PNAS RESEARCH ARTICLE | Election

Voter Turnouts Govern Key Electoral Statistics

Ritam Pal[a], Aanjaneya Kumar[b,c], and M. S. Santhanam[a]

This manuscript was compiled on December 24, 2025

Elections are central institutions in all the democratic societies. They are regarded as complex system due to varied human interactions and interests that shape. Quantitative analysis of election data has provided us insights into existence of common patterns in elections. In this work, we show that the voter turnout – the number of people who actually vote on the election day and hence represents their faith and interest in the process – contains crucial information that can help to accurately predict several key electoral statistics. Using empirical election data from 12 countries spanning multiple decades and random voting model, we show that the distributions of votes secured by winners and runner-ups are strongly correlated with turnout distributions. The former can be predicted from the knowledge of turnout distribution. This new direction in the quantitative study of elections can provide newer ways to diagnose elections.

elections | turnouts | voteshare | model

Free and fair elections play a pivotal role in functioning democracies, ensuring that governing bodies reflect the people's mandate. Understanding elections from a political, social and quantitative perspective is essential to safeguard their integrity and uncover potential universal patterns that can emerge across different countries and contexts.

In the last three decades, aided by the availability of extensive election data, many studies (1–6) had identified patterns in the distributions of vote shares garnered by candidates (7–10), victory margins (16–21) and voter turnouts (11, 12). However, these patterns exhibit universality (1, 2, 4), usually limited to some geographical regions or to some types of elections. In this backdrop, a recent work by the authors analyzed a large corpus of election data from 34 countries spanning multiple decades and showed that a suitably scaled distribution of the ratio M/T, where $M = V_w - V_r$ is the victory margin and T is the turnout, with V_w and V_r denoting the votes secured by the winner and runner-up, exhibits a robust universality independent of the details of the electoral processes (22). Beyond universality, turnouts play a fundamental role in driving the victory margins. In conjunction with Random Voting Model (RVM), turnouts can accurately predict the scaled distribution of victory margins (22).

These findings naturally raise the question: Does voter turnouts contain information about other key electoral statistics? And can we recover those statistics using RVM? To explore these questions, we employ the RVM and empirical data from 12 countries to demonstrate that the voter turnout distribution $g(T)$, combined with the effective number of candidates (defined later (31)), is sufficient to accurately predict two fundamental electoral statistics: the scaled vote distributions of the winner $Q_1(V_w)$ and the runner-up $Q_2(V_r)$. This prediction holds across all electoral scales—from large parliamentary constituencies ($\sim 10^5 - 10^6$ voters) down to the smallest polling booth levels ($\sim 10^2 - 10^3$ voters). Thus, irrespective of electoral system or size, voter turnout emerges as a fundamental metric determining key electoral outcomes and offering a new diagnostic for election health.

1. Materials and methods

Framework: An *election* happens at all the N electoral units following the first-post-the-past principle (30). Let

Author affiliations: [a]Department of Physics, Indian Institute of Science Education and Research, Pune 411008, India.; [b]Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA; [c]High Meadows Environmental Institute, Princeton University, Princeton, NJ, 08544, USA

DRAFT

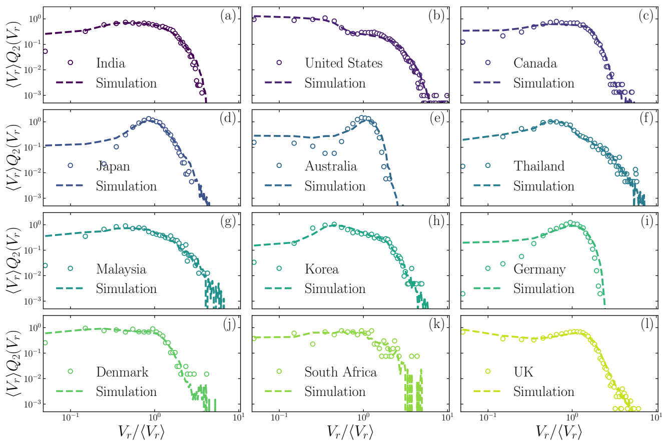

Fig. 1. Scaled distribution of the votes V_r received by the runner-up for 12 countries. Empirical election data (circles) agrees with RVM simulations (dashed line) obtained using turnout distribution $g(T)$ as input.

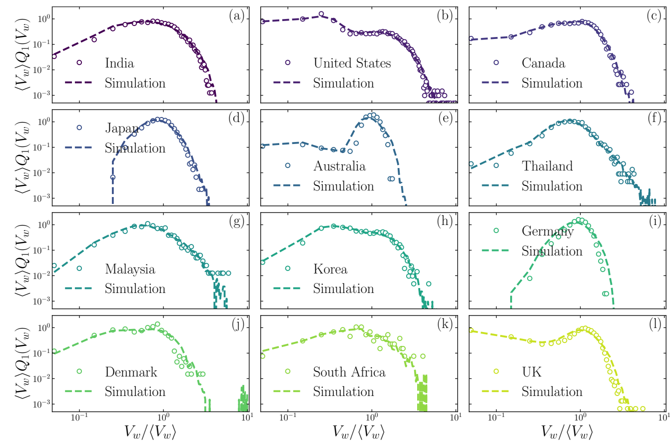

Fig. 2. Scaled distribution of the votes V_w received by the winners for 12 countries. Empirical data (circles) agrees with RVM simulations (dashed line) obtained using turnout distribution $g(T)$ as input.

the i-th electoral unit (constituency, Congressional District, polling booth etc.) have c_i candidates and n_i voters, where $i = 1, 2, \ldots N$. Usually, only a fraction of the voters cast their votes. This is termed the turnout $T_i \le n_i$, and indicates people's interest in the electoral process. Let $V_{i,1}, V_{i,2}, \ldots, V_{i,c_i}$ be the votes secured by c_i candidates such that $\sum_{j=1}^{c_i} V_{i,j} = T_i$. The candidate securing most votes, $V_w = \max(V_{i,1}, V_{i,2}, \ldots, V_{i,c_i})$ is declared the winner, and the candidate with the second-highest number of votes, V_r, is the runner-up.

Model: The random voting model (22), denoted as RVM(T, c), is an abstraction of the election framework described above. It takes $(T = \{T_1, T_2, \ldots, T_N\}, c = \{c_1, c_2, \ldots, c_N\})$ as input. Then, the probability that j-th candidate in i-th electoral unit attracts electors' votes is:

$$p_{ij} = \frac{w_{ij}}{\sum_{k=1}^{c_i} w_{ik}}, \quad w_{ij} \sim \mathcal{U}(0,1), \quad (j = 1, 2, \cdots c_i). \quad \textbf{[1]}$$

In this, $\mathcal{U}(0,1)$ is the uniform distribution. This model captures the statistical features of empirical election data and the universality embedded in the election data (22).

Effective Number of Candidates: In most large elections, many candidates enter the fray but only a few corner most of the votes. To quantify this concentration, we use the effective number of candidates (eNoC) introduced in (31) and defined for each electoral unit i as:

$$c_i^{\text{eff}} = \frac{1}{\sum_{k=1}^{c_i} (V_{i,k}/T_i)^2}, \qquad i = 1, 2, \ldots, N. \qquad \textbf{[2]}$$

This measure reflects how many candidates are meaningfully competitive. If a single candidate receives all the votes, then $c_i^{\text{eff}} = 1$. However, if the votes are shared equally among $\mathcal{C}$ candidates, $c_i^{\text{eff}} = \mathcal{C}$. Thus, Eq. 2 provides a natural and quantitative way to characterize the eNoC in the i-th electoral unit.

Data: Empirical national election data from 12 countries and polling booth level data for Canada and India are obtained from []. Basic statistical features of these 12 data sets and other processing details are given in SI.

2. Information embedded in the voter turnouts

Scaled vote distributions for winners and runners-up were examined using national election data from 12 countries, each spanning multiple decades. For each country, we consider the quantity $\langle V_w \rangle Q_1(V_w)$ as a function of the scaled winner votes $V_w/\langle V_w \rangle$. Although the resulting curves (circles) in figure 2 vary between countries, the overall structure shows a broad central region with relatively rapid decay in the tails, except in the case of Japan and Germany. We find a similar structure when analyzing the Runner-up votes. The Plots of $\langle V_r \rangle Q_2(V_r)$ as a function of $V_r/\langle V_r \rangle$, display analogous shapes as demonstrated in figure 1. Together, these distributions capture the features of two essential electoral statistics – the scaled vote distribution of the winners and the runner-ups.

A key insight emerges when these empirical profiles are compared with predictions from the Random Voting Model (RVM). The RVM requires only two empirical inputs: (a) the observed turnout in each electoral unit and (b) the corresponding effective number of candidates c_i^{eff}, rounded to the closest integer. No other country-specific assumptions are imposed. Using these inputs, the model generates vote totals for all candidates, from which the winner and runner-up distributions can be constructed. The simulated curves closely mimic the empirical ones for all countries examined. This agreement indicates that turnout statistics encode quantitative information about both $Q_1(V_w)$ and $Q_2(V_r)$, and that this information is rich enough for the RVM to reproduce the observed shapes without additional assumptions about the underlying political processes.

We further investigate this turnout-driven behavior analytically and derive the expressions for the vote-share distributions of winners and runners-up by solving 3 candidate RVM in the limit of large turnout, $T \gg 1$. In this regime, with $c_i = 3; i = 1, 2, \ldots, N$, general forms for the distributions $P_1(v_w)$ and $P_2(v_r)$ can be derived where $v_w = V_w/T$ and $v_r = V_r/T$ are the vote shares of the winners and the runner-ups. By combining them with the empirical turnout distribution $g(T)$ for each country, we compute the numerical predictions for the corresponding winner and runner-up vote distributions, $Q_1(V_w)$ and $Q_2(V_r)$. After suitable scaling the resulting analytical curves coincide closely with both the empirical distributions and the simulated RVM outputs for countries such as the United States and India as demonstrated by

DRAFT

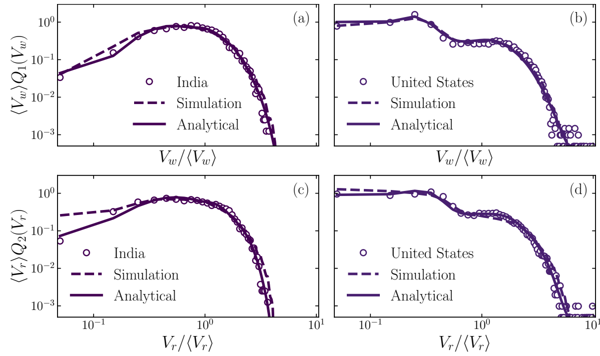

Fig. 3. Scaled distribution of the votes received by the winner and the runner-up. The analytical prediction (solid line) closely agrees with the RVM simulations (dashed line) and the empirical distributions (circles) for India and the USA.

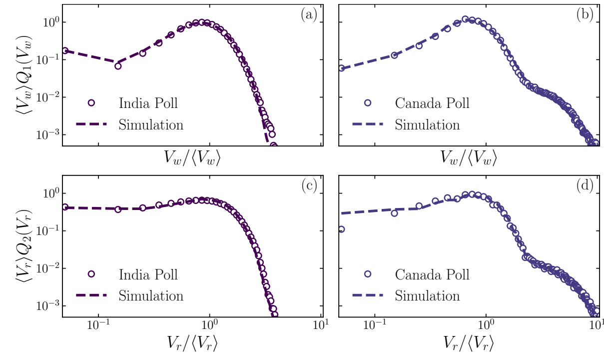

Fig. 4. Scaled distribution of the votes received by the winner and the runner-up in polling level data for India and Canada. The empirical data (circles) closely agree with the RVM simulations (dashed line) that takes turnout data as input.

the solid lines in figure 3. Full derivations and explicit expressions for the analytical forms are provided in the SI. This consistency across data, simulations, and analytics confirms that the turnout distribution plays the central role in shaping the observed electoral statistics.

Do these results in Figs. 2-3, obtained for national elections, hold good at smaller electoral scales such as at the polling booth levels? To answer this question, we use parliamentary constituency (PC) and polling booth (PB) level data from India and Canada. In Canada, the average electorate size in PC and PB is $\sim 10^4$ and 10^2, while the numbers for India are 10^5 and 10^2 – both displaying a clear separation of scales in terms of electorate size. Recall that Fig. 2(a,c) shows scaled distribution of winner votes in PC level data, respectively, for India and Canada, and the corresponding runner-up distributions in Fig. 1(a,c), empirical data in good agreement with RVM simulations. Remarkably, at PB levels too, in Fig. 4(a,c) for India and in Fig. 4(b,d) for Canada a similar agreement is observed between the empirical distribution and RVM simulations using the actual turnout data. This reinforces the central result that, in fair elections, turnouts govern statistics of winner and runner-up votes, and this feature is valid at all electoral scales.

In summary, elections datasets are excellent sources for exploring collective decision making by people. The voter turnout is usually interpreted as an indicator of the public trust and interest in the electoral process. Going beyond this convention, we have demonstrated that the voter turnout distribution $g(T)$ encodes significant information about several crucial election statistics, namely, quantitative information about (scaled) distribution of votes, $Q_1(V_w)$ and $Q_2(V_r)$, received by winner and runner-up in fair elections. We demonstrate this using empirical data from 12 countries of multiple elections held over many decades. This is further confirmed tby use of Random Voting Model, which takes empirical turnout data and an effective number of candidates as input, and correctly reproduces the empirically observed distributions $Q_1(V_w)$ and $Q_2(V_r)$. This is further strengthened by analytical solution to RVM, which also agrees with $Q_1(V_w)$ and $Q2(V_r)$. Remarkably, all these results hold good at all electoral scales – from large constituencies/Congressional District levels down to small polling booth levels. These results point to a new direction in the quantitative study of elections: electoral turnout data contains much more information than is usually attributed to it, and it can provide new ways to diagnose elections. Voter turnout is not just a statistical curiosity of public interest, but an indicator of the health of fair elections. Hence, election conducting bodies across the world must report turnout data at all levels to help analyze the fairness of the process.

DRAFT

ACKNOWLEDGMENTS. R.P. and A.K. thank the Prime Minister's Research Fellowship of the Government of India for financial support. The authors acknowledge the National Supercomputing Mission for the use of PARAM Brahma at IISER Pune.

1. R. N. C. Filho, M. P. Almeida, J. S. Andrade, and J. E. Moreira, Scaling behavior in a proportional voting process, Phys. Rev. E **60**, 1067 (1999) .
2. S. Fortunato and C. Castellano, Scaling and universality in proportional elections, Phys. Rev. Lett. **99**, 138701 (2007) .
3. M. Mantovani, H. Ribeiro, M. Moro, S. Picoli, and R. Mendes, Scaling laws and universality in the choice of election candidates, Europhysics Letters **96**, 48001 (2011) .
4. A. Chatterjee, M. Mitrović, and S. Fortunato, Universality in voting behavior: an empirical analysis, Scientific reports **3**, 1049 (2013) .
5. E. Bokányi, Z. Szállási, and G. Vattay, Universal scaling laws in metro area election results, PLOS ONE **13**, 1 (2018) .
6. V. Hösel, J. Müller, and A. Tellier, Universality of neutral models: Decision process in politics, Palgrave Communications **5**, 1 (2019) .
7. A. M. Calvão, N. Crokidakis, and C. Anteneodo, Stylized facts in brazilian vote distributions, PLOS ONE **10**, 1 (2015) .
8. K. Burghardt, W. Rand, and M. Girvan, Competing opinions and stubborness: Connecting models to data, Phys. Rev. E **93**, 032305 (2016) .
9. S. Mori, M. Hisakado, and K. Nakayama, Voter model on networks and the multivariate beta distribution, Phys. Rev. E **99**, 052307 (2019) .
10. A. Kononovicius, Modeling of the parties' vote share distributions, Acta Physica Polonica A **133**, 1450 (2018) .
11. C. Borghesi and J.-P. Bouchaud, Spatial correlations in vote statistics: a diffusive field model for decision-making, Eur. Phys. J. B **75**, 395 (2010) .
12. C. Borghesi, J.-C. Raynal, and J.-P. Bouchaud, Election turnout statistics in many countries: Similarities, differences, and a diffusive field model for decision-making, PLOS ONE **7**, 1 (2012) .
13. P. Klimek, Y. Yegorov, R. Hanel, and S. Thurner, Statistical detection of systematic election irregularities, Proceedings of the National Academy of Sciences **109**, 16469 (2012) .
14. R. Jimenez, M. Hidalgo, and P. Klimek, Testing for voter rigging in small polling stations, Science advances **3**, e1602363 (2017) .
15. A. Rozenas, Detecting election fraud from irregularities in vote-share distributions, Political Analysis **25**, 41 (2017) .
16. G. C. Jacobson, The marginals never vanished: Incumbency and competition in elections to the us house of representatives, 1952-82, American Journal of Political Science , 126 (1987) .
17. S. J. McCann, Threatening times," strong" presidential popular vote winners, and the victory margin, 1824–1964., Journal of personality and social psychology **73**, 160 (1997) .
18. C. B. Mulligan and C. G. Hunter, The empirical frequency of a pivotal vote, Public Choice **116**, 31 (2003) .
19. T. R. Magrino, R. L. Rivest, and E. Shen, Computing the margin of victory in {IRV} elections, in *2011 Electronic Voting Technology Workshop/Workshop on Trustworthy Elections (EVT/WOTE 11)* (2011) .
20. L. Xia, Computing the margin of victory for various voting rules, in *Proceedings of the 13th ACM conference on electronic commerce* (2012) pp. 982–999 .
21. A. Bhattacharyya and P. Dey, Predicting winner and estimating margin of victory in elections using sampling, Artificial Intelligence **296**, 103476 (2021) .
22. R. Pal, A. Kumar, and M. Santhanam, Universal statistics of competition in democratic elections, arXiv preprint arXiv:2401.05065 (2024) .
23. G. Brigaldino, Elections in the imperial periphery: Ethiopia hijacked, Review of African political economy **38**, 327 (2011) .
24. M. Frear, The parliamentary elections in belarus, september 2012, Electoral Studies **33**, 350 (2014) .

25. Report of organization for security and co-operation in europe (osce), https://www.osce.org/odihr/elections/belarus (2020) .
26. A. Czwołek and J. Kołodziejska, Belarusian parliamentary election in 2019, The Copernicus Journal of Political Studies , 81 (2021) .
27. S. Bedford, The 2020 presidential election in belarus: Erosion of authoritarian stability and re-politicization of society, Nationalities Papers **49**, 808–819 (2021) .
28. Election data of india, https://www.eci.gov.in .
29. Lok dhaba - a repository of indian election results, https://lokdhaba.ashoka.edu.in/ .
30. L. Johnston, *Politics: An introduction to the modern democratic state* (University of Toronto Press, 2007) .
31. M. Laakso and R. Taagepera, "effective" number of parties: a measure with application to west europe, Comparative political studies **12**, 3 (1979) .
32. B. C. Arnold, N. Balakrishnan, and H. N. Nagaraja, *A First Course in Order Statistics* (Society for Industrial and Applied Mathematics, 2008) .
33. See Supplemental Material [URL] for (1) the description of RVM, (2) theoretical calculations for RVM and other related discussions, (3) data summary, and (4) figures.

DRAFT