Essays in Development Economics: Incumbency Disadvantage, Political Competition, and Remedial Education in India

by

Leigh L. Linden

B.S. Mathematics
B.A. Economics
The University of Texas at Austin, 1997

SUBMITTED TO THE DEPARTMENT OF ECONOMICS IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN ECONOMICS
AT THE
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

SEPTEMBER 2004

Signature of Author: ______________________
Department of Economics
August 15, 2004

Certified by: ______________________
Esther Duflo
Castle Krob Career Development Associate Professor of Economics
Thesis Supervisor

Accepted by: ______________________
Peter Temin
Elisha Gray II Professor of Economics
Chairperson, Department Committee on Graduate Students

Essays in Development Economics: Incumbency Disadvantage, Political Competition, and Remedial Education in India

by

Leigh L. Linden

Submitted to the Department of Economics
on August 15, 2004 in Partial Fulfillment of the
Requirements for the Degree of
Doctor of Philosophy in Economics

ABSTRACT

This dissertation comprises three separate empirical studies. Using a non-parametric regression discontinuity design that compares candidates who barely win an election to those who barely lose, the first study estimates the effect of incumbency on a candidate's electoral prospects in India. Starting in 1991, I estimate that, rather than being at an advantage, incumbents are actually fourteen percent *less* likely to win an election than similar non-incumbents. While the available data prevent a formal test, the dominance of a single political party (the Indian National Congress) before 1991 may have provided a framework in which experience was valuable because incumbents who gained experience under the Congress system would interact with the same system when reelected. Starting in 1991, however, no party could be counted on to control parliament, making experience under the previous regime potentially less valuable.

The second study estimates the effects of new competitors on existing candidates in India by taking advantage of a change in the election laws in 1996 that made it more difficult for candidates to run for office. The law affected constituencies differently, allowing the use of both across time and between constituency variation in the number of candidates to estimate the impact of restricting the number of new candidates in an election. The resulting estimates suggest that the reduction in the number of new candidates had a small, but measurable effect on the probability that the average existing candidates would win election. However, there is evidence of heterogeneity in the effect across candidates.

Finally, the third study presents the results of a two-year randomized evaluation of a remedial education program in India. The remedial education program hires young women from the community to provide remedial assistance to third and fourth grade children who have fallen behind their peers. The program is extremely cheap (five dollars per child per year), and is easily replicable. We find the program to be very effective, increasing learning by 0.15 standard deviations in the first year, and 0.25 in the second year. The results are similar in the two grade levels, and in the two cities.

Thesis Supervisor: Esther Duflo
Title: Castle Krob Career Development Associate Professor of Economics

Thesis Supervisor: Sendhil Mullainathan
Title: Associate Professor of Economics

Acknowledgements

I am indebted to many individuals for their assistance during my tenure as a graduate student. I am, of course, indebted to my advisors, Esther Duflo and Sendhil Mullainathan who spent many hours patiently counseling me on the progress of my research. I am particularly indebted to Esther for being willing to take a chance by sending me, a young American student with no experience traveling internationally, halfway across the world to set up what eventually became the project evaluation in Chapter 3. Before she sent me, she made me promise that, if I did not like the experience of traveling to a developing country, I would give India a week to grow on me before coming home. I not only lasted more than a week – I kept going back and found a career. I will always be grateful. While not formally an advisor, Abhijit Banerjee generously provided guidance at many times during the writing of my thesis. I am also indebted to Susan Athey, Robert Gibbons, and Rebecca Henderson for guidance early in my career.

Each project in the dissertation could not have been completed with out the individual assistance of many people. For the political competition papers, I thank Daron Acemoglu, Joshua Angrist, David Autor, Kanchan Chandra, Victor Chernozhukov, Shawn Cole, Richard Eckaus, Michael Greenstone, Dean Karlan, Stuti Khemani, Rohini Pande, Michael Piore, James Snyder, the staff at the Election Commission of India. I am also indebted to the various seminar participants at MIT, NERA, NYU Wagner School, and Columbia University for helpful suggestions and comments.

The program evaluation in Chapter 3 was an enormous undertaking. I am, of course, indebted to my co-authors. I am also indebted to Rukmini Banerji and Madhav Chavan for helpful conversion, inspiration and guidance. I am also indebted to the Pratham team members, who made the evaluation possible and put up with endless changes in plans and new requests: Pratima Bandekar, Lekha Bhatt, Shekhar Hardikar, and Rajashree Kabare. I thank Jim Berry, Gauri Kartini Shastry, and Marc Shotland, for their excellent research assistance and work coordinating the fieldwork in Vadodara and Mumbia. They were helped by Nandit Bhat, Mukesh Prajapati and many others. For financial support, I thank the ICICI Corporation, the World Bank, and the MacArthur Foundation Network on the Costs of Inequality.

For emotional support and constant encouragement, I thank my parents Leslie and Lucille Linden, my brother Laine, and all of my many friends who put up with my ignoring them for so long during the job market. I am, however most indebted to Melissa Antenucci. Without her companionship and unwavering support over the last few years, I probably would not be where I am today. She has picked me up when I was most down, provided logistical support during my many travels, thought of details that would have never occurred to me, and most importantly, believed in my when even I doubted.

Finally, I am grateful to the Jacob K. Javitts Fellowship program, the MacArthur Foundation, and the Schultz Fund at MIT for financial support.

Contents

Chapter 1: Are Incumbents Really Advantaged? The Preference for Non-Incumbents in Indian National Elections

1.1 Introduction

A growing body of literature emphasizes the importance of political institutions in the process of economic development. While the exact nature of economically viable institutions remain unidentified, it is clear that the distribution of political power can determine the distribution of resources, even in a democracy (Acemoglu, 2002). In India, for example, villages where leadership positions are randomly reserved for female candidates are more likely to invest in infrastructure that meets the needs of women (Chattopadhyay and Dulfo, 2003a). Similarly, India reserves a large number of state and national parliament positions for historically disadvantaged groups known as scheduled castes and tribes. When candidates from these groups are elected, they use their influence to redirect resources to their constituents (Pande, 2003). The fact that political power makes such a difference is the reason that democratic governments are founded on the principle that voters should ultimately decide which representatives are chosen to wield power.

The major risk in a democracy is that elected officials will become entrenched or that running for office may simply prove too expensive. By the nature of the democratic system, incumbents are given access to resources and decision processes that non-incumbent challengers do not have. If elected officials are able to use their political influence to remain in power, voters will have no way to influence their policy decisions. Once entrenched, politicians are free to take distortionary actions such as simple graft or the evisceration of property rights. Stronger incumbents also raise the cost of entering politics because challengers must have sufficient resources to overcome the advantage of incumbency. These costs may then skew the allocation of political power even further towards the wealthy and effectively disenfranchise the poor, who may not have the resources to support candidates to represent their views.

Though mostly about the United States, a large literature has investigated the degree to which holding office affects a candidate's electoral prospects. These efforts have yielded two sets of empirical facts. First, incumbents enjoy significant advantages compared to their non-incumbent competitors. In terms of the raw probability of re-election, incumbents in national congressional elections are fifty percent more likely than similar non-incumbent peers to be re-elected (Lee, 2001). Second, the margin of victory of incumbents has increased significantly over time (Alford and Hibbing, 1989; Collie, 1981; Garand and Gross, 1984). The notable exception is Miguel and Zaidi's (2003) investigation of national elections in Ghana in which they find no evidence to support an incumbency advantage.

In this paper, I study the effects of incumbency on candidates' prospects for election in the Indian national parliament over the last fifteen years. Like in most developing democracies, critics have claimed that entrenched politicians and social elites control the political process. Running for office is expensive, and corruption seems to be rampant. Belying these criticisms, however, is the vibrancy of the Indian political system. Compared to elections in the US, a large percentage of eligible voters visit the polls, those voters have a dizzying array of candidates and political parties from which to choose, and an active and free press informs their vote (Besley and Burgess, 2002).

To estimate the effects of holding office on a candidate's probability of reelection, I use a non-parametric regression discontinuity design. First, I calculate the margin of victory for winning candidates as the winner's vote share less the vote share of the second place candidate. Similarly, for losing candidates, I calculated the margin of victory by subtracting the winner's vote share from the losing candidate's voter share. The election rules then guarantee that each candidate wins (and thus becomes an incumbent for the subsequent election) if the candidate gains a sufficient percentage of the vote that his or her margin of victory is greater than zero. Stated differently, assignment to incumbency status is discontinuous at zero: Candidates with a positive margin of victory win and those with a negative margin of victory lose. Because of the discontinuity, it is possible to infer the causal effect of incumbency status by comparing candidates that are just barely elected to those that just barely lose as long as all other candidate characteristics that could influence the probability of being elected vary, on average, continuously at zero (Lee, 2001; Miguel and Zaidi, 2003). Mathematically, I estimate the effect of incumbency by non-parametrically estimating the relationship between the probability of winning a given election and the margin of victory in the previous election while allowing for a discontinuity at zero. The size of the discontinuity is then the causal effect of being an incumbent prior to a given election.

The last fifteen years in India provide a particularly interesting environment for conducting such a comparison. First, the Congress party which dominated Indian politics since independence, lost its dominance in 1989. At the same time, the political system became much more competitive with large increases in both the number of nationally competitive parties and the number of candidates contesting in each constituency. Second, unlike in the U.S., constituencies in which candidates are barely elected make up over ninety percent of the constituencies in India. As a set, the marginal constituencies are more than large enough to change the balance of power in parliament, making an understanding of their dynamics important in their own right. They are, however, also very similar to constituencies in which candidates did not win by narrow margins suggesting that the marginal constituencies can shed light on the size of the incumbency effect in general.

The estimate from the regression discontinuity design yields a very surprising result – incumbents in India face an enormous disadvantage. Compared to similar non-incumbents, incumbents are fourteen percent less likely to be elected to office.[1] To make up this gap an incumbent would have to have won the previous election by an additional five and a half percent of the popular vote, a change equivalent to an incumbent moving from the first to the thirty-fifth percentile of elected officials ranked by margin of victory. For comparison, I also estimate the effects of incumbency before 1991. While the interpretation of the results are not as clear, the

[1] To avoid confusion, I will not report differences as relative changes unless explicitly stated. As a result, all estimates of the difference will reflect the raw change in the probability of winning resulting from holding office prior to entering the election.

data suggest that prior to the last fifteen years, incumbents enjoyed advantages of six to eleven percent.

While there is insufficient data to identify clearly the cause of the change, the shift in the effects of incumbency seem to be related to the Indian National Congress party losing its dominance of national elections. The largest drops in the effects of incumbency are concentrated in constituencies that have either an incumbent affiliated with Congress or elected a Congress candidates in at least half of the elections prior to 1991. Prior to 1991, experienced incumbents are also more likely to be reelected than inexperienced incumbents and experienced incumbents that belong to Congress are the most likely to be reelected. Starting in 1991, however, being an incumbent is almost equally disadvantageous for all candidates, with congress candidates slightly more disadvantaged. It is thus possible that the expectation of continued control of the parliament by a single party meant that experience working with the party in the past would be valuable in the party's subsequent parliamentary term. Incumbents were thus more valuable to voters than non-incumbents by virtue of their past experience working in a parliament controlled by the Congress party. After the fall of Congress in 1989, no party could be expected to hold power after a given election, limiting the value of having learned to work within the previous political regime and reducing the value of being an incumbent.

The remainder of this paper is organized as follows. The next section describes the empirical strategy and provides a simple structural model of the electoral process. Section III provides a brief description of the electoral process and history of Indian national parliamentary elections. Section IV describes the available data, and Section V discusses the importance of understanding the performance of incumbents in marginal elections in India. I present the empirical results in Section VI. Finally, I conclude in Section VII.

1.2 Empirical Estimation Strategy

In this paper, I intend to estimate the degree to which being an incumbent before standing for election affects a candidate's probability of election. This requires the estimation of the difference in performance of a candidate in two states of the world: one in which the candidate is an incumbent and one in which the candidate is not an incumbent. Since I can not observe a candidate as both an incumbent and a non-incumbent in the same election, I instead resort to comparing the performance of candidates that are identical except for the fact that one holds office immediately prior to the election. The identification problem in this context is to develop a method that ensures that compared candidates differ only in their incumbency status.

Unlike researchers who focus on the effects of incumbency in the U.S., I focus on the impact of incumbency on the probability of winning rather than the subsequent vote share. There are several reasons for this difference. First, the meaning of the vote share is not well defined in elections with varying numbers of candidates. Five percent of the popular vote, for example, may be meaningless in an election between two candidates, but might be a large margin in a first-past-the-post election with four candidates or more. Second, when considering models of incumbent behavior, only the probability of winning influences the candidates. Vote shares are important to the extent that they reflect these underlying probabilities, but it is the underlying probability of reelection that ultimately matters. It is possible, for example, for the margins of victory of two candidates to differ while their underlying probability of re-election remains the same (Jacobson, 1987) despite the fact that the two statistics are strongly correlated.

The first problem in estimating the incumbency effects is to identify a pool of eligible candidates. The candidates contesting the election of interest are an obvious choice, but since there is often little information available about them, it is difficult to control for other differences that may account for the differential performance of incumbents and non-incumbents. A common solution is to use the pool of candidates in the election immediately preceding the election of interest and to use the results of this first election as a means of controlling for differences between the candidates in the subsequent election. This methodology can then be refined by restricting the sample to specific types of elections or by supplementing with additional information.

1.2.1 A Simple Model of Elections

The problems involved with this estimation strategy can be illustrated with a simple structural model. First, I will restrict attention to a single constituency with N voters. I will refer to the election of interest as election t and the preceding election as election $t-1$. For simplicity, I will assume that C candidates contest in election $t-1$ and that the same candidates contest in election t.[2] I will assume that voters vote for the candidate that in expectation will generate the highest value return for them, and to allow for the possibility of strategic voting and other possible correlations of values across voters, I will not restrict the initial distribution of these values.

Let $\overline{VS}_j^{t'}$ be the actual vote share of candidate j in election t'. If V_j^i is the initial valuation of candidate j by voter i, then $\overline{VS}_j^{t-1}$ can be calculated according to the following equation:

$$\overline{VS}_j^{t-1} = \frac{1}{N}\sum_{i=1}^{N} 1(V_j^i \geq V_{(C)}^i) \quad (1)$$

where $V_{(C)}^i$ is the valuation of the highest valued candidate for voter i, and $1(\cdot)$ is an indicator variable set to one if the given statement is true and zero if false. The vote share of candidate j is simply the fraction of voters that value voting for candidate j the most. The "true" vote share, however, will differ from the vote share observed in the election because random factors (i.e. the placement of polling booths, traffic problems, etc.) that randomly affect voters voting for a particular candidate. I will thus assume that the observed vote share received by candidate j in election t' is given by $VS_j^{t'} = \overline{VS}_j^{t'} + \varepsilon_j^{t'}$ where $\varepsilon_j^{t'}$ is a random variable with a zero mean.

To identify candidates, I will rank them by their observed vote share in election $t-1$ with candidate C being the candidate with the largest actual vote share and candidate one having the lowest. In a first-past-the-post election, candidate C will win the election $t-1$ and will be the incumbent entering election t. I then define the margin of victory for candidate j, MV_j^{t-1}, as follows. For the winning candidate, let $MV_C^{t-1} = VS_C^{t-1} - VS_{C-1}^{t-1}$. For every other candidate, $j < C$, let $MV_j^{t-1} = VS_j^{t-1} - VS_C^{t-1}$. By definition, the margin of victory for the winning candidate is non-negative and the margin of victory for losing candidates is non-positive.

At election t, voters' valuations of the candidates will change and the change in valuation will change the underlying vote share. For expositional purposes, I will assume that

[2] The primary disadvantage of identifying candidates through the prior election is that many of them will not contest the election of interest. This is not a problem if the probability of rerunning is unrelated to whether or not the candidate wins the first election, but this may not always hold true. This is not a problem over the last fifteen years in India as I will discuss in Section VI.

the change in vote share $CVS_j^t = \overline{VS}_j^t - \overline{VS}_j^{t-1}$ can be partitioned into an additively separable function given by the following equation:

$$CVS_j = CVS_j(e) + CVS(MV_j^{t-1})$$

where e is the change in experience of the incumbent elected in election $t-1$. The point of this partition is to highlight that there are two paths through which a candidate's underlying vote share can change. First, voter's valuation of the candidate can change as a result of the candidate holding office for a term. Second, however, the vote share could change for reasons not related to the candidate's experience, and these effects are likely to differ depending on the vote share in the previous election. If an incumbent wins by a large margin, for example, opposition parties may decide to reduce the amount of campaign money they spend to support their candidates. The probability of the incumbent winning election t is then given by

$$\Pr[MV_C^{t-1} + (\overline{VS}_j^{t-1} - \overline{VS}_{C-1}^{t-1}) + \Delta CVS_j^C(e) + \Delta CVS(MV_C^{t-1}, MV_j^{t-1}) > \Delta\varepsilon_j^C \forall j < C] \tag{2}$$

and the probability of any other candidate $j < C$ winning the election is given by

$$\Pr[MV_j^{t-1} + (\overline{VS}_{j'}^{t-1} - \overline{VS}_C^{t-1}) + \Delta CVS_{j'}^j(e) + \Delta CVS(MV_j^{t-1}, MV_{j'}^{t-1}) > \Delta\varepsilon_{j'}^j \forall j < C] \tag{3}$$

where $\Delta CVS_{j''}^{j'}(e) = CVS_{j'}(e) - CVS_{j''}(e)$ and

$\Delta CVS(MV_j^{t-1}, MV_{j'}^{t-1}) = CVS(MV_j^{t-1}) - CVS(MV_{j'}^{t-1})$. These equations describe a candidate's probability of election by four terms: MV_j^{t-1} represents the relative strength of candidate j, $(\overline{VS}_{j'}^{t-1} - \overline{VS}_C^{t-1})$ gives the relative strength of other candidates, $\Delta CVS_{j'}^j(e)$ is the change in voter share due to experience, and $\Delta CVS(MV_j^{t-1}, MV_{j'}^{t-1})$ is the change in vote share due to factors other than experience. These equations also illustrate the challenge of comparing incumbents and non-incumbents to estimate the effects of incumbency. Within this framework, the effects of either being an incumbent or not being an incumbent are given by $\Delta CVS_j^C(e)$ and $\Delta CVS_{j'}^j(e)$. It is clear from these equations, however, that differences in the probability of reelection are driven both by the effects of incumbency and by pre-existing vote share differences and other factors that could change the vote share a candidate receives.

Most researchers seek to hold these effects constant by controlling for differences in observable characteristics or by taking advantage of natural experiments that limit the degree to which incumbents can differ from non-incumbents. Levitt and Wolfram (1997) for example, focus in part on districts in which an incumbent won an open seat in the first election and correct for attrition bias through a structural assumption regarding the functional form of the election process. Ansolabehere and Snyder (2002) perform a similar analysis but focus on districts in which the incumbent prior to the first election was forced from office due to term limits in order to control for the endogeneity of an incumbent's decision about when to contest an election. Because of the function form assumptions, however, these studies are limited in the degree to which they can control for a candidate's quality or the quality of a candidate's competitors. For example, if incumbents are at an advantage (disadvantage) in an election relative to non-incumbents, then parties may place more (less) resources or nominate better (worse) candidates in elections without incumbents, biasing estimates of the incumbency advantage upwards. Similarly, candidates themselves differ in their strength. If a candidate's probability of being elected can be estimated ex ante (perhaps because of matches between the candidate and the characteristics of the district or because of a party's historical strength in a geographic area) and opposition parties then field weaker candidates, candidates may be re-elected based upon their strength as a candidate rather than any effects associated with holding office prior to an election.

1.2.2 Estimating the Effects of Being an Incumbent

As an alternative, I employ a regression discontinuity methodology. The general framework is provided by Hahn, Todd, and Van der Klaauw (2001), and the application to elections is suggested by Lee (2001)[3] This approach is based upon the insight that while two candidates with the same margin must also have the same incumbency status, it is possible to compare candidates that have a different status as incumbents but have arbitrarily close margins of victory by comparing incumbents who are barely elected to non-incumbents that barely lose. In the context of the previous model, this amounts to comparing the probability of reelection as MV_j^{t-1} approaches zero. The critical assumption of this methodology is that all of the other characteristics of candidates that could affect their probability of reelection on average vary continuously as a function of the margin of victory at zero. This allows incumbents and non-

[3] Miguel and Zaidi (2003) have recently performed a similar analysis in Ghana. While they focus on political patronage as well as the effects of incumbency, they find that on average incumbent and non-incumbents fair similarly.

incumbents to differ systematically, and requires only that on average those candidates that are barely elected are similar to those candidates that barely lose. The major limitation of this approach, however, is that it relies on marginal elections to estimate the incumbency effects and these elections may, of course, not be representative of all elections.

In the structural model, this would only require that $CVS(MV_j^{t-1})$ is continuous at $MV_j^{t-1} = 0$. To illustrate the effect of this assumption, consider the limits of equations (2) and (3) as MV_j^{t-1} approaches zero for some candidate j. The fact that MV_j^{t-1} approaches zero for some candidate $j < C$ implies that both MV_{C-1}^{t-1} and MV_C^{t-1} also approach zero.[4] At the limit, the vote shares of candidates j, $C-1$, and C are all equal in election $t-1$. I will denote the limit of the vote shares as VS^{t-1}. Finally to further simplify the exposition, I will assume without loss of generality that $C = 4$ and that MV_3^{t-1} converges to zero. In other words, I will assume that four candidates are contesting an election and consider the case where the voter shares of three candidates converge to a three-way tie.

The limit of (2) as the vote shares of the top three candidates converge is then given by the following value:

$$\Pr[\Delta CVS_3^4(e) > \Delta\varepsilon_3^4, \Delta CVS_2^4(e) > \Delta\varepsilon_2^4, (\overline{VS}_1^{t-1} - VS^{t-1}) + \Delta CVS_1^4(e) + \Delta CVS_1^4(0, MV_1^{t-1}) > \Delta\varepsilon_1^4] \tag{4}$$

Similarly, the value of (3) can be rewritten in the following form for $j, j' = 2,3$, $j \neq j'$:

$$\Pr[\Delta CVS_4^j(e) > \Delta\varepsilon_4^j, \Delta CVS_{j'}^j(e) > \Delta\varepsilon_{j'}^j, (\overline{VS}_4^{t-1} - VS^{t-1}) + \Delta CVS_j^4(e) + \Delta CVS_j^4(0, MV_1^{t-1}) > \Delta\varepsilon_1^j] \tag{5}$$

The assumption that $CVS(MV_j^{t-1})$ is continuous at zero then guarantees that as the vote shares converge $\Delta CVS_1^4(0, MV_1^{t-1}) = \Delta CVS_j^4(0, MV_1^{t-1})$ where $j = 2,3$. This leaves the changes due the additional experience of the incumbent as the only differences between the probability of reelection for the incumbent and the second and third ranked candidates. The other differences that biased the comparison converge to zero with the margin of victory. Since the only difference between the candidates is that one of them is an incumbent, this allows me to infer the causal effect of incumbency on the probability of winning election t for candidates that barely win or lose an election. Relying on these marginal candidates, however, is the main shortcoming

[4] If $MV_{C-1}^{t-1} \to 0$, then $MV_C^{t-1} \to 0$ since $MV_{C-1}^{t-1} = -MV_C^{t-1}$. So, if $j = C-1$, the result follows directly. If $j < C-1$, $MV_{C-1}^{t-1} \to 0$ follows from the fact that $MV_{C-1}^{t-1} < 0$ and $MV_j^{t-1} < MV_{C-1}^{t-1}$.

of the methodology because the effects of incumbency on candidates that win or lose by larger margins need not be the same. These marginal elections alone may be interesting, as they are in Indian parliamentary elections, but extrapolating the estimates to non-marginal candidates and constituencies requires the assumption that these marginal elections are not systematically different. In Section V, I demonstrate that in the current context, marginal and non-marginal constituencies are in fact similar.

To estimate the effect of being an incumbent on the probability of winning an election, I estimate the relationship between the probability of winning election t and the margin of victory in election $t-1$ separately for incumbents and non-incumbents and then estimate the size of the discontinuity that results for a margin of victory of zero. Formally, I estimate the following relationship:

$$\mathrm{Win}_{j,k}^{t} = \mathrm{f}_1(\mathrm{MV}_{j,k}^{t-1} \mid \mathrm{MV}_{j,k}^{t-1} > 0) + \mathrm{f}_2(\mathrm{MV}_{j,k}^{t-1} \mid \mathrm{MV}_{j,k}^{t-1} < 0) + \varepsilon_{j,k}^{t-1} \tag{6}$$

where $\mathrm{Win}_{j,k}^{t}$ is an indicator variable for whether or not candidate j, in constituency k won the election in period t. The effect of being an incumbent is then measure by taking the following difference:

$$\lim_{MV \to 0^+} \hat{\mathrm{f}}_1(\mathrm{MV} \mid \mathrm{MV}>0) - \lim_{MV \to 0^-} \hat{\mathrm{f}}_2(\mathrm{MV} \mid \mathrm{MV}<0) \tag{7}$$

where $\hat{\mathrm{f}}_1$ and $\hat{\mathrm{f}}_2$ are estimates of f_1 and f_2 respectively.

1.2.3 Incumbency Advantage or Disadvantage?

Empirically in the United States, incumbents do fare better than non-incumbents, and as a result, a number of explanations have been offered to explain such an advantage. These explanations range from models of asymmetric information (Banks and Sundaram, 1993; Austen-Smith and Banks, 1989; Besley and Case, 1995; Rogoff, 1990) to the beneficial aspects of office entitlements such as franking privileges, the ability to shape redistricting plans, seniority privileges, and the potential for increased name recognition (Alford and Hibing; 1989). Nothing in these models, however, guarantees an advantage, and many of them could just as easily generate a disadvantage with minor changes in the underlying assumptions.

Returning to the structural model from above, recall that voter i has an initial valuation of candidate j that I am denoting V_j^i. The true vote share for candidate j is then the fraction of voters who receive the highest expected value by voting for candidate j and is given by equation (1). The general assumption behind the incumbency advantage is that after spending a term in office the expected value of the incumbent, candidate C, increases. Let $V(e) = S(e)$ be the value of the incumbent due to experience so that the value of the incumbent for an arbitrary voter is given by $V_C^i + V(e)$. Let e be the number of terms the incumbent has held office. The basic assumption in the models that generate an incumbency advantage is that $S'(e) > 0$ so that after spending a term in office voters expect that voting for the candidate is more valuable than they expected before election $t-1$. This change in expectation could result from any of the explanations listed above. The incumbent may have more experience allowing her to work more effectively within the established political system, she may have gotten access to resources that allow her to improve or expand her public image, or by being in office voters may have learned more about her abilities.

The change in vote share of each candidate due to the increased experience is then given by the following equations:

$$CVS_C(e) = \frac{1}{N}\sum_{i=1}^{N} 1(V_C^i + V(e+1) \geq V_{(C)}^i \,\&\, V_C^i + V(e) < V_{(C)}^i) \tag{8}$$

$$CVS_j(e) = -\frac{1}{N}\sum_{i=1}^{N} 1(V_C^i + V(e+1) \geq V_{(C)}^i \,\&\, V_j^i = V_{(C)}^i), j < C \tag{9}$$

The incumbent gains votes on average from voters for whom the increase value of voting for the incumbent is enough that they gain more by voting for the incumbent than their previously highest ranked candidate (C). Similarly, candidates that were previously the most valuable vote for those voters proportionally lose votes. The higher vote shares mean that on average the incumbent also receives a higher measured vote share compared with other candidates and the increased expected value translates into a higher probability of winning than non-incumbents. For marginal candidates where the probability of victory is given by (4) and (5), this change is the only difference between the candidates.

The assumption, however, that the expected value of a candidate increases after a period in office is questionable. While incumbents certainly develop experience working within the existing political system they may use that experience to pursue activities that are not in the best interests of voters like pet causes or project or even corruption. To account for these divergent effects, consider a different specification of voters' valuation of experience, $\tilde{V}(e)=(1-G(e))(\tilde{S}(e))$ where the increased service provided to voters is given by $\tilde{S}(e)$ and the negative effect, which I will refer to as the graft effect, is given by $G(e)$. As before, I assume that candidates become more adept at serving voters as they gain experience ($\tilde{S}'(e)>0$). I also, however, assume that the graft effect becomes larger with experience ($G'(e)>0$) as candidates get more experienced influencing the civil system and develop the ability to pursue their own agenda. Simple differentiation yields the following relationship between a change in experience and the average expected value of voters:

$$\tilde{V}'(e)=(1-G(e))\tilde{S}'(e)-G'(e)\tilde{S}(e) \tag{10}$$

If $\tilde{V}'(e)>0$, then the valuation of the incumbent follows the process laid out in equations (8) and (9). If $\tilde{V}'(e)<0$, the process runs in the opposite direction with voters that previously valued voting for the incumbent the most switching to other candidates, making incumbents less likely to win election t.

In what follows, incumbents switch from being at an advantage before the 1991 election and then fall to a disadvantage starting with the 1991 election. Equation (10) suggests that this can occur in two ways. First, the levels of provided service and graft could change even if the marginal ability to engage in the activities due to increased experience remains the same. Second, the relative marginal returns to experience could change. Specifically, either candidates might become relatively more effective at diverting resources to personal gain or they may become less effective at serving their constituents.

While the data is insufficient to identify the actual cause of the change, I will argue that the data seems consistent with a model in which Congress' dominance provided a relatively stable structure in which because Congress could always be expected to control parliament, voters could expect that experience in the previous term would prove valuable systemic experience in the subsequent term. With the end of Congress's dominance in 1989, control of parliament switched between parties making experience under the previous regime less valuable.

This change in the political structure then resulted in a system in which as politicians gained more experience and influence they become more likely to pursue activities that are not in the best interest of voters.

1.3 Indian Political System and Lok Sabha Elections

The current political institutions took shape shortly after India gained independence from Great Britain in 1947. The Indian form of government is very similar to both the US and British systems. Like the US, India has a federal system with thirty-two semi-autonomous states and union territories joined under a single central government. The legislature comprises two houses of parliament: an upper house (Rajya Sabha) and a lower house (Lok Sabha) which is the focus of this study. All members of the upper house are elected by state legislatures except for twelve members appointed by the president by virtue of their expertise in particular subjects like art, history, science, etc. Except for two seats, appointed by the president to represent the Anglo-Indian community (should he or she feel that the community is not adequately represented by the other elected members), all members of the lower house are directly elected by all citizens over the age of eighteen. A fraction of these seats can only be contested by members of historically disadvantaged groups designated as either scheduled castes or scheduled tribes.[5] The number of seats in both houses allotted to each state is proportional to the population of the state. There are currently 545 members of the Lok Sabha elected from 542 constituencies and 235 members of the Rajya Sabha. The president, elected by members of parliament and state officials, serves as the head of state while the Prime Minister, appointed by the president from the party or coalition enjoying majority support in the Lok Sabha, implements the directives of the legislature and oversees the day-to-day functioning of the government.

Constituencies for the lower house are drawn by an independent, legislatively created Delimitation Commission. The original intent was to redraw the boundaries after each decennial census to ensure that each constituency had roughly the same population. In practice, constituencies have been redefined twice since they were originally set for the 1951-1952 election: once in 1963 and again in 1973. Under a 1976 constitutional amendment, constituency

[5] The goal of this policy is to counter caste based discrimination and the historically disadvantaged socio-economic positions of these groups.

boundaries were frozen until after the 2001 census to ensure that states' family planning efforts did not jeopardize their representation in the national legislature.

Elections are organized by a constitutionally empowered organization called the Election Commission of India. The commission is staffed by three members appointed by the president, and has the responsibility of overseeing the election of all state and national parliaments and elections to the offices of the president and vice-president. Shortly before the target date for an election, the Commission formally announces the up-coming elections at which time special rules governing the behavior of political parties and candidates come into effect. Candidates are then nominated within each constituency, at least two weeks are provided for campaigning, and finally the polling begins. Results for a given constituency are usually announced a day after polling. All races are first-past-the-post with the candidate receiving the largest percentage of eligible votes winning regardless of the proportion of votes received. National elections are an enormous undertaking. Due the size of the country, national elections occur over at least three days. There are currently over 600 million eligible voters that vote in 800,000 polling stations. In 1996 general election, the commission employed almost 4 million people, counted over 2.5 million ballot boxes, and spent the equivalent of over $US 100 million (ECI, 2003).

In practice, national parliamentary elections were dominated by the Congress party who, except for a brief three year period starting in 1977,[6] held power by wide margins until the 1989 election. Before the 1989 election, a large scandal broke, in which officials received kickbacks for the purchase of military hardware. The new scandal reopened an old wound. Congress had always been beset by corruption charges, and Sanjay Gandhi, the older brother of the Prime Minister Rajiv Gandhi, had been notorious for profiting at the public's expense. Possibly due to these revelations, Congress lost in 1989 for the first time to an un-united opposition, signaling the potential for a multi-party democracy in India. To make matters worse, charges of vote rigging emerged in Rajiv Gandhi's district of Amethi. As one Indian scholar notes, "The November 1989 General Elections in India put an end to what was widely accepted as the unavoidable and permanent monopoly of power by a single party in power which ruled the

[6] In 1975, the Prime Minister Indira Gandhi imposed martial law and initiated a period known as the Emergency. During the next twenty-one months, the government initiated a policy of clearing slum areas, relocating their inhabitants, and forcing many poor individuals to undergo sterilization procedures. Partly as a result of alienating the poor, a core constituency, and also a result of the opposition parties merging to form a single united party, Congress lost the first post-Emergency election in 1977. While there is considerable debate about the sources of Congress support in the next two elections, Congress won both of them by very wide margins. Until the 1989 elections, the loss in 1977 was seen by many as a possible anomaly.

country for 42 years, except for a brief interlude, when the Janata Government was at the Centre (Ahuja, 1992)." Since 1989, no party, including congress, has ever again won an outright majority in a national election.[7]

After the 1989 elections, a number of political trends emerged. First, the Congress party, while still powerful, generally fared as well as other major parties over the period in stark contrast to its previous success (Figure 1). Second, the political environment became much more competitive. As Figure 2 shows, the average number of candidates in each district increased to the point where the Election Commission felt it necessary to increase the deposits and nomination requirements in 1996.[8] The number of nationally competitive political parties also increased significantly (Figure 3) as previously state and local parties vied for a larger role in national politics. Starting in 1989 control of parliament began to shift between parties and coalitions and no one coalition or party could expect to win. Finally, while always significant in India, communal conflicts became increasingly prominent in national politics. Hindu-Muslim conflicts have sparked many large scale riots over a mosque that is reputed to have been built centuries ago by a Moghal invader on the ruins of Hindu temple in Ayodhya and propelled a number of religiously oriented parties onto the national political scene. The once ardently (and now somewhat moderated) Hindu nationalist party, the Bharatiya Janata Party (BJP) has even managed to form a coalition government after the national elections in 1998 and 1999.

1.4 Description of the Data

In addition to holding all elections in India, the Election Commission of India also releases comprehensive reports on the outcome of every state and national election. From the ECI, I was able to obtain the results for all thirteen national parliamentary elections (1951, 1957, 1962, 1967, 1971, 1977, 1980, 1984, 1989, 1991, 1996, 1998, and 1999). For each candidate in each constituency, the ECI releases the individual's party, the number of votes the candidate receives, and the candidate's gender. For each constituency, the ECI also releases the number of candidates that were nominated, rejected, and who eventually contested along with the number

[7] Though in the Tenth Lok Sabha they would build an outright majority for the last time through bye-elections and other members joining the Congress party.
[8] I am currently investigating the role that increase political competition played in the change in the effects of incumbency status.

that officially withdrew and the number that forfeited their deposit. They also report the number of polls, number of eligible voters, the number of actual votes, and the number of votes rejected.

While comprehensive, these data do have a number of short comings. Most unfortunate is the fact that the ECI does not consistently record the names of candidates. Variation occurs along four primary dimensions. First, for a given name, the spelling can vary. Some of the discrepancies are clearly the result of the ambiguity resulting from the transliteration of names into English. However, there are also a number of spelling variations that are just mistakes. Second, there is variation in which names are reported. In some elections, the ECI may list two names and in another three or four. Third, names that are reported separately in one election can be combined into an individual name in another. Finally, descriptors like the name of a candidate's relation, the candidate's profession, and honorifics are also inconsistently recorded.

The methodology that I employ requires me to track each candidate's political career through each election, and the fact that names are inconsistently recorded complicates this. My solution is a simple algorithm that allows me to mechanically match candidates over time within a given constituency. I developed this algorithm by manually matching the names of candidates in five constituencies, generalizing the characteristics of the differences in the way that names were recorded, and applying this methodology to all candidates. Specifically, I first match all names as given in the data across all thirteen elections. I then iteratively relax the definition of a match by allowing for, in order, omitted or mis-ordered names, in-accurate divisions between names, and finally spelling differences of a character or less. Because there are over 40,000 names in the data set, the potential for mis-matched names is large. To control for this, I match only within a given constituency over time which is possible because the definitions of the constituencies were frozen by parliamentary mandate after the Emergency.[9] I can thus use the results from any constituency for which I have regularly reported data for each election from 1977 to 1999.

In addition to the name issues, however, there are a few inconsistencies in the timing of elections as well as the reporting of results from some constituencies. First, the ECI has not consistently reported the occurrence or outcomes of bye elections.[10] As a result, I am forced to ignore these elections. Second, even regular elections in some areas have not been consistently

[9] It is, of course, possible that candidates will change constituencies over time, and I check for bias resulting from the effect below.
[10] These are elections held to fill a position vacated prior the subsequent national election.

held or reported. Elections in the states of Jammu and Kashmir, Punjab, and Assam have at one time either been canceled or delayed because of communal unrest. Additionally, the following districts do not report results for each year after the emergency: Purnea and Patna in the state of Bihar, Mizoram in the state of Mizoram, Meerut in Uttar Pradesh, Shillong in Meghalaya, and Daman and Diu in the state of Daman and Diu. Eliminating the districts with inconsistent reports for regular elections then leaves me with a data set reflecting 504 of the 543 constituencies in India.

1.5 Representativeness of Marginal Elections

As mentioned in Section II, the major limitation of the regression discontinuity approach is that it relies on candidates that barely win or lose elections to estimate the incumbency effects for candidates in general. Results from these marginal elections are potentially valuable for two reasons. First, if marginal constituencies are common enough to determine the control of parliament, then understanding the effects of incumbency in these constituencies is important in its own right. Second, however, if marginal constituencies do not differ significantly from non-marginal constituencies, then it is reasonable to assume that the measured incumbency effects are applicable to all constituencies.

In this study, both of these justifications support the importance of the regression discontinuity results. In India, constituencies in which the decision is marginal could easily determine which party or coalition controls the parliament. Starting in 1991, over half of the constituencies in an average election have a margin of victory of less than ten percent. A party's ability to win these constituencies alone would be enough to control the Lok Sabha. Through 1989, the outcomes of twenty-nine percent of constituencies were determined by a margin of less than ten percent. While smaller this is still a significant percentage of constituencies and could make any party a formidable presence in the parliament.

Additionally, however, marginal constituencies are very similar to non-marginal constituencies. Table 1 shows the descriptive characteristics of marginal and non-marginal constituencies. The first two columns display the average characteristics for the period starting in 1991. On average these constituencies are very similar, and do not support the contention that marginal and non-marginal constituencies differ systematically. This should not be surprising, however, since over ninety percent of constituencies at some point experience an election where the margin of victory is less than ten percent. The third and fourth columns of Table 1 list the differences for the period immediately after the Emergency.[11] The results are generally the same. The lack of any systematic difference between marginal and non-marginal constituencies suggests that the results for marginal constituencies are likely to reflect the effects of incumbency in general.

1.6 Estimation of Incumbency Effects

1.6.1 Estimation Method

Figure 4 graphically presents the central result of this paper – a graph of the relationship between the probability of re-election and the margin of victory in the previous election (described by equation (6)).[12] The figure depicts a local polynomial estimation of the probability of winning a given election (y-axis) as a function of the margin of victory in the previous election (x-axis). The black dots represent the average probability of election for all candidates that fall within a two and a half percent interval centered at the point on the x-axis at which the dot is located. The size of each dot is directly proportional to the number of observations that fall in the respective interval. Generally, as one would expect, the estimated function fits the average probabilities of winning quite well, and the probability of winning increases in the margin of victory – except for candidates with a zero margin of victory. At zero, the right and left hand limits are visibly different suggesting that incumbents are about fourteen percent less likely to win than their non-incumbent competitors.

Figure 5 depicts the relationship using the entire support, highlighting the importance of the regression discontinuity design. A simple comparison of the reelection rates of all candidates, listed in Table 2, suggests that incumbents are at a thirty-seven percent advantage compared to non-incumbents. As explained previously, however, this sample includes both very strong incumbents that derive their power from sources other than their office and conversely,

[11] In comparison to the period starting in 1991, candidates before 1991 seem to have relatively less experience. This result, however, is probably an artifact of my inability to measure candidate's experience before the emergency. As a result, all candidates' experience is measured starting in 1977.
[12] Candidates that run in the first election but not the second are recorded as having lost the second election.

very poor candidates who lose by very large margins and have little chance of building a vote base. The first three columns in Table 3 illustrate the differences between these candidates. On average, incumbents are more likely to be female, have substantially more political experience, and are more likely to belong to one of the major parties. In addition to the difference in incumbency status, the difference in the reelection rates thus likely reflects other differences between the candidates.

Restricting the comparison to only candidates that win or lose by less than a fifteen percent margin eliminates the poorest and strongest candidates, reducing the measured advantage to twenty-two percent. Reducing the sample further, however, to those winning or losing by less than two and a half percent, yields an incumbency disadvantage of five percent. If the continuity assumption holds, then these candidates will differ less along both observable and unobservable characteristics, making their relative performance more indicative of the effects of incumbency in general. The last three columns in Table 3 confirm this conjecture for characteristics observable in the data. The differences between incumbents and non-incumbents in this range are quite small. Compared to non-incumbents, incumbents, for example, are six tenths of a percent less likely to be female, have marginally more political experience, and are almost equally likely to belong to a major party.

While these comparisons of candidates that barely won and loss are suggestive, the comparison ignores both the slope of the data near the discontinuity and the information available in the rest of the data set. A more precise comparison can be made by estimating equation (6) and comparing the left and right hand limits directly. In most cases, the discontinuity can be estimated simply by specifying a parametric relationship between the variables of interest and including an indicator variable for whether or not a candidate was an incumbent. This approach, however, is problematic in the context of Indian parliamentary elections. As Figure 5 shows, a large number of candidates run for office, perform poorly, and have very little hope of winning the subsequent election. This mass of candidates, far from the discontinuity, exerts tremendous influence over the estimate of the left hand limit of the function at zero. As a result, the estimates are very sensitive to the order of the polynomial used in the specification, making it difficult to estimate the endpoints without additional information regarding the true functional form. The same is true for estimations using probit and logit models.

Given the uncertainty regarding the functional form, the natural solution to this problem is to use a semi-parametric estimate for the relationship which would allow more flexibility in the specification of the functional form. In regression discontinuity context, however, the discontinuity itself becomes a problem since smoothness is one of the underlying characteristics usually assumed by semi-parametric estimators. One way to resolve this problem is to estimate the semi-parametric relationship on either side of the discontinuity and take the difference of the conditional expectations at the discontinuity point itself. Unfortunately, many estimators are biased at the boundaries requiring the use of smaller bandwidths that slow the rate of convergence of the point estimates.

Following the work of Hahn, Tadd, and Van der Klaauw (2001) who suggest avoiding these biases by using local polynomial regression techniques developed by Fan (1992; Fan and Gijbels 1996), Porter (2002) develops a general estimator that achieves the optimal rate of convergence in the non-parametric framework. Using the local polynomial regression estimator, I estimate the following function:

$$\mathrm{Win}^{t}_{j,k} = f_1(\mathrm{MV}^{t-1}_{j,k} \mid \mathrm{MV}^{t-1}_{j,k} > 0) + f_2(\mathrm{MV}^{t-1}_{j,k} \mid \mathrm{MV}^{t-1}_{j,k} < 0) + \sigma(\mathrm{MV}^{t-1}_{j,k})\varepsilon$$

where ε is assumed to be independent and distributed according to the standard normal distribution. Within this framework, the value of f_1 or f_2 at x is estimated by finding β_0 where β_0 and β_1 minimize the following function at x:

$$\sum_{t,j,k} \{\mathrm{Win}^{t}_{j,k} - \beta_0 - \beta_1(\mathrm{MV}^{t-1}_{j,k} - x)\}^2 K_h(\mathrm{MV}^{t-1}_{j,k} - x)$$

where $K_h(z) = 0.75(1-z^2)$ is the Epanechnikov Kernel.[13] I select a uniform bandwidth according to the following two-step process. First, I make a preliminary estimate of the bandwidth using a cross-validation estimator. Second, using this initial bandwidth estimate, I then estimate the mean integrated square error of the estimated function and choose the bandwidth that minimizes the estimate. I can then estimate the discontinuity of the function at zero by subtracting the left hand limit at zero of the estimate of f_2 from the estimated right hand limit of f_1 at zero as in equation (7).

[13] Unlike other kernel estimators, the Epanechnikov kernel is always the optimal choice for Local Polynomial Regression Estimator.

The bandwidth estimate chosen by this selection algorithm then converges to zero fast enough as the sample size grows that the distribution of the estimated discontinuity converges to the following distribution:

$$\sqrt{nh}(\hat{\alpha}-\alpha)\xrightarrow{d} N(0,\frac{\sigma^{2+}(0)+\sigma^{2-}(0)}{g_0(0)}e_1'\Gamma^{-1}\Delta\Gamma^{-1}e_1)$$

where α is the true size of the discontinuity, $\hat{\alpha}$ is the estimated discontinuity, $g_0(0)$ is the density of $MV_{j,k}^{t-1}$ evaluated at zero, and $\sigma^{2+}(0)$ and $\sigma^{2-}(0)$ are the right and left hand limits of the variance estimate.[14] The terms e_1, Γ, and Δ are defined as $e_1=(1,0)'$, $\Gamma=\begin{bmatrix}\gamma_0 & \gamma_1\\ \gamma_1 & \gamma_2\end{bmatrix}$, and $\Delta=\begin{bmatrix}\delta_0 & \delta_1\\ \delta_1 & \delta_2\end{bmatrix}$ where $\gamma_k=\int_0^\infty K_h(u)u^k du$ and $\delta_k=\int_0^\infty K_h^2(u)u^k du$.

Compared with other estimation techniques, this process has two potential problems. First, as with most non-parametric estimators, my estimates are not unbiased, but any reduction in bias would come at the cost of an increase in the variance of the estimate. This is further complicated by the fact that the discontinuity in question occurs at an inflection point of the function. Since the function is concave to the left of zero and convex to the right, this bias should cause me to under-estimate slightly the size of the discontinuity. Second, when estimating the variance of the error terms, I control for nothing more than heteroskedasticity as a function of the margin of victory. This is problematic since a misspecification of the correlation of error across treated and untreated observations can bias the estimated standard deviation of the point estimates.

To gauge the magnitude of these potential problems, I also estimate the incumbency advantage using a spline estimator with knots at -45, -35, -25, -15, -5, 0, 5, 10, 15, 25, 35, and 45 and including both year and constituency fixed effects. All of the estimates are clustered at the state level to avoid making strong assumptions about the form of variance/covariance matrix of the error terms and to allow for the possible non-independence of error terms. Since the spline estimator is more efficient and generates results consistent with the local polynomial regression, I use the spline estimator to estimate the ancillary results which are often based upon subsets of

[14] I estimate the variance of the disturbance term at the right and left hand limits (i.e. allowing for heteroskedasticity conditioned on the margin of victory) using the density weighted process outlined in both Porter (2002) and Hardle (1990).

the data. Finally, for reference, I include two additional estimates: a simple linear fit of the data that lie within a ten percent interval around zero and a quadratic polynomial fit using all of the data. The linear fit summarizes the data directly adjacent to the point of discontinuity, but is inefficient because it ignores the remaining data.

1.6.2 Incumbency Effects: 1991 – Present

The first six rows in Table 4 exhibit the results of these estimations, which confirm the graphical depiction of the discontinuity in Figure 4. These results are not conditioned on a candidate contesting the current election and candidates that fail to contest are assumed to have lost. The first column is the right hand limit of the function in Figure 4 or the probability of a candidate that was barely elected in the first period winning in the subsequent period. The second column contains the probability of victory for candidates that barely lost the first election. Finally, the third column is the difference between these values. As usual, the standard deviation of each of the point estimates is in parenthesis below the estimate. The bias in the non-parametric estimates due to the inflection point does not seem to be a problem since all of the estimates are similar. On average, being an incumbent makes a candidate about fourteen percent less likely than a non-incumbent to win an election. The difference is also statistically significant under each treatment of the error terms, though the clustered standard errors are much larger than those of the non-parametric estimate (5.98 versus 1.92).

As explained in Section II, the ability to interpret this difference as a causal relationship depends critically on the assumption that all other characteristics that affect a candidate's probability of being reelected vary continuously as a function of the previous margin of victory at zero (i.e. that $CVS(MV_j^{t-1})$ is continuous at $MV_j^{t-1} = 0$). While it is impossible to completely verify the validity of this assumption, I can verify it for candidate characteristics observable in my data set. The seventh through eleventh rows of Table 4 show this comparison using the spline estimator for four characteristics: membership in the Congress party, number of pervious electoral victories since 1977, number of years experience as a member of the Lok Sabha since 1977, and number of elections contested since 1977. All of the estimated differences are statistically and practically insignificant, suggesting that, at least at the margin, the candidates are comparable.

The remaining rows of Table 4 decompose the source of the incumbency effect into the probability of running in the second election and the probability of winning conditional on running. Row twelve shows the estimated probability of running in the subsequent election for marginal incumbents and non-incumbents. In previous studies that find advantages to incumbency status, incumbents are much more likely than non-incumbents to run for reelection. This can bias the estimates of the incumbency effects if a systematic relationship exists between a candidate's probability of running again and the candidate's probability of reelection. As row twelve illustrates, over the period in question, incumbents and non-incumbents are equally likely to run for re-election. This relationship is depicted in Figure 6 using the spline estimate of the function. As one would expect the probability of contesting in the subsequent election increases with one's margin of victory, but there is almost no difference in the probabilities of those candidates just barely elected or not elected.

This allows for a straightforward interpretation of row thirteen, which reports the probability of victory conditional on a candidate contesting the current election. Assuming that a candidate runs a second time, not holding office increases ones chances of reelection by almost twenty-eight percent. Both this and the unconditional estimate are enormous effects. In practice, it would take significant effort to make up such a disadvantage. On average, for example, an incumbent would have to win by an additional five and a half percent of the popular vote, a change equivalent to an incumbent moving from the first to the thirty-fifth percentile of elected officials ranked by margin of victory.

1.6.3 Robustness

Because I do not know with certainty which names correspond to the same candidate, I had to use the name matching algorithm explained in Section IV. This algorithm, however, can create biases in two different ways. First, if the equivalence relationships are too general, then matches between names will occur randomly. This effect will attenuate any measurable differences between incumbents and non-incumbents because both are treated in the same increasingly random fashion. To check for this, I estimated the effects of incumbency using the spline estimator for three different sets of equivalence relationships. The results of which are displayed in the first three rows of Table 5.

In the first row, I consider two names to match if and only if the exact string provided by the Election commission matches. In the second row, I also identify individual names within the given string, allow for the possibility that all names are not reported in each period, and also allow for the possibility that the names were mis-ordered (i.e. by being listed last name then first with the comma omitted). The third row presents the equivalence relationships used for the estimates in Table 4, and is the same as the relationships for row two with the exception that I allow the names to differ by one character. The results suggest that the relationships are in fact not too general since the estimated difference is the same for each set of equivalence relationships. Generalizing the equivalence relationships from full string matching to the matching of individual names increases the estimated probability that incumbents or non-incumbents will be elected, but it does not affect estimated differences in these probabilities. Allowing the single character deviation has almost no effect either on the estimation of the levels or the difference.

Another problem arises from the possibility that candidates may play upon name recognition. If competitors, for example, seek to confuse voters by nominating candidates with names similar to widely recognized or particularly strong candidates (for example, a particularly strong incumbent), the matching of incumbents' names may be more likely to generate an ambiguous match, making it more likely that my algorithm will match the incumbent to a weaker competitor that loses the subsequent election. An ambiguous match can occur in three ways. First, two candidates in the first election could be matched to the same candidate in the subsequent election. Second, a single candidate can be matched to two candidates in the subsequent elections. Because I am interested in a candidate's performance over time, however, it is also possible that allowing a match between a name in the first election and the subsequent election will create an inconsistency with another name already matched to one of the two names. In practice, if this effect exists, it is negligible. Row four of Table 5 displays the results of the model depicted by equation (6) but with probability of candidate's name being involved in an ambiguous match as the dependent variable. As the fourth row indicates, however, the probability of an ambiguous match is very low for marginal candidates and at most, this effect could change the estimated incumbency effects by 1.36 percent.

Finally, it is not uncommon for candidates in India to change constituencies. If this behavior is more likely for incumbents than non-incumbents, then because I match only within each constituency, I will be more likely to record an incumbent as having lost an election due to not rerunning when in fact, the candidate contested in another constituency. An easy solution would be to match candidates across all constituencies, but unfortunately, it becomes too likely at that point that a given match will prove ambiguous. Instead, I perform another matching of the candidates, matching within party rather than within constituency.[15] I can then record whether or not a match occurred within the same constituency or across constituencies, leaving open only the rarer possibility that a candidate changes both parties and contests in a different constituency. The fifth row of Table 5 presents the results of an estimation of the differential relationship between incumbency and the probability of changing constituency. While the difference is almost statistically significant at the ten percent level, the point estimate is only 2.21 percent, suggesting that at most 2.21 percent of the estimated incumbency disadvantage results from candidates switching constituencies.

1.6.4 Previous Periods

Because of the large differences in the political climate prior to 1991, it is useful to ask whether or not the effects of incumbency were different before 1991. Table 6 shows the estimated advantage of being an incumbent for each national election, except for those in 1951 and 1977.[16] While not all of the point estimates are significant, it is clear that there was a break between the 1989 election and the 1991 elections. Before 1991, incumbents enjoy advantages of varying sizes, but starting in 1991, being an incumbent becomes a disadvantage in each election.

Figures 7 and 8 depict the non-parametric estimation of the relationship between the probability of victory in a given election and the margin of victory in the previous election. These figures differ from the Figure 4 in two ways. First, allowing for the discontinuity, there is at best a weak increasing relationship between the margin of victory and the probability of reelection. Second, however, the behavior at the discontinuity is the opposite of that shown in

[15] In India political parties tend to divide frequently. So, in the process of matching names within parties, I take care to group all of the splinter parties together when I encounter a party that has split.

[16] These years are omitted because there is no prior election to use to identify a sample of candidates. The election in 1951 is simply the first election ever. The election in 1977 is the first election after the emergency. Matching candidates from the 1971 election to outcomes in the 1977 election is prevented both by the imposition of martial law which made the experience of holding office over this period qualitatively different than holding office in other terms and by the fact that the constituency definitions changed after the 1971 elections.

Figure 4 with incumbents being more likely to be elected than non-incumbents. The differences in the discontinuity are formalized in Tables 7 and 8 which present, for the pre-Emergency[17] and pre-1991 periods respectively, the same specifications that were presented in Table 4. Prior to the Emergency, incumbents enjoyed an advantage of being eleven percent more likely to win than non-incumbents. In the 1980, 1984 and 1989 elections, this advantage falls to about six to nine percent and becomes statistically insignificant. Finally, despite the insignificance of the level, the change that occurs after the 1989 election (column four) is statistically significant, highlighting the importance of the political shift that occurred between the 1989 and 1991 elections.

The results for both the pre-Emergency and pre-1991 periods are strikingly similar to the results found for the United States Congressional elections by Lee (2001). First, incumbents are at an advantage when compared to non-incumbents. Second, part of the source of the incumbency advantage comes from the fact that non-incumbents are less likely than incumbents to rerun. As the twelfth rows in Tables 7 and 8 show, incumbents are seventeen (Pre-Emergency) and thirteen percent (Pre-1991) more likely to contest the subsequent election. In fact, the effect is large enough to constitute the entire advantage, raising questions about the cause and size of the effect. If, for example, winning candidates find it easier to garner support for a subsequent bid for election, then on average incumbents will prove more likely to win, not because they are stronger candidates, but because they are simply more likely to appear in the subsequent election.[18]

[17] Unlike the period after the emergency, the political boundaries did change before the emergency. Rather than match names within constituency, I match instead by state which is feasible since, relative to after the Emergency, the number of total candidates is small. Additionally, for elections in 1951 and 1957, a number of constituencies elected two or three candidates. In these cases, all elected candidates are assumed to be incumbents. The margin of victory for winners is then calculated relative to the unelected candidate with the largest percentage of votes and the margin of victory for losers is calculated relative to the winner with the least votes.

[18] The interpretation of this effect as a bias depends critically on the definition of the incumbency effect. It could be argued, for example, that garnering support for re-election could be a component of the incumbency advantage. The goal of this estimation procedure, however, is not to estimate the relative strength of incumbents and non-incumbents over time, but to answer the counter-factual question, "How strong would an incumbent have been in an election had he or she not held office prior to the election?" Under this interpretation, anything that differentially affects the rate at which incumbents and non-incumbents run for office could bias the estimated incumbency effects.

1.6.5 Why the Change?

Why were incumbents at an advantage prior to 1991 and at a disadvantage starting in 1991? The available data is not sufficient to identify a single cause but does suggest a possible explanation. The major political difference between the periods prior to 1991 and the period starting in 1991 is that in the first period, Congrėss could be expected to hold power in the Lok Sabha. Starting in 1991, control of the parliament oscillated between parties potentially making experience in the previous term less valuable. Using the structural model presented in Section II, it is possible that the uncertainty about which party controlled the parliament caused the degree to which voters benefited from a candidate's experience, $\tilde{S}'(e)$, to fall. Making a voter's valuation of additional experience, $\tilde{V}'(e)$, negative.[19]

To probe for potential causes of the disadvantage, I estimate the effects of incumbency status on different types of incumbents. I focus on constituencies with different observable characteristics. Specifically, constituencies are chosen conditional on whether the winner of the first election matches the identified criteria. This requires care when interpreting the results. The regression discontinuity approach allows me to estimate the causal relationship between the probability of reelection and a candidate holding office prior to the election for each sub-sample. It is not correct, however, to interpret the differences between the incumbency effects for different subgroups as being caused by the differences in candidate characteristics. It is possible, for example, that the opposition that candidates face could vary depending on their characteristics. The purpose of this exercise is simply to compare average differences in the estimate incumbency effects for various subsets of the candidates. For each sub-sample, I estimate the incumbency effects using the spline estimator described above.

Congress candidates should be relatively more effective at working within the Congress system. If the stability of Congress's dominance made experience valuable, then incumbents affiliated with the Congress party should have a larger incumbency advantage than those not affiliated with Congress. Second, starting in 1991, Candidates from all parties should fare the

---

[19] A critical assumption in this explanation is that the graft and service effects are affected to different degrees by the collapse of the stable political system. As Besley and Case (1995), however, make clear, voters can monitor and hence to some degree control more public activities like setting tax rates and enacting legislation. These are precisely the kinds of activities that would be affected by the loss of Congress's dominance. The less public activities that derive from holding office in the constituency and that are reportedly prone to corruption, like exerting influence in the awards of local government contracts and civil service appointments, were probably less affected.

same since no party can be counted on to control the parliament over the next term. As Table 9 shows this is roughly what happed. Row one lists the incumbency effect for incumbents that belong to the Congress party and the second row shows the effect for non-Congress candidates. Before 1991, Congress incumbents enjoyed an advantage of almost twenty percent compared to non-Congress incumbents who fared just as well as the non-incumbents that they faced. Starting in 1991, both groups experience large disadvantages with Congress incumbents seventeen percent less likely than their non-incumbent competitors to win an election and non-Congress candidates ten percent less likely.

Rows three and four of Table 9 divide constituencies by whether or not Congress dominated the constituency prior to 1991. An advantage of this approach is that, unlike in the previous comparison, constituencies cannot change groups over time. The results are consistent with those in rows one and two. Row three contains constituencies in which Congress won over half of the four elections between 1977 and 1989 and row four contains the remaining constituencies. Up to 1989, incumbents in Congress dominated constituencies enjoyed advantages of about thirty-five percent while incumbents in other constituencies fare as well as their non-incumbent competitors. Starting in 1991, however, all incumbents suffered a disadvantage with those from Congress dominated constituencies fourteen percent less likely to win an election and those from other constituencies eleven percent less likely to win.

Rows five through eight further divide the constituencies in rows one and two by the incumbent's level of experience. Specifically, I divide the constituencies based upon whether or not the incumbent has been an incumbent for two periods (a "serial" incumbent) or for only one period. Referring back to equation (10), if the change in the incumbency effects were driven by a change in $\tilde{S}'(e)$ due to Congress's decline, The experience results should reflect the same pattern as the one that exists for party membership. Prior to 1991, experienced incumbents should outperform inexperienced incumbents and experience should be more valuable for incumbents that belong to Congress. Starting in 1991, all the effects of incumbency should decline and all incumbents should fare equally well. If on the other hand, the disadvantage results from the fact that $G'(e)$ increased because voters became more likely to expect that candidates would divert funds to for their own benefit, then experienced candidates should fare worse than inexperienced candidates starting in 1991.

The results are largely consistent with those in the first four rows. Prior to 1991, experienced incumbents outperformed inexperienced incumbents with experienced Congress incumbents having the largest advantage (fifty-six percent compared to about eleven percent for non-Congress incumbents). Starting in 1991, the incumbency effects for all types of incumbents except experienced Congress incumbents declined to a six to eleven percent disadvantage. Congress incumbents, on the other hand, fare much worse, falling to a disadvantage of twenty-seven percent. This could reflect that, at least for incumbents affiliated with Congress, the scandals that broke around the 1989 election made voters somewhat more suspicious of experienced candidates, possibly reflecting an increase in $G'(e)$.

Finally, I check a possible alternative explanation for these results. In the preceding discussion, I have assumed that voters choose to vote for a candidate based upon value that voters expect to derive from voting for the particular candidate. A vote for a candidate, however, is also a vote for a party, and it is possible that voters' valuation of a candidate is dominated by the voters' valuation of the party that the candidate represents. If this is true, then the poor performance of non-Congress incumbents prior to 1991 could be explained by voters' general preference for Congress, and the incumbency disadvantage that began in 1991 could simply reflect voters' changing preferences for candidates of different parties. Elected when their party was in favor, incumbents lost in subsequent election since voters' preferences increasingly switched to an opposing party.

Row nine of Table 9 contains incumbents that belong to a major party that either wins an outright majority or joins the coalition formed in the given election. Row ten on the other hand contains only incumbents of major parties that fall from power in the given election. If the cycling effect dominates, then incumbents in row nine should enjoy an advantage while incumbents in row ten should suffer a disadvantage. This patter, however, is not evident in the data. Prior to 1991, incumbents that belong to a party that is coming to power enjoy a large advantage. This is consistent with the results in the previous rows since except for the 1989 election, this row contains primarily Congress candidates. Starting in 1991 when power switched first to and then away from Congress, all incumbents suffer a similar disadvantage. Those that belong to parties that are coming to power are seventeen percent less likely to win than their non-incumbent competitors and those falling from power are twelve percent less likely to win.

In general, the results seem to be consistent with a model in which Congress's dominance before 1989 provided a stable political system in which voters could expect that the experience candidates developed while holding office would make them more valuable in future terms. Starting in 1991, this experience became less valuable since the party that held office in a given term could not be counted on to win the next election. The effect does not seem to be the result of parties cycling through power, and there is some evidence that voters find experienced Congress incumbents less valuable than inexperienced incumbents or experienced incumbents from other parties, possibly as a result of Congress's reputation for corruption over the period in question.

1.7 Conclusion

This study documents the surprising fact that unlike the results of studies that investigate the effects of incumbency in the U.S. and other countries, incumbents in Indian national parliamentary elections starting in 1991 are at a disadvantage compared to those candidates that do not hold office prior to contesting an election. I estimate the causal effects of incumbency using a non-parametric regression discontinuity design which relies on comparing incumbents that are barely elected to non-incumbents that barely lose. The results indicate that in marginal elections, incumbents are fourteen percent less likely to win an election than non-incumbents. This is an enormous deficit. To make up such a loss, an incumbent would have to have won the previous election by over five and a half percent of the popular vote, a change equivalent to an incumbent moving from the first to the thirty-fifth percentile of elected officials ranked by margin of victory.

In addition to other contexts, the disadvantage is also in contrast to results from before the 1991 election when incumbents seem to have enjoyed advantages almost as large in magnitude. Congress dominated the Indian national parliament prior to 1989 while afterwards control of the parliament oscillated between political parties. Comparing subsets of constituencies over time reveals two suggestive facts. First, the change in the effect of incumbency seems to be concentrated in constituencies that were represented by incumbents affiliated with the Congress party. Second, after 1989, incumbents from all parties seem to fair equally poorly. The results are generally consistent with a model in which prior to 1991, voters valued the experience of incumbents because they had experience working within the Congress system and Congress was likely to continue holding power. Starting in 1991, however, because control of parliament was likely to change hands, experience over the previous term proved less valuable to voters.

1.8 References

Acemoglu, Daron (2002) "Why Not a Political Coase Theorem? Social Conflict, Commitment a Politics" MIT Mimeo.

Acemoglu, Daron, Simon Johnson and James Robinson (2001) "The Colonial Origins of Comparative Development: An Empirical Investigation," *American Economic Review.* 91:1369-1401.

Ahuja, M. L. and Sharda Paul (1992). *1989-1991 General Elections in India.* Associated Publishing House: New Delhi, India.

Alfrod, John R. and John R. Hibbing. (1981) "Increased Incumbency Advantage in the House," *The Journal of Politics.* 43(4): 1042-1061.

Ansolabehere, Stephen and James M. Snyder, Jr. (2002) "Using Term Limits to Estimate Incumbency Advantages when Office Holders Retire Strategically," MIT Mimeo.

Austin-Smith, David and Jeffrey Banks (1989) "Electoral Accountability and Incumbency," *Models of Strategic Choice in Politics.* Ed. Peter C. Ordeshook. Ann Arbor: The University of Michigan Press.

Banks, Jeffrey S. and Rangarajan K. Sundaram (1993) "Adverse Selection and Moral Hazard in Repeated Elections Model" *Political Economy: Institutions, Competition, and Representation.* Ed. William A. Barnett, Melvin J. Hinich, and Normal J. Schofield. Cambridge: Cambridge University Press.

Banerjee, Abhijit and Lakshmi Iyer (2002) "History, Institutions and Economic Perfomance: The Legacy of Colonial Land Tenure Systems in India," MIT Mimeo.

Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan (2003). "How Much Should We Trust Differences-In-Differences Estimates?" MIT Mimeo.

Besley, Timothy and Anne Case, "Incumbent Behavior: Vote-Seeking, Tax-Setting, and Yardstick Competition," *American Economic Review,* 85(1): 25-45.

Chattopadhyay, Raghabendra and Esther Duflo (2003). "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India," *NBER working paper #8615.*

Chattopadhyay, Raghabendra and Esther Duflo (2003). "Efficiency and Rent Seeking in Local Governments," MIT Mimeo.

Collie, Melissa P. (1981) "Incumbency, Electoral Safety, and Turnover in the House of Representatives, 1952-1976," *The American Political Science Review.* 75(1): 119-131.

Delimitation Commission of India (2003). "Delimitation." http://www.delimitation-india.com.

Election Commission of India (2003). http://www.eci.gov.in.

Fan, Jianqing (1992). "Design-Adaptive Non-Parametric Regression," *Journal of the American Statistical Association*. 87(420): 998-1004.

Fan, Jianqing and I. Gijbels (1996). *Local Polynomial Modelling and Its Applications*. Chapman and Hall: London, 1997.

Frank, Katherine (2001). *Indira: The Life of Indira Nehru Gandhi.* Harper Collins Publishers: London, England.

Garand, James C. and Donald A. Gross (1984) "Changes in the Vote Margins for Congressional Candidates: A Specification of Historical Trends," *The American Political Science Review.* 78(1): 17-30.

Gelman, Andrew and Gary King (1990) "Estimating Incumbency Advantage without Bias," *American Journal of Political Science*. 34(4) 1142-1164.

Gould, Harold A. and Sumit Ganguly (1993) *India Votes: Alliance Politics and Minority Governments in the Ninth and Tenth General Elections.* Boulder: Westview Press.

Gupta, R. L. (1985) *Electoral Politics in India.* Discovery: Delhi, India.

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

Indian Parliament (2003). http://www.parliamentofindia.nic.in.

Jacobson, Gary C. (1987) "The Marginals Never Vanished: Incumbency and Competition in Elections to the U.S. House of Representatives, 1952-1982," *American Journal of Political Science*. 31(1): 126-141.

Jai, Janak Raj (1998) *Voter's Delima: Indian Political Scene.* Regency Publications: New Delhi, India.

Kashyap, Subhash (1988). *Parliament of India: Myths and Realities*. National Publishing House: New Delhi, India.

Kashyap, Subhash (1992). *The Ten Lok Sabhas: From the First to the Tenth, 1952-1991.* Shipra Publications: New Delhi, India.

Lee, David (2001). "The Electoral Advantage to Incumbency and Voters' Valuation of Politicians' Experience: A Regression Discontinuity Analysis of Elections to the U.S. House," NBER Working Paper #8441.

Levitt, Steven D., and Catherine D. Wolfram. (1997) "Decomposing the Sources of Incumbency Advantage in the U.S. House," *Legislative Studies Quarterly.* 22: 45-60.

Mayhew, David R. (1975) "Congressional Elections: The Case of the Vanishing Marginals" *Polity.* 6:310-312.

Majeed, Akhtar (2000) *Coalition Politics and Power Sharing.* Manak Publications: New Delhi, India.

Miguel, Edward and Farhan Zaidi (2003). "Do Politicians Reward their Supporters? Public Spending and Incumbency Advantage in Ghana," UC Berkeley Mimeo.

Mirchandani, G. G. and K. S. R. Murthi (1985) *Massive Mandate for Rajiv Gandhi: 1984 Lok Sabha Elections.* Sterling: New Delhi, India.

Pande, Rohini (2003) "Can Mandated Political Representation Provide Disadvantaged Minorities Policy Influence? Theory and Evidence from India," *American Economic Review.* 93(4): 1132-1151.

Porter, Jack R. (2002). "Asymptotic Bias and Optimal Convergence Rates for Semiparametric Kernel Estimators in the Regression Discontinuity Model," HIER Discussion Paper No. 1989.

Rana, Mahendra Singh (2000) *India Votes: Lok Sabha and Vidhan Sabha Elections 1999, 2000: Poll Analysis, Election Data, and Party Manifestos.* New Delhi: B.R. Publishing Corp.

Roy, Meenu (1996) *India Votes, Elections 1996: A Critical Analysis.* New Delhi: Deep and Deep Publications.

Roy, Ramashray and Paul Wallace (1999) *Indian Politics and the 1998 Election: Regionalism, Hidutva, and the State Politics.* New Delhi: Sage Publications.

Sharda, Paul (1990) *1989 General Elections in India.* New Delhi: Associated Publishing House.

Tufte, Edward R. (1973). "The Relationship Between Seats and Votes in Two-Party Systems," *American Political Science Review.* 67:549-553.

Weiner, Myron (1983) *India at the Polls, 1980: A Study of the Parliamentary Election.* American Enterprise Institute for Public Policy Research: Washington, DC.

Figure 1: Membership in Lok Sabha

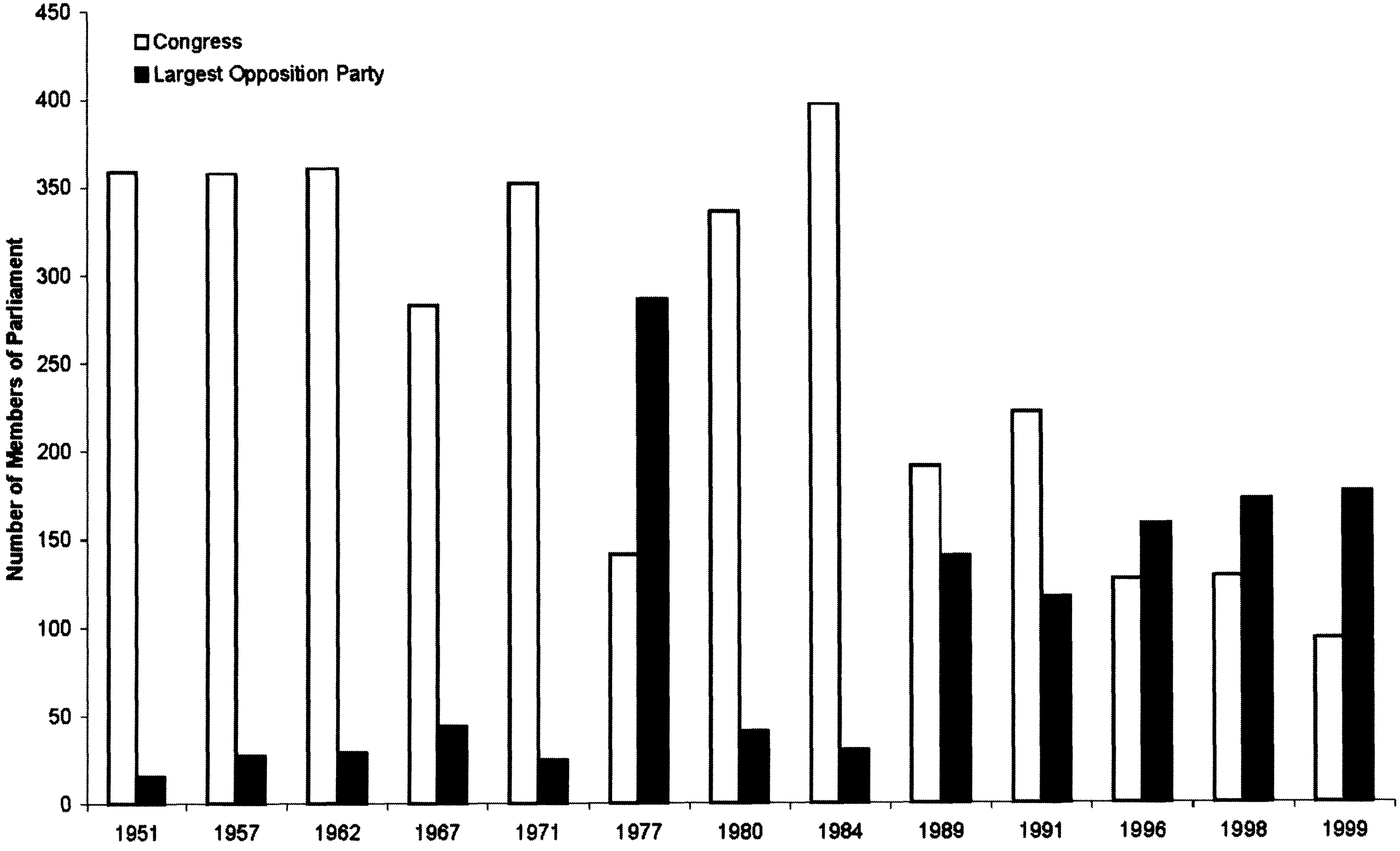

Figure 2: Average Number of Candidates per Constituency

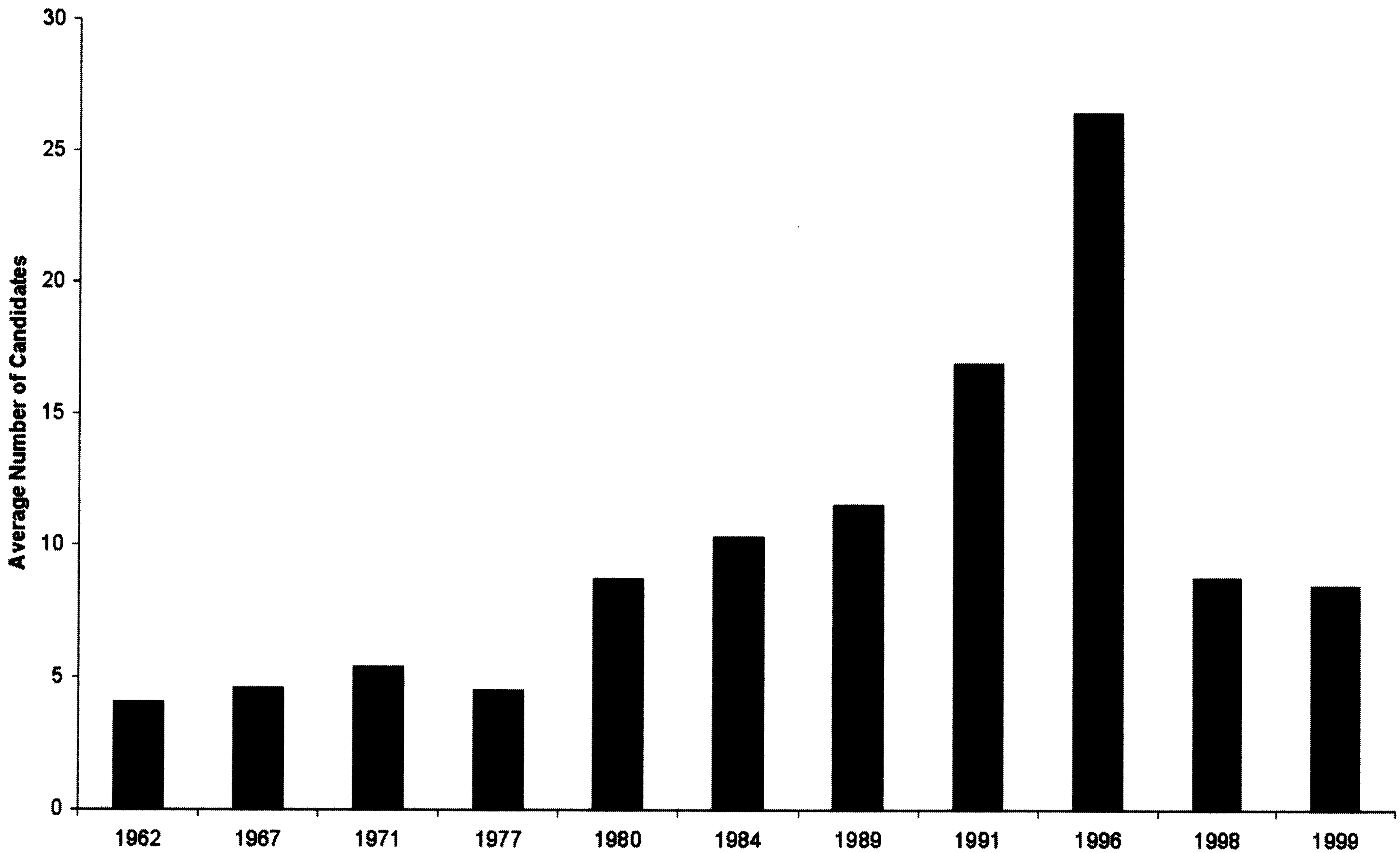

Figure 3: Number of Parties with at Least 10 Candidates

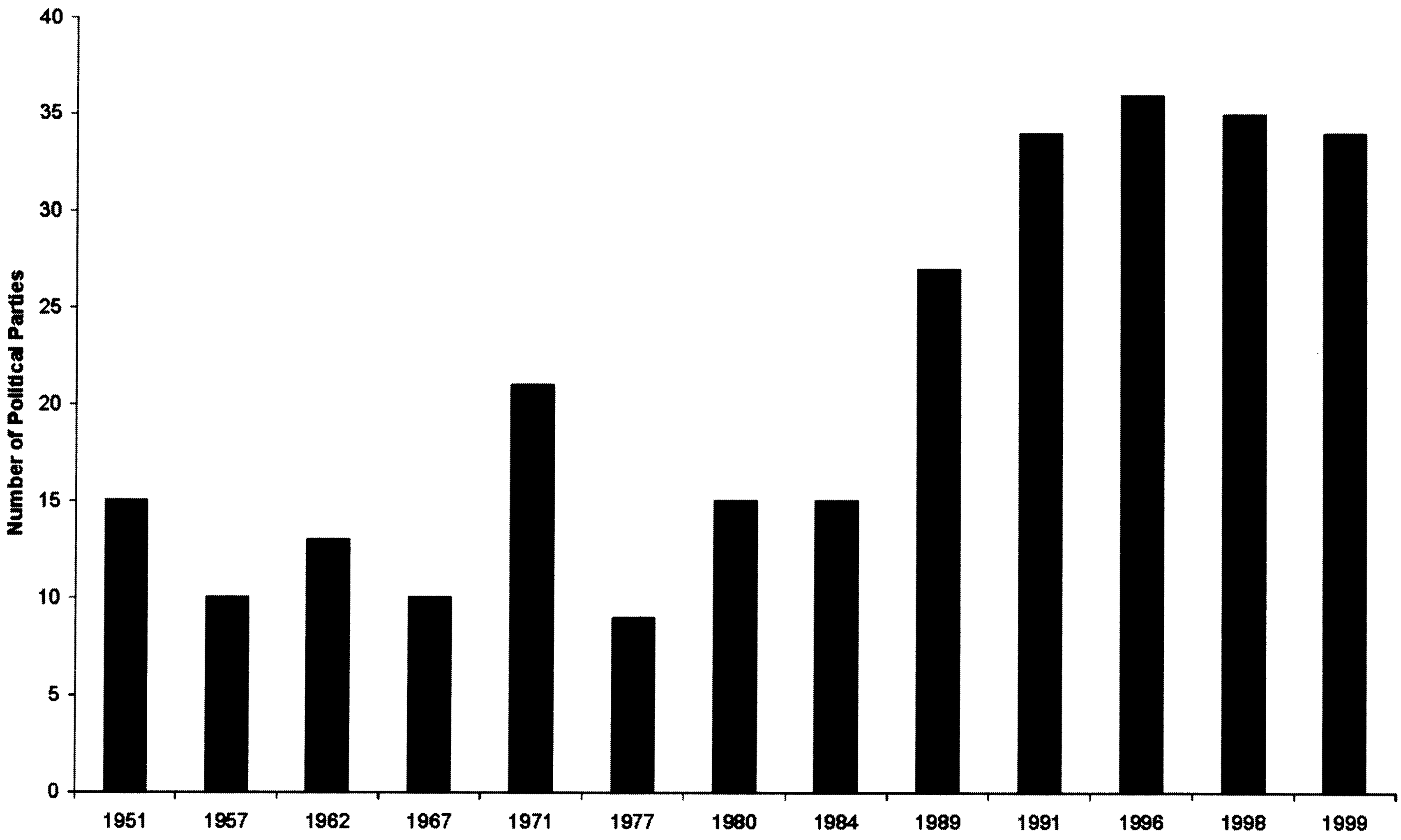

Figure 4: Probability of Re-election, 1991-1999

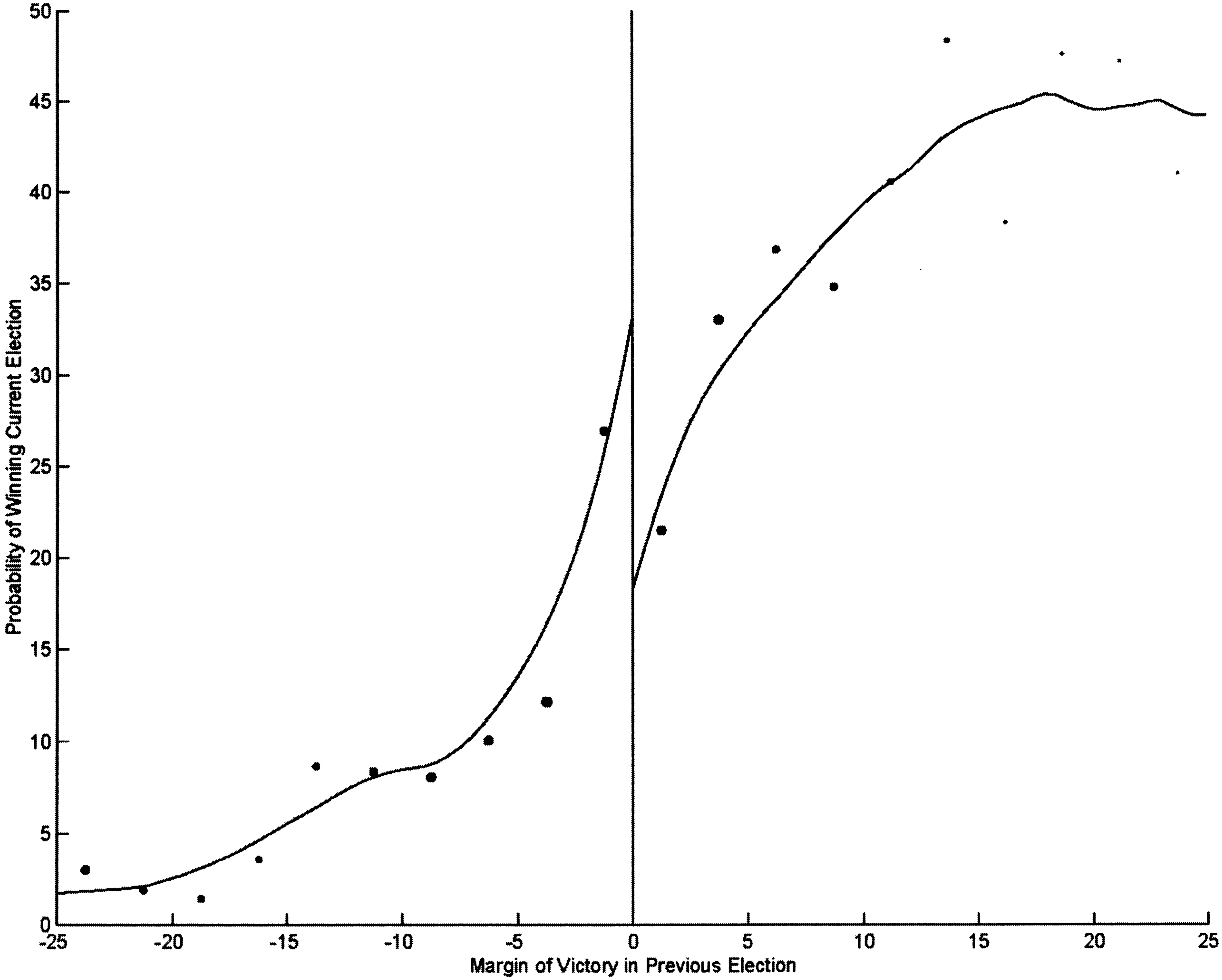

Figure 5: Probability of Re-election, 1991-1999, Full Support

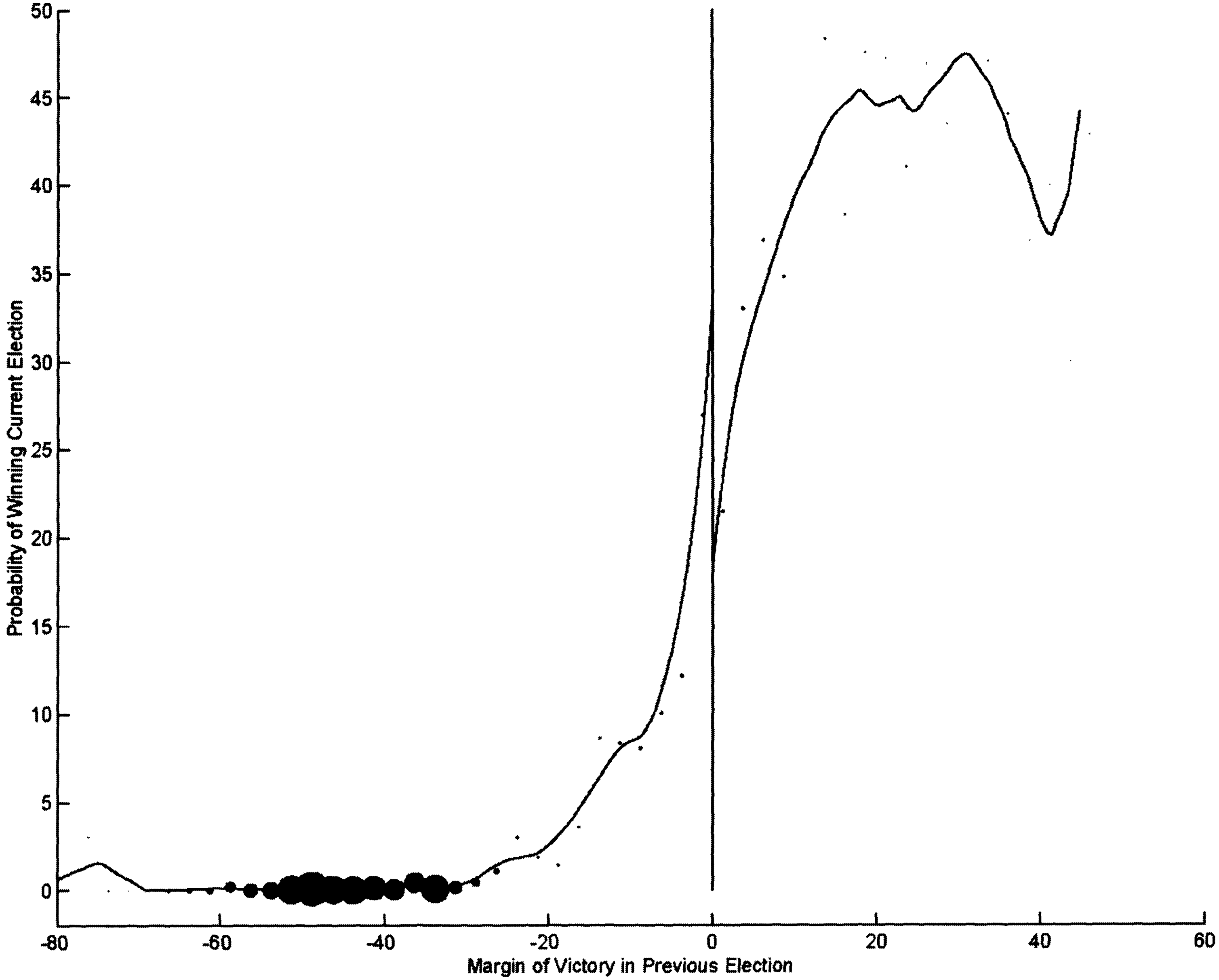

Figure 6: Probability of Re-Running, 1991-1999

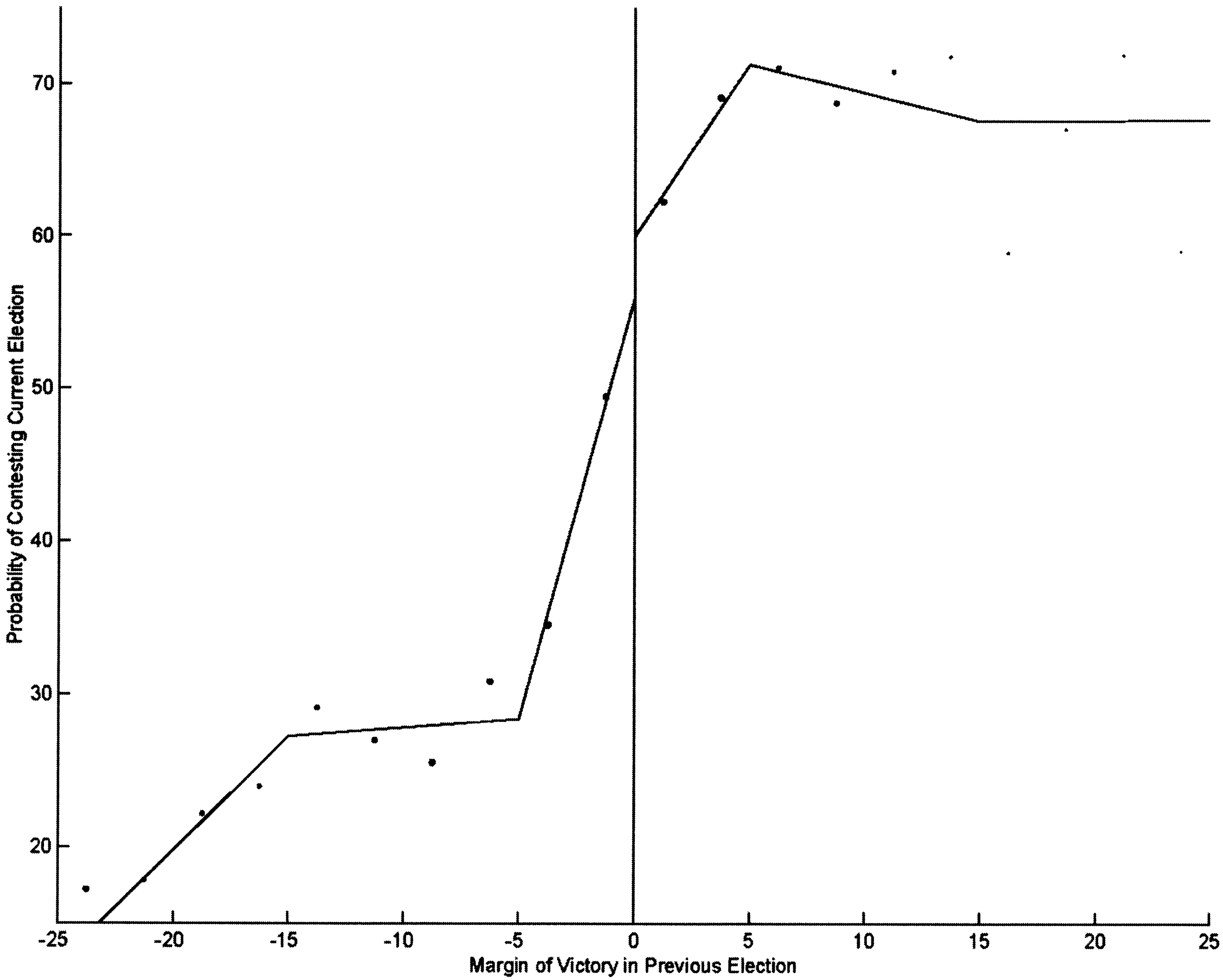

Figure 7: Probability of Re-election, 1957-1971

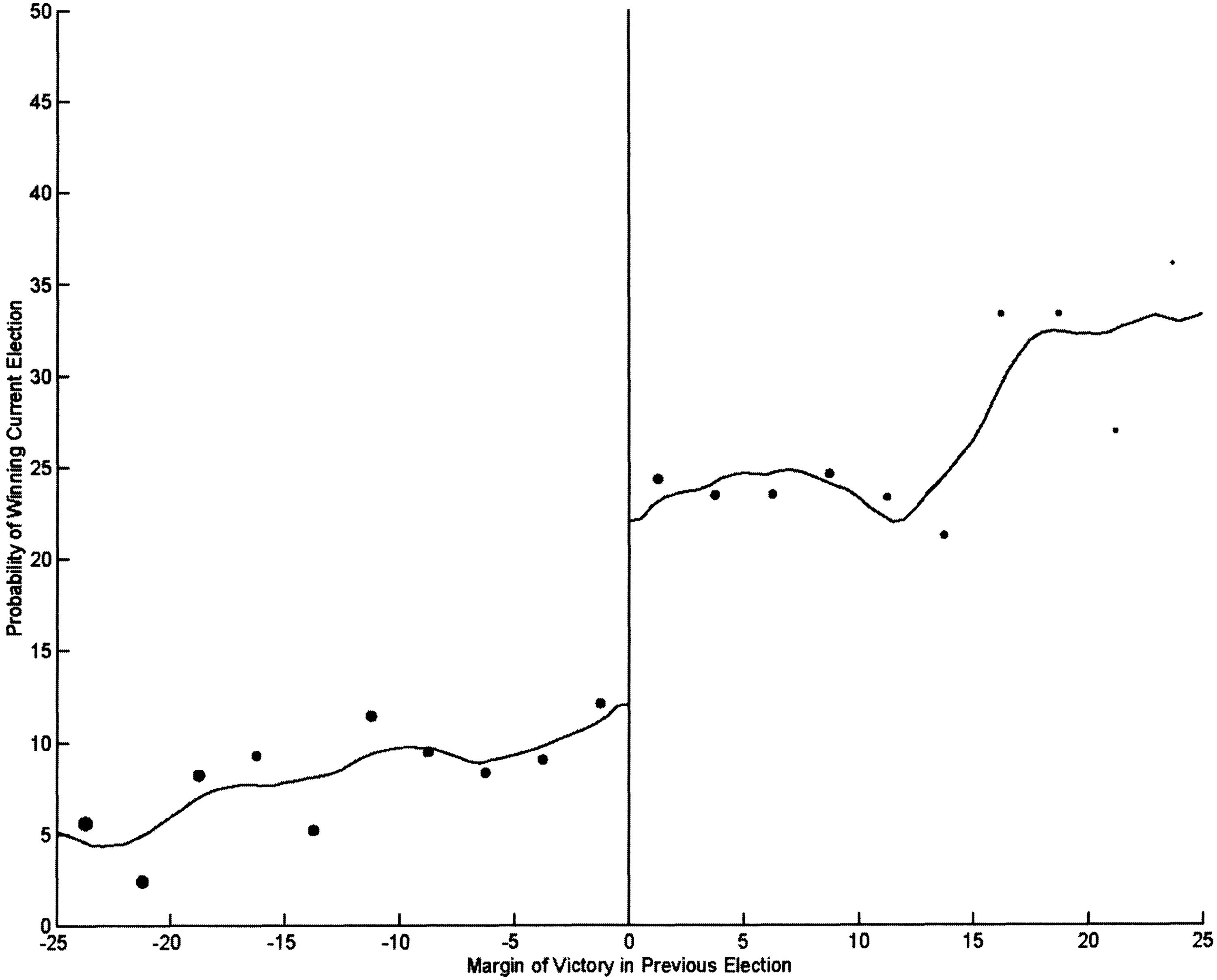

Figure 8: Probability of Re-election, 1980-1989

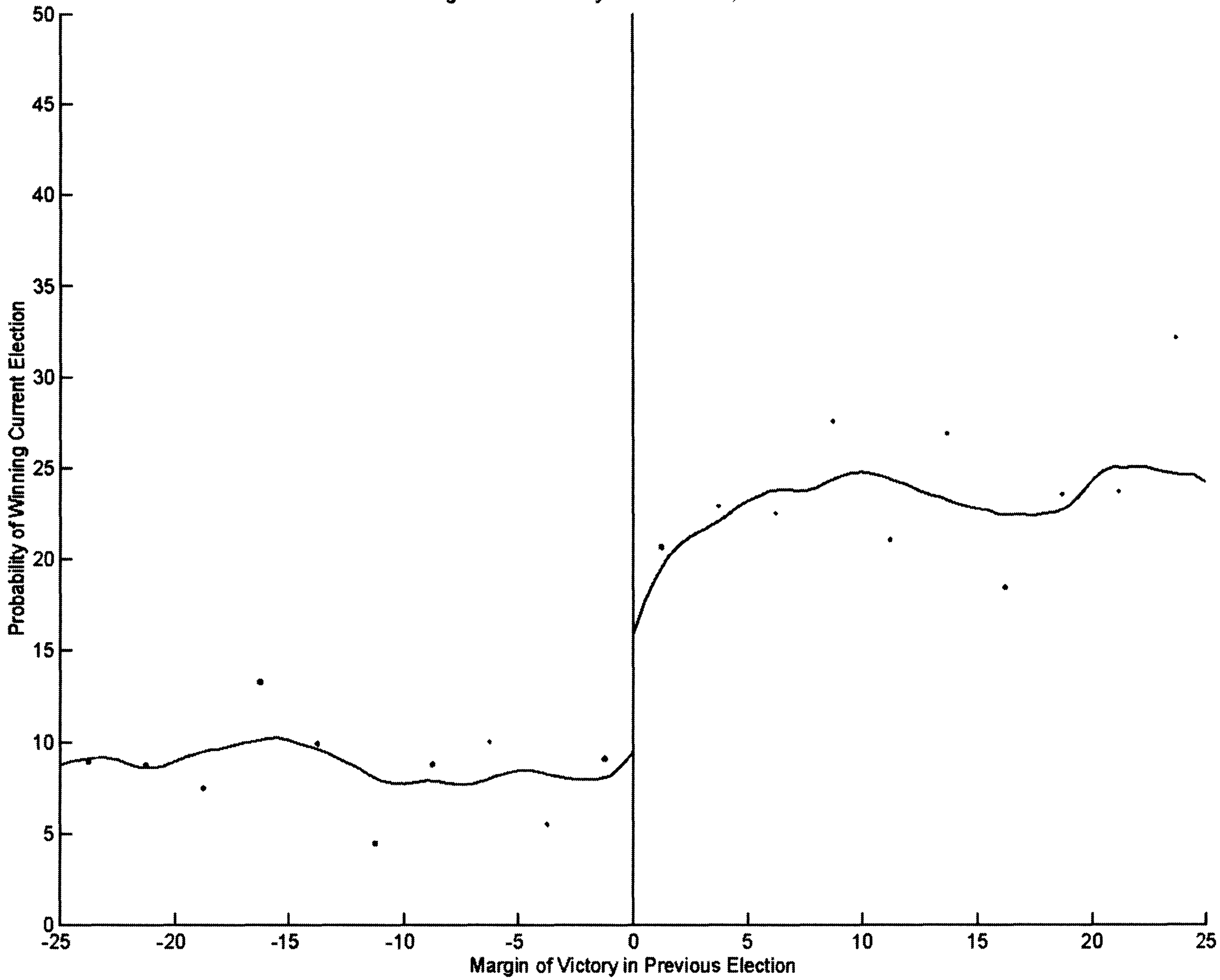

Table 1: Average Characteristics of Constituencies

Constituency Characteristic	1991-1999 Marginal	1991-1999 Non-Marginal	1980-1989 Marginal	1980-1989 Non-Marginal
% Female	4.8%	4.7%	3.1%	2.9%
# Previous Elections Won	0.19	0.22	0.09	0.08
# Years Experience	0.59	0.73	0.33	0.29
# Previous Elections Contested	0.50	0.50	0.21	0.18
% Belong to Congress	11.0%	9.9%	13.6%	16.0%
% Belong to BJP	7.3%	7.6%	2.2%	1.7%
% Reserved for Tribes	6.9%	8.4%	6.4%	8.0%
% Reserved for Castes	14.5%	15.3%	13.8%	15.3%
Number Electors	1,087,370	1,081,797	771,193	726,926
Number Voters	650,423	628,965	488,943	436,029
Number of Candidates	14.8	15.5	9.1	8.6
% of Voters Polled	60.0%	58.8%	63.3%	59.9%
% with margin of victory less than 5% at least once		90.1%		73.2%

Table 2: Average Election Rates of Candidates, 1991-1999

Subset of Candidates	Average Election Rates Inc	Non-Inc	Diff
All Candidates	37.78***	0.86***	36.92***
	(3.28)	(0.07)	(3.30)
Won or Lost by Less than 15 Percent	34.42***	12.46***	21.95***
	(4.04)	(1.17)	(4.71)
Won or Lost by Less than 2.5 Percent	21.48***	26.91***	-5.43
	(2.23)	(4.32)	(5.60)

Table 3: Average Characteristics of Candidates, 1991-1999

Candidate Characteristic	All Candidates Inc	Non-Inc	Diff	|Margin of Victory| < 2.5* Inc	Non-Inc	Diff
% Female	6.9%	3.9%	3.1%	7.8%	8.4%	-0.6%
# Previous Elections Won	0.75	0.05	0.70	0.69	0.54	0.15
# Years Experience	2.62	0.19	2.44	2.38	1.81	0.57
# Previous Elections Contested	1.09	0.27	0.82	1.09	0.93	0.16
% Belong to Congress	33.2%	4.1%	29.1%	35.6%	33.1%	2.5%
% Belong to BJP	26.3%	3.1%	23.2%	21.9%	25.5%	-3.6%

* Includes only candidates who win or lose an election by less than five percent of the popular vote

Table 4: Incumbency Effects, 1991 to 1999

	Estimates at Discontinuity		
	Inc	Non-Inc	Diff
Probability of Winning			
Non-Parametric	18.37***	33.24***	-14.87***
	(1.34)	(1.24)	(1.82)
Local Linear Fit	19.42***	33.71***	-14.29**
	(3.00)	(5.46)	(5.98)
Quartic Polynomial			
No Fixed Effects	17.39***	36.65***	-19.26***
	(4.57)	(5.31)	(6.77)
Fixed Effects	20.49***	39.7***	-19.21***
	(6.77)	(6.85)	(7.33)
Spline			
No Fixed Effects	19.72***	32.05***	-12.34**
	(2.82)	(4.91)	(6.15)
Fixed Effects	19.46***	31.71***	-12.25**
	(2.65)	(5.23)	(6.19)
Continuity Checks			
Female	7.532***	8.42***	-0.888
	-2.012	-1.452	-3.071
Membership of Congress	34.454***	34.845***	-0.39
	(4.93)	(7.08)	(5.31)
Num Previous Elections Won	.68***	.563***	0.117
	(0.09)	(0.06)	(0.08)
Num Years MP Experience	2.344***	1.921***	0.423
	(0.33)	(0.24)	(0.31)
Num Elections Contested	1.063***	.953***	0.11
	(0.11)	(0.09)	(0.11)
Probability of Rerunning	59.98***	55.98***	4
	(4.38)	(6.16)	(5.63)
Probability of Winning	33.23***	60.3***	-27.07***
Conditional of Rerunning	(4.29)	(5.88)	(8.57)

Note: Standard deviations of point estimates, clustered by states, in parenthesis.
MISE minimizing bandwidth for Local Polynomial Regression is 6.27.
Unless otherwise stated, estimates generated using spline estimtor.

Table 5: Sources of Bias in Estimation of Incumbency Effects

	Estimates at Discontinuity		
	Inc	**Non-Inc**	**Diff**
Levels of Name Matching			
Entire String	16.68***	27.83***	-11.15***
	-1.75	-3.79	-3.67
Reorder and Missing Names	20.99***	31.86***	-10.87**
Exact Match Required	(2.65)	(3.46)	(4.83)
Reorder of Names	19.46***	31.71***	-12.25**
and Partial Matching	(2.65)	(5.23)	(6.19)
Ambiguous Matches	0.27	1.63	1.36
	(0.61)	(1.04)	(0.99)
Constituency Switching	5.84***	3.63***	2.21
	-1.52	-1.56	-1.3

Note: Standard deviations of point estimates, clustered by states, in parenthesis.
Estimates generated using spline estimator.

Table 6: Incumbency Effects by Year

Year	Inc	Non-Inc	Diff
1957	16.48	14.25	2.23
	(10.36)	(9.21)	(18.08)
1962	27.65***	10.73*	16.92**
	(9.24)	(6.08)	(7.16)
1967	21.12***	10.08**	11.04
	(4.93)	(4.69)	(7.21)
1971	32.13***	9.79*	22.34**
	(5.97)	(5.88)	(10.23)
1980	27.65***	15.18*	12.47
	(8.97)	(8.70)	(12.75)
1984	12.13*	6.2	5.93
	(6.90)	(8.46)	(11.46)
1989	16.66***	6.91*	9.75
	(5.06)	(3.60)	(7.47)
1991	21.28**	40.92***	-19.64
	(8.79)	(7.24)	(12.09)
1996	15.32**	19.23***	-3.92
	(6.00)	(6.50)	(5.07)
1998	19.57**	27***	-7.43
	(7.97)	(6.00)	(11.27)
1999	21.75***	38.47***	-16.71
	(5.46)	(10.07)	(14.12)

Note: Standard deviations of point estimates, clustered by states, in parenthesis.
Estimates generated using spline estimator.

Table 7: Incumbency Effects, 1951 to 1967

	Estimates at Discontinuity		
	Inc	Non-Inc	Diff
Probability of Winning			
Non-Parametric	23.01***	11.95***	11.05***
	(1.25)	(1.66)	(2.08)
Local Linear Fit	23.59***	12.38***	11.22**
	(3.26)	(3.60)	(5.25)
Quartic Polynomial			
No Fixed Effects	23.66***	12.06***	11.6
	(5.33)	(3.57)	(7.42)
Fixed Effects	24.6***	13***	11.6
	(5.04)	(3.87)	(7.44)
Spline			
No Fixed Effects	24.79***	11***	13.79**
	(4.32)	(3.45)	(6.23)
Fixed Effects	26.7***	12.98***	13.72**
	(4.17)	(3.48)	(6.21)
Continuity Checks*			
Female	-2.09	-0.35	-1.74
	(1.48)	(2.48)	(2.91)
Membership of Congress	53.125***	51.21***	1.914
	(3.10)	(3.53)	(5.99)
Num Previous Elections Won	-.261***	-.289***	0.029
	(0.05)	(0.05)	(0.06)
Num Years MP Experience	-1.384***	-1.54***	0.156
	(0.26)	(0.27)	(0.30)
Num Elections Contested	-.721***	-.743***	0.022
	(0.10)	(0.10)	(0.07)
Probability of Rerunning	70.43***	53.04***	17.39*
	(7.63)	(5.50)	(9.96)
Probability of Winning	40.3***	24.57***	15.72*
Conditional of Rerunning	(4.59)	(5.87)	(8.55)

Note: Standard deviations of point estimates, clustered by states, in parenthesis.
MISE minimizing bandwidth for Local Polynomial Regression is 4.37.
Unless otherwise stated, estimates generated using spline estimtor.
* Negative point estimates are the result of using an estimator that does not restrict the range to the [0,1] interval.

Table 8: Incumbency Effects, 1980 to 1989

	Estimates at Discontinuity			1980-89 less
	Inc	Non-Inc	Diff	1991-99
Probability of Winning				
Non-Parametric	15.9***	9.46***	6.44***	-21.31***
	(1.37)	(1.89)	(2.33)	(4.15)
Local Linear Fit	16.16***	8.77**	7.39	-21.68***
	(4.74)	(4.08)	(7.14)	(6.26)
Quartic Polynomial				
No Fixed Effects	14.98***	5.49	9.48	-28.75***
	(4.30)	(5.07)	(6.99)	(7.21)
Fixed Effects	40.06***	30.31***	9.75	-29.4***
	(5.04)	(6.54)	(8.17)	(7.61)
Spline				
No Fixed Effects	17.39***	8.3***	9.08	-21.42***
	(4.36)	(3.13)	(5.66)	(5.08)
Fixed Effects	18.62***	9.44***	9.18	-21.45***
	(4.70)	(3.11)	(5.78)	(5.09)
Continuity Checks				
Female	7.54***	5.83**	1.71	
	(2.65)	(2.38)	(3.22)	
Membership of Congress	54.519***	47.792***	6.727	
	(4.99)	(7.09)	(9.74)	
Num Previous Elections Won	.281***	.303***	-0.022	
	(0.06)	(0.06)	(0.08)	
Num Years MP Experience	1***	1.003***	-0.003	
	(0.20)	(0.19)	(0.28)	
Num Elections Contested	.246***	.347***	-0.101	
	(0.07)	(0.07)	(0.07)	
Probability of Rerunning	39.35***	26	13.34	-7.88
	(10.77)	(17.14)	(19.01)	(7.96)
Probability of Winning	35.23***	23.85***	11.38	-42.63***
Conditional of Rerunning	(5.16)	(5.35)	(8.87)	(14.12)

Note: Standard deviations of point estimates, clustered by states, in parenthesis.
MISE minimizing bandwidth for Local Polynomial Regression is 5.43.
Unless otherwise stated, estimates generated using spline estimtor.

Table 9: Incumbency Effects by Incumbent Characteristics

Sub-Division	Pre-1991 (1980-1989) Inc	Non-Inc	Diff	Post-1989 (1991-1999) Inc	Non-Inc	Diff	Pre-'91 vs. Post-'89
Congress Party	25.83***	6.34*	19.49*	11.26***	28.44***	-17.18**	-36.67***
	(8.72)	(3.50)	(10.35)	(4.24)	(5.36)	(8.12)	(10.84)
Non-Congress Party	9.3**	10.05	-0.75	24.46***	34.03***	-9.56	-8.81
	(4.31)	(6.58)	(6.72)	(3.76)	(5.34)	(7.64)	(8.84)
Congress Party Held Constituency For 3 or 4 Terms before 1991	37.28***	1.84	35.44***	18.57***	32.96***	-14.4	-49.84***
	(8.28)	(2.18)	(8.84)	(4.89)	(6.61)	(8.77)	(9.54)
Congress Party Held Constituency For less than 3 Terms before 1991	8.33***	11.11***	-2.77	20.29***	31.57***	-11.27	-8.5
	(3.19)	(4.13)	(4.01)	(4.62)	(6.11)	(7.72)	(6.58)
Serial† Congress Party Incumbent	60.26***	3.98	56.28**	7.93	35.27***	-27.34**	-83.62***
	(22.21)	(8.25)	(26.19)	(6.88)	(10.36)	(13.76)	(28.04)
Non-Serial Congress Party Incumbent	19.42**	6.69	12.73	12.87***	23.77***	-10.9	-23.62**
	(7.92)	(4.17)	(9.89)	(4.96)	(6.01)	(8.95)	(10.10)
Serial Non-Congress Party Incumbent	8.24*	-2.44	10.68**	30.73***	36.56***	-5.84	-16.52
	(4.62)	(2.58)	(5.35)	(9.00)	(7.70)	(13.01)	(13.38)
Non-Serial Non-Congress Party Incumbent	10.42*	13.61	-3.19	22.78***	33.17***	-10.39	-7.2
	(5.78)	(8.26)	(8.02)	(4.07)	(6.13)	(7.98)	(9.99)
Incumbent Belongs to Party Coming to Power	21.35***	3.86	17.5***	14.3***	30.92***	-16.61*	-34.11***
	(5.02)	(3.14)	(6.46)	(4.46)	(7.31)	(9.34)	(9.63)
Incumbent Belongs to Party Falling from Power	17.39***	8.3***	9.08	19.72***	32.05***	-12.34**	-21.42***
	(4.36)	(3.13)	(5.66)	(2.82)	(4.91)	(6.15)	(5.08)

Note: Standard deviations of point estimates, clustered by states, in parenthesis.
Estimates generated using spline estimator.
† Serial incumbents refers to incumbents that hold office for two terms prior to an election

Chapter 2: Could Have Been a Contender?

The Effects of Limiting the Number of Candidates in Indian State Parliamentary Elections

2.1 Introduction

It is hard to overemphasize the importance of competition in the democratic process. For voters, it provides a range of policy options from which to choose. It gives other candidates an incentive to articulate their positions and inform voters about the differences in their respective platforms. And once in office, the knowledge that the official will face competition in the next election provides an incentive to take actions that are in the interests of voters while in office.

The difficult question, however, is to determine what constitutes political competition. Candidates that have a significant chance of winning an election undoubtedly provide meaningful competition, but candidates do not have to win to influence an election. Washington (2003), for example, provides evidence that just the existence of an African American candidate causes existing congressional members to vote more conservatively.

The theoretical literature describing how democratic electoral processes relate to a candidate's behavior in office emphasizes that the just the opportunity of being able to vote for other candidates, even if not exercised, can influence the behavior of public officials. Most models use the standard moral hazard (Austin-Smith and Banks, 1989) and adverse selection models (Banks and Sundaram, 1993) to reconcile the large advantages that incumbents posses with the fact that incumbents do seem responsive to the electoral process. In all of these models, incumbents take observable actions that are in the interests of voters because, within voters' equilibrium strategies, these actions allow the incumbents to increase their probability of re-election. Implicit in these models, however, is the idea that voters do choose who they would prefer to be in office. Furthermore, the choice to retain an incumbent is a choice in which the benefits of providing incentives to elected officials trades off with the opportunity cost of forgoing the election of other candidates.

In this body of research, two relationships shape the incentives of incumbents. First, voters must act such that proper action on the part of the incumbent increases the probability of election. Providing a "carrot" for proper action, however, is meaningless without a "stick" in hand to punish deviations from the equilibrium strategy. So, it is also necessary that the same voters can credibly commit to punishing the incumbents for non-action by choosing another candidate if incumbents choose to take actions that are not in the interests of voters. The more valuable the opportunity cost of not electing another candidate, the stronger the incentives that voters can provide to incumbents.

The first relationship has received a fair amount of empirical attention. For example, Besley and Case (1995), Case (2001), Chattopadhyay and Dulfo (2003), Miguel and Zaidi (2003), and Pande (2003), all demonstrate that electing a specific candidate increases the flow of resources to the candidate's constituents. This paper, on the other hand, relates to the second question. By asking whether a change in election laws that reduces the number of candidates changes the probability of electing existing candidates, I ask whether or not these candidates provide opportunities to voters that they would not have if restricted to existing candidates. Based upon the existing theory, the question implicitly determines also whether or not these candidates serve a role in setting the incentives of candidates during the election and when in office.

More importantly, however, empirically investigating this relationship also sheds light on the more subtle question of whether additional candidates can constitute political competition and the effects of laws that attempt to limit competition to "viable" candidates. This simple question has profound implications for the design of electoral systems because, by necessity, democracies limit the number of candidates in order to control the cost of providing elections. In the U.S. for example, some consider the existence of only two political parties (and the resulting fact that most elections only have two viable candidates) to suggest a system plagued by a lack of competition that strongly favors incumbent candidates. In short, voters can not properly discipline incumbents because they have almost no one else to elect. This has motivated some observers to argue that requirements for listing candidates on state and national ballots should be relaxed to allow additional, though weaker, candidates to participate.

Compared to the U.S., India seems the opposite. The dizzying number of political parties and independent candidates at both the state and national levels suggest not a dearth but an excess of competition. Such an extraordinary number of people contesting for the relatively limited number of elected offices have caused many to wonder if these candidates actually have a role in the electoral process or are simply taking advantage of a public forum for reasons unrelated to the potential of being elected to office. After the most recent national election, for example, the commission in charge of Indian elections concluded that of the 5,435 candidates contesting at the national level, only 4,190 were serious contenders (NewKerala.com, 2004).

I focus on the degree to which the number of new candidates in a constituency affects the probability that existing candidates win office in Indian state parliamentary elections. Unfortunately, the causal relationship between the number of candidates and the probability of election is particularly difficult to estimate because the relationship is potentially endogenous. First, while additional candidates may make existing candidates less likely to be elected, constituencies where existing potential candidates are weak are also likely to attract many new competitors. So, even an inverse relationship between the number of new candidates and the probability of election would be ambiguous.

The relationship can also, however, work in the opposite direction. There is a significant amount of serial correlation in the number of candidates that contest an election in a given Indian constituency, meaning that a constituency that experiences an increase in the number of new candidates in one period will have a large number of new candidates in the next. If providing voters with more candidates from which to choose yields strong incumbents, then incumbents from constituencies that on average have large numbers of new candidates may also have higher rates of re-election than incumbents from constituencies with lower levels of competition.

My identification strategy utilizes a change in the Indian election laws enacted after the 1996 national election in response to the ever growing length of ballots in state and national parliamentary elections. The reforms made it more difficult for a candidate to contest an election by increasing the requirements for being nominated and by doubling the size of the deposit that candidates must risk when running for office.[20] These changes significantly reduced the number of candidates running for office after 1996 and had differing effects across constituencies depending on the prior level of competition. As one would expect, constituencies with the largest average number of candidates experienced the largest drop in the number of candidates contesting, providing exogenous variation in the number of new candidates both over time and between constituencies.

The results suggest that the number of new candidates have a measurable effect on the probability that existing candidates win office. On average, the addition of one new candidate reduces the probability that an existing candidate wins by three fourths of a percent. The probability of the average existing candidate winning falls by three hundredths of a percentage point from a baseline probability of four percentage points. This suggests that the change in electoral laws that decreased the number of new candidates in each constituency by an average of six candidates, increased the electoral prospects of the average existing candidate by 4.5 percent or two tenths of a percentage point. There is, however evidence of heterogeneity in the impact of the law on existing candidates, suggesting that the law could have impacted particular types of candidates more severely.

The remaining sections of the paper are organized as follows. Section II describes both the state parliamentary system in India and the change in the electoral rules in 1996. Section III, outlines the data used in the analysis, and Section IV describes the general empirical strategy. Section V provides the empirical results, and the paper concludes with section VI.

[20] Indian election rules require candidates to post a deposit when running for office. The deposit is only returned to a candidate if the candidate either wins the election or captures at least a sixth of the popular vote. The rule is design to deter individuals with no hope of winning election from running.

2.2 Indian State Elections and the 1996 Rule Change

2.2.1 The Role of the Vidhan Sabha

The Indian constitution provides for a federalist system much like the system in the U.S. – made up of national and state level representative bodies – but with a stronger center. Each state and two union territories (Pondicherry and the National Capital Territory of Delhi) have one directly elected house of parliament that mirrors the lower house of parliament at the national level, the Lok Sabha. These bodies are known as the Vidhan Sabhas or legislative assemblies, and their members are referred to as a Member of the Legislative Assembly or MLA. Some states also have unelected upper parliament similar to the Rajya Sabha at the national level.

As in most federalist systems, the relationship between the center (or national government) and India's has 32 state and union territories is often precarious. Unlike the U.S., however, the framers of the Indian constitution deliberately tipped the balance in the favor of the central government. The Indian constitution explicitly enumerates three lists of subjects upon which the center, the state, and both bodies can legislate, but the state parliaments must defer to the center on shared subjects and the center can legislate on matters reserved for the states as long as the measure receives the support of two-thirds of the national upper house of parliament. The autonomy of the state parliaments is thus limited relative to the national parliament, but the state parliaments do still play a crucial roll in the governance of Indian voters.

2.2.2 The Election of State Representatives

State parliamentary elections must occur either every five years or when the current state government loses majority control. Once called, the election is scheduled and run by the independent Election Commission of India. The commission runs all national and state parliamentary elections as well as the election of the President and Vice-President. The commission comprises three individuals: two Election Commissioners and the Chief Election Commissioner who can only be removed from office by parliamentary impeachment. In addition to running the elections, the commission also maintains the polling records that identify who can and can not vote in a given election.

The election itself follows the same first-past-the-post system as the elections to the national parliament. After all votes are cast, the candidate with the largest percentage of the vote (regardless of the magnitude of the fraction) wins the election. Votes are cast by secret ballot, and anyone over the age of 18 is allowed to vote (21 before 1989). To ensure accessibility, the ECI tries to locate polling stations so that no elector is more than 2km away from the nearest polling station and so that no station has to serve more than 1,500 voters. Finally, the election itself is usually spread over several days due to the sheer number of people and geographic distances involved.

Anyone over the age of 25 can contest a parliamentary election for the state in which they are a resident. Candidates must be nominated by one registered elector if standing from a registered political party[21] or ten registered electors if standing as an independent or member of an un-registered party. Finally, candidates must also post a bond, called the "deposit", when contesting. The candidate forfeits the deposit if they do not receive more than a sixth of the votes cast.

2.2.3 Change in Election Laws in 1996

Indian elections attract a surprisingly large number of political parties and candidates. Indian elections have always been relatively large. In the 1980's, most states had an average number of between 4 and 6 candidates per state. These average grew with each subsequent until in the early 1990's four states averaged over ten candidates a constituency. The most extreme example occurred in Tamil Nadu where in 1996, 1033 candidates contested the same parliamentary seat in Modakurichi.

Since many of these candidates were considered potentially frivolous and because the size of the ballot created administrative difficulties, the electoral rules were changed in August of

[21] Candidates can contest as an independent or as a member of a political party. In India, political parties are officially recognized as national or state level registered parties if they have performed sufficiently well in the last election. This recognition is critical because candidates are identified on a ballot by the largely illiterate population by symbols rather than the candidates' names. State parties are allowed to reserve a single symbol for use across the state and national parties have exclusive use of their symbol across the entire country. The allocation of all other symbols varies by constituency.

1996 after the 1996 general election. The number of people that must nominate candidates from non-registered parties was increased, but more importantly the size of the deposit, which is forfeited by weak candidates, was also increased. Before the law, candidates were required to deposit 500 Rupees ($US 10) to contest national elections and 250 Rupees for state elections. The law increased the sums to 10,000 Rupees and 5,000 Rupees respectively. Individuals contesting from constituencies in which representation is reserved for members of scheduled castes and scheduled tribes pay half of the amount required of regular candidates (both before and after the rule change). Despite this change, many still feel that too many candidates contest the elections.

2.3 Data

The data for this paper were acquired from the Election Commission of India. The ECI is responsible for holding all parliamentary elections in India, and after each election, provides an official report listing summary statistics on the campaigns for each state and national constituency as well as the actual results for each candidate in the campaign. These data include the name of each candidate, the candidate's gender, the candidate's political party, and the number of votes that the candidate received.

The political boundaries for both the state and national parliaments were set in 1973 by a special delimitation commission such that each state was represented in proportion to their share of the national population. These constituency definitions where then frozen in 1976 by constitutional amendment to avoid providing a disincentive for state implementation of family planning measures. The new constituency definitions will be redrawn by the current delimitation commission based upon the results of the most recent decennial census. Despite this general policy, some state level constituencies did change between 1977 and the present. Most notably, Uttar Pradesh completely changed constituencies for the state's last election in 2002, and the configurations of constituencies in Arunachal Pradesh and Goa changed in 1990 and 1989 respectively. Finally, a small number of constituencies do not report results in particular years.

The ECI maintains and releases the election data primarily to ensure transparency and accountability in the electoral process, and the organization and quality of the data are quite remarkable. From the perspective of this study, however, the major short coming of the available data is the inability to track the performance of candidates over time. The spelling of names, in most cases, is a transliteration from other Indian languages which generates variation in spelling of the same name from one election to the next. In addition, however, there is very little consistency in the reporting of name formats. When comparing the name given to the same candidate in various elections, the commission reports names in various orders, drops first or middle names entirely, and inconsistently reports various modifiers like the candidate's father, occupation, or other modifiers that the candidate associates with his or her name.

The solution employed in this paper is to use the same algorithm employed by Linden (2004). The algorithm seeks to match candidates over time within a given constituency. For a given constituency, the algorithm iterates through decreasing levels of specificity, searching through the names of candidates contesting an election in each year and matching the names across each election. The criteria for a match begin with exact matches between the names provided by the ECI. The algorithm then allows omitted or mis-ordered names, inaccurate division of names, and finally spelling differences of a character or less.

2.4 Empirical Strategy

2.4.1 General Approach

The goal of this paper is to estimate the effect of changes in the level of competition on the electability of candidates using the change in election laws in 1996 to provide exogenous variation in the number of candidates. Using a common strategy, I take the set of candidates in the election prior to the election of interest as the basis of observation. So, if the election of interest is in period t, then I use the election at period t-1 to identify a pool of existing candidates, and measure their probability of winning the election in period t as a function of the number of new candidates appearing in period t.

This strategy allows both for the identification of a set of likely candidates before the change in election laws in 1996 and allows for the use of the first period election results to control for the strength of each candidate in the election of interest. The disadvantage is that it is impossible to identify only candidates who intend to run in the subsequent election. While part of this decision is certainly determined by the number of new candidates likely to enter in the subsequent period, the inclusion of candidates with no intention of running in the subsequent period will bias the estimated effect of new candidates towards zero.

The potential endogeneity problems with this strategy are also clear. I propose to test an inverse relationship between the probability of winning the election in period t by a candidate from the election in period t-1 and the number of new candidates in period t. However, strong candidates in the prior period are likely to deter potential new candidates from entering an election, generating the same inverse relationship. Conversely, the relationship could also be direct. For example, it is possible that constituencies that have large numbers of candidates in elections produce strong candidates that are likely to win subsequent elections.

My solution is to use the change in electoral laws in 1996 to identify exogenous variation in the number of new candidates contesting. Figure 1 displays the average number of new candidates per constituency for the five largest states in the data set. The elections are ordered relative to 1996 so that the election in period 0 is the last election in the specified state to operate under the old deposit requirements, and the election in period -1 is the first election after the new deposit requirements. The change in election rules not only significantly reduced the number of new candidates contesting in the last election, but they also reversed a strong trend towards larger elections, suggesting that the laws in fact did constrain the number of new candidates contesting in the last election.

Figures 2 and 3 provide grater detail on the effects of the change in election laws. Figure 2 displays the distribution of constituencies by the number of new candidates for the election just prior to and just after the 1996 election laws. After the election laws, not a single constituency has more than 25 new candidates and the majority of constituencies have between zero and seven new candidates. Figure 3 demonstrates that the magnitude of the change in the number of new candidates is not uniformly distributed but rather varies inversely with the average number of candidates in each constituency before the change in the law. The constituencies that experienced the largest declines in the number of new candidates were the constituencies that, on average, had the largest number of candidates.

The problem with using a single uniform change in the laws is that it is difficult to isolate the changes in the electoral process caused by the change in the law from other factors that change over time. In Figure 1, for example, the elections in period 3 have about the same number of new candidates as the election after the change in electoral laws. Even controlling for the number of new candidates, however, the probability of candidates being elected is different because the political climate changed significantly in intervening 15-20 years. In fact, a very significant rupture occurs at the national level between 1989 and 1991 after the Congress party loses for the first time to a disparate opposition and incumbency changes from a political asset to a liability (Linden, 2004).

To ensure both comparability between elections before and after the change in election laws and robustness of the empirical results, I use two sets of election in the following analysis. First, I use the election just before and just after the change, and second, I use the first and last two elections after and before the change to check robustness. Since there are so few states with a second election after the change in the election laws, the second set primarily adds a second group of elections before the change in election laws. Table 1 presents the summary statistics of the variables used in the following statistical analysis.

2.4.2 Model

I employ the following basic model:

$$y_{ijkt} = \beta + f(NewCands_{ijkt}) + \delta' X_{ijkt} + \varepsilon_{ijkt}$$

y_{ijkt} is a dummy variable for whether or not candidate i won the election in period t, state j, and constituency k. ε_{ijkt} is a random disturbance term, and X_{ijkt} is a set of control variables. The set of control variables always contains a dummy for the gender of the candidate and a cubic polynomial of the margin of victory. The cubic polynomial is motivated by Figure 4 which provides a graph of the relationship between the probability of election and the margin of victory allowing for a discontinuity at zero due to incumbency effects. Over this period, the average incumbency effect is relatively small and the entire relationship can be well approximated with a cubic polynomial. The estimated standard errors for all coefficients are clustered at the constituency level.

The difference between the individual specifications is simply the variables that I use to instrument for the number of new candidates and then the control variables that I use to isolate the effects of those candidates. The goal is to test the robustness of the estimates of the effect of changes in the number of new candidates using different types of variation surrounding the change in the election laws in 1996. I use four specifications in what follows:

1. OLS: An Ordinary Least Squares estimate of the relationship between the probability of election of existing candidates and the number of new candidates.
2. IV1: Uses a dummy variable for the period after the change in deposit levels as an instrument for the number of new candidates.
3. IV2: Uses the interaction between a dummy variable for the period after the change and the average number of total candidates in each constituency before the change in election laws.
4. IV3: Uses both the average number of candidates before the change in laws and the interaction of the average with a dummy variable for the period after the change in electoral laws.

The results of each of the instrumental variable specifications are consistent. For the average existing candidate, an additional new candidate reduces the probability of election by about a half a percent.

2.5 Empirical Results

2.5.1 OLS Estimates

Table 2 contains the ordinary least squares estimates of the relationship between the probability of election and the number of new candidates. The first two columns contain estimates using only the first election before and after the change in the deposit requirements while the last two are estimated using two elections before and after the change. The estimates are very consistent in demonstrating a negative correlation between the probability of winning the election and the number of new candidates. On average, one additional new candidate decreases the probability of existing candidates winning the election by about two hundredths of a percentage point. Taking the curvature of the relationship into account with the quadratic specification suggests that the impact of additional candidates declines in the number of new candidates with the first new candidate reducing the probability of victory by eight hundredths of a percent. The implied effect is small, but not negligible since the probability of an existing candidate winning election is only four percent points. So, in the linear specification, each candidate changes the relative probability of election by a half a percent, and, in the quadratic specification, the impact of the first new candidate is two percent.

2.5.2 First Stage Regressions

To control for the potential endogeneity of the posited relationship, I use the exogenous variation generated by the change in the election laws in 1996. Table 3 contains the results of the first stage regressions for the subsequent instrumental variables estimates. Again, the first two columns contain estimates for the first election before and after the change in the law, and the second column adds the second elections before and after the change. Both pairs of estimates, however, show roughly the same relationship with the magnitude of the coefficients in the second pairs slightly smaller than the first. Columns one and three confirm the relationship displayed in Figure 1 – after the change in the deposit requirements, five and a half less new candidates contest the average election. Adding the elections two years out reduces this difference by over a candidate since the election just prior to 1996 was usually the largest.

As Figure 3 demonstrates, the effects of the change in electoral rules is not uniform, but rather is more significant in constituencies that, on average, had larger numbers of candidates prior to the change in laws. Columns two and four quantify this relationship. Constituencies that, on average, have one more candidate before the rule change experience about one less new candidate after the change in laws. This is consistent with the fact that these constituencies contain more candidates on average that will fail to obtain at least a sixth of the popular vote and will loose their deposits.

2.5.3 Variation Over Time

The first instrumental variables specification (IV1) uses a dummy variable for the period after the election as an instrument. The effect of this specification is to use only the portion of the variation in the number of new candidates that is correlated with the average number of new candidates before or after the rule change. Specifically, it only uses changes in the average number of new candidates over time to estimate the causal relationship between the number of new candidates and an existing candidate's probability of election.

Table 4 contains the results for this specification. Columns 1 and 3 contains the reduced form estimates for one and two elections before and after the change in election laws respectively. The average probability of election by existing candidates increases by three tenths of a percentage point after the change in election laws using both data sets. Columns 2 and 4 contain the IV estimates using the dummy variable to instrument for the number of new candidates. The implied effects are about 50 percent larger than the effects generated by the OLS correlation, but still very consistent.

2.5.4 Differences between Constituencies

There are two disadvantages to the preceding specification. First, it does not allow for an estimate of more than a linear function of the number of new candidates. Second, it does not control for possible year effects in the changes in the probability of existing candidates winning office. If this probability increased (decreased) for a reason unrelated to the changes in the number of new candidates, then this specification will attribute the change to the number of new candidates, over (under) estimating the coefficient on the number of new candidates.

To check for changes in the probability of election not caused by changes in the number of new candidates, I use the fact that the election laws had larger effects in constituencies with larger numbers of average candidates. This variation in the effects of the law allows for the use of between constituency variation in the number of new candidates rather than the average effects of the law over time. Column 1 of Table 5 shows the reduced form regressions. While not statistically significant, candidates in constituencies that average more candidates before the change in election laws have higher rates of election afterwards.

Columns 2 through 4 contain the results of specifications that use only between constituency variation after the change in election laws. In these specifications, the interaction between the dummy variable for the period after the change in deposit requirements and the average number of candidates is used as an instrument for the number of new candidates. To control for baseline levels of new candidates' electability, I include the average number of candidates in the second stage regression. Mathematically, this instrumentation strategy only uses the variation in the number of new candidates that both occurs after the change in election laws and is correlated with the average number of candidates in the constituency before the change in election laws. While this allows one to control for changes in the level of electability, the exclusion restrictions are more severe. In addition to accepting the exogeneity of the change in the election laws, we must also accept that the variation in the number of candidates correlated with the average number of candidates before the change is determined by the change in the electoral laws rather than the strength of the existing candidates in a given location.

The results across all specifications are consistent. Column 2 estimates the relationship using both the dummy for after the legal change and the average number of candidates as control variables in the second stage regression. While the coefficient cannot be statistically distinguished from zero, the point estimate is that a change in the number of new candidates reduces the probability of existing candidates being elected by three hundredths of a percent – identical to the estimate using only the average variation in the number of candidates over time. Absent the change in the number of new candidates, however, there does not seem to have been much change in the probability of existing candidates being elected – the coefficient on "After 1996" is small in magnitude and statistically insignificant. This allows us to improve the precision of the estimated coefficient on the number of new candidates by using the dummy variable as an instrument. Column 3 contains these results. While this increases the level of variation in the number of candidates by including the average number before and after the legal change, the estimate of the coefficient is identical to that in column 2 but now is statistically significant at the one percent level. Column 4 contains the estimate of the quadratic relationship. While the coefficient suggests the same declining effect of the number of candidates over time as the estimates in Table 2, the magnitude of the linear component is much smaller and consistent with the estimates of the linear specification. However, the coefficients on neither the linear nor the quadratic terms are statistically significant.

Finally, Columns 5 and 6 present the results of regressions using both the average number of candidates and the interaction with the period after 1996 dummy variable as instruments. These are identical to the estimations in columns 2 through 4 except that the average number of candidates is used as an instrument rather than as a control variable. Although the exclusion restrictions on this specification are severe – requiring one to accept that all variation in the number of new candidates correlated with the average number of candidates before the change in the law is unrelated to the strength of individual candidates – the point estimates are consistent with the previous estimates of the effects of additional new candidates.

2.5.5 Effect of the Number of Original Candidates

In the regression results presented above, I have implicitly assumed that the margin of victory has controlled for the strength of existing candidates. This is critical because otherwise the fact that the number of candidates is correlated over time will mechanically generated an inverse relationship between the probability of an existing candidate being elected and the number of new candidates that cannot be excluded using the previous instrumentation strategy. The problem is that only one candidate from each constituency can be elected to office. If one candidate wins, the other candidates necessarily lose. More candidates in a constituency force more candidates to lose, generating a negative correlation between the number of candidates in a constituency and the probability of a candidate winning election.

To check this, I estimate the results of the original ordinary least squares and the main two instrumental variables regressions using only the top five candidates in an election for the elections just before and after the change in election laws. Because this fixes the number of existing candidates per constituency, it eliminates the mechanical relationship described above from generating the observed inverse relationship. However, because there is a significant correlation between the rank of an original candidate and the probability of winning the next election, these candidates have a higher baseline probability of winning the next election, and may be more affected by the number of new candidates.

Table 6 contains the re-estimation of the model using the four basic specifications for the elections directly before and after the change in election laws. The estimated coefficients of the instrumental variables regressions are consistent with each other, estimating that the addition of one new candidate reduces the probability of one of the top existing candidates from winning by about seven to eight percent. As with the estimates using all of the candidates, these estimates are about twice the size of the OLS estimate of the model. Rather than reflecting the mechanical correlation that motivated the estimations, however, the larger coefficients suggest heterogeneity in the effects of the new candidates that is correlated with the strength of the existing candidates.

2.6 Conclusion

All of the estimates suggest that the addition of additional candidates for Indian state parliamentary elections had a very small effect on the electability of the average existing candidate. The addition of a new candidate reduces the probability that an existing candidate wins office by only three fourths of a percent. The probability of election falls from about four percentage points to only 3.7 percent. This suggests that the increase in deposit requirements had almost no effect on the average existing candidate. On average, the number of new candidates fell by about six candidates per constituency, increasing the probability of the average candidate being elected by only 0.18 percentage points.

The appropriate level of restrictions on candidates depends not only on the probability that the excluded candidates have an effect on the election but also the cost of allowing them to contest and the magnitude of the benefits derived from their small effects. These results suggest that without significant social benefits, however, relatively small administrative costs would justify the additional restrictions imposed in 1996. These results, however, consider only the effects on the average candidates. They also suggest that the impact of new candidates may be larger for stronger existing candidates. This means that, while not affecting the average existing candidates, new candidates could have larger impacts on specific groups of candidates.

2.7 References

Aleaz, Bonita. (1997). *Struggles of Indian Federalism*. Punthi-Pustak: Calcutta.

Austin-Smith, David and Jeffrey Banks (1989) "Electoral Accountability and Incumbency," *Models of Strategic Choice in Politics.* Ed. Peter C. Ordeshook. Ann Arbor: The University of Michigan Press.

Banks, Jeffrey S. and Rangarajan K. Sundaram (1993) "Adverse Selection and Moral Hazard in Repeated Elections Model" *Political Economy: Institutions, Competition, and Representation.* Ed. William A. Barnett, Melvin J. Hinich, and Normal J. Schofield. Cambridge: Cambridge University Press.

Besley, Timothy and Anne Case (1995), "Incumbent Behavior: Vote-Seeking, Tax-Setting, and Yardstick Competition," *American Economic Review,* 85(1): 25-45.

Case, Anne (2001), "Election Goals and Income Redistribution: Recent Evidence from Albania," *European Economic Review.* 45(2001) 405-423.

Chattopadhyay, Raghabendra and Esther Duflo (2003). "Women as Policy Makers: Evidence from a Randomized Policy Experiment in India," *NBER working paper #8615.*

Delimitation Commission of India (2003). "Delimitation." http://www.delimitation-india.com.

Election Commission of India (2003). http://www.eci.gov.in.

Indian-Elections.com (2004). "Election FAQ's". http://www.indian-elections.com.

Katuria, H. S. (1990). *President's Rule in India, 1967-1989.* Uppal Publishing House: Delhi.

Linden, Leigh (2004), "Are Incumbents Really Advantaged? The Preference for Non-Incumbents in Indian National Elections" *MIT Job Market Paper.*

NewKerala.com (2004). "More than 77 per cent were 'non-serious candidates' in Indian Elections, EC," May 15, 2004.

Noorani, A. G. (2000). *Constitutional Questions in India.* Oxford University Press: New Delhi.

Miguel, Edward and Farhan Zaidi (2003). "Do Politicians Reward their Supporters? Public Spending and Incumbency Advantage in Ghana," UC Berkeley Mimeo.

Pande, Rohini (2003) "Can Mandated Political Representation Provide Disadvantaged Minorities Policy Influence? Theory and Evidence from India," *American Economic Review.* 93(4): 1132-1151

Washington, Ebonya (2003) "How Black Candidates Affect Voter Turnout and Candidate Positioning" *MIT Job Market Paper*.

Figure 1: Average Number of New Candidates by State

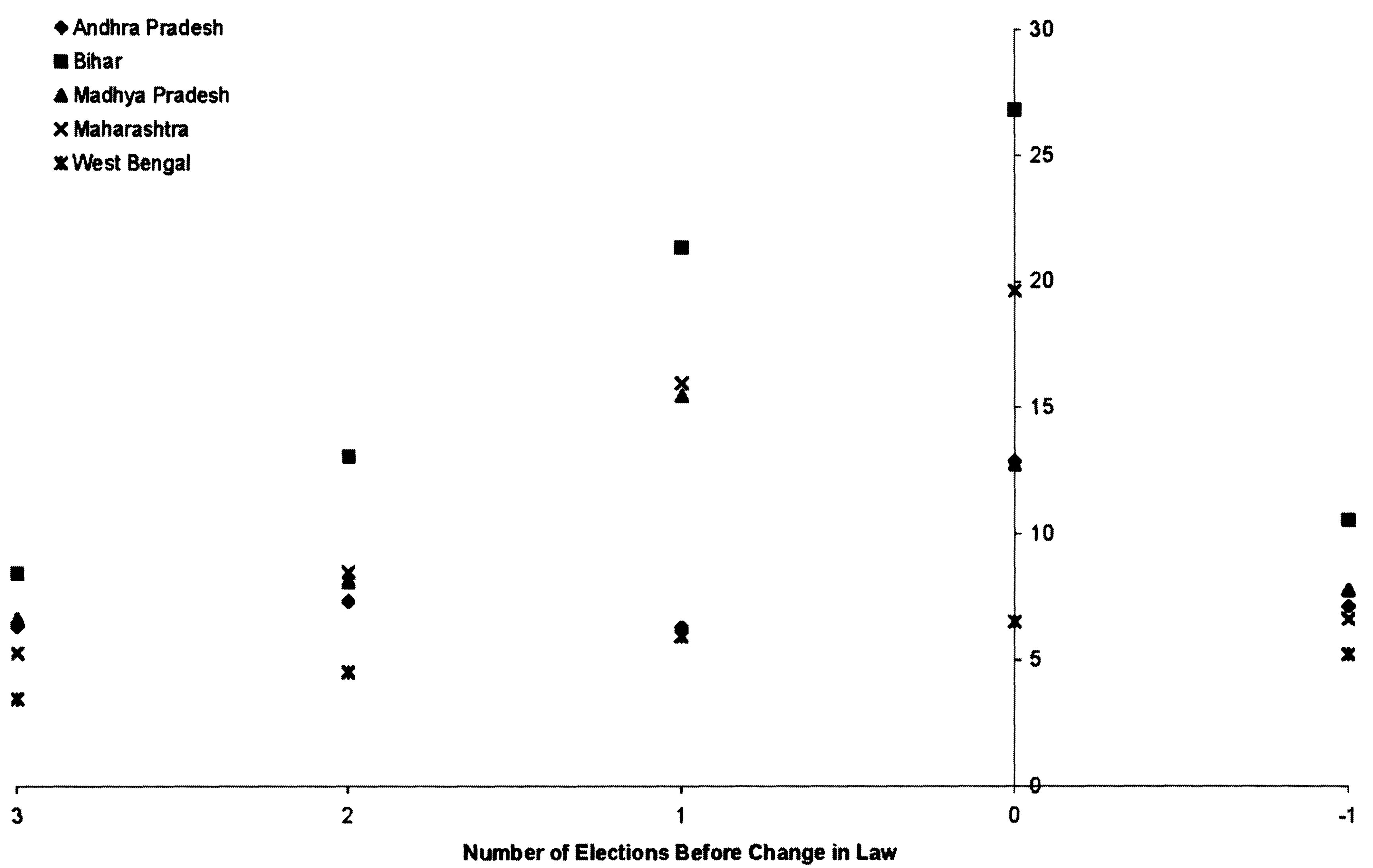

Figure 2: Distribution of Number of New Candidates

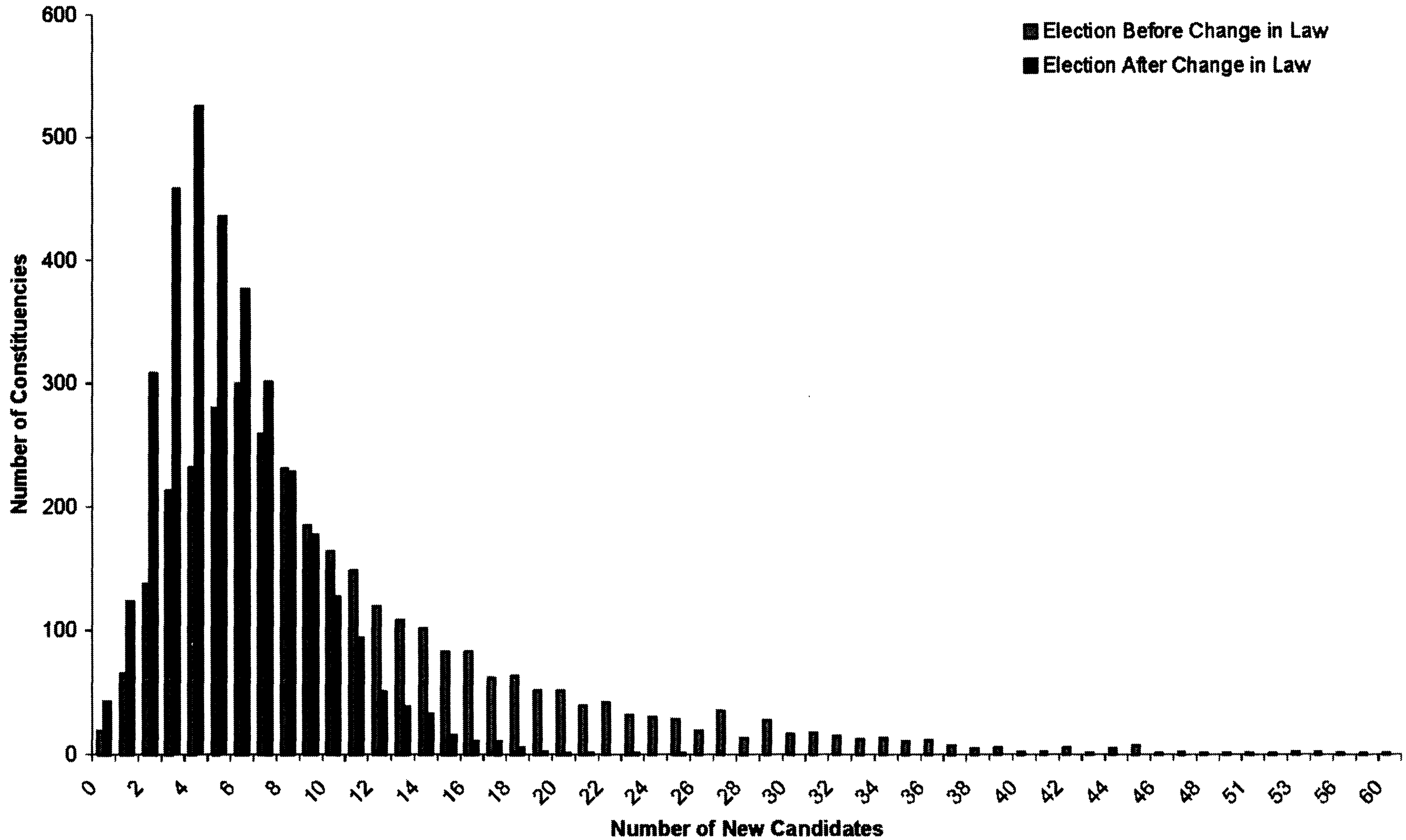

Figure 3: Difference in Number of New Candidates Immediately Before and After Change in Election Law

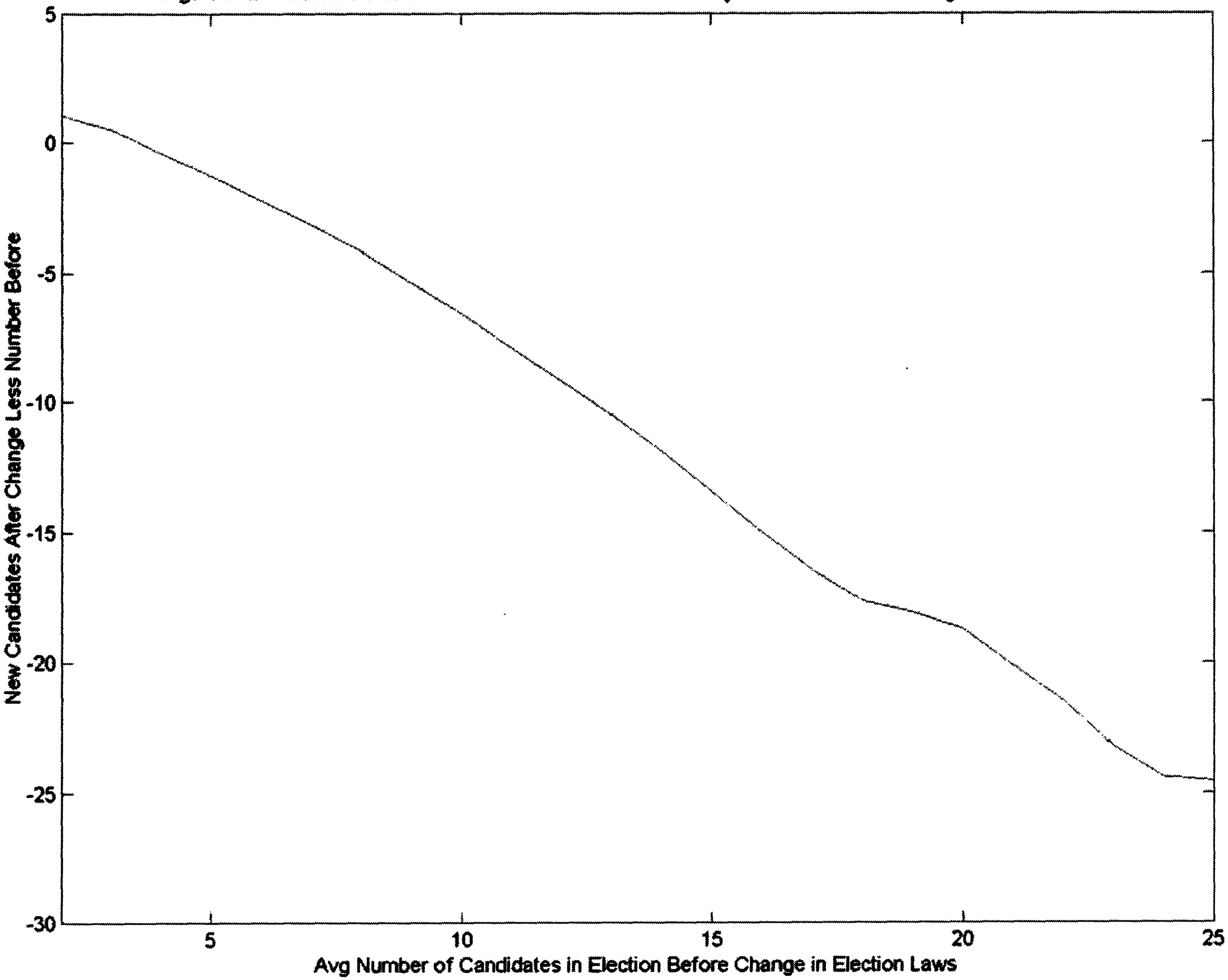

Figure 4: Probability of Election, State Elections 1977-2002

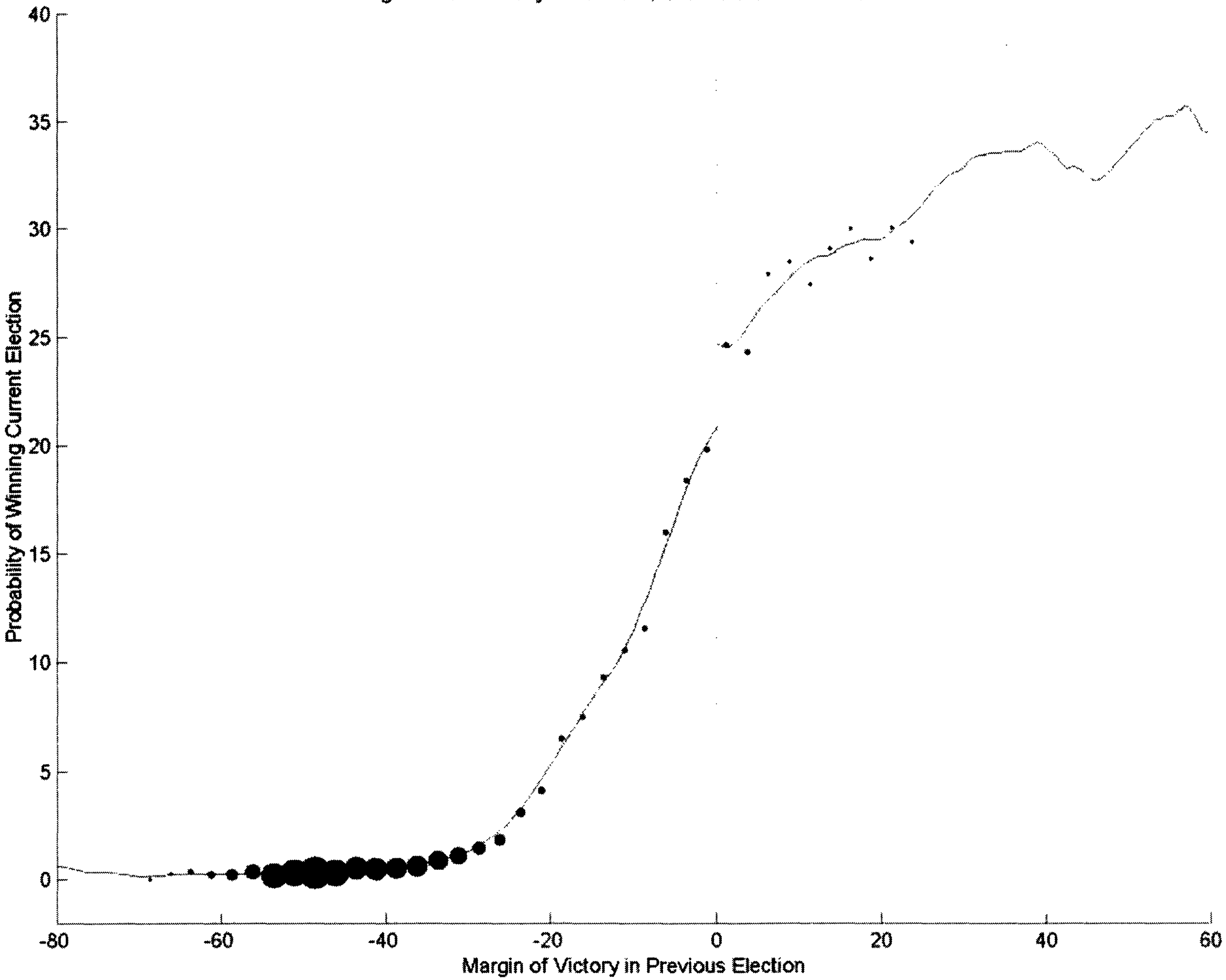

Table 1: Summary Statistics

	1 Election before and after Change in Law	2 Elections before and after Change in Law
Candidate Level Variables		
Won Election	4.09	4.56
	(19.8)	(20.9)
Margin of Victory	-35.7	-34.9
	(21.4)	(22.7)
Male	0.96	0.96
	(0.2)	(0.2)
Number of Observations	79,151	108,997
Constituency Level Variables		
Number of Candidates in First Period	12	10.4
	(15.8)	(13.3)
Number of New Candidates	8.43	8.4
	(14.3)	(12.6)
Number of Observations After 1996	3,322	3,923
Number of Constituencies	6,644	10,567

Table 2: OLS Estimates of Probability of Victory as a Function of the Number of New Candidates in Next Election

	1 Election Before and After Change		2 Elections Before and After Change	
	(1)	(2)	(3)	(4)
Specification	OLS	OLS	OLS	OLS
Number of New Candidates	-0.017	-0.076***	-0.021	-0.074***
	(0.012)	(0.008)	(0.013)	(0.008)
Number of New Candidates^2		0.00007***		0.00007***
		(0.000)		(0.000)
Margin of Victory	0.606***	0.605***	0.564***	0.563***
	(0.013)	(0.013)	(0.011)	(0.011)
Margin of Victory^2	0.0002	0.0002	0.0002	0.0002
	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Margin of Victory^3	-0.00008***	-0.00008***	-0.00007***	-0.00007***
	(0.000005)	(0.000005)	(0.000003)	(0.000003)
Male	0.973***	1.013***	1.385***	1.418***
	(0.349)	(0.349)	(0.320)	(0.320)
Constant	18.731***	19.152***	17.764***	18.096***
	(0.539)	(0.531)	(0.474)	(0.463)
Observations	79151	79151	108997	108997
R-squared	0.19	0.19	0.18	0.18

Standard errors are clustered at the constituency level

* significant at 10%; ** significant at 5%; *** significant at 1%

Table 3: First Stage, Determinants of Number of New Candidates

	1 Election Before and After Change		2 Elections Before and After Change	
	(1)	(2)	(3)	(4)
After 1996	-5.670***	1.18	-4.393***	1.928***
	(0.330)	(0.781)	(0.178)	(0.557)
After 1996 * Avg Num Cands		-1.143***		-1.065***
		(0.118)		(0.085)
Avg Num Cands		1.324***		1.213***
		(0.091)		(0.057)
Constant	10.742***	3.145***	8.365***	1.500***
	(0.288)	(0.775)	(0.194)	(0.509)
Observations	6644	6644	10567	10567
R-squared	0.11		0.13	

Standard errors are clustered at the constituency level
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 4: IV Regression of Probability of Victory using Post-1996 as an Instrument

	1 Election Before and After Change		2 Elections Before and After Change	
	(1)	(2)	(3)	(4)
Specification	OLS	IV	OLS	IV
After 1996	0.265**		0.283***	
	(0.122)		(0.110)	
Number of New Candidates		-0.029**		-0.039***
		(0.013)		(0.015)
Margin of Victory	0.606***	0.606***	0.565***	0.564***
	(0.013)	(0.013)	(0.011)	(0.011)
Margin of Victory^2	0	0	0	0
	0.000	0.000	0.000	0.000
Margin of Victory^3	-0.000***	-0.000***	-0.000***	-0.000***
	0.000	0.000	0.000	0.000
male	0.987***	0.973***	1.402***	1.386***
	(0.349)	(0.348)	(0.320)	(0.320)
Constant	18.413***	18.825***	17.483***	17.886***
	(0.533)	(0.536)	(0.466)	(0.467)
Observations	79151	79151	108997	108997
R-squared	0.19		0.18	

Standard errors are clustered at the constituency level
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 5: IV Regressions of Probability of Victory using Average Numver of Candidates before Change in Law as Instrument

		Avg Num Cands as Control			Avg Num Cands as Instrument		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Specification	OLS	IV1	IV1	IV1	IV2	IV2	IV2
After 1996	-0.011	-0.002			-0.168		
	(0.284)	(0.272)			(0.223)		
Number of New Candidates		-0.03	-0.029***	-0.031	-0.048**	-0.036***	-0.027
		(0.025)	(0.011)	(0.037)	(0.020)	(0.010)	(0.032)
Number of New Candidates^2				0.00002			-0.0001
				(0.0004)			(0.0004)
After 1996*Avg Number of Candidates	0.038						
	(0.033)						
Avg Number of Candidates	-0.072**	-0.029	-0.029	-0.117*			
	(0.030)	(0.030)	(0.025)	(0.070)			
Average Number of Candidates^2				0.004			
				(0.003)			
Margin of Victory	0.605***	0.605***	0.605***	0.605***	0.605***	0.606***	0.607***
	(0.013)	(0.013)	(0.013)	(0.014)	(0.013)	(0.013)	(0.014)
Margin of Victory^2	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Margin of Victory^3	-0.00008***	-0.00008***	-0.00008***	-0.00008***	-0.00008***	-0.00008***	-0.00008***
	(0.000005)	(0.000005)	(0.000005)	(0.000005)	(0.000005)	(0.000005)	(0.000005)
male	0.983***	0.970***	0.970***	0.985**	0.964***	0.973***	0.907**
	(0.349)	(0.348)	(0.348)	(0.426)	(0.348)	(0.348)	(0.409)
Constant	18.863***	18.991***	18.988***	19.338***	19.085***	18.873***	18.868***
	(0.572)	(0.604)	(0.543)	(0.622)	(0.600)	(0.531)	(0.532)
Observations	79151	79151	79151	79151	79151	79151	79151
R-squared	0.19	0.19	0.19	0.19	0.19	0.19	0.18

Standard errors are clustered at the constituency level
* significant at 10%; ** significant at 5%; *** significant at 1%

Table 6: Estimations Using Only the Top Five Candidates in Each Constituency

	(1)	(2)	(3)	(4)
Specification	OLS	IV1	IV2	IV3
Number of New Candidates	-0.041	-0.069*	-0.071**	-0.084***
	(0.031)	(0.039)	(0.034)	(0.032)
Average Number of Candidates			-0.046	
			(0.065)	
Margin of Victory	0.596***	0.596***	0.596***	0.596***
	(0.017)	(0.017)	(0.017)	(0.017)
Margin of Victory^2	-0.001**	-0.001**	-0.001**	-0.001**
	-0.0005	-0.0005	-0.0005	-0.0005
Margin of Victory^3	-0.00009***	-0.00009***	-0.00009***	-0.00009***
	-0.000008	-0.000008	-0.000008	-0.000008
Male	2.241***	2.206***	2.208***	2.188***
	(0.791)	(0.795)	(0.794)	(0.796)
Constant	18.860***	19.116***	19.395***	19.245***
	(0.980)	(1.011)	(1.025)	(0.989)
Observations	28985	28985	28985	28985
R-squared	0.15	0.15	0.15	0.15

Standard errors are clustered at the constituency level

* significant at 10%; ** significant at 5%; *** significant at 1%

Chapter 3: Remedying Education:
Evidence from Two Randomized Experiments in India

3.1 Introduction

There has been a lot of interest recently in the question of how to effectively deliver education to the poor in developing countries and a corresponding burgeoning of high quality research on the subject. A lot of the research focuses on the effects of reducing the cost of schooling, with the view that the important goal is to get the children into school. Examples of this kind of work include Banerjee, Jacob and Kremer (2002) on school meals in India, Duflo (2001) on school construction in Indonesia, Glewwe, Kremer and Moulin (1997) on school uniforms in Kenya, Spohr (1999) on compulsory schooling laws in Taiwan and Vermeersch (2002) on school meals for preschoolers in Kenya. The primary metric by which success is judged in these studies is attendance, and in each of these cases a significant impact was found.

Are students also learning measurably more as a result of these interventions? There is no obvious reason why they would. The influx of new students probably makes learning harder for the children who were already in school, simply because there are more demands on existing resources.[22] And while the newcomers will presumably learn more, just by the fact that they are now attending school, it is not clear that there is anyone with whom we could compare them.

At the other extreme are interventions that focus directly on improving test scores for students who are already in school. These are interventions where students are explicitly rewarded for doing well on tests: Angrist, et. al., (2002) study a program in Colombia that offers private school vouchers to students who keep their scores above a certain level. A recent study by Kremer, Miguel and Thornton (2002) looks at the impact of offering scholarships to students in Kenya who do well on a standardized test. Both studies find an impact on test scores, though in such cases the existence of an impact is perhaps less interesting than whether the gains are commensurate with the money spent.

Perhaps the most interesting case is the one in between: Interventions that purport to improve the quality of the learning experience, but for which no evidence exists that they actually do improve learning. Examples include increasing the teacher-student ratio (Banerjee, Jacob and Kremer, 2002), subsidized textbooks (Glewwe, Kremer and Moulin, 1997), free flip-charts (Glewwe, Kremer, Moulin and Zitzewitz, 1997); and then the interventions that improve the health of school children (for example, deworming, as in Kremer and Miguel, 2002), incentives for teachers (Glewwe, Kremer and Moulin, 2002), and blackboards and other school inputs (Chin, 2001), etc. By improving school quality, these programs can increase attendance. One ought also to expect an improvement in test scores among those who were already in school. Nevertheless, it is notable that relatively few of the studies from developing countries report a positive impact on test scores for those who were already in school.[23] Moreover, one cannot rule out the idea that there is no impact on children's educational achievement *a priori*, because the quality of teaching in many schools leaves much to be desired. Or it could even be the case that the children do not learn because they do not want to: The returns are just not high enough.

[22] Indeed this is what Banerjee, Jacob and Kremer (2002) find for mid-day meals, and Glewwe, Kremer and Moulin (1997) find for a program that offered both free textbooks and free school uniforms.

[23] The one exception we of we are aware is the study of a program that provides incentives for teachers in Kenya that is reported in Glewwe, Kremer and Moulin (2002b), though even in this case the authors seem to be somewhat disappointed by the lack of a more robust impact. Chin (2001) finds that Operation Blackboard in India did increase school completion rates for girls, which implies that there must have been an increase in test scores, but she cannot tell whether those who would have completed school in any case learn more as a result of the intervention. Vermeersch (2002) also finds an impact on test scores of a school meals program in schools where the teachers were trained, but she too cannot distinguish between those who were already in school and the newcomers.

This paper reports on a randomized evaluation of an intervention in urban India focused on improving the learning environment in public schools. The intervention is motivated by the belief that children often drop out because they fall behind and feel lost in class. The program, which is run Pratham, a Mumbai-based Non-Governmental Organization, provides remedial education, in small groups, to children that are lagging behind. To keep costs low and ensure a good instructor-student relationship, the program hires young women (the "Balsakhis") who have the equivalent of a high school degree from the local slum communities in which the schools are located.

The evaluation of the remedial education (Balsakhi) program offered an opportunity to implement an evaluation design that is often recommended but rarely, if ever, utilised.[24] First, it was a randomized evaluation. We can therefore be relatively confident of the absence of confounding factors.[25] Second, the program we study was run on a very large-scale (over 15,000 students were included in the study), and had already clearly demonstrated the ability to scale up in other cities, as the description below will make clear. In other words, there is no risk that what we are evaluating cannot be reproduced elsewhere. Third, we simultaneously carried out randomized evaluations of the program in two different cities, each of which had its own management team. This reinforces our confidence in the external validity of these results. Finally, we conducted the study over two years, using several tests, making it less likely that the results are a consequence of the newness of the program, or the effect of implementing an evaluation.

Finally, though we find no effect on attendance, we find that the program has a substantial positive effect on children's academic achievement. This impact is remarkably stable across years and cities, especially when we take into account the instability of the environment---there was a major riot and a catastrophic earthquake while the program was running. Moreover, the weaker students, who are the primary target of the program, gained the most. This study demonstrates both the efficacy of the remedial education program, and more generally, the feasibility of dramatically impacting test scores at very low cost. We also make an attempt to

[24] We are currently in the middle of a two-year evaluation of a Computer Assisted Learning program in Baroda whose research design is very similar to the one described here. However, we are unaware of any other evaluation of an educational program that meets the criteria listed below.

[25] In Baroda, the Computer Aided Learning (CAL) program was implemented in the second year of the Balaskhi study, for fourth standard children only. However, the experimental design randomly assigned the CALprogram to half of the balsakhi-treatment and half of the balsakhi-comparison schools, thus allowing us to estimate unbiased balsakhi-only effects. The results of the CAL program will be reported once the study is complete.

distinguish the direct effect of the program (on children who worked with the balsakhi) and the indirect (on those who did not). The estimates suggest that the reducing class size by hiring a balsakhi is at least twice as effective as reducing class size by keeping children with regular teachers.

3.2 The Balsakhi Program

Pratham was established in Mumbai in 1994, with support from UNICEF, and has since expanded to several other cities in India. Pratham now reaches over 220,000 children in 34 cities in India, and employs about 10,000 individuals. Pratham works closely with the government: Most of its programs are conducted in the municipal schools, and Pratham also provides technical assistance to the government.

One of Pratham's core programs is a remedial education program, called the Balsakhi program. This program, in place in many municipal schools, provides a teacher (usually a young woman, recruited from the local community, who has herself finished secondary school) to work with children identified as falling behind their peers. While the exact details vary according to local conditions, the typical instructor meets with a group of approximately 15-20 children in the morning for two hours, and with another group of the same size in the afternoon. Instruction focuses on the core competencies the children should have learned in the second and third standards, primarily basic numeracy and literacy skills. The instructors are provided with a standardized curriculum that was developed by Pratham. They receive two weeks of training at the beginning of the year and ongoing reinforcement while school is in session. The program has been implemented in twenty Indian cities, reaching tens of thousands of students. It was started in Mumbai in 1994, and expanded to Vadodara in 1999.

According to Pratham, the main benefit of the program is to provide individualized, non-threatening attention to children who are lagging behind in the classroom and are not capable of following the standard curriculum. Children may feel more comfortable with women from their own communities than teachers, who are often from different backgrounds. As the balsakhi's class size is relatively small, she may tailor the curriculum to the children's specific needs. Furthermore, because Pratham's program takes children out of the classroom, it may even benefit children who were not directly targeted by the intervention. Removing children from the classroom for two hours means the effective student-teacher ratio in the main classroom drops, and the teacher may be able to focus on more advanced material. Finally, if the balsakhis are indeed effective, even when the children are returned to the main classroom, the teacher may not need to keep re-teaching remedial material.

An important characteristic of this program is the ease which with it can be scaled up. Because Pratham relies on local personnel, trained for a short period of time, the program is very low-cost (each teacher is paid 500-750 rupees, or 10-15 dollars, per month) and is easily replicated. There is rapid turnover among the balsakhis (each of them staying for an average of one year, typically until they get married or get another job), indicating that the success of the program does not depend on a handful of very determined and enthusiastic individuals. Finally, since the balsakhis use whatever space is available (free classrooms, or even hallways when necessary), the program has very low overhead and capital costs.

3.3 Evaluation Design

3.3.1 Sample: Vadodara

In 2000, when Pratham decided to expand their remedial education (balsakhi) program to cover the entire city of Vadodara, they decided to take advantage of the expansion to evaluate the effectiveness of the program in the remaining 98 eligible schools in the city. In November, 2000, they administered an academic test (designed by the Pratham team) to all children in the third standard. They then hired and trained balsakhis, which were sent to half of the schools in Vadodara. Assignment was random, with schools stratified by medium of instruction, gender, and pupil-teacher ratios. Unfortunately, the school year was disrupted by an earthquake in Gujarat, and children received only a few weeks of instruction between November and March. This year of the program is best understood as a pilot program.[26]

In July, 2001, the group of schools that had received a balsakhi in the previous year of the program received the balsakhi in the fourth standard, and the remaining schools received a

[26] Throughout the paper, we will refer to the year 2001-2002 as "year 1" and 2002-2003 as "year 2."

balsakhi in the third standard. Children in the standard that did not receive the balsakhi in a given grade form the comparison group for children who did receive the balsakhi.

The program was continued during the school year 2002-2003, with the addition of the 25 remaining primary schools. Schools where the balsakhi was assigned in standard three in the year 2001-2002 were now assigned a balsakhi in standard four, so that in year 2, standard 4 children in the treatment group benefitted from two years of the balsakhi program. Schools where the balsakhi was assigned in standard four in the year 1 received balsakhi assistance for standard three in year 2. The new schools were randomly assigned to either group with equal probability in the same way that the original schools were assigned. The number of schools and divisions in the two groups are given in Table 1.

3.3.2 Sample: Mumbai

To ensure the results from the Vadodara study would be generalizable, the Balsakhi program in Mumbai was also evaluated, in 2001-2002 and 2002-2003. Mumbai was Pratham's birthplace, and Pratham is currently operating various programs throughout the city. We selected one ward (the L-ward) to implement a design similar to the design in Vadodara, including all Gujarati, Hindi, and Marathi schools. In total, 62 schools are included in the study. Schools were stratified according to their scores in a pre-test, as well as by the medium of instruction. Half the schools were randomly selected to receive a balsakhi in standard two, and half the schools were randomly selected to receive a balsakhi in standard three. In 2001-2002, data were collected only for standard three children, while in 2002-2003, data were collected for standards three and four. As in Vadodara, children kept their treatment assignment status as the moved from standard two to three (or three to four).

In the second year of the study, the Mumbai program experienced some administrative difficulties. A decision to require balsakhis to pass a competency test resulted in the firing of many balsakhis. Hiring new recruits was complicated by the fact that the administrative staff in L-Ward turned over between year 1 and year 2, and the new staff lacked community contacts necessary for recruitment. Finally, the principals of a couple of schools, hearing that the study was being conducted by a group of Americans, refused balsakhis. Thus, only two thirds of the schools assigned Balsakhis actually received them. (Schools could not refuse testing, because

Pratham had obtained written permission for testing from the city administration). Throughout the paper, the schools that were assigned balsakhis but did not get them are included in the treatment group.

3.3.3 Outcomes

The main outcome of interest is whether the interventions resulted in any improvement in cognitive skills. In the Vadodara pilot year, children were given a pretest in November, 2000, and post-test in March, 2001. In the first full year, the Vadodara pretest was at the beginning of the school year (August 2001), and the post test in March 2002. In the second full year, children were tested at the beginning of the school year (August 2002), in November 2002, and again in March, 2003. In the first year in Mumbai, children were tested in October, 2001 and March, 2002; in the second year tests were given in August, 2002, and February 2003.

In Vadodara, the same test is used for standard three and four children, so that the scores can be directly compared across grades. Scores on the pre- and post-test can also be directly compared, as the format of the questions and the competencies tested remain the same. The exam comprises two parts: A math section and a language section. In Vadodara, both parts focused on competencies that the Vadodara Municipal Corporation (VMC) prescribe for children in standards one through four. On the math exam, for example, tasks ranged from basic number recognition, counting, and ordering of single digit numbers to ordering of two digit numbers, addition of single and two digit numbers, and basic word problems. Tests were similar in Mumbai. In the first year, tests focused on competencies in standards one through three, while in the second year they included standards one through four. In the second year, the same test was used for third and fourth standard children.

The "pilot" year of the program (2000-2001) allowed Pratham to make significant progress in developing a testing instrument (the initial test was too difficult) and effective testing procedures to prevent cheating and exam anxiety. The test was administered in both cities by Pratham, with the authorization of the municipal corporation. At least three Pratham employees were present in the classroom during each test to minimize cheating. [27]To minimize attrition,

[27] In Mumbai, since administration of the pre-test was less than satisfactory at the first attempt, we conducted a second pre-test, which we use as the basis for the analysis.

Pratham returns to the schools multiple times, and children who still failed to appear and who could be tracked down were administered a make-up test outside of school.

In Vadodara, the school year 2001-2002 was disturbed by massive inter-communal riots following an attack on a train carrying Hindus. Although a post-test was conducted in March (after the riots had receded), attrition was high. We thus use the October 2001 mid-test as the "post-test" for year one in Vadodara.[28] The year one pretests were in August (Vadodara) and September (Mumbai) 2001, while the Mumbai year one post-test was in March 2001. In year 2, the pretests were in August 2002 (both cities), and the post-tests in February (Bombay) and March (Vadodara) 2003.

Another outcome of interest is attendance and school dropout rates, which are collected weekly by Pratham employees, who made randomly timed weekly appearances in classrooms to take attendance. (Data from the official rolls was also collected, but administrators have incentives to inflate the attendance data).[29]

Finally, in the second year of the program, in both cities, data were collected on which specific children were sent to the Balsakhi. (Balsakhis work with, on average, about 20 children per school).

3.3.4 Statistical Framework

Effect of the Balsakhi Program on test scores

Given the randomized allocation of both programs, we expect the 2001 pre-test results in the treatment schools to be similar between those in the control. The results of the 2002 pre-test may be different in the treatment and control schools in standard four in Vadodara, as well as standard three and four in Mumbai, since they may reflect long-lasting benefits of the previous year's program for the children who were in the same school in the previous year. In both cities, the experimental design (in which each school was both in the treatment and comparison group, with one standard in each group) is such that even if a "good school" were in the treatment group for a

[28] The results of the first year of the program do not significantly change if we use the mid- or post-test: There was no further improvement (or deterioration) of the performance in the treatment schools relative to the control schools between the mid-test and the post-test.
[29] We report results from Mumbai in this draft. In Mumbai 2001-2002, teachers in some schools often refused to let the research assistants count the number of children present, resulting in biased data.

given standard, the other standard of that "good school" would be in the comparison group, ensuring that the averages across the standard are likely to be very similar.

Denoting y_{igjk} the test score of child i in grade g in school j in test k (k is either "PRE" or "POST"), we start by comparing test scores in the treatment and comparison schools, in each city and standard. First, we check that there is no difference between treatment and control schools before the program was run:

$$y_{igjPRE} = \alpha + \beta D_{jg} + \varepsilon_{igjPRE}, \tag{3.1}$$

where D_{jg} is a dummy indicating whether school j is in the treatment group in that particular year in standard g, and ε_{igjPRE} the error term.

This regression is run separately in each standard, year and city. It is run separately for the math exam, the verbal exam, and the total score on the exam. The standard errors are clustered at the school level.[30]

We then run the same regression in the post-period ($k = POST$):

$$y_{igjPOST} = \alpha + \beta D_{jg} + \varepsilon_{igjPOST}. \tag{3.2}$$

This provides a first estimate of the effect of being assigned to the treatment group. For all cities and year, except for Mumbai in year 2, this is also an estimate of the average effect of being a student in a school that was assigned a balsakhi. However, in Mumbai in year 2, because not all schools received a balsakhi (and not all classes within schools where treated), to obtain the average effect of receiving a balsakhi, we use the assignment to the treatment group (D_{jg}) as an instrument for whether or not the class of a specific child actually received the balsakhi (B_{jg}). In practice, we are estimate the following equation:

$$y_{igjPOST} = \alpha + \beta B_{jg} + \varepsilon_{igjPOST}. \tag{3.3}$$

[30] If we instead use a nested random effects model (with a classroom effect nested within a school effect), the point estimates are very similar, and the the standard errors are smaller. Clustering is a more conservative approach.

The first stage is the equation for whether or not a child's class was actually assigned a balsakhi:

$$B_{jg} = \alpha_1 + \delta_1 D_{jg} + \eta_{igjPOST} \,. \tag{3.4}$$

Because tests scores are very strongly auto-correlated, the precision of the estimate is increased by controlling for the child's test score in the pre-test.

We do so using the following difference in difference specification:

$$y_{igjk} = \lambda + \delta B_{jg} + \theta POST_k + \gamma(B_{ig} * POST_k) + \varepsilon_{igjk} \,, \tag{3.5}$$

where $POST_k$ is a dummy indicating whether the test is the post test. For all years and samples except Mumbai in year 2 $B_{jg} = D_{jg}$, and equation (3.5) is estimated with OLS. However, for Mumbai in year two (and when both cities are pooled), equation (3.5) is estimated with instrumental variables, with D_{jg} (the initial assignment to the treatment group), $POST_k$, and $D_{jg} * POST_k$ used as instruments.

We also present an alternative way to estimate the treatment effect in Mumbai, as a specification check. Since every school was supposed to receive a balsakhi in either standard 3 or standard 4, we keep in the sample only the schools that did receive a balsakhi. This means that a school will not be in the comparison school for one standard if the other standard did not receive a balsakhi. In this reduced sample, B_{jg} is equal to D_{jg}, and equation (3.5) is estimated by OLS. The assumption underlying this specification is that the characteristics that make the school more likely to have a balsakhi have the same influence on the test scores of children in standard 3 and standard 4.

To gain more insight about the impact of the program, we also present estimates of specifications similar to equations (3.3) and (3.5) using for y_{igjk} a binary variable indicating whether the child correctly answered the questions indicating competencies for standard 1, 2 and 3, respectively. Finally, we estimate the impact of being in the program for 2 years (for children who were in the treatment group in standard 3 in year 1, and whom the balsakhi has followed in year 2 when they moved to standard 4), by estimating equation (3.5) using the pre-test of year 1 as the pre-test, and the post test of year 2 as the post test.

Disentangling Class size and balsakhi effect

Estimating equations (3.3) and (3.5) generates estimates of the average impact of the program on all children who whose standard-school received a balsakhi. The program may impact the children in a treated school in two ways: directly, for children who were assigned to work with the balsakhi, or indirectly, because the weakest children are removed from the classroom for part of the day. This indirect effect can potentially work through two mechanisms: through a reduced number of students in the class (class size effect), and through the higher average quality of their classmates (tracking effect).

To separate the balsakhi and the indirect effects, an ideal experiment would have identified the children who would work with the balsakhi in all schools, *before* randomly assigning treatment and comparison groups (and to not allow substitution after the initial allocation). The balsakhi effect could then be estimated by comparing children at risk of working with the balsakhi in the treatment and the comparison group. The indirect effect would have been estimated by comparing the children who were not at risk of working with the balsakhi in the treatment and the comparison group. Unfortunately, this design was not practical in this setting.

We do know, however, that the assignment to the balsakhi group was based in part on pre-test score, and that a maximum of twenty children per school in Vadodara, and twenty per class in Mumbai were assigned to a balsakhi. In schools in the treatment group, we start by predicting assignment to the balsakhi as a function of the number of students (in the school in Vadodara, in the class in Mumbai), the sum of the math and verbal score at the pre-test, and a variable indicating whether the child is among the bottom 20 children in his group.

$$P_{ijg} = \pi_1 + \pi_2 S_{ijg} + \pi_3 y_{ijgPRE} + \pi_4 R_{ijg} + \pi_5 Z_{ijg} + \omega_{ijg} \tag{3.6}$$

where S_{igj} is the number of student in the class or the school, y_{ijgPRE} is the score of the child at the pre-test, R_{ijg} is the rank of the child in the class (starting from the bottom), and Z_{ijg} is a dummy indicating whether the child is among the bottom 20 children in the class. We will show that, even after controlling linearly for the class rank, the dummy Z_{ijg} predicts whether or not the child was assigned to the balsakhi.

Denoting X_{ijg} the vector $[S_{ijg} \quad y_{ijgPRE} \quad R_{ijg}]$, the following equation (which interacts the variables in equation (3.6) with a dummy for whether the child is in the balsakhi group) predicts assignment to the balsakhi in the whole sample.

$$P_{ijg} = \alpha + \gamma D_{ijg} + \beta(Z_{ijg} * D_{ijg}) + \mu Z_{ijg} + X'_{ijg}\kappa + \lambda(X'_{ijg} * D_{ijg}) + \varepsilon_{ijg} \tag{3.7}$$

We can then regress the post test scores on the same variables, and examine whether being one of the bottom 20 children is associated with a bigger effect for those in the balsakhi group:

$$y_{ijgPOST} - y_{ijgPRE} = \alpha + \beta Z_{ijg} * D_{ijg} + \gamma D_{ijg} + \mu Z_{ijg} + X'_{ijg}\kappa + X'_{ijg} * D_{ijg}\lambda + \varepsilon_{ijg} \tag{3.8}$$

Equation (3.7) and (3.8) form the first stage and the reduced form of a instrumental variables estimation of the following equation:

$$y_{ijgPOST} - y_{ijgPRE} = \alpha + \beta P_{ijg} + \gamma B_{ijg} + \mu Z_{ijg} + X'_{ijg}\kappa + X'_{ijg} * D_{ijg}\lambda + \varepsilon_{ijg} \tag{3.9}$$

where D_{ijg} and $Z_{ijg} * D_{ijg}$ are the excluded instruments. The identification assumption underlying this estimation strategy is that the only reason why the treatment effect varies with the variable Z_{ijg} is because Z_{ijg} makes it more likely that the child is sent to the balsakhi group. However, the effect of the treatment is allowed to vary with class size, the test score, and the rank of the child. We also estimate an alternative specification which controls for a fourth-order polynomial in the rank of the child. In this equation, the effect of being assigned to the balsakhi group is given by $\beta + \gamma$, and the effect of being in a balsakhi school, but not assigned to the balsakhi group, is given by γ.

3.3.5 Disentangling tracking and class size effects

The effect of the program on children who were not assigned to the balsakhi can be further attributed to either a class size effect or a tracking effect: by removing the 20 weakest children in the class, the program might allow the teacher to spend more time on more advanced topics, or less time reviewing or disciplining children who do not follow the class.

To gain some insight on whether the two effects can be separated, we estimate an equation where we interact the treatment dummy with the average pre-test score of children who were sent to the balsakhi: if the test scores of children who go to the Balskahi are lower, the tracking benefits to the remaining children should be bigger. To instrument for the test score of the children who went to the Balsakhi, we use the average test score of the bottom twenty children in the class. The reduced form equation is thus:

$$y_{ijgPOST} = \alpha + \beta Z_{ijg} * D_{ijg} + \gamma D_{ijg} + \delta D_{ijg} * T20_{jg} + \mu Z_{ijg} + X'_{ijg}\kappa + X'_{ijg} * D_{ijg}\lambda + \varepsilon_{ijg} \qquad (3.10)$$

where $T20_{jg}$ is the average test score among the bottom twenty children of school j. The structural equation is:

$$y_{ijgPOST} = \alpha + \beta P_{ijg} + \gamma B_{ijg} + \delta B_{ijg} * TB_{jg} + \mu Z_{ijg} + X'_{ijg}\kappa + X'_{ijg} * D_{ijg}\lambda + \varepsilon_{ijg} \qquad (3.11)$$

where TB_{jg} is the average score among children who are assigned to the balsakhi.

3.4 Results

3.4.1 Descriptive Statistics: Level of Competencies and Pre-intervention Differences

Tables 2 through 5 present the descriptive statistics of the test scores for all samples used in this analysis (year 1 and 2 in Vadodara and Mumbai). The scores are presented in three different ways: percentage of points scored, normalized scores (relative to the distribution of the pre-test score in the comparison group in each city and year),[31] and as the percentage of children correctly answering the questions in the test relating to the competencies in each standard.

The randomization appears to have been successful: Neither in Mumbai nor in Vadodara are there any large or systematic differences between the pre-test score and the post-test score. None of the differences between the groups prior to the implementation of the program are significant.

The raw scores, and the percentage of children correctly answering the questions relating to the curriculum in each standard give an idea of how little these children actually know. In standard three in Vadodara in the second year, for example, the average student in math scores about 16%, both in the control and treatment groups. Since one math question is multiple-choice, on average a student who knows nothing will score 1.8% points. If a student can consistently order two numbers and add two single digit numbers, she earns the additional 14% needed to achieve the average third standard performance. Only 5.4% of third standard children in Vadodara pass the standard 1 competencies in maths in standard 3 in Vadodara (and 14% in Mumbai). Standard one competencies cover number recognition, counting and one digit addition and subtraction.

The results are more encouraging in verbal competencies: 50% of the standard 3 children pass the standard 1 competencies in Vadodara (reading a single word, choosing the right spelling among different possible spelling for a word), and 65% do so in Mumbai.

[31] Scores are normalized for each standard, year, and city, such that the mean and standard deviation of the comparision group is zero and one, respectively. (We subtract the mean of the control group in the pre-test, and divide by the standard deviation.) This allows for comparison across samples, as well with results from other studies.

3.4.2 Attrition

Table 6 presents the levels of attrition in Mumbai and Vadodara. We present attrition that occurred between the pre-test and post-test for both cities in both years, as well as the two-year attrition (in Mumbai, for standard 4 only), broken down by treatment status. Attrition was generally very low, except for Vadodara in year 1. This is likely attributable to the civil unrest. The post-test was conducted after the riots, and the research team attempted to track down all of the children who did not appear for the exam. Attrition rates are not different in the comparison group than in the treatment group: In year 1 in Vadodara, attrition was 19% in the balsakhi treatment group, and 18% in the comparison group. In year 2, attrition was 4% in the balsakhi in both the treatment and the comparison group. In Mumbai year 1, attrition was 7% in the treatment group, and 7.5% in the comparison group, while in year 2 it was 7.7% in the treatment group and 7.3% in the comparison group.

The fact that there was no differential attrition rate in the treatment and control groups suggests that the estimate of the treatment effects will not be biased, unless different types of people drop out from the sample in the treatment and the control groups (Angrist, 1995). This does not seem to occur in our study: The second row in each panel presents the difference between the score at the pre-test of children who were not present at the post-test, by treatment status. The third column of each sample group present the differences-in-differences in the treatment and comparison groups. Children who will eventually leave the sample tend to be at the bottom of the distribution of the pre-test scores. However, the difference is very similar in the treatment and control groups, in both years and for both programs. In Mumbai in the second year, the difference-in-difference is almost statistically significant (p-value .067), with the attritors in the treatment group seeming to perform relatively better than those in the comparison group; this could be because the program encouraged weaker children to stay in school, making it more likely that they could be tracked down for the post-test. If anything, this would bias the estimates of the treatment effect downward.

Finally, both the attrition and the difference in test scores are also similar among the bottom 20 children in each school, the group of children who were the most likely to be assigned to a balsakhi (these results are not reported).

3.4.3 Effects of the Balsakhi Program

Effect on attendance

Part of the goal of the program was to make it easier for parents to play a role in their children's education, by serving as an intermediary between parents and the school environment. One could therefore have expected an impact of the program on attendance. In practice, there does not seem to be any: table 11 show two estimates of the program on attendance: a simple difference (comparing average attendance in the treatment and the comparison schools) and a difference in difference estimate (where we compare the change in attendance between the month of September-October and January-February in the treatment and the comparison group). Using either specification (and either the official roster data or data collected by our assistant) we do not find any impact of the program on children's attendance.

Overall effect on test score

Tables 2, 3, 4, and 5 present the first estimates of the effect of the balsakhi program, as simple differences between the post-test scores in the treatment and control groups. In all years and standards, for both tests, and in both cities, and for all subgroups, the difference in post-test scores between treatment and control groups is positive. In the first year in Vadodara (table 2), the difference in post-test score between treatment and control groups was 0.17 standard deviations in standard three for math, .15 in standard 3 for language, and .14 and .06 in standard 4, for math and language respectively.[32] The results in Mumbai (table 4) are remarkably similar, with the math and language test scores improving by 0.15 and 0.16 standard deviation, respectively.

In the second year of the program, the effects are larger: In Vadodara (table 3), the difference in total test scores is .44 for math and 0.25 for language in standard three, and .33 and 0.30 in standard four, for math and language respectively. In Mumbai in year two (table 5), the IV estimate of the impact of the program on test scores differences in are .25 and .09 in standard

[32] Throughout the paper, test results, differences, and differences-in-differences are presented in terms of standard deviations, unless otherwise specified.

3 (for math and language respectively) an .43 and .17 in standard 4 (for math and language respectively). In year two in Vadodara, all of the differences between treatment and control groups are statistically significant, while for Mumbai, the standard four results are significant.

Because test scores have a strong persistent component, the precision of these estimates can be improved significantly, however, by turning to a differences-in-differences specification (equation (3.5)). Since the randomization appeared to be successful, and attrition was low in both the treatment and comparison groups, the point estimates should be similar in the simple differences and the differences-in-differences specification. Table 7 presents differences-in-differences estimates of the effects of the balsakhi program, in various years, cities, standards, and sub-groups. For Mumbai in year 2, we estimate the treatment effect in two ways: first, we instrument for the dummy indicating whether or not the school received a balsakhi with a dummy for whether the school was assigned to the treatment group; second, we include only schools that got a balsakhi in at least one standard in the sample. The estimates using either specification are very similar.

As expected, the point estimates suggest a substantial treatment effect, and the standard errors are lower than the simple differences. Pulling both cities and standard together (in the first two rows of table 7), the impact of the program was 0.14 standard deviation overall in the first year, and 0.25 standard deviation in the second year. Both estimates for total score are significant at the 99% confidence level.

The impact is bigger in the second year, and bigger for math than for language in both years (0.19 standard deviations versus 0.069 in the first year, and 0.32 versus 0.15 standard deviations in the second year; all but first-year verbal scores are significant at the 99% level.) For both years and both subjects, the effect is larger in Vadodara than in Mumbai (with a total-score effect of 0.16 standard deviation versus 0.15 in the first year (standard 3 only), and 0.31 versus 0.15 in the second year (both standards)). The difference is the strongest for language, where there is a significant impact in both years for Vadodara (0.11 and 0.23 standard deviation respectively), but no significant impact in either year in Mumbai (0.06 standard deviation for standard 3 in year 1, and 0.032 standard deviation in year 2). For both cities and both subjects, the effects are very similar in standard 3 and standard 4. We also computed all those estimates for both genders separately, and found the impact to be very similar (results not reported).

In the last panel of the table, we display our estimate of the impact of the program for two years (for children who were in a treatment school in standard 3, and stayed in the treatment school). First, it appears that the effect of the first year does not seem to persist over the summer: at the pre-test in year 2, children who were in a treatment class in year 1 do not seem to know more than children who weren't. However, the effect of two years of treatment (from year1 pre-test score to year 2 post-test score) is substantially larger than that of either individual year (0.52 standard deviation in math, for example, versus 0.38 for one year): it seems likely that the foundation laid in the first year of the program helped the children benefit from the second year of the program.

Compared to the other educational interventions, this program thus appear to be quite effective. The Tennessee STAR experiment, for example, where class size was reduced by 7 to 8 children (from 22 to about 15), improved test scores by about 0.21 standard deviation. This program improved test scores by 0.25 standard deviations in the second year, by reducing effective class size from 35 to 20 children on average (averaging over the balsakhi and the non-balsakhi group) for part of the day, but doing so by hiring an assistant paid a fraction of the teacher's salary.

Effect on specific competencies

The test was designed to cover competencies from standard 1 to 3. In Mumbai in year 2, it also covered some standard 4 competencies. Table 8 offers more details on the level at which the program was effective. Estimates in this table suggest that, for math, the biggest effect was on the competencies of standard 1: in Vadodara for example, the program increased the fraction of children who mastered the competencies of the first standard in math by 5.0% in the first year, and 6.4% in the second year. In Mumbai the effect was 5.8% and 8.8% respectively. The effect on the fraction of children demonstrating knowledge of standard 2 and 3 competencies is much smaller. In language, the most important effect seems to be to help children master the competencies of standard 2 This may not be surprising, since many children seemed to have already mastered the competencies of standard 1. The effect of the program may thus be the strongest on the easiest competencies not already mastered by many pupils. These results correspond well with the stated role of the program, which was to work with children on basic competencies.

Distributional effects

The balsakhi program was primarily intended to help children at the lower end of the ability distribution, by providing especially targeted instruction. However, as we already mentioned, it could still help the higher scoring children, either because they are assigned to the balsakhi, or because they benefit from smaller classes when their classmates are with the balsakhi. In table 9, we directly ask whom the program benefitted, by breaking down the sample into thirds in the distribution of initial test scores. The effect of the program is indeed the strongest for children in the bottom of the pre-test score distribution. Taking both cities together, the impact in the first year on overall test scores was 0.21 in the bottom third, 0.12 in the middle third, and 0.08 in the top third. In the second year, the impact was respectively 0.32, 0.18 and -0.003. Children in the bottom of the distribution were also much more likely to see the balsakhi: 22% from the bottom third, compared to 16% from the middle third and 6% from the top third were sent to the balsakhi. By increasing scores of the lowest-scoring children more than their peers, the program has an equalizing effect on pupil achievement. It is worth noting that this occurs even in the Indian education system, which, like many in developing countries, places particular emphasis on the children who perform well The effect of the program on the bottom third of the distribution are impressive: in the second year, the math test scores increased by 0.46 standard deviation in Vadodara, and 0.51 in Mumbai. The language test score increased by 0.31 standard deviation in Vadodara, and 0.13 in Mumbai.

3.4.4 Inside the box: Balsakhi, class size, and tracking effects

These results lead to our next question: to what extent is the program effect due to a direct effect of the balsakhi teacher (affecting only the children who got assigned to the balsakhi group) and to an indirect effect, affecting children who were *not* assigned to the balsakhi group. The fact that both the program impact and the probability of being assigned to a balsakhi declines with a child's position in the test score distribution suggest that the impact of the program may have been larger for those who were actually assigned to the balsakhi. However, an alternative plausible explanation for this patter is that the direct (or indirect) effects of the test score are lower for children with higher test scores. This question therefore necessitates further investigation.

As we explained above, we propose using a dummy for whether a child belongs to the bottom 20 children of a class as an instrument for whether he is assigned to the balsakhi group. Columns 1 to 4 in table 10 show that in both Mumbai and in Vadodara, a dummy for whether a child belongs to the bottom 20 in his class predicts his assignment to the balsakhi, even after controlling for her rank, her score at the pre-test, and the number of students in her class (which are all negatively and significantly associated with assignment to the balsakhi). Not surprisingly, because some schools in Bombay were not assigned a balsakhi, all coefficients are smaller. In column 4, we include a fourth-order polynomial of rank (interacted with treatment); the dummy for child rank below 20 remains a significant predictor of assignment to the balsakhi group. In columns 5 to 8, we present the reduced form estimates for test score gain. The coefficient of the interaction between the dummy for belonging to the bottom 20 children in the class and belonging to a treatment school is significant in all of these columns, which indicates that, conditional on being in a school assigned to the treatment group, the treatment effect is actually bigger if the child is more likely to be assigned to the balsakhi. The treatment effect also appears to be declining with initial test score and the number of students in the class, though, controlling for the rank of the child, this effect is not very large.

In panel B of table 10, we present instrumental variables estimates of the direct and indirect impact of being in a balsakhi group. The direct effect is the sum of the coefficient on the dummy for whether the school gets a balsakhi and on the dummy for whether the child is assigned to the balsakhi group. In both Mumbai and Vadodara, the coefficient for whether or not the child is assigned to the balsakhi is positive, significant, and large (0.89 standard deviations in Vadodara, and 1.3 standard deviations in Mumbai.) This suggests that the children actually assigned to the balsakhi benefit much more than others in their school. In Mumbai, the treatment effect on those who were not sent to the balsakhi is negligible. In Vadodara, it is 0.21 standard deviations. Unfortunately, the estimate is too noisy to be distinguished from zero. Pooling both sample, we find an effect of 0.23 for children not assigned to the balsakhi in treatment school, and an additional 0.95 standard deviation for children who are assigned.

Clearly, while there is some (not very conclusive) evidence of an indirect effect for children who remained in class, the main reason the program is effective seems to be that it delivers more effective education to the children who were actually with the balsakhi. Part of this is probably due to the fact that the children assigned to work with the balsakhi ended up in much smaller classes than those who weren't, given the initial large class sizes in both cities. The reduction in class size was somewhat larger for children who were assigned to the balsakhi, because they were moved to classes whose average size was 20. However, if one believes the point estimate, the effect of the program for children assigned to the balsakhi group is more than five time the size of the effect for other children, which suggest that the balsakhi is a more effective teacher for these children.

In Vadodara, a mid-test was conducted, as well as a pre- and post- test. We conduct the analysis for both tests separately. It appears that, during the period between the pre and the mid test, both the balsakhi and the non-balskahi children benefitted equally from the program (the balsakhi dummy is not significant). Between the mid-test and the post-test, however, only children allocated to the balsakhi group benefitted.

Finally, we examine whether there is any evidence of a positive (or negative) impact of tracking by interacting the dummy for being in the treatment group with the average score of the bottom 20 children in the school. In Vadodara pre to mid-test, the effect of being in a balsakhi school for children who were not allocated to the balsakhi group declines with the initial test score of the children, which is consistent with a positive impact of tracking.

3.5 Cost Benefit Analysis

In seeking to improve the academic performance of schoolchildren, governments could potentially hire additional teachers, or hire balsakhis. Data on the cost of teachers, combined with the results presented above, give an idea of the cost effectiveness of each option. Table 12 shows the cost per student per year of the balsakhi program. While the data available from the municipalities about the cost of the program in Vadodara and Mumbai are not directly comparable to each other (for Vadodara we have the total cost of schoolteachers, while for Mumbai we know the starting salary of a new teacher), we use measures of the Pratham program cost that are comparable within cities.[33] The program cost 91 rupees, (about 2 dollars) per student[34] in Vadodara, and 54 Rupees (or 1 dollar) per child in Mumbai, while the teacher salary cost per student of the Vadodara and Mumbai Municipal Corporations are 3,168 Rupees and 1318 Rupees, respectively. (The cost per student/year increased in Mumbai in year 2 because the wage of the balsakhi was increased from 500 to 750 Rs./month).

Table 13 combines these numbers with the test score improvements over the pre- to post-test in years 1 and 2. Since the post-test used the same structure as the pre-test, improvement of the control children between the pre- and the post-test provides a measure of the effect of being in school for one year, which can be compared with the effect of having a balsakhi over the same period. Clearly, this is only suggestive, since children's competency over a year may change for reasons other than being in school. In year 1 in Vadodara, the estimate of the treatment effect was about 40% as large as the improvement between the pre-test and the post-test in the comparison group. The ratio of the cost per child is however 35. This calculation suggests that the average balsakhi is about 12-16 times more cost effective (in terms of improvement in test scores) than the average teacher. In the second year, the balsakhi program appears to be 11-12 times more cost effective than the average teacher. The results in Mumbai also suggest the Balsakhi program is dramatically more cost effective. In the first year, the treatment effect was half as large as the gains made by the comparison group, while the cost ratio was 24, suggesting Balsakhis are thirteen times more cost effective than teachers. In the second year, the treatment effect was 1/10 (respectively 1/4) as large in the third (fourth) standard as the comparison groups gains, giving a cost advantage of 2 in the third standard, and of six in the fourth standard. It is important to note that this results *do not* suggest that teachers should be replaced by balsakhis, since balsakhis are always complementing the teachers. The results do provide evidence that if the Mumbai and Vadodara Municipalities wanted to spend additional resources, hiring balsakhis may be a more effective way to do it than hiring additional teachers.

---

[33] We also have cost data from Vadodara in 2001-2002, and from Mumbai in 2002-2003. As costs did not change appreciably from one year to the next, we use these figures for cost-benefit comparisons for both years.
[34] The denominator includes all students in standards three and four in the treatment schools, since we will also use average test scores.

3.6 Conclusion

This paper reports the preliminary results of a remedial education program. The program has already shown that it can be brought to scale, since it is already reaching tens of thousands of children across India. Evaluations conducted in two cities over two years suggest that this is a remarkably effective and cost effective program: Test scores of children whose schools benefited from the program improved by .12 to 0.16 standard deviations in the first year, and .15 to 0.3 standard deviations in the second year. At the margin, the program is up to 12-16 times more effective than resources spent on teachers. Results are even stronger for children in the bottom of the distribution (in the bottom third of the distribution, the program improved tests score by 0.22 standard deviations in the first year, and 0.58 in the second year).

3.7 References

Angrist, Joshua D. (1995) "Conditioning on the Probability of Selection to Control Selection Bias," NBER Technical Paper #181

Angrist, Joshua, et. al. "Vouchers for Private Schooling in Colombia: Evidence from a Randomized Natural Experiment." American Economic Review 92 (2002): 1535-1558

Banerjee, Abhijit, Suraj Jacob and Michael Kremer. "Promoting School Participation in Rural Rajasthan: Results from Some Prospective Trials." MIT Department of Economics Working Paper (2002): .

Chin, Amiee (2002) "The Returns to School Quality When School Quality is Very Low: Evidence from Operation Blackboard in India." University of Houston

Duflo, Esther, (2001). "Schooling and Labor Market Consequences of School Construction in Indonesia: Evidence from an Unusual Policy Experiment." American Economic Review 91 (2001): 795-813

Glewwe, Paul, Michael Kremer and Sylvie Moulin (1997) "Textbooks and Test Scores: Evidence from a Prospective Evaluation in Kenya." Manuscript, Harvard University Department of Economics

Glewwe, Paul, Michael Kremer, Sylvie Moulin and Eric Zitzewitz. (2003) "Retrospective vs. Prospective Analyses of School Inputs: The Case of Flip Charts in Kenya." Journal of Development Economics, forthcoming

Kremer, Michael and Ted Miguel (2003) "Worms: Identifying Impacts on Education and Health in the Presence of Treatment Externalities." Econometrica, forthcoming

Kremer, Miguel, and Thornton (2002) "Merit Scholarships," manuscript, Harvard Economics.

Spohr, Chirstopher. (2003) "Formal schooling and workforce participation in a rapidly developing economy: evidence from `compulsory' junior high school in Taiwan." Journal of Development Economics 70: 291-321

Vermeersch, Christel (2002) "School Meals and Educational Achievement: Evidence from a Randomized Evaluation,." Manuscript, Harvard University Department of Economics

Table 1: Sample Design

	Standard	Study Group	Number of Schools	Number of Divisions	Number of Children	Number of Schools Treated	Number of Children With Balsakhi
Vadodara							
Year 1	Three	Balsakhi	48	78	2596	-	-
		No Balsakhi	48	80	2527	-	-
	Four	Balsakhi	48	73	2414	-	-
		No Balsakhi	49	77	2619	-	-
Year 2	Three	Balsakhi	61	101	3146	61	952
		No Balsakhi	61	93	2906	0	0
	Four	Balsakhi	61	96	3167	61	1011
		No Balsakhi	61	94	3170	0	0
Mumbai							
Year 1	Three	Balsakhi	32	70	2592	-	-
		No Balsakhi	35	65	2182	-	-
Year 2	Three	Balsakhi	39	74	2530	28	636
		No Balsakhi	38	79	2943	.	.
	Four	Balsakhi	38	77	2812	27	688
		No Balsakhi	39	71	2460	.	.

Table 2: Summary Statistics: Vadodara Year 1

	PRE TEST			MID TEST			POST TEST		
	Treatment	Control	Difference	Treatment	Control	Difference	Treatment	Control	Difference
STANDARD 3									
A. OBSERVATIONS	2596	2527	69	2204	2082	122	2094	2069	25
B. SCORES (PERCENTAGE)									
Math	0.262	0.264	-0.002	0.407	0.382	0.025	0.354	0.315	0.039
			(0.019)			(0.020)			(0.023)
Verbal	0.233	0.222	0.011	0.400	0.380	0.020	0.385	0.356	0.030
			(0.017)			(0.019)			(0.021)
C. NORMALIZED TEST SCORES									
Math	-0.008	0.000	-0.008	0.619	0.511	0.108	0.391	0.221	0.170
			(0.084)			(0.089)			(0.100)
Verbal	0.059	0.000	0.059	0.916	0.814	0.102	0.840	0.688	0.152
			(0.089)			(0.099)			(0.106)
D. PERCENTAGE OF CHILDREN PASSING EACH STANDARD COMPETENCY									
Math Standard 1	0.177	0.159	0.017	0.338	0.287	0.051	0.262	0.216	0.046
			(0.024)			(0.028)			(0.027)
Math Standard 2	0.012	0.018	-0.005	0.047	0.047	0.001	0.038	0.024	0.014
			(0.006)			(0.009)			(0.008)
Math Standard 3	0.042	0.031	0.012	0.132	0.113	0.020	0.091	0.064	0.027
			(0.010)			(0.020)			(0.017)
Verbal Standard 1	0.238	0.214	0.024	0.529	0.526	0.003	0.519	0.496	0.023
			(0.028)			(0.032)			(0.035)
Verbal Standard 2	0.158	0.150	0.008	0.331	0.317	0.014	0.284	0.252	0.032
			(0.024)			(0.028)			(0.031)
Verbal Standard 3	0.038	0.030	0.008	0.134	0.139	-0.005	0.095	0.076	0.019
			(0.011)			(0.022)			(0.020)
STANDARD 4									
A. OBSERVATIONS	2414	2619	-205	2127	2341	-214	1960	2178	-218
B. SCORES (PERCENTAGE)									
Math	0.441	0.452	-0.011	0.533	0.509	0.024	0.506	0.473	0.033
			(0.019)			(0.022)			(0.022)
Verbal	0.340	0.353	-0.013	0.511	0.497	0.014	0.498	0.485	0.013
			(0.017)			(0.022)			(0.023)
C. NORMALIZED TEST SCORES									
Math	-0.046	0.000	-0.046	0.349	0.247	0.102	0.231	0.088	0.142
			(0.080)			(0.096)			(0.095)
Verbal	-0.059	0.000	-0.059	0.738	0.672	0.067	0.677	0.617	0.060
			(0.078)			(0.101)			(0.108)
D. PERCENTAGE OF CHILDREN PASSING EACH STANDARD COMPETENCY									
Math Standard 1	0.365	0.386	-0.021	0.471	0.469	0.002	0.434	0.404	0.030
			(0.028)			(0.034)			(0.033)
Math Standard 2	0.064	0.057	0.007	0.108	0.097	0.011	0.084	0.078	0.006
			(0.010)			(0.016)			(0.013)
Math Standard 3	0.106	0.106	0.000	0.225	0.193	0.032	0.174	0.145	0.029
			(0.022)			(0.026)			(0.027)
Verbal Standard 1	0.437	0.459	-0.022	0.677	0.630	0.047	0.717	0.670	0.046
			(0.027)			(0.031)			(0.033)
Verbal Standard 2	0.274	0.316	-0.042	0.482	0.452	0.030	0.444	0.433	0.010
			(0.029)			(0.035)			(0.038)
Verbal Standard 3	0.111	0.123	-0.012	0.261	0.251	0.010	0.216	0.212	0.004
			(0.021)			(0.034)			(0.031)

Table 3: Summary Statistics: Vadodara Year 2

	PRE TEST			MID TEST			POST TEST		
	Treatment	Control	Difference	Treatment	Control	Difference	Treatment	Control	Difference
STANDARD 3									
A. OBSERVATIONS	3146	2906	240	2572	2409	163	3019	2790	229
B. SCORES (PERCENTAGE)									
Math	0.167	0.160	0.007	0.423	0.343	0.080	0.478	0.396	0.082
			(0.014)			(0.022)			(0.022)
Verbal	0.221	0.216	0.004	0.467	0.399	0.068	0.428	0.386	0.042
			(0.014)			(0.018)			(0.018)
C. NORMALIZED TEST SCORES									
Math	0.039	0.000	0.039	1.407	0.980	0.427	1.698	1.259	0.439
			(0.074)			(0.119)			(0.116)
Verbal	0.025	0.000	0.025	1.470	1.070	0.400	1.245	0.998	0.247
			(0.082)			(0.106)			(0.103)
D. PERCENTAGE OF CHILDREN PASSING EACH STANDARD COMPETENCY									
Math Standard 1	0.258	0.249	0.009	0.587	0.483	0.104	0.616	0.541	0.075
			(0.025)			(0.030)			(0.032)
Math Standard 2	0.016	0.018	-0.002	0.144	0.086	0.058	0.183	0.119	0.064
			(0.005)			(0.020)			(0.023)
Math Standard 3	0.001	0.000	0.001	0.023	0.009	0.014	0.057	0.034	0.023
			(0.001)			(0.005)			(0.010)
Verbal Standard 1	0.508	0.500	0.008	0.820	0.710	0.110	0.776	0.740	0.036
			(0.035)			(0.027)			(0.028)
Verbal Standard 2	0.073	0.088	-0.015	0.322	0.247	0.075	0.332	0.319	0.013
			(0.016)			(0.029)			(0.029)
Verbal Standard 3	0.016	0.009	0.007	0.137	0.124	0.013	0.146	0.120	0.026
			(0.005)			(0.020)			(0.021)
STANDARD 4									
A. OBSERVATIONS	3167	3170	-3	2893	2935	-42	3003	3007	-4
B. SCORES (PERCENTAGE)									
Math	0.309	0.297	0.012	0.529	0.436	0.094	0.575	0.498	0.076
			(0.018)			(0.016)			(0.020)
Verbal	0.348	0.332	0.017	0.561	0.476	0.085	0.517	0.457	0.060
			(0.017)			(0.017)			(0.018)
C. NORMALIZED TEST SCORES									
Math	0.051	0.000	0.051	1.002	0.598	0.405	1.197	0.869	0.329
			(0.077)			(0.069)			(0.087)
Verbal	0.083	0.000	0.083	1.135	0.716	0.418	0.916	0.621	0.295
			(0.083)			(0.083)			(0.089)
D. PERCENTAGE OF CHILDREN PASSING EACH STANDARD COMPETENCY									
Math Standard 1	0.470	0.449	0.021	0.737	0.642	0.095	0.779	0.695	0.084
			(0.030)			(0.020)			(0.029)
Math Standard 2	0.055	0.051	0.004	0.232	0.144	0.088	0.289	0.221	0.069
			(0.010)			(0.022)			(0.029)
Math Standard 3	0.015	0.012	0.003	0.052	0.040	0.012	0.110	0.082	0.028
			(0.003)			(0.009)			(0.014)
Verbal Standard 1	0.749	0.740	0.009	0.897	0.811	0.086	0.859	0.848	0.011
			(0.026)			(0.017)			(0.016)
Verbal Standard 2	0.186	0.181	0.005	0.441	0.325	0.116	0.481	0.363	0.118
			(0.020)			(0.026)			(0.028)
Verbal Standard 3	0.082	0.079	0.003	0.260	0.196	0.064	0.223	0.180	0.043
			(0.015)			(0.024)			(0.024)

Table 4: Summary statistics, Mumbai year 1

	PRE TEST			POST TEST		
	Treatment	Control	Difference	Treatment	Control	Difference
STANDARD 3						
A. OBSERVATIONS	2592	2182	410	2417	2027	390
B. SCORES (PERCENTAGE)						
Math	0.470	0.470	0.001	0.571	0.530	0.041
			(0.029)			(0.033)
Verbal	0.596	0.569	0.027	0.666	0.626	0.040
			(0.029)			(0.028)
C. NORMALIZED TEST SCORES						
Math	0.002	0.000	0.002	0.383	0.227	0.156
			(0.108)			(0.126)
Verbal	0.100	0.000	0.100	0.359	0.210	0.149
			(0.108)			(0.102)
D. PERCENTAGE OF CHILDREN PASSING EACH STANDARD COMPETENCY						
Math Standard 1	0.326	0.337	-0.012	0.397	0.357	0.040
			(0.038)			(0.039)
Math Standard 2	0.126	0.147	-0.021	0.211	0.195	0.017
			(0.025)			(0.036)
Math Standard 3	0.024	0.023	0.001	0.089	0.091	-0.003
			(0.007)			(0.022)
Verbal Standard 1	0.856	0.837	0.019	0.937	0.913	0.025
			(0.025)			(0.018)
Verbal Standard 2	0.486	0.473	0.013	0.577	0.526	0.050
			(0.045)			(0.042)
Verbal Standard 3	0.517	0.470	0.047	0.631	0.584	0.047
			(0.039)			(0.039)

Table 5: Summary Statistics: Mumbai Year 2

	PRE TEST			POST TEST			
	Treatment	Control	Difference	Treatment	Control	Difference	Implied Difference
STANDARD 3							
A. OBSERVATIONS	2530	2943	-413	2337	2731	-394	
B. SCORES (PERCENTAGE)							
Math	0.221	0.233	-0.012	0.502	0.470	0.031	0.044
			(0.016)			(0.028)	(0.038)
Verbal	0.351	0.344	0.007	0.588	0.569	0.018	0.025
			(0.022)			(0.025)	(0.034)
C. NORMALIZED TEST SCORES							
Math	-0.070	0.000	-0.070	1.509	1.333	0.176	0.245
			(0.087)			(0.155)	(0.213)
Verbal	0.025	0.000	0.025	0.898	0.831	0.067	0.093
			(0.082)			(0.091)	(0.126)
D. PERCENTAGE OF CHILDREN PASSING EACH STANDARD COMPETENCY							
Math Standard 1	0.137	0.167	-0.030	0.421	0.339	0.082	0.114
			(0.025)			(0.043)	(0.058)
Math Standard 2	0.082	0.090	-0.008	0.412	0.412	0.001	0.001
			(0.015)			(0.053)	(0.073)
Math Standard 3	0.003	0.006	-0.003	0.136	0.099	0.037	0.052
			(0.003)			(0.023)	(0.031)
Math Standard 4	0.007	0.013	-0.006	0.123	0.088	0.035	0.048
			(0.004)			(0.024)	(0.034)
Verbal Standard 1	0.653	0.648	0.005	0.820	0.817	0.004	0.005
			(0.036)			(0.022)	(0.031)
Verbal Standard 2	0.165	0.147	0.017	0.388	0.363	0.024	0.034
			(0.022)			(0.033)	(0.046)
Verbal Standard 3	0.137	0.131	0.005	0.317	0.307	0.010	0.013
			(0.021)			(0.034)	(0.047)
STANDARD 4							
A. OBSERVATIONS	2812	2460	352	2635	2290	345	
B. SCORES (PERCENTAGE)							
Math	0.409	0.396	0.013	0.642	0.564	0.079	0.106
			(0.019)			(0.027)	(0.034)
Verbal	0.555	0.530	0.025	0.721	0.683	0.038	0.051
			(0.021)			(0.021)	(0.026)
C. NORMALIZED TEST SCORES							
Math	0.053	0.000	0.053	0.995	0.678	0.317	0.426
			(0.076)			(0.111)	(0.136)
Verbal	0.083	0.000	0.083	0.641	0.513	0.127	0.171
			(0.071)			(0.069)	(0.086)
D. PERCENTAGE OF CHILDREN PASSING EACH STANDARD COMPETENCY							
Math Standard 1	0.300	0.240	0.060	0.474	0.387	0.087	0.117
			(0.031)			(0.036)	(0.046)
Math Standard 2	0.245	0.243	0.003	0.554	0.464	0.090	0.121
			(0.023)			(0.055)	(0.072)
Math Standard 3	0.042	0.041	0.001	0.241	0.171	0.069	0.093
			(0.010)			(0.033)	(0.042)
Math Standard 4	0.074	0.063	0.011	0.335	0.242	0.093	0.125
			(0.013)			(0.035)	(0.044)
Verbal Standard 1	0.825	0.796	0.029	0.923	0.900	0.023	0.030
			(0.022)			(0.014)	(0.017)
Verbal Standard 2	0.338	0.333	0.005	0.576	0.512	0.064	0.086
			(0.027)			(0.034)	(0.043)
Verbal Standard 3	0.355	0.317	0.038	0.532	0.485	0.047	0.064
			(0.031)			(0.033)	(0.043)

Table 6 : Attrition patterns

	Bombay, year 1			Bombay, year 2			Bombay 2 years			Vadodara, year 1			Vadodara, year 2		
	Balsakhi	No Balsakhi	Difference	Balsakhi	No Balsakhi	Difference	Balsakhi	No Balsakhi	Difference	Balsakhi	No Balsakhi	Difference	Balsakhi	No Balsakhi	Difference
Standard 3, All															
Percent attrition	0.070	0.075	-0.004	0.077	0.073	0.005	0.255	0.250	-0.006	0.193	0.181	0.012	0.040	0.040	0.000
			(0.015)			(0.010)			(0.009)			(0.020)			(0.008)
Difference in score at pretest attriters-stayers	-0.146	-0 274	0.128	-0.330	-0.193	-0.137	-0.445	-0.594	0.149	-0.130	-0.222	0 092	-0.060	-0.075	0.015
			(0.169)			(0.129)			(0.151)			(0.093)			(0.131)
Standard 4, All															
Percent attrition				0.063	0.070	-0.006				0.188	0.168	0.020	0.052	0.051	0.000
						(0.010)						(0.021)			(0.020)
Difference in score at pretest attriters-stayers				-0.180	-0.427	0.247				-0.178	-0.176	-0 002	-0.098	0.015	-0.113
						(0.139)						(0.077)			(0.155)

Table 7: Differences in differences estimate of the impact of the balsakhi program, by city and sample

	Number of Observations	Math	Verbal	Total
Mumbai and Vadodara together Year 1	12730	0.188	0.069	0.138
		(0.047)	(0.056)	(0.047)
Mumbai and Vadodara together Year 2	21805	0.319	0.153	0.250
		(0.067)	(0.050)	(0.059)
Pooling Both Standards				
Vadodara Year 1	8301	0.196	0.109	0.164
		(0.059)	(0.058)	(0.058)
Vadodara Year 2	11819	0.342	0.225	0.309
		(0.077)	(0.064)	(0.073)
Mumbai Year 2	9986	0.279	0.032	0.150
		(0.124)	(0.076)	(0.099)
Mumbai Year 2 Specification Check	9986	0.285	0.063	0.173
		(0.112)	(0.067)	(0.088)
Standard 3				
Vadodara Year 1	4163	0.198	0.101	0.162
		(0.092)	(0.090)	(0.089)
Vadodara Year 2	5809	0.399	0.224	0.342
		(0.111)	(0.094)	(0.103)
Mumbai Year 1	4429	0.163	0.060	0.118
		(0.072)	(0.072)	(0.067)
Mumbai Year 2	5063	0.312	0.043	0.163
		(0.164)	(0.108)	(0.133)
Mumbai Year 2 Specification Check	5063	0.276	0.073	0.168
		(0.149)	(0.097)	(0.121)
Standard 4				
Vadodara Year 1	4138	0.186	0.114	0.160
		(0.074)	(0.075)	(0.072)
Vadodara Year 2	6010	0.269	0.211	0.258
		(0.089)	(0.073)	(0.081)
Mumbai Year 2	4923	0.376	0.084	0.232
		(0.111)	(0.077)	(0.093)
Mumbai Year 2 Specification Check	4923	0.403	0.104	0.257
		(0.099)	(0.066)	(0.081)
Two Year 01-03				
Mumbai Pre-test Year 1 to Pre-test Year 2	3124	-0.129	0.003	-0.066
		(0.093)	(0.100)	(0.094)
Mumbai Pre-test Year 1 to Post-test Year 2	3299	0.521	0.145	0.351
		(0.142)	(0.113)	(0.115)

Table 8: Differences in differences for standard competencies, by city and year

	Math Competencies for				Verbal Competencies for		
	Standard 1	Standard 2	Standard 3	Standard 4	Standard 1	Standard 2	Standard 3
Vadodara							
Year 1							
Both Standards	0.050	0.011	0.021	-	0.039	0.040	0.014
	(0.022)	(0.007)	(0.019)	-	(0.023)	(0.019)	(0.015)
Standard Three	0.047	0.018	0.016	-	0.007	0.024	0.011
	(0.033)	(0.009)	(0.019)	-	(0.036)	(0.027)	(0.017)
Standard Four	0.051	0.004	0.027	-	0.070	0.058	0.019
	(0.031)	(0.011)	(0.033)	-	(0.028)	(0.027)	(0.022)
Year 2							
Both Standards	0.064	0.063	0.022	-	0.019	0.073	0.031
	(0.023)	(0.019)	(0.009)	-	(0.023)	(0.021)	(0.015)
Standard Three	0.065	0.066	0.022	-	0.030	0.029	0.020
	(0.032)	(0.022)	(0.010)	-	(0.034)	(0.028)	(0.021)
Standard Four	0.061	0.064	0.024	-	0.002	0.115	0.042
	(0.033)	(0.029)	(0.013)	-	(0.023)	(0.030)	(0.021)
Mumbai							
Year 1							
Standard Three	0.058	0.037	-0.003	-	0.004	0.045	0.007
	(0.032)	(0.030)	(0.020)	-	(0.021)	(0.030)	(0.032)
Year 2							
Both Standards	0.088	0.058	0.078	0.094	-0.013	0.041	0.009
	(0.038)	(0.047)	(0.025)	(0.032)	(0.030)	(0.025)	(0.037)
Standard Three	0.147	0.005	0.053	0.054	-0.008	0.003	0.003
	(0.056)	(0.066)	(0.029)	(0.032)	(0.049)	(0.035)	(0.043)
Standard Four	0.042	0.125	0.094	0.110	-0.003	0.084	0.016
	(0.048)	(0.064)	(0.039)	(0.045)	(0.030)	(0.036)	(0.061)

Table 9: Differences in differences, by third on the initial test score distribution

	Year 1				Year 2					Two Year Analysis			
	N	Math	Verbal	Total	Fraction of Children who go to the Balsakhi	N	Math	Verbal	Total	N	Math	Verbal	Total
Vadodara and Mumbai together													
Bottom Third	4147	0.250	0.146	0.211	0.22	7293	0.507	0.133	0.316				
		(0.055)	(0.061)	(0.057)			(0.155)	(0.093)	(0.123)				
Middle Third	4271	0.179	0.036	0.115	0.16	7086	0.319	0.034	0.167				
		(0.057)	(0.068)	(0.054)			(0.136)	(0.089)	(0.111)				
Top Third	4312	0.127	0.016	0.079	0.06	7426	0.039	-0.038	-0.003				
		(0.062)	(0.072)	(0.060)			(0.151)	(0.076)	(0.106)				
Vadodara													
Bottom Third	2661	0.224	0.120	0.185	0.22	4007	0.458	0.308	0.417				
		(0.059)	(0.064)	(0.061)			(0.088)	(0.074)	(0.084)				
Middle Third	2784	0.186	0.140	0.173	0.18	3836	0.407	0.217	0.342				
		(0.075)	(0.070)	(0.069)			(0.094)	(0.072)	(0.083)				
Top Third	2856	0.156	0.060	0.118	0.09	3976	0.205	0.183	0.210				
		(0.084)	(0.083)	(0.081)			(0.083)	(0.086)	(0.082)				
Mumbai													
Bottom Third	1486	0.297	0.212	0.269	0.22	3286	0.507	0.133	0.316	949	0.754	0.329	0.572
		(0.109)	(0.099)	(0.102)			(0.155)	(0.093)	(0.123)		(0.196)	(0.146)	(0.167)
Middle Third	1487	0.147	-0.032	0.060	0.14	3250	0.319	0.034	0.167	1079	0.504	0.057	0.295
		(0.083)	(0.087)	(0.069)			(0.136)	(0.089)	(0.111)		(0.160)	(0.128)	(0.110)
Top Third	1456	0.049	0.005	0.029	0.03	3450	0.039	-0.038	-0.003	1096	0.342	0.069	0.217
		(0.069)	(0.070)	(0.063)			(0.151)	(0.076)	(0.106)		(0.121)	(0.101)	(0.093)

Table 10: Disentangling balsakhi and class size effects

	Balsakhi Assignment				Improvement in Test scores Pre to Post				Improvement in Test scores Vadodara	
	Mumbai	Vadodara	Both		Mumbai	Vadodara	Both		Pre to Mid	Mid to Post
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
A. First stages and reduced form										
Treatment School	0.259	0.457	0.363	0.471	0.434	0.636	0.590	0.524	1.039	-0.393
	(0.064)	(0.045)	(0.037)	(0.052)	(0.194)	(0.185)	(0.132)	(0.176)	(0.238)	(0.226)
Treatment * Rank <20	0.075	0.177	0.138	0.063	0.124	0.152	0.132	0.158	-0.047	0.201
	(0.029)	(0.022)	(0.017)	(0.020)	(0.086)	(0.080)	(0.063)	(0.065)	(0.082)	(0.096)
Treatment * Rank	-0.002	0.001	-0.001	-0.003	-0.003	0.003	0.000	0.005	-0.021	0.023
	(0.001)	(0.001)	(0.001)	(0.003)	(0.004)	(0.005)	(0.003)	(0.009)	(0.012)	(0.013)
Treatment * Pre-test score	-0.083	-0.088	-0.082	-0.077	0.004	-0.100	-0.055	-0.057	-0.152	0.072
	(0.027)	(0.013)	(0.013)	(0.014)	(0.097)	(0.087)	(0.065)	(0.066)	(0.076)	(0.070)
Treatment * Number of students	0.001	-0.004	-0.001	-0.001	-0.002	-0.007	-0.005	-0.005	-0.008	0.001
	(0.001)	(0.001)	(0.001)	(0.001)	(0.003)	(0.004)	(0.003)	(0.003)	(0.004)	(0.002)
Rank <20					-0.219	-0.031	-0.097	0.002	0.065	-0.047
					(0.063)	(0.059)	(0.045)	(0.046)	(0.058)	(0.066)
Rank					0.003	0.004	0.004	0.015	0.033	-0.008
					(0.002)	(0.004)	(0.002)	(0.005)	(0.008)	(0.008)
Pre-test score					-0.368	-0.340	-0.349	-0.356	-0.297	-0.032
					(0.069)	(0.060)	(0.042)	(0.043)	(0.061)	(0.043)
Number of students					0.000	-0.002	0.000	0.000	0.002	-0.004
					(0.001)	(0.003)	(0.001)	(0.001)	(0.004)	(0.001)
B. Instrumental variable estimates										
Sawa Balsakhi					1.383	0.891	0.954	2.785	0.105	0.833
					(0.753)	(0.490)	(0.408)	(1.644)	(0.385)	(0.504)
Balsakhi School					0.080	0.213	0.237	-0.877	0.753	-0.576
					(0.363)	(0.330)	(0.212)	(0.991)	(0.276)	(0.339)
Treatment * Rank					-0.001	0.003	0.001	0.015	0.003	-0.001
					(0.004)	(0.005)	(0.003)	(0.014)	(0.004)	(0.003)
Treatment * Pre-test score					0.134	-0.019	0.031	0.147	-0.149	0.160
					(0.128)	(0.104)	(0.079)	(0.134)	(0.086)	(0.096)
Treatment * Number of students					-0.003	-0.004	-0.004	-0.003	-0.007	0.004
					(0.002)	(0.005)	(0.002)	(0.002)	(0.004)	(0.003)
C. Instrumental Variable estimate (with interactions with the average pre-test score of balsakhi children)										
Sawa Balsakhi					1.281	0.813	1.029		0.335	0.524
					(0.717)	(0.538)	(0.447)		(0.412)	(0.511)
Balsakhi School					-0.211	0.299	0.128		0.469	-0.211
					(0.498)	(0.389)	(0.277)		(0.297)	(0.376)
Treatment school * Pre-test score					-0.168	0.150	-0.013		-0.365	0.486
					(0.356)	(0.283)	(0.203)		(0.224)	(0.242)

Notes. Column (4) has as additional controls Treatment*Rank Squared, Treatment*Rank Cubed, and Treatment*Rank^4
Column (8) includes fourth order polynomials in rank, and in treatment * rank. Space limitations preclude reporting of the polynomial coefficients

Table 11: Attendance

		Standard 3			Standard 4			Standard 3 and 4 Together		
		Avg Attendance Sept-Oct	Simple Differences	Difference in Difference	Avg Attendance Sept-Oct	Simple Differences	Difference in Difference	Avg Attendance Sept-Oct	Simple Differences	Difference in Difference
Year 1	**RA Attendance**									
	Whole Sample	0.88	-0.002	0.006						
			(0.012)	(0.017)						
	Roster Attendance									
	Whole Sample	0.93	-0.004	-0.014						
			(0.010)	(0.011)						
Year 2	**RA Attendance**									
	Whole Sample	0.88	0.013	0.005	0.88	-0.016	0.000	0.88	-0.001	0.001
			(0.017)	(0.023)		(0.014)	(0.021)		(0.011)	(0.016)
	Roster Attendance									
	Whole Sample	0.91	0.001	0.009	0.91	-0.021	0.011	0.91	-0.010	0.010
			(0.013)	(0.013)		(0.012)	(0.011)		(0.009)	(0.009)

Table 12: Cost Comparison

		Cost Per Year (Dollars)	Cost Per Year (Rupees)	Students	Rps/student per year
Vadodara					
year 1 & 2	Balsakhi	11830	520384	5730	91
year 1 & 2	Primary School Teachers	3924300	172740888	54525	3168
Mumbai					
year 1	Primary School Teachers (L Ward)	89488.6364	3937500	2988	1318
year 1	Balsakhi	4200	184800	3433	54
year 2	Balsakhi	7770	341880	4225	81

Note: Cost of teachers for Vadodara is calculated by dividing the total wage bill for teachers by the number of students, using data from year 1. In Mumbai, it is calculated using the starting salary of a teacher (7500 Rs/month) and the number of divisions in the schools in the study. In Vadodara, balsakhis were paid 500 Rs./month in both years. In the first year in Mumbai, balsakhis were paid 500 Rs./month, while in the second they were paid 750 Rs./month) (Schoolteacher salary figures are preliminary.)

Table 13: Cost Benefit Analysis

		Improvements in Test Scores		Rupees per standard deviation	
		Standard 3	Standard 4	Standard 3	Standard 4
Vadodara	Year 1				
	Balsakhi treatment effect	0.162	0.16	561	568
	Pre to Post difference (comparison)	0.46	0.36	6887	8800
	Year 2				
	Balsakhi treatment effect	0.356	0.249	255	365
	Pre to Post difference (comparison)	1.144	0.731	2769	4334
Mumbai	Year 1				
	Balsakhi treatment effect	0.115	n/a	468	n/a
	Pre to Post difference (comparison)	0.215	n/a	6129	n/a
	Year 2				
	Balsakhi treatment effect	0.151	0.226	536	358
	Pre to Post difference (comparison)	1.094	0.607	1205	2171