Heidelberg Papers

in South Asian

and Comparative Politics

Emergence of Independent Candidates: A Negative Binomial Regression Model of an Indian Parliamentary Election

by

Kaushik Bhattacharya

Working Paper No. 57

August 2010

South Asia Institute

Department of Political Science

Heidelberg University

HEIDELBERG PAPERS IN SOUTH ASIAN AND COMPARATIVE POLITICS

ISSN: 1617-5069

About *HPSACP*

This occasional paper series is run by the Department of Political Science of the South Asia Institute at the University of Heidelberg. The main objective of the series is to publicise ongoing research on South Asian politics in the form of research papers, made accessible to the international community, policy makers and the general public. HPSACP is published only on the Internet. The papers are available in the electronic pdf-format and are designed to be downloaded at no cost to the user.

The series draws on the research projects being conducted at the South Asia Institute in Heidelberg, senior seminars by visiting scholars and the world-wide network of South Asia scholarship. The opinions expressed in the series are those of the authors, and do not represent the views of the University of Heidelberg or the Editorial Staff.

Potential authors should consult the style sheet and list of already published papers at the end of this article before making a submission.

Emergence of Independent Candidates: A Negative Binomial Regression Model of an Indian Parliamentary Election[1]

Kaushik Bhattacharya[2]

Keywords: Independent Candidates, Election, Prisoners' Dilemma, Negative Binomial

ABSTRACT:

The paper specifies a model of the first-past-the-post (FPTP) electoral system in which political parties themselves float independent candidates to gain electoral advantage, leading to a Prisoners' Dilemma type game where each party tries to out-maneuver one another. Imposing some intuitively appealing assumptions on this game, we show that the total number of independent candidates across constituencies would follow a Negative Binomial distribution. Empirical results for the 2004 parliamentary election in India reveal a good fit of the Negative Binomial model to data. Results also help to identify a few major determinants of the spatial distribution of independent candidates in India. Results point out that the number of independent candidates across constituencies in a State in India is strongly influenced by political fractionalization in that State, with metropolitan and urban constituencies on an average having more independent candidates. We also find that elite politicians and their family members, ceteris paribus, face more independent candidates. Finally, results establish that number of independent candidates is typically less in reserved constituencies due to reduced number of potential candidates. Our results suggest that FPTP electoral systems as in India need to put appropriate institutional constraints that increase transaction costs of electoral participation for independent candidates.

SECTION 1: INTRODUCTION

In politics, an "independent" is defined as one who is not affiliated with any political party. Generally such a person runs in an election without the support of any political party. Internationally, independent candidates have

[1] The views expressed in this paper are personal and not necessarily of the institution the author belongs to. The author is grateful to Harihar Bhattacharyya, Sukumar Nandi and an anonymous referee whose comments and suggestions greatly improved an earlier exposition. The author bears full responsibility for any error.
[2] The author can be contacted at: kbhattacharya@iiml.ac.in

played important roles in different countries in different stages of political development. In old and established democracies like the UK or Canada, independent candidates have won elections frequently. In relatively new democracies like India, independents have actively participated in the electoral process. In recent years, independents have also contributed significantly in the political developments in South Africa, Russia and transitional economies in East Europe.

This paper attempts to explain the reasons behind the emergence of independent candidates and also attempts to trace its spatial determinants. Without the institutional support of a political party, independent candidates not only have to bear the risk of losing, but often also have to bear the risk of forfeiture of deposit. Therefore, a natural question to ask is: what is his/her incentive? More importantly, we ask: what type of political environments in general, and what type of constituencies in particular, induce individuals to run as independent candidates? We argue that these questions are important from both theoretical and practical perspectives and that existing political and economic literature have not addressed them in sufficient detail.

In an ideal world, independents would hold a centrist viewpoint in a relatively polarized political environment. They may also have a viewpoint on specific issues or policies. Therefore, one reason for their taking part in electoral politics could be grievance against existing politicians and/or policies. Independent candidates could also be a former member of a political party and stand in election as rebels. A third category of independents are those who may support a political party but believe that they should not formally commit to be its member. Fourth category could be free-riders who, by paying a limited cost, enjoy and utilize the free publicity that elections offer them. Finally, independent candidates could also be cranks who run for idiosyncratic reasons (Canon, 1993).[3] Unfortunately, neither standard voting theories nor the existing theories of coalition formation in democracies provide clear and unambiguous characterization of situations that would lead to the emergence of independent candidates in elections.

The starting point of the literature on voting is the median voter theory. As per this theory, in a first-past-the-post (FPTP) electoral system, if voters' preferences can be represented by a uni-dimensional single-peak distribution and if there are two political parties, to win elections both the political parties would then tend to pick up the preferences of voters in or around the median of this distribution. Depending upon the purpose of the study, the simple median voter framework has been extended in different ways. In a seminal review article, Osborne (1995) has shown that if there are more than two (potential) candidates, then the basic incentive inherent in the two-party model is significantly diluted. If information is perfect and voting is costless, then enough of the incentive may survive so that all candidates who enter the competition choose similar policy positions. In such cases, if the distribution reflecting voters' preference is highly heterogeneous, one reason of the emergence of smaller parties and independent candidates would be to fill up

[3] Theoretically, there could be a method in madness. In economies with weak institutional structure, a "crank" may run as independent (i) if the transaction cost of running as independent is low compared to "donations" expected from him by the existing political parties, or, (ii) to avoid the "hazards" of anticipated election duties.

ideological vacuums created by major party competition (Rosenstone *et al*, 1996). However, it may be noted that when information is imperfect, results are unclear.

Existing studies have pointed out that presence of a third party or independent candidates could distort the results in an election. In the context of the US, Abramson *et al* (1995) has reviewed the experience with third party and independent candidates in the context of US presidential elections. Also, Heckelman and Yates (2008) have shown both theoretically and empirically that in the presence of independent candidates, two state senators will generally not be from opposite parties and will be closer in ideological space than if they were elected under strict two-party competition. Majority of these studies are, however, in the context of developed countries and implicitly assume the existence of two major political parties. Further, these studies focus on the outcomes of elections and not on the factors that lead to the emergence of independent candidates.

It is well known that in a politically charged atmosphere, independent candidates could play a decisive role in an election. Therefore, in a FPTP electoral system where even a single vote could make the difference between the winner and the losers, for independent candidates, the very motive of standing in an election may not be winning *per se*, but to influence the outcome through participation. A major incentive here is to engage in bargaining with the mainstream political parties and eventually, to cut a deal with one of them in the long-run.

In the context of the US, recently Lem and Dowling (2006) have specified a model that attempts to explain the emergence of smaller parties and independent candidates in gubernatorial elections. In the empirical part, they have specified and estimated a Negative Binomial model to account for the count data properties of the number of such candidates. Lem and Dowling (2006) have identified several factors (e.g., political competition, constituency size, population density, legal and institutional hurdles across States etc.) that explain the emergence of independent candidates. However, this study – although it enhances our understanding significantly – suffers from two major limitations. First, it has not addressed the complex interaction among mainstream political parties and independent candidates in sufficient detail. Such interactions might have important implications in all democracies, especially in developing and emerging market economies with a FPTP electoral system in place. Second and more importantly, the Negative Binomial specification in the empirical part of their study is *ad hoc* and is based on statistical convenience. In particular, it does not emerge out of a set of deductive arguments.

In the theoretical framework proposed in this paper, we argue that independent candidates could be dummies floated by political parties themselves to gain electoral advantage. That mainstream political parties have incentives to float dummy independent candidates has been recognized for a long time (McKnight, 1999). FPTP electoral systems are likely to be more vulnerable to this type of electoral behavior because in such systems, difference of even a few votes may turn out to be crucial.

The dummies floated by political parties could be “clones” of its major rivals, the underlying idea being to confuse voters intending to vote for its

rivals. Floating independents may also help parties to have additional electoral agents in different polling booths in a constituency. In many developing and emerging market economies, sometimes it is difficult to establish identities of voters. Support or objection of a "neutral" agent may convince electoral officers about identities of voters more than the agents of mainstream parties. Having additional agents in booths may, therefore, help parties to "pass on" false voters as true ones or prevent genuine voters from voting.

Whatever be the underlying motives of the political parties, we attempt to show that this unsavory behavior, if carried out by all political parties, leads to a prisoners' dilemma type political game. We also show that when this game unfolds, it has certain implications on the spatial distribution of independent candidates. Although in principle this type of political game may start in any FPTP electoral system any time, standard results in game theory indicate that they are more likely to start in a fragmented polity because in such a situation all political parties have incentives to float dummies. Further, common sense also suggests that economies with weak institutional practices would be more vulnerable from this type of political behavior. Taken together, an implication of these observations is that for potential applications of this theory, one should study voting patterns of developing and emerging market economies rather than developed countries with relatively stronger institutional structure and with two major political parties in place.

In this paper an attempt is made to model the spatial distribution of independent candidates in India. To convince readers, we now argue why India would be a good case study. In terms of scale of operation, logistics and active involvement of number of people, Indian parliamentary election is one of the biggest events in the world. Election of the lower house (*Lok Sabha*) of the parliament in India is carried out under FPTP rule. Since independence in 1947, *Lok Sabha* elections have been held in India fifteen times. However, even with more than sixty years of democratic experience, results of these elections suggest that a clear bi-party system in India is yet to emerge at the central level.

Because of its lack of an effective bi-party system at the central level, India would be a good case to examine the role political fragmentation plays in the emergence of independent candidates. Being a large country with a lot of diversities, it is no surprise that different parts of India react differently to parliamentary elections. While many smaller units like States show signs of emergence of a bi-party system, politics at the central level in India is still guided by complex political coalitions involving both national and regional parties.

In the Indian context, there are several benchmark studies on party system and coalition making. In an early and seminal study, Kothari (1970) identified some of the specialties and complexities of the Indian political process. Studies like Weiner (1978) analyzed the socio-political implications of the benchmark 1977 *Lok Sabha* election in India. Subsequent studies like Mitra and Singh (1999) and Chhibber (1999) have examined the linkage between social cleavages and political parties and their gradual changes over time. These later studies attempt to explain the complex coalition politics in India in terms of multiple cleavages like caste, religion, language, region and

social and economic classes. However, besides the relational dynamics of different political parties in India over space and time, a major feature of Indian parliamentary elections is the large number of independent candidates. None of these studies offer a systematic explanation of this phenomenon.[4] This is surprising because common sense suggests that when coalition governance is the order of the day, that governance structure may itself lead to participation of too many independent candidates and, in turn, enhance the complexity of political management.

In the specific context of India, political importance of independent candidates has varied substantially over the years. In the *Lok Sabha* elections in 1962 and and 1967 in India, independent candidates won 20 and 35 seats respectively. However, since 1970s, number of seats won by independent candidates has decreased substantially. During early years of Indian democracy, the high number of independent candidates in an election used to cost the public exchequer in India severely.[5] Other costs like security costs were also high. If for example, an independent candidate died (or, was murdered), election in that constituency would be postponed. In a politically volatile or charged atmosphere, such a situation – though infrequent – happened in India.[6] Because of these institutional weaknesses, during the 1996 parliamentary election in India, a total of 10,635 independent candidates appeared in a total of 543 parliamentary seats. Electoral reforms and the use of electronic voting machines led to a decrease in independent candidates subsequently. Still, Lok Sabha elections in 1998, 1999 and 2004 had 1,915, 1,945 and 2,385 independent candidates respectively. These numbers are still high, especially if one considers one of the earliest democracies like the United Kingdom as benchmark.[7]

This paper attempts to provide a systematic explanation of the emergence of independent candidates in the Parliamentary election in India

[4] To put the lack of discussion on this area in perspective, we note that the very word "Independent" is not present in the "Index" part of the books of Kothari (1970), Mitra and Singh (1999) and Chhibber (1999). Winer (1978) too has discussed this aspect scantily, observing that "In any event, the number of votes received by most independent candidates was very small, often only 1 or 2 per cent of the total vote" (p. 83).

[5] For example, during the assembly election in Tamil Nadu, India, in 1996, Modaurichi assembly constituency had 1033 candidates. During Parliamentary elections in the same year, Nalgonda constituency in Andhra Pradesh and Belgaum constituency in Karnataka had 480 and 456 candidates respectively. In each of these cases, overwhelming majority were independent candidates. Election Commission of India was forced to print ballot booklets instead of ballot papers.

[6] In the parliament polls of 1985, Palakondarayudu, the candidate of Telugu Desam at Raychoti in Cuddapah district, Andhra Pradesh, was unsure of the support of the two main local factions. So, allegedly he got the election postponed by killing Guvvala Subbarayudu, an independent candidate. He thus gained time to rope in the two factions, and succeeded in winning the election held later. In 1989, polls were held simultaneously for assembly and parliament in Andhra Pradesh. Palakondarayudu was this time a candidate for the assembly. Apprehensive that he may repeat his victorious performance, Nagi Reddy, his main opposition, set up a pliant man of their own faction, Avula Subba Reddy by name, as an independent candidate. To get the election postponed, Nagi Reddy allegedly have him killed the day before the election (Balagopal, 2004).

[7] The number of independent candidates in the 2005 parliamentary election in the United Kingdom was only 162 for a total of 646 seats.

in the year 2004. In the 2004 *Lok Sabha* election, only 5 out of 2385 independent candidates in India could win. In contrast, as many as 2370 independent candidates lost their deposits. However, these figures do not truly reflect the importance of independent candidates in the political process. Independent candidates as a group got 4.25% of the total valid votes in the 2004 parliamentary election in India. In a fragmented polity, these figures are not negligible. More importantly, in 116 constituencies out of a total of 543, the total number of votes that independent candidates got as a group was more than the gap between the winner and his/her nearest rival.

To empirically explain the spatial distribution of independent candidates in India in the 2004 *Lok Sabha* election, we specify a Poisson and a Negative Binomial regression framework. Poisson and Negative Binomial regression models accommodate the integer property of the count data directly. Further, the equations that are used for estimation of parameters in these models are similar to more traditional regression models. Because of these attractive features, these models have been applied in hundreds of studies involving count data.[8] . In most such studies, however, specification of Poisson and Negative Binomial models were *ad hoc* and were based on statistical convenience.[9] In this study, however, attempts are made to show through a simple behavioral model why the distribution of independent candidates could be Negative Binomial.

The plan of the paper is as follows: Section 2 presents the analytical framework. Section 3 discusses the data, carries out a brief descriptive analysis and presents the estimated Poisson and Negative Binomial models. To check the stability of the estimated coefficients, it also carries out a brief bootstrap analysis. Section 4 discusses policy implications of the results. Finally, Section 5 summarizes the main findings, identifies some limitations of this study and suggests a few possible generalizations. The paper has two appendices. Appendix A provides the mathematical details about the optimization carried out by political parties in the model. It also offers a heuristic justification of the Negative Binomial model. Appendix B summarizes the process of econometric specification and estimation of the model parameters.

SECTION 2: THE ANALYTICAL FRAMEWORK

Suppose there are two political parties (say, P_1 and P_2) fighting in an election in a constituency (say, *C*). Suppose, both P_1 and P_2 perceive that they have equal chance of winning the election. Both parties now realize that if they float one or more independent candidates cloning its rival, each voter will then vote for each such independent candidate with a small probability at the cost of its rival. Both parties, therefore, have incentive to compete with each other to float more independent candidates than the rival. As floating such candidates is not costless, the ultimate number of independents floated by them, however, depends on the resources in their possession.

[8] See Cameron and Trivedi (1998) for references.
[9] This includes the study by Lem and Dowling (2006) in the specific context of independent candidates.

This entire process of floating clone independent candidates may be thought of as a Prisoners' dilemma type game. The game is simultaneous because although in principle, each party can observe what the other one is doing for some time, crucial adjustments will take place only at the last moment of filing nomination. To maximize the chance of winning in this game, each political party has to guess the number of independent candidates that would be floated by its rival and then need to put one more clone independent candidate in addition, provided the party has sufficient resources to do that. Appendix A contains a more sophisticated mathematical specification that outlines the optimizing behavior of political parties. As both political parties would like to outmaneuver each other at the crucial last moment of filing nominations, any information obtained prior to that on the number of candidates floated by its rival will not be of any use. Appendix A also offers a heuristic justification of the Negative Binomial distribution through the "memorylessness" associated with this process.

We now discuss some implications and generalizations of this simple model. First, if both parties have unequal strength in a constituency and both perceive that cloning is unlikely to change the outcome, then cloning may not take place at all. In our model, political parties will float clone candidates only if they are within budget and floating such candidates lead to an expected win.[10]

Second, the same framework can be generalized to take care of "rebel" candidates. Political parties are not homogeneous entities. If a particular person in a political party is not selected as its candidate and decides to run as independent, the same candidate may be interpreted as an independent floated by the other political party at zero cost.

Third, standard results on Prisoners' Dilemma indicate that any tacit understanding between the two political parties is likely to be unstable. The incentive to "cheat" is strong, especially in political games of this type which is repeated infrequently and where the stake is high. Further, even if a tacit pact is there, if a free rider decides to take advantage of his/her similarity with one of the candidates, the fragile pact may break down due to misunderstanding.

Fourth, with three or more political parties having equal strength, the situation becomes more complex. Each political party will now have to field independent candidates against all its rivals. Like in the classical Prisoners' Dilemma with multiple players, detection of the cheater becomes more difficult because even if "clones" are identified, one may not know which political party has floated the clone.

Fifth, all political parties have strong incentives to not to disclose their hands early in this game. Therefore, we hypothesize that unsavory political competition will lead to a rush in filing nominations of independent candidates at the last moment.

[10] Political parties sometimes may think of the "big picture". If one party is sure to win majority of seats, then that party may not be interested to take part in a cloning battle in a constituency where the chance of winning is same for all parties. Similarly, desperation to win the majority of seats may force some party to start cloning battle in a constituency in unfavorable circumstances.

Sixth, in this model, constituency size does not play any role. Rather, the important roles are played by the number of non-independent candidates in fray. An observable implication of the model is that *ceteris paribus*, number of independent candidates will be unaffected by constituency size, but will have an increasing relationship with the number of non-independent candidates.

SECTION 3: DATA AND EMPIRICAL ANALYYSIS

The data for this paper have been downloaded and compiled from the website of the Election Commission of India (ECI) [http://www.eci.gov.in]. The ECI website contains detailed election results of all constituencies for all parliamentary elections in India. In this study, we focus on the year 2004 only. For each constituency, the name, age, sex, political affiliation and the number of votes obtained by each candidate have been recorded. For each candidate, we also know their caste status. In some cases, a constituency itself is declared a reserve constituency – sometimes for scheduled castes (SC) and sometimes for scheduled tribes (ST).

A conceptual difficulty encountered here is to define independent candidates empirically. This is because many bizarre candidates stand in Indian elections under the garb of a political party.[11] Election Commission publications list 173 parties categorized as "Registered (Unrecognised) Parties", sometimes with strange names like *Bharatiya Muhabbat Party (All India)* or *Vijeta Party*. Together, these 173 parties fielded a total of 898 candidates in the parliamentary election of 2004. [12] Barring a few exceptions, majority of them did not join any coalition involving National or State parties. Further, many of them floated candidates in a single constituency only. As majority of these candidates behaved like independent candidates and economic or political theories do not provide clear answer to their status, in this paper we define independent candidates as per official Election Commission categorization.

Table 1: Descriptive Statistics on Number of Candidates per Parliamentary Seat

Series	Mean	Standard Deviation	Skewness	Kurtosis	Minimum	Maximum
(1)	(2)	(3)	(4)	(5)	(6)	(7)
INDCAN	4.39	3.77	1.89 (#)	6.74 (#)	0.00	30.00
NONIND	5.62	2.25	1.04 (#)	1.56 (#)	1.00	16.00
TOTCAN	10.01	4.92	1.30 (#)	3.04 (#)	2.00	35.00

Table 1 compares the statistical features of independent (INDCAN), non-independent (NONIND) and total number of candidates (TOTCAN) per

[11] The situation is not unique in case of India. Bizarre candidates stood under such labels as BELLS (Ban Every Licensing Law Society) or SBILP (Southport Back in Lancashire Party) in the 1983 parliamentary election in the United Kingdom and could actually manage to get 75 and 374 votes respectively (Moores, 1987).

[12] Individually, only 8 among these 173 parties managed to get more than 0.10% of the total votes polled in the 2004 election.

constituency. Table 1 reveals that on average, a constituency in India had more than four independent candidates in the 2004 parliamentary election. This number is smaller than the average number of non-independent candidates per constituency. However, Table 1 also shows that variation in the number of independent candidates is high. Further, all important summarized measures other than mean (e.g., range, standard deviation, skewness and kurtosis) are higher for INDCAN than for NONIND.

We now examine to what extent the variation in INDCAN across constituencies could be explained through covariates. The first covariate of interest is the size of the total electorate. In the context of the US, Lem and Dowling (2006) have recognized that the impact of population size on emergence of smaller parties and independent candidates would be ambiguous (p. 473). This observation is consistent with our theoretical framework, in which, the number of independent candidates, *ceteris paribus*, will not depend upon population size but on political fragmentation proxied by the average number of non-independent candidates participating in an election. Lem and Dowling (2006) have, however, obtained a positive impact of constituency size on the emergence of independent candidates. In their estimated Negative Binomial regression models, both population size and population density turn out to be strongly significant.

The year 2004 elections in India give us a rare opportunity to examine the impact of population size because in that year assembly elections were also held in several of the Indian States. These States are Andhra Pradesh, Arunachal Pradesh, Karanataka, Maharashtra, Orissa and Sikkim. Other than Arunachal Pradesh and Sikkim, the remaining four States in India are large States. Together the six States cover 143 constituencies in the Indian Parliament.

Note that the simultaneous or near-simultaneous holding of assembly and parliamentary elections in these States work like a controlled experiment where impacts of other factors are likely to be similar, so that the partial impact of increase in population size could be examined. One exception which weakens our argument is the difference in the security deposit money in case of parliamentary (Rs.10,000/- for general and Rs.5,000/- for SC and ST candidates) and assembly elections (Rs.5,000/- for general and Rs.2,500/- for SC and ST candidates. We, however, stress that this difference is small and symbolic and hence is unlikely to play an important role in elections. Further, it may be noted that besides the security deposit, all major electoral formalities in India (e.g., signature requirements etc.) are same for assembly and parliamentary elections. It may be noted that roles of these other formalities in increasing transaction costs cannot be ignored. Further, international evidence suggests that indirect transaction costs are more important determinants than deposit money per se. For example, Stratmann (2005) observes that collection of signatures is likely to involve larger monetary and nonmonetary costs than the average filing fee.

Table 2 presents the number of independent candidates per constituency in case of both the parliament and the assembly elections in select States. Note that average population size in parliamentary constituencies would be about six times larger than that in assemblies for all States in Table 2. Still, except in Arunachal Pradesh and Sikkim, in all cases the number of independent candidates per constituency is slightly more in case of assembly

elections in Table 2. The stability of these numbers appears to be consistent with an implication of our model that population size may not be an important determinant of independent candidates.

To estimate the impact of the population in the Negative Binomial model, we have defined a variable named TOTVOT which measures the total number of voters (in millions) in a constituency. It may be noted that constituency sizes in India were based on the distribution of population in the year 1971. As subsequent population growth rates in different parts of India were different, TOTVOT varies substantially across constituencies. The unequal constituency size enables us to test the impact of population size in the estimated Negative Binomial regression models in a similar manner as in Lem and Dowling (2006).

Table 2: Number of Independent Candidates in States where Elections in Both Parliament and Legislative Assembly Took Place in the Year 2004

State	Parliamentary Election			Legislative Assembly Election		
	Total Seats	Independent Candidates [Total Candidates]	Independent Candidates per Seat	Total Seats	Independent Candidates [Total Candidates]	Independent Candidates per Seat
(1)	(2)	(3)	(4)	(5)	(6)	(7)
Andhra Pradesh	42	114 [279]	2.71	294	872 [1896]	2.97
Arunachal Pradesh	2	6 [13]	3.00	60	48 [168]	0.80
Karnataka	28	40 [172]	1.43	224	442 [1715]	1.97
Maharashtra	48	151 [412]	3.15	288	1083 [2678]	3.76
Sikkim	1	0 [4]	0.00	32	16 [91]	0.50
Orissa	21	26 [100]	1.24	147	295 [802]	2.01

Note: The numbers within third bracket are total number of contestants .in that state.

An important source of heterogeneity in the number of independent candidates that is related to population size is the reservation status of the constituency itself. In the 2004 Parliamentary election, 79 seats were reserved for SC and 41 for ST candidates. ECI data on candidates reveal that the numbers of independent candidates in General, SC and ST constituencies are 4.88, 3.30 and 1.44 on average. In the econometric models, two dummies named SC and ST are specified to capture the impact of reserved constituencies.

Another important source of variation in INDCAN is the extent of urbanization in a constituency. In this paper, we have categorized the status of urbanization of a constituency based on the CCA and the HRA classification. The dummy variable METRO reflects the highest category and consists of six cities, viz., Delhi, Bombay, Kolkata, Chennai, Bangalore

and Hyderabad. Together, they consist of 26 parliamentary seats.[13] TIER II cities are cities listed as category B cities in both CCA and HRA classification. Seven constituencies match with these cities, viz., Ahmedabad, Pune, Kanpur, Surat, Jaipur, Lucknow and Nagpur. TIER III cities are category C and D cities in terms of CCA and HRA classification. Together, they span 35 constituencies.[14] It may be noted that together the CCA and HRA classifications A – D cover almost all the major State capitals. All remaining constituencies are categorized as OTHERS. We observe that constituencies in OTHERS have only 3.99 independent candidates on average. Similar figures for TIER III, TIER II and METRO cities are 5.97, 8.57 and 8.41 respectively.

Empirically, defining constituencies that attract more media attention is not straightforward. Typically, media attention is more in constituencies where an elite politician or a celebrity is a candidate. Our theory is that these constituencies will have more independent candidates because: (i) a certain section of the population may have disillusions against incumbent politicians and their families, and, (ii) free riders would target these candidates to get mileage out of the extra media attention.

An empirical problem in constructing the variable is the lack of reliable data on media attention or recall factor for each candidate. An occasional choice pursued currently is to proxy the recall factor by the number of Google hits. Such a measure, however, is not without problems – especially if candidate names are common.[15] The number of Google hits, unless suitably refined, will not therefore be appropriate. Alternatives like lists maintained by experts, governmental agencies or reputed NGOs etc are also not readily available for the 2004 election in usable format.

In this study, we define a dummy variable named ELITE that takes the value unity for a constituency if a candidate in that constituency satisfies any one of the following conditions:

(1) A current or ex-cabinet minister or current or ex-speaker of the Lok Sabha

(2) A current or ex-chief minister of a State in India

Note that the variable ELITE basically characterizes senior politicians. To capture the spirit, we have, therefore, made a few exceptions to the above rule. Four senior politicians, viz., Suresh Kalmadi (Pune), Mani Shankar Aiyar (Mayiladuturai), Somnath Chatterjee (Bolpur) and Shibu Soren (Dumka) have been considered elite politicians based on judgment. Each of the four maintained a high profile in the media either due to their position or due to their action. Suresh Kalmadi was an important sports official and headed the Indian Olympic Association for a long time. Mani Shankar Aiyar was a career bureaucrat whose views and articles used to appear in the media

[13] We have also included a few congested suburbs (e.g., Jadavpur and Dum Dum in case of Kolkata) as parts of these metros. In case of Hyderabad and Kolkata, their twin counterparts (e.g., Secundrabad and Howrah respectively) have also been included as METRO constituencies.

[14] Bhattacharya (2010) presents the detailed lists of constituencies belonging to METRO, TIER II and TIER III.

[15] In the Indian context, duplication of candidate names could be a conscious electoral choice to confuse the voters – an issue that we discuss in this paper later.

on regular basis. Somnath Chatterjee was the leader of the Left in the parliament for a number of years. Shibu Soren headed *Jharkhand Mukti Morcha* and had important influence in the coalition politics at the central level. Including the above four, the total number of ELITE constituencies in our database is 63.

The second dummy variable constructed to capture a similar impact is called DYNAST. The variable DYNAST is unity for constituencies in which family members (e.g., wife/son/daughter/brother/sister/in-laws etc.) of elite politicians were candidates. Our data set has 29 such constituencies.

Finally, another dummy variable constructed to capture the impact of VIPs is called CELEB. CELB is unity for constituencies in which noted celebrities (e.g., film-stars, artists, musicians, players etc.) have appeared as candidates. Our data set provides 17 such constituencies.[16] It may be noted that although in theory a constituency could ELITE, DYNAST and CELEB constituency simultaneously, instances of duplication are rare.

We note that ELITE constituencies on an average have about 5.5 independent candidates vis-à-vis 4.2 in case of non-ELITE ones. A similar pattern is observed for DYNAST where the average numbers are 6.3 and 4.3 respectively. For CELEB, the averages are 5.6 and 4.4 respectively. Interestingly, if one considers the constituencies for the leader and the deputy leader of the incumbent government (A. B. Vajpayee in Lucknow and L. K. Advani in Gandhinagar) as well as the leader of the opposition (Sonia Gandhi in Rae Bareli)[17], these three constituencies have 18, 8 and 9 independent candidates respectively.

In so far as political fractionalization is concerned, it can be measured in alternative ways. One approach is to use traditional indices of political competition as in Holbrook and Van Dunk (1993).[18] Holbrook and Van Dunk's indicator is an average of four components, viz., (1) the winning candidate's percentage of the popular vote, (2) the winning candidate's margin of victory, (3) whether or not the seat is "safe" (winning percentage of 55 percent or more), and (4) whether or not the race was contested. These four components are then averaged into a single index.

Another approach to measure political fractionalization would be through shares of votes of different political parties. Mimicking market share indices, one may define vote shares of top few (say, 2 or 3) political parties as an indicator of fractionalization. Alternatively, one may also compute Herfindahl-Hirschman indices for different constituencies or States based on vote shares.

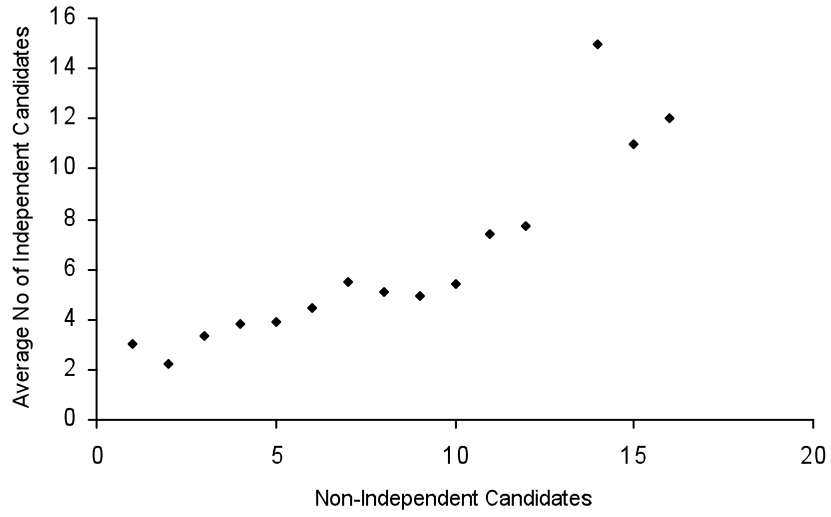
[16] See Bhattacharya (2010) for a detailed list of ELITE, DYNAST and CELEB constituencies.
[17] Officially, there was no clear deputy leader of the opposition prior to the election.
[18] Another popular measure of political competition is the Ranney Index. However, it is sometimes felt that Holbrook and Van Dunk's measure is a more reliable indicator of party competition. For example, Lem and Dowling (2006) has argued that the Ranney Index taps more directly into party control and is not as exhaustive as the measure proposed by Hollbrook and Van Dunk (1993).

While these indices would reflect different aspects of political fractionalization in a constituency, a common econometric argument against their use will be the endogeneity problem. In case of 2004 parliamentary election in India, if one uses the data on that same election to compute fractionalization, that would be an *ex post* computation. Ideally, one needs *ex ante* measures of political fractionalization. One problem with *ex ante* measures is that the data on earlier parliamentary elections that took place in India in 1999 would not reflect the current situation if not anything else but for the popular saying that a week is a long time in politics.

In this study, we attempt to measure political fractionalization by the number of non-independent candidates. Note that while this definition of political fractionalization may not be an ideal one, this is consistent with our theoretical framework because our theory suggests an increasing relationship between NONIND and INDCAN. In case of the 2004 parliamentary election, the correlation coefficient between INDCAN and NONIND turns out to be 0.29.

Empirically, political fractionalization may be measured at the constituency level as well as at the State level. Several studies in the past have pointed out the States as important cleavages in the Indian context (Mitra and Singh, 1999; Chhibber, 1999). In this study, we, therefore, attempt to measure fractionalization both at the State and at the constituency level.

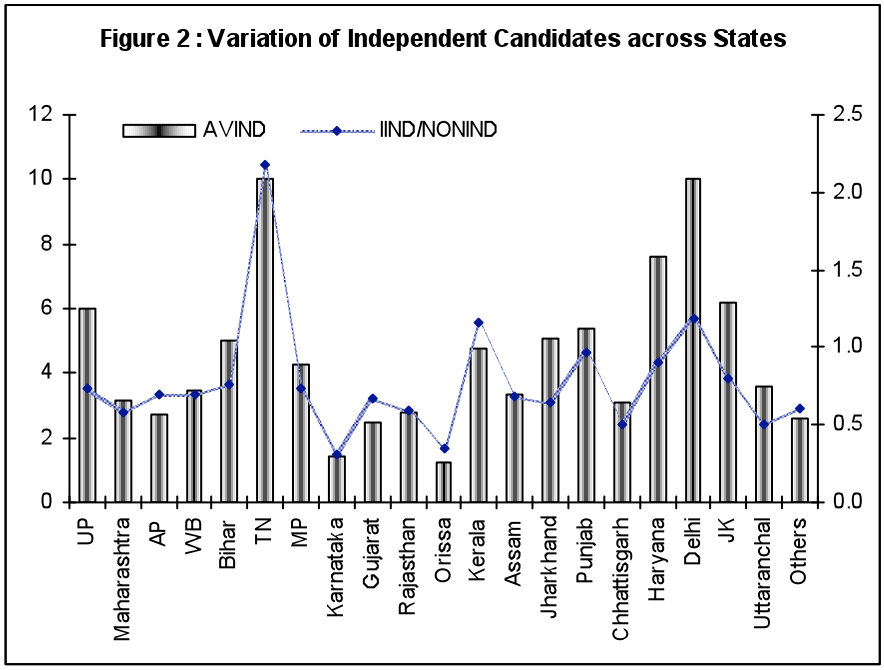
Figure 1:
Average Number of Independent Candidates in Constituencies with Specific Number of Non-Independent Candidates



When NONIND is averaged at the State level, the correlation of INDCAN with that variable (AVNONIND) increases marginally to 0.30. To capture fractionalization at the constituency level, we compute a variable named CONSTITUT which is the deviation between NONIND and AVNONIND. The variable CONSTITUT reflects the excess number of non-independent

candidates in the constituency compared to the average number of non-independent candidates in a State. Interestingly, correlation coefficient between INDCAN and CONSTITUT turns out to be only 0.10, indicating that INDCAN is more affected by fractionalization at the State level rather than at the constituency itself.

Figure 1 plots the average number of independent candidates for constituencies with specific number of non-independent candidates. Figure 1 reveals a strong increasing relationship between the two variables.[19] The weighted correlation coefficient between the two variables in Figure 1 is about 0.88.



The estimated weighted least squares equation is as follows:

(3.1) $$\text{AVINDCAN} = 1.6771 + 0.4834\ \text{NONIND}$$
$$(3.91) \quad (6.82)$$
$$R^2 = 0.93$$

[Note: The bracketed figures are estimated t-statistics]

Finally, INDCAN also fluctuates widely across States. Figure 2 plots the spatial distribution of average number of independent candidates (AVIND, bars in Figure 2) along with the ratio of independent to non-independent candidates (IND/NONIND, lines in Figure 2) across States.

[19] In Figure 1, the number of constituencies with 1 non-independent candidate was only one. Similarly, the number of constituencies with more than 12 independent candidates was only 4. Hence, the rightmost point and the three leftmost points in Figure 1, the averages are computed on small number of observations. Therefore, these points are not reliable.

Figure 2 highlights that while the average numbers of independent candidates are only 1.24 and 1.43 for Orissa and Karnataka respectively, the same numbers are as high as 10.00 and 10.02 in case of Delhi and Tamil Nadu. Further observations reveal that the IND/NONIND ratio in Figure 2 appears to be fairly stable and generally, is within the range of 0.5–0.7 for most States. The ratio is low in case of Orissa and Karnataka (less than 0.5), but high in case of Tamil Nadu (2.2).

The case of Tamil Nadu deserves special attention because it provides additional circumstantial evidence of cloning. A particular way of confusing the voter would be to float candidates with the same name. It appears that this practice had been more prevalent in the state of Tamil Nadu than in other States. Because of certain features in Hindu Tamil naming convention, obtaining candidates with same name is easier in case of Tamil Nadu than in other States of India. Out of 39 parliamentary constituencies in Tamil Nadu, we encounter the phenomenon of duplication of name of at least one of the non-independent candidates with an independent candidate in ten constituencies. These constituencies are: Madras North, Chengalpattu, Arakkonam, Tiruppattur, Vandavasi, Tiruchengode, Pollachi, Karur, Sivaganga and Sivakasi. Out of these, the case of Arakkonam highlights the malaise. In this constituency, the main fight was between "Velu, R" of PMK and "Shanmugam, M" of ADMK. Twenty candidates fought in that constituency. Among these, eight were non-independent and twelve were independent. Interestingly, there were 4 independent candidates with names "Velu" and 2 independent candidates with names "Shanmugam". "Velu R" won by a margin of well over 1,00,000 votes. Despite these observations, one cannot still prove that the independent Velu's and Shanmugan's were clones and not just free riders. However, given that the 1999 parliamentary election in the same State did not experience any duplication of candidate names of this proportion, the phenomenon certainly deserves more scientific attention along with sociological explanation.

Because of the high number of independent candidates with significant duplication of candidate names, Tamil Nadu is being treated as a separate and somewhat anomalous case in our study. The dummy TN reflects constituencies in the state of Tamil Nadu. We have also added two other dummies for the States of Karnataka (KARNATAKA) and Orissa (ORISSA) for our model.

Table 3 presents the results of the fitted Poisson and Negative Binomial models. Results in Table 3 appear to be consistent with the descriptive data analysis. All the coefficients have expected signs. Further, except the variables TOTVOT, CONSTITUT and CELEB, all other estimated coefficients turn out to be statistically significant in case of both the Poisson and the Negative Binomial model. A point to note is that all common coefficients of both Poisson and Negative Binomial models other than the INTERCEPT are close in Table 3. As these common coefficients attempt to explain the mean, this is not surprising.

Table 3: Estimated Poisson and Negative Binomial Regression Models

Variables	Poisson		Negative Binomial	
	Coefficients	*t*-statistic	Coefficients	*t*-statistic
(1)	(2)	(3)	(4)	(5)
INTERCEPT	0.2101	2.14**	0.8373	3.88 ***
TOTVOT	0.0769	1.34	0.1001	1.16
SC	-0.3506	-4.96 ***	-0.3143	-3.07 ***
ST	-0.8917	-7.10 ***	-0.8631	-4.94 ***
METRO	0.4566	7.27 ***	0.4605	4.84 ***
TIER2	0.5885	5.20 ***	0.5991	3.50 ***
TIER3	0.2509	4.07 ***	0.2507	2.67 ***
ELITE	0.1921	4.18 ***	0.1758	2.35 **
DYNAST	0.2318	4.08 ***	0.2093	2.30 **
CELEB	0.0976	0.53	0.1293	0.49
AVNONIND	0.1827	14.62 ***	0.1840	9.81***
CONSTITUT	0.0099	0.93	0.0075	0.47
TN	1.1129	24.07 ***	1.1124	15.41 ***
ORISSA	-0.5986	–3.12 ***	–0.5733	–2.10 **
KARNATAKA	-0.8948	–5.16 ***	–0.8374	–3.67 ***
DELTA			1.9511	6.15 ***
Sample Size	543		543	
Log-Likelihood	-1503.15		-1213.54	
AIC	3036.30		2459.08	
SIC	3100.76		2527.83	
TAPE	0.16 [0.53]		0.13 [0.15]	

Notes:

1. Here ***, ** and * denote significance at 1.0, 5.0 and 10.0 per cent level of significance only.

2. The numbers in third bracket for TAPE denotes the computed values of TAPE for the respective simple models without covariates.

Results in Table 3 help us to identify three major factors behind variations in the number of independent candidates. First, a major determinant of the number of independent candidates is political fractionalization. Results suggest that the number of non-independent candidates would typically lead to more independent candidates in the fray. Results further corroborate that the major determinant appears to be fractionalization at the State level rather than at the constituency itself.

Second, Table 3 highlights that *ceteris paribus*, constituencies with VIPs or their family members have more independent candidates. Interestingly, constituencies with other celebrity candidates do not reveal this pattern. As veteran politicians and their families tend to chose "safe" seats, it is likely that due to lack of sufficient incentives, these seats would be relatively free from the competition of floating independent candidates. Therefore, one possible explanation is that this phenomenon could be a manifestation of disillusion of a certain section of the population about the existing political incumbents. For other celebrities, who are relatively non-veterans, the "anger" may not be as much. A second possible explanation is that as good many among the other celebrities are first-time entrants in

electoral politics, typical anti-establishment arguments may not be an adequate "selling point" for free riders to snatch the limelight away from celebrities who are often veterans in media management themselves. It may be noted that for all practical purpose, the set of disillusioned independent candidates who specifically target political veterans and families behave like free riders. It is, therefore, difficult to trace the exact motive unless further data on independent candidates become available.

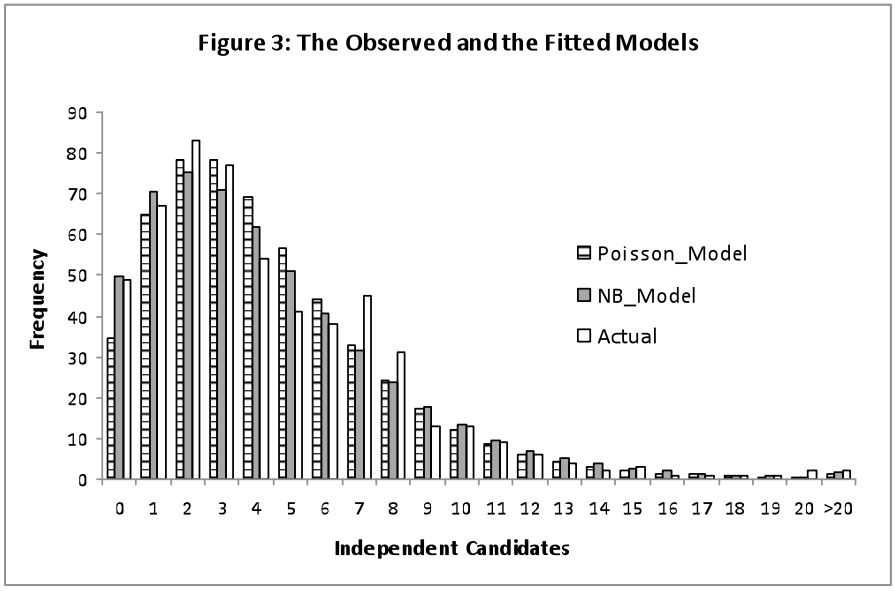
Third, our results suggest that although population size is not a significant determinant of independent candidates in India, SC and ST constituencies had significantly lower number of independent candidates. This is because potential supply of independent candidates is much less in these constituencies compared to unreserved ones.

It may be noted that for goodness of fit measures, the standard chi-square test cannot be carried out because the possible values taken by the dependent variable may not be high compared to the number of parameters. One can increase the number of cells by separately looking across various covariate groups. However, there could be too many possible covariate combinations. Hence, if the number of observations is not too high, there would be many cells with only a few observations and this will make the chi-square test invalid. In this paper, we measure goodness of fit by Total Absolute Prediction Error (TAPE) as defined by Cooil (1991).[20]

For both the Poisson and the Negative Binomial models, the computed values of TAPE are reported in Table 3. Expectedly, the Negative Binomial model fits better than the Poisson model. Poisson models grossly underestimate the probability of zero independent candidates and overestimate that of one candidate. Both the models are improvements over their simpler counterparts. However, though TAPE for the Negative Binomial model with covariates is the lowest in Table 3, the gain in terms of log-likelihood and other model fitting criteria is small vis-à-vis its simple counterpart. In general, the estimated *t*-statistics for the coefficients in the Negative Binomial model also turn out to be smaller than those of the Poisson model. Together, they imply that even the simple Negative Binomial model would be a reasonably good model to begin with.

Figure 3 presents the actual and the fitted models with covariates. Figure 3 indicates a bi-modal pattern in the observed frequency distribution of the independent candidates. The bimodality suggests that the observed distribution could be weighted sum of two standard unimodal distributions. Importantly, the theoretical framework in Section 2 suggests that mixture Negative Binomial models could be a more general choice compared to simple ones. In this paper, though the fits of both the Poisson and the Negative Binomial models good, they are unimodal models and, therefore, they fail to explain the local peaks in INDCAN around 7 and 8.

[20] See Appendix A for a brief discussion.



To examine the stability of the estimates, we carry out a small bootstrap analysis for the Negative Binomial model. We randomly divide constituencies into two parts, viz., Subsamples (S-S) 1 and 2 respectively. We estimate the model separately for each of the two subsamples. This exercise is replicated four times.

Table 4 presents the results of the bootstrap exercise. Table 4 reflects that barring a few exceptions for statistically insignificant coefficients, all other estimated coefficients are of the same sign with the corresponding coefficients in Table 3 and are close to them in general. Further, most of the estimated coefficients in Table 3 lie between the intervals created by the corresponding coefficients in a particular bootstrap. Even in a few cases where an estimated coefficient in Table 3 does not lie in such intervals, either the interval is too narrow or the coefficient in Table 3 misses the interval narrowly.

Table 4 also presents the in-sample and out-of sample forecast performances based on the bootstrap sub-samples. In sample TAPE's are calculated based on the estimated parameters of a sub-sample on the observed values in that sub-sample. Out-of-sample measures, in contrast, are computed using the observations in the other sub-sample as the actual ones. Table 4 reveals that the computed out-of-sample TAPE statistic has fluctuated between 0.12 and 0.26, with median values in the range of 0.20. These values are reasonable and together they imply that the model will not be affected too much by sampling fluctuation. The average in-sample TAPE value in Table 4 is also in the range of 0.16–0.17, not too distant from the TAPE value of 0.13 of the Negative Binomial model in Table 3. The goodness of fit statistics obtained from bootstrap forecast measures, therefore, enhance the credibility of the model.

Table 4: Results on Bootstrap

	Bootsrtap 1		Bootstrap 2		Bootstrap 3		Bootstrap 4	
	S-S 1.1	S-S 1.2	S-S 2.1	S-S 2.2	S-S 3.1	S-S 3.2	S-S 4.1	S-S 4.2
(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
Intercept	0.64	1.25	0.66	1.14	1.09	0.64	1.66	0.15
TOTVOT	0.37	-0.08	0.12	0.10	**#0.14**	**#0.11**	-0.01	0.35
SC	-0.29	-0.37	-0.25	-0.39	-0.35	-0.28	-0.28	-0.40
ST	-1.13	-0.65	-0.81	-0.91	-0.68	-1.03	-1.03	-0.66
Metro	0.33	0.54	0.48	0.43	0.58	0.33	0.43	0.61
TIER2	0.49	1.03	0.53	0.64	0.72	0.56	0.63	0.53
TIER3	0.16	0.48	0.08	0.30	0.31	0.19	0.25	0.21
ELITE	0.21	0.14	0.28	0.09	0.18	0.18	0.12	0.19
DYNAST	-0.10	0.44	0.19	0.30	0.06	0.32	0.29	0.17
CELEB	0.14	-0.02	0.16	0.06	0.09	0.19	0.13	0.24
AVNONIND	0.15	0.21	0.20	0.16	0.18	0.19	0.16	0.21
CONSTITUT	-0.01	0.03	0.02	0.00	0.00	0.01	0.01	0.01
TN	0.97	1.23	1.20	1.01	1.07	1.12	1.01	1.11
ORISSA	-0.64	-0.51	-0.63	-0.57	-0.32	-0.77	-1.13	-0.20
KARNATAKA	-1.07	-0.67	-0.81	-0.86	-1.05	-0.59	-0.68	-0.93
DELTA	1.69	2.98	**#2.02**	**#2.15**	2.54	1.69	3.29	1.59
Sample Size	272	271	272	271	272	271	272	271
Log-Likelihood	-605.5	-573.9	-579.0	-607.6	-591.1	-595.8	-583.9	-598.4
In-Sample TAPE	0.18	0.21	0.23	0.10	0.18	0.13	0.14	0.19
Out-of-Sample TAPE	0.22	0.23	0.26	0.20	0.20	0.14	0.17	0.12

Notes:

1. S-S i.j reflects results based on the data in Subsample 1 in the first random allocation exercise.
2. Here, # before the estimated coefficients (in bold font) in bootstrap i (i=1,2,,3,4) indicates that the corresponding coefficient does not lie between the interval created by the estimated coefficients in S-S 1 and S-S 2 for that bootstrap.

SECTION 4: POLICY IMPLICATIONS

From social perspective, emergence of too many independent candidates is not a healthy sign in a democracy. McKnight (1999), in the following paragraph summarizes the rationale of regulating independent candidates succinctly:

> "There are two major reasons why a state may want to regulate candidates who run without party affiliation. First, a state has an interest in protecting the integrity of the political party process. This interest includes the goals of (1) stopping intraparty fights after the primary in order to have a settled contest, and (2) preventing one party from "raiding" another's votes by running a party candidate disguised as an independent. Second, a state

> has an interest in protecting the integrity of the electoral process. This includes such concerns as verifying that a candidate is truly independent, keeping the ballot a manageable length, and ensuring that candidates on the ballot are serious and have a minimal level of support that would make it possible for them to govern if elected."

Political scientists have suggested different policies to prevent the emergence of independent candidates in elections. Some, especially in the context of the US, have actually been implemented. Although these policies vary in detail, the underlying idea is to increase the transaction costs of independent candidates to a level so that competing in an election becomes "unprofitable". McKnight (1999) discusses three basic types of statutes used in the US to regulate independent candidates, viz., (i) the early filing deadline, (ii) the "sore loser" statute, and, (iii) a disaffiliation statute.

Early filing deadline forces candidates to declare themselves as independents or members of a specific party at an early date. The apparent theory is that requiring early filing will prevent someone who is really a disappointed party member from seeking office later in the election as an "independent" (McKnight, 1999). A "sore loser" statute prevents a defeated primary candidate from appearing on the general election ballot as an independent. Under disaffiliation statute, a candidate must certify that she or he does not belong to any political party in order to run as an independent.

In the context of the US, these regulations have increased the transaction costs of small party and independent candidates substantially. However, it may be noted that some of these regulations like disaffiliation etc. require open access to political party membership data and are not feasible in the context of developing countries. Some other electoral restrictions that are popular in existing democracies involve increasing the requirements minor party candidates must meet to appear on the ballot. These restrictions typically include filing fees and signature requirements.

Among these two, increasing the signature requirements appears to be more effective in controlling the entry of independent candidates (Stratmann, 2005, p.62). The signature requirements in case of some of the states in the US gubernatorial elections could be very high. For example, Alabama requires that an independent candidate or new party obtain signatures equal to 3 percent of the previous gubernatorial vote (Lem and Dowling, 2006). For ballot access in the 2002 election, this would require 39,536 signatures. New Jersey electoral law, on the other hand, requires only 800 signatures, which is a static requirement (Lem and Dowling, 2006).

Based on this experience, a few possible policies in the Indian context could be:

- To increase the signature requirements for independent candidates substantially. The signature requirements could be relaxed for those independent candidates who had demonstrable success in earlier local or assembly elections.[21]

[21] Independent candidates may directly enter as a candidate in a Parliamentary election with the support of a political party.

- To empower the Election Commission to impose a differential deposit scheme for different types of constituencies (e.g., as the problem due to free riders is more in urban constituencies, deposit money charged for these constituencies could be more) in so far as independent candidates are concerned.

Given the incentive structure in the political game, it is also necessary to examine the detailed asset data and income tax returns of all independent candidates for a few subsequent years after participation in elections.

Incidentally, while all major parties currently have incentive to cheat each other under the existing rules of the game reported in this study, they also have incentives to change the rules of the game in their favor vis-à-vis smaller political parties and independent candidates.[22] A classic example of changing the rule of such political games in the Indian political context is the enactment of the anti-defection legislation. Similar to anti-defection enactment, if major political parties in India favor long-term gain vis-à-vis short term ones, they can collude together to redesign the institutional constraints in their favor by increasing the transaction cost of participation of independent candidates.

SECTION 5: CONCLUSION

The paper attempted to explain the emergence of independent candidates in Indian parliamentary election in the year 2004. We specified a model of an FPTP electoral system in which political parties themselves float dummy independent candidates to gain electoral advantage. We showed that in a fragmented political milieu with relatively weak institutional infrastructure, such unsavory behavior of political parties could lead to a Prisoners' dilemma type game where each party tries to guess the number of independent candidates that would be pitted against it by others. We proposed that in this game, the perceived probability distribution of the number of independent candidates floated against one political party by another would be "memoryless". Using this property and a few other simplifying assumptions, we showed that the spatial distribution of independent candidates would follow a Negative Binomial distribution. While many studies in the past modeled count data like independent candidates in this way, in most such studies, specification of Poisson and Negative Binomial models were ad hoc and were based on statistical convenience. In this study, however, attempts were made to show through a simple behavioral model why the distribution of independent candidates could be Negative Binomial.

Our empirical results revealed that the distribution of independent candidates in India could be explained well with a Negative Binomial probability model. Our results highlighted the roles three major factors played in the emergence of independent candidates. First, a major determinant of the number of independent candidates was political

[22] For example, it is often said that Democrats and Republicans do not often agree, but one thing they do agree on is that they both dislike third parties (Lem and Dowling, 2006).

fractionalization. As suggested by our theory, empirical analysis revealed that more number of non-independent candidates in a constituency would typically lead to more number of independent candidates. Interestingly, results pointed out that in case of India, the major determinant was the political fractionalization at the State level rather than at the level of the constituency. Second, we found some evidence that *ceteris paribus*, constituencies with VIPs or their family members had more independent candidates. Interestingly, constituencies with other non-incumbent celebrities did not reveal this pattern. Third, our results indicated that although population size was not a significant determinant of independent candidates in India, reserved constituencies had significantly lower number of independent candidates due to lower potential supply of independent candidates compared to unreserved ones.

We now discuss a few limitations and possible generalizations of this study. First, our theory indicates the possibility that mixture Negative Binomial models could be a more general choice of probability distribution compared to simple ones (Appendix A). Empirical observations on the spatial distribution of independent candidates also revealed a bi-modal pattern that is typical for mixture distributions.

Second, it may be noted that in this paper, we have maintained the formal distinction of small political parties and independent candidates in India. As candidates of many small parties in India behave like independents, one may treat them as independents. However, incentives of smaller parties, especially if they float candidates in more than one constituency, may be more general and not necessarily focused on a particular constituency. Working with them, therefore, brings many complex issues like coalition formation in a democracy. Still, it is important to examine whether including them as independents would change some of the main results reported in this paper.

Third, the paper's coverage of the role of incumbency in the emergence of independent candidates was partial. Our results on VIP constituencies indicate the possibility that incumbent and established politicians disillusion a section of the population more, leading to more independent candidates in such constituencies. One can generalize this concept and test whether incumbent members of parliament or "tainted" politicians (e.g., politicians with criminal record) attract more independent candidates. To have the data on incumbency, one needs to match the electoral records of earlier elections in India, which was beyond the scope of this study. Similarly, to identify "tainted" politicians, one needs to prepare the detailed affidavit data on candidates in clear database format.

Fourth, one factor whose impact on the emergence of independent candidates could not be tested was the heterogeneity of the population in a constituency. Many constituencies in India are heterogeneous in terms of caste, religion, language and income or wealth. In these constituencies with multiple cleavages, there could be several interest groups based on complex combinations of these factors. Therefore, it is likely that the number of independent candidates in these constituencies would be more. At this stage, detailed data on all the variables across constituencies are not available in India. Common sense, however, suggests that cities will be more heterogeneous than villages. Therefore, the observed significance of

METRO, TIER2 and TIER3 constituencies in our study highlights the need to have such data for better identification of factors.

Fifth, in this paper, we hypothesized that when there is unsavory competition among political parties in floating independent candidates, there will be a last minute rush in filing nominations. Unfortunately, we could not test the hypothesis due to lack of data. A verification or refutation of the hypothesis would be an important progress in our understanding of electoral behavior and strategies adopted by political parties, especially in developing and emerging market economies. Our study, therefore, highlights the need for having this data in usable format in the public domain.

Sixth, more research is needed on the backgrounds of the independent candidates themselves. In particular, we need to know the detailed history of their occupations and interests so that the reasons for their fighting in elections and the incentives they have could be clearly quantified.

Seventh, a weakness of the paper is that a good fit of the Negative Binomial model does not provide a direct proof that political parties float independent candidates for electoral benefits. Our theories are based on incentives and empirical results are based on circumstantial evidence rather than rigorous proof. However, given the widespread actual existence of "strange bedfellows" in politics across countries and over time, more research is needed on the actual behavior of candidates in elections in India. For example, our study indicates the importance of electoral budgets in floating independent candidates. It, therefore, highlights the need for more credible auditing of the accounts of political parties and candidates, including independent candidates. Extending this argument further, a crucial part of our understanding of the incentives of all major players would also come from studies that observe post-election behavior and asset accumulation and not just during the election period.

Future research on these aspects will not only help in understanding the spatial distribution of independent candidates in India further, it will also improve transparency in the electoral process and through this, cleaner political behavior in India.

Finally, the problem addressed in this paper and the issues raised are not just relevant for India, but for all democracies with an FPTP electoral system in place. The process of globalization and emergence of multicultural societies in many old and established democracies in recent years may lead to multiple cleavages in their polity. In such situations, competition among the mainstream parties may create a political vacuum leading to the emergence of smaller parties (e.g., Liberal Democrats in case of the UK) or independent candidates (e.g., Ross Perot in case of the US). If institutional structures in democracies allow this trend to continue, it has the potential to turn an established bi-party system into a complex multi-party system. Our results indicate that such situations would lead to the emergence of more independent candidates, enhancing the complexity of political management. In extreme case, some of these independent candidates could be floated discreetly by the mainstream political parties to gain electoral advantage. Because of the nature of the incentives of political parties, FPTP electoral systems are especially susceptible to such phenomena. Such electoral systems, therefore, need to put appropriate institutional constraints in place

by increasing the transaction costs of smaller parties and independent candidates.

APENDIX A

THE OPTIMIZATION PROBLEM OF POLITICAL PARTIES AND ITS IMPLICATIONS

Suppose, there are *N* voters in a constituency (say, C). Suppose two political parties, say P_1 and P_2, are fighting an election in C. Each party assumes that in a direct contest, each of the voters will vote for both the candidates with probability (1/2). Both parties also know that if they float one or more independent candidates cloning its rival, each voter will then vote for each such independent candidate with probability δ (where δ is small and close to zero) at the cost of its rival. For example, if P_1 floats K independent candidates (viz., I_{11}, I_{12}, … I_{1K}) and P_2 floats L independent candidates (say, I_{21}, I_{22}, …, I_{2L}) then each neutral voter will vote with a (K+L+2) dimensional Multinomial probability distribution (½ - Lδ, ½ - Kδ, δ, δ, …δ). The first two components of this vector reflect probabilities in favor of P_1 and P_2 respectively and the remaining (K+L) components are in favor of independent candidates.

The objective of each political party is to win the election at minimum cost. Suppose, party P_i could afford to spend at most F_i amount on floating independent candidates. Suppose each independent candidate costs D, which is same for both the parties. However, budgets on independent candidates are closely guarded secret and therefore, both P_1 and P_2 do not know how much its rival can spend on floating independent candidates.

Then the optimization problem of party P_1 is:

$$(A.1) \qquad M_1 = m_1 \quad \text{for} \quad m_1 > 0, \; m_1 D \leq F_1, \text{ and, } E(Z_1) - E(Z_2) > 0.$$
$$= 0 \quad \text{otherwise}$$

where M_1 is the number of independent candidates floated by P_1 and $E(Z_i)$ is the expected number of votes in favor of party P_i (as perceived by P_1). The optimization problem of P_2 may also be expressed in a similar manner.

As both political parties would like to outmaneuver each other at the crucial last moment of filing nominations, any information obtained prior to that on the number of candidates floated by its rival will not be of any use implying that the number of independent candidates floated by one's rival is likely to be "memoryless", e.g.,

$$(A.2) \qquad P [M_i > (a+b) \mid M_i > b] = P [M_i > a] \quad i=1,2$$

It is well known that the property of memoryless-ness in (A.2) characterizes the Geometric distribution. Therefore, the number of independent candidates floated by each political party, in this case, will be a Geometric distribution. Since two political parties come to the decision of floating "clone"

candidates independently in our model, the resulting distribution of the total number of independent candidates would be Negative Binomial.

Generalizing, when *s* parties are having equal committed vote shares and voters vote for each political party with probability (1/*s*), each of these *s* political parties will have to float "clones" for all its (*s*-1) rivals. Note that in theory, if s political parties have equal strength in a constituency and each one floats at least one clone independent candidates against others, the total number of independent candidates would be at least s(s-1) in that constituency. The actual number in reality is likely to be less. The case of more than two political parties having equal real or perceived strength in a constituency is rare. Therefore, our model cannot suggest an exact functional relationship between the number of non-independent and independent candidates other than that the last is an increasing function of the first. The relationship, however, provides upper bound for some parameters in the perceived probability distributions. When s political parties have equal strength in a constituency and each one floats independent candidates against others randomly, total number of independent candidates becomes a random variable that is a sum of s(s-1) independent Geometric random variables. This observation suggests that even the parameters of Negative Binomial distribution could be heterogeneous across constituencies.

APENDIX B

ECONOMETRIC SPECIFICATION

Let X_k be a Bernoulli indicator random variable reflecting the choice of the k-th "neutral" voter in C, with $P[X_k = 1] = p$ and $P[X_k=0] = (1-p)$. Define $Y = \sum_{k=1}^{n} X_k$. Then, Y denotes the number of independent candidates in C.

Clearly, $Y \sim$Binomial (n,p). However, common sense suggests that in this case n would be very large and p would be very small. Hence, the probability mass function of Y can be approximated by a Poisson (λ) distribution, e.g.,

(B.1) $$P[Y = y] = (e^{-\lambda}\lambda^{y})/y!$$

where $\lambda \approx np$ and $E(Y) = Var(Y) = \lambda$.

When the dependent variable is a count variable, a Poisson specification like (B.1) is often the starting point. Empirical studies, however, suggest that in case of count variables, $E(Y) = Var(Y)$ is a restrictive specification because in cross-section, one often encounters the problem of over-dispersion, e.g., $E(Y) < Var(Y)$.

The problem of over-dispersion occurs if the λ parameter in (B.1) is not homogeneous in the population. Negative Binomial distribution is a probability distribution that satisfies the property of over-dispersion and becomes a natural specification in such cases. Negative Binomial specification may also be treated as a generalization of (B.1). If the Poisson parameter λ is treated like a random variable that follows $\Gamma(\phi, \delta)$ [e.g., a Gamma distribution with parameters ϕ, and δ respectively], then the resulting distribution turns out to be a Negative Binomial distribution with the following probability mass function:

$$P[Y = y] = \frac{\Gamma(\phi + y)}{\Gamma(\phi)\Gamma(y+1)}\left(\frac{\delta}{1+\delta}\right)^{\phi}(1+\delta)^{-y} \tag{B.2}$$

Here, the first two moments of Y are: $E(Y) = \phi/\delta$ and $Var(Y) = [\phi(1+\delta)]/\delta^2$. Therefore, the variance to mean ratio of this distribution is:

$$Var(Y) / E(Y) = (1+\delta)/\delta > 1 \tag{B.3}$$

Thus, the parameter δ takes care of the phenomenon of over-dispersion mentioned above. Note that as $\delta \to \infty$, the probability mass function of Negative Binomial in (B.2) tends to that of Poisson distribution.

In empirical studies the parameters λ in case of Poisson or ϕ and δ in case of Negative Binomial distribution may be estimated by standard techniques like method of moments or maximum likelihood.

For more general Poisson regression models with covariates, we assume that Y_i, the number of independent candidates in constituency *i,* follows Poisson (λ_i). We specify,

$$\lambda_i = e^{\mathbf{Z_i\beta}} \tag{B.4}$$

where $\mathbf{Z_i}$ is a K× 1 vector of characteristics of the *i*-th constituency and $\boldsymbol{\beta}$ is the corresponding parameter vector. The log-likelihood function for the *n* constituencies is then written as:

$$L(\boldsymbol{\beta}) = \sum_{i=1}^{n}\left[-\ln(y_i!) - e^{\mathbf{Z_i'\beta}} + y_i\mathbf{Z_i'\beta}\right] \tag{B.5}$$

The gradient and the Hessian can be written as:

$$\frac{\partial L}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n}\left[\mathbf{Z_i'}(y_i - e^{\mathbf{Z_i'\beta}})\right] \tag{B.6}$$

$$\frac{\partial^2 L}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} = -\sum_{i=1}^{n}\left[\mathbf{Z_i}\mathbf{Z_i'}e^{\mathbf{Z_i'\beta}}\right] \tag{B.7}$$

The expressions in (B.6) and (B.7) reveal some special advantages of the specification in (B.4). First, an implication of equation (B.4) is that equation (B.6) becomes analogous to the more familiar normal equations in a standard

regression specification. The regression property of this specification comes from the fact that $E(Y_i) = \lambda_i$, so that $(Y_i - \lambda_i)$ may be interpreted as the residual. Second, the specification in equation (B.4) guarantees that the estimated value of λ_i is always nonnegative, thus ensuring meaningful results (Hausman *et al*, 1984).

In case of Negative Binomial regression, we specify

(B.8) $$\phi_i = e^{\mathbf{Z}i\boldsymbol{\beta}}$$

Where $\mathbf{Z_i}$ and $\boldsymbol{\beta}$ are as in (B.4). The log-likelihood function, the gradient and the Hessian are once again standard and the method of estimation of parameters is as in the case of the Poisson regression (Hausman *et al*, 1984).

The parameter vector $\boldsymbol{\beta}$ is estimated by the method of maximum likelihood. Alternatively, an iterative nonlinear weighted least squares method may also be used. Under some mild conditions, the likelihood function becomes globally concave and convergence takes place rapidly (Hausman *et al*, 1984).

To measure the goodness of fit of the models, we follow Cooil (1991). Let p_j be the observed proportion of constituencies with j independent candidates. We define the corresponding model estimates as the average unconditional predictive probabilities

(B.9) $$\hat{p}_j = \frac{1}{n}\sum_{i=1}^{n} \hat{f}_i(j)$$

Here $\hat{f}_i(j)$ is the estimated conditional probability that constituency *i*, the number of independent candidates is *j*. These probabilities are averaged across all *n* constituencies.

Following Cooil (1991), we measures goodness of fit by the Total Absolute Prediction Error (TAPE) in estimation as:

(B.10) $$\text{TAPE} = \sum_j | p_j - \hat{p}_j |$$

BIBLIOGRAPHY

Abramson PR, JH Aldrich, P Paolino and DW Rohde, 1995: 'Third-Party and Independent Candidates in American Politics: Wallace, Anderson, and Perot', *Political Science Quarterly*, 110(3), 349 -- 367.

Alvarez RM, J Nagler and S Bowler, 2000: 'Economics, and the Dynamics of Multiparty Elections: The British 1987 General Election', *American Political Science Review*, 94(1), 131 – 149.

Balagopal K, 2004: 'Andhra Pradesh: Beyond Media Images', *Economic and Political Weekly*, June 12.

Cameron AC and PK Trivedi, 1998: *Regression Analysis of Count Data*, Cambridge University Press, Cambridge.

Canon DT, 1993: 'Sacrificial Lambs or Strategic Politicians? Political Amateurs in U.S. House Elections', *American Journal of Political Science,* 37(4), 1119-41.

Chibber PK, 1999: *Democracy without Associations: Transformation of the Party System and Social Cleavages in India,* Vistaar Publications, New Delhi.

Cooil B, 1991: 'Using Medical Malpractice Data to Predict the Frequency of Claims: A Study of the Poisson Process Models with Random Effects', *Journal of the American Statistical Association*, 86, 295–295.

Downs A, 1957: *An Economic Theory of Democracy*, Harper and Row, New York.

Grogger J, 1990: 'The Deterrent Effect of Capital Punishment: An Analysis of Daily Homicide Counts', *Journal of the American Statistical Associations*, 85, 295–403.

Gerber ER and JB Lewis, 2004: 'Beyond the Median: Voter Preferences, District Heterogeneity, and Political Representation', *Journal of Political Economy*, 112(6), 1364-1383.

Hausman J et al, 1984: 'Econometric Models of Count Data with an Application to the Patents and R & D Relationship', *Econometrica*, 52, 909–938.

Heckelman JC and AJ Yates, 2008: 'Senate Elections with Independent Candidates', *Journal of Theoretical Politics*, 20(1), 31–46.

Holbrook TM and E Van Dunk, 1993: 'Electoral Competition in the American States', *American Political Science Review*, 87 (4), 955-62.

Kothari R, 1970: *Politics in India,* Orient Longman.

Lawless JF, 1987: 'Regression Models for Poisson Process Data', *Journal of the American Statistical Association*, 82, 808–815.

Lem SB and CM Dowling, 2006: 'Picking Their Spots: Minor Party Candidates in Gubernatorial Elections', *Political Research Quarterly,* 59(3), 471-480.

Lokniti Team, 2004: 'National Election Study 2004: An Introduction', *Economic and Political Weekly*, 39(51), 5373–5382.

McKnight D, 1999: 'Constitutional Issues in Regulating Independent Candidates', Mimeo, House Research, Minnesota State Relay Service (www.house.leg.state.mn.us/hrd/hrd.htm).

Mitra SK and VB Singh, 1999: *Democracy and Social Change in India: A Cross-Sectional Analysis of the National Electorate,* Sage Publications, New Delhi.

Moores B, 1987: 'Some Consequences of a Three-Party Split in a UK General Election', *Journal of the Operational Research Society*, 38(7), 569 -- 576.

Mullahy J, 1986: 'Specification and Testing of Some Modified Count Data Models', *Journal of Econometrics*, 33, 341–365.

Osborne MJ, 1995: 'Spatial Models of Political Competition under Plurality Rule: A Survey of Some Explanations of the Number of Candidates and the Positions They Take', *Canadian Journal of Economics*, 28(2), 261-301.

Rosenstone SJ, RL Behr, and EH Lazarus, 1996: *Third Parties in America,* Princeton University Press, Princeton.

Stratmann T, 2005: 'Ballot Access Restrictions and Candidate Entry in Elections', *European Journal of Political Economy,* 21 (1), 59-71.

Wedel M et al, 1993: 'A Latent Class Poisson Regression Model for Heterogeneous Count Data', *Journal of Applied Econometrics*, 8, 397–411.

Weiner M, 1978: *India at the Polls: The parliamentary Elections of 1977,* American Enterprise Institute for Public Policy Research, Washington DC.