Bayesian lasso estimation of treatment effects in regression discontinuity designs

L. Jason Anastasopoulos

j.andronici@gmail.com

July 17, 2018

Abstract

The regression discontinuity design is one of the most popular and credible methods available for causal inference with observational data. Estimation of local average treatment effects in RDDs are typically based on local linear regressions using the outcome variable, a treatment assignment variable and a continuous running variable. In political science research, treatment effects using RDDs are often estimated with a small number of observations and when the correct functional form of local regressions are unknown. Covariates are typically added to increase the efficiency of treatment effect estimates or to adjust for known imbalances, but model selection can lead to instability in treatment effect estimates, particularly in the small-n context. In this paper, we propose a Bayesian lasso approach to RDD treatment effect estimation as a solution to these problems. Simulations demonstrate that Bayesian lasso estimation of treatment effects in the small-n context minimizes Type I Error and Type II Error relative to local linear regression and produces more accurate treatment effect estimates. Using an illustrative example with close U.S. House of Representative primaries, we show how Bayesian lasso treatment effect estimates can be incorporated into RDDs as a robustness check when fewer than 50 observations are available.

Keywords: regression discontinuity design, causal inference, treatment effect, Bayesian lasso, machine learning

Word count: 4,786.

*This draft was prepared for the Polmeth 2018 conference at Brigham Young University in Provo, Utah. I would like to thank Justin Grimmer for providing me with the inspiration for this paper and the Center for Information Technology Policy at Princeton University for providing me the time and resources to think through these ideas. This is a rough draft, please do not cite without permission.
The regression discontinuity design is one of the most popular and credible methods in political science for causal inference with observational data (see eg. Caughey and Sekhon (2011); Erikson and Rader (2017); Green et al. (2009); Imai (2011); Skovron and Titiunik (2015)). The premise of these designs is simple and intuitive. Around some small window of a threshold which assigns a treatment, individuals or units on either side of this small window are as if randomly assigned to a treatment. While the RDD is simple in principle, in practice, treatment effect estimation can become complicated, often requiring the use of optimal bandwidth selection algorithms which use cross-validated MSE minimization and regularization (Imbens and Kalyanaraman 2012) and a wide variety of parametric and non-parametric model selection choices, even before the introduction of covariates.

Adding to these problems are issues of data sparsity around the threshold of the forcing variable when local linear regressions are used to estimate treatment effects. The nature of political data is such that we must often make due with estimating treatment effects in RDDs with few observations leading to violations of the asymptotic assumptions required by local linear regression (LLRs) estimators which can result in significant treatment effect bias (see eg Kenward and Roger (1997). Table 1 contains a sample of recent studies employing regression discontinuity designs in which treatment effects were estimated at least once with $N < 50$ observations. In addition to this, while the inclusion of covariates for treatment effect estimation in RDDs tends to increase the efficiency of estimates (Calonico et al. 2016), this property, like other properties of frequentist estimators, is only asymptotically correct and potential biases resulting from the inclusion of covariates in small N contexts are less clear.

A solution to these problems requires considering methods which do not rely on asymptotic assumptions and perform well in small-n contexts. For this task Bayesian
methods have been found to perform particularly well for small-n applications relative to generalized linear models in a variety of contexts (see eg. Browne, Draper et al. (2006); Stegmueller (2013)) In particular, we argue here that the Bayesian lasso (BL), a Bayesian variant of the traditional Lasso model, a machine learning method which performs regularization and variable selection, provides a solution to some of the problems tied to estimating treatment effects in RDDs when there are few observations available.

Using a series of simulations, we demonstrate that the Bayesian lasso has a number of desirable properties which make it an ideal method for estimating treatment effects in RDDs when few observations are available. The Bayesian lasso consistently outperforms local linear regression methods in terms of accuracy and estimates credible bounds which minimize Type I and Type II when little data are available for estimation. For these reasons, we recommend incorporating the Bayesian lasso as a routine part of RDD treatment effect estimation when the number of observations around the cutpoint is less than 50. Using an illustrated example with close U.S. House of Representative primaries in which a male candidate ran against a female candidate, we show how this method can be incorporated into the reporting of RDD estimates and can serve as a robustness check in these cases.
The remainder of this paper is as follows. Section 1 provides a brief introduction to treatment effect estimation for sharp RDDs; Section 2 introduces the Bayesian lasso and discusses how it can be incorporated into the estimation of sharp RDDs; Section 3 provides evidence from simulations demonstrating each of the desirable properties of the Bayesian lasso in the RDD context; Section 4 provides an illustrated example of how the Bayesian lasso can be used as a robustness check for treatment effect estimation in close House primaries and Section 5 concludes with a discussion.

1 Treatment effect estimation in regression discontinuity designs

Regression discontinuity designs provide a framework for the causal estimation of treatment effects with observational data. This is accomplished through a forcing variable \( F_i; i = 1, \ldots, n \) which assigns treatment \( T_i \) using a some threshold value \( f \) such that if \( F_i > f \), a unit (individual, geographic unit etc) is assigned to treatment \( T_i = 1 \) and is not assigned to treatment otherwise \( T_i = 0 \). Assuming continuity of the forcing variable, the sharp RDD leverages this mechanism by allowing for the causal estimation of treatment effects around a narrow window of the threshold \( f - \epsilon < f < f + \epsilon \) by making the assumption that, in the limit of this window, units are as as if randomly assigned to a treatment (Hahn, Todd, and Van der Klaauw 2001).

Under the potential outcomes framework (Rubin 2005), define \( Y_i \) as the outcome, \( Y_i(1) \) as the outcome had unit \( i \) received treatment and \( Y_i(0) \) as the outcome assuming unit \( i \) had not received treatment, RDDs allow us to estimate the local average treatment effects (LATE) at the threshold \( F_i = f \). For purposes of illustration, we
assume that $f = 0$:

\[
LATE = \tau = \lim_{F_i \downarrow 0} E[Y(1)_i | F_i = f] - \lim_{F_i \uparrow 0} E[Y(0)_i | F_i = 0]
\]  

(1)

Estimation of $\tau$ is accomplished through a local linear regression in a neighborhood of the cutpoint $F_i \in [c-h, c+h]$ which is often determined through optimal bandwidth selection procedures which minimizes cross-validated MSE \cite{imbens2012}. 

\[
\hat{Y}_i = \beta_0 + \hat{\tau} T_i + f(T_i, F_i) + \eta_i
\]  

(2)

In Equation 2, $\hat{\tau}$ is the estimated local average treatment effect, $T_i$ is a binary treatment indicator function which equals 1 when $F_i > 0$ and $f(T_i, F_i)$ is a function of the forcing variable which often takes the form of a non-parametric kernel or $p^{th}$ order polynomial. Triangular kernels are typically used as the default method of estimation in software programs because they are boundary optimal \cite{cheng1997, mccrary2008}. Despite the fact that $f(T_i, F_i)$ is non-parametric, Equation 2 is essentially a linear regression and, as such, the estimated treatment effect $\hat{\tau}$ is only asymptotically unbiased such that $E[\hat{\tau} | T_i, F_i] = \tau$ only in the limit as $n \to \infty$. In smaller sample sizes, the asymptotic assumptions of linear regression break down.

This is especially problematic in the political science literature as RDDs in the context of close elections often require estimation of $\hat{\tau}$ in Equation 2 with significantly fewer than $n = 100$ observations around the cutpoint and often fewer than $n = 50$ observations (see Table 1). Compounding this problem, is the practice of including covariates in Equation 2 to increase the precision of the estimates of $\hat{\tau}$ or to make the key identifying assumption that $\eta_i \perp T_i$ more plausible. While asymptotic analyses by \cite{calonico2016} find that the inclusion of covariates does indeed increase
precision under many scenarios, in the context of sparse support around a window of the cutpoint, the disproportionate decrease in degrees of freedom and lack of asymptotic stability of the treatment effect estimate renders this practice questionable as well.

Below we argue that a practical solution to both of these problems is to neither to discard RDDs which are estimated with a small number of observations around cut-point, nor to eliminate the practice of including covariates in the model. Rather, we propose supplementing treatment effect estimation in these circumstances with Bayesian lasso estimates first introduced by [Tibshirani (1996)], further developed by [Park and Casella (2008)] and successfully employed in subgroup analyses in the political science context by [Ratkovic and Tingley (2017)]. As we demonstrate below, the Bayesian lasso consistently produces more accurate and stable treatment effects than local linear regression in the small $n$ context ($n < 50$) regardless of the model estimated.

In addition to this, the BL tends to estimate credible bounds on estimates which result in significantly lower Type I and Type II errors when $\tau = 0$ and $|\tau| > 0$, respectively. Below we provide a brief introduction to the BL and adapt it for the RDD setting. This work can be seen as an approach which serves as a middle ground between those calling for a fully Bayesian estimation strategy for RDDs [Branson et al. (2017); Chib and Jacobi (2016)] and those skeptical of Bayesian approaches to treatment effect estimation. Indeed, this approach calls for LLR estimation as the primary method for treatment effect estimation in RDDs with BL estimation as only a supplement to it.
2 Treatment effect estimation with the Bayesian lasso

2.1 The Lasso

The Bayesian lasso is a Bayesian variant of the Lasso, or least absolute shrinkage and selection operator, technique first introduced by [Tibshirani (1996)] a Bayesian formulation of the popular Lasso regularization technique that he pioneered. As its name suggests, the Lasso was introduced as a model selection and dimensionality reduction method which improves the predictive ability of linear models through the inclusion of a regularization term, or penalty on additional parameters in a linear model. Take, for example, an OLS model with \( p \) parameters:

\[
Y_i = \beta_0 + \sum_{j=1}^{p} \beta_j X_{ij} + \epsilon_i
\]  

The least squares method of parameter estimation in OLS involves estimation of \( \beta_0, \ldots, \beta_p \) through minimizing the squared difference between \( Y_i \) and \( \beta_0 + \sum_{i=1}^{p} \beta_i X_i \), w.r.t the parameters:

\[
\arg\min_{\beta} \sum_{i=1}^{N} \left[ Y_i - \beta_0 - \sum_{j=1}^{p} \beta_j X_{ij} \right]^2
\]  

As a result of the bias-variance trade off, in the context of using OLS as a predictive model/machine learning algorithm in high-dimensional settings when there is little theoretical guidance, “kitchen sink” approaches to model selection which involve estimating linear models with many covariates will tend to result in over-fitting while the exclusion of potentially relevant covariates will tend to result in non-optimal test MSE. The Lasso solves these problems through the inclusion of a \( L_1 \) regularization
term which penalizes, or shrinks, parameter values in the objective function:

$$\arg \min_{\beta} \sum_{i=1}^{N} \left[ Y_i - \beta_0 - \sum_{j=1}^{p} \beta_j X_{ij} \right]^2 + \lambda \sum_{j=1}^{p} |\beta_j|$$  \hspace{1cm} (5)

The value of $\lambda$ is determined through cross-validation.

The Lasso was found to be a significant improvement over its predecessor, ridge regression, which employed $L_2$ regularization because unlike ridge regression, it was able to shrink parameter values to 0, leading to greater model interpretability and greater reduction in test MSE. From the perspective of statistical inference, however, one of the major issues with the Lasso is that, because the Lasso produces asymptotically biased parameter estimates by definition, it was practically impossible to derive unbiased standard errors and confidence bounds on point estimates generated by the it. In recent years, however, several researchers have demonstrated that the Lasso can be fruitfully employed as an inferentially valid model selection tool when paired with OLS (see eg. Bloniarz et al. (2016)).

### 2.2 The Bayesian Lasso

In his original paper on the Lasso Tibshirani (1996) notes that parameter estimates in ordinary Lasso had a Bayesian equivalent which can be estimated as the mode of the posterior distribution of $\beta$:

$$\hat{\beta} = \arg \max_{\beta} p(\beta | Y, \sigma^2, \lambda)$$  \hspace{1cm} (6)

using a Laplace prior on each of the $p$ regression coefficients:

$$2\lambda \sum_{j=1}^{p} \beta_j^2$$

Also referred to as a double-exponential prior
\[ p(\beta | \lambda, \sigma^2) = \left[ \frac{\lambda}{2 \sigma} \right]^p \exp \left[ \lambda \sigma^{-1} \sum_{j=1}^{p} |\beta_j| \right] \] \tag{7}

and a likelihood \( p(Y | \beta, \sigma) = N(Y | X\beta, \sigma I_n) \) which is effectively the likelihood from the OLS model.

Park and Casella (2008) were the first to explicitly deal with estimation of the Bayesian version of the lasso in the regression context through placement of priors on the hyperparameters \( \sigma^2 \) and \( \lambda \). Point estimates of the regression coefficients are then estimated using the posterior median of the distribution of \( \beta \) which is derived via a three step Gibbs sampling process. To arrive at this conclusion, Park and Casella (2008) observed that the Laplace prior of Tibshirani (1996) can be represented as a mixture of normal distributions through the introduction of a set of positive parameters \( \tau = \{\tau_1, \cdots, \tau_p\} \). Specifically, they take advantage of the work developed by Andrew and Mallows (1974) who demonstrated that the Laplace can be represented as mixture of normals thus suggesting the following hierarchical structure which can estimated via Gibbs sampling:

\[
\begin{align*}
Y | \mu, \beta, X, \sigma^2 & \sim N_n(\mu 1_n + X \beta, \sigma^2 I_n) \\
\beta | \sigma^2, \tau & \sim N_p(0_p, \sigma^2 D_\tau) \\
\sigma^2 & \sim \pi(\sigma^2) \\
\tau_j & \sim Exp \left[ \frac{\lambda^2}{2} \right] \\
\end{align*}
\]

\[
\frac{a}{2} \exp(-a|z|) = \int_0^\infty \left[ \frac{1}{\sqrt{2\pi s}} \exp(-z^2/(2s)) \right] \left[ \frac{a^2}{2} \exp(-a^2 s/2) \right] ds
\]
where $D_\tau = \text{Diag}(\tau_1, \cdots, \tau_p)$. The full conditional model for $\beta$ is then a multivariate normal with the following form:

$$
\beta|\sigma^2, \tau, Y, X \sim N_p(A^{-1}_\tau X^T \tilde{Y}, \sigma^2 A^{-1}_\tau)
$$

(8)

where $A_\tau = X^T X + D_\tau^{-1}$ and $\tilde{Y}$ is the mean centered transformation of the outcome $Y$. These conditional distributions are then used to construct a Gibbs sampler which draws from the posterior distributions of $\beta, \sigma^2$ and $\tau$.

One of the truly impressive things about this framework is that the median posterior estimates from the Bayesian lasso are identical to the original Lasso in the linear regression framework but with the added benefit of producing valid credible intervals on these estimates, allowing for statistical inference with the benefits of Bayesian methods and regularization (Park and Casella 2008; Kyung et al. 2010).

### 2.3 Treatment effect estimation with the Bayesian Lasso

Given the general structure of the model above, we now describe treatment effect estimation with the BL in the context of a simple RDD where $F$ is the forcing variable, $T$ is a binary treatment indicator and $[c-h, c+h]$ represents the interval on which the treatment effect is estimated as determined by the optimal bandwidth parameter $h$ and the cutpoint of the forcing variable $c$. Note here that the $MSE$ used to determine the optimal bandwidth is not based on predictions generated by the BL model. While the BL estimation can, theoretically, be used to determine the optimal bandwidth, this topic is beyond the scope of this paper but is discussed in some greater detail in the Appendix. Indeed, to be clear, the point that we wish to demonstrate here is that BL estimates of treatment effects for RDDs are useful quantities to have available to the researcher when treatment effects are estimated in the ordinary way but there are
a relatively small number of observations available for estimation conditional on the optimal bandwidth. As above, BL estimation of treatment effects involves estimating \( \tau_{RD} \) in the local linear regression where:

\[
Y_i = \beta_0 + \tau_{RD}T_i + \beta_1 F_i + \beta_2 (F_i \cdot T_i) + \epsilon_i
\]

where \( i \in [c - h, c + h] \).

Applying the BL framework as discussed by Park and Casella (2008) by defining \( \Theta = [\tau_{RD}, \beta_1, \beta_2] \) and we have the following hierarchical model:

\[
Y|\mu, \Theta, X, \sigma^2 \sim N_n(\mu_1 + X \Theta, \sigma^2 I_n) \\
\Theta|\sigma^2, \tau \sim N_p(0, \sigma^2 \tau)
\]

When we explore the distributions of each of the parameters contained within \( \Theta \) in this model, however, we notice something interesting, namely that the treatment effect \( \tau_{RD} \) is estimated using a prior which assumes that it is null:

\[
\tau_{RD}|\sigma^2 \sim N(0, \sigma^2 \tau_1)
\]

In the remaining sections, we use a series of simulations to compare BL treatment effect estimates with those from local linear regressions under different sample sizes in the RDD context and demonstrate that the BL has a number of desirable properties when compared with local linear regression in small \( n \) contexts.
3 Simulations

As mentioned above, we argue for the using the Bayesian lasso as a means of estimating treatment effects in RDDs when small sample sizes are available for three reasons: (1) the BL estimates more accurate treatment effects; (2) the BL provides us with credible bounds on treatment effects estimates in these situations and; (3) BL credible bounds minimize Type I and II errors relative to LLR. To provide evidence for each of these three points, we run a series of experiments under two possible RDD treatment effect scenarios based on simulations similar to those conducted by Imbens and Kalyanaraman (2012) and Calonico, Cattaneo, and Titiunik (2014) in their RDD studies. In the first scenario, the true treatment effect is 1 and in the second scenario the true treatment effect is 0. In both cases, the outcome $Y$ is modeled directly as a function of the forcing variable and the treatment effect indicator:

\[
Y = T + (T \cdot F) + F + \epsilon
\]

\[
\epsilon \sim N(0, 0.1)
\]

\[
F \sim N(0, 1)
\]

\[
T = 1(F > 0)
\]

Figure 1 contains a typical plot from a local linear regression in each of the scenarios. For every simulation with $N$ observations, a new forcing variable, treatment effect indicator and error term are generated 1000 times, allowing for variation in the density around the cutpoint which is defined here as 0. In addition to this, several noisy covariates are added to the estimated model to simulate situations in which the model is incorrectly specified by the researcher. Thus, the final model estimated by
Figure 1 – Discontinuity plots from one simulated data set with $N = 30$
the BL and local linear regression is:

$$Y = \beta_0 + \tau_{RD} T + \beta_1 (T \cdot F) + \beta_2 F + f(X\gamma) + \epsilon$$  \hspace{1cm} (9)

3.1 Accuracy of Bayesian Lasso v local linear linear regression treatment effects

In order to assess the first claim, we estimate the $MSE_{BL}^N$ for treatment effects produced by the Bayesian lasso and $MSE_{LL}^N$ by the local linear regression method for a number of observations around the cutpoint ranging from a minimum of 15 to a maximum of 50. The MSE estimated for each sample size $N$ around the cutpoint is defined as the average error over the $S = 1000$ variations of the data constructed:

$$MSE_{BL}^N = \frac{1}{S} \sum_{i=1}^{S} (\hat{\tau}_{BL,i}^N - \tau_{BL,i}^N)^2$$  \hspace{1cm} (10)

and

$$MSE_{LLR}^N = \frac{1}{S} \sum_{i=1}^{S} (\hat{\tau}_{LLR,i}^N - \tau_{LLR,i}^N)^2$$  \hspace{1cm} (11)

Here the MSE for each model is a measure of how accurately the model estimates the true treatment effect over a range of observations and under a variety of distributions of the forcing variable around the cutpoint with a mis-specified model that includes covariates.

Figure 2 contains plots of $MSE_{BL}^N$ and $MSE_{LLR}^N$ for treatment effect by sample size $N$. Overall, we notice that for each treatment effect scenario, the Bayesian lasso outperforms local linear regression with smaller MSEs for each sample size but dramatically outperforms local linear regression as the sample size decreases below
Figure 2 – Plots of $MSE_{BL}^N$ and $MSE_{LL}^N$ using Bayesian lasso and local linear regression estimation when the true treatment effects are 1 (a) and 0 (b). MSE for each sample size (N) is calculated as an average over 1000 simulated samples.
3.2 Treatment effect bounds

As mentioned above, the implementation by Park and Casella (2008) of the Bayesian Lasso in the regression context included the ability to estimate properly specified credible intervals for each of the parameter values estimated in the model. When considering the use of the BL for treatment effect estimation in the context of political science research, however, it is important to consider not only whether the treatment effect point estimates are accurate, but also whether the bounds on those estimates have desirable properties relative to LLR. For our purposes, we define desirable properties of bounds as those which 1) contain the true treatment effect under a wide variety of circumstances; 2) minimize false positives (Type I Error) when the true treatment effect is 0; 3) minimize false negatives (Type II Error) when the absolute value of the true treatment effect is greater than zero. Evidence from our simulations demonstrates that the Bayesian Lasso outperforms local linear regression in terms of each of these desirable properties.

Figure 3 contains plots of the proportion of simulations in which bounds on the treatment effect estimate contain the true treatment effect when it is null and not null. The BL outperforms LLR for each RDD sample size. When the treatment effect is null, the credible bounds estimated with the Bayesian Lasso capture the true treatment effect consistently, roughly 99% – 100% of the time. This of course, makes sense since the Bayesian Lasso is effectively regularized regression and, as shown above, treatment effect priors are assumed to be zero to begin with.

Given this fact, there might be a fear that the Bayesian lasso has undesirable properties when true treatment effects are, in fact, not null as BL intervals might tend
Figure 3 – % of times the estimated interval contains the true treatment effect (correct coverage) in the case where $\tau_{RD} = 0$ and $\tau_{RD} = 1$ using the Bayesian Lasso and local linear regression estimation. For each sample size (N), correct coverage is calculated as the % of intervals containing the true treatment effect.
toward worse coverage and higher false negative rates in these cases. We find that this fear is unfounded for any number of observations that we study below 50. Figure 3 (a) plots the proportion of intervals containing the true treatment effect of 1000 simulations for each sample size ranging from 15 to 50. This plot clearly demonstrates that the BL consistently estimates intervals containing the true treatment effect more often than local linear regression over the entire range of sample sizes. Consistent with these results, we find that the BL also tends to have significantly lower false negative rates (Type II Error) when the true treatment effect is 1 as shown in Figure 4(a).

Of course, Type I and II error calculations are performed under the commonly understood assumption that the null hypothesis of the treatment effect is zero:

\[ H_0 : \tau_{RD} \]
Figure 4 – False positive and False negatives...

For each sample size (N), correct coverage is calculated as the % of intervals containing the true treatment effect.

(a) False negative rate (Type II Error) when \( \tau_{RD} = 1 \)

(b) False positive rate (Type I Error) when \( \tau_{RD} = 0 \)
4 Application: Estimating the causal effect of candidate gender in House of Representative Elections

Here we show how the Bayesian lasso can be incorporated into RDD treatment effect estimation when there are fewer than 50 observations available for any given bandwidth. To demonstrate how the Bayesian lasso can be incorporated into treatment effect estimation for RDDs, we estimate the causal effect of candidate gender on general election vote share using a RDD developed by Anastasopoulos (2016). Anastasopoulos (2016) uses U.S. House of Representatives primary vote share as a forcing variable in close 2-candidate elections in which a female candidate barely beat a male candidate as a means of estimating the causal effect of gender on general election vote share between 1982-2012 and finds no evidence of a “gender penalty” faced by female candidates.

4.1 Setup

The majority vote share requirement in partisan US House of Representative primary elections allows for RDD estimation. When a candidate is challenged in a primary election, their ability to run in the general election is determined by their vote share in the primaries. In a narrow subset of 2-candidate primaries in which a male candidate runs against a female candidate, a female candidate will represent her party in the general election if she receives over 50% of the vote share.

\footnote{present, there are two implementations of the Bayesian Lasso in \textbf{R} which are more \cite{Ratkovic2017} and less \cite{Park2008} user friendly. As part of this paper, we hope to release a package tailored to Bayesian lasso estimation of RDD treatment effects which estimates both the ordinary Bayesian Lasso of \cite{Park2008} and the LASSOPlus of \cite{Ratkovic2017}.}
In Equation (12) $\Pi_{ipt}^f$ represents primary vote share for the female candidate in two-candidate male/female primaries in district $i$ for party $p = \text{Democratic, Republican}$ in year $t$ and $\Pi_{ipt}^m$ represents primary vote share for the male candidate in the same primary contest. A female candidate will represent her party in her district in the general election in year $t$ when $\Pi_{ipt}^f - \Pi_{ipt}^m > 0$ while a male candidate will represent his district when $\Pi_{ipt}^f - \Pi_{ipt}^m < 0$. Figure 5 provides a graphical illustration of the RDD setup.

Figure 5 – Regression Discontinuity Design Setup (reproduced from Anastasopoulos (2016))

For these analyses, we divide the data by party to estimate the causal effect of gender among Republican and Democratic female winners of primaries, a situation in which there are few observations available at optimal bandwidth choices, analyses which were not conducted originally in Anastasopoulos (2016). The treatment effect
estimated using conventional, robust[^6] and Bayesian lasso methods are all based on the following model in which \( G_{ipt} \) is the general election vote share for party \( p \) in district \( i \) at time \( t \), \( B_{ipt} \) is a binary treatment indicator as defined above, \( F_{ipt} \) is the female primary vote share forcing variable such that \( F_{ipt} = \Pi_{ipt}^f - \Pi_{ipt}^m \):

\[
G_{ipt} = \alpha + \tau_{RD} B_{ipt} + \beta_1 (B_{ipt} \cdot F_{ipt}) + \beta_2 F_{ipt} + \eta_{ipt}
\] (13)

The primary quantity of interest in Equation 13 is \( \hat{\tau}_{RD} \) which represents an estimate of the causal effect of candidate gender on general election vote share if assumptions are met. Figure 6 contains standard RDD plots in which observations are binned and the forcing variable is plotted against the outcome along with locally weighted predictions of the outcome.

Table 2 – Conventional, Robust and Bayesian treatment effect estimates with RDD with female primary vote share as the forcing variable, 1982–2012

| Method               | Estimate | SE  | \( z \)  | \( P > |z| \) | Bounds         |
|----------------------|----------|-----|---------|---------------|----------------|
| Conventional (LLR)   | 0.047    | 0.179 | 0.264 | 0.792         | [-0.304, 0.399]|
| Robust (LLR)         | -        | -    | 0.110  | 0.913         | [-0.381, 0.426]|
| Bayesian (BL)        | 0.000    | -    | -      | -             | [-0.016, 0.056]|

| Method               | Estimate | SE  | \( z \)  | \( P > |z| \) | Bounds         |
|----------------------|----------|-----|---------|---------------|----------------|
| Conventional (LLR)   | -0.006   | 0.078 | -0.071 | 0.944         | [-0.159, 0.148]|
| Robust (LLR)         | -        | -    | 0.132  | 0.895         | [-0.172, 0.197]|
| Bayesian (BL)        | 0.000    | -    | -      | -             | [-0.032, 0.014]|

Table 2 contains treatment effect estimates using each method. Point estimates across using all three methods clearly suggest that the effect of gender on general election vote share are null, a result that is confirmed by Bayesian lasso point estimates which shrunk treatment effects to 0.

[^6]: Calonico, Cattaneo, and Titiumik (2014)
(a) Female Republican v. Male Republican RDD

(b) Female Democrat v. Male Democrat RDD

Figure 6 – Female House primary vote margin ($F_{ipt}$) vs. House General Election Vote Share separated by party, 1982–2012.

5 Discussion
References

Anastasopoulos, Lefteris. 2016. “Estimating the gender penalty in House of Representa-


Appendix

Discussion of optimal bandwidth selection with the Bayesian Lasso.

Examples of treatment effect estimation through 1000 simulations
Figure 7 – Simulated estimates of treatment effects and bounds using the Bayesian lasso (left) and local linear regression (right) when the true treatment effect is 1.