Independent Candidates in a Parliamentary Election in India: A Poisson Regression Model

Bhattacharya, Kaushik

Indian Institute of Management Lucknow

May 2010

Working Paper Series: 2009/10 – 034

Independent Candidates in a Parliamentary Election in India: A Poisson Regression Model

Kaushik Bhattacharya

Indian Institute of Management Lucknow
Prabandh Nagar, Off Sitapur Road
LUCKNOW – 226013
INDIA

Independent Candidates in a Parliamentary Election in India: A Poisson Regression Model

By

Kaushik Bhattacharya[†]

Indian Institute of Management Lucknow

May 2010

Abstract

The paper attempts to explain the number of independent candidates in Indian parliamentary election in the year 2004. The statistical models developed are applications and generalizations of Poisson and Negative Binomial distributions. Our results suggest that the distribution of independent candidates can be explained well with a negative binomial probability model or its generalizations. Our results also help to identify three major factors behind the 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. Interestingly, our analysis points out that the major determinant appears to be political fractionalization at the State level rather than at the constituency itself. Second, we find some indirect evidence of presence of free riders. Free riders typically stand in urban constituencies and against the so called VIP candidates. Third, our results suggest that SC and ST constituencies would have typically lower number of independent candidates due to lack of potential candidates as compared to general constituencies.

Keywords: Independent Candidates, Election, Poisson, Negative Binomial

Journal of Economic Literature Classification: D72

Running Title: Independent candidates

Address for Correspondence:

Kaushik Bhattacharya

Indian Institute of Management Lucknow;

Prabandh Nagar; Off Sitapur Road;

LUCKNOW – 226 013

INDIA

Tel: 91-522-273-6623, 91-522-273-6521

Mob : 91-94522-97642 ;

E-mails: kbhattacharya@iiml.ac.in

[†] The views expressed in this paper are personal and not necessarily of the institution the author belongs to. The author bears full responsibility for any error.

Independent Candidates in a Parliamentary Election in India: A Poisson Regression Model

1 Introduction

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. Since independence in 1947, parliamentary 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.

Besides the relational dynamics of different political parties, a major feature of Indian parliamentary elections is the number of independent candidates. During the 1996 parliamentary election in India, a total of 10,635 independent candidates appeared in a total of 543 parliamentary seats. Although in subsequent elections in 1998 and 2004, the numbers decreased to 1,915 and 2,385 respectively, they are still high, especially if one considers one of the earliest democracies like the United Kingdom as benchmark.[1]

This paper attempts to explain the reasons behind such high number of independent candidates in India and also attempts to trace its determinants. To motivate the readers, we argue that the problem is important from both theoretical and practical perspectives. We also argue that the problem is not just relevant for India only, but for all democracies.

In politics, an independent politician 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 played important roles in different countries in different stages of political development. In traditional democracies like the UK or Canada, independent candidates have won elections frequently. Independents have contributed significantly in the political developments in South Africa. In recent years, independent candidates have also played important functional roles in Russia and some other transitional economies in East Europe.

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 is: what is his/her incentive? More importantly, we ask: what type of political environments in general, and

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

what type of constituencies in particular, induce individuals to run as independent candidates? Neither standard voting theories like the median voter theory, nor the theories of coalition formation in democracies provide clear and unambiguous answers to these questions.

Theories in political science and public economics have offered different reasons behind their success like ideological gaps created by major party competition (Rosenstone *et al*, 1996).

In an ideal world, independents would hold a centrist viewpoint in a polarized political environment. They may also have a viewpoint based on issues that they do not feel that any major party addresses. Therefore, one reason for their taking part in electoral politics could be grievance against existing politicians and/or policies.[2] 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 they should not formally represent it and thus be subject to its policies. Fourth category of independent candidates could be free-riders who, by paying a limited cost, enjoy and utilize the free publicity that elections offer them. Finally, independent candidates could be cranks who run for idiosyncratic reasons that are not rational at all (Canon, 1993)

Whatever be the reason behind participation in an election, it is well known that in a politically charged atmosphere, independent candidates could play a decisive role in an election. Therefore, in a first-past-the-post 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 major political parties and eventually, to cut a deal with one of them in the long-run.

Further, a more common-sense and traditional view is that independent candidates could be dummies floated by political parties themselves. These dummies would “clone” the major rival, the underlying idea being to create confusions in the mind of voters planning to vote for its rival and through this process, divide its votes.

Internationally, there is a voluminous literature on the problems that independent candidates have caused. There is also a vast literature on the roles that independent candidates actually have or could have played in policy in different stages of political and economic development in a country. Majority of these studies are, however, in the context of developed countries with two major

[2] 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 and the “ballot papers” were in the form of a booklet.

political parties.[3] Theoretical results point out that presence of independent candidates could distort the results in an election. In the context of US, Heckelman and Yates (2008) have shown 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. Empirical analysis of senate composition carried out by the authors themselves from 1991 to 2002 seems to support the theory.

It may, however, be noted that focus of Heckelman and Yates (2008) is on the outcomes of elections and not on the factors that lead to the emergence of independent candidates. In this paper an attempt is made to model the number of independent candidates in a parliamentary constituency, both theoretically and econometrically.

Because of its lack of an effective bi-party system at the central level, India would be a good case to examine the roles that different factors play 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. One major objective of the paper is to examine the role political fragmentation plays in the emergence of independent candidates, both theoretically as well as empirically. In a first-past-the-post system, all political parties have incentive to float "clones". We attempt to show that a major consequence of this could be a prisoners' dilemma type political game leading to a loss-loss situation for all political parties. Political fragmentation could worsen the situation. Therefore, common sense suggests that, *ceteris paribus*, more political fragmentation would lead to more number of independent candidates in a first-past-the-post electoral system.

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. In the 2004 Lok Sabha election, only 5 out of 2385 independent candidates 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 the independent candidates got as a group was more than the gap between the winner and his/her nearest rival.

[3] For example, Abramson et al (1995) reviews the experience with third party and independent candidates in the context of US presidential elections.

During early years of Indian democracy, the high number of independent candidates in an election used to cost the public exchequer in India severely.[4] 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.[5] Although during recent years, electoral reforms and the use of electronic voting machines in India have taken care of many of these problems, it may be remembered that when coalition governance is the order of the day, participation of too many independent candidates may enhance complexity of political management.

This paper attempts to provide a systematic explanation of the emergence of independent candidates in an election. The theoretical framework proposed by us focuses on "cloning" of rivals carried out by political parties through dummy independent candidates. We attempt to show that the "cloning", if carried out by all political parties, would lead to a prisoners' dilemma type political game. So far as individual incentives are concerned, we attempt to identify a few factors which would induce an individual to contest a parliamentary election as an independent candidate.

Empirically, we specify a Poisson regression framework that can take care of the count data properties of a series well. Poisson regression models have a number of attractive features for econometric applications involving count data. They accommodate the integer property of the count data directly and justify aggregation of the count variable over time. Further, the equations that are used for estimation of parameters in these models are surprisingly similar to more traditional regression models. Because of these advantages, these models and its generalizations (e.g., Negative Binomial regression models) have been applied in hundreds of studies involving count data. Earlier, these models have been used to explain firms' patents (Hausman et al, 1984), number of doctor consultations (Cameron and Trivedi, 1986), daily beverage consumption (Mullahy, 1986), daily homicide counts (Grogger, 1990), number of malpractice claims (Cooil, 1991) and number of book orders (Wedel et al, 1993). Cameron and Trivedi (1998) list many other comparatively recent references.

---

[4] For example, in a few cases, election commission in India had to print out ballot booklets rather than ballot papers.

[5] Earlier, a rule followed by the Election Commission was that in the event of death of any candidate, the election would be postponed. Killing defenceless candidates to get the poll postponed was a method not unknown in India. For example, in the assembly polls of 1989, Nagi Reddy fought the Telugu Desam's Palakondarayudu at Raychoti in Cuddapah district. In the parliament polls of 1985, Palakondarayudu, who was then a candidate for parliament, was unsure of the support of the two main local factions that ruled Raychoti town. So he is said to have got an independent candidate, Guvvala Subbarayudu killed and got the election postponed. 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. Palakondarayudu was this time a candidate for the assembly. Apprehensive that he may repeat his victorious performance, YSR's man Nagi Reddy set up a pliant man of their own faction, Avula Subba Reddy by name, as an independent candidate, and allegedly killed him the day before the election to get the election to the assembly postponed (Balagopal, 2004).

The plan of the paper is as follows: Section 2 presents the basic 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. Finally, Section 4 summarizes the main findings, analyzes the policy implications and discusses some limitations and possible generalizations of this study .

2 The Analytical Framework

Consider a constituency (say, C) with n "neutral" voters.[6] For simplicity, we assume that (i) each "neutral" voter in C is eligible to run as an independent candidate, and, (ii) independent candidates in C would emerge from "neutral" voters within C only. [7] Further, the decision on whether to run in the election as an independent candidate is taken by each "neutral" voter in C independently of one another.

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, $Y \sim \text{Binomial}(n,p)$. 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.,

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

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

Traditionally, when the dependent variable is a count variable, a Poisson specification like (1) is often the starting point. Empirical studies, however, suggest that in case of count variables, mean of that variable rarely equals variance. Hence, a probability model with $E(Y) = Var(Y)$ is considered too restrictive a specification. Rather, in empirical studies involving count data in cross-section, one generally encounters the problem of over-dispersion, e.g., $E(Y) < Var(Y)$.

In cross-section, the problem of over-dispersion could occur if the λ parameter in (1) is not homogeneous in the population. To understand the possible factors that

---

[6] By "neutral" voter, we mean voters who are not members of any political party and are not guided by their rules and regulations.

[7] Theoretically, any person from anywhere of India can stand in any constituency provided some standard formalities are met.

could lead to heterogeneity across constituencies, the relationship $\lambda \approx np$ would be of help. Clearly, for a given p, if the size of "neutral" voters is different in any two constituencies, expected numbers of independent candidates in them would be different. Therefore, *ceteris paribus*, the more actively the population in a constituency takes part in the mainstream political process, the less will be the number of independent candidates in that constituency. Without constituency-wise detailed data on different party memberships, it is difficult to measure the true impact of n on λ precisely. In a few cases, however, one can observe the impact of n indirectly. For example, if some constituencies are reserved for certain categories of people, the number of potential independent candidates among "neutral" voters in that constituency decreases, leading to a smaller value of λ . In case of India, some constituencies are randomly declared as SC or ST constituencies, reserving them respectively for scheduled castes or scheduled tribes. Our theory, therefore, suggests that *ceteris paribus*, these constituencies will have less number of independent candidates.

The role of p as a determinant of independent candidates, in contrast, is complex because it interacts with n and also depends on many other factors. This is because individuals have different motives or incentives to run as independent candidates in election. Whatever be the motive, p depends on the perception of an individual running as an independent candidate that he/she would be able to affect the result and would gain from the process. It may be noted that effectiveness of individuals with fixed or nominal local support would increase as n decreases. To that extent, smaller constituencies may have more independent candidates as well.

An important determinant of p, however, would be political fragmentation. Common sense suggests that if political fragmentation is high, it may even be possible for individuals with limited popular support to affect the outcome in an election as an independent candidate. Given the heterogeneity of political fragmentation between and within States in India, it is expected that p will vary substantially across constituencies.

A third source of variation in p could be due to preference of particular types of constituencies by free riders. For example, free riders would like to target constituencies with VIP candidates in fray. They may also target metro or urbanized constituencies where media attention is likely to be more. Because of all these factors, the λ parameter may experience considerable variation across constituencies, leading to over-dispersion.

It may be noted that Negative Binomial distribution is a probability distribution that satisfies the property of over-dispersion. Hence, when one encounters over-dispersion, Negative Binomial distribution becomes a natural starting point in specification. Negative Binomial specification may also be treated as a generalization of (1). The generalization is obtained by introducing another source of randomness in the Poisson distribution. If we allow the Poisson parameter λ to be randomly distributed across the population and assume that $\lambda \sim \Gamma(\phi, \delta)$ [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:

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

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:

(3) $$Var(Y) / E(Y) = (1+\delta)/\delta > 1$$

Thus, the parameter δ takes care of the phenomenon of over-dispersion mentioned above. Note that using the expression of the probability mass function of Negative Binomial in (3), one can show that the limiting case of Poisson distribution is reached when $\delta \to \infty$.

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. Because of the availability of an easily interpretable closed-form solution with standard estimation techniques, the Negative Binomial model had been used frequently in the literature to model count data (Cameron and Trivedi, 1998). However, in most cases, the specification of a Negative Binomial model is ad hoc and based on statistical convenience.

In this paper, we show that there could also be another justification of the Negative Binomial specification if the problem of emergence of independent candidates is addressed from the perspective of political parties. To illustrate this point, we shall now specify a very simple model of political behavior.

Suppose, there are N voters in a constituency (say, C). Suppose there are only two political parties, viz., P_1 and P_2, fighting in an election in C. For simplicity, assume that the number of committed voters is m for both P_1 and P_2, the remaining (N–2m) voters in the constituency, say n, are “neutral” voters. We assume that each party knows its own committed voters as well as those of their rival. An implication of this assumption is that they know that the neutral voters will be the deciding factor in the election. Each party assumes that in a direct contest, whatever be the amount spent in campaigning on behalf of their own candidates, each of the “neutral” voters will vote for both the candidates with probability (1/2). As both P_1 and P_2 do not know which neutral voter will vote for which party, they take the total number of votes that their respective parties will obtain in the election as random variables Z_1 and Z_2 respectively. Obviously, due to symmetry, P[$Z_1 > Z_2$] = P[$Z_1 < Z_2$], implying both P_1 and P_2 perceive that they have equal chance of winning the election.

Each party now reasons in the following manner. They know that if they can float one or more independent candidates cloning its rival, each neutral voter will then vote for each such independent candidate with probability δ (where δ is very small and close to zero) at the cost of its rival. For example, if P_1 floats an independent candidate I_{11} cloning the candidate for P_2, each neutral voter will then vote with a Multinomial probability distribution (1/2, ½ - δ, δ), where the probabilities respectively pertain to (P_1, P_2, I_{11}). Thus, if P_1 floats K such independent candidates, viz., I_{11}, I_{12}, … I_{1K}, then (P_1, P_2, I_{11}, …, I_{1k}) will then have a (K+2) dimensional Multinomial probability distribution (1/2, ½ - Kδ, δ, δ, …δ). Of course, P_2 may also think in a similar fashion and may float L independent candidates against P_1, say, I_{21}, I_{22}, … I_{2L}. Clearly, success in the election would then depend upon which political party would float more "clone" 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:

$$(4) \qquad M_1 = \begin{cases} m_1 & \text{such that} \quad m_1 > 0, \quad m_1 D \le F_1, \text{ and, } E(Z_1) - E(Z_2) > 0. \\ 0 & \text{otherwise} \end{cases}$$

where M_1 is the number of independent candidates floated by P_1. The optimization problem of P_2 may also be expressed in a similar manner.

This entire process of floating clone independent candidates may be thought of as a two-person simultaneous move 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.

It may be noted that the way we have specified our model, the process of guessing the number of independent candidates floated by the rival is likely to be "memoryless". To elaborate, suppose that through its own information network, at some point during the period of filing the nomination papers, P_1 has come to know that P_2 has so far floated *b* "clone" candidates. This information will not be of much use to P_1 because:

(5) $$P [M_2 > (a+b) \mid M_2 > b] = P [M_2 > a]$$

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.

It may be noted that the property of memoryless-ness in (5) characterizes the Geometric distribution. Therefore, the number of independent candidates floated by each political party, in this case, will be a Geometric probability 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 distribution.

Undoubtedly, the above model is stylized and simple. However, we now discuss implications of some of the assumptions and also attempt some generalizations.

First, if both parties have unequal strength in a constituency and both the political parties perceive that cloning is unlikely to change the outcome but will unnecessarily inflate the cost, then cloning may not take place at all. In this model, both the political parties cannot float unlimited number of independent candidates due to budget constraints. They will float clone candidates only if they are within budget and floating such candidates changes the probability of winning in their favor. Common sense suggests that cloning battles will take place when both the parties perceive that they have equal or nearly equal strength in the constituency.

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 a "clone" floated by the other political party at zero cost.

Third, the way the model is posed, the game of floating clone independent candidates is a classic Prisoners' dilemma type game. Ideally, each political party would like to fight directly as it reduces its cost. However, any tacit understanding between the 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 tacit 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. Generalizing this, when s parties are having equal committed vote shares and "neutral" 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. Like in the classical Prisoners’ Dilemma with multiple players, detection of the cheater becomes even more difficult because even if “clones” are identified, one may not know which political party has floated the clone.

Fifth, in this model, constituency size does not play any role. Rather, the important roles are played by the perceived “gap” between the winners and losers along with the budget constraints. To outsiders, the budget constraints would be unobservable. However, an observable implication of the model is that *ceteris paribus*, number of independent candidates will not vary for parliamentary and assembly elections.

We now discuss the more general specifications involving covariates in the Poisson and the Negative Binomial model. In the presence of covariates, Y_i, the number of independent candidates in constituency *i* may follow Poisson (λ_i). Assume that λ_i varies across constituencies in the following manner, viz.,

$$\lambda_i = e^{\mathbf{Z_i}\boldsymbol{\beta}} \tag{6}$$

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:

$$\mathrm{L}(\boldsymbol{\beta}) = \sum_{i=1}^{n}\left[-\ln(y_i!) - e^{\mathbf{Z}_i'\boldsymbol{\beta}} + y_i\mathbf{Z}_i'\boldsymbol{\beta}\right] \tag{7}$$

The gradient and the Hessian can be written as:

$$\frac{\partial \mathrm{L}}{\partial \boldsymbol{\beta}} = \sum_{i=1}^{n}\left[\mathbf{Z}_i'(y_i - e^{\mathbf{Z}_i'\boldsymbol{\beta}})\right] \tag{8}$$

$$\frac{\partial^2 \mathrm{L}}{\partial \boldsymbol{\beta}\partial \boldsymbol{\beta}'} = -\sum_{i=1}^{n}\left[\mathbf{Z}_i\mathbf{Z}_i' e^{\mathbf{Z}_i'\boldsymbol{\beta}}\right] \tag{9}$$

The expressions in (8) and (9) reveal some special advantages of the specification in (6). First, an implication of equation (6) is that equation (8) 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) = λ_i, so that (Y_i - λ_i) may be interpreted as the residual. Second, the specification in equation (6) guarantees that the estimated value of λ_i is always nonnegative, thus ensuring meaningful results (Hausman *et al*, 1984).

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

In case of Negative Binomial distribution, we specify

(10) $$\phi_i = e^{\mathbf{Z_i}\boldsymbol{\beta}}$$

Where $\mathbf{Z_i}$ and $\boldsymbol{\beta}$ are as in (6). 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).

3 Data and Empirical Analysis

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 the schedules castes (SC) and sometimes for the scheduled tribes (ST).

A difficulty encountered prior to empirical analysis is to define independent candidates empirically. While the election commission website and publications have an explicit category called "Independent", many bizarre candidates stand in Indian elections under the garb of a political party.[8] 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 parties fielded a total of 898 candidates in the parliamentary election of 2004. The difficulty in defining an independent candidate arises because barring a few exceptions, majority of them did not join any coalition involving National or State parties. Therefore, for all practical purpose, majority of the candidates in this category behaved like independent candidates.[9] Despite their presence, in this paper, we stick to the official Election Commission categorization of Independent candidates and do not add them to our list of independents.

---

[8] 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).

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

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

Series (1)	Mean (2)	Standard Deviation (3)	Skewness (4)	Kurtosis (5)	Minimum (6)	Maximum (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 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, an important point to note is that if "Registered (Unrecognised) Parties" are considered as independent, then mean of INDCAN and NONIND would be 6.05 and 3.97 respectively.

Table 2: Observed and Fitted Poisson and Negative Binomial Models

INDCAN (1)	Observed (2)	Fitted Models: Poisson (3)	Fitted Models: Negative Binomial (4)
0	49	6.73	54.20
1	67	29.56	73.44
2	83	64.88	75.15
3	77	94.95	68.58
4	54	104.21	58.80
5	41	91.49	48.46
6	38	66.94	38.87
7	45	41.98	30.56
8	31	23.04	23.67
9	13	11.24	18.11
10	13	4.93	13.73
11	9	1.97	10.32
12	6	0.72	7.71
> 12	17	0.36	21.40
Log-Likelihood		-1606.37	-1368.93

Table 1 also shows that variation in the number of independent candidates is high. For example, Table 1 indicates that while some constituencies may not have independent candidates at all, in some there could be as many as 30 independent candidates. In fact, Table 1 reveals that all important summarized measures other than mean (e.g., range, standard deviation, skewness and kurtosis) are higher for INDCAN than for NONIND.

Table 2 analyzes the detailed distributional patterns of independent candidates. Column 2 of Table 2 provides the observed frequency distribution of the number of independent candidates.[10] Column 3 provides the fitted Poisson model without any covariates. A comparison of columns 2 and 3 reveals a poor fit, with the value of the computed χ^2 statistic being 1235.33 whereas the critical value at 5.0 per cent level of significance for 12 degrees of freedom being 21.03 only. This is because the distribution of independent candidates appears to have a long-tail. This concentration has led to the problem of "over-dispersion". The variable INDCAN has a variance-mean ratio of 3.24. The high variance-mean ratio indicates that the basic Poisson model is not an appropriate probability model here and a more general model is needed. To that end, a simple negative binomial model has been fitted in column 4 of Table 2. A comparison of columns 2 and 4 reveals a much closer fit. The value of the computed χ^2 statistic is only 16.50 for the negative binomial model vis-à-vis a critical value of 19.68 at 5.0 per cent level of significance for 11 degrees of freedom. Therefore, Table 2 indicates that even a simple negative binomial model without any covariates would be able to explain the variations of independent candidates across constituencies reasonably well.

We now examine to what extent the variation in INDCAN across constituencies could be explained through covariates. This is also important because it helps us to specify and test some hypotheses based on the analytical framework in Section 2. The first among them is that the number of independent candidates would typically be less in SC and ST constituencies. This is because the supply of potential independent candidates in such constituencies would be less. Second, the free riders would typically stand for election in "cities" and against the so called "VIP candidates". Third, political fractionalization would be an important determinant of the number of independent candidates in a constituency.

An important source of heterogeneity is the reservation status of the constituency itself. 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.

Another important source of variation in INDCAN would be the extent of urbanization in the constituency. Out of the 543 constituencies in the data set, some are predominantly rural, some are predominantly urban, and some are part-rural-and-part-urban. The rural-urban divide is also an important determinant of the number of independent candidates. In this paper, we have categorized constituencies 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

[10] It may be noted that the frequency distribution is truncated so that observed frequencies for each cell are at least 5 so that they can be used to compute goodness-of fit.

consist of 26 parliamentary seats.[11] The TIER II cities are the cities listed as category B cities in both CCA and HRA classification. Seven constituencies match with these cities and are included in this category. These are: Ahmedabad, Pune, Kanpur, Surat, Jaipur, Lucknow and Nagpur. TIER III cities are the category C and D cities in India in terms of CCA and HRA classification. Together, they span 35 constituencies.[12] All remaining constituencies are categorized as OTHERS. It is interesting to note that while OTHERS constituencies 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, the constitution of the variable VIP is not straightforward. The variable VIP consists of the constituencies when a candidate in that constituency satisfies any one of the following conditions:

- A cabinet minister in the current or in the immediately preceding cabinet (MINS)
- A son or a daughter of an “influential” political family (SONS)
- Celebrities (Filmstars, artists, players etc.) (CELEBS)

Our data set identifies a total of 78 such VIP constituencies. Appendix A presents the list of such constituencies along with the VIP candidate.

We note that constituencies with a VIP candidate typically have about 2 more independent candidates compared to non-VIP constituencies. 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 opposition (Sonia Gandhi in Rae Bareli and Rahul Gandhi in Amethi respectively), these four constituencies have 18, 8, 9 and 7 independent candidates respectively.

In this study, the fractionalization in a constituency is measured in two ways. In India, States are one of the major units. We identify the average number of non-independent candidates in a state as an indicator of political fractionalization in the state itself (AVNONIND). The deviation of the number of non-independent candidates in a constituency from this state-level average (CONSTITUT) is also used as a measure of fractionalization of polity in the constituency itself. The correlation coefficient between INDCAN and AVNONIND is 0.30 and that between INDCAN and CONSTITUT is 0.10.

Finally, the average number of independent candidates also fluctuates widely across States. For example, the average number of independent candidates in

[11] It may be noted that 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.

[12] Appendix A presents the detailed lists of constituencies belonging to METRO, TIER II and TIER III.

the state of Tamil Nadu is as high as 10.02, more than double the all-India average. Our analytical framework suggests that political parties may sometimes engage in floating "clone" independent candidates to confuse the voter. A particular way of confusing the voter is to have candidates with the same name. It appears that this practice had been more prevalent in the state of Tamil Nadu, due to certain features in Hindu Tamil naming convention. That is why Tamil Nadu is being treated as a separate and somewhat anomalous case. The dummy TN reflects constituencies in the state of Tamil Nadu.

Table 3: Estimated Poisson and Negative Binomial Regression Models

Variables	Poisson		Negative Binomial	
	Coefficients	*t*-statistic	Coefficients	*t*-statistic
Intercept	0.1191	1.60	0.0958	0.88
SC	-0.3682	-5.37 ***	-0.3568	-3.49 ***
ST	-0.9040	-7.40 ***	-0.9020	-6.06 ***
Metro	0.4875	7.87 ***	0.4832	3.78 ***
Tier 2	0.6785	6.27 ***	0.6840	2.88 ***
Tier 3	0.2619	4.48 ***	0.2942	2.92 ***
VIP	0.1349	3.31 ***	0.1356	1.94 *
AVNONIND	0.2098	18.28 ***	0.2126	11.55 ***
CONSTITUT	0.0157	1.57	0.0138	0.79
TN	1.1791	27.03 ***	1.1806	12.96 ***
DELTA			7.6204	6.30 ***
Log-Likelihood	-1503.1504		-1213.5400	
Sample Size	543		543	

Note: ***, ** and * denote significance at 1.0, 5.0 and 10.0 per cent level of significance only.

Table 3 presents the results of the fitted Poisson and Negative Binomial models. Results in Table 3 are consistent with the descriptive data analysis. All the coefficients have expected signs. Further, except the intercept and the variable CONSTITUT, 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 the common coefficients of both Poisson and Negative Binomial models are close in Table 3. As these common coefficients attempt to explain the mean, this is not surprising.

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. Interestingly, *t*-statistics of the estimated coefficients in Table 3 point out that in so far as the impact of fractionalization is concerned, the major determinant appears to be fractionalization at the State level rather than at the constituency itself. Second, we find some indirect evidence of presence of free riders. Free riders typically stand in urban constituencies and against the so called VIP candidates. Third, our results suggest that SC and ST constituencies would have typically lower number of independent candidates due to lack of potential candidates as compared to general constituencies.

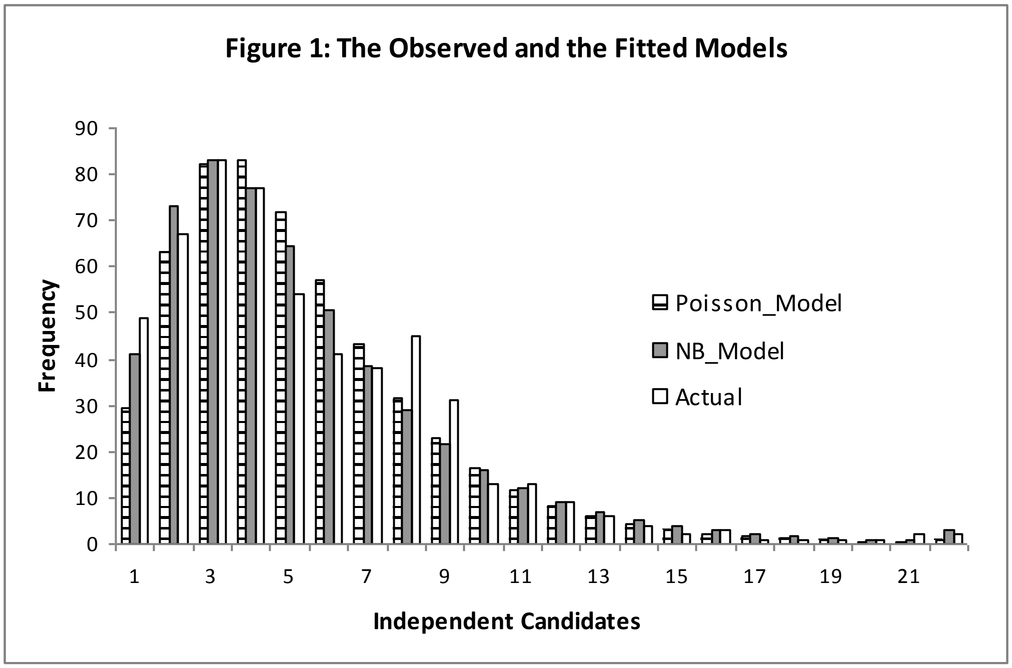

Figure 1 presents the actual and the fitted models. Figure 1 reveals a close fit of the Negative Binomial model to the data. Expectedly, the Negative Binomial model provides a much better fit than the Poisson model. Both the models are improvements over their simpler counterparts in Table 2. However, interestingly the gain in terms of log-likelihood for the Negative Binomial model in Table 3 is small vis-à-vis its simple counterpart in Table 2. 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 in Table 2 would be a reasonably good model to begin with.

For more rigorous goodness of fit measures, the standard test is the chi-square test, which compares the observed and the expected probabilities of the cells in the model. However, here the possible values taken by the variable is not high, although there are ten and eleven parameters in the Poisson and the Negative Binomial model respectively. Therefore, chi-square test cannot be conducted for the combined sample. One can of course increase the number of cells by separately looking across various covariate groups. However, the total number of possible covariate combinations is around one hundred. Hence, there will be many cells with only a few observations and this will make the chi-square test invalid. We, therefore, measure goodness of fit by other criteria.

To measure the goodness of fit rigorously, we consider an alternative approach followed by Cooil (1991) in which the distributional assumption is checked directly by seeing whether the models are able to generate the right predictive frequencies for each covariate group. 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

$$(11) \qquad \hat{p}_j = \sum_{i=1}^{n} \hat{f}_i(j)$$

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 constituencies.

Following Cooil (1991), two measures of goodness of fit are considered to compare the performance of the proposed models:

- The proportional prediction error (PPE_0) which gives the relative error in fit corresponding to the probability of zero jumps, i.e.,

$$(12) \qquad \mathrm{PPE}_0 = \frac{|p_0 - \hat{p}_0|}{\hat{p}_0}$$

- The total absolute prediction error (TAPE) in estimation, i.e.,

$$(13) \qquad \mathrm{TAPE} = \sum_j |p_j - \hat{p}_j|$$

For both the Poisson and the Negative Binomial models, the measures of PPE_0 and TAPE are reported in Table 4 respectively.

Table 4: Measures of Goodness of Fit

	Simple		With Covariates	
	Poisson	Negative Binomial	Poisson	Negative Binomial
PPE0	6.29	0.10	0.68	0.20
TAPE	0.53	0.15	0.19	0.13

From Table 4 it is clear that the negative binomial model provides a better fit to the data. While the values of TAPE for the Poisson model is 0.19, the corresponding range for the negative binomial model is about 0.13. The Poisson model grossly underestimates the probability of zero independent candidates and overestimates that of one candidate. Compared to the Poisson model, the performance of negative binomial model is far more superior. The error in estimation in the probability of no independent candidate is small for negative binomial candidates, as is being reflected by the values of PPE_0 from Table 4. The tail probabilities are also better estimated by the negative binomial model as the problem of over-dispersion is more prominent.

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

Table 5: Results on Bootsrap

	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
Intercept	-0.11	0.31	0.27	-0.04	0.15	0.01	0.26	-0.04
SC	-0.40	-0.33	-0.41	-0.31	-0.37	-0.34	-0.30	-0.44
ST	-0.71	-1.08	-0.89	-0.91	-0.70	-1.12	-1.10	-0.75
Metro	0.55	0.41	0.46	0.49	0.53	0.41	0.43	0.62
Tier 2	1.02	0.55	0.71	0.65	0.85	0.63	**#0.66**	**#0.66**
Tier 3	0.46	0.18	0.32	0.11	**#0.32**	**#0.29**	0.33	0.23
VIP	0.11	0.18	0.14	0.12	0.08	0.20	**#0.14**	**#0.15**
AVNONIND	0.24	0.18	0.20	0.23	**#0.21**	**#0.22**	0.19	0.24
CONSTITUT	0.02	0.01	0.01	0.02	0.00	0.03	0.01	0.03
TN	1.32	1.08	1.11	1.24	1.15	1.22	**#1.16**	**#1.16**
DELTA	9.64	6.87	8.99	6.95	9.01	6.94	10.99	6.05
Log-Likelihood	-589.09	-620.06	-618.02	-591.65	-608.10	-602.23	-593.47	-615.80
Sample Size	272	271	272	271	272	271	272	271

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.

Table 5 presents the results of the bootstrap exercise. Table 5 reflects that in each case, the estimated coefficients are of the same sign with the corresponding coefficients in Table 3 and are close to them in general. In general, 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 the corresponding interval in a bootstrap, either the interval is too narrow or the coefficient in Table 3 misses the interval narrowly.

Table 6 presents the in-sample and out-of sample forecast performances based on the bootstrap sub-samples. Once again, as in Table 4, the focus is on the statistics PPE_0 and TAPE. In sample PPE_0's and 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 6 reveals that the computed out-of-sample TAPE statistic has fluctuated between 0.0779 and 0.2660, with median values in the range of 0.17–0.18. 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 6 is also in the range of 0.17–0.18. It may be noted that in Table 4, negative binomial model with covariates yielded a TAPE value of about 0.13 for the whole data.

In-sample and out-of-sample PPE_0 values are also reasonable, though in one or two cases the values obtained are on the higher side. The goodness of fit statistics obtained from bootstrap forecast measures therefore enhance the credibility of the model.

Table 6: In-Sample and Out-of-Sample Forecast Performance of the Negative Binomial Model for Different Bootstrap Sub-Samples

Bootstrap and Subsamples		In Sample		Out of Sample	
		PPE_0	TAPE	PPE_0	TAPE
Bootstrap 1	S-S 1.1	0.0037	0.1713	0.1597	0.2029
	S-S 1.2	0.4233	0.2510	0.2318	0.2660
Bootstrap 2	S-S 2.1	0.1678	0.2238	0.2373	0.2600
	S-S 2.2	0.2043	0.1260	0.8439	0.1849
Bootstrap 3	S-S 3.1	0.2828	0.2019	0.0061	0.1710
	S-S 3.2	0.1236	0.1479	0.4501	0.1725
Bootstrap 4	S-S 4.1	0.2017	0.1497	0.0407	0.1442
	S-S 4.2	0.1765	0.2045	0.3584	0.0779

4 Discussion and Policy Implications

It may be noted that the political game of cloning independent candidates is a classic prisoners' dilemma game leading to a loss-loss situation for all. From social perspective, emergence of too many independent candidates is perhaps not a healthy sign in a democracy. 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 cost of the independent candidate to a level so that competing in an election becomes "unprofitable". Some of these policies like disaffiliation etc. require open access to political party membership data and are not feasible in the context of emerging market economies. However, others like increasing deposit or signature requirements have more potential in developing countries.

A few policies in the Indian context could be:

- To barr entry as independent candidate in parliamentary elections unless the candidate has demonstrable success in earlier local or assembly elections.[13]
- Another approach could be 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.

5 Conclusion

The paper attempted to explain the emergence of independent candidates in Indian parliamentary election in the year 2004. Independent candidates in a constituency being a count variable, a natural starting point in this case was to specify and estimate Poisson and Negative Binomial models and their generalizations. While many studies in the past modeled count data 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.

We showed that in a first-past-the-post electoral system, political parties have the incentive to float independent candidates who would "clone" their major rivals. In a fragmented political milieu, such behavior would lead to a simultaneous-move 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 where political parties try to out-maneuver one another, 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 more simplifying assumptions, we showed that the total number of independent candidates across constituencies would follow a Negative Binomial distribution.

Our empirical results revealed that the distribution of independent candidates in India could be explained well with a Negative Binomial probability model. Our results also 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 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

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

the political fractionalization at the State level rather than at the level of the constituency. Second, we found some indirect evidence of the presence of free riders. Free riders, typically “picking their spots”, stood in the metropolitan and Tier II constituencies and chose to fight against the so called VIP and celebrity candidates. Third, our results indicated that SC and ST constituencies had lower number of independent candidates, because potential independent candidates were much less in these constituencies compared to unreserved ones.

The problem addressed in this paper and the issues raised are not just relevant for India, but for all democracies. Even in older democracies in developed countries, competition among the two mainstream parties could sometimes create a political vacuum leading to the emergence of a third party (e.g., Liberal Democrats in case of the UK) or independent candidates (e.g., Ross Perot in case of the US). Our results indicate that such situations may lead to the emergence of more independent candidates. Some of these candidates could be floated discreetly by the mainstream political parties to clone their rivals. Because of the nature of the incentives of the political parties, first-past-the-post electoral systems are especially susceptible to such phenomena. Such electoral systems, therefore, need to put appropriate institutional constraints in place. The institutional constraints could be different for different countries. However, the main idea behind them is to increase the transaction cost of smaller parties and independent candidates.

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 take part 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. However, despite all these, we need to know whether their inclusion in independents would change some of the main results obtained in this paper.

Finally, while the paper could give some justification of applying negative binomial model through “cloning”, a major weakness of the paper is that the converse may not be true. Fit of the Negative Binomial model does indicate heterogeneity and over-dispersion in the spatial distribution of independent candidates but provides no direct evidence of cloning. 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 are needed on the emergence of independent candidates in a political game.

Reference

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.

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.

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.

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.

Miller JM and JA Krosnick, 1998: 'The Impact of Candidate Name Order on Election Outcomes', *The Public Opinion Quarterly,* **62(3)**, 291-330.

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.

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

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

Appendix A: Data Definitions

A.1 List of VIP Constituencies along with the VIP Candidate

01. Karimnagar, K. CHANDRA SHAKHER RAO (Sic!) 02. Khammam, RENUKA CHOWDHURY; 03. Miryalguda, JAIPAL REDDY SUDINI; 04. Silchar, SONTOSH MOHAN DEV; 05. Gauhati, BHUPEN HAZARIKA; 06. Hajipur, RAM VILAS PASWAN; 07. Vaishali, RAGHUBANSH PRASAD SINGH; 08. Muzaffarpur, GEORGE FERNANDES; 09. Barh, NITISH KUMAR; 10. Madhepura, LALU PRASAD & SHARAD YADAV; 11. Kishanganj, SYED SHAHNAWAZ HUSSAIN; 12. Begusarai, ; 13. Patna, C. P. THAKUR; 14. Sasaram, MIRA KUMAR; 15. Gandhinagar, L. K. ADVANI; 16. Kapadvanj, VAGHELA SHANKERSINH LAXMANSINH; 17. Surat, KASHIRAM RANA; 18. Kurukshetra, NAVEEN JINDAL & ABHAY SINGH CHAUTALA; 19. Kangra, SHANTA KUMAR; 20. Srinagar, OMAR ABDULLAH; 21. Anantnag, MEHBOOBA MUFTI; 22. Bangalore_S, ANANTH KUMAR; 23. Mangalore, M VEERAPPA MOILY; 24. Hassan, H. D. DEVEGOWDA; 25. Shimoga, S. BANGARAPPA; 26. Guna, JYOTIRADITYA MADHAVRAO SCINDIA; 27. Chhindwara, KAMALNATH; 28. Ujjain, DR. SATYANARAYAN JATIYA; 29. Rajapur, SURESH PRABHAKAR PRABHU; 30. Mumbai_S, MILIND MURLI DEORA; 31. Mumbai_NC, MANOHAR GAJANAN JOSHI; 32. Mumbai_NW, SUNIL DUTT; 33. Mumbai_N, GOVINDA; 34. Latur, PATIL SHIVRAJ VISHWANATH; 35. Pune, KALMADI SURESH; 36. Shillong, SANBOR SWELL LYNGDOH; 37. Sundargarh, JUAL ORAM; 38. Gurdaspur, VINOD KHANNA; 39. Amritsar, NAVJOT SINGH SIDHU; 40. Bikaner, DHARMENDRA; 41. Churu, BAL RAM JAKHAR; 42. Jhunjhunu, SHISH RAM OLA; 43. Dausa, SACHIN PILOT; 44. Jhalwar, DUSHYANT SINGH; 45. Jalore, BUTA SINGH; 46. Madras_C, DAYANIDHI MARAN; 47. Madras_S, BAALU, T. R.; 48. Perambalur, RAJA, A.; 49. Mayiladuturai, MANI SHANKAR AIYAR; 50. Sivaganga, CHIDAMBARAM, P.; 51. Rampur, P. JAYA PRADA NAHATA; 52. Pilibhit, MANEKA GANDHI; 53. Lucknow, ATAL BIHARI VAJPAYEE; 54. RaeBareli, SONIA GANDHI; 55. Amethi, RAHUL GANDHI; 56. Akbarpur, MAYA WATI; 57. Bansgaon, MAHAVEER PRASAD; 58. Ballia, CHANDRA SHEKHAR; 59. Allahabad, DR. MURALI MANOHAR JOSHI; 60. Kannauj, AKHILESH YADAV; 61. Mainpuri, MULAYAM SINGH YADAV; 62. Raiganj, PRIYA RANJAN DASMUNSI; 63. Malda, A. B. A. GHANI KHAN CHOUDHURY; 64. Jangipur, PRANAB MUKHERJEE; 65. Calcutta_S, MAMATA BANERJEE; 66. Bolpur, CHATTERJEE SOMNATH; 67. Mahasamund, AJIT JOGI; 68. Dumka, SHIBU SOREN; 69. Ranchi, SUBODH KANT SAHAY; **70. Khunti**, ; 71. Hazaribagh, YASHWANT SINHA; 72. New Delhi, JAGMOHAN; 73. Outer Delhi, SAJJAN KUMAR & SAHIB SINGH; 74. East Delhi, SANDEEP DIKSHIT; 75. Chandni Chowk, KAPIL SIBAL & SMRITI Z. IRAANI; 76. Delhi Sadar, JAGDISH TYTLER; 77. Lakshadweep, P. M. SAYEED; 78. Pondicherry, LALITHA KUMARAMANGALAM;

A.2 Details on Metropolitan, Tier II and Tier III Constituencies

Table A.2: Metropolitan, Tier II and Tier III Constituencies

Constituency Category (1)	**Constituencies (2)**
Metro [26]	Hyderabad, Secunderabad, Bangalore North, Bangalore South, Mumbai South, Mumbai South-Central, Mumbai North-Central, Mumbai North-East, Mumbai North West, Mumbai North, Madras North, Madras Central, Madras South, Jadavpur, Dum Dum, Calcutta North-West, Calcutta North-East, Calcutta South, Howrah, New Delhi, South Delhi, Outer Delhi, East Delhi, Chandni Chowk, Delhi Sadar, Karol Bagh,
Tier II [7]	Ahmedabad, Surat, Nagpur, Pune, Jaipur, Lucknow, Kanpur
Tier III [35]	Vishakhapatnam, Vijayawada, Guntur, Gauhati, Patna, Rajkot, Baroda, Srinagar, Mysore, Mangalore, Trivandrum, Gwalior, Jabalpur, Bhopal, Indore, Nashik, Aurangabad, Sholapur, Bhubaneswar, Amritsar, Jullundur, Ludhiana, Jodhpur, Coimbatore, Madurai, Tiruchirapalli, Varanasi, Allahabad, Agra, Asansol, Durg, Dhanbad, Ranchi, Jamshedpur, Chandigarh.

Note: Numbers within third bracket in Column 1 indicate total number of such constituencies in Column 2.