Gaussian Process Regression Discontinuity*

Joseph T. Ornstein†
JBrandon Duck-Mayr‡

July 7, 2021

Abstract

In applied settings, regression discontinuity (RD) designs often suffer from noisy data and low power. This tends to produce exaggerated causal effect estimates, typified by implausibly large slope and/or concavity parameters. We introduce a new method for estimating causal effects in RD designs called Gaussian Process Regression Discontinuity (GPRD). This approach overcomes the major disadvantages of global polynomial estimators and does so with lower variance than local linear estimators. When applied to a large set of empirical examples from the published literature, GPRD yields more modest and plausible treatment effect estimates. We make this new method available through the `R` package `gprd`.

*Working paper. Thanks to Danny Hidalgo, Jacob Montgomery, and participants at the International Methods Colloquium and WashU Data Lab for helpful comments on earlier drafts.

†Assistant Professor of Political Science, University of Georgia
‡PhD Candidate in Political Science, Washington University in St. Louis

1 Introduction

Regression discontinuity (RD) is an approach to causal inference that leverages a discontinuous change in treatment at a cutoff. The key identification assumption of the RD design is the continuity of other pre-treatment covariates; so long as treatment status is the only variable that changes discontinuously at the cutoff, the causal effect of treatment is identifiable at that point (Hahn, Todd and Van Der Klaauw, 2001). Because such thresholds, cutoffs, and boundaries are a common feature of political institutions, RD has proven a popular research design in political science over the past two decades (de la Cuesta and Imai, 2016).

Estimating the average treatment effect at the cutoff requires finding the limits of the outcome variable as it approaches from the left and right. Traditionally, researchers have estimated these limits in one of two ways. The first approach is to fit a high-order polynomial regression on either side of the cutoff, then take the difference between the two regressions' predictions at the cutoff. This approach suffers from a disadvantage common to any parametric estimation strategy: if the true data generating process is not captured by the researcher's model specification, then any causal effect estimate is likely to be biased. Gelman and Imbens (2019) catalogue three other disadvantages of this approach: global methods tend to overfit to observations far away from the cutoff, estimates are sensitive to the researcher's choice of polynomial degree, and confidence intervals have poor coverage.

In response to these problems, Calonico, Cattaneo and Titiunik (2014), hereafter CCT, developed a local linear regression approach to RD estimation that keeps observations within a bandwidth (h) that minimizes the mean squared error of the RD estimator. Their approach estimates the limits approaching the cutoff using local linear regression weighted by a triangular kernel and adjusting for a bias-correction term, and the authors derive robust standard errors for inference. Confidence intervals from this method produce the best empirical coverage of any method proposed to date, and its estimates perform well in a wide array of simulations.

Because the CCT estimator relies on a smaller number of observations close to the cutoff, it reduces bias at the expense of higher variance. As a result, datasets that are sparse or noisy in the neighborhood of the cutoff can yield estimates that overfit to chance patterns near the cutoff. As the ongoing replication crisis in experimental sciences has demonstrated, such low-powered studies are pernicious when combined with a publication bias towards statistically significant results (Button et al., 2013). The estimated treatment effects from published low-powered studies tend to significantly overestimate the true treatment effect, a distortion that Gelman and Carlin (2014) refer to as Type M errors.

To illustrate this problem in the RD case, consider the simulated data in Figure 1, distributed $X \sim U(-1, 1)$, $Y \sim \mathcal{N}(0, \sigma) + \mathbf{1}(X > 0)\tau$, and $\tau = 0$.

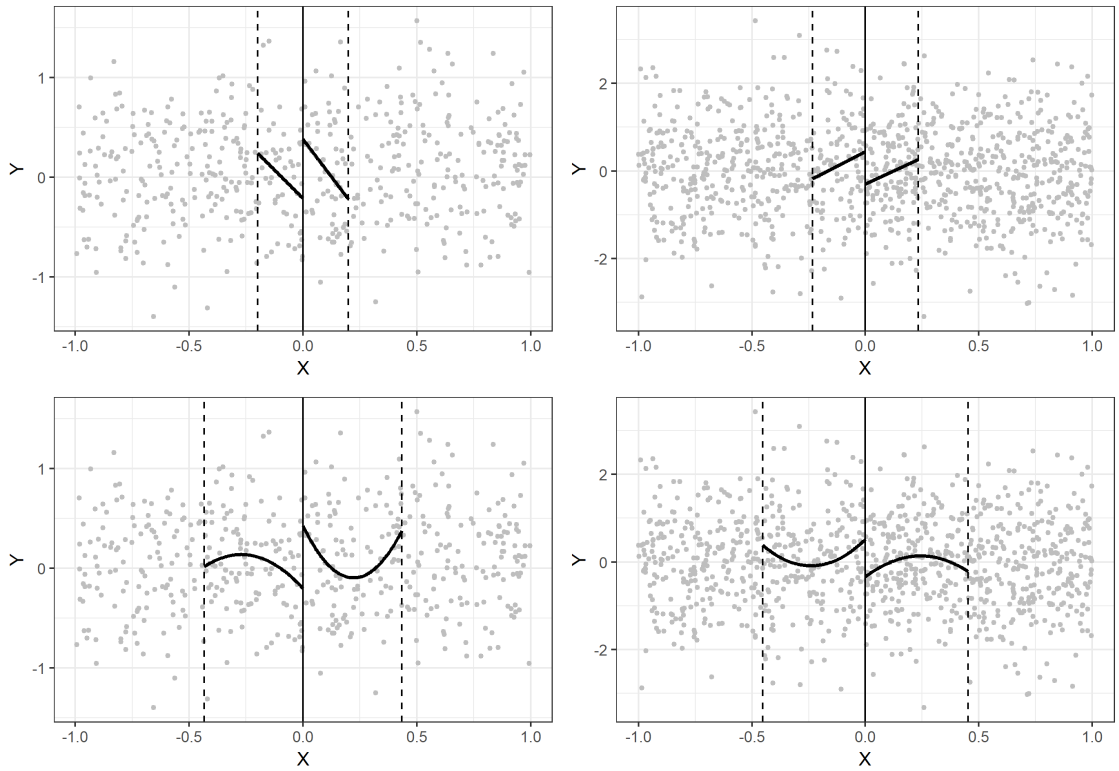


Figure 1: Simulated false positive RD estimates from local polynomial approaches. Dashed vertical lines are the CCT MSE-optimal bandwidth. Solid vertical line is the cutoff. Gray points are the raw data, and black lines are the local polynomial fits. Left two figures generated with parameters $\tau = 0$, $\sigma = \frac{1}{2}$, $n = 500$. Right two figures generated with parameters $\tau = 0$, $\sigma = 1$, $n = 1000$.

Although there is no simulated discontinuity at the cutoff, local polynomial estimators produce large and statistically significant estimates of $\hat{\tau}$. When $\tau = 0$, CCT 95% confidence intervals reject the null hypothesis ($H_0 : \tau = 0$) roughly 7% of the time. Of these rejections, over 90% display a characteristic "zig zag" pattern illustrated in the figure – steep-sloped regression functions on either side of the cutoff and a treatment effect of the opposite sign.[1] This pattern is emblematic of false positive or exaggerated claims from noisy RD data, and will be present in many of the empirical applications we present below.

When faced with an overfitting problem, the principled remedy is *regularization*, penalizing extreme estimates to guard against overfitting to noisy data. In this paper we suggest a Bayesian approach to regularized RD estimation first proposed by Branson et al. (2019), which we call Gaussian Process Regression Discontinuity (GPRD). Because it places a prior on the smoothness of the conditional expectation function, this method is particularly useful for moderating the exaggerated claims from low-powered RD studies. In the following two sections, we introduce GPRD and show that it performs well in simulations, overcoming the disadvantages of traditional global polynomial approaches while producing lower variance estimates than local linear approaches. In section 4 we apply GPRD to a large set of published RD studies where low power has yielded exaggerated treatment effects. We conclude with an agenda for future research, and provide open source software for researchers (the `R` package `gprd`).

2 Gaussian Process Regression for RD Designs

In regression discontinuity (RD) designs, we attempt to estimate the causal effect of a treatment that is assigned at a specific value of some *running variable*, x. We assume the outcomes y are a noisy function of the running variable x, and we are interested in the discontinuity in

[1] Or, when estimated using a local quadratic regression, a concave regression function on one side of the cutoff, and a convex function on the other side.

$f(x)$ induced by the treatment. We propose using Gaussian process regression (Neal, 1998) to learn this conditional expectation function $f(x)$ and estimate causal effects in RD designs.

To formalize the RD problem we start by assuming the outcomes are normally distributed,

$$y \sim \mathcal{N}\left(f\left(x\right), \sigma_y^2 I\right), \tag{1}$$

where $f\left(x\right)$ is an unknown function of the running variable x. A treatment occurs at a cutoff value c of x; that is, all outcomes where $x \geq c$ receive the treatment. For simplicity here we will assume $c = 0$.

Alternatively, we might assume that the outcomes are normally distributed but with different mappings between input and output on either side of the cutoff;

$$y_+ \sim \mathcal{N}\left(f_+\left(x_+\right), \sigma_y^2 I\right), \tag{2}$$

$$y_- \sim \mathcal{N}\left(f_-\left(x_-\right), \sigma_y^2 I\right), \tag{3}$$

We will propose two different methods to estimate the treatment effect τ of the intervention using Gaussian process (GP) regression. The first, which we call the *global GPRD estimator*, fits a single GP model for all observations, with a dummy variable D to indicate treatment, and estimates the treatment effect using the difference in prediction when $x = c$ and $D = 1$ and $D = 0$. The second, which we call the *piecewise GPRD estimator*, models the data separately on either side of the cutoff and estimates the treatment effect using the difference between predictions at the cutoff from the "right equation" and "left equation."

2.1 Gaussian Process Regression

GP regression is a method used to learn the mapping from x to y when its functional form is not known, accomplished by placing a GP prior over the function space. As this methodology

is uncommon in the political science literature, we provide a brief overview here.

A Gaussian process (GP) is an infinite dimensional generalization of the normal distribution. More specifically, it is "collection of random variables, any finite number of which have a joint Gaussian distribution" (Rasmussen and Williams, 2006, 13). The mean and covariance of this normal distribution is given as functions of the inputs, so that we say

$$f(x) \sim \mathcal{GP}(m(x), K(x)). \tag{4}$$

Common examples for the mean function are the "mean zero" function $m(x) = 0$ and the "linear mean" function $m(x) = x\beta$. A common example of a covariance function (also called a kernel) is the isometric squared exponential covariance function, aka the radial basis function,

$$K\left(x, x'\right) = \sigma_f^2 \exp\left(-0.5\frac{\left(x - x'\right)^2}{\ell^2}\right), \tag{5}$$

where σ_f is a hyperparameter called the scale factor, which scales the entire covariance matrix, and ℓ is a hyperparameter called the length scale, which influences how quickly f varies in x. Then the i, j element of the covariance matrix is given by $K(x_i, x_j)$.

This setup allows us to model distributions over *functions* rather than simply distributions over *variables*. Then, rather than assume we know the form of the mapping between the input variables $\mathbf{X}$ and outcomes $\mathbf{y}$—such as a polynomial of a particular order—we can instead use a GP to place a probability distribution over all possible mappings from $\mathbf{X}$ to $\mathbf{y}$. With a Gaussian likelihood for the data given $\mathbf{X}$ and $f(x)$, a posterior distribution over $f(x)$ is then given by application of Bayes' rule utilizing Gaussian identities.[2] Crucially for our purposes, well known results provide the posterior predictive distribution for a test

[2]For a more detailed derivation of this posterior distribution, see Rasmussen and Williams (2006), Chapter 2.

observation x^* as

$$f(x^*) \sim \mathcal{N}(m^*, C^*) \tag{6}$$

$$m^* = m(x^*) + K(x^*, \mathbf{x})[K(\mathbf{x}) + \sigma_y^2 I]^{-1}(\mathbf{y} - m(\mathbf{x})) \tag{7}$$

$$C^* = K(x^*) - K(x^*, \mathbf{x})[K(\mathbf{x}) + \sigma_y^2 I]^{-1} K(\mathbf{x}, x^*). \tag{8}$$

Note that Bayesian linear regression is a special case of Gaussian process regression, using the linear covariance function $K(x, x') = x \cdot x'$.[3] When using other kernels, we simply allow non-linearity in the mapping from input to response. In other words, GP regression is a flexible extension to Bayesian linear regression to account for uncertainty in the functional form mapping predictors to response by assuming the covariance between outcomes is a function of the predictor variables. In the case of the common squared exponential covariance function (and its extension discussed in Section 2.2, the automatic relevance determination kernel), we assume that covariance between outcomes is a function of distance in the covariate space.

2.2 The Global GPRD Estimator

For the global GPRD estimator, we place a Gaussian process (GP) prior on $f(x)$,

$$p(f) = \mathcal{GP}(X\beta, K(X)), \tag{9}$$

where K is the squared exponential automatic relevance determination covariance function

$$K(X, X') = \sigma_f^2 \exp\left(-0.5 \sum_j \frac{\left(X_{\cdot,j} - X'_{\cdot,j}\right)^2}{\ell_j^2}\right) \tag{10}$$

with hyperparameters σ_f, the scale factor as in the squared exponential covariance func-

[3] A constant hyperparameter is added to K if an intercept is desired.

tion from Equation 5, and ℓ is a length scale *vector* with a separate length scale for each predictor variable, and where $X = \Big[\mathbf{1}|x|D\Big]$, with

$$D = \begin{cases} 1 & \text{if } x \geq c, \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

Note that there is never a difference between observations on the value of the intercept column, so the X that goes into the mean function $X\beta$ will include that column, but the X going into the covariance function does not (and accordingly ℓ is of length two, not three); we suppress this difference in the notation for simplicity.

We use a Gaussian prior for β, with mean $\mathbf{b}$ and covariance B. Then the posterior over β is given by

$$\beta \mid \mathbf{y}, X \sim \mathcal{N}\left(\bar{\beta}, \Sigma_\beta\right), \tag{12}$$

$$\bar{\beta} = \Sigma_\beta \left(X^T K_y^{-1} \mathbf{y} + B^{-1}\mathbf{b}\right), \tag{13}$$

$$\Sigma_\beta = \left(B^{-1} + X^T K_y^{-1} X\right)^{-1}, \tag{14}$$

$$K_y = K(X) + \sigma_y^2 I \tag{15}$$

(see Rasmussen and Williams (2006), section 2.7). Note that for the common case where $\mathbf{b} = \mathbf{0}$, $\bar{\beta}$ simplifies to $\Sigma_\beta(X^T K_y^{-1}\mathbf{y})$.

Then the mean and variance of f at a test point X_* is

$$\bar{f}(X_*) = X_*\bar{\beta} + K_* K_y^{-1}(\mathbf{y} - X\bar{\beta}), \tag{16}$$

$$\text{cov}(f_*) = K_{**} - K_* K_y^{-1} K_*^T + R^T (X^T K_y^{-1} X)^{-1} R, \tag{17}$$

$$K_* = K(X_*, X), \tag{18}$$

$$K_{**} = K(X_*), \tag{19}$$

$$R = X_*^T - X^T K_y^{-1} K_*^T \tag{20}$$

(See Equations 2.24, 2.38, and 2.41 in Rasmussen and Williams 2006). These differ from Equations 7 and 8 because we have incorporated uncertainty in the mean function parameters.[4]

So we are interested in the treatment effect

$$\tau_{GPRD-G} \overset{\text{def}}{=} f\left(\begin{bmatrix} 0 & 1 \end{bmatrix}\right) - f\left(\begin{bmatrix} 0 & 0 \end{bmatrix}\right), \tag{21}$$

or the difference between $f\left(x = 0, D = 1\right)$ and $f\left(x = 0, D = 0\right)$, which is distributed

$$\tau_{GPRD-G} \sim \mathcal{N}\left(\mu_*, \Sigma_*\right), \tag{22}$$

$$\mu_* = \bar{f}\left(\begin{bmatrix} 0 & 1 \end{bmatrix}\right) - \bar{f}\left(\begin{bmatrix} 0 & 0 \end{bmatrix}\right), \tag{23}$$

$$\Sigma_* = \text{cov}\left(f\left(\begin{bmatrix} 0 & 1 \end{bmatrix}\right)\right) + \text{cov}\left(f\left(\begin{bmatrix} 0 & 0 \end{bmatrix}\right)\right). \tag{24}$$

Note the key differences between the global GPRD estimator and the global polynomial RD estimator. In the global polynomial model, the treatment effect is taken to be the coefficient on D, which gives the difference in expected outcomes for the treated and untreated

[4]One could instead select mean function coefficients by maximizing the marginal log likelihood as we discuss for the covariance function hyperparameters in Section 2.4, in which case $\bar{f}(X_*)$ and $\text{cov}(f_*)$ would again be given by Equations 7 and 8.

observations at any point x. In the GP model, this is not true; the coefficient on D impacts the expected outcome, but the effect of the treatment also runs through the kernel (see Equation 16).

2.3 The Piecewise GPRD Estimator

Non-parametric regression discontinuity designs fit polynomials on either side of the discontinuity, and estimate the treatment effect as the difference of the limit of the polynomials at the cutoff. The piecewise GP estimation strategy extends this approach by placing Gaussian process (GP) priors over the functions on either side of the cutoff, which will be much more flexible than the polynomial regression approach and less sensitive to overfitting predictions close to the cutoff to observations far away from the cutoff. By design, the fit near the cutoff will rely more on inputs close to the cutoff than those far away.

To formalize, x_+ will be the inputs to the right of the cutoff and y_+ the corresponding outcomes, and analogously for x_- and y_-. Then we will learn two functions, $f_+ : x_+ \to y_+$ and $f_- : x_- \to y_-$. To do so, we place GP priors over the functions,

$$f_+ \sim \mathcal{GP}(X_+\beta_+, K(x_+)), \tag{25}$$

$$f_- \sim \mathcal{GP}(X_-\beta_-, K(x_-)), \tag{26}$$

where $X_\cdot$ prepends the column vector $\mathbf{1}$ to $x_\cdot$, $\beta_\cdot$ is a vector giving the intercept and slope of the linear mean function, and $K(\cdot)$ is the isometric squared exponential covariance function from Equation 5.

We again use a Gaussian prior for β, with mean $\mathbf{b}$ and covariance B, and the mean and variance of $f_\cdot$ at a test point x_* is as given in Equations 16 and 17. Then we are interested in the treatment effect

$$\tau_{GPRD-P} \stackrel{\text{def}}{=} f_+(0) - f_-(0), \tag{27}$$

which is distributed

$$\tau_{GPRD-P} \sim \mathcal{N}(\bar{f}_+(0) - \bar{f}_-(0), \text{cov}(f_+(0)) + \text{cov}(f_-(0))). \tag{28}$$

Note the key differences between the piecewise GPRD estimator and local linear or polynomial RD estimators. The local polynomial estimators use only a portion of the data to either side of the cutoff to avoid undue influence of observations far from the cutoff. In contrast, the GP model is able to use all of the available data because the covariance between outputs decreases with distance in the covariate space, a natural and smooth way to decrease undue influence of observations far from the cutoff while still borrowing *some* information from them. Additionally, the local methods require specification of the degree of the local polynomials, and this researcher-imposed model restriction can significantly impact inference. By contrast, in the GP model, we place a prior over the possible mappings from x to y and learn $f(x)$ from the data.

Also note the differences between the piecewise GPRD estimator we introduce and the approach taken in Branson et al. (2019). While Branson et al. largely rely on a shared covariance assumption, we relax this assumption, selecting different covariance function hyperparameters for the treatment and control groups. When the covariance function hyperparameters in fact should be shared, parameter selection or sampling routines should be able to recover that. When they are not, and the mapping from runnning variable to outcomes is wholly different between the treatment and control groups, the piecewise GPRD estimator we introduce may be most appropriate.

Taken together, our global and piecewise GPRD estimators can be seen as capturing a fairly wide range of restrictiveness of prior assumptions in GP regression estimation of treatment effects in RD designs. It is also useful to acknowledge one important assumption shared by our approaches, that of *stationarity*, or that covariance hyperparameters do not themselves also vary as a function of the running variable. We leave the relaxation of that

assumption to future work.

2.4 Choosing Hyperparameters

Because our inferences are strongly affected by the choice of hyperparameters σ_y, σ_f, and ℓ, we need a theoretically-grounded, automatic procedure for selecting their values. Our preferred approach is to place priors over the covariance function hyperparameters, propagating uncertainty over their values into the posterior distribution of the treatment effect. Note that in this case, however, we cannot use exact inference but must instead resort to simulating the posterior with an MCMC sampler. In the empirical applications below, we take this approach using the following prior distributions:

$$\ell \sim \text{InvGamma}(5, 5) \tag{29}$$

$$\sigma_f \sim \text{HalfNormal}(0, 1) \tag{30}$$

$$\sigma_y \sim \text{HalfNormal}(0, 1) \tag{31}$$

The inverse gamma prior on the length scale is a boundary-avoiding prior (Gelman, 2014, 313) which penalizes infinitesimal values, and the half-normal prior on σ_f ensures that some prior weight is placed on the zero function. These priors are most sensible for x and y variables that have been standardized so that their variance equals 1.

A practical impediment to the fully Bayesian approach is computation time, which scales super-linearly with n. As of writing, fitting a Gaussian process regression with 1,000 observations by Hamiltonian Monte Carlo takes roughly half an hour. Though this is not too time-consuming for applied work, it is prohibitive for simulations. So in the following section, we speed up computation by plugging in hyperparameters that maximize the log marginal

likelihood. In the case of a linear mean, the log marginal likelihood is

$$\log p(y \mid X, \mathbf{b}, B) = -\frac{1}{2} M^T Q^{-1} M - \frac{1}{2} \log |Q| - \frac{n}{2} \log 2\pi, \tag{32}$$

$$M = X\mathbf{b} - \mathbf{y}, \tag{33}$$

$$Q = K_y + XBX^T, \tag{34}$$

(see Equation 2.43 in Rasmussen and Williams 2006). Again, note that for the common case $\mathbf{b} = \mathbf{0}$, M reduces to $-\mathbf{y}$.

Then the gradient of the log marginal likelihood with respect to the hyperparameters $\theta = (\sigma_y, \sigma_f, \ell)$ in the case of the isometric covariance function is

$$\frac{\partial}{\partial \theta_i} \log p(y \mid X, \mathbf{b}, B) = \frac{1}{2} M Q^{-1} \frac{\partial Q}{\partial \theta_i} Q^{-1} M - \frac{1}{2} \operatorname{Tr}\left(Q^{-1} \frac{\partial Q}{\partial \theta_i} \right), \tag{35}$$

$$\frac{\partial Q}{\partial \theta} = \begin{bmatrix} 2\sigma_y I \\ 2\sigma_f \exp\left(-0.5 K_0\right) \\ \sigma_f^2 \exp\left(-0.5 K_0\right) \circ K_0 \end{bmatrix}, \tag{36}$$

where the i, j element of the matrix K_0 is given by

$$K_{0\,i,j} = \frac{(x_i - x_j)^2}{\ell^2}, \tag{37}$$

and we can use (e.g.) the conjugate gradient method to optimize the hyperparameters. The gradient for the automatic relevance determination case is analogous, but just extended with an element for each element of ℓ accordingly.

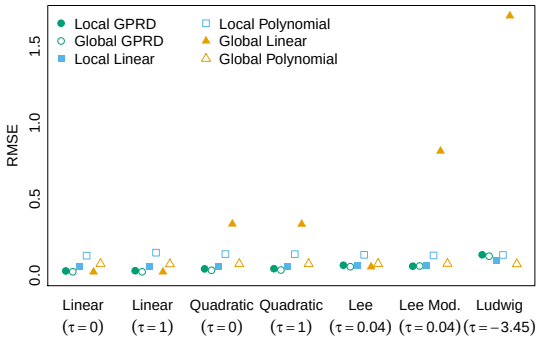
3 Comparing Models: Simulation Evidence

To assess performance of the GPRD estimators compared to existing methods, we engage in two simulation exercises. First, we use a common set of simulations from the RD literature, where the running variable x is given by $2z - 1$, with $z \sim \mathcal{B}(2, 4)$, and $y = f_j(x) + \varepsilon$, with $\varepsilon \sim \mathcal{N}(0, 0.1295^2)$. To specify the shape function $f_j(x)$, we use the three functions explored in Calonico, Cattaneo and Titiunik (2014), themselves taken from Lee (2008) and Ludwig and Miller (2007). We additionally use global linear and quadratic functions, $f(x) = x + \tau I(x > 0)$ and $f(x) = x^2 + \tau I(x > 0)$, with $\tau = 0$ and $\tau = 1$, for a total of seven tested data generating processes. For each DGP, we simulate 1,000 datasets of 500 observations.

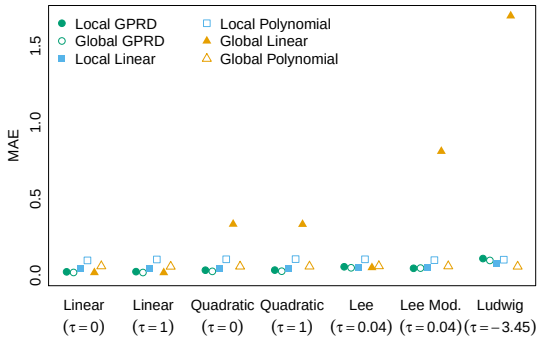
We estimate treatment effects and confidence intervals for global GPRD, piecewise GPRD, local linear, and fifth-degree global polynomial approaches (the Lee and Ludwig and Miller DGPs both use five-degree polynomials). Figure 2 depicts the root mean squared error, mean absolute error, confidence interval length, and confidence interval coverage averaged by simulation condition. The GPRD methods generally outperform other methods in terms of error and confidence interval length; when averaged across conditions, the GPRD methods outperform all other methods on these metrics, and for some DGPs, this difference is more pronounced.

In the Lee and Ludwig-Miller simulations, the GPRD estimates are biased; this echoes findings in Branson et al. (2019), who note that with these DGPs, the functions' slope increases markedly near the cutoff, violating the stationarity assumption we leverage. As with any regularizing estimate, one can think of our method as deliberately introducing a bias in small samples to reduce variance and overfitting. As sample size increases, this bias shrinks. The results averaged across all conditions for all data generating processes, as well as only stationary data generating processes, are given in Table 1.

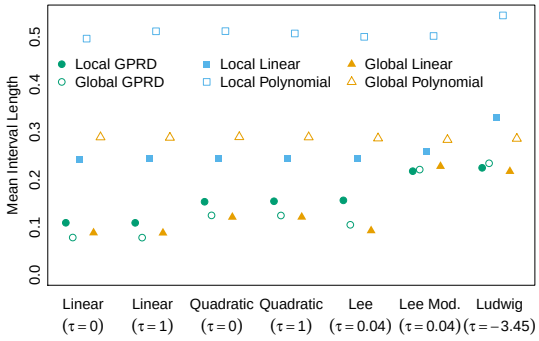
In the absence of a violation of the stationarity assumption, the GPRD methods increases precision while retaining coverage. Importantly, note that these simulations, standard in



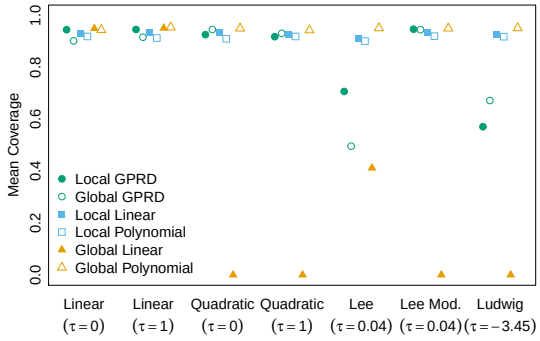
(a) Root mean squared error of treatment effect estimates.



(b) Mean absolute error of treatment effect estimates.



(c) Mean confidence interval length for treatment effect estimates.



(d) Mean 95% confidence interval coverage of true treatment effects.

Figure 2: Fit statistics for treatment effect estimates from the global and piecewise GPRD models (plotted in empty and filled green circles respectively), local linear and polynomial regression (plotted in filled and empty blue squares respectively), and global linear and polynomial regression (plotted in filled and empty yellow triangles).

the literature, assume very low observation noise. When σ^2 is large, we see an even greater difference in performance between the methods. We explore the settings where $\varepsilon \sim \mathcal{N}(0, 0.5^2)$ and $\varepsilon \sim \mathcal{N}(0, 1)$ for $f(x) = x + \tau I(x > 0)$ for $\tau = 0$ and $\tau = 1$ to consider the difference in performance in (perhaps more realistic) noisier environments. We display the RMSE, MAE, interval length, and coverage by condition in Figure 3 and report the pooled results in Table 2.

In this setting, the mean absolute error, root mean squared error, and confidence interval length are all more than *twice* as large for the local linear regression method than the piecewise GPRD method, and the global GPRD method edges out the piecewise GPRD

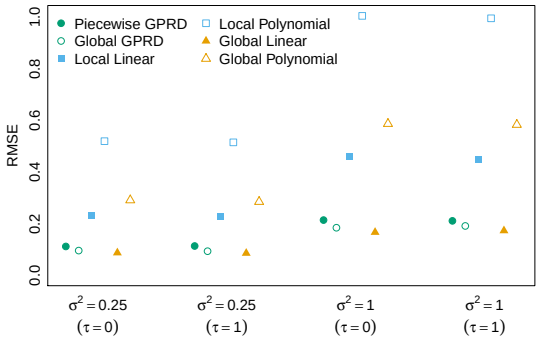
Table 1: Fit statistics averaged across replicates and data generating processes.

Estimator	MAE	RMSE	Mean CI Length	Mean CI Coverage
All DGPs				
Global GPRD	0.041	0.061	0.138	0.830
Piecewise GPRD	0.047	0.067	0.161	0.851
Local Linear	0.052	0.068	0.259	0.928
Local Polynomial	0.103	0.136	0.511	0.914
Global Linear	0.466	0.735	0.137	0.330
Global Polynomial	0.059	0.075	0.288	0.949
Stationary DGPs				
Global GPRD	0.028	0.038	0.126	0.928
Piecewise GPRD	0.032	0.043	0.149	0.936
Local Linear	0.047	0.060	0.247	0.932
Local Polynomial	0.102	0.137	0.506	0.916
Global Linear	0.303	0.422	0.129	0.380
Global Polynomial	0.059	0.075	0.288	0.948

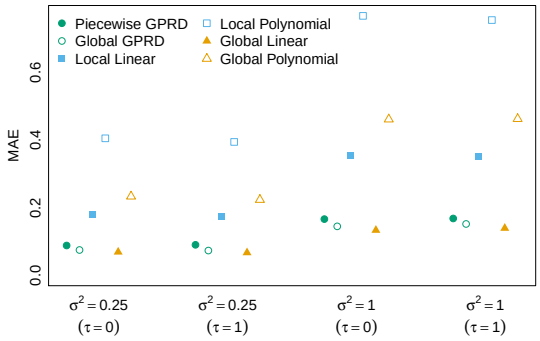
Table 2: Fit statistics averaged across replicates and noise and effect sizes.

Estimator	MAE	RMSE	Mean CI Length	Mean CI Coverage
Global GPRD	0.111	0.149	0.476	0.913
Piecewise GPRD	0.128	0.170	0.627	0.946
Local Linear	0.265	0.360	1.410	0.931
Local Polynomial	0.580	0.796	2.880	0.910
Global Linear	0.102	0.135	0.511	0.951
Global Polynomial	0.345	0.461	1.660	0.949

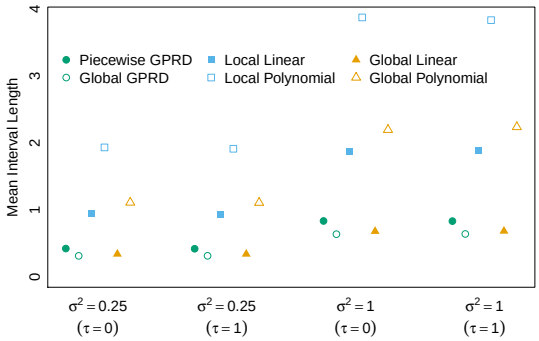
method further still. OLS, the global linear model, does very well, as may be expected when the simulation's DGP matches exactly the OLS assumptions. Remarkably, the GPRD methods perform similarly to the model whose assumptions match the DGP. In the common scenario where noise is appreciable, the GPRD methods outperform existing local methods substantially and do no worse than when the researcher can correctly intuit the precise functional form of the mapping from running variable to outcomes.



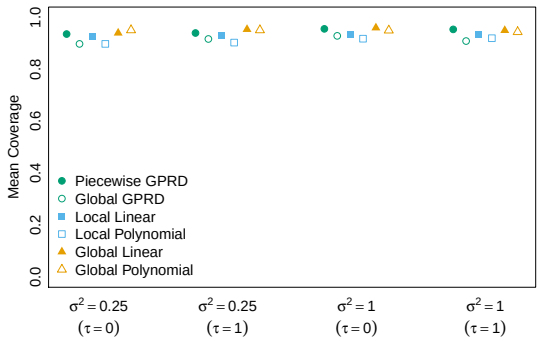
(a) Root mean squared error of treatment effect estimates.



(b) Mean absolute error of treatment effect estimates.



(c) Mean confidence interval length for treatment effect estimates.



(d) Mean 95% confidence interval coverage of true treatment effects.

Figure 3: Fit statistics for treatment effect estimates from the global and piecewise GPRD models (plotted in empty and filled green circles respectively), local linear and polynomial regression (plotted in filled and empty blue squares respectively), and global linear and polynomial regression (plotted in filled and empty yellow triangles).

4 Empirical Applications

We selected the following empirical applications based on three criteria. First, they are all published in top political science and economics journals over the past five years. Second, they all adhere to current best practices for RD studies, employing local polynomial estimators with automated bandwidth selectors and robust standard errors, and conducting extensive falsification and robustness tests. Third, the authors provide sufficient replication materials to reproduce their work. In short, we select these studies not because they are *poor* examples of applied RD, but because they are good examples of careful and rigorous empirical work. Nevertheless, many of these studies employ datasets that are too low-powered

near the cutoff to reliably detect plausible treatment effects.

In total, we replicate 29 studies published since 2015 that employ a regression discontinuity design. We re-estimate the treatment effects using both CCT local linear regression and piecewise GPRD. These estimates are summarized in Figure 4, standardized so as to appear on the same scale. For studies with a large number of observations near the cutoff, the estimates from the two methods do not differ significantly. But for studies with low power near the cutoff, the two estimates often differ noticeably. Typically, the GPRD estimate is closer to zero than the CCT estimate, but not universally – see, for example, the replications of Carson and Sievert (2017) and Horowitz et al. (2019). The reader may find descriptions of each replication in the appendix, and below we discuss three in detail.

4.1 The Radical Right and Party Manifestos

Abou-Chadi and Krause (2018) study how the presence of radical right parties influence the platforms of mainstream parties. Their causal identification strategy is based on electoral thresholds in parliamentary systems; typically it is required that a party clear some percentage of the total vote before gaining representation in parliament, where the particular threshold varies by country. The authors compare elections where radical right parties barely exceeded the threshold (gaining representation) and barely missed the threshold, observing how mainstream political parties respond.

The dependent variable is change in a measure of Cultural Protection in the party's manifesto during the following election. These data are compiled by the Comparative Manifestos Project (Volkens, Pola Lehmann and Werner, 2015), and the outcome variable is a function of the difference between the number of favorable mentions of cultural diversity and encouragement of integration and cultural homogeneity in the party's platform (Lowe et al., 2011). In their paper, Abou-Chadi and Krause (2018) present estimates from a diverse range of specifications, which range from 3.1 to 4.9. Replicating these results using CCT

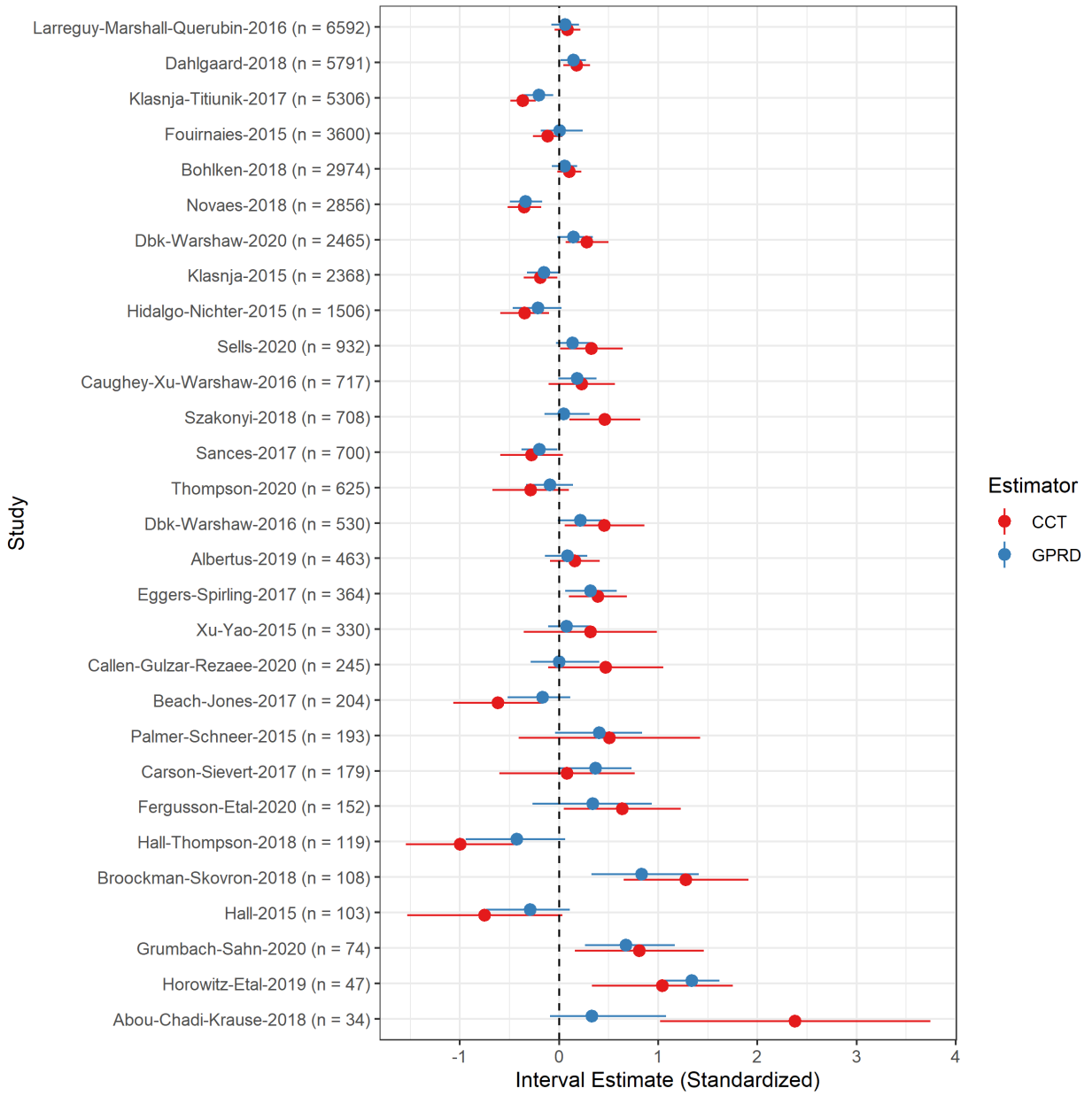


Figure 4: Estimates and 95% confidence intervals across 30 replications, sorted by sample size within the CCT bandwidth. Estimates are standardized in order to plot them on the same scale.

bias-corrected standard errors and robust confidence intervals yields an estimate of 3.96, with 95% confidence interval $[1.7, 6.2]$.

To get a sense of the relative magnitude of this estimated effect, consider Figure 5. This plots the average value of the Cultural Protection score, by country, for each family of political party since 1980. Although the Cultural Protection score is noisy from election year to election year, averaging across years yields predictable patterns. Right-leaning parties tend to score higher on the measure than left-leaning parties, and Nationalist parties – where they exist – typically score 2 to 3 points higher than the average mainstream party.

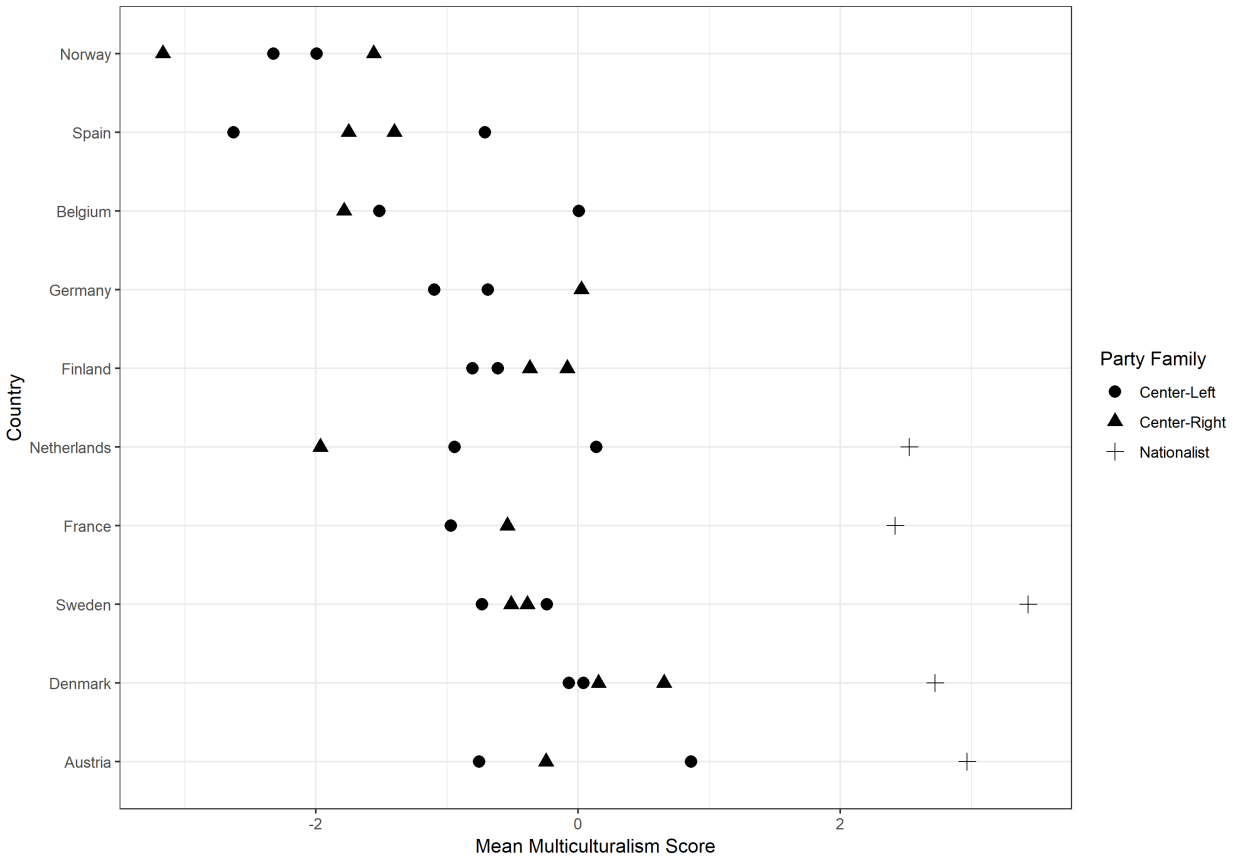


Figure 5: Mean Cultural Protection score by country and party family (all elections post-1980; Center-Left includes Social Democrats and Liberals, Center-Right includes Christian Democrats and Conservatives).

In this context, an estimated effect size of 3.9 is enormous. If true, it suggests that not only do mainstream parties respond to Radical Right representation by moving their plat-

forms to the right, but they do so in such a way that their rhetoric completely closes or even overtakes the average gap between mainstream and rightwing nationalist party positions.

By comparison, GPRD yields a more modest treatment effect estimate, with a 95% posterior interval that covers zero. Table 3 compares the effect sizes estimated by the piecewise GPRD and CCT estimators, and Figure 6 visualizes those estimates. We can see that the local linear estimator yields a steep prediction line just to the left of the cutoff, within a bandwidth that contains only 24 observations. Because the GPRD estimator makes use of observations outside that bandwidth, it yields a mean function that is less overfit to observations near the cutoff. The estimated treatment effect is roughly 0.33 standard deviations, or 0.5 units on the cultural protectionism scale, a more *a priori* plausible effect size.

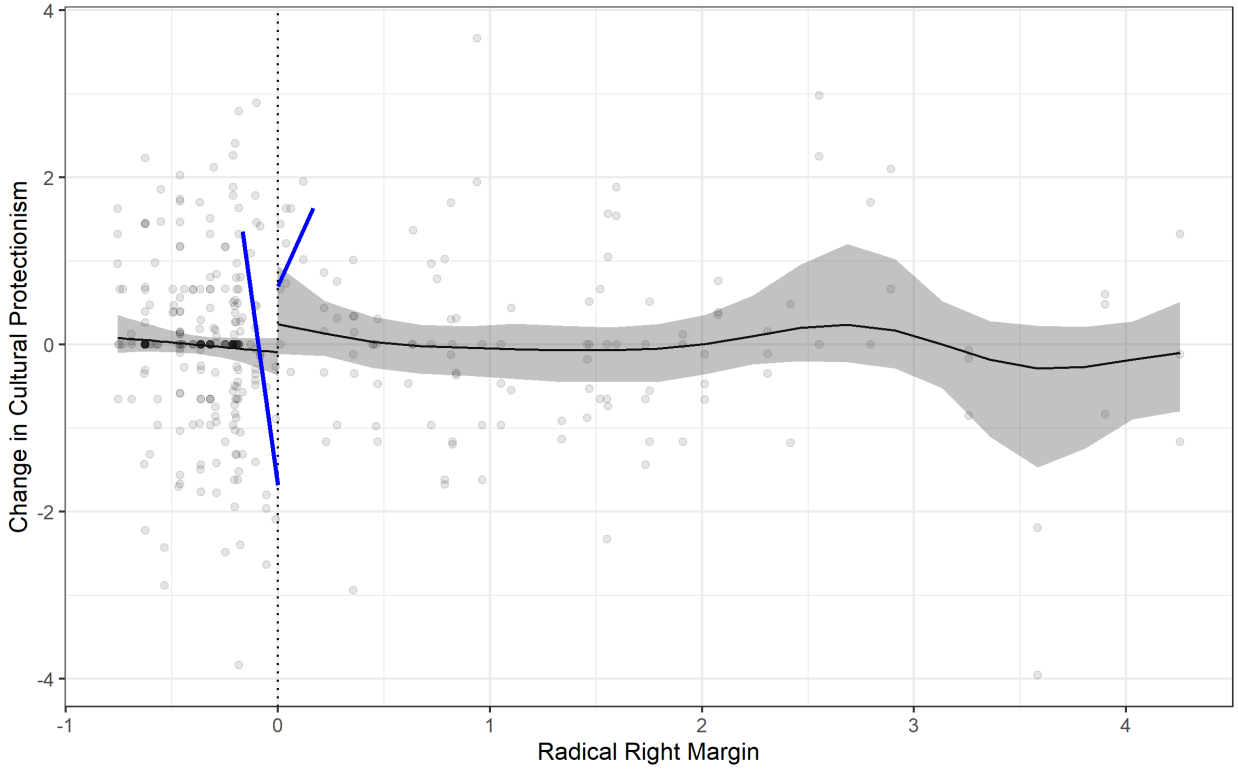


Figure 6: Fit comparison between GPRD (black with shaded 95% posterior interval) and CCT (blue) estimators for the Abou-Chadi and Krause (2018) application. All variables standardized with x centered around the RD cutoff and y centered around its mean.

Table 3: Treatment effect comparison between GPRD and local linear regression for the Abou-Chadi and Krause (2018) application.

Estimator	Treatment effect	95% CI
Local Linear	2.38	$[1.02, 3.74]$
Piecewise GPRD	0.33	$[-0.09, 1.08]$

4.2 Ethnic Diversity and Municipal Public Spending

The second empirical illustration comes from Beach and Jones (2017), who study the effect of diverse city council representation on municipal-level public goods spending. There is a large literature on this topic, dating back to Alesina, Baqir and Easterly (1999) and continuing through Hopkins (2011) and Trounstine (2015), finding that in ethnically diverse and segregated cities, municipal governments devote less spending to public goods than in ethnically homogeneous cities. All of these studies rely on cross-sectional or longitudinal regression analysis of public finance data, and so lack a credible causal identification. Beach and Jones (2017) approach the problem through an RD analysis of city council elections in California. Because narrowly elected city councilors from a 'non-modal' ethnicity should increase the diversity of a city council, the discontinuity at the plurality margin allows for the causal identification of an effect of council diversity on public spending.

By regressing per capita public goods spending on the vote margin of 'non-modal' candidates for races where a non-modal candidate faced a modal candidate, Beach and Jones (2017) estimate a treatment effect of -0.31 on a log scale, implying that the election of a single councilmember that increases ethnic diversity causes a 31% drop in public goods spending.

To consider this estimate in context, a 31% cut in public goods spending amounts to roughly $463 per capita for the average city. This is equivalent to eliminating all spending on police (13% in the average city), fire protection (9%), and roads (9%). By comparison, other studies of municipal spending suggest much more modest (or null) effects. For example, in an RD analyis of Democratic vs. Republican mayoral candidates, de Benedictis-Kessner

and Warshaw (2016) find that narrowly-elected Democratic mayors increase public spending by 5% on average. In their original study of US cities, Alesina, Baqir and Easterly (1999) find that ethnically diverse cities spend 6% to 9% lower shares on "productive" public goods than ethnically homogeneous cities.[5] All of these estimates are several multiples smaller than 30%.

Even assuming that the effect of a non-modal city councilmember is comparable to that of a Democratic mayor,[6] the RD analysis would be too underpowered to reliably detect a more plausible effect. The minimum detectable effect size given the number of observations in this study is roughly 32% (Bloom, 1995)[7], and the RD plots in Figure 7 makes this problem clear: although a substantial fraction of the sample lies within the bandwidth, the variation in public goods spending is very large relative to any plausible effect size. A sufficiently high powered study to detect an effect size of 5% would require many times more observations within the neighborhood of the cutoff.

Table 4 and Figure 7 compare the GPRD and local linear RD estimates. The GPRD estimate is four times smaller than the local linear estimate, and not reliably different than zero.

Table 4: Treatment effect comparison between GPRD and local linear regression for the Beach and Jones (2017) application.

Estimator	Treatment effect	95% CI
Local Linear	-0.618	$[-1.07, -0.167]$
Piecewise GPRD	-0.165	$[-0.519, 0.113]$

[5]This is the estimated effect of changing the Herfindahl ethnic fractionalization index from 0 to 1 (i.e. from perfect homogeneity to perfect heterogeneity).

[6]Not an unreasonable assumption, since in many municipalities the mayor does not wield strong executive power, and is essentially the chair of the city council.

[7]See Cattaneo, Titiunik and Vazquez-Bare (2019) for derivation of power calculations in the RD context

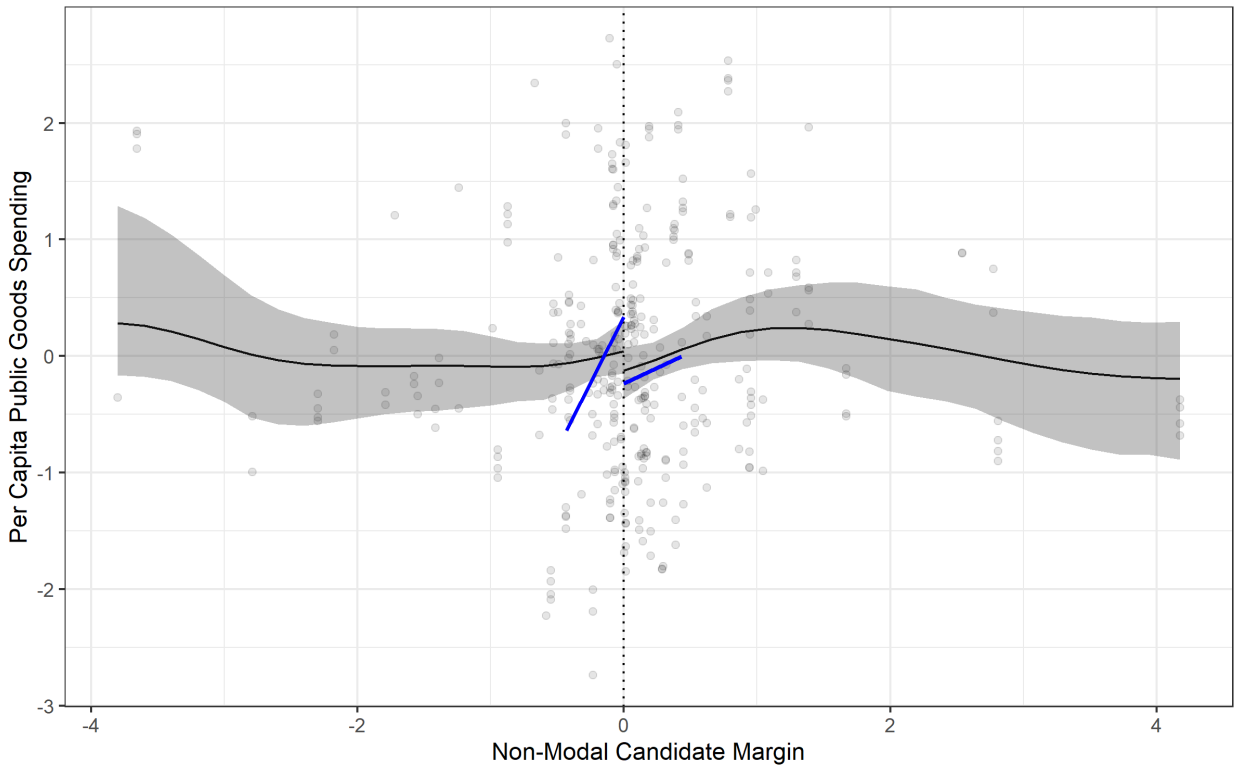


Figure 7: Fit comparison between GPRD (black with shaded 95% posterior interval) and CCT (blue) estimators for the Beach and Jones (2017) application. All variables standardized with x centered around the RD cutoff and y centered around its mean.

4.3 Political Connections and Firm Profitability in Russia

Szakonyi (2018) investigates close elections to regional legislatures in Russia, testing whether firms with regional political connections subsequently see higher profit margins. The local linear RD estimate is strongly positive, suggesting that the election a board member causes as much as a 15% increase in firm profits. This is lucrative effect! As the author notes, "The presence of a political connection can spell the difference between an impressively profitable firm and one that barely breaks into the black."

By comparison, the GPRD treatment effect estimate is approximately zero. Eyeballing Figure 8, it is clear that, unlike the previous two examples, there are plenty of observations within the CCT bandwidth. But, like the examples from Figure 1 a chance pattern in observations near the cutoff yields steep-sloped prediction lines that appear to exaggerate

the discontinuity.

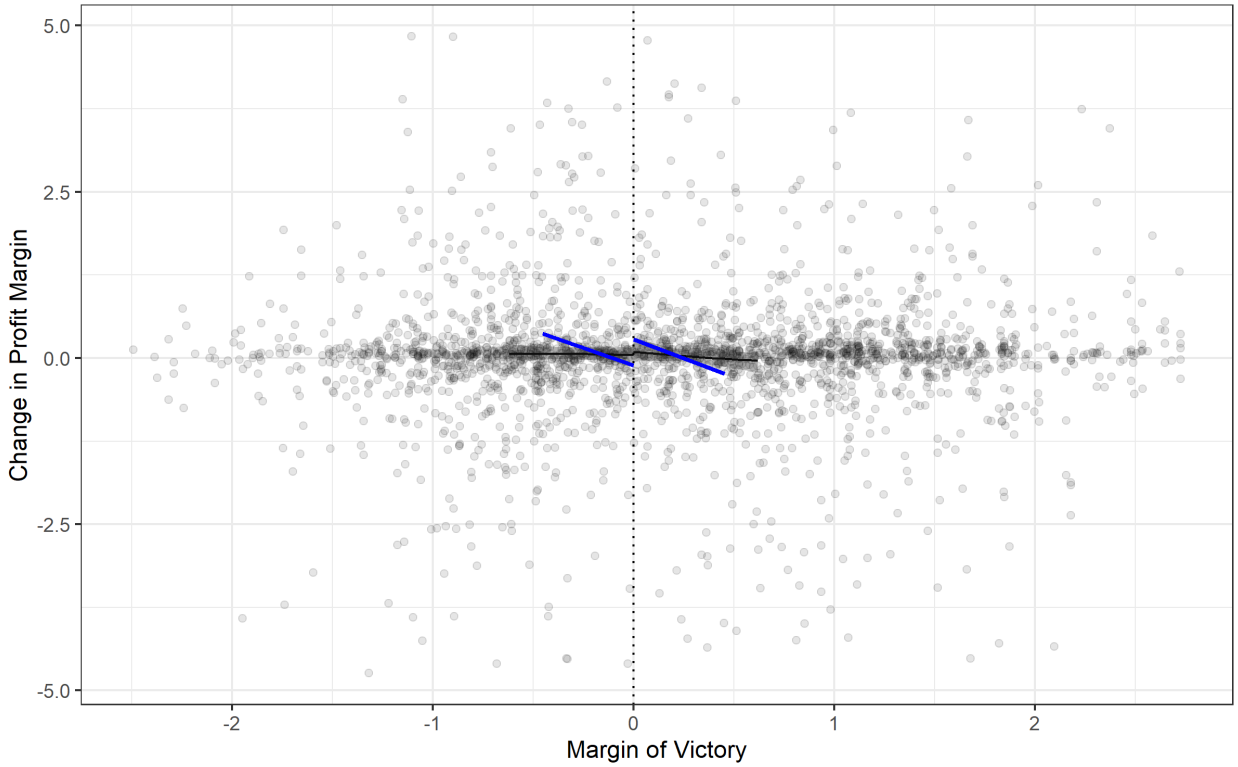


Figure 8: Fit comparison between GPRD (black with shaded 95% posterior interval) and CCT (blue) estimators for the Szakonyi (2018) application. All variables standardized with x centered around the RD cutoff and y centered around its mean.

Table 5: Treatment effect comparison between GPRD and local linear regression for the Szakonyi (2018) application.

Estimator	Estimate	95% CI
CCT	0.46	[0.105, 0.816]
GPRD	0.05	[-0.146, 0.306]

5 Conclusion

In this paper, we have demonstrated how low-powered regression discontinuity analyses can yield misleading and exaggerated causal effect estimates, characterized by an implausibly large divergence in the slope of the conditional expectation function in the neighborhood of

the cutoff. We propose Gaussian Process Regression Discontinuity (GPRD) as a principled method to address these problems. GPRD performs well in simulations and provides more plausible treatment effect estimates in empirical applications.

In future work we plan to expand GPRD to handle fuzzy RD designs, clustered data, and the inclusion of additional pre-treatment covariates. All of these will be made available for researchers in the forthcoming `R` package `gprd`.

References

Abou-Chadi, Tarik and Werner Krause. 2018. "The Causal Effect of Radical Right Success on Mainstream Parties' Policy Positions: A Regression Discontinuity Approach." *British Journal of Political Science* pp. 1–19.

Alesina, Alberto, Reza Baqir and William Easterly. 1999. "Public Goods and Ethnic Divisions." *The Quarterly Journal of Economics* 114(4):1243–1284.

Beach, Brian and Daniel B. Jones. 2017. "Gridlock: Ethnic Diversity in Government and the Provision of Public Goods." *American Economic Journal: Economic Policy* 9(1):112–136.

Bloom, Howard S. 1995. "Minimum Detectable Effects: A Simple Way to Report the Statistical Power of Experimental Designs." *Evaluation Review* 19(5):547–556.

Branson, Zach, Maxime Rischard, Luke Bornn and Luke W. Miratrix. 2019. "A Nonparametric Bayesian Methodology for Regression Discontinuity Designs." *Journal of Statistical Planning and Inference* 202:14–30.

Button, Katherine S., John P. A. Ioannidis, Claire Mokrysz, Brian A. Nosek, Jonathan Flint, Emma S. J. Robinson and Marcus R. Munafò. 2013. "Power Failure: Why Small Sample Size Undermines the Reliability of Neuroscience." *Nature Reviews Neuroscience* 14(5):365–376.

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

Carson, Jamie L. and Joel Sievert. 2017. "Congressional Candidates in the Era of Party Ballots." *The Journal of Politics* 79(2):534–545.

Cattaneo, Matias D., Rocío Titiunik and Gonzalo Vazquez-Bare. 2019. "Power Calculations for Regression-Discontinuity Designs." *The Stata Journal* 19(1):210–245.

de Benedictis-Kessner, Justin and Christopher Warshaw. 2016. "Mayoral Partisanship and Municipal Fiscal Policy." *The Journal of Politics* 78(4):1124–1138.

de la Cuesta, Brandon and Kosuke Imai. 2016. "Misunderstandings About the Regression Discontinuity Design in the Study of Close Elections." *Annual Review of Political Science* 19:375–396.

Gelman, Andrew. 2014. *Bayesian Data Analysis*. Chapman & Hall/CRC Texts in Statistical Science third edition ed. Boca Raton: CRC Press.

Gelman, Andrew and Guido Imbens. 2019. "Why High-Order Polynomials Should Not Be Used in Regression Discontinuity Designs." *Journal of Business & Economic Statistics* 37(3):447–456.

Gelman, Andrew and John Carlin. 2014. "Beyond Power Calculations: Assessing Type S (Sign) and Type M (Magnitude) Errors." *Perspectives on Psychological Science* pp. 1–11.

Gibbs, Mark N. 1998. Bayesian Gaussian Processes for Regression and Classification Dissertation University of Cambridge.

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

Hopkins, Daniel J. 2011. "The Limited Local Impacts of Ethnic and Racial Diversity." *American Politics Research* 39(2):344–379.

Horowitz, Michael, Brandon M. Stewart, Dustin Tingley, Michael Bishop, Laura Resnick Samotin, Margaret Roberts, Welton Chang, Barbara Mellers and Philip Tetlock. 2019. "What Makes Foreign Policy Teams Tick: Explaining Variation in Group Performance at Geopolitical Forecasting." *The Journal of Politics* 81(4):1388–1404.

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

Lowe, Will, Kenneth Benoit, Slava Mikhaylov and Michael Laver. 2011. "Scaling Policy Preferences from Coded Political Texts." *Legislative Studies Quarterly* 36(1):123–155.

Ludwig, J and D L Miller. 2007. "Does Head Start Improve Children's Life Chances?" *The Quarterly Journal of Economics* 122:159–208.

Neal, Radford M. 1998. "Regression and Classification Using Gaussian Process Priors." *Bayesian Statistics* 6:16.

Rasmussen, Carl Edward and Christopher K. I. Williams. 2006. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning Cambridge, Mass: MIT Press.

Szakonyi, David. 2018. "Businesspeople in Elected Office: Identifying Private Benefits from Firm-Level Returns." *American Political Science Review* 112(2):322–338.

Trounstine, Jessica. 2015. "Segregation and Inequality in Public Goods." *American Journal of Political Science* 60(3):709–725.

Volkens, Andrea, Sven Regel Theres Matthieß Pola Lehmann, Nicolas Merz and Annika Werner. 2015. *The Manifesto Data Collection: Manifesto Project (MRG/CMP/MARPOR). Version 2015a.* Berlin: .

Appendices

A Empirical Applications

Since 2015, over sixty studies employing a regression discontinuity design have been published in *American Political Science Review*, *American Journal of Political Science*, and *Journal of Politics*. We are able to replicate twenty-eight of the studies listed in Table 6 using both local linear and Gaussian Process RD. The appendices that follow contain details about each study's design and the estimated treatment effect using both methods. Throughout, we report standardized effect sizes, centering the running variable on the RD cutoff and the outcome variable on its mean. In the figures, the local linear fit is plotted in blue and the GPRD fit and 95% posterior interval is plotted in black. Raw data are semi-transparent circles, though for some visualizations it is clearer to present the averages by bins instead of the raw data. These binned data are represented as squares. The GPRD posterior is sampled via Hamilton Monte Carlo (`rstan`), and to speed computation we keep at most 1,000 observations closest to the cutoff.

	author_year	journal	outcome	treatment
1	Grumbach & Sahn (2020)	APSR	Minority Campaign Contributions	Minority Candidate
2	Thompson (2020)	APSR	Sheriff Compliance with Federal Immigration Requests	Sheriff Partisanship
3	Dynes & Holbein (2020)	APSR	Policy Outcomes	US State Party Control
4	Cavaille & Marshall (2019)	APSR	Anti-Immigrant Attitudes	Extra Year of Schooling
5	Mo & Conn (2018)	APSR	Beliefs About Disadvantage	National Service (TFA)
6	Broockman & Skovron (2018)	APSR	Contacts from Republicans	Republican Politician
7	Hall & Thompson (2018)	APSR	In-Party Turnout	Extremist Nominee
8	Dahlgaard (2018)	APSR	Turnout	Child Eligibility
9	Fiva & Smith (2018)	APSR	Dynasty Formation	Incumbency
10	Szakonyi (2018)	APSR	Firm Profits	Connected Politician
11	Nellis & Siddiqui (2018)	APSR	Religious Violence	Secular Party Rule
12	Hyytinen et al. (2018)	APSR	Public Spending	Municipal Employee Politician
13	Clinton & Sances (2018)	APSR	Election Turnout	Medicaid Expansion
14	Hainmueller, Hangartner, & Pietrantuono (2017)	APSR	Immigrant Social Integration	Naturalization
15	Klasnja & Titiunik (2017)	APSR	Vote Share	Incumbency
16	Gulzar & Pasquale (2017)	APSR	Employment Program Implementation	Number of Political Principals
17	Folke, Persson, & Rickne (2016)	APSR	Becoming a Party Leader	Primary Win
18	Croke, Grossman, Larreguy, & Marshall (2016)	APSR	Election Turnout	Education
19	Holbein (2016)	APSR	School Board Election Turnout	NCLB Failure Signal
20	Larreguy, Marshall, & Querubin (2016)	APSR	Turnout and PRI/PAN Vote Share	New Polling Station
21	Xu & Yao (2015)	APSR	Local Public Goods Expenditure	Large Clan Leader
22	Hall (2015)	APSR	General Election Vote Share	Extremist Nominee
23	Broockman & Ryan (2016)	AJPS	Contacting Legislators	Copartisan Legislator
24	Coppock & Green (2016)	AJPS	Downstream Voting	Upstream Voting
25	Hidalgo & Nichter (2016)	AJPS	Turnout	Voter Audits
26	Holbein & Hillygus (2016)	AJPS	Youth Turnout	Preregistration
27	Gilardi (2015)	AJPS	Women Candidates	Women Politicians
28	Rueda (2017)	AJPS	Vote Buying	Polling Place Size
29	Lopes da Fonseca (2017)	AJPS	Vote Share	Incumbency
30	Novaes (2018)	AJPS	Party Switching	Mayoral Win
31	Nyhan, Skovron, & Titiunik (2017)	AJPS	Youth Turnout	Voting Eligibility
32	Velez & Newman (2019)	AJPS	Turnout	Access to Spanish-language TV
33	Larreguy, Montiel Olea, & Guerubin (2017)	AJPS	Support for SNTE Machine	Polling Place Size
34	Bohlken (2018)	AJPS	Project Expenditure	Co-partisan MP
35	Kim (2019)	AJPS	Women's Turnout	Direct Democracy
36	Albertus (2019)	AJPS	Conflict	Land Reform
37	Fergusson, Guerubin, Ruiz, & Vargas (2020)	AJPS	Right-Wing Paramilitary Violence	Left-Wing Party Control
38	Bischof & Wagner (2019)	AJPS	Ideological Polarization	Radical Right Representation
39	Palmer & Schneer (2015)	JOP	Corporate Board Membership	Elected to Office
40	Lerman & McCabe (2017)	JOP	Attitudes toward ACA	Personal Experience with Public Healt
41	Erikson, Folke, & Snyder (2015)	JOP	Presidential Vote Share	Copartisan Governor
42	Caughey, Warshaw, & Xu (2017)	JOP	Policy Liberalism	Democratic Governor
43	Klasnja (2015)	JOP	Vote Share	Incumbency
44	de Benedictis-Kessner & Warshaw (2016)	JOP	Municipal Fiscal Policy	Mayoral Partisanship
45	Eggers & Spirling (2017)	JOP	Vote Share	Incumbency
46	Rozenas, Schutte, & Zhukov (2017)	JOP	Pro-Russian Vote Share	Deportations
47	Marshall (2016)	JOP	Voting Conservative	Extra Year of Schooling
48	de Benedictis-Kessner (2018)	JOP	Vote Share	Incumbency
49	de Benedictis-Kessner & Warshaw (2020)	JOP	County Spending	Democratic Legislator
50	Sances (2017)	JOP	Presidential Vote Share	Tax Increases
51	Schafer & Holbein (2020)	JOP	Election Turnout	One Extra Hour Of Daylight
52	Feierhard (2019)	JOP	Presidential Vote Share	Copartisan Municipal Politician
53	Agira (2015)	JOP	Vote Share	Incumbency
54	de Kadt (2017)	JOP	Election Turnout	Voting Eligibility
55	RozenasÂ & Stukal (2019)	JOP	Media Coverage (Censorship)	Good or Bad News
56	Callen, Gulzar, & Rezaee (2020)	JOP	Assigned Doctors	Governing Party Constituency
57	Fouirnaies & Mutlu-Eren (2015)	JOP	Intergovernmental Transfers	Copartisan Local Councilor
58	Imai, King, & Rivera (2020)	JOP	Incumbent Vote Share	Antipoverty Program
59	Horowitz, et al. (2019)	JOP	Year 2 Prediction Accuracy	Joined Superforecaster Team
60	Fouirnaies & Hall (2020)	JOP	General Election Vote Share	Primary Runoff
61	Carson & Sievert (2017)	JOP	Presidential Vote Share	Congressional Copartisan
62	Cooper, Kim, & Urpelainen (2018)	JOP	Pro-Environmental Voting	Shale Endowment

Table 6: Applied regression discontinuity papers in top political science journals between 2015 and 2020.

A.1 Albertus (2019)

This paper estimates the effect of land reform in Peru on subsequent civil conflict, leveraging the discontinuity in treatment at the geographic boundary between agricultural zones.

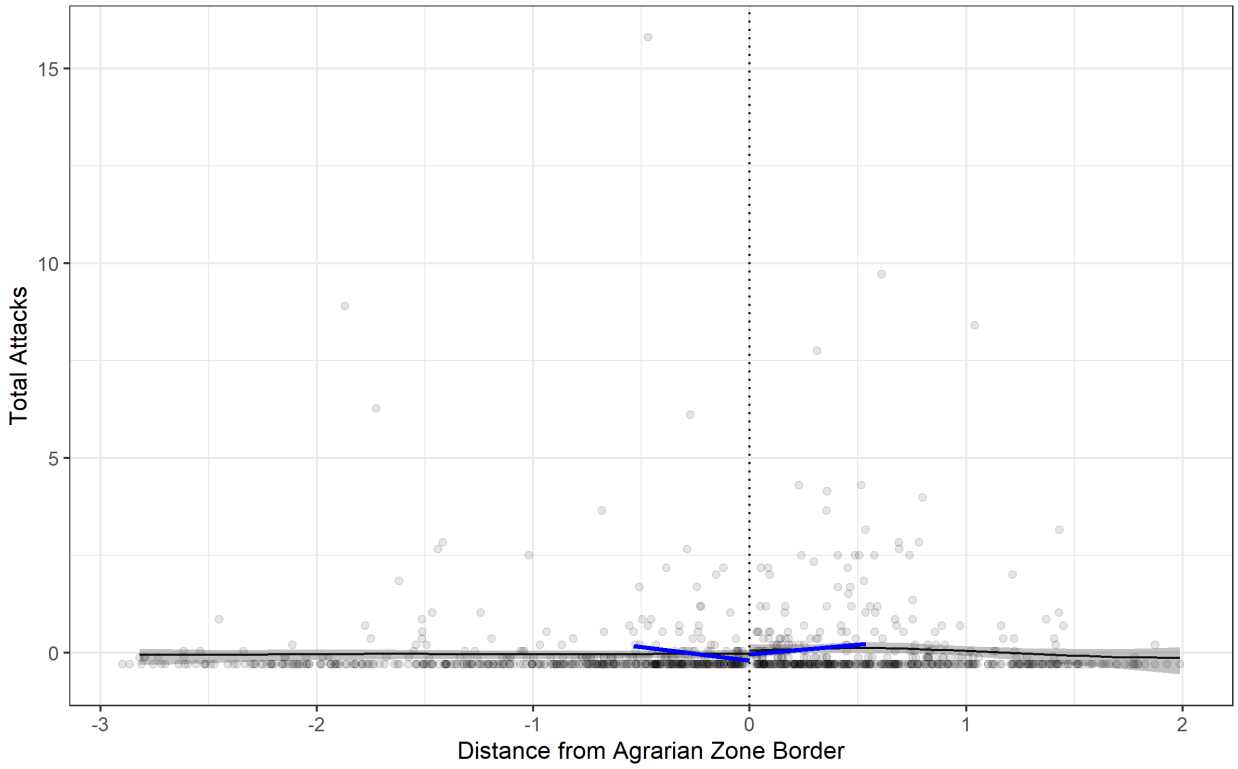


Figure A.1: Local linear and GPRD estimates. Dotted vertical line is the cutoff, solid black lines are the GPRD fit, and solid blue lines are the CCT local linear fit.

	Estimator	Estimate	95% CI
1	CCT	0.16	[-0.092, 0.411]
2	GPRD	0.08	[-0.144, 0.284]

A.2 Beach & Jones (2017)

This paper estimates the effect of city council diversity on public spending, leveraging the discontinuity in treatment when a non-modal candidate's vote margin exceeds zero.

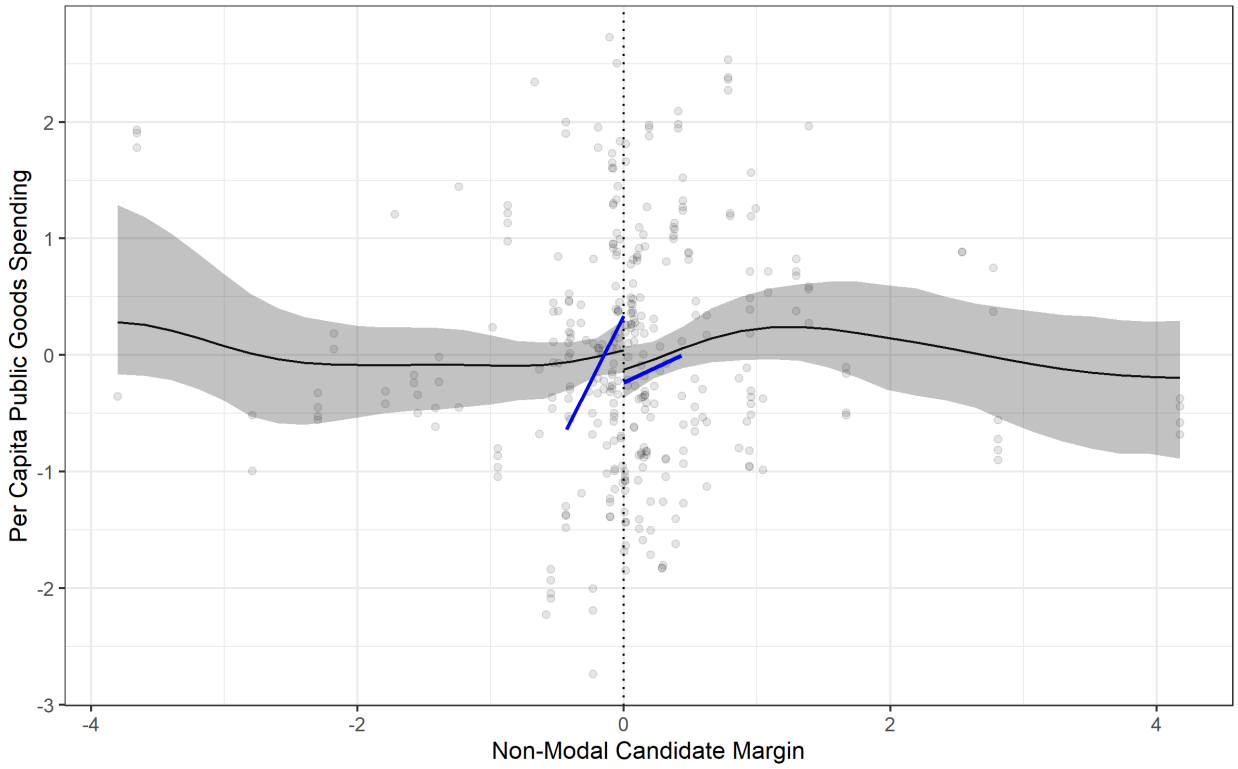


Figure A.2: Local linear and GPRD estimates. Dotted vertical line is the cutoff, solid black lines are the GPRD fit, and solid blue lines are the CCT local linear fit.

	Estimator	Estimate	95% CI
1	CCT	-0.62	[-1.069, -0.167]
2	GPRD	-0.17	[-0.52, 0.114]

A.3 Callen, Gulzar, & Rezaee (2020)

This paper estimates the effect of the governing party in Pakistan on the assignment of doctors, leveraging the discontinuity in treatment when a candidate's margin of victory exceeds zero.

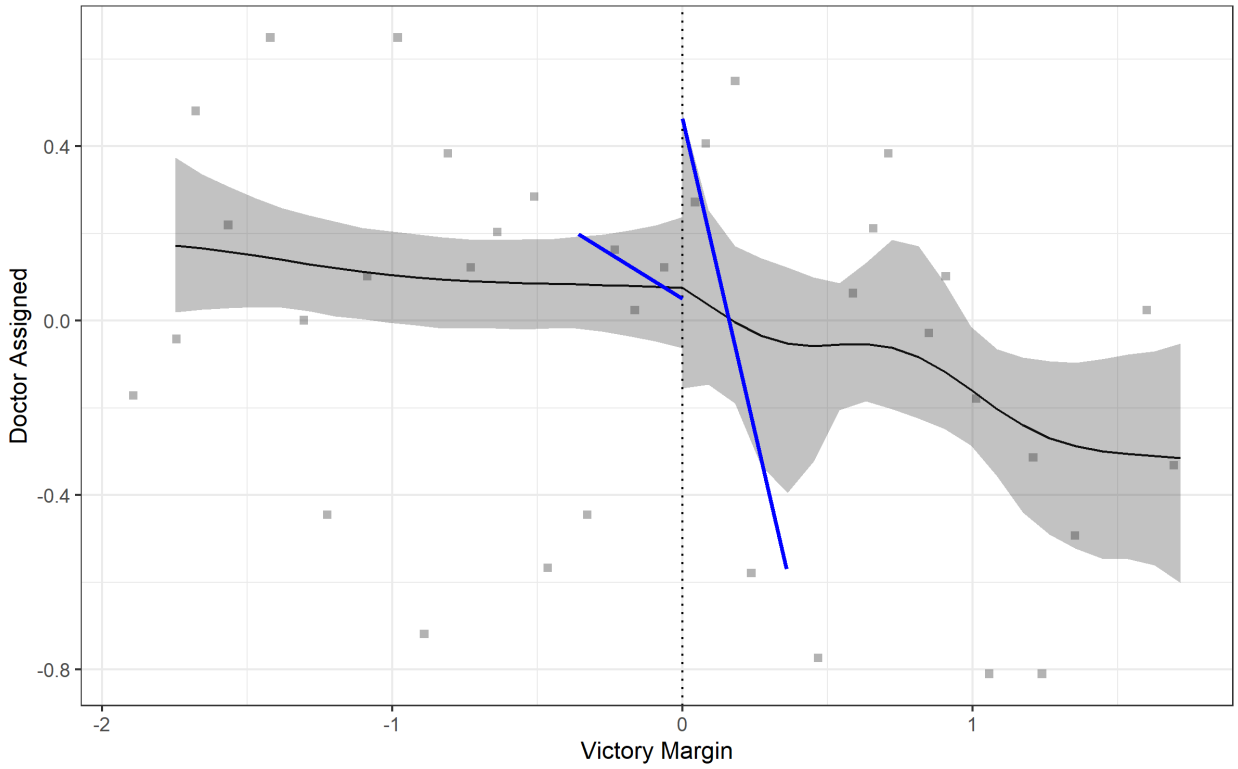


Figure A.3: Local linear and GPRD estimates. Dotted vertical line is the cutoff, solid black lines are the GPRD fit, and solid blue lines are the CCT local linear fit.

	Estimator	Estimate	95% CI
1	CCT	0.47	[-0.111, 1.052]
2	GPRD	-0.00	[-0.288, 0.407]

A.4 Dahlgaard (2018)

This paper estimates the effect of a child's eligibility to vote on the election turnout of parents, leveraging the discontinuity in treatment when the child's birthdate falls on the registration cutoff.

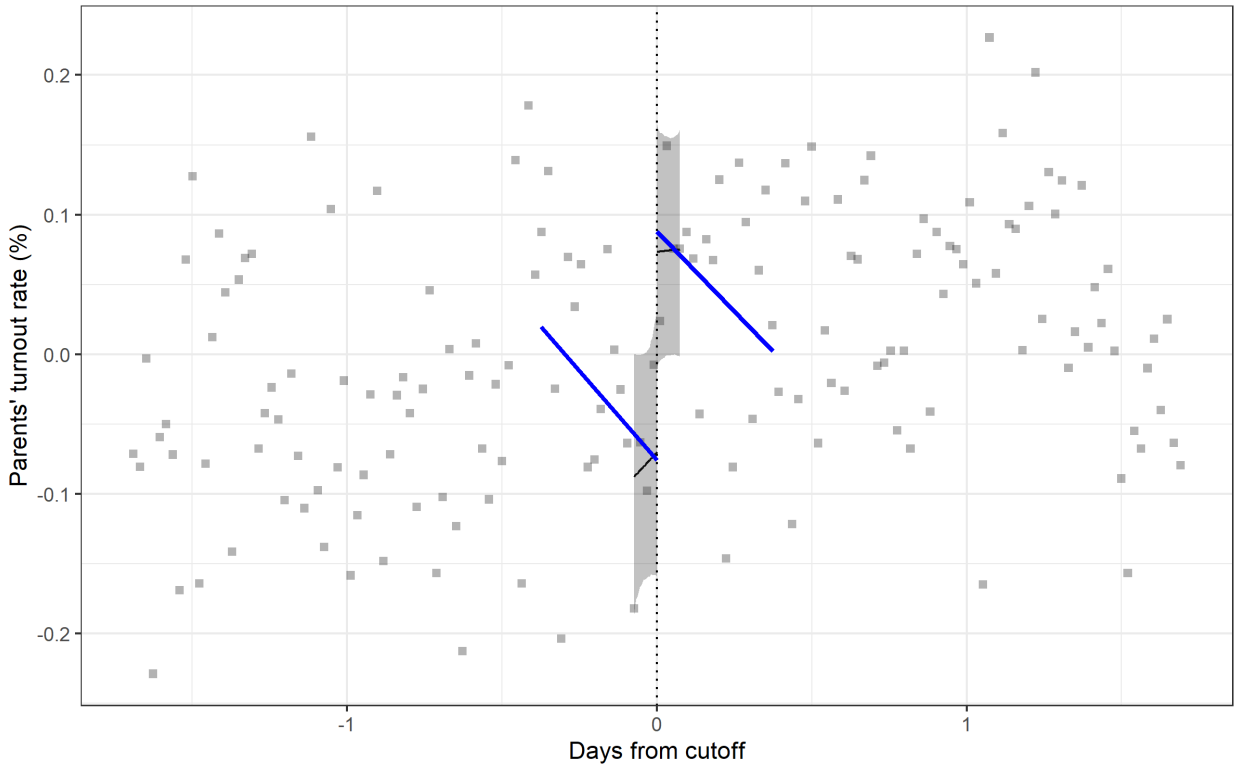


Figure A.4: Local linear and GPRD estimates. Dotted vertical line is the cutoff, solid black lines are the GPRD fit, and solid blue lines are the CCT local linear fit.

	Estimator	Estimate	95% CI
1	CCT	0.18	[0.043, 0.313]
2	GPRD	0.14	[0.017, 0.268]

A.5 de Benedictis-Kessner & Warshaw (2016)

This paper estimates the effect a Democratic mayor on total expenditures in US cities, leveraging the discontinuity in treatment at the plurality threshold of vote share.

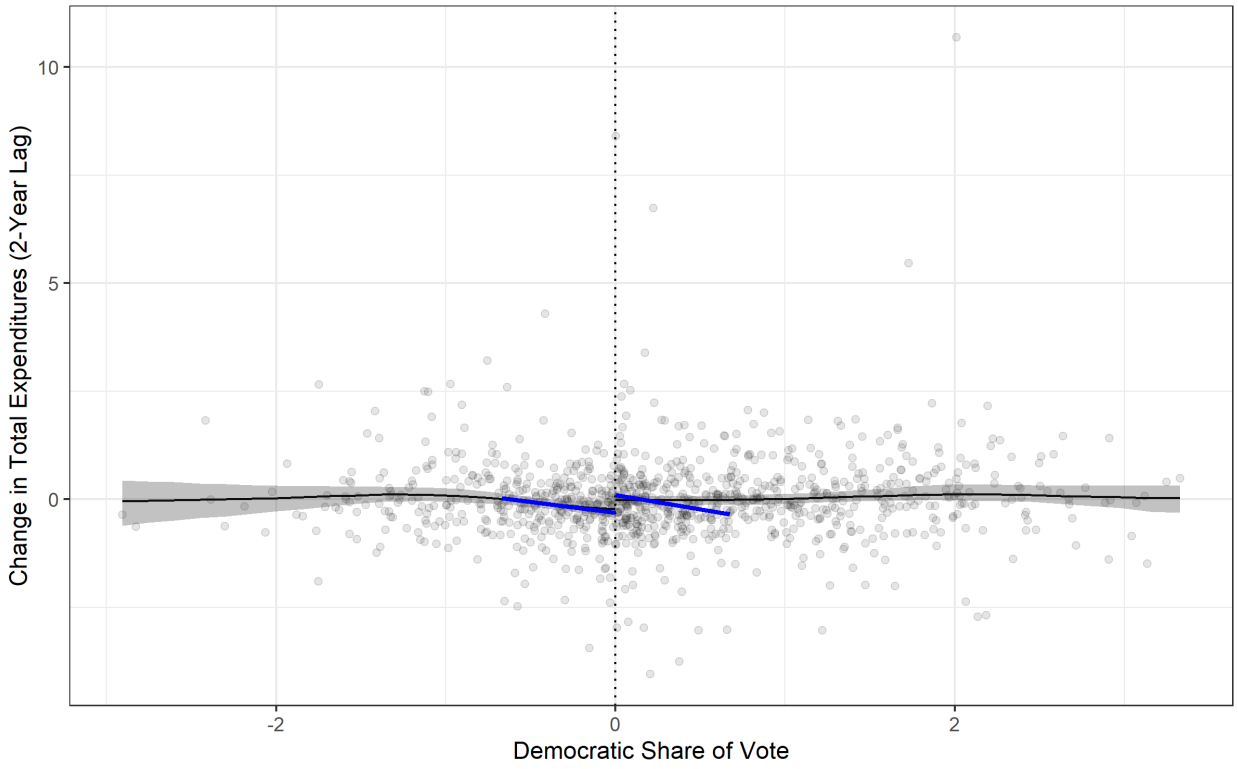


Figure A.5: Local linear and GPRD estimates. Dotted vertical line is the cutoff, solid black lines are the GPRD fit, and solid blue lines are the CCT local linear fit.

	Estimator	Estimate	95% CI
1	CCT	0.46	[0.057, 0.858]
2	GPRD	0.21	[-0.014, 0.452]

A.6 Eggers & Spirling (2017)

This paper estimates the incumbency effect for UK Conservatives vis a vis Liberal Democrats, leveraging the discontinuity in treatment at the plurality threshold of vote share.

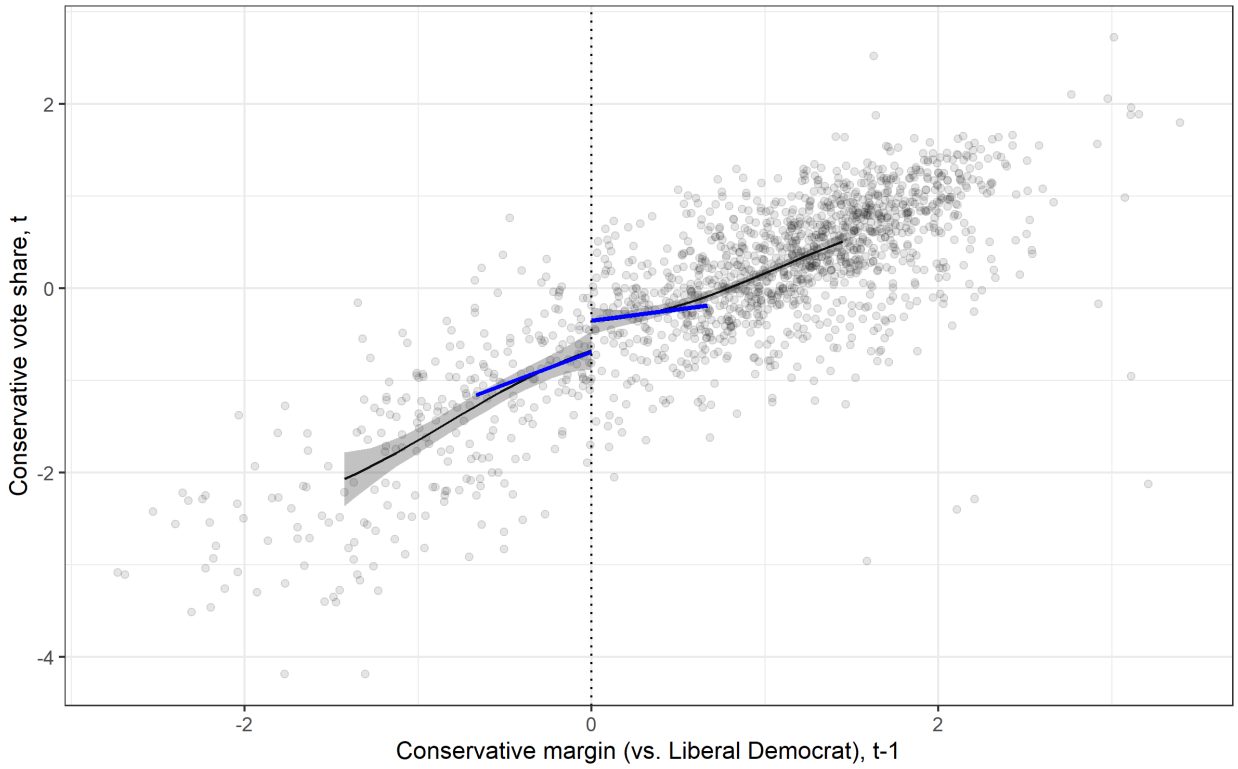


Figure A.6: Local linear and GPRD estimates. Dotted vertical line is the cutoff, solid black lines are the GPRD fit, and solid blue lines are the CCT local linear fit.

	Estimator	Estimate	95% CI
1	CCT	0.39	[0.101, 0.684]
2	GPRD	0.32	[0.06, 0.583]

A.7 Fergusson, Guerubin, Ruiz, & Vargas (2020)

This paper estimates the effect of left-wing party control on right-wing paramilitary violence in Colombia, leveraging the discontinuity in treatment at the plurality threshold of vote share in local elections.

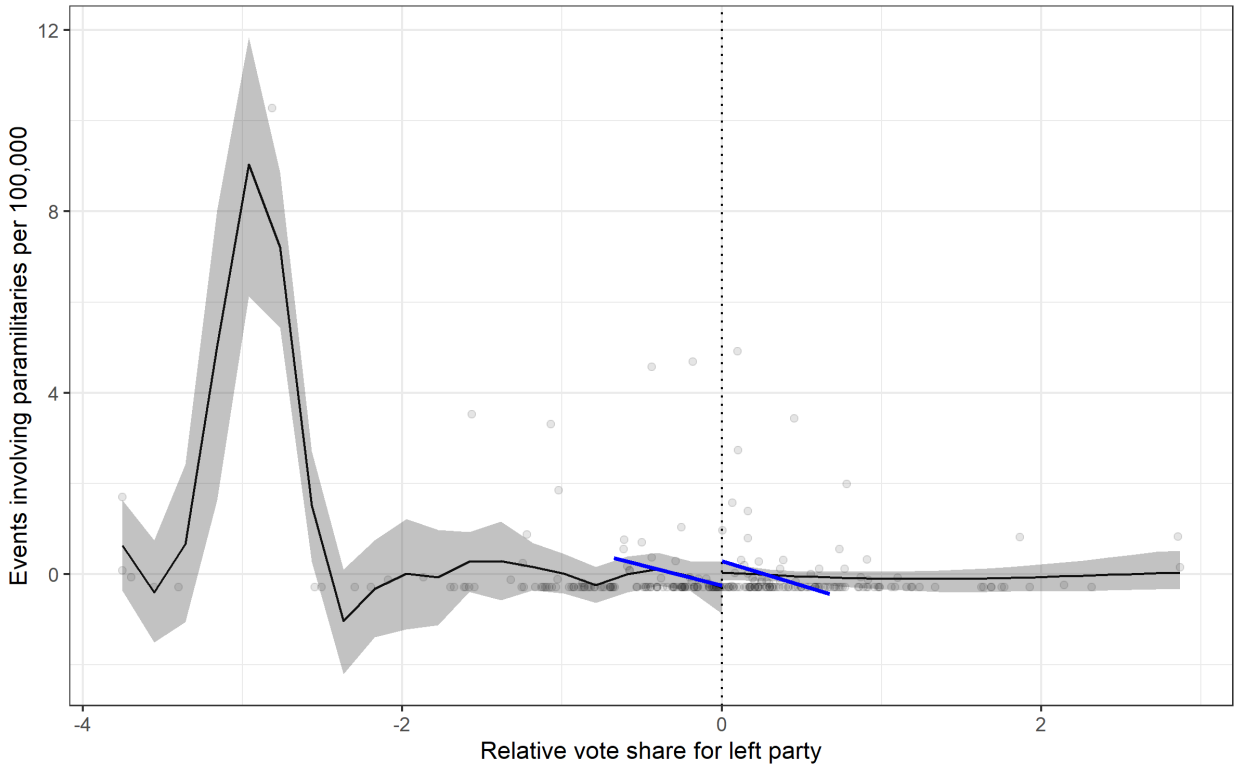


Figure A.7: Local linear and GPRD estimates. Dotted vertical line is the cutoff, solid black lines are the GPRD fit, and solid blue lines are the CCT local linear fit.

	Estimator	Estimate	95% CI
1	CCT	0.64	[0.047, 1.226]
2	GPRD	0.34	[-0.269, 0.935]

A.8 Grumbach & Sahn (2020)

This paper estimates the effect of minority candidates on campaign contributions from minorities, leveraging the discontinuity in treatment at the plurality threshold in primary vote share. There are a series of RD estimates in the paper; below we report the results for Latino Democratic candidates, the largest reported effect size.

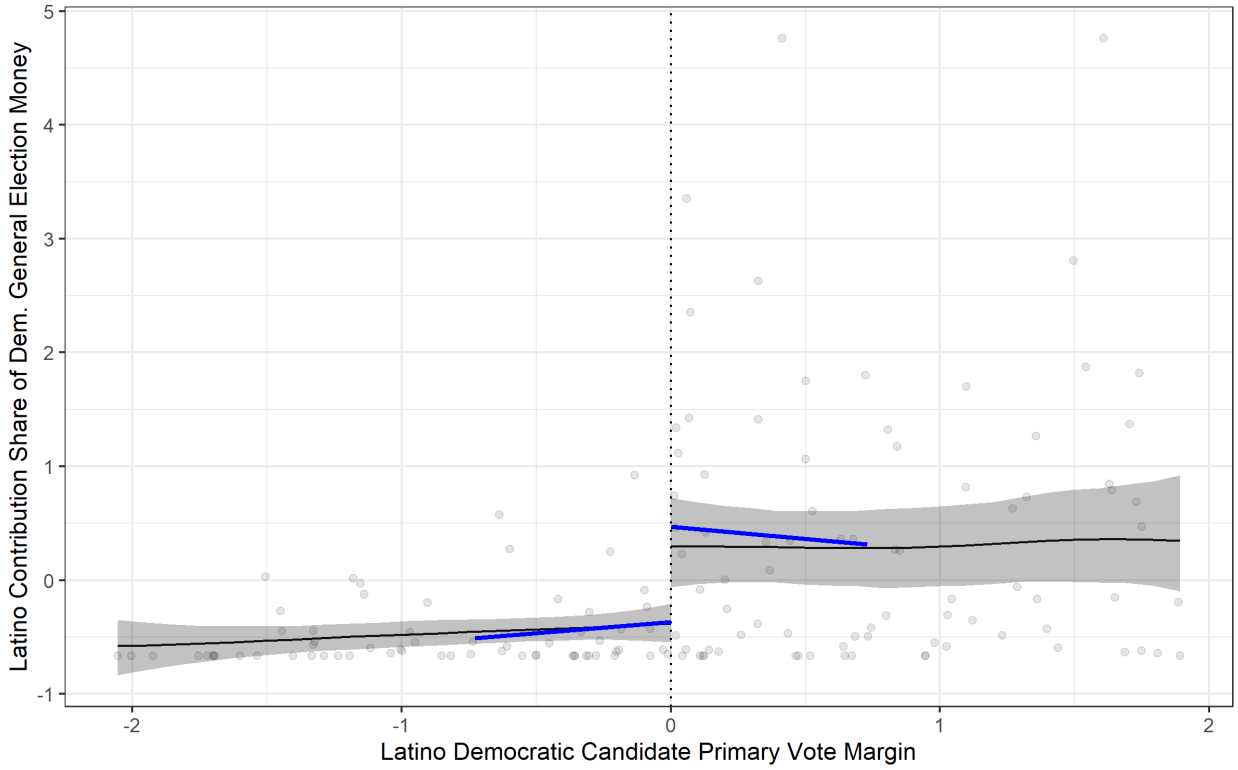


Figure A.8: Local linear and GPRD estimates. Dotted vertical line is the cutoff, solid black lines are the GPRD fit, and solid blue lines are the CCT local linear fit.

	Estimator	Estimate	95% CI
1	CCT	0.81	[0.16, 1.461]
2	GPRD	0.67	[0.263, 1.165]

A.9 Hall (2015)

This paper estimates the effect of nominating an extremist candidate on general election vote share, leveraging the discontinuity in treatment when an extremist candidate's primary vote margin exceeds zero.

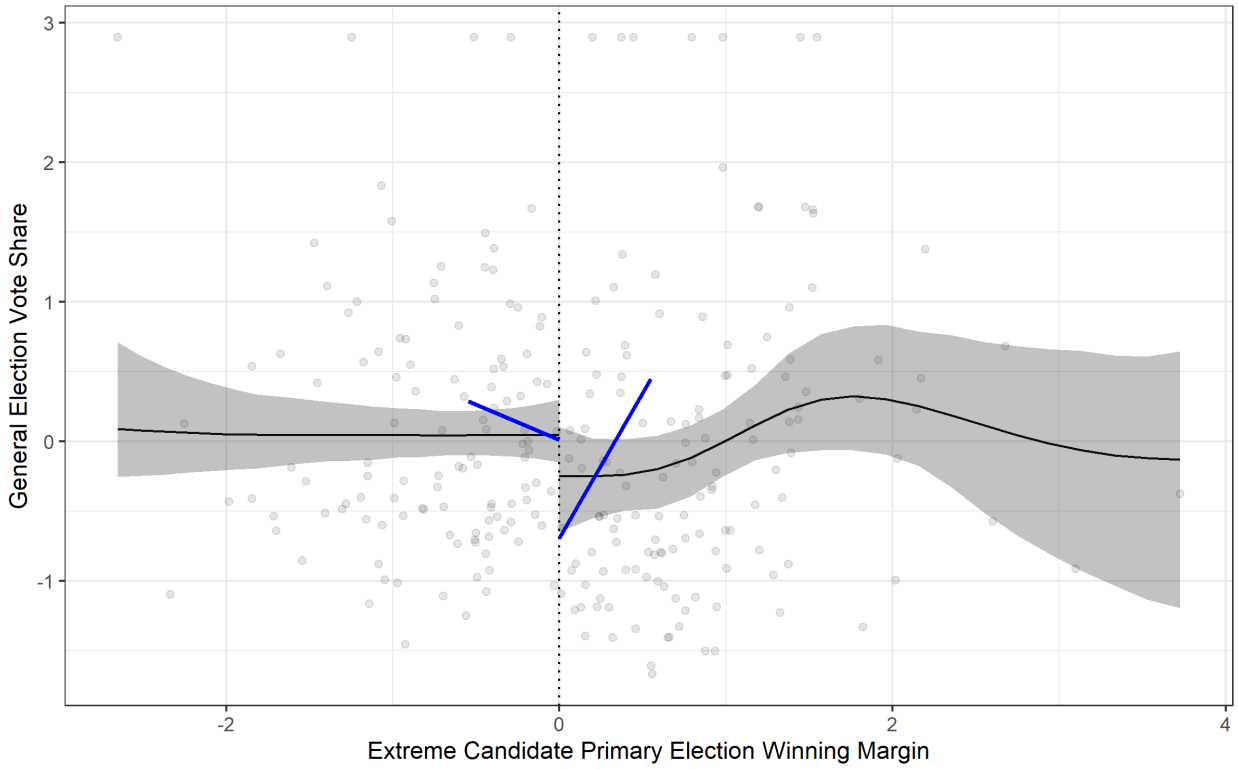


Figure A.9: Local linear and GPRD estimates. Dotted vertical line is the cutoff, solid black lines are the GPRD fit, and solid blue lines are the CCT local linear fit.

	Estimator	Estimate	95% CI
1	CCT	-0.75	[-1.53, 0.032]
2	GPRD	-0.29	[-0.736, 0.108]

A.10 Hall & Thompson (2018)

This paper estimates the effect of nominating an extremist candidate on in-party turnout, leveraging the discontinuity in treatment when an extremist candidate's primary vote margin exceeds zero.

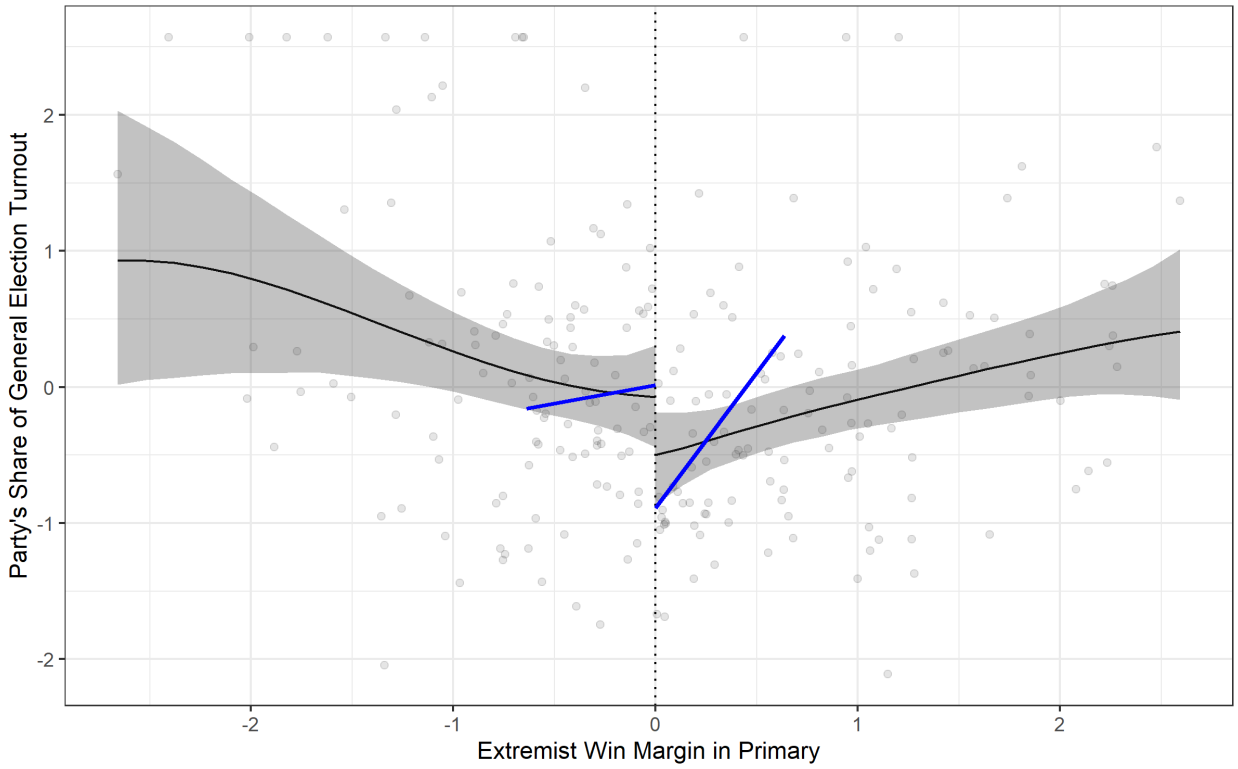


Figure A.10: Local linear and GPRD estimates. Dotted vertical line is the cutoff, solid black lines are the GPRD fit, and solid blue lines are the CCT local linear fit.

	Estimator	Estimate	95% CI
1	CCT	-1.00	[-1.544, -0.447]
2	GPRD	-0.43	[-0.94, 0.063]

A.11 Hidalgo & Nichter (2015)

This paper estimates the effect of voter audits on the re-election probability of incumbents, leveraging the discontinuity in treatment when the electorate as a percentage of the population exceeds 80%.

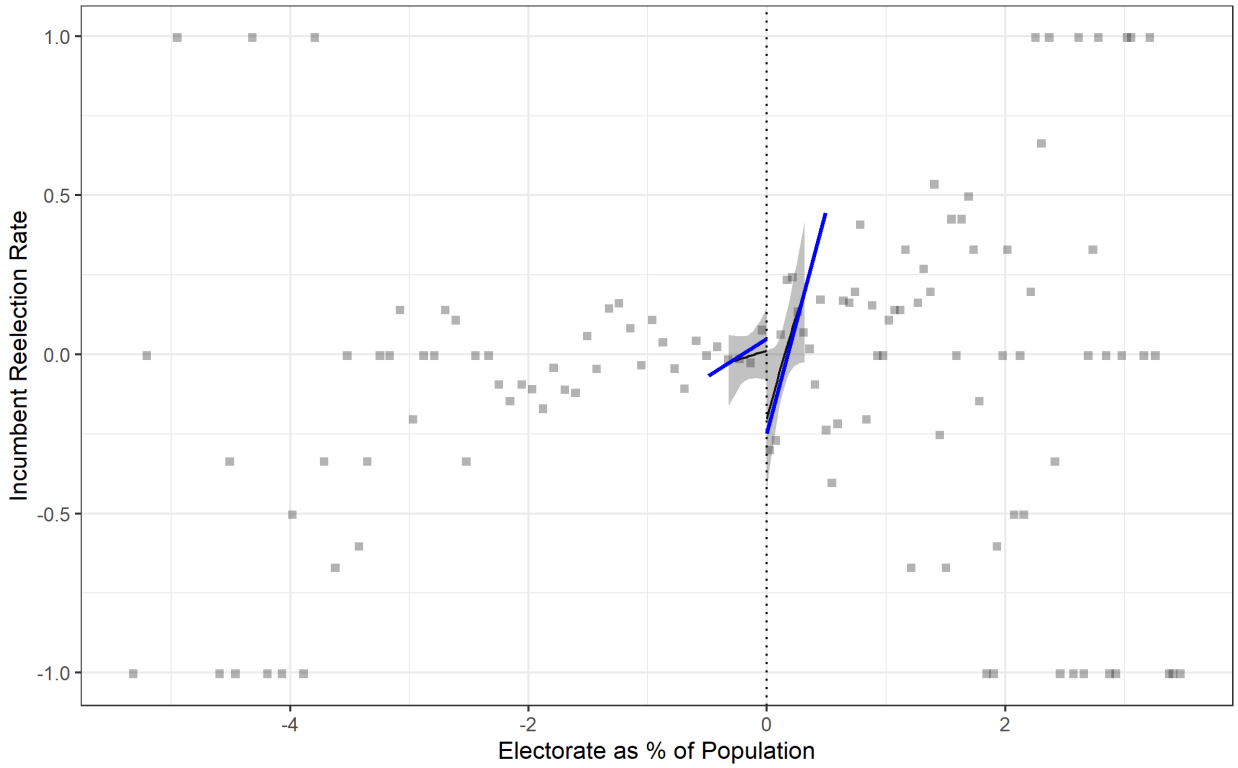


Figure A.11: Local linear and GPRD estimates. Dotted vertical line is the cutoff, solid black lines are the GPRD fit, and solid blue lines are the CCT local linear fit.

	Estimator	Estimate	95% CI
1	CCT	-0.35	[-0.595, -0.099]
2	GPRD	-0.21	[-0.468, 0.026]

A.12 Horowitz, et al. (2019)

This paper estimates the effect of joining a Superforecaster Team on subsequent prediction accuracy, leveraging the discontinuity in treatment when prediction accuracy in Year 1 exceeds a threshold.

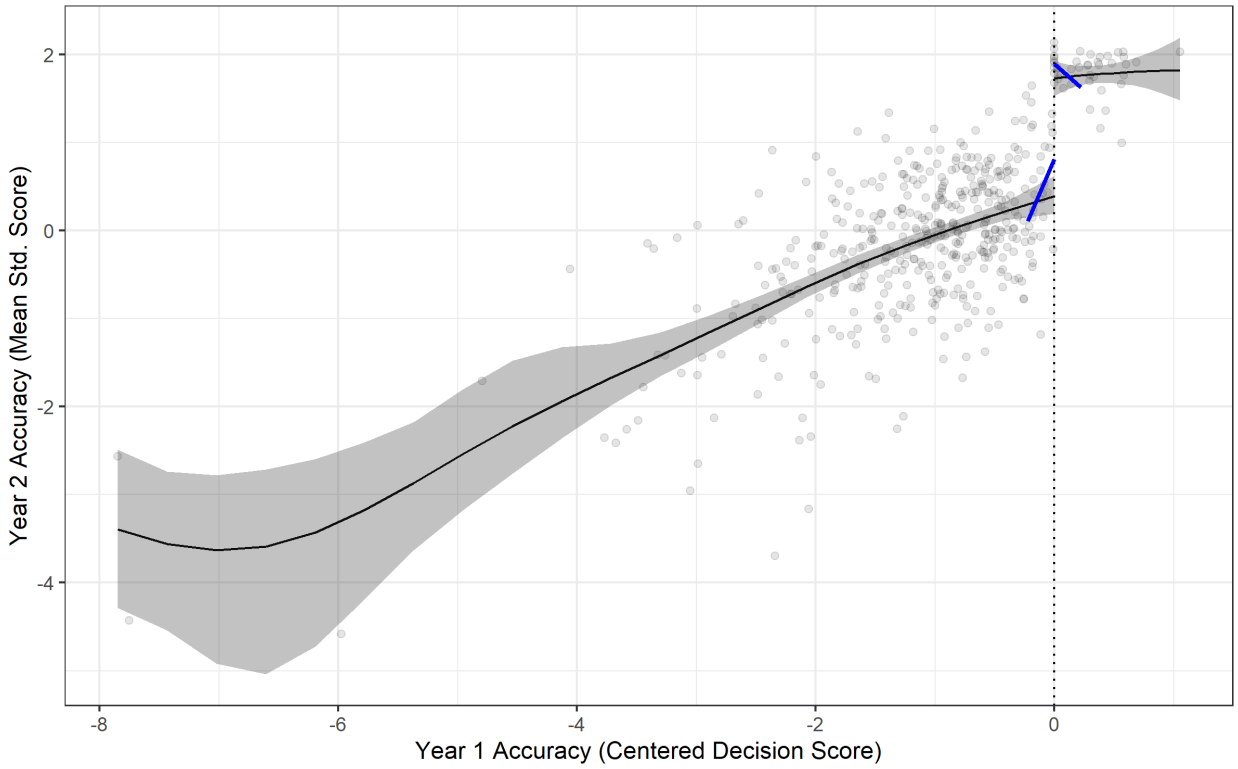


Figure A.12: Local linear and GPRD estimates. Dotted vertical line is the cutoff, solid black lines are the GPRD fit, and solid blue lines are the CCT local linear fit.

	Estimator	Estimate	95% CI
1	CCT	1.04	[0.329, 1.751]
2	GPRD	1.34	[1.052, 1.619]

A.13 Klasnja & Titiunik (2017)

This paper estimates the incumbency disadvantage in Brazil, leveraging the discontinuity in treatment when a mayoral candidate's margin of victory exceeds zero.

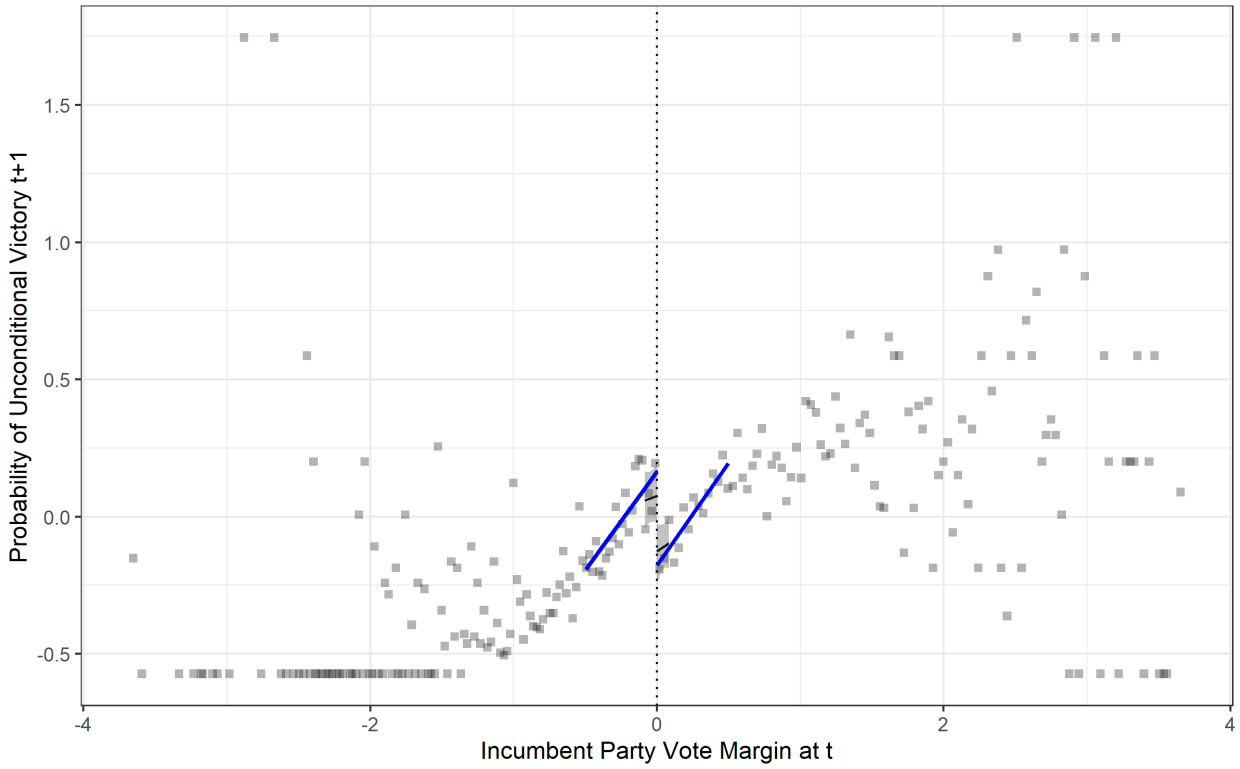


Figure A.13: Local linear and GPRD estimates. Dotted vertical line is the cutoff, solid black lines are the GPRD fit, and solid blue lines are the CCT local linear fit.

	Estimator	Estimate	95% CI
1	CCT	-0.36	[-0.493, -0.235]
2	GPRD	-0.20	[-0.367, -0.06]

A.14 Novaes (2018)

This paper estimates the effect of winning a mayoral election on subsequent party switching, leveraging the discontinuity in treatment when a mayoral candidate's margin of victory exceeds zero.

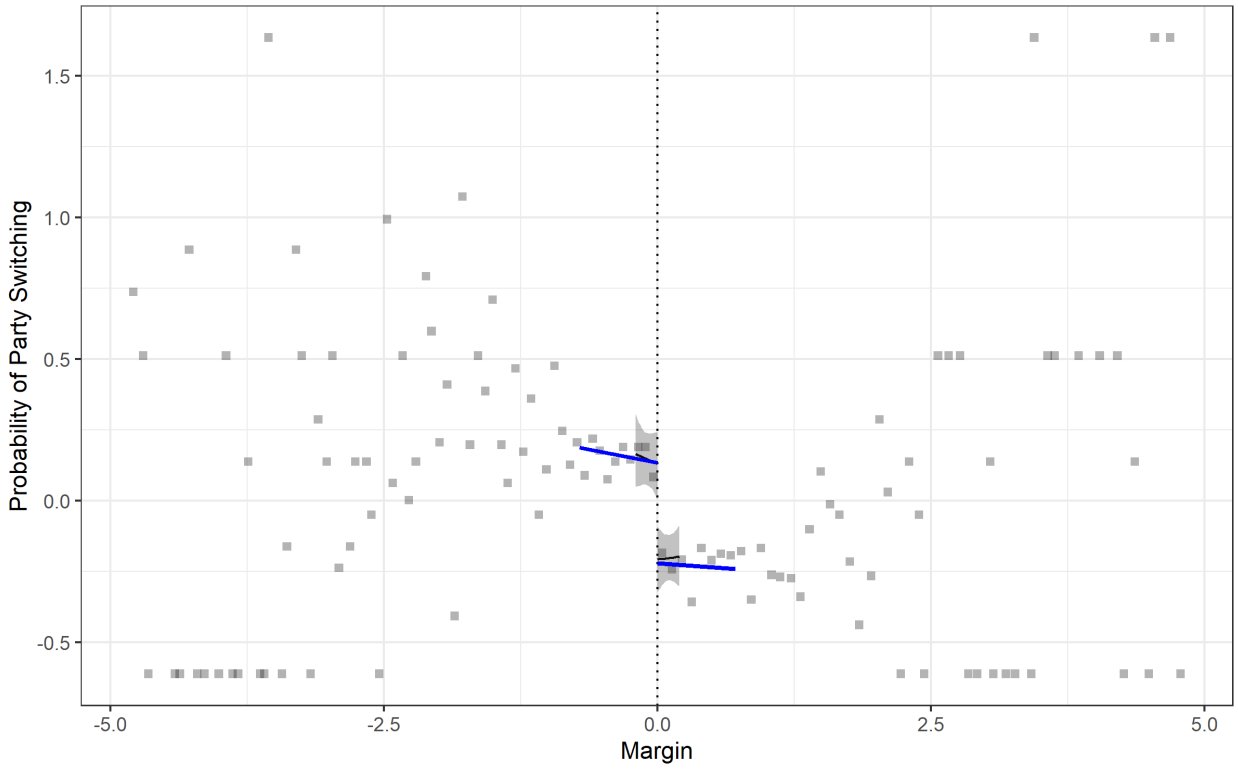


Figure A.14: Local linear and GPRD estimates. Dotted vertical line is the cutoff, solid black lines are the GPRD fit, and solid blue lines are the CCT local linear fit.

	Estimator	Estimate	95% CI
1	CCT	-0.35	[-0.52, -0.182]
2	GPRD	-0.34	[-0.495, -0.173]

A.15 Sances (2017)

This paper estimates the effect of unrelated municipal tax increases on the incumbent President's vote share, leveraging the discontinuity in treatment when a tax-increasing referendum's margin of victory exceeds zero.

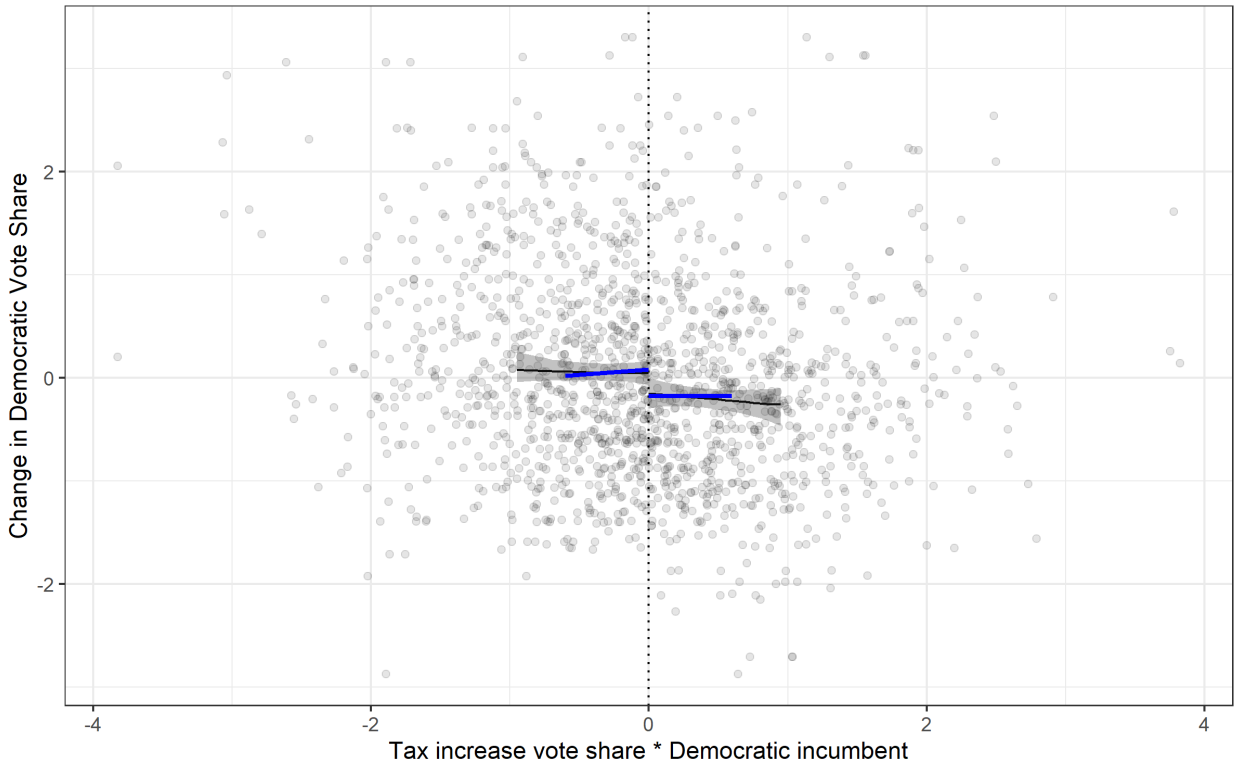


Figure A.15: Local linear and GPRD estimates. Dotted vertical line is the cutoff, solid black lines are the GPRD fit, and solid blue lines are the CCT local linear fit.

	Estimator	Estimate	95% CI
1	CCT	-0.28	[-0.595, 0.036]
2	GPRD	-0.20	[-0.378, -0.017]

A.16 Szakonyi (2018)

This paper estimates the effect of Russian firms' political connections on their subsequent profitability, leveraging the discontinuity in treatment when the vote margin of a politician connected to the firm exceeds zero.

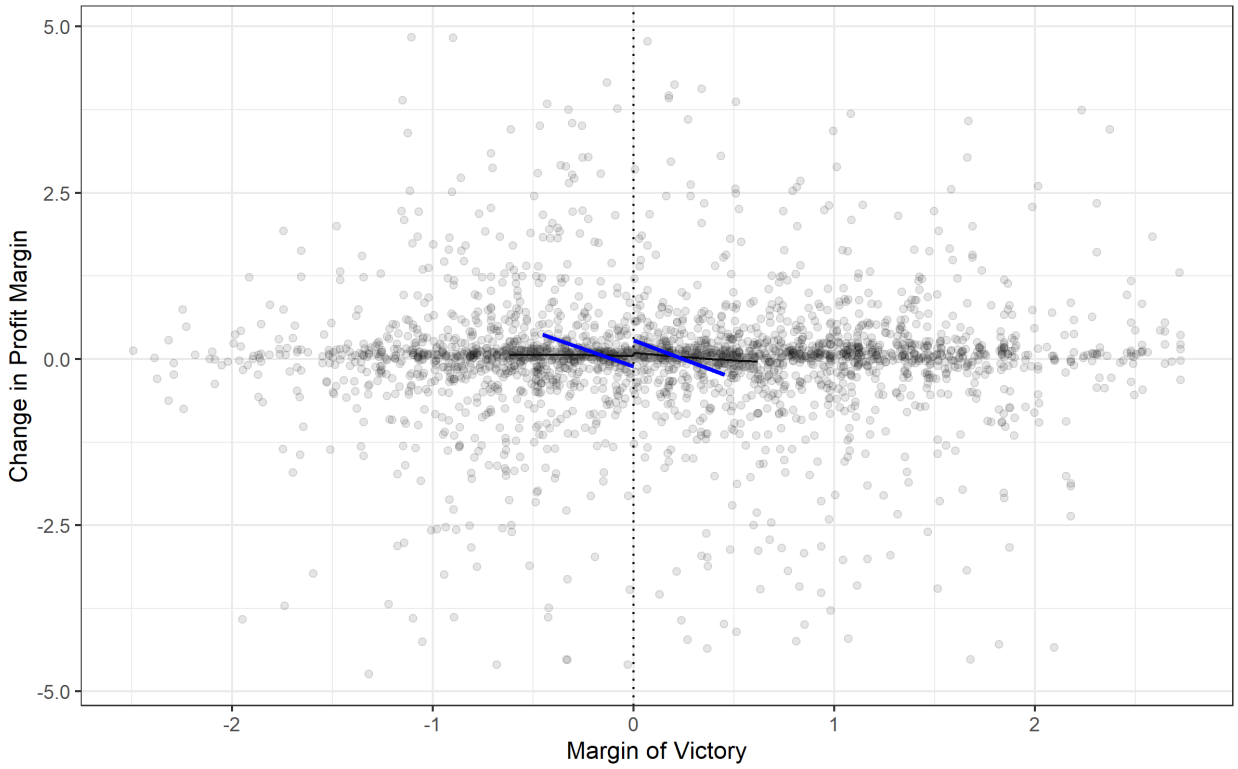


Figure A.16: Local linear and GPRD estimates. Dotted vertical line is the cutoff, solid black lines are the GPRD fit, and solid blue lines are the CCT local linear fit.

	Estimator	Estimate	95% CI
1	CCT	0.46	[0.105, 0.816]
2	GPRD	0.05	[-0.146, 0.306]

A.17 Thompson (2020)

This paper estimates the effect of sheriff partisanship on the probability of complying with federal immigration requests, leveraging the discontinuity in treatment when a sheriff candidate's margin of victory exceeds zero.

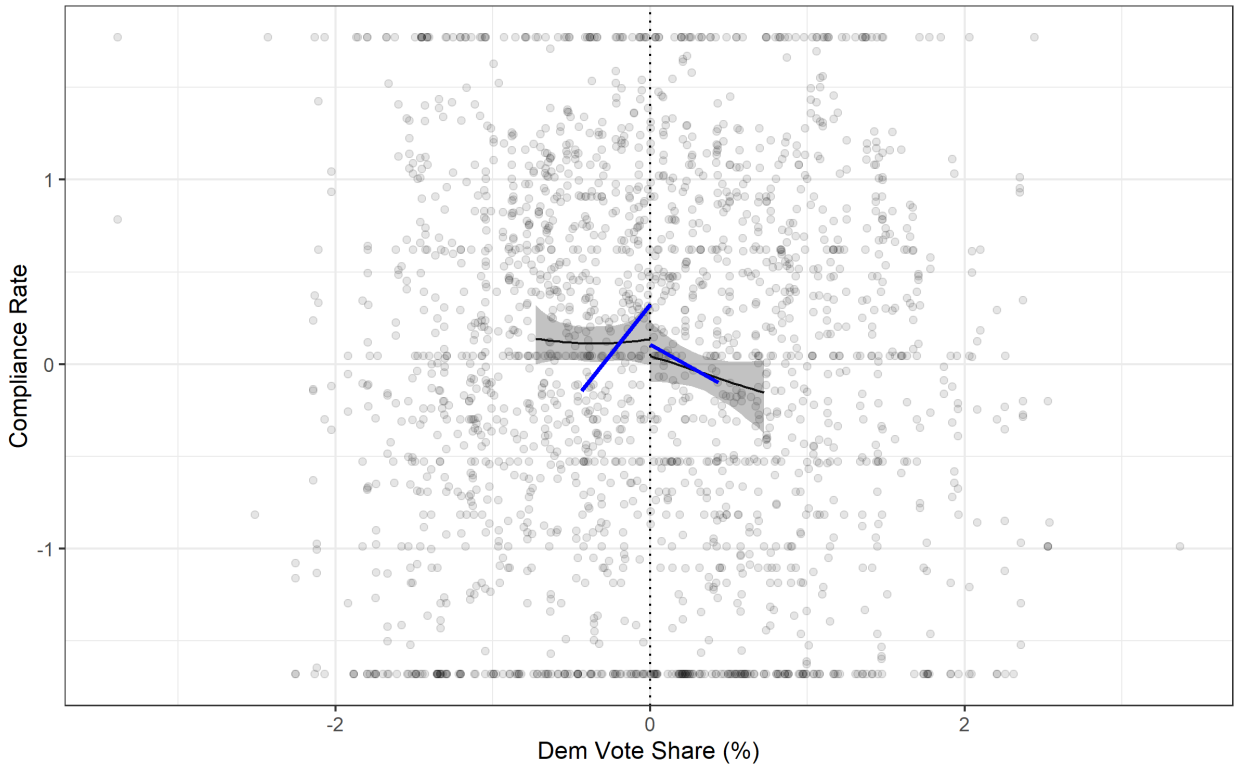


Figure A.17: Local linear and GPRD estimates. Dotted vertical line is the cutoff, solid black lines are the GPRD fit, and solid blue lines are the CCT local linear fit.

	Estimator	Estimate	95% CI
1	CCT	-0.29	[-0.674, 0.1]
2	GPRD	-0.09	[-0.334, 0.138]