# Monotone Regression Discontinuity

## Partial Identification, Estimation, and Extrapolation*

Matthew Estes†

August 15, 2026

**Click here for current version of paper**

**Abstract**

This paper develops a partial identification method to bound treatment effects in regression discontinuity designs. Unlike the standard regression discontinuity approach, the partial identification results here rely on monotonicity and boundedness assumptions rather than a continuity assumption. This paper develops two novel estimation strategies to bound the local average treatment effect in such designs where the data generating processes are monotone near the running variable cutoff. In simulation studies, these estimators perform well with good coverage properties and are robust to bandwidth selection. But there is an efficiency cost, particularly in missing data or measurement error settings. Next, the paper extends the monotone partial identification approach to fuzzy designs and to extrapolate bounds on the average treatment effect away from the cutoff. Finally, I apply Monotone RDD to a prior regression discontinuity study, investigating the effect of energy efficiency ratings on property prices (Sejas-Portillo, Moro, and Stowasser, 2025).

**Keywords:** regression discontinuity, partial identification, bounds, extrapolation, monotonicity

*I am grateful to Mike Alvarez, Ransi Clark, Jim Greiner, Jonathan Katz, Bob Sherman, and other participants at Caltech seminars for helpful feedback on this work.

†California Institute of Technology. Email: mestes@caltech.edu

# 1 Introduction

Regression discontinuity is one of the most popular observational causal inference designs in applied economics. Goldsmith-Pinkham (2026) estimates that more than 10% of all applied microeconomics NBER papers employ a regression discontinuity design ("RDD"). These studies generally rely on one of two untestable assumptions to identify an average treatment at the boundary or cutoff: a continuity assumption or a local randomization assumption. Both assumptions may fail to hold in a variety of settings. For example, in the most common "sharp" RDD, treatment is determined entirely by a forcing or running variable, but if agents manipulate the forcing variable this can violate the continuity assumption (McCrary, 2008). Consequently, the RDD will not in general identify a causal effect. Placebo tests identifying discontinuities at non-treatment running variable values are also said to invalidate the design (Imbens and Lemieux, 2008).

This paper considers an alternative identifying assumption compatible with non-standard data generating processes (DGPs). The paper shows that in a setting where the average potential outcome functions are monotone in the forcing variable and outcomes are bounded, the cutoff or boundary average treatment effect is partially identified. Sharp bounds on the cutoff average treatment effect are derived. The bounds are compatible with jumps or discontinuities in the potential outcome conditional expectation functions (CEFs) at the cutoff that determines treatment. Discontinuous but monotone data generating processes (e.g. discrete monotone processes) generally may have partially identified cutoff treatment effects provided potential outcomes or averages of potential outcomes are also bounded locally.

Next, the paper proposes two novel estimators for the monotone setting. Using the monotonicity of the CEFs, the paper defines *local extremum* and *local reverse kernel* estimators. The estimators are designed for robustness to potentially confounding features of data generating processes and researcher free parameters (bandwidth selection). For example, the estimators can be used in RDD studies with: discontinuous DGPs, trend reversals, discrete running variables, measurement error, or missing data near the cutoff (e.g. "donut RDD"). Both estimators are also bandwidth-robust: they cover the true treatment effect bounds at a variety of selected bandwidths. Instead, bandwidth selection, which is the most important free-parameter in RDD estimation, becomes an issue of efficiency: certain bandwidths may achieve tighter treatment effect bounds or narrower partially-identified regions.

The paper illustrates the performance of the two estimators in comparison to the local linear estimator—the most popular RDD estimation strategy—via simulations. The local extremum and local reverse kernel estimators have good coverage properties in finite samples across a range of data generating processes. The local linear estimator often performs less well at estimating bounds, but may still cover the true treatment effect when there is a discontinuity at the cutoff in certain cases.

The identification power of monotonicity in other designs targeting other estimands is then demonstrated. First, results are derived to show that monotonicity and outcome boundedness partially identify extrapolated bounds away from the boundary or cutoff: assuming there is some window around the cutoff determining treatment where both CEFs are monotone, the average treatment effect (ATE) can be bounded in that entire window. Such bounds may be of more policy relevance than point identification of *merely local* cutoff average treatment effects. Second, the partial identification in fuzzy RDD settings is discussed, under the additional "monotone treatment response" assumption (Manski, 2007). The fuzziness of the assignment can tighten the bounds away from the boundary, which is demonstrated via an example.

Finally, the method is applied to the RDD study from Sejas-Portillo, Moro, and Stowasser (2025). Using discrete energy efficiency ratings, UK properties receive a simplified, low-information energy band rating (A to G). Prices are generally monotone increasing in the running variable, and Sejas-Portillo, Moro, and Stowasser (2025) document price discontinuities at the energy band cutoffs. The application here reproduces those estimates and applies the estimators to obtain bounds on cutoff effects and placebo effects at other values of the running variable. The results here show that the treatment effect estimates at the cutoffs hold, even under the different counterfactual assumption, alternative estimators, and different bandwidths. Placebo tests further suggest that treatment effects at the energy band cutoffs are large under various estimators.

### 1.1 Prior Literature

I contribute to different strands of the RDD literature. First, I contribute to the literature concerned with continuity failures near the boundary. For example, manipulation of the forcing variable can invalidate the continuity assumption (McCrary, 2008). Several ad hoc strategies—e.g. the so-called "donut hole" RDD (Noack and Rothe, 2023)—have been suggested, but monotone RDD may provide bounds that are informative in such settings. Second, I contribute to a growing literature concerned with continuity-type assumptions. Korting et al. (2023) show that evaluating

CEFs for discontinuities can be difficult, even for expert evaluators. By contrast, the monotone RDD method advanced here relies on a monotonicity assumption, whose plausibility may be simpler for researchers to recognize and evaluate. In particular, the plausibility of the monotonicity assumption may be tested on the observable data. Additional work on estimation and inference in monotone regression discontinuity settings has also been performed for various estimators, including isotonic regression (Babii and Kumar, 2023). Third, this work similarly contributes to studying how certain estimators and specifications perform in various settings (Gelman and Imbens, 2018), with emphasize on bandwidth selection, discrete data generating processes, and other applied issues (e.g. missing data, running variable measurement error). Finally, I contribute to the vast literature on partial identification. For example, I extend Manski (2007) type monotone bounds to the RDD setting.

The rest of the paper is organized as follows. Section 2 reviews continuity-based point identification, states the main partial identification result, and discusses estimation strategies. Section 3 surveys simulation study results and evaluates small, finite sample coverage properties for the three types of estimators. Section 4 shows how to bound the ATE under global monotonicity and gives identification results in the fuzzy regression discontinuity design. Section 5 is the application to Sejas-Portillo, Moro, and Stowasser (2025). Further details and results can be found in the Appendices.

## 2 RDD Identification: Local Treatment Effect

This section explains a new regression discontinuity methodology, which exploits variation stemming from the discontinuous treatment assignment to partially identify the local (or cutoff) treatment effect in a sharp regression discontinuity design. I first consider the typical sharp RDD identification strategy and contrast it to our partial identification method. In the method, I give upper and lower bounds on the local treatment effect for a data generating process (DGP) that is discontinuous but monotone at the treatment cutoff. Finally, I discuss estimation and inference strategies, including model specification and bandwidth selection.

### 2.1 Standard Point Identification: Identifying $\tau_C$ Under Continuity in the Forcing Variable

To begin, I assume a potential outcomes framework following the setup in Wooldridge (2010), Imbens and Lemieux (2008), and Cattaneo, Idrobo, and Titiunik (2020). Let

$Y_{0i}$ and $Y_{1i}$ be the counterfactual outcomes for an individual $i$ without and with treatment, respectively. The binary treatment $W_i$ is determined from the "forcing variable" $X_i$ and a known cutoff or threshold $C$ by $W_i = \mathbf{1}\{X_i \geq C\}$. The observed outcome is given by the switching equation $Y_i = (1 - W_i)Y_{0i} + W_i Y_{1i}$. Define further the counterfactual conditional expectation functions (CEFs):

$$\mu_g(x) = \mathbb{E}[Y_g|X = x]$$

for $g = 0, 1$. In the usual RDD setting, the potential outcome CEFs are assumed to be continuous in $x$. Moreover, because $W_i$ is a deterministic function of $X_i$ note that ignorability (conditional mean independence) holds:

$$\mathbb{E}[Y_g|X, W] = \mathbb{E}[Y_g|X]$$

for $g = 0, 1$.

The usual estimand of interest in a sharp RDD is the local average treatment effect (or cutoff treatment effect), defined as:

$$\tau_C = \mathbb{E}[Y_1 - Y_0|X = C]$$

Following Wooldridge (2010), I first relate the potential outcome CEFs to the observed CEF as follows. By continuity of $\mu_g(\cdot)$, ignorability, and the treatment assignment rule, I find that:

$$\underbrace{\mathbb{E}[Y_0|X = C]}_{\equiv \mu_0(C)} = \lim_{x \to C^-} \mathbb{E}[Y_0|X = x] = \lim_{x \to C^-} \mathbb{E}[Y_0|W = 0, X = x] = \lim_{x \to C^-} \mathbb{E}[Y|X = x]$$

By a similar argument, identify $\mu_1(C)$ with the limit from the right:

$$\mu_1(C) = \lim_{x \to C^+} \mathbb{E}[Y|X = x]$$

It follows that the local average treatment effect $\tau_C$ is identified by a difference in one-sided limits of the observed CEF:

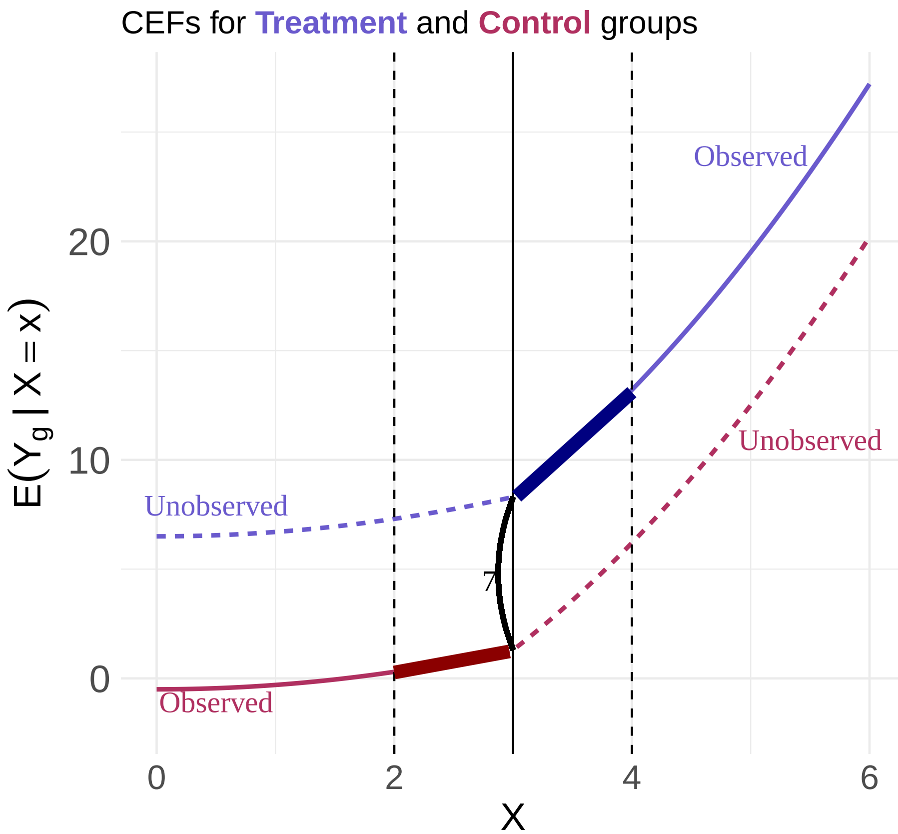
$$\tau_C = \mu_1(C) - \mu_0(C) = \lim_{x \to C^+} \mathbb{E}[Y|X = x] - \lim_{x \to C^-} \mathbb{E}[Y|X = x]$$

In general, note that $\mathbb{E}[Y|X = x, W = 0]$ is estimable from the observable data for all values $x < C$ (and similarly for $W = 1$ with $x \geq C$). Therefore, we can estimate $\tau_C$ by estimating their limits. The most common strategy is local linear regression:[1] for observations within a small bandwidth $h > 0$, regress $Y_i$ on $1, (X_i - C)$ for two subsets of data, one with $X_i \in (C - h, C)$ and the other with $X_i \in [C, C + h)$. The difference in the estimated intercepts gives $\hat{\tau}_C$.

An example of a sharp regression discontinuity is given below in Figure 1. The dashed lines are unobservable in the sharp design, whereas the solid lines are observable.[2] In the Figure 1 example, there is a constant treatment effect of 7 across the support of $X$, which also means that $\tau_C = ATE$.

[1] Local polynomial regression can also be used to estimate the limit of the conditional means. See Wooldridge (2010) or Imbens and Lemieux (2008) for further details, including information on bandwidth selection. In general, to estimate the local treatment effect we do not use global polynomial fits to estimate the limits of the CEFs following results in Gelman and Imbens (2018). In Gelman and Imbens (2018), the authors find *inter alia* that (global) higher order polynomial estimation attaches noisy weights to observations and may not achieve nominal coverage rates.

[2] That is, the solid lines are, in principle, observable by virtue of the sharp treatment assignment. In practice, of course, the solid lines can only be (consistently) estimated from the data.




The thick dark red and navy blue lines are linear fits using data with x−values in [2,3) and [3,4], respectively. The true treatment effect at the boundary is 7.0, which is also the ATE.

**Figure 1.** Sharp RDD: Continuous CEFs with Constant Treatment Effect

### 2.2 Partial Identification: Bounding $\tau_C$ Under Monotonicity in the Forcing Variable and Bounded Outcomes

The example in Figure 1—which assumes continuity and a constant treatment effect across the entire support of forcing variable $X$—may not hold in practice. For example, manipulation of the forcing variable $X$ can invalidate the crucial CEF continuity assumption (McCrary, 2008). Here, I detail simple assumptions to obtain bounds on the cutoff treatment effect while allowing for discontinuities in the CEFs. Specifically, I assume monotonicity in the forcing variable and boundedness in the outcome-of-interest to partially identify the local treatment effect or the cutoff effect $\tau_C$.

Again assume the standard (sharp) RDD setting, with treatment $W_i = \mathbf{1}\{X_i \geq C\} \in \{0, 1\}$ as above, for forcing variable $X_i$ and cutoff $C$. Recall the CEFs were written as:

$$\mu_1(x) = \mathbb{E}[Y_1|X = x]$$
$$\mu_0(x) = \mathbb{E}[Y_0|X = x]$$

Consider the following monotonicity[3] and boundedness[4] assumption on the CEFs:

**Assumption 1** (Monotonicity)**.** *The conditional expectation functions $\mu_1, \mu_0$ are weakly monotone increasing in $x$ for all $x \in supp(X)$.*

**Assumption 2** (Boundedness)**.** *The conditional expectation functions $\mu_1, \mu_0$ are bounded above by $M_1$ and bounded below by $M_0$ for all $x \in supp(X)$.*

Some initial remarks on these assumptions and what is and is not assumed. First, the continuity of $\mu_1(x), \mu_0(x)$ in $x$ is not assumed. Second, global monotonicity (Assumption 1) is strictly stronger than required for bounding $\tau_C$: it is assume for exposition, but it is sufficient for bounding $\tau_C$ that there be a small bandwidth $h > 0$ s.t. in the interval $(C - h, C + h)$ around the cutoff $C$ the CEFs $\mu_1, \mu_0$ are monotone. Third, it is assumed for exposition that $\mu_1(x), \mu_0(x)$ are weakly monotone increasing, but it is not required that $\mu_1, \mu_0$ are monotone *increasing*, or even that they are *both increasing* or *both decreasing*.[5] Fourth, in some settings, the upper bound on $\mu_0(C)$ may be given by $\mu_1(C)$, which corresponds to Manski's "monotone treatment response" (Manski, 2007).[6] Fifth, observe that ignorability (conditional mean independence) still holds because $W_i = \mathbf{1}\{X_i \geq C\}$ is a deterministic function of $X_i$, so that for $g = 0, 1$ we have $\mathbb{E}[Y_g|X, W] = \mathbb{E}[Y_g|X]$.

Under the assumptions, we can partially identify the usual estimand in an RD setting, the average treatment effect at the cutoff:

$$\tau_C = \mathbb{E}[Y_1 - Y_0|X = C] = \mu_1(C) - \mu_0(C)$$

[3] Technically, we only require monotonicity on the unobservable portion of the CEF. However, it would be odd to make an assumption on only the unobservable half of a CEF in the RDD design if, on the observable portion of the CEF, monotonicity does not hold near the cutoff. See discussion below about motivating this assumption in context and using the observable portion of the CEFs to test for monotonicity.

[4] Note that boundedness is only necessary to establish a lower bound in what follows. As with monotonicity, we also only need some form of local boundedness for the unobservable half of the CEFs in the sharp design. We assume boundedness over the entire support of the forcing variable $X$ in the exposition though.

[5] That is, $\mu_1$ can be monotone increasing, and $\mu_0$ can be monotone decreasing.

[6] Otherwise, Assumption 2 may be particularly relevant when the outcomes of interest are rates that are bounded above by 1 and below by 0, or other outcomes that are guaranteed to be bounded above and below.

To discuss identification, it will be useful to write the CEFs as separate functions to the left and right of the cutoff $C$:

$$\mu_1(x) = \mathbf{1}\{x < C\}\mu_1^L(x) + \mathbf{1}\{x \geq C\}\mu_1^R(x)$$
$$\mu_0(x) = \mathbf{1}\{x < C\}\mu_0^L(x) + \mathbf{1}\{x \geq C\}\mu_0^R(x)$$

Discontinuity at the cutoff $C$ in this notation means, for example, that $\mu_0^L(C) \neq \mu_0^R(C)$. The identification problem for $\tau_C$ is that $\mu_0(C) = \mu_0^R(C)$ is unobservable by the assignment rule ($W_i = \mathbf{1}\{X_i \geq C\}$).

Although we cannot point identify the average treatment effect at the cutoff, we can partially identify it. The (sharp) bounds are given by Theorem 1:

**Theorem 1** (Local Treatment Effect ($\tau_C$) Bounds)**.** *Let $\mu_1, \mu_0$ be the potential outcome CEFs defined as above, satisfying Assumption 1 and Assumption 2. Then bounds for the local treatment effect $\tau_C$ are $\tau_C \in [\underline{\tau_C}, \overline{\tau_C}]$, where:*

$$\overline{\tau_C} = \mu_1^R(C) - \lim_{x \to C^-} \mu_0^L(x)$$
$$\underline{\tau_C} = \mu_1^R(C) - M_1$$

*Proof.* By the sharp RDD treatment assignment rule, $\mu_1(C) = \mu_1^R(C)$ is observable but the identification problem is that $\mu_0(C) = \mu_0^R(C)$ is unobservable.[7] However, by monotonic increasing assumption, we have that $\mu_0(x) \leq \mu_0(C)$ for all $x < C$, which implies that

$$\mu_0(C) \geq \sup_{x<C} \mu_0(x) = \sup_{x<C} \mu_0^L(x) = \lim_{x \to C^-} \mu_0^L(x)$$

where the one-sided (left) limit is equal to the supremum, guaranteed to exist by monotonicity, and finite by boundedness assumption.[8] The lower bound requires

[7]Although $\mu_1(C) = \mu_1^R(C)$ can also bounded from above by $\lim_{x \to C^+} \mu_1^R(x)$, it is observable under the assignment rule and so does not pose an identification problem. That is, the identification and estimation tasks are kept conceptually distinct. Below, we generally use the estimator of $\mu_1^R(C)$ from local linear regression, even though this is in principle observable—i.e. can be consistently estimated by the sample mean for units with $X_i = C$—under the treatment assignment rule. This is consistent with the standard continuity-based RDD practice. For identification, we technically only need the inequalities on $\mathbb{E}[Y_0|X = C]$.

[8]Suppose $f(x)$ is a monotone increasing function on interval $(0, C)$. If it is bounded above on domain—as we've assumed—then:

$$\lim_{x \to C^-} f(x) = \sup_{x \in (0,C)} f(x)$$

More generally, bounded monotone functions always have well-defined and finite one-sided limits on

Assumption 2: to identify the lower bound $\underline{\tau_C}$, use $\mu_0(x)$ bounded above on supp$(X)$ by $M_1$ to obtain

$$\mu_0(C) \leq M_1$$

These inequalities together imply that the local treatment effect $\tau_C = \mu_1(C) - \mu_0(C)$ is bounded by:

$$\underbrace{\mu_1^R(C) - M_1}_{\underline{\tau_C}} \leq \tau_C \leq \underbrace{\mu_1^R(C) - \lim_{x \to C^-} \mu_0^L(x)}_{\overline{\tau_C}}$$

The bounds are sharp because, under the assumptions, $\mu_0(C) = \mu_0^R(C)$ can take any value in $[\lim_{x \to C^-} \mu_0(x), M_1]$ and remain compatible with the observed data. $\square$

The plots in Figure 2 give an example DGP compatible with our assumptions. In the left plot, the true CEFs for the treatment and control groups are plotted. The solid lines are observable (by which we mean we can estimate them from data), whereas the dashed lines are unobservable because of the sharp treatment assignment. The typical RDD assumption that both CEFs $(\mu_1(x), \mu_0(x))$ are continuous in $x$ **does not hold.** The functions for the potential outcome CEFs are given by:

$$\mu_1(x) = \mathbf{1}\{x < 3\}(0.2x^2 + 2) + \mathbf{1}\{x \geq 3\}(0.7x^2 + 2)$$
$$\mu_0(x) = \mathbf{1}\{x < 3\}(0.2x^2 - 2) + \mathbf{1}\{x \geq 3\}(0.7x^2 - 5)$$

their domain.

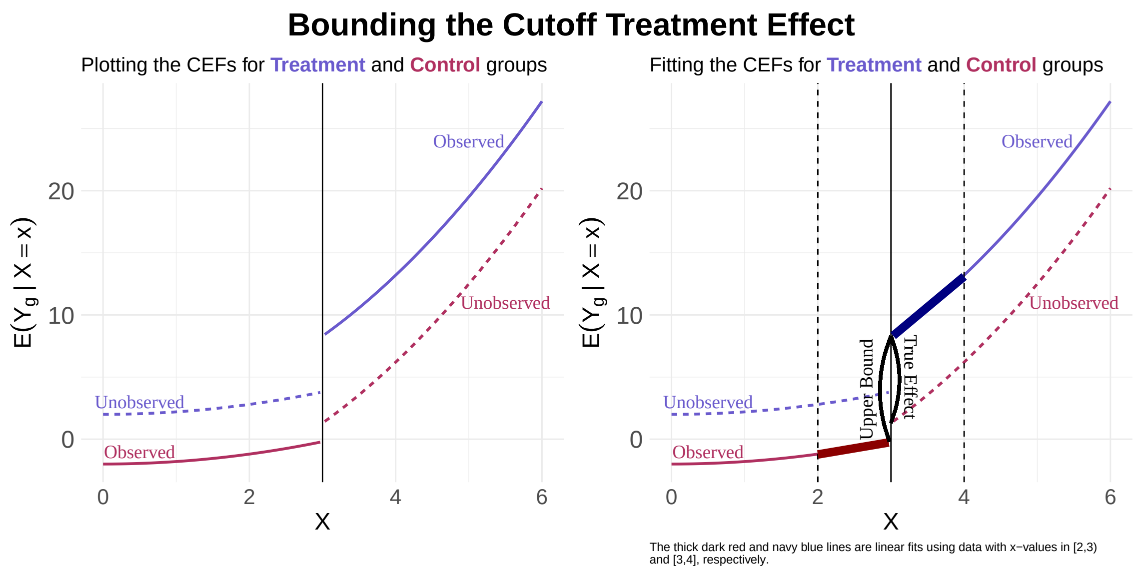



**Figure 2.** Bounding the Treatment Effect at the Cutoff $C = 3$

In the right panel of Figure 2, we show how the usual estimation procedure—obtain linear fit in a small bandwidth around the cutoff $C = 3$, e.g.— estimates the upper bound $\overline{\tau_C}$ rather than $\tau_C$. For this particular example, the true effect at the cutoff $C = 3$ is given by the difference between the solid blue line (observable $\mu_1$) and the dashed red line (unobservable $\mu_0$):

$$\tau_C = \mu_1(3) - \mu_0(3) = (0.7(3)^2 + 2) - (0.7(3)^2 - 5) = \boxed{7}$$

But the upper bound $\overline{\tau}_C$ at the cutoff is given by taking the difference between $\mu_1(3)$ and the lower bound for $\mu_0(3)$. In Figure 2 above, this results in the upper bound:

$$\overline{\tau_C} = (0.7(3)^2 + 2) - (0.2(3)^2 - 2) = 8.3 - (-0.2) = \boxed{8.5}$$

found from the known CEFs above.[9]

[9]In practice, we obtain estimates on $\hat{\mu_1}(3)$ and $\hat{\mu_0}(3)$ from the local linear regression in a small window around the cutoff, as in the rightmost plot. For the lower bound, $M_1 = 27.2$ bounds both $\mu_0, \mu_1$ from above on $[0, 6]$, so that a lower bound on $\tau_C$ is:

$$\underline{\tau_C} = \mu_1(C) - M_1 = (0.7(3)^2 + 2) - (0.7(6)^2 + 2) = 8.3 - 27.2 = \boxed{-18.9}$$

This bound is sharp if the researcher does not assume a stronger form of boundedness. Note that if we instead assume $\mu_0(C)$ is bounded above by $\mu_1(C)$ (e.g. “monotone treatment response”), then $M_1 = \mu_1(C)$ and the lower bound is zero. See Manski (2007).

This example shows that for the class of DGPs considered here—monotone potential outcome CEFs with discontinuities at the cutoff—the standard RDD estimation is biased for $\tau_C$: in this example, the standard estimation procedure for a sharp RDD will produce an estimated $\hat{\tau}_C$ that is 20% larger than the true effect although, in DGPs like the example, local linear regression still consistently estimates *the bounds* $(\overline{\tau_C}, \underline{\tau_C})$.[10]

### 2.3 Estimation & Inference

Monotonicity on the entire support of the forcing variable is a strong shape restriction. However, as discussed above, we only require a small bandwidth $h > 0$ s.t. $\mu_1, \mu_0$ are monotone in $(C - h, C + h)$. Although we leave the optimal estimation and inference procedures for future work, I study here two novel estimation and inference procedures: local extremum and local reverse kernel estimators. In Section 3, I investigate the finite-sample performance of these estimators in comparison to the widely used local linear estimator via simulations.

Recall that our estimands are given by:

$$\overline{\tau_C} = \mu_1^R(C) - \lim_{x \to C^-} \mu_0^L(x)$$
$$\underline{\tau_C} = \mu_1^R(C) - M_1$$

Assuming that $M_1$ is known to the researcher, the estimation task is twofold: we must estimate $\mu_1^R(C)$ and $\lim_{x \to C^-} \mu_0^L(x)$. If there are sufficiently many treated units with $X_i = C$, then $\mu_1^R(C)$ can be estimated with the sample average for units with $X_i = C$. Assuming this is not the case, we consider three ways to estimate these quantities: (1) local linear estimators, (2) local extremum estimators, and (3) reverse kernel estimators.

#### 2.3.1 Local Linear Estimation

The local linear estimator is well-known and oft-applied in regression discontinuity designs. The estimation procedure is straightforward: obtain linear fits within a

[10]For example, local linear regression in a window around the cutoff will consistently estimate $\overline{\tau_C}$ when the potential outcome CEFs are monotone increasing with only one discontinuity at the cutoff and are otherwise smooth. For more general monotone DGPs with many discontinuities, a modified estimation procedure may be necessary and bandwidth selection will be much more important.

bandwidth $h > 0$ on either side of the cutoff $C$:

$$\min_{\alpha_L,\beta_L} \sum_{i:C-h<X_i<C} (Y_i - \alpha_L - \beta_L(X_i - C))^2 \quad \text{and} \quad \min_{\alpha_R,\beta_R} \sum_{i:C\leq X_i<C+h} (Y_i - \alpha_R - \beta_R(X_i - C))^2$$

Then estimate $\mu_1^R(C)$ and $\lim_{x\to C^-} \mu_0^L(x)$ as follows:

$$\widehat{\mu_1^R(C)} = \hat{\alpha_R}$$
$$\lim_{x\to C^-} \widehat{\mu_0^L}(C) = \hat{\alpha_L}$$

which implies that estimated upper and lower bounds are given by:

$$\widehat{\overline{\tau_C}} = \hat{\alpha_R} - \hat{\alpha_L}$$
$$\widehat{\underline{\tau_C}} = \hat{\alpha_R} - M_1$$

As noted in Imbens and Lemieux (2008), standard least squares methods of inference may be used. When the bandwidth $h$ goes to zero at the appropriate rate as sample sizes increase, the estimates are asymptotically normal so that EHW standard errors may be used. As shown in Kolesar and Rothe (2018), the EHW standard errors have good coverage properties even in small samples where the running variable is discrete; by contrast, CIs based on clustering by the running variable ("CRV") may fail to achieve 95 percent coverage.

### 2.3.2 Local Extremum Estimation

Although local linear regression is the most commonly used RDD estimation strategy, for discontinuous but monotone DGPs—which are allowed by our identification strategy—local linear regression may fail to provide good estimates of the upper and lower bound. Although it is possible higher order polynomials may work, in practice the introduction of higher order polynomials may introduce specification bias.

To allow for discontinuous processes and alleviate specification bias, we consider a local extremum estimator of the estimands of interest. Specifically, our estimator is motivated by rewriting the left limit of $\mu_0^L(x)$ as follows:

$$\lim_{x\to C^-} \mu_0^L(x) = \lim_{h\to 0^+} \left( \inf_{0<C-x<h} \mu_0^L(x) \right)$$

which suggests, e.g., taking a local minimum (when $\mu_0$ is monotone increasing) over a shrinking window $(C-h, C)$ where $h$ is a bandwidth. In particular, for $\mu_0$ monotone increasing and bandwidth $h > 0$ we will have that:

$$\min_{x \in (C-h,C)} \mu_0(x) \leq \lim_{x \to C^-} \mu_0^L(x)$$

Accordingly, for $\mu_0, \mu_1$ both monotone increasing we can estimate, for bandwidth $h > 0$, the two quantities of interest as:

$$\widehat{\mu_1^R(C)} = \max_{x \in [C,C+h)} \frac{1}{n_{x,1}} \sum_{i:X_i=x,W_i=1} Y_i$$
$$\widehat{\lim_{x \to C^-} \mu_0^L(C)} = \min_{x \in (C-h,C)} \frac{1}{n_{x,0}} \sum_{i:X_i=x,W_i=0} Y_i$$

where $n_{x,w}$ is the number of units with $X_i = x$ and $W_i = w$. For inference, we use a simple bootstrap procedure to obtain standard errors and construct confidence intervals, which performs well in our Section 3 simulation study.

The local extremum estimator is illustrated below in Figure 3. In the left panel, the solid red line is the true control outcome CEF for DGP1 (discussed below) and each point is the population mean $\mathbb{E}[Y|X = x]$: the vertical gray and brown lines show the estimated minimums with bandwidths of $h = 1.5$ and $h = 1.0$, resp. In the right panel, we show the estimated minimum $\widehat{\lim_{x \to C^-} \mu_0^L(C)} = \min_{x \in (C-h,C)} \mathbb{E}[Y_0|X = x]$ varying $h$ from 3 to 0: as the bandwidth $h$ goes to 0, the local extremum estimator $\widehat{\lim_{x \to C^-} \mu_0^L(C)}$ approaches the true value $\lim_{x \to C^-} \mu_0^L(C)$.

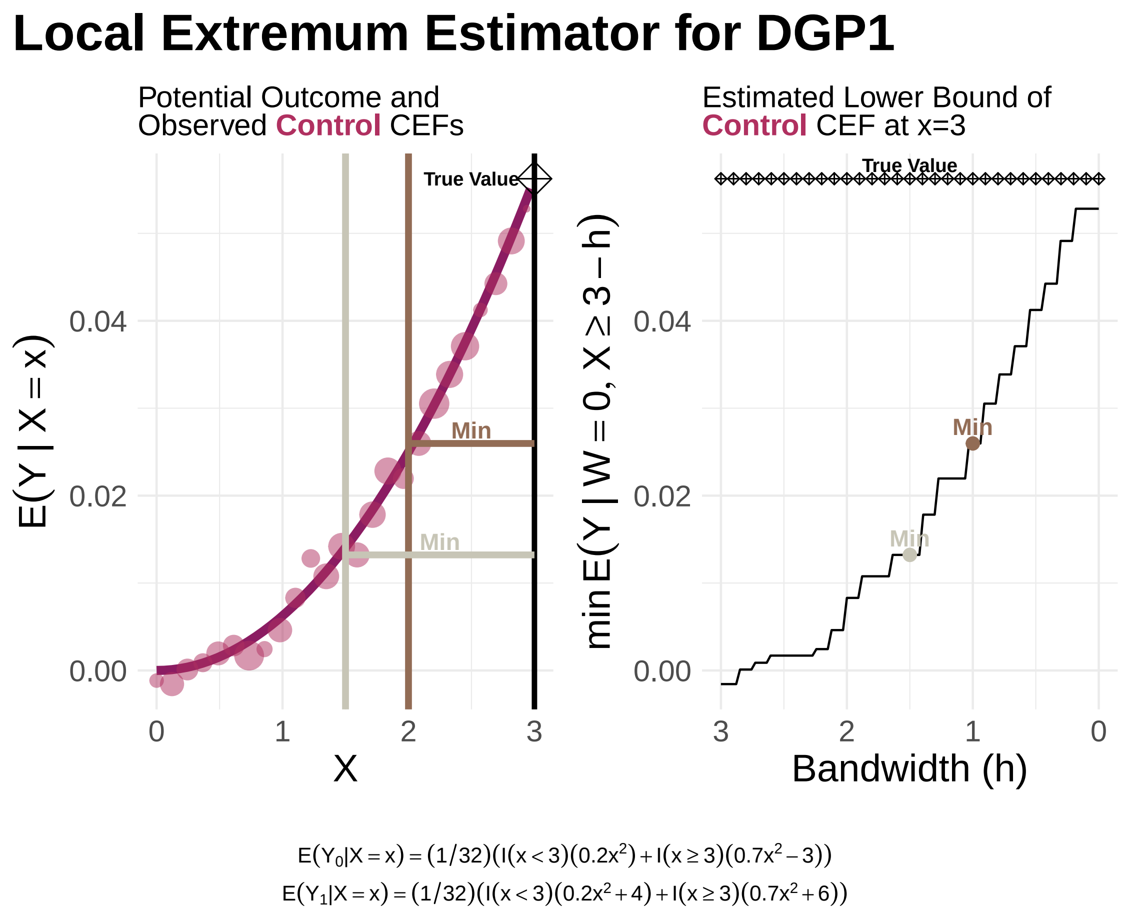



**Figure 3.** Local Extremum Estimation (DGP1 Example)

*Note:* In the left plot, the dark red, monotonic curve is the true potential outcome CEF, with the true limit value at x=3 represented by the diamond cross point. The points are the observed CEF with volume proportional to the share of observations with X=x. The vertical lines mark bandwidth windows (bandwidths h=1.5 and h=1.0) with the horizontal lines denoting the estimated minimum of the observed control CEF within each window. The right plot is the estimated minimum of the control CEF by bandwidth. The tan and gray points correspond to the minimum estimates from the left plot.

One advantage of the local extremum estimator is evident from Figure 3: the estimator ($\widehat{\lim_{x\to C^-} \mu_0^L(C)}$) is a lower bound for the true left-sided limit ($\lim_{x\to C^-} \mu_0^L(C)$) for every bandwidth value $h \in [0,3]$. The local extremum estimator is, in this sense, robust to bandwidth selection although it may be overly conservative for certain bandwidths. For instance, the larger bandwidths (e.g. $h = 2.5$) in the right panel of Figure 3 give estimated lower bounds near zero, when the true left-sided limit value at $x = 3$ is near 0.056. Moreover, bandwidth selection is often the most important researcher decision in an RDD study, which recent work from Korting et al. (2023) shows is consequential for evaluating the validity of the regression discontinuity design.

Another thing to note is how this estimator uses the monotonicity assumption. For a globally monotone increasing function as in Figure 3, the estimator increases in a stepwise but discontinuous fashion as bandwidth $h \to 0$. Again, we do not require global monotonicity of the potential outcome CEFs, but the estimator will perform similarly provided there is a window around the cutoff $C$ where the function is monotone. In a setting with a trend reversal, e.g., the step-like behavior of the local extremum estimator may only hold in a smaller window around the cutoff $C$. For that reason, it makes sense to show results for a variety of bandwidths, even if that is not strictly necessary for identification.

#### 2.3.3 Local Reverse Kernel Estimation

The next estimator we consider is a local reverse kernel estimator. I motivate this estimator similarly: it is a conservative estimator that uses the monotonicity assumption and is robust to bandwidth selection. As is well-known, kernel estimators often behave poorly at point identification near the boundary.[11] Nevertheless, kernel estimators can *bound* or partially identify the boundary.

To begin, I therefore consider reverse kernel weighting functions. For example, unlike the traditional parabolic kernel which equals $K(u) = \frac{3}{4}(1-u^2)$ for $|u| \leq 1$ and is zero elsewhere, we consider a *reverse* parabolic kernel defined by:

$$K(u) = \begin{cases} \frac{3}{2}u^2 & \text{if } |u| \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

This kernel places relatively more weight on further away observations. Using this kernel function $K(u)$, our estimator for a given bandwidth $h > 0$ is then given by:

$$\widehat{\lim_{x \to C^-} \mu_0^L}(C) = \frac{\sum_{i:W_i=0} K\left(\frac{C-X_i}{h}\right) Y_i}{\sum_{i=1}^{n} K\left(\frac{C-X_i}{h}\right)}$$

The estimator is analogous for estimating $\mu_1^R(C)$.

An example of how the local reverse kernel estimator works is given below in Figure 4. In the top-left of the left panel, the inset plot shows the parabolic and reverse parabolic kernel weights as a function of $u$ ($K(u)$). The left panel, shows a noisy sine curve data generating process: the solid blue (resp. dashed green) curve is the reverse parabolic (resp. parabolic) fitted kernel regression line with bandwidth

[11]See, e.g., Imbens and Lemieux (2008) for a discussion.

$h = 0.5$. In the right panel is the estimated boundary value by bandwidth $h$: for example, the solid green and blue boxes are the estimates for the right boundary value, whereas the horizontal line is the true right boundary value. As the bandwidth $h \to 0$, both estimates generally serve as a good upper bound on the true right boundary value. However, the reverse parabolic kernel estimator remains an upper bound for almost all bandwidth choices: by contrast, the parabolic kernel estimator is not an upper bound for bandwidth $h < 0.25$.

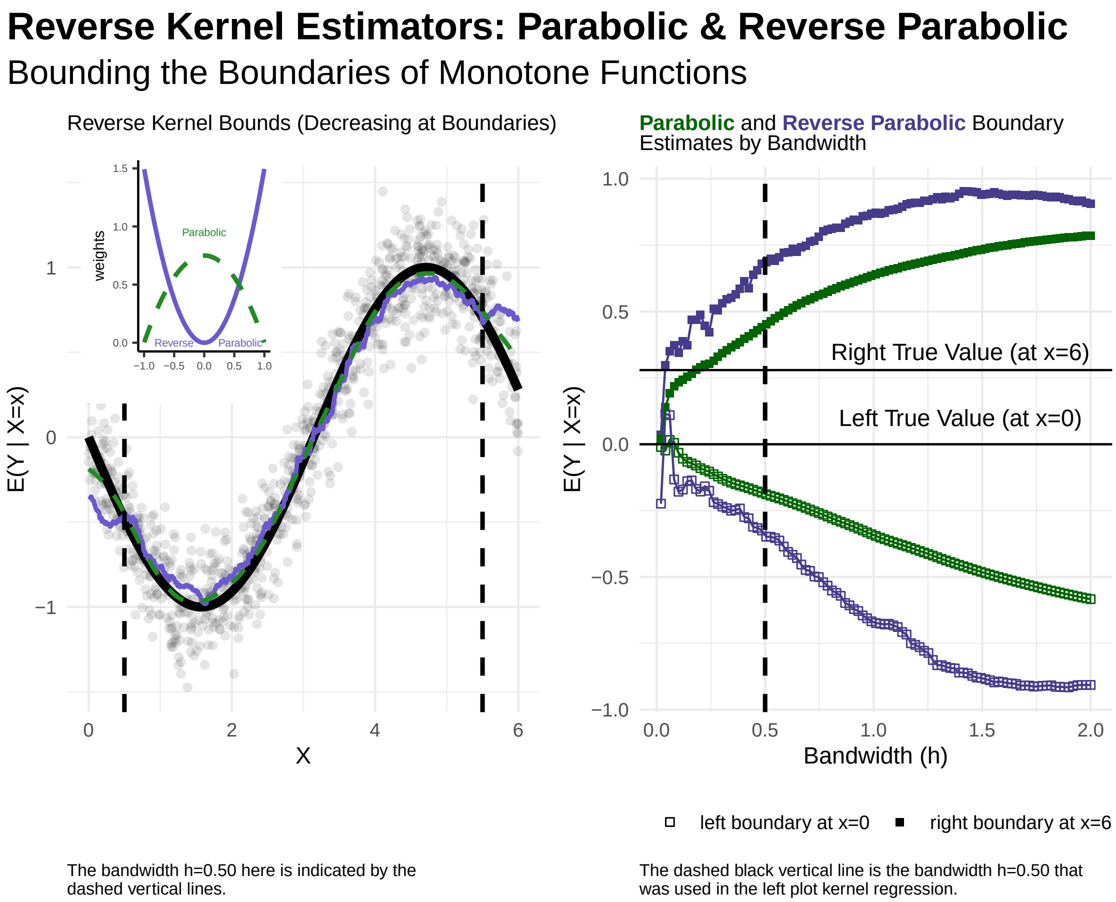



**Figure 4.** Kernel Estimation: Parabolic & Reverse Parabolic Kernels

*Note:* The left plot is the unknown function (solid black) with reverse parabolic (solid blue) and parabolic (dashed green) kernel estimates. The right plot are the kernel estimates by bandwidth for the same decreasing-at-boundaries DGP. The inset plot in the top-left of the left plot give the weights for the parabolic (green) and reverse parabolic (blue) kernels.

When the CEF is increasing near the left boundary, then the reverse kernel estimator serves as an upper bound on the true left boundary value. Conversely, when the CEF is decreasing near the left boundary (as it is in Figure 4), then the estimators serve as a lower bound on the true value at the left boundary. Similarly, when the CEF is increasing (resp. decreasing) near the right boundary, the reverse kernel estimator is a lower (resp. upper) bound on the true right boundary value.

It should also be apparent that the local reverse kernel estimator will likely only work well for either the upper or lower bound. And whether it is a good upper or lower bound will depend on the monotonicity near the boundary. Indeed, we verify this in the Section 3 simulations below. The reason is, for example, that if the control CEF is decreasing near its right boundary, then the kernel estimate at the right boundary will be larger than the true left-sided limit $\lim_{x\to C^-}\mu_0^L(x)$. Accordingly, since we subtract that estimate—which is larger than the true left-sided limit value—to compute the upper bound, the estimated upper bound will be smaller the true upper bound. However, this will not affect the lower bound: subtracting a larger quantity will simply result in an estimated lower bound that is less than the true lower bound and, hence, it may still serve as a bound. Both the local extremum and local reverse kernel estimators are, in this sense conservative and may sacrifice some efficiency (in terms, e.g., of partially-identified region width) for higher coverage rates.[12]

#### 2.3.4 Summary: Three Estimators

In the rest of the discussion, we will use the following notation for our three different estimators—local linear, local extremum, and local reverse kernel—of $\widehat{\lim_{x\to C^-}\mu_0^L(C)}$ and $\widehat{\mu_1^R(C)}$:

$$\widehat{\mu_{0,LL}^L(C)}(h) = \hat{\alpha_L} \qquad \widehat{\mu_{1,LL}^R(C)}(h) = \hat{\alpha_L}$$

$$\widehat{\mu_{0,LE}^L(C)}(h) = \min_{x\in(C-h,C)} \frac{1}{n_{x,0}} \sum_{i:X_i=x,W_i=0} Y_i \qquad \widehat{\mu_{1,LE}^R(C)}(h) = \max_{x\in[C,C+h)} \frac{1}{n_{x,1}} \sum_{i:X_i=x,W_i=1} Y_i$$

$$\widehat{\mu_{0,LRK}^L(C)}(h) = \frac{\sum_{i:W_i=0} K\left(\frac{C-X_i}{h}\right) Y_i}{\sum_{i=1}^n K\left(\frac{C-X_i}{h}\right)} \qquad \widehat{\mu_{1,LRK}^R(C)}(h) = \frac{\sum_{i:W_i=1} K\left(\frac{C-X_i}{h}\right) Y_i}{\sum_{i=1}^n K\left(\frac{C-X_i}{h}\right)}$$

where $K(\cdot)$ is a reverse kernel function, $n_{x,w}$ is the number of units with $X_i = x$ and $W_i = w$, $\hat{\alpha_L}$ comes from the minimization problem $\min_{\alpha_L,\beta_L}\sum_{i:C-h<X_i<C}(Y_i - \alpha_L - \beta_L(X_i - C))^2$, and $\hat{\alpha_R}$ comes from the minimization problem $\min_{\alpha_R,\beta_R}\sum_{i:C\le X_i<C+h}(Y_i -$

[12] Confidence intervals with coverage probability "at least the nominal level regardless of the size of the sample" are referred to as conservative. (Finkelstein, Tucker, and Veeh, 2000). In some cases, "traditional and conservative intervals provide nearly equal coverage probability" while remaining "competitive" in terms of the lengths of the conservative and traditional CIs. (Finkelstein, Tucker, and Veeh, 2000).

$\alpha_R - \beta_R(X_i - C))^2$. Note that each estimator is a function of the bandwidth $h$.

## 3 Simulation Results

In this section, we evaluate the three estimators—local linear, local extremum, and local reverse kernel—with simulations. We assess the coverage rates for each estimator as well as show how the two more conservative estimators (local extremum and local reverse kernel) give conservative bounds across more bandwidths and for relatively small sample sizes.

The simulation uses data generated from four data generating processes ("DGPs"). In all example DGPs, the true potential outcome CEFs have $M_1 = 1$ and $M_0 = 0$ upper and lower bounds, resp., near the cutoff of $C = 3$. In each DGP, treatment is determined by a sharp cutoff rule $W_i = \mathbf{1}\{X_i \geq 3\}$ and the forcing variable is always $X_i \in [0, 6]$. The potential outcome CEFs and the population data for each DGP are shown below in Figure 5. The treated outcome for control units are not observable for units with $X_i < 3$, whereas the control outcome for treated units is not observable for units with $X_i \geq 3$. There are 10,000 or 100,000 observations in each DGP, and the number of unique values in supp$(X)$ ranges from 30 to 1000. Please see Appendix C for further plots and information on each DGP and associated population data.

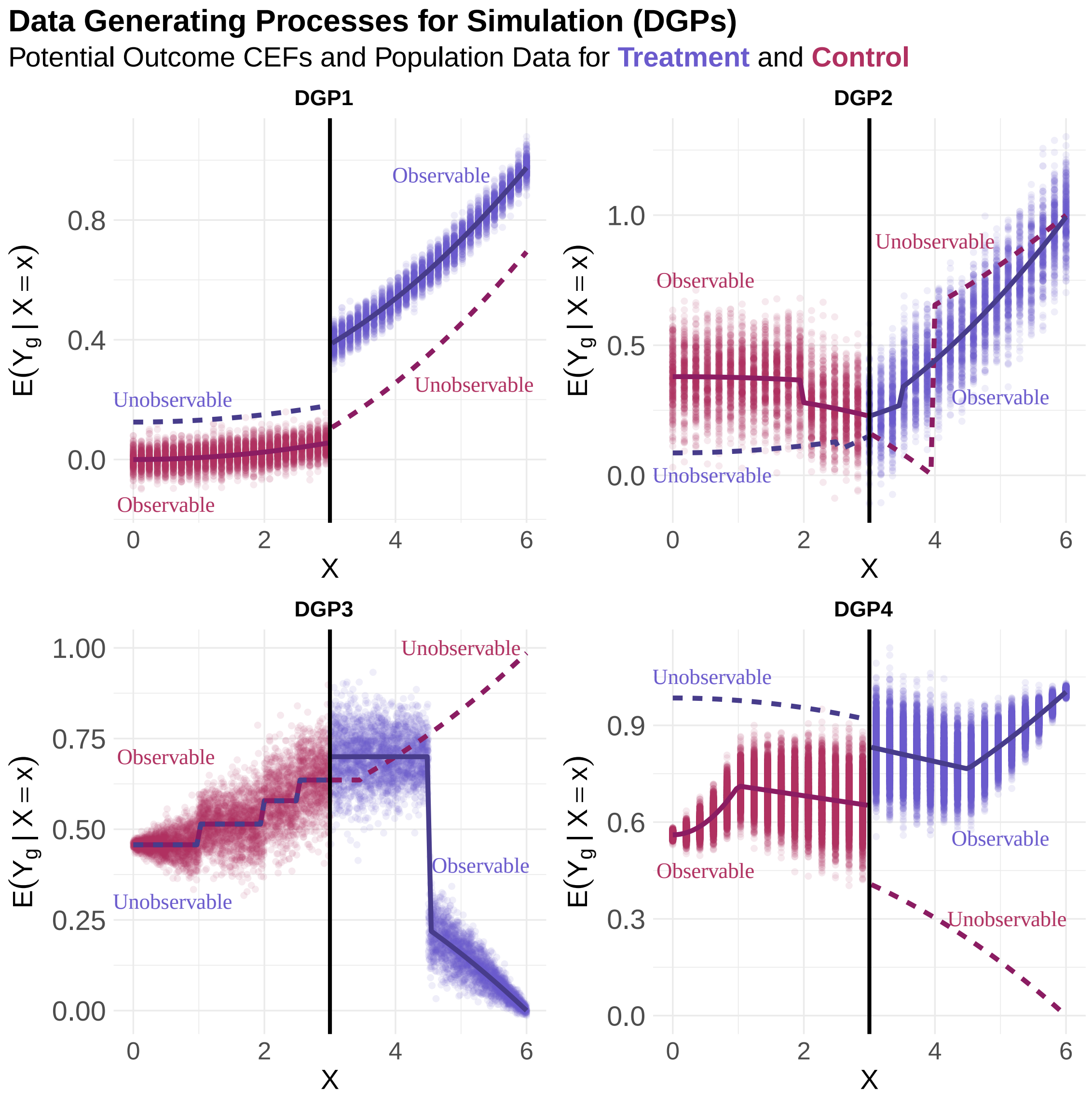



**Figure 5.** Data Generating Processes for Simulation (DGPs)

*Note:* The solid (dashed) lines are the observable (unobservable) potential outcome CEFs. The full population data with added noise is plotted as points.

For each population dataset, we take 1000 simulation samples of sample size $s \in \{100, 500, 1000\}$ with bandwidths $h \in \{0.5, 1.0, 1.5, 3.0\}$ for each DGP. In each simulation run, we estimate the local cutoff treatment effect bounds as discussed above. This involves, for example, estimating the left-sided limit $\lim_{x\to C^-} \mu_0^L(x) = \lim_{x\to C^-} \mathbb{E}[Y_0|X = x]$ and $\mu_1^R(C)$.

For each each DGP, bandwidth, and sample size combination, I evaluate the distribution of estimates. For example, in Appendix Figure E1 I show the estimated upper bounds on the local average treatment effect for DGP1. The estimates generally improve as bandwidth shrinks and sample size increases. For nearly all simulation runs, the Local Kernel and Local Extremum cover the true upper bound on the local average treatment effect (the black vertical line) although the estimates are frequently upward biased. In this sense, the Local Kernel and Local Extremum estimators are "conservative". By contrast, the local linear estimator is less biased but has worse coverage properties. See Appendix E for full local simulation results for each DGP plotted separately.

I next present information on each estimator across all simulations. In Table 1, I show results from the simulation study for each DGP, bandwidth, and sample size combination for the upper bound on the cutoff average treatment effect. I compute for each combination the finite sample bias, standard deviation, coverage of the upper bound, and coverage of the true treatment effect.

As seen in Table 1, the local kernel and local extremum estimators often achieve higher coverage of the upper bound across DGPs, bandwidths, and sample sizes. This is indicated by the green shading in the cells: for each row, the estimator(s) with green shaded cells in the coverage bound ("cov bound") column(s) achieve the highest coverage. As expected, the bias typically decreases as sample size increases and bandwidth decreases for each estimator and DGP. Note, however, that for some bandwidths and sample sizes—typically larger bandwidths and smaller samples—the bias for the local kernel and local extremum estimators is larger. Generally speaking, this corresponds to larger partially identified intervals and it is, in this sense, that the local kernel and local extremum estimators are conservative. The direction of the bias, moreover, is often in the expected direction. For example, because both DGP1 CEFs are monotone increasing, it is unsurprising that the upper bound estimates from the local extremum estimator are higher than the true upper bound and are decreasing as a function of the bandwidth.[13]

[13] That is, an estimate of the local minimum (at some bandwidth to the left of the cutoff) will result in larger upper bound than true if the estimated local minimum is smaller than $\lim_{x \to C^-} \mu_0(x)$, since it is being subtracted from the estimate of $\mu_1(C)$.

| | | | Local Linear (LL) | | | | Local Reverse Kernel (LRK) | | | | Local Extremum (LE) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **dgp** | **bandwidth** | **sample size** | **bias** | **sd** | **cov bound** | **cov true** | **bias** | **sd** | **cov bound** | **cov true** | **bias** | **sd** | **cov bound** | **cov true** |
| **DGP1** | 0.5 | 100 | 0.041 | 0.019 | 0.996 | 1.000 | 0.064 | 0.018 | 1.000 | 1.000 | 0.122 | 0.014 | 1.000 | 1.000 |
| | | 500 | 0.017 | 0.008 | 0.996 | 1.000 | 0.057 | 0.007 | 1.000 | 1.000 | 0.094 | 0.006 | 1.000 | 1.000 |
| | | 1000 | 0.012 | 0.004 | 1.000 | 1.000 | 0.056 | 0.002 | 1.000 | 1.000 | 0.088 | 0.004 | 1.000 | 1.000 |
| | 1.0 | 100 | 0.036 | 0.015 | 0.992 | 1.000 | 0.119 | 0.011 | 1.000 | 1.000 | 0.233 | 0.019 | 1.000 | 1.000 |
| | | 500 | 0.014 | 0.006 | 0.991 | 1.000 | 0.118 | 0.004 | 1.000 | 1.000 | 0.192 | 0.007 | 1.000 | 1.000 |
| | | 1000 | 0.010 | 0.004 | 0.993 | 1.000 | 0.118 | 0.003 | 1.000 | 1.000 | 0.184 | 0.005 | 1.000 | 1.000 |
| | 1.5 | 100 | 0.030 | 0.015 | 0.982 | 1.000 | 0.184 | 0.011 | 1.000 | 1.000 | 0.344 | 0.022 | 1.000 | 1.000 |
| | | 500 | 0.010 | 0.006 | 0.960 | 1.000 | 0.184 | 0.005 | 1.000 | 1.000 | 0.304 | 0.008 | 1.000 | 1.000 |
| | | 1000 | 0.005 | 0.004 | 0.905 | 1.000 | 0.184 | 0.004 | 1.000 | 1.000 | 0.294 | 0.006 | 1.000 | 1.000 |
| | 3.0 | 100 | 0.014 | 0.014 | 0.837 | 1.000 | 0.419 | 0.025 | 1.000 | 1.000 | 0.724 | 0.032 | 1.000 | 1.000 |
| | | 500 | -0.008 | 0.006 | 0.112 | 1.000 | 0.419 | 0.011 | 1.000 | 1.000 | 0.714 | 0.015 | 1.000 | 1.000 |
| | | 1000 | -0.012 | 0.004 | 0.002 | 1.000 | 0.420 | 0.007 | 1.000 | 1.000 | 0.691 | 0.009 | 1.000 | 1.000 |
| **DGP2** | 0.5 | 100 | -0.024 | 0.087 | 0.605 | 0.911 | -0.061 | 0.118 | 0.599 | 0.651 | 0.019 | 0.094 | 0.647 | 0.999 |
| | | 500 | -0.030 | 0.087 | 0.740 | 0.797 | -0.025 | 0.099 | 0.797 | 0.797 | 0.006 | 0.090 | 0.797 | 0.842 |
| | | 1000 | -0.021 | 0.075 | 0.743 | 0.868 | -0.008 | 0.083 | 0.868 | 0.868 | 0.014 | 0.078 | 0.868 | 0.868 |
| | 1.0 | 100 | -0.016 | 0.089 | 0.654 | 0.867 | 0.069 | 0.109 | 0.773 | 0.978 | 0.240 | 0.070 | 1.000 | 1.000 |
| | | 500 | -0.011 | 0.050 | 0.482 | 0.953 | 0.115 | 0.055 | 0.953 | 1.000 | 0.201 | 0.045 | 0.981 | 1.000 |
| | | 1000 | -0.009 | 0.026 | 0.223 | 0.989 | 0.124 | 0.028 | 0.989 | 1.000 | 0.198 | 0.025 | 0.989 | 1.000 |
| | 1.5 | 100 | 0.032 | 0.031 | 0.858 | 0.999 | 0.205 | 0.023 | 0.999 | 1.000 | 0.407 | 0.048 | 1.000 | 1.000 |
| | | 500 | 0.003 | 0.012 | 0.595 | 1.000 | 0.207 | 0.009 | 1.000 | 1.000 | 0.337 | 0.017 | 1.000 | 1.000 |
| | | 1000 | -0.005 | 0.009 | 0.291 | 1.000 | 0.207 | 0.006 | 1.000 | 1.000 | 0.323 | 0.012 | 1.000 | 1.000 |
| | 3.0 | 100 | 0.016 | 0.038 | 0.711 | 0.993 | 0.484 | 0.045 | 1.000 | 1.000 | 0.863 | 0.070 | 1.000 | 1.000 |
| | | 500 | -0.015 | 0.013 | 0.111 | 1.000 | 0.487 | 0.016 | 1.000 | 1.000 | 0.827 | 0.028 | 1.000 | 1.000 |
| | | 1000 | -0.023 | 0.009 | 0.004 | 1.000 | 0.487 | 0.011 | 1.000 | 1.000 | 0.806 | 0.020 | 1.000 | 1.000 |
| **DGP3** | 0.5 | 100 | 0.334 | 0.304 | 0.992 | 0.992 | 0.280 | 0.314 | 0.751 | 0.751 | 0.530 | 0.239 | 1.000 | 1.000 |
| | | 500 | 0.237 | 0.297 | 0.989 | 0.989 | 0.212 | 0.300 | 0.622 | 0.622 | 0.527 | 0.214 | 1.000 | 1.000 |
| | | 1000 | 0.172 | 0.272 | 0.977 | 0.977 | 0.154 | 0.274 | 0.496 | 0.496 | 0.477 | 0.198 | 1.000 | 1.000 |
| | 1.0 | 100 | 0.067 | 0.060 | 0.982 | 0.982 | 0.042 | 0.052 | 0.990 | 0.990 | 0.345 | 0.058 | 1.000 | 1.000 |
| | | 500 | 0.017 | 0.013 | 0.909 | 0.909 | 0.039 | 0.007 | 1.000 | 1.000 | 0.418 | 0.032 | 1.000 | 1.000 |
| | | 1000 | 0.008 | 0.009 | 0.798 | 0.798 | 0.039 | 0.005 | 1.000 | 1.000 | 0.429 | 0.027 | 1.000 | 1.000 |
| | 1.5 | 100 | 0.056 | 0.030 | 0.965 | 0.965 | 0.086 | 0.016 | 1.000 | 1.000 | 0.382 | 0.044 | 1.000 | 1.000 |
| | | 500 | 0.010 | 0.013 | 0.773 | 0.773 | 0.087 | 0.007 | 1.000 | 1.000 | 0.458 | 0.037 | 1.000 | 1.000 |
| | | 1000 | -0.001 | 0.009 | 0.452 | 0.452 | 0.087 | 0.005 | 1.000 | 1.000 | 0.482 | 0.032 | 1.000 | 1.000 |
| | 3.0 | 100 | 0.292 | 0.051 | 1.000 | 1.000 | -0.300 | 0.039 | 0.000 | 0.000 | 0.374 | 0.040 | 1.000 | 1.000 |
| | | 500 | 0.223 | 0.019 | 1.000 | 1.000 | -0.303 | 0.017 | 0.000 | 0.000 | 0.442 | 0.036 | 1.000 | 1.000 |
| | | 1000 | 0.207 | 0.013 | 1.000 | 1.000 | -0.304 | 0.012 | 0.000 | 0.000 | 0.467 | 0.033 | 1.000 | 1.000 |
| **DGP4** | 0.5 | 100 | -0.178 | 0.302 | 0.617 | 0.633 | -0.250 | 0.317 | 0.139 | 0.633 | -0.152 | 0.285 | 0.633 | 0.634 |
| | | 500 | -0.106 | 0.259 | 0.788 | 0.798 | -0.142 | 0.264 | 0.025 | 0.798 | -0.103 | 0.254 | 0.798 | 0.798 |
| | | 1000 | -0.064 | 0.216 | 0.866 | 0.873 | -0.093 | 0.220 | 0.005 | 0.873 | -0.065 | 0.213 | 0.873 | 0.873 |
| | 1.0 | 100 | -0.055 | 0.234 | 0.816 | 0.840 | -0.138 | 0.246 | 0.006 | 0.840 | 0.027 | 0.201 | 0.840 | 0.898 |
| | | 500 | 0.005 | 0.107 | 0.970 | 0.973 | -0.049 | 0.108 | 0.000 | 0.973 | 0.028 | 0.100 | 0.973 | 0.973 |
| | | 1000 | 0.014 | 0.046 | 0.990 | 0.995 | -0.035 | 0.047 | 0.000 | 0.995 | 0.026 | 0.045 | 0.995 | 0.995 |
| | 1.5 | 100 | -0.015 | 0.188 | 0.890 | 0.906 | -0.107 | 0.196 | 0.000 | 0.906 | 0.072 | 0.160 | 0.906 | 0.948 |
| | | 500 | 0.020 | 0.023 | 0.990 | 0.999 | -0.044 | 0.022 | 0.000 | 0.999 | 0.055 | 0.024 | 0.999 | 0.999 |
| | | 1000 | 0.016 | 0.006 | 0.992 | 1.000 | -0.043 | 0.003 | 0.000 | 1.000 | 0.036 | 0.009 | 1.000 | 1.000 |
| | 3.0 | 100 | -0.692 | 0.050 | 0.001 | 0.003 | -0.596 | 0.040 | 0.003 | 0.003 | -0.347 | 0.042 | 0.003 | 0.998 |
| | | 500 | -0.745 | 0.015 | 0.000 | 0.000 | -0.597 | 0.009 | 0.000 | 0.000 | -0.366 | 0.017 | 0.000 | 1.000 |
| | | 1000 | -0.756 | 0.010 | 0.000 | 0.000 | -0.597 | 0.006 | 0.000 | 0.000 | -0.382 | 0.004 | 0.000 | 1.000 |

**Table 1.** Results of Simulation Study (Upper Bound)

*Note:* The table shows the finite-sample bias, standard deviation (sd), and coverage of Local Linear (LL), Local Reverse Kernel (LRK), and Local Extremum (LE) estimators for each DGP, bandwidth, and sample size 3-tuple. "cov bound" is the coverage rate for whether the true upper bound on the treatment effect is less than or equal to the estimated upper bound on the treatment effect, whereas "cov true" is coverage of whether the true treatment effect at the cutoff is less than or equal to the estimated upper bound on the treatment effect. For each row, if the estimator covers the bound with at least 0.95 probability, it is highlighted in green.

We also show the simulation results for the lower bound on the cutoff average treatment effect in Table 2. See Appendix F for coverage information in plot formats for both the "cov bound" (coverage of the bound) and "cov true" (coverage of the true treatment effect).

| | | | Local Linear (LL) | | | | Local Reverse Kernel (LRK) | | | | Local Extremum (LE) | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **dgp** | **bandwidth** | **sample size** | **bias** | **sd** | **cov bound** | **cov true** | **bias** | **sd** | **cov bound** | **cov true** | **bias** | **sd** | **cov bound** | **cov true** |
| **DGP1** | 0.5 | 100 | 0.139 | 0.352 | 0.815 | 0.834 | 0.204 | 0.356 | 0.000 | 0.832 | 0.146 | 0.346 | 0.712 | 0.838 |
| | | 500 | 0.008 | 0.126 | 0.970 | 0.982 | 0.062 | 0.127 | 0.000 | 0.982 | 0.018 | 0.125 | 0.336 | 0.982 |
| | | 1000 | -0.006 | 0.003 | 0.985 | 1.000 | 0.045 | 0.002 | 0.000 | 1.000 | 0.004 | 0.003 | 0.069 | 1.000 |
| | 1.0 | 100 | 0.008 | 0.162 | 0.963 | 0.970 | 0.125 | 0.165 | 0.000 | 0.970 | 0.009 | 0.156 | 0.878 | 0.985 |
| | | 500 | -0.011 | 0.004 | 0.998 | 1.000 | 0.096 | 0.003 | 0.000 | 1.000 | -0.002 | 0.006 | 0.617 | 1.000 |
| | | 1000 | -0.008 | 0.003 | 1.000 | 1.000 | 0.096 | 0.002 | 0.000 | 1.000 | 0.001 | 0.004 | 0.354 | 1.000 |
| | 1.5 | 100 | -0.019 | 0.086 | 0.987 | 0.992 | 0.161 | 0.086 | 0.000 | 0.992 | -0.015 | 0.083 | 0.903 | 0.997 |
| | | 500 | -0.016 | 0.004 | 1.000 | 1.000 | 0.153 | 0.004 | 0.000 | 1.000 | -0.005 | 0.007 | 0.744 | 1.000 |
| | | 1000 | -0.014 | 0.003 | 1.000 | 1.000 | 0.153 | 0.003 | 0.000 | 1.000 | 0.000 | 0.005 | 0.476 | 1.000 |
| | 3.0 | 100 | -0.055 | 0.011 | 1.000 | 1.000 | 0.372 | 0.024 | 0.000 | 1.000 | -0.016 | 0.021 | 0.755 | 1.000 |
| | | 500 | -0.043 | 0.005 | 1.000 | 1.000 | 0.372 | 0.011 | 0.000 | 1.000 | -0.012 | 0.011 | 0.889 | 1.000 |
| | | 1000 | -0.041 | 0.003 | 1.000 | 1.000 | 0.372 | 0.007 | 0.000 | 1.000 | -0.005 | 0.007 | 0.742 | 1.000 |
| **DGP2** | 0.5 | 100 | -0.373 | 0.334 | 0.996 | 0.999 | -0.262 | 0.359 | 0.646 | 0.997 | -0.386 | 0.333 | 1.000 | 1.000 |
| | | 500 | -0.213 | 0.297 | 0.997 | 1.000 | -0.149 | 0.302 | 0.508 | 1.000 | -0.222 | 0.293 | 1.000 | 1.000 |
| | | 1000 | -0.148 | 0.252 | 1.000 | 1.000 | -0.096 | 0.254 | 0.411 | 1.000 | -0.156 | 0.249 | 1.000 | 1.000 |
| | 1.0 | 100 | -0.301 | 0.296 | 1.000 | 1.000 | -0.080 | 0.308 | 0.227 | 0.389 | -0.408 | 0.258 | 1.000 | 1.000 |
| | | 500 | -0.115 | 0.157 | 1.000 | 1.000 | 0.057 | 0.156 | 0.047 | 0.056 | -0.164 | 0.146 | 1.000 | 1.000 |
| | | 1000 | -0.073 | 0.079 | 1.000 | 1.000 | 0.084 | 0.077 | 0.011 | 0.011 | -0.113 | 0.074 | 1.000 | 1.000 |
| | 1.5 | 100 | -0.119 | 0.051 | 0.999 | 1.000 | 0.106 | 0.036 | 0.001 | 0.077 | -0.405 | 0.075 | 1.000 | 1.000 |
| | | 500 | -0.051 | 0.020 | 0.996 | 1.000 | 0.109 | 0.013 | 0.000 | 0.002 | -0.254 | 0.024 | 1.000 | 1.000 |
| | | 1000 | -0.035 | 0.014 | 0.994 | 1.000 | 0.109 | 0.009 | 0.000 | 0.000 | -0.221 | 0.017 | 1.000 | 1.000 |
| | 3.0 | 100 | -0.199 | 0.072 | 1.000 | 1.000 | 0.345 | 0.065 | 0.007 | 0.007 | -0.477 | 0.089 | 1.000 | 1.000 |
| | | 500 | -0.124 | 0.020 | 1.000 | 1.000 | 0.349 | 0.018 | 0.000 | 0.000 | -0.341 | 0.036 | 1.000 | 1.000 |
| | | 1000 | -0.109 | 0.014 | 1.000 | 1.000 | 0.350 | 0.012 | 0.000 | 0.000 | -0.278 | 0.022 | 1.000 | 1.000 |
| **DGP3** | 0.5 | 100 | 0.102 | 0.167 | 0.546 | 1.000 | 0.156 | 0.178 | 0.308 | 0.882 | -0.085 | 0.118 | 0.671 | 1.000 |
| | | 500 | 0.098 | 0.166 | 0.650 | 1.000 | 0.119 | 0.171 | 0.450 | 0.930 | -0.165 | 0.098 | 0.925 | 1.000 |
| | | 1000 | 0.072 | 0.153 | 0.734 | 0.999 | 0.086 | 0.156 | 0.558 | 0.963 | -0.170 | 0.099 | 0.901 | 1.000 |
| | 1.0 | 100 | -0.041 | 0.038 | 0.967 | 1.000 | 0.000 | 0.034 | 0.563 | 0.993 | -0.168 | 0.047 | 0.993 | 1.000 |
| | | 500 | -0.020 | 0.009 | 0.985 | 1.000 | -0.002 | 0.005 | 0.675 | 1.000 | -0.217 | 0.039 | 1.000 | 1.000 |
| | | 1000 | -0.014 | 0.007 | 0.985 | 1.000 | -0.002 | 0.004 | 0.741 | 1.000 | -0.231 | 0.038 | 1.000 | 1.000 |
| | 1.5 | 100 | -0.042 | 0.022 | 0.973 | 1.000 | 0.000 | 0.011 | 0.509 | 1.000 | -0.157 | 0.038 | 1.000 | 1.000 |
| | | 500 | -0.021 | 0.010 | 0.985 | 1.000 | -0.001 | 0.005 | 0.545 | 1.000 | -0.205 | 0.037 | 1.000 | 1.000 |
| | | 1000 | -0.015 | 0.006 | 0.989 | 1.000 | 0.000 | 0.003 | 0.533 | 1.000 | -0.225 | 0.037 | 1.000 | 1.000 |
| | 3.0 | 100 | 0.100 | 0.036 | 0.001 | 1.000 | -0.444 | 0.038 | 1.000 | 1.000 | -0.694 | 0.011 | 1.000 | 1.000 |
| | | 500 | 0.138 | 0.016 | 0.000 | 1.000 | -0.447 | 0.017 | 1.000 | 1.000 | -0.705 | 0.005 | 1.000 | 1.000 |
| | | 1000 | 0.146 | 0.011 | 0.000 | 1.000 | -0.448 | 0.012 | 1.000 | 1.000 | -0.708 | 0.003 | 1.000 | 1.000 |
| **DGP4** | 0.5 | 100 | -0.197 | 0.152 | 0.997 | 1.000 | -0.145 | 0.163 | 0.940 | 1.000 | -0.256 | 0.132 | 1.000 | 1.000 |
| | | 500 | -0.102 | 0.134 | 0.991 | 1.000 | -0.088 | 0.136 | 0.999 | 1.000 | -0.140 | 0.125 | 1.000 | 1.000 |
| | | 1000 | -0.068 | 0.112 | 0.995 | 1.000 | -0.062 | 0.113 | 1.000 | 1.000 | -0.101 | 0.107 | 1.000 | 1.000 |
| | 1.0 | 100 | -0.127 | 0.117 | 0.995 | 1.000 | -0.106 | 0.121 | 1.000 | 1.000 | -0.336 | 0.086 | 1.000 | 1.000 |
| | | 500 | -0.042 | 0.056 | 0.997 | 1.000 | -0.062 | 0.053 | 1.000 | 1.000 | -0.172 | 0.046 | 1.000 | 1.000 |
| | | 1000 | -0.024 | 0.025 | 0.991 | 1.000 | -0.055 | 0.024 | 1.000 | 1.000 | -0.135 | 0.023 | 1.000 | 1.000 |
| | 1.5 | 100 | -0.107 | 0.095 | 0.995 | 1.000 | -0.104 | 0.094 | 1.000 | 1.000 | -0.355 | 0.069 | 1.000 | 1.000 |
| | | 500 | -0.035 | 0.017 | 0.993 | 1.000 | -0.074 | 0.012 | 1.000 | 1.000 | -0.208 | 0.017 | 1.000 | 1.000 |
| | | 1000 | -0.023 | 0.009 | 0.997 | 1.000 | -0.073 | 0.005 | 1.000 | 1.000 | -0.168 | 0.010 | 1.000 | 1.000 |
| | 3.0 | 100 | -0.478 | 0.033 | 1.000 | 1.000 | -0.306 | 0.024 | 0.997 | 1.000 | -0.519 | 0.034 | 1.000 | 1.000 |
| | | 500 | -0.450 | 0.011 | 1.000 | 1.000 | -0.306 | 0.007 | 1.000 | 1.000 | -0.501 | 0.022 | 1.000 | 1.000 |
| | | 1000 | -0.443 | 0.008 | 1.000 | 1.000 | -0.306 | 0.005 | 1.000 | 1.000 | -0.465 | 0.011 | 1.000 | 1.000 |

**Table 2.** Results of Simulation Study (Lower Bound)

*Note:* The table shows the finite-sample bias, standard deviation (sd), and coverage of Local Linear (LL), Local Reverse Kernel (LRK), and Local Extremum (LE) estimators for each DGP, bandwidth, and sample size 3-tuple. "cov bound" is the coverage rate for whether the true lower bound on the treatment effect is greater than or equal to the estimated lower bound on the treatment effect, whereas "cov true" is coverage of whether the true treatment effect at the cutoff is greater than or equal to the estimated lower bound on the treatment effect. For each row, if the estimator covers the bound with at least 0.95 probability, it is highlighted in green.

## 3.1 Further Simulations: More Bandwidths, Missing Data, Measurement Error

Next, I conduct additional simulations to estimate bounds on local treatment effects. Specifically, I conduct three sets of additional simulations: more bandwidths, missing data, and measurement error.

The first set of additional simulations use the same design as before but more bandwidths. Using the four DGPs, more bandwidths allows us to better see how coverage and efficiency (partial ID region widths) vary as the bandwidth varies. The local extremum and local reverse kernel estimators may have better coverage properties but result in larger partial ID regions and are, in this sense, conservative estimation strategies. Indeed, in Appendix Figure F1 and Appendix Figure F2 we see that the local extremum and local reverse kernel estimators often obtain high coverage rates for one or both bounds, but often at the cost of efficiency for higher bandwidth selections. In settings where coverage or bandwidth-robustness are important, estimators adapted to bounding may be practically useful.

The second set of additional simulations uses the four DGPs specified above subject to missing data. Specifically, we conduct simulations where data near the cutoff (running variable values $|X - 3| \leq 0.5$) are missing. Such missingness may occur in practice or in an econometric exercise where data near a cutoff is possibly manipulated (donut RDD). Coverage and partial ID region widths are shown in Appendix Figure F3 and Appendix Figure F4. In general, the results suggest coverage rates for both bounds are low for each estimator, with some success, however, in covering either the upper or lower bound. In such settings, missingness near the boundary may result in severe extrapolation errors if one proceeds with the continuity-based approach and a local linear estimator (depending on the DGP). In such situations, proceeding in the alternative (i.e. assuming DGP is monotone increasing, and then assuming DGP is monotone decreasing) in the missing region, and reporting resulting bounds may be more credible than extrapolation.

The third set of additional simulations uses the original DGPs subject to running variable measurement error. The coverage and partial ID region widths are shown in Appendix Figure F5 and Appendix Figure F6. In the four DGPs, only the local extremum estimator achieves nominal 95% coverage rates at any bandwidth value. Upper and lower bounds alone may be estimated adequately, however, for certain DGPs using the local linear or local reverse kernel estimators. Region widths are generally much larger than in our baseline simulations (esp. for the local extremum estimator), suggesting that in situations which across-the-board measurement error, high coverage of partially-identified bounds comes at an efficiency cost.

## 4 Extensions: Extrapolation to ATE & Fuzzy Monotone RDD

Although regression discontinuity designs are frequently lauded for a high degree of internal validity, policy-makers are not only interested in the effect of a policy for individuals near the cutoff. But point-identifying the most-common estimands—the average treatment effect (ATE) or average treatment effect on the treated ("ATT")—usually requires heroic extrapolation, which may be implausible in a sharp regression discontinuity design where there is a complete failure of overlap. Existing methods to extrapolate treatment effects in an RD setting generally assume constant treatment effects,[14] polynomial functional forms on the potential outcome CEFs,[15] or otherwise restrict the heterogeneity of treatment effects across individual units.[16]

But under our two assumptions, we can partially identify the more general estimand ATE by extrapolation. Indeed, a partial identification approach more fully exploits the modeled information to get lower and upper bounds for larger sub-populations. In the case of global monotonicity, this is the entire population; but in the case of monotonicity on only some of the support of $X$, this corresponds to the sub-population with running variable $x$-values that are compatible with monotone potential outcome CEFs $(\mu_1(x), \mu_0(x))$.

I also briefly consider the fuzzy regression discontinuity design. In this case, further assumptions are needed to bound the local cutoff treatment effect (or to bound the ATE under global monotonicity assumptions). We consider bounds when there is monotone treatment selection ("MTS"), a partial identification assumption discussed in Manski (2007).

### 4.1 ATE Partial Identification

Under the assumptions—global monotonicity ($\mu_1, \mu_0$ both weakly increasing) and boundedness—we can obtain bounds on ATE. Define $\mu_1, \mu_0$ as before:

$$\mu_1(x) = \mathbb{E}[Y_1|X = x] = \mathbf{1}\{x < C\}\mu_1^L(x) + \mathbf{1}\{x \geq C\}\mu_1^R(x)$$
$$\mu_0(x) = \mathbb{E}[Y_0|X = x] = \mathbf{1}\{x < C\}\mu_0^L(x) + \mathbf{1}\{x \geq C\}\mu_0^R(x)$$

[14]See, e.g., Wooldridge (2010) discussing constant treatment effect.

[15]See, e.g., Angrist and Rokkanen (2012) for a good parametric extrapolation example, which considers parametric extrapolation by assuming a specific functional form of the CEF. As noted therein, extrapolation based on functional form restrictions on the CEF is unlikely to achieve point identification. By contrast, here we are focused on partial identification and more flexible shape restrictions.

[16]See Cattaneo, Idrobo, and Titiunik (2024) for further details.

and suppose that the forcing variable $X$ is discrete.[17] Recall that, by the sharp assignment rule $W_i = \mathbf{1}\{X_i \geq C\}$, only $\mu_1^R(x)$ (for $x \geq C$) and $\mu_0^L(x)$ (for $x < C$) are observable.

Theorem 2 gives bounds on the ATE under the monotonicity and boundedness assumptions:

**Theorem 2** (ATE Bounds). *Suppose $\mu_1(x), \mu_0(x)$ are monotone increasing for all $x \in supp(X)$. Let $M_0, M_1$ be global bounds as in Assumption 2. Then the average treatment effect (ATE) is bounded by:*

$$\underline{ATE} \leq ATE \leq \overline{ATE}$$

*where:*

$$\overline{ATE} = \sum_{x<C} Pr(X = x) \cdot \left(\mu_1^R(C) - \mu_0^L(x)\right) + \sum_{x \geq C} Pr(X = x) \cdot \left(\mu_1^R(x) - \lim_{x \to C^-} \mu_0^L(x)\right)$$

$$\underline{ATE} = \sum_{x<C} Pr(X = x) \cdot (M_0 - \mu_0^L(x) + \sum_{x \geq C} Pr(X = x) \cdot (\mu_1^R(x) - M_1)$$

The proof (in Appendix B) is a straightforward application of the LIE and the assumptions. The results for other cases (e.g. both $\mu_1, \mu_0$ monotone decreasing) are similar but omitted. Appendix B also works through a specific example (with a visual example) on bounding the ATE under our assumptions.

### 4.1.1 Partial Identification of $ATE(\Omega)$ Under Local Monotonicity

Monotonicity on the entire support of the forcing variable is a strong shape restriction. However, there may be a region $\Omega$ for which a monotone CEF is consistent with the observable data. The goal, then, is to obtain an estimated region $\hat{\Omega} \subseteq \text{supp}(X)$ that is consistent with monotone CEFs from the observable data.

When $\mu_1, \mu_0$ are monotone for a proper subset $\Omega \subsetneq \text{supp}(X)$, we can partially identify the average treatment effect on that subset. Because each CEF in a sharp RDD is only observable on one side of the cutoff $C$, we also use a symmetry assumption on the monotonicity of each potential outcome CEF across the cutoff. We consider the following local monotonicity and symmetry assumption:

[17] A similar approach works for continuously distributed forcing variables. Replace the summations by integrals in the obvious way and use the pdf of $X$. We use discrete $X$ for expositional purposes and translation into practical settings.

**Assumption 3** (Local Symmetric Monotonicity)**.** *The conditional mean functions $\mu_1, \mu_0$ are (locally) weakly monotone increasing in $x$ for all $x \in \Omega_1, \Omega_0 \subsetneq supp(X)$, respectively, where $\Omega_1 = [C-a, C+a]$ and $\Omega_0 = [C-b, C+b]$ for $a, b \in \mathbb{R}_+$.*

Under Assumption 3, we cannot partially identify the ATE, but we can partially identify the average treatment effect for units with $X \in \Omega = \Omega_1 \cap \Omega_0$:

$$ATE(\Omega) = \mathbb{E}[Y_1 - Y_0 | X \in \Omega]$$

The proof is analogous to the proof of Theorem 2.

If only local monotonicity holds, the econometrician must know in advance or estimate the monotone region $\Omega$. By symmetry, we only need to find the $a, b$ values for which the observable part of $\mu_1, \mu_0$ is compatible with monotonicity. Then we can partially identify $ATE(\Omega)$ for $\Omega = \Omega_1 \cap \Omega_0$. If $\Omega$ is estimated, the estimated monotone region $\hat{\Omega}$ is dependent on the model specification and research context. One possibility is to proceed conservatively and consider the smallest estimated interval $\hat{\Omega}$ that holds under several plausible models.[18]

#### 4.1.2 Nonparametric Estimation & Inference: ATE Bounds

The quantities $M_0, M_1$ are assumed known. Estimating the bounds $\overline{ATE}$ and $\underline{ATE}$ can be done non-parametrically. To estimate the upper bound $\overline{ATE}$, for example, the econometrician needs to estimate $\mu_1^R(C)$, $\lim_{x \to C^-} \mu_0^L(x)$, $\mu_0^L(x)$ for $x < C$, and $\mu_1^R(x)$ for $x \geq C$. I considered estimation methods for the first two estimands above. For the last two estimands, a variety of estimation strategies are available. For instance, if the running variable $X$ is discretely distributed and there are many observations for each unique $x$-value, one can estimate $\mu_1^R(x) = \mathbb{E}[Y_1 | W = 1, X = x]$ for each $x \geq C$ with the sample analogue and use bootstrap procedures to obtain CIs.

### 4.2 Fuzzy RDD

I also consider identification in a fuzzy RDD setting where the potential outcome CEFs are monotone. As before, we assume the running variable $X$ is discrete for expositional simplicity. In the fuzzy RDD, there are treated and control observations

[18] In Manski (2007), for example, monotonicity in the observable CEF is a testable implication of the monotone treatment response and monotone selection assumptions. There, the author suggests using the bootstrap to compute a 95 percent confidence band for the observed CEF. If the confidence band contains everywhere monotone functions, the data is compatible with a monotone CEF on its domain.

on both sides of a cutoff $C$. Although the treatment probability does not change from zero to one at the cutoff, there is still a discontinuous change in the assignment probability at the cutoff. Importantly, this means that ignorability may not hold, i.e.:

$$\mathbb{E}[Y_g|W=w,X=x] \neq \mathbb{E}[Y_g|X=x]$$

for $g=0,1$ and $w=0,1$. Because treatment $W$ is not a deterministic function of $X$, there are therefore four CEFs of interest:

$$\begin{aligned}
\mu_{00}(x) &= \mathbb{E}[Y_0|W=0,X=x] \\
\mu_{01}(x) &= \mathbb{E}[Y_0|W=1,X=x] \\
\mu_{10}(x) &= \mathbb{E}[Y_1|W=0,X=x] \\
\mu_{11}(x) &= \mathbb{E}[Y_1|W=1,X=x]
\end{aligned}$$

Of course, we can estimate (or observe for $x \in \text{supp}(X)$) $\mu_{00}(x), \mu_{11}(x)$ for $x$-values less than $C$ and $x$-values greater than or equal to $C$, respectively. But $\mu_{01}(x), \mu_{10}(x)$ are not observed for any $x$-values: for instance, $\mu_{10}(x)$ is the conditional mean of the treated potential outcome among untreated units with $X=x$.

I consider the following assumption to partially identify the ATE in a fuzzy design. The assumption is a conditional form of monotone treatment selection ("MTS") as in Manski (2007):

**Assumption 4** (Conditional Monotone Treatment Selection)**.** *For $g=0,1$ and $w=0,1$ and any $x \in supp(X)$, if $w \geq w'$, then*

$$\mathbb{E}[Y_g|W=w,X=x] \geq \mathbb{E}[Y_g|W=w',X=x]$$

This assumption relates the unobservable CEFs $(\mu_{01}, \mu_{10})$ to the observable ones. I further assume that $\mu_{11}, \mu_{00}$ are monotonic increasing and that outcomes are uniformly bounded above and below by $M_1, M_0$, resp., as before. In this case, local treatment effect bounds may be obtained, which are formally discussed in Appendix I.

Essentially, the fuzziness allows for "tightening" the bounds on either side of the cutoff. This is visualized with an example in Appendix Figure I1. The two large dots are examples of treated units with $x < C$ (green dot) and control units with $x > C$ (purple dot). The fact that we can observe some treated units to the left of the cutoff and some control units to the right of the cutoff results in step functions that more closely approximate the unobservable CEFs. The bounds are, therefore, tighter in the fuzzy case than the sharp case because the bounds utilize this extra information. See Appendix I for further discussion of this example.

# 5 Application: Energy Efficiency and Property Price Discontinuities

This section turns to an application on the impact of energy efficiency ratings on property prices. Specifically, this section estimates bounds on property prices as a function of discontinuous energy efficiency rating bands (Sejas-Portillo, Moro, and Stowasser, 2025). Using data on UK residential properties, Sejas-Portillo, Moro, and Stowasser (2025) estimate the effects of simplified energy efficiency bands (ranging from A to G) which are themselves determined by underlying energy efficiency scores (1-100). The bands are "arbitrary and technically redundant," yet Sejas-Portillo, Moro, and Stowasser (2025) find "sizeable price discontinuities at rating band thresholds." As described in Sejas-Portillo, Moro, and Stowasser (2025), the result is consistent with a behavioral effect, where less informative rating bands induce price premiums that "are not warranted by actual gains in energy efficiency."

For our purposes, the application is attractive due to the CEF being plausibly a plausibly monotone function of a discrete running variable. Indeed, as visualized in Sejas-Portillo, Moro, and Stowasser (2025)(Figure 3) and reproduced in Appendix Figure F1, the average property price per meter squared is increasing as a function of the running SAP (Standard Assessment Procedure) score. In Appendix Figure F1, the multiple cutoffs as a function of the running SAP score are plotted as red dashed vertical lines and labeled for each energy band rating (A to G). Average prices are plotted at each SAP score in blue points, with larger points representing more available data.

The cutoff effect estimates in the baseline specification are reproduced in Table 3. The local treatment effect estimates $\tau$ at the energy band cutoffs, as well as estimation details, are provided for the first three energy band cutoffs.

| | E–D | F–E | G–F |
|---|---|---|---|
| $\tau$ | 0.019 | 0.029 | 0.023 |
| Robust SE | (0.002) | (0.003) | (0.003) |
| RBC 95% CI | [0.017, 0.027] | [0.024, 0.041] | [0.018, 0.035] |
| RBC $p$-value | [0.000] | [0.000] | [0.000] |
| BW estimate | 3.710\|3.710 | 4.196\|4.196 | 3.816\|3.816 |
| BW bias | 5.470\|5.470 | 6.807\|6.807 | 6.996\|6.996 |
| $N$ | 1,325,863\|3,413,478 | 299,568\|1,325,863 | 65,293\|299,568 |
| Effective $N$ | 404,837\|740,804 | 113,639\|245,331 | 16,631\|31,631 |

**Table 3.** Local Linear RDD Price Discontinuity Estimates (Sejas-Portillo, Moro, and Stowasser (2025) Table 3)

Next, using the data for these first three cutoffs (E–D, F–E, and G–F), I estimate the bounds on treatment effects as a function of the bandwidth parameter. The results from each estimator and cutoff are shown in Figure 6. The horizontal black dashed line is the `rdrobust` (Calonico, Cattaneo, and Titiunik, 2015, 2014) estimate with default selections. As seen in the right-most G–F cutoff panel of Figure 6, for example, the estimated bounds from the local extremum and local reverse kernel estimators are higher than the `rdrobust` estimates and the estimates in Sejas-Portillo, Moro, and Stowasser (2025). This means that, under a running variable monotonicity assumption, the true treatment effect at the G–F cutoff may be even larger than the baseline estimate in Sejas-Portillo, Moro, and Stowasser (2025). Upper bound treatment effect estimates from the smallest bandwidths are similar (only slightly larger) to the rdrobust and Sejas-Portillo, Moro, and Stowasser (2025) estimates. In other cases, the upper bound estimates and the estimated effects from Sejas-Portillo, Moro, and Stowasser (2025) are similar, though the local linear upper bound estimates at many bandwidths are below the rdrobust estimates.

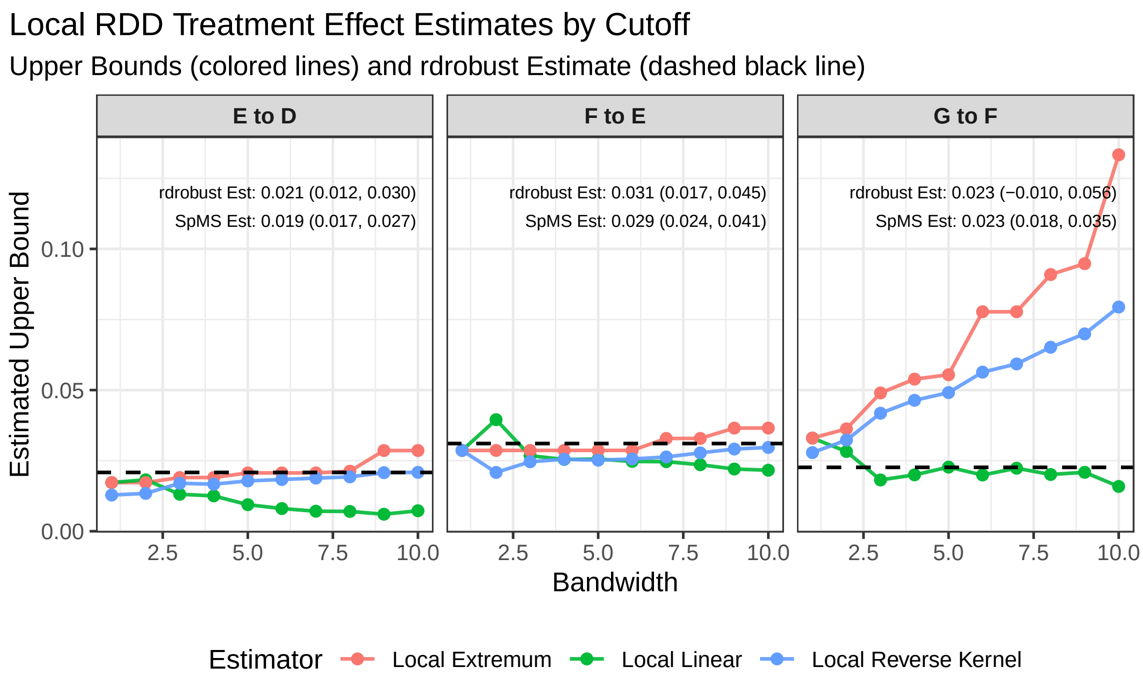



**Figure 6.** Local RDD Treatment Effect Estimates by Cutoff

*Note:* The estimates from rdrobust are plotted as black horizontal dashed line. Robust estimates and estimates Sejas-Portillo, Moro, and Stowasser (2025) are also shown as annotations in each panel.

Finally, I further conducted placebo tests at non-treatment values of the running variable. In continuity-based designs, the idea is to look for continuity violations at running variable values to test the underlying CEF continuity assumptions (Imbens and Lemieux, 2008).[19] Here, placebo tests are also used (with each estimator) to see if jumps at the cutoffs are "large" in the sense of having larger estimated treatment effects than jumps at nearby placebo running variable values. Placebo jump distributions are plotted in Appendix H for running variable values near the three cutoffs (E–D, F–E, and G–F).[20] Under each estimator and in each of the three cutoffs, the estimated upper bounds (and in many cases the Sejas-Portillo, Moro, and Stowasser (2025) estimates themselves) are larger than similarly estimated effect sizes at placebo running variable values. This lends further credibility to the Sejas-Portillo, Moro, and Stowasser (2025) findings that effects at cutoff values are both substantively large even under the alternative identifying assumptions here.

[19] Another possibility is placebo testing in time (Lee, Moretti, and Butler, 2004), where we expect to see "no systematic differences in any variable determined prior to the experiment."

[20] That is, all running variable values on either side of a given cutoff but not beyond the next cutoff.

## 6 Conclusion

This paper develops new methodology that applies generally to regression discontinuity designs (RDD). The usual continuity assumption in the discontinuity design is modified to allow for discontinuities in the potential outcome conditional expectation functions (CEFs) at the cutoff determining treatment. Instead, monotonicity in the CEFs and boundedness in the outcome-of-interest is assumed. Under these assumptions, researchers can partially identify the treatment effect at the cutoff. Additional bounds in fuzzy designs or extrapolated bounds on ATE-type parameters is also permitted under certain assumptions.[21]

Monotone counterfactual assumptions may be useful for a variety of empirical settings. First, the monotonicity assumption is attractive in the RDD setting where the underlying data generating process is well-approximated by a polynomial model.[22] Second, monotonicity may hold where continuity-based identification is implausible: for instance, continuity may be untenable in discrete data generating processes, where there is manipulation of the running variable (McCrary, 2008), or where placebo tests of the continuity assumption at non-cutoff values are violated. Third, continuity may be difficult to visually infer or statistically verify (Korting et al., 2023), whereas monotonicity may be theoretically motivated and empirically tested.[23] Finally, the robustness of causal conclusions from different counterfactual assumptions yields credibility to a study and may strengthen resulting policy recommendations.

As with all counterfactual assumptions, monotonicity is no panacea and the reasonableness of certain assumptions are context-dependent. Partially identified sets of treatment effects or policy parameters under monotonicity may not be helpful where precise estimation is a first-order concern, or where more informative bounds are avail-

[21] Appendix B and Appendix I further discuss possible methods for extrapolation to ATE and partial identification in the fuzzy RDD setup.

[22] There is guaranteed to be a region $\Omega$ where the CEFs are monotone for correctly specified polynomial models around a treatment cutoff. Moreover, in that region we can obtain bounds on $ATE(\Omega) = \mathbb{E}[Y_1 - Y_0|X_i \in \Omega]$, as discussed above. Although continuity is also a plausible assumption in a polynomial model, continuity alone does not permit extrapolation on non-cutoff treatment effects (e.g. $ATE(\Omega)$).

[23] As discussed in Korting et al. (2023), there is strong experimental evidence that detecting continuities or discontinuities is visually difficult, even for experts. Whether a particular RDD passes the "eye test" is highly dependent on data visualization choices under the control of the researcher. This means that, even for the observable portion of each potential outcome CEF, it is often difficult to detect discontinuities. Moreover, despite recommendations that applied researchers test for discontinuities away from the cutoff (Imbens and Lemieux, 2008), it is unclear how widespread the practice. By contrast, it may be possible to visually infer whether a CEF is compatible with monotonicity, and there are formal tests for whether a confidence band contains a monotone function.

able under other tenable assumptions. Plausible theoretical or structural models, for instance, may incorporate monotonicity assumptions, ground counterfactual analysis, and output estimated bounds on treatment effect parameters. Policymakers may use the resulting causal bounds to gauge high-payoff reform areas and researchers may show estimated effects are plausible directionally and in magnitude under bounding assumptions.

## References


Angrist, Joshua and Miikka Rokkanen. 2012. "Wanna Get Away? RD Identification Away from the Cutoff." *NBER* Working Paper 18662. URL https://www.nber.org/system/files/working_papers/w18662/w18662.pdf.

Babii, Andrii and Rohit Kumar. 2023. "Isotonic Regression Discontinuity Designs." *Journal of Econometrics* 234 (2):371–393. URL https://doi.org/10.1016/j.jeconom.2021.01.008.

Calonico, Sebastian, Matias D. Cattaneo, and Rocio Titiunik. 2014. "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs." *Econometrica* 82 (6):2295–2326. URL https://doi.org/10.3982/ECTA11757.

———. 2015. "rdrobust: An R Package for Robust Nonparametric Inference in Regression-Discontinuity Designs." *R Journal* 7 (1):38–51. URL https://rdpackages.github.io/references/Calonico-Cattaneo-Titiunik_2015_R.pdf.

Cattaneo, Matias D., Nicolas Idrobo, and Rocío Titiunik. 2024. *A Practical Introduction to Regression Discontinuity Designs: Extensions*. Elements in Quantitative and Computational Methods for the Social Sciences. Cambridge University Press. URL https://doi.org/10.1017/9781009441896.

Cattaneo, Matias D., Nicolás Idrobo, and Rocío Titiunik. 2020. *A Practical Introduction to Regression Discontinuity Designs: Foundations*. Elements in Quantitative and Computational Methods for the Social Sciences. Cambridge University Press. URL https://doi.org/10.1017/9781108684606.

Finkelstein, Mark, Howard G Tucker, and Jerry Alan Veeh. 2000. "Conservative Confidence Intervals for a Single Parameter." *Communications in Statistics-Theory and Methods* 29 (8):1911–1928. URL https://doi.org/10.1080/03610920008832585.

Gelman, Andrew and Guido Imbens. 2018. "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs." *Journal of Business & Economic Statistics* 37 (3):447–456. URL https://doi.org/10.1080/07350015.2017.1366909.

Goldsmith-Pinkham, Paul. 2026. "Tracking the Credibility Revolution Across Fields." *NBER* Working Paper 35051. URL https://www.nber.org/papers/w35051.

Imbens, Guido W. and Thomas Lemieux. 2008. "Regression Discontinuity Designs: A Guide to Practice." *Journal of Econometrics* 142 (2):615–635. URL https://doi.org/10.1016/j.jeconom.2007.05.001.

Kolesar, Michal and Christoph Rothe. 2018. "Inference in Regression Discontinuity Designs with a Discrete Running Variable." *American Economic Review* 108 (8):2277–2304. URL https://doi.org/10.1257/aer.20160945.

Korting, Christina, Carl Lieberman, Jordan Matsudaira, Zhuan Pei, and Yi Shen. 2023. "Visual Inference and Graphical Representation in Regression Discontinuity Designs." *Quarterly Journal of Economics* 138 (3):1977–2019. URL https://doi.org/10.1093/qje/qjad011.

Lee, David S., Enrico Moretti, and Matthew J. Butler. 2004. "Do Voter Affect or Elect Policies? Evidence from the U.S. House." *Quarterly Journal of Economics* 119 (3):807–859. URL https://www.princeton.edu/~davidlee/wp/voterspolicies.pdf.

Manski, Charles F. 2007. *Identification for Prediction and Decision*. Harvard University Press.

McCrary, Justin. 2008. "Manipulation of the Running Variable in the Regression Discontinuity Design: A Density Test." *Journal of Econometrics* 142 (2):698–714. URL https://doi.org/10.1016/j.jeconom.2007.05.005.

Noack, Claudia and Christoph Rothe. 2023. "Donut Regression Discontinuity Designs." URL https://arxiv.org/pdf/2308.14464.

Sejas-Portillo, Rodolfo, Mirko Moro, and Till Stowasser. 2025. "The Simpler the Better? Threshold Effects of Energy Labels on Property Prices and Energy Efficiency Investments." *American Economic Journal: Applied Economics* 17 (1):41–89. URL https://doi.org/10.1257/app.20220329.

Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross Section and Panel Data*. The MIT Press, 2nd ed.

# APPENDICES

# A Appendix A: Sharp RDD Constant Treatment Effect

Appendix A gives regression results for the constant treatment effect sharp RDD example.

| | *Dependent variable:* | |
|---|---|---|
| | Y | |
| | (1) | (2) |
| X-3 | 3.268*** | 2.893*** |
| | (2.981,3.555) | (2.552,3.234) |
| Treated | 7.000*** | 7.000*** |
| | (6.290,7.710) | (6.713,7.287) |
| (X-3):Treated | 0.000 | -0.000 |
| | (-0.406,0.406) | (-0.482,0.482) |
| Constant | 4.950*** | 2.232*** |
| | (4.448,5.452) | (2.029,2.434) |
| Observations | 200 | 68 |
| $R^2$ | 0.875 | 0.978 |
| Adjusted $R^2$ | 0.873 | 0.977 |
| F Statistic | 456.389*** (df = 3; 196) | 948.619*** (df = 3; 64) |
| *Note:* | | *p<0.1; **p<0.05; ***p<0.01 |

**Appendix Table A1.** Linear Regression Results

As seen above, the estimates on "Treated" recover the true treatment effect of 7 from the example plot. Appendix Table A1 reports estimates of the local treatment effect. The first column uses all observations, whereas the second column uses only observations with $X_i \in [2, 4]$, which is the local linear regression plotted in thick dark red and dark blue lines between the dashed vertical lines in Figure 1. In both columns, we regress $Y_i$ on $1, (X_i - 3), W_i, W_i(X_i - 3)$ and recover the true $\tau_C = 7.0$; however, the local linear regression (Column (2)) gives smaller CIs on $\tau_C$.

# B Appendix B: Extrapolation to ATE

Appendix B considers extrapolation to the ATE. For simplicity, we focus on the case where the running variable $X$ is discrete.

## B.1 Extrapolation: Proof of the ATE Bounds

We first obtain the upper bound by applying the LIE, the global monotonicity assumption, and observing the following:

$$
\begin{aligned}
ATE &= \mathbb{E}[Y_1 - Y_0] \\
&= \sum_{x \in \text{supp}(X)} Pr(X = x) \cdot \mathbb{E}[Y_1 - Y_0 | X = x] \\
&= \sum_{x<C} Pr(X = x) \cdot \mathbb{E}[Y_1 - Y_0 | X = x] + \sum_{x \geq C} Pr(X = x) \cdot \mathbb{E}[Y_1 - Y_0 | X = x] \\
&\leq \sum_{x<C} Pr(X = x) \cdot (\mathbb{E}[Y_1 | W = 1, X = C] - \mathbb{E}[Y_0 | W = 0, X = x]) \\
&\qquad + \sum_{x \geq C} Pr(X = x) \cdot \left( \mathbb{E}[Y_1 | W = 1, X = x] - \lim_{x \to C^-} \mathbb{E}[Y_0 | W = 0, X = x] \right) \\
&\leq \underbrace{\sum_{x<C} Pr(X = x) \cdot \left(\mu_1^R(C) - \mu_0^L(x)\right) + \sum_{x \geq C} Pr(X = x) \cdot \left( \mu_1^R(x) - \lim_{x \to C^-} \mu_0^L(x) \right)}_{=\overline{ATE}}
\end{aligned}
$$

where we used that $\mu_1(x), \mu_0(x)$ are increasing functions of $x$. Note as well that $\mu_1^R(x)$ for $x \geq C$ is observable by the sharp assignment rule; and similarly for $\mu_0^L(x)$ for $x < C$.

The lower bound on ATE uses boundedness and monotonicity. Again, we apply the LIE and the assumptions:

$$
\begin{aligned}
ATE &= \mathbb{E}[Y_1 - Y_0] \\
&= \sum_{x<C} Pr(X = x) \cdot \mathbb{E}[Y_1 - Y_0 | X = x] + \sum_{x \geq C} Pr(X = x) \cdot \mathbb{E}[Y_1 - Y_0 | X = x] \\
&\geq \underbrace{\sum_{x<C} Pr(X = x) \cdot (M_0 - \mu_0^L(x)) + \sum_{x \geq C} Pr(X = x) \cdot (\mu_1^R(x) - M_1)}_{=\underline{ATE}}
\end{aligned}
$$

where we use that $\mu_1(x)$ is bounded below by $M_0$ for $x < C$ and that $\mu_0(x)$ is bounded above by $M_1$ for $x > C$.

Thus, the upper and lower bounds for the ATE are given by the quantities:

$$\underline{ATE} \leq ATE \leq \overline{ATE}$$

where

$$\overline{ATE} = \sum_{x<C} Pr(X=x) \cdot \left(\mu_1^R(C) - \mu_0^L(x)\right) + \sum_{x\geq C} Pr(X=x) \cdot \left(\mu_1^R(x) - \lim_{x\to C^-} \mu_0^L(x)\right)$$
$$\underline{ATE} = \sum_{x<C} Pr(X=x) \cdot (M_0 - \mu_0^L(x) + \sum_{x\geq C} Pr(X=x) \cdot (\mu_1^R(x) - M_1)$$

### B.2 ATE: Bounding Example

To fix ideas, I also include an example of bounding the ATE. In the example, the forcing variable $X$ is discrete and the CEFs are globally monotone increasing. I consider Figure B1 below, which demonstrates the idea for the upper bound on $ATE$. First, in the top, left-most plot, we display the same CEFs for the treated (blue) and control (maroon) groups as above in Figure 2. The specific CEFs defined on $\text{supp}(X) = [0, 6]$ are recorded below again:

$$\mu_1(x) = \mathbf{1}\{x < 3\}(0.2x^2 + 2) + \mathbf{1}\{x \geq 3\}(0.7x^2 + 2)$$
$$\mu_0(x) = \mathbf{1}\{x < 3\}(0.2x^2 - 2) + \mathbf{1}\{x \geq 3\}(0.7x^2 - 5)$$

The treatment assignment cutoff occurs at $C = 3$. The solid lines are observed: because of the assignment rule $W_i = \mathbf{1}\{X_i \geq 3\}$, $\mu_1(x) = \mathbb{E}[Y_1|X = x]$ is observable for $x \geq 3$, while $\mu_0(x) = \mathbb{E}[Y_0|X = x]$ is observable for $x < 3$.[24] The dashed lines are the true values for the CEFs, but they are unobservable.

As before, note that both CEFs are discontinuous at the regression cutoff value $C = 3$. The typical RDD assumption that both conditional means $\mu_1(x), \mu_0(x)$ are continuous in $x$ **does not hold.** However, under (global) monotonicity we can obtain an upper bound on the ATE. Consider the top, right-most panel in Figure B1: the dotted lines for each potential outcome CEF are those that are compatible with monotonicity for the purposes of calculating an upper bound on the ATE.

Since we know the DGP, we can calculate the upper bound on the ATE as follows. First, we assume the $X$ are uniformly distributed on $[0, 6]$ and we calculate the highest

[24] Again, we use "observable" to mean "can be estimated consistently from the data."

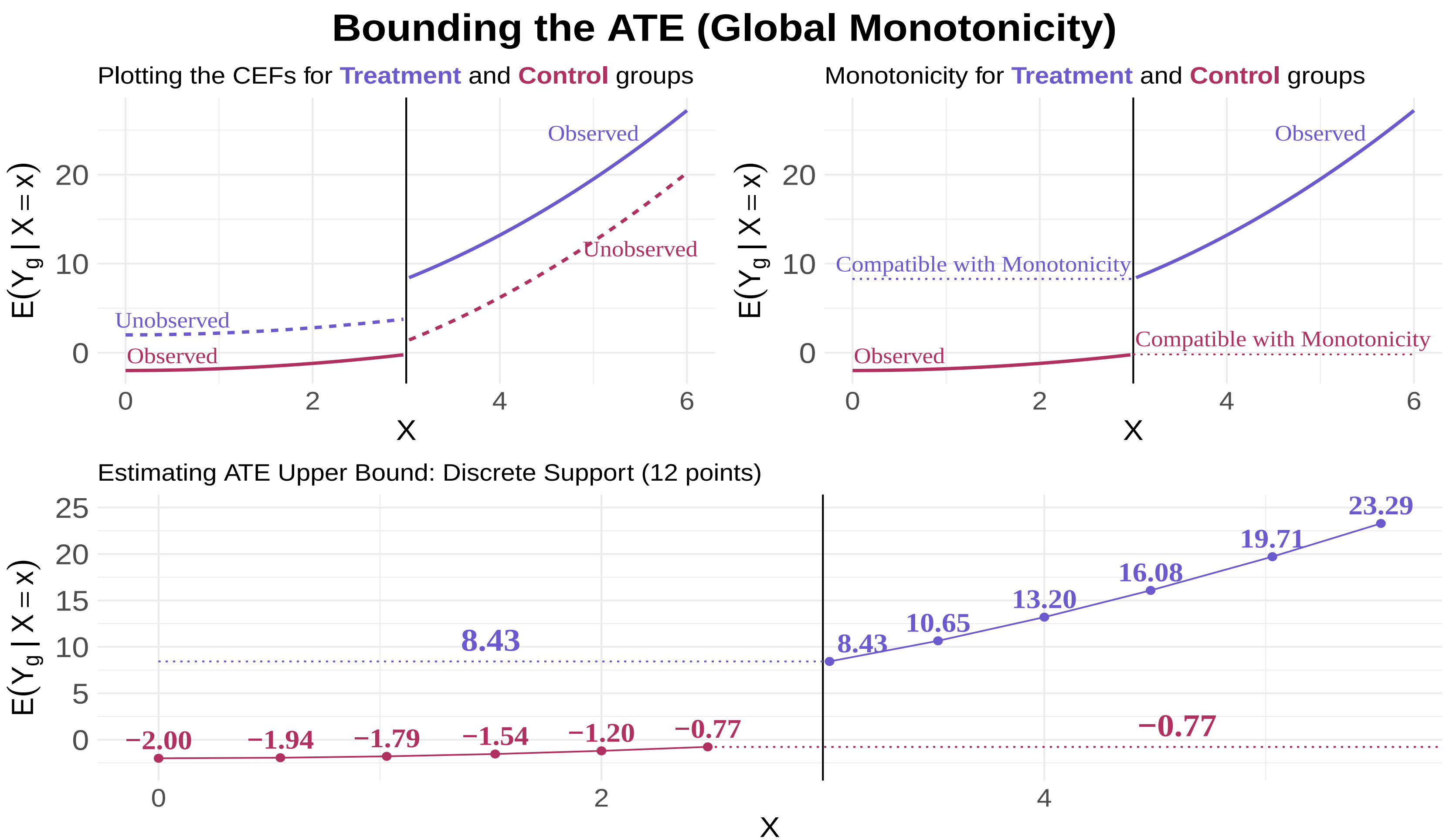
values of $\mu_1(x)$ for each $x$ and the lowest values of $\mu_0(x)$ that are compatible with monotonicity and the observed data. The ATE is the difference between these values integrated over the support of the forcing variable. In our Figure B1 example, this means, for $x \in [0, 3)$, taking the difference between the dotted blue line and the solid maroon line; whereas, for $x \in [3, 6]$, we take the difference between the solid blue line and the dotted maroon line. This results in an upper bound for the ATE of:

$$\begin{aligned}\overline{ATE} &= \frac{1}{6}\int_{x=0}^{3}(8.3 - (0.2x^2 - 2))dx + \frac{1}{6}\int_{x=3}^{6}(0.7x^2 + 2 - (-0.2))dx \\ &= \frac{1}{6}(29.1 + 50.7) = \boxed{13.3}\end{aligned}$$

Note that this can also be done for RDD cases where there are only a few discrete values that the forcing variable $X$ can take. For example, in the bottom panel of Figure B1, we keep the same CEFs but suppose we have only 6 support points on each side of the cutoff $C$. The values for $\mathbb{E}[Y|X = x]$ are displayed above each point in the panel. Supposing that $X$ is uniformly distributed, this gives an estimated upper bound on the ATE of 12.98, which is close to the 13.3 above in the continuously distributed $X$ scenario.

Compare this to the true ATE, which is, for $x \in [0, 3)$, given by taking the difference between the dashed blue line and the solid maroon line; whereas, for $x \in [3, 6]$, we take the difference between the solid blue line and the dashed maroon line, and calculated as follows:

$$\begin{aligned}ATE &= \frac{1}{6}\left(\int_{x=0}^{3}\big((0.2x^2 + 2) - (0.2x^2 - 2)\big)\,dx + \int_{x=3}^{6}\big((0.7x^2 + 2) - (0.7x^2 - 5)\big)\,dx\right) \\ &= \frac{1}{6}(3 * 4 + 3 * 7) = \boxed{5.5}\end{aligned}$$

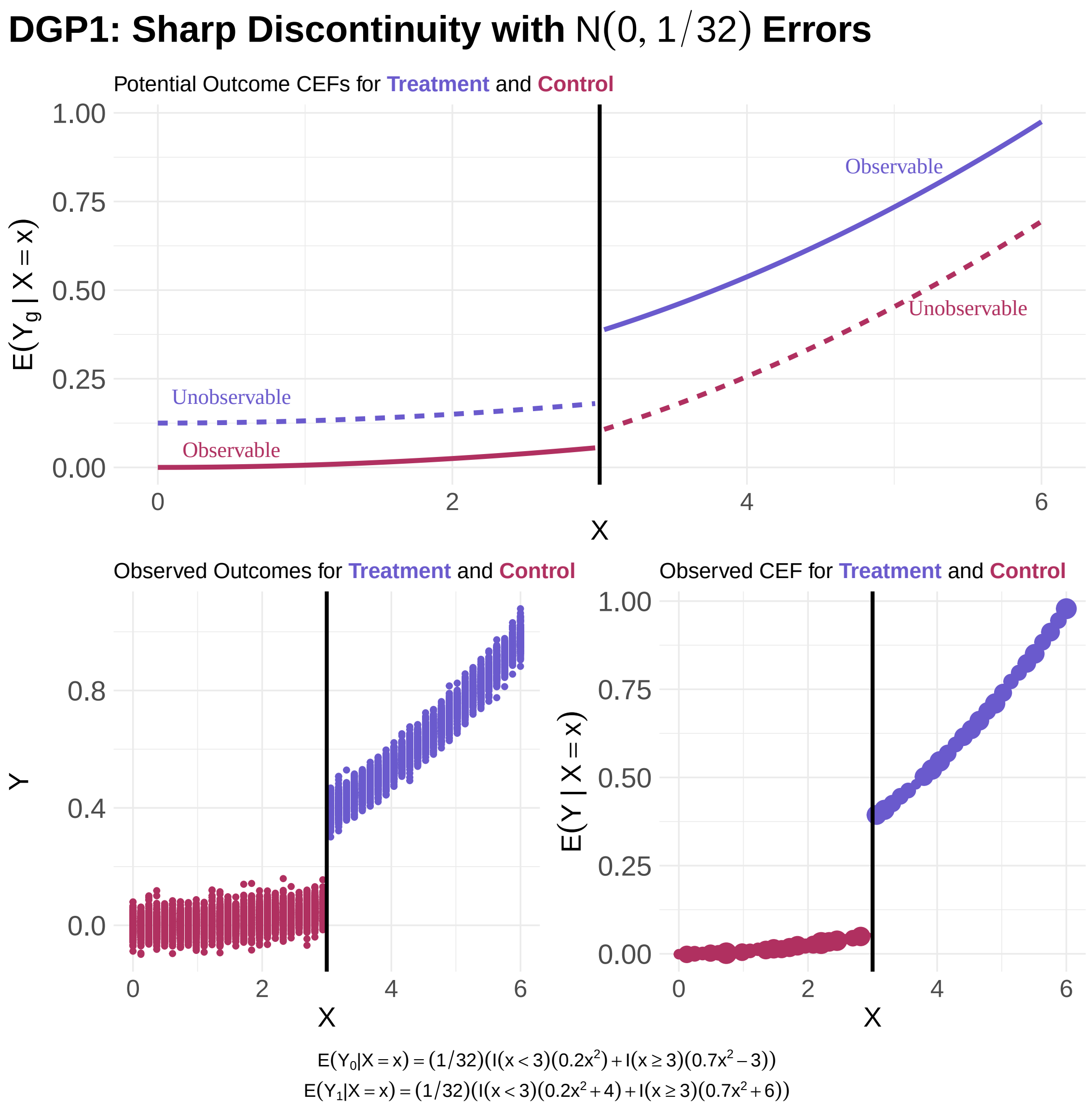



**Appendix Figure B1.** Bounding the ATE Under Global Monotonicity

*Note:* The top left panel shows the DGP. The top right panel shows the imputed values for each CEF compatible with monotonicity for the purposes of computing the upper bound on ATE. The bottom panel is included for a computational example: using these values there, and if the forcing variable X is uniformly distributed, the ATE upper bound is approximately 12.98.

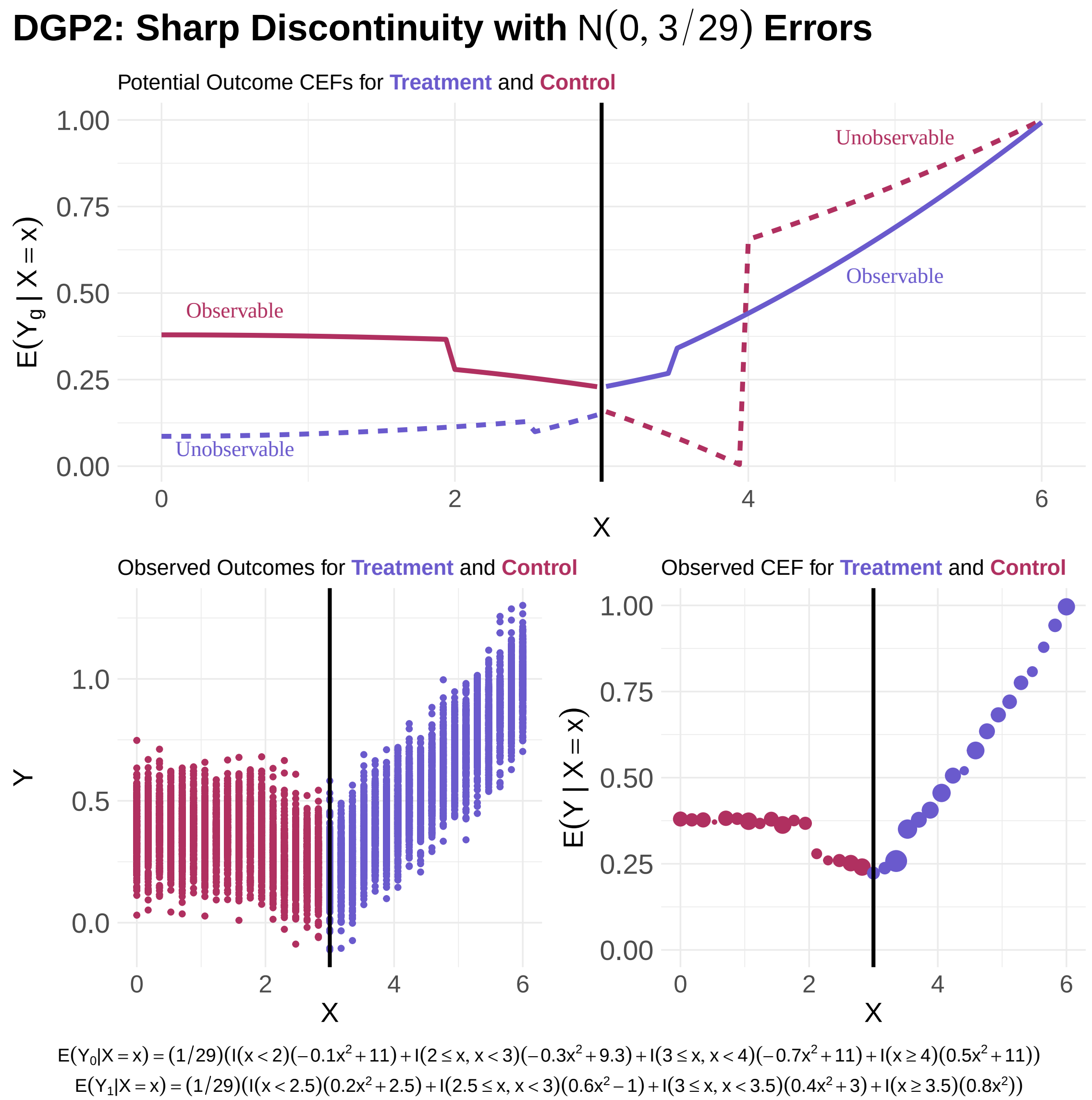
# C Appendix C: Data Generating Processes (DGPs)

Appendix C gives more information on the data generating processes used in the simulation study. Each plot belows shows the true potential outcome CEFs, the population data with added noise, and the observable population CEF.



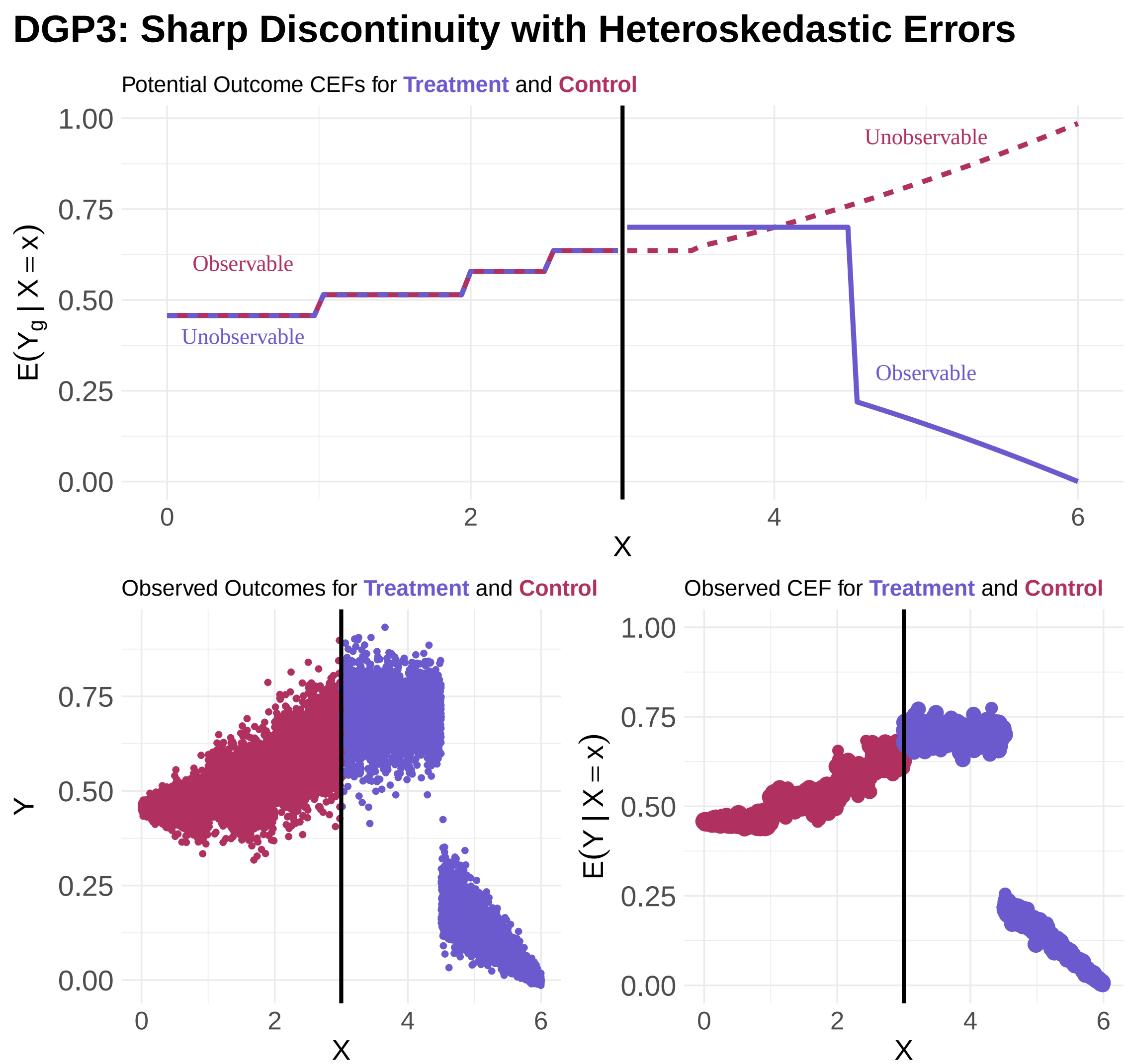

**Appendix Figure C1.** Data Generating Process 1 (DGP1)

*Note:* By the sharp design, the solid lines of the potential outcome CEFs are observable for treated units (control units) to the right (left) of the cutoff $C = 3$. The dashed lines are the unobservable part of the potential outcome CEFs because of the sharp design. The equations for the potential outcome CEFs are given below, where I() is the indicator function. The true local treatment effect at $C = 3$ is approximately 0.281. The bottom-left panel shows that there are 50 unique $X$ values in the simulated dataset. In the bottom-right panel, the volume of the points is proportional to the share of observations with $X = x$.




**Appendix Figure C2.** Data Generating Process 2 (DGP2)

*Note:* By the sharp design, the solid lines of the potential outcome CEFs are observable for treated units (control units) to the right (left) of the cutoff $C = 3$. The dashed lines are the unobservable part of the potential outcome CEFs because of the sharp design. The equations for the potential outcome CEFs are given below panels, where I() is the indicator function. The true local treatment effect at $C = 3$ is approximately 0.066. The bottom-left panel shows that there are 35 unique $X$ values in the simulated dataset. In the bottom-right panel, the volume of the points is proportional to the share of observations with $X = x$.

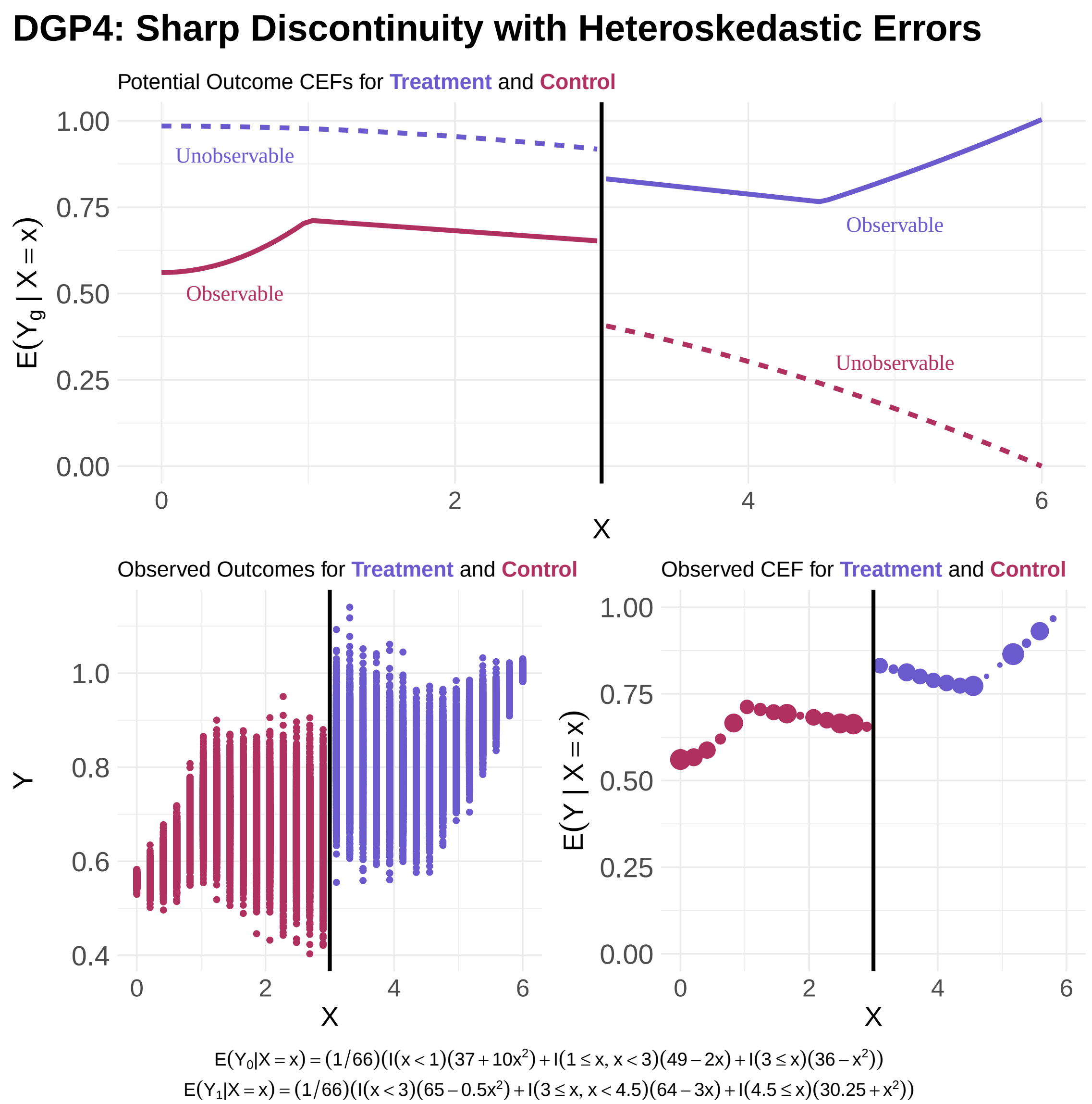



$$E(Y_0|X=x) = (1/70)(I(x<1)32 + I(1 \leq x, x<2)36 + I(2 \leq x, x<2.5)40.5 + I(2.5 \leq x, x<3.5)44.5 + I(3.5 \leq x)(x^2+33))$$

$$E(Y_1|X=x) = (1/70)(I(x<1)32 + I(1 \leq x, x<2)36 + I(2 \leq x, x<2.5)40.5 + I(2.5 \leq x, x<3)44.5 + I(3 \leq x, x<4.5)49 + I(4.5 \leq x)(-x^2+36))$$

**Appendix Figure C3.** Data Generating Process 3 (DGP3)

*Note:* By the sharp design, the solid lines of the potential outcome CEFs are observable for treated units (control units) to the right (left) of the cutoff $C = 3$. The dashed lines are the unobservable part of the potential outcome CEFs because of the sharp design. The equations for the potential outcome CEFs are given below panels, where I() is the indicator function. The true local treatment effect at $C = 3$ is approximately 0.064. The bottom-left panel shows that there are 1000 unique $X$ values in the simulated dataset. In the bottom-right panel, the volume of the points is proportional to the share of observations with $X = x$.

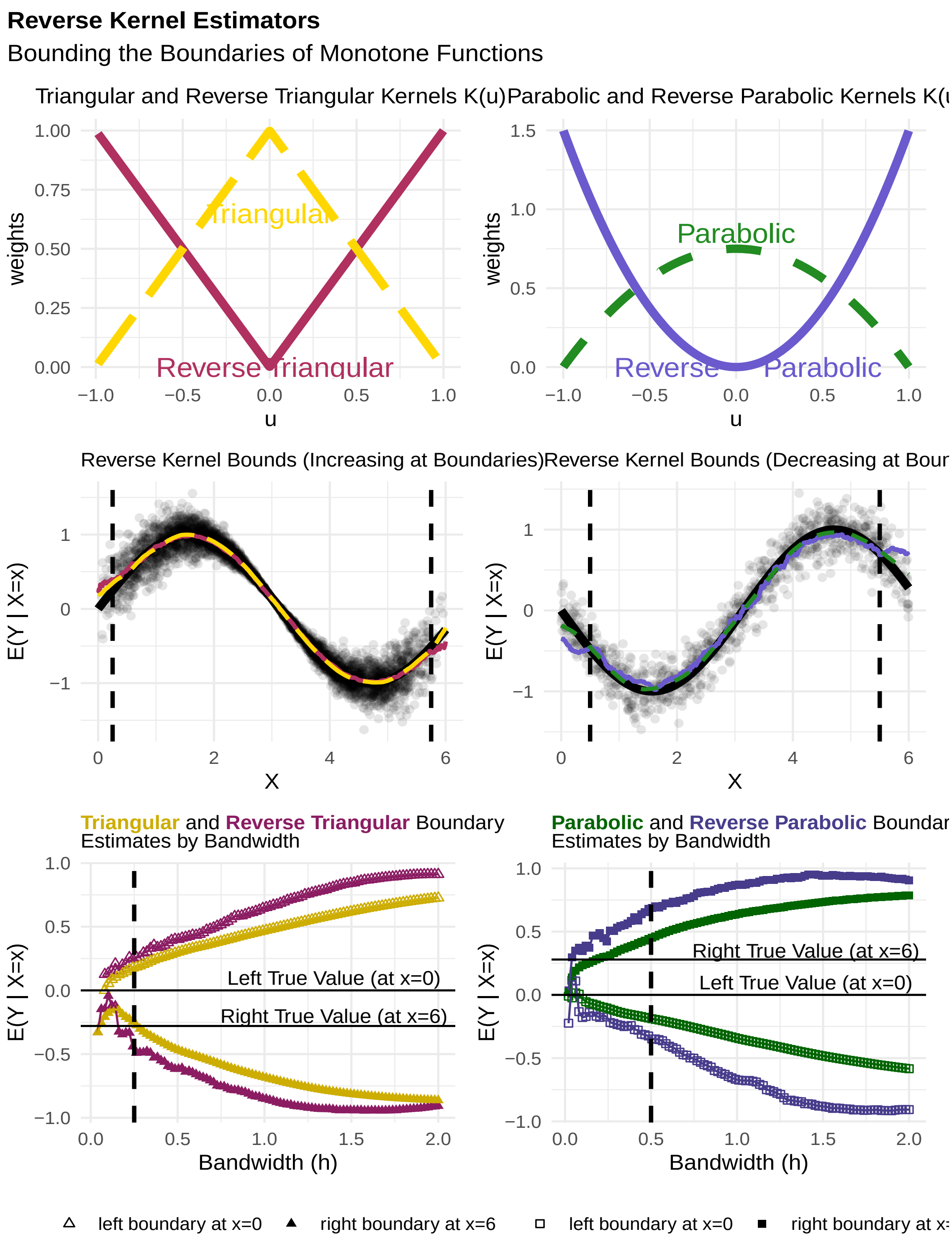



**Appendix Figure C4.** Data Generating Process 4 (DGP4)

*Note:* By the sharp design, the solid lines of the potential outcome CEFs are observable for treated units (control units) to the right (left) of the cutoff $C = 3$. The dashed lines are the unobservable part of the potential outcome CEFs because of the sharp design. The equations for the potential outcome CEFs are given below panels, where I() is the indicator function. The true local treatment effect at $C = 3$ is approximately 0.424. The bottom-left panel shows that there are 30 unique $X$ values in the simulated dataset. In the bottom-right panel, the volume of the points is proportional to the share of observations with $X = x$.

# D Appendix D: Reverse Kernel Examples

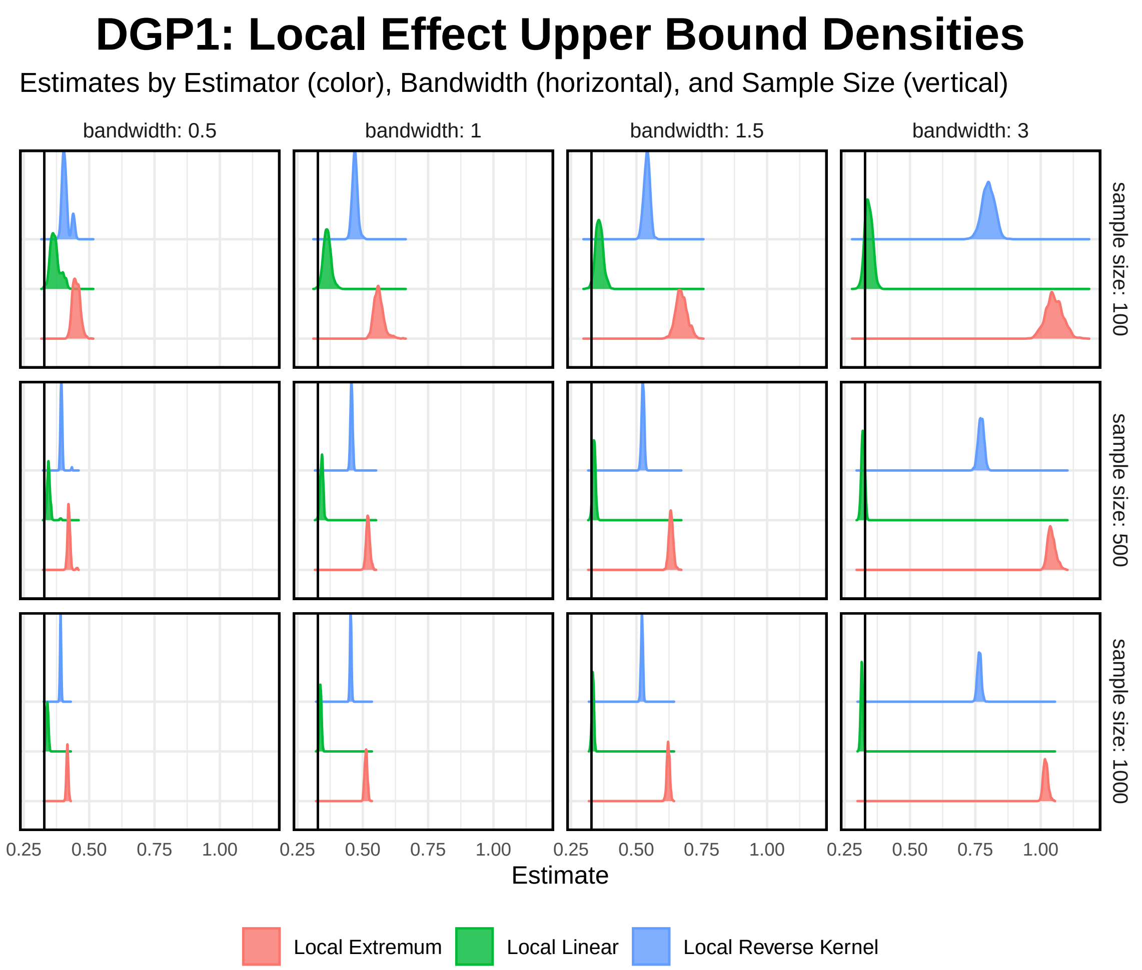
Appendix D shows further examples of reverse kernels. The top plots are the kernels used (weights). The middle plots are two unknown functions with estimates based on the above kernels: the solid red (reverse triangular) and solid blue (reverse parabolic) lines are the reverse kernel estimates, whereas the dashed yellow (triangular) and dashed green (parabolic) lines are the standard kernel estimates. For viewability, we use the triangular kernels in the left-side plots and the parabolic kernels in the right-side plots. The bottom two plots are kernel estimates by bandwidth for the increasing (bottom-left) and decreasing (bottom-right) DGPs, and for viewability we only include either the parabolic and reverse parabolic (resp. triangular and reverse triangular) estimates in the bottom-left (resp. bottom-right) plot. The bandwidths (left h=0.25 and right h=0.50) are indicated by dashed vertical lines in the middle and bottom panels.



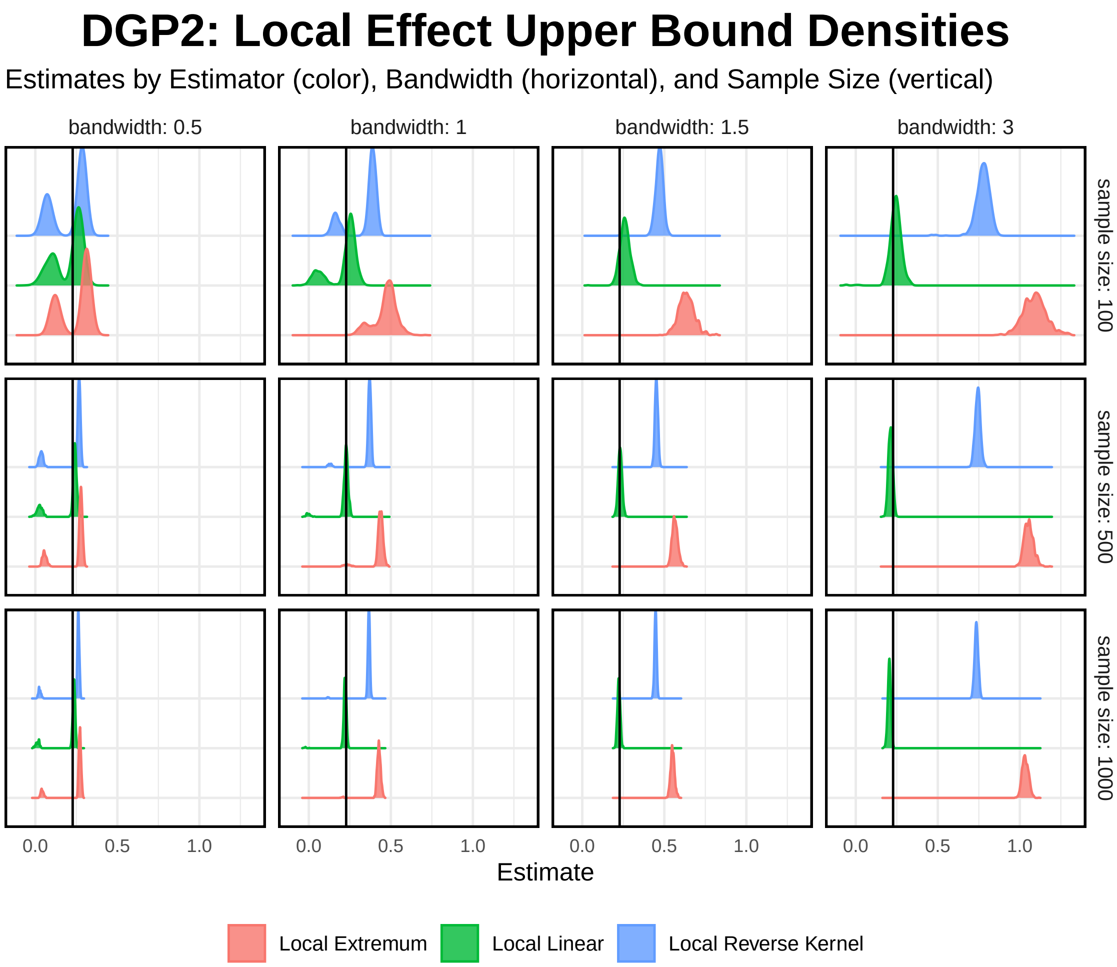

**Appendix Figure D1.** Reverse Kernel Estimation

# E Appendix E: Local Simulation Study Results

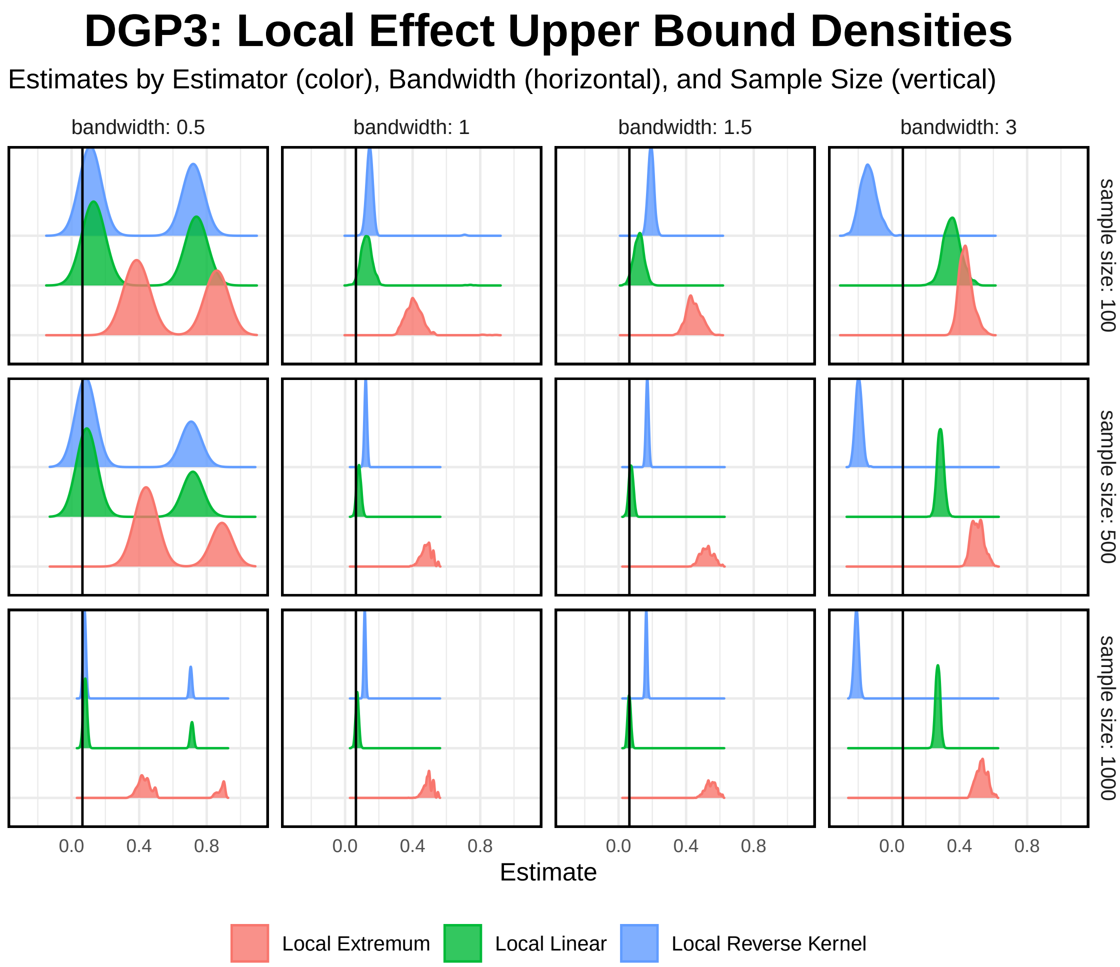
Appendix E shows the local simulation results as density plots for each DGP and both the upper and lower bounds.




**Appendix Figure E1.** DGP1: Local Effect Upper Bound Simulation Results
*Note:* The true local upper bound is given by the black vertical line in each panel.

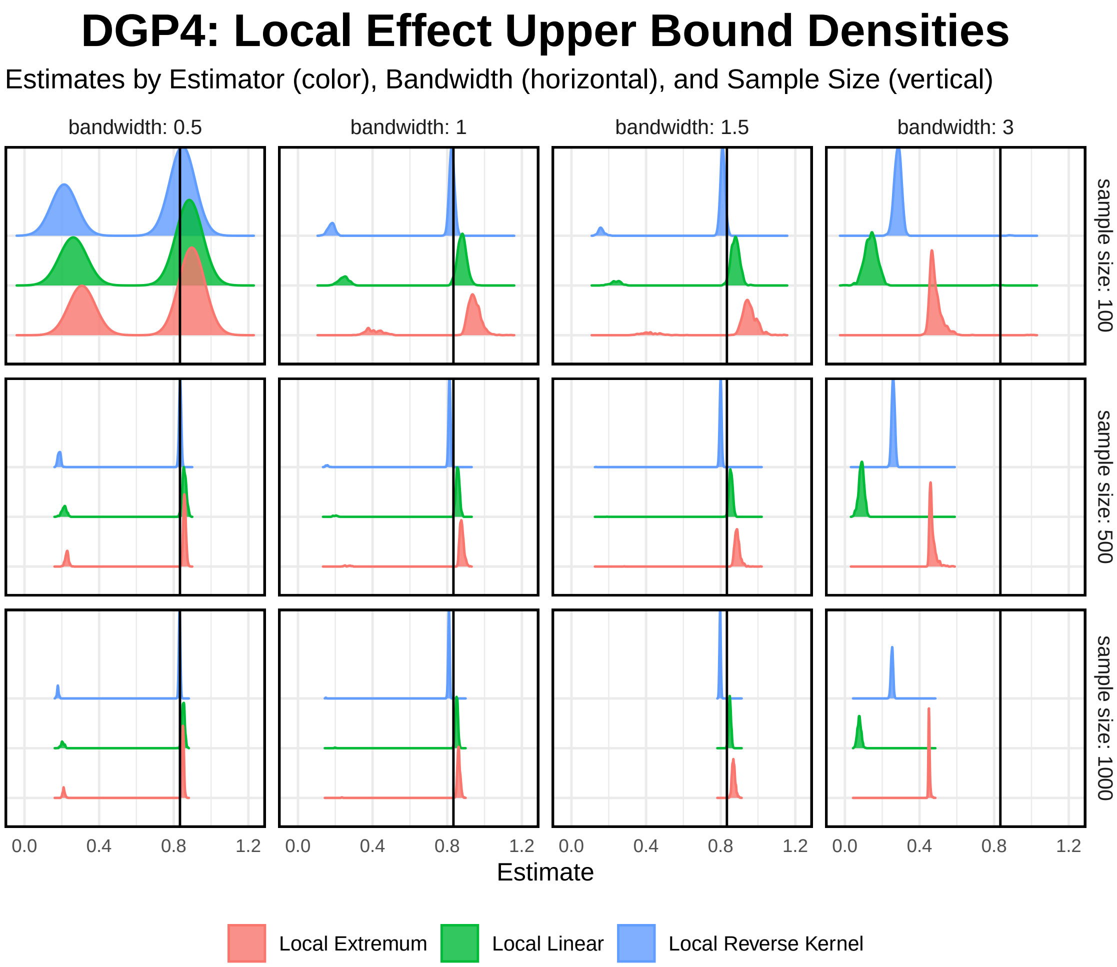



**Appendix Figure E2.** DGP2: Local Effect Upper Bound Simulation Results
*Note:* The true local upper bound is given by the black vertical line in each panel.



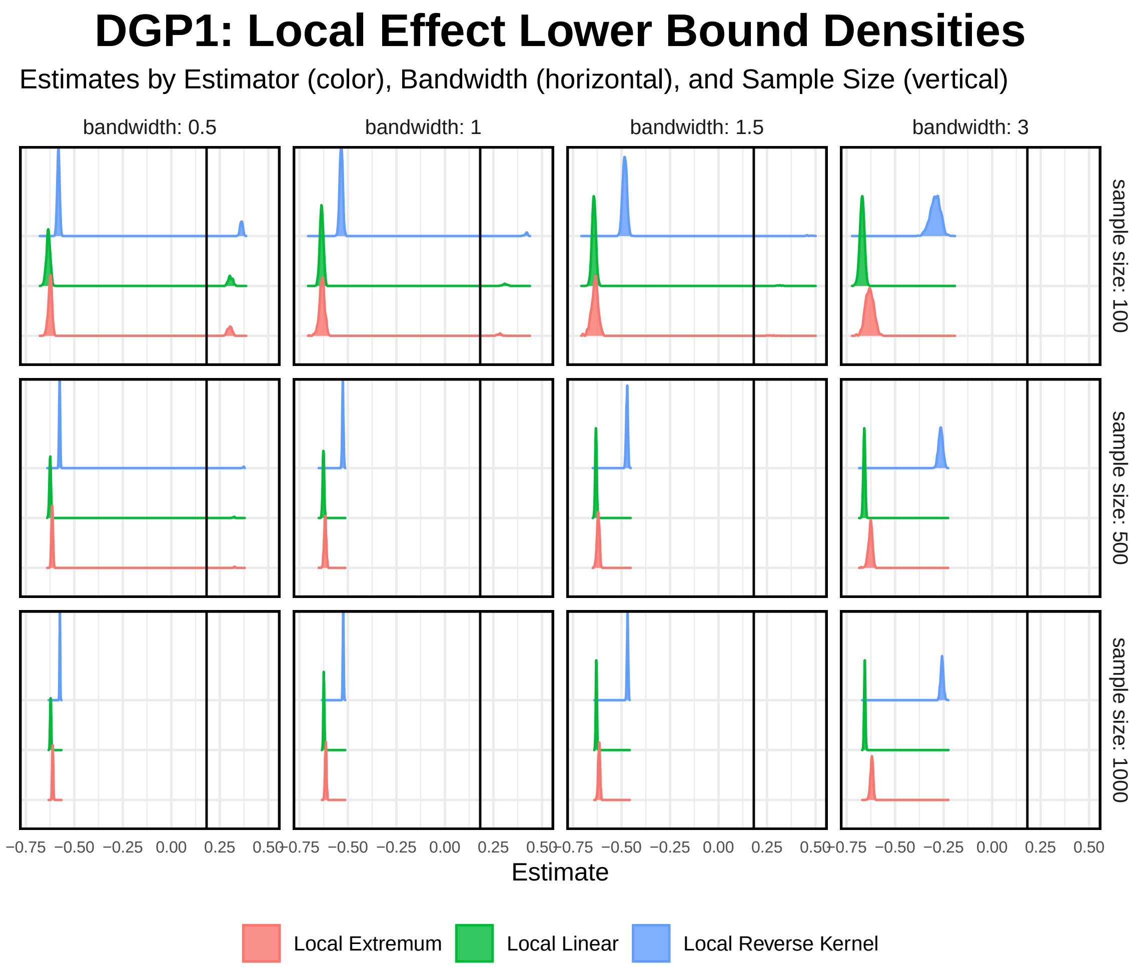

**Appendix Figure E3.** DGP3: Local Effect Upper Bound Simulation Results

*Note:* The true local upper bound is given by the black vertical line in each panel.



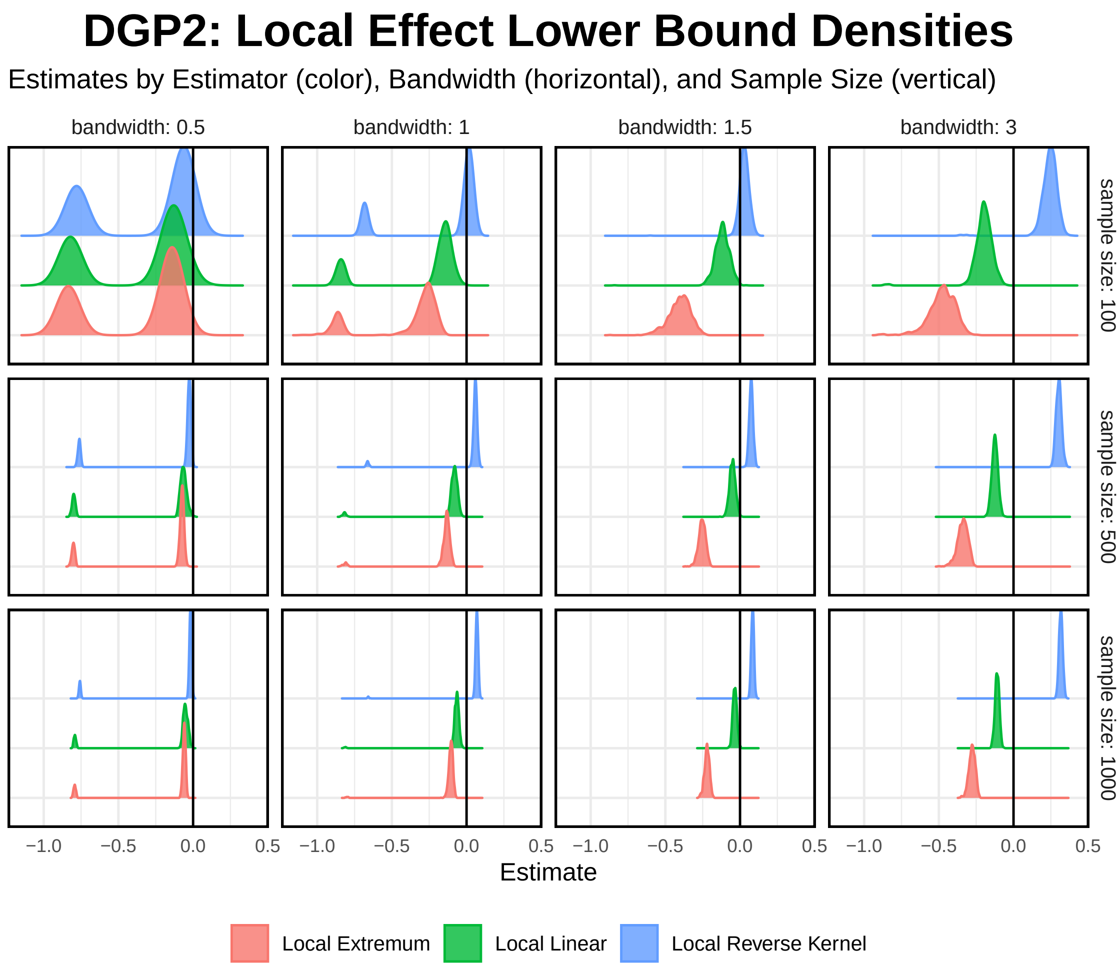

**Appendix Figure E4.** DGP4: Local Effect Upper Bound Simulation Results

*Note:* The true local upper bound is given by the black vertical line in each panel.



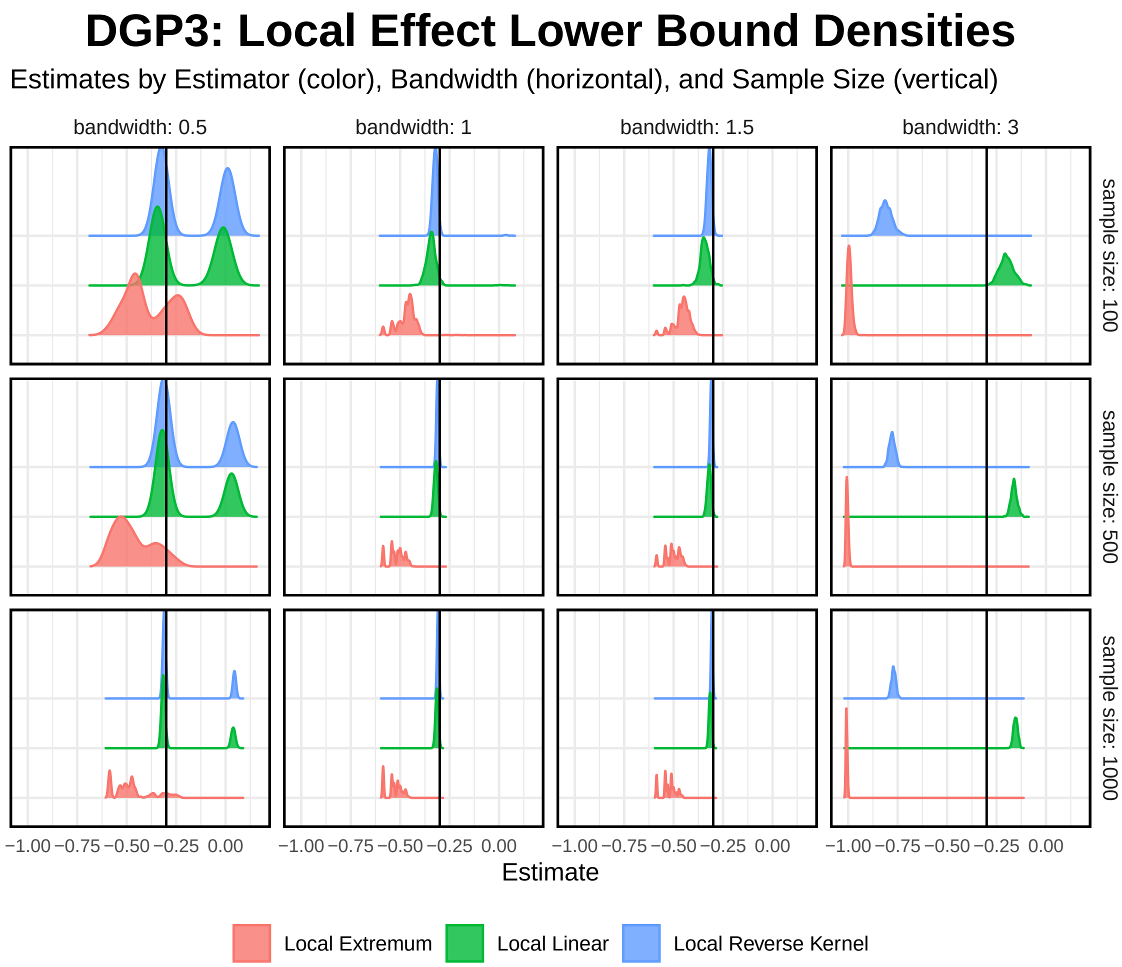

**Appendix Figure E5.** DGP1: Local Effect Lower Bound Simulation Results

*Note:* The true local lower bound is given by the black vertical line in each panel.



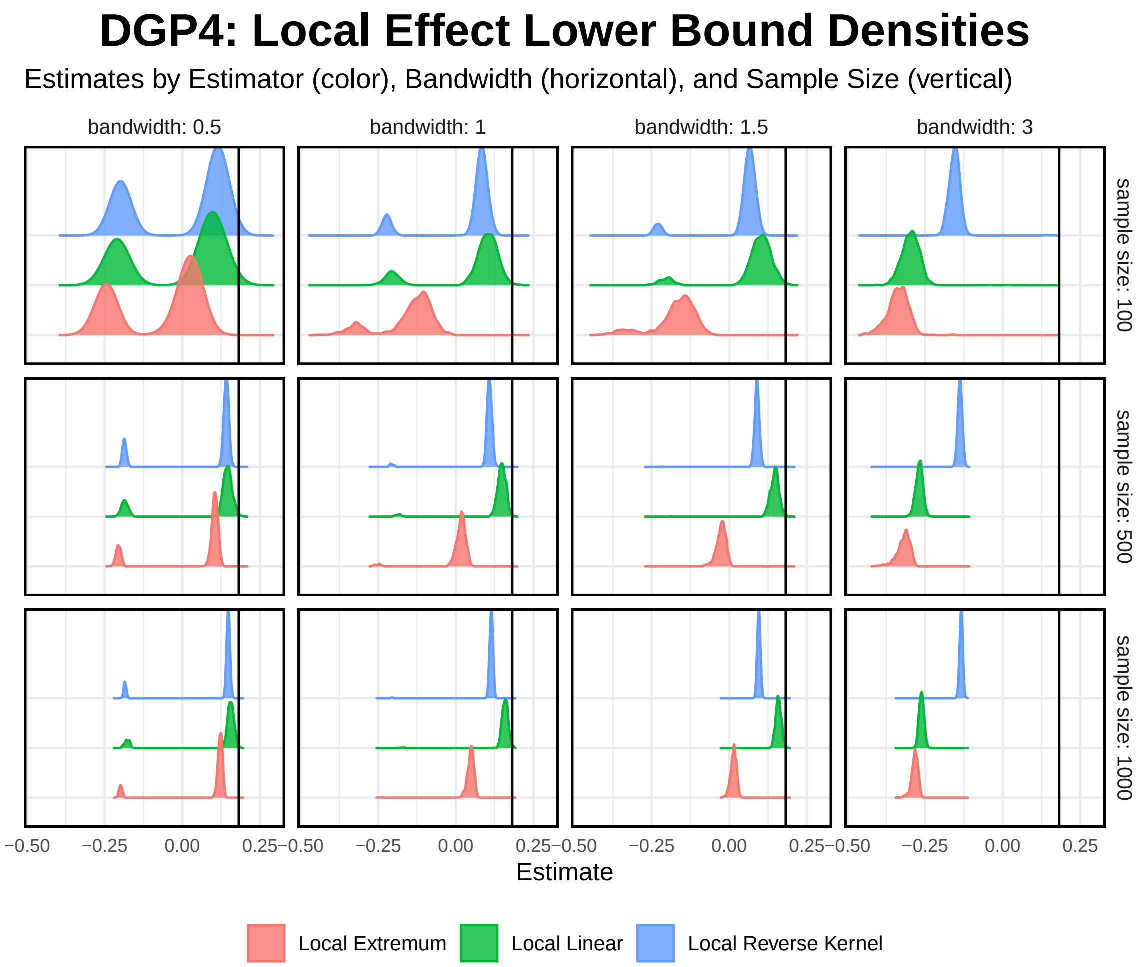

**Appendix Figure E6.** DGP2: Local Effect Lower Bound Simulation Results

*Note:* The true local lower bound is given by the black vertical line in each panel.




**Appendix Figure E7.** DGP3: Local Effect Lower Bound Simulation Results

*Note:* The true local lower bound is given by the black vertical line in each panel.

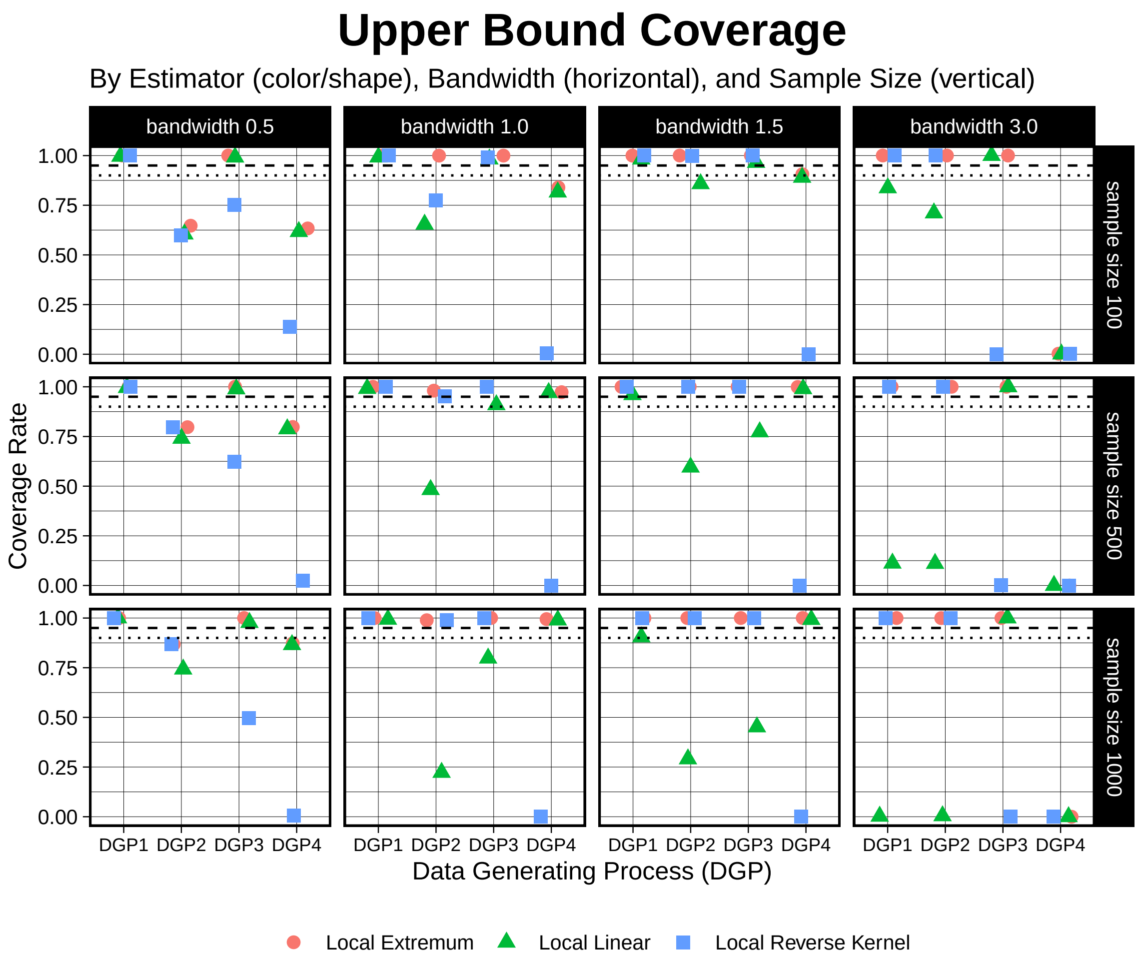



**Appendix Figure E8.** DGP4: Local Effect Lower Bound Simulation Results

*Note:* The true local lower bound is given by the black vertical line in each panel.

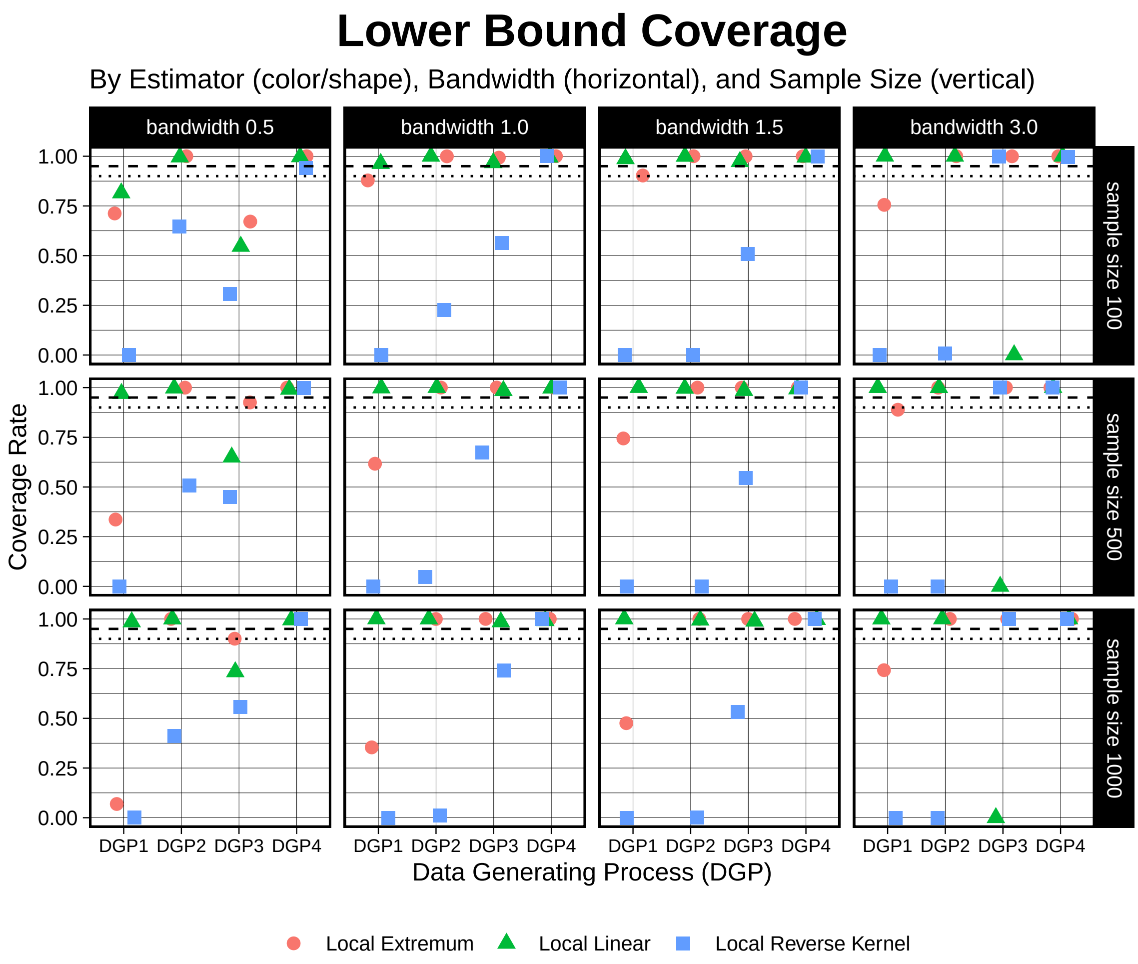
# F Appendix F: Local Simulation Study Coverage Results

Appendix F shows coverage results for the local simulation study in plots rather than a table.



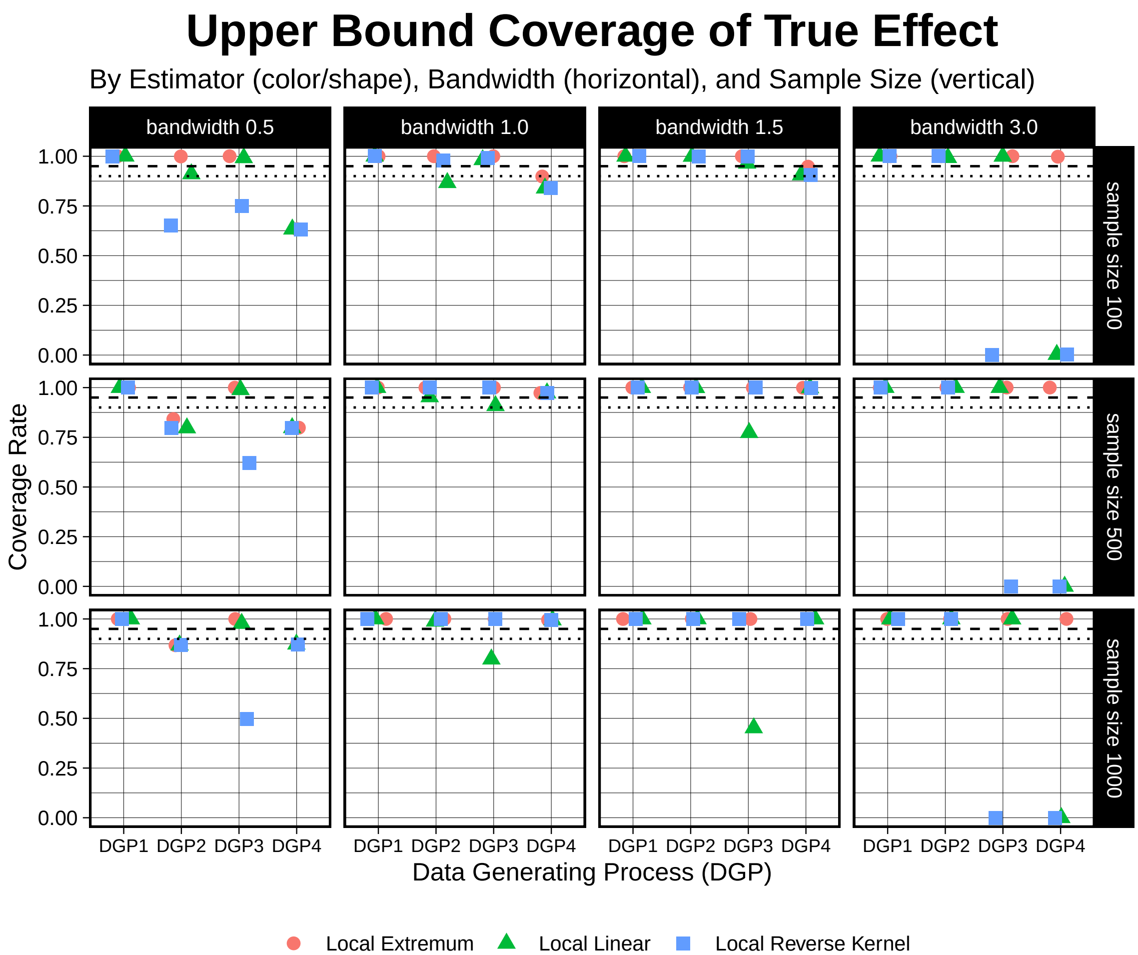

**Appendix Figure F1.** Upper Bound Coverage

*Note:* The coverage rate for the upper bound is the number of simulation runs where estimated upper bound CIs for a given estimator include the true upper bound. For each bandwidth and sample size pair, there were 1000 simulation runs.



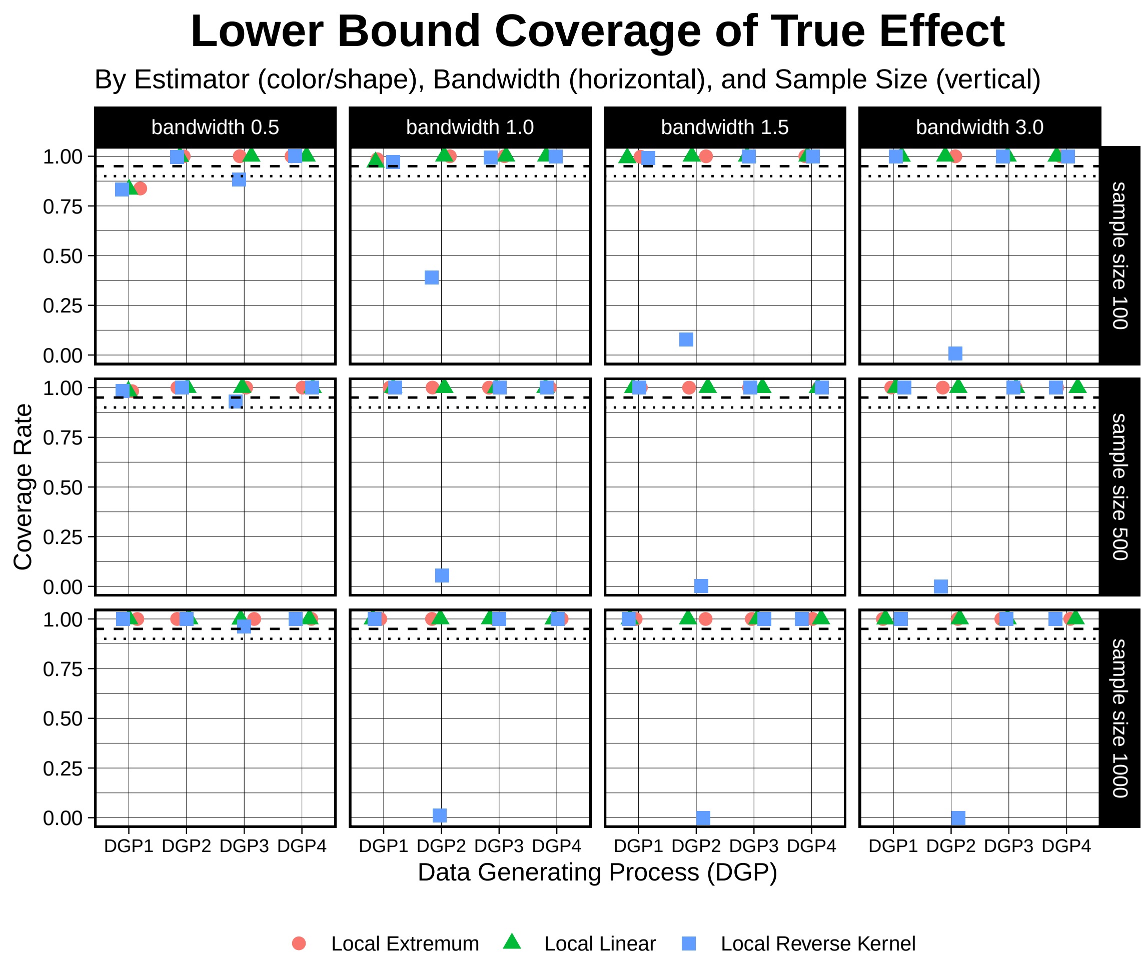

**Appendix Figure F2.** Lower Bound Coverage

*Note:* The coverage rate for the lower bound is the number of simulation runs where estimated lower bound CIs for a given estimator include the true lower bound. For each bandwidth and sample size pair, there were 1000 simulation runs.




**Appendix Figure F3.** Upper Bound Coverage of True Effect

*Note:* The upper bound coverage of the true effect rate is the number of simulation runs where the estimated upper bound CI for a given estimator is greater than or equal to the true treatment effect. For each bandwidth and sample size pair, there were 1000 simulation runs.




**Appendix Figure F4.** Lower Bound Coverage of True Effect

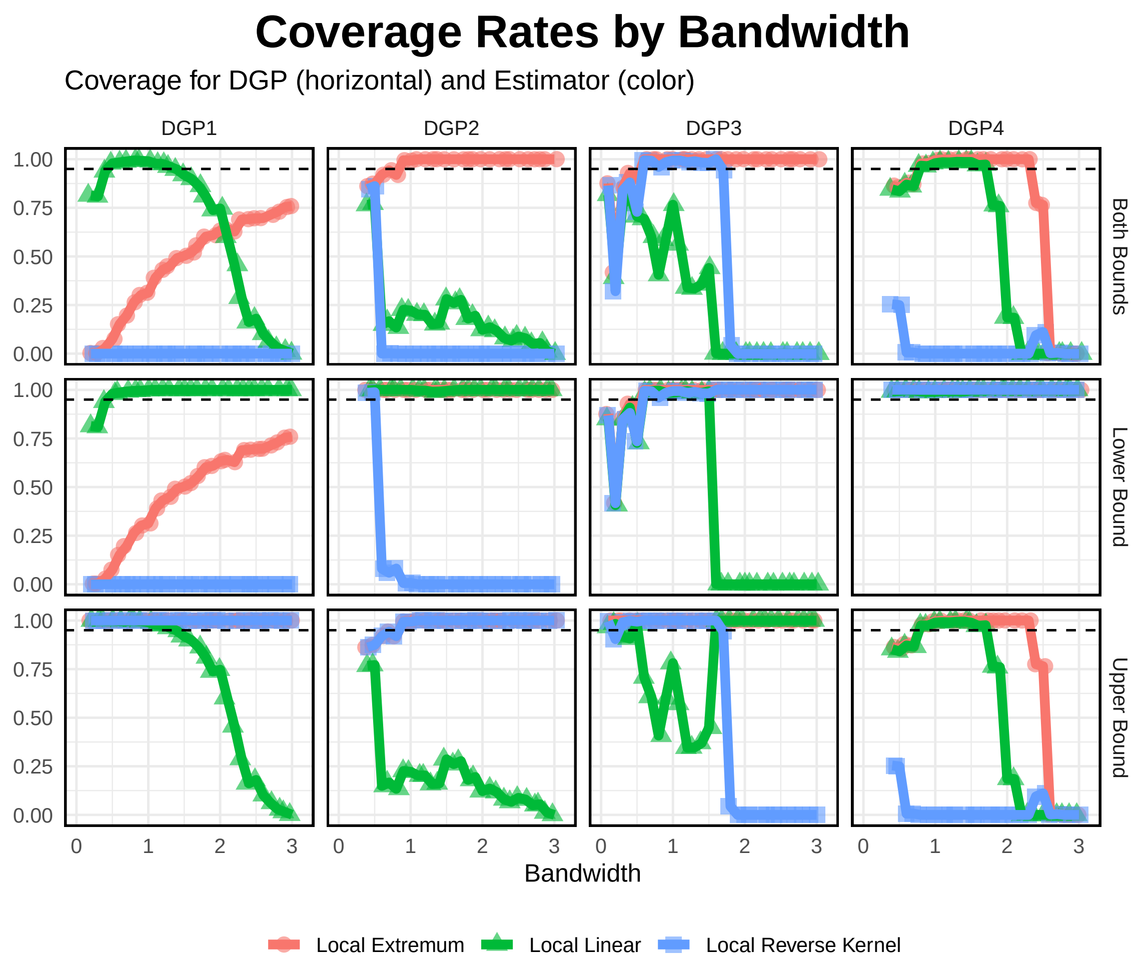
*Note:* The lower bound coverage of the true effect rate is the number of simulation runs where the estimated lower bound CI for a given estimator is less than or equal to the true treatment effect. For each bandwidth and sample size pair, there were 1000 simulation runs.

# G Appendix G: Additional Simulation Results

Appendix G shows additional simulation results, including: baseline simulations with more bandwidths, missing data simulations, and measurement error simulations.

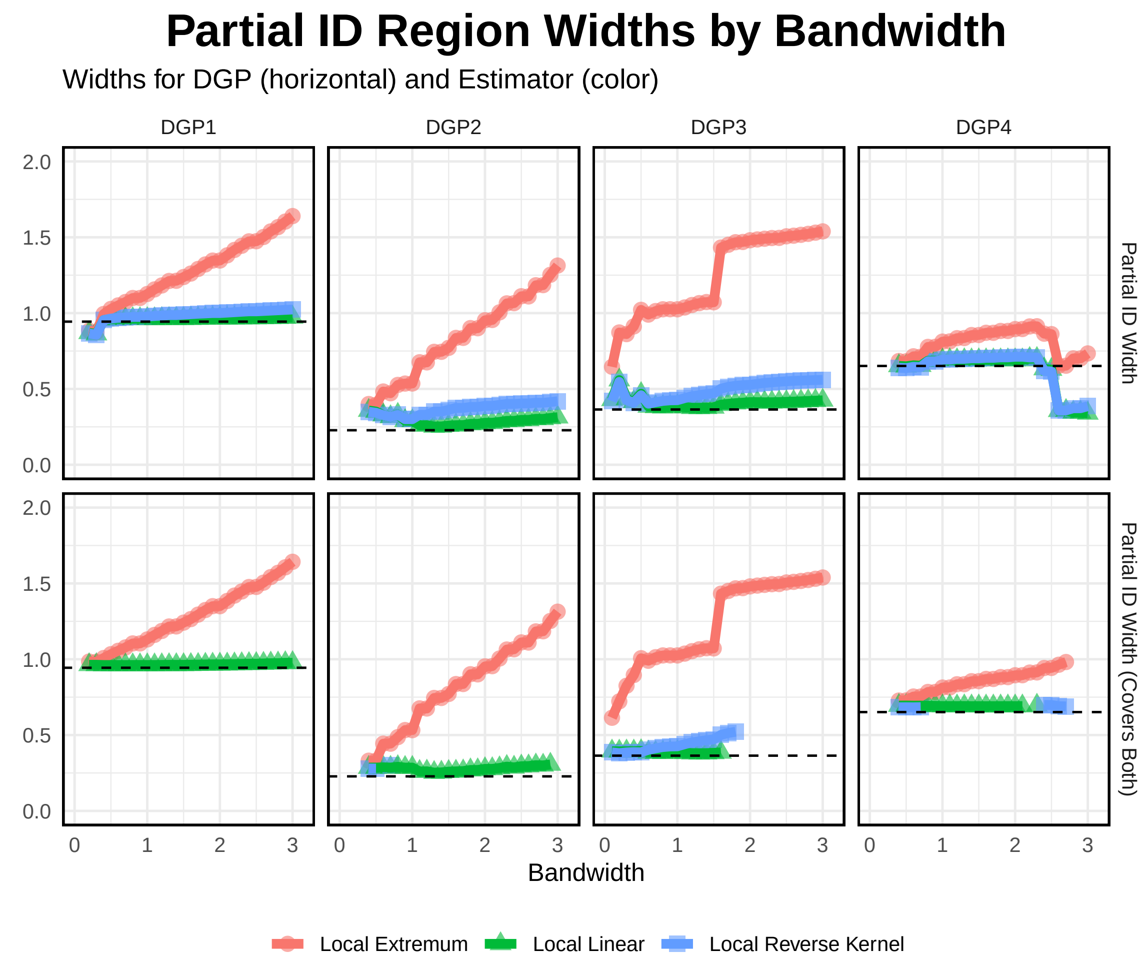
## G.1 More Bandwidths Simulations

The more bandwidths simulation use sample sizes of $n = 1000$ and 1000 simulation runs for each DGP, as specified above. Then I re-run the local effect simulations as before with more bandwidths in 0.10 increments from, e.g., $h = 0.20, 0.30, \ldots, 3.00$ for DGP1. Coverage rates (for upper bounds, lower bounds, and both bounds) and partial ID region widths are shown by estimator, bandwidth, and DGP.




**Appendix Figure F1.** More Bandwidths: Coverage Rates

*Note:* The 95% nominal coverage rate is plotted as the dashed black horizontal line in each panel.

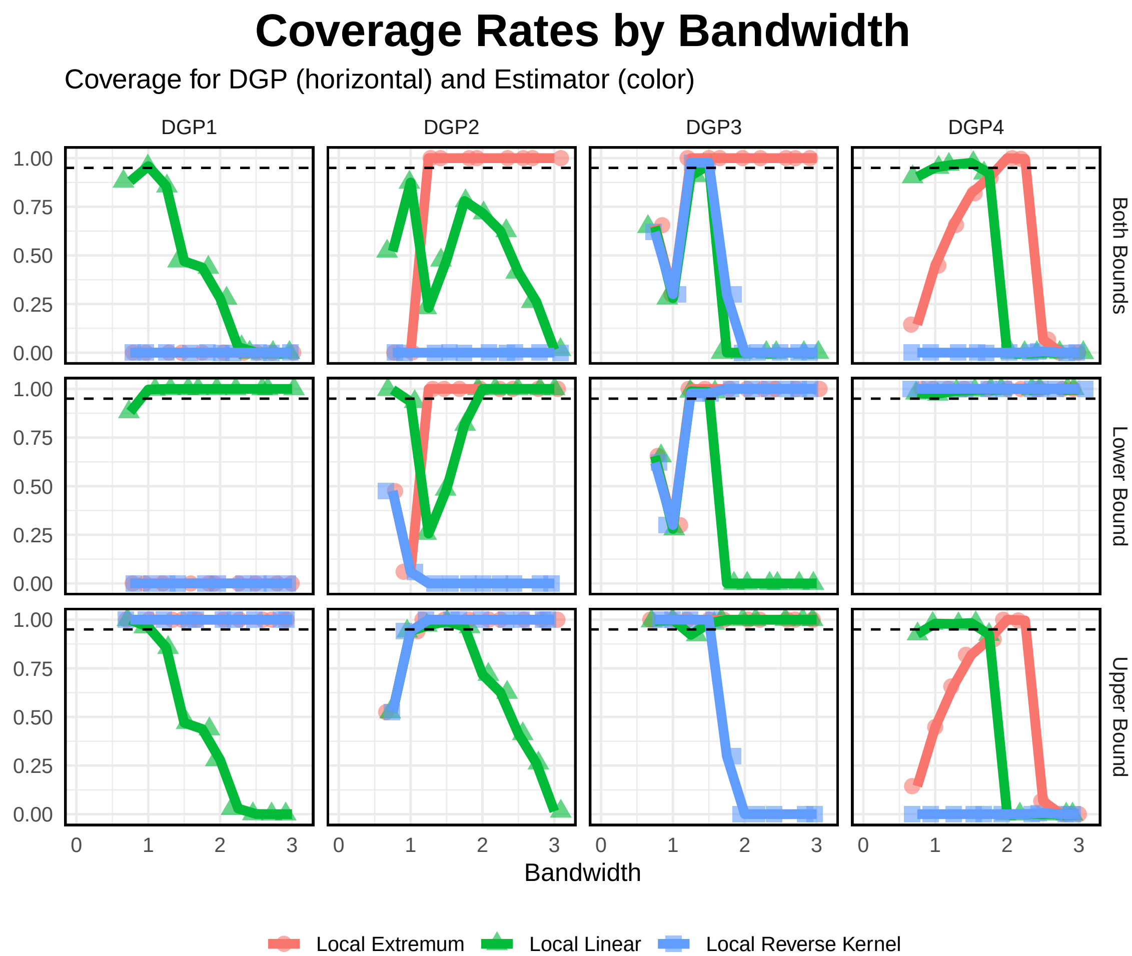



**Appendix Figure F2.** More Bandwidths: Partial ID Region Widths

*Note:* The true partially identified region width is plotted as the dashed black horizontal line in each panel.

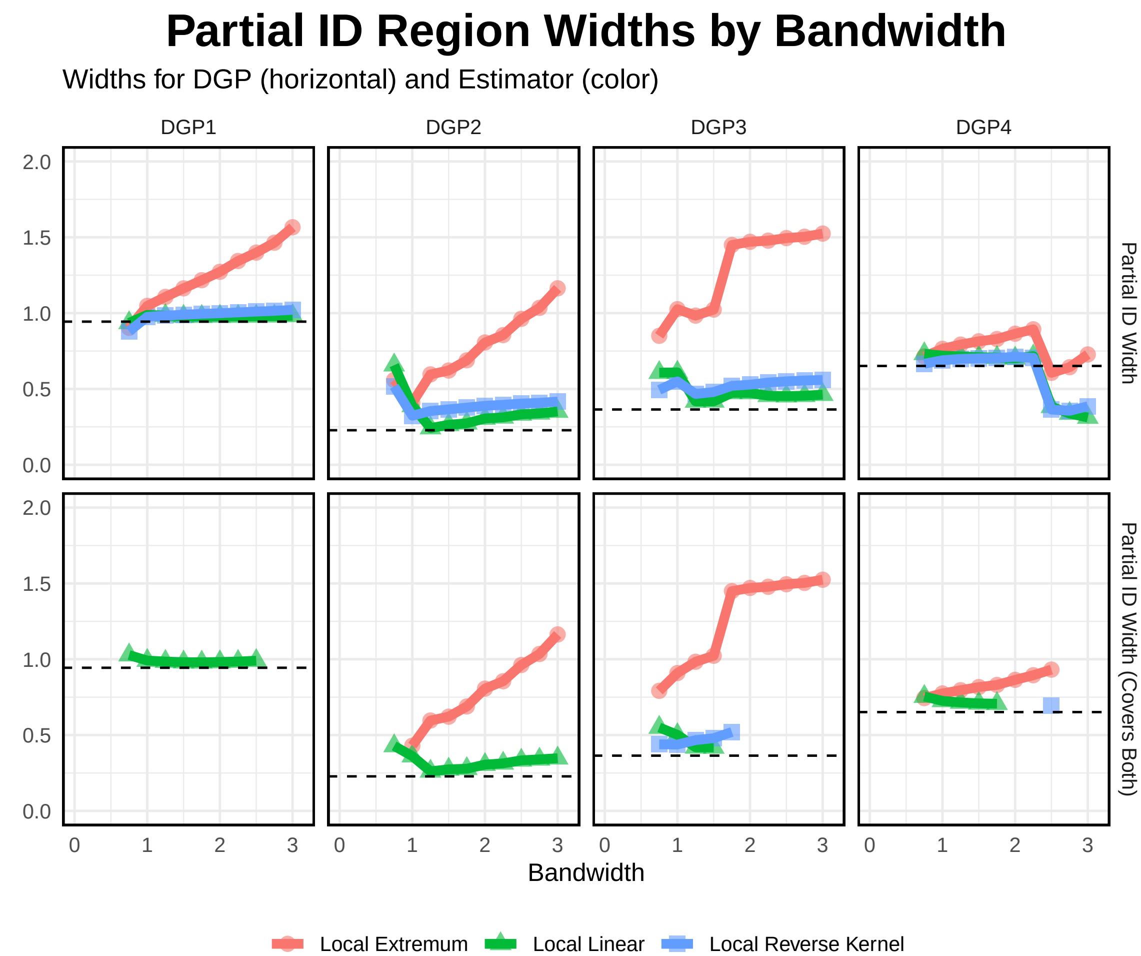
## G.2 Missing Data Simulations

The missing data simulations use sample sizes of $n = 1000$ and 1000 simulation runs for each DGP, as specified above. Then I re-run the local effect simulations removing observations with running variable values near the cutoff: those with $|X - 3| \leq 0.5$ are removed from the dataset. Such missingness may occur either due to true missingness or as a robustness check as in donut RDD designs (Noack and Rothe, 2023). Coverage rates (for upper bounds, lower bounds, and both bounds) and partial ID region widths are shown by estimator, bandwidth, and DGP.




**Appendix Figure F3.** Missing Data: Coverage Rates

*Note:* The 95% nominal coverage rate is plotted as the dashed black horizontal line in each panel.

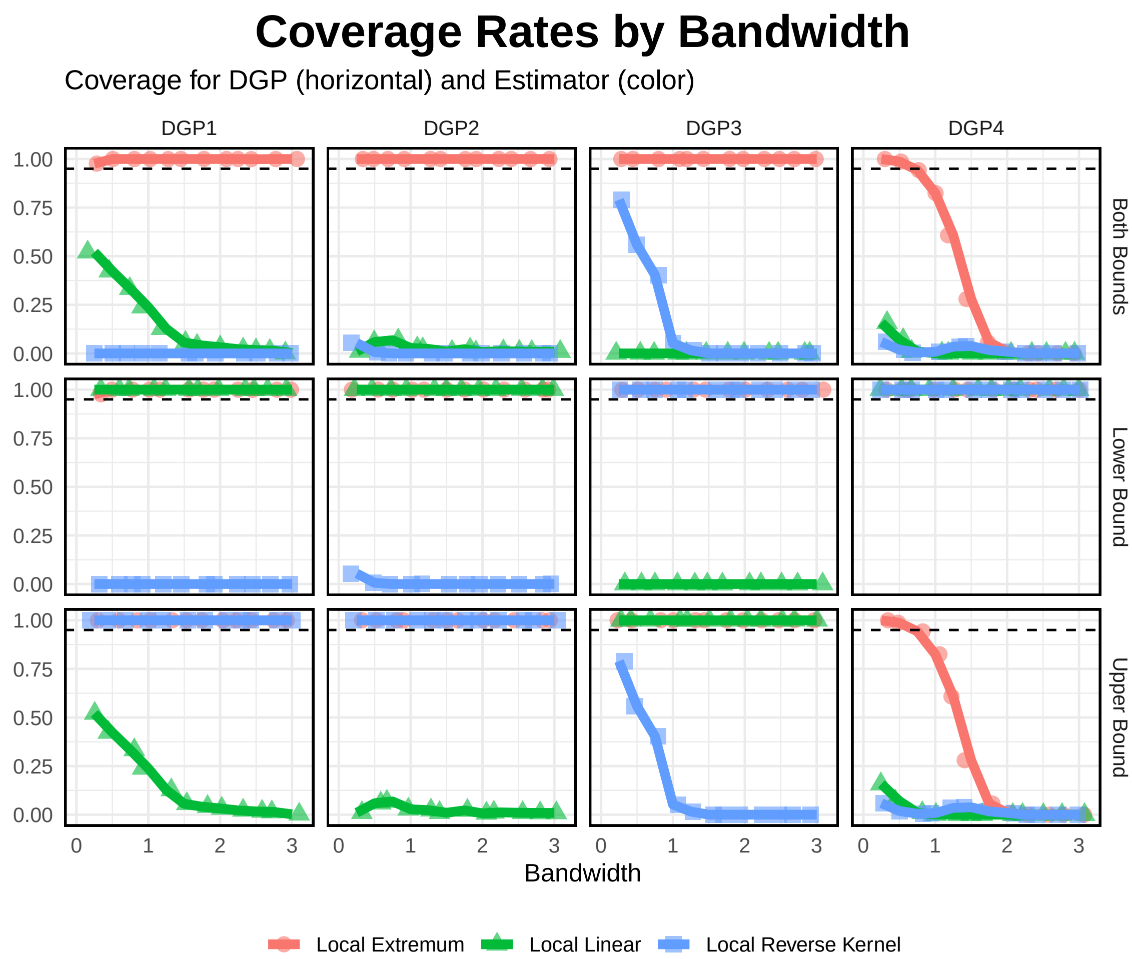



**Appendix Figure F4.** Missing Data: Partial ID Region Widths

*Note:* The true partially identified region width is plotted as the dashed black horizontal line in each panel.

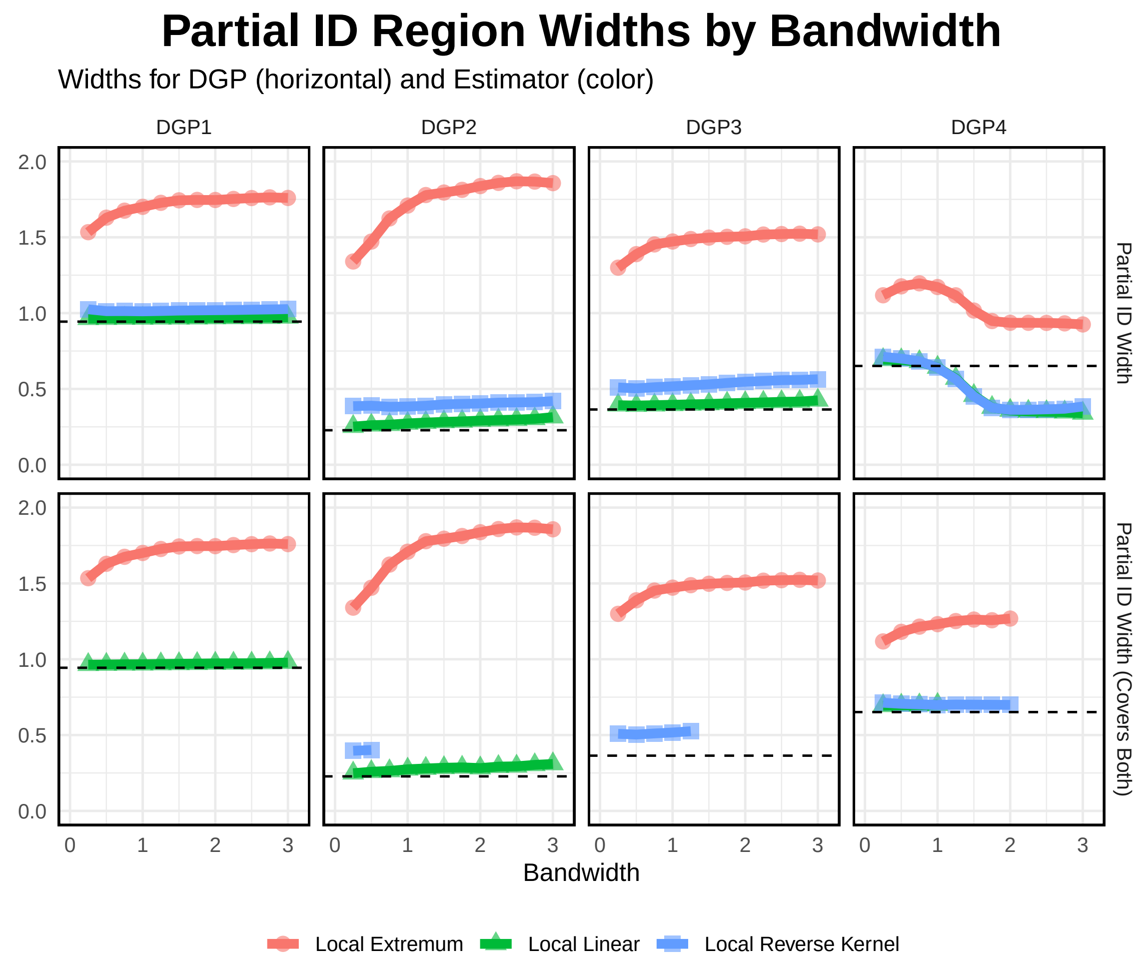
### G.3 Measurement Error Simulations

The measurement error simulations use sample sizes of $n = 1000$ and 1000 simulation runs for each DGP, as specified above. Then I re-run the local effect simulations adding noise (standard normal) and rounding (to nearest two decimals) to each running variable value, while ensuring running variable values lie in $[0, 6]$ and do not shift observations to the other side of the cutoff. This is distortion of the running variable rather than treatment status manipulation. Rounding and classical measurement error may occur in a variety of settings where the running variable is imperfectly reported (by agents or agencies) or estimated by the econometrician. Coverage rates (for upper bounds, lower bounds, and both bounds) and partial ID region widths are shown by estimator, bandwidth, and DGP.




**Appendix Figure F5.** Measurement Error: Coverage Rates

*Note:* The 95% nominal coverage rate is plotted as the dashed black horizontal line in each panel.

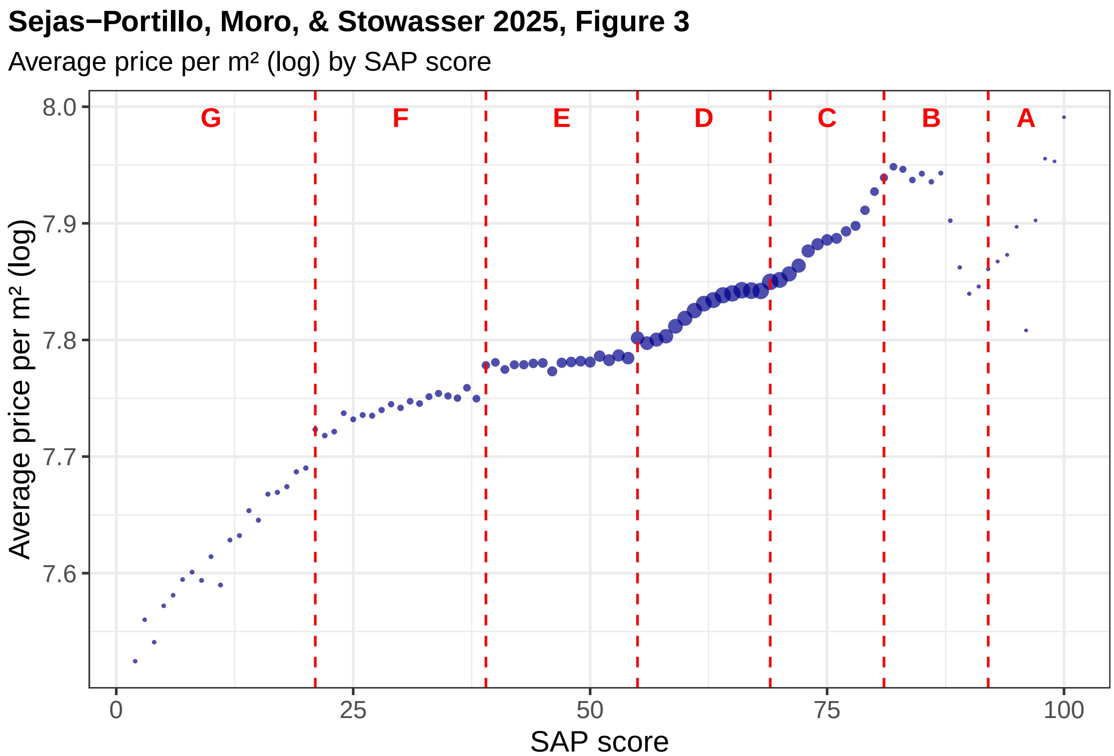



**Appendix Figure F6.** Measurement Error: Partial ID Region Widths

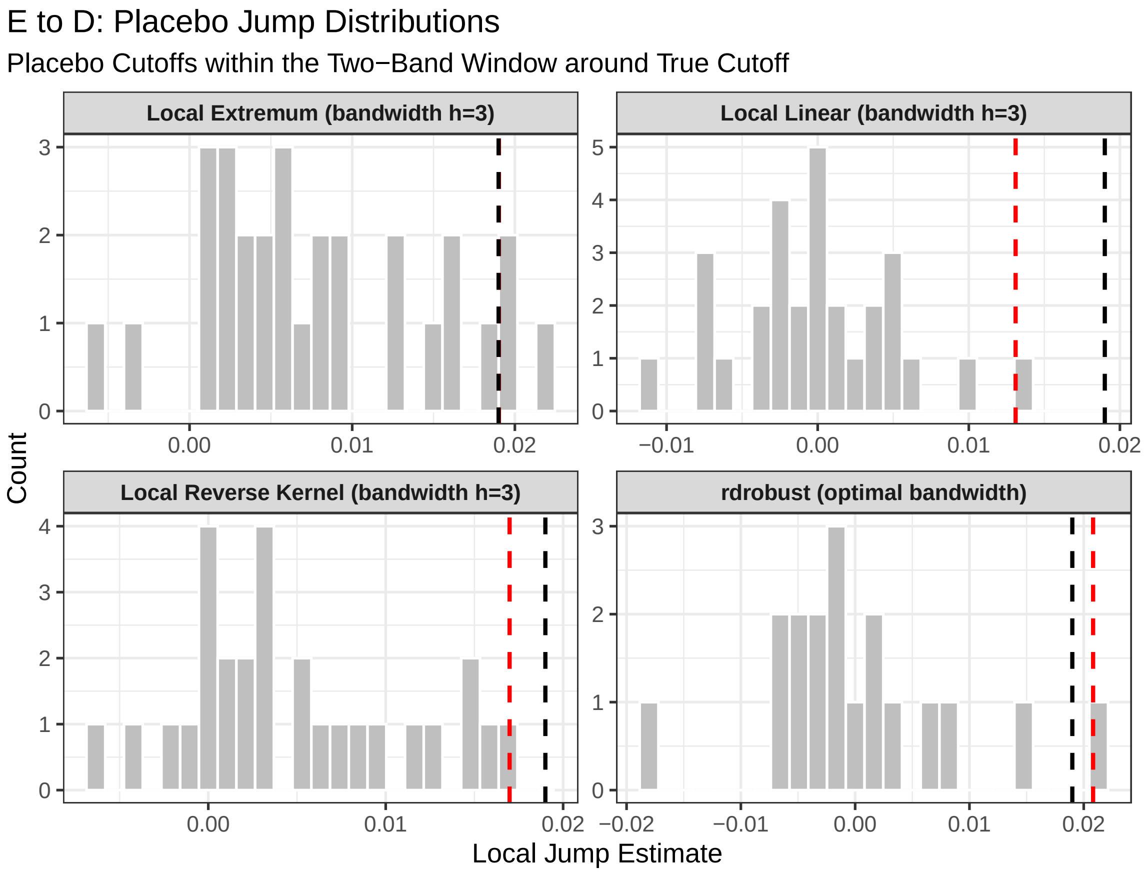
*Note:* The true partially identified region width is plotted as the dashed black horizontal line in each panel.

# H Appendix H: Additional Application Results

Appendix H contains additional application figures and results.

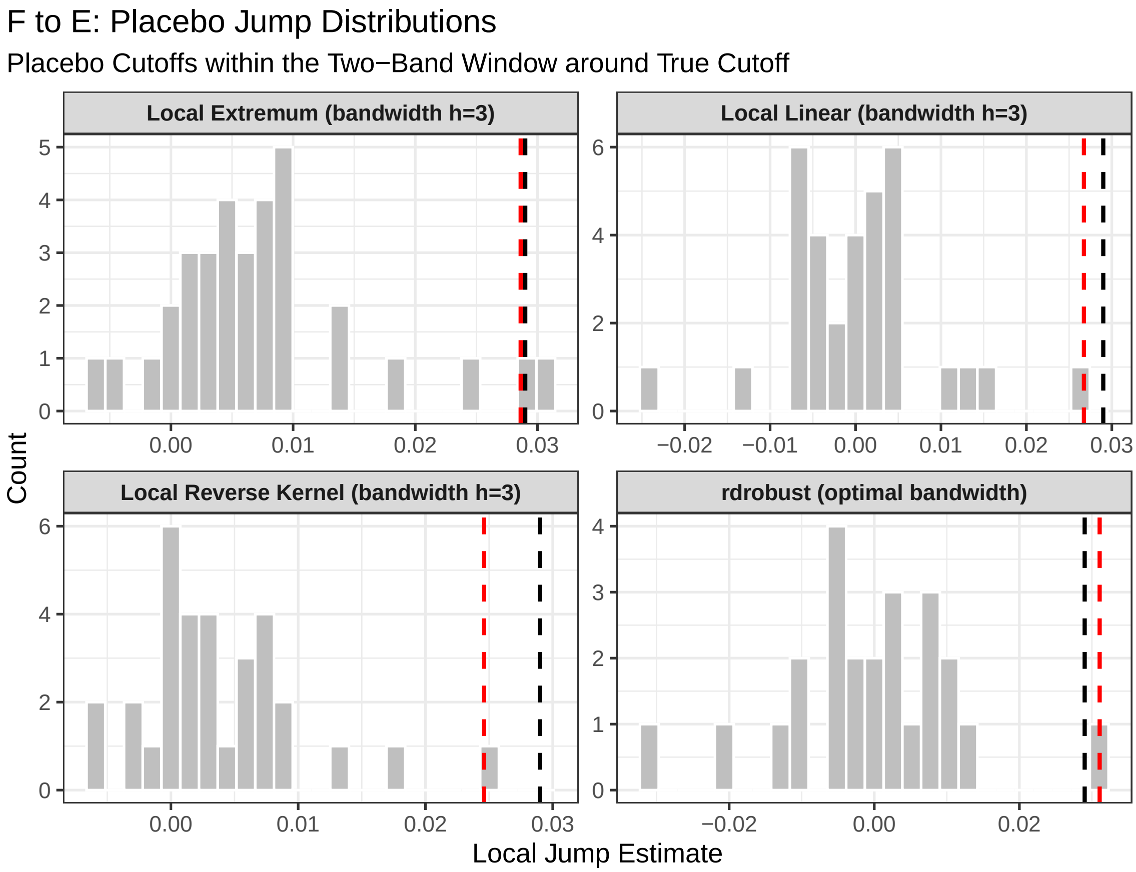



**Appendix Figure F1.** Sejas-Portillo, Moro, and Stowasser 2025, Figure 3 Reproduction



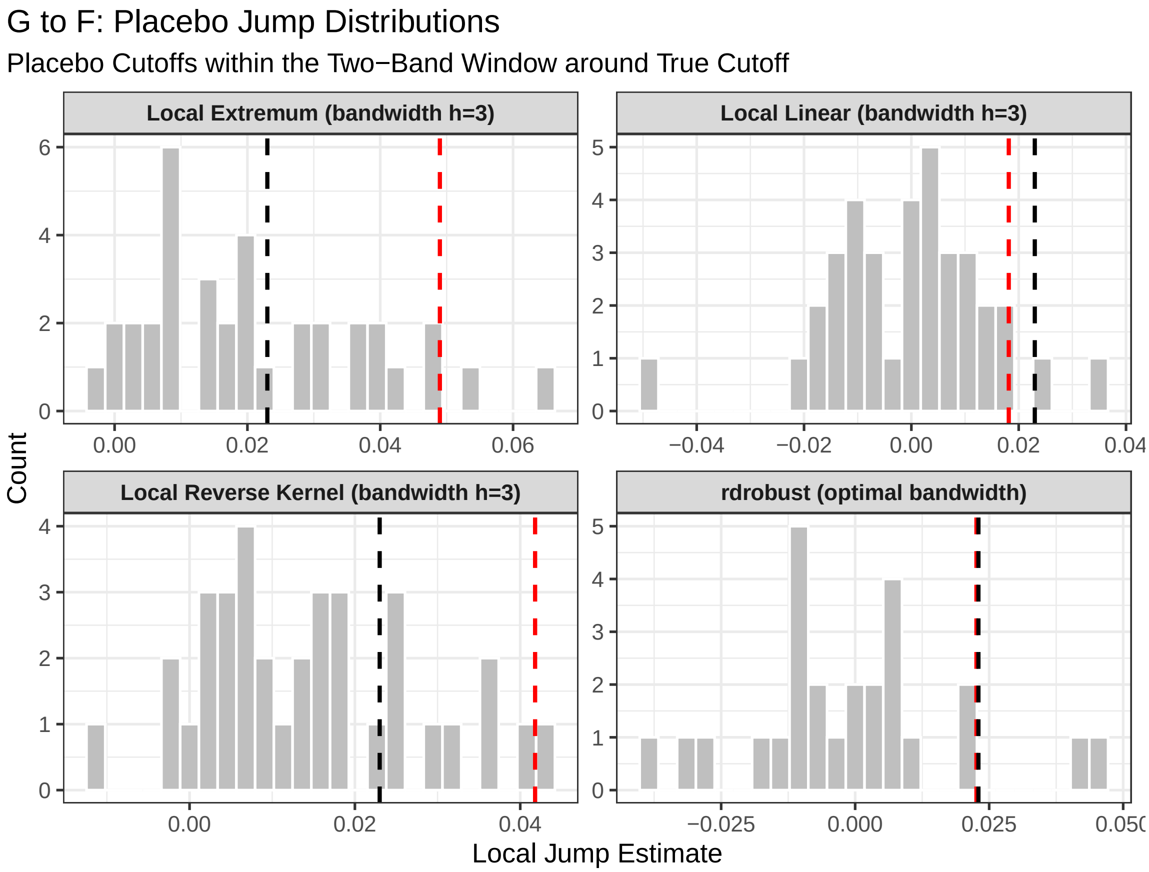

**Appendix Figure F2.** Placebo Jump Distributions near E to D Cutoff

*Note:* Estimated placebo effects at each running variable value near the cutoff (but not beyond the next cutoff) are shown for each estimator. The dashed red line shows the estimated cutoff effect at the true cutoff value, and the dashed black line shows the baseline estimate from SpMS 2025.




**Appendix Figure F3.** Placebo Jump Distributions near F to E Cutoff

*Note:* Estimated placebo effects at each running variable value near the cutoff (but not beyond the next cutoff) are shown for each estimator. The dashed red line shows the estimated cutoff effect at the true cutoff value, and the dashed black line shows the baseline estimate from SpMS 2025.




**Appendix Figure F4.** Placebo Jump Distributions near G to F Cutoff

*Note:* Estimated placebo effects at each running variable value near the cutoff (but not beyond the next cutoff) are shown for each estimator. The dashed red line shows the estimated cutoff effect at the true cutoff value, and the dashed black line shows the baseline estimate from SpMS 2025.

# I Appendix I: Fuzzy RDD Results

Appendix I considers fuzzy regression discontinuity designs. This section illustrates how causal bounds may be obtained under certain counterfactual assumptions. An example visualization and data generating process are also included.

## I.1 Fuzzy Monotone RDD: Partial Identification of ATE under Monotone Treatment Selection (MTS)

In the fuzzy RDD, recall that there are treated and control observations on both sides of a cutoff $C$. Because treatment $W$ is not a deterministic function of $X$, there are therefore four CEFs of interest:

$$\begin{aligned}
\mu_{00}(x) &= \mathbb{E}[Y_0|W = 0, X = x] \\
\mu_{01}(x) &= \mathbb{E}[Y_0|W = 1, X = x] \\
\mu_{10}(x) &= \mathbb{E}[Y_1|W = 0, X = x] \\
\mu_{11}(x) &= \mathbb{E}[Y_1|W = 1, X = x]
\end{aligned}$$

We consider an assumption to extrapolate to the $ATE$ (or $ATE(\Omega)$). Bounds on the local or cutoff average treatment effect may also be obtained in the obvious way.[25] The assumption does not exploit the jump in treatment probability at the cutoff directly, but rather relies on the relationship between the different $\mu_{gg'}$ functions.

Under the following three assumptions, we can similarly bound the $ATE$ by marginalizing and exploiting monotonicity. The first assumption is needed to deal with the fuzziness of the design:

**Assumption 5** (Conditional Monotone Treatment Selection)**.** *For $g = 0, 1$ and $w = 0, 1$ and any $x \in supp(X)$, if $w \geq w'$, then*

$$\mathbb{E}[Y_g|W = w, X = x] \geq \mathbb{E}[Y_g|W = w', X = x]$$

The monotone treatment selection ("MTS") assumption is well-known in the partial identification literature.[26] Here, we only need it to hold conditional on $X = x$, so

[25] In particular, if three assumptions below hold, the bound is the same as that from the sharp RDD case.

[26] See, e.g., Manski (2007).

we call this Conditional Monotone Treatment Selection ("CMTS"). This assumption implies that $\mu_{11}(x) \geq \mu_{10}(x)$ and $\mu_{01}(x) \geq \mu_{00}(x)$ for all $x$-values.

We further assume monotonicity and boundedness assumptions as in the sharp RDD case:

**Assumption 6** (Global Monotonicity of $\mu_{11}, \mu_{00}$)**.** *The conditional mean functions $\mu_{11}(x)$,$\mu_{00}(x)$ are weakly monotone increasing in $x$ for all $x \in supp(X)$.*

**Assumption 7** (Boundedness)**.** *The conditional mean functions $\mu_{gg'}$ for any $g, g' \in \{0, 1\}$ are uniformly bounded above by $M_1$ and below by $M_0$.*

Finally, before proceeding to obtain the sharp upper bound on $ATE$ under these assumptions, define the following sets: let $V_1 = \{v_1, \ldots, v_n\}$ be the set of $x$-values less than the cutoff $C$ for which $Pr(W = 1|X \in V_1) \neq 0$; similarly, let $U_0 = \{u_0, \ldots, u_k\}$ be the set of $x$-values greater than or equal to $C$ for which $Pr(W = 0|X \in U_0) \neq 0$. These are the sets of $x$-values to the left and right of the cutoff $C$, respectively, for which treated and control units are both observed (i.e. overlap is satisfied).[27]

[27] Technically, when $X$ is continuously distributed there may still be a failure of overlap if, for example, we have a point $v_1 < C$ and $Pr(W = 1|X = v_1) = 1$. If there are treated and control units with $X_i = v_1$, then there is overlap (i.e. the probability of treatment assignment is not zero or one.) We can handle this case in the obvious way, but we simplify and assume overlap at all $v \in V_1$ and $u \in U_0$ for notational simplicity.

Consider then the following upper bound on $ATE$ under these assumptions:

$$\begin{aligned}
ATE &= \mathbb{E}[Y_1 - Y_0] \\
&= \sum_{w\in\{0,1\}} \sum_{x\in\text{supp}(X)} \mathbb{E}[Y_1 - Y_0|W = w, X = x]Pr(W = w|X = x)Pr(X = x) \\
&= \sum_{x\in\text{supp}(X)} \mathbb{E}[Y_1 - Y_0|W = 0, X = x]Pr(W = 0|X = x)Pr(X = x) \\
&\quad + \sum_{x\in\text{supp}(X)} \mathbb{E}[Y_1 - Y_0|W = 1, X = x]Pr(W = 1|X = x)Pr(X = x) \\
&\leq \sum_{x\in\text{supp}(X)} \left(\mathbb{E}[Y_1|W = 1, X = x] - \mathbb{E}[Y_0|W = 0, X = x]\right) Pr(W = 0|X = x)Pr(X = x) \\
&\quad + \sum_{x\in\text{supp}(X)} \left(\mathbb{E}[Y_1|W = 1, X = x] - \mathbb{E}[Y_0|W = 0, X = x]\right) Pr(W = 1|X = x)Pr(X = x) \\
&\leq \sum_{x\in\text{supp}(X)} \left(\mathbb{E}[Y_1|W = 1, X = x] - \mathbb{E}[Y_0|W = 0, X = x]\right) Pr(X = x) \\
&\leq \sum_{x\leq v_1} \left(\mathbb{E}[Y_1|W = 1, X = x] - \mathbb{E}[Y_0|W = 0, X = x]\right) Pr(X = x) \\
&\quad + \sum_{v_1<x\leq v_2} \left(\mathbb{E}[Y_1|W = 1, X = x] - \mathbb{E}[Y_0|W = 0, X = x]\right) Pr(X = x) + \cdots \\
&\quad + \sum_{v_n<x<C} \left(\mathbb{E}[Y_1|W = 1, X = x] - \mathbb{E}[Y_0|W = 0, X = x]\right) Pr(X = x) \\
&\quad + \sum_{C\leq x\leq u_0} \left(\mathbb{E}[Y_1|W = 1, X = x] - \mathbb{E}[Y_0|W = 0, X = x]\right) Pr(X = x) \\
&\quad + \sum_{u_0<x\leq u_1} \left(\mathbb{E}[Y_1|W = 1, X = x] - \mathbb{E}[Y_0|W = 0, X = x]\right) Pr(X = x) + \cdots \\
&\quad + \sum_{u_k\leq x} \left(\mathbb{E}[Y_1|W = 1, X = x] - \mathbb{E}[Y_0|W = 0, X = x]\right) Pr(X = x) \\
&\leq \sum_{x\leq v_1} \left(\mu_{11}(v_1) - \mu_{00}(x)\right) Pr(X = x) + \cdots + \sum_{v_n<x<C} \left(\mu_{11}(C) - \mu_{00}(x)\right) Pr(X = x) \\
&\quad + \sum_{C\leq x\leq u_0} \left(\mu_{11}(x) - \lim_{x\to C^-} \mu_{00}(x)\right) Pr(X = x) + \cdots + \sum_{u_k\leq x} \left(\mu_{11}(x) - \mu_{00}(u_k)\right) Pr(X = x)
\end{aligned}$$

In the first inequality, we use Assumption 5; in the second inequality, we grouped like terms and used $Pr(W = 0|X = x) + Pr(W = 1|X = x) = 1$; in the third inequality, we split the support into $x$-values split by the values in $V_1, U_0$; finally, in the fourth inequality we use the observed conditional means for $x$-values in sets

$V_1, U_0$ and Assumption 6 to get bounds on the conditional means $\mu_{11}, \mu_{00}$ for all $x \in \text{supp}(X)$. What remains in the final line is an estimable upper bound for $ATE$ under global monotonicity of $\mu_{11}$ and $\mu_{00}$ using observable data alone.

Obtaining the lower bound is similar but requires also using Assumption 7.

### I.2 Fuzzy RDD Example

To reinforce ideas, we provide an example for the fuzzy RDD upper bound estimation on the following page. We assume that $Pr(X = x)$ is uniformly distributed. In our simulated dataset, $X$ is a discrete distribution with 100 different $x$-values, but for computing the exact ATE we will assume $X$ is continuously uniformly distributed over $[0, 6]$. The treatment assignment is as follows:

$$Pr(W = 0|X = x) = \begin{cases} 1 & \text{if } 0 \leq x < 2 \\ \frac{1}{2} & \text{if } x = 2 \\ 1 & \text{if } 2 < x < 3 \\ 0 & \text{if } 3 \leq x < 4 \\ \frac{1}{2} & \text{if } x = 4 \\ 0 & \text{if } 4 \leq x \leq 6 \end{cases}$$

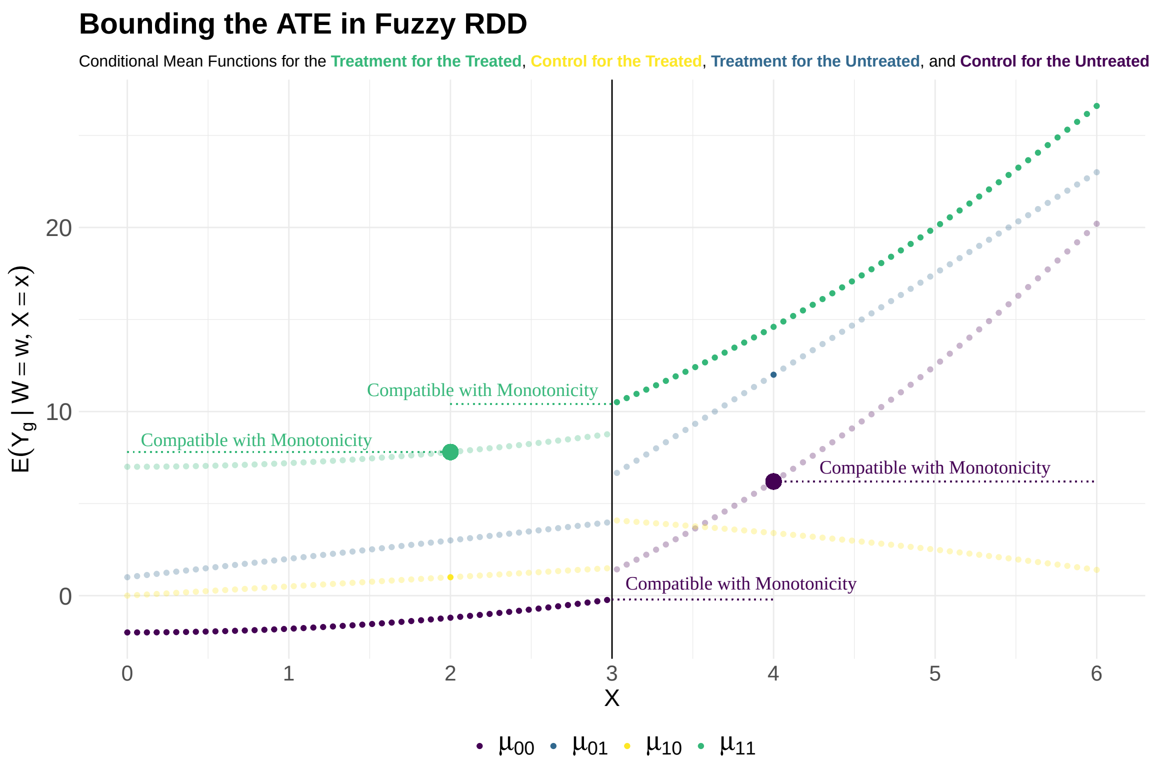
This means that there is overlap at $x = 2$ and $x = 4$, but otherwise units with $X_i < 3$ are control units and units with $X_i \geq 3$ are treated units.

The four functions that define the CEFs are as follows:

$$\mu_{00}(x) = \begin{cases} 0.2x^2 - 2 & \text{if } 0 \leq x < 3 \\ 0.7x^2 - 5 & \text{if } 3 \leq x \leq 6 \end{cases} \qquad \mu_{01}(x) = \begin{cases} x + 1 & \text{if } 0 \leq x < 3 \\ 5.5x - 10 & \text{if } 3 \leq x \leq 6 \end{cases}$$

$$\mu_{10}(x) = \begin{cases} 0.5x & \text{if } 0 \leq x < 3 \\ -0.1x^2 + 5 & \text{if } 3 \leq x \leq 6 \end{cases} \qquad \mu_{11}(x) = \begin{cases} 0.2x^2 + 7 & \text{if } 0 \leq x < 3 \\ 0.6x^2 + 5 & \text{if } 3 \leq x \leq 6 \end{cases}$$

Observe that all four CEFs are discontinuous at the cutoff $C = 3$. This is plotted in the figure below:




**Appendix Figure I1.** Bounding ATE in Fuzzy RDD Under Global Monotonicity

*Note:* This plot shows the four conditional mean functions in the fuzzy discontinuity design, with the lower transparency points unobserved. The two large points are for observed treated units to the left of the cutoff and observed control units to the right of the cutoff, at 2 and 4, respectively. The true treatment effect at the boundary is 3.9, whereas the upper bound is 10.75. The true ATE is 2.49, whereas the upper bound is approximately 11.81.

In the plot, the CMTS assumption is satisfied because the green points lie above the yellow points, and the purple points lie below the blue points. The monotonicity assumption holds for $\mu_{11}(x)$ and $\mu_{00}(x)$ on the entire support of $X$, but observe that in this example $\mu_{10}(x)$ is not monotone increasing on the entire domain. Recall that we do not require that the unobservable functions $\mu_{10}(x)$ and $\mu_{01}(x)$ be monotone.

In this particular example, the true treatment effect at the cutoff is 3.9, which can be found analytically as follows for the cutoff $C = 3$:

$$\begin{aligned}
\tau_C &= \mathbb{E}[Y_1 - Y_0 | X = C] \\
&= \mathbb{E}[Y_1 - Y_0 | W = 0, X = C] \underbrace{Pr(W = 0 | X = C)}_{=0} + \mathbb{E}[Y_1 - Y_0 | W = 1, X = C] \underbrace{Pr(W = 1 | X = C)}_{=1} \\
&= \mathbb{E}[Y_1 | W = 1, X = C] - \mathbb{E}[Y_0 | W = 1, X = C] \\
&= \mu_{11}(C) - \mu_{01}(C) \\
&= (0.6 * (3)^2 + 5) - (5.5 * 3 - 10) \\
&= \boxed{3.9}
\end{aligned}$$

The upper bound at the cutoff is found by using that $\mu_{01}(C) \geq \mu_{00}(C)$ and then replacing $C$ with the last observable value of $x$ for control units to the left of the cutoff. Similarly, since we don't observe $\mu_{11}(C)$ but we do observe $\mu_{11}(3.03)$ we will use that for bounding. In this case, this results in the following upper bound:

$$\begin{aligned}
\overline{\tau}_C &= \mu_{11}(C) - \mu_{01}(C) \\
&\leq \mu_{11}(3.03) - \mu_{00}(2.97) = 10.51 - (-0.24) = \boxed{10.75}
\end{aligned}$$

The ATE is given by the following computation (for $X$ continuously uniformly distributed on $[0, 6]$ so that $Pr(X = x) = \frac{1}{6}$ for all $x$):

$$\begin{aligned}
ATE &= \frac{1}{6}\left(\int_{x=0}^{3} (\mu_{10}(x) - \mu_{00}(x))dx + \int_{x=3}^{6} \mu_{11}(x) - \mu_{01}(x)dx\right) \\
&= \frac{1}{6}(6.45 + 8.55) \\
&= \boxed{2.5}
\end{aligned}$$

We are using that $Pr(W = 0 | X = x)$ is zero or one except at two points, which doesn't matter for the integration. In the caption below the plot, we note that ATE is 2.49 for the dataset generated from the $\mu_{gg'}(x)$ functions with 100 $x$-values evenly spaced. Finally, we also compute the upper bound on ATE from the generated dataset and get an upper bound of

$$\overline{ATE} = \boxed{11.81}$$