Safer Elections and Women Turnout: Evidence from India

Divya Singh*†

November 26, 2017

PLEASE DO NOT CITE WITHOUT PERMISSION

Abstract

Differences in group turnouts affects distribution of public goods in equilibrium when groups have heterogeneous preferences. In this paper, I measure the effect of improving security at the polling booths intended to prevent electoral malpractices, on gender composition in turnout and political outcomes, during an election in one of the major states in India. In particular, I use Regression Discontinuity Design. Using a novel dataset, I find that women turnout increases by 2-6 percentage points per booth with no effect on male turnout. As a consequence, incumbent's vote share falls and the main challenger party gains. The findings indicate that the vote share of the candidates with some indication of corruption in the office falls. I also find that female candidates fare worse and 'criminal' candidates gain vote share. Overall, the evidence is consistent with the hypothesis of women being 'change agents' and anti-corruption. Moreover, these findings are also evidence of gender and ability bias in womens' voting.

*Divya Singh is an Economics Ph.D. Candidate at Columbia University in the City of New York (divya.singh@columbia.edu)

†I would like to thank Wojciech Kopczuk, Suresh Naidu, S.Y. Quraishi, Francois Gerard, Miikka A.T. Rokkanen, Eric Verhoogen, Rodrigo R. Soares, Johannes Urpelainen, Anja Karolina Tolonen and Gilles Verniers for their precious comments. I am also grateful to Raphael Susewind for providing additional Polling booths data. This paper was awarded Harriss Prize for Best Second Year paper at Columbia University, 2016. This project has greatly benefited from research grants provided by Columbia Development Colloquium and Center for Development Economics and Policy. All errors are my own.

1 Introduction

World History provides many examples of elections coupled with violence. Nineteenth century English Voters found elections so threatening that some 'voluntarily' attempted to disenfranchise themselves. (Rapoport & Weinberg 2000). Parts of present day Africa have 'democracy' riddled with severe instances of violence, voter intimidation, ballot fraud. (Collier and Vicente 2010). Elections in India have historically been associated with malpractices such as booth capturing, clientelism, vote buying, election violence (Austin 1996). At the same time, participation of women in political and electoral process is limited. Women form only ten percent of the representatives in the most recent elections to the Indian Parliament. Credible causal evidence linking election malpractice and inadequate women participation is lacking.

This paper looks at one particular measure of women political participation: women turnout in elections and points to poll booth safety during election as one possible reason for low women turnouts. For this purpose, I look at the effect of enhancing security at the Polling booths during elections in India on turnout by gender. I find that turnout for women relative to men goes up when extra security is deployed at the polling booths. This translates in lower vote share of candidates belonging to incumbent party, female candidates and for re-contesting Ministers with significant asset growth in office.

To provide credible causal estimates, I use Regression Discontinuity Design which is made possible due to the strategy employed by Election Commission of India in identifying 'critical' booths. In order to prevent violence and electoral fraud, in elections post 2008 the Election Commission of India declared certain Polling Booths 'critical' and deployed extra security there in the form of trained Police personnel video surveillance, flying squads etc. One of the criteria [1] used for identifying 'critical' booths was: if more than 75% of votes cast in the Polling booth in the past election were polled for one candidate. This provides an opportunity to implement Regression Discontinuity Design with maximum vote share for any candidate at the booth in the previous election as the running variable. Uttar Pradesh (UP) provides a good context to study since this state has been associated with history of electoral malpractices like Booth capturing, voter intimidation and vote-stuffing.

I combine election results at the level of the Polling booth with the characteristics of the candidates running in the election to create a unique dataset of election results at the level of the Polling Booth. Using maximum vote share in *previous* election as the running variable, I find that increasing security at the Polling booths increases women turnout by around 2-6

[1] The other criteria were: high share of registered voters with no photo; instances of violence and repolls at the booth in previous election; turnout percentage in the booth in previous election greater than 75 %.

percentage points which is equivalent to 3000-9000 additional women votes per constituency. Additionally, there is no significant impact on male turnout. Next, I look at the effect of extra security on vote share of candidates with different characteristics. The recent evidence has pointed that women are more left-wing, anti-incumbent, anti corruption and anti-crime (Edlund and Pande (2002), Kapoor and Ravi (2014), Swamy et al (2001), Dollar et al(2001), Iyer et al (2012)). First, I look at the effect of extra security in the polling booths on the incumbent party. Consistent with the hypothesis that women are 'change agents', I find that vote share of incumbent party goes down by 1.5 percentage points while the main challenger party gains vote share of about 1.8 percentage points.. Second, I look at the effect on vote share of female candidates. Beaman et al. (2009) discuss that there is gender bias against women in positions of power. If all else equal, female votes prefer female candidates (gender affinity effect), then one would expect the female candidates to gain when there is higher turnout of women. However, if women voters suffer from the gender bias in ability of women leaders, then this effect could be negative. I find some evidence consistent with latter. In particular, female candidates' vote share falls when women turnout is higher, suggesting that gender bias dominates gender affinity effect. Third, as in Fisman et al(2015), I look at the effect of increased security on vote share of candidates with some indication of corruption in office. In particular, I focus on candidates who were in office in the last election, are re-contesting, and have disclosed positive growth in their net worth at the end of their term. I find that such candidates witness fall in their vote-shares. Lastly, I look at the vote-share of candidates who have at least one 'criminal' case against them. I find that the vote share of such candidates increases by about 3 percentage points. This can be explained by 'ability' bias effect dominating anti-crime effect among women voters. The 'ability' bias emerges when 'criminality' of candidates serves as signal of candidates' ability when social divisions among voters are highly pronounced (Vaishnav 2012)

I perform four robustness checks to bolster confidence in the observed treatment effect for women turnout. First, I use alternate criterion of determining a polling booth critical, turnout share in past election, as running variable. Estimates are similar but ill-powered. Second, I combine cutoffs together by using distance to the cutoff as running variable. I find that for different specifications, estimate of the treatment effect is around 4 percentage points. Third, I rule out the possibility that the observed increase in women turnout represents gender-biased reduction in ballot-stuffing, by proposing a simple rule for ballot-stuffing. I estimate the effect of change in ballot-stuffing rule around the cutoff on treatment effect and I find that estimated treatment effect is similar to earlier estimates. Finally, I perform permutation test by moving the cutoff to hypothetical points and find that estimated treatment effect at true cutoff is an outlier on the empirical distribution graph of simulated treatment effects.

The above analysis suggests that electoral safety is key to female voter participation. The resulting change in composition of voters [2] can have substantial effects on political economy equilibrium of India. In traditional voting models, politicians are usually assumed to be vote-seeking, motivated by career incentives and re-election for themselves or their parties (e.g., Downs 1957, Arrow 1963, Mayhew 1974, Fenno, Jr. 1978, Kingdon 1989, Cox and McCubbins 2005). If groups who turn out to vote have different preferences than the ones who don't, then the Median Voter Theorem predicts that outcome would favour the group that actually votes. Then increased turnout of women could have consequences on the types of public goods provided in the equilibrium and/or who holds the office. (Chattopadhyay & Duflo (2004), Miller(2008), Brookman (2014)). Additionally, reserving seats for certain groups in the legislative bodies to provide them representation might not be effective if candidates representing these groups are not supported by members of their group during elections. (Jensenius (2015)).

This paper contributes to the literature on institutional changes to voting, election malpractices and women political participation. Miller (2008), Fujiwara (2015) have shown how enfranchisement of a group of voters affects the provision of public goods. Among election malpractices, Baland and Robinson (2008) study vote-buying, Anderson (2015) study clientalism in India. Few studies have looked at the effect of empowering women in the election process through greater representation in the legislative process. There is some consensus that greater representation of women in office affects the type of public goods provided and could have leadership effects (Chattopadhyay & Duflo (2004), Beaman et al. (2009)).

This paper deviates from previous literature in the following ways: it analyzes different malpractice, booth capturing and provides polling booth as alternate unit of analysis. In addition, it serves as first step to analyzing how *de-facto* enfranchisement of women due to greater security at the polling booths may affect the political economy of India whereas previous literature has focused more on quotas for women in local government bodies.

The paper is organized as follows: First, I discuss the related literature. Second, I discuss the institutional background and possible hypotheses. Third, I discuss data and empirical strategy. Fourth, I discuss the results and possible mechanisms. Finally, I conclude and discuss the next steps.

[2] such as more women voters in this case.

2 Related Literature

This paper is broadly related to three main literatures. First, it is related to incentives to voting. This literature has identified three main channels through which turnout in elections could be affected: pecuniary, non-pecuniary and institutional changes. Leon (2012) showed that in Peru, reduction in penalty for not voting decreased total turnout. Among non-pecuniary incentives, Gerber, Green, and Larimer (2008) show that social pressure increased voting turnout in US. DellaVigna et al (2014) show that individuals turnout to vote 'because others will ask'. Institutional changes can affect turnout as well. In US, Miller(2008) shows that extension of suffrage rights to American women in the nineteenth century led to increases in public health spending and decline in child mortality. This relates to *de-jure* enfranchisement of women. However, *de-jure* enfranchisement is not a necessary condition. Fujiwara (2015) shows that introduction of voting machines in Brazil led to *de-facto* enfranchisement of especially less educated individuals.

Second, it is related to literature on electoral malpractices. Few papers theoretical papers that look at how election violence affects electoral outcomes are Ellman and Wantchekon (2000); Chaturvedi (2005); Collier and Vicente (2012). Few field experiments look at elections in low-income countries. For instance, Wantchekon (2009) shows that information about policy and public debates in town hall meetings can overcome clientelism. In India, Banerjee, Kumar, Pande & Su (2010) show that informed voters make better choices and reduces likelihood of vote-buying. Pande (2011) summarises evidence on if informed voters enforce better governance. Among quasi experimental studies, in Chile, Baland and Robinson (2008) provide evidence that landowners buy the support of their workers for higher wages. In India, Anderson (2015) show that evidence of ingrained clientelist vote-trading structures maintained through extra-political means in rural Maharashtra undermine policies that redistribute income toward the majority poor. Fujiwara and Wantchekon (2013), Vicente and Watchekon (2009), Wantchekon (2003) explore clientelist policies as well.

Third, it is related to the literature on the effect of women electoral participation on various outcomes. Chattopadhyay and Duflo (2004) show that women village leaders invest more in infrastructure relevant to the needs of women. Beaman et al (2009) show that prior exposure to women in leadership positions is associated with electoral gains for women. They provide evidence that exposure to females in leadership positions weakens the stereotype about gender roles in public spheres. In addition, Beaman et al. (2012) find direct evidence that local female leaders in India raise the aspirations and educational attainment of girls (and their parents' aspirations for them), most likely through a role model effect. Iyer et al. (2012) find that an increase in female representation in local government induces a large and significant rise in documented crimes against women in India. The authors interpret this as good news, driven primarily by greater reporting rather than greater incidence of such crimes. However, in developed country context, Brookman (2014) finds no evidence of female officeholding on women's turnout in US.

The next section provides details about the institutional background before discussing the empirical strategy.

3 Institutional Background

Election Process in India

In this paper, I focus on the election for the state assembly of UP in 2012. The previous state election was held in 2007. In between, in 2009 the elections were held for the federal government. Each state is divided into constituencies where each constituency elects representative for the State Assembly through plurality voting rule. All registered voters in a constituency go to their designated polling booths to cast their ballot. There are on average 300 polling booths per constituency and all polling booths in a constituency vote for contestants to the constituency. [3] [4]. Contestants either belong to a party or chose to run independently. From a party's point of view, there are two main decisions to be made at the level of the constituency: i) whether to nominate a candidate for the party and if yes then ii) who to nominate. At the state level, the party with the majority of elected representatives in the assembly [5] is declared the winner.

Political Climate of UP

In this paper, I focus attention to the largest state in India, Uttar Pradesh. Uttar Pradesh (UP) is the most populous state in India with more than a million villages. In terms of GDP, ranks 26th out of 32 states in India. Historically, the state has been associated with violent elections and fielding of candidates who often use force and intimidation to win elections.

The political atmosphere in UP is intense. Criminalization of politics is rampant. In 2007 elections (two years before my period of study), 20% of the candidates had self disclosed criminal charges against them and 10% had serious criminal charges These figures were

[3] For instance, Uttar Pradesh, our analysis state has 71 districts and 403 assembly constituencies.

[4] A state is divided into many districts and each district is composed of many assembly constituencies. There are on average 300 Polling Booths in each Assembly Constituency. A village/town could have one or two Polling booths depending on its population.

[5] and not majority of votes. This is similar to US Presidential elections. This system has often led to situations where a party wins majority without sufficient mandate of the electorate (in terms of vote share in the state). For instance, Samajwadi Party, the winner in UP 2012 Elections won 56% of the districts but had only 29% of the total votes in the state. This results in intense political competition often warranting the use of force and fielding of candidates who can make the party 'win' the elections..

almost the same in 2012. Moreover, 46% of the winners in 2012 had criminal charges against them compared to 35% in 2007. [6] Local political parties are dominant. National parties like Bhartiya Janta Party and Indian National Congress have seen a drop in their influence in the state over the years while regional parties like Bahujan Samaj Party (BSP) and Samajwadi Party (SP) have witnessed increased dominance. Caste plays a major role in the state. BSP has traditionally relied on votes from Scheduled Caste and Schedule Tribes (extreme lower caste) whereas SP has enjoyed the support of Other Backward Classes. Challenger SP defeated incumbent BSP in 2012 by winning 224 out of 403 seats in UP. Since there are many parties active in the state, any analysis based on party is complex. I consequently focus on two main parties, the incumbent BSP and challenger SP which won the 2012 elections.

Booth Capturing and Violent Intimidation

The most crucial part of the election process is a polling booth which is the designated location where voters exercise their right to vote. Historically, the booths are sensitive regions and are easy target by political parties for intimidating voters, ballot fraud, booth capturing etc. Witsoe (2013) documents the interaction between local leaders and caste based territorial dominance which makes it easier for parties to identify 'friendly' and 'enemy' booths in a village. In booth capturing, politicians use goons and powerful men to restrict or intimidate voters in 'enemy' booths. It can take many forms such as i) manage booths where local leaders stand near the voting machine and direct voters to vote for specific candidates, ii) armed gunmen take control of the polling booth, iii) or steal the voting machine, iv) or artificially stamp the ballots, v) and in worst case, use bombs and gunfire to cause panic and abandonment. This is even more prominent in the rural areas which are often cut off from the media coverage and the police.

Treatment of Violent Intimidation

One outcome of Booth Capturing is the concentration of violence around the Polling Booths. To deal with this problem, the Election commission of India (ECI) decided to ehance security at polling booths in all elections post 2008. Since the required amount of security forces was far greater than the supply (as there are millions of Polling Booths throughout India), the ECI decided to identify vulnerable booths and deploy extra forces in such 'critical' booths only. Once a booth was declared critical, extra security was deployed in those booths in the form of Police forces, flying squads, micro-observers and CCTV monitoring on the day of election. Few of the parameters which decide whether a particular booth was 'critical' are:

1. If more than 75% of votes cast in the past election were for one candidate and,

[6] Figure provided by Association of Democratic Reform.

2. If the turnout as ratio of registered voters at the polling booth in the previous election was more than 75%.

The other criteria included share of registered voters with no photo registration; share of missing voters under two categories: missing voters with family links and missing voters with no family links; booths which went to repoll in the previous election. Quraishi (2014) points out that an extraordinarily high turnout flags the possibility of booth capturing, hence an excel sheet using all the data mentioned above was often used for sensitivity analysis of such polling booths, for instance in elections in Bihar 2010.

Womens' Political Participation in India

Traditionally, women in India have lagged behind their counterparts in other parts of the world, in terms of their political participation. This includes lack of representation in the legislature and public office, lower women turnout; candidacy and lobbying. The first Lower House at the Center had only 4.5% women ministers. These figures haven't changed much and in 2014, it stands at 12.15%. Closer to my analysis, Figure 9.3 shows that turnout percentage of women in 2009 elections in UP is far lower than that of men. [7]

The literature has proposed many reasons for low women political participation. Ravi and Kapoor (2014) summarize these as: i) socio-historic forces inherited from nationalist movements, current social policies and the gendered nature of citizenship in hampering womens' political participation in government structures,elections and community organizations (Vissandjee et al 2006); ii) lack of reservation of seats for women in the Parliament and state legislatures; [8] (Rai 2011) iii) lack of national consensus and willingness among political parties to give more tickets to women in elections (Basu 1992); iv)perpetuation of a patriarchal political structure together with class, caste and gender subordination acting as strong deterrents to women contesting elections; and v) lack of awareness and knowledge of electoral politics combined with a lack of support from the family and political parties in resources severely affects women's chances to contest and win elections.

Analyzing the effect of electoral safety in the present context

This paper focuses on women turnout ignoring other possible avenues of low participation and tries to establish link between safety and low women turnout. In particular, the two main questions that this project tries to answer is: i) how does increased security at the polling booths affects the composition (by gender) of the voters who turn out to vote and ii) how does this change in composition (if it indeed exists) affects the policy outcomes.

[7] This also reflects gender imbalances in voter registration due to the skewed sex ratio in the population. In India, there are 934 females per 1000 men.

[8] However, reservation for women exists at the grass root level which is the Village Council head. Chattopadhyay and Duflo (2004) have used the randomization in reservation of seats for women to show that this leads to provision of public goods more suited to womens' needs.

The first question can be assessed by looking at the effect of the increased security at the polling booths on the turnout for both genders. It is natural to assume that increased security at the polling booths provides some check on violent practices on the election day at the polling booths. If both men and women are equally affected by violence, then one should expect no change in composition of turnout. This can be directly tested by looking at the effect of treatment on turnout. However, if the effects differ by gender and women turn up in larger numbers then this provides evidence for change in composition of voters.

The second question is more indirect and requires some thought. If women indeed turn out in bigger numbers when booths are safer, then how can one test if this affects policy? Without directly jumping to the actual policy outcomes, one can analyze how treatment affects the vote share for different candidates with different characteristics. As long as men and women have heterogeneous preferences, higher women turnout will affect the vote share of candidates with characteristics that women generally prefer.[9]

There is some evidence that women tend to care more about crime and justice and corruption.[10] Ravi and Kapoor (2014) discuss that women act as 'change agents' and states with higher women turnout are associated with lower vote shares for incumbent party. In that case, higher women turnout should lead to lower vote share for incumbent party and more for challenger party. Gender bias also plays a big role in electoral politics especially in India. Beaman et al. (2009) discuss inherent bias in womens' ability to lead which is malleable to seeing women in office. On one hand, if women voters prefer women candidates (*gender affinity effect*), one would expect vote share of female candidates to go up when women share of turnout goes up, all else equal. On the other hand, if there is gender bias even among female voters, the vote share of female candidates will go down (*gender bias effect*). Empirically, the effect is unclear. [11]

Similarly, the effect on vote shares for 'criminal' candidates is unclear. On one hand, since women on average are more anti-crime (Iyer, 2012) than men, it should lead to lower vote shares for 'criminal' candidates *(anti-crime effect).* On the other hand, 'criminality' may serve as signal of candidates' ability to lead when social divisions are highly pronounced *(ability bias effect).* (Vaishnav 2012), which could lead to higher vote share for 'criminal'

[9]There is some evidence that women (on average) have different preferences than men and this is reflected in policy outcomes when women hold office (Chattopadhyay and Duflo (2004), Duflo (2003), Miller (2008)). Women also seem to have different risk taking preferences. (Byrnes, Miller and Schafer 1999; Niederle and Vesterlund 2007)

[10]For instance, Iyer et al.(2012) show that reported crime against women goes up when there is an increase in female representation in local government.

[11]Additionally, *Womanifesto* [12] proposed by women's movements, journalists, economists, academics and lawyers for the election in 2014 states six main goals for policy makers to commit resources for womens'rights: educate for equality, make laws count towards protection of women who are victims of violence, put women in power, more police for the people, swift certain justice and economic flourishing.

candidates as they might be judged more capable.

Finally, the effect on vote shares of ‘corrupt’ candidates is expected to be negative if women are the ‘fairer’ sex (Dollar, Fisman and Gatti (2001), Swamy et al. (2001)).

Keeping the above in mind, the reduced form analysis in this paper looks at the effect of increasing security at the polling booths on the following outcomes:

1. Vote share of incumbent and the challenger.
2. Vote share of female candidates
3. Vote share of candidates with at least one criminal case against them.
4. Vote share of candidates holding office in previous election and are re-contesting with reported positive net growth in office.

If women are anti-incumbent then 1) is expected to go down. If gender bias effect dominates gender affinity effect, then 2) should go down. If ‘ability’ bias effect among women dominates anti-crime effect then 3) could increase. If women care more about corruption, one should expect 4) to go down;

The next section discusses data and the empirical strategy to test the above.

4 Data and Empirical Strategy

4.1 Data

Data on Turnout by gender

Since the Election Commission of India began to further tighten security in elections post 2008, I obtain data of booth-level election outcomes for elections in 2007, 2009 and 2012 in UP. UP has 403 Assembly Constituencies and around 120,000 Polling booths. For each booth, ECI provides the total number of registered voters in that booth; the turnout for men and women voters; the total votes polled in favour of each candidate at each booth, the party affiliation of each candidate for elections of 2009 and 2012.

Since the intervention happened in 2008, the ideal set up would be to look at 2009 election outcomes by using 2007 as the baseline. However, there was delimitation in 2008 due to which the constituencies and consequently the booth ids changed. [13] This makes it impossible to merge booths in 2007 with 2009, therefore ‘critical’ booths in 2009 can not

[13] Delimitation is the process by which boundaries of the Assembly constituency are re-drawn to make the population sizes equal in each constituency. The last delimitation happened in 2002.

be identified. Hence, I focus on 2012 elections and identify the 'critical' booths for 2012 elections based on 2009 elections. This requires 2009 data to be merged with 2012 which I do by using *booth id*. However, some of the files in both years 2009 and 2012 are corrupted which doesn't allow perfect merge. In the final dataset, I have data for around 109,619 polling booths for both the years which is 91% of the total number of polling booths in UP. Furthermore, all the booths in 2009 where more than 75 per cent of votes were polled in favour of one candidate are tagged as 'treated' in 2012. Figure 9.4 plots the distribution of the running variable. The candidate with the highest vote share at most of the booths gets around 50% of the votes on average. [14] In the final merged dataset, around 10% of the polling booths are treated in 2012. I use 2007 data to test for balance in baseline.

Additionally, since the effect of the treatment is expected to be greater in regions which are more remotely located and hence more prone to electoral malpractices, I restrict my analysis to booths which are located in rural areas. For this purpose, I use the dataset generated by Schneider et al (2003) who identify rural locations all over the world by using 'boosting' and combining MODIS data with nighttime lights data and gridded population dataset. This improves the accuracy of this data over traditional nighttime light datasets. [15] This restricts the sample to 85% of the polling booths merged across years. This is not surprising since UP is a rural state. [16]

Data on Characteristics of Candidates

To show that the change in composition of turnout translates into any policy change, it is crucial to look at the voting patterns for different candidate at each polling booth. For this, I combine this election data with candidates characteristics provided by Association of Democratic Reform (ADR) who have digitized affidavits filed by candidates at the time of their nomination. Beginning 2002, the Supreme court made it mandatory for all candidates contesting elections to disclose criminal, financial and educational background prior to the polls by filing an affidavit with the Election Commission. [17]. [18] The election data provided by ECI contains the names of the party and the votes polled in favor of each party. Merging the two datasets is not fairly simple process. The Polling booth election data provides the candidate name and the party and vote share while ADR also provides candidate name and

[14] Note however, that he might not be the ultimate winner of the constituency since who obtains maximum votes in *constituency* and not *booth*.

[15] I am highly grateful to Raphael Susewind for providing me the latitude and longitude of polling booths in 2009 merged with the MODIS dataset

[16] The results are unchanged when I at the whole sample though they are smaller.

[17] See `http://myneta.info`

[18] Fisman et al (2014) have used this dataset to show that the annual asset growth of winners is 3–5 percent higher than that of runners-up. Additionally, a follow up paper Fisman et al. (2015) provides evidence that asset disclosure led to exit of low ability candidates post disclosure and also possibily reduced rent extraction by office-holders.

party names. However, most of the candidates have similar names which is compounded by the fact that the same candidate is spelled differently in the ADR data and in the Poll Booth data. Hence simply merging through candidate names is very risky. However, since each party fields one candidate at each constituency, it is possible to merge through party names. I merge ADR data with ECI by fuzzy merge using party names. I am able to merge around 85 % of the observations.[19] This merged dataset allows me to look at the effect of the treatment on the vote share of different candidates.

Summary Statistics

Table 9.1 discusses some of the summary statistics of the combined dataset. A typical booth in 2007 has about 1025 registered voters. This figure goes down to 949 in 2012 which could be attributed to the delimitation in 2008. In both the years, average women turnout has been lower than that of men. Part of this could be gender disparity in registration of voters. The average turnout (as share of registered voters) at a booth is around 46% in 2007 but has gone up in 2012 to 61%. Both average women and male turnout has gone up in 2012 as well. However, on average maximum vote share for a candidate at a booth ranges from 25% of the total votes polled to about 99%.

In terms of candidate characteristics, a polling booth has on average 18 contestants out of which there are about 1.37 women candidates. A candidate has on average 12 years of education. Interestingly, on average, less than one candidate is a recontesting minister i.e held office in previous election. This is consistent with the literature that has pointed to 'incumbency' disadvantage in India which makes it worthless to recontest elections. The variation in net assets held by contesting candidates is quite high with some candidates reporting significant debts. Note that re-contesting candidates are on average richer than other candidates by about 13 million rupees (216,000 dollars). Additionally, the re-contesting ministers have reported on average asset growth of about thirty hundred thousand dollars (500,000 dollars). These are huge figures and are consistent with significant private returns to holding public office as discussed in Fisman et al (2014).

The next section discusses the empirical strategy and the underlying assumptions.

4.2 Empirical Strategy

The program design provides the opportunity to implement Regression Discontinuity Design in two dimensions as there are two running variables. However, exploiting the two dimensions together is not very straightforward. To circumvent this issue, in this section, I use maximum

[19] Note that there are independents as well who are not associated with any party. There is possibility of incorrect merge in the constituencies where there are more than one independent. Therefore, I also perform a robustness check where I drop the independents from the sample. The results are robust.

vote share for a candidate as running variable. This converts the problem into standard RD exercise in one dimension and lends visual appeal. The disadvantage is that using only one running variable disregards information contained in the other running variable and may not be efficient. As robustness check in Section 6, alternatively, I collapse the two dimensions into one by using distance from the 2-D cutoff as running variable. The distance variable exploits both the running variables but it comes at the cost that the interpretation of the coefficient is not the same as typical RD because distance variable is always positive. Additionally, I also check for the second dimension, turnout percentage in past election.

Treatment in 2012 election can be defined as:

$$T_b = \left\{ \begin{array}{ll} 1 & v_b \geq 0.75 \\ 0 & v_b < 0.75 \end{array} \right.$$

where v_b is the maximum vote share for any candidate in booth b in the election of 2009. In short, treatment is positive when either the candidate with maximum votes in past election got more than 75% of the votes.

The main empirical strategy in this case follows Imbens and Lemieux (2008) who propose that Local Linear Regression around the cutoff has better properties in terms of bias and precision. [20] In particular, I run the following regression:

$$y_b = \alpha_0 + \alpha_1 \tilde{v_b} + \beta T_b \times \tilde{v_b} + \epsilon_b \tag{1}$$

where $\tilde{v_b} = v_b - 0.75$ is the running variable centered at 0. [21]Note that T_b is defined on the basis of election results at booth b in *previous* election. y_b is turnout by gender, vote share of the incumbent party, vote share of the female candidates, vote share of 'criminal' candidates, vote share of re-contesting candidates at booth b. I weigh (1) by using *triangular* or *edge* kernel following Meng-Yen-Ching, Fan, James S Marron (1997) who show that *triangular* kernel is boundary optimal. Intuitively, this bandwidth allows for greater weight to observations close to the cutoff and the weights decrease as distance to cutoff increases. Specifically, the edge kernel is expressed as:

$$K_h(r_{ik}) = \mathbb{1}\left(|\frac{r_{ik}}{h}| \leq 1\right) . \left(1 - |\frac{r_{ik}}{h}|\right) \tag{2}$$

where h denotes the bandwidth. Unless specified, I use optimal bandwidth proposed by Imbens and Kalyanaraman (2012). This is a data-dependent bandwidth and is obtained by minimizing mean squared error loss function with a regularization adjustment.

[20]Cheng et al. (1997) show that for a boundary point the local linear method is 100% efficient among linear estimators in a minimax sense.

[21]Since both the running variables have the same cutoff

Our main estimate of interest is $\hat{\beta}$. Equation (1) gives unbiased estimate under the following assumptions:

1. **Assumption 1**: The conditional expectation of counterfactual outcomes is continuous in the running variable around the cutoff.

2. **Assumption 2**: There is no manipulation of the running variable around the cutoff.

Assumption 1 implies that pre-determined characteristics should be smooth around the cutoff. One way to check this in the present analysis is to look at the effect of the cutoff in the baseline election, 2009 in our context. In figure A.1, I plot female turnout, male turnout and total turnout in 2009 on the maximum vote share running variable in 2009. There appears to be no significant jump at the threshold for either variable. The slight negative effect is also shown to be insignificant in Table A.2. In addition, I check if there is something special about this cutoff in pre-treatment year 2007. Table A.1 confirms that none of these variables jump at the cutoff. This reinforces my belief that Assumption 1 is likely to be true.

Assumption 2 can be partially tested by using McCrary Density test (2008) which looks at discontinuity in the local linear density estimator of the density function of the running variable around the cutoff. If there is no *monotonic* manipulation, then the local linear density estimator should be smooth around the cutoff. Figure C.10 shows no evidence of such discontinuity in the density of the running variable at the cutoff [22]

One word of caution is in order. Since we don't have any information on which booths were actually treated, β identifies *reduced-form* or *intention-to-treat* effect, that is the effect of being above the cutoff on the outcomes. Since it is possible that many booths above the cutoff may not be treated, the reduced form effect is going to be smaller. If we had information on which booths were actually treated, then we would have the first-stage and would allow us to calculate Local Average Treatment Effect (LATE) which is *intention-to-treat* effect scaled down by probability of treatment. In the absence of first-stage, β could be interpreted as the *lower* bound of the actual effect when the treatment effect is positive.

[22] One should keep in mind that this test is only informative when the manipulation is *monotonic*. That is the existence of the program induces agents to adjust the running variable in one direction only. It is possible that the design passes the McCrary Density test yet there is manipulation. This could be the case if because of the fear of declaring a booth critical, the Election authorities reduce the highest vote share received by the candidate in the previous election for some booths but also at the same time, increase the highest vote share received by the candidate in the previous election for other booths. If the mass of booths that go to the right of the cutoff is the same as the mass that goes to the left, then the McCrary Density Test will fail to detect any manipulation even though there is one. However, in the current context it seems unlikely that there is *non-monontonic* manipulation in the elections. At best, one should expect political parties engaged in electoral fraud/violent intimidation reduce the vote share to reduce the probability of a booth being treated.

The next section discusses the results of the effect of treatment on turnout by gender and vote shares for different candidates.

5 Results and Discussion

5.1 Effect on turnout by gender

To begin with, as a baseline check, I plot in Figure A.2 for turnout by gender in 2009 as in Lee(2008). The dots represent unconditional binned averages in bin width of 0.05. The line represents local linear fit estimated using IK Bandwidth. There is no evidence of any significant jump in the female or total turnout around the cutoff. The drop in male turnout is insignificant. Another baseline check is to look for jump in registered voters around the cutoff in 2012. This is shown in figure A.2. Again there is no evidence of any discontinuity. This provides some evidence in favor of Assumption 1.

Figure B.1 plots the effect of extra security in the polling booths on female and total turnout. In order to reduce noise in the plots due to fewer observations in each bandwidth, I plot turnouts in 2012 as share of registered voters in 2012. Both total turnout and female turnout go up at the cutoff, however there is no jump in male turnout. Table B.1 reports the regression estimates for turnout by gender using IK bandwidth. For female turnout, the jump is significant and positive for almost all reported bandwidths and it ranges from 1.5% to 1.96%. The male turnout is insignificant for all bandwidths. Total turnout is significant only at 1.2IK bandwidth. This implies that one way in which increased security affects the composition of voters is through greater share of women voters. This is consistent with the hypothesis that women are generally more affected by electoral violence and malpractices therefore cleaning up elections impacts them more.

5.2 Effect on Political Outcomes

It is crucial to analyze the effect of the change in the composition of voters on voting outcomes. To this end, I use the polling data merged with candidate characteristics. As emphasized before, the following discussion only reports *reduced form* effect of increasing security at the polling booths. This is since I don't have information on which booths were actualy treated. In particular, I run the following regression for each political outcome varaible:

$$y_b = \alpha_0 + \alpha_1 \tilde{v}_b + \beta T_b \times \tilde{v}_b + \epsilon_b \tag{3}$$

And for purpose of brevity and visual appeal, I limit my analysis to RD in one dimension by using the maximum vote share of candidate in past election as the running variable. (denoted by $\tilde{v_b}$ here)

First, I look at the effect of increased security on vote share of candidates who have at least one self-disclosed criminal case against them. The candidates are required by law to provide information about pending criminal cases against them. The definition of 'criminal' cases is fairly broad and covers all cases under the Indian Penal Code. Some of the cases under which the candidate could be charged are criminal conspiracy, offences against public tranquility, offences relating to public servants, false evidence against public justice, offences affecting life, wrongful restraint and confinement, defamation etc. It is worth noting that these cases are not allegations but indictment. There is substantial discourse in the media on criminality in Indian politics. Figure C.1 plots the vote share of 'criminal' candidates against the running variable. There is clear significant jump in the vote share. There could two possible reasons why enhancing security at the polling booths and hence higher women turnout instead helps 'criminal' politicians. On one hand since women are more anti-crime, higher women turnout should hurt the 'criminal' candidates. On the other hand, the literature has pointed to 'ability bias' (Vaishnav (2012) whereby 'criminality' of the candidate serves as signal of the ability to lead and to protect the interests of his supporters. If latter effect dominates the former among women voters, it is possible that 'criminal' candidates benefit from greater share of women voters.

Second, figure C.2 and C.3 plot the vote shares for incumbent and challenger party. There is clear significant drop in vote share of the incumbent party and similar gain for challenger party. Table C.1 reports that the results are significant for alternate bandwidths. The loss to incumbent party is between 1.3% to 1.6% while the gain for challenger party is between 1.7% to 2.1%. This is consistent with the hypothesis that women are 'change agents' and anti-incumbent. Ravi and Kapoor (2014) discuss that incumbent parties perform worse in states with higher women turnout. Third, a natural outcome to consider is to look at the effect on female candidates. Surprisingly, female candidates witness a drop in their vote share as in Figure C.4. However as Table C.1 reports this is not significant. This maybe due to lack of power as there are very few female candidates. Hence if we condition only on the booths where there is at least female candidate, it increases power (row 5) and there is a drop of about 1.3% but only at 1.2IK bandwidth. This implies that *gender bias* effect dominates *gender affinity* effect. Beaman et al(2009) have shown that women are generally not perceived to be capable of holding local leadership positions. The present results show that there is possibility of presence of *gender bias* even among women voters for elections to higher levels of office, though one must keep in mind that this is significant only at 3 percent.

Finally, I look at the effect on vote share of re-contesting candidates who have reported positive growth in net assets while in office (C.5). Fisman et al. (2015) have used this as proxy of corruption in office. There is negative jump at the cutoff but this is not significant in full sample. This is because there are very few candidates who actually re-contest elections, therefore results are significant at IK bandwidth when I restrict the sample to include only those booths where there is at least one candidate re-contesting election, thereby increasing power. The treatment effect is around 0.01% which is quite small since there are very few re-running candidates to begin with. This result is consistent with the hypothesis that women are generally more anti-corruption than men. Women are generally main beneficiaries of redistribution programs like NREGA, subsidized grains, pensions etc and are more likely to be affected by corrupt politicians and bureaucrats. Hence, when women turnout in larger numbers, they are less likely to vote for 'corrupt' politicians. Anand (2001) provides micro and macro evidence regarding women being less tolerant to corrupt practices like bribery etc. Using cross-section of studies, Dollar (2001) has shown that greater representation of women in parliament is associated with lower levels of corruption.

6 Robustness Checks

6.1 Alternative running variable: turnout share in 2009

Even though the quasi experiment provides two running variables, in my analysis I mainly focus on first running variable: maximum vote share for any candidate in 2009. This is because this regression has more power. As robustness check, I use turnout share in 2009 as the running variable. As Table B.2 shows, the magnitude of coefficient of the effect on turnout by gender is mostly similar to those in Table B.1, however, none of them are significant. Therefore, I focus on maximum vote share as the running variable for all regressions in this paper.

6.2 Using distance to the cutoff as running variable

Another concern is that using only one dimension of running variable leads to loss of information. Hence, alternatively as in Dell(2010), I use distance to the two dimensional cutoff as running variable. In particular, I run the following regression:

$$y_b = \alpha_0 + \alpha_1 treat_b + \alpha_2 treat_b \times f(dis_b) + f(dis_b) + \epsilon_b \tag{4}$$

where $treat_b$ takes value 1 if either maximum vote share or turnout in 2009 at booth b is more than 0.75. dis is the Euclidean distance of two points from the cutoff. $f(dis_b)$ is polynomial in distance to the cutoff. Table B.4 shows the results for many cutoffs and linear, quadratic and cubic polynomial of distance. To conserve space, I focus on female turnout as share of total turnout as the outcome variable. I find that appropriate bandwidth in terms of stability of estimates is around 0.05 and effect varies from 1.8% in linear specification to 6.7% in cubic specification. This estimate is slightly higher than 2% obtained by using maximum vote share as running variable. However, results are mostly significant in appropriate bandwidths and specifications which increases confidence on using maximum vote share as the running variable.

6.3 Is the increase in women turnout an actual increase or gender-biased reduction in ballots-stuffing?

In the above sections, I have argued that increase in security at polling booths increased women turnout by about 2-6% which is likely to be the mediator of effect of increased security on various political outcomes. One concern is that the 2-6% increase in 'reported' turnout of women may not be increase in actual numbers of women who voted but instead reflect a gender-biased reduction in ballot stuffing. This is a relevant concern since prior to treatment of booths by the Election Commission, one of the key features of electoral malpractice was illegal casting of ballots in the absence of actual voters. I refer to this act as 'ballot-stuffing'. Ballot-stuffing is more likely to happen in situations where booths are captured by strong men and voters are intimidated to enter the polling booth. Substantial gender gap in voter registration and voter turnout over time has been documented in India (Kapoor and Ravi (2014)) that may encourage delinquents to stuff greater proportion of votes for males than females to avoid raising suspicion. Whether ballot-stuffing biases the treatment effect on turnout of women depends on whether it is gender neutral.

If ballot-stuffing is independent of gender-composition of voters and increase in security at the polling booths leads to reduction in probability of occurrence of ballot-stuffing, then effect of treatment around the threshold should capture increase in actual turnout of women. This is because booths towards the right of the threshold are likely to be treated and are likely to have have lesser ballot stuffing than the booths that are towards the left. If vote-stuffing is gender neutral, then there is no reason to believe why ballots stuffed for females in the booths to the right of threshold are less than for booths to the left of threshold.

However if ballot-stuffing is not gender neutral, then part of the increase in women turnout could be increase in average share of ballots stuffed for females in the treated booths.

One reason why ballot-stuffing may not be gender neutral is when delinquents stuff ballots to ensure that gender distribution in final count resembles the distribution in voter registration. Since in general, sex ratio in voter registration is skewed towards men, there will be more votes stuffed for males than for females. In that case if treatment of a booth with higher security leads to reduction in ballot stuffing, that may imply higher composition of ballots stuffed for females on average to the right than to the left and that may be mistaken for higher turnout of women. One counter-argument for that is that if ballot stuffing is indeed gender-biased then this should reflect in *lower* reported turnout for men among treated booths. However as reported in Section 5.1, effect on male turnout is insignificant. Nevertheless, it is also possible that actual increase in turnout for men balances decrease in 'turnout' due to reduction in biased stuffing in treated booths resulting in insignificant overall effect on male turnout.

If delinquents target stuffed votes to be certain proportion of registered voters then it is possible to analyze the effect of treatment on ballot-stuffing. Consider the following model of relationship between turnout and registered voters:

$$t_i^b = t_a^i + p_s^i V^b \tag{5}$$

where t_i^b is 'reported' turnout at the booth for gender i, V^b is total registered voters at the booth. t_a^i are real votes and p_s^f is the proportion of registered voters which are stuffed of gender i.[23] This model assumes that delinquents follow a rule for ballot-stuffing and that leads to systematic correlation between turnout percentage and registered voters. [24]

A simple exercise would be to regress (5) on either side of the cutoff to analyze the average relationship for treated vs non-treated booths. Panel A of Table B.3 shows this relationship for each gender on either side of cutoff. There are two points worth mentioning here. First, female turnout coefficient is always less than male coefficient on either side. This follows from low share of women in registered voters, their lower turnout percentage and maybe lower proportion of ballots stuffed for female with respect to males. Second, the voting coefficient is smaller to the right (where booths are likely to be treated) than to the left for each gender. However, the drop for females is twice that of males (0.4 versus 0.2). This could in part reflect difference in votes stuffed for each gender on either sides. If we assume

[23]Note that this exercise uses *total* registered voters and not *female* voters. This is because gender decomposition of registered voters at the booth level is not available.

[24]Note that there are many factors like history, institutions, economic and political conditions that determine turnout percentage, hence in reality correlation between actual turnout percentage and registered voters is likely to be weak. However, if registration of voters is correlated with these factors then when regressing (5), the coefficient on registered voters captures this correlation. This implies that empirical counterpart of (5) does not have causal interpretation.

that increasing security at the polling booths reduces likelihood of ballot stuffing then the coefficients show that if anything, ballots stuffed for females drop higher than for males. In other words, treatment leads to gender-biased reduction in ballot-stuffing which reduces ratio of female-to-male in total votes polled. This intuitively shows that if gender-biased ballot-stuffing were to bias the estimate, it is likely to be biased downwards.

To see this clearly, I run the following regression:

$$y_b = \alpha_0 + \alpha_1 treat_b + \alpha_2 treat_b \times f(dis_b) + \alpha_3 treat_b \times RegVoters + f(dis_b) + \epsilon_b \qquad (6)$$

Then α_3 captures differential relationship between turnout and registered voters on either side of the cutoff. Increasing security at the polling booths is likely to alter this relationship since average number of treated booths is going to be higher on the right side of the cutoff and consequently average number of ballots stuffed is likely to be lower. Panel B of Table B.3 reports results for equation 6 and for cubic polynomial in distance to cutoff. Outcome variable is female turnout share. Coefficient of treatment remains unchanged after inclusion of registered voters variable. Interestingly, α_3 is negative and significant for larger bandwidths. This implies that correlation between turnout and registered voters is weaker when booths are more likely to be treated. This implies that reduction in ballot-stuffing as a result of increased security reduces votes stuffed for females. This increases confidence that increase in women turnout represents increase in number of real women voters.

6.4 Permutation Test

Is it possible that the estimated treatment coefficient at the cutoff reflects some unusual characteristic of the cutoff and not real effect? To check for this possibility, I perform a permutation test as in Chetty et al. (2009). I randomly draw a 'hypothetical' cutoff from $(0, 0.5)$. I then calculate treatment effect at this cutoff as in equation 4 using IK Bandwidth. Intuitively, if treatment had real effect on women turnout, we would expect estimated coefficient at the real cutoff to be in the tails of the distribution function.

Figure C.11 plots empirical cumulative distribution of treatment coefficients for 5000 such simulations. Vertical line denotes treatment coefficient (average figure for different empirical specifications) at the true cutoff of 0.75. Clearly, original treatment coefficient is an outlier. The p-value of estimated coefficient at true cutoff in simulated distribution is 0.78. This increases confidence that estimated treatment coefficient at original cutoff is not usual and reflects real treatment effect.

7 Conclusions

In his book, S.Y. Quraishi, Chief of the Election Commission who spearheaded major election reforms, explains that while studying gender behavior among voters in Bihar it was found that women felt the need for a more safe and secure environment for voting as violence at the polling booth was a common occurrence. This paper aims to provide concrete evidence in this regard and show that increasing security at the polling booths increases women turnout. This in turn has implications on political outcomes. Using RD Design, it is shown that women turnout goes up by 2-5 per cent points with no change in male turnout. Since the aim of enhancing security at Polling booths was to reduce electoral violence, this provides some evidence of women being victims of electoral violence.

The incumbent party's vote share is lowered as a result of the treatment by about 1.5 percentage points consistent with the hypothesis of women being 'change agents'. Female candidates fare worse suggesting that *gender bias* effect dominates *gender affinity* effect for women voters. Also, consistent with descriptive studies which point to women as being 'fairer' sex, I find that increase in women turnout lowers the vote share of 'corrupt' candidates. Finally, increasing security at the polling booths also leads to increase in vote share for candidates with at least one 'criminal' case against them. There are two possible factors which could be driving this result: on one hand, since women are more anti-crime, higher turnout of women should hurt the 'criminal' candidates, on the other hand, if women suffer from the ability bias where 'criminality' of the candidate serves as a signal of the ability to lead and protect the interests of his people, this could benefit 'criminal' candidates. The findings indicate that the latter effect dominates the former.

Some descriptive studies have highlighted the phenomenon of rising female turnout in recent years. For instance, just 3 years after the period in this study, in Bihar elections (another state known for electoral violence), women outnumbered men in turnout by 4 percentage points for the first time. Media reported such women as 'agents of change' and likely responsible for election of then Chief Minister with full majority. This *de-facto* enfranchisement of women of India can have substantial effects on political economy equilibrium of the country and this paper highlights some of the ways in which women turnout can affect this equilibrium such as election of less 'corrupt' candidates and challengers.

To the extent that political outcomes affect provision of public goods in the economy, the analysis in this paper suggests that significant economic effects can be achieved through higher security at the polling booths. Conduct of free and fair elections can help in bridging political representation gap between men and women.

8 References

Anderson, Siwan, Patrick Francois, and Ashok Kotwal. "Clientelism in Indian villages." The American Economic Review 105.6 (2015): 1780-1816.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. "Identification of causal effects using instrumental variables." *Journal of the American statistical Association* 91.434 (1996): 444-455.

Arulampalam, Wiji, et al. "Electoral goals and center-state transfers: A theoretical model and empirical evidence from India." *Journal of Development Economics* 88.1 (2009): 103-119.

Asher, Sam, and Paul Novosad. "Politics and Local Economic Growth: Evidence from India." Cambridge, MA: Harvard University, unpublished (2013).

Austin, Dennis. "Democracy and Violence in India and Sri Lanka", *Contemporary Southeast Asia* (1996).

Baland, Jean-Marie, and James A. Robinson. Land and power: theory and evidence from Chile. No. w12517. National Bureau of Economic Research, 2006.

Banerjee, Abhijit, Lakshmi Iyer, and Rohini Somanathan. "History, social divisions, and public goods in rural india." *Journal of the European Economic Association* 3.2?3 (2005): 639-647.

Banerjee, Abhijit, et al. "Do informed voters make better choices? Experimental evidence from urban India." Unpublished manuscript. http://www. povertyactionlab. org/node/2764 (2010).

Banerjee, Abhijit, et al. "Can voters be primed to choose better legislators? Experimental evidence from rural India." Presented and the Political Economics Seminar, Stanford University. 2010.

Beaman, L., et al. "Powerful women: does exposure reduce bias?." Quarterly Journal of Economics 124.4 (2009): 1497-1540.

Besley, Timothy J., and Robin Burgess. "The political economy of government responsiveness: Theory and evidence from India." (2001).

Broockman, David E. "Do female politicians empower women to vote or run for office? A regression discontinuity approach." *Electoral Studies* 34 (2014): 190-204.

Collier, Paula, and Vicente, Pedro C. "Violence, bribery, and fraud: the political economy of elections in Sub-Saharan Africa", *Public Choice*

Chattopadhyay, Raghabendra, and Esther Duflo. "Women as policy makers: Evidence from a randomized policy experiment in India." *Econometrica* 72.5 (2004): 1409-1443.

Chaturvedi, A. (2005). Rigging elections with violence. *Public Choice*, 125, 189?202.

Chaves, Isaias N., Leopoldo Fergusson, and James A. Robinson. "He who counts elects: Determinants of fraud in the 1922 Colombian Presidential Election". No. w15127. National Bureau of Economic Research, 2009.

Chetty, Raj, Adam Looney, and Kory Kroft. 2009. "Salience and Taxation: Theory and Evidence." *American Economic Review*, 99(4): 1145-77.

Cheng, M.-Y., Fan, J., Marron, J.S., 1993. Minimax efficiency of local polynomial fit estimators at boundaries. Unpublished manuscript Series # 2098, Institute for Statistics, University of North Carolina.

Dell, Melissa. "The persistent effects of Peru's mining mita." *Econometrica* 78.6 (2010): 1863-1903.

DellaVigna, Stefano, et al. Voting to tell others. No. w19832. National Bureau of Economic Research, 2014.

Dollar, David, Raymond Fisman, and Roberta Gatti. "Are women really the "fairer" sex? Corruption and women in government." *Journal of Economic Behavior & Organization* 46.4 (2001): 423-429.

Ellman, Mathew and Wantchekon, Leonard, "Electoral Competition under the Threat of Political Unrest", *Quarterly Journal of Economics* (2000)

Fujiwara, Thomas. "Voting technology, political responsiveness, and infant health: evidence from Brazil." Econometrica 83.2 (2015): 423-464.

Gerber, Alan S., Donald P. Green, and Christopher W. Larimer. "Social pressure and voter turnout: Evidence from a large-scale field experiment." American Political Science Review 102.01 (2008): 33-48.

Angrist, Joshua D., Guido W. Imbens, and Donald B. Rubin. "Identification of causal effects using instrumental variables." Journal of the American statistical Association 91.434 (1996): 444-455. Imbens, Guido W., and Thomas Lemieux. "Regression discontinuity designs: A guide to practice." Journal of econometrics 142.2 (2008): 615-635.

Iyer, Lakshmi et al.. ?The Power of Political Voice: Women's Political Representation and Crime in India?. *American Economic Journal: Applied Economics* 4.4 (2012): 165?193.

Miller, Grant. "Women?s suffrage, political responsiveness, and child survival in American history." The Quarterly Journal of Economics 123.3 (2008): 1287.

Jensenius, Francesca Refsum. "Development from representation? A study of quotas for the scheduled castes in India." *American Economic Journal: Applied Economics* 7.3 (2015): 196-220.

Kapoor, Mudit, and Shamika Ravi. "Women voters in Indian democracy: A silent revolution." Available at SSRN 2231026 (2013).

Kumar, Sanjay, and Pranav Gupta. "Changing Patterns of Women's Turnout in Indian

Elections." *Studies in Indian Politics* 3.1 (2015): 7-18.

Lee, David S., and Thomas Lemieux. "Regression Discontinuity Designs in Economics." Journal of Economic Literature 48 (2010): 281-355.

Leon, Gianmarco. "Turnout, political preferences and information: Experimental evidence from Peru." Available at SSRN 1965669 (2011).

Lindert, Peter H. Growing public: Volume 1, the story: Social spending and economic growth since the eighteenth century. Vol. 1. *Cambridge University Press*, 2004.

McCrary, Justin. "Manipulation of the running variable in the regression discontinuity design: A density test." *Journal of Econometrics* 142.2 (2008): 698-714.

Pande, Rohini. "Can informed voters enforce better governance? Experiments in low-income democracies." Annu. Rev. Econ. 3.1 (2011): 215-237.

Quraishi, S.Y. "An Undocumented Wonder: The Making of The Great Indian Election", 2014 Rapoport, David C. & Weinberg,Leonard Elections and violence, Terrorism and Political Violence, 12:3-4, 15-50, DOI: 10.1080/09546550008427569 (2000)

Raymond Fisman, Florian Schulz, and Vikrant Vig, "The Private Returns to Public Office," *Journal of Political Economy* 122, no. 4 (August 2014): 806-862.

Fisman, Raymond, Florian Schulz, and Vikrant Vig. "Asset disclosure and political selection: Evidence from India." (2015).

Robinson, James A., and Ragnar Torvik. "The Real Swing Voter's Curse." American Economic Review 99.2 (2009): 310-15.

Schneider, Annemarie, et al. "Mapping urban areas by fusing multiple sources of coarse resolution remotely sensed data." Photogrammetric Engineering & Remote Sensing 69.12 (2003): 1377-1386.

Snyder, Jack. "From Voting to Violence: Democratization and Nationalist Conflict", New York: W.W. Norton, 2000, 382 pp

Swamy, Anand, et al. "Gender and corruption." *Journal of development economics* 64.1 (2001): 25-55.

Wilkinson, S. I. (2004). *Votes and violence: electoral competition and ethnic riots in India.* Cambridge: Cambridge University Press.

Zvesper, John. "From Bullets to Ballots:The Election of 1800 and the First Peaceful Transfer of Political Power", *Claremont Institute* (2003).

A Baseline Checks

Table A.1: Polling Booth Outcomes in 2007

	0.8IK	IK	1.2IK
Female Turnout	0.0314	0.0288	0.0261
	(1.56)	(1.59)	(1.57)
N	10748	14389	18495
Male Turnout	0.00956	0.00709	0.00544
	(0.63)	(0.51)	(0.43)
N	14222	19362	25471
Total Turnout	0.0192	0.0158	0.0136
	(1.15)	(1.05)	(0.98)
N	12049	16184	21026

The outcome variable is in log terms. Each term denotes the estimate of the coefficient using the bandwidth in the column. The term in parentheses denotes the t-statistic. Robust Standard errors are used. *** Significant at 1%, **Significant at 5%, *Significant at 10%. Triangular Kernel

Table A.2: Polling Booth Outcomes in 2009

	(1)	(2)	(3)
	Female	Male	Total
treat	-0.00123	-0.00302	-0.00483
	(-0.53)	(-1.02)	(-1.01)
N	17117	13005	14533

Outcome variable is expressed in log terms. IK Bandwidth is reported.
t statistics in parentheses. Robust standard errors.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Figure A.1: Turnout by gender in 2009

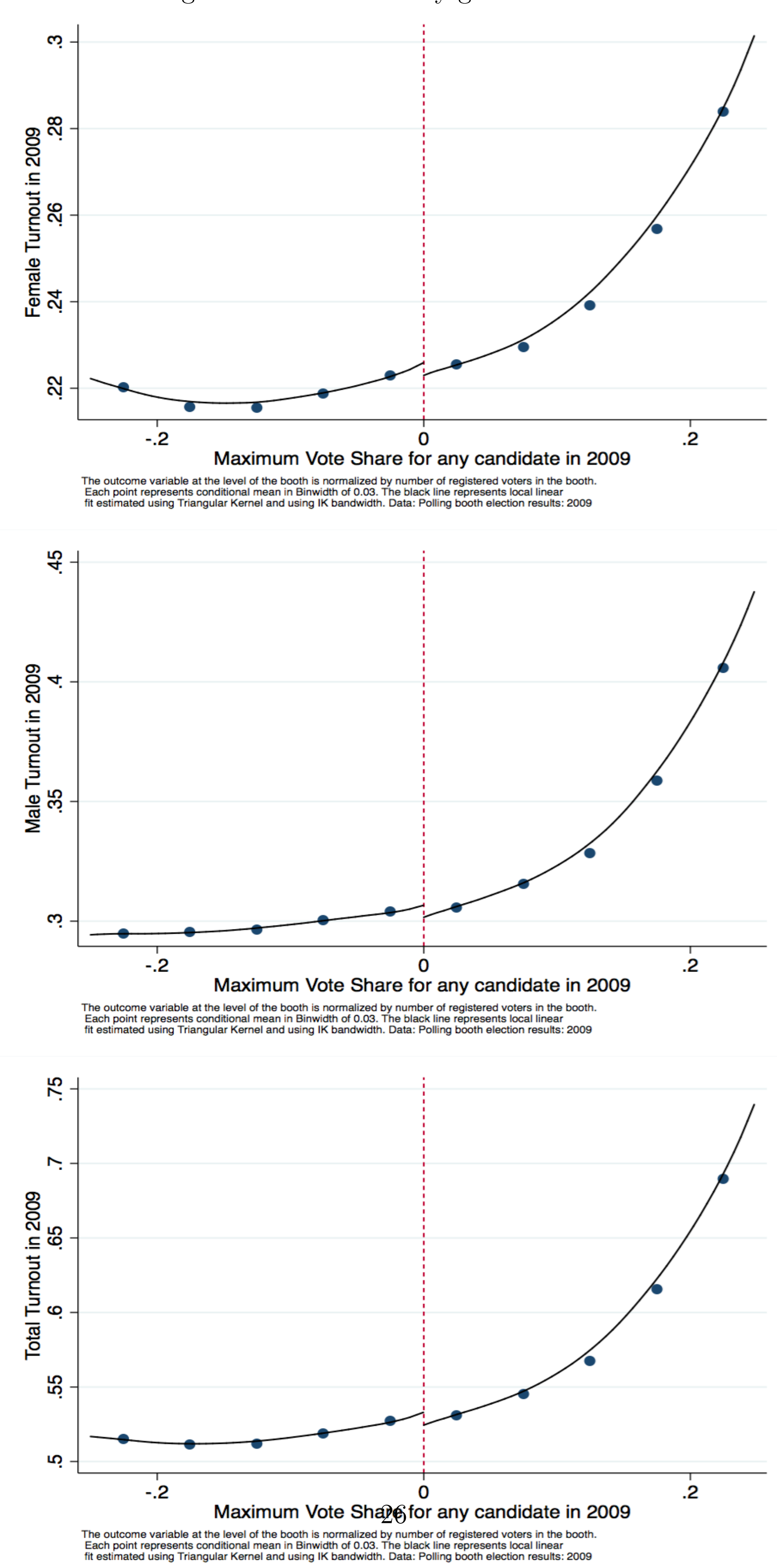
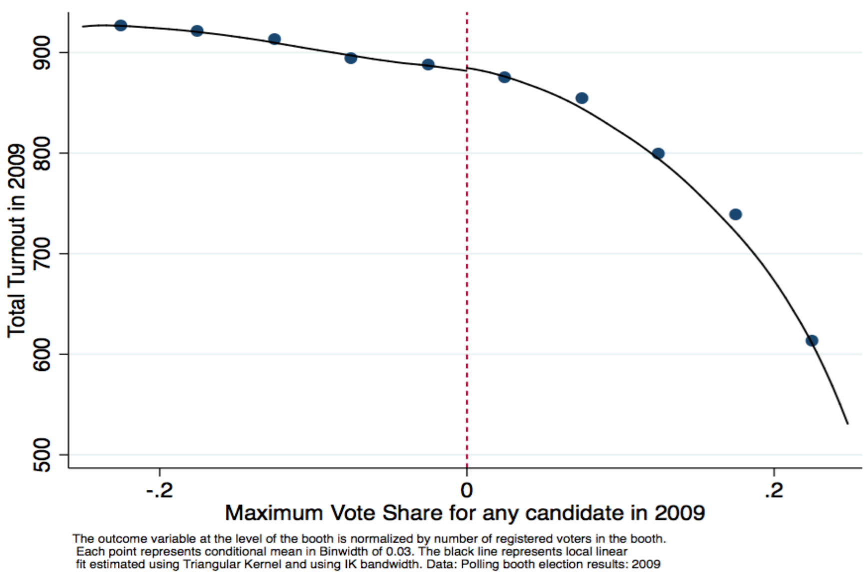


Figure A.2: Registered voters in 2012



B Electoral Outcomes

Table B.1: Outcomes using Maximum Vote Share in past election as Running Variable

	0.8IK	IK	1.2IK
Female Turnout	0.0154	0.0187*	0.0196**
	(1.35)	(1.81)	(2.05)
N	26861	35477	45805
Male Turnout	0.00704	0.00955	0.0134
	(0.48)	(0.73)	(1.11)
N	16051	20419	25261
Total Turnout	0.0133	0.0164	0.0178*
	(1.20)	(1.63)	(1.92)
N	26479	34945	45103

The outcome variable is turnout expressed in log terms. Running variable is the maximum vote share at the booth in past election. Each term denotes the estimate of the treatment effect using the bandwidth in the column. The term in parentheses denotes the t-statistic. Robust Standard errors are used. *** Significant at 1%, **Significant at 5%, *Significant at 10%. Triangular Kernel

Table B.2: Outcomes using Turnout share in past election Running Variable

	0.8IK	IK	1.2IK
Female Turnout	0.0222	0.0214	0.0181
	(1.11)	(1.17)	(1.06)
N	11423	15522	20508
Male Turnout	0.0265	0.0306	0.0306
	(1.04)	(1.33)	(1.44)
N	6078	7780	9654
Total Turnout	0.0251	0.0253	0.0243
	(1.10)	(1.23)	(1.28)
N	7554	9812	12492

The outcome variable is turnout expressed in log terms. Running variable is the share of voters who voted in past election. Each term denotes the estimate of the treatment effect using the bandwidth in the column. The term in parentheses denotes the t-statistic. Robust Standard errors are used. *** Significant at 1%, **Significant at 5%, *Significant at 10%. Triangular Kernel

Table B.3: Robustness Check for Section 6.3

A: Voting coefficient by gender		
Gender	To the left of cutoff	To the right of cutoff
Female	0.22***	0.166***
	(0.0011)	(0.003)
Male	0.30***	0.28***
	(0.0009)	(0.002)
Total	0.52***	0.45***
	(0.001)	(0.005)
B: Treatment adjusted for Vote Stuffing		
Bandwidth	Treat	TreatX Voters12
0.05	0.07**	-0.0000119
	(0.03)	(8.27×10^{-6})
0.06	0.09***	-0.0000205***
	(0.02)	(7.44×10^{-6})
0.07	0.08***	-0.0000251***
	(0.02)	(7.44×10^{-6})

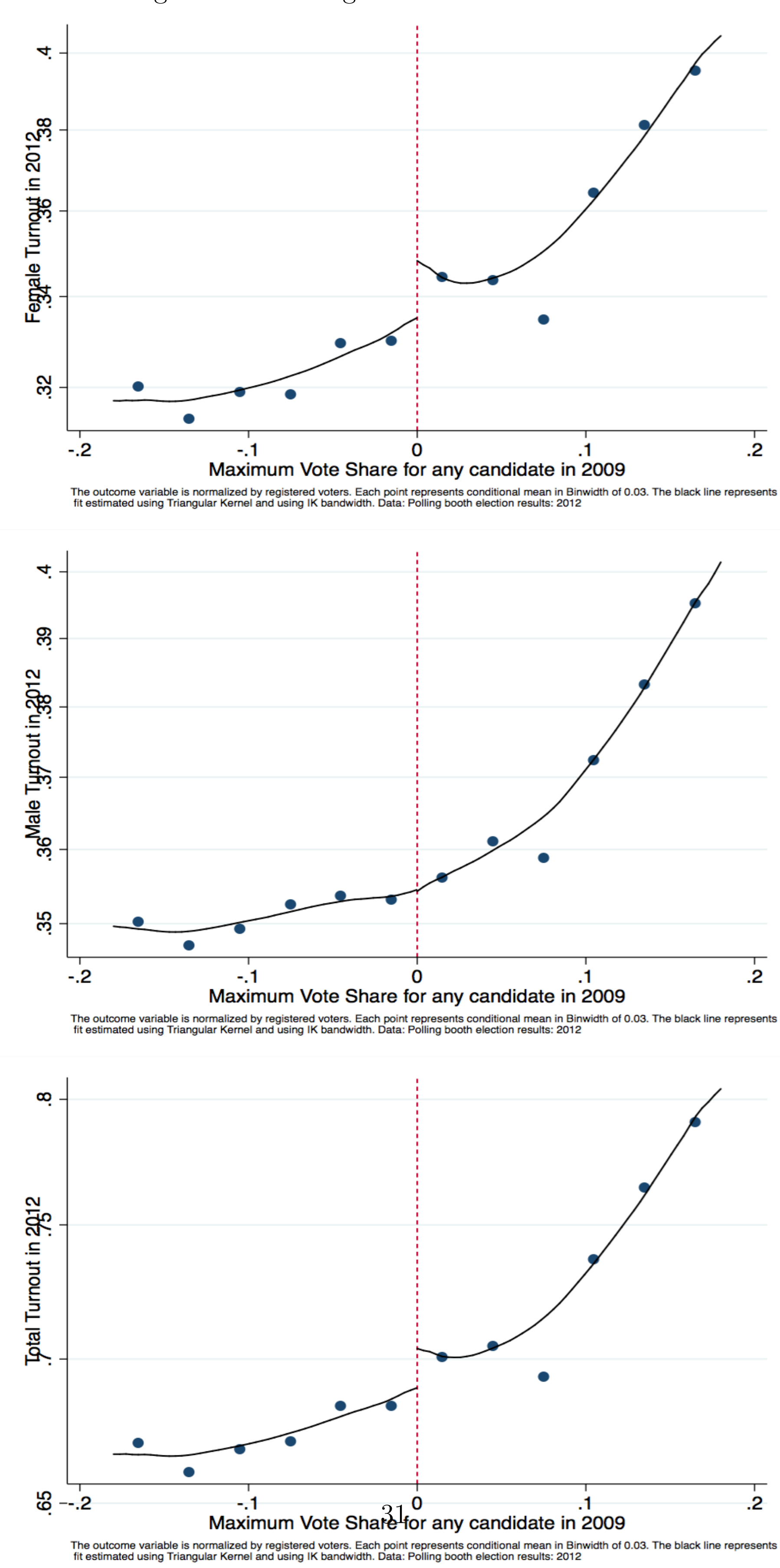
The term in parentheses denotes the t-statistic. Robust Standard errors are used. *** Significant at 1%, **Significant at 5%, *Significant at 10%. Panel A denotes coefficients of regression (5) and (6). Observations to the left of cutoff include those which have maximum vote share less than 0.75 and observations to the right are those which have maximum vote share at the booth above 0.75. Outcome Variable in Panel B is female turnout share. Regression is based on equation 6 and controls for cubic polynomial.

Table B.4: Outcomes using distance to 2D Cutoff as Running Variable under different specifications

	(1)	(2)	(3)	(4)	(5)	(6)
			Bandwidth=0.03			
treat	0.0144*	0.0137	0.0141*	0.0794***	0.0218	0.0500
	(1.72)	(1.61)	(1.65)	(3.71)	(0.51)	(0.71)
N	74	74	74	74	74	74
			Bandwidth=0.04			
treat	0.0112*	0.0123*	0.0129**	0.0455**	0.0778**	0.0467
	(1.71)	(1.92)	(1.98)	(2.56)	(2.20)	(0.99)
N	173	173	173	173	173	173
			Bandwidth=0.05			
treat	0.00366	0.00304	0.00252	0.0181	0.0565***	0.0677**
	(0.83)	(0.69)	(0.58)	(1.51)	(2.77)	(2.06)
N	477	477	477	477	477	477
			Bandwidth=0.06			
treat	0.00498	0.00510	0.00421	0.0109	0.0420**	0.0794***
	(1.42)	(1.47)	(1.22)	(1.03)	(2.30)	(2.88)
N	680	680	680	680	680	680

The outcome variable is ratio of female turnout to total turnout in 2012. Running variable is the distance to the two dimensional cutoff of maximum vote share and turnout share in past election. Each term denotes the estimate of the treatment effect using the bandwidth in the panel.(1), (2) and (3) specifications control for linear, quadratic and cubic polynomials in distance respectively. (4), (5) and (6) also control for such polynomials interacted with treatment dummy. The term in parentheses denotes the t-statistic. Robust Standard errors are used. *** Significant at 1%, **Significant at 5%, *Significant at 10%. Triangular Kernel

Figure B.1: Polling Booth Outcomes in 2012



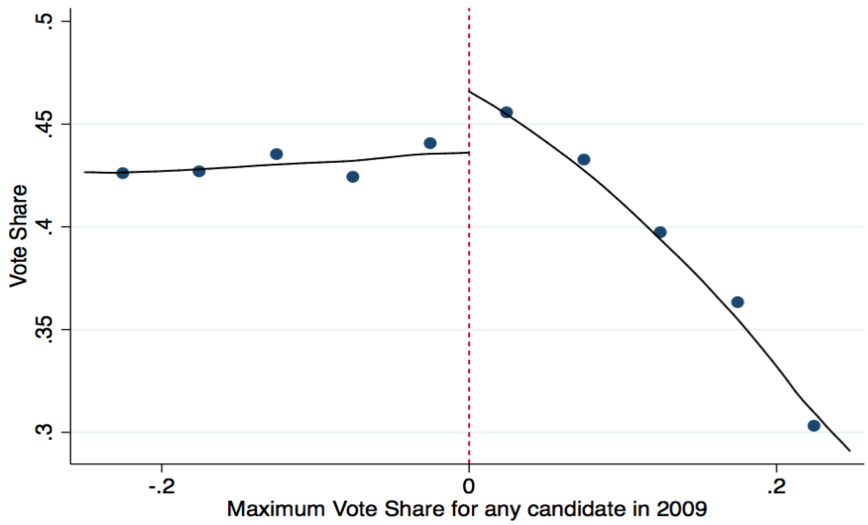
C Political Outcomes

Table C.1: Political Outcomes 2012: Vote shares

	0.8IK	IK	1.2IK
Incumbent	-0.0136**	-0.0157***	-0.0168***
	(-2.07)	(-2.65)	(-3.04)
N	24727	32455	41164
Challenger	0.0213**	0.0187**	0.0177**
	(2.02)	(1.97)	(2.03)
N	12393	15750	19401
Atleast one self-reported	0.0268*	0.0299**	0.0314***
‘criminal’ case	(1.92)	(2.37)	(2.72)
N	14480	18549	23148
Female candidates	-0.00816	-0.00724	-0.00641
(Unconditional)	(-1.40)	(-1.39)	(-1.33)
N	16373	21290	26628
Female candidates	-0.0146	-0.0144	-0.0138*
(Conditional)	(-1.45)	(-1.58)	(-1.65)
N	8139	10415	12929
Rerunning candidates positive net	-0.0000878	-0.000107	-0.0000973
asset growth (unconditional)	(-1.33)	(-1.40)	(-1.18)
N	17339	22675	28584
Rerunning candidates positive net	-0.0000963	-0.000151**	-0.000182**
asset growth (conditional)	(-1.07)	(-2.19)	(-2.13)
N	7569	9564	11641

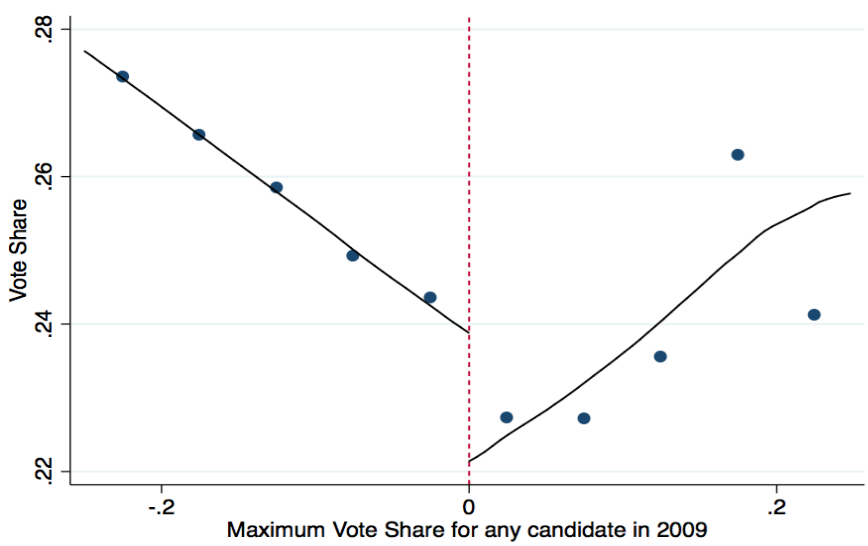
Each term denotes the estimate of the treatment effect using the bandwidth in the column. The term in parentheses denotes the t-statistic. Robust Standard errors are used. *** Significant at 1%, **Significant at 5%, *Significant at 10%. Triangular Kernel

Figure C.1: Vote Share of 'Criminal' Candidate



Each point represents conditional mean in Binwidth of 0.05. The black line represents local linear fit estimated using Triangular Kernel and using IK bandwidth. Data: Polling booth election results: 2012 merged with politicians characteristics 2012.

Figure C.2: Vote Share of Incumbent



Each point represents conditional mean in Binwidth of 0.03. The black line represents local linear fit estimated using Triangular Kernel and using IK bandwidth. Data: Polling booth election results: 2012

Figure C.3: Vote Share of Challenger Candidate

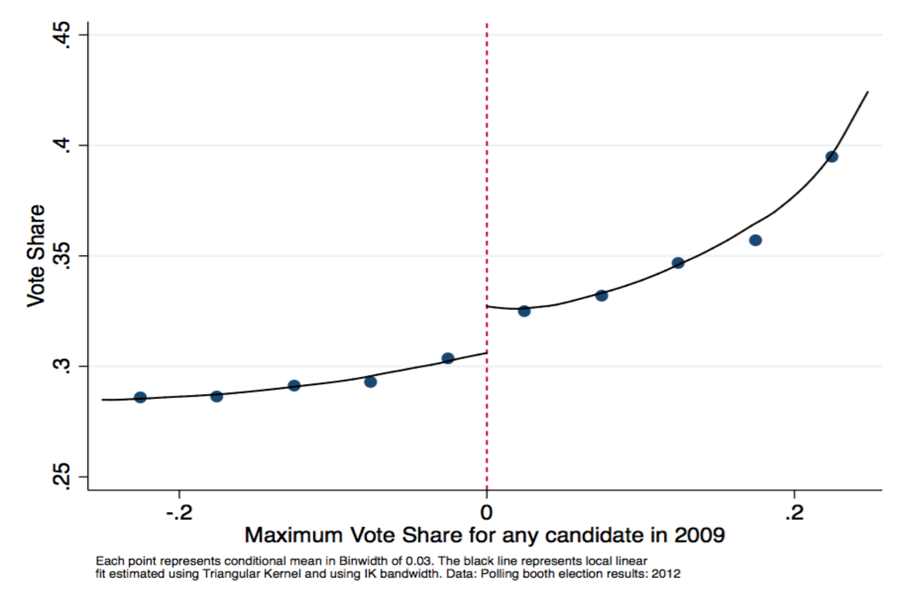


Figure C.4: Vote Share of Female Candidates (Unconditional)

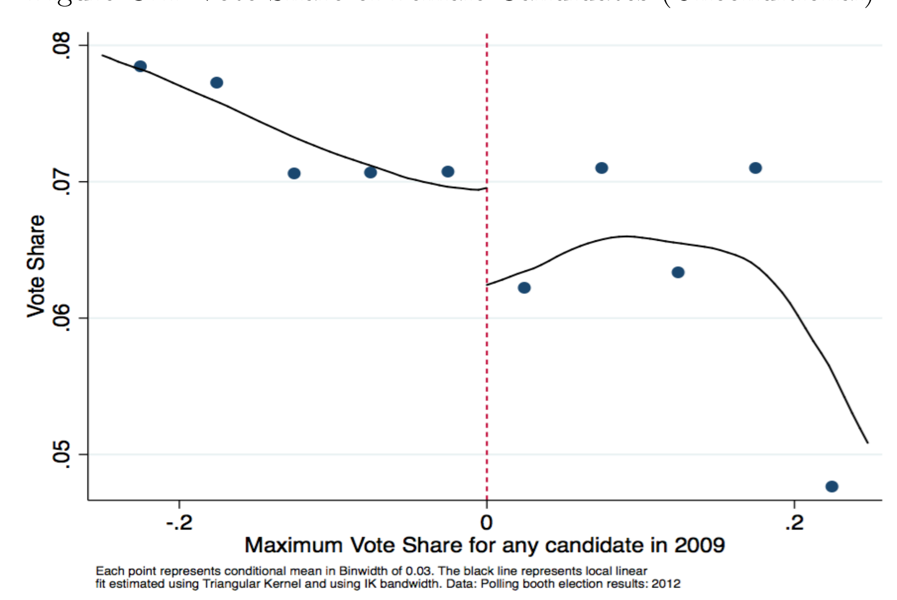
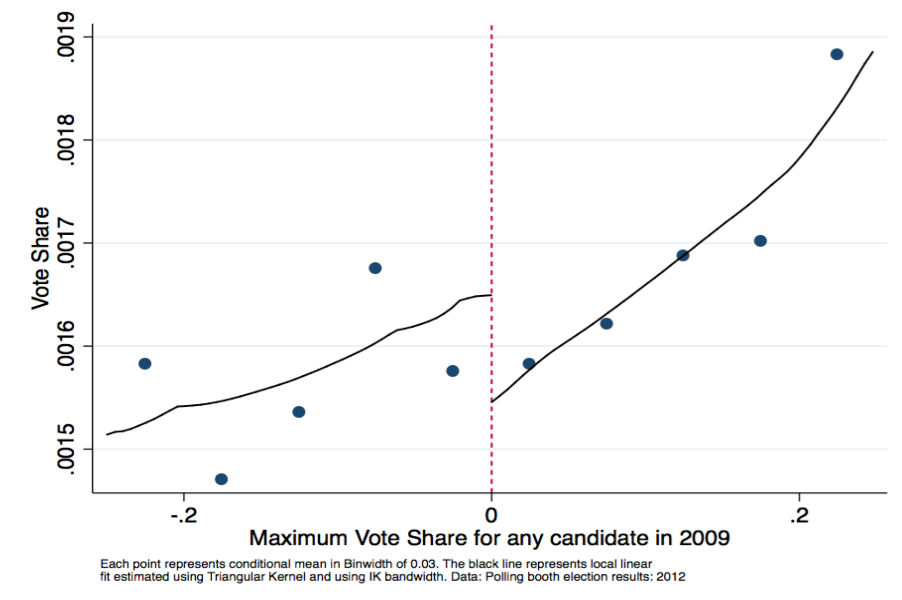


Figure C.5: Vote Share of Rerunning Candidate and positive growth in Assets (Unconditional)



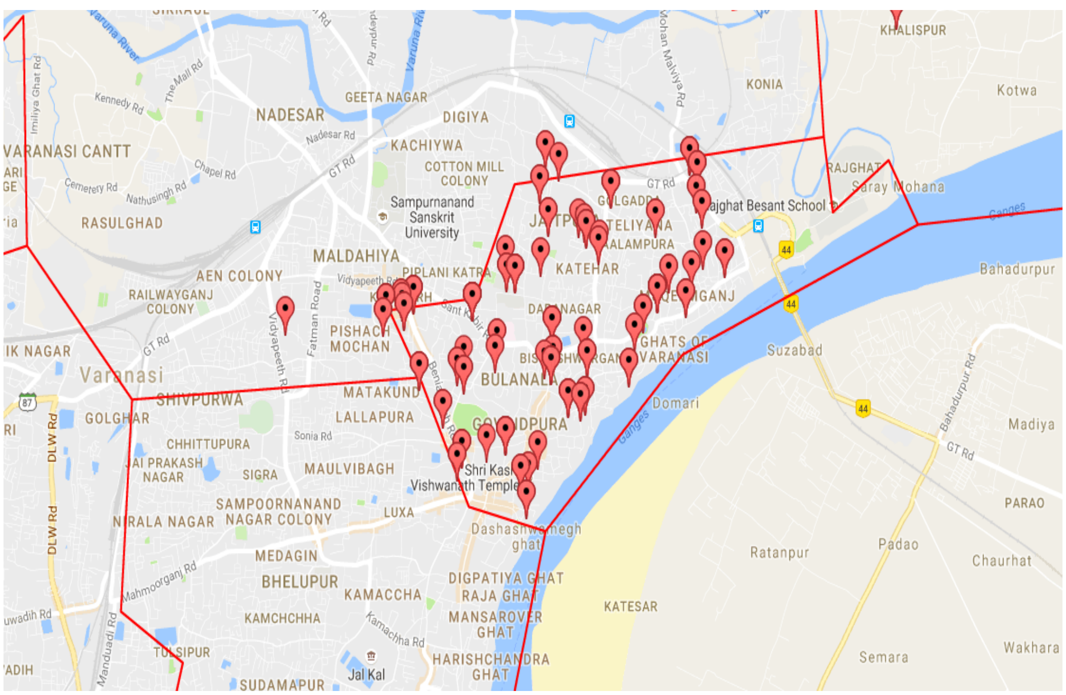


Figure C.6: This figure shows the location of Polling booths in 'Varanasi South' Assembly Constituency in Varanasi District of Uttar Pradesh as an example. The red lines depict the boundary of the constituency.

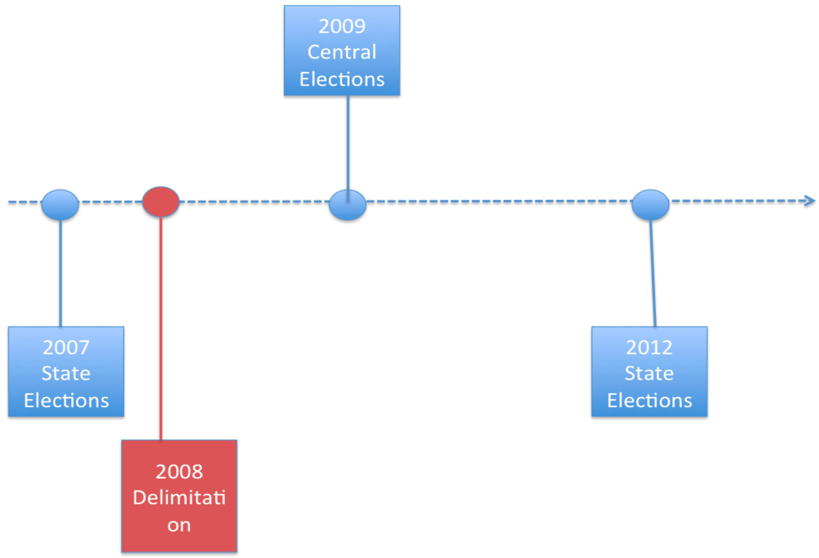


Figure C.7: Timeline in UP

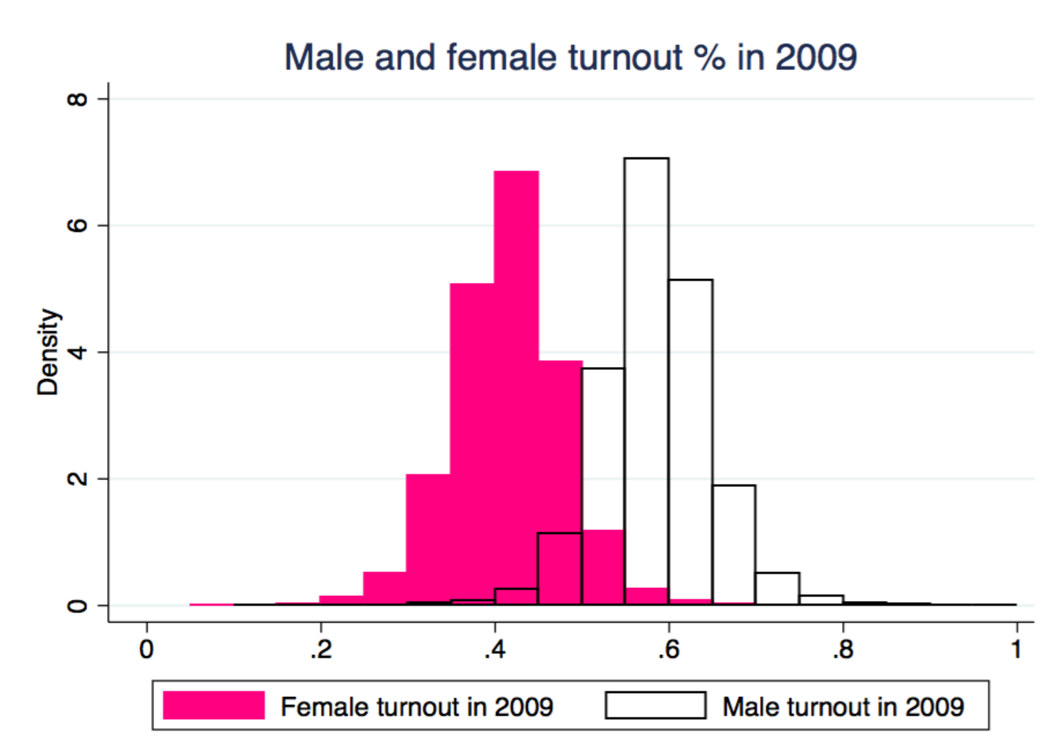


Figure C.8: This figure plots the density of turnout by gender as share of total turnout. Bin width=0.05

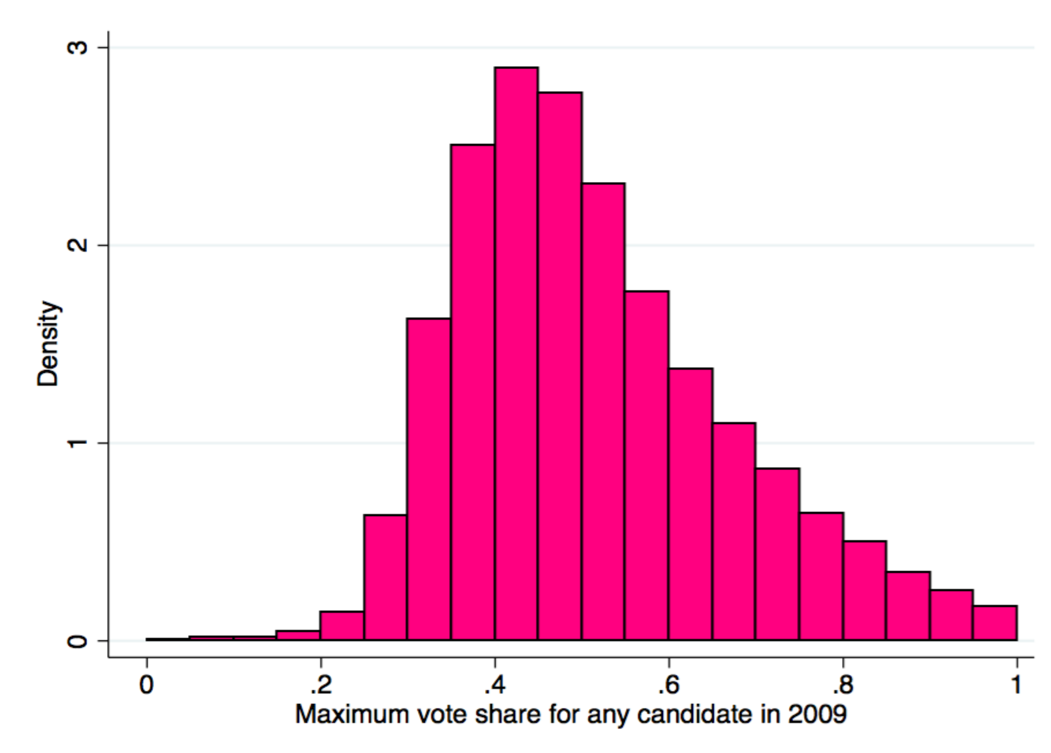


Figure C.9: This figure plots the distribution of the running variable: maximum vote share for any candidate in a polling booth. Bin width=0.05

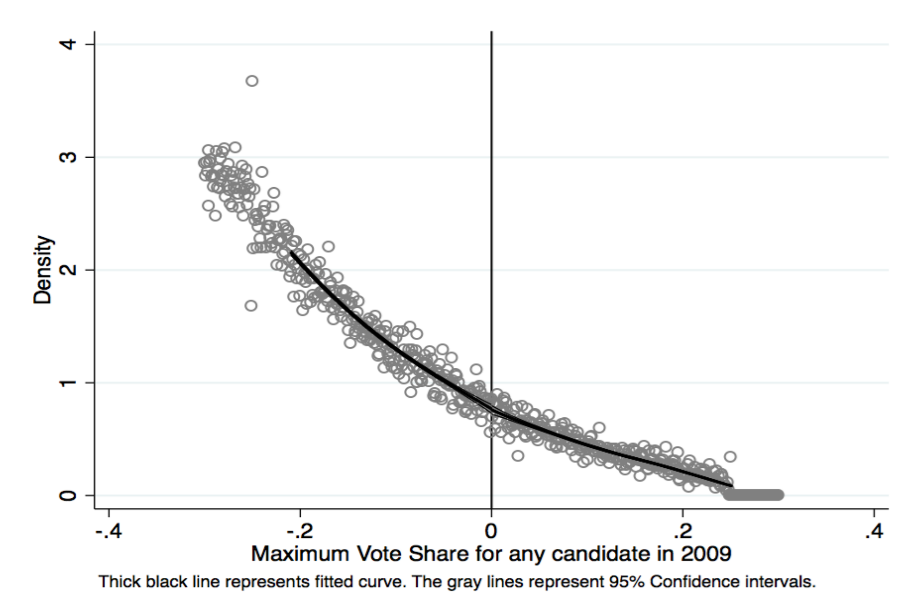


Figure C.10: McCrary Density Test

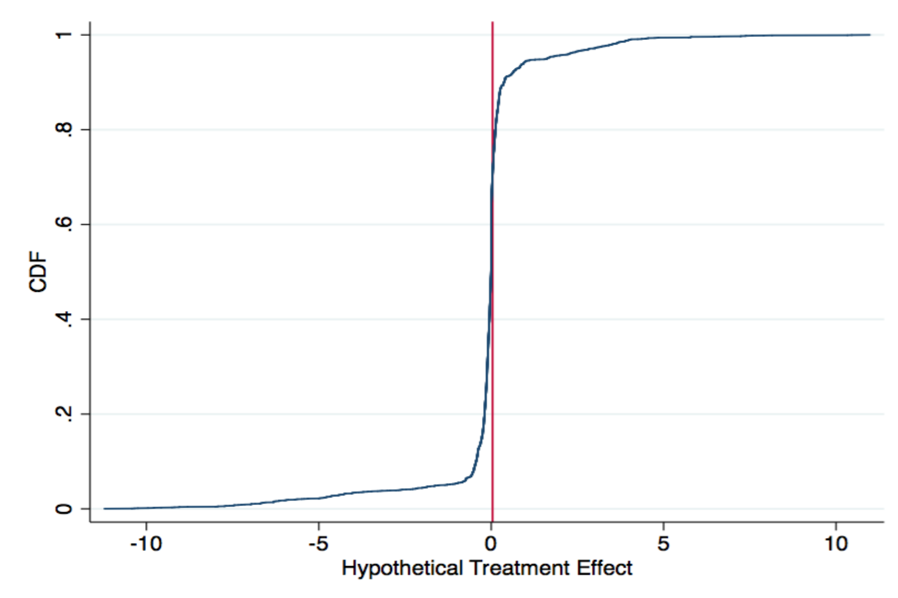


Figure C.11: Permutation Test

This figure plots the cumulative distribution of treatment coefficients with hypothetical cutoff points $\in (0, 0.5)$. Treatment effects are calculated as in equation 1 using IK Bandwidth. Vertical line represents treatment effect at actual cutoff= 0.75. Total simulations=5,000.

Figure C.12: Vote Share of Female Candidates (Conditional)

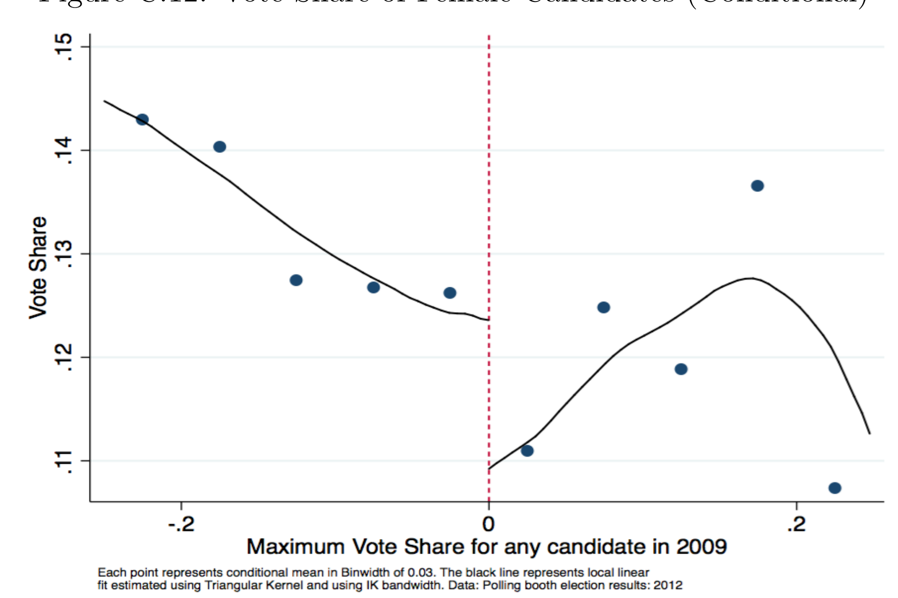


Table C.2: Summary Statistics

Variable	Mean	Std Dev	Min	Max	N
Booth data: 2007					
Female Turnout	194.26	73.83	43	387	86163
Male Turnout	273.66	88.92	82	498	86163
Total Turnout	468.10	153.59	133	852	86163
Registered Voters	1025.03	253.81	445	1546	85254
Maximum Vote share for any candidate	0.4857	0.1346	0.254	0.9967	86163
Booth data: 2012					
Female Turnout	266.41	94.70	87	554	109212
Male Turnout	316.12	113.75	98	638	109212
Total Turnout	582.86	198.35	195	1152	109212
Registered Voters	949.77	312.18	254	1578	108282
Maximum vote share for any candidate	0.4821	0.1511	0.212	0.99822	109212
Candidate Characteristics 2012					
No. of candidates	17.92	4.93	6	32	6717
No. of female candidates	1.37	1.24	0	6	6719
Years of Schooling	11.96	5.01	0	21	6719
Criminal cases	0.4832	1.77	0	44	6719
Age	43.81	11.86	0	87	6719
Net assets	10,729,268	113487383	-34,971,200	6,314,637,824	6719
No. of rerunning candidates	0.7827	0.6438	0	3	6719
Net Assets of rerunning candidates	23,615,654	57,140,518	-29,367,000	799,350,273	313
Asset growth of rerunning candidates	29,593,855	49,709,592	491,808	563,602,368	313

The election variables are at the booth level.The candidates data is at the level of assembly constituency. There are 403 Assembly Constituencies in Uttar Pradesh. All assets are reported in local currency. Candidate Characteristics obtained from affidavits filed by candidates. Polling booth level data obtained from Election Commission of India.

Figure C.13: Vote Share of Rerunning Candidate and positive growth in Assets (Conditional)