paper joins a burgeoning body of work that considers violations of SUTVA under various research designs, such as experiments (e.g. Hudgens and Halloran 2008; Aronow and Samii 2017; Sävje et al. 2021; Hu et al. 2022; Leung 2022a; Li and Wager 2022; Gao and Ding 2023; Vazquez-Bare 2023b; Auerbach et al. 2025a), differences-in-differences (e.g. Clarke 2017; Butts 2021; Xu 2023), synthetic control (Cao and Dowd 2019), instrumental variables (Sobel 2006; Vazquez-Bare 2023a) and other observational settings (Forastiere et al. 2020). RDD poses unique challenges relative to these other settings because it concerns parameters that are local to the cutoff and estimation is nonparametric. Additionally, we consider endogenous spillovers, which has received relatively less attention in this literature, though exceptions include, in the experiment context, Leung (2022a); Munro et al. (2021); Li et al. (2023); Munro (2023) and Faridani and Niehaus (2024). In the presence of endogenous spillovers, a

unit's outcome may depend on the treatment status of the entire population, even if spillovers are assumed to have a small radius. The interaction of this dependence with the discontinuity of the observed outcome function poses novel technical challenges that we address.

The rest of the paper is organized as follows. Section 2 describes our econometric framework. Section 3 characterizes the estimand of local linear regression under spillovers. Section 3.4 in particular considers the use of donut-hole RDs. Section 4 presents the local spillovers regression, our proposal for recovering direct treatment and spillover effects. Section 5 applies our method to the setting of Gonzalez (2021). Section 6 concludes. Proofs are contained in the Appendix. Simulation results and the auxiliary lemmas used in the proofs are available in the Supplemental Appendix.

The remainder of this paper uses the following notation. We write $A_n \gg B_n$ if $A_n/B_n \to \infty$, $A_n \propto B_n$ if $A_n/B_n \to c$ where $0 < c < \infty$, and $A_n \ll B_n$ if $A_n/B_n \to 0$. Let $\iota$ and $\mathbf{0}$ be constant functions that take values 1 and 0 on $[-1, 1]$ respectively.

## 2  Econometric Framework

In this section, we introduce the econometric framework for studying RDD with spillovers. Section 2.1 presents our model for potential outcomes. Section 2.2 defines the parameters of interest. Section 2.3 discusses our assumed neighborhood structure. Section 2.4 presents the sampling process.

### 2.1  Potential Outcomes

Consider a continuum of agents that are indexed by their coordinates $z \in \mathcal{Z} = [-1, 1]$. For convenience, we will assume that agents are uniformly distributed according to $F = \text{Uniform}(\mathcal{Z})$, so that their density with respect to the Lebesgue measure is $f(z) = \frac{1}{2}$. Our results extend easily to any density bounded away from 0 at the cutoff, although our expressions for the bias of RDD need not apply in the more

general setting.

We work in the usual potential outcomes framework with binary treatment, where the observed outcome at $z$ satisfies

$$Y_d(z) = d(z)Y_d^+(z) + (1 - d(z))Y_d^-(z) .$$

Here, $d : \mathcal{Z} \to \{0, 1\}$ is the treatment assignment function. $Y_d^+(z)$ and $Y_d^-(z)$ are the potential outcomes under treatment and control respectively. Because of the treatment assignment rule in RDD, defined below, we will denote quantities related to treatment with "+" and those related to control with "−".

Potential outcomes are indexed by $d$ because as a result of spillovers, they may depend on the entire treatment assignment function. Let every agent $z$ have the set of relevant neighbors $R(z) \subset \mathcal{Z}$ with measure $|R(z)|$. These are agents whose realized outcomes affect $z$. Let potential outcomes be defined as follows:

**Assumption 1** (Potential Outcomes). *For a given treatment assignment function $d : \mathcal{Z} \to [0, 1]$ and neighborhood structure $R : \mathcal{Z} \to \mathcal{P}(\mathcal{Z})$, the potential outcome of agent $z$ under treatment $d$ is:*

$$Y_d^+(z) = m^+(z) + \delta(z)\mu_d(z) + \gamma(z)\nu_d(z)$$
$$Y_d^-(z) = m^-(z) + \delta(z)\mu_d(z) + \gamma(z)\nu_d(z)$$

*where*

$$\mu_d(z) = \frac{1}{|R(z)|} \int_{R(z)} Y_d(u)f(u)\, du$$
$$\nu_d(z) = \frac{1}{|R(z)|} \int_{R(z)} d(u)f(u)\, du .$$

*Furthermore, let $m^+(z)$ and $m^-(z)$ be Lipschitz continuous on $\mathcal{Z} = [-1, 1]$ with Lipschitz constant $C$. Let $\delta(z)$ be continuously differentiable with $|\frac{d\delta}{dz}| < C_\delta$, $\sup_{z \in \mathcal{Z}} |\delta(z)| \leq$*

$\bar{\delta}$, $\sup_{z_1, z_2 \in \mathcal{Z}} |\delta(z_1) - \delta(z_2)| \le \bar{\delta}$, $\bar{\delta} < 1$. *Finally, let $\gamma(z)$ be continuously differentiable with $|\frac{d\gamma}{dz}| < C_\gamma$.*

The above model extends the continuity-based framework of Hahn et al. (2001) to include two sources of spillovers, both of which are linear-in-means. The exogenous spillover term is $\gamma(z)\nu_d(z)$, where $\nu_d(z)$ is the mean treatment status of $z$'s neighbors, and $\gamma(z)$ is the effect of this term on $z$'s outcome. The endogenous spillover term is $\delta(z)\mu_d(z)$, where $\mu_d(z)$ is the mean outcome of $z$'s neighbors, and $\delta(z)$ is the effect of this term. Spillovers are assumed to enter the treated and control potential outcomes in the same way. As will become clear in Section 2.2, this leads to a particularly simple formula for the parameters of interest.

The linear-in-means assumption is potentially restrictive. However, it is the workhorse model in the peer effects (e.g. Manski 1993; Bramoullé et al. 2009; De Giorgi et al. 2010; Goldsmith-Pinkham and Imbens 2013; De Paula et al. 2024) and spatial autoregression (e.g. Cliff and Ord 1973; Kelejian and Prucha 1998; 2010; Lee 2004; 2007) literatures. We consider it to be a reasonable first step for analyzing endogenous spillovers, a task which is challenging in many settings. The linear-in-means assumption is in fact stronger than necessary for our characterization of the estimands. In particular, the qualitative results in Section 3 does not require linearity in the effects of spillovers, and allows neighbors to have different weights in treated and control outcomes. The main requirement being that neighborhoods have approximately bounded support. However, as will become clear in Section 4, the specific form of spillovers needs to be assumed in order to recover target parameters in our preferred asymptotic regime. We therefore focus on the above model. Nonetheless, it is more general than the standard peer effects model in that the effects of spillovers, $\delta$ and $\gamma$, is allowed to vary with $z$.

The remaining conditions in Assumption 1 concerns continuity of the functions $m^+, m^-, \delta$ and $\gamma$ as well as the boundedness of $\delta$. Continuity of $m^+$ and $m^-$ around the cutoff is the main identifying assumption in RDD. Standard RDD requires only that

$m^+(z)$ and $m^-(z)$ are continuous at $z = 0$. However, due to spillovers, continuity of potential outcomes also requires continuity of the two functions around the boundaries of $R_n(0)$ as well as the continuity of $\gamma$ and $\delta$. Since $R_n(0)$ is fixed under some of the regimes we consider, we assume continuity over $\mathcal{Z}$ for simplicity. We also slightly strengthen the condition to Lipschitz continuity. To ensure that $Y_d^+$ and $Y_d^-$ are well-defined, we require $\sup_{z \in \mathcal{Z}} |\delta(z)| < 1$, as well as for $\sup_{z_1, z_2 \in \mathcal{Z}} |\delta(z_1) - \delta(z_2)| \leq 1$.

## 2.2 Parameters of Interest

In a setting with spillovers, any two treatment assignment functions $d$ and $d'$ can give rise to a treatment effect

$$Y_d(z) - Y_{d'}(z)$$

that is potentially of interest. Following the literature on causal inference with spillovers (see e.g. Hudgens and Halloran 2008), we focus on the following two parameters:

**Definition 1** (Treatment Effects).

- The direct treatment effect at $z = 0$ is

$$\tau_{\mathrm{DIR}} := Y_d^+(0) - Y_d^-(0) = m^+(0) - m^-(0) .$$

- The total treatment effect at $z = 0$ is

$$\tau_{\mathrm{TOT}} := Y_\iota^+(0) - Y_{\mathbf{0}}^-(0) = m^+(0) - m^-(0) + \delta(0)\left(\mu_\iota(0) - \mu_{\mathbf{0}}(0)\right) + \gamma(0) .$$

The direct treatment effect on a given agent is the effect of treatment on their outcomes, keeping the treatment assignment of all other agents unchanged at some $d$. This parameter is useful for thinking about selective rollout of some policy, under which it is plausible to consider only the direct treatment effect, treating the spillovers

9

as fixed. Here, we abuse notation in using $d$ to refer to two different treatment assign functions, although they are identical on $\mathcal{Z} \setminus 0$. In principle, direct treatment effect depends on $d$. In RDD, it is natural to focus on:

**Definition 2** (RDD Treatment Assignment). $d(z) = \mathbf{1}\{z \geq 0\}.$

The above treatment assignment function is the natural reference point for evaluating direct treatment effects in RDD. However, under Assumption 1, the agent at $z = 0$ experiences the same spillover whether or not they are treated for two reasons. Firstly, $\delta(z)$ and $\gamma(z)$ are assumed to be the same regardless of treatment status. Secondly, the agent $z = 0$ is infinitesimal, so that changing their treatment status does not affect $\mu_d(z)$ or $\nu_d(z)$ for any $z \in \mathcal{Z}$. Implicitly, this is a model involving dense network asymptotics. The result of these two factors is that $\tau_{\text{DIR}} = m^+(0) - m^-(0)$ regardless of $d$. Additionally, $\tau_{\text{DIR}}$ is equal to local average treatment effect in the standard setting with no spillovers.

With infinitesimal agents, direct treatment effects are easy to define since we can change the treatment status of one individual but still keep spillovers fixed. Nonetheless, there is a coherent notion of spillovers in this model: if two treatment functions $d$ and $d'$ differ on a set of positive measure, then $\mu_d(z)$ and $\nu_d(z)$ need not be equal to $\mu_{d'}(z)$ and $\nu_{d'}(z)$. This point is also evident in the definition of our next parameter.

The total treatment effect on a given agent is the effect on their outcome when we switch the treatment status of the entire population from control to treatment. This corresponds to the effect on $z = 0$ from a large-scale rollout of treatment. In our notation, the subscripts $\boldsymbol{\iota}$ and $\mathbf{0}$ denote treatment assignments in which everyone and no one is treated respectively. As before, we will focus on the agent at the cutoff, $z = 0$. With total treatment effect, we are evaluating potential outcomes under different treatment assignment functions. Consequently, spillovers no longer cancel out. We remark that $\nu_{\boldsymbol{\iota}}(z) = 1$ and $\nu_{\mathbf{0}}(z) = 0$ so that exogenous spillovers at $z = 0$ is $\gamma(0)$.

## 2.3 Neighborhood Structure

In the presence of spillovers, the neighborhood structure is critical in determining the estimands of RDD. We focus on the case where spillovers occur along the running variable:

**Assumption 2** (Neighborhood).

$$R_n(z) = \{u \, : \, \|z - u\| \leq r_n\} \ .$$

Here, $\|\cdot\|$ is taken to be the Euclidean distance. As such, $z$ is affected only by units whose running variable takes value within $r_n$ of $z$. All other units exert no effect. Together with an i.i.d sampling assumption introduced in the next section, the above assumption implies that the exposure map between sampled units is a random geometric graph (see e.g. Penrose 2003). This is a well-studied model that is commonly used for spillovers and interference (see e.g. Leung 2020).

When neighborhoods are defined by the running variable, nearby units that are comparable in their conditional means $(m^+, m^-)$ can experience spillovers that are not comparable. Consequently, RDD may not recover any meaningful treatment effect parameters, as our analysis in Section 3 shows. This fundamental tension is absent when neighborhoods are defined by another variable, say $w$, that is suitably continuous with respect to $z$. We will refer to such a $w$ as a background variable. Suppose for now that units are defined by their values of $(z, w)$ and that neighborhoods are defined as

$$R_n(z, w) = \{(\tilde{z}, \tilde{w}) \, : \, |\tilde{w} - w| \leq r_n\} \ .$$

Additionally, suppose $z$ and $w$ have a joint density in a neighborhood containing the line $z = 0$. Then, local to the cutoff, we have that $z$ is uncorrelated $w$, so that neighborhoods, and in turn spillovers, are uncorrelated to the RDD treatment assignment. The result is that RDD always recovers the direct treatment effect,

provided that the correlation between $z$ and $w$ remain fixed in the asymptotics.[1]

In light of the above discussion, spillovers in RDD may appear pathological. However, there are many settings in which neighborhoods appear strongly correlated with the running variable. In spatial RDDs, it is often plausible that nearby units interact. For example, Jardim et al. (2022) studies the effect of minimum wage policies on labor market outcomes at the census tract level. However, it is clear that census tracts compete in the same labor market as all other census tracts within a reasonable commute. Gonzalez (2021) provides another example based on cell phone coverage and electoral fraud which we take up in greater detail in Section 5. Leung (2022b) provides many more examples. The same is true with test score RDDs, such as the classic Thistlethwaite and Campbell (1960) which studies the effect of merit scholarships on academic achievements. Here, scholarships are assigned based on test score cutoffs and it is plausible that students who are close in test scores interact. This could happen because of homophily, because students take classes and therefore socialize with those of similar abilities, or because they compete for the same set of career opportunities. In instances like the above, our model is a useful starting point for understanding the effect of spillovers.

## 2.4 Sampling

Having defined a sequence of models for generating potential outcomes, we close this section by considering sampling. We will assume that our data takes the following form:

**Assumption 3.** *Suppose we observe the i.i.d. sample $\{(Y_i, Z_i)\}_{i=1}^n$, where $Z_i \sim F$,*

$$Y_i = Y_d(Z_i) + \varepsilon_i \quad , \quad E[\varepsilon_i | Z_i] = 0$$

---

[1]The first version of this paper presented the above intuition for spillovers on background variables (Auerbach et al. 2024, Section 3.3.3). Dal Torrione et al. (2024) proves this result in a related setting, drawing a connection to multi-score RDD. Also see the discussion after Theorem 1.

and $d(z) = \mathbf{1}\{z \geq 0\}$ *is the RDD treatment assignment function. Moreover, suppose that* $E[\varepsilon_i^2 | Z_i = z] \leq \bar{\sigma}^2$ *for all* $z \in \mathcal{Z}$ *and that* $\mathbb{E}[\varepsilon^2 \mid Z = z]$ *is continuous in a neighborhood of* $0$.

In words, the data-generating process operates as follows. A continuum of agents interact and their outcomes are determined by the values of their running variables as well as spillovers. The econometrician then samples agents randomly and observes their outcomes, possibly with an error that has 0 conditional mean. We also assume that the error has conditional variance that is continuous in $z$ – a standard assumption. Our framework resembles De Paula et al. (2018) in defining interactions to occur in the population.

# 3 Local Linear Regression in the Presence of Spillovers

In this section, we show that the local linear regression estimator may not recover meaningful treatment effect parameters in the presence of spillovers, necessitating proposals such as that in Section 4. Section 3.1 defines the local linear regression estimator and Section 3.2 describes its estimands in the presence of spillovers. Our results suggest that the ratio of $r_n/h_n$ is a key modeling assumption and we argue for using $r_n \propto h_n$ in Section 3.3. Finally, Section 3.4 considers the use of donut-hole designs as a solution to spillovers.

## 3.1 The Local Linear Regression Estimator

The local linear regression estimator for the local average treatment (equivalently, $\tau_{\mathrm{DIR}}$) is defined as follows:

**Definition 3** (Local Linear Regression Estimator)**.** Let $K$ be a bounded second-order kernel so that $\int_{-1}^{1} K(u)du = 1$ and $\int_{-1}^{1} uK(u)du = 0$. Furthermore, assume that for all $u$, $0 \leq K(u) \leq \bar{K}$, and that $K(u) = 0$ if $u < -1$ or $u > 1$. Let $h_n$ be a sequence

of bandwidths such that $h_n \to 0$ and $n h_n \to \infty$. Then, define

$$\left( \hat{\beta}_0^+, \hat{\beta}_1^+ \right) = \arg \min_{b \in \mathbb{R}^2} \sum_{Z_i \geq 0} K \left( \frac{Z_i}{h_n} \right) (Y_i - b_0 - b_1 Z_i)^2$$

$$\left( \hat{\beta}_0^-, \hat{\beta}_1^- \right) = \arg \min_{b \in \mathbb{R}^2} \sum_{Z_i < 0} K \left( \frac{Z_i}{h_n} \right) (Y_i - b_0 - b_1 Z_i)^2$$

The local linear estimator for the local average treatment effect is $\hat{\tau}_{\mathrm{RDD}} := \hat{\beta}_0^+ - \hat{\beta}_0^-$.

Second order kernels are commonly used in local linear regression. For a discussion on the properties of higher order kernels, see e.g. Wand and Jones (1995). We additionally assume that $K$ has finite support and is bounded and non-negative.

In the absence of spillovers, it is standard practice to estimate the local average treatment effect using $\hat{\tau}_{\mathrm{RDD}}$ (see e.g. Hahn et al. 2001; Cattaneo and Titiunik 2022). In this setting, $\hat{\tau}_{\mathrm{RDD}}$ is consistent for $\tau_{\mathrm{DIR}}$, though inference is complicated by the presence of an asymptotic bias of order $h_n^2$. Various solutions are available for the problem of inference (see e.g. Calonico et al. 2014; Armstrong and Kolesár 2018). In sum, the properties of RDD are relatively well-understood in settings without spillovers.

## 3.2 Estimands of RDD

Our first result characterizes the estimands of RDD when spillovers occur along the running variable.

**Theorem 1.** *Suppose Assumptions 1, 2 and 3 hold.*

(a) *If $r_n \gg h_n$, then $\hat{\tau}_{RDD} \xrightarrow{p} \tau_{DIR}$*

(b) *If $r_n \ll h_n$, then $\hat{\tau}_{RDD} \xrightarrow{p} \tau_{TOT}$*

(c) *If $r_n / h_n \to c$ where $0 < c < \infty$, then $\hat{\tau}_{RDD} \xrightarrow{p} \tau_*$*

When spillovers occur along the running variable, the estimand of RDD exhibits phase transition, with the phases depending on the ratio of $r_n$ to $h_n$. When $r_n \gg h_n$, RDD estimates direct treatment effect. On the other hand, when $r_n \ll h_n$, RDD estimates total treatment effect. In the intermediate regime where $r_n \propto h_n$, it estimates a parameter $\tau_*$, which is displayed in full in Equation (16). $\tau_*$ depends on the $c$, the limiting ratio of $r_n$ and $h_n$, even though this is suppressed in the notation. Generally speaking, $\tau_*$ does not have a clear interpretation. In particular, it is not equal to either $\tau_{\mathrm{DIR}}$ or $\tau_{\mathrm{TOT}}$ unless either $\delta(0) = \gamma(0) = 0$ (i.e. no spillovers) or $\tau_{\mathrm{DIR}} = \gamma(0) = 0$ (i.e. treatment has 0 effect), in which case $\tau_{\mathrm{DIR}} = \tau_{\mathrm{TOT}} = \tau_*$. It may be close to 0 even when $\tau_{\mathrm{DIR}}$ and $\tau_{\mathrm{TOT}}$ are large in magnitude – a common concern expressed by empirical papers regarding spillovers.

The above result is intuitive. The local linear regression estimator essentially treats units within the bandwidth as being comparable. If $r_n$ is large relative to $h_n$, units in the bandwidth have neighborhoods that overlap almost completely so that they experience similar spillovers. As such, units on different sides of the cutoff differ only in their treatment status and $\hat{\tau}_{\mathrm{RDD}}$ estimates direct treatment effect. Conversely, when $r_n$ is small relative to $h_n$, units to the left of the cutoff essentially only has control neighbors, while units to the right of the boundary only has treated neighbors. Comparing these units therefore reveals total treatment effect. When $r_n \propto h_n$, RDD compares a mix of units, some of which have similar neighborhoods, some of which do not. The resulting estimand is therefore a complicated interpolation of direct and spillover effects.

**Remark 1.** Although the exact form of $\tau_*$ depends on the structure of the spillovers, changing assumptions about the functional form or the relative weights of neighbors will not qualitatively affect the results in Theorem 1. In particular (a) and (b) will continue to hold as long as spillovers occur over approximately bounded neighborhood. Similarly, when $r_n \propto h_n$, the limit of the $\hat{\tau}_{\mathrm{RDD}}$ will be combination of terms with no clear interpretation.

**Relation to Existing Work**

Theorem 1, in particular cases (b) and (c), stands in stark contrast to existing work on spillovers within an RDD. Aronow et al. (2017) studies spillovers under the local randomization framework of Cattaneo et al. (2015) and find that RDD always recovers a weighted average of direct treatment effect. Under local randomization, RDD is exactly an experiment under some neighborhood of the cutoff. Consequently, treatment assignment is orthogonal to spillovers, reproducing the results seen in the literature on experiments with spillovers (see e.g. Hudgens and Halloran 2008; Aronow and Samii 2017; Leung 2022a). Dal Torrione et al. (2024) provides a similar analysis under the continuity-based approach of Hahn et al. (2001). They assume that the running variable and the variable determining neighborhood structure have a continuous joint density, ruling out the case we consider. Given continuous joint density, they find that treatment assignment is approximately orthogonal to spillovers. As a consequence, RDD recovers a weighted average of direct treatment effect even under the continuity framework.

Our paper studies RDD under a continuity-based framework and we focus on the case where neighborhoods are completely determined by the running variable. In this setting, the extent of orthogonality between treatment assignment and spillovers depends on the ratio $r_n/h_n$. Our approach – particularly the $r_n \propto h_n$ case – leads to an asymptotic model in which spillovers can affect the estimands of RDD. This is not possible in the framework of Aronow et al. (2017) and Dal Torrione et al. (2024). Our model therefore provides a basis for using data to speak to the effects of spillovers, a task which we take up in Section 4. Finally, we note that this paper is also unique relative to the above two papers in considering endogenous spillovers.

## 3.3 Choice of Asymptotic Framework

Theorem 1 shows that how we model the relative rates between $r_n$ and $h_n$ affects the interpretation of $\hat{\tau}_{\mathrm{RDD}}$. We argue that the $r_n \propto h_n$ regime is the natural choice when

researchers are concerned about spillovers.

In any given application, spillovers occur over some radius that is a finite ratio $c$ of the bandwidth. The regimes $r_n \gg h_n$ or $r_n \ll h_n$ are therefore unrealistic edge cases that take $c = 0$ or $c = \infty$. Consequently, they assert that $\hat{\tau}_{\mathrm{RDD}}$ must either recover direct treatment effect or total treatment effect. There is no room for the data to speak to the amount of direct effects relative to spillovers.

In contrast, the estimand of RDD depends explicitly on the ratio of $c$ under the regime $r_n \propto h_n$, which is a better approximation to how spillovers manifest in practice. It leads to an asymptotic model that preserves more features of the finite sample distribution and therefore provides greater scope for reasoning about spillovers in the data. For this reason, $r_n \propto h_n$ is our preferred framework for analyzing RDD with spillovers. We do not claim $r_n$ to be a quantity that changes with sample size, only that taking it to 0 leads to a useful approximation. Such an approach is also taken in traditional analysis of boundary bias for kernel methods (see e.g. Section 5.5 in Wand and Jones 1995). Here, the target point is modeled as drifting towards the boundary at rate $h_n$ to capture the notion that these two points are close.

In our preferred regime, RDD is potentially inconsistent for either treatment effects parameter. Section 4 proposes a method for recovering both direct treatment effects and spillovers.

## 3.4   Donut-Hole Designs

In the remainder of this section, we informally discuss donut-hole designs as viewed through the lens of our model. Often suggested as a solution for estimating $\tau_{\mathrm{DIR}}$ in the presence of spillovers, donut-hole designs involve excluding observations close to the cutoff from estimation. The idea is that these units are contaminated by spillovers, so that removing them should lead to the recovery of $\tau_{\mathrm{DIR}} = m^+(0) - m^-(0)$. This approach does not work under our model, but we state an alternative model of spillovers under which donut-hole RD can be justified, provided that there are no

endogenous spillovers.

To define the donut-hole RD, let $h_n^o > 0$ be such that $h_n^o < h_n$. Then donut-hole RD estimates $\hat{\beta}^+$ by local linear regression on observations for which $Z_i \in [h_n^o, h_n]$. Similarly for $\hat{\beta}^-$ and observations for which $Z_i \in [-h_n, -h_n^o]$. Let the estimand of the donut-hole RD be denoted $\tau_o$.

Under our linear-in-means model, suppose $h_n^o = r_n$. That is, we exclude all observations that have neighbors on the other side of the cutoff. Suppose there are no endogenous spillovers ($\delta(0) = 0$). Then $\tau_o = \tau_{\text{TOT}}$ and not $\tau_{\text{DIR}}$. In our model, agents closest to the cutoff experience the most similar spillovers, so that comparing them leads to the best estimate of $\tau_{\text{DIR}}$. However, after excluding units in the donut hole $[-h_n^o, h_n^o]$, the remaining units only have neighbors who have the same treatment status as them so that RDD estimates $\tau_{\text{TOT}}$. In this sense, a donut-hole RD is exactly the wrong thing to do if researchers are interested in $\tau_{\text{DIR}}$.

However, suppose $\gamma(z) = \check{\gamma}(z)\mathbf{1}\{z \leq 0\}$. In other words, only control units experience spillovers based on how many of their neighbors are treated. This is reasonable, for example, in the case of information treatment, where information can diffuse from the treated to control units but not vice versa. In this case, if $h_n^o = r_n$ and $\delta(0) = 0$, the intuition in the above paragraph still applies so that $\tau_o$ estimates total treatment effect. However, in this model, total treatment effect is also equal to direct treatment effect under the reference treatment assignment function $d = 0$. Note that direct treatment effect is now dependent on $d$, since spillovers enter the treated and control potential outcomes asymmetrically. Donut-hole RDs can therefore be used to recover $\tau_{\text{DIR}}$ under this model.

Finally, we note that once $\delta(0) \neq 0$, donut-hole RDs do not recover $\tau_{\text{DIR}}$ or $\tau_{\text{TOT}}$ in either model. This is because under endogenous spillovers, the outcome of a given unit is affected by all other units in the domain: a given unit $z$ has an outcome that depends on the outcomes of $[z - r_n, z + r_n]$. In turn, the outcome for $z + r_n$ depends on the outcomes on $[z, z + 2r_n]$, and so on. As such, it is not possible to isolate units

that are "contaminated" by spillovers using a donut design.

# 4 Local Spillover Regression

As we argued in Section 3.3, our preferred asymptotic regime features $r_n \propto h_n$, in which the RDD estimand need not have a causal interpretation. In order to disentangle direct treatment effect from spillovers, Section 4.1 proposes a local analog of the peer effects regression that we term the local spillover regression (LSR). We discuss estimation consistency and inference in Section 4.2. LSR requires researchers to specify a radius $r_n$. We consider choosing $r_n$ as well as $h_n$ in Section 4.3.

## 4.1 Proposed Method

Since spillovers are the cause of inconsistency, our proposal is to estimate $\mu_d(z)$ and $\nu_d(z)$, and then to incorporate them into the local linear regression.

Suppose for now that $r_n$ is known. For a given $Z_i$, define its observed neighbors to be:

$$\tilde{R}(Z_i) = \{j \neq i \,:\, \|Z_j - Z_i\| \leq r_n\} \,.$$

To be consistent with our econometric model, $\|\cdot\|$ is taken to be the Euclidean distance. We then estimate the mean outcome of $Z_i$'s neighbors using a leave-$i$-out estimator:

$$\tilde{\mu}_d(Z_i) = \frac{1}{|\tilde{R}(Z_i)|} \sum_{j \in \tilde{R}(Z_i)} Y_j \,.$$

Additionally, we define

$$\tilde{\mu}_d(0) = \frac{1}{|\tilde{R}(0)|} \sum_{j \in \hat{R}(0)} Y_j \quad , \quad \tilde{R}(0) = \{j \neq i \,:\, \|Z_j\| \leq r_n\} \,.$$

where the only difference is that $z = 0$ is not observed. We focus on the case when the distribution of $Z$ is uniform, so that $\nu_d(Z_i)$ known. When $f$ is not uniform, it is

19

straightforward to form the analogous leave-$i$-out estimator for $\nu_d(Z_i)$.

We propose to estimate $\tau_{\mathrm{DIR}}$, $\gamma(0)$ and $\delta(0)$ using the following:

**Definition 4** (Local Spillover Regression Estimator). For a given kernel $K$ and bandwidths $h_n$, let

$$\tilde{\beta} = \underset{b}{\mathrm{argmin}} \sum_{i=1}^{n} K\left(\frac{Z_i}{h_n}\right) \left(Y_i - \tilde{X}_i' b\right)^2$$

where $D_i = \mathbf{1}\{Z_i \geq 0\}$,

$$
X_i = \begin{pmatrix}
1 \\
Z_i \\
D_i \\
Z_i D_i \\
\mu_d(Z_i) - \mu_d(0) \\
Z_i(\mu_d(Z_i) - \mu_d(0)) \\
\nu_d(Z_i) - \nu_d(0) \\
Z_i(\nu_d(Z_i) - \nu_d(0))
\end{pmatrix}
\quad \text{and} \quad
\tilde{X}_i = \begin{pmatrix}
1 \\
Z_i \\
D_i \\
Z_i D_i \\
\tilde{\mu}_d(Z_i) - \tilde{\mu}_d(0) \\
Z_i(\tilde{\mu}_d(Z_i) - \tilde{\mu}_d(0)) \\
\nu_d(Z_i) - \nu_d(0) \\
Z_i(\nu_d(Z_i) - \nu_d(0))
\end{pmatrix}.
$$

$\tilde{X}_i$ is our estimate of the unobserved $X_i$. The estimator $\tilde{\beta}$ targets the following:

$$
\beta = \begin{pmatrix}
m^-(0) + \delta(0)\mu_d(0) + \gamma(0)\nu_d(0) \\
(m^-)'(0) + \delta'(0)\mu_d(0) + \gamma'(0)\nu_d(0) \\
m^+(0) - m^-(0) \\
(m^+)'(0) - (m^-)'(0) \\
\delta(0) \\
\delta'(0) \\
\gamma(0) \\
\gamma'(0)
\end{pmatrix}.
$$

The first four components of $\tilde{X}_i$ are terms coming from the standard LLR. They are obtained by rewriting the two separate local regressions in Definition 3 as a

single regression. As such, the coefficient of $D_i$ targets $\tau_{\mathrm{DIR}} = m^+(0) - m^-(0)$. The remaining components address spillovers. $\tilde{\mu}_d(Z_i) - \tilde{\mu}_d(0)$ and $\nu_d(Z_i) - \nu_d(0)$ are our estimated mean outcome and mean treatment status in the neighborhood of $Z_i$. As such, their coefficients target $\delta(0)$ and $\gamma(0)$ respectively. $Z_i(\tilde{\mu}_d(Z_i) - \tilde{\mu}_d(0))$ and $Z_i(\nu_d(Z_i) - \nu_d(0))$ are not needed for consistency but they reduce the asymptotic bias when $\delta(z)$ and $\gamma(z)$ are not constant, analogous to how introducing $Z$ reduces the bias of LLR relative to the Nadaraya-Watson estimator.

In view of the above discussion, we will also use the notation $\tilde{\tau}_{\mathrm{DIR}} := \tilde{\beta}_3$, $\tilde{\delta}(0) := \tilde{\beta}_5$, $\tilde{\gamma}(0) := \tilde{\beta}_7$. When $r_n \to 0$, $\tau_{\mathrm{TOT}} \to \frac{\tau_{\mathrm{DIR}} + \gamma(0)}{1 - \delta(0)}$. As such, we also define $\tilde{\tau}_{\mathrm{TOT}} := \frac{\tilde{\tau}_{\mathrm{DIR}} + \tilde{\gamma}(0)}{1 - \tilde{\delta}(0)}$.

## 4.2  Theoretical Properties of LSR

Section 4.2.1 presents the consistency properties of LSR. Section 4.2.2 provides the asymptotic distribution for $\tilde{\beta}$. Section 4.2.3 covers confidence intervals based on the bias-aware approach of Armstrong and Kolesár (2020).

### 4.2.1  Consistency

This section presents results on the consistency of LSR. We first discuss consistency of $\tau_{\mathrm{DIR}}$, which obtains under fairly unrestrictive conditions. We then turn to $\gamma(0)$ and $\delta(0)$, which are more challenging to estimate and will require more assumptions.

For direct treatment effect,

**Theorem 2.** *Suppose Assumptions 1, 2 and 3 hold. Additionally, suppose $n \to \infty$, $h_n \to 0$, $nh_n \to \infty$ and $r_n \propto h_n$. Then $\tilde{\tau}_{DIR} \xrightarrow{p} \tau_{DIR}$.*

The main condition for consistency of $\tilde{\tau}_{\mathrm{DIR}}$ is the Lipschitz continuity of $m^+$, $m^-$, $\delta$ and $\gamma$. This is a slight strengthening of the conditions needed for LLR, which typically assumes continuity of $m^+$ and $m^-$ at the cutoff. Including estimated spillover terms is therefore a simple and effective method to ensure that the target parameter is $\tau_{\mathrm{DIR}}$.

Recovering spillover effects requires more assumptions:

**Assumption 4.** *Suppose the following hold:*

(a) *Suppose $r_n/h_n \to c < 1$.*

(b) *$\delta(0) \neq 0$ and $\tau_{DIR} \neq 0$ or $\gamma(0) \neq 0$.*

Then, we have the following:

**Theorem 3.** *Suppose Assumptions 1, 2, 3 and 4 hold. Additionally, suppose $n \to \infty$, $h_n \to 0$, $nh_n \to \infty$ and $r_n \propto h_n$. Then $\tilde{\beta} \xrightarrow{p} \beta$. In particular,*

$$\tilde{\delta}(0) \xrightarrow{p} \delta(0) \quad , \quad \tilde{\gamma}(0) \xrightarrow{p} \gamma(0) \quad , \quad \tilde{\tau}_{TOT} \xrightarrow{p} \tilde{\tau}_{TOT} \ .$$

The additional assumptions here address various sources of non-identification in the model. Condition (a) is specific to $z \in \mathbb{R}$. In this case, when $c > 1$, $\nu_d(z) = z$ under the assumption of uniform $f$. Having $c < 1$ is therefore necessary for the identification of $\gamma(0)$. This collinearity does not arise when units are in e.g. $\mathbb{R}^2$. Condition (b) is needed to ensure that $\mu_d(z)$ is not collinear with $\nu_d(z)$. Our approximation result in Lemma 9 shows that locally,

$$\mu_d(z) - \mu_d(0) \approx (\nu_d(z) - \nu_d(0)) \cdot \big(\tau_{\text{DIR}} \cdot (\text{first order term smoothing due to } \delta(0))$$
$$\gamma(0) \cdot (\text{second order term smoothing due to } \delta(0))\big)$$

If $\delta(0) = 0$, the first and second order terms are constant, so that $\mu_d(z) - \mu_d(0) = (\nu_d(z) - \nu_d(0))$. Meanwhile, if both $\tau_{\text{DIR}}$ and $\gamma(0)$, then $\mu_d(z) - \mu_d(0) \approx 0$. Intuitively, to learn about $\delta(0)$, we need treatment to create some effect ($\tau_{\text{DIR}}$ or $\gamma(0) \neq 0$) that decays in $z$.

Condition (b) is relatively strong but it is reminiscent of the assumption that $\tau\gamma + \delta \neq 0$ common in the peer effects literature (see e.g. Proposition 1 in Bramoullé et al. 2009). Finally, since the collinearities pertain to some combination of $z$, $\mu_d(z)$ and $\nu_d(z)$, $\tilde{\tau}_{\text{DIR}}$ is consistent for $\tau_{\text{DIR}}$ even when these conditions are not satisfied.

### 4.2.2 Asymptotic Distribution

We next state a central limit theorem for $\tilde{\beta}$. For ease of exposition, we will maintain the full set of assumptions in Theorem 3. We focus on the case when the bandwidth is assumed to have the MSE-optimal rate of $n^{-1/5}$ and introduce the following assumption:

**Assumption 5.** *Suppose that*

*(a)* $h_n, r_n \propto n^{-1/5}$.

*(b) For some $s > 0$,*

$$\sup_{z \in (-s,s)} \mathbb{E}[|\varepsilon|^3 \mid Z = z] < \infty .$$

Then,

**Theorem 4.** *Suppose Assumptions 1, 2, 3 4 and 5 hold. Then, as $n \to \infty$, we have that*

$$\mathbf{V}^{-1/2} \sqrt{n} \left( \tilde{\beta} - \beta - \mathbf{B}_n \right) \xrightarrow{d} \mathbf{N} \left( 0, \mathbb{I}_8 \right) , \tag{1}$$

*where $\mathbf{B}_n = \widetilde{\mathbf{Q}}^{-1} \widetilde{\mathbf{T}}_\rho$ is a bias term, $\widetilde{\mathbf{T}}_\rho$ is defined in (A.2) in the appendix, $\mathbf{V}^{-1/2} = \left( \mathbf{V}^{1/2} \right)^{-1}$, $\mathbf{V}^{1/2} = \widetilde{\mathbf{Q}}^{-1} \mathbf{\Omega}^{1/2}$, and*

$$\begin{aligned}
\widetilde{\mathbf{Q}} &= \frac{1}{n} \sum_{i=1}^{n} K\left( \frac{Z_i}{h} \right) \widetilde{X}_i \widetilde{X}_i', \\
\mathbf{\Omega} &= \mathrm{Var}\left[ K\left( \frac{Z_i}{h} \right) X_i \varepsilon_i + \eta_X \left( \xi_i \right) \right].
\end{aligned} \tag{2}$$

*In the above expression,*

$$\begin{aligned}
\eta_X \left( \xi_1 \right) = &\; \mathbf{E}\left[ K\left( \frac{Z_0}{h} \right) (\delta(0) + \delta'(0) Z_0) X(Z_0) (Y_1 - \mu_d(0)) \left. \frac{R\left( \frac{Z_1}{r} \right)}{\pi_r(0)} \right| \xi_1 \right] \\
&- \mathbf{E}\left[ K\left( \frac{Z_0}{h} \right) (\delta(0) + \delta'(0) Z_0) X(Z_0) (Y_1 - \mu_d(Z_0)) \left. \frac{R\left( \frac{Z_1 - Z_0}{r} \right)}{\pi_r(Z_0)} \right| \xi_1 \right] .
\end{aligned} \tag{3}$$

*where $\xi_i \sim P$ independently. Furthermore, $R(z) = \mathbf{1}_{[-1,1]}(z)$ and $\pi_r(z) = \mathbf{E}\left[R\left(\frac{Z_1-z}{r}\right)\right]$.*

When $h_n \propto n^{-1/5}$, there is an asymptotic bias term $\mathbf{B}_n$ that we address below. $\widetilde{\mathbf{Q}}$ is the regression design matrix. The matrix $\mathbf{\Omega}$ in the definition of $\mathbf{V}$ has two components. The term $K(Z_i/h)X_i\varepsilon_i$ captures the usual source of estimation error in weighted regression if the "ideal" regressors $X_i$ were known. In this case, the only source of estimation error would be due to the unobserved error term $\varepsilon_i$. The feasible regressors $\widetilde{X}_i$, used in place of $X_i$, contain an additional source of error: estimation of the unknown neighborhood average outcomes $\mu_d(Z_i)$. The term $\eta_X(\xi_i)$ captures this additional source of estimation error, i.e. the difference $\widetilde{X}_i - X_i$. It is straightforward to see that the sample analogue of $\mathbf{\Omega}$ combined with $\widetilde{\mathbf{Q}}$ is a consistently estimates $\mathbf{V}$. Let the plug-in estimator be denoted $\widetilde{\mathbf{V}}_n$.

### 4.2.3    Bias-Aware Confidence Intervals

The bias term in Theorem 4, $\mathbf{B}_n$, is asymptotically non-negligible when $h_n \propto n^{-1/5}$. To form valid confidence intervals, we follow the bias-aware approach of Armstrong and Kolesár (2020). This entails restricting the classes of functions to which $m^+, m^-, \delta$ and $\gamma$ can potentially belong, so that an upper bound for $\mathbf{B}_n$ can be obtained. Confidence intervals can then be appropriately widened to ensure coverage under this worst-case bias.

To that end, define the Taylor class of order 2:

$$\mathcal{F}_{T,2}(M) = \left\{f: \ |f(z) - f(0) - f'(0) \cdot z| \leq M \cdot \frac{z^2}{2} \ , \ z \in \mathcal{Z}\right\}$$

This is the class of functions whose first order Taylor approximation has error less than $Mz^2/2$. We can loosely think of the tuning parameter $M$ as the upper bound for the second derivative of $f$ at 0. We will now assume that:

**Assumption 6.** *Suppose*

$$m^+, m^- \in \mathcal{F}_{T,2}(M_m) \quad , \quad \delta \in \mathcal{F}_{T,2}(M_\delta) \quad , \quad \gamma \in \mathcal{F}_{T,2}(M_\gamma) \ .$$

It is then straightforward to see that $|\mathbf{B}_{n,j}| \leq \overline{\mathbf{B}}_{n,j}$, where

$$\overline{\mathbf{B}}_{n,j} = \sum_{i=1}^{n} |w(Z_i)| \cdot \frac{Z_i^2}{2} \left( M_m + M_\delta |\mu_d(Z_i) - \mu_d(0)| + M_\gamma |\nu_d(Z_i) - \nu_d(0)| \right) \quad (4)$$

and $w_j(Z_i) = e_j' \hat{Q}_n^{-1} \frac{1}{h_n} K\left(\frac{Z_i}{h_n}\right) \tilde{X}_i$. Since $\mu_d(Z_i)$ and $\mu_d(0)$ are unobserved, we replace them with $\tilde{\mu}_d(Z_i)$ and $\tilde{\mu}_d(0)$. Denote the resulting estimator $\widetilde{\mathbf{B}}_{n,j}$. We can then form asymptotically valid confidence intervals as follows:

**Theorem 5.** *Suppose Assumptions 1, 2, 3, 4, 5 and 6 hold. Let $t_j = \frac{\sqrt{n}\widetilde{\mathbf{B}}_{n,j}}{\sqrt{e_j'\widetilde{\mathbf{V}}_n e_j}}$ where $e_j$ be an $8 \times 1$ vector with 1 in the j-th component and 0 everywhere else. Denote by $c_{1-\alpha}$ the $1 - \alpha$ quantile of the folded normal distribution $|N(t_j, 1)|$ and let*

$$C_j = \left[ \tilde{\beta}_j - c_{1-\alpha}\sqrt{\frac{e_j'\widetilde{\mathbf{V}}_n e_j}{n}} \ , \ \tilde{\beta}_j + c_{1-\alpha}\sqrt{\frac{e_j'\widetilde{\mathbf{V}}_n e_j}{n}} \right] \ .$$

*Then,*

$$\liminf_{n\to\infty} \mathbb{P}(\beta_j \in C_j) \geq 1 - \alpha \ .$$

The above result is an immediate consequence of Theorem 4 and the fact that $|\mathbf{B}_{n,j}| \leq \overline{\mathbf{B}}_{n,j}$ under Assumption 6. It yields asymptotically valid $1 - \alpha$ confidence intervals for $\tau_{\text{DIR}}$, $\delta(0)$ and $\gamma(0)$. It is then straightforward to obtain valid confidence interval for $\tau_{\text{TOT}}$ by combining these confidence intervals together with a Bonferroni correction.

## 4.3 Choice of $r_n$ and $h_n$

Results in the previous section are predicated on researcher's choices of $r_n$ and $h_n$. This section presents heuristics for choosing these parameters.

### 4.3.1 Choice of $r_n$

In principle, researchers should choose $r_n$ based on their beliefs about likely sources of spillovers. In practice, however, such a choice may not be obvious. Researchers may instead consider choosing $r_n$ via cross-validation as follows.

Let the fold number $L$ be specified by the user and suppose we are given some initial bandwidth $h_n^0$. Partition the data set $D$ into $L$ equal-sized subsets $D_1, ..., D_L$ and let $\tilde{\beta}^{(l)}(r)$ be the LSR estimator computed on $D \setminus D_l$ under radius $r$. Define the out-of-sample MSE as:

$$\text{MSE}^{(l)}(r) = \sum_{i \in D_l} K\left(\frac{Z_i}{h_n^0}\right) \left(Y_i - \tilde{X}_i'\tilde{\beta}^{(l)}(r)\right)^2 \bigg/ \sum_{i \in D_l} K\left(\frac{Z_i}{h_n^0}\right)$$

We can then choose

$$\tilde{r}_n = \operatorname*{argmin}_r \sum_{l \in L} \text{MSE}^{(l)}(r) \ .$$

Intuitively, if spillovers exist, getting the true radius correct should lead to good out-of-sample fit. It would be difficult to do better than the true radius in MSE since this would require the misspecified model to overfit on out-of-sample $\varepsilon_i$'s, which is independent of the training sample.

Finally, to choose the initial bandwidth $h_0$, we might consider a rule of thumb such as

$$h_n^0 = n^{-1/5} \cdot \max_i \|Z_i\| \tag{5}$$

Here, the $\max\|Z_i\|$ scaling ensures that $h_n^0$ is invariant to the units of measurement of the running variable.

The simulations in Section B explores the finite sample performance of the above

method. We find that cross-validation can increase the MSE of various parameters by around 4 times relative to the oracle procedure in which $r_n$ is known. This appears to be a reasonable trade-off when researchers have only weak priors about spillovers in a given setting.

### 4.3.2 Choice of $h_n$

The rule of thumb in Equation (5) leads to a reasonable sequence of bandwidths, but it does not take MSE into consideration. Here, we present an intuitive adaptation of the procedure in Armstrong and Kolesár (2018), which aims to reduce finite sample version of the worst-case MSE.

Given an initial bandwidth $h_n^0$ and a radius that is either pre-specified or selected via cross validation, we can perform LSR and obtain an initial estimate of the asymptotic variance based on the sample analog of $V_n$ in Theorem 4. Call this estimate $\widetilde{\mathbf{V}}_n^0$. For any given bandwidth $h$, we can also obtain an estimate for the maximum bias for $\tilde{\beta}_j$ using the sample analog of $\overline{\mathbf{B}}_{j,n}$ in (4). Call this term $\widetilde{\mathbf{B}}_{j,0}(h)$. We can then choose bandwidth by solving:

$$h_n^* = \underset{h}{\operatorname{argmin}}\ \widetilde{\mathbf{B}}_{j,0}^2 + \frac{h_n^0 \left[\widetilde{\mathbf{V}}_n^0\right]_{jj}}{h \cdot n}$$

While any $\tilde{\beta}_j$ can potentially work with this procedure, we recommend the use of $\tilde{\beta}_3 = \tilde{\tau}_{\text{DIR}}$. This is because $\tilde{\tau}_{\text{DIR}}$ targets a parameter of immediate interest and it is consistent under weaker conditions than $\tilde{\delta}(0)$ and $\tilde{\gamma}(0)$.

## 5  Empirical Demonstration

In this section, we revisit Gonzalez (2021), which studies the effect of cell phone coverage on fraud in the 2009 Afghan presidential election. The analysis is conducted at the polling station level and the outcomes of interest are (1) indicator for whether fraud has likely occurred and (2) likely share of fraudulent votes. Using a spatial

regression discontinuity design, the author finds that cell phone coverage reduces the probability that fraud occurs by 8 percentage points and decreases the share of fraudulent votes by 4 percentage points. Gonzalez (2021) is concerned with spillovers, highlighting for example that voters may walk from a non-covered to a covered area to report fraud. To test for spillovers, the author invokes donut-hole type reasoning and compares the fraud level at non-covered polling stations between 0–2 km of the boundary to those that are 2–4 km, 4–6 km or 6–8 km away. The idea is that spillovers should lead to spikes in fraud levels at polling stations closer to the boundary. They find that the differences in outcomes are not statistically significant and conclude that there is likely no spillovers.

In the spirit of the above exercise, we apply our method in the setting of Gonzalez (2021) as a robustness check. We first replicate the paper's result using either the method of Armstrong and Kolesár (2020) (AK) or Calonico et al. (2014) (CCT). All tuning parameters are selected using the default procedures in the authors' respective R packages. Relative to the original specification, we do not include spatial fixed effects since they fall outside the scope of conventional theory. We then proceed with local spillover regression (LSR) taking the bandwidth choices of AK and CCT as given. Following Gonzalez (2021), we consider circular neighborhoods that are between 2 to 8 km in radius. In principle, we can compute $\nu_i$ given the exact boundaries of cell phone coverage. Since that is not available, we estimate $\nu_i$ using the number of treated polling station in a given station's neighborhood, although we treat it as known when conducting inference. With the running variable measured in kilometers, we chose $M_m$ to be 0.01, which implies that the local linear approximation to the conditional probability of fraud has an error up to 1 (i.e. the entire support of the variable) at the 10 km radius. Similarly, we chose $M_\delta$ and $M_\gamma$ to be 0.01. Finally, we consider the case in which we select the spillover radius by cross-validation and choose bandwidth using our rule-of-thumb (LSR-CV).

Table 1 presents the results when using the indicator for fraud as outcome and

the sample is the full set of polling stations. This is comparable to Panel A Column (1) in Table 2 of Gonzalez (2021). Across the various bandwidth and radius choices, we find $\delta(0)$ to be positive and significant at the 5% level. The estimates are also relatively stable across parameter values hovering between 0.6 to 0.75. Positive exogenous spillovers would arise, for example, if fraud is driven by local norms of corruption (Fisman and Miguel 2007; Barr and Serra 2010 and references therein). On the other hand, we do not find conclusive evidence of exogenous spillovers. These results continue to hold with data-driven choices of radius and bandwidth. Our cross-validation procedure suggests that spillovers occur across a radius of 6.1 km, which is within the plausible range of radii considered by Gonzalez (2021). Results when using fraud share as outcome are similar.

The above exercise suggests that spillovers can be a threat to the identification of treatment effect parameters in RDD. In this case, our method can be useful as a robustness check. Additionally, the parameters for endogenous and exogenous spillovers may also be of interest since they can shed light on how social interactions mediate a variable of interest.

# 6    Conclusion

Regression Discontinuity Design allows researchers to obtain causal estimates of treatment effects at the cutoff. It relies on the assumption of no spillovers, which may not hold in practice. When neighborhoods are determined by the running variable, we find that the estimand of RDD is sensitive to the ratio of two terms: (1) radius over which spillovers occur and (2) the bandwidth used for local linear regression. In our preferred approximation, those two terms are of similar order so that an alternative to local linear regression is needed to recover direct treatment effects and spillovers. We propose the local spillover regression – the local analog of peer effects regression – and show that it can be a useful tool for addressing spillovers in RDD.

| Bandwidth | Eff. Obs | Method | Radius | $\tau_{DIR}$ | $\delta(0)$ | $\gamma(0)$ |
|---|---|---|---|---|---|---|
| | | AK | - | -0.028 [-0.093, 0.038] | | |
| | | LSR | 2 | -0.077 [-0.372, 0.219] | 0.585 [0.226, 0.944] | 0.103 [-0.182, 0.388] |
| 10.8 | 1249 | LSR | 4 | -0.018 [-0.326, 0.291] | 0.640 [0.240, 1.039] | -0.011 [-0.309, 0.286] |
| | | LSR | 6 | -0.021 [-0.333, 0.290] | 0.719 [0.309, 1.129] | 0.000 [-0.276, 0.277] |
| | | LSR | 8 | -0.029 [-0.344, 0.286] | 0.742 [0.307, 1.178] | -0.032 [-0.330, 0.267] |
| | | CCT | - | 0.014 [-0.060, 0.088] | | |
| | | LSR | 2 | -0.061 [-0.244, 0.123] | 0.586 [0.347, 0.825] | 0.119 [-0.043, 0.281] |
| 6.7 | 1026 | LSR | 4 | -0.008 [-0.191, 0.174] | 0.642 [0.378, 0.906] | 0.002 [-0.158, 0.162] |
| | | LSR | 6 | -0.009 [-0.180, 0.162] | 0.722 [0.471, 0.973] | 0.009 [-0.141, 0.159] |
| | | LSR | 8 | -0.024 [-0.191, 0.143] | 0.752 [0.500, 1.004] | -0.025 [-0.186, 0.136] |
| 3.2 | 591 | LSR-CV | 6.10 | -0.004 [-0.103, 0.095] | 0.744 [0.611, 0.876] | 0.023 [-0.067, 0.112] |

Table 1: Comparison of AK, CCT and LSR in the setting of Gonzalez (2021). Outcome is indicator for whether any fraud has occurred. This analysis uses the full sample of polling stations. Square brackets contain 95% confidence intervals. Eff. Obs. is number of polling stations within the bandwidth.

# References

**Armstrong, Timothy B and Michal Kolesár**, "Optimal inference in a class of regression models," *Econometrica*, 2018, *86* (2), 655–683.

_ **and** _ , "Simple and honest confidence intervals in nonparametric regression," *Quantitative Economics*, 2020, *11* (1), 1–39.

**Aronow, Peter M and Cyrus Samii**, "Estimating average causal effects under general interference, with application to a social network experiment," 2017.

_ **, Nicole E Basta, and M Elizabeth Halloran**, "The regression discontinuity design under interference: a local randomization-based approach," *Observational Studies*, 2017, *3* (2), 129–133.

**Auerbach, Eric, Hongchang Guo, and Max Tabord-Meehan**, "The Local Approach to Causal Inference under Network Interference," *arXiv. org*, 2025.

_ **, Yong Cai, and Ahnaf Rafi**, "Regression discontinuity design with spillovers," *arXiv preprint arXiv:2404.06471v1*, 2024.

_ **, ** _ **, and ** _ , "Regression discontinuity design with spillovers," *arXiv preprint arXiv:2404.06471*, 2025.

**Barr, Abigail and Danila Serra**, "Corruption and culture: An experimental analysis," *Journal of Public economics*, 2010, *94* (11-12), 862–869.

**Borusyak, K and M Kolerman-Shemer**, "Regression Discontinuity Aggregation, with an Application to the Union Effects on Inequality," *Working Paper*, 2024.

**Bramoullé, Yann, Habiba Djebbari, and Bernard Fortin**, "Identification of peer effects through social networks," *Journal of econometrics*, 2009, *150* (1), 41–55.

**Butts, Kyle**, "Difference-in-differences estimation with spatial spillovers," *arXiv preprint arXiv:2105.03737*, 2021.

**Calonico, Sebastian, Matias D Cattaneo, and Rocio Titiunik**, "Robust non-parametric confidence intervals for regression-discontinuity designs," *Econometrica*, 2014, *82* (6), 2295–2326.

**Cao, Jianfei and Connor Dowd**, "Estimation and inference for synthetic control methods with spillover effects," *arXiv preprint arXiv:1902.07343*, 2019.

**Cattaneo, Matias D and Rocio Titiunik**, "Regression discontinuity designs," *Annual Review of Economics*, 2022, *14*, 821–851.

**_ , Brigham R Frandsen, and Rocio Titiunik**, "Randomization inference in the regression discontinuity design: An application to party advantages in the US Senate," *Journal of Causal Inference*, 2015, *3* (1), 1–24.

**Clarke, Damian**, "Estimating difference-in-differences in the presence of spillovers," 2017.

**Cliff, Andrew David and J Keith Ord**, "Spatial autocorrelation," 1973.

**Faridani, Stefan and Paul Niehaus**, "Linear Estimation of Global Average Treatment Effects," Technical Report, National Bureau of Economic Research 2024.

**Fisman, Raymond and Edward Miguel**, "Corruption, norms, and legal enforcement: Evidence from diplomatic parking tickets," *Journal of Political economy*, 2007, *115* (6), 1020–1048.

**Forastiere, Laura, Davide Del Prete, and Valerio Leone Sciabolazza**, "Causal inference on networks under continuous treatment interference," *arXiv preprint arXiv:2004.13459*, 2020.

**Gao, Mengsi and Peng Ding**, "Causal inference in network experiments: regression-based analysis and design-based properties," *arXiv preprint arXiv:2309.07476*, 2023.

**Giorgi, Giacomo De, Michele Pellizzari, and Silvia Redaelli**, "Identification of social interactions through partially overlapping peer groups," *American Economic Journal: Applied Economics*, 2010, *2* (2), 241–275.

**Goldsmith-Pinkham, Paul and Guido W Imbens**, "Social networks and the identification of peer effects," *Journal of Business & Economic Statistics*, 2013, *31* (3), 253–264.

**Gonzalez, Robert M**, "Cell phone access and election fraud: evidence from a spatial regression discontinuity design in Afghanistan," *American Economic Journal: Applied Economics*, 2021, *13* (2), 1–51.

**Hahn, Jinyong, Petra Todd, and Wilbert Van der Klaauw**, "Identification and estimation of treatment effects with a regression-discontinuity design," *Econometrica*, 2001, *69* (1), 201–209.

**Hu, Yuchen, Shuangning Li, and Stefan Wager**, "Average direct and indirect causal effects under interference," *Biometrika*, 2022, *109* (4), 1165–1172.

**Hudgens, Michael G and M Elizabeth Halloran**, "Toward causal inference with interference," *Journal of the American Statistical Association*, 2008, *103* (482), 832–842.

**Jardim, Ekaterina S, Mark C Long, Robert Plotnick, Emma van Inwegen, Jacob L Vigdor, and Hilary Wething**, "Boundary Discontinuity Methods and Policy Spillovers," Technical Report, National Bureau of Economic Research 2022.

**Kelejian, Harry H and Ingmar R Prucha**, "A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances," *The journal of real estate finance and economics*, 1998, *17*, 99–121.

_ **and** _ , "Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances," *Journal of econometrics*, 2010, *157* (1), 53–67.

**Lee, Lung-Fei**, "Asymptotic distributions of quasi-maximum likelihood estimators for spatial autoregressive models," *Econometrica*, 2004, *72* (6), 1899–1925.

_ , "GMM and 2SLS estimation of mixed regressive, spatial autoregressive models," *Journal of Econometrics*, 2007, *137* (2), 489–514.

**Leung, Michael P**, "Treatment and spillover effects under network interference," *Review of Economics and Statistics*, 2020, *102* (2), 368–380.

_ , "Causal inference under approximate neighborhood interference," *Econometrica*, 2022, *90* (1), 267–293.

_ , "Rate-optimal cluster-randomized designs for spatial interference," *The Annals of Statistics*, 2022, *50* (5), 3064–3087.

**Li, Shuangning and Stefan Wager**, "Random graph asymptotics for treatment effect estimation under network interference," *The Annals of Statistics*, 2022, *50* (4), 2334–2358.

_ , **Ramesh Johari, Xu Kuang, and Stefan Wager**, "Experimenting under stochastic congestion," *arXiv preprint arXiv:2302.12093*, 2023.

**Manski, Charles F**, "Identification of endogenous social effects: The reflection problem," *The review of economic studies*, 1993, *60* (3), 531–542.

**Munro, Evan**, "Efficient Estimation of Causal Effects under Interference through Designed Markets," 2023.

_ , **Stefan Wager, and Kuang Xu**, "Treatment effects in market equilibrium," *arXiv preprint arXiv:2109.11647*, 2021.

**Paula, Aureo De, Imran Rasul, and Pedro CL Souza**, "Identifying network ties from panel data: Theory and an application to tax competition," *Review of Economic Studies*, 2024, p. rdae088.

**Paula, Áureo De, Seth Richards-Shubik, and Elie Tamer**, "Identifying preferences in networks with bounded degree," *Econometrica*, 2018, *86* (1), 263–288.

**Penrose, Mathew**, *Random geometric graphs*, Vol. 5, OUP Oxford, 2003.

**Pollard, David**, *Convergence of Stochastic Processes* Springer Series in Statistics, Springer New York, 1984.

**Sävje, Fredrik, Peter Aronow, and Michael Hudgens**, "Average treatment effects in the presence of unknown interference," *Annals of statistics*, 2021, *49* (2), 673.

**Sobel, Michael E**, "What do randomized studies of housing mobility demonstrate? Causal inference in the face of interference," *Journal of the American Statistical Association*, 2006, *101* (476), 1398–1407.

**Thistlethwaite, Donald L and Donald T Campbell**, "Regression-discontinuity analysis: An alternative to the ex post facto experiment.," *Journal of Educational psychology*, 1960, *51* (6), 309.

**Torrione, Elena Dal, Tiziano Arduini, and Laura Forastiere**, "Regression Discontinuity Designs Under Interference," *arXiv preprint arXiv:2410.02727*, 2024.

**van der Vaart, Aad and Jon Wellner**, *Weak Convergence and Empirical Processes: With Applications to Statistics* Springer Series in Statistics, 2 ed., Springer New York, 2023.

**van der Vaart, Aad W.**, *Asymptotic Statistics* Asymptotic Statistics, Cambridge University Press, 1998.

**Vazquez-Bare, Gonzalo**, "Causal spillover effects using instrumental variables," *Journal of the American Statistical Association*, 2023, *118* (543), 1911–1922.

_ , "Identification and estimation of spillover effects in randomized experiments," *Journal of Econometrics*, 2023, *237* (1), 105237.

**Wand, Matt P and M Chris Jones**, *Kernel smoothing*, CRC press, 1995.

**Xu, Ruonan**, "Difference-in-Differences with Interference: A Finite Population Perspective," *arXiv preprint arXiv:2306.12003*, 2023.

# A   Appendix

This Appendix provides proofs for the theorems in the main text. They draw on auxillary lemmas that are stated in Section C of the Supplemental Appendix.

## A.1   Proof of Theorem 1

We consider the regimes case-by-case. The proofs are similar across the cases except we use a different approximation for $\mu_d(xh_n) - \mu_d(0)$ depending on the regimes.

**Case 1:** $r_n \gg h_n$.

We first derive the bias for $\hat{\beta}^+$. A similar argument follows for $\hat{\beta}^-$. The two together yield the bias for $\hat{\tau}_d$. By Assumption 1 and the standard expansion:

$$
\begin{aligned}
Y_i &= m^+(Z_i) + \delta(Z_i)\mu_d(Z_i) + \gamma(Z_i)\nu_d(Z_i) + \varepsilon_i \\
&= m^+(0) + \delta(0)\mu_d(0) + \gamma(0)\nu_d(0) + m_z^+(0) \cdot Z_i + \frac{1}{2}m_{zz}^+(0) \cdot Z_i^2 \\
&\quad + \delta(0)\left(\mu_d(Z_i) - \mu_d(0)\right) + \delta_z(0)Z_i\mu_d(Z_i) + \frac{1}{2}\delta_{zz}(0)Z_i^2\mu_d(Z_i) \\
&\quad + \gamma(0)\left(\nu_d(Z_i) - \nu_d(0)\right) + \gamma_z(0)Z_i\nu_d(Z_i) + \frac{1}{2}\gamma_{zz}(0)Z_i^2\nu_d(Z_i) + \varepsilon_i + O(h_n^3)
\end{aligned}
$$

In matrix form,

$$
\mathbb{E}\left[\hat{\beta}_0^+ \mid \mathbf{Z}\right] = e_1^T \left(\tilde{\mathbf{Z}}^T \mathbf{W} \tilde{\mathbf{Z}}\right)^{-1} \tilde{\mathbf{Z}} \mathbf{W} \mathbf{M}
$$

where only observations with $Z_i \geq 0$ enter the following matrices:

$$
\tilde{\mathbf{Z}} = \begin{pmatrix} 1 & Z_1 \\ \vdots & \vdots \\ 1 & Z_{n_+} \end{pmatrix} \quad , \quad \mathbf{M} = \begin{pmatrix} m^+(Z_1) + \delta(Z_1)\mu_d(Z_1) + \gamma(Z_1)\nu_d(Z_1) \\ \vdots \\ m^+(Z_{n_+}) + \delta(Z_{n_+})\mu_d(Z_{n_+}) + \gamma(Z_{n_+})\nu_d(Z_{n_+}) \end{pmatrix}
$$

$$
\mathbf{W} = \operatorname{diag}\left(K_h(Z_1)\cdots K_h(Z_{n_+})\right) .
$$

As such, we can write

$$
\mathbf{E}\left[\hat{\beta}_0^+ \mid \mathbf{Z}\right] = m^+(0) + \delta(0)\mu_d(0) + \gamma(0)\nu_d(0) + A_n + B_n + R
$$

where $R = O_p(h_n^2/r_n)$ and

$$
A_n := \delta(0) \cdot e_1^T \left(\tilde{\mathbf{Z}}^T \mathbf{W} \tilde{\mathbf{Z}}\right)^{-1} \tilde{\mathbf{Z}} \mathbf{W} \begin{pmatrix} \mu_d(Z_1) - \mu_d(0) \\ \vdots \\ \mu_d(Z_{n_+}) - \mu_d(0) \end{pmatrix} \quad , \quad B_n := \gamma(0) \cdot e_1^T \left(\tilde{\mathbf{Z}}^T \mathbf{W} \tilde{\mathbf{Z}}\right)^{-1} \tilde{\mathbf{Z}} \mathbf{W} \begin{pmatrix} \nu_d(Z_1) - \nu_d(0) \\ \vdots \\ \nu_d(Z_{n_+}) - \nu_d(0) \end{pmatrix} .
$$

We first analyze $A_n$. Note that the denominator is standard. From Wand and

Jones (1995),

$$\left(\frac{1}{n_+}\tilde{\mathbf{Z}}^T\mathbf{W}\tilde{\mathbf{Z}}\right)^{-1} = \left[h_n\gamma_{2,1}\gamma_{0,1} - h_n\gamma_{1,1}^2 + o_p(h_n^2)\right]^{-1}\begin{pmatrix} h_n^2\gamma_{2,1} + o_p(h_n^2) & -h_n\gamma_{1,1} + o_p(h_n) \\ * & \gamma_{0,1} + o_p(h_n^2) \end{pmatrix} \quad (6)$$

where have normalized $f(0)$ to 1. For the numerator,

$$\frac{1}{n_+}\tilde{\mathbf{Z}}\mathbf{W}\begin{pmatrix} \mu_d(Z_1) - \mu_d(0) \\ \vdots \\ \mu_d(Z_{n_+}) - \mu_d(0) \end{pmatrix} = \begin{pmatrix} \frac{1}{n_+}\sum_{i=1}^{n_+} K_h(Z_i)\left(\mu_d(Z_i) - \mu_d(0)\right) \\ \frac{1}{n_+}\sum_{i=1}^{n_+} K_h(Z_i)Z_i\left(\mu_d(Z_i) - \mu_d(0)\right) \end{pmatrix}$$

We evaluate the expectation of each term with respect to $\mathbf{Z}$. First,

$$\mathbb{E}\left[\frac{1}{n_+}\sum_{i=1}^{n_+} K_h(Z_i)\left(\mu_d(Z_i) - \mu_d(0)\right)\right] = \frac{h_n^2}{r_n}\left(\Gamma(r_n)\int_0^1 K(x)dF(x) + o(1)\right) = \frac{h_n^2}{r_n}\left(\Gamma(r_n)\cdot\gamma_{0,1} + o(1)\right),$$

$$(7)$$

where the first equality follows from the usual change of variables argument and last equality follows from Lemma 9, which also provides the exact form of $\Gamma(r_n)$. By a similar set of manipulations,

$$\mathbb{E}\left[\frac{1}{n_+}\sum_{i=1}^{n_+} K_h(Z_i)Z_i\left(\mu_d(Z_i) - \mu_d(0)\right)\right] = \frac{h_n^2}{r_n}\left(\Gamma(r_n)\int_0^1 xK(x)dF(x) + o(1)\right) = \frac{h_n^2}{r_n}\left(\Gamma(r_n)\cdot\gamma_{1,1} + o(1)\right).$$

$$(8)$$

Combining equations (6), (7) and (8), we have that $A_n = O_p\left(\frac{h_n^2}{r_n}\right)$. Finally, observe that $\frac{1}{2}m_{zz}^+(0)Z_i^2$ contributes to the standard $O_p(h_n^2)$ bias term in RDD. It is also straightforward to show that the residual corresponding to $\delta_z(0)Z_i\mu_d(Z_i)$ and $\delta_z(0)Z_i^2\mu_d(Z_i)$ are $h_n$ times smaller than the above display.

Next, consider $B_n$. The denominator is the same as for $A_n$ and it's limit is given

by Equation (6). To analyze the numerator, first observe that

$$\nu_d(Z_i) - \nu_d(0) = \begin{cases} -\frac{1}{2} & \text{if } Z_i \leq -r_n \\ \frac{Z_i}{2r_n} & \text{if } -r_n \leq Z_i \leq Z_i \\ \frac{1}{2} & \text{if } Z_i \geq r_n \end{cases} \tag{9}$$

As such, if $h_n \ll r_n$, $\nu_d(Z_i) - \nu_d(0) = O_p(h_n)$ conditional on $Z_i \in [-h_n, h_n]$. It is therefore immediate that analogs of Equations 7 and 8 hold for $B_n$. Therefore, we also have that $B_n = O_p\left(\frac{h_n^2}{r_n}\right)$. It remains to note that the residual error from approximating $\gamma(Z_i)$ at $\gamma(0)$ is $O_p(h_n^2)$, as in the case with $\delta(\cdot)$ above.

Conclude that $\hat{\beta}_0^+ \xrightarrow{p} m^+(0) + \delta(0)\mu_d(0) + \gamma(0)\nu_d(0)$. The stated properties on $\hat{\tau}_{\text{RDD}}$ follow immediately by a similar set of manipulations for $\hat{\beta}_0^-$

**Case 2:** $r_n/h_n \to 0$

Again, start by considering $A_n$. For $Z_i \geq 0$, the change of variables $x = z/h_n$ yields:

$$\mathbb{E}\left[\frac{1}{n_+}\sum_{i=1}^{n_+} K_h(Z_i)\left(\mu_d(Z_i) - \mu_d(0)\right)\right] = h_n \int_0^1 K(x)\left(\mu_d(xh_n) - \mu_d(0)\right) dF(x)$$
$$= h_n\left(\frac{\tau_d + \delta(0)\gamma(0)}{2(1 - \delta(0))} \cdot \gamma_{0,1} + o(1)\right) \tag{10}$$

where the first equality follows from Lemma 9 and the fact that the integrand is bounded. And similarly,

$$\mathbb{E}\left[\frac{1}{n_+}\sum_{i=1}^{n_+} K_h(Z_i)Z_i\left(\mu_d(Z_i) - \mu_d(0)\right)\right] = h_n^2\left(\frac{\tau_d + \delta(0)\gamma(0)}{2(1 - \delta(0))}\gamma_{1,1} + o(1)\right) \tag{11}$$

Combining equations (6), (10) and (11) gives us $A_n = \delta(0) \cdot \frac{\tau_d + \delta(0)\gamma(0)}{2(1-\delta(0))} + O_p(h_n)$.

Next, consider $B_n$. From Equation 9, we have that $|Z_i| \in [w_n, 1]$, $\nu_d(Z_i) - \nu_d(0) =$

$\text{sgn}(Z_i)/2$. On $[-w_n, w_n]$, the term is bounded. As such, we can write:

$$\nu_d(Z_i) - \nu_d(0) = \frac{1}{2}\text{sgn}(Z_i) + O(w_n)$$

uniformly over $Z_i \in \mathcal{Z}$. By similar manipulations as for $A_n$, but using the approximation above, $B_n = \frac{\gamma(0)}{2} + O_p(h_n)$. As such,

$$\hat{\beta}_0^+ = m^+(0) + \delta(0)\mu_d(0) + \gamma(0)\nu_d(0) + \delta(0) \cdot \frac{\tau_d + \gamma(0)}{2(1 - \delta(0))} + \frac{\gamma(0)}{2} + O_p(h_n)$$

Similarly for $\hat{\beta}^1$. Therefore $\hat{\tau}_{\text{RDD}} \xrightarrow{p} \frac{\tau_d + \gamma(0)}{1 - \delta(0)}$ $\hspace{2cm}$ $\square$

**Case 3:** $r_n = \frac{1}{2}ch_n$

As before, using Lemma 9 yields:

$$\mathbb{E}\left[\frac{1}{n_+}\sum_{i=1}^{n_+} K_h(Z_i)\left(\mu_d(Z_i) - \mu_d(0)\right)\right] = h_n\left(\tau_d\Lambda_{0,1,1}^+ + \gamma(0)\tilde{\Lambda}_{0,1,1}^+ + o(1)\right) \quad (12)$$

$$\mathbb{E}\left[\frac{1}{n_+}\sum_{i=1}^{n_+} K_h(Z_i)Z_i\left(\mu_d(Z_i) - \mu_d(0)\right)\right] = h_n^2\left(\tau_d\Lambda_{1,1,1}^+ + \gamma(0)\tilde{\Lambda}_{1,1,1}^+ + o(1)\right) \quad (13)$$

Combining equations (6), (12) and (13) gives us

$$A_n = \frac{\tau_d}{\gamma_{2,1}\gamma_{0,1} - \gamma_{1,1}^2} \cdot \left(\gamma_{2,1}\Lambda_{0,1,1}^+ - \gamma_{1,1}\Lambda_{1,1,1}^+\right) + \frac{\gamma(0)}{\gamma_{2,1}\gamma_{0,1} - \gamma_{1,1}^2} \cdot \left(\gamma_{2,1}\tilde{\Lambda}_{0,1,1}^+ - \gamma_{1,1}\tilde{\Lambda}_{1,1,1}^+\right) + O_p(h_n).$$

Next, for $B_n$:

$$\mathbb{E}\left[\frac{1}{n_+}\sum_{i=1}^{n_+} K_h(Z_i)\left(\nu_d(Z_i) - \nu_d(0)\right)\right] = h_n\left(\Gamma_{0,1,1}^+ + o(1)\right) \quad (14)$$

$$\mathbb{E}\left[\frac{1}{n_+}\sum_{i=1}^{n_+} K_h(Z_i)Z_i\left(\nu_d(Z_i) - \nu_d(0)\right)\right] = h_n^2\left(\Gamma_{1,1,1}^+ + o(1)\right). \quad (15)$$

40

By Equations (6), (14) and (15), we have that $B_n = \frac{\gamma_{2,1}\Gamma_{0,1,1}^+ - \gamma_{1,1}\Gamma_{1,1,1}^+}{\gamma_{2,1}\gamma_{0,1} - \gamma_{1,1}^2} + O_p(h_n)$. Noting that the usual bias term coming from $m_{zz}$ is smaller than the above, we therefore have that:

$$
\begin{aligned}
\hat{\beta}_0^+ = {}& m^+(0) + \delta(0)\mu_d(0) + \gamma(0)\nu_d(0) \\
& + \frac{\delta(0)\tau_d}{\gamma_{2,1}\gamma_{0,1} - \gamma_{1,1}^2} \cdot \left( \gamma_{2,1}\Lambda_{0,1,1}^+ - \gamma_{1,1}\Lambda_{1,1,1}^+ \right) \\
& + \frac{\delta(0)\gamma(0)}{\gamma_{2,1}\gamma_{0,1} - \gamma_{1,1}^2} \cdot \left( \gamma_{2,1}\tilde{\Lambda}_{0,1,1}^+ - \gamma_{1,1}\tilde{\Lambda}_{1,1,1}^+ \right) \\
& + \gamma(0) \cdot \frac{\gamma_{2,1}\Gamma_{0,1,1}^+ - \gamma_{1,1}\Gamma_{1,1,1}^+}{\gamma_{2,1}\gamma_{0,1} - \gamma_{1,1}^2} + O_p(h_n) .
\end{aligned}
$$

Similarly for $\hat{\beta}^-$. As such, $\hat{\tau}_{\mathrm{RDD}} = \tau_* + O_p(h_n)$, where

$$
\begin{aligned}
\tau_* - \tau_d = {}& \frac{\delta(0)\tau_d}{\gamma_{2,1}\gamma_{0,1} - \gamma_{1,1}^2} \cdot \left( \gamma_{2,1}\left( \Lambda_{0,1,1}^+ - \Lambda_{0,1,1}^- \right) - \gamma_{1,1}\left( \Lambda_{1,1,1}^+ - \Lambda_{1,1,1}^- \right) \right) \\
& + \frac{\delta(0)\gamma(0)}{\gamma_{2,1}\gamma_{0,1} - \gamma_{1,1}^2} \cdot \left( \gamma_{2,1}\left( \tilde{\Lambda}_{0,1,1}^+ - \tilde{\Lambda}_{0,1,1}^- \right) - \gamma_{1,1}\left( \tilde{\Lambda}_{1,1,1}^+ - \tilde{\Lambda}_{1,1,1}^- \right) \right) \qquad (16) \\
& + \frac{\gamma(0)}{\gamma_{2,1}\gamma_{0,1} - \gamma_{1,1}^2} \left( \gamma_{2,1}\left( \Gamma_{0,1,1}^+ - \Gamma_{0,1,1}^+ \right) - \gamma_{1,1}\left( \Gamma_{1,1,1}^- - \Gamma_{1,1,1}^- \right) \right) .
\end{aligned}
$$

and $\Lambda, \tilde{\Lambda}, \Gamma$ are as in Definition 12. $\qquad\square$

## A.2    Proof of Theorems 2 and 3

We prove Theorem 3. Theorem 2 follows from the usual Frisch-Waugh-Lovell argument after observing that the projection residuals of $Y_i$ and $D_i$ on $\tilde{\mu}(Z_i) - \tilde{\mu}(0)$, $\tilde{\nu}(Z_i) - \tilde{\nu}(0)$ and $Z_i$ are well-defined even if they are collinear. The proof for Theorem 3 largely follows standard arguments as in Chapter 5 of Wand and Jones (1995). There are two complications: (1) $\tilde{X}_i$ is measured with error. We use a concentration inequality in Lemma 21 to show that this error is neglible. (2) We require a uniform approximation for $\mu(Z_i) - \mu(0)$. This is provided in Lemma 9.

Under our assumptions,

$$\widetilde{\beta} - \beta = \left( \frac{1}{n} \sum_{i=1}^{n} K_h\left(Z_i\right) \widetilde{X}_i \widetilde{X}_i' \right)^{-1} \frac{1}{n} \sum_{i=1}^{n} K_h\left(Z_i\right) \widetilde{X}_i \left\{ \left(X_i - \widetilde{X}_i\right)' \beta + \rho_i + \varepsilon_i \right\},$$

which can be rewritten as $\widetilde{\beta} - \beta = \widetilde{\mathbf{Q}}^{-1} \left\{ \mathbf{B}_n + \widetilde{\mathbf{T}}_X + \widetilde{\mathbf{T}}_\varepsilon \right\}$, where

$$\widetilde{\mathbf{Q}} = \frac{1}{n} \sum_{i=1}^{n} K\left(Z_i/h_n\right) \widetilde{X}_i \widetilde{X}_i' \qquad\qquad \mathbf{B}_n = \frac{1}{n} \sum_{i=1}^{n} K\left(Z_i/h_n\right) \widetilde{X}_i \rho_i$$

$$\widetilde{\mathbf{T}}_X = \frac{1}{n} \sum_{i=1}^{n} K\left(Z_i/h_n\right) \widetilde{X}_i \left(X_i - \widetilde{X}_i\right)' \beta \qquad\qquad \widetilde{\mathbf{T}}_\varepsilon = \frac{1}{n} \sum_{i=1}^{n} K\left(Z_i/h_n\right) \widetilde{X}_i \varepsilon_i.$$

We start with the terms in $\widetilde{\mathbf{Q}}$. First note that for $1 \leq i, j \leq 4$, $\widetilde{\mathbf{Q}}_{ij}$ are terms associated with the usual local linear regression. Standard arguments and noting that $f'(0) = 0$ give us:

$$\frac{1}{h_n} \widetilde{\mathbf{Q}}_{1:4,1:4} = f(0) \cdot \begin{pmatrix} 1 + o_p(1) & o_p(h_n^2) & \int_0^1 K(x)dx + o_p(1) & h_n\left(\int_0^1 xK(x)dx + o_p(1)\right) \\ * & h_n^2\left(\int_{-1}^1 x^2 K(x)dx + o_p(1)\right) & h_n\left(\int_0^1 xK(x)dx + o_p(1)\right) & h_n^2\left(\int_0^1 x^2 K(x)dx + o_p(1)\right) \\ * & * & \int_0^1 K(x)dx + o_p(1) & h_n\left(\int_0^1 xK(x)dx + o_p(1)\right) \\ * & * & * & h_n^2\left(\int_0^1 x^2 K(x)dx + o_p(1)\right) \end{pmatrix}$$

Next, consider

$$\frac{1}{h_n} \widetilde{\mathbf{Q}}_{1,5} = \frac{1}{nh_n} \sum_{i=1}^{n} K\left(Z_i/h\right)\left(\mu(Z_i) - \mu(0)\right) + \frac{1}{nh_n} \sum_{i=1}^{n} K\left(Z_i/h_n\right)\left(\tilde{\mu}(Z_i) - \mu(Z_i)\right) + \frac{1}{nh_n} \sum_{i=1}^{n} K\left(Z_i/h\right)\left(\tilde{\mu}(0) - \mu(0)\right) .$$

By the Lemma 21,

$$\frac{1}{nh_n} \sum_{i=1}^{n} K\left(Z_i/h_n\right)\left(\tilde{\mu}(Z_i) - \mu(Z_i)\right) = o(1) \cdot \left(f(0) + o_p(1)\right) \qquad \text{w.p.a. } 1$$

Next, denote the uniform approximation in the $r_n \propto h_n$ regime in Lemma 9 by $u(x)$. Using $u(x)$ and applying the standard arguments yields:

$$\frac{1}{nh_n} \sum_{i=1}^{n} K\left(Z_i/h_n\right)\left(\mu(Z_i) - \mu(0)\right) = \int_{-1}^{1} u(x)K(x)dx \cdot f(0) + o_p(1)$$

As such, we conclude that $\frac{1}{h_n} \widetilde{\mathbf{Q}}_{1,5} = \int_{-1}^{1} u(x)K(x)dx \cdot f(0) + o_p(1)$. Similar arguments apply with the remaining components of $\widetilde{X}_i$. Note that $\nu(xh_n) - \nu(0) = \max\left\{-1, \min\left\{1, \frac{x}{c}\right\}\right\} =:$

$v(x)$ is exact. This yields:

$$\frac{1}{h_n}\widetilde{\mathbf{Q}}_{1:8,5:8} = f(0)\cdot$$

$$\begin{pmatrix}
\int_{-1}^{1} u(x)K(x)dx + o_p(1) & h_n\left(\int_{-1}^{1} xu(x)K(x)dx + o_p(1)\right) & \int_{-1}^{1} v(x)K(x)dx + o_p(1) & h_n\left(\int_{-1}^{1} xv(x)K(x)dx + o_p(1)\right) \\
h_n\left(\int_{-1}^{1} xu(x)K(x)dx + o_p(1)\right) & h_n^2\left(\int_{-1}^{1} x^2u(x)K(x)dx + o_p(1)\right) & h_n\left(\int_{-1}^{1} xv(x)K(x)dx + o_p(1)\right) & h_n^2\left(\int_{-1}^{1} x^2v(x)K(x)dx + o_p(1)\right) \\
\int_{0}^{1} u(x)K(x)dx + o_p(1) & h_n\left(\int_{0}^{1} xu(x)K(x)dx + o_p(1)\right) & \int_{0}^{1} v(x)K(x)dx + o_p(1) & h_n\left(\int_{0}^{1} xv(x)K(x)dx + o_p(1)\right) \\
h_n\left(\int_{0}^{1} xu(x)K(x)dx + o_p(1)\right) & h_n^2\left(\int_{0}^{1} x^2u(x)K(x)dx + o_p(1)\right) & h_n\left(\int_{0}^{1} xv(x)K(x)dx + o_p(1)\right) & h_n^2\left(\int_{0}^{1} x^2v(x)K(x)dx + o_p(1)\right) \\
\int_{-1}^{1} u^2(x)K(x)dx + o_p(1) & h_n\left(\int_{-1}^{1} xu^2(x)K(x)dx + o_p(1)\right) & \int_{-1}^{1} u(x)v(x)K(x)dx + o_p(1) & h_n\left(\int_{-1}^{1} xu(x)v(x)K(x)dx + o_p(1)\right) \\
* & h_n^2\left(\int_{-1}^{1} x^2u^2(x)K(x)dx + o_p(1)\right) & h_n\left(\int_{-1}^{1} xu(x)v(x)K(x)dx + o_p(1)\right) & h_n^2\left(\int_{-1}^{1} x^2u(x)v(x)K(x)dx + o_p(1)\right) \\
* & * & \int_{-1}^{1} v^2(x)K(x)dx + o_p(1) & h_n\left(\int_{-1}^{1} xv^2(x)K(x)dx + o_p(1)\right) \\
* & * & * & h_n^2\left(\int_{-1}^{1} x^2v^2(x)K(x)dx + o_p(1)\right)
\end{pmatrix}$$

Invertibility of $\widetilde{\mathbf{Q}}$ follows from observing that $v(x)$ is piecewise linear (but not linear) while $u(x)$ is an infinite sum of piecewise polynomials given our assumed identification conditions.

Next, consider $\mathbf{B}_n$. Since $m^+, m^-, \delta$ and $\gamma$ are all twice continuously differentiable, $\rho_i = O(h_n^2)$. Write:

$$\frac{1}{h_n}\mathbf{B}_n = \frac{1}{nh_n}\sum_{i=1}^{n} K(Z_i/h_n)\,\widetilde{X}_i\rho_i = \frac{1}{nh_n}\sum_{i=1}^{n} K(Z_i/h_n)\,X_i\rho_i + \frac{1}{nh_n}\sum_{i=1}^{n} K(Z_i/h_n)\left(\widetilde{X}_i - X_i\right)\rho_i$$

Now,

$$\frac{1}{nh_n}\sum_{i=1}^{n} K(Z_i/h_n)\left(\widetilde{X}_i - X_i\right)\rho_i = O(h_n^2)\cdot\frac{1}{nh_n}\sum_{i=1}^{n} K(Z_i/h_n)\left(\widetilde{X}_i - X_i\right)$$

$$= O(h_n^2)\cdot o_p(1)\cdot\frac{1}{nh_n}\sum_{i=1}^{n} K(Z_i/h_n)\,\iota = o_p(h_n^2)$$

Meanwhile, by the preceding analysis for $\widetilde{\mathbf{Q}}$,

$$\frac{1}{nh_n}\sum_{i=1}^{n}K\left(Z_i/h_n\right)X_i\rho_i = O(h_n^2)\cdot\begin{pmatrix}1+o_p(1)\\o_p(h_n^2)\\\int_0^1 K(x)dx+o_p(1)\\h_n\left(\int_0^1 xK(x)dx+o_p(1)\right)\\\int_{-1}^1 u(x)K(x)dx+o_p(1)\\h_n\left(\int_{-1}^1 xu(x)K(x)dx+o_p(1)\right)\\\int_{-1}^1 v(x)K(x)dx+o_p(1)\\h_n\left(\int_{-1}^1 xv(x)K(x)dx+o_p(1)\right)\end{pmatrix}\quad\Rightarrow\quad\frac{1}{h_n}\mathbf{B}_n=\begin{pmatrix}o_p(h_n)\\o_p(h_n^2)\\o_p(h_n)\\o_p(h_n^2)\\o_p(h_n)\\o_p(h_n^2)\\o_p(h_n)\\o_p(h_n^2)\end{pmatrix}.$$

Next, consider $\widetilde{\mathbf{T}}_X$. By Lemma 21, we again have that:

$$\frac{1}{h_n}\widetilde{\mathbf{T}}_X = \frac{1}{n}\sum_{i=1}^{n}K\left(Z_i/h_n\right)X_i\left(X_i-\widetilde{X}_i\right)'\beta + \frac{1}{n}\sum_{i=1}^{n}K\left(Z_i/h_n\right)\left(X_i-\widetilde{X}_i\right)\left(X_i-\widetilde{X}_i\right)'\beta$$

$$= \begin{pmatrix}o_p(h_n) & o_p(h_n^2) & o_p(h_n) & o_p(h_n^2) & o_p(h_n) & o_p(h_n^2) & o_p(h_n) & o_p(h_n^2)\end{pmatrix}'.$$

Finally, consider $\widetilde{\mathbf{T}}_\varepsilon$:

$$\frac{1}{h_n}\widetilde{\mathbf{T}}_\varepsilon = \frac{1}{nh_n}\sum_{i=1}^{n}K\left(Z_i/h_n\right)X_i\varepsilon_i + \frac{1}{nh_n}\sum_{i=1}^{n}K\left(Z_i/h_n\right)\left(X_i-\widetilde{X}_i\right)\varepsilon_i$$

Standard arguments based on the continuous differentiability of $\sigma^2(z)$ yields:

$$\frac{1}{nh_n}\sum_{i=1}^{n}K\left(Z_i/h_n\right)X_i\varepsilon_i = O_p\left(\frac{1}{\sqrt{nh_n}}\right)\cdot\begin{pmatrix}O_p(1) & O_p(h_n) & O_p(1) & O_p(h_n) & O_p(1) & O_p(h_n) & O_p(1) & O_p(h_n)\end{pmatrix}'$$

$$= \begin{pmatrix}o_p(h_n) & o_p(h_n^2) & o_p(h_n) & o_p(h_n^2) & o_p(h_n) & o_p(h_n^2) & o_p(h_n) & o_p(h_n^2)\end{pmatrix}'\quad\text{since } h_n\gg n^{-1/3}.$$

By another application of Lemma 21,

$$\frac{1}{nh_n}\sum_{i=1}^{n}K\left(Z_i/h_n\right)\left(X_i-\widetilde{X}_i\right)\varepsilon_i = o(h_n)\cdot\begin{pmatrix}0 & 0 & 0 & 0 & O_p\left(\frac{1}{\sqrt{nh_n}}\right) & O_p\left(\frac{h_n}{\sqrt{nh_n}}\right) & 0 & 0\end{pmatrix}'\quad\text{w.p.a. 1}$$

$$\Rightarrow\quad\frac{1}{h_n}\widetilde{\mathbf{T}}_\varepsilon = \begin{pmatrix}o_p(h_n) & o_p(h_n^2) & o_p(h_n) & o_p(h_n^2) & o_p(h_n) & o_p(h_n^2) & o_p(h_n) & o_p(h_n^2)\end{pmatrix}'.$$

Given the above analysis, let $W_n = \text{diag}(1,h_n,1,h_n,1,h_n,1,h_n)$ and write $\overline{\mathbf{Q}} = W_n\widetilde{\mathbf{Q}}W_n$.

44

The entries of $\overline{\mathbf{Q}}$ converge to non-zero values in probability as $n \to \infty$ and it is asymptotically full-rank. Then, $\widetilde{\beta} - \beta = W_n^{-1} \overline{\mathbf{Q}}^{-1} W_n^{-1} \left( \widetilde{\mathbf{T}}_\rho + \widetilde{\mathbf{T}}_X + \widetilde{\mathbf{T}}_\varepsilon \right)$ and

$$W_n^{-1} \frac{1}{h_n} \mathbf{B}_n = o_p(h_n) \cdot \iota \quad \Rightarrow \quad \overline{\mathbf{Q}}^{-1} W_n^{-1} \frac{1}{h_n} \mathbf{B}_n = o_p(h_n) \cdot \iota \quad \Rightarrow \quad W_n^{-1} \overline{\mathbf{Q}}^{-1} W_n^{-1} \frac{1}{h_n} \mathbf{B}_n = o_p(1) \cdot \iota \,.$$

By a similar argument for $\widetilde{\mathbf{T}}_X$ and $\widetilde{\mathbf{T}}_\varepsilon$, we are done.

## A.3 Proof of Theorem 4

Henceforth, we denote $h := h_n$ and $r := r_n$ and drop the subscript $n$ unless there is risk of ambiguity. Use (A.2) and write

$$\widetilde{\mathbf{T}} = \widetilde{\mathbf{Q}} \left( \widetilde{\beta} - \beta - \mathbf{B}_n \right) \quad \text{and} \quad \widetilde{\mathbf{T}}_* = \frac{1}{n} \sum_{i=1}^n \left[ K \left( \frac{Z_i}{h} \right) X_i \varepsilon_i + \dot{\eta}_X \left( \xi_i \right) \right], \qquad (17)$$

where

$$\dot{\eta}_X \left( \xi_i \right) = \eta_X \left( \xi_i \right) - \mathbf{E} \left[ \eta_X \left( \xi_1 \right) \right],$$

and $\eta_X(\cdot)$ is given in (3). Note that $\mathbf{E} \left[ \sqrt{n} \widetilde{\mathbf{T}}_* \right] = \mathbf{0}$ and $\text{Var} \left[ \sqrt{n} \widetilde{\mathbf{T}}_* \right] = \mathbf{\Omega}$. The claim (1) is equivalent to $\mathbf{\Omega}^{-1/2} \sqrt{n} \widetilde{\mathbf{T}} \overset{d}{\to} \mathbf{N} \left( 0, \mathbb{I}_8 \right)$ and by Slutsky's theorem, is implied by

$$\mathbf{\Omega}^{-1/2} \sqrt{n} \widetilde{\mathbf{T}}_* \overset{d}{\to} \mathbf{N} \left( 0, \mathbb{I}_8 \right) \quad \text{and} \quad \mathbf{\Omega}^{-1/2} \sqrt{n} \left( \widetilde{\mathbf{T}} - \widetilde{\mathbf{T}}_* \right) = o_p(1). \qquad (18)$$

The asymptotic normality result is due to Lemma 10. The asymptotic negligibility result is due to Lemma 11. $\qquad \square$

# Supplemental Appendix

# B Simulations

This section compares the performance of LSR to prevailing methods popular for RDD in the context of a simple DGP. We first show that LLR with a naively chosen bandwidth exhibits behavior best described by the $r_n \propto h_n$ case in Theorem 1. We then compare the performance of LSR with the methods of Armstrong and Kolesár (2020) (AK hereinafter) and Calonico et al. (2014) (CCT hereinafter) which involve more sophisticated rules for bandwidth choice.

Consider the data-generating process described in Section 2 with the following values:

$$m^+(z) = 2 \quad , \quad m^-(z) = 1 \quad , \quad \tau_{\text{DTE}} = 1$$

$$\delta(z) = \delta \quad , \quad \gamma(z) = \gamma$$

The researcher observes

$$Y_i = D_i Y_d^+(Z_i) + (1 - D_i) Y_d^-(Z_i) + \varepsilon_i \quad , \quad \varepsilon_i \sim \text{N}(0, \sigma^2)$$

where $\sigma = 0.2$. The values of $n$, $r_n$, $\delta$ and $\gamma$ will vary depending on the set up. We estimate LLR using a triangular kernel. LLR has no bias in this setting. We set $M_m = M_\delta = M_\gamma = 0.5$ for conservative inference. The following results are based on 1,000 draws for each set of parameter values.

## B.1 Accuracy of the Preferred Regime

We first present simulations pertaining to Theorem 1. Figure 1 plots the mean of $\hat{\tau}_{\text{RDD}}$ under various values of $\delta$ and $\gamma$ when $n = 2000$ and bandwidth is chosen based on the rule-of-thumb in (5) ($\approx 0.22$). We see that as $c = r_n/h_n$ decreases to 0, the LLR estimate goes from $\tau_{\text{DIR}} = 1$ to $\tau_{\text{TOT}} = \frac{\tau_{\text{DIR}} + \gamma}{1 - \delta}$. This is predicted by the $r_n \propto h_n$ case in our Theorem 1. Approximations under the regime $r_n \gg h_n$ or $r_n \ll h_n$ would predict either end point but would miss the transition. As such, our preferred regime $r_n \propto h_n$ leads to a more accurate

approximation.

In practice, LLR is implemented based on either AK or CCT, each of which entail its own rules for bandwidth choice. Figures 2 and 3 plots the same diagrams when bandwidths are chosen based on the recommendations of AK and CCT respectively. We see that both methods are more robust to spillovers than naive LLR in that the estimates are very close to $\tau_{\text{DIR}}$. Spillovers only distort these estimators when $r_n/h_n \leq 0.2$. These plots do not give the full picture in that they omit MSE and coverage. We will return to these points below. Interestingly, distortion under AK seems to be increasing in the amount of spillovers. On the other hand, CCT seems to do worse under intermediate values of spillovers, with little distortion when spillovers are large.

The above simulations suggest that an alternative is needed to simple LLR under spillovers. AK and CCT appear more robust, but to our knowledge, this performance is not currently supported by theory, motivating consideration of LSR.
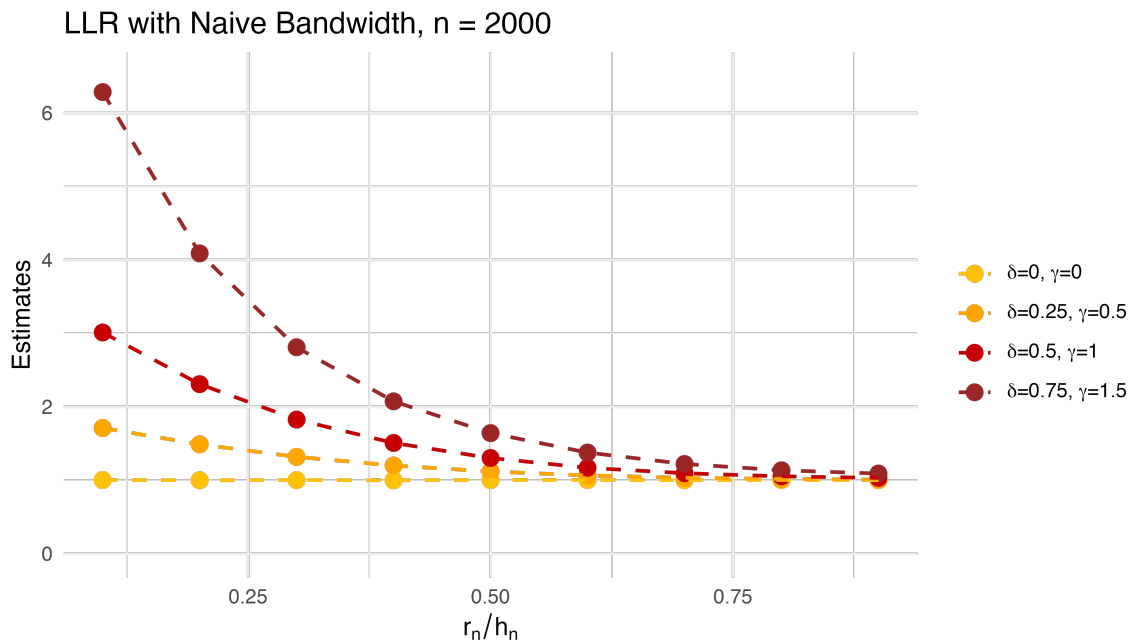


Figure 1: Estimates of LLR when bandwidth is chosen based on the rule-of-thumb (5). The behavior of $\hat{\tau}_{\text{RDD}}$ is best described by the $r_n \propto h_n$ case in Theorem 1.
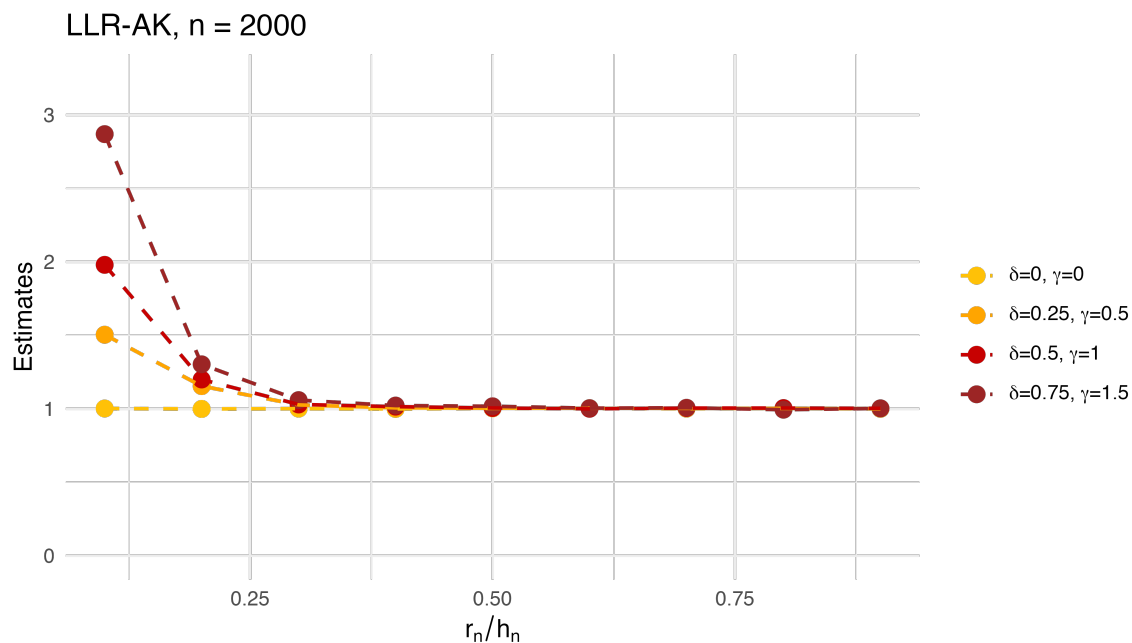
Figure 2: Estimates of LLR when bandwidth is chosen based on Armstrong and Kolesár (2018), as implemented in the R package `RDHonest`.
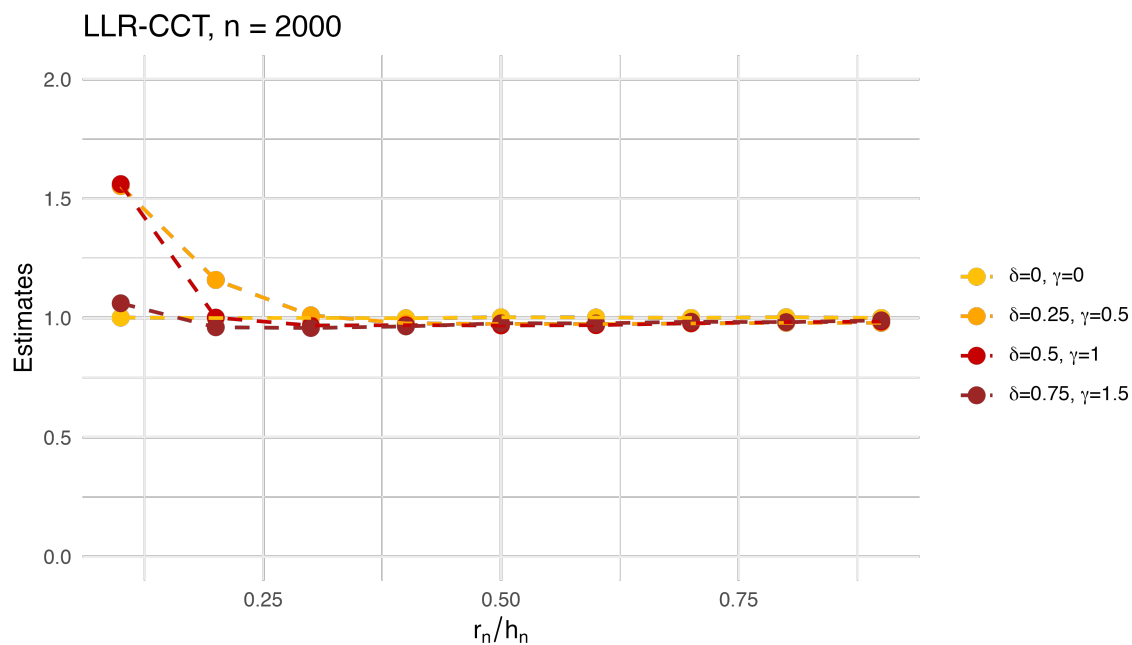


Figure 3: Estimates of LLR when bandwidth is chosen based on Calonico et al. (2014), as implemented in the R package `rdrobust`.

## B.2 Local Spillover Regression

This section investigates the performance of LSR for recovering the various parameters of interest. We first compare its MSE and coverage for $\tau_{\mathrm{DIR}}$ to various implementations of LLR. We then consider estimation of $\delta$ and $\gamma$. The remainder of this section maintains $r_n = 0.05$ fixed, with $n \in \{500, 2000, 10000\}$. This sequence of models is better described by the asymptotic regime $r_n \gg h_n$, which favors LLR over LSR.

Table 2 and 3 present results on $\tau_{\mathrm{DIR}}$. In the first panel in Table 2, $\delta = \gamma = 0$. LLR (with naive bandwidth given in (5)), AK and CCT all perform well. In this setting, $\tilde{\delta}(0)$ and $\tilde{\gamma}(0)$ are inconsistent. However, $\tilde{\tau}_{\mathrm{DIR}}$ is consistent, although it has MSE that can be close to 5 times higher than methods designed for settings with no spillovers. Our method leads to conservative coverage even relative to AK. We conjecture that the conservatism can be reduced with data-driven choices of $M_m, M_\delta$ and $M_\gamma$, similar to the approach in AK.

Panel 2 in Table 2 considers the case where $\delta = 0.25$ and $\gamma = 0.50$. As we explained at the start of this section, this sequence of models is not strictly comparable since $r_n$ is fixed as $n \to \infty$ and $h_n \to 0$. The asymptotics here favors LLR over LSR since $c = r_n/h_n \to \infty$ implies that $\hat{\tau}_{\mathrm{RDD}} \overset{p}{\to} \tau_{\mathrm{DIR}}$. As such, we see the largest gains of LSR over various implementations of LLR for smaller values of $n$. When $n = 500$, LSR has MSE that is $1/2$ of AK and $1/3$ of CCT. The two methods also have coverage of 60% and 40% respectively, whereas LSR has coverage that is consistently above 90%. LSR-CV has relatively higher MSE and lower coverage but performance appears reasonable. We also see that it appears to deliver consistent estimates for $r_n$.

The picture remains similar when we consider Table 3. The top panel features intermediate levels of spillovers with $\delta = 0.5, \gamma = 1$ while the bottom panel features relatively high levels of spillovers with $\delta = 0.75, \gamma = 1.5$. Comparing across panels, we see that the MSE of AK increases as spillovers increase. However, CCT does the worst under small values of spillovers. Performance of LSR with known radius seems to suffer under high levels of spillovers, but LSR-CV surprisingly does better.

We next turn to results on $\tilde{\gamma}(0)$ and $\tilde{\delta}(0)$, presented in Table 4. The top panel considers the case when $\delta = \gamma = 0$. Here, our estimators are not consistent for $\gamma(0)$ and $\delta(0)$ and MSE indeed appears to diverge. However, it is encouraging that coverage is 1, even if our theory

49

cannot speak to inference in this case. The remaining panels in Table 4 covers cases when $\delta$ and $\gamma$ are not 0. We see that estimation of these parameters is more challenging that for $\tau_{\mathrm{DIR}}$, although we achieve reasonable performance with sufficiently large sample size. The confidence intervals exhibit considerable over-coverage, possibly because of upward bias in our various plug-in estimators in addition to the built-in conservativeness of the bias-aware confidence intervals. Finally, we see that estimation of $\delta$ and $\gamma$ is much more sensitive to selection of the radius compared to $\tau_{\mathrm{DIR}}$, although it can still be a compelling alternative when $r_n$ is unknown.

In sum, our simulations show that naive LLR – and to a lesser extent AK and CCT – exhibits the phase transition described in Theorem 1. LSR deliver more precise estimates of $\tau_{\mathrm{DIR}}$, particularly for smaller sample sizes. It also yields estimates of $\delta(0)$ and $\gamma(0)$, which the other approaches cannot provide. These parameters are interesting in and of themselves, although their estimation requires more assumptions and larger sample sizes. Finally, cross-validation seems to be a reasonable approach to selecting radius, even though it is not without loss. Taken together, LSR and LSR-CV provides a useful set of tools for researchers who are concerned about spillovers in the RDDs .

# C    Auxillary Lemmas

This section states lemmas which are used in Appendix A. Proofs of these lemmas can be found in Auerbach et al. (2025b). We sometimes suppress the dependence of variables (e.g. $r_n, h_n$) on $n$ to reduce notational clutter.

## C.1    Preliminaries

The following definitions and results will be useful. In particular, Lemma 9 is the technical core of the results in Sections 3 and 4.

**Definition 5.** Let the spillover operator $G_n$ be defined as

$$(G_n \circ f)(z) = \frac{\delta(z)}{|R_n(z)|} \int \mathbf{1}\left\{|u - z| < r_n\right\} f(z)\frac{1}{2}dz$$

|  | $n$ | Method | Radius | Bandwidth | Eff. Obs. | MSE | Coverage |
|---|---|---|---|---|---|---|---|
| $\delta = 0$, $\gamma = 0$ | 500 | LLR | - | 0.289 | 72.1 | 0.005 | - |
|  |  | AK | - | 0.175 | 43.8 | 0.011 | 0.959 |
|  |  | CCT | - | 0.303 | 75.7 | 0.009 | 0.932 |
|  |  | LSR | 0.050 | 0.622 | 155.4 | 0.032 | 0.991 |
|  |  | LSR-CV | 0.105 | 0.645 | 161.2 | 0.032 | 0.996 |
|  | 2000 | LLR | - | 0.219 | 218.7 | 0.002 | - |
|  |  | AK | - | 0.176 | 176.5 | 0.002 | 0.973 |
|  |  | CCT | - | 0.308 | 308.2 | 0.002 | 0.951 |
|  |  | LSR | 0.050 | 0.489 | 489.2 | 0.008 | 0.997 |
|  |  | LSR-CV | 0.089 | 0.537 | 537.3 | 0.009 | 0.999 |
|  | 10000 | LLR | - | 0.158 | 792.4 | 0.000 | - |
|  |  | AK | - | 0.176 | 878.3 | 0.000 | 0.971 |
|  |  | CCT | - | 0.309 | 1543.3 | 0.000 | 0.947 |
|  |  | LSR | 0.050 | 0.330 | 1649.1 | 0.002 | 0.999 |
|  |  | LSR-CV | 0.092 | 0.386 | 1931.3 | 0.002 | 0.999 |
| $\delta = 0.25$, $\gamma = 0.50$ | 500 | LLR | - | 0.289 | 72.1 | 0.305 | - |
|  |  | AK | - | 0.122 | 30.4 | 0.062 | 0.628 |
|  |  | CCT | - | 0.164 | 41.0 | 0.098 | 0.383 |
|  |  | LSR | 0.050 | 0.559 | 139.8 | 0.032 | 0.992 |
|  |  | LSR-CV | 0.079 | 0.579 | 144.7 | 0.061 | 0.975 |
|  | 2000 | LLR | - | 0.219 | 218.7 | 0.186 | - |
|  |  | AK | - | 0.094 | 93.7 | 0.014 | 0.750 |
|  |  | CCT | - | 0.102 | 102.3 | 0.017 | 0.632 |
|  |  | LSR | 0.050 | 0.396 | 396.5 | 0.006 | 0.985 |
|  |  | LSR-CV | 0.058 | 0.420 | 420.1 | 0.014 | 0.958 |
|  | 10000 | LLR | - | 0.158 | 792.4 | 0.087 | - |
|  |  | AK | - | 0.070 | 348.7 | 0.002 | 0.922 |
|  |  | CCT | - | 0.060 | 301.9 | 0.002 | 0.937 |
|  |  | LSR | 0.050 | 0.291 | 1457.4 | 0.001 | 0.982 |
|  |  | LSR-CV | 0.050 | 0.299 | 1494.5 | 0.002 | 0.947 |

Table 2: MSE and coverage for $\tau_{\text{DIR}}$ under zero to low spillovers. LLR uses the naive bandwidth (5) and we do not consider inference. AK is the implementation of Armstrong and Kolesár (2020). CCT is the implementation of Calonico et al. (2014). LSR is local spillover regression when the true radius is known. LSR-CV selects the radius by cross-validation, as described in Section 4.3.1. Both variants of LSR choose bandwidth using the procedure described in Section 4.3.2.

| | $n$ | Method | Radius | Bandwidth | Eff. Obs. | MSE | Coverage |
|---|---|---|---|---|---|---|---|
| $\delta = 0.50$, $\gamma = 1.00$ | 500 | LLR | - | 0.289 | 72.1 | 2.281 | - |
| | | AK | - | 0.094 | 23.6 | 0.107 | 0.659 |
| | | CCT | - | 0.083 | 20.6 | 0.064 | 0.840 |
| | | LSR | 0.050 | 0.501 | 125.2 | 0.034 | 0.994 |
| | | LSR-CV | 0.057 | 0.499 | 124.7 | 0.061 | 0.966 |
| | 2000 | LLR | - | 0.219 | 218.7 | 1.328 | - |
| | | AK | - | 0.074 | 74.0 | 0.018 | 0.844 |
| | | CCT | - | 0.049 | 49.5 | 0.009 | 0.933 |
| | | LSR | 0.050 | 0.418 | 417.6 | 0.008 | 0.999 |
| | | LSR-CV | 0.050 | 0.421 | 421.2 | 0.013 | 0.942 |
| | 10000 | LLR | - | 0.158 | 792.4 | 0.583 | - |
| | | AK | - | 0.053 | 266.9 | 0.002 | 0.958 |
| | | CCT | - | 0.031 | 155.8 | 0.003 | 0.925 |
| | | LSR | 0.050 | 0.384 | 1918.0 | 0.002 | 0.999 |
| | | LSR-CV | 0.049 | 0.386 | 1929.1 | 0.002 | 0.996 |
| $\delta = 0.75$, $\gamma = 1.50$ | 500 | LLR | - | 0.289 | 72.1 | 13.106 | - |
| | | AK | - | 0.088 | 22.0 | 0.270 | 0.746 |
| | | CCT | - | 0.051 | 12.7 | 0.068 | 0.992 |
| | | LSR | 0.050 | 0.653 | 163.4 | 0.273 | 1.000 |
| | | LSR-CV | 0.056 | 0.608 | 152.0 | 0.166 | 0.990 |
| | 2000 | LLR | - | 0.219 | 218.7 | 7.059 | - |
| | | AK | - | 0.066 | 65.7 | 0.035 | 0.946 |
| | | CCT | - | 0.029 | 29.2 | 0.017 | 0.931 |
| | | LSR | 0.050 | 0.567 | 566.9 | 0.039 | 1.000 |
| | | LSR-CV | 0.049 | 0.563 | 563.4 | 0.036 | 1.000 |
| | 10000 | LLR | - | 0.158 | 792.4 | 2.787 | - |
| | | AK | - | 0.042 | 211.8 | 0.004 | 0.975 |
| | | CCT | - | 0.020 | 99.9 | 0.004 | 0.932 |
| | | LSR | 0.050 | 0.422 | 2107.9 | 0.007 | 1.000 |
| | | LSR-CV | 0.049 | 0.424 | 2118.6 | 0.007 | 1.000 |

Table 3: MSE and coverage for $\tau_{\mathrm{DIR}}$ under moderate to high spillovers. LLR uses the naive bandwidth (5) and we do not consider inference. AK is the implementation of Armstrong and Kolesár (2020). CCT is the implementation of Calonico et al. (2014). LSR is local spillover regression when the true radius is known. LSR-CV selects the radius by cross-validation, as described in Section 4.3.1. Both variants of LSR choose bandwidth using the procedure described in Section 4.3.2.

| | $n$ | Method | Radius | Bdwth | Eff. Obs. | $\delta(0)$ | | $\gamma(0)$ | |
|---|---|---|---|---|---|---|---|---|---|
| | | | | | | MSE | Cov. | MSE | Cov. |
| $\delta = 0$, $\gamma = 0$ | 500 | LSR | 0.050 | 0.622 | 155.4 | 0.837 | 1.000 | 0.945 | 1.000 |
| | | LSR-CV | 0.105 | 0.645 | 161.2 | 5.583 | 1.000 | 5.738 | 1.000 |
| | 2000 | LSR | 0.050 | 0.489 | 489.2 | 1.485 | 1.000 | 1.536 | 1.000 |
| | | LSR-CV | 0.089 | 0.537 | 537.3 | 6.587 | 1.000 | 6.650 | 1.000 |
| | 10000 | LSR | 0.050 | 0.330 | 1649.1 | 2.865 | 1.000 | 2.887 | 1.000 |
| | | LSR-CV | 0.092 | 0.386 | 1931.3 | 9.726 | 1.000 | 9.768 | 1.000 |
| $\delta = 0.25$, $\gamma = 0.50$ | 500 | LSR | 0.050 | 0.559 | 139.8 | 0.455 | 1.000 | 1.732 | 1.000 |
| | | LSR-CV | 0.079 | 0.579 | 144.7 | 1.567 | 1.000 | 5.623 | 1.000 |
| | 2000 | LSR | 0.050 | 0.396 | 396.5 | 0.200 | 1.000 | 0.720 | 1.000 |
| | | LSR-CV | 0.058 | 0.420 | 420.1 | 0.644 | 1.000 | 2.258 | 1.000 |
| | 10000 | LSR | 0.050 | 0.291 | 1457.4 | 0.040 | 1.000 | 0.134 | 1.000 |
| | | LSR-CV | 0.050 | 0.299 | 1494.5 | 0.120 | 1.000 | 0.419 | 1.000 |
| $\delta = 0.50$, $\gamma = 1.00$ | 500 | LSR | 0.050 | 0.501 | 125.2 | 0.095 | 0.982 | 1.309 | 0.973 |
| | | LSR-CV | 0.057 | 0.499 | 124.7 | 0.257 | 0.972 | 3.333 | 0.959 |
| | 2000 | LSR | 0.050 | 0.418 | 417.6 | 0.020 | 0.998 | 0.255 | 0.996 |
| | | LSR-CV | 0.050 | 0.421 | 421.2 | 0.051 | 0.992 | 0.642 | 0.985 |
| | 10000 | LSR | 0.050 | 0.384 | 1918.0 | 0.003 | 0.997 | 0.037 | 0.997 |
| | | LSR-CV | 0.049 | 0.386 | 1929.1 | 0.010 | 0.997 | 0.131 | 0.997 |
| $\delta = 0.75$, $\gamma = 1.50$ | 500 | LSR | 0.050 | 0.653 | 163.4 | 0.015 | 1.000 | 1.179 | 1.000 |
| | | LSR-CV | 0.056 | 0.608 | 152.0 | 0.051 | 0.987 | 3.488 | 0.985 |
| | 2000 | LSR | 0.050 | 0.567 | 566.9 | 0.003 | 1.000 | 0.194 | 1.000 |
| | | LSR-CV | 0.049 | 0.563 | 563.4 | 0.009 | 1.000 | 0.746 | 1.000 |
| | 10000 | LSR | 0.050 | 0.422 | 2107.9 | 0.001 | 1.000 | 0.046 | 1.000 |
| | | LSR-CV | 0.049 | 0.424 | 2118.6 | 0.002 | 1.000 | 0.195 | 1.000 |

Table 4: MSE and coverage for $\delta(0)$ and $\gamma(0)$. LSR is local spillover regression when the true radius is known. LSR-CV selects the radius by cross-validation, as described in Section 4.3.1. Both variants of LSR choose bandwidth using the procedure described in Section 4.3.2.

where $|R_n(z)| = \int \mathbf{1}\{|u - z| < r_n\} \frac{1}{2} dz$.

**Definition 6.** For a function $g : [-1, 1] \to \mathbf{R}$, let its sup-norm be denoted $\|g\| := \sup_{z \in [-1,1]} |g(z)|$.

Let the "structural" and "reduced form" functions be, respectively,

**Definition 7.**

$$m(z) := \begin{cases} m^+(z) & \text{if } z \geq 0 \\ m^-(z) & \text{if } z < 0 \end{cases},$$

$$g_n(z) := \left( (I - G_n)^{-1} \circ m \right)(z) .$$

It will also be useful to split $m(\cdot)$ into the part that is Lipschitz continuous and the part that is not:

**Definition 8.** Let $m(x) = m^c(x) + \mathbf{1}\{z \geq 0\}\tau_d$ where

$$m^c(z) = \begin{cases} m^+(z) - \tau_d & \text{if } z \geq 0 \\ m^-(z) & \text{otherwise} \end{cases}$$

is Lipschitz continuous with Lipschitz constant $C$.

**Definition 9.** Let the difference in the spillover neighborhoods at $xh_n$ and 0 be:

$$R^+(xh_n) := R(xh_n) \setminus R(0)$$

$$R^-(xh_n) := R(0) \setminus R(xh_n)$$

**Definition 10.** Let $\tilde{d}$ be

$$\tilde{d}(a) = \begin{cases} 0 & \text{if } a \leq -1 \\ \frac{a+1}{2} & \text{if } -1 \leq a \geq 1 \\ 1 & \text{if } a \geq -1 \end{cases}$$

Observe that if $\delta(z) = 1$, $\tilde{d}(a/r_n) = (G_n \circ \mathbf{1}\{z \geq 0\})(a)$.

**Definition 11.** Let $G_\infty$ be the pointwise limit of $G_n$ when $r_n \to 0$. That is,

$$(G_\infty \circ g)(z) = \lim_{n \to \infty} \frac{\delta(z)}{R_n(z)} \int_{R(z)} g(y)dy = \delta(z)g(z) \;,$$

where the last equality follows from the Lebesgue Differentiation Theorem. Furthermore, let $g_\infty(z) = \left((I - G_\infty)^{-1} \circ g\right)(z) = \frac{g(z)}{1 - \delta(z)}$.

**Lemma 1.** *Suppose $\|f - g\| \leq \varepsilon$. Then $\|(G_n \circ f) - (G_n \circ g)\| \leq \bar{\delta}\varepsilon$ . In particular, if $\|g(z)\| < \varepsilon$, then $\left\|G_n^k \circ g\right\| < \bar{\delta}^k \varepsilon$.*

**Lemma 2.** *Under Assumptions 1 and 2, suppose $r_n \to 0$. Then, there exists a function $\lambda$ such that for $a \neq 0$,*

$$|((I - G_n)^{-1} \circ \mathbf{1}\{z \geq 0\})(ar_n) - \lambda(a)| \leq \frac{2C_\delta r_n}{a\left(1 - \bar{\delta}\right)^2} + \frac{\bar{\delta}^{1/r_n - |a|}}{1 - \bar{\delta}} \;,$$

*where $\lambda(a) > 0$ if $\delta(z) > 0$ for all $z \in \mathcal{Z}$. Furthermore, if $\delta(0) \neq 0$, then*

$$\left|\left((I - G_n)^{-1} \circ \gamma(z)\tilde{d}(z/r_n)\right)(ar_n) - \frac{\gamma(0)}{\delta(0)}(\lambda(a) - \mathbf{1}\{a \geq 0\})\right| \leq$$

$$\frac{2C_\delta r_n}{a\left(1 - \bar{\delta}\right)^2} + \frac{\bar{\delta}^{1/r_n - |a|}}{1 - \bar{\delta}} + \frac{C_\gamma \bar{\delta}}{1 - \bar{\delta}} \cdot ar_n \;.$$

*If $\delta(0) = 0$, then*

$$\left((I - G_n)^{-1} \circ \gamma(z)\tilde{d}(z/r_n)\right)(ar_n) = \gamma(ar_n)\tilde{d}(a) = \gamma(0)\tilde{d}(a) + O(r_n) \;.$$

In fact, we know more about the structure of $\lambda(a)$. Let $G_*$ be operator that smooths over $f \in \mathbf{R} \to \mathbf{R}$:

$$(G_* \circ f)(z) = \frac{1}{2} \int \mathbf{1}\{|u - z| < 1\} f(z)\frac{1}{2}dz$$

Then $\lambda(a)$ is approximately $((I - \delta(0)G_*)^{-1} \circ \mathbf{1}\{z \geq 0\})(a)$. Note that $\lambda$ depends on $\delta$, though this is suppressed in the notation. Given Lemma 2, define:

**Definition 12.** Let $\tilde{R}^+(x) = \left[\max\left\{1, \frac{2x}{c} - 1\right\}, 1 + \frac{2x}{c}\right]$ and $\tilde{R}^-(x) = \left[-1, \min\left\{1, \frac{2x}{c} - 1\right\}\right]$.

Define:

$$\lambda^+(x) := \frac{1}{|\tilde{R}^+(x)|} \int_{\tilde{R}^+(x)} \lambda(a)\, dF(a) \quad , \quad \lambda^-(x) := \frac{1}{|\tilde{R}^-(x)|} \int_{\tilde{R}^-(x)} \lambda(a)\, dF(a)$$

$$\tilde{\lambda}^+(x) := \frac{1}{|\tilde{R}^+(x)|} \int_{\tilde{R}^+(x)} \lambda(a) - \mathbf{1}\{a \geq 0\}\, dF(a) \quad , \quad \tilde{\lambda}^-(x) := \frac{1}{|\tilde{R}^-(x)|} \int_{\tilde{R}^-(x)} \lambda(a) - \mathbf{1}\{a \geq 0\}\, dF(a) \, .$$

And:

$$\Lambda^+_{p,q,s} := \int_0^1 x^p \left(\lambda^+(x) - \lambda^-(x)\right)^q K^s(x)ds \quad , \quad \Lambda^-_{p,q,s} := \int_{-1}^0 x^p \left(\lambda^+(-x) - \lambda^-(-x)\right)^q K^s(x)ds$$

$$\tilde{\Lambda}^+_{p,q,s} := \int_0^1 x^p \left(\tilde{\lambda}^+(x) - \lambda^-(x)\right)^q K^s(x)ds \quad , \quad \tilde{\Lambda}^-_{p,q,s} := \int_{-1}^0 x^p \left(\tilde{\lambda}^+(-x) - \lambda^-(-x)\right)^q K^s(x)ds$$

$$\Gamma^+_{p,q,s} := \int_0^1 x^p \left(\tilde{d}\left(\frac{x}{c}\right)\right)^q K^s(x)ds \quad , \quad \Gamma^-_{p,q,s} := \int_{-1}^0 x^p \left(\tilde{d}\left(\frac{x}{c}\right)\right)^q K^s(x)ds \, .$$

**Lemma 3.** *Suppose $g$ is Lipschitz on $[-1,1]$ with Lipschitz constant $C$ and $\|g\| \leq M$. Then $(G_n \circ g)$ is Lipschitz on $[-1,1]$ with Lipschitz constant $C_\delta \cdot M + \bar{\delta}C$. Moreover, $(I - G_n)^{-1} \circ g$ is Lipschitz on $[-1,1]$ with Lipschitz constant*

$$\frac{C_\delta \cdot M}{\left(1 - \bar{\delta}\right)^2} + \frac{C}{1 - \bar{\delta}} \, .$$

**Lemma 4.** *Suppose $g$ is an odd function. If $\delta(z)$ is an even function, then $(G_n \circ g)$ is an odd function.*

**Lemma 5.** *$g_n(z) = g_n^c(z) + \tau_d \mathbf{1}\{z \geq 0\}$ where $g_n^c(z)$ is Lipschitz with Lipschitz constant*

$$\frac{C_\delta \cdot (1 + \bar{\delta})M}{\left(1 - \bar{\delta}\right)^2} + \frac{C}{1 - \bar{\delta}} + \frac{C\bar{\delta}}{2r_n\left(1 - \bar{\delta}\right)}$$

**Lemma 6.** *Let $g$ be a Lipschitz continuous function with Lipschitz constant $C$ and suppose $r_n \to 0$. Then:*

$$\sup_{z \in [-1/2, 1/2]} \left|\left((I - G_n)^{-1} \circ g\right)(z) - \left((I - G_\infty)^{-1} \circ g\right)(z)\right| \leq \frac{Cr_n}{2} \frac{\bar{\delta}}{\left(1 - \bar{\delta}\right)^2} \, .$$

**Lemma 7.** *Let $h_n$ be such that $h_n > K r_n$. Let $\tilde{G}_n$ be $G_n$ except with $\delta(z)$ set to $\delta(0)$. Then for all $x \in [-h_n, h_n]$, we have that*

$$\left( G_n^k \circ \mathbf{1}\{z \geq 0\} \right)(x) = \left( \tilde{G}_n^k \circ \mathbf{1}\{z \geq 0\} \right)(x) + \sum_{j=1}^{k} \left( G_n^{k-j} \circ R_j \right)(x)$$

*where*

$$\sup_{x \in [-h_n, h_n]} \left| \left( G_n^{k-j} \circ R_j \right)(x) \right| \leq 2 \bar{\delta}^k C_\delta h_n \ .$$

**Lemma 8.** *Under Assumptions 1 and 2, suppose $r_n \to 0$. The following holds for all $\eta \in (0, 1)$:*

$$\mu_d(0) = \frac{m^+(0) + m^-(0) + \gamma(0)}{2(1 - \delta(0))} + O\left( \frac{C_\delta r_n^\eta}{\left(1 - \bar{\delta}\right)^2} + \frac{\bar{\delta}^{1/r_n^\eta}}{1 - \bar{\delta}} \right) \ .$$

**Lemma 9.** *Under Assumptions 1 and 2, the following approximations are uniform in $x \in [0, 1]$ as $n \to \infty$:*

$$\mu_d(x h_n) - \mu_d(0) = \begin{cases} \frac{x h_n}{r_n} \left( g_1(r) - g_1(-r) + l_1(r) - l_1(-r) + o(1) \right) & \text{if } r_n = r \\ \frac{x h_n}{r_n} \cdot \left( \tau_d \cdot (\lambda(1) - \lambda(-1)) + \gamma(0) \cdot ((\lambda(1) - 1) - \lambda(-1)) + o(1) \right) & \text{if } \frac{h_n}{r_n} \to 0 \\ \min\left\{1, \frac{|x|}{c}\right\} \cdot \left( \tau_d \cdot (\lambda^+(x) - \lambda^-(x)) + \gamma(0) \cdot \left( \tilde{\lambda}^+(x) - \tilde{\lambda}^-(x) \right) + o(1) \right) & \text{if } r_n = \frac{1}{2} c h_n \end{cases} \ .$$

*If $r_n / h_n \to 0$, let $w_n$ be such that $w_n / r_n \to \infty$ and $w_n / h_n \to 0$. Then for $x \in [w_n, 1]$,*

$$\mu_d(x h_n) - \mu_d(0) = \frac{\tau_d + \gamma(0)}{2(1 - \delta(0))} + O\left( h_n + \bar{\delta}^{w_n / r_n} \right) \ .$$

*The same statements hold for $x \in [-1, 0]$ mutatis mutandis.*

Finally, we define the (mixed) incomplete moments:

**Definition 13** (Incomplete Moments). The incomplete moment of order $p$ at 0 is

$$\gamma_{p,s} := \int_0^1 x^p K^s(x) dx$$

57

and the incomplete mixed moment of order $p$ at 0 is

$$\phi(p,q,r,s) := \int_0^1 x^p \left(\min\left\{\frac{x}{c}, 1\right\} \left(\tau_d \cdot \left(\lambda^+(x) - \lambda^-(x)\right) + \gamma(0) \cdot \left(\left(\lambda^+(x) - 1\right) - \lambda^-(x)\right)\right)\right)^q$$
$$\cdot \left(\min\left\{\frac{x}{c}, \frac{1}{2}\right\}\right)^r K^s(x)\, dx$$

Incomplete moments are standard objects in the analysis of local polynomial regressions at the boundary (see e.g. Wand and Jones (1995)).The incomplete mixed moment is so named since also includes terms that arise from the approximations of $\mu_d(Z_i)$ and $\nu_d(Z_i)$ around the cutoff. We suppress the dependence of $\phi$ on $c$ for convenience.

**Lemma 10.** *Under the conditions of Theorem 4, $\boldsymbol{\Omega}^{-1/2}\sqrt{n}\widetilde{\mathbf{T}}_* \overset{\mathrm{d}}{\to} \mathbf{N}\left(0, \mathbb{I}_8\right)$, where $\widetilde{\mathbf{T}}_*$ is as defined in (17).*

**Lemma 11.** *Let the premises of Theorem 4 hold and $\widetilde{\mathbf{T}}$ and $\widetilde{\mathbf{T}}_*$ be as defined in (17). Then $\boldsymbol{\Omega}^{-1/2}\sqrt{n}\left(\widetilde{\mathbf{T}} - \widetilde{\mathbf{T}}_*\right) = o_{\mathrm{p}}(1)$.*

**Lemma 12.** *$T_g$ in (43) has the following representation:*

$$
\begin{aligned}
T_g &= T_{*,g} + U_g + R_g, \\
\text{where} \quad T_{*,g} &= \frac{1}{n}\sum_{i=1}^n \dot{\eta}_{g,1}\left(\xi_i\right), \\
U_g &= \frac{1}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^n \dot{\eta}_{g,2}\left(\xi_i, \xi_j\right) - T_{*,g}, \\
R_g &= \frac{1}{n}\sum_{i=1}^n K\left(Z_i/h\right) g\left(\xi_i\right)\left(\widehat{\mathrm{Rem}}_{i,r}\left(Z_i\right) - \widehat{\mathrm{Rem}}_{i,r}(0)\right).
\end{aligned}
\tag{19}
$$

**Lemma 13.** *Let* $T_{*,g}, U_g, R_g$ *be as in* (19). *Then* $\mathbf{E}\left[T_{*,g}\right] = 0$, $\mathbf{E}\left[U_g\right] = 0$. *Furthermore,*

$$
\mathbf{E}\left[T_{*,g}^2\right] = \frac{\mathbf{E}\left[\dot{\eta}_{g1}\left(\xi_1\right)^2\right]}{n}
$$

$$
\mathbf{E}\left[\dot{\eta}_{g1}\left(\xi_1\right)^2\right] = h\frac{h}{4r}(1 + o(1))\left(\int f(vr)\sigma^2(vr)\varphi_{\alpha,1}(v;h,r)^2\mathrm{d}v + \int \varphi_{\alpha,2}(v;h,r)^2 f(vr)\,\mathrm{d}v\right)
$$

$$
\mathbf{E}\left[\left|\dot{\eta}_{g1}\left(\xi_1\right)\right|^3\right] \leq 16h\frac{h^2}{8r^2}\left\{\int f(vr)\rho_3(vr)\left|\varphi_{\alpha,1}(v;h,r)\right|^3\mathrm{d}v + \int \left|\varphi_{\alpha,2}(v;h,r)\right|^3 f(vr)\,\mathrm{d}v\right\}
$$

$$
+ o\left(h\frac{h^2}{r^2}\right) + O\left(h^3\right),
$$

$$
\tag{20}
$$

*for* $\rho_3(z) \equiv \mathbf{E}\left[\varepsilon_1 | Z_1 = z\right]$ *and measurable bounded functions* $\varphi_{\alpha,j}$.

**Lemma 14.** *Let* $T_{*,g}, U_g, R_g$ *be as in* (19). *For a bounded measurable function* $C_v(\cdot)$ *and* $m_{g^2}(z) \equiv \mathbf{E}\left[g\left(\xi_1\right)^2 \big| Z_1 = z\right]$.

$$
\mathbf{E}\left[U_g^2\right] = \frac{2}{n(n-1)}\mathrm{E}\left[\mathrm{Var}\left[\dot{\eta}_{g,2}\left(\xi_1,\xi_2\right)|\xi_1\right]\right] - \mathrm{Var}\left[\mathrm{E}\left[\dot{\eta}_{g,2}\left(\xi_1,\xi_2\right)|\xi_1\right]\right]
$$

$$
\leq \frac{8}{n(n-1)} \cdot \frac{h}{r}\int K(u)^2 f(uh)m_{g^2}(uh)C_v(u;h,r)\,\mathrm{d}u, \tag{21}
$$

*so that* $U_g = o_{\mathrm{p}}\left(T_{*,g}\right)$ *if* $nh \to \infty$.

*Furthermore, if* $nh \to \infty$ *and* $(\log n)/(nr) \to 0$, *then*

$$
R_g = O_{\mathrm{p}}\left(\sqrt{\mathbf{E}\left[U_g^2\right]}\right). \tag{22}
$$

# D   Proofs of Auxiliary Lemmas

## D.1   Proof of Lemma 1

For any $z \in [-1, 1]$,

$$|(G_n \circ f)(z) - (G_n \circ g)(z)| \leq \frac{|\delta(z)|}{|R(z)|} \int_{R(z)} |f(x) - g(x)| \, dz \leq \bar{\delta}\varepsilon \; .$$

Letting $g = \mathbf{0}$ and the bound for $G_n^k$ follows by induction.   $\square$

## D.2   Proof of Lemma 2

We will prove that the sequence $\left( (I - G_n)^{-1} \circ \mathbf{1}\{ar_n\} \right)_{n=1}^{\infty}$ is Cauchy. Let $m, n$ be given, where $m > n$. We consider the Neumann series: Writing $(I - G_n)^{-1}$ in terms of its Neumann series:

$$(I - G_n)^{-1} = I + G_n + G_n^2 + \cdots$$

First observe that for all $a$ such that $|a| < 1/r_n$, we have that

$$\mathbf{1}\{z \geq 0\}(ar_n) = \mathbf{1}\{z \geq 0\}(ar_m) \; .$$

Next, suppose $\delta = 1$ is a constant function. Then,

$$(G_n \circ \mathbf{1}\{z \geq 0\})(ar_n) = (G_m \circ \mathbf{1}\{z \geq 0\})(ar_m)$$

for all $a$ such that $|a| < 1/r_n$. This is because the average of $G_n \circ \mathbf{1}\{z \geq 0\}$ in the $r_n$ neighborhood is around $ar_n$ is exactly the average of $G_m \circ \mathbf{1}\{z \geq 0\}$ in the $r_m$ neighborhood is around $ar_m$. This is true also for all $a$ such that $|a| < 1/r_n - 1$. This is because our last result established equality on $[1, 1]$, and this guarantees equality of the local averages on $[-1 + r_n, 1 - r_n]$.

Now suppose $\delta(x)$ is changing. We can write

$$\left(G_n \circ \mathbf{1}\{z \geq 0\}\right)(ar_n) = \left(G_m \circ \mathbf{1}\{z \geq 0\}\right)(ar_m) + R_1$$

where $\|R_1\| \leq |\delta(ar_n) - \delta(ar_m)| \leq C_\delta |a|(r_n + r_m)$ by Lemma 1.

Suppose for induction that for $|a| < 1/r_n - k$,

$$\left(G_n^k \circ \mathbf{1}\{z \geq 0\}\right)(ar_n) = \left(G_m^k \circ \mathbf{1}\{z \geq 0\}\right)(ar_m) + \sum_{j=1}^{k-1} G_n^j R_{k-j}$$

where $\|R_k\| \leq C_\delta |a|(r_n + r_m)$ for all $k$. Now,

$$\left(G_n^{k+1} \circ \mathbf{1}\{z \geq 0\}\right)(ar_n) = G_n \circ \left(G_m^k \circ \mathbf{1}\{z \geq 0\}\right)(ar_m) + G_n \circ \sum_{j=1}^{k-1} G_n^j R_{k-j}$$

for $|a| < 1/r_n - (k+1)$. Now for a fixed $a^*$, choose $K$ such that $(K + |a|)r_n \leq 1$. Then we have that:

$$\left| (I - G_n)^{-1} \mathbf{1}\{z \geq 0\})(a^* r_n) - (I - G_m)^{-1} \mathbf{1}\{z \geq 0\})(a^* r_m) \right|$$

$$= \left| \sum_{j=0}^{\infty} \left(G_n^k \circ \mathbf{1}\{z \geq 0\}\right)(a^* r_n) - \left(G_m^k \circ \mathbf{1}\{z \geq 0\}\right)(a^* r_m) \right|$$

$$\leq \left| \sum_{j=0}^{K} \left(G_n^k \circ \mathbf{1}\{z \geq 0\}\right)(a^* r_n) - \left(G_m^k \circ \mathbf{1}\{z \geq 0\}\right)(a^* r_m) \right|$$

$$+ \left| \sum_{j=K+1}^{\infty} \left(G_n^k \circ \mathbf{1}\{z \geq 0\}\right)(a^* r_n) - \left(G_m^k \circ \mathbf{1}\{z \geq 0\}\right)(a^* r_m) \right|$$

$$\leq \sum_{j=1}^{K-1} \left\| k G_n^j R_{k-j} \right\| + \frac{\bar{\delta}^{K+1}}{1 - \bar{\delta}}$$

$$\leq \frac{C_\delta(r_n + r_m)}{a^* \left(1 - \bar{\delta}\right)^2} + \frac{\bar{\delta}^{1/r_n - |a^*|}}{1 - \bar{\delta}} \ .$$

where the second to last inequality follows from our inductive arguments earlier as well as repeated application of Lemma 1. Applying Lemma 1 again, and rewriting $K = \lfloor 1 - /r_n - |a| \rfloor$ yields the final inequality. As such, the sequence is Cauchy for fixed $a^*$. The limit

61

therefore exists with the final term above being the rate of convergence.

For the next part, first suppose $\delta(0) \neq 0$. Then, we can write:

$$
\begin{aligned}
&\left( (I - G_n)^{-1} \circ \gamma(z) \tilde{d}(z/r_n) \right)(ar_n) \\
&\quad = \frac{\gamma(0)}{\delta(0)} \left( (I - G_n)^{-1} \circ (G_n \circ \mathbf{1}\{z \geq 0\}) \right)(ar_n) \\
&\qquad + \frac{\gamma(0)}{\delta(0)} \left( (I - G_n)^{-1} \circ (\gamma(z) - \gamma(0)) \left( G_n \circ \mathbf{1}\{z \geq 0\} \right) \right)(ar_n) + R
\end{aligned}
$$

and

$$
|R| \leq \frac{C_\delta(r_n + r_m)}{a^* \left(1 - \bar{\delta}\right)^2} + \frac{\bar{\delta}^{1/r_n - |a^*|}}{1 - \bar{\delta}} \ .
$$

This is because as before, the error arises solely from approximating $\delta(z)$ with $\delta(0)$ on the relevant part of the domain.

By the first part of this lemma,

$$
\left| \frac{\gamma(0)}{\delta(0)} \left( (I - G_n)^{-1} \circ (G_n \circ \mathbf{1}\{z \geq 0\}) \right)(ar_n) - \frac{\gamma(0)}{\delta(0)} \left( \lambda(a) - \mathbf{1}\{a \geq 0\} \right) \right| \leq \frac{2C_\delta r_n}{a \left(1 - \bar{\delta}\right)^2} + \frac{\bar{\delta}^{1/r_n - |a|}}{1 - \bar{\delta}} \ .
$$

Next observe that

$$
(\gamma(z) - \gamma(0)) \left( G_n \circ \mathbf{1}\{z \geq 0\} \right))(ar_n) = (\gamma(ar_n) - \gamma(0)) \, \delta(ar_n) \tilde{d}(a)
$$

where $\tilde{d}$ is given in Definition 10.

As such,

$$
\left| (\gamma(z) - \gamma(0)) \left( G_n \circ \mathbf{1}\{z \geq 0\} \right))(ar_n) \right| \leq C_\gamma \bar{\delta} \cdot ar_n \ .
$$

By the Neumann expansion and repeatedly applying Lemma 1 as above,

$$
\left| (I - G_n)^{-1} \circ (\gamma(z) - \gamma(0)) \left( G_n \circ \mathbf{1}\{z \geq 0\} \right))(ar_n) \right| \leq \frac{C_\gamma \bar{\delta}}{1 - \bar{\delta}} \cdot ar_n
$$

Finally, when $\delta(0) = 0$, $G_n = \mathbf{0}$ and the result follows immediately.

## D.3 Proof of Lemma 3

Let $x, y \in [-1, 1]$ be given.

$$|(G_n \circ g)(x) - (G_n \circ g)(y)| = \left| \frac{\delta(x)}{|R(x)|} \int_{R(x)} g(z) dz - \frac{\delta(y)}{|R(y)|} \int_{R(y)} g(z)\, dz \right|$$

$$= \left| \frac{\delta(x)}{|R(x)|} \int_{R(x)} g(z)\, dz - \frac{\delta(y)}{|R(x)|} \int_{R(x)} g(z + x - y)\, dz \right|$$

$$\leq \left| \frac{\delta(x) - \delta(y)}{|R(x)|} \int_{R(x)} g(z)\, dz \right| + \left| \frac{\delta(y)}{|R(x)|} \int_{R(x)} g(z) - g(z + x - y)\, dz \right|$$

$$\leq M \cdot C_\delta |x - y| + \bar{\delta} C |x - y|$$

Now suppose for induction $\left(G_n^k \circ g\right)$ has Lipschitz constant $k \bar{\delta}^{k-1} C_\delta \cdot M + \bar{\delta}^k C$. We know from the above that $\left(G_n^{k+1} \circ g\right)$ has Lipschitz constant

$$\bar{\delta}^k C_\delta \cdot M + \bar{\delta} \left( k \bar{\delta}^{k-1} C_\delta \cdot M + \bar{\delta}^k C \right) = (k+1) \bar{\delta}^k C_\delta \cdot M + \bar{\delta}^{k+1} C \ .$$

Writing $(I - G_n)^{-1}$ in terms of its Neumann series:

$$(I - G_n)^{-1} = I + G_n + G_n^2 + \cdots$$

gives us that

$$\left| \left( (I - G_n)^{-1} \circ g \right)(x) - \left( (I - G_n)^{-1} \circ g \right)(y) \right|$$

$$\leq \sum_{k=0}^{\infty} \left| \left( G_n^k \circ g \right)(x) - \left( G_n^k \circ g \right)(y) \right|$$

$$\leq |x - y| \sum_{k=0}^{\infty} k \bar{\delta}^{k-1} C_\delta \cdot M + \bar{\delta}^k C$$

$$\leq |x - y| \cdot \left( \frac{C_\delta \cdot M}{(1 - \bar{\delta})^2} + \frac{C}{1 - \bar{\delta}} \right)$$

$\square$

## D.4 Proof of Lemma 4

Suppose $R(z) = [z - a, z + b]$. Then $R(-z) = [-z - b, -z + a]$. Since, we have that

$$
\begin{aligned}
(G_n \circ g)(z) &= \frac{\delta(z)}{|R(z)|} \int_{z-a}^{z+b} g(x)\, dF(x) \\
&= -\frac{\delta(-z)}{|R(-z)|} \int_{-z+a}^{-z-b} g(-y)\, dF(y) \text{ by change of variable: } y = -x \\
&= -\frac{\delta(-z)}{|R(-z)|} \int_{-z-b}^{-z+a} g(y)\, dF(y) \text{ since } g(z) = -g(-z) \\
&= (G_n \circ g)(-z)
\end{aligned}
$$

$\square$

## D.5 Proof of Lemma 5

Using Definition 8, write

$$
g_n(z) = \left( (I - G_n)^{-1} \circ m^c \right)(z) + \left( (I - G_n)^{-1} \circ \mathbf{1}\{z \ge 0\} \tau_d \right)(z)
$$

From Lemma 3, $\left( (I - G_n)^{-1} \circ m^c \right)(z)$ is Lipschitz with Lipschitz constant

$$
\frac{C_\delta \cdot M}{\left(1 - \bar{\delta}\right)^2} + \frac{C}{1 - \bar{\delta}} \ .
$$

Next, by the Neumann expansion,

$$
\begin{aligned}
\left( (I - G_n)^{-1} \circ \mathbf{1}\{z \ge 0\} \tau_d \right)(z) &= \mathbf{1}\{z \ge 0\} \tau_d + (G_n \circ \mathbf{1}\{z \ge 0\} \tau_d)(z) + \left( G_n^2 \circ \mathbf{1}\{z \ge 0\} \tau_d \right)(z) + \cdots \\
&= \mathbf{1}\{z \ge 0\} \tau_d + \left( (I - G_n)^{-1} \circ (G_n \circ \mathbf{1}\{z \ge 0\} \tau_d) \right)(z)
\end{aligned}
$$

It is easy to see that $(G_n \circ \mathbf{1}\{z \ge 0\})$ is Lipschitz continuous with Lipschitz constant $\bar{\delta}/2r_n$.

We are done after applying Lemma 3 again. $\square$

## D.6 Proof of Lemma 6

In this proof only, let $m^c$ be any Lipschitz continuous function with Lipschitz constant $C$. By Lipschitz continuity, no point in the $r_n$ neighborhood of $z$ can be larger than $m^c(z) + Cr_n$. Choose $z \in [-\frac{1}{2}, \frac{1}{2}]$. Then,

$$(G_n \circ m^c)(z) \le \delta(z) \left( m^c(z) + \frac{Cr_n}{2} \right) \ .$$

Similarly,

$$\left( G_n^2 \circ m^c \right)(z) \le \delta(z) \left( (G_n \circ m^c)(z) + \bar{\delta} \cdot \frac{Cr_n}{2} \right)$$

$$\le (\delta(z))^2 \, m^c(z) + \frac{Cr_n}{2} \cdot 2 \left( \bar{\delta} \right)^2$$

where the first inequality follows from the Lemma 3 and the second inequality follows from substituting in our bound for $(G_n \circ m^c)(z)$. By induction,

$$\left( G_n^k \circ m^c \right)(z) \le (\delta(z))^k \, m^c(z) + \frac{Cr_n}{2} \cdot k \left( \bar{\delta} \right)^k$$

This is true as long as for all $k \le K := \lfloor 1/2r_n \rfloor$. Summing the terms in the Neumann expansion, we can write

$$\left( (I - G_n)^{-1} \circ m^c \right)(z) - \frac{m^c(z)}{1 - \delta(z)} \le \frac{Cr_n}{2} \cdot \sum_{K=1}^{\infty} k \bar{\delta}^k + 2 \cdot \frac{\bar{\delta}^{K+1} \cdot M}{1 - \bar{\delta}}$$

$$= \frac{Cr_n}{2} \cdot \frac{\bar{\delta}}{\left( 1 - \bar{\delta} \right)^2} + 2 \cdot \frac{\bar{\delta}^{1/2r_n} \cdot M}{1 - \bar{\delta}}$$

where the second term above comes from bounding terms corresponding to $k \ge K + 1$. An analogous lower bound gives us that

$$\left| \left( (I - G_n)^{-1} \circ m^c \right)(z) - \frac{m^c(z)}{1 - \delta(z)} \right| \le \frac{Cr_n}{2} \cdot \frac{\bar{\delta}}{\left( 1 - \bar{\delta} \right)^2} + 2 \cdot \frac{\bar{\delta}^{1/2r_n} \cdot M}{1 - \bar{\delta}} \ .$$

Note here that for $z \geq 0$,

$$\frac{m^c(z)}{1 - \delta(z)} = \left( (I - G_\infty)^{-1} \circ m^c \right)(z) .$$

Putting the above together with a similar argument for $z < 0$ gives us that for $z \in [-\frac{1}{2}, \frac{1}{2}]$,

$$\left| \left( (I - G_n)^{-1} \circ m^c \right)(z) - \left( (I - G_\infty)^{-1} \circ m^c \right)(z) \right| \leq \frac{Cr_n}{2} \cdot \frac{\bar{\delta}}{\left(1 - \bar{\delta}\right)^2} .$$

In other words, $\left( (I - G_n)^{-1} \circ m^c \right)$ converges uniformly to $\left( (I - G_\infty)^{-1} \circ m^c \right)$ on $[-\frac{1}{2}, \frac{1}{2}]$.

$\square$

## D.7 Proof of Lemma 7

The argument here is essentially the same as the inductive argument in Lemma 2. First note that for a given $h_n$, we have by Lipschitz continuity of $\delta(z)$ that for all $x \in [-2h_n, 2h_n]$, $|\delta(x) - \delta(0)| \leq 2C_\delta h_n$.

Next, observe that on $x \in [-2h_n + r_n, 2h_n - r_n]$

$$\left( G_n \circ \mathbf{1}\{z \geq 0\} \right)(x) = \left( \tilde{G}_n \circ \mathbf{1}\{z \geq 0\} \right)(x) + R_1(x)$$

where the requisite bound on $R_1$ holds by Lipschitz continuity of $\delta(z)$. We propagate this bound forward. On $x \in [-2h_n + 2r_n, 2h_n - 2r_n]$,

$$\left( G_n^2 \circ \mathbf{1}\{z \geq 0\} \right)(x) = \left( G_n \circ \left( \tilde{G}_n \circ \mathbf{1}\{z \geq 0\} \right) \right)(x) + \left( G_n \circ R_1 \right)(x)$$
$$= \left( \tilde{G}_n^2 \circ \mathbf{1}\{z \geq 0\} \right)(x) + R_2(x) + \left( G_n \circ R_1 \right)(x) .$$

The bounds on $G_n^{k-j} R_j$ follows from Lemma 1. Note that the domain over which the property holds shrinks by $r_n$ in each step of the induction. Since $h_n > Kr_n$, however, we can perform the above inductive step $K$ times and still have the property hold for $[-h_n, h_n]$. Lemma 7 therefore follows.

## D.8 Proof of Lemma 8

We decompose $\mu_d(0)$ into the parts corresponding to the structural function $m$ and the exogenous spillover:

$$\mu_d(0) = A_n + B_n ,$$

where

$$A_n = \frac{1}{|R_n(0)|} \int_{-r_n}^{r_n} g_n(x)\, dF(x) \quad , \quad B_n = \frac{1}{|R_n(0)|} \int_{-r_n}^{r_n} l_n(x)\, dF(x) .$$

We start by evaluating the limit of $A_n$. Using Definition 8 write:

$$A_n = \frac{1}{|R_n(0)|} \int_{-r_n}^{r_n} \left( (1 - G_n)^{-1} \circ m \right)(x)\, dF(x)$$

$$= \frac{1}{|R_n(0)|} \int_{-r_n}^{r_n} \left( (1 - G_n)^{-1} \circ m^c \right)(x)\, dF(x) + \frac{\tau_d}{|R_n(0)|} \int_{-r_n}^{r_n} \left( (1 - G_n)^{-1} \circ \mathbf{1}\left\{ z \geq 0 \right\} \right)(x)\, dF(x)$$

By Lemma 6,

$$\frac{1}{2r_n} \int_{-r_n}^{r_n} \left( (1 - G_n)^{-1} \circ m^c \right)(x)\, dF(x) = \frac{m^-(0)}{1 - \delta(0)} + O(r_n)$$

For the second term in $A_n$, we use the Neumann series to write:

$$\frac{1}{2r_n} \int_{-r_n}^{r_n} \left( (1 - G_n)^{-1} \circ \mathbf{1}\left\{ z \geq 0 \right\} \right)(x)\, dF(x)$$

$$= \frac{1}{2r_n} \int_{-r_n}^{r_n} \sum_{k=0}^{K} \left( G_n^k \circ \mathbf{1}\left\{ z \geq 0 \right\} \right)(x)\, dF(x) + \frac{1}{2r_n} \int_{-r_n}^{r_n} \sum_{k=K+1}^{\infty} \left( G_n^k \circ \mathbf{1}\left\{ z \geq 0 \right\} \right)(x)\, dF(x)$$

$$\leq \frac{1}{2r_n} \int_{-r_n}^{r_n} \sum_{k=0}^{K} \left( \tilde{G}_n^k \circ \mathbf{1}\left\{ z \geq 0 \right\} \right)(x)\, dF(x) + \frac{C_\delta r_n^{1-\eta}}{(1 - \bar{\delta}^2)} + \frac{1}{2r_n} \int_{-r_n}^{r_n} \sum_{k=K+1}^{\infty} \left( G_n^k \circ \mathbf{1}\left\{ z \geq 0 \right\} \right)(x)\, dF(x)$$

$$\leq \frac{1}{2r_n} \int_{-r_n}^{r_n} \sum_{k=0}^{K} \left( \tilde{G}_n^k \circ \mathbf{1}\left\{ z \geq 0 \right\} \right)(x)\, dF(x) + \frac{C_\delta r_n^{1-\eta}}{(1 - \bar{\delta}^2)} + \frac{\bar{\delta}^{1/r_n^\eta}}{1 - \bar{\delta}} \tag{23}$$

where the first inequality above follows from Lemma 7 with $h_n$ in the lemma set to $r_n^{1-\eta}/2$ and $K$ set to $\lfloor 1/r_n^\eta \rfloor$. The second inequality follows from Lemma 1 and the definition of $K$. A similar argument leads to an analogous lower bound.

67

Now observe that observe that $\mathbf{1}\{z \geq 0\} - \frac{1}{2}$ is an odd function. By repeated application of Lemma 4, we have that $\tilde{G}_n^k \circ \left(\mathbf{1}\{z \geq 0\} - \frac{1}{2}\right)$ is an odd function. Since odd functions integrate to 0 on intervals symmetric around 0,

$$\frac{1}{2r_n} \int_{-r_n}^{r_n} \sum_{k=0}^{K} \left(\tilde{G}_n^k \circ (\mathbf{1}\{z \geq 0\}) - \frac{1}{2}\right)(x)\, dF(x) = 0 \ .$$

In other words,

$$\frac{1}{2r_n} \int_{-r_n}^{r_n} \sum_{k=0}^{K} \left(\tilde{G}_n^k \circ \mathbf{1}\{z \geq 0\}\right)(x)\, dF(x) = \frac{1}{2r_n} \int_{-r_n}^{r_n} \sum_{k=0}^{K} \left(\tilde{G}_n^k \circ \frac{1}{2}\mathbf{I}\right)(x)\, dF(x)$$

$$= \frac{1}{2} \sum_{k=0}^{K} (\delta(0))^k \tag{24}$$

where $\mathbf{I}$ is the identity function. The final equality above follows from the fact that $G_n \circ \mathbf{I} = \mathbf{I}$. Conclude that

$$A_n = \frac{m^-(0)}{1 - \delta(0)} + \frac{\tau_d}{2(1 - \delta(0))} + O\left(\frac{C_\delta r_n^{1-\eta}}{(1 - \bar{\delta}^2)} + \frac{\bar{\delta}^{1/r_n^\eta}}{1 - \bar{\delta}}\right)$$

$$= \frac{m^+(0) + m^-(0)}{2(1 - \delta(0))} + O\left(\frac{C_\delta r_n^{1-\eta}}{(1 - \bar{\delta}^2)} + \frac{\bar{\delta}^{1/r_n^\eta}}{1 - \bar{\delta}}\right) \ .$$

Next, consider $B_n$:

$$B_n = \frac{1}{|R_n(0)|} \int_{-r_n}^{r_n} \left((I - G_n)^{-1} \circ \left(\gamma(z) \cdot \tilde{d}(z/r_n)\right)\right)(x)\, dF(x)$$

$$= \frac{1}{|R_n(0)|} \int_{-r_n}^{r_n} \left((I - G_n)^{-1} \circ ((\gamma(z) - \gamma(0)) \cdot \mathbf{1}\{z \geq 0\})\right)(x)\, dF(x)$$

$$- \frac{1}{|R_n(0)|} \int_{-r_n}^{r_n} \left((I - G_n)^{-1} \circ \left((\gamma(z) - \gamma(0)) \cdot \left(\mathbf{1}\{z \geq 0\} - \tilde{d}(z/r_n)\right)\right)\right)(x)\, dF(x)$$

$$+ \frac{1}{|R_n(0)|} \int_{-r_n}^{r_n} \left((I - G_n)^{-1} \circ \left(\gamma(0) \cdot \tilde{d}(z/r_n)\right)\right)(x)\, dF(x)$$

Observe that $(\gamma(z) - \gamma(0)) \mathbf{1}\{z \geq 0\}$ is Lipschitz continuous with Lipschitz constant $C_\gamma$. As

such, by Lemma 6,

$$\frac{1}{|R_n(0)|} \int_{-r_n}^{r_n} \left( (I - G_n)^{-1} \circ ((\gamma(z) - \gamma(0)) \cdot \mathbf{1}\{z \geq 0\}) \right)(x) \, dF(x) = O(r_n) \ .$$

Furthermore, since $|\gamma(z) - \gamma(0)| \leq C_\gamma r_n$, we also have by Lemma 1 that

$$\frac{1}{|R_n(0)|} \int_{-r_n}^{r_n} \left( (I - G_n)^{-1} \circ \left( (\gamma(z) - \gamma(0)) \cdot \left( \mathbf{1}\{z \geq 0\} - \tilde{d}(z) \right) \right) \right)(x) \, dF(x) = O(r_n)$$

Next, observe that when $\delta(0) \neq 0$,

$$\frac{1}{|R_n(0)|} \int_{-r_n}^{r_n} \left( (I - G_n)^{-1} \circ \tilde{d}(z/r_n) \right)(x) \, dF(x)$$

$$= \frac{1}{\delta(0)} \frac{1}{|R_n(0)|} \int_{-r_n}^{r_n} \left( (I - G_n)^{-1} \circ (G_n \circ \mathbf{1}\{z \geq 0\}) \right)(x) \, dF(x) + O\left( \frac{C_\delta r_n^{1-\eta}}{(1 - \bar{\delta}^2)} + \frac{\bar{\delta}^{1/r_n^\eta}}{1 - \bar{\delta}} \right)$$

$$= \frac{1}{\delta(0)} \left( \frac{1}{|R_n(0)|} \int_{-r_n}^{r_n} \left( (I - G_n)^{-1} \circ \mathbf{1}\{z \geq 0\} \right)(x) \, dF(x) - \frac{1}{|R_n(0)|} \int_{-r_n}^{r_n} \mathbf{1}\{x \geq 0\} \, dF(x) \right)$$

$$\quad + O\left( \frac{C_\delta r_n^{1-\eta}}{(1 - \bar{\delta}^2)} + \frac{\bar{\delta}^{1/r_n^\eta}}{1 - \bar{\delta}} \right)$$

$$= \frac{1}{\delta(0)} \left( \frac{1}{2(1 - \delta(0))} - \frac{1}{2} \right) + O\left( \frac{C_\delta r_n^{1-\eta}}{(1 - \bar{\delta}^2)} + \frac{\bar{\delta}^{1/r_n^\eta}}{1 - \bar{\delta}} \right)$$

$$= \frac{1}{2(1 - \delta(0))} + O\left( \frac{C_\delta r_n^{1-\eta}}{(1 - \bar{\delta}^2)} + \frac{\bar{\delta}^{1/r_n^\eta}}{1 - \bar{\delta}} \right)$$

where the first equality comes from approximating $\delta(z)$ with $\delta(0)$. The second to last equality above follows from the same arguments as in equations (23) and (24). When $\delta(0) = 0$,

$$\frac{1}{|R_n(0)|} \int_{-r_n}^{r_n} \left( (I - G_n)^{-1} \circ \tilde{d}(z/r_n) \right)(x) \, dF(x)$$

$$= \frac{1}{|R_n(0)|} \int_{-r_n}^{r_n} \left( \left( I - \tilde{G}_n \right)^{-1} \circ \tilde{d}(z/r_n) \right)(x) \, dF(x) + O\left( \frac{C_\delta r_n^{1-\eta}}{(1 - \bar{\delta}^2)} + \frac{\bar{\delta}^{1/r_n^\eta}}{1 - \bar{\delta}} \right)$$

$$= \frac{1}{|R_n(0)|} \int_{-r_n}^{r_n} \tilde{d}(x/r_n) \, dF(x) + O\left( \frac{C_\delta r_n^{1-\eta}}{(1 - \bar{\delta}^2)} + \frac{\bar{\delta}^{1/r_n^\eta}}{1 - \bar{\delta}} \right)$$

$$= \frac{1}{2} + O\left( \frac{C_\delta r_n^{1-\eta}}{(1 - \bar{\delta}^2)} + \frac{\bar{\delta}^{1/r_n^\eta}}{1 - \bar{\delta}} \right) \ .$$

69

As such,

$$B_n = \frac{\gamma(0)}{2(1 - \delta(0))} + O\left(\frac{C_\delta r_n^{1-\eta}}{(1 - \bar{\delta}^2)} + \frac{\bar{\delta}^{1/r_n^\eta}}{1 - \bar{\delta}}\right).$$

$\square$

## D.9 Proof of Lemma 9

Write:

$$\mu_d(xh_n) - \mu_d(0) = A_n + B_n$$

where

$$\begin{aligned}
A_n &= \frac{1}{|R_n(xh_n)|} \int_{R_n(xh_n)} g_n(z)\, dF(z) - \frac{1}{|R_n(0)|} \int_{R_n(0)} g_n(z)\, dF(z) \\
&= \frac{|R_n^+(xh_n)|}{|R_n(xh_n)|} \cdot \frac{1}{|R_n^+(xh_n)|} \int_{R_n^+(xh_n)} g_n(z)\, dF(z) \\
&\quad - \frac{|R_n^-(xh_n)|}{|R_n(xh_n)|} \cdot \frac{1}{|R_n^-(xh_n)|} \int_{R_n^-(xh_n)} g_n(z)\, dF(z)
\end{aligned}$$

and

$$\begin{aligned}
B_n &= \frac{1}{|R_n(xh_n)|} \int_{R_n(xh_n)} l_n(z)\, dF(z) - \frac{1}{|R_n(0)|} \int_{R_n(0)} l_n(z)\, dF(z) \\
&= \frac{|R_n^+(xh_n)|}{|R_n(xh_n)|} \cdot \frac{1}{|R_n^+(xh_n)|} \int_{R_n^+(xh_n)} l_n(z)\, dF(z) \\
&\quad - \frac{|R_n^-(xh_n)|}{|R_n(xh_n)|} \cdot \frac{1}{|R_n^-(xh_n)|} \int_{R_n^-(xh_n)} l_n(z)\, dF(z).
\end{aligned}$$

Here,

$$l_n(z) = \left[(I - G_n)^{-1} \circ (\gamma \cdot (G_n \circ \mathbf{1}\{u \geq 0\}))\right](z).$$

$A_n$ and $B_n$ are the parts of $\mu_d$ coming from (1) the structural function $m$ and (2) the exogenous spillover $\nu_d = G_n \circ \mathbf{1}\{z \geq 0\}$ respectively. We analyze these term by considering cases separately.

**Case 1: $r_n = r$ constant**

In this case, $G_n$, $g_n$ and $R_n$ do not change with $n$. Suppose first that $z \geq 0$. Then,

$$R_n^+(xh_n) = [r, r + xh_n] \quad , \quad R_n^-(xh_n) = [-r - xh_n, -r]$$

which shrink to $r$ and $-r$ respectively.

We start with $A_n$. By Lemma 5,

$$\frac{1}{|R_n^+(xh_n)|} \int_{R_n^+(xh_n)} g_n(z) \, dF(z) = \frac{1}{|R_n^+(xh_n)|} \int_{R_n^+(xh_n)} g_n^c(z) \, dF(z) + \frac{\tau_d}{|R_n^+(xh_n)|} \int_{R_n^+(xh_n)} \mathbf{1}\{z \geq 0\} \, dF(z)$$

$$\leq g_n^c(r) + \tilde{C}xh_n + \frac{\tau_d}{|R_n^+(xh_n)|} \int_{R_n^+(xh_n)} \mathbf{1}\{z \geq 0\} \, dF(z)$$

$$= g_n^c(r) + \tilde{C}xh_n + \tau_d$$

$$= g_n(r) + \tilde{C}xh_n$$

where the inequality above follows from the fact that $g_n^c$ is Lipschitz, with $\tilde{C}$ being the relevant constant in Lemma 5. Together with a similar argument for the lower bound, we have that

$$\left| \frac{1}{|R_n^+(xh_n)|} \int_{R_n^+(xh_n)} g_n(z) \, dF(z) - g_n(r) \right| \leq \tilde{C}xh_n$$

Furthermore, since $h_n/r_n \to 0$, we have that for $n$ large enough,

$$\frac{|R_n^+(xh_n)|}{|R_n(xh_n)|} = \frac{|R_n^-(xh_n)|}{|R_n(xh_n)|} = \frac{xh_n}{r_n} \quad . \tag{25}$$

Hence,

$$A_n = \frac{xh_n}{r_n} \left( g_n(r) - g_n(-r) + O(h_n) \right) \ .$$

where the constants in $O(h_n)$ do not depend on $x$. Now, if $z < 0$, a similar argument to the

71

one above gives us

$$A_n = \frac{xh_n}{r_n} \left( g_n(-r) - g_n(r) + O(h_n) \right) .$$

Next consider $B_n$. By Lemma 3, $l_n(z)$ is Lipschitz continuous with Lipschitz constant

$$\tilde{C}_1 = \frac{C_\delta}{\left(1 - \bar{\delta}\right)^2} + \frac{C_\gamma}{1 - \bar{\delta}}$$

As such,

$$\frac{1}{|R_n^+(xh_n)|} \int_{R_n^+(xh_n)} l_n(z) \, dF(z) \leq l_n(r) + \tilde{C}_1 x h_n ,$$

with an analogous lower bound. We conclude that

$$B_n = \frac{xh_n}{r_n} \left( l_n(r) - l_n(-r) + O(h_n) \right) .$$

when $z \geq 0$ and

$$B_n = \frac{xh_n}{r_n} \left( l_n(-r) - l_n(r) + O(h_n) \right) .$$

when $z < 0$.

**Case 2: $r_n \to 0$, $h_n/r_n \to 0$**

Start with $A_n$. Suppose again that $z \geq 0$. Since $h_n/r_n \to 0$, equation (25) continues to hold. Write:

$$\frac{1}{|R_n^+(xh_n)|} \int_{R_n^+(xh_n)} g_n(z) \, dF(z) = \frac{1}{|R_n^+(xh_n)|} \int_{R_n^+(xh_n)} \left( (I - G_n)^{-1} \circ m^c \right)(z) \, dF(z)$$

$$+ \frac{\tau_d}{|R_n^+(xh_n)|} \int_{R_n^+(xh_n)} \left( (I - G_n)^{-1} \circ \mathbf{1} \left\{ z \geq 0 \right\} \right)(z) \, dF(z)$$

$$(26)$$

We start with first term above:

$$\frac{1}{|R_n^+(xh_n)|} \int_{R_n^+(xh_n)} \left( (I - G_n)^{-1} \circ m^c \right)(z)\, dF(z)$$

$$= \frac{1}{|R_n^+(xh_n)|} \int_{R_n^+(xh_n)} \left( (I - G_\infty)^{-1} \circ m^c \right)(z)\, dF(z) + O(r_n) \text{ by Lemma 6}$$

$$= \frac{m^c(r_n)}{1 - \delta(r_n)} + O(h_n) + O(r_n) \text{ by Definition 11 and the fact that } m^c \text{ is Lipschitz}$$

$$= \frac{m^c(0)}{1 - \delta(0)} + O(h_n) + O(r_n) \tag{27}$$

where the last equality follows again from Lipschitz continuity of $m^c(z)$ and $\delta(z)$.

Next, consider

$$\frac{\tau_d}{|R_n^+(xh_n)|} \int_{R_n^+(xh_n)} \left( (I - G_n)^{-1} \circ \mathbf{1}\{x \geq 0\} \right)(z)\, dF(z)$$

$$= \frac{\tau_d}{|R_n^+(xh_n)|} \int_{R_n^+(xh_n)} \lambda(z/r_n)\, dF(z) + O\left( \frac{2\bar{\delta}^{\lfloor 1/r_n \rfloor - |a|}}{1 - \bar{\delta}} \right)$$

which follows by an application of Lemma 2 with $|a| = 2$ since $h_n \ll r_n$. Next consider the change of variable $a = z/r_n$. This gives us

$$\frac{\tau_d}{|R_n^+(xh_n)|} \int_{R_n^+(xh_n)} \left( (I - G_n)^{-1} \circ \mathbf{1}\{x \geq 0\} \right)(z)\, dF(z)$$

$$= \frac{\tau_d}{r_n(xh_n/r_n)} \int_{[1, 1+2xh_n/r_n]} \lambda(a) \cdot r_n\, dF(a) + O\left( \frac{2\bar{\delta}^{\lfloor 1/r_n \rfloor - |a|}}{1 - \bar{\delta}} \right)$$

$$= \frac{\tau_d}{xh_n/r_n} \int_{[1, 1+2xh_n/r_n]} \lambda(a)\, dF(a) + O\left( \frac{2\bar{\delta}^{\lfloor 1/r_n \rfloor - |a|}}{1 - \bar{\delta}} \right)$$

$$= \tau_d \cdot \lambda(1) + o(1) \tag{28}$$

where the last equality follows from the Lebesgue Differentiation Theorem. The domain of integration is a convex interval and trivially of bounded eccentricity.

Putting equations (27) and (28) together with equation (25) yields:

$$A_n = \frac{xh_n}{r_n} \cdot \tau_d \cdot (\lambda(1) - \lambda(-1) + o(1)) \ .$$

Finally, suppose $z < 0$. Note that analysis of the continuous part is identical. For the discontinuous part, the first term is now zero. The second term is expanded around $-r_n$, giving us

$$A_n = \frac{xh_n}{r_n} \cdot \tau_d \cdot (\lambda(-1) - \lambda(1) + o(1)) \ .$$

Next, consider $B_n$. Using the second part of Lemma 2, and identical argument to the one in equation (28) gives us that

$$\frac{1}{|R_n^+(xh_n)|} \int_{R_n^+(xh_n)} l_n(z) \, dF(z) = \gamma(0) \cdot (\lambda(1) - \mathbf{1}\{1 \geq 0\}) + o(1) \tag{29}$$

Consequently, if $z \geq 0$,

$$B_n = \frac{xh_n}{r_n} \cdot \gamma(0) \cdot ((\lambda(1) - 1) - \lambda(-1) + o(1)) \ .$$

On the other hand, if $z \leq 0$,

$$B_n = \frac{xh_n}{r_n} \cdot \gamma(0) \cdot (\lambda(-1) - (\lambda(1) - 1) + o(1)) \ .$$

**Case 3:** $r_n = \frac{1}{2}ch_n$

Under this asymptotic regime,

$$\frac{|R_n^+(xh_n)|}{|R_n(xh_n)|} = \frac{|R_n^-(xh_n)|}{|R_n(xh_n)|} = \min\left\{1, \frac{|x|}{c}\right\} \ . \tag{30}$$

To see this, note that when $|x|h_n > 2r_n = ch_n$, there is no overlap between $R_n(xh_n)$ and $R_n(0)$. When $|x| < c$, the area of

$$R_n^+(xh_n) = R_n^-(xh_n) = (c - |x|)h_n$$

while $R_n(xh_n) = 2r_n = ch_n$, yielding the expression above.

Suppose again that $z \geq 0$. We consider the same decomposition as in equation (26). Equation (27) continues to be valid since we did not make any assumption about $h_n/r_n$ in its derivation.

Now, by a change of variable $a = z/r_n$, we have that

$$\frac{\tau_d}{|R_n^+(xh_n)|} \int_{R_n^+(xh_n)} \left( (I - G_n)^{-1} \circ \mathbf{1}\{x \geq 0\} \right)(z) \, dF(z)$$

$$= \frac{\tau_d}{r_n |\tilde{R}(x)|} \int_{\tilde{R}(x)} \lambda(a) \cdot r_n \, dF(a) + O \left( \frac{2\bar{\delta}^{\lfloor 1/r_n \rfloor - |a|}}{1 - \bar{\delta}} \right)$$

$$= \tau_d \cdot \lambda^+(x) + O \left( \frac{2\bar{\delta}^{\lfloor 1/r_n \rfloor - |a|}}{1 - \bar{\delta}} \right) \tag{31}$$

In first equality above, we applied Lemma 2 and the fact that

$$R_n^+(xh_n) = [\max\{r_n, xh_n - r_n\}, r_n + xh_n] = r_n \left[ \max\left\{ 1, \frac{2x}{c} - 1 \right\}, 1 + \frac{2x}{c} \right] = r_n \tilde{R}^+(x) \ .$$

Since $|x| \leq 1$, the above bound is uniform in $x$ once we let $|a| = 1 + \frac{2}{c}$.

By the same reasoning,

$$R_n^- = r_n \underbrace{\left[ -1, \min\left\{ \frac{2x}{c} - 1, 1 \right\} \right]}_{=: \tilde{R}^-(x)}$$

and

$$\frac{\tau_d}{|R_n^-(xh_n)|} \int_{R_n^-(xh_n)} \left( (I - G_n)^{-1} \circ \mathbf{1}\{x \geq 0\} \right)(z) \, dF(z)$$

$$= \tau_d \cdot \lambda^-(x) + O \left( \frac{2\bar{\delta}^{\lfloor 1/r_n \rfloor - |a|}}{1 - \bar{\delta}} \right) \tag{32}$$

Putting equations (27), (31) and (32) together with equation (30) yields:

$$A_n = \min\left\{ 1, \frac{|x|}{c} \right\} \cdot \tau_d \cdot \left( \lambda^+(x) - \lambda^-(x) + o(1) \right) \ .$$

A corresponding derivation for $x \leq 0$ yields

$$A_n = \min\left\{ 1, \frac{|x|}{c} \right\} \cdot \tau_d \cdot \left( \lambda^+(-x) - \lambda^-(-x) + o(1) \right) \ .$$

Next consider $B_n$. By derivations that is similar to (31), we have that

$$\frac{1}{|R_n^+(xh_n)|} \int_{R_n^+(xh_n)} \left( (I - G_n)^{-1} \circ l_n(z) \right) dF(z) = \gamma(0) \cdot \left( \lambda^+(x) - \mathbf{1}\{x \geq 0\} \right) + O\left( \frac{2\bar{\delta}^{\lfloor 1/r_n \rfloor - |a|}}{1 - \bar{\delta}} \right)$$

which implies that

$$B_n = \min\left\{ 1, \frac{|x|}{c} \right\} \cdot \gamma(0) \cdot \left( \tilde{\lambda}^+(x) - \tilde{\lambda}^-(x) + o(1) \right) .$$

when $z \geq 0$. When $z < 0$,

$$B_n = \min\left\{ 1, \frac{|x|}{c} \right\} \cdot \gamma(0) \cdot \left( \tilde{\lambda}^+(-x) - \tilde{\lambda}^-(-x) + o(1) \right) .$$

**Case 4:** $r_n/h_n \to 0$

Write

$$\begin{aligned}
\mu_d(xh_n) = {}& \frac{1}{|R_n(xh_n)|} \int_{R_n(xh_n)} g_n^c(z) \, dF(z) \\
& + \frac{\tau_d}{|R_n(xh_n)|} \int_{R_n(xh_n)} \left( (I - G_n)^{-1} \circ \mathbf{1}\{y \geq 0\} \right)(z) \, dF(z) \\
& + \frac{1}{|R_n(xh_n)|} \int_{R_n(xh_n)} l_n(z) \, dF(z)
\end{aligned}$$

As before, the limit of the first term above is given by Equation (27).

Next, for the second term, we apply Lemma 7, with $h_n$ in the lemma set to bandwidth. Next, let $K = \lfloor w_n/r_n \rfloor + 1$. Then, uniformly for $x \in [-1, 1]$,

$$\begin{aligned}
& \frac{\tau_d}{|R_n(xh_n)|} \int_{R_n(xh_n)} \left( (I - G_n)^{-1} \circ \mathbf{1}\{y \geq 0\} \right)(z) \, dF(z) \\
& = \frac{\tau_d}{|R_n(xh_n)|} \int_{R_n(xh_n)} \left( \sum_{k=0}^{K} \tilde{G}_n^k \circ \mathbf{1}\{y \geq 0\} \right)(z) \, dF(z) + O\left( \frac{C_\delta h_n}{(1 - \bar{\delta})^2} + \frac{\bar{\delta}^K}{1 - \bar{\delta}} \right)
\end{aligned}$$

Furthermore, for $|z| > Kr_n$, we have that

$$\left( \tilde{G}_n^k \circ \mathbf{1}\{y \geq 0\} \right)(z) = \delta(0)^k \mathbf{1}\{z \geq 0\} .$$

76

As such, for $x \in [w_n, 1]$, we have that:

$$\frac{\tau_d}{|R_n(xh_n)|} \int_{R_n(xh_n)} \left( \sum_{k=0}^{K} \tilde{G}_n^k \circ \mathbf{1}\{y \geq 0\} \right) (z) \, dF(z)$$

$$= \tau_d \cdot \sum_{k=0}^{K} \delta(0)^k = \frac{\tau_d}{1 - \delta(0)} + O\left( \frac{\bar{\delta}^k}{1 - \bar{\delta}} \right) . \qquad (33)$$

As such, the second term of $\mu_d(xh_n)$ is

$$\frac{\tau_d}{|R_n(xh_n)|} \int_{R_n(xh_n)} \left( (I - G_n)^{-1} \circ \mathbf{1}\{y \geq 0\} \right) (z) \, dF(z)$$

$$= \frac{\tau_d}{1 - \delta(0)} + O\left( \bar{\delta}^K \right) + O\left( h_n \right)$$

Finally, for the third term in $\mu_d(xh_n)$. Applying Lemma 7 again, we have that when $\delta(0) \neq 0$, that for $x \in [-1, 1]$,

$$\frac{1}{|R_n(xh_n)|} \int_{R_n(xh_n)} l_n(z) \, dF(z)$$

$$= \frac{1}{\delta(0)} \frac{1}{|R_n(xh_n)|} \int_{R_n(xh_n)} \left( \left( I - \tilde{G}_n \right)^{-1} \circ \left( \gamma(y) \cdot (G_n \circ \mathbf{1}\{y \geq 0\}) \right) \right) (z) \, dF(z)$$

$$= \frac{1}{\delta(0)} \frac{1}{|R_n(xh_n)|} \int_{R_n(xh_n)} \left( \sum_{k=0}^{K} \tilde{G}_n^k \circ \left( \gamma(y) \cdot \left( \tilde{G}_n \circ \mathbf{1}\{y \geq 0\} \right) \right) \right) (z) \, dF(z) + O\left( h_n + \bar{\delta}^K \right)$$

$$= \frac{1}{\delta(0)} \frac{1}{|R_n(xh_n)|} \int_{R_n(xh_n)} \left( \sum_{k=0}^{K} \tilde{G}_n^k \circ \left( \gamma(0) \cdot \left( \tilde{G}_n \circ \mathbf{1}\{y \geq 0\} \right) \right) \right) (z) \, dF(z) + O\left( h_n + \bar{\delta}^K \right)$$

where the last equality follows because for $k \leq h_n/r_n$, $\tilde{G}_n^k$ evaluated on $[-h_n, h_n]$ does not depend on values of $\gamma(y)$ outside of $[-2h_n, 2h_n]$. The error from replacing $\gamma(y)$ with $\gamma(0)$ on this latter interval is $O(h_n)$ uniformly in $x \in [-1, 1]$. By the same reasoning, so is the

error from replacing $\delta(y)$ with $\delta(0)$. Now, suppose $\delta(0) \neq 0$. Then write

$$\frac{1}{|R_n(xh_n)|} \int_{R_n(xh_n)} \left( \sum_{k=0}^{K} \tilde{G}_n^k \circ \left( \gamma(0) \cdot \left( \tilde{G}_n \circ \mathbf{1}\{y \geq 0\} \right) \right) \right)(z) \, dF(z)$$

$$= \frac{\gamma(0)}{|R_n(xh_n)|} \int_{R_n(xh_n)} \left( \sum_{k=1}^{K+1} \tilde{G}_n^k \mathbf{1}\{y \geq 0\} \right)(z) \, dF(z)$$

$$= \gamma(0) \left( \frac{1}{1 - \delta(0)} - \underbrace{\frac{1}{|R_n(xh_n)|} \int_{R_n(xh_n)} \mathbf{1}\{z \geq 0\} \, dF(z)}_{=1} \right) + O\left(\bar{\delta}^{K+1}\right)$$

$$= \frac{\delta(0)\gamma(0)}{1 - \delta(0)} + O\left(\bar{\delta}^{K+1}\right)$$

where the second to last equality above follows from Equation (33). That is, when $\delta(0) \neq 0$,

$$\frac{1}{|R_n(xh_n)|} \int_{R_n(xh_n)} l_n(z) \, dF(z) = \frac{\gamma(0)}{1 - \delta(0)} + O\left(\bar{\delta}^{K+1}\right)$$

uniformly for $x \in [w_n, 1]$. When $\delta(0) = 0$,

$$\frac{1}{|R_n(xh_n)|} \int_{R_n(xh_n)} l_n(z) \, dF(z) = \frac{1}{|R_n(xh_n)|} \int_{R_n(xh_n)} \gamma(z)\tilde{d}(z/r_n) \, dF(z) + O\left(h_n + \bar{\delta}^K\right)$$

$$= \gamma(0) + O\left(h_n + \bar{\delta}^K\right) .$$

again uniformly for $x \in [w_n, 1]$.

Putting our bounds together, we therefore have that

$$\mu_d(xh_n) = \frac{m^+(0) + \gamma(0)}{1 - \delta(0)} + O\left(h_n + \bar{\delta}^{w_n/r_n}\right) .$$

With Lemma 8, we then have that for $x \in [w_n, 1]$,

$$\mu_d(xh_n) - \mu_d(0) = \frac{\tau_d + \gamma(0)}{2(1 - \delta(0))} + O\left(h_n + \bar{\delta}^{w_n/r_n}\right)$$

By a similar argument, when $x \in [-1, -w_n]$

$$\mu_d(xh_n) - \mu_d(0) = -\frac{\tau_d + \gamma(0)}{2(1 - \delta(0))} + O\left(h_n + \bar{\delta}^{w_n/r_n}\right) .$$

$\square$

## D.10   Proof of Lemma 10

Define

$$\eta_{X,1}\left(\xi_1\right) = \varepsilon_1 \mathbf{E}\left[K\left(\frac{Z_0}{h}\right)\left(\delta(0) + \delta'(0)Z_0\right)X\left(Z_0\right)\left\{\frac{R\left(\frac{Z_1}{r}\right)}{\pi_r(0)} - \frac{R\left(\frac{Z_1 - Z_0}{r}\right)}{\pi_r\left(Z_0\right)}\right\}\Bigg|\xi_1\right]$$

$$= \varepsilon_1 \varphi_{\eta,1}\left(Z_1\right),$$

$$\varphi_{\eta,1}\left(Z_1\right) = \mathbf{E}\left[K\left(\frac{Z_0}{h}\right)\left(\delta(0) + \delta'(0)Z_0\right)X\left(Z_0\right)\left\{\frac{R\left(\frac{Z_1}{r}\right)}{\pi_r(0)} - \frac{R\left(\frac{Z_1 - Z_0}{r}\right)}{\pi_r\left(Z_0\right)}\right\}\Bigg|Z_1\right],$$

$$\eta_{X,2}\left(\xi_1\right) = \mathbf{E}\left[K\left(\frac{Z_0}{h}\right)\left(\delta(0) + \delta'(0)Z_0\right)X\left(Z_0\right)\left(Y\left(Z_1\right) - \mu_d(0)\right)\frac{R\left(\frac{Z_1}{r}\right)}{\pi_r(0)}\Bigg|\xi_1\right]$$

$$- \mathbf{E}\left[K\left(\frac{Z_0}{h}\right)\left(\delta(0) + \delta'(0)Z_0\right)X\left(Z_0\right)\left(Y\left(Z_1\right) - \mu_d\left(Z_0\right)\right)\frac{R\left(\frac{Z_1 - Z_0}{r}\right)}{\pi_r\left(Z_0\right)}\Bigg|\xi_1\right],$$

$$\dot{\eta}_{X,2}\left(\xi_1\right) = \eta_{X,2}\left(\xi_1\right) - \mathbf{E}\left[\eta_{X,2}\left(\xi_1\right)\right].$$

$$\tag{34}$$

By (3),

$$\eta_X = \eta_{X,1} + \eta_{X,2}, \quad \text{and} \quad \dot{\eta}_X = \eta_{X,1} + \dot{\eta}_{X,2}.$$

By mean independence ($\mathbf{E}\left[\varepsilon_1 | Z_1\right] = 0$),

$$\text{Cov}\left(K\left(\frac{Z_i}{h}\right)X_i\varepsilon_i, \eta_{X,2}\left(\xi_i\right)\right) = 0, \quad \text{and} \quad \text{Cov}\left(\eta_{X,1}\left(\xi_i\right), \eta_{X,2}\left(\xi_i\right)\right) = 0.$$

Thus,

$$\boldsymbol{\Omega} = \text{Var}\left[K\left(\frac{Z_i}{h}\right)X_i\varepsilon_i + \eta_X\left(\xi_i\right)\right] = \text{Var}\left[K\left(\frac{Z_i}{h}\right)X_i\varepsilon_i + \eta_{X,1}\left(\xi_i\right)\right] + \text{Var}\left[\eta_{X,2}\left(\xi_i\right)\right]$$

$$\succeq \text{Var}\left[K\left(\frac{Z_i}{h}\right)X_i\varepsilon_i + \eta_{X,1}\left(\xi_i\right)\right] =: \boldsymbol{\Omega}_1,$$

$$\tag{35}$$

where $\succeq$ denotes the Loewner order ($A \succeq B$ iff $A - B$ is positive semi-definite). This implies

that $\left\| \boldsymbol{\Omega}^{-1/2} t \right\| \leq \left\| \boldsymbol{\Omega}_1^{-1/2} t \right\|$ for any vector $t$. In what follows, define

$$t_{*,i} = K\left(\frac{Z_i}{h}\right) X_i \varepsilon_i + \dot{\eta}_X(\xi_i) \quad \text{and} \quad \alpha_{*,i} = \left(\frac{\boldsymbol{\Omega}}{n}\right)^{-1/2} t_i. \tag{36}$$

Clearly $\mathbf{E}[\alpha_i] = \mathbf{0}_8$ and $\sum_{i=1}^{n} \mathrm{Var}[\alpha_{*,i}] = \mathbb{I}_8$. Furthermore, $\boldsymbol{\Omega}^{-1/2}\sqrt{n}\widetilde{\mathbf{T}}_* = \sum_{i=1}^{n} \alpha_{*,i}$. It suffices to show that Liapunov's condition holds:

$$\lim_{n \to \infty} \sum_{i=1}^{n} \mathbf{E}\left[\|\alpha_{*,i}\|^3\right] = 0 \tag{37}$$

Following calculations similar to those for $\widetilde{\mathbf{Q}}$ in the proofs of Theorems 2 and 3 in section A.2,

$$\boldsymbol{\Omega}_1 = h \cdot f(0)\sigma_\varepsilon^2(0) \cdot H\left(\boldsymbol{\Omega}_1^* + o(1)\right) H, \tag{38}$$

where $\boldsymbol{\Omega}_1^*$ is a limiting positive definite matrix, and

$$H = \mathbb{I}_4 \otimes \mathrm{diag}(1, h) = \begin{bmatrix} 1 & 0 & \cdots & 0 & 0 \\ 0 & h & \cdots & 0 & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & \cdots & 0 & h \end{bmatrix}.$$

By (36), (35), and (38) it follows that

$$\sum_{i=1}^{n} \mathbf{E}\left[\|\alpha_{*,i}\|^3\right] = n\mathbf{E}\left[\|\alpha_{*,1}\|^3\right]$$

$$= \frac{1}{\sqrt{n}}\mathbf{E}\left[\left\|\boldsymbol{\Omega}^{-1/2} t_{*,1}\right\|^3\right]$$

$$= \frac{1}{\sqrt{n}}\mathbf{E}\left[\left\|\boldsymbol{\Omega}_1^{-1/2} t_{*,1}\right\|^3\right]$$

$$= \frac{f(0)\sigma_\varepsilon^2(0)}{\sqrt{nh}}\mathbf{E}\left[\left\|\left(H\left(\boldsymbol{\Omega}_1^* + o(1)\right)H\right)^{-1/2} t_{*,1}\right\|^3\right]$$

$$\leq \frac{f(0)\sigma_\varepsilon^2(0)}{\left(\sqrt{\mathrm{eigmin}\left(\boldsymbol{\Omega}_1^*\right)} + o(1)\right) \cdot \sqrt{nh}}\mathbf{E}\left[\left\|H^{-1} t_{*,1}\right\|^3\right].$$

Denoting $\boldsymbol{\omega}_1^* = \sqrt{\operatorname{eigmin}\left(\boldsymbol{\Omega}_1^*\right)}$, we get

$$\sum_{i=1}^n \mathbf{E}\left[\|\alpha_{*,i}\|^3\right] = \left(\frac{f(0)\sigma_\varepsilon^2(0)}{\boldsymbol{\omega}_1^*} + o(1)\right)\frac{1}{\sqrt{nh}}\mathbf{E}\left[\left\|H^{-1}t_{*,1}\right\|^3\right] = O(1)\cdot\frac{1}{\sqrt{nh}},$$

since $\mathbf{E}\left[\left\|H^{-1}t_{*,1}\right\|^3\right] = O(1)$ when $\mathbf{E}\left[\varepsilon^3\big|Z = z\right]$ is uniformly bounded in a neighborhood of zero. Thus Liapunov's condition (37) holds, and so $\boldsymbol{\Omega}^{-1/2}\sqrt{n}\widetilde{\mathbf{T}}_* \rightsquigarrow \mathbf{N}\left(\mathbf{0}_8, \mathbb{I}_8\right)$.

## D.11    Proof of Lemma 11

Write

$$\widetilde{\mathbf{T}}_{*,\varepsilon} = \frac{1}{n}\sum_{i=1}^n K\left(\frac{Z_i}{h}\right)X_i\varepsilon_i, \quad \text{and} \quad \widetilde{\mathbf{T}}_{*,X} = \frac{1}{n}\sum_{i=1}^n \left\{\eta_X\left(\xi_i\right) - \mathbf{E}\left[\eta_X\left(\xi_2\right)\right]\right\}. \tag{39}$$

Then,

$$\begin{aligned} \widetilde{\mathbf{T}}_\varepsilon &= \widetilde{\mathbf{T}}_{*,\varepsilon} + \widetilde{\mathbf{S}}_\varepsilon, \\ \text{where} \quad \widetilde{\mathbf{S}}_\varepsilon &= \frac{1}{n}\sum_{i=1}^n K\left(Z_i/h\right)\left(\widetilde{X}_i - X_i\right)\varepsilon_i. \end{aligned} \tag{40}$$

Similarly,

$$\begin{aligned} \widetilde{\mathbf{T}}_X &= \widetilde{\mathbf{T}}_{*,X} + \widetilde{\mathbf{T}}_{*,X}^{(1)} - \widetilde{\mathbf{T}}_{*,X} + \widetilde{\mathbf{S}}_X^{(1)}, \\ \text{where} \quad \widetilde{\mathbf{T}}_{*,X}^{(1)} &= \frac{1}{n}\sum_{i=1}^n K\left(Z_i/h\right)X_i\left(X_i - \widetilde{X}_i\right)'\beta \\ \widetilde{\mathbf{S}}_X^{(1)} &= -\frac{1}{n}\sum_{i=1}^n K\left(Z_i/h\right)\left(\widetilde{X}_i - X_i\right)\left(\widetilde{X}_i - X_i\right)'\beta. \end{aligned} \tag{41}$$

We wish to then show that

$$\widetilde{\mathbf{S}}_\varepsilon = o_{\mathrm{p}}\left(\widetilde{\mathbf{T}}_{*,\varepsilon}\right) \quad \text{and} \quad \underbrace{\widetilde{\mathbf{T}}_{*,X}^{(1)} - \widetilde{\mathbf{T}}_{*,X} + \widetilde{\mathbf{S}}_X^{(1)}}_{\widetilde{\mathbf{s}}_X} = o_{\mathrm{p}}\left(\widetilde{\mathbf{T}}_{*,X}\right). \tag{42}$$

81

Define the following statistic

$$T_g = \frac{1}{n} \sum_{i=1}^{n} K\left(\frac{Z_i}{h}\right) g\left(\xi_i\right) \left(\widehat{\mu}_i\left(Z_i\right) - \mu\left(Z_i\right) - \widehat{\mu}_i(0) + \mu(0)\right). \tag{43}$$

Then, $\widetilde{\mathbf{T}}_{*,X} + \widetilde{\mathbf{S}}_X = T_{g_X}$, where

$$g_X(\xi) = -\left(\delta(0) + \delta'(0)Z\right) \cdot X\left(Z, \mu_r(Z), \mu_r(0), \nu_r(Z), \nu_r(0)\right). \tag{44}$$

Similarly, $\widetilde{\mathbf{S}}_\varepsilon$ in (40) can be written as $\widetilde{\mathbf{S}}_\varepsilon = T_{g_\varepsilon}$, where

$$g_\varepsilon(\xi) = (0, 0, 0, 0, Y - Y(Z), Z \cdot (Y - Y(Z)), 0, 0). \tag{45}$$

We can show that $T_g$ is the sum of mean-zero i.i.d. random variables with an additional remainder term (see Lemma 12). To that end, define

$$\dot{\eta}_{g,2,h,r}\left(\xi_1, \xi_2\right) = K\left(Z_1/h\right) g\left(\xi_1\right) \left(\dot{\varphi}_r\left(\xi_2, Z_1\right) - \dot{\varphi}_r\left(\xi_2, 0\right)\right)$$
$$+ K\left(Z_2/h\right) g\left(\xi_2\right) \left(\dot{\varphi}_r\left(\xi_1, Z_2\right) - \dot{\varphi}_r\left(\xi_1, 0\right)\right),$$
$$\text{and} \quad \dot{\eta}_{g,1,h,r}\left(\xi_1\right) = \mathbf{E}\left[\dot{\eta}_{g,2,h,r}\left(\xi_1, \xi_2\right)|\xi_1\right] \tag{46}$$
$$= \mathbf{E}\left[K\left(Z_2/h\right) g\left(\xi_2\right) \left(\dot{\varphi}_r\left(\xi_1, Z_2\right) - \dot{\varphi}_r\left(\xi_1, 0\right)\right)|\xi_1\right],$$
$$\text{and for brevity,} \quad \dot{\eta}_{g,j} \equiv \dot{\eta}_{g,j,h,r} \quad \text{for } j \in \{1, 2\}.$$

$\dot{\eta}_{g,1}$ is the first-order Hajek projection of $\dot{\eta}_{g,2}$, and $\dot{\varphi}_r$ (defined in (84)) is an influence function for the difference $\widehat{\mu}_i(z) - \mu(z)$ based on a first order Taylor approximation of reciprocals.

The negligibility claim in (42) follows from Lemmas 12, 13 and 14 applied componentwise to each component of $g_\varepsilon$ and $g_X$.

## D.12  Proof of Lemma 12

For this proof, write $\varphi \equiv \varphi_r$, $\dot{\varphi} \equiv \dot{\varphi}_r$, $\mu \equiv \mu_r$, $\nu \equiv \nu_r$ and $\pi \equiv \pi_r$.

### D.12.1 Proof of Lemma 12: break down of $T_g$

Break $T_g$ in (43) into parts and use the representation in (93),

$$
\begin{aligned}
T_g &= \frac{1}{n} \sum_{i=1}^{n} K\left(Z_i/h\right) g\left(\xi_i\right) \left(\widehat{\mu}_i\left(Z_i\right) - \mu\left(Z_i\right)\right) - \frac{1}{n} \sum_{i=1}^{n} K\left(Z_i/h\right) g\left(\xi_i\right) \left(\widehat{\mu}_i(0) - \mu(0)\right) \\
&= \frac{1}{n} \sum_{i=1}^{n} K\left(Z_i/h\right) g\left(\xi_i\right) \left(\widehat{P}_{-i}\left[\dot{\varphi}\left(\cdot, Z_i\right)\right] + \widehat{\mathrm{Rem}}_{i,r}\left(Z_i\right)\right) \\
&\quad - \frac{1}{n} \sum_{i=1}^{n} K\left(Z_i/h\right) g\left(\xi_i\right) \left(\widehat{P}_{-i}\left[\dot{\varphi}(\cdot, 0)\right] + \widehat{\mathrm{Rem}}_{i,r}(0)\right).
\end{aligned}
$$

Thus,

$$
\begin{aligned}
T_g &= T_g^{(1)} + R_g, \\
\text{where} \quad T_g^{(1)} &= \frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} K\left(Z_i/h\right) g\left(\xi_i\right) \left(\dot{\varphi}\left(\xi_j, Z_i\right) - \dot{\varphi}\left(\xi_j, 0\right)\right), \\
R_g &= \frac{1}{n} \sum_{i=1}^{n} K\left(Z_i/h\right) g\left(\xi_i\right) \left(\widehat{\mathrm{Rem}}_{i,r}\left(Z_i\right) - \widehat{\mathrm{Rem}}_{i,r}(0)\right)
\end{aligned} \tag{47}
$$

Turn the expression for $T_g^{(1)}$ into that of a second order U-statistic with a symmetric kernel: combine (47) with the identity $\sum_{i,j:i \neq j} A_{ij} = \sum_{i,j:i<j} \left(A_{ij} + A_{ji}\right) = 2 \sum_{i,j:i<j} \frac{1}{2} \left(A_{ij} + A_{ji}\right)$ and $\dot{\eta}_{g,2}$ in (46) to get

$$
\begin{aligned}
T_g &= T_g^{(1)} + R_g, \\
\text{where} \quad T_g^{(1)} &= \frac{1}{n(n-1)} \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} \dot{\eta}_{g,2}\left(\xi_i, \xi_j\right), \\
R_g &= \frac{1}{n} \sum_{i=1}^{n} K\left(Z_i/h\right) g\left(\xi_i\right) \left(\widehat{\mathrm{Rem}}_{i,r}\left(Z_i\right) - \widehat{\mathrm{Rem}}_{i,r}(0)\right).
\end{aligned} \tag{48}
$$

Now (19) follows from (48) and noting that $U_g = T_g^{(1)} - T_{*,g}$ (by definition in (19)) so that $T_g^{(1)} = T_{*,g} + U_g$.

## D.13  Proof of Lemma 13

Recall from (19) and (46) that

$$T_{*,g} = \frac{1}{n} \sum_{i=1}^{n} \dot{\eta}_{g,1}\left(\xi_i\right), \text{ where } \dot{\eta}_{g,1}\left(\xi_1\right) = \mathbf{E}\left[K\left(Z_2/h\right) g\left(\xi_2\right) \left(\dot{\varphi}\left(\xi_1, Z_2\right) - \dot{\varphi}\left(\xi_1, 0\right)\right)|\xi_1\right]$$

Since $\mathbf{E}\left[\dot{\eta}_{g,1}\left(\xi_1\right)\right] = 0$ and $\xi_i$ are independent, we get

$$\mathbf{E}\left[T_{*,g}^2\right] = \frac{\mathbf{E}\left[\dot{\eta}_{g1}\left(\xi_1\right)^2\right]}{n}.$$

Define the "un-centered" version of $\dot{\eta}_{g,1}$ by

$$\eta_{g,1}\left(\xi_1\right) = \mathbf{E}\left[K\left(Z_2/h\right) g\left(\xi_2\right) \left(\varphi\left(\xi_1, Z_2\right) - \varphi\left(\xi_1, 0\right)\right)|\xi_1\right] \tag{49}$$

Recall from (90) that

$$\varphi\left(\xi_1, z\right) = \left(Y_1 - \mu(z)\right) \frac{R\left(\frac{Z_1 - z}{r}\right)}{\pi(z)}.$$

Hence combining (90) with (49) yields

$$\eta_{g,1}\left(\xi_1\right) = \mathbf{E}\left[K\left(\frac{Z_2}{h}\right) g\left(\xi_2\right) \left(\left(Y_1 - \mu\left(Z_2\right)\right) \frac{R\left(\frac{Z_1 - Z_2}{r}\right)}{\pi\left(Z_2\right)} - \left(Y_1 - \mu(0)\right) \frac{R\left(\frac{Z_1}{r}\right)}{\pi(0)}\right)\bigg|\xi_1\right]. \tag{50}$$

Break this down as follows

$$\eta_{g,1,h,r}\left(\xi_1\right) = \alpha_{1,h,r}\left(\xi_1\right) + \alpha_{2,h,r}\left(\xi_1\right) - \alpha_{3,h,r}\left(\xi_1\right),$$

$$\alpha_{1,h,r}\left(\xi_1\right) = \varepsilon_1 \mathbf{E}\left[K\left(\frac{Z_2}{h}\right) g\left(\xi_2\right) \left(\frac{R\left(\frac{Z_1 - Z_2}{r}\right)}{\pi\left(Z_2\right)} - \frac{R\left(\frac{Z_1}{r}\right)}{\pi(0)}\right)\bigg|Z_1\right],$$

$$\alpha_{2,h,r}\left(\xi_1\right) = \mathbf{E}\left[K\left(\frac{Z_2}{h}\right) g\left(\xi_2\right) \left(Y\left(Z_1\right) - \mu\left(Z_2\right)\right) \left(\frac{R\left(\frac{Z_1 - Z_2}{r}\right)}{\pi\left(Z_2\right)} - \frac{R\left(\frac{Z_1}{r}\right)}{\pi(0)}\right)\bigg|Z_1\right], \tag{51}$$

$$\alpha_{3,h,r}\left(\xi_1\right) = \mathbf{E}\left[K\left(\frac{Z_2}{h}\right) g\left(\xi_2\right) \left(\mu\left(Z_2\right) - \mu(0)\right)\right] \frac{R\left(\frac{Z_1}{r}\right)}{\pi(0)}.$$

As before, we suppress reference to $h, r$ and furthermore write $\alpha_j = \alpha_{j,h,r}\left(\xi_1\right)$. We now derive moments of the $\alpha_j$'s and their approximate representations. By $\mathbf{E}\left[\varepsilon_1|Z_1\right] = 0$, $\mathbf{E}\left[\alpha_1\right] = 0$ and $\mathbf{E}\left[\alpha_1 \cdot \alpha_j\right] = 0$ for each $j \in \{2, 3\}$.

Throughout, the following approximation result will be useful:

$$f \text{ bounded and uniformly continuous} \implies \lim_{r \to 0} \sup_z \left| \frac{\pi(z)}{2r} - f(z) \right| = 0. \qquad (52)$$

Our use of (52) will be that normalization by division by a factor $2r$ results in an "asymptotic stabilization" of $\pi$.

### D.13.1 Moments for $\alpha_1$

Use (51) to write

$$\alpha_1(\xi_1) = \varepsilon(\xi_1) \cdot \dot{\varphi}_{\alpha,1}(Z_1),$$
$$\text{where } \dot{\varphi}_{\alpha,1}(Z_1) = \mathbf{E}\left[ K\left( \frac{Z_2}{h} \right) g(\xi_2) \left( \frac{R\left( \frac{Z_1 - Z_2}{r} \right)}{\pi(Z_2)} - \frac{R\left( \frac{Z_1}{r} \right)}{\pi(0)} \right) \middle| Z_1 \right]. \qquad (53)$$

In what follows, define

$$m_g(z) = \mathbf{E}[g(\xi)|Z = z]. \qquad (54)$$

For fixed $z_1$, by a standard change of variables,

$$\dot{\varphi}_{\alpha,1}(z_1) = \frac{h}{2r} \int f(uh) K(u) m_g(uh) \left( \frac{R\left( \frac{z_1}{r} - u\frac{h}{r} \right)}{\pi(uh)/2r} - \frac{R\left( \frac{z_1}{r} \right)}{\pi(0)/2r} \right) \mathrm{d}u. \qquad (55)$$

Then

$$\dot{\varphi}_{\alpha,1}(vr) = \frac{h}{2r} \int f(uh) K(u) m_g(uh) \left( \frac{R\left( v - u\frac{h}{r} \right)}{\pi(uh)/2r} - \frac{R(v)}{\pi(0)/2r} \right) \mathrm{d}u.$$

Define

$$\varphi_{\alpha,1}(v; h, r) = \int f(uh) K(u) m_g(uh) \left( \frac{R\left( v - u\frac{h}{r} \right)}{\pi(uh)/2r} - \frac{R(v)}{\pi(0)/2r} \right) \mathrm{d}u,$$
$$\text{so that} \quad \dot{\varphi}_{\alpha,1}(vr) = \frac{h}{2r} \varphi_{\alpha,1}(v; h, r). \qquad (56)$$

Along sequences $(h, r) = (h_n, r_n)$ such that $\liminf(h/r) \in (0, \infty)$, the quantity $\varphi_{\alpha,1}$ is asymptotically stable.

85

Denote $\sigma^2(Z_1) = \mathbf{E}\left[\varepsilon_1^2 \middle| Z_1\right]$. Then

$$
\begin{aligned}
\operatorname{Var}\left[\alpha_1\right] = \mathbf{E}\left[\alpha_1^2\right] &= \mathbf{E}\left[\varepsilon_1^2 \cdot \dot\varphi_{\alpha,1}\left(Z_1\right)^2\right] = \mathbf{E}\left[\mathbf{E}\left[\varepsilon_1^2 \middle| Z_1\right] \cdot \dot\varphi_{\alpha,1}\left(Z_1\right)^2\right] \\
&= \mathbf{E}\left[\sigma^2\left(Z_1\right)\dot\varphi_{\alpha,1}\left(Z_1\right)^2\right] = \int f\left(z_1\right)\sigma^2\left(z_1\right)\dot\varphi_{\alpha,1}\left(z_1\right)^2 \; \mathrm{d}z_1 \\
&= r \int f(vr)\sigma^2(vr)\dot\varphi_{\alpha,1}(vr)^2 \; \mathrm{d}v.
\end{aligned}
$$

Combine this with (56) to get

$$
\operatorname{Var}\left[\alpha_1\right] = h\frac{h}{4r}\int f(vr)\sigma^2(vr)\varphi_{\alpha,1}(v; h, r)^2 \mathrm{d}v. \tag{57}
$$

Next denote $\rho_3(Z_1) = \mathbf{E}\left[|\varepsilon_1|^3 \middle| Z_1\right]$. Then

$$
\begin{aligned}
\mathbf{E}\left[|\alpha_1|^3\right] &= \mathbf{E}\left[|\varepsilon_1|^3 \, |\dot\varphi_{\alpha,1}\left(Z_1\right)|^3\right] = \mathbf{E}\left[\rho_3\left(Z_1\right)|\dot\varphi_{\alpha,1}\left(Z_1\right)|^3\right] \\
&= \int f\left(z_1\right)\rho_3\left(z_1\right)|\dot\varphi_{\alpha,1}\left(z_1\right)|^3 \, \mathrm{d}z_1 = r\int f(vr)\rho_3(vr)\,|\dot\varphi_{\alpha,1}(vr)|^3 \, \mathrm{d}v \\
&= r\int f(vr)\rho_3(vr)\,|\dot\varphi_{\alpha,1}(vr)|^3 \, \mathrm{d}v.
\end{aligned}
$$

Combine this with (56) to get

$$
\mathbf{E}\left[|\alpha_1|^3\right] = h\frac{h^2}{8r^2}\int f(vr)\rho_3(vr)\,|\varphi_{\alpha,1}(v; h, r)|^3 \, \mathrm{d}v. \tag{58}
$$

Thus combining $\mathbf{E}\left[\alpha_1\right] = 0$ with (57) and (58), we get

$$
\begin{aligned}
\mathbf{E}\left[\alpha_1\right] = 0, \qquad \mathbf{E}\left[\alpha_1^2\right] &= h\frac{h}{4r}\int f(vr)\sigma^2(vr)\varphi_{\alpha,1}(v; h, r)^2 \mathrm{d}v, \\
\mathbf{E}\left[|\alpha_1|^3\right] &= h\frac{h^2}{8r^2}\int f(vr)\rho_3(vr)\,|\varphi_{\alpha,1}(v; h, r)|^3 \, \mathrm{d}v.
\end{aligned} \tag{59}
$$

### D.13.2   Moments for $\alpha_2$

Use (51) and (54) to write

$$
\alpha_2 = \alpha_2\left(Z_1\right) := \mathbf{E}\left[K\left(\frac{Z_2}{h}\right)m_g\left(Z_2\right)\left(Y\left(Z_1\right) - \mu\left(Z_2\right)\right)\left(\frac{R\left(\frac{Z_1-Z_2}{r}\right)}{\pi\left(Z_2\right)} - \frac{R\left(\frac{Z_1}{r}\right)}{\pi(0)}\right)\middle| Z_1\right] \tag{60}
$$

Then

$$\alpha_2\left(z_1\right) = \int K\left(\frac{z_2}{h}\right) m_g\left(z_2\right)\left(Y\left(z_1\right) - \mu\left(z_2\right)\right)\left(\frac{R\left(\frac{z_1-z_2}{r}\right)}{\pi\left(z_2\right)} - \frac{R\left(\frac{z_1}{r}\right)}{\pi(0)}\right) f\left(z_2\right) \, dz_2$$

By the change of variables $z_2 = uh$,

$$\alpha_2\left(z_1\right) = \frac{h}{2r}\int K(u)f(uh)m_g(uh)\left(Y\left(z_1\right) - \mu(uh)\right)\left(\frac{R\left(\frac{z_1}{r} - u\frac{h}{r}\right)}{\pi(uh)/2r} - \frac{R\left(\frac{z_1}{r}\right)}{\pi(0)/2r}\right) \, du.$$

Therefore,

$$\alpha_2(vr) = \frac{h}{2r}\int K(u)f(uh)m_g(uh)\left(Y(vr) - \mu(uh)\right)\left(\frac{R\left(v - u\frac{h}{r}\right)}{\pi(uh)/2r} - \frac{R(v)}{\pi(0)/2r}\right) \, du.$$

Let

$$\varphi_{\alpha,2}(v;h,r) = \int K(u)f(uh)m_g(uh)\left(Y(vr) - \mu(uh)\right)\left(\frac{R\left(v - u\frac{h}{r}\right)}{\pi(uh)/2r} - \frac{R(v)}{\pi(0)/2r}\right) \, du.$$

Then

$$\alpha_2(vr) = \frac{h}{2r}\varphi_{\alpha,2}(v;h,r). \tag{61}$$

Along sequences $(h,r) = (h_n, r_n)$ such that $\liminf(h/r) \in (0,\infty)$, the quantity $\varphi_{\alpha,2}$ is asymptotically stable.

By the change of variables $z_2 = vr$,

$$\mathbf{E}\left[\alpha_2\right] = \int \alpha_2\left(z_1\right) f\left(z_1\right) \, dz_1 = r\int \alpha_2(vr)f(vr) \, dv.$$

Therefore using (61),

$$\mathbf{E}\left[\alpha_2\right] = \frac{h}{2}\int \varphi_{\alpha,2}(v;h,r)f(vr) \, dv.$$

For the second moment

$$\mathbf{E}\left[\alpha_2^2\right] = \int \alpha_2\left(z_1\right)^2 f\left(z_1\right) \, dz_1 = r\int \alpha_2(vr)^2 f(vr) \, dv,$$

and so, (61) yields

$$\mathbf{E}\left[\alpha_2^2\right] = h\frac{h}{4r}\int \varphi_{\alpha,2}(v;h,r)^2 f(vr)\,\mathrm{d}v.$$

A similar calculation for the third absolute moment using (61) yields

$$\mathbf{E}\left[|\alpha_2|^3\right] = h\frac{h^2}{8r^2}\int |\varphi_{\alpha,2}(v;h,r)|^3 f(vr)\,\mathrm{d}v.$$

Combine all of the above to get

$$\mathbf{E}\left[\alpha_2\right] = \frac{h}{2}\int \varphi_{\alpha,2}(v;h,r)f(vr)\,\mathrm{d}v,\quad \mathbf{E}\left[\alpha_2^2\right] = h\frac{h}{4r}\int \varphi_{\alpha,2}(v;h,r)^2 f(vr)\,\mathrm{d}v$$
$$\mathbf{E}\left[|\alpha_2|^3\right] = h\frac{h^2}{8r^2}\int |\varphi_{\alpha,2}(v;h,r)|^3 f(vr)\,\mathrm{d}v. \tag{62}$$

### D.13.3 Moments for $\alpha_3$

From (51),

$$\alpha_3 = \dot{\varphi}_{\alpha,3}(h,r)\cdot\frac{R(Z/r)}{\pi(0)},\quad \text{where}\quad \dot{\varphi}_{\alpha,3} = \mathbf{E}\left[K\left(\frac{Z}{h}\right)m_g(Z)\left(\mu(Z)-\mu(0)\right)\right].$$

Therefore, since $\mathbf{E}\left[\frac{R(Z/r)}{\pi(0)}\right] = 1$,

$$\mathbf{E}\left[\alpha_3\right] = \dot{\varphi}_{\alpha,3},\quad \mathbf{E}\left[\alpha_3^2\right] = \dot{\varphi}_{\alpha,3}^2\frac{\mathbf{E}\left[R(Z/r)^2\right]}{\pi(0)^2},\quad \text{and}\quad \mathbf{E}\left[|\alpha_3|^3\right] = |\dot{\varphi}_{\alpha,3}|^3\frac{\mathbf{E}\left[R(Z/r)^3\right]}{\pi(0)^3}.$$

Furthermore

$$\dot{\varphi}_{\alpha,3} = \mathbf{E}\left[K\left(\frac{Z}{h}\right)m_g(Z)\left(\mu(Z)-\mu(0)\right)\right] = \int K\left(\frac{z}{h}\right)m_g(z)\left(\mu(z)-\mu(0)\right)f(z)\mathrm{d}z,$$

and so by change of variables $z = uh$,

$$\dot{\varphi}_{\alpha,3} = h\int K(u)m_g(uh)\left(\mu(uh)-\mu(0)\right)f(uh)\mathrm{d}u.$$

Therefore

$$\dot{\varphi}_{\alpha,3} = h \cdot \frac{\pi(0)}{2r} \cdot \varphi_{\alpha,3}(h,r), \quad \text{where}$$
$$\varphi_{\alpha,3}(h,r) = \frac{2r}{\pi(0)} \int K(u) m_g(uh) \left(\mu(uh) - \mu(0)\right) f(uh) \mathrm{d}u. \tag{63}$$

For the second moment, from (D.13.3),

$$\mathbf{E}\left[\alpha_3^2\right] = \frac{h^2}{4r^2} \varphi_{\alpha,3}^2(h,r) \int R\left(\frac{z}{r}\right)^2 f(z) \,\mathrm{d}z,$$

and so

$$\mathbf{E}\left[\alpha_3^2\right] = h\frac{h}{4r} \varphi_{\alpha,3}^2(h,r) \int R(v)^2 f(vr) \,\mathrm{d}v. \tag{64}$$

For the third moment, essentially the same calculation as that for the second moment yields

$$\mathbf{E}\left[|\alpha_3|^3\right] = h\frac{h^2}{8r^2} |\varphi_{\alpha,3}(h,r)|^3 \int R(v)^3 f(vr) \,\mathrm{d}v. \tag{65}$$

Combine (63), (64) and (65):

$$\mathbf{E}\left[\alpha_3\right] = h\frac{\pi(0)}{2r}\varphi_{\alpha,3}(h,r), \quad \mathbf{E}\left[\alpha_2^2\right] = h\frac{h}{4r}\varphi_{\alpha,3}(h,r)^2 \int R(v)^2 f(vr) \,\mathrm{d}v,$$
$$\mathbf{E}\left[|\alpha_3|^3\right] = h\frac{h^2}{8r^2} |\varphi_{\alpha,3}(h,r)|^3 \int R(v)^3 f(vr) \,\mathrm{d}v. \tag{66}$$

## D.14 Proof of Lemma 14

**Proof of Lemma 14: Variance of $U_g$**

For (21), first note that $\dot{\eta}_{g,1}$ is the first order projection of $\dot{\eta}_{g,2}$, since

$$\mathbf{E}\left[\dot{\eta}_{g,2}\left(\xi_1,\xi_2\right)|\xi_1\right] = \mathbf{E}\left[K\left(Z_1/h\right) g\left(\xi_1\right)\left(\dot{\varphi}\left(\xi_2,Z_1\right) - \dot{\varphi}\left(\xi_2,0\right)\right)|\xi_1\right]$$
$$+ \mathbf{E}\left[K\left(Z_2/h\right) g\left(\xi_2\right)\left(\dot{\varphi}\left(\xi_1,Z_2\right) - \dot{\varphi}\left(\xi_1,0\right)\right)|\xi_1\right]$$
$$= K\left(Z_1/h\right) g\left(\xi_1\right)\left(\mathbf{E}\left[\dot{\varphi}\left(\xi_2,Z_1\right)|Z_1\right] - \mathbf{E}\left[\dot{\varphi}\left(\xi_2,0\right)|\xi_1\right]\right)$$
$$+ \mathbf{E}\left[K\left(Z_2/h\right) g\left(\xi_2\right)\left(\dot{\varphi}\left(\xi_1,Z_2\right) - \dot{\varphi}\left(\xi_1,0\right)\right)|\xi_1\right].$$

Since $\mathbf{E}\left[\dot\varphi\left(\xi_2, z\right)\right] = 0$ for every $z$,

$$\mathbf{E}\left[\dot\eta_{g,2}\left(\xi_1, \xi_2\right)|\xi_1\right] = \mathbf{E}\left[K\left(Z_2/h\right)g\left(\xi_2\right)\left(\dot\varphi\left(\xi_1, Z_2\right) - \dot\varphi\left(\xi_1, 0\right)\right)|\xi_1\right] = \dot\eta_{g,1}\left(\xi_1\right).$$

Thus $\dot\eta_{g,1}$ is the first order projection of $\dot\eta_{g,2}$. Hence (21) follows from (67) in Lemma 15.

**Lemma 15.** *Let $P$ be a probability measure and for $m \in \mathbb{N} \setminus \{1\}$, let $\eta \in \mathcal{L}_2\left(P^m\right)$ be a $m^{th}$ order kernel that is permutation symmetric. In the case of $m = 2$, we have the equality*

$$\mathrm{E}\left[\left\|U_n(\eta) - \widehat{U}_n(\eta)\right\|^2\right] = \frac{2}{n(n-1)} \mathrm{tr}\left(\mathrm{E}\left[\mathrm{Var}\left[\eta\left(\xi_1, \xi_2\right)|\xi_1\right]\right] - \mathrm{Var}\left[\mathrm{E}\left[\eta\left(\xi_1, \xi_2\right)|\xi_1\right]\right]\right). \quad (67)$$

*More generally, for any $m \geq 2$, we have the bound*

$$\begin{aligned}
\mathrm{E}\left[\left\|U_n(\eta) - \widehat{U}_n(\eta)\right\|^2\right] &\leq \frac{m^2(m-1)^2\left(1 + \frac{m-1}{n-m+1}\right)}{n^2}\zeta_1(\eta) \\
&\quad + \sum_{c=2}^{m} \frac{1}{\left|\mathrm{Inj}_{n,c}\right|} \cdot \frac{(m!)^2}{((m-c)!)^2 c!}\left(1 + \delta_{n,m,c}\right)\zeta_c(\eta),
\end{aligned} \quad (68)$$

*where for each $c \in \{2, \ldots, m\}$, the quantities $\delta_{n,m,c}$ do not depend on $\eta$, $P$ or the dimension of $\xi_i$ (except possibly through $n$ and $m$) and satisfy $\lim_{n\to\infty} \delta_{n,m,c} = 0$.*

The result (67) and the definitions of the covariances $\zeta_c$ can be found in the chapter on U-statistics in van der Vaart (1998). The result (68) is not difficult to derive.

**Proof of Lemma 14: rate of convergence of $U_g$**

Now for the rate of convergence, first write

$$\dot\eta_{g,2}\left(\xi_1, \xi_2\right) = \dot\eta_*\left(\xi_1, \xi_2\right) + \dot\eta_*\left(\xi_2, \xi_1\right),$$

$$\dot\eta_*\left(\xi_1, \xi_2\right) = \dot\eta_{*,h,r}\left(\xi_1, \xi_2\right) = K\left(Z_1/h\right)g\left(\xi_1\right)\left(\dot\varphi\left(\xi_2, Z_1\right) - \dot\varphi\left(\xi_2, 0\right)\right).$$

Since $\mathbf{E}\left[\dot\varphi\left(\xi_2, z_1\right)\right] = 0$ for every $z_1$ (so that $\mathbf{E}\left[\dot\varphi\left(\xi_2, Z_1\right)|Z_1\right] = 0$),

$$\mathrm{Var}\left[\dot\eta_{g,2}\right] = \mathbf{E}\left[\dot\eta_{g,2}^2\right] \quad \text{and} \quad \mathrm{Var}\left[\dot\eta_*\right] = \mathbf{E}\left[\dot\eta_*^2\right].$$

By the $C_r$ inequality,

$$\text{Var}\left[\dot{\eta}_{g,2}\left(\xi_1,\xi_2\right)\right] = \mathbf{E}\left[\dot{\eta}_{g,2}\left(\xi_1,\xi_2\right)^2\right] \leq 2\left(\mathbf{E}\left[\dot{\eta}_*\left(\xi_1,\xi_2\right)^2\right] + \mathbf{E}\left[\dot{\eta}_*\left(\xi_2,\xi_1\right)^2\right]\right).$$

and so

$$\text{Var}\left[\dot{\eta}_{g,2}\left(\xi_1,\xi_2\right)\right] \leq 4\mathbf{E}\left[\dot{\eta}_*\left(\xi_1,\xi_2\right)^2\right]. \tag{69}$$

Denote

$$m_{g^2}\left(Z_1\right) = \mathbf{E}\left[g\left(\xi_1\right)^2\Big|Z_1\right].$$

Expanding the moment in the upper bound in (69),

$$\mathbf{E}\left[\dot{\eta}_*\left(\xi_1,\xi_2\right)^2\right] = \mathbf{E}\left[K\left(\frac{Z_1}{h}\right)^2 g\left(\xi_1\right)^2\left\{\dot{\varphi}\left(\xi_2,Z_1\right) - \dot{\varphi}\left(\xi_2,0\right)\right\}^2\right]$$

$$= \mathbf{E}\left[K\left(\frac{Z_1}{h}\right)^2 m_{g^2}\left(Z_1\right)\left\{\dot{\varphi}\left(\xi_2,Z_1\right) - \dot{\varphi}\left(\xi_2,0\right)\right\}^2\right]$$

$$= \mathbf{E}\left[K\left(\frac{Z_1}{h}\right)^2 m_{g^2}\left(Z_1\right)\text{Var}\left[\varphi\left(\xi_2,Z_1\right) - \varphi\left(\xi_2,0\right)|Z_1\right]\right].$$

Denote

$$v^2\left(Z_1\right) = \text{Var}\left[\varphi\left(\xi_2,Z_1\right) - \varphi\left(\xi_2,0\right)|Z_1\right]. \tag{70}$$

Then,

$$\mathbf{E}\left[\dot{\eta}_*\left(\xi_1,\xi_2\right)^2\right] = \mathbf{E}\left[K\left(\frac{Z_1}{h}\right)^2 m_{g^2}\left(Z_1\right)v^2\left(Z_1\right)\right] = \int f\left(z_1\right)K\left(\frac{z_1}{h}\right)m_{g^2}\left(z_1\right)v^2\left(z_1\right)\,\mathrm{d}z_1,$$

and so by the change of variables $z_1 = uh$,

$$\mathbf{E}\left[\dot{\eta}_*\left(\xi_1,\xi_2\right)^2\right] = h\int K(u)^2 f(uh)m_{g^2}(uh)v^2(uh)\,\mathrm{d}u. \tag{71}$$

Recall from (90) that

$$\varphi\left(\xi_2,z_1\right) = \left(Y_2 - \mu\left(z_1\right)\right)\frac{R\left(\frac{Z_2-z_1}{r}\right)}{\pi\left(z_1\right)}.$$

With some algebra, we can show

$$
\begin{aligned}
\varphi\left(\xi_2, z_1\right) - \varphi\left(\xi_2, 0\right) = \varepsilon_2 & \left\{\frac{R\left(\frac{Z_2-z_1}{r}\right)}{\pi\left(z_1\right)} - \frac{R\left(\frac{Z_2}{r}\right)}{\pi(0)}\right\} \\
& + \left\{Y\left(Z_2\right) - \mu\left(z_1\right)\right\}\left\{\frac{R\left(\frac{Z_2-z_1}{r}\right)}{\pi\left(z_1\right)} - \frac{R\left(\frac{Z_2}{r}\right)}{\pi(0)}\right\} \\
& + \left\{\mu(0) - \mu\left(z_1\right)\right\} \frac{R\left(\frac{Z_2}{r}\right)}{\pi(0)}
\end{aligned}
\tag{72}
$$

Revisiting (70),

$$
v^2\left(z_1\right) = \operatorname{Var}\left[\varphi\left(\xi_2, Z_1\right) - \varphi\left(\xi_2, 0\right) \mid Z_1 = z_1\right] = \operatorname{Var}\left[\varphi\left(\xi_2, z_1\right) - \varphi\left(\xi_2, 0\right)\right],
$$

and so,

$$
v^2\left(z_1\right) = \mathbf{E}\left[\left\{\varphi\left(\xi_2, z_1\right) - \varphi\left(\xi_2, 0\right)\right\}^2\right] - \left(\mathbf{E}\left[\varphi\left(\xi_2, z_1\right) - \varphi\left(\xi_2, 0\right)\right]\right)^2.
\tag{73}
$$

Therefore using (72), the two moments in (73) can be written as

$$
\begin{aligned}
\mathbf{E}\left[\varphi\left(\xi_2, z_1\right) - \varphi\left(\xi_2, 0\right)\right] = \mathbf{E} & \left[\left(Y\left(Z_2\right) - \mu\left(z_1\right)\right)\left(\frac{R\left(\frac{Z_2-z_1}{r}\right)}{\pi\left(z_1\right)} - \frac{R\left(\frac{Z_2}{r}\right)}{\pi(0)}\right)\right] \\
& + \left(\mu(0) - \mu\left(z_1\right)\right) \mathbf{E}\left[\frac{R\left(\frac{Z_2}{r}\right)}{\pi(0)}\right]
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{E}\left[\left\{\varphi\left(\xi_2, z_1\right) - \varphi\left(\xi_2, 0\right)\right\}^2\right] = \mathbf{E} & \left[\sigma^2\left(Z_2\right)\left(\frac{R\left(\frac{Z_2-z_1}{r}\right)}{\pi\left(z_1\right)} - \frac{R\left(\frac{Z_2}{r}\right)}{\pi(0)}\right)^2\right] \\
& + \mathbf{E}\left[\left\{\begin{array}{c}\left(Y\left(Z_2\right) - \mu\left(z_1\right)\right)\left(\frac{R\left(\frac{Z_2-z_1}{r}\right)}{\pi\left(z_1\right)} - \frac{R\left(\frac{Z_2}{r}\right)}{\pi(0)}\right) \\ + \left(\mu(0) - \mu\left(z_1\right)\right)\frac{R\left(\frac{Z_2}{r}\right)}{\pi(0)}\end{array}\right\}^2\right]
\end{aligned}
$$

We can then bound the second moment so that

$$\mathbf{E}\left[\varphi\left(\xi_2, z_1\right) - \varphi\left(\xi_2, 0\right)\right] = \mathbf{E}\left[\left(Y\left(Z_2\right) - \mu\left(z_1\right)\right)\left(\frac{R\left(\frac{Z_2 - z_1}{r}\right)}{\pi\left(z_1\right)} - \frac{R\left(\frac{Z_2}{r}\right)}{\pi(0)}\right)\right]$$

$$+ \mathbf{E}\left[\left(\mu(0) - \mu\left(z_1\right)\right)\frac{R\left(\frac{Z_2}{r}\right)}{\pi(0)}\right]$$

$$\mathbf{E}\left[\left\{\varphi\left(\xi_2, z_1\right) - \varphi\left(\xi_2, 0\right)\right\}^2\right] \leq \mathbf{E}\left[\sigma^2\left(Z_2\right)\left(\frac{R\left(\frac{Z_2 - z_1}{r}\right)}{\pi\left(z_1\right)} - \frac{R\left(\frac{Z_2}{r}\right)}{\pi(0)}\right)^2\right] \tag{74}$$

$$+ 2\mathbf{E}\left[\left(Y\left(Z_2\right) - \mu\left(z_1\right)\right)^2\left(\frac{R\left(\frac{Z_2 - z_1}{r}\right)}{\pi\left(z_1\right)} - \frac{R\left(\frac{Z_2}{r}\right)}{\pi(0)}\right)^2\right]$$

$$+ 2\left(\mu(0) - \mu\left(z_1\right)\right)^2\frac{\mathbf{E}\left[R\left(\frac{Z_2}{r}\right)^2\right]}{\pi(0)^2}.$$

For the first moment,

$$\mathbf{E}\left[\varphi\left(\xi_2, z_1\right) - \varphi\left(\xi_2, 0\right)\right] = \int f\left(z_2\right)\left(Y\left(z_2\right) - \mu\left(z_1\right)\right)\left(\frac{R\left(\frac{z_2 - z_1}{r}\right)}{\pi\left(z_1\right)} - \frac{R\left(\frac{z_2}{r}\right)}{\pi(0)}\right)\, \mathrm{d}z_2$$

$$+ \left(\mu(0) - \mu\left(z_1\right)\right)$$

$$= r\int f(v)\left(Y(vr) - \mu\left(z_1\right)\right)\left(\frac{R\left(v - \frac{z_1}{r}\right)}{\pi\left(z_1\right)} - \frac{R(v)}{\pi(0)}\right)\, \mathrm{d}v$$

$$+ \left(\mu(0) - \mu\left(z_1\right)\right),$$

where the last line follows by the change of variables $z_2 = vr$. Normalizing $\pi$ to $\pi/2r$ for stability, we get

$$\mathbf{E}\left[\varphi\left(\xi_2, z_1\right) - \varphi\left(\xi_2, 0\right)\right] = \frac{1}{2}\int f(v)\left(Y(vr) - \mu\left(z_1\right)\right)\left(\frac{R\left(v - \frac{z_1}{r}\right)}{\pi\left(z_1\right)/2r} - \frac{R(v)}{\pi(0)/2r}\right)\, \mathrm{d}v \tag{75}$$

$$+ \left(\mu(0) - \mu\left(z_1\right)\right)$$

For the variance, from (74)

$$\mathbf{E}\left[\{\varphi\left(\xi_2, z_1\right) - \varphi\left(\xi_2, 0\right)\}^2\right] \leq \int f\left(z_2\right) \sigma^2\left(z_2\right) \left(\frac{R\left(\frac{z_2-z_1}{r}\right)}{\pi\left(z_1\right)} - \frac{R\left(\frac{z_2}{r}\right)}{\pi(0)}\right)^2 \mathrm{d}z_2$$

$$+ 2 \int f\left(z_2\right) \left(Y\left(z_2\right) - \mu\left(z_1\right)\right)^2 \left(\frac{R\left(\frac{z_2-z_1}{r}\right)}{\pi\left(z_1\right)} - \frac{R\left(\frac{z_2}{r}\right)}{\pi(0)}\right)^2 \mathrm{d}z_2$$

$$+ 2\left(\mu(0) - \mu\left(z_1\right)\right)^2 \frac{\int f\left(z_2\right) R\left(\frac{z_2}{r}\right)^2 \mathrm{d}z_2}{\pi(0)}$$

$$= r \int f(vr)\sigma^2(vr) \left(\frac{R\left(v-\frac{z_1}{r}\right)}{\pi\left(z_1\right)} - \frac{R(v)}{\pi(0)}\right)^2 \mathrm{d}v$$

$$+ 2r \int f(vr)\left(Y(vr) - \mu\left(z_1\right)\right)^2 \left(\frac{R\left(v-\frac{z_1}{r}\right)}{\pi\left(z_1\right)} - \frac{R(vr)}{\pi(0)}\right)^2 \mathrm{d}v$$

$$+ 2\left(\mu(0) - \mu\left(z_1\right)\right)^2 \frac{r \int f(vr) R(v)^2 \mathrm{d}v}{\pi(0)^2}$$

Thus, normalizing $\pi$ to $\pi/2r$,

$$\mathbf{E}\left[\{\varphi\left(\xi_2, z_1\right) - \varphi\left(\xi_2, 0\right)\}^2\right] = \frac{1}{4r} \int f(vr)\sigma^2(vr) \left(\frac{R\left(v-\frac{z_1}{r}\right)}{\pi\left(z_1\right)/2r} - \frac{R(v)}{\pi(0)/2r}\right)^2 \mathrm{d}v$$

$$+ \frac{1}{2r} \int f(vr)\left(Y(vr) - \mu\left(z_1\right)\right)^2 \left(\frac{R\left(v-\frac{z_1}{r}\right)}{\pi\left(z_1\right)/2r} - \frac{R(vr)}{\pi(0)/2r}\right)^2 \mathrm{d}v$$

$$+ \frac{1}{2r}\left(\mu(0) - \mu\left(z_1\right)\right)^2 \frac{\int f(vr) R(v)^2 \mathrm{d}v}{(\pi(0)/2r)^2}$$

$$(76)$$

Thus, plugging (75) and (76), into (73)

$$
\begin{aligned}
v^2\left(z_1\right) \leq & \frac{1}{4r} \int f(vr)\sigma^2(vr)\left(\frac{R\left(v-\frac{z_1}{r}\right)}{\pi\left(z_1\right)/2r} - \frac{R(v)}{\pi(0)/2r}\right)^2 \mathrm{d}v \\
& + \frac{1}{2r} \int f(vr)\left(Y(vr) - \mu\left(z_1\right)\right)^2 \left(\frac{R\left(v-\frac{z_1}{r}\right)}{\pi\left(z_1\right)/2r} - \frac{R(vr)}{\pi(0)/2r}\right)^2 \mathrm{d}v \\
& + \frac{1}{2r}\left(\mu(0) - \mu\left(z_1\right)\right)^2 \frac{\int f(vr)R(v)^2\,\mathrm{d}v}{(\pi(0)/2r)^2} \\
& - \left(\begin{array}{l} \frac{1}{2}\int f(v)\left(Y(vr) - \mu\left(z_1\right)\right)\left(\frac{R\left(v-\frac{z_1}{r}\right)}{\pi(z_1)/2r} - \frac{R(v)}{\pi(0)/2r}\right)\,\mathrm{d}v \\ + \left(\mu(0) - \mu\left(z_1\right)\right) \end{array}\right)^2.
\end{aligned}
\tag{77}
$$

Thus evaluate the upper bound in (77) along sequences $z_1 = uh$ to conclude that

$$
\begin{aligned}
v^2\left(uh\right) \leq & \frac{1}{r}C_v(u;h,r) \\
C_v(u;h,r) = & \frac{1}{4} \int f(vr)\sigma^2(vr)\left(\frac{R\left(v-u\frac{h}{r}\right)}{\pi(uh)/2r} - \frac{R(v)}{\pi(0)/2r}\right)^2 \mathrm{d}v \\
& + \frac{1}{2} \int f(vr)\left(Y(vr) - \mu(uh)\right)^2 \left(\frac{R\left(v-u\frac{h}{r}\right)}{\pi(uh)/2r} - \frac{R(vr)}{\pi(0)/2r}\right)^2 \mathrm{d}v \\
& + \frac{1}{2}\left(\mu(0) - \mu(uh)\right)^2 \frac{\int f(vr)R(v)^2\,\mathrm{d}v}{(\pi(0)/2r)^2} \\
& - r\left(\begin{array}{l} \frac{1}{2}\int f(v)\left(Y(vr) - \mu(uh)\right)\left(\frac{R\left(v-u\frac{h}{r}\right)}{\pi(uh)/2r} - \frac{R(v)}{\pi(0)/2r}\right)\,\mathrm{d}v \\ + \left(\mu(0) - \mu\left(uh\right)\right) \end{array}\right)^2.
\end{aligned}
\tag{78}
$$

Combine (78) with (71) to get

$$
\mathbf{E}\left[\dot{\eta}_*\left(\xi_1, \xi_2\right)^2\right] \leq \frac{h}{r} \int K(u)^2 f(uh)m_{g^2}(uh)C_v(u;hr)\,\mathrm{d}u.
\tag{79}
$$

Furthermore, by (69)

$$
\operatorname{Var}\left[\dot{\eta}_{g,2}\left(\xi_1, \xi_2\right)\right] \leq 4\frac{h}{r} \int K(u)^2 f(uh)m_{g^2}(uh)C_v(u;hr)\,\mathrm{d}u.
$$

95

The inequality in (21) now follows from the above and the fact that

$$\mathrm{E}\left[\mathrm{Var}\left[\dot{\eta}_{g,2}\left(\xi_1,\xi_2\right)|\xi_1\right]\right] - \mathrm{Var}\left[\mathrm{E}\left[\dot{\eta}_{g,2}\left(\xi_1,\xi_2\right)|\xi_1\right]\right]$$

$$\leq \mathrm{E}\left[\mathrm{Var}\left[\dot{\eta}_{g,2}\left(\xi_1,\xi_2\right)|\xi_1\right]\right] + \mathrm{Var}\left[\mathrm{E}\left[\dot{\eta}_{g,2}\left(\xi_1,\xi_2\right)|\xi_1\right]\right]$$

$$= \mathrm{Var}\left[\dot{\eta}_{g,2}\left(\xi_1,\xi_2\right)\right].$$

## Proof of Lemma 14: rate of convergence of $U_g$ relative to $T_{g,*}$

Take the ratio of second moments: By the expression for $\mathbf{E}\left[U_g^2\right]$ in (21) and for $\mathbf{E}\left[T_{g,*}^2\right]$ in (20),

$$\frac{\mathbf{E}\left[U_g^2\right]}{\mathbf{E}\left[T_{*,g}^2\right]} \leq \frac{8}{(n-1)h} \cdot \frac{\int K(u)^2 f(uh) m_{g^2}(uh) C_v(u; hr)\,\mathrm{d}u}{(1+o(1))\left(\int f(vr)\sigma^2(vr)\varphi_{\alpha,1}(v;h,r)^2\mathrm{d}v + \int \varphi_{\alpha,2}(v;h,r)^2 f(vr)\,\mathrm{d}v\right)}.$$

Therefore, $U_g = o_{\mathrm{p}}\left(T_{*,g}\right)$ if $nh \to \infty$ since the variance ratio tends to zero.

## Proof of Lemma 14: rate of convergence of $R_g$ relative to $U_{g,*}$

Write

$$\mathcal{E} = \left\{\max_{i=1,\dots,n} \sup_{z \in [-1,1]} |\widehat{\pi}_i(z) - \pi(z)| \leq \frac{r}{2} \inf f\right\}.$$

We want to apply (100) in Lemma 19 with $y = \frac{r}{16} \inf f$ (so that $8y = \frac{r}{2} \inf f$). First, the side condition $y \geq \sqrt{\frac{Cr}{n}}$ is satisfied for $n$ large enough since $\frac{r}{16} \inf f \geq \sqrt{\frac{Cr}{n}}$ if and only if $\frac{1}{\sqrt{nr}} \leq \frac{\inf f}{16\sqrt{C}}$ and the latter occurs since $1/(nr) \to 0$. Then by (100) with $y = \frac{r}{16} \inf f$,

$$\Pr\left\{\mathcal{E}^c\right\} \leq A \exp\left\{-\frac{(n-1)(\inf f)^2 r}{265 C_*} + 2\log(116/(r \inf f)) + \log n\right\}$$

$$+ A \exp\left\{-C(n-1)r + 2\log(1/(Cr)) + \log n\right\}$$

$$\to 0.$$

since $(\log n)/(nr) \to 0$.

Now split

$$|R_g| = |R_g|\mathbf{1}_{\mathcal{E}} + |R_g|\left(1 - \mathbf{1}_{\mathcal{E}}\right) = (1 + o_{\mathrm{p}}(1))|R_g|\mathbf{1}_{\mathcal{E}}, \tag{80}$$

where the latter equality follows since $1 - \mathbf{1}_{\mathcal{E}} \xrightarrow{\mathrm{P}} 0$ and $1 - \mathbf{1}_{\mathcal{E}}$ is a 0-1 sequence, the

second term in the first equality above is negligible at any rate (i.e. for arbitrary norming sequences).

Recall from (19) that

$$R_g = \frac{1}{n} \sum_{i=1}^{n} K\left(Z_i/h\right) g\left(\xi_i\right) \left(\widehat{\text{Rem}}_{i,r}\left(Z_i\right) - \widehat{\text{Rem}}_{i,r}(0)\right),$$

and from (86):

$$\text{If} \quad |\widehat{\pi}_{i,r}(z) - \pi_r(z)| = \left|\left(\widehat{P}_{-i} - P\right)[\psi_{2,r,z}]\right| \leq \frac{r}{2}\inf f,$$

$$\text{then} \quad \left|\widehat{\text{Rem}}_{i,r,z}\right| \leq 2\left|\overline{\psi}_{1,r}(z)\right| \cdot \frac{\left[\left(\widehat{P}_{-i} - P\right)[\psi_{2,r,z}]\right]^2}{r^3(\inf f)^3}$$

$$+ \frac{\left[\left(\widehat{P}_{-i} - P\right)[\psi_{1,r,z}]\right]^2 + \left[\left(\widehat{P}_{-i} - P\right)[\psi_{2,r,z}]\right]^2}{r^2(\inf f)^2}.$$

Combine (80) with the above to get

$$|R_g| = (1 + o_{\text{p}}(1))\,\mathbf{1}_{\mathcal{E}}\,|R_g| \leq \frac{(1 + o_{\text{p}}(1))\,\mathbf{1}_{\mathcal{E}}}{n} \sum_{i=1}^{n} K\left(\frac{Z_i}{h}\right) |g\left(\xi_i\right)| \left(\left|\widehat{\text{Rem}}_{i,r}\left(Z_i\right)\right| + \left|\widehat{\text{Rem}}_{i,r}(0)\right|\right)$$

$$\leq \frac{(2 + o_{\text{p}}(1))\,\mathbf{1}_{\mathcal{E}}}{n} \sum_{i=1}^{n} K\left(\frac{Z_i}{h}\right) |g\left(\xi_i\right)| \left|\overline{\psi}_{1,r}\left(Z_i\right)\right| \cdot \frac{\left[\left(\widehat{P}_{-i} - P\right)[\psi_{2,r}\left(\cdot; Z_i\right)]\right]^2}{r^3(\inf f)^3}$$

$$+ \frac{(2 + o_{\text{p}}(1))\,\mathbf{1}_{\mathcal{E}}}{n} \sum_{i=1}^{n} K\left(\frac{Z_i}{h}\right) |g\left(\xi_i\right)| \left|\overline{\psi}_{1,r}\left(0\right)\right| \cdot \frac{\left[\left(\widehat{P}_{-i} - P\right)[\psi_{2,r}\left(\cdot; 0\right)]\right]^2}{r^3(\inf f)^3}$$

$$+ \frac{(1 + o_{\text{p}}(1))\,\mathbf{1}_{\mathcal{E}}}{n} \sum_{i=1}^{n} K\left(\frac{Z_i}{h}\right) |g\left(\xi_i\right)| \frac{\left[\left(\widehat{P}_{-i} - P\right)[\psi_{1,r}\left(\cdot; Z_i\right)]\right]^2}{r^2(\inf f)^2}$$

$$+ \frac{(1 + o_{\text{p}}(1))\,\mathbf{1}_{\mathcal{E}}}{n} \sum_{i=1}^{n} K\left(\frac{Z_i}{h}\right) |g\left(\xi_i\right)| \frac{\left[\left(\widehat{P}_{-i} - P\right)[\psi_{2,r}\left(\cdot; Z_i\right)]\right]^2}{r^2(\inf f)^2}$$

$$+ \frac{(1 + o_{\text{p}}(1))\,\mathbf{1}_{\mathcal{E}}}{n} \sum_{i=1}^{n} K\left(\frac{Z_i}{h}\right) |g\left(\xi_i\right)| \frac{\left[\left(\widehat{P}_{-i} - P\right)[\psi_{1,r}\left(\cdot; 0\right)]\right]^2}{r^2(\inf f)^2}$$

$$+ \frac{(1 + o_{\text{p}}(1))\,\mathbf{1}_{\mathcal{E}}}{n} \sum_{i=1}^{n} K\left(\frac{Z_i}{h}\right) |g\left(\xi_i\right)| \frac{\left[\left(\widehat{P}_{-i} - P\right)[\psi_{2,r}\left(\cdot; 0\right)]\right]^2}{r^2(\inf f)^2}$$

Each term is a $V$ statistic which can be separated into a $U$ statistic and a term $V - U$ both of which have variances that evolve at maximal rate of $n^{-2}O(h/r)$. This proves (22).

## D.15 Proofs of auxiliary lemmas used for the proof of Theorem 4

### D.15.1 A useful leave-one-out result

Let $\widehat{P}_{-i}$ denote the empirical leave-$i$-out distribution, i.e.

$$\widehat{P}_{-i}[g(\xi)] = \int g(x)\widehat{P}_{-i}(\mathrm{d}x) = \frac{1}{n-1}\sum_{j=1, j\neq i}^{n} g\left(\xi_j\right).$$

Following usual empirical process notation, we denote $Pg = P[g(\xi)] = \int g(x)P(\mathrm{d}x)$.

To estimate $\mu$, we can use its empirical analogue. To that end, define

$$
\begin{aligned}
\psi_{1,r,z}\left(\xi_1\right) &:= \psi_{1,r}\left(\xi_1, z\right) = Y_1 R\left(\frac{Z_1 - z}{r}\right), \\
\psi_{2,r,z}\left(\xi_1\right) &:= \psi_{2,r}\left(\xi_1, z\right) = R\left(\frac{Z_1 - z}{r}\right).
\end{aligned}
\tag{81}
$$

Furthermore,

$$
\begin{aligned}
\text{define} \quad \overline{\psi}_{1,r}(z) &= P\left[\psi_{1,r,z}\right] = \int \psi_{1,r}\left(x_1, z\right) P\left(\mathrm{d}x_1\right) \\
\text{and note that} \quad \pi_r(z) &= P\left[\psi_{2,r,z}\right] = \int \psi_{2,r}\left(x_1, z\right) P\left(\mathrm{d}x_1\right) \\
\text{and} \quad \mu_r(z) &= \frac{\overline{\psi}_{1,r}(z)}{\pi_r(z)} \quad \text{by } \mathbf{E}_P[Y - Y(Z; D)|Z] = 0.
\end{aligned}
\tag{82}
$$

An analogue estimator for $\mu(z)$ is then

$$\widehat{\mu}_i(z) = \frac{\widehat{P}_{-i}\left[\psi_{1,r,z}\right]}{\widehat{P}_{-i}\left[\psi_{2,r,z}\right]}.
\tag{83}$$

The following lemma helps us to characterize the behavior of the difference $\widehat{\mu}_i(z) - \mu(z)$.

**Lemma 16.** *Let*

$$\varphi_{r,z}\left(\xi_1\right) := \varphi_r\left(\xi_1, z\right) := \frac{1}{\pi_r(z)} \cdot \psi_{1,r,z}\left(\xi_1\right) - \frac{\overline{\psi}_{1,r}(z)}{\pi_r(z)^2} \cdot \psi_{2,r,z}\left(\xi_1\right), \tag{84}$$

*and let* $\widehat{\mathrm{Rem}}_{i,r,z}$ *be defined by*

$$\widehat{\mathrm{Rem}}_{i,r,z} := \widehat{\mathrm{Rem}}_{i,r}(z) := \widehat{\mu}_i(z) - \mu(z) - \left(\widehat{P}_{-i} - P\right)[\varphi_r(\cdot, z)]. \tag{85}$$

*If* $\inf f := \inf_{z \in [-1,1]} f(z) > 0$, *then* $\widehat{\mathrm{Rem}}_{i,r,z}$ *satisfies the following:*

$$\mathrm{If} \quad \left|\left(\widehat{P}_{-i} - P\right)[\psi_{2,r,z}]\right| \le \frac{r}{2}\inf f,$$

$$\mathrm{then} \quad \left|\widehat{\mathrm{Rem}}_{i,r,z}\right| \le 2\left|P\psi_{1,r,z}\right| \cdot \frac{\left|\left(\widehat{P}_{-i} - P\right)[\psi_{2,r,z}]\right|^2}{r^3(\inf f)^3} \tag{86}$$

$$+ 2\frac{\left|\left(\widehat{P}_{-i} - P\right)[\psi_{1,r,z}]\right| \cdot \left|\left(\widehat{P}_{-i} - P\right)[\psi_{2,r,z}]\right|}{r^2(\inf f)^2}.$$

**Lemma 17.** *If* $\inf f := \inf_{z \in [-1,1]} f(x) > 0$, *then for every* $r \in (0,1]$,

$$\inf_{z \in [-1,1]} P\left[\psi_{2,r,z}\right] \ge r\inf f.$$

Lemma 18 below is an error bound on the Taylor approximation for reciprocals of positive numbers. We omit its straightforward proof.

**Lemma 18.** *Let* $k \in \mathbb{N}$ *and* $a_0 \ge a_{\min} > 0$. *Then for any* $a > 0$,

$$\frac{1}{a} - \frac{1}{a_0} - \frac{1}{a_0}\sum_{j=1}^{k}\left(\frac{a_0 - a}{a_0}\right)^j = \left(\frac{a_0 - a}{a_0}\right)^k\left(\frac{1}{a} - \frac{1}{a_0}\right). \tag{87}$$

*Thus for* $a > 0$ *such that* $|a - a_0| < a_{\min}$,

$$\left|\frac{1}{a} - \frac{1}{a_0} - \frac{1}{a_0}\sum_{j=1}^{k}\left(\frac{a_0 - a}{a_0}\right)^j\right| \le \frac{1}{1 - (|a - a_0|/a_{\min})} \cdot \frac{|a - a_0|^{k+1}}{a_{\min}^{k+2}}. \tag{88}$$

*Furthermore, if* $|a - a_0| \leq \frac{1}{2} a_{\min}$,

$$\left| \frac{1}{a} - \frac{1}{a_0} - \frac{1}{a_0} \sum_{j=1}^{k} \left( \frac{a_0 - a}{a_0} \right)^j \right| \leq \frac{2 |a - a_0|^{k+1}}{a_{\min}^{k+2}}. \tag{89}$$

$\varphi_r$ in (84) has the following alternative representation:

$$\varphi_r (\xi_1, z) = \left( Y_1 - \frac{\overline{\psi}_{1,r}(z)}{\pi_r(z)} \right) \frac{\psi_{2,r} (\xi_1, z)}{\pi_r(z)} = (Y_1 - \mu_r(z)) \frac{R \left( \frac{Z_1 - z}{r} \right)}{\pi_r(z)}. \tag{90}$$

Define the centered version of $\varphi_r$ by

$$\dot{\varphi}_{r,z}(\xi) := \dot{\varphi}_r(\xi, z) := \varphi_r(\xi, z) - P \left[ \varphi_r(\cdot, z) \right]. \tag{91}$$

Then

$$P \left[ \dot{\varphi}_{r,z} \right] = \mathbf{E}_{\xi \sim P} \left[ \dot{\varphi}_{r,z}(\xi) \right] = 0 \quad \text{for all } z \text{ and } r, \tag{92}$$

and $\widehat{\mathrm{Rem}}_{i,r}(z)$ in (85) also satisfies

$$\widehat{\mu}_i(z) - \mu(z) = \widehat{P}_{-i} \left[ \dot{\varphi}_r(\cdot, z) \right] + \widehat{\mathrm{Rem}}_{i,r}(z). \tag{93}$$

### D.15.2   Proof of Lemma 16

For the purpose of this proof, let $r, z$ be fixed and denote $\psi_j = \psi_{j,r,z}$ for $j \in \{1, 2\}$. From (84) and (85),

$$\widehat{\mathrm{Rem}}_{i,r,z} = \frac{\widehat{P}_{-i} \left[ \psi_1 \right]}{\widehat{P}_{-i} \left[ \psi_2 \right]} - \frac{P \left[ \psi_1 \right]}{P \left[ \psi_2 \right]} - \frac{\left( \widehat{P}_{-i} - P \right) \left[ \psi_1 \right]}{P \left[ \psi_2 \right]} + \frac{P \left[ \psi_1 \right]}{\left( P \left[ \psi_2 \right] \right)^2} \cdot \left( \widehat{P}_{-i} - P \right) \left[ \psi_2 \right]$$

This is a first order Taylor approximation remainder. For the bound in (86), combine Lemmas 17 and 18 below.

### D.15.3 Proof of Lemma 17

Assume $0 < r \leq 1$. For every $z \in [-1, 1]$,

$$
\begin{aligned}
P\left[\psi_{2,r,z}\right] &= \int R\left(\frac{u-z}{r}\right) f(u)\, \mathrm{d}u \\
&\geq (\inf f) \int_{-1}^{1} R\left(\frac{u-z}{r}\right) \mathrm{d}u \\
&= r \inf f \int_{(-1-z)/r}^{(1-z)/r} R(v)\, \mathrm{d}v.
\end{aligned}
$$

Switching variables back to $u$, we have

$$
P\left[\psi_{2,r}(\xi, z)\right] \geq r \inf f \int_{(-1-z)/r}^{(1-z)/r} R(u)\, \mathrm{d}u.
$$

By $R(u) = \mathbf{1}_{[-1,1]}(u)$,

$$
\begin{aligned}
P\left[\psi_{2,r,z}\right] &\geq r \inf f \int_{\max\{-1,(-1-z)/r\}}^{\min\{1,(1-z)/r\}} \mathrm{d}u \\
&= r \inf f \cdot \left(\min\{1, (1-z)/r\} - \max\{-1, (-1-z)/r\}\right),
\end{aligned}
$$

Start with $z \geq 0$. Then $\max\{-1, (-1-z)/r\} = -1$. Hence

$$
P\left[\psi_{2,r,z}\right] \geq r \inf f \cdot (1 + \min\{1, (1-z)/r\})
$$

$$
\geq r \inf f,
$$

since $\min\{1, (1-z)/r\} \geq 0$ for every $z \in [0, 1]$.

Next consider $-1 \leq z < 0$. Then $\min\{1, (1-z)/r\} = 1$. Hence

$$
\begin{aligned}
P\left[\psi_{2,r,z}\right] &\geq r \inf f \cdot (1 - \max\{-1, (-1-z)/r\}) \\
&= r \inf f \cdot (1 + \min\{1, (1+z)/r\}) \\
&\geq r \inf f
\end{aligned}
$$

since $\min\{1, (1+z)/r\} \geq 0$ for every $z \in [-1, 0)$.

# E    Concentration Inequalities

**Theorem 6.** *Let $\widehat{F}_n$ denote empirical measure under an iid sample of size $n$ from a probability measure $F$. Let $\mathcal{G}$ be a Vapnik-Červonenkis (VC) class of measurable functions with VC-dimension $\mathcal{V}(\mathcal{G})$ such that $\sup_{g \in \mathcal{G}} \|g\|_\infty \leq 1$. Then there exists a universal positive constant $K \in (0, \infty)$ not depending on $\mathcal{G}$, $P$ and $n$ such that*

$$
\begin{aligned}
&\Pr\left\{ \left\| \widehat{F}_n - F \right\|_{\mathcal{G}} > 8y \right\} \\
&\leq 16K \cdot \mathcal{V}(\mathcal{G})(16e)^{\mathcal{V}(\mathcal{G})} \left[ \exp\left\{ -\frac{ny^2}{128t^2} + \mathcal{V}(\mathcal{G})\log(1/y) \right\} + \exp\left\{ -nt^2 + 2\mathcal{V}(\mathcal{G})\log(1/t) \right\} \right],
\end{aligned}
\tag{94}
$$

*for every $y, t > 0$ such that*

$$
y \geq \frac{1}{\sqrt{8n}} \sup_{g \in \mathcal{G}} \sqrt{F\left[g^2\right] - (Fg)^2} \quad and \quad t \geq \sup_{g \in \mathcal{G}} \sqrt{F\left[g^2\right]}.
\tag{95}
$$

**Remark 2.** Recall that $\|\widehat{F}_n - F\|_{\mathcal{G}} = \sup_{g \in \mathcal{G}} |\widehat{F}_n[g] - F[g]|$. In cases where measurability of this quantity is of concern, (94) still holds with outer probability $\Pr^*\{\cdot\}$ in place of $\Pr\{\cdot\}$.

*Proof of Lemma 6.* Combine the covering number results for VC classes in Theorem 2.6.7 in van der Vaart and Wellner (2023, p. 206) with the proof of Theorem II.37 in Pollard (1984, pp. 34 and 35). $\qquad\square$

For our use, the following parameterized classes will be of interest:

$$
\mathcal{G}_r = \{g_{z,r} : z \in [-1, 1]\} \quad \text{where} \quad g_{z,r}(z_1) = R\left(\frac{z_1 - z}{r}\right).
\tag{96}
$$

Recall that $R(u) = \mathbf{1}\{|u| \leq 1\}$, but all subsequent arguments apply to functions of the form $R(u) = \psi(|u|)$ for $\psi(\cdot)$ non-increasing on $[0, \infty)$. We parameterize the classes by $r$ so that subsequent concentration inequalities depend on $r$. For any fixed $r > 0$, $\mathcal{G}_r$ satisfies $\mathcal{V}(\mathcal{G}_r) \leq 2$ — see Pollard (1984, Chapter II Problem 28). Now, denote

$$
\widehat{\pi}_{i,r}(z) = \frac{1}{n-1} \sum_{j=1, j \neq i}^{n} \mathbf{1}\{|Z_j - z| \leq r\} \quad \text{and} \quad \pi_r(z) = \int R\left(\frac{z_1 - z}{r}\right) f(z_1) \, \mathrm{d}z_1.
\tag{97}
$$

**Remark 3.** In (97),

1. $\widehat{\pi}_{i,r}(z)$ is the empirical probability of being $i$'s neighbor if $i$ had $z$;

2. $\pi_r(z)$ is the population probability of being neighbors with someone with running variable $z$.

**Remark 4.** Denote the probability measure with Lebesgue density $f$ by $F$. Denote the leave-$i$-out empirical measure integrating $Z_j$'s with $j \in \{1, \ldots, n\} \setminus \{i\}$ by $\widehat{F}_{-i}$, i.e.

$$\widehat{F}_{-i}[g] := \frac{1}{n-1} \sum_{j=1, i \neq j}^{n} g\left(Z_i\right).$$

Then $\widehat{\pi}_{i,r}(z)$ and $\pi_r(z)$ in (97) can be rewritten in empirical process notation via (96) as

$$\pi_{i,r}(z) = \widehat{F}_{-i}\left[g_{z,r}\right] \quad \text{and} \quad \pi_r(z) = F\left[g_{z,r}\right]. \tag{98}$$

**Lemma 19.** *Assume $\|f\|_{\infty} < \infty$. Take $C \geq \|f\|_{\infty} \int R(u)^2 \, \mathrm{d}u$, $C_* = 128C$ and $y > 0$ such that $y \geq \sqrt{\frac{Cr}{8n}}$. Then there is a universal constant $A$ not depending on $r$ or $n$ such that*

$$\begin{aligned}
&\max_{i=1,\ldots,n} \Pr\left\{ \sup_{z \in [-1,1]} |\widehat{\pi}_{i,r}(z) - \pi_r(z)| > 8y \right\} \\
&\leq A\left[ \exp\left\{ -\frac{(n-1)y^2}{C_* r} + 2\log(1/y) \right\} + \exp\left\{ -C(n-1)r + 2\log(1/(Cr)) \right\} \right].
\end{aligned} \tag{99}$$

*and therefore,*

$$\begin{aligned}
&\Pr\left\{ \max_{i=1,\ldots,n} \sup_{z \in [-1,1]} |\widehat{\pi}_{i,r}(z) - \pi_r(z)| > 8y \right\} \\
&\leq A\exp\left\{ -\frac{(n-1)y^2}{C_* r} + 2\log(1/y) + \log n \right\} \\
&\quad + A\exp\left\{ -C(n-1)r + 2\log(1/(Cr)) + \log n \right\}.
\end{aligned} \tag{100}$$

*For sequences $r = r_n$ such that $r \downarrow 0$ and $\log n/(nr) \downarrow 0$, the first exponential term (dependent on $y$) dominates in both (99) and (100).*

*Proof of Lemma 19.* It is straightforward to show that

$$\sup_{g \in \mathcal{G}} \left( F\left[g^2\right] - (Fg)^2 \right) \le \sup_{g \in \mathcal{G}} F\left[g^2\right] \le Cr. \tag{101}$$

Following the notation conventions in (98), and using (94),

$$\Pr\left\{ \left\| \widehat{F}_{-i} - F \right\|_{\mathcal{G}_r} > 8y \right\}$$
$$\le 32K(16e)^2 \left[ \exp\left\{ -\frac{(n-1)y^2}{128t^2} + 2\log(1/y) \right\} + \exp\left\{ -(n-1)t^2 + 4\log(1/t) \right\} \right],$$

for every $y, t > 0$ such that

$$y \ge \sqrt{\frac{Cr}{8n}} \quad \text{and} \quad t \ge \sqrt{Cr},$$

by (95) and (101). Since this holds for every $i \in \{1, \dots, n\}$, set $A = 32K(16e)^2$, $t = \sqrt{Cr}$, $C_* = 128C$ to get

$$\max_{i=1,\dots,n} \Pr\left\{ \left\| \widehat{F}_{-i} - F \right\|_{\mathcal{G}_r} > 8y \right\}$$
$$\le A \left[ \exp\left\{ -\frac{(n-1)y^2}{C_* r} + 2\log(1/y) \right\} + \exp\left\{ -C(n-1)r + 2\log(1/(Cr)) \right\} \right],$$

which is exactly (99).

Then, (100) follows from $\Pr\left\{ \max_{j=1,\dots,N} V_j > \varepsilon \right\} \le N \max_{j=1,\dots,N} \Pr\left\{ V_j > \varepsilon \right\}$ for arbitrary random variables $V_1, \dots, V_N$. The claim about the dominant first term stems from the fact that for $r_n \downarrow 0$,

$$\frac{\log n}{n r_n} \to 0 \quad \text{if and only if} \quad \frac{\log(1/r_n)}{n r_n} \to 0.$$

Therefore for the second term independent of $y$ for example in (100)

$$\exp\left\{ -C(n-1)r + 2\log(1/(Cr)) + \log n \right\}$$
$$= \exp\left\{ -C(n-1)r \left( 1 + \frac{2\log(1/(Cr))}{n} + \frac{\log n}{nr} \right) \right\} = \exp\left\{ -C(n-1)r\,(1 + o(1)) \right\},$$

and $\log n/(nr) \downarrow 0$ implies $nr \uparrow \infty$. This term tends to zero regardless of the chosen $y > 0$

with $y \geq \sqrt{(Cr)/(8n)}$. $\qquad\square$

**Lemma 20.** *Assume $\|f\|_\infty < \infty$. Let $r_n$ be a sequence such that $r_n \downarrow 0$ and $\log n / (nr_n) \downarrow 0$. Then for any $\varepsilon > 0$,*

$$\Pr\left\{\max_{i=1,\dots,n} \sup_{z\in[-1,1]} |\widehat{\pi}_{i,r_n}(z) - \pi_{r_n}(z)| > r\varepsilon\right\} \to 0, \quad as\ n \to \infty.$$

*Therefore, $\max_{i=1,\dots,n} \sup_{z\in[-1,1]} |\widehat{\pi}_{i,r_n}(z) - \pi_{r_n}(z)| = o_p(r_n)$.*

*Proof of Lemma 20.* Write $r = r_n$ and let $A$, $C$ and $C_*$ be as defined in Lemma 19. Set $y = (\varepsilon r)/8$. Then $y \geq \sqrt{\frac{Cr}{8n}}$ for $n$ sufficiently large. To see this note that $y \geq \sqrt{\frac{Cr}{8n}}$ iff $\frac{\varepsilon}{\sqrt{8C}} \geq \sqrt{\frac{1}{nr}}$. By $\log n/(nr) \to 0$, $1/(nr) \to 0$ and so, $y \geq \sqrt{\frac{Cr}{8n}}$ for $n$ sufficiently large. By Lemma 19,

$$\Pr\left\{\max_{i=1,\dots,n} \sup_{z\in[-1,1]} |\widehat{\pi}_{i,r_n}(z) - \pi_{r_n}(z)| > r\varepsilon\right\}$$

$$= \Pr\left\{\max_{i=1,\dots,n} \sup_{z\in[-1,1]} |\widehat{\pi}_{i,r_n}(z) - \pi_{r_n}(z)| > 8y\right\}$$

$$\leq A\exp\left\{-\frac{(n-1)y^2}{C_*r} + 2\log(1/y) + \log n\right\}$$

$$+ A\exp\left\{-C(n-1)r + 2\log(1/(Cr)) + \log n\right\}$$

$$= A\exp\left\{-\frac{(n-1)\varepsilon r}{64C_*} + 2\log(8/(\varepsilon r)) + \log n\right\}$$

$$+ A\exp\left\{-C(n-1)r + 2\log(1/(Cr)) + \log n\right\}.$$

Both terms in the final two lines above tend to zero by $\log n / (nr_n) \to 0$ (which is equivalent to $\log(1/r_n)/(nr_n) \to 0$ when $r_n \to 0$). $\qquad\square$

**Lemma 21.** *Suppose $h_n \gg n^{-1/3}$. Then*

$$\max_{i:Z_i\in[-h_n,h_n]} |\tilde{\mu}(Z_i) - \mu(Z_i)| = o(h_n) \quad w.p.a.\ 1\ .$$

*Proof.* Let $\varepsilon = \frac{f(0)h_n}{2r_n}$. Let $\Gamma$ be the event that all $Z_i$ have between $f(0) \cdot nh_n/2$ and $f(0) \cdot 3nh_n/2$ neighbors. Then by Lemma 20, $\mathbb{P}(\Gamma) \to 1$. Let $K$ be the upper bound of

$Y_d(z)$. That is, $Y_d(z) \leq K$ for all $z \in [-1, 1]$. Applying Hoeffding's Inequality to the conditional distribution of $Y_j \mid Z_j \in R(Z_i)$ and on the event $\Gamma$, we have:

$$\mathbb{P}\left(\left|\frac{1}{(n-1)\hat{\pi}_{i,r}} \sum_{j=1, j\neq i}^{n} Y_j \mathbf{1}\{|Z_j - z| \leq r_n\} - \mu(Z_i)\right| \geq (nh_n)^{-1/2+\eta}\right)$$

$$\leq \exp\left(-\frac{(nh_n)^{1+2\eta}}{(n-1)\hat{\pi}_{i,r}K^2}\right)$$

$$\leq \exp\left(-\frac{2(nh_n)^{1+2\eta}}{(n-1)h_n K^2}\right) \quad \text{on } \Gamma \ .$$

Now, w.p.a. 1, there are no more than $3nh_n/2$ observations on $[-h_n, h_n]$. On this event and $\Gamma$, the union bound gives us that:

$$\mathbb{P}\left(\max_{i:Z_i \in [-h_n, h_n]} |\tilde{\mu}(Z_i) - \mu(Z_i)| \geq (nh_n)^{-1/2+\eta}\right) \leq \frac{3f(0)}{2} \cdot nh_n \exp\left(-\frac{2}{K^2}(nh_n)^{2\eta}\right) \to 0$$

Since $h_n \gg n^{-1/3}$, choose $\eta$ such that $(nh_n)^{-1/2+\eta} \ll h_n$ and we are done. $\qquad \square$