Indirect Effects and Causal Inference: Reconsidering Regression Discontinuity

Gary Cornwall*
Bureau of Economic Analysis
and
Beau Sauley
University of Cincinnati

May 16, 2019

Abstract

Causal inference models, like regression discontinuity (RD) design, rely upon some variation of the no-interference assumption, where peer effects or spatial spillovers are null. Given the increased application of network, spatial, and peer effects models, this paper reconsiders RD design when this assumption is not satisfied, yielding indirect effects of the treatment in addition to the traditionally measured direct effects. Using a combination of residualization and numerical integration we develop a method – using the Spatial Durbin framework – which retains the full adjacency matrix and allows for a full accounting of these cross-sectional interactions. As an application, we revisit a well-known RD design using U.S. House of Representatives election results from 1945-1995, finding close election wins have substantial indirect effects which previously were unaccounted.

Keywords: Bayesian, Durbin Models, Regression Discontinuity

*We would like to thank the Taft Research Center for providing generous research support. The authors would like to acknowledge James P. LeSage, Olivier Parent, Jeff Mills, Huiben Weng, Justine Mallatt, Marina Gindelsky, Scott Wentland, and participants of the Midwest Econometrics Group, Southern Regional Science Association Conference, and Southern Economic Association Meetings for their helpful comments. The views expressed here are those of the authors and do not represent those of the U.S. Bureau of Economic Analysis or the U.S. Department of Commerce. First version 05/19/2017.

1 Introduction

Regression discontinuity (RD) design has, over the previous two decades, become a mainstay of the empirical economics research toolbox (see Lee and Lemieux (2010) for a survey). Due to the difficulty, both in cost and ethical considerations, of randomized control trials (RCTs) economists have turned to this quasi-experimental design method to better understand the [causal] impacts of policy changes. First introduced by Thistlethwaite and Campbell (1960) RD has been used to identify the effect of passing school levies on new residential construction (Brasington, 2017), school-board composition on student segregation (Macartney and Singleton, 2017), implementation of gifted or high achieving classrooms on minority student achievement (Card and Giuliano, 2016), traffic congestion following a cessation of public transit services (Anderson, 2014), and class size on student achievement (Angrist and Lavy, 1999), among many other topics. Peer effects, that is the impact changing one's own outcome or characteristics has on neighboring outcomes, is quickly becoming a topic of great interest with respect to causal inference (Athey and Imbens, 2017). The importance of these indirect effects is made clear through violations of the *no-interference* assumption (see Imbens and Rubin (2015) for a discussion) which is critical to the potential outcome framework commonly used in economic literature. This paper reconsiders RD design when this assumption is not satisfied, yielding indirect effects in addition to the traditionally measured direct effects.

RD exploits the common use of pre-specified treatment rules which are linked – through a threshold of eligibility – to some continuous measurement (running variable) in order to estimate the direct effect of treatment on the treated. Since agents cannot precisely control or manipulate their relative position around the threshold, they can be thought of as pseudo-randomized. This randomization produces, at least in theory, comparison groups which differ in expectation only with respect to the treatment effect (Rubin, 1978). The RD framework has been shown to provide close approximations to RCT results in many cases (Cook and Wong, 2008) and thus has become omnipresent in applied economics literature. Examples of continuous measurements used in determining eligibility for treatment include: vote share (Hainmueller and Kern, 2008; Pettersson-Lidbom, 2008; Eggers et al., 2015), test scores and academic ability (Thistlethwaite and Campbell, 1960; Van der Klaauw, 2002; Jacob and Lefgren, 2004), and poverty rates (Ludwig and Miller, 2007; Meng, 2013), among others.

While the RD framework has turned into a nearly ubiquitous structure from which to draw causal statements it is not without its weaknesses; particularly the cross-sectional dependence commonly found in spatial and network models. It is well known that, in the face of such cross-sectional dependence, parameter estimates are potentially both biased and inefficient (Anselin, 1988; LeSage and Pace, 2009; Pace et al., 2011). Additionally, it has been shown that common responses to this dependence, *e.g.* fixed effects and/or clustered standard errors, can exacerbate model misspecification issues including bias, efficiency, and spuriousness of results (Anselin and Arribas-Bel, 2013). The primary issue with these types of models in an RD context is a clear violation of the *no-interference* assumption which underlies much of the potential outcome framework upon which RD is based. In this paper we extend the local-linear RD to a generalized network RD framework – one which nests

both network and non-network base specifications as a special case – using a combination of Bayesian sampling methods, residualization, and numeric integration. Following LeSage and Pace (2017) we use a Monte Carlo study to show the resulting model specification produces estimates with lower bias and mean-square-error (MSE) relative to its peers, even in the nested, network free case. Moreover, the approach contained herein allows for a full examination of both partial and cross-partial derivatives common to both spatial and network models (LeSage and Pace, 2009).

These cross-partial derivatives, known as indirect effects, can be sizable depending on the strength of cross-sectional dependence. In a standard RD framework it is assumed that these cross-partial derivatives are equal to zero and that the total and direct effects of the treatment are equal. Put bluntly, this assumption prohibits the treatment from spilling over to other treated or non-treated units. While this may be a suitable assumption in some limited RCT settings – especially those with strict compliance protocols – in a quasi-experimental framework it is overly restrictive.

Let us abstract away from the RD framework for a moment in an effort to build some intuition. Consider the following example: we have a class of students who are scheduled to take an exam. We would like to know what impact, if any, a tutor would have on the students' final scores. One way this can be done is to randomize which students get access to the treatment; the tutor in this example. As the students enter the classroom we might flip a fair coin and assign a tutor to all students who flip heads. For simplicity assume these tutors are homogeneous in their quality and that the treated students fully comply with the tutor and the tutor's methods. After the exam we compare the scores of treated and non-treated students to estimate the average treatment effect (ATE).

Is the ATE measuring only the effect of tutors on test scores? Given the information outlined above, the answer is no. Cross-sectional dependence is introduced through a number of channels, though the most obvious in retrospect is the students ability to cheat off of one another. An untreated student's test score, in the presence of cheating, depends upon the test score of the treated and untreated students which sat next to him, a violation of the *no-interference* assumption. Moreover, this is not the only channel by which cross-sectional dependence can emerge in our simple example. In addition to cheating students may also form outside study groups comprised of both treated and untreated students. The interaction within these groups confers some of the benefits from the tutor on those that were not selected for treatment. The first channel is a relatively easy one to fix, we could just produce a different exam for every student and thus prevent them from cheating. Alternatively we could have students take the exam in isolation.[1] The second channel is a bit harder to deal with though it is not clear we would want to prevent it. Students studying is a good thing!

How does our RCT example, flawed as it may be, shed light on the issue of cross-sectional dependence in quasi-experimental frameworks such as RD? The data we use in RDs are not collected in an RCT setting. As a result, we may not know which cross-sectional dependence

[1]It is not clear that even if we gave each student a personalized exam that this channel would not produce any cross-sectional dependence. Students may still try to cheat and rather than improve their score from cheating off of their neighbor they may instead make their score worse. In such a situation we would have negative cross-sectional dependence.

inducing channels have been closed. For example, substitute a draw from a standard normal distribution in place of the coin flip in the above example, with a cutoff of zero. Any value drawn at or above the cutoff is treated while any below the cutoff is untreated. This is a classic RD setup in which we could examine, within a certain bandwidth, students around the cutoff and calculate a local average treatment effect (LATE). Without knowing if the aforementioned channels have been closed (*e.g.* Did the professor randomize exams?) we have left ourselves open to bias, inefficiency, and potential spurious results by assuming those channels had been closed. Moreover, as is often the case, we may be evaluating the impact of tutors specifically so that we can make a policy recommendation on public subsidy of tutors. If we do not fully account for the cross-sectional dependence we may overstate (or understate - depending on the direction of the indirect effects) the effectiveness of tutors on test scores, recommending a subsidy which is too large (or too small) and thus create inefficiencies in public financing.

There are plenty of instances in which we would expect cross-sectional dependence to exist in causal studies. First, consider the significant evidence pointing to strong cross-sectional dependence in housing markets (Kim and Goldsmith, 2009; Bin et al., 2011; Mihaescu and vom Hofe, 2012; Wong et al., 2013; Lazrak et al., 2014; vom Hofe et al., 2019) where prices are dependent not only upon the characteristics of one's own home, but also the price and characteristics of neighboring homes. It follows then that any treatment which affects the price of homes would have an effect on surrounding (treated and untreated) homes. Hidano et al. (2015) use an RD framework to examine how buyers in Tokyo evaluate seismic risk via the price premium on the property. They are able to show that properties in a low-risk zone are at a price premium relative to those that are outside of the low-risk zone. While their work acknowledges the presence of cross-sectional dependence, and utilizes a spatial hedonic model in the form of Kelejian and Prucha (1998, 1999) to support their main conclusions, they do not incorporate the dependence into the RD framework itself and admit "... accounting for such cross-sectional interactions in a quasi-experimental framework is challenging and an interesting topic for future research (pp. 121)". A similar example comes in the form of Moulton et al. (2016) which examines the benefits of targeted property tax relief measures in Virginia, using an RD framework, and finds that increased demand stemming from the measure lead to a 5% increase in home values. Again, acknowledging the possibility of these cross-sectional interactions, they employ a spatial fixed effects specification as a robustness check.

Second, and perhaps more well known to most applied economists, is the use of RD to tease out measures of incumbency advantage. Several studies have shown that vote share in the current election is positively impacted by close wins, as measured by margin of victory, in previous elections (see Lee (2008); Chib and Greenberg (2014); Cattaneo et al. (2015) for example). Meanwhile, there is evidence that voting exhibits positive cross-sectional dependence and thus would be exposed to indirect effects (Kim et al., 2003; Cho and Rudolph, 2008; Cutts and Webber, 2010; Lacombe et al., 2014). We can think of several channels by which this cross-sectional dependence may manifest. First, while elections are district specific, the local media coverage area may overlap with multiple districts either in part or whole. Political advertisements would thus have an impact over multiple constituencies, even if those advertisements are for a particular candidate.

Moreover, get out the vote initiatives by local political parties may cross district boundaries as they push for voter turnout in the area more generally rather than in a specific district. This is particularly relevant in situations where several tiers of government may be having simultaneous elections (*e.g.* House and Senate races). We find it interesting that use of RD in this context has become a bit controversial as researchers question if the basic assumptions of RD are being met in the aforementioned studies (see Caughey and Sekhon (2011); Eggers et al. (2015); De la Cuesta and Imai (2016) for a discussion). Yet, none of these criticisms take into account the potential for cross-sectional dependence which would violate the Stable Unit Treatment Value Assumption (Rosenbaum and Rubin, 1983).

By generalizing the RD framework we are able to overcome some of these limitations. We limit our generalization to the local-linear framework for two reasons. First, recent research has shown that RD models using higher-order polynomials (*e.g.* third or fourth degree) in the running variable produce poor results in practice with noisy point estimates and confidence intervals with ill defined coverage (Gelman and Imbens, 2018). Second, and perhaps more importantly, we find that interval construction in the local-linear method has a simple, clean form in a Bayesian context with bandwidth selection – though important for identification purposes – playing a marginal role. Since the distribution of the dependence parameter is of unknown form, we utilize numeric integration to include uncertainty about both initial parameters and the bandwidth in our estimate of the local average treatment effect (LATE). In each iteration we are recalculating the bandwidth based on our uncertainty about the residualization parameters. This creates a frontier of marginal observations which move in and out of the estimation sample throughout the algorithm. For computational efficiency we limit ourselves to bandwidth calculations as outlined by Imbens and Kalyanaraman (2012), although in our empirical example we relax this requirement and show bandwidth is not particularly relevant in that context.

Finally, we illustrate this new model specification using the now canonical example of close elections in the U.S. House of Representatives (Lee, 2008; Imbens and Kalyanaraman, 2012; Calonico et al., 2014). We show that U.S. House districts are in fact spatially correlated and that this correlation impacts the estimates in a material way. Our results show that close wins in a district during time t leads to an increase in vote share at time $t+1$ of approximately 15% with roughly 9% attributable to the direct effects of treatment on the treated, and approximately 6% the result of indirect effects. Original work by Lee (2008) and follow up work by Caughey and Sekhon (2011) yield estimates of a direct effect between 7% and 9% depending on the specification. Possible mechanisms which would produce the indirect effects include a shift in ad-buying resources, increased voter turnout following a nearby win, and voter migration. These dimensions are not captured in the data so the channels by which the indirect effects manifest themselves are speculative.

The remainder of this paper is organized as follows. Section 2 outlines the proposed model specification including interpretation of the resulting parameter(s). Section 3 includes a Monte Carlo study to examine the properties of the proposed specification. In Section 4 we contextualize the generalized RD framework using close election in the U.S. House of Representatives. Finally, Section 5 concludes and offers avenues for additional research.

2 Generalizing the RD Design

This section is presented through three subsections. The first outlines the general RD framework with focus on the local-linear structures. The second provides background on network models. Additionally, we utilize these first two subsections to establish the mechanical framework and notation used in outlining our proposed model specification outlined in the third subsection.

2.1 Regression Discontinuity Design

RD is a method which has received quite a bit of renewed attention in recent years. Its primary purpose is to draw out estimates which are causal in nature by exploiting a deterministic relationship between a continuous running variable (RV) and a dichotomous treatment. For concreteness, suppose there is some treatment which is a monotonic, deterministic, and discontinuous function of some continuous variable, z_i, where $i = (1, \ldots, N)$. For our purposes we refer to z_i as the value of the RV for the i^{th} individual, and Z to denote the $N \times 1$ vector of values for each of the observations in a sample. Further, suppose that the treatment is of the form,

$$q_i = \begin{cases} 1 & z_i \geq Z_0 \\ 0 & z_i < Z_0 \end{cases} \tag{1}$$

where Z_0 is a sharp cutoff for treatment eligibility and application. Again, we use q_i to denote treatment of the i^{th} agent and Q refers to an $N \times 1$ vector of indicators for the entire sample. Note that we, for ease of exposition, confine ourselves to the sharp RD case.

The interest here is to estimate,

$$E[(y_i|q_i = 1) - (y_i|q_i = 0)|z_i = z], \tag{2}$$

where $(y_i|q_i = 1)$ is the observed outcome and $(y_i|q_i = 0)$ is the corresponding counterfactual. Since the counterfactual is unobserved, RD relies on estimating the average treatment effect by utilizing the discontinuity of Y at the cutoff, Z_0. Adopting notation from Gelman and Imbens (2018) this can be written as,

$$\gamma = \lim_{Z \downarrow Z_0} E(y_i^{obs}|z_i = z) - \lim_{Z \uparrow Z_0} E(y_i^{obs}|z_i = z). \tag{3}$$

Surprisingly there is a number of ways one could estimate the treatment effect, γ, including local-linear methods, kernel regression, high-order polynomials, and others. For our purposes we focus on local-linear methods, that is a pair of linear regressions on either side of the cutoff, for two reasons. First, this allows for the slope on either side of the cutoff to vary (Lee and Lemieux, 2010), a parameter we see no need to constrain as equal. Second, it has recently been shown that inferences drawn from high-order polynomial models are untrustworthy, with noisy point estimates and confidence intervals with ill-defined coverage (Gelman and Imbens, 2018). Using a local-linear model in a Bayesian framework produces transparent credible intervals using standard probability theory, a point to which we will return later.

Prior to estimating a local-linear model it is necessary to establish a bandwidth, b, which is the subset of observations deemed to be quasi-randomized, also known as *compliers*. Observations outside of the bandwidth are considered to be either *always-takers*, if above the threshold, or *never-takers* if below, and consequently are discarded to remove potential bias (Angrist et al., 1996; Hahn et al., 2001). As one would expect, since adherence to the bandwidth requires a researcher to discard data, there is quite the discussion regarding selection of an optimal bandwidth. The methodology here is presented using the bandwidth constructed by Imbens and Kalyanaraman (2012), hereafter IK, though Calonico et al. (2014), hereafter CCT, is more commonly used in current empirical work. The reason for our use of IK will become apparent later. It is important to note that construction of the bandwidth and subsequent discarding of data is done solely for identification purposes. It facilitates the imitation of an RCT and allows for the parameter estimate to be viewed as causal rather than correlative.

Going forward, and for concreteness, we will use $\bar{\cdot}$ to denote observations above the cutoff and $\underline{\cdot}$ for those below. Local-linear methods estimate the treatment parameter, γ, using only those observations in b, through the use of two equations:

$$\bar{Y} = \bar{\boldsymbol{\gamma}}\bar{\boldsymbol{Z}} + \bar{\epsilon}, \quad \bar{\epsilon} \sim N(0, \bar{\sigma}^2), \tag{4}$$

$$\underline{Y} = \underline{\boldsymbol{\gamma}}\underline{\boldsymbol{Z}} + \underline{\epsilon}, \quad \underline{\epsilon} \sim N(0, \underline{\sigma}^2), \tag{5}$$

where $\bar{\boldsymbol{Z}} = [\iota_{\bar{n}} \ (\bar{Z} - Z_0)]$, $\bar{\boldsymbol{\gamma}} = [\bar{\gamma} \ \bar{\zeta}]'$, $\underline{\boldsymbol{Z}} = [\iota_{\underline{n}} \ (\underline{Z} - Z_0)]$, $\underline{\boldsymbol{\gamma}} = [\underline{\gamma} \ \underline{\zeta}]'$. By centralizing the RV through $(\bar{Z} - Z_0)$ and $(\underline{Z} - Z_0)$ the intercept of each equation provides the value at Z_0, or more directly, the difference between them, $\gamma = \bar{\gamma} - \underline{\gamma}$, outlines the size of the discontinuity at the threshold, which is the parameter of interest. Again, it is important to recognize that this is a local average treatment effect for the *compliers* and any generalizations to the average treatment effect on the population should be tempered by the lack of external validity in the method.

2.2 Network Models

Peer effects, that is the impact changing one's own outcome or characteristics has on neighboring outcomes, is quickly becoming a topic of great interest with respect to causal inference (Athey and Imbens, 2017). The importance of peer effects is made clear through violations of the *no-interference* assumption (see Imbens and Rubin (2015) for a discussion) which is critical to the potential outcome framework commonly used in economic literature. Our main purpose in this section is to provide the notational framework we will use going forward. Networks are characterized in a variety of ways, however a few attributes are relatively ubiquitous throughout the literature. First, each network is typically represented by a graph, $\boldsymbol{\mathcal{G}}$. this graph is made up of vertices, $\boldsymbol{\mathcal{V}}$, and edges, $\boldsymbol{\mathcal{E}}$. The vertices are indexed by $i = 1, \ldots, N$, are typically a finite set, and are the unit of observation in the analysis (*e.g.* agents, firms, regions, etc.). An edge is formed when two agents are connected, that is $ij \in \boldsymbol{\mathcal{E}}$.

Algebraically these graphs are represented through an adjacency or weight matrix which we will denote as $\boldsymbol{G}$. Let $g_{ij} = 1$ if an edge exists between agents i and j and zero otherwise. In a directed graph it is not necessary that the connection be reciprocated, that

is $g_{ij} \neq g_{ji}\ \forall i, j$, while in an undirected graph $g_{ij} = g_{ji}\ \forall\ i, j$. We assume that agents cannot be connected to themselves, that is $g_{ii} = 0\ \forall\ i$, and as a result the maximum number of connections is $N^2 - N$, though in practice this matrix tends to be quite sparse. Finally, we assume that $\boldsymbol{G}$ is right-stochastic for reasons which will be made clear shortly.

Using the above notation we can write the following model,

$$y_i = \alpha + \rho \sum_{j=1}^{N} g_{ij} y_j + \sum_{k=1}^{K} \beta^k x_i^k + \phi \sum_{j=1}^{N} g_{ij} x_j + \epsilon_i,\ \epsilon_i \sim \mathrm{N}(0, \sigma^2). \tag{6}$$

Which we can write in matrix form as,

$$Y = \rho \boldsymbol{G} Y + X\beta + \boldsymbol{G} X \phi + \epsilon,\ \epsilon\ \mathrm{N}(0, I_N \sigma^2), \tag{7}$$

with the data generating process written as,

$$Y = (I_N - \rho \boldsymbol{G})^{-1}(X\beta + \boldsymbol{G} X \phi + \epsilon). \tag{8}$$

As defined earlier Y is the $N \times 1$ vector of outcomes, X is an $X \times K$ information matrix. In this structure both $\boldsymbol{G}Y$ and $\boldsymbol{G}X$ represent the linear combination of connected agent outcomes and characteristics respectively. Astute readers will recognize this as the Durbin class of model commonly used in spatial econometrics (LeSage, 2014). The parameter ρ is a scalar which informs on the strength of the network dependence and is bound to the interval $(1/\psi_{min}, 1)$ where ψ_{min} is the minimum real eigenvalue of $\boldsymbol{G}$. Both β and ϕ are $K \times 1$ vectors of parameters to be estimated. Finally, the noise, ϵ is assumed to be independently and identically distributed (i.i.d) following a Normal distribution with mean zero and variance $I_N \sigma^2$.[2]

It is important to note that the partial derivative from the network model outlined above differs dramatically from those of a standard linear regression. For the k^{th} variable the partial derivative for the form can be expressed as,

$$\delta y / \delta x^k = (I_N - \rho \boldsymbol{G})^{-1}(I_N \beta^k + \boldsymbol{G} \phi^k), \tag{9}$$

while the latter is $n^{-1} tr(I_N \beta^k)$, where $tr(\cdot)$ is the trace operator; this simplifies to β. Since (9) is a dense matrix representing a global spillover structure it is useful to convert the information into scalar summaries (LeSage and Pace, 2009). Letting $S_k(\boldsymbol{G}) = \frac{\delta y}{\delta x^k}$, the total impact to an observation are $n^{-1} \iota_n' (S_k(\boldsymbol{G}) \iota_n)$ while the total impact from an observation is $n^{-1}(\iota_n' S_k(\boldsymbol{G})) \iota_n$, where ι_n is a N-dimensional vector of ones (LeSage and Pace, 2009). In practice, these are the same value and are known as the *total* effects. Total effects can be decomposed into direct and indirect effects:

$$\underbrace{n^{-1}(\iota_n' S_k(\boldsymbol{G})) \iota_n}_{\text{Total Effects}} = \underbrace{n^{-1} tr(S_k(\boldsymbol{G}))}_{\text{Direct Effects}} + \underbrace{\left(n^{-1}(\iota_n' S_k(\boldsymbol{G})) \iota_n - n^{-1} tr(S_k(\boldsymbol{G}))\right)}_{\text{Indirect Effects}} \tag{10}$$

[2]For ease of exposition we will confine the discussion to the homoskedastic case however this can be expanded for heteroskedastic errors without loss of generality, see LeSage and Pace (2009) for a full discussion.

The *direct* effects, that is the impact of changes in one's own characteristics as well as feedback stemming from changes in one's neighbors, are $n^{-1}tr(S_k(\boldsymbol{G}))$, where $tr(\cdot)$ is the trace operator. The *indirect* effects are simply the *total* minus the *direct* effects and represent the average cross-partial derivative. These summary measures reflect how changes propagate through the simultaneous dependence structure and result in a new steady state. It is also important to consider that, at least inferentially, it is not necessarily vital for any single parameter of the three outlined to be significant on its own. The effect of the non-linear relationship implies that even if ρ is not statistically distinguishable from zero, which many would take as evidence of no dependence, the outlined effects can have well defined distributions which do not include zero within any reasonable credible interval.

2.3 Network Regression Discontinuity Design

The previous two sections have shown that in order to properly identify the treatment effect in an RD setting it is preferable, if not necessary, to cut away the data which represents *always* and *never* takers. This leaves a quasi-randomization of the remaining observations around the cutoff – thereby approximating an RCT – which is then subject to a local-linear model to estimate the treatment effect. This is in direct conflict with the concept of network models, which shows that the connectivity structure of the network is important to accurate inference from any model.

The root issue of incorporating these cross-sectional interactions lays with the structure of the adjacency or weight matrix. This matrix defines the connectivity structure – how information is transmitted between the observations – and likely includes interactions between *always* takers, *never* takers, and compliers. Removing these observations introduces bias in the parameter estimates since the outcomes of compliers may be dependent upon those of the discarded data or their characteristics.

We solve this problem through a multi-step estimation procedure which incorporates elements of residualization and numeric integration nested within a Bayesian sampler. The first stage requires us to residualize the outcome by estimating parameters of interest in the model outlined by 7; think of this as netting out the variation from characteristics which are not our treatment while at the same time estimating the scalar parameter governing strength of the network. Residualization is by no means novel in and of itself, it has become quite popular over recent years (Sales and Hansen, 2014; Chernozhukov et al., 2017; Terrier and Ridley, 2017; Cooper, 2018). Since the residuals are a function of an unknown parameter, ρ, and an observed connectivity structure, $\boldsymbol{G}$, we employ numerical integration to marginalize ρ as a nuisance parameter.

Consider the following,

$$p(\beta, \phi, \sigma^2, \rho | Y, X, \boldsymbol{G}) \propto \pi(\beta, \phi)\pi(\sigma^2)\pi(\rho)p(Y, X, \boldsymbol{G}|\beta, \phi, \sigma^2, \rho), \tag{11}$$

where $p(Y, X, \boldsymbol{G}|\beta, \phi, \sigma^2, \rho)$ is the likelihood, $\pi(\beta, \phi, \sigma^2)$ and $\pi(\rho)$ represent suitable priors over the listed parameters, and $p(\beta, \phi, \sigma^2, \rho|Y, X, \boldsymbol{G})$ is the joint posterior distribution. Assuming the process is linear, or is appropriately modeled as such, and utilizing the

common Normal Inverted-Gamma (NIG) structure we can write (11) as,

$$p(\beta,\phi,\sigma^2,\rho) \propto \sigma^{-n-1}|A| \\ \times \exp\left[-\frac{1}{2\sigma^2}\big((AY-X\beta-\boldsymbol{G}X\phi)'(AY-X\beta-\boldsymbol{G}X\phi)\big)\right]\pi(\rho), \tag{12}$$

where $A = (I_N - \rho\boldsymbol{G})$ and $\pi(\rho)$ is the prior distribution on ρ. Typically, this is either a Uniform or Beta distribution (see LeSage and Pace (2009) for a discussion). Note we have included the intercept as a column of ones in X. This is the posterior form for equation (7). First, let $B = [\beta', \phi']'$, $X_{\boldsymbol{G}} = [\iota_N \; X \; \boldsymbol{G}X]$ $\pi(B) \sim N(b_0, V_0)$, $\pi(\sigma^2) \sim IG(a_0, c_0)$, and $\pi(\rho) \sim Beta(a1, a2)$. This leads to a sampler based on the following conditional distributions,

$$\begin{aligned} p(B_{(m)}|\sigma^2_{(0)},\rho_{(0)}) &\sim N(D_B d_B, \sigma^2 D_B) \qquad (13)\\ D_B &= (X'_{\boldsymbol{G}}X_{\boldsymbol{G}} + V_B^{-1})^{-1} \\ d_B &= X'_{\boldsymbol{G}}(I-\rho\boldsymbol{G})Y + V_B b_B \\ p(\sigma^2_{(m)}|\rho_{(0)},\beta_{(m)}) &\sim IG(a,c) \qquad (14)\\ a &= a_0 + N/2; \\ c &= c_0 + (AY - X_{\boldsymbol{G}}B)'(AY - X_{\boldsymbol{G}}B)/2 \\ A &= (I-\rho\boldsymbol{G}) \end{aligned}$$

$$p(\rho_{(m)}|\beta_{(m)},\sigma^2_{(m)}) \propto |I-\rho\boldsymbol{G}| \exp\left(\frac{1}{2\sigma^2}(AY - X_{\boldsymbol{G}}B)'(AY - X_{\boldsymbol{G}}B)\right) \tag{15}$$

Note that Equation (15) is of unknown form leading to either a draw by integration and inversion (aka 'Griddy Gibbs') or a Metropolis-Hastings (M-H) step (LeSage and Pace, 2009). So far, we have broken no new ground and these results are well known; it is here where we depart from the standard approach.

Having obtained estimates for B, σ^2, and ρ we residualize the outcome by subtracting the conditional mean, $(I-\rho\boldsymbol{G})^{-1}X_{\boldsymbol{G}}B$. However, we cannot simply use this residualized outcome since it is still in fact a function of ρ and $\boldsymbol{G}$. Consider the following,

$$\epsilon = Y - (I-\rho_{(m)}\boldsymbol{G})^{-1}X_{\boldsymbol{G}}B_{(m)}, \tag{16}$$

which implies that,

$$\epsilon \sim N\big(0, (I-\rho_{(m)}\boldsymbol{G})^{-1}(I-\rho_{(m)}\boldsymbol{G})^{-1'}\tau\big). \tag{17}$$

Using this, we filter ϵ by numerically integrating over ρ and B. That is to say for each draw (m) of ρ and B in the sampler we create a new vector of filtered residuals by,

$$\tilde{\epsilon}_{(m)} = (I-\rho_{(m)}\boldsymbol{G})\big(Y - (I-\rho_{(m)}\boldsymbol{G})^{-1}X_{\boldsymbol{G}}B_{(m)}\big), \tag{18}$$

which results in,

$$\tilde{\epsilon}_{(m)} \sim N\big(0, I\tau\big), \tag{19}$$

where $m = 1, \ldots, M$ and M is a suitable number of post burn-in iterations through the sampler.

Here we have residualized the outcome as mentioned in Lee and Lemieux (2010) and then, recognizing that the residuals are correlated, transformed them through our draw of the dependence parameter. Subjecting ϵ to a Moran I test for cross-sectional correlation would tend to reject the null of no correlation, while the opposite is true for $\tilde{\epsilon}$. It is important to remember that, since ρ is an unknown parameter with unknown form, the filtering of residuals with each draw from the joint posterior ensures uncertainty from the initial parameter estimates will propagate through remainder of the sampler.

Now that we have a residual, $\tilde{\epsilon}$, which is orthogonal to $\boldsymbol{G}$, X, $\boldsymbol{G}X$, and ρ we can consider establishing a bandwidth around the cutoff Z_0. Here, for the sake of brevity and computational efficiency, we use methods from IK to calculate a bandwidth. Note that we do this M number of times over the course of our sampler, and as a result our estimate of the bandwidth will change at each iteration. This creates a frontier of marginally relevant observations which move into and out of the sampler over the draws. Since these bandwidths are data driven and each time our data is changing based on uncertainty around the residualizing parameters, we are marginalizing the bandwidth as a nuisance parameter much like we did with the network itself.

Again, we assume that the resulting treatment can be modeled using a local-linear structure and using a NIG structure, see Equations (4) and (5), for both observations above and below the cutoff we sample from the following conditional distributions:

$$
\begin{aligned}
p\left(\bar{\boldsymbol{\gamma}}_{(m)}|\bar{\tau}_{(0)}, \bar{\tilde{\epsilon}}_{(m)}, (\bar{Z} - Z_0)\right) &\sim N(D_{\bar{\boldsymbol{\gamma}}} d_{\bar{\boldsymbol{\gamma}}}, \bar{\tau} D_{\bar{\boldsymbol{\gamma}}}) \qquad (20)\\
D_{\bar{\boldsymbol{\gamma}}} &= (\bar{\boldsymbol{Z}}'\bar{\boldsymbol{Z}} + V_{\bar{\boldsymbol{\gamma}}}^{-1})^{-1}\\
d_{\bar{\boldsymbol{\gamma}}} &= \bar{\boldsymbol{Z}}'\bar{\tilde{\epsilon}} + V_{\bar{\boldsymbol{\gamma}}} b_{\bar{\boldsymbol{\gamma}}}\\
p\left(\bar{\tau}_{(m)}|\bar{\tilde{\epsilon}}_{(m)}, (\bar{Z} - Z_0), \bar{\boldsymbol{\gamma}}_{(m)}\right) &\sim IG(\bar{a}, \bar{c}) \qquad (21)\\
\bar{a} &= \bar{a}_0 + \bar{N}/2\\
\bar{c} &= \bar{c}_0 + (\bar{\tilde{\epsilon}} - \bar{\boldsymbol{Z}}\bar{\boldsymbol{\gamma}})'(\bar{\tilde{\epsilon}} - \bar{\boldsymbol{Z}}\bar{\boldsymbol{\gamma}})/2\\
p\left(\underline{\boldsymbol{\gamma}}_{(m)}|\underline{\tau}_{(0)}, \underline{\tilde{\epsilon}}_{(m)}, (\underline{Z} - Z_0)\right) &\sim N(D_{\underline{\boldsymbol{\gamma}}} d_{\underline{\boldsymbol{\gamma}}}, \underline{\tau} D_{\underline{\boldsymbol{\gamma}}}) \qquad (22)\\
D_{\underline{\boldsymbol{\gamma}}} &= (\underline{\boldsymbol{Z}}'\underline{\boldsymbol{Z}} + V_{\underline{\boldsymbol{\gamma}}}^{-1})^{-1}\\
d_{\underline{\boldsymbol{\gamma}}} &= \underline{\boldsymbol{Z}}'\underline{\tilde{\epsilon}} + V_{\underline{\boldsymbol{\gamma}}} b_{\underline{\boldsymbol{\gamma}}}\\
p\left(\underline{\tau}_{(m)}|\underline{\tilde{\epsilon}}_{(m)}, (\underline{Z} - Z_0), \underline{\boldsymbol{\gamma}}_{(m)}\right) &\sim IG(\underline{a}, \underline{c}) \qquad (23)\\
\underline{a} &= \underline{a}_0 + \underline{N}/2\\
\underline{c} &= \underline{c}_0 + (\underline{\tilde{\epsilon}} - \underline{\boldsymbol{Z}}\underline{\boldsymbol{\gamma}})'(\underline{\tilde{\epsilon}} - \underline{\boldsymbol{Z}}\underline{\boldsymbol{\gamma}})/2
\end{aligned}
$$

where $\bar{\boldsymbol{\gamma}}$ and $\underline{\boldsymbol{\gamma}}$ represent the vector of parameters, γ and ζ, above and below the cutoff first introduced in Equations (5) and (4). We use $\bar{\boldsymbol{Z}}$ to denote the matrix $[\bar{\iota} \;\; \bar{Z}]$, where $\bar{\iota}$ is

a vector of ones above the cutoff, and $\boldsymbol{Z}$ below. Draws from these conditional distributions produce a filtered local average treatment effect (LATE), $\gamma = \bar{\gamma} - \underline{\gamma}$.

Since there is a great deal of consternation with respect to appropriate inference from RD type models (Calonico et al., 2014; Gelman and Imbens, 2018) we would like to take a few moments and point out what is rather transparent in a Bayesian sense but not so much when using frequentist estimation methods. The LATE, $\gamma = \bar{\gamma} - \underline{\gamma}$, is the difference between two conditional Gaussian distributions, which produces a new Gaussian distribution with mean $\mu_1 - \mu_2$ and variance $\sigma_1^2 + \sigma_2^2 - 2\text{Cov}(\cdot)$, where $\text{Cov}(\cdot)$ is the covariance. In Section 3 we show, through simulations, that the intervals produced for γ by the process outlined herein are nearly equivalent to those found by using interval estimates from Calonico et al. (2014).

Obtaining a filtered estimate of the LATE puts us in similar position to current RD methodology. However, as mentioned earlier we are interested in being able to evaluate a policy which may produce spillovers relevant to its evaluation.

To accomplish this we return to the original model in our sampler and, conditional upon the data and the drawn LATE, sample from the following conditional distributions:

$$p\big(B_{(m)}|\tilde{\sigma}^2_{(0)}, \tilde{\rho}_{(0)}, \gamma_{(m)} = (\bar{\gamma}_{(m)} - \underline{\gamma}_{(m)})\big) \sim N(\tilde{D}_B \tilde{d}_B, \tilde{\sigma}^2 \tilde{D}_B) \tag{24}$$

$$\tilde{D}_B = (X'_{\boldsymbol{G}} X_{\boldsymbol{G}} + V_B^{-1})^{-1}$$

$$\tilde{d}_B = X'_{\boldsymbol{G}}(AY - \gamma Q) + V_B b_B$$

$$p\big(\tilde{\sigma}^2_{(m)}|\tilde{\rho}_{(0)}, \gamma_{(m)} = (\bar{\gamma}_{(m)} - \underline{\gamma}_{(m)}\big), B_{(m)}) \sim IG(\tilde{a}, \tilde{c}) \tag{25}$$

$$\tilde{a} = a_0 + N/2$$

$$\tilde{c} = c_0 + (AY - X_{\boldsymbol{G}} B - \gamma Q)'(AY - X_{\boldsymbol{G}} B - \gamma Q)/2$$

$$p\big(\tilde{\rho}_{(m)}|\gamma_{(m)} = (\bar{\gamma}_{(m)} - \underline{\gamma}_{(m)}\big), B_{(m)}, \tilde{\sigma}^2_{(m)}\big) \propto |I - \tilde{\rho}\boldsymbol{G}|$$

$$\times \exp\left(\frac{1}{2\tilde{\sigma}^2}(AY - X_{\boldsymbol{G}} B - \gamma Q)'(AY - X_{\boldsymbol{G}} B - \gamma Q)\right). \tag{26}$$

Recall that $B = [\beta', \phi']'$, Q is the vector of treatment indicators outlined in Equation (1), and the estimates for both β and ϕ are conditional upon the $\gamma = \bar{\gamma} - \underline{\gamma}$ values. This is particularly relevant since the estimates which produce γ come from a smaller sample size, and as mentioned earlier, the variance of the LATE is equal to the sum of its component variances. By conditioning upon these draws from equations (20) and (22) we carry that uncertainty into the full sample estimates of the other parameters.

The full algorithm can be written as:

Algorithm 1:

1. Establish parameters on prior distributions, arbitrary starting values for each parameter, and number of iterations with suitable burn-in period.
2. Sample from $p(B_{(1)}|\sigma^2_{(0)}, \rho_{(0)}) \sim N(D_B d_B, \sigma^2_{(0)} D_B)$
3. Sample from $p(\sigma^2_{(1)}|\rho_{(0)}, B_{(1)}) \sim IG(a, c)$

4. Sample from

$$p(\rho_{(1)}|B_{(1)},\sigma^2_{(1)}) \propto |I-\rho_{(1)}\boldsymbol{G}| \times \exp\left(\frac{1}{2\sigma^2_{(1)}}(AY-X_{\boldsymbol{G}}B_{(1)})'(AY-X_{\boldsymbol{G}}B_{(1)})\right) \quad (27)$$

using "Griddy Gibbs" or M-H.
5. Create vector of filtered residuals via $\tilde{\epsilon}_{(1)} = (I-\rho_{(1)}\boldsymbol{G})(Y-(I-\rho_{(1)}\boldsymbol{G})^{-1}XB_{(0)})$
6. Calculate the bandwidth using chosen method conditional on $\tilde{\epsilon}_{(1)}$.
7. Sample from $p\big(\bar{\boldsymbol{\gamma}}_{(1)}|\bar{\tau}_{(0)},\bar{\epsilon}_{(1)},(\bar{Z}-Z_0)\big) \sim N(D_{\bar{\gamma}}d_{\bar{\gamma}},\bar{\tau}D_{\bar{\gamma}})$
8. Sample from $p\big(\underline{\boldsymbol{\gamma}}_{(1)}|\underline{\tau}_{(0)},\underline{\epsilon}_{(1)},(\underline{Z}-Z_0)\big) \sim N(D_{\underline{\gamma}}d_{\underline{\gamma}},\underline{\tau}D_{\underline{\gamma}})$
9. Sample from $p\big(\bar{\tau}_{(1)}|\bar{\boldsymbol{\gamma}}_{(1)},\tilde{\epsilon}_{(1)},(\bar{Z}-Z_0)\big) \sim IG(\bar{a},\bar{c})$
10. Sample from $p\big(\underline{\tau}_{(1)}|\underline{\boldsymbol{\gamma}}_{(1)},\tilde{\epsilon}_{(1)},(\underline{Z}-Z_0)\big) \sim IG(\underline{a},\underline{c})$
11. Sample from $p\big(B_{(1)}|\tilde{\sigma}^2_{(0)},\tilde{\rho}_{(0)},\gamma_{(1)}=(\bar{\gamma}_{(1)}-\underline{\gamma}_{(1)})\big) \sim N(\tilde{D}_B\tilde{d}_B,\tilde{\sigma}^2\tilde{D}_B)$
12. Sample from $p\big(\tilde{\sigma}^2_{(1)}|\tilde{\rho}_{(0)},\gamma_{(1)}=(\bar{\gamma}_{(1)}-\underline{\gamma}_{(1)}),B_{(1)}\big) \sim IG(\tilde{a},\tilde{c})$
13. Sample from

$$p\big(\tilde{\rho}_{(1)}|\gamma_{(1)}=(\bar{\gamma}_{(1)}-\underline{\gamma}_{(1)}),B_{(1)},\tilde{\sigma}^2_{(1)}\big) \propto |I-\tilde{\rho}_{(1)}\boldsymbol{G}| \times \exp\left(\frac{1}{2\tilde{\sigma}^2_{(1)}}(AY-\bar{X}B_{(1)}-\gamma_{(1)}Q)'(AY-\bar{X}B_{(1)}-\gamma_{(1)}Q)\right)$$

using "Griddy Gibbs" or M-H.
14. Return to 2. until M number of draws is complete where M is a suitable post burn-in value.

The result of this sampler is the ability to evaluate the LATE estimate within the context of cross-sectional interactions. We want to highlight an important assumption that the reader shouldn't forget. The estimate produced using Equations (20) and (22) is a local average, not an average, treatment estimate. By conditioning the full sample estimates on the LATE we are using the LATE as a proxy for the ATE, specifically so that we can evaluate the partial derivatives of the Durbin model.

Consider that, from equation (9), the partial derivative with respect to the treatment effect can be written as,

$$\Delta y/\Delta q = S_\gamma(\boldsymbol{G}) = (I_n-\tilde{\rho}\boldsymbol{G})^{-1}(I_n\gamma). \quad (28)$$

This partial derivative is a dense matrix which outlines the long-run equilibrium effects of changes on treatment status. Note that this partial derivative is different from that outlined in Equation 9. Since Q is a deterministic function of Z it stands to reason that any network based lag effect of treatment would be contained in $\boldsymbol{G}Z$ rather than $\boldsymbol{G}Q$. For example, if Z were test scores as in Thistlethwaite and Campbell (1960) and Q was the resulting scholarship award contingent upon Z, then the test scores of neighbors would be important, not the scholarship receipt. One could include this effect through the structure of $\boldsymbol{Z}$ used in the second stage of the sampler; since $\boldsymbol{G}Z$ does not rely on ρ it would just be

an additional parameter in the local-linear structure. Even so it would not appear in the partial derivative above because what is carried through to the third stage is the difference in intercept from the local linear equations, not the additional parameters.

From this matrix we construct scalar summaries to characterize the effects (LeSage and Pace, 2009). Total effects, $n^{-1}\iota_n'(S_\gamma(\boldsymbol{G})\iota_n)$, can be decomposed into direct and indirect effects as outlined in Equation 10. Direct effects are the average of the main diagonal, $n^{-1}tr(S_\gamma(\boldsymbol{G}))$, where $tr(\cdot)$ is the trace operator; which represents the direct effect of treatment on the treated. It should be noted that, since Equation (28) can be written as a Taylor expansion, the main diagonal includes feedback from higher order terms. The indirect effects are the average of off diagonal terms, $n^{-1}\iota_n'(S_\gamma(\boldsymbol{G})\iota_n) - n^{-1}tr(S_\gamma(\boldsymbol{G}))$, and represent the spillovers from treatment on treated and untreated agents alike. Current RD frameworks assume that these indirect effects are zero.

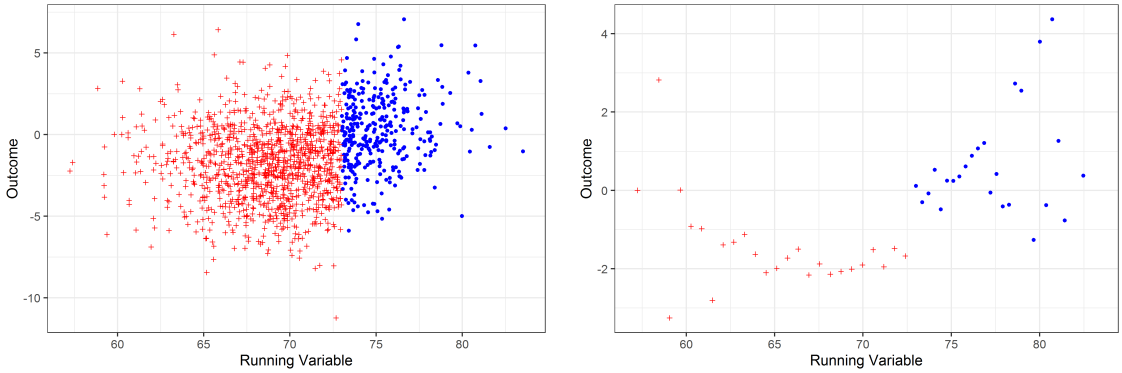
In this section we outlined how to obtain causal estimates of a treatment effect in the presence of cross-sectional interactions commonly found in network models. In the next section we will present Monte Carlo study results, which show that the proposed method acts as a generalized RD framework not only allowing for the evaluation of richer partial derivative but also producing estimates with reduced bias and mean square error at $\rho = 0$.

3 Simulations

To study the efficacy of the proposed model specification we performed a Monte Carlo study where the primary data generating process can be written as:

$$
\begin{aligned}
Y &= (I - \rho\boldsymbol{G})^{-1}(X\beta + \boldsymbol{G}X\phi + \gamma Q + \epsilon), \\
q_i &= \begin{cases} 1 & z_i \geq Z_0 \\ 0 & z_i < Z_0 \end{cases}, \\
\epsilon &\sim N(0, \sigma^2).
\end{aligned}
\tag{29}
$$

In each simulation the sample size is $N = 1,500$ and the number of covariates, K, is set to four. The network is constructed by distributing the sample across a plane and using Voronoi Tesselation to construct an undirected adjacency matrix. We use a queen contiguity design – that is connections are determined through shared borders and vertices – and while the results presented herein use only this structure, additional simulations have been done using other common structures (*e.g.* rook or bishop contiguity, K-nearest neighbor, minimum distance, etc.) as well as directed graph structures more commonly found in network analysis; code for this is available upon request. The information matrix, X, is drawn from a multivariate normal distribution such that $X \sim \text{MVN}(\mathbf{0}, I\sigma^2)$ with $\beta_1 = \ldots = \beta_k = 1$ and $\phi_1 = \ldots = \phi_k = 0$. The running variable is distributed as $Z \sim \text{N}(70, 16)$ with $Z_0 = 73$ resulting in 21.4% of the observations being treated. We set $\gamma = 2$ for all simulations. Figure 1 provides a visual reference for one data set generated under the outlined parameters. Note that this data set is one which exhibits no cross-sectional dependence (*i.e.* both $\rho = 0$ and $\phi = 0$).



(a) Here we plot one example data set using the DGP outlined above. Since $\rho = 0$ this is a data set which exhibits no cross-sectional dependence.

(b) Here we bin the raw data so that the discontinuity is clearly visible. The bins are generating using the *rdplot* function from the *rdrobust* package in R.

Figure 1: Example of Simulated Data: $\rho = 0.00$

The dependence parameter is partitioned into 91 segments over the interval $(-0.90, 0.90)$ with $1,008$ simulations completed for each value of ρ. Since $\boldsymbol{G}$ is fixed the domain of ρ remains fixed over the interval $(-2.005, 1)$; our simulations are over a smaller interval because in practice negative values are rare, let alone large negative values. For each simulation we fix X, $\boldsymbol{G}$, R, and Q, drawing a new ϵ with σ^2 such that the signal-to-noise ratio is maintained between 0.70 and 0.85 as calculated in LeSage and Pace (2017).

With these MC simulations we have two primary goals. First, we want to demonstrate that, over the domain of ρ, the proposed method is well-behaved with respect to recovering the true treatment parameter.[3] For comparison, we will repeat the simulations using methods outlined by IK, CCT (both Conventional and Robust), and Chib and Greenberg (2014); the latter of which is a Bayesian nonparametric take on sharp and fuzzy RD structures. Second, and more importantly, we want to illustrate the differences in partial derivatives between the proposed and alternative methods. In all cases we employ a set of uninformative, but proper, priors. Table 1 outlines the treatment parameter at $\rho = (-0.50,\ 0.00,\ 0.50)$ using each of the five methods.

[3]We present – for brevity – only the results for the parameter of interest, however the additional estimates (*e.g.* β or ϕ are available upon request).

Table 1: Comparison of Parameter Estimates: A Selected Look

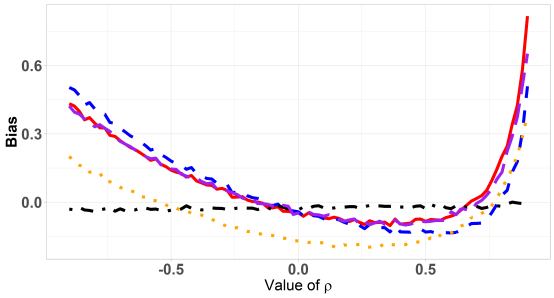
		IK	CCTc	CCTr	CG	Prop
$\rho = 0.00$	Bias	-0.17	-0.05	-0.04	-0.05	-0.03
	SE/SD	0.19	0.20	0.23	0.24	0.25
	MSE	0.06	0.04	0.06	0.05	0.03
	BW	3.41	2.33	2.33	-	3.15
$\rho = 0.50$	Bias	-0.17	-0.08	-0.13	-0.09	-0.03
	SE/SD	0.21	0.26	0.31	0.30	0.25
	MSE	0.06	0.08	0.07	0.07	0.03
	BW	3.46	2.10	2.10	-	3.17
$\rho = -0.50$	Bias	0.03	0.14	0.18	0.15	-0.04
	SE/SD	0.21	0.24	0.28	0.28	0.25
	MSE	0.03	0.08	0.11	0.08	0.03
	BW	3.52	2.22	2.22	-	3.13

It is important to note several subtle differences in this table between the established methods and the proposed. First, we choose to characterize the posterior distribution of the proposed through its mean. This choice is made because methodology being evaluated is a sharp RD structure and it is expected that the distribution of γ is distributed in a Gaussian manner. This may not always be the case. For instance, in a Fuzzy RD structure the distribution of γ is a ratio of normal distributions, and it would likely be more appropriate to characterize that posterior through its mode. In both cases the value presented is the average of point estimates or posterior mean values where applicable. Additionally, while the values are similar, the standard deviation of the marginal posterior distribution is not equivalent to the standard error of the frequentist point estimate. Under very specific circumstances (e.g. known variance, flat priors) these two values can be equivalent, however in practice they generally are very different. Finally, under the proposed estimation method a bandwidth is calculated with every iteration of the algorithm. The value reported in Table 1 is the average bandwidth calculated over all the simulations.

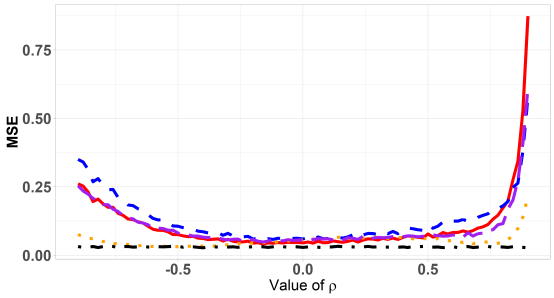
Since the non-spatial model is a special case of an SDM, the first block in Table 1 outlines estimates where the true DGP matches the assumptions of existing RD models. Here we can see that, despite $\rho = 0$, the model uncertainty is such that the standard deviation of the posterior for the treatment is slightly larger than that of the robust CCT estimate. The downward bias comes from the numeric integration of the initial parameters used to residualize, and it is clear from these results that even at $\rho = 0$ both the estimator bias and mean-squared error (MSE) are lowest in the proposed specification as compared to the alternative.

Recall that earlier research indicated that, in the presence of cross-sectional dependence, estimates failing to control for the relationships are often biased and/or inefficient and prone to spuriousness of results (Anselin, 1988; LeSage and Pace, 2009; Pace et al., 2011; Anselin and Arribas-Bel, 2013). This is supported in Table 1 and Figure 2 which show a curvature to the bias for non-spatial specifications over the domain of ρ. Inefficiencies crop up as the value of ρ moves toward the edges of the domain and as a result the standard errors

for non-spatial specifications increase, sometimes dramatically, leading to increases in the MSE.



(a) In this figure we are plotting the bias of each estimator over the domain of ρ. The sold (red) line and dashed (blue) lines represent the CCT conventional and robust estimators respectively. The dot line (orange) represents the IK estimate and the long-dash (purple) line represents estimates from CG. Finally, the dot-dash (black) line is the proposed estimator.



(b) In this figure we are plotting the MSE of each estimator over the domain of ρ. The sold (red) line and dashed (blue) lines represent the CCT conventional and robust estimators respectively. The dot line (orange) represents the IK estimate and the long-dash (purple) line represents estimates from CG. Finally, the dot-dash (black) line is the proposed estimator.

Figure 2: Bias/MSE over domain of ρ

For the frequentist estimators we take the average of standard errors to construct the confidence interval for those simulations, outlined in Table 2. At $\rho = 0$, and indeed for each of the four values, this confidence interval contains the true value on average. Though we caution the reader that this says little about the coverage properties for these estimators in the face of spatial dependence; a topic we will address shortly. Moreover, despite containing the true value on average, these intervals are getting wider as ρ moves away from zero. The intervals – while similar in value for $\rho = 0$ – differ through their interpretation. Intervals for both CG and the proposed method are the 95% highest posterior density (HPD), which is the shortest of possible Bayesian intervals (Casella and Berger, 2002). Unlike the confidence interval, an HPD is a probability statement specifically indicating that the true value lies within the interval with probability 0.95. Going forward we will refer to these intervals in a comparative, and interchangeable fashion though we encourage the reader to see Casella and Berger (2002) and Gelman et al. (2013) for a full discussion of the differences.

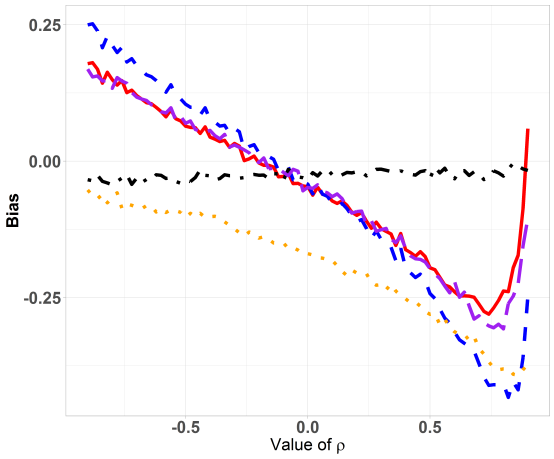
It is with this in mind that we would like to point out the similarities, at $\rho = 0$, between the constructed CIs and HPDs found in Table 2. Note that, according to Calonico et al. (2014), the main difference between the CI of conventional RD estimation and their robust estimator is the bias correction in the bandwidth and new standard error calculations based on that correction. Their method ultimately produces a smaller bandwidth with larger standard errors that have better coverage properties. Indeed, our simulations confirm these results and both CCT estimators do in fact have simulated coverage closer to 95% when $\rho = 0$. However, over the domain of ρ, coverage for both the CCT and IK estimates varies wildly from a low of 80% to a high of 100%.

Table 2: Interval Estimates: A Selected Look

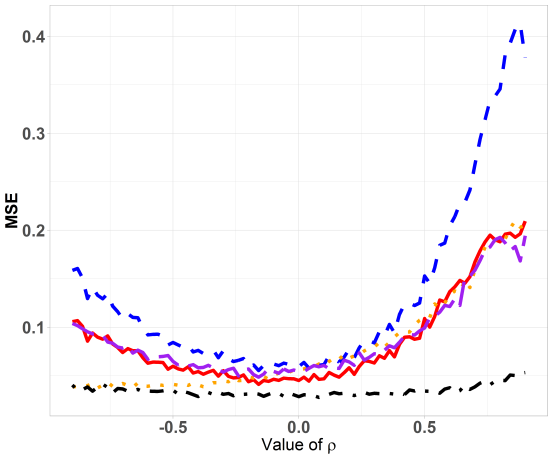
	$\rho = 0.00$			$\rho = 0.50$			$\rho = -0.50$		
Method	Mean	L95	U95	Mean	L95	U95	Mean	L95	U95
IK	1.83	1.40	2.26	1.83	1.40	2.26	1.97	1.54	2.40
CCTc	1.95	1.57	2.34	1.92	1.40	2.44	2.14	1.67	2.61
CCTr	1.96	1.50	2.41	1.87	1.26	2.48	2.18	1.63	2.73
CG	1.95	1.49	2.42	1.91	1.33	2.50	2.15	1.59	2.70
Proposed	1.97	1.47	2.46	1.97	1.48	2.46	1.97	1.47	2.46

The true value of the treatment effect is $\gamma = 2$.

Of course, the parameter itself in a Durbin model is of little importance since it is only one input into the partial derivative. It may then be useful to consider comparing the direct effects. By assumption there are no spillovers in typical RD estimation methods meaning that the direct effects are explicitly assumed to be equivalent to the total effects. Figures 3a and 3b outline the bias and MSE if the parameters from non-spatial estimation methods are considered to be direct effects themselves. A similar story emerges under this view; the bias and MSE are both U-shaped as before, however many of the curves have shifted down with the absolute bias being smaller in magnitude. Interestingly for estimates produced using the IK algorithm at no point under the domain examined is there not a bias where as when comparing the parameters the bias disappeared at $\rho \approx -0.50$. Meanwhile the cross for other alternative estimation methods still occurs at $\rho \approx 0$ showing that they have an unbiased estimate of the treatment even with no spatial dependence.



(a) The sold (red) line and dashed (blue) lines represent the CCT conventional and robust estimators respectively. The dot line (orange) represents the IK estimate and the long-dash (purple) line represents estimates from CG. Finally, the dot-dash (black) line is the proposed estimator.



(b) The sold (red) line and dashed (blue) lines represent the CCT conventional and robust estimators respectively. The dot line (orange) represents the IK estimate and the long-dash (purple) line represents estimates from CG. Finally, the dot-dash (black) line is the proposed estimator.

Figure 3: Bias and MSE of Direct Effects

Table 3 shows that, if $\rho \neq 0$ the direct effect, as estimated by both CCT and IK, are

close to the true value but exhibit noticeable bias with relatively wide confidence intervals (as compared to $\rho = 0$). More striking is the estimate of the total effect. For $\rho = -0.50$ the estimate of the total effects from both CCT and IK estimators overstate the welfare effect of the treatment by $\approx 64\%$. For values of $\rho = 0.50$ the opposite is true, the estimators of CCT and IK understate the total welfare effect by $\approx 100\%$. Remember that for conventional estimators there are no indirect effects by construction. As a result the total effect is the direct effect. This is particularly pertinent to policy decisions based on the effect size since inaccurate assessment of effect size can create inefficiencies in public spending on that treatment.

Table 3: Partial Effects Estimates

	Estimator	Total	L95	U95	Direct	L95	U95	Indirect	L95	U95
$\rho = 0.00$	True Value	2.000	-	-	2.000	-	-	0.000	-	-
	IK	1.831	1.400	2.261	1.831	1.400	2.261	-	-	-
	CCTc	1.953	1.569	2.338	1.953	1.569	2.338	-	-	-
	CCTr	1.958	0.232	2.333	1.958	0.232	2.333	-	-	-
	CG	1.952	1.488	2.416	1.952	1.488	2.416	-	-	-
	Proposed	1.965	1.466	2.470	1.967	1.474	2.458	-0.002	-0.103	0.105
$\rho = 0.50$	True Value	4.000	-	-	2.111	-	-	1.889	-	-
	IK	1.831	1.401	2.261	1.831	1.401	2.261	-	-	-
	CCTc	1.916	1.397	2.435	1.916	1.397	2.435	-	-	-
	CCTr	1.869	1.256	2.481	1.869	1.256	2.481	-	-	-
	CG	1.914	1.327	2.501	1.914	1.327	2.501	-	-	-
	Proposed	3.926	2.920	4.963	2.081	1.563	2.597	1.845	1.333	2.409
$\rho = -0.50$	True Value	1.333	-	-	2.076	-	-	-0.743	-	-
	IK	1.973	1.543	2.403	1.973	1.543	2.403	-	-	-
	CCTc	2.140	1.670	2.611	2.140	1.670	2.611	-	-	-
	CCTr	2.181	1.626	2.735	2.181	1.626	2.735	-	-	-
	CG	2.146	1.595	2.696	2.146	1.595	2.696	-	-	-
	Proposed	1.309	0.976	1.644	2.041	1.526	2.556	-0.732	-0.932	-0.540

4 Empirical Example: U.S. House Elections

To ground the proposed method in real data we turn to what has become a common example used in RD econometrics literature; election results from U.S. legislative bodies. These studies often look at party incumbency effects by the U.S. House of Representatives (Lee, 2008; Broockman, 2009; Butler, 2009; Caughey and Sekhon, 2011; Cho et al., 2012; Calonico et al., 2014) or U.S. Senate (Albouy, 2013; Chib and Greenberg, 2014). Here we look at results from elections in the U.S. House of Representatives from 1946 until 1996 and, similar to both Lee (2008) and Caughey and Sekhon (2011), we look at the effect of a close (Democrat) win, as measured by margin of victory (DifDPct), in time period t on (Democrat) vote share in time period $t + 1$ (DPctNxt).[4]

While additional information is not strictly necessary (Lee and Lemieux, 2010) we use

[4]It is important to note that while we have chosen to look at the effect of close Democrat wins we only do this to link our analysis to the previous literature. The model looking at close Republican wins would be a near mirror image of this model and provide qualitatively the same results.

some of the data collected by Caughey and Sekhon (2011) as controls.[5] Specifically we use a dummy indicator for the party affiliation of the Secretary of State (SoSDem), the margin of victory in presidential elections averaged over the decade (DifPVDec), the percentage of government workers in a district (GovWkPct), the percentage of population considered to be living in an urban setting (UrbanPct), the percentage of black voters in the district (BlackPct), and the percentage of population which is foreign born (ForgnPct). Table 4 outlines the descriptive statistics of the data.

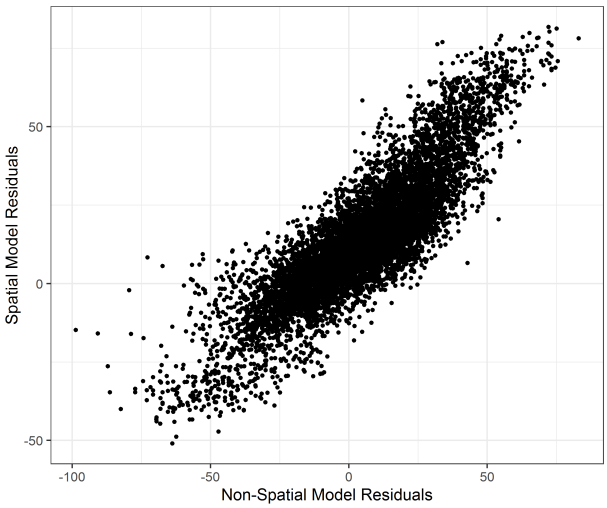
Table 4: Descriptive Statistics: U.S. House Election Results

Variable	Mean	Std. Dev.	Minimum	Maximum
DPctNxt	56.43	24.01	0.00	100.00
DifDPct	14.66	47.09	-100.00	100.00
SoSDem	0.63	0.48	0.00	1.00
DifPVDec	-0.01	0.26	-0.87	0.91
GovWkPct	4.97	2.55	0.04	19.63
UrbanPct	65.72	26.59	0.00	100.00
BlackPct	10.70	14.09	0.02	92.07
ForgnPct	5.31	6.47	0.00	58.52
N	7,949			

We obtain the weight matrix by combining the above data with latitude and longitude from historical congressional district maps (Lewis et al., 2013) using a queen contiguity design. This is done for each congress separately and then the complete weight matrix is of a block diagonal form. We chose this structure for its simplicity rather than its strict validity. It is important to note that while U.S. House districts are adjusted following the Decennial Census there is a great deal of correlation in neighboring districts across the election years. Many times a district will have the same neighbors throughout the sample.[6] A Moran's I test, with a null hypothesis of no correlation, produces a statistic of 51.9902 indicating substantial evidence against the null. This is further buttressed by a naive look at model residuals using simple OLS and an SDM model estimated using maximum likelihood. Figure 4a shows that, while there is a high degree of correlation between the spatial and non-spatial residuals, it is no where near perfect. This, combined with the results of the Moran's I, indicate that a Durbin model is likely a correct specification.

[5]We would like to thank the authors for making this data available through replication files hosted at `https://dataverse.harvard.edu/dataset.xhtml?persistentId=hdl:1902.1/16357&version=1.0`.

[6]These adjustments are what led to Lee (2008) omitting the years ending in '0' and '2'. We have found however that inclusion of these years does not materially change the results and we can provide supplementary tables upon request.



(a) Here the we plot the residuals from both a spatial and non-spatial model using the U.S. House of Representatives data. If there were no spatial dependence then the residuals should be near perfectly correlated since a basic OLS framework is nested within the SDM estimation.

Figure 4: Residuals: Spatial and Non-Spatial

Figures 5a through 5f plots the outcome and running variable in a traditional RD type plot. Along the left column we plot the raw data from the outcome (Figure 5a), non-spatial (Figure 5c), and spatial residuals (Figure 5e). The discontinuity is apparent both in Figure 5a and 5e though far less convincing in 5c. The plots themselves provide no informative power regarding the significance (or lack thereof) of subsequent estimates but the difference in the plots is such that the jump is clearer. Along the right column we have plotted the binned values over the running variable. The story repeats itself in these plots with the outcome 5b and 5f displaying a clear jump at zero. Moreover, in all three cases the area around the cutoff looks decidedly linear despite the non-linearity of the full sample.

As a reminder, while the original study by Lee (2008) was completed using a fourth order polynomial in the margin of victory, our pseudo-replication will be limited to a local linear specification. Our decision is supported by recent literature which casts doubt on higher order polynomial use in the RD setting (Gelman and Imbens, 2018) as well as graphic evidence from Figures 5a through 5f. Table 5 provides the results of our proposed specification and a comparative set of specifications.[7] For all results the dependent variable is the residualized rather than raw outcome so as to keep the results as consistent as possible for comparison. The conventional estimates, those in the top block, are consistent with those found in Lee (2008) where winning a close election in time t provides a roughly 7.5% boost to vote share in time $t + 1$. [8]

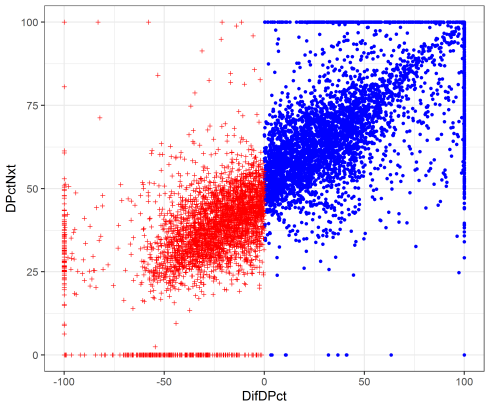
[7]Conventional IK was estimated using functions written by the authors in Matlab while both CCT estimates and the CG estimates were produced using their respective packages in R.

[8]Interestingly Caughey and Sekhon (2011) find close to a 9.25% increase using the same data however that was on the raw rather than the residualized outcome. The conventional estimates we obtain using the raw outcome variable are quantitatively similar to this value.

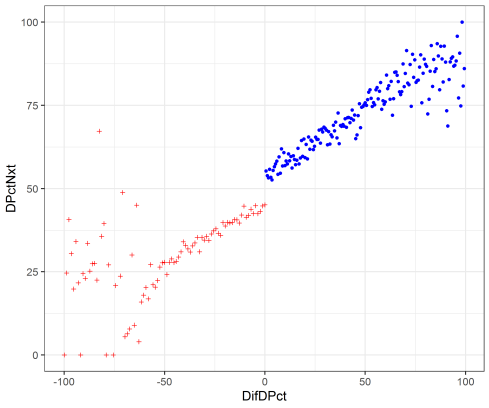
Table 5: Effect of Close House Elections in %: Residualized Outcome (DPctNxt)

Method	Mean	SE/SD	L95	U95	LBW	RBW
Conventional (IK)	7.873	1.090	5.736	10.010	29.920	29.920
Conventional (CCT)	7.402	1.757	3.958	10.847	18.962	18.962
Robust (CCT)	7.177	2.112	3.038	11.317	18.962	18.962
Bayes (CG)	7.857	1.426	5.083	10.634		
Proposed (IK)	8.916	1.416	6.169	11.717	26.525	26.525
ρ	0.405	0.014	0.378	0.431		
Total	14.985	2.379	10.368	19.693		
Direct	9.220	1.464	6.380	12.117		
Indirect	5.765	0.915	3.989	7.576		
Proposed (CCT)	9.109	1.383	6.398	11.824	18.287	18.287
ρ	0.403	0.014	0.375	0.430		
Total	15.413	2.341	10.826	20.006		
Direct	9.427	1.432	6.622	12.370		
Indirect	5.985	0.909	4.204	7.769		
Proposed (AI)	9.088	1.322	6.437	11.671	18.436	32.048
ρ	0.403	0.014	0.377	0.430		
Total	14.826	2.157	10.501	19.039		
Direct	9.368	1.363	6.635	12.030		
Indirect	5.458	0.794	3.866	7.009		

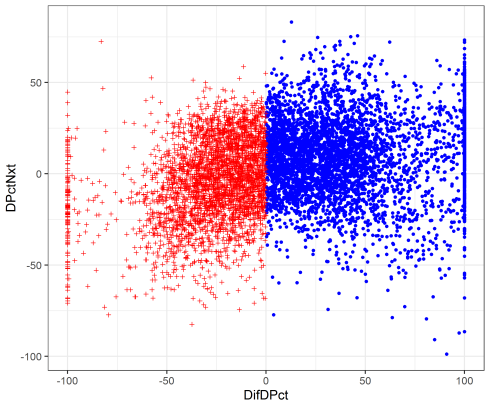
Estimates with the raw outcome variable are available upon request. While the specification here for the proposed method does not include temporal fixed effects the addition of them does not change the estimates in a significant fashion. These results are also available upon request. In all of the specifications there is no clustering of standard errors or standard deviations, and all information included in the residualization process is kept consistent across the specifications. For the proposed estimation procedure an average of $3,472$ (63.26) observations were used to calculate the treatment parameter when IK bandwidths were employed. For the proposed method using CCT and AI bandwidths this number is $2,436$ (10.76) and $3,250$ (264.27) respectively. As a reminder, the full sample size is $7,949$.



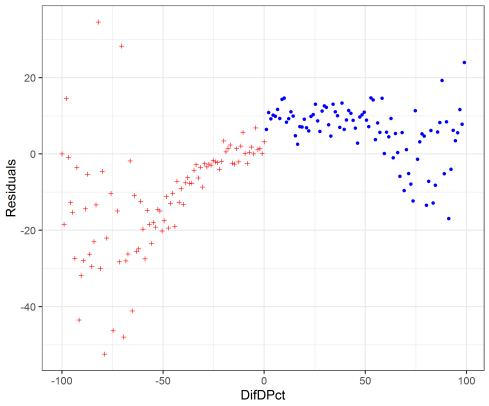
(a) Here the raw data is plotted to show a clear discontinuity at zero without binning the data. Vote share in time period $t+1$ is on the Y-axis with the plus (red) icons referring to Republican vote share and dot (blue) referring to Democrat vote share.



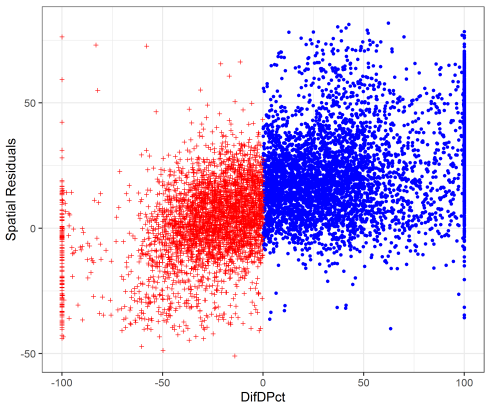
(b) Here the data is binned in a more traditional RD plot to show a clear discontinuity at zero. Vote share in time period $t+1$ is on the Y-axis with the plus (red) icons referring to Republican vote share and dot (blue) referring to Democrat vote share.



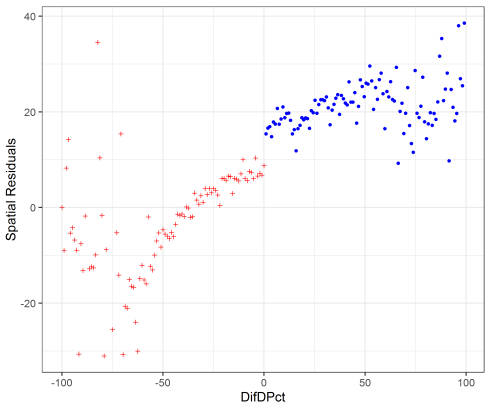
(c) Here the residualized outcome (using OLS) is plotted to show a clear discontinuity at zero without binning the data. Vote share in time period $t+1$ is on the Y-axis with the plus (red) icons referring to Republican vote share and dot (blue) referring to Democrat vote share.



(d) Here the residualized outcome (using OLS) is binned in a more traditional RD plot to show a clear discontinuity at zero. Vote share in time period $t+1$ is on the Y-axis with the plus (red) icons referring to Republican vote share and dot (blue) referring to Democrat vote share.



(e) Here the residualized outcome (at the conditional mean for all posterior parameter estimates using an SDM) is plotted to show a clear discontinuity at zero without binning the data. Vote share in time period $t+1$ is on the Y-axis with the plus (red) icons referring to Republican vote share and dot (blue) referring to Democrat vote share.



(f) Here the residualized outcome (at the conditional mean for all posterior parameter estimates using an SDM) is binned in a more traditional RD plot to show a clear discontinuity at zero. Vote share in time period $t+1$ is on the Y-axis with the plus (red) icons referring to Republican vote share and dot (blue) referring to Democrat vote share.

Figure 5: Regression Discontinuity Plots

While the point estimates in the first block of Table 5 are similar, the standard errors (or standard deviation in the case of CG) vary significantly and produce different confidence intervals. The widest of intervals in this block are produced by robust CCT methods, which are consistent with previously published results. The second through fourth blocks of this table outline the empirical results using the proposed specification with variation in how the bandwidth is calculated. In our simulations we used the IK method of bandwidth calculation primarily for its computational efficiency. In this particular example the choice of bandwidth under the proposed model is irrelevant as all three options produce nearly identical results both in point estimates and standard deviation. Results from the original study (Lee, 2008) show between a 7.7% and 8.1% average effect while Caughey and Sekhon (2011) shows approximately 9% using a similar specification. While the results of proposed method are similar with respect to the parameter estimate it is important to note that the most direct comparison is the parameter estimate of non-spatial specifications to the *direct* effects from the spatial models. With this in mind we see that the estimated direct effect, 9.2%, is larger than any of the non-spatial specifications. This is consistent with the simulations provided earlier given the estimated $\rho \approx 0.4$.

Results using the proposed method indicate that the direct effect of winning a close election in district A, in time period t, is approximately a 9.2% increase in Democrat vote share in time $t+1$ for district A. Furthermore, neighbors of district A experience an average increase of approximately 5.4% in Democrat vote share in time period $t+1$ following a close Democrat win in district A during time period t. For example, consider the map in Figure 6a. The darker shaded (blue) districts indicate a district which, in 1980, was won by a Democrat with a margin sufficiently small that it appears in the (CCT) bandwidth. In 1982 these districts, according to our results, can expect an average increase of $\approx 9\%$ in Democrat vote share. The lightly shaded (gray) districts are the direct neighbors which, in 1982, can expect an average increase of $\approx 5.4\%$ in Democrat vote share stemming from the win in 1980. One possible explanation for such an increase is that, given a close win, the resources in subsequent campaigns can be shifted away from that district to neighboring districts due to the incumbency advantage.

Figure 6: The Effect of Close Elections: California 1980

In this section we have shown that elections in the U.S. House of Representatives from 1946-1996 do exhibit spatial dependence. We then compared the local-linear estimations for the treatment effect put forth by Imbens and Kalyanaraman (2012), Calonico et al. (2014), and Arai and Ichimura (2018) to those of the proposed methods. We find significant and material indirect effects from close (Democrat) wins which support geographic clustering by voters. We vary the bandwidth calculation in our proposed method and show that the standard deviation of the marginal posterior distribution for the treatment has less variation than the frequentist standard errors. This leads to interval estimates which are roughly equivalent across each of the three alternatives and is informative in the sense that additional data admitted through a wider bandwidth are not materially impacting the results.

5 Conclusion

Regression discontinuity design has become a favored technique for identifying a causal relationship from secondary data. As these types of data continue to evolve and include measures of both space and network position it is important to consider how these cross-sectional interactions interact with the RD structure. In this paper we have outlined one method for estimating, in an RD framework, the effect of cross-sectional dependence on treatment. Failure to account for this dependence produces estimates of treatment which are potentially both biased and inefficient. We use a combination of residualization, numeric integration, and Bayesian sampling methods to filter out the cross-sectional dependence and get a clean look at the local average treatment effect through a local-linear specification.

We provide simulated evidence that the omission of cross-sectional interactions from the standard RD framework can lead to inferential problems associated with the aforementioned bias and inefficiencies. Moreover, we show that in the special case of $\rho = 0$, that is a

standard RD model, the proposed estimation method produces estimates with slightly lower bias and mean-squared-error than existing methods. This improvement comes from the incorporation of uncertainty both from the observations admitted to the estimation of treatment and the residualization parameters. We finished by demonstrating the proposed estimation method using the common RD example of close elections in the U.S. House of Representatives. We show that the direct effect of the treatment on the treated is $\approx 9\%$ while the average indirect effect from the treatment is $\approx 6\%$. While the mechanisms which cause these indirect effects are not outlined in the data, one such avenue is a deployment of ground resources around a closely won district rather than within it in anticipation of the incumbency effect.

References

1. Albouy, D. (2013). Partisan representation in congress and the geographic distribution of federal funds. *Review of Economics and Statistics 95*(1), 127–141.

2. Anderson, M. L. (2014). Subways, strikes, and slowdowns: The impacts of public transit on traffic congestion. *The American Economic Review 104*(9), 2763–2796.

3. Angrist, J. D., G. W. Imbens, and D. B. Rubin (1996). Identification of causal effects using instrumental variables. *Journal of the American statistical Association 91*(434), 444–455.

4. Angrist, J. D. and V. Lavy (1999). Using maimonides' rule to estimate the effect of class size on scholastic achievement. *The Quarterly Journal of Economics 114*(2), 533–575.

5. Anselin, L. (1988). *Spatial Econometrics: Methods and Models*. Dordrecht: Kluwer Academic Publishers.

6. Anselin, L. and D. Arribas-Bel (2013). Spatial fixed effects and spatial dependence in a single cross-section. *Papers in Regional Science 92*(1), 3–17.

7. Arai, Y. and H. Ichimura (2018). Simultaneous selection of optimal bandwidths for the sharp regression discontinuity estimator. *Quantitative Economics 9*(1), 441–482.

8. Athey, S. and G. W. Imbens (2017). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives 31*(2), 3–32.

9. Bin, O., B. Poulter, C. F. Dumas, and J. C. Whitehead (2011). Measuring the impact of sea-level rise on coastal real estate: A hedonic property model approach. *Journal of Regional Science 51*(4), 751–767.

10. Brasington, D. M. (2017). School spending and new construction. *Regional Science and Urban Economics 63*(Supplement C), 76 – 84.

11. Broockman, D. E. (2009). Do congressional candidates have reverse coattails? evidence from a regression discontinuity design. *Political Analysis 17*(4), 418–434.

12. Butler, D. M. (2009). A regression discontinuity design analysis of the incumbency advantage and tenure in the us house. *Electoral Studies 28*(1), 123–128.

13. Calonico, S., M. D. Cattaneo, and R. Titiunik (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica 82*(6), 2295–2326.

14. Card, D. and L. Giuliano (2016). Can tracking raise the test scores of high-ability minority students? *The American Economic Review 106*(10), 2783–2816.

15. Casella, G. and R. L. Berger (2002). *Statistical inference*, Volume 2. Duxbury Pacific Grove, CA.

16. Cattaneo, M. D., B. R. Frandsen, and R. Titiunik (2015). Randomization inference in the regression discontinuity design: An application to party advantages in the us senate. *Journal of Causal Inference 3*(1), 1–24.

17. Caughey, D. and J. S. Sekhon (2011). Elections and the regression discontinuity design: Lessons from close us house races, 1942–2008. *Political Analysis 19*(4), 385–408.

18. Chernozhukov, V., M. Goldman, V. Semenova, and M. Taddy (2017). Orthogonal machine learning for demand estimation: High dimensional causal inference in dynamic panels. *arXiv preprint arXiv:1712.09988*.

19. Chib, S. and E. Greenberg (2014). Nonparametric bayes analysis of the sharp and fuzzy regression discontinuity designs. *St. Louis, MO: Washington University in St Louis*.

20. Cho, W. K. T., J. G. Gimpel, D. R. Shaw, et al. (2012). The tea party movement and the geography of collective action. *Quarterly Journal of Political Science 7*(2), 105–133.

21. Cho, W. K. T. and T. J. Rudolph (2008). Emanating political participation: untangling the spatial structure behind participation. *British Journal of Political Science 38*(2), 273–289.

22. Cook, T. D. and V. C. Wong (2008). Empirical tests of the validity of the regression discontinuity design. *Annales d'Economie et de Statistique*, 127–150.

23. Cooper, J. (2018). How political competition can increase corruption: Electoral cycles in police extortion in west africa.

24. Cutts, D. and D. J. Webber (2010). Voting patterns, party spending and relative location in england and wales. *Regional Studies 44*(6), 735–760.

25. De la Cuesta, B. and K. Imai (2016). Misunderstandings about the regression discontinuity design in the study of close elections. *Annual Review of Political Science 19*, 375–396.

26. Eggers, A. C., A. Fowler, J. Hainmueller, A. B. Hall, and J. M. Snyder Jr (2015). On the validity of the regression discontinuity design for estimating electoral effects: New evidence from over 40,000 close races. *American Journal of Political Science 59*(1), 259–274.

27. Gelman, A. and G. Imbens (2018). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, 1–10.

28. Gelman, A., H. S. Stern, J. B. Carlin, D. B. Dunson, A. Vehtari, and D. B. Rubin (2013). *Bayesian data analysis*. Chapman and Hall/CRC.

29. Hahn, J., P. Todd, and W. Van der Klaauw (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica 69*(1), 201–209.

30. Hainmueller, J. and H. L. Kern (2008). Incumbency as a source of spillover effects in mixed electoral systems: Evidence from a regression-discontinuity design. *Electoral studies 27*(2), 213–227.

31. Hidano, N., T. Hoshino, and A. Sugiura (2015). The effect of seismic hazard risk information on property prices: Evidence from a spatial regression discontinuity design. *Regional Science and Urban Economics 53*, 113–122.

32. Imbens, G. and K. Kalyanaraman (2012). Optimal bandwidth choice for the regression discontinuity estimator. *The Review of Economic Studies 79*(3), 933–959.

33. Imbens, G. W. and D. B. Rubin (2015). *Causal inference in statistics, social, and biomedical sciences*. Cambridge University Press.

34. Jacob, B. A. and L. Lefgren (2004). Remedial education and student achievement: A regression-discontinuity analysis. *Review of Economics and Statistics 86*(1), 226–244.

35. Kelejian, H. H. and I. R. Prucha (1998). A generalized spatial two-stage least squares procedure for estimating a spatial autoregressive model with autoregressive disturbances. *The Journal of Real Estate Finance and Economics 17*(1), 99–121.

36. Kelejian, H. H. and I. R. Prucha (1999). A generalized moments estimator for the autoregressive parameter in a spatial model. *International Economic Review 40*(2), 509–533.

37. Kim, J., E. Elliott, and D.-M. Wang (2003). A spatial analysis of county-level outcomes in us presidential elections: 1988–2000. *Electoral Studies 22*(4), 741–761.

38. Kim, J. and P. Goldsmith (2009). A spatial hedonic approach to assess the impact of swine production on residential property values. *Environmental and Resource Economics 42*(4), 509–534.

39. Lacombe, D. J., G. J. Holloway, and T. M. Shaughnessy (2014). Bayesian estimation of the spatial durbin error model with an application to voter turnout in the 2004 presidential election. *International Regional Science Review 37*(3), 298–327.

40. Lazrak, F., P. Nijkamp, P. Rietveld, and J. Rouwendal (2014). The market value of cultural heritage in urban areas: an application of spatial hedonic pricing. *Journal of Geographical Systems 16*(1), 89–114.

41. Lee, D. S. (2008). Randomized experiments from non-random selection in us house elections. *Journal of Econometrics 142*(2), 675–697.

42. Lee, D. S. and T. Lemieux (2010). Regression discontinuity designs in economics. *Journal of economic literature 48*(2), 281–355.

43. LeSage, J. (2014). What regional scientists need to know about spatial econometrics. *The Review of Regional Studies 44*(1), 13–32.

44. LeSage, J. and R. K. Pace (2009). *Introduction to spatial econometrics*.

45. LeSage, J. P. and R. K. Pace (2017). Spatial econometric monte carlo studies: raising the bar.

46. Lewis, J. B., B. Devine, L. Pitcher, and K. C. Martis (2013). Digital boundary definitions of united states congressional districts, 1789-2012. [data file and code book]. [Online; accessed: 2018-12-21].

47. Ludwig, J. and D. L. Miller (2007). Does head start improve children's life chances? evidence from a regression discontinuity design. *The Quarterly Journal of Economics 122*(1), 159–208.

48. Macartney, H. and J. D. Singleton (2017). School boards and student segregation. Technical report, National Bureau of Economic Research.

49. Meng, L. (2013). Evaluating china's poverty alleviation program: a regression discontinuity approach. *Journal of Public Economics 101*, 1–11.

50. Mihaescu, O. and R. vom Hofe (2012). The impact of brownfields on residential property values in Cincinnati, Ohio: A spatial hedonic approach. *Journal of Regional Analysis & Policy 42*(3), 223.

51. Moulton, J. G., B. D. Waller, and S. Wentland (2016). Who benefits from targeted property tax relief? evidence from virginia elections.

52. Pace, R. K., J. P. LeSage, and S. Zhu (2011). Spatial dependence in regressors and its effect on estimator performance. *Available at SSRN 1801241*.

53. Pettersson-Lidbom, P. (2008). Do parties matter for economic outcomes? a regression-discontinuity approach. *Journal of the European Economic Association 6*(5), 1037–1056.

54. Rosenbaum, P. R. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*(1), 41–55.

55. Rubin, D. B. (1978). Bayesian inference for causal effects: The role of randomization. *The Annals of statistics*, 34–58.

56. Sales, A. C. and B. B. Hansen (2014). Limitless regression discontinuity. *arXiv preprint arXiv:1403.5478*.

57. Terrier, C. and M. W. Ridley (2017). Fiscal and education spillovers from charter expansion.

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

59. Van der Klaauw, W. (2002). Estimating the effect of financial aid offers on college enrollment: A regression–discontinuity approach. *International Economic Review 43*(4), 1249–1287.

60. vom Hofe, R., O. Parent, and M. Grabill (2019). What to do with vacant and abandoned residential structures? the effects of teardowns and rehabilitations on nearby properties. *Journal of Regional Science 59*(2), 228–249.

61. Wong, S. K., C. Yiu, and K. Chau (2013). Trading volume-induced spatial autocorrelation in real estate prices. *The Journal of Real Estate Finance and Economics 46*(4), 596–608.