



Why do voters elect criminal politicians?

Abhinav Khemka

Department of Economics, University of Barcelona and IEB, C/ de John Maynard Keynes, 1, 11, 08034 Barcelona, Spain

ARTICLE INFO

JEL classification:

D72
D73
O12

Keywords:

Ethnicity
Criminality
India
Voter behavior
Elections

ABSTRACT

Voters across the world are often complicit in electing bad-quality legislators to political office. This problem is particularly salient in India, where candidates accused of criminality often succeed at the polls. Why do voters show a willingness to cast their ballots for candidates accused of wrongdoing? Using primary individual-level voter survey data from the Bihar 2020 state assembly elections, this paper examines whether ethnic voting can explain this surprising voter behavior. Contrary to voter preference theory, I find that voters exhibit a stronger negative response to candidates accused of criminality when they belong to their preferred ethnic party. Voter support for the non-ethnic falls by 89.2% for violent charges. Coethnicity further reduces electoral support by 67% for violent criminals. This pattern holds regardless of the voters' level of news consumption, political knowledge, education status, and income. These findings suggest that the electoral success of criminal politicians could be attributed to other factors such as a lack of proper institutions or lower state capacity rather than the voters' underlying ethnic preferences.

“So what if he is a criminal? I vote for him because at least he is from my jati (caste).”

— Voter Interview, 16th March 2022

1. Introduction

In theory, free competitive elections would prevent corrupt or criminal politicians from being elected to office and discourage those in power from committing acts of misconduct (Caselli and Morelli, 2004; Besley, 2007). However, the empirical evidence is inconclusive: while some studies find that voters hold bad politicians accountable (Klašnja, 2017), others do not (Chang and Kerr, 2017). What is prompting voters to routinely select candidates of disrepute for public office?

In this respect, India provides an ideal setting to examine why corrupt or criminal politicians win. Although holding massive democratic elections with multiple parties, voters often elect and re-elect criminally accused candidates at the polls. For example, in the last *Lok Sabha* (national) election of 2019, 43% of the Members of Parliament (MPs) faced criminal allegations, out of which 30% were accused of “serious” crimes such as rape, kidnapping, and murder.^{1,2} In comparison, in the previous 2014 national elections, 34% of the MPs faced criminal charges, of which 22% were accused of committing serious offenses.

One plausible explanation for this steady rise in the electoral success of criminal politicians could be attributed to the voters' ethnic preferences, where they select candidates based on ethnic profiles rather than their qualifications. There are several

E-mail address: akhemka@ub.edu.

¹ The data on the candidates' criminal records is collected from MyNeta, an open data platform run by Association for Democratic Reform (ADR). ADR has compiled the information available from the original candidate affidavits for all national and state elections. Retrieved from <https://myneta.info>.

² The classification of serious crimes is taken from ADR. An explanation of serious crimes along with the related IPCs is available on the ADR website: <https://adrindia.org/content/criteria-categorization-serious-criminal-cases>.

mechanisms that directly or indirectly provide support for why voters might consider ethnicity in their decision-making. First, in societies that have many social groups, individuals identify with only a few of them because they get a sense of belonging with other members of their group (Klor and Shayo, 2010). This shared identity can come in various forms such as race, language, or ethnicity, and voters observe some form of “in-group” positive externality from such candidates, perhaps explaining why they vote for them (Ansolabehere and Puy, 2016). Second, in weak institutional settings, citizens might be more prone to make their vote choices based on ethnic ties (Horowitz, 2001). Since these countries often tend to exhibit high ethnic cleavages and low information availability, citizens use ethnic cues rather than the qualification or electoral performance of the candidate in making their voting decisions. A third explanation argues that voters elect candidates from shared ethnic groups because they believe that such candidates would promote their interests or provide them with better access to resources (Posner, 2005; Chandra, 2007). This belief comes from the assumption that promises of patronage from coethnic candidates might be more credible since they have some form of moral obligation towards them. Regardless of the mechanism at play, the central theme in these theories is that voters might consider ethnicity in their decision-making. Hence, we can easily construe that if voters put a higher relative importance on the candidates’ ethnicity over their quality, this could perhaps explain why they fail to punish such politicians.

This leads to two main questions: Do voters make electoral decisions based on their ethnic preferences? If so, is this ethnic alliance that strong that voters are willing to ignore allegations of criminality? To address these questions, this article extends the standard deterministic voting model to include both ethnicity and criminal allegations in the voters’ decision-making (provided in Appendix A). The novelty of this approach is the inclusion of the interaction between these characteristics. Using this setup, I show that while voters display a general aversion towards criminality, this punishment is often mitigated if the voter considers ethnicity in their underlying preferences. To empirically test this prediction, I use primary survey data in which 2000 voters in the Indian state of Bihar were asked questions regarding their vote choice for the assembly elections held in 2020. Using a conditional logit model, I examine whether voters elect candidates based on their ethnic identities and how this alters their support towards criminally accused candidates.

To further explore whether the electoral support for coethnic candidates accused of criminality depends on certain conditions, I test for two underlying mechanisms: first, I examine whether voters with lower levels of political information are more likely to side with the ethnic party, even when they put forward candidates accused of criminality. It is frequently argued that citizens vote for tainted candidates because they lack sufficient evidence of their wrongdoings (Ferraz and Finan, 2008; Costas-Pérez et al., 2012). Despite these studies providing sufficient credence to the information hypothesis, the impact of information campaigns on altering voter behavior in the Indian context has been inconclusive.³ One explanation for why information campaigns have failed to have the desired effect could be that criminality is a highly salient trait in Indian politics. In the days leading up to the election, this issue is often extensively covered in the local and national news, providing voters with ample credible information on the candidates’ criminal backgrounds.⁴ However, one concern is that voters with limited knowledge about the candidates’ criminal activities might be more inclined to support their ethnic party. For example, Chandra (2007) argues that under low information settings, citizens forgive politicians for their misconduct because they rely on ethnic uses in making their voting decisions or they believe that the criminal allegations against their preferred candidates are more likely to be false. I test for this using two main instruments: the news consumption habits and the political knowledge of the voter.⁵

Another argument often made is that voters are willing to vote for criminal politicians because they might be more likely to provide them with patronage. Various scholars argue that voters view criminal politicians as “Robin Hoods”, who are capable of delivering public resources and that is why they elect them (Berenschot, 2011a; Vaishnav, 2017). This theory holds that in contexts with a lack of state capacity, criminal politicians can use their vast funds acquired through illegal activities and muscle power to either bribe or intimidate bureaucrats into delivering resources to their constituents, where others have failed. Another related explanation argues that criminal politicians might be more prone to engage in acts of providing side payments or vote buying (Wade, 1985; Bratton, 2008). Thus, if voters value tangible benefits over probity, we might expect them to have a stronger preference for their ethnic party. This is because they believe that once in power, their ethnic candidate would provide them with better access to public goods (Posner, 2005). To test for this, I examine whether poorer or less educated voters are more likely to forgive coethnic candidates accused of criminality. Allegedly, if voters were making decisions solely based on clientelistic reasons or the immediate patronage on offer, we should observe this effect to be the most prominent amongst the poorest and least educated segments of society (Banerjee et al., 2014; Kyriacou, 2023).

The findings of this paper challenge the common notion that Indian voters prefer criminal politicians not despite, but precisely because of this characteristic (Vaishnav, 2017). Voters show a stronger aversion towards candidates accused of criminality when they belong to their ethnically preferred party. This pattern holds regardless of the voters’ attributes such as their income, education, news consumption habits, and political knowledge. These findings suggest that even in the context of Bihar, a region well known for its caste politics and reputation for electing criminal legislators, the distaste for politicians of disrepute seems to be widespread.

³ Banerjee et al. (2011) using an experimental setting in the slums of Delhi find no conclusive evidence that voters alter their behavior when presented with report cards on the performance of incumbents. Vaishnav (2017) using qualitative data across India argues that voters were not told anything they did not already know. His work shows that politicians often boast about their criminal reputation, and voters are well aware of the criminal allegations against them.

⁴ Numerous articles in prominent local and national newspapers from *The Telegraph* to *The Times of India-Bihar* with headlines such as “Gangsters flex muscles in poll-bound Bihar” and “68% of elected MLAs in Bihar face criminal charges” are commonly published citing the number of candidates contesting with criminal allegations.

⁵ The use of these instruments is not novel to this paper. Various existing studies have exploited these characteristics to examine the information hypothesis (Aarts and Smetko, 2003; Klačnja, 2017).

This paper contributes to several strands of literature. Foremost, more narrowly, it contributes to the literature that evaluates the role of ethnic voting in the selection of bad-quality legislators in India. Although few existing studies have examined the ethnic voting hypothesis in the Indian context, the findings are mixed. At the aggregate level, several studies find that voters select parties solely based on their ethnic profile rather than policy platforms or candidate quality.⁶ In contrast, individual-level studies using experimental designs have found that voters respond negatively to coethnic politicians when criminal accusations are leveled against them.⁷

This paper provides the first empirical evidence using actual survey election data rather than an experimental approach. Using non-experimental data in estimating individual voting behavior might be of interest for several reasons. First, field experiments due to design limitations generally tend to present voters with a binary option, where they have a clean uncontested choice between a good and a bad-quality hypothetical candidate. However, multiple candidates with criminal allegations often run in Indian elections and these accusations of criminality are often met with denials and counter-accusations. Unlike in experimental settings, in real-world elections, this makes the voting decision much more complex than simply choosing between the good versus the bad quality candidate. Second, such studies rely on providing explicit ethnic cues to voters to reveal their preferences, either via using popular ethnic surnames (Banerjee et al., 2014) or imagery (Chauchard, 2016). Previous research has shown that although voters might be considering ethnicity in their decision-making at the polls, when made conscious about their choice they are unwilling to disclose their ethnic preferences (Carlson, 2016). A related concern is that ethnic voting can be highly sensitive, and any slight changes in ethnic cues or the voters' perception can alter the response significantly (Adida, 2015). Thus, using non-experimental data allows me to directly test the voters' response toward coethnic candidates and whether this can explain the electoral success of criminal politicians on the ballot.

Second, this paper contributes to the burgeoning body of literature examining the effects of information campaigns in altering voter response in the Indian context (see, Banerjee et al., 2011; George et al., 2018). This paper modestly adds to this literature by providing further evidence indicating that voters punish criminal politicians regardless of their news consumption habits or political knowledge.

Beyond India, these findings have wider applications. There has been a recent rise in the number of corrupt or criminal officials in various developing countries such as (not limited to) Brazil, Pakistan, the Philippines, and Nepal. Since these countries often have similar institutional settings, these results are particularly relevant to policymakers trying to come up with solutions to solve this perplexing problem.

Lastly, this paper adds to the comparative literature that has previously examined the role that ethnic voting plays in the selection of bad-quality officials in other developing and mature democracies across the world (Lindberg and Morrison, 2008; Carlson, 2015; Chang and Kerr, 2017; Adida et al., 2017). The findings in this paper highlight that although ethnicity plays an integral role in the voting decision, this is often mediated when the coethnic candidate is accused of committing serious acts of misconduct.

The rest of the article is structured as follows: Section 2 describes the electoral context. Sections 3 and 4 describe the data and the empirical strategy, respectively. Section 5 discusses the results. Section 6 provides some policy implications and concludes.

2. Electoral context

2.1. Electoral context in Bihar

Bihar is arguably one of the most important political states in India. With a population of over 100 million (8.6% of the nation's share), it is the third largest in the country and holds one of the largest state elections in the world.⁸ It is also one of the poorest states in India, with about 34% having income levels below the national average. About two-thirds of the population in Bihar are engaged in agricultural activities, and nearly 89% reside in rural areas. Literacy rates are close to 80% among men and 70% among women. Bihar's total GDP was approximately 80 billion dollars in 2019 contributing to 2.8% of the national GDP (Reserve Bank of India, 2022).

2.2. Criminality in Bihar politics

The "criminalization" of Indian politics is hardly a new phenomenon, where political parties are often guilty of fielding tainted candidates and voters of electing them. Although there has been a steady upward trend in the criminalization of politics in the country, the severity of the problem was unknown until recently. In 2003, the Indian Supreme Court in a landmark judgment made it mandatory for political candidates to submit an affidavit that included a comprehensive detail of their criminal records.

⁶ Chandra (2007) using data from 1984 to 1998 finds that Schedule Caste voters are more likely to vote for their ethnic party in Indian states with higher representation of coethnics in elite positions. She concludes India can be described as a "patronage democracy", where voters choose parties by counting the heads of coethnics in positions of power and prestige within the party organization. Banerjee and Pande (2007) find that in Uttar Pradesh, regions with increased voter ethnicization tend to elect lower-quality candidates. This is because the ethnically dominant party has a comparative advantage, irrespective of electoral performance or candidate attributes.

⁷ Both Banerjee et al. (2014) and Chauchard (2016) using an experimental setting in Uttar Pradesh show that although voters display a strong preference towards their ethnically preferred party or candidate, this ethnic voting advantage is often mitigated if the candidate is of lower quality.

⁸ The data for Bihar demographics is collected from the Census of India (2011). Retrieved from <http://www.censusindia.gov.in/pca/default.aspx>.

Remarkably, the release of these affidavits revealed that criminal politicians existed at all levels of government and the problem was even more acute in certain states such as Bihar and Uttar Pradesh.

While the law aimed at providing better information to voters, the recent uptake in criminal politicians in India suggests that the judgment has fallen short of having the desired effect of decriminalizing politics. For example, as Fig. C.1 shows the number of criminal MLAs rose from 49% in the 2005 elections to 58% in 2010 and 2015. Surprisingly, the number of MLAs with serious offenses was even higher and increased from 29% in 2005 to 35% in 2010 and 40% in 2015. Similar trends can be observed across the country at both the state and national levels. In 2020, 54% of the 4,676 sitting MLAs and MPs had some form of criminal allegations against them.

Although Section 8 of the Representation of People's Act bars individuals who have been convicted for more than two years from participating in elections for at least six years after their incarceration, there is no such bar forbidding candidates facing trial from contesting. Since these court cases can drag on for years, candidates can freely compete in elections, making Section 8 almost ineffective. This also incentivizes criminally accused candidates to run for political office, since once in power, they can manipulate the judiciary in throwing out the charges against them. Subsequently, the Supreme Court in 2020, citing the recent rise in criminal candidates, ordered parties to highlight candidate criminal records on their social media platforms in various vernacular languages. However, since all parties are equally complicit, this law has had little effect in curbing the criminalization of Indian politics. This was highlighted in the Bihar 2020 assembly elections, where the state recorded the second highest number of legislators facing criminal cases in India, with nearly 68% of MLAs being accused of criminal charges, including more than 50% of them being charged with committing serious crimes. The staggering number of politicians with criminal or corruption cases in Bihar politics is indeed disturbing. Although recent government measures are a step in the right direction, crime is intertwined in the fabric of Indian politics, and the steady uptake in criminal politicians suggests that there are other crucial factors at play in explaining why voters continue to support criminal candidates on the ballot.

2.3. Caste in Bihar politics

Politics in Bihar is closely intertwined with the caste system in India. Generally, caste refers to a hierarchical social group in which affiliation is determined by birth. Caste plays an integral role in an individual's identity and ability to acquire resources. It is well-known that voters identify with politicians based on their religion or caste. Parties are often complicit in engaging with a few ethnic groups to gain a comparative advantage and further secure their vote banks (Chandra, 2007; Banerjee and Pande, 2007). Elections in Bihar are dominated by the Hindu voter, which comprises more than 80% of the population. However, in certain regions, Muslim voters can influence election results. The Hindu society can be further divided into three broad caste groups: Upper Caste, Other Backward Caste (middle caste), and Schedule Caste/Tribe (lower caste). Amongst the Other Backward Caste (OBCs), arguably, the caste can be further segregated into two sub-castes: *Yadav* and *non-Yadav OBCs*.⁹

Political parties generally cater to one or two of these caste groups, and these relationships are traditionally well-established. From the five main parties contested in the Bihar 2020 elections: BJP, BSP, JD (U), RJD, and LJP, their ticket allocation and the electoral representation at the assembly legislature highlight the persistence of ethnic voting. BJP predominately catering to the upper caste, fielded about 50% upper caste candidates, even though the group only comprises 16% of the state population.¹⁰ In terms of representation, 30% of upper caste candidates that were elected more than half belonged to BJP.¹¹ Likewise, RJD, whose core voter base is the Muslim and *Yadav* community, fielded 12.5% and 33.3% of their candidates from these two groups respectively. Out of which, 45.4% were electorally successful. JD (U) gave tickets to two-thirds of candidates belonging to non-Yadav OBCs, with a maximum of their winning seats coming from these communities. Finally, BSP and LJP gave most of their tickets to candidates belonging to SC/ST caste groups. These trends seem to suggest that parties regularly field specific ethnic candidates and citizens seem to routinely vote for them, highlighting the extent of polarization in Bihar politics.

3. Data

3.1. Voter data

The Bihar state assembly elections were held during the months of October–November 2020. I collected primary voter data during the months of January–April 2022 in two districts of Bihar: Muzaffarpur and Samastipur.¹² The sampling strategy was as follows: in total, there are 21 constituencies in both these districts, out of which about one-fifth of the constituencies are reserved for Schedule Caste/Tribe (SC/ST) candidates. Since there is no choice for voters but to select a candidate from a reserved ethnicity in these constituencies, they were dropped from the sampling procedure. This provided a sample of 17 constituencies to choose from. From the list of general constituencies, 5 constituencies were chosen at random from each district. Next, using polling station coverage maps provided by the Bihar Election Commission, 2 polling booths were randomly chosen within each of the sampled

⁹ *Yadav* is a prominent *jati* or sub-caste accounting for about 14% of Bihar's total population. While the *Yadav* community traditionally belonged to the lower caste mainly involved in agricultural activities, they often tend to wield significant political power (Gupta, 1992).

¹⁰ The data on candidate caste background has been retrieved from the SPINPER Project and TCPD-ILD Dataset.

¹¹ Similarly, in the assembly polls of 2015, BJP fielded 65 upper caste candidates with 53 of them securing electoral victory.

¹² See Fig. D.1(a)–(b) for the location of Bihar in India and the constituency-level map of Bihar.

constituencies.¹³ Finally, within each polling booth, 100 respondents were randomly chosen for the survey using the voter list generated by the Bihar Election Commission for the 2020 assembly elections.¹⁴ This procedure provided a complete size of 2000 voters across the 10 assembly constituencies. A brief description of the sample size is provided in Table D.1.

In addition, Bihar being predominantly an agricultural state faces a high level of seasonal labor migration and some voters in the sample size were unavailable to be surveyed. For this reason, the survey team was provided with 50 additional substitutes in each polling booth. To avoid the risk of pre-selecting the easiest respondents, the substitutes were randomly selected beforehand. Surveyors were instructed to first attempt to locate the original respondent and failing to do so to note the reason and then move on to the substitute list. A month before the survey, I recruited and trained the field team comprising of eight surveyors and two supervisors. A few days leading up to the survey, a pilot study was conducted in a non-sampled polling area to test the responses to the questionnaire. After making the relevant changes, the survey was translated and coded using a standard Open Data Kit (ODK) platform. In the final sample, 76% of the respondents were from the original list and 24% were substitutes. About 90% of the substitutions were made because the respondent had permanently moved or was temporarily unavailable. Once the respondent was located, the response rate was extremely high, with less than 1% refusing to participate in the survey. Amongst the respondents who participated, 27 of them did not vote in the 2020 elections mainly due to personal reasons or having voter ID issues and 4% did not recall who they voted for.¹⁵ Therefore, these observations were dropped, providing a final sample size of 1892 voters.

To further ensure responses were not impacted by any observer effects, enumerators were instructed to conduct the survey only when the respondent was alone. If this was not possible, respondents were asked for an alternative time for when the survey could be conducted. Additionally, to make female respondents comfortable in giving the survey alone, they were only surveyed by female enumerators. Lastly, to check the accuracy of the data, 20% of the respondents who participated in the main survey were randomly selected to be surveyed again. For this purpose, two independent back-checkers were hired to contact them a few days later remotely. The back-checkers asked a few pre-determined questions from the main survey and reported to the supervisor if any errors were detected. The supervisor then personally visited the respondent to make further inquiries and if required the survey was conducted again or the respondent was replaced using the substitute list.

A few sections included in the survey were on the demographics and the socio-economic characteristics of the respondent, their news consumption habits, general political knowledge, voting preferences, and how political parties engaged with them during the election campaign. The relevant version of the survey questionnaire is provided in Appendix D. In terms of demographics, as presented in Table 1, on average more than half of the respondents were female, a bit higher than the state average.¹⁶ 48% of the respondents were literate and only 6% possessed a college degree. The average monthly household income was just above 10,000 Indian Rupees (approximately 123 US\$). Most of the respondents were Hindu, with 13% upper caste, 14% *Yadav*, 48% *non-Yadav* OBCs, and 18% belonging to the SC/ST category. In terms of the caste break, the sample was mostly similar to the state average.

To reveal the voting preferences of the respondents in the 2020 assembly elections, they were provided with a full list of candidates that contested in their respective constituency, along with the names and symbols of the parties they represented. The respondents were asked to mark their choice and hand over the list to the surveyor. Each of the candidates was given a unique code, which the surveyor then entered into the survey. For respondents who did vote in the 2020 state (Fig. D.2), 33% of them claimed to have made their choice based on various development issues such as government performance, unemployment, and economic growth, 14% of voters claimed to have voted for having some form of loyalty to the candidate or party, while the largest fraction of voters chose candidates based on some clientelistic issues such as getting better access to resources or government schemes.

The survey further included questions on news consumption habits and political knowledge to gauge how informed the voter was. First, to measure news consumption, respondents were asked questions on how much political news they saw, listened to, and read using various platforms such as newspapers, television, and the internet. Using this information, I constructed a binary variable news frequency, which takes a value of 1 if the respondent consumed political news on a daily or weekly basis and 0 otherwise. The data revealed that only 39% of the respondents in the sample size consumed news frequently. Second, to assess the political knowledge of the respondents, they were asked nine simple questions such as the MLA's name and the party they were affiliated to.¹⁷ I defined a dummy variable political knowledge which equals 1 if the respondent at least answered more than half of the questions correctly and 0 otherwise. Remarkably, given the simplicity of the questions, only 43% of the respondents in the survey were found to have a high level of political knowledge.

Lastly, since the main interest of this study is to check whether ethnicity played a role in the selection of criminal candidates, a respondent was defined as a coethnic if they voted for their ethnically preferred party. Specifically, the variable scores 1 if an upper caste respondent selected a BJP candidate, a *Yadav* or Muslim respondent selected a RJD candidate, a *non-Yadav* OBC respondent selected a JD(U) candidate, and a SC/ST respondent selected a candidate from BSP or LJP, otherwise the variable scores a 0. While some of these parties contested under an alliance, this was not taken into consideration to avoid any over-estimation issues.

¹³ The polling and voter list is provided in the Bihar Chief Electorate Office. Retrieved from <http://ele.bihar.gov.in/pdfsearch>.

¹⁴ The sampled polling booth comprised of about 1000 registered voters, out of which 10%–12% were randomly picked for the survey.

¹⁵ The high voter turnout in the survey could be attributed to several factors. First, the survey was conducted only in rural areas where voter turnout generally tends to be high. Second, the elections took place right after the COVID lockdown allowing migrant workers to return home and participate in the electoral process. Third, it could be that a fraction of the respondents who did not recall who they voted for or were substituted included voters who did not turn out. Fourth, a post-poll survey conducted by Lokniti CSDS reported a similar voter turnout of 97% with a sample size of 3612 respondents. However, one concern with the Lokniti CSDS survey was that the reported turnout might suffer from overestimation due to sampling errors.

¹⁶ The state average data is taken from the Lokniti CSDS Bihar Post-poll 2020-Survey Findings. Lokniti is a department at the Center for the Study of Development Societies specializing in conducting election-related surveys. The full report is available at https://www.lokniti.org/media/PDF-upload/1606577835_22658400_download_report.pdf.

¹⁷ A full breakdown of the voter response to all the questions is provided in Table D.2.

Table 1
Voter profile.

Variable	Mean	S.D.	Min	Max	State Avg.
Age	46.48	15.82	18	100	43.22
Male	0.48	0.50	0	1	0.52
Literate	0.48	0.50	0	1	0.48
College degree	0.06	0.24	0	1	0.09
Household income (in thousands)	10.22	10.20	0.03	150	7.40
Hindu	0.93	0.25	0	1	0.84
Upper caste	0.13	0.34	0	1	0.24
Yadav	0.14	0.34	0	1	0.15
Non-Yadav OBC	0.48	0.50	0	1	0.43
SC/ST	0.18	0.47	0	1	0.18
NREGA worker	0.17	0.38	0	1	0.08
Attended political rally	0.17	0.38	0	1	0.18
Joined party media platform	0.07	0.26	0	1	0.07
Read political manifesto	0.19	0.39	0	1	NA
High news frequency	0.39	0.49	0	1	NA
High political knowledge	0.43	0.50	0	1	NA

Notes: The voter profile is based on the final sample size of 1892 respondents. The state average data is taken from the Lokniti CSDS Bihar Post-poll 2020 survey that included a sample size of 3612 respondents from 37 assembly constituencies. NA indicates data for which the state averages are not available.

3.2. Election outcomes and candidate data

Data on election outcomes for the Bihar state assembly elections held in 2020 was collected from the Trivedi Centre for Political Data (TCPD).¹⁸ The data includes various election-related information such as the constituency names, their reservation status, candidate names, their affiliated party, electoral size and turnout, registered number of voters, and the vote share received by each candidate.

The data for criminal accusations and background of the candidates was collected from MyNeta, an open data depository run by ADR. This information is available on the ECI website in the form of affidavits in PDF form (Fig. E.1). ADR has re-entered and compiled this data, making it freely available to the public to provide better access and transparency.

Although all the information about the candidates was collected, the primary interest was the information about their criminal background. Using this data, criminal charges were separated into two types of crimes: violent and non-violent. In this paper, violent crimes refer to those that are related to bodily harm.¹⁹ Using this definition, I construct a dummy variable that equals 1 if the candidate is accused of violent allegations and 0 otherwise. Likewise, non-violent crimes take a value of 1 if the candidate faced any form of charges other than violent ones and 0 otherwise.

In the ten assembly constituencies sampled, the average electoral turnout was similar and the average margin of victory between the top two candidates was below 10% regardless of the number of criminal candidates contesting. Of the 185 competing candidates, as presented in Table C.1, on average, most of them were men and belonged to the Hindu community, and less than half of them possessed a college degree. In terms of criminal accusations, a simple look at the distribution of candidates with criminal charges and the different types of charges highlights the severity of the problem (Tables C.2–C.3). Of the 185 candidates, more than 31% of the candidates faced at least one criminal charge, of which more than 50% faced violent allegations. The problem is even more acute for the top-finishing candidates. Amongst the top 3 candidates, more than 75% had criminal charges and 50% were facing violent charges. Among the winners, 9 out of 10 sitting MLAs had a criminal record, with 2 of them facing allegations of violent offenses. Looking at the criminal candidates by political representation, a high number of them belonged to the most popular parties. From the 30 candidates selected by the popular parties, 20 of them had some form of criminal record, out of which 5 were accused of committing violent offenses (Table C.4). Lastly, a comparison between non-criminal versus criminal candidates revealed that criminal candidates had a higher likelihood of belonging to a national party or being an incumbent (Table C.5). However, when restricting the sample to only the top three finishing candidates, there was no evidence any statistical differences in their characteristics. Since the analysis in this paper is limited to the top three finishing candidates, this provided some assurance that the findings capture the effect of criminal allegations rather than that of any other correlated candidate characteristics.

4. Empirical strategy

The main aim of this paper is to investigate if coethnic voting can explain why voters support criminal politicians at the polls. Do voters elect politicians along ethnic lines? Is this ethnic preference so strong that voters are willing to overlook allegations of wrongdoing?

¹⁸ TCPD has compiled the data for all the elections held both at the national and state level from the original reports available from Election Commission of India (Agarwal et al., 2021). The data is available at <https://lokdhaba.ashoka.edu.in/>.

¹⁹ Violent crimes include murder, attempt to murder, rape, kidnapping, extortion, and armed robbery.

There are several challenges involved in estimating the effect of candidate traits on voting behavior in multi-candidate elections, which make common econometric choice models such as multinomial logit inadvisable (see, [Alvarez and Nagler, 1998](#)). One solution is the conditional logit model, which allows for an estimation of choices when multiple parties or candidates are available. There are several advantages of using this methodology: first, as the theoretical model predicts, the vote choice for one candidate is dependent on the characteristics of the other. The conditional logit precisely accounts for this by estimating the likelihood of the selection of one candidate conditional on the attributes of the alternatives. A second advantage of the conditional logit is that it allows grouping the data for each decision-maker, meaning it controls for any individual-level voter fixed effects. To estimate the conditional logit model, the data is stacked, such that, for each voter, there are multiple rows, where each row indicates the alternatives available. In formal terms, the model can be specified as:

$$\Pr(y_i = p/x_i, z_i) = \frac{\exp(z_{ip}\psi + x_i\beta_p)}{\sum_{j=1}^J \exp(z_{ij}\psi + x_i\beta_j)}$$

where y_i is the outcome measuring the probability of an individual i choosing an alternative party p from a set of choices. In this case, it is a binary variable that takes a value of 1 for the party the voter chose in the Bihar 2020 assembly elections and 0 otherwise. z_{ip} indicates a variable measuring the characteristics of alternative p relative to case i . In this model, the alternatives p are represented by the parties that finished in the top three positions in their respective constituency. In general, for each variable z_k , there are J values of the variable for each case, but only the single parameter ψ . x_i contains case-specific independent variables for case i and β_p represents coefficients for the effects of alternative p relative to the base alternative (in this case the clean non-ethnic).

To measure the ethnic preferences of the voter, the model includes a dummy variable *coethnic* which takes the value of 1 if the voter chose a party belonging to their ethnic group and 0 otherwise. To account for criminal accusations, I include two types of charges: *violent* and *non-violent*. To capture whether coethnic voting strengthens or weakens the electoral support towards criminality, interaction terms between each type of allegation and coethnicity are included in the model. The model further controls for various observable candidate characteristics such as their age, income in logs, education status, and incumbency. Lastly, since Indian elections are mostly party-centric, I account for this by including party fixed effects, defined as a dummy variable indicating the party the candidate belongs to.

To further explore whether the voters' response towards criminal candidates belonging to their ethnically preferred party alters under certain conditions, I test for two main alternative explanations. First, I test whether the strength of coethnic voting towards criminal candidates weakens depending on the voters' news consumption habits and level of political knowledge. Second, I examine whether poor and illiterate voters show a higher tendency to forgive coethnic politicians accused of criminality.

There are two potential concerns to consider when using the above empirical strategy. First, it could be that the parties' decision to field criminal candidates in certain constituencies over others depends on their winning probability. However, in the sampled constituencies, fewer than half of the parties contested in the previous elections and largely ran under a different alliance, making it less likely that they could rely on their past voter support. This provided some assurance that the results are not endogenous to the party's expectations. Another concern is that the conditional logit model imposes the IIA property on the individual voter. For the IIA property to hold requires an assumption that the voter has clear distinct choices. One solution is the use of probit models that relax the assumption of IIA, allowing an examination of the full choice set available to voters, while explicitly allowing voters to observe some parties as close alternatives ([Alvarez and Nagler, 1998](#)). Hence, to check the robustness of the results, I re-examine the model using probit regressions with individual voter fixed effects.

5. Results

5.1. Main results

Table 2 provides the estimates for the conditional logit, where the main outcome measures the vote choice of the respondent conditional on the alternative available. Columns (1)–(2) provide estimates without the inclusion of the interaction term between coethnicity and criminality. Columns (3)–(5) present the results with the inclusion of the interaction term. Columns (2), (4), and (5) include candidate controls for their age, income, education, and incumbency. All columns include controls for party-fixed effects.

The initial findings suggest that voters display a strong tendency to vote for their ethnically preferred party. When candidates are accused of criminality, the coefficient for both types of criminal charges is statistically significant and negative, implying a fall in voter support. Note that this negative voter response is substantially higher for violent charges than for non-violent ones. This suggests that respondents in the survey were not only able to evaluate candidates based on their criminal records but were also cognizant of the different types of criminal charges. In column (2), the results remain consistent with the inclusion of candidate controls.

Looking at the main specification of interest in columns (3)–(4), we see that the voters' response is stronger for criminal candidates belonging to the ethnically preferred party. In the absence of criminal allegations, voters are 2.58 times more likely to side with their ethnic party.²⁰ However, coethnicity reduces electoral support for candidates accused of violent charges. The

²⁰ The magnitude of the results can be interpreted using the odds ratio, which is the natural metric for conditional logit models. To extract the odds ratio, we can simply take the exponential of the coefficient term, where an odds ratio of value greater than one represents a positive effect, while an odds ratio of value less than one represents a negative effect. Column 5 reflects the odds ratio for the coefficients presented in column 4.

Table 2
Conditional logit estimates on vote choice in Bihar 2020 state assembly elections.

	(1)	(2)	(3)	(4)	(5) <i>Odds ratio</i>
Coethnic	1.003*** (0.081)	1.083*** (0.090)	0.958** (0.173)	0.948*** (0.203)	2.581*** (0.524)
Violent criminal	−1.268*** (0.224)	−2.116*** (0.275)	−1.298*** (0.236)	−2.228*** (0.301)	0.108*** (0.0324)
Coethnic × Violent criminal			−0.875*** (0.307)	−0.844** (0.331)	0.430** (0.142)
Non-violent criminal	−0.362*** (0.137)	−0.425** (0.212)	−0.436*** (0.162)	−0.511** (0.231)	0.600** (0.139)
Coethnic × Non-violent criminal			0.225 (0.218)	0.394 (0.252)	1.483 (0.374)
Party fixed effects	Yes	Yes	Yes	Yes	Yes
Candidate controls	No	Yes	No	Yes	Yes

Notes: The table reports conditional logit estimates for 1892 respondents. The dependent variable takes the value of 1 if the respondent chose the candidate and 0 otherwise. The data is grouped at the voter level accounting for individual voter-fixed effects. In columns (1) and (3) the estimates do not include any candidate controls. In columns (2) and (4) the estimates include candidate controls for their age, income in logs, education, and incumbency. Column (5) reflects the odds ratio for the coefficients estimated in column (4). All models include party-fixed effects. Standard errors are given in parentheses. Asterisks denotes the significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

probability of voting for the non-ethnic falls by 89.2% for violent crimes. This electoral punishment is stronger for the coethnic where voter support plummets further by 67%. This finding is contrary to the existing literature that tends to find that the voters' response towards corruption or criminality is often mitigated by the politicians' ethnicity (Banerjee and Pande, 2007; Chauchard, 2016). In this respect, these results are in line with recent literature that shows that local politicians perceive that a scandal in their national party has a larger impact on the party image relative to scandal in other parties (Schönhage and Geys, 2022). Likewise, from the voter perspective, Reuter and Szakonyi (2021) find that voters withdraw their support from ruling parties when they commit election fraud, and this effect is larger amongst core voters. Frederiksen (2023) show that voters punish partisan politicians equally for undemocratic behavior, whereas the punishment is even larger when the parties' policies are aligned with that of the voter. These studies theorize that voters have an *ex-ante* belief that either their aligned party is more likely to run fairer elections or have a higher expectation from them to be democratically compliant.

In contrast, voters do not seem to show the same willingness to punish coethnics for non-violent accusations. This could be attributed to the fact that criminality is a highly visible trait in Indian politics, and voters might be willing to trade probity for their ethnically preferred party for so-called “weaker” allegations. For example, in the sample, 60% of the candidates faced some form of criminal charges. Thus, if voters expect all politicians to be equally complicit, it makes sense for them to side with the coethnic even if they have a general distaste towards criminality. However, when these same ethnic parties field violent criminals, the electoral punishment might be larger (relative to the non-ethnic violent criminals) because voters have a higher expectation from their aligned party to field relatively cleaner candidates. While this might prove some intuition behind the results, the data does not allow us to further disentangle if this effect is a result of the voters' having some prior expectations from their ethnic party or if any alternative factors are influencing their decision-making. Lastly, when looking at the estimates for the non-ethnic parties, voter support falls by 40% for non-violent charges. These results are consistent with the estimates of a probit regression (reported in Table B.1 of the Appendix). Coethnicity reduces electoral support for candidates with violent allegations by 48.47% ($p < .001$).

To summarize, the baseline findings suggest that voters show a strong distaste for candidates accused of criminality regardless of their ethnic affiliation. Voters withdraw their support for the non-ethnic candidate when they face criminal allegations. However, this electoral punishment is much larger for the coethnic violent criminal candidates relative to that of the violent non-ethnic criminal candidates. Lastly, the respondents in the survey not only display a strong contempt for criminal politicians but seem to be highly aware of the severity of different charges. The voters' negative response towards violent charges is always more pronounced relative to non-violent charges.

5.2. Alternative explanations

In this subsection, I examine whether certain conditions mitigate the voters' negative response toward the coethnic candidate accused of criminality. First, I test whether the voter is more likely to side with their ethnically preferred party if they have limited knowledge of the candidates' criminal activities. Under low information environments, voters might place a higher weight on the candidates' ethnicity. For example, several studies have found that in contexts with limited access to information, voters use ethnic cues as a shortcut in making their voting decisions (Chandra, 2007). Second, in contexts where candidates face multiple accusations that are often met with counter-accusations and denials, voters might be more likely to believe that their ethnically preferred candidate is falsely accused. To test this theory, I examine whether voters with lower levels of news consumption or political knowledge are more likely to support their ethnic party, even when they present candidates accused of criminality on the

Table 3
Impact of voter characteristics on vote choice in Bihar 2020 state assembly elections.

	News frequency		Political knowledge		Literacy		Income	
	Low	High	Low	High	Illiterate	Literate	Poor	Not poor
Coethnic	1.220*** (0.227)	-0.592* (0.330)	1.253*** (0.232)	-0.639* (0.331)	1.242*** (0.244)	-0.553* (0.328)	0.971*** (0.181)	-0.0370 (0.578)
Violent criminal	-1.010*** (0.274)	-0.652** (0.308)	-1.055*** (0.271)	-0.496 (0.314)	-1.191*** (0.290)	-0.218 (0.307)	-1.308*** (0.241)	0.187 (0.545)
Coethnic × Violent criminal	-1.125*** (0.360)	0.537 (0.546)	-0.937** (0.375)	0.0586 (0.511)	-1.046*** (0.397)	0.323 (0.499)	-0.891*** (0.316)	0.0939 (1.042)
Non-violent criminal	-0.279 (0.189)	-0.335* (0.203)	-0.227 (0.193)	-0.416** (0.205)	-0.420** (0.195)	-0.227 (0.204)	-0.393** (0.166)	-0.441 (0.325)
Coethnic × Non-violent criminal	-0.0370 (0.285)	0.576 (0.429)	-0.272 (0.285)	-0.0289 (0.301)	1.164*** (0.432)	0.489 (0.428)	0.235 (0.230)	-0.0796 (0.714)
Candidate controls	Yes		Yes		Yes		Yes	
Party FE	Yes		Yes		Yes		Yes	
N	5499		5499		5499		5499	

Notes: The table reports conditional logit estimates with voter fixed effects for 1892 respondents. The dependent variable takes the value of 1 if the respondent chose the candidate and 0 otherwise. In columns (1)–(2) a voter is defined to have low news frequency if they do not consume news on a daily or weekly basis and high news frequency otherwise. In columns (3)–(4) a voter is defined as having low political knowledge if they responded to less than half of the political questions incorrectly and high political knowledge otherwise. In columns (5)–(6) a voter is defined as illiterate if they cannot read and write in the vernacular language and literate otherwise. In columns (8)–(9) a voter is defined as being poor if they own a below poverty line card or have a household income below 10,000 Rupees and not poor otherwise. All models include party fixed effects and candidate controls for their age, income in logs, education, and incumbency. Standard errors are given in parentheses. Asterisks denotes the significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

ballot. Table 3 provides no support for this argument. Just as before, the voters' negative response is stronger for violent criminal candidates belonging to their ethnic party. Voter support for the non-ethnic falls by 63.58% for violent accusations. Coethnicity further reduces support by 67.54%. In the same vein, the overall pattern holds when looking at the estimates for political knowledge. Coethnicity further plummets support by 60.82% for candidates accused of violent crimes. The probit regressions confirm these results, where coethnicity reduces the probability of voting for criminal candidates, even when the respondents exhibit low levels of news consumption or political knowledge. Overall, these results do not suggest that the effect of ethnic voting for criminal candidates varies by political information.²¹ If this were the case, we would observe a very different pattern where less informed voters would be more likely to support their ethnically aligned candidate, despite the criminal allegations against them.

Next, I examine whether patronage voting can mitigate the voters' electoral punishment towards their ethnically preferred party for acts of criminality. To test for this, I examine whether less educated or poorer voters are more likely to vote for the coethnic candidate. Several scholars argue that patronage voting can be strongest amongst these segments of society, as they have more immediate needs, making them more likely to exchange votes for any resources that might be on offer (Stokes et al., 2013; Banerjee et al., 2014). Since these voters also tend to exhibit stronger ethnic preferences, they might willing to excuse their ethnically preferred candidate for their criminal conduct, because they believe that they would provide them with some form of patronage (Posner, 2005). Columns (5)–(8) do not seem to suggest that the voters' education or income reduces the electoral punishment towards the coethnic accused of criminality. Coethnicity further reduces electoral support by 64.88% for candidates accused of violent charges regardless of their education level. Similar trends can be observed when looking at the level of income: coethnicity reduced the electoral support for criminal conduct. Again, these results are confirmed by the probit estimates. Violent allegations against both the coethnics and non-ethnics are negative and significantly distinguishable from zero ($p < .001$). On average, these results do not provide any evidence in support of the argument that voters are willing to excuse their ethnic party for giving tickets to criminally accused candidates because they believe that they would provide them with some form of patronage. Instead, the distaste for politicians accused of hard-hitting crimes seems to be widespread amongst a diverse set of voters.

Lastly, as a robustness check, I examine whether any other voter characteristics such as gender, age, and the reason for voting can alter their response toward criminal candidates (reported in Table B.4 of the Appendix). All things equal, the results do not suggest that the coethnic voting effect is dependent on these factors. The only interesting aspect of these results is that respondents who make their vote choice based on candidate or party loyalty have a higher likelihood to forgive coethnics alleged of violent crimes. This result supports the predictions of the theoretical model that voters with sufficiently strong ethnic preferences might be more likely to excuse their coethnic party, even when they put forward criminally accused candidates. All things equal, these findings put together with the baseline estimates suggest that voters in general show a strong contempt for criminal candidates regardless of ethnicity. However, sometimes this punishment effect can be mitigated if the voter is so deeply committed to their caste-preferred party, that they seem to value ethnic alignment over the probity of the candidate.

²¹ As a robustness check, I examine the effect of ethnic voting on electoral support for criminal politicians using the full range of both the information variables. The results of this exercise are provided in Tables B.2–B.3. In both cases, the estimates are qualitatively and quantitatively similar to those of the main findings.

6. Conclusion

In this paper, I estimate the effects of ethnic voting on the selection of criminally accused political candidates in the Indian state of Bihar during the assembly elections of 2020. The main findings suggest that voters show a strong distaste for their ethnically preferred party when they give tickets to candidates accused of criminality. This negative response is larger when the coethnic candidate faces violent charges relative to the non-ethnic violent candidate. This pattern holds even when voters are separated by education, income, news consumption habits, and political knowledge. Although Indian voters are often criticized for evaluating candidates solely on parochial issues, be it caste, patronage, or clientelism, the findings in this paper show a very different pattern. Instead, a distaste for candidates of disrepute seems to be the norm amongst a diverse type of voters, even in a region well-known for electing candidates of disrepute to political office and its communal politics.

These results are contrary to the vast body of literature that finds ethnic voting to promote clientelism (Chandra, 2007), increase racial division (Ferree, 2010), and harm democracy (Horowitz, 2001). Similarly, research indicates that the public's reaction to corruption, whether from party members or voters, is frequently influenced by the ethnic affiliation of politicians (Banerjee and Pande, 2007; Chang and Kerr, 2017) or their partisan alignment (Anduiza et al., 2013; Schönhage and Geys, 2023). In this respect, the findings of this paper add to the small but growing experimental comparative politics literature, which shows that the effect of ethnic voting can be mediated by candidates' quality or other non-ethnic factors (Banerjee et al., 2014; Carlson, 2015; Chauchard, 2016; Chauchard et al., 2019).

How do we then explain the steady rise of criminal politicians in the Indian legislature? A supply-side issue, where there is a lack of availability of viable clean alternatives, could contribute to why voters are willing to look the other way when it comes to so-called "weaker" charges. However, it does not account for the steady rise of candidates accused of heinous crimes being elected to political office. The results in this paper show that voters display a strong negative response to candidates accused of committing violent offenses, regardless of their ethnic background. However, in practice, over the past two Bihar state assembly elections, 45% of the winning candidates have had serious allegations against them. Then why is this willingness to punish criminal candidates not translating into actual electoral behavior?

It is often claimed that since the Indian voter is less educated or poorer, this is purely an information constraint problem, or the voter is simply making their decision based on clientelistic issues. However, the results provide no support for this claim. This is consistent with the ethnographic literature on India, where several scholars have shown that voters are quite aware of the criminal records and candidates do not attempt to mask their reputation (Berenschot, 2011a; Vaishnav, 2017). Thus, information campaigns are not likely to inform voters about anything that they do not already know. Likewise, other studies have argued that voters elect corrupt or criminal politicians regardless of the expectation of *quid pro quo* exchange (Vaishnav, 2017; Auerbach et al., 2022). These scholars theorize that when elections are conducted using secret ballots and there are a diverse set of voters, the ability for politicians to monitor voters is often challenging and probabilistic at best.

Two plausible alternative mechanisms could be contributing to the rise of criminal legislators in India: first, partly it might simply come down to hard cold cash. Criminal politicians tend to be substantially wealthier than clean candidates. Indian elections are fiercely competitive, with the average margin of victory often being less than 10%. Thus, any comparative advantage can be telling in terms of election outcomes. Criminal politicians can not only use their excess wealth to pay political rent but also run expensive electoral campaigns, which can sometimes translate into votes. This comparative advantage is further exacerbated by the lack of oversight in election financing in India. For example, Kapur and Vaishnav (2013) find that the key difference between developed and developing countries is the lack of accountability in electoral finance, making the use of illicit funds a crucial factor in determining election outcomes.

Another contributing factor could be that voters truly believe that criminality serves as a signal of competency. This is particularly relevant in the Indian context. With a lack of state capacity and weak rule of law, criminal politicians can be often viewed as having the ability to "get things done" (Vaishnav, 2017). There is a large body of qualitative works across India that show criminal politicians are viewed as effective strongmen who are willing to go above the law to protect individual rights and deliver resources to their constituents (see, Berenschot, 2011a,b; Martin and Michelutti, 2017). On the other hand, the recent empirical studies do not support this claim and find that criminal politicians worsen overall economic activity (for e.g., Chemin, 2012; Prakash et al., 2019; Gehring et al., 2019; Zakharov, 2019). However, these studies are limited to looking only at broad constituency outcomes that the local politician might not have much control over or voters might care less about.

The empirical question is then, under settings where the formal state has failed, whether criminal politicians are truly able to provide certain targeted resources to their constituents. If so, the challenge remains for reformers to think about ways to further strengthen local state capacity and improve governance to change the voters' perception and translate their willingness to punish criminal candidates into actual electoral outcomes.

CRedit authorship contribution statement

Abhinav Khemka: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Supervision, Validation, Visualization, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Abhinav Khemka reports financial support was provided by Spanish Ministry of Science, Innovation and Education.

Data availability

Data will be made available on request.

Acknowledgments

I am grateful for the helpful comments from Jordi Munõz, Pilar Sorribas-Navarro, and two anonymous reviewers. I would also like to thank participants at the European Public Choice Society Annual Meeting 2023, International Corruption Research Forum 2023, IV UB PhD Workshop on Empirical Political Science, and the UB PhD in Economics Workshop for their valuable comments and feedback. This work received financial support from the Spanish Ministry of Science, Innovation and Education [PID2021-126517NB-I00].

Appendix A. Theoretical model

This section develops a theoretical model using a two-candidate framework to illustrate the effect of the voters' ethnic preferences on the selection of criminal candidates.²² The model approach is broadly based on two strands of literature: first, it draws on the literature that extends the conventional voting models to include identity issues such as race, language, or religion (see, [Akerlof and Kranton, 2000](#); [Ansolabehere and Puy, 2016](#); [Karakas and Mitra, 2021](#)). Second, it draws on the literature that evaluates voting and election outcomes with the inclusion of valence components, where all voters have a similar position on certain issues such as higher candidate quality, crime rates, or economic growth (for e.g., [Schofield, 2004](#); [Ansolabehere and Snyder, 2000](#); [Grosceclose, 2001](#)). I add to this literature by analyzing the interaction between both identity and valence issues in the voters' decision-making. Identity issues in the model refer to the ethnic preferences of the voter, such that they evaluate ethnically aligned candidates or parties in one way, but other candidates or parties differently. Valence issues in the model indicate candidate quality, such that voters hold a homogeneous distaste for candidates accused of criminality. The model aims to examine whether voters consider both these dimensions in their decision-making and how this consideration affects voting and electoral outcomes.

A.1. Setup

Assume that the political environment has two candidates j and k belonging to a vector $C = \{j, k\}$. The total population of voters denoted as n is divided into two groups: μ representing the share of ethnic voters and $1 - \mu$ denoting the share of neutral voters. Ethnic voters are assumed to value the ethnic background of the candidate, such that they derive a positive utility from a candidate belonging to their ethnicity denoted as $e > 0$, and a negative utility from a non-ethnic denoted as $e < 0$. A neutral voter does not consider ethnicity when making their choice such that $e = 0$. Additionally, following [Banerjee et al. \(2014\)](#), I assume that the voter is sincere, meaning they prefer good quality candidates. The voters derive a positive utility of 1 from a non-criminal candidate and a negative utility of -1 from a criminal candidate. In particular, the voter observes a vector denoted as (c_j, c_k) that can either equal to $(1, 1)$, $(-1, 1)$, $(1, -1)$, or $(-1, -1)$. Voters further value several candidate characteristics such as incumbency, party affiliation, education, and so on denoted as x . Lastly, the ethnic profile of the candidates is assumed to be pre-determined, where j is always the coethnic, meaning that they belong to the same ethnic background as the voter, and k is the non-ethnic. Thus, for an individual voter i from a set of voters $V = \{1, 2, \dots, n\}$, the utility derived from voting for candidate j can be expressed as:

$$u_{ij} = \alpha_i e_{ij} + \beta_j c_j + \gamma_i x_j \quad (\text{A.1.1})$$

where e_{ij} denotes the ethnic preference the voter i has for candidate j and this is assumed to be known to the voter. This seems plausible since ethnic alliances are historically established, making it difficult for parties to switch. c_j indicates if candidate j has a criminal record or not. x_j denotes the utility the voter gets from the inherent characteristics of candidate j . $\alpha, \beta, \gamma > 0$ are the weights the voter attaches to each issue and are known to the voter as well. The voter knows what they like.

The voting game occurs as follows: Each party chooses a single candidate to run from a constituency, and then voting occurs. The best response is to choose the candidate that maximizes their observed utility, such that a voter i votes for candidate j if $\Delta u_i > 0$ and k if $\Delta u_i < 0$, and is indifferent when $\Delta u_i = 0$, where $\Delta u_i = u_{ij} - u_{ik}$. Thus, a voter i chooses candidate j if

$$\begin{aligned} \Delta u_i &= \alpha_i (e_{ij} - e_{ik}) + \beta_j (c_j - c_k) + \gamma_i (x_j - x_k) > 0 \\ &= \alpha_i \Delta e + \beta_j \Delta c + \gamma_i \Delta x > 0 \end{aligned} \quad (\text{A.1.2})$$

where, $\Delta e = e_{ij} - e_{ik}$, $\Delta c = c_j - c_k$, and $\Delta x = x_j - x_k$. To examine the effect of ethnic voting on criminality, for simplicity' sake Δx is normalized to zero, meaning that candidates only differ on two dimensions: ethnicity and criminality. This assumption allows us to directly evaluate the effects of ethnic voting on the voter response by considering different cases, where one or both candidates have criminal charges against them.

²² The model is simplified by using only two candidates for illustrative purposes. Extending the model with more candidates would not qualitatively affect the main conclusions of the model and introduce unnecessary complexity.

A.2. Benchmark model

In the first case, assume that both candidates j and k have the same criminal background. Using Eq. (A.1.2), a voter i chooses candidate j only if $\alpha_i \Delta e > 0$. Thus, a neutral voter always remains indifferent between both candidates since $\Delta u_i = 0$. In contrast, since $\Delta u_i > 0$, an ethnic voter will always select candidate j . The basic intuition here is straightforward: Given two candidates of similar quality, the coethnic will always have a comparative advantage. This advantage is dependent on the value of μ : an increase in μ or a rise in the ethnic population leads to a corresponding increase in the voter support for the coethnic candidate j and vice versa. Likewise, if candidate k is accused of criminality, it is easy to construe that no one will ever vote for them, since for all voters $\Delta u_i > 0$.

Now moving on to the most interesting case, when only the coethnic candidate j is accused of criminality. Thus, a voter i will choose candidate j if $\alpha_i \Delta e > 2\beta_i$. This means that the choice is dependent on the weight the voter attaches to the two dimensions and how much they value e . Note, I do not make any assumption on e . Therefore, for a constant given weight, since $\Delta e > 0$, as the $\lim \Delta e \rightarrow \infty$, the voter will always vote for the coethnic. This means that if voters value ethnicity more than quality, they might vote for the coethnic regardless of their criminal record. Similarly, as the voter moves towards neutrality ($\lim \Delta e \rightarrow 0$), the higher the probability the voter will bite the bullet and shift towards the non-criminal candidate. The intuition here is simple: if ethnicity matters more than probity, then this should trump candidate quality, including even those voters who dislike criminality. Lastly, the neutral voter will behave similarly as before, where their alliance will completely shift towards candidate k .

This result has two main implications: first, ethnic voting always plays a mitigating effect on criminality. This means that even though voters have a general distaste towards criminality, their relative utility derived from voting for a criminal coethnic is always higher in comparison to a criminal non-ethnic. Second, note that while in the previous case, the non-ethnic criminal had zero chance of winning, in this case, the criminal ethnic candidate could win. These predictions put together show that the quality threshold required for an ethnic candidate is always lower in comparison to that of a non-ethnic.

To summarize, the model under some general assumptions and conditions, predicts that voters on average favor coethnics. When their ethnically preferred candidate faces criminal charges, it reduces the likelihood that the voter chooses them. Regardless of this distaste for criminality, in certain circumstances, the criminal coethnic can win depending on the fraction of ethnic voters in the region. On the other hand, the neutral voter always prefers the better-quality candidate. Thus, the benchmark model shows how the ethnic preferences of the voter can sometimes mitigate their response towards criminal allegations.

A.3. Moving to a probabilistic voting model

A voter i utility for candidate j can be expressed as $u_{ij} + \epsilon_{ij}$, where ϵ_{ij} is assumed to be a realization of a random variable $\epsilon_{ij} \in (-\infty, \infty)$. As before, a voter chooses candidate j over k if $\Delta u_i + \epsilon_{ij} - \epsilon_{ik} > 0$. Let $\epsilon = \epsilon_{ij} - \epsilon_{ik}$, implying that the probability voter i votes for candidate j over candidate k is measured by $Pr(\Delta u_i > \epsilon)$. Thus, Eq. (A.1.2) can be re-written as:

$$Pr(Vote_i = j | j \text{ or } k) = Pr(\alpha_i \Delta e + \beta_i \Delta c + \gamma_i \Delta x > \epsilon) \quad (A.3.1)$$

Since the dependent variable is a binary choice, Eq. (A.3.1) can be expressed in probability terms, such that a voter i chooses candidate j over candidate k if

$$Pr(Vote_i = j | j \text{ or } k) = \frac{P_{ij}}{P_{ik}} = \frac{e^{u_{ij}}}{e^{u_{ik}}} = \frac{e^{\psi Z_{ij}}}{e^{\psi Z_{ik}}} \quad (A.3.2)$$

where Z represents all the observed explanatory variables and ψ represents the parameters obtained from the model. Note, since ϵ is assumed to be independent and identically distributed, the probability ratio depends only on the attributes of j and k and is independent of all other alternatives available. This is referred to as the independence of irrelevant alternatives or the IIA property. Thus using Eq. (A.1.2), we can directly estimate the likelihood of voting for a binary option given a set of observed explanatory variables. This can be further generalized to multiple candidates, such that let $C = \{1, 2, \dots, m\}$. McFadden (1973) proves that if IIA property holds or if $(\epsilon_1, \epsilon_2, \dots, \epsilon_m)$ is independently and identically distributed, then the addition of other candidates would not alter the probability. Thus, the probability that a voter i chooses candidate j over every other pair of choices $j, l \in C$ can be expressed as:

$$Pr(Vote_i = j | (1, 2, \dots, m)) = \frac{P_{ij}}{P_{il}} = \frac{e^{u_{ij}}}{e^{u_{ik}}} = \frac{e^{\psi Z_{ij}}}{e^{\psi Z_{il}}} \quad (A.3.3)$$

where, for every pair of candidates $j, l \in C$, we can estimate the coefficients by using a conditional logit model. Hence, this setup provides a clear intuition behind the use of a conditional model approach with various directly testable implications and guidelines for the selection variables.

A.4. Theoretical model extensions

A.4.1. Introducing the uniformed voter

The benchmark model relies on the assumption that the voter has perfect information on the candidates' criminal records. We can easily construe that this information might be noisy. For example, a voter might have to get access to electronic prints or the affidavits on the candidates' criminal charges, and then should have the ability to interpret this information correctly. We can expect this noise might be dependent on several factors, such as the literacy of the voter or their general interest in politics. Now assume

a voter still has a distaste for criminal candidates but might not know the true value of c . In particular, the voter now has a prior belief that (c_j, c_k) can either equal to $(1, 1)$, $(-1, 1)$, $(1, -1)$, or $(-1, -1)$ but this information is noisy. Hence, a voter i 's expected knowledge of the criminality status of the candidate can be written as:

$$c_i^e = \eta c_i + (1 - \eta)c_i^* \quad (\text{A.4.1.1})$$

where η is the probability the voter is informed. Assume that $\eta \neq 1$, meaning that the voter has imperfect information on criminality. c_i^* is the mean difference between the actual level of criminality and what the uninformed voter believes. Therefore, using Eq. (A.1.2), we can express that voter i will vote for candidate j if:

$$\Delta u_i = \alpha_i \Delta e + \beta_i \Delta c^e > 0 \quad (\text{A.4.1.2})$$

where, $\Delta c^e = c_{ij}^e - c_{ik}^e$. As before, we can re-evaluate the effect of information by looking at the three cases again. In the first case, there are two plausible scenarios: first, the ethnic voter has similar beliefs on the criminality status of both candidates such that $\eta_j = \eta_k$. Thus, the probability of voting for candidate j does not change regardless of how informed the voter is. Alternatively, the voter might believe that accusations against the non-ethnic candidate are more likely to be true than against the candidate of their ethnic preference such that $\eta_k > \eta_j$. This belief comes from the assumption that a voter with strong ethnic preferences might be more prone to believe that their preferred candidate might be falsely accused. This would lead to a rise in Δu . This implies that in comparison to the case with perfect information, the likelihood of voting for the ethnic candidate would increase. Note that regardless of the voter's expected knowledge of the criminality status, the voting outcome does not change if the voter values ethnicity ($e > 0$).

Moving to the second case, where candidate k faces criminal charges does not alter the voter response. This result holds due to two reasons: first, for the ethnic voter $\Delta u_i > 0$, so they will never choose candidate k . Second, the neutral voters will always shift their preferences towards the clean candidate if $\Delta c^e > 0$. The intuition here is straightforward: since the neutral voter only evaluates candidates based on their quality, even a modest indication of criminal allegations against one candidate, should completely shift their alliance towards the clean alternative.

Again we can observe the most significant effect of information when the coethnic candidate j faces criminal charges. Now voter i prefers candidate j if $\alpha_i \Delta e > \beta_i \Delta c^e$. Since $\eta \neq 1$, the expected belief of the voter is always lower than the true value ($\Delta c^e < 2$). This implies that imperfect information always increases the likelihood that a voter chooses the criminal coethnic. The intuition here is that as the voter's expected belief gets closer to the true criminality status (η rises), the $\lim \Delta c^e \rightarrow 2$, lowers the probability the voter selects the criminal candidate. Likewise, as η falls, the $\lim \Delta c^e \rightarrow 0$, the voter will solely base their decision on the ethnicity of the candidate. Again, this result ought to be obvious: the lower the information the voter has on the criminality of the candidate, the higher the weight they will put on ethnicity. Lastly, the neutral voter will act as before and completely shift their alliance towards the clean candidate k . Therefore, all things equal, the model predicts that low information availability increases the relative importance of ethnic cues in the voter's decision-making. Also note that in comparison to the case with perfect information, a reduction in information availability further reduces the quality threshold required for a coethnic to win.

A.4.2. Introducing the clientelistic voter

Now consider the voter response towards criminality when they consider clientelistic issues in their decision-making. As discussed earlier, voters might value the patronage on offer either in terms of resources or direct benefit. In such a case, voters' preferences might be more malleable towards the ethnic party if they believe that electing them to office would provide them with better access to public resources. Thus, the utility of voter i for candidate j can be expressed as:

$$u_{ij} = \alpha_i e_{ij} + \beta_i c_j + \gamma_i x_j + \delta_i B_{ij}^e \quad (\text{A.4.2.1})$$

where, $\delta > 0$ and B_{ij}^e is the expected benefit that voter i believes they can get from candidate j . I assume that the voter believes that

$$B_{ij}^e = \sigma e_{ij} - \kappa c_j \quad (\text{A.4.2.2})$$

where the voter considers both the ethnicity and candidate quality within themselves to acquire resources from the government. Note, that the voter evaluates both ethnicity and criminality as a positive signal for the delivery of the resources. This assumption is based on two main theoretical foundations: first, there is sufficient literature that theorizes that voters believe that candidates belonging to their ethnic group would be more likely to provide them with patronage or promote their best interest once in power (Posner, 2005). Second, previous studies have shown that criminal politicians might be more prone to engage in clientelistic strategies or might be better suited to "get things done" (Vaishnav, 2017). Thus, if criminal politicians are perceived to be more competent, voters might observe a positive expected benefit from criminality. As before, a voter i will vote for candidate j if:

$$\Delta u_i = \alpha_i \Delta e + \beta_i \Delta c + \delta_i \Delta B^e > 0 \quad (\text{A.4.2.3})$$

where, $\Delta B^e = B_{ij}^e - B_{ik}^e$. Using this setup, we can re-evaluate the three cases. Case one does not need much interpretation. Since both candidates have the same criminal record, the ethnic voter will always choose candidate j and the neutral voter would remain indifferent between both candidates. However, note that the model predicts that the bias towards the coethnic is even stronger. Since $e_{ij} > 0$, this directly implies that $\Delta B^e > 0$.

In case two, when only the non-ethnic candidate k faces criminal allegations, we can easily construe that the ethnic voter will not alter their response unless $\delta_i \kappa > \beta$ and e_{ij} is sufficiently small. Likewise, a neutral voter will vote for candidate j if $\beta_i > \delta_i \kappa$.

This implies that the neutral voter's decision is based on how much they value probity over getting resources and their expectation that the criminal candidate would be better at delivering these resources. This result shows that if the voter thinks that the criminal politician is more competent, this mitigates their distaste for criminality. Also note that in comparison to the benchmark model, where no one ever chose the criminal candidate, now some neutral voters will shift their preference towards the criminal.

Moving to the third case, the implications of patronage voting are quite straightforward. Holding other things constant, in comparison to the benchmark model, the ethnic voter will always vote for candidate j unless $\delta_i \Delta B^e > 2\beta_i$. This implies that the weight the voter attaches to candidate quality has to be larger than in the model without patronage. This result should be intuitively clear: since $B_{ij}^e > 0$, it has a positive effect on the voters' utility through both dimensions. First, the stronger the ethnic bias of the voter, the larger will be their expectation that a candidate belonging to their ethnic group would serve in their best interest. Second, if the voter believes that criminal politicians are more likely to distribute resources (κ rises), this would lead to a corresponding increase in the expected benefit (B^e). To offset this effect, β_i needs to be much larger in comparison to the benchmark model. To conclude, all things equal, we can observe that clientelistic voting plays a mitigating effect on criminality regardless of the candidates' ethnicity.

Appendix B. Robustness checks

See [Tables B.1–B.4](#).

Table B.1
Probit estimates on vote choice in Bihar 2020 state assembly elections.

	All	News frequency		Political knowledge		Literacy		Income	
		Low	High	Low	High	Illiterate	Literate	Poor	Not poor
Coethnic	0.780*** (0.107)	0.946*** (0.137)	−0.383* (0.202)	0.863*** (0.140)	−0.172 (0.202)	0.787*** (0.148)	−0.015 (0.201)	0.773*** (0.112)	0.0985 (0.333)
Violent criminal	−1.629*** (0.134)	−1.481*** (0.149)	−0.353** (0.155)	−1.559*** (0.153)	−0.114 (0.153)	−1.510*** (0.154)	−0.291 (0.153)	−1.626*** (0.136)	−0.00977 (0.246)
Coethnic × Violent criminal	−0.663*** (0.177)	−0.865*** (0.212)	0.485 (0.342)	−0.599*** (0.223)	−0.187 (0.321)	−0.579** (0.239)	−0.167 (0.316)	−0.656*** (0.183)	−0.0362 (0.626)
Non-violent criminal	−0.435*** (0.0949)	−0.326*** (0.114)	−0.240* (0.138)	−0.354*** (0.117)	−0.160 (0.137)	−0.252** (0.119)	−0.382*** (0.138)	−0.412*** (0.0982)	−0.188 (0.212)
Coethnic × Non-violent criminal	0.311** (0.128)	0.190 (0.164)	0.265 (0.248)	0.165 (0.166)	0.328 (0.248)	0.361** (0.176)	−0.114 (0.246)	0.327** (0.135)	−0.151 (0.393)
Candidate controls	Yes	Yes		Yes		Yes		Yes	
Party FE	Yes	Yes		Yes		Yes		Yes	
N	5499	5499		5499		5499		5499	

Notes: The table reports probit estimates with voter fixed effects for 1892 respondents. The dependent variable takes the value of 1 if the respondent chose the candidate and 0 otherwise. Column (1) provides estimates for the baseline specification. In columns (2)–(3) a voter is defined to have low news frequency if they do not consume news on a daily or weekly basis and high news frequency otherwise. In columns (4)–(5) a voter is defined as having low political knowledge if they responded to less than half of the political questions incorrectly and high political knowledge otherwise. In columns (6)–(7) a voter is defined as illiterate if they cannot read and write in the vernacular language and literate otherwise. In columns (8)–(9) a voter is defined as being poor if they own a below poverty line card or have a household income below 10,000 Rupees and not poor otherwise. All models include party fixed effects and candidate controls for their age, income in logs, education, and incumbency. Standard errors are given in parentheses. Asterisks denotes the significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.2
Impact of news on vote choice in Bihar 2020 state assembly elections.

	News consumption			
	Never	Daily/weekly	Monthly	>Month
Coethnic	1.194*** (0.269)	−0.568 (0.360)	0.405 (0.565)	−0.628 (0.737)
Violent criminal	−1.001*** (0.302)	−0.664** (0.333)	0.231 (0.526)	−0.891 (0.848)
Coethnic × Violent criminal	−1.008** (0.397)	0.416 (0.572)	−1.134 (0.890)	0.580 (0.210)
Non-violent criminal	−0.383 (0.210)	−0.244 (0.223)	0.503 (0.365)	−0.114 (0.467)
Coethnic × Non-violent criminal	−0.021 (0.344)	0.562 (0.470)	−0.697 (0.692)	1.428 (0.960)
Candidate controls		Yes		
Party FE		Yes		
N		5676		

Notes: The table reports conditional logit estimates for 1892 respondents. The dependent variable takes the value of 1 if the respondent chose the candidate and 0 otherwise. The data is grouped at the voter level accounting for individual voter-fixed effects. In columns (1)–(4) the estimates indicate the full range for the frequency of news consumption. All estimates include party fixed effects and candidate controls for their age, income in logs, education, and incumbency. Standard errors are given in parentheses. Asterisks denotes the significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B.3
Impact of political knowledge on vote choice in Bihar 2020 state assembly elections.

	Political knowledge								
	Zero	One	Two	Three	Four	Five	Six	Seven	Eight
Coethnic	1.289*** (0.425)	1.377 (0.866)	−0.212 (0.638)	−0.541 (0.679)	−0.0601 (0.679)	−0.363 (0.659)	−0.640 (0.575)	−0.618 (0.731)	−1.308* (0.727)
Violent criminal	−1.447*** (0.421)	1.045 (0.707)	0.665 (0.623)	0.533 (0.606)	0.250 (0.582)	−0.284 (0.672)	0.132 (0.529)	0.033 (0.642)	−0.980 (0.765)
Coethnic × Violent criminal	−1.061 (0.659)	−0.977 (1.125)	0.218 (0.941)	0.531 (0.992)	0.005 (1.082)	0.581 (0.989)	−0.081 (0.896)	0.048 (1.040)	−16.27 (4451)
Non-violent criminal	−0.162 (0.312)	−0.224 (0.463)	−0.068 (0.440)	0.066 (0.468)	−0.217 (0.426)	−0.291 (0.452)	−0.560 (0.358)	−0.830* (0.444)	−0.262 (0.446)
Coethnic × Non-violent criminal	−0.420 (0.508)	−1.257 (1.000)	0.489 (0.802)	0.994 (0.874)	−0.013 (0.835)	0.702 (0.847)	1.611** (0.727)	0.809 (0.954)	2.054** (0.959)
Candidate controls					Yes				
Party FE					Yes				
N					5676				

Notes: The table reports conditional logit estimates for 1892 respondents. The dependent variable takes the value of 1 if the respondent chose the candidate and 0 otherwise. The data is grouped at the voter level accounting for individual voter-fixed effects. In columns (1)–(9) the estimates indicate the number of total questions the voter answered correctly. All models include party fixed effects and candidate controls for their age, income in logs, education, and incumbency. Standard errors are given in parentheses. Asterisks denotes the significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

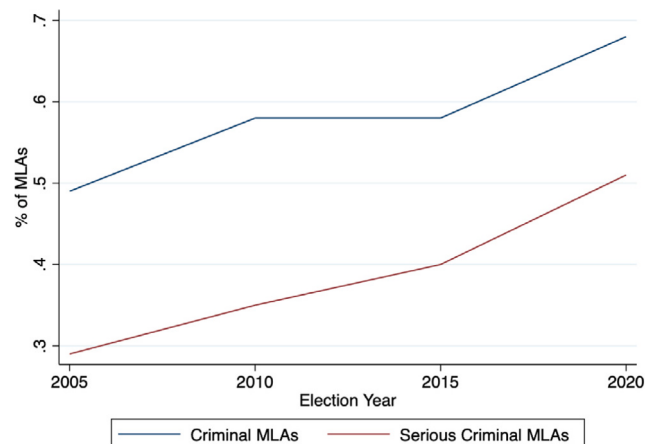
Table B.4
Voter heterogeneous effects on vote choice in Bihar 2020 state assembly elections.

	Gender		Age		Voting Reason		
	Female	Male	Young	Old	Development	Clientelism	Loyalty
Coethnic	0.993*** (0.289)	-0.139 (0.374)	1.033*** (0.252)	-0.215 (0.379)	0.825** (0.323)	0.348 (0.407)	-0.524 (0.632)
Violent criminal	-2.226*** (0.362)	-0.043 (0.348)	-2.417*** (0.330)	0.469 (0.351)	-2.199*** (0.384)	-0.071 (0.378)	0.102 (0.587)
Coethnic × Violent criminal	-0.723* (0.419)	-0.258 (0.536)	-0.915** (0.391)	0.180 (0.544)	-1.119** (0.503)	0.128 (0.585)	3.307** (1.311)
Non-violent criminal	-0.612** (0.276)	0.167 (0.246)	-0.510** (0.254)	-0.002 (0.243)	-0.391 (0.293)	-0.169 (0.269)	-0.125 (0.394)
Coethnic × Non-violent criminal	0.317 (0.350)	0.213 (0.464)	0.281 (0.315)	0.275 (0.468)	0.495 (0.401)	-0.478 (0.507)	1.068 (0.761)
Candidate controls	Yes		Yes		Yes		
Party FE	Yes		Yes		Yes		
N	5676		5676		5676		

Notes: The table reports conditional logit estimates for 1892 respondents. The dependent variable takes the value of 1 if the respondent chose the candidate and 0 otherwise. The data is grouped at the voter level accounting for individual voter-fixed effects. In column (1)–(2) the estimates are generated by gender. In columns (2)–(3) the estimates are generated by age where a young voter is defined as having an age below 50 and an old otherwise. In columns (4)–(6) the estimates are generated using the response of voters on their main voting reason in the Bihar 2020 elections (See Figure D.2 for more details). All models include party fixed effects and candidate controls for their age, income in logs, education, and incumbency. Standard errors are given in parentheses. Asterisks denotes the significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix C. Candidate profile and criminal backgrounds

See Fig. C.1 and Tables C.1–C.5.



Data Source: Association for Democratic Reform (ADR)

Fig. C.1. % of MLAs with criminal records in Bihar state assembly elections.
Data Source: Association for Democratic Reform (ADR).

Table C.1
Profile of candidates in Bihar 2020 state assembly election.

Variable	All candidates	Top 3	Winner
Age	43.13 (11.55)	51.60 (10.32)	47.40 (8.401)
Male	0.908 (0.290)	0.933 (0.254)	1 (0)
College degree	0.459 (0.50)	0.533 (0.507)	0.600 (0.516)
Hindu	0.914 (0.282)	0.900 (0.305)	0.800 (0.422)
Income (in Thousands)	16 693.50 (42 872.20)	48 956.80 (87 537)	34 260.8 (28 730.90)
Liabilities (in Thousands)	1134.30 (3192.20)	2217.80 (2885.30)	2246.9 (2587.90)
National Party	0.054 (0.227)	0.20 (0.407)	0.30 (0.483)
<i>N</i>	185	30	10

Notes: The table reports the candidate profile for all candidates contesting in the Bihar 2020 state assembly elections with mean coefficients and the standard deviation in parentheses.

Table C.2
Distribution of candidates by number of criminal charges in Bihar 2020 state assembly election.

Frequency	All candidates	Top 3	Winner
0	127	7	1
1	20	6	1
2–4	26	10	4
4–6	6	4	3
>6	6	3	1
<i>N</i>	185	30	10

Notes: The table reports the number of criminal charges for all candidates contesting in the Bihar 2020 state assembly elections.

Table C.3
Distribution of candidates by type of criminal charges in Bihar 2020 state assembly election.

Type	All candidates	Top 3	Winner
None	127	7	1
A. Non-violent corruption	14	4	1
Other charges	22	3	6
B. Violent	22	6	2
<i>N</i>	185	20	10

Notes: The table reports the distribution of candidates by the type of criminal accusations for all the candidates contesting in the Bihar 2020 assembly elections. The definition of violent and non-violent crimes is provided in Section 3.2. Corruption is defined as charges that lead to a financial loss to the government using the classification provided by the Indian Penal Code system.

Table C.4
Criminal candidates running from popular parties in Bihar 2020 state assembly election.

	Non-criminal	Non-violent	Violent	Total
BJP	0	2	1	3
BSP	2	1	1	4
CPI(M)	0	0	1	1
INC	1	0	0	1
JD(U)	3	2	1	6
LJP	4	3	0	7
RJD	0	7	1	8
<i>N</i>	10	15	5	30

Notes: The table reports the distribution of criminal candidates running from popular parties in the Bihar 2020 assembly elections from the ten assembly constituencies included in the sample size.

Table C.5
Non-criminal vs. Criminal.

Variable	Non-Criminal	Criminal	Diff
Age	42.29 (11.53)	44.97 (11.49)	-2.674 (-1.465)
Education (in years)	11.49 (5.210)	11.83 (4.867)	-0.339 (-0.419)
Income (in thousands)	14 675.70 (46 762.4)	21 111.80 (32 719.1)	-6436.10 (-0.947)
National party	0.0236 (0.152)	0.121 (0.329)	-0.097** (-2.749)
Incumbency	0.016 (0.125)	0.138 (0.348)	-0.122*** (-3.503)
N	127	58	185

Notes: The table reports a comparison of criminal versus non-criminal candidates for candidates contesting in the Bihar 2020 assembly elections. The coefficients presented are means along with the standard deviation in parentheses. Asterisks denotes the significance levels: * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Appendix D. Survey sampling and questionnaire

See Figs. D.1–D.2 and Tables D.1–D.2.

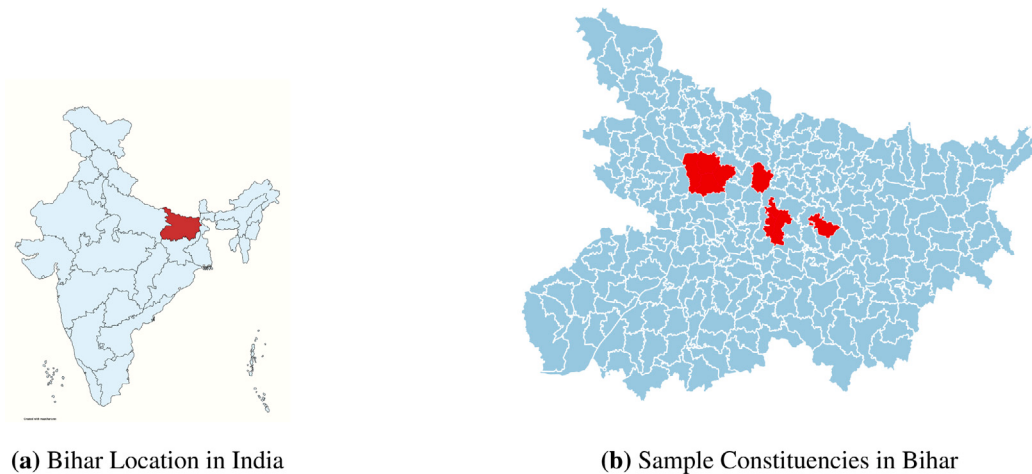


Fig. D.1. Survey sample area. **Notes:** Figure (a) shows the location of the state of Bihar in India. Figure (b) shows the 10 assembly constituencies within Bihar where the survey was conducted.

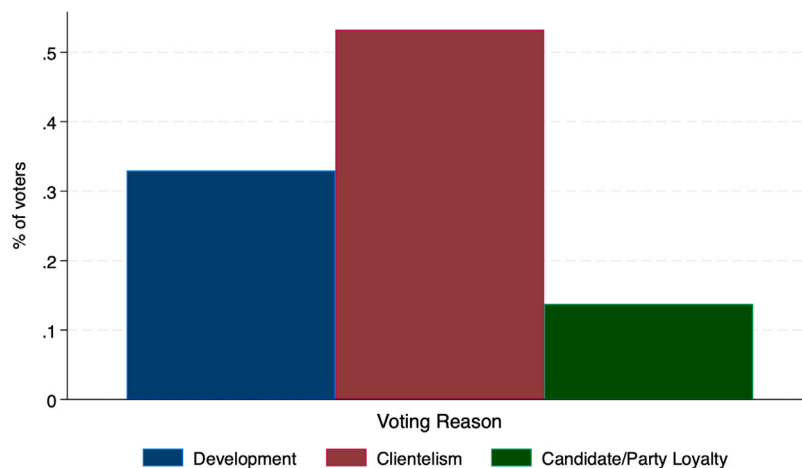


Fig. D.2. Main reason for voting in Bihar 2020 state assembly election. **Notes:** The figure reported is based on responses of 1892 respondents.

Table D.1

Sample size.

District	AC name	Obs.
Muzaffarpur	Baruraj	177
Muzaffarpur	Gaighat	194
Muzaffarpur	Kanti	137
Muzaffarpur	Paroo	194
Muzaffarpur	Sahebganj	196
Samastipur	Bhibutpur	198
Samastipur	Hasanpur	198
Samastipur	Samastipur	198
Samastipur	Sarairanjan	200
Samastipur	Ujiarpur	200

Notes: The table reports the responses of 1892 voters from the ten assembly constituencies across the two districts. The final sample size only includes those voters who completed the survey and voted in the Bihar 2020 elections.

Table D.2

Voter response to political awareness questions.

Variable	Mean	S.D.
MLA name	0.64	0.48
MLA party	0.48	0.50
MLA term limit	0.93	0.25
Chief minister name	0.65	0.48
Chief minister party	0.41	0.49
Prime minister name	0.74	0.44
Prime minister party	0.57	0.50
Lok Sabha vs. Rajya Sabha difference	0.16	0.36
Who elects Rajya Sabha members	0.07	0.26

Notes: The table reports the responses of 1892 respondents who completed the survey. The structure of the questions along with the choices are provided in [Appendix D](#) of the survey module.

Survey Questionnaire

Bihar 2020 Assembly Elections Voter Survey

Location

Surveyor ID:

District Name: ☐ Muzzafarpur ☐ Samastipur

PC ID:

AC ID:

Polling ID:

Voter Name: _____

Consent

Hello, My name is _____. I have come from IEB a research organization at the University of Barcelona. We want to find out the opinions of people on politics and elections, and for this purpose, we are interviewing voters. The findings of this study will be used for research purposes only. This survey is an independent study and is not linked to any political party or government agency. All the information you provide will be kept confidential. This survey will take about 20 to 25 minutes.

Do you agree to participate in this interview? ☐ Yes ☐ No → End Survey

Voter Contact No.: _____

A. Demographics

A1. What is your gender? ☐ Male ☐ Female

A2. What is your age in years? Surveyor: Keep in mind that age is recorded in completed years. Enter -90 if not tell or know. _____

A3. How many years have you been living here for? _____

A4. What is your religion? ☐ Hindu ☐ Muslim ☐ No Response

A5. What is your caste community?

- ☐ General/Upper Caste ☐ Other Backward Caste (OBC)
☐ Schedule Caste (SC) ☐ Schedule Tribe (ST)
☐ No Response

A6. Which sub-caste or *jati* you belong to? _____

A7. Can you read and write? ☐ Yes ☐ No

A8. What is your occupation?

- ☐ Laborer ☐ Farmer ☐ Shop Owner ☐ Medium/Big businessman
☐ Skilled Worker ☐ Semi-Skilled/ Unskilled Worker ☐ Clerical Jobs
☐ Government Employee ☐ Service/Professional Job ☐ Student
☐ Dairy/Fishery/Animal Farming ☐ Housewife ☐ Unemployed
☐ Retired ☐ Other (Please Specify): _____ ☐ No Response

A9. What is your monthly household income (including all family members currently living with you)? Rs. _____

B. Voting

B1. Did you vote in the assembly elections held between October and November of 2020?

- ☐ Yes → B1.1 ☐ No → B1.0 ☐ No Response → B1.0

B1.0 What was the main reason which you could not vote for in the assembly elections of 2020? Surveyor: Do not read options aloud.

- ☐ Out of Station ☐ Personal issues ☐ Voter ID issues
☐ Health issues ☐ Bad candidates ☐ Fear of coronavirus
☐ Voting does not make a difference ☐ No interest in voting
☐ Other (Please Specify): _____ ☐ No Response

B1.1 What was the most important factor for you in deciding whom to vote for in the assembly elections of 2020? Surveyor: Do not read options aloud.

- ☐ Government Performance ☐ Economic Growth/Development ☐ Inflation
☐ Improvement of public goods (Roads, electricity, water, toilet) ☐ Access to development program
☐ Voting for a particular party ☐ Removing NDA
☐ Bringing back NDA ☐ Caste of candidate/party ☐ Dissatisfied with current MLA performance
☐ Crime or corruption scandals ☐ Voting for a particular ideology ☐ Other (Please Specify): _____ No Response

B1.2 Which party/candidate did you vote for in the assembly elections of 2020? Surveyor: Hand over the list of candidates and parties to the respondent and enter the relevant code. Enter -90 if no response. _____

C. Information

How often do you read/listen/watch news using the following methods?

	Daily/Weekly	Monthly	Over a month	Never
Newspapers	☐	☐	☐	☐
Television	☐	☐	☐	☐
Social media/internet	☐	☐	☐	☐
WhatsApp	☐	☐	☐	☐

D. Political Knowledge

D1. Can you please tell me the name of the MLA in your assembly constituency?

D2. Can you please tell me which party the MLA in your assembly constituency belongs to? _____

D3. Can you please tell me what is the term limit of a MLA? ☐ Five Years ☐ Others/ Don't Know

D4. Can you please tell me the name of the Chief Minister (CM) of Bihar?

☐ Nitish Kumar ☐ Others/ Don't Know

D5. Can you please tell me which party the CM of Bihar belongs to?

☐ Janata Dal (United) (JD(U)) ☐ Others/ Don't Know

D6. Can you please tell me the name of the Prime Minister (PM) of India?

☐ Narendra Modi ☐ Others/ Don't Know

D7. Can you please tell me which party the PM of India belongs to?

☐ Bhartiya Janta Party (BJP) ☐ Others/ Don't Know

D8. Can you please tell me the difference between *Lok Sabha* and *Vidhan Sabha*?

☐ Lok Sabha is the legislative body at the central level and Vidhan Sabha is the legislative body at the state level. ☐ Others/ Don't Know

D9. Can you please tell me who elects members to the *Rajya Sabha*?

☐ Lok Sabha MP ☐ Others/ Don't Know

END OF SURVEY.

See Fig. E.1.

प्ररूप 26
(नियम 4क देखिए)

१४- साहेबगंज

(निर्वाचन क्षेत्र का नाम)

से बिहार विधान सभा (निर्वाचन का नाम) के निर्वाचन के लिए रिटर्निंग आफिसर के समक्ष अग्र्यर्थी द्वारा नाम-निर्देशन पत्र के साथ प्रस्तुत किया जाने वाला शपथ पत्र

भाग-क

मैं, राजू कुमार सिंह पुत्र/पुत्री/पत्नी श्री उदय प्रताप सिंह आयु 50 वर्ष, जो ग्राम- आनन्दपुर खसौली, पीठ- बाड़ाडाउड, थाना- पार, जिला- मुजफ्फरपुर, बिहार, पिन- 843112 (डाक का पूरा पता लिखें) का/की निवासी हूँ, और उपरोक्त निर्वाचन के लिए अग्र्यर्थी हूँ सत्यनिष्ठा से प्रतिज्ञा करता हूँ/करती हूँ, शपथ पर निम्नलिखित कथन करता हूँ/करती हूँ :-

(1) मैं विकाशशील इंसान पार्टी (राजनैतिक दल का नाम) द्वारा खड़ा किया गया अग्र्यर्थी/एक स्वतंत्र अग्र्यर्थी के रूप में लड़ रहा हूँ।

(जो लागू न हो उसे काट दें)

(2) मेरा नाम १४- साहेबगंज (निर्वाचन-क्षेत्र और राज्य का नाम) में भाग सं० 201 के क्रम सं० 1301 पर प्रविष्ट है।

(3) मेरा/मेरे 7004761388 संपर्क दूरभाष संख्या/संख्याएं हैं/हैं और शुभ्य मेरा ईमेल पता (यदि कोई हो) है तथा मेरा/मेरे सोशल मीडिया खाता/खाते (यदि कोई हो) निम्नलिखित हैं/हैं।

(i) शुभ्य
(ii) शुभ्य
(iii) शुभ्य

सिंह

Fig. E.1. Example of candidate affidavit. **Notes:** The figure shows the first page and the relevant page with criminal charges from the candidate affidavit for the Bihar 2020 assembly elections. The full version of the affidavit is available on the ECI website.

ANNEXURE-(I)
राजू कुमार सिंह
मुकदमा का व्योरा:-

क्र० सं०	न्यायालय का नाम	मामला सं०	संज्ञान लेने के आदेश की तारीख	द्वारा एवं अपराध
1	एस०डी०जे०एम०, प०	पारु थाना काण्ड सं०-67/06	24.01.2008	323,504 भा०द०वि०
2	दीपक कुमार, जे०एम० प०	नगर थाना काण्ड सं०-394/10	05.11.2012	188 भा०द०वि०
3	एस०डी०जे०एम०, प०	देवरिया थाना काण्ड संख्या-91/10	07.09.2012	341, 323, 504, 379/34 भा०द०वि०
4	एस०डी०जे०एम०-3	पारु थाना काण्ड सं०-115/15	अनुसंधान जारी	147, 148, 143, 341, 353, 504/34 भा०द०वि० एवं 27 आर्मस एक्ट
5	न्या० दण्डाधिकारी, प्रथम श्रेणी	पारु थाना काण्ड सं०-116/15	अनुसंधान जारी	147, 148, 149, 341, 323, 447, 307, 504, 506, भा०द०वि० एवं 27 आर्मस एक्ट
6	एस०डी०जे०एम० प०	पारु थाना काण्ड सं०-14/17	एस०डी०जे०एम०प०, मुज० जी०आर०नं०-117315	302/34 भा०द०वि० एवं 27 आर्मस एक्ट
7	एस०डी०जे०एम०-1, प०	साहेबगंज थाना काण्ड सं०-165/05	एस०डी०जे०एम०-1, पूर्वी, मुज०।	188 भा०द०वि०
8	जे०एम०प्रथम श्रेणी, पटना सिविल कोर्ट	न्यायालय मुकदमा संख्या-1426/सी०डब्ल्यू-20/12	श्रीमती आरती जयसवाल, प्रथम न्यायिक दण्डाधिकारी, पटना।	420, 426, 120 (बी०) भा०द०वि० एवं 138 एन०आई०एक्ट
9	श्रीमती स रीबन बहालिया, उप० न्या०द०-14, एस०डी०जे०एम०, पटना।	गोंधी मैदान पटना काण्ड संख्या-491/2017	एस०डी०जे०एम०-14, पटना,	341,342, 323,330,504,506/34 भा०द०वि०
10	साकेत न्यायालय, नई दिल्ली।	फतेहपुर बेरी, नई दिल्ली, काण्ड सं०-01/2019	अनुसंधान जारी जमानत पर	302 भा०द०वि० एवं 27 आर्मस एक्ट

Fig. E.1. (continued).

References

- Aarts, K., Semetko, H.A., 2003. The divided electorate: Media use and political involvement. *J. Politics* 65 (3), 759–784. <http://dx.doi.org/10.1111/1468-2508.00211>.
- Adida, C., 2015. Do african voters favor coethnics? Evidence from a survey experiment in Benin. *J. Exp. Political Sci.* 2 (1), 1–11. <http://dx.doi.org/10.1017/XPS.2014.26>.
- Adida, C., Gottlieb, J., Kramon, E., McClendon, G., 2017. Overcoming or reinforcing coethnic preferences? An experiment on information and ethnic voting. *Q. J. Political Sci.* 12 (4), 437–477. <http://dx.doi.org/10.1561/100.00017018>.
- Agarwal, A., Agrawal, N., Bhogale, S., et al., 2021. TCPD Indian elections data v2.0. Trivedi Centre for Political Data, Ashoka University.
- Akerlof, G.A., Kranton, R.E., 2000. Economics and identity. *Q. J. Econ.* 115 (3), 715–753. <http://dx.doi.org/10.1162/003355300554881>.
- Alvarez, M.R., Nagler, J., 1998. When politics and models collide: Estimating models of multiparty elections. *Am. J. Political Sci.* 55–96. <http://dx.doi.org/10.2307/2991747>.
- Anduiza, E., Gallego, A., Muñoz, J., 2013. Turning a blind eye: Experimental evidence of partisan bias in attitudes toward corruption. *Comp. Political Stud.* 46 (12), 1664–1692. <http://dx.doi.org/10.1177/001041401348908>.
- Ansolabehere, S., Puy, S.M., 2016. Identity voting. *Public Choice* 169 (1), 77–95. <http://dx.doi.org/10.1007/s11127-016-0371->.
- Ansolabehere, S., Snyder, Jr., J.M., 2000. Valence politics and equilibrium in spatial election models. *Public Choice* 103 (3), 327–336. <http://dx.doi.org/10.1023/A:1005020606153>.
- Auerbach, A.M., Bussell, J., Chauchard, S., et al., 2022. Rethinking the study of electoral politics in the developing world: Reflections on the Indian case. *Perspect. Politics* 20 (1), 250–264. <http://dx.doi.org/10.1017/S1537592721000062>.
- Banerjee, A., Green, D.P., McManus, J., Pande, R., 2014. Are poor voters indifferent to whether elected leaders are criminal or corrupt? A vignette experiment in rural India. *Political Commun.* 31 (3), 391–407. <http://dx.doi.org/10.1080/10584609.2014.914615>.
- Banerjee, A., Kumar, S., Pande, R., Su, F., 2011. Do informed voters make better choices? Experimental evidence from urban India. Harvard University, Unpublished manuscript.
- Banerjee, A., Pande, R., 2007. Parochial politics: Ethnic preferences and politician corruption. CEPR discussion paper no. DP6381.
- Berenschot, W., 2011a. On the usefulness of *Goondas* in Indian politics: 'Moneypower' and 'Musclepower' in a Gujarati locality. *South Asia: J. South Asian Stud.* 34 (2), 255–275.
- Berenschot, W., 2011b. The spatial distribution of riots: Patronage and the instigation of communal violence in Gujarat, India. *World Dev.* 39 (2), 221–230.
- Besley, T., 2007. The new political economy. *Econ. J.* 117 (524), F570–F587.
- Bratton, M., 2008. Vote buying and violence in Nigerian election campaigns. *Elect. Stud.* 27 (4), 621–623.
- Carlson, E., 2015. Ethnic voting and accountability in africa: A choice experiment in Uganda. *World Politics* 67 (2), 353–385.
- Carlson, E., 2016. Identifying and interpreting the sensitivity of ethnic voting in Africa. *Public Opinion Q.* 80 (4), 837–857.
- Caselli, F., Morelli, M., 2004. Bad politicians. *J. Public Econ.* 88 (3–4), 759–782.
- Chandra, K., 2007. Why ethnic parties succeed: Patronage and ethnic head counts in India. Cambridge University Press.
- Chang, E.C., Kerr, N.N., 2017. An insider-outsider theory of popular tolerance for corrupt politicians. *Governance* 30 (1), 67–84.
- Chauchard, S., 2016. Unpacking ethnic preferences: Theory and micro-level evidence from north India. *Comp. Political Stud.* 49 (2), 253–284.
- Chauchard, S., Klačnja, M., Harish, S.P., 2019. Getting rich too fast? Voters' reactions to politicians' wealth accumulation. *J. Politics* 81 (4), 1197–1209.
- Chemin, M., 2012. Welfare effects of criminal politicians: a discontinuity-based approach. *The Journal of Law and Economics* 55 (3), 667–690.
- Costas-Pérez, E., Solé-Ollé, A., Sorribas-Navarro, P., 2012. Corruption scandals, voter information, and accountability. *Eur. J. Political Econ.* 28 (4), 469–484.
- Ferraz, C., Finan, F., 2008. Exposing corrupt politicians: The effects of Brazil's publicly released audits on electoral outcomes. *Q. J. Econ.* 123 (2), 703–745.
- Ferree, K.E., 2010. Framing the race in South Africa: The political origins of racial census elections. Cambridge University Press.
- Frederiksen, K.V.S., 2023. Do partisanship and policy agreement make citizens tolerate undemocratic behavior? Working Paper.
- Gehring, K., Kauffeldt, T.F., Vadlamannati, K.C., 2019. Crime, incentives and political effort: Evidence from India. *Eur. J. Political Econ.* 59, 1–20.
- George, S., Gupta, S., Neggers, Y., 2018. Coordinating voters against criminal politicians: Evidence from a mobile experiment in India. Unpublished Manuscript.
- Groseclose, T., 2001. A model of candidate location when one candidate has a valence advantage. *Am. J. Political Sci.* 862–886.
- Gupta, T.D., 1992. Yadav ascendancy in Bihar politics. *Econ. Political Wkly.* 1304–1306.
- Horowitz, D.L., 2001. Ethnic groups in conflict. University of California Press.
- Kapur, D., Vaishnav, M., 2013. Quid pro quo: Builders, politicians, and election finance in India. Center for Global Development Working Paper, 276.
- Karakas, L.D., Mitra, D., 2021. Electoral competition in the presence of identity politics. *J. Theor. Politics* 33 (2), 169–197.
- Klačnja, M., 2017. Uninformed voters and corrupt politicians. *Am. Politics Res.* 45 (2), 256–279.
- Klor, E.F., Shayo, M., 2010. Social identity and preferences over redistribution. *J. Public Econ.* 94 (3–4), 269–278.
- Kyriacou, A.P., 2023. Clientelism and fiscal redistribution: Evidence across countries. *Eur. J. Political Econ.* 76, 102234.
- Lindberg, S.I., Morrison, M.K., 2008. Are African voters really ethnic or clientelistic? Survey evidence from Ghana. *Political Sci. Q.* 123 (1), 95–122.
- Martin, N., Michelutti, L., 2017. Protection rackets and party machines: Comparative ethnographies of "Mafia Raj" in north India. *Asian J. Soc. Sci.* 45 (6), 693–723.
- McFadden, D., 1973. Conditional logit analysis of qualitative choice behavior. In: Zarembka, P. (Ed.), *Frontiers in Econometrics*. Academic Press, New York, (chap. 4).
- Posner, D.N., 2005. Institutions and ethnic politics in Africa. Cambridge University Press.
- Prakash, N., Rockmore, M., Uppal, Y., 2019. Do criminally accused politicians affect economic outcomes? Evidence from India. *J. Dev. Econ.* 141, 102370.
- Reserve Bank of India, 2022. Handbook of statistics on Indian states. Retrieved from www.rbi.org.in.
- Reuter, O.J., Szakonyi, D., 2021. Electoral manipulation and regime support: Survey evidence from Russia. *World Politics* 73 (2), 275–314.
- Schofield, N., 2004. Equilibrium in the spatial 'Valence' model of politics. *J. Theor. Politics* 16 (4), 447–481.
- Schönhage, N.L., Geys, B., 2022. Partisan bias in politicians' perception of scandals. *Party Politics* 28 (4), 691–701.
- Schönhage, N.L., Geys, B., 2023. Politicians and scandals that damage the party brand. *Legis. Stud. Q.* 48 (2), 305–331.
- Stokes, S.C., Dunning, T., Nazareno, M., Brusco, V., 2013. Brokers, Voters, and Clientelism: The Puzzle of Distributive Politics. Cambridge University Press.
- Vaishnav, M., 2017. When crime pays: Money and muscle in Indian politics. Yale University Press.
- Wade, R., 1985. The market for public office: Why the Indian state is not better at development. *World Dev.* 13 (4), 467–497.
- Zakharov, N., 2019. Does corruption hinder investment? Evidence from Russian regions. *Eur. J. Political Econ.* 56, 39–61.