Voter Turnouts Govern Key Electoral Statistics

Ritam Pal,[1, *] Aanjaneya Kumar,[1, 2, 3, †] and M. S. Santhanam[1, ‡]

[1]*Department of Physics, Indian Institute of Science Education and Research, Pune 411008, India.*
[2]*Santa Fe Institute, 1399 Hyde Park Road, Santa Fe, NM 87501, USA*
[3]*High Meadows Environmental Institute, Princeton University, Princeton, NJ, 08544, USA*

Elections, the cornerstone of democratic societies, are usually regarded as unpredictable due to the complex interactions that shape them at different levels. In this work, we show that voter turnouts contain crucial information that can be leveraged to predict several key electoral statistics with remarkable accuracy. Using the recently proposed random voting model, we analytically derive the scaled distributions of votes secured by winners, runner-ups, and margins of victory, and demonstrating their strong correlation with turnout distributions. By analyzing Indian election data – spanning multiple decades and electoral scales – we validate these predictions empirically across all scales, from large parliamentary constituencies to polling booths. Further, we uncover a surprising scale-invariant behavior in the distributions of scaled margins of victory, a characteristic signature of Indian elections. Finally, we demonstrate a robust universality in the distribution of the scaled margin-to-turnout ratios.

The institution of election plays a pivotal role in every functioning democracy to ensure that the governing bodies are based on the people's mandate. While the rules for choosing candidates are often simple at the microscopic level, the effects arising from complex interactions among individuals can make electoral processes and their final outcomes unpredictable on larger scales. To address these complexities, tools from statistical physics and complex systems were applied to analyze and uncover patterns in electoral outcomes [? ? ? ? ? ? ? ? ? ? ? ?].

Over the last few decades, aided by the availability of extensive election data, many earlier studies [? ? ? ? ? ?] have attempted to identify universal patterns to characterize and simplify the complexities of electoral processes irrespective of microscopic details. While the distributions of vote shares garnered by candidates [? ? ? ?] and voter turnouts[? ?] have been extensively studied, they exhibit limited universality at best [? ? ?]. Nevertheless, these distributions have proven valuable in flagging irregularities and detecting fraudulent practices in elections [? ? ?].

Among the many statistics of interest, the *margin* of victory, defined as $M = V_w - V_r$, where V_w and V_r are the votes secured by the winner and runner-up, respectively, encodes key information about the competitiveness of elections. While margins of victory have been previously studied [? ? ? ? ? ?], often independently of voter turnouts, our recent work [?] suggests that voter turnouts, in combination with margins, provide deeper insight into electoral dynamics. This extensive analysis of election data spread over many decades of elections held in 34 countries demonstrated that the scaled distribution of the margin-to-turnout ratio exhibits a universal form [?]. Significant deviations from this universal behavior

can indicate potential electoral malpractice [? ? ? ? ? ?]. Furthermore, we demonstrated that voter turnouts play a fundamental role in driving the margins and, together with a proposed random voting model (RVM), can accurately predict the scaled distribution of victory margins. This robust connection between turnouts and margins holds across multiple elections and electoral scales. These findings naturally raise the question: Can the distribution of other relevant electoral statistics be uncovered using voter turnout distributions?

To explore this question, large datasets covering a range of different electoral scales are required. India, the world's largest democracy, regularly conducts elections involving vast electorates (960 million in 2024), and its publicly available election data spans multiple decades and electoral scales. The diversity of India's linguistic and cultural landscape further adds to the complexities of its electoral outcomes, making it an excellent testing ground for assessing the robustness of the RVM framework.

In this Letter, using empirical data from Indian elections [? ?] spanning several decades and vastly different electoral scales, we demonstrate a strong correlation between the distributions of votes received by winners and runner-ups and voter turnouts. Leveraging this correlation and the RVM, we analytically predict the scaled distributions of the votes secured by winners and runner-ups using the corresponding turnout distributions. This prediction remarkably holds good at all the electoral scales – from large parliamentary constituencies ($\sim 10^6$ voters) down to the smallest polling booth levels ($\sim 10^2 - 10^3$ voters). Further, we show a rather surprising scale invariance of the margin distributions, a characteristic typical of Indian elections. Finally, a robust universality in the distribution of scaled margin-to-turnout ratio is demonstrated, strengthening our recent proposition.

We formalize our framework as follows: An *election* happens at all the N electoral units following the first-post-the-past principle [?]. Let the i-th electoral unit have n_i^c candidates and n_i^v eligible voters, where

$i = 1, 2, \ldots N$. Usually, only a fraction of the eligible voters cast their votes. This is termed the turnout $T_i \leq n_i^v$. It is a direct indicator of the people's interest in the electoral process, and its distribution $g(T)$ encodes information about electoral statistics [**?**]. Let $v_{i,1}, v_{i,2}, \ldots, v_{i,n_i^c}$ be the votes secured by n_i^c candidates such that $\sum_j v_{i,j} = T_i$. The candidate securing the highest number of votes, V_w, is declared the winner, while the candidate with the second-highest votes, V_r, is the runner-up. By definition $V_w > V_r$. Further, the *margin* of victory is defined as $M_i = V_{i,w} - V_{i,r}$, and indicates the extent of electoral competition.

The electoral units in this work have three distinct scales in terms of the size of the electorate: (i) parliamentary constituency (PC, largest scale), (ii) assembly constituency (AC, intermediate scale), and (iii) polling booth (PB, smallest scale). Table **??** describes the granularity of election data used in this work – electoral units and their typical electorate size. While for the national-level general elections, we employ data from 3 different electoral scales (PC, AC, and PB), we use AC-level data for state elections. Depending on the electoral level considered, the winner and runner-up vote distributions have vastly different scales, with the winner vote distribution having wider support than the runner-up. However, when both distributions are scaled by their respective mean values (mean taken over all elections for which data is available), the winner and runner-up vote distributions – $Q_{\widetilde{V_w}}(\widetilde{V_w})$ and $Q_{\widetilde{V_r}}(\widetilde{V_r})$ – explicitly display a strong correlation with the corresponding scaled turnout distributions $Q_{\widetilde{T}}(\widetilde{T})$. Note that any variable Y, scaled by their mean $\langle Y \rangle$, is denoted by $\widetilde{Y} = Y/\langle Y \rangle$. Figure **??** displays this key result at four electoral units described in Table **??**. Remarkably, at larger electoral scales (PC and AC), not only the tail but the entire scaled distributions of winner and runner-up votes mimic the corresponding scaled turnout distribution as seen in Fig. **??**(b-d). This strong correlation indicates that the turnout distribution contains crucial information about different election statistics and can be leveraged to predict the scaled vote distributions of the winner and the runner-up. To explore this possibility, we employ our recently proposed random voting model (RVM) [**?**], which is demonstrably effective at predicting the scaled distribution of several election statistics, such as the *margin* of victory.

The random voting model, later denoted as RVM (T, n^c), is built around the framework of elections described above and consists of N electoral units with n_i^c

TABLE I. Electoral units in various elections in India

Type	Type of electoral unit	Type of election	Size
GE-PB	Polling booth	General election	10^3
GE-AC	Assembly constituency	General election	10^5
GE-PC	Parliamentary constituency	General election	10^6
SE-AC	Assembly constituency	State election	10^5

FIG. 1. Winner, runner-up vote distributions, and turnout distributions, scaled by their respective means. Notably, at larger electoral scales (AC / PC), the winner and runner-up distributions mimic the corresponding turnout distribution.

number of candidates contesting for the votes of T_i electors who have voted in the i-th electoral unit. Then, the probability that j-th candidate attracts electors' votes is:

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

In this, $\mathcal{U}(0,1)$ is the uniform distribution. This model was shown to capture the various statistical features of empirical election data, such as the margin distributions and the universal features embedded in the election data [**?**]. In particular, the random voting model with $n^c = 3$ candidates – RVM $(T, 3)$ – predicts the scaled distribution of *margin* remarkably well, irrespective of electoral scales and countries [**?**]. In this work, we employ a refined approach based on the notion of the effective number of parties defined as [**?**]:

$$^{(E)}n_i^c = \frac{1}{\sum_{k=1}^{n_i^c} (V_{ik}/T_i)^2}. \quad (2)$$

In large elections exercise such as that in India, even though many candidates join the fray, a few corner most of the votes. For instance, if all the votes are garnered by just one candidate, then $V_{i1} = T_i$, and $V_{ij} = 0$ for $j = 2, \ldots n^c$. In this case, $^{(E)}n_i^c = 1$. However, if all the votes are split equally among two candidates, $^{(E)}n_i^c = 2$, thus, Eq. **??** captures the idea of an effective number of candidates in i-th electoral unit. Further, by averaging

FIG. 2. Winner and runner-up vote distributions scaled by their respective means. Panels (a, b), (c, d), and (e, f) depict, respectively, the scaled winner and runner-up vote distribution at the polling booth, assembly constituency, and parliamentary constituency level for Indian general elections. Panels (g, h) correspond to the distributions for the state elections at the assembly constituency level. The analytical predictions (solid lines) are in remarkable agreement with the empirical distributions (open circle). Predictions from RVM simulations (dashed line) closely follow the analytical curves.

over all the electoral units, we obtain:

$$^{(E)}\tilde{n}^c = \left[\frac{1}{N}\sum_{k=1}^{N} {}^{(E)}n_k^c\right], \qquad (3)$$

where $[\,*\,]$ denotes the operation of extracting the closest integer value. And $^{(E)}\tilde{n}^c$ indicates the effective number of candidates for an entire election at different electoral scales. From empirical data of Indian elections at different scales, we find $^{(E)}\tilde{n}^c = 2$ at the polling booth (PB-GE) level for the General Elections. However, for all the other three cases of AC-GE, PC-GE, and AC-SE, we obtain $^{(E)}\tilde{n}^c = 3$. Now, we shall solve RVM (T, n^c)

for $n^c = 2$ and $n = 3$ to analytically describe the results observed in Fig. **??**.

The primary object of interest is the distribution of the votes received by the winner and the runner-up. In the large turnout limit, $T >> 1$, the votes received by j-th candidate can be approximated as $V_j \approx p_j T$ (index i is dropped as we focus on a single electoral unit). Consequently, the vote share is defined as:

$$v_j = V_j/T \ \text{ with } j = 1, 2, \ldots n^c. \tag{4}$$

Thus, in this limit, the vote share distribution is effectively the same as the distribution of p_j. Recall that p_j can be constructed from random weights using Eq. **??**. These weights are n^c *i.i.d.* random variables $\{w_1, w_2, \cdots, w_{n^c}\}$ drawn from the uniform distribution $\mathcal{U}(0, 1)$. When arranged in ascending order, the random variable at the k-th place is defined as the k-th order statistics and is denoted by $w_{(k)}^{n^c}$ [**?**]. Specifically, $w_{(1)}^{n^c}$ and $w_{(n^c)}^{n^c}$ represent the smallest and the largest weights, respectively. Then, the joint probability distribution function (jpdf) or all the order statistics is given by:

$$\mathbb{P}\left(w_{(1)}^{n^c}, w_{(2)}^{n^c}, \ldots w_{(n^c)}^{n^c}\right) = n^c!. \tag{5}$$

Finally, the winner's vote share v_w and runner-up's vote share v_r can be expressed in terms of order statistics as:

$$v_w = \frac{w_{(n^c)}^{n^c}}{\sum_{k=1}^{n^c} w_{(k)}^{n^c}} \quad \text{and} \quad v_r = \frac{w_{(n^c-1)}^{n^c}}{\sum_{k=1}^{n^c} w_{(k)}^{n^c}}. \tag{6}$$

Their distributions can be obtained from the jpdf in Eq. **??** by integrating out the other variables (for detailed calculations, see Supplementary Material [**?**]). When the number of candidates $n^c = 2$, the distribution of the winner's vote share $P_{v_w}(v_w)$ is found to be:

$$P_{v_w}(v_w) = \begin{cases} \dfrac{1}{v_w^2}, & \text{if } \dfrac{1}{2} < v_w < 1, \\ 0, & \text{otherwise.} \end{cases} \tag{7}$$

The vote share distribution of the runner-up can also be calculated similarly and is given by:

$$P_{v_r}(v_r) = \begin{cases} \dfrac{1}{(1-v_r)^2}, & \text{if } 0 < v_r \le \dfrac{1}{2}, \\ 0, & \text{otherwise.} \end{cases} \tag{8}$$

The conditions in Eqs. **??-??** reflect the intuitive idea that when there are only two candidates, the winner's vote share cannot be less than $1/2$, and the runner-up cannot exceed $1/2$. For $n^c = 3$, the winner's vote share distribution is:

$$P_{v_w}(v_w) = \begin{cases} \dfrac{3v_w - 1}{v_w^3}, & \text{if } \dfrac{1}{3} < v_w \le \dfrac{1}{2} \\ \dfrac{1 - v_w}{v_w^3}, & \text{if } \dfrac{1}{2} < v_w < 1 \\ 0, & \text{otherwise,} \end{cases} \tag{9}$$

and the distribution of the runner-up's vote share is:

$$P_{v_r}(v_r) = \begin{cases} \dfrac{v_r(2 - 3v_r)}{(1 - v_r)^2(1 - 2v_r)^2}, & \text{if } 0 < v_r \le \dfrac{1}{3} \\ \dfrac{1 - 2v_r}{v_r^2(1 - v_r)^2}, & \text{if } \dfrac{1}{3} < v_r \le \dfrac{1}{2} \\ 0, & \text{otherwise.} \end{cases} \tag{10}$$

The vote share distribution of the winner and the runner-up are defined as piecewise functions. They are non-zero when $\frac{1}{3} < v_w < 1$ and $0 < v_r \le \frac{1}{2}$, respectively. Based on Eq. **??**, the distribution of unscaled variables $Y = (V_w, V_r)$, given T, is related to the scaled variables $y = (v_w, v_r)$ via:

$$\mathcal{P}(Y|T) = \frac{1}{T} P_y\left(\frac{Y}{T}\right), \tag{11}$$

where $P_y(y)$ is the probability density function for the scaled variable, $y = (v_w, v_r)$. The distribution of Y for arbitrary turnout distribution $g(T)$ is:

$$Q_Y(Y) = \int g(T) \, \mathcal{P}(Y|T) \, dT, \tag{12}$$

with $\langle Y \rangle = \int Y \, Q_Y(Y) \, dY$. Finally, for the scaled variable $\widetilde{Y} = Y/\langle Y \rangle$, the distribution is:

$$Q_{\widetilde{Y}}(\widetilde{Y}) = \langle Y \rangle \, Q_Y(\widetilde{Y}\langle Y \rangle). \tag{13}$$

Using the empirical turnout distribution $g(T)$ from elec-

FIG. 3. Margin distributions scaled by their respective means. (a) Data collapse in the scaled margin distributions of Indian elections at four electoral scales. (b) In contrast, such collapse is absent in the election data from the USA.

tion data, Eq. **??** is numerically integrated. The resulting distribution is then scaled using Eq. **??** to obtain the scaled distributions for the winner and runner-up vote shares, $Q_{\widetilde{V_w}}(\widetilde{V_w})$ and $Q_{\widetilde{V_r}}(\widetilde{V_r})$, respectively. As demonstrated in Fig. **??**, the analytical prediction (solid lines) is remarkably consistent with the empirical vote share distributions. The predictions from RVM simulations, which use the raw turnout data and $n^c = {}^{(E)}\tilde{n}^c$ as inputs, closely follow the analytical distributions in Fig. **??**. The scaled distributions of winner and runner-up votes depicted across all electoral scales, in Fig. **??**, typically exhibit a power-law behavior in the tails for $\widetilde{V_w}, \widetilde{V_r} \gg 1$. Conversely, for $\widetilde{V_w}, \widetilde{V_r} \ll 1$, the distributions display different profiles. Remarkably, these differences are well captured by RVM predictions: RVM $(T, 2)$ accurately predicts distribution at the GE-PB level, while RVM $(T, 3)$ closely matches the distributions at the GE-AC, GE-PC, and SE-AC levels. Hence, the effective number of candidates (Eq. **??**) and the turnout distribution $g(T)$, when used within the RVM framework, successfully predict the winner and runner-up vote share distributions across distinct electoral scales.

Next, we consider the effect of voter turnouts T on the *margin* of victory M. To do this, firstly we define *specific margin* as $\mu = M/T = (V_w - V_r)/T$. In the large turnout $(T >> 1)$ limit, the specific margin can be expressed in terms of the order statistics of w as:

$$\mu = \frac{w_{(n^c)}^{n^c} - w_{(n^c-1)}^{n^c}}{\sum_{k=1}^{n^c} w_{(k)}^{n^c}}, \tag{14}$$

where n^c is the number of candidates. Using the jpdf in Eq. **??**, for RVM $(T, 2)$ the distribution of specific margin can be obtained as (see Ref. [**?**] for more details):

$$P_\mu(\mu) = \frac{2}{(1+\mu)^2}, \tag{15}$$

and for RVM $(T, 3)$, the distribution becomes:

$$P_\mu(\mu) = \frac{(1-\mu)(5+7\mu)}{(1+\mu)^2(1+2\mu)^2}. \tag{16}$$

Using Eq. **??** - **??** and the empirical turnout distributions $g(T)$, the scaled margin distribution $Q_{\widetilde{M}}(\widetilde{M})$ can be obtained for the four different electoral scales which match the corresponding empirical distributions closely (see Ref. [**?**]). Remarkably, the scaled distributions for the margin for Indian elections at four different scales collapse onto a single curve, as shown in Fig. **??**(a). This data collapse is a direct consequence of the similarity in tail behavior in the corresponding turnout distributions (see Fig. **??**). This appears to be a characteristic of Indian elections and is not observed in most other countries. For instance, this data collapse is absent in the US elections for the empirical data at the County and Congressional district levels, as demonstrated in Fig. **??** (b).
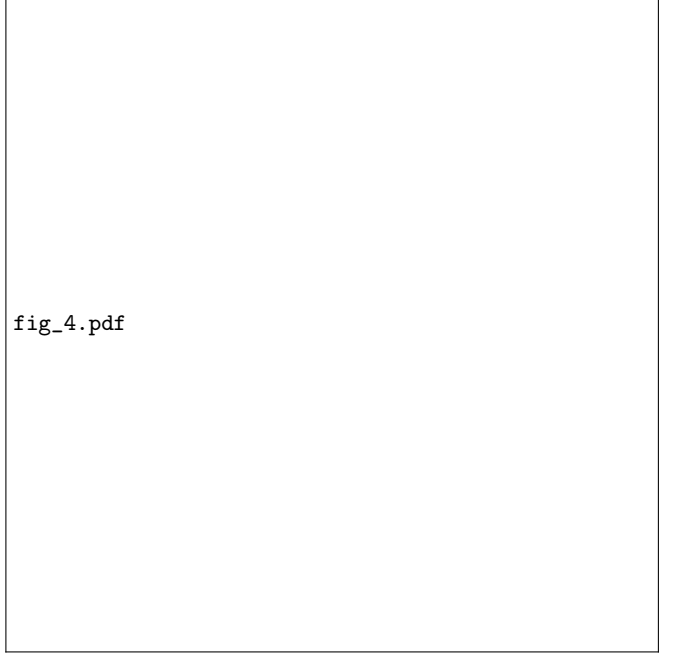

FIG. 4. Specific margin distributions scaled by their respective means at distinct electoral scales in different types of Indian elections. The scaled specific margin distributions collapse on the analytical universal curve.

Recently, we have shown that the scaled distribution of specific margins $Q_{\widetilde{\mu}}(\widetilde{\mu})$, is universal irrespective of electoral scales and countries [**?**]. This universal distribution can be analytically obtained by rescaling Eq. **??** with $\langle \mu \rangle = 1/2 + \ln\left(9\sqrt[4]{3}/16\right)$ (see Ref. [**?**]). The empirical distributions from large election datasets are proposed to exhibit a better collapse to the analytical curve, as finite-size effects are suppressed. The presence of such universality helps us distill the complexity of the electoral processes in terms of universal behavior. As the electorate size in India is large (960 million in 2024) and datasets are available at various electoral scales, Indian elections provide the best test-bed to demonstrate such universality. Fig. **??** demonstrates such universality. The distributions $Q_{\widetilde{\mu}}(\widetilde{\mu})$ at four distinct electoral scales (colored open circles) remarkably collapse onto the analytically predicted universal curve (solid line). This excellent collapse strengthens the proposition that in fairly conducted elections, the apparent deviation from universality can originate from finite data size.

In summary, elections are a source of excellent datasets for exploring collective decision-making by millions of people (interacting agents) at *distinct electoral scales*. However, most of the earlier works on elections have ignored the effects arising from differences in electoral scales. The voter turnout distribution $g(T)$ is a good indicator of the public trust and interest in the electoral process. Remarkably, we show that they also encode information about several crucial election statistics. Aided by election data from India – the largest democ-

racy in the world – we demonstrate our results for different types of elections at multiple *distinct electoral scales.* Given the empirical $g(T)$ and an effective number of candidates for each electoral scale, we analytically obtain the scaled distributions of the winner and the runner-up votes using the framework of our recently proposed random voting model. The analytical predictions and the simulations are in excellent agreement with the empirical election data. Further, we demonstrate that the random voting model is effective in predicting the scaled margin distributions of Indian elections at vastly different electoral scales. Surprisingly, the scaled margin distribution remains invariant with respect to changes in electoral scales, making it a characteristic feature of Indian elections. Finally, the scaled specific margin distributions of Indian elections show a remarkable data collapse, strengthening the recently proposed universality [**?**]. This work paves the way for a wider understanding of electoral statistics from turnout distributions.

Supplemental Material for "Voter Turnouts Govern Key Electoral Statistics"

This Supplemental Material provides further discussion and derivations which support the findings reported in the Letter, and provides details of the models and simulations used to validate the results.

CONTENTS

S1. DESCRIPTION OF RANDOM VOTING MODEL (RVM)

In the Random Voting Model, N electoral units are considered, with n_i^c number of candidates contesting to win votes from T_i voters present to cast their votes in the i-th electoral unit. Each of the n_i^c candidates is assigned a random weight w_{ij}. These weights are drawn independently from a uniform distribution between 0 and 1. The corresponding probability p_{ij} of receiving votes is calculated by normalizing these weights. Hence, we have the following,

$$w_{ij} \sim \mathcal{U}(0,1) \text{ and } p_{ij} = \frac{w_{ij}}{\sum_{k=1}^{3} w_{ik}}; \text{ with } j = 1, 2 \cdots n_i^c. \tag{S1}$$

For the rest of the analysis, we focus on a single (i-th) electoral unit with voter turnout T and drop the corresponding index i for brevity. Hence,

$$w_{ij} := w_j \text{ and } p_{ij} := p_j. \tag{S2}$$

S2. ANALYTICAL CALCULATIONS OF DIFFERENT ELECTION STATISTICS

For large turnout ($T \gg 1$), it is reasonable to assume the number of votes received by j-th candidate is proportional to their probability p_j, in particular, $V_j \approx p_j T$. Hence, for $T \gg 1$, the votes received by the winner V_w can be approximated as,

$$V_w \approx p_{max} T, \tag{S3}$$

and the votes received by the runner-up V_r as,

$$V_r \approx p_{2nd\,max} T, \tag{S4}$$

and the margin $M = V_w - V_r$ can also be written as,

$$M \approx (p_{max} - p_{2nd\,max}) T, \tag{S5}$$

where p_{max} and $p_{2nd\,max}$ correspond to the largest and the second largest probabilities assigned to the candidates. For example, if the number of candidates $n^c = 3$ and the probabilities p_1, p_2, and p_3 assigned to those 3 candidates are $0.2, 0.5$, and 0.3, then $p_{max} = p_2 = 0.5$ and $p_{2nd\,max} = p_3 = 0.3$.

A. Order statistics and its connection to winner and runner-up vote share and margins

Consider n *iid* random variables $\{X_1, X_2 \ldots X_n\}$ drawn from a distribution $\rho(x)$. When arranged in ascending order, the random variable at the k-th spot is defined as the k-th order statistics. In particular, n-th and 1-st order statistics correspond to the maximum and minimum of those n random variables, respectively. The k-th order statistics of the random variable X is denoted by $X_{(k)}$.

The joint probability density of all the order statistics of the above-mentioned n random variables, $\mathbb{P}\left(x_{(1)}, x_{(2)}, \ldots x_{(n)}\right)$, defined as the probability density that the random variable $X_{(k)}$ takes the value $x_{(k)}$ for $k \in \{1, 2, \ldots, n\}$, is

$$\mathbb{P}\left(x_{(1)}, x_{(2)}, \ldots x_{(n)}\right) = n! \prod_{k=1}^{n} \rho\left(x_{(k)}\right). \tag{S6}$$

Now as described in Eq. ?? the probabilities p_j can be expressed in terms of w_j. Hence the winner vote share $v_w = V_w/T$, runner-up vote share $v_r = V_r/T$ and the specific margin $\mu = M/T$ can be expressed in terms of w as the following,

$$v_w = \frac{V_w}{T} \approx p_{max} = \frac{w_{max}}{\sum_{k=1}^{n^c} w_k} = \frac{w_{(n^c)}}{\sum_{k=1}^{n^c} w_{(k)}} \tag{S7}$$

$$v_r = \frac{V_r}{T} \approx p_{2nd\,max} = \frac{w_{2nd\,max}}{\sum_{k=1}^{n^c} w_k} = \frac{w_{(n^c-1)}}{\sum_{k=1}^{n^c} w_{(k)}} \tag{S8}$$

$$\mu = \frac{M}{T} \approx p_{max} - p_{2nd\,max} = \frac{w_{max} - w_{2nd\,max}}{\sum_{k=1}^{n^c} w_k} = \frac{w_{(n^c)} - w_{(n^c-1)}}{\sum_{k=1}^{n^c} w_{(k)}}, \tag{S9}$$

where $w_{(k)}$ is the k-th order statistics [?].

B. Random Voting Model with two candidates

In the two-candidate Random Voting Model, we have $n = n^c = 2$ and $p(x) = \mathcal{U}(0,1)$. Hence, the joint probability distribution of all the order statistics have the following form,

$$\mathbb{P}\left(w_{(1)}, w_{(2)}\right) = 2! = 2; \quad \text{with } 0 < w_{(1)} < w_{(2)} < 1, \tag{S10}$$

and $\mathbb{P}\left(w_{(1)}, w_{(2)}\right) = 0$ otherwise, with the following normalization:

$$\int_0^1 dw_{(2)} \int_0^{w_{(2)}} 2dw_{(1)} = 1. \tag{S11}$$

1. Winner vote share distribution

From the joint probability distribution of all the order statistics (Eq. ??), the approximate vote share distribution of the winner can be obtained as,

$$
\begin{aligned}
P_{v_w}(v_w) &= 2\int_0^1 dw_{(2)} \int_0^{w_{(2)}} \delta\left(v_w - \frac{w_{(2)}}{w_{(1)} + w_{(2)}}\right) dw_{(1)}, \\
&= 2\int_0^1 \frac{2w_{(2)}}{v_w^2} \mathbb{1}_{1/2 \le v_w < 1} dw_{(2)},
\end{aligned}
\tag{S12}
$$

or,

$$
P_{v_w}(v_w) = \begin{cases} \dfrac{1}{v_w^2} & \text{if } \dfrac{1}{2} \le v_w < 1 \\[2mm] 0, & \text{otherwise.} \end{cases}
\begin{aligned} &\tag{S13} \\ &\tag{S14} \end{aligned}
$$

2. Runner-up vote share distribution

We can similarly calculate the probability density function of the runner-up vote share as the following,

$$
\begin{aligned}
P_{v_r}(v_r) &= 2\int_0^1 dw_{(2)} \int_0^{w_{(2)}} \delta\left(v_r - \frac{w_{(1)}}{w_{(1)} + w_{(2)}}\right) dw_{(1)}, \\
&= 2\int_0^1 \frac{2w_{(2)}}{(1 - v_r)^2} \mathbb{1}_{0 < v_r < 1/2} dw_{(2)},
\end{aligned}
\tag{S15}
$$

or,

$$P_{v_r}(v_r) = \begin{cases} \dfrac{1}{(1-v_r)^2} & \text{if } 0 < v_r < \dfrac{1}{2} \quad\text{(S16)} \\ 0, & \text{otherwise.} \quad\text{(S17)} \end{cases}$$

3. Specific margin distribution

Similarly, the distribution of the specific margin $\mu = M/T$ can be obtained as,

$$\begin{aligned} P_\mu(\mu) &= 2 \int_0^1 dw_{(2)} \int_0^{w_{(2)}} \delta\left(\mu - \frac{w_{(2)} - w_{(1)}}{w_{(1)} + w_{(2)}}\right) dw_{(1)}, \\ &= 2 \int_0^1 \frac{4w_{(2)}}{(1+\mu)^2} dw_{(2)}, \end{aligned} \tag{S18}$$

or,

$$P_\mu(\mu) = \begin{cases} \dfrac{2}{(1+\mu)^2} & \text{if } 0 < \mu < 1 \quad\text{(S19)} \\ 0, & \text{otherwise.} \quad\text{(S20)} \end{cases}$$

C. Random Voting Model with three candidates

In the three-candidate Random Voting Model, we have $n = n^c = 3$ and $p(x) = \mathcal{U}(0,1)$. Then, the joint probability distribution of all the order statistics is,

$$\mathbb{P}\left(w_{(1)}, w_{(2)}, w_{(3)}\right) = 3! = 6; \text{ with } 0 < w_{(1)} < w_{(2)} < w_{(3)} < 1, \tag{S21}$$

and $\mathbb{P}\left(w_{(1)}, w_{(2)}, w_{(3)}\right) = 0$ otherwise, with the following normalization:

$$\int_0^1 dw_{(3)} \int_0^{w_{(3)}} dw_{(2)} \int_0^{w_{(2)}} 6 dw_{(1)} = 1. \tag{S22}$$

1. Winner vote share distribution

From the joint probability distribution of all the order statistics, we calculate the approximate probability density function of the winner vote share $v_w = V_w/T$ as follows,

$$\begin{aligned} P_{v_w}(v_w) &= 6 \int_0^1 dw_{(3)} \int_0^{w_{(3)}} dw_{(2)} \int_0^{w_{(2)}} \delta\left(v_w - \frac{w_{(3)}}{w_{(1)} + w_{(2)} + w_{(3)}}\right) dw_{(1)}, \\ &= 6 \int_0^1 dw_{(3)} \int_0^{w_{(3)}} \frac{w_{(3)}}{v_w^2} \mathbb{1}_{0 < \frac{w_{(3)} - v_w\left(w_{(2)} + w_{(3)}\right)}{v_w} < w_{(2)}} dw_{(2)}, \end{aligned} \tag{S23}$$

or,

$$P_{v_w} = \begin{cases} 6 \displaystyle\int_0^1 w_{(3)}^2 \frac{3v_w - 1}{2v_w^3} dw_{(3)}, & \text{if } \dfrac{1}{3} < v_w \le \dfrac{1}{2} \quad\text{(S24)} \\ 6 \displaystyle\int_0^1 w_{(3)}^2 \frac{1 - v_w}{2v_w^3} dw_{(3)}, & \text{if } \dfrac{1}{2} < v_w \le 1 \quad\text{(S25)} \\ 0, & \text{otherwise.} \quad\text{(S26)} \end{cases}$$

Finally, after performing the integral, we get

$$
P_{v_w} = \begin{cases} \dfrac{3v_w - 1}{v_w^3} & \text{if } \dfrac{1}{3} < v_w \leq \dfrac{1}{2} & \text{(S27)} \\[3mm] \dfrac{1 - v_w}{v_w^3}, & \text{if } \dfrac{1}{2} < v_w < 1 & \text{(S28)} \\[3mm] 0, & \text{otherwise.} & \text{(S29)} \end{cases}
$$

2. Runner-up vote share distribution

Similarly, the probability density function of the runner-up vote share $v_r = V_r/T$ can be obtained as follows,

$$
\begin{aligned}
P_{v_r}(v_w) &= 6 \int_0^1 dw_{(3)} \int_0^{w_{(3)}} dw_{(2)} \int_0^{w_{(2)}} \delta\left(v_r - \frac{w_{(2)}}{w_{(1)} + w_{(2)} + w_{(3)}} \right) dw_{(1)}, \\
&= 6 \int_0^1 dw_{(3)} \int_0^{w_{(3)}} \frac{w_{(2)}}{v_r^2} \mathbb{1}_{0 < (1/v_r - 1)w_{(2)} - w_{(3)} < w_{(2)}} dw_{(2)},
\end{aligned}
\tag{S30}
$$

or,

$$
P_{v_r}(v_r) = \begin{cases} 6 \displaystyle\int_0^1 w_{(3)}^2 \dfrac{v_r(2 - 3v_r)}{2(1 - v_r)^2(1 - 2v_r)^2} dw_{(3)}, & \text{if } 0 < v_r \leq \dfrac{1}{3} & \text{(S31)} \\[3mm] 6 \displaystyle\int_0^1 w_{(3)}^2 \dfrac{1 - 2v_r}{2v_r^2(1 - v_r)^2} dw_{(3)}, & \text{if } \dfrac{1}{3} < v_r < \dfrac{1}{2} & \text{(S32)} \\[3mm] 0, & \text{otherwise.} & \text{(S33)} \end{cases}
$$

Finally, after performing the integral, we get

$$
P_{v_r}(v_r) = \begin{cases} \dfrac{v_r(2 - 3v_r)}{(1 - v_r)^2(1 - 2v_r)^2} & \text{if } 0 < v_r \leq \dfrac{1}{3} & \text{(S34)} \\[3mm] \dfrac{1 - 2v_r}{v_r^2(1 - v_r)^2}, & \text{if } \dfrac{1}{3} < v_r < \dfrac{1}{3} & \text{(S35)} \\[3mm] 0, & \text{otherwise.} & \text{(S36)} \end{cases}
$$

3. Specific margin distribution

We obtain the distribution of specific margin $\mu = M/T$ as follows,

$$
\begin{aligned}
P_\mu(\mu) &= 6 \int_0^1 dw_{(3)} \int_0^{w_{(3)}} dw_{(2)} \int_0^{w_{(2)}} \delta\left(\mu - \frac{w_{(3)} - w_{(2)}}{w_{(1)} + w_{(2)} + w_{(3)}} \right) dw_{(1)}, \\
&= 6 \int_0^1 dw_{(3)} \int_0^{w_{(3)}} \frac{w_{(3)} - w_{(2)}}{\mu^2} \mathbb{1}_{0 < \frac{w_{(3)} - w_{(2)} - \mu\left(w_{(3)} + w_{(2)}\right)}{\mu} < w_{(2)}} dw_{(2)},
\end{aligned}
\tag{S37}
$$

or,

$$
P_\mu(\mu) = \begin{cases} 6 \displaystyle\int_0^1 w_{(3)}^2 \dfrac{5 + 2\mu - 7\mu^2}{2(1 + \mu)^2(1 + 2\mu)^2} dw_{(3)}, & \text{if } 0 < \mu < 1 & \text{(S38)} \\[3mm] 0, & \text{otherwise.} & \text{(S39)} \end{cases}
$$

Finally, after performing the integral, we get

$$
P_\mu(\mu) = \begin{cases} \dfrac{(1 - \mu)(5 + 7\mu)}{(1 + \mu)^2(1 + 2\mu)^2} & \text{if } 0 < \mu < 1 & \text{(S40)} \\[3mm] 0, & \text{otherwise.} & \text{(S41)} \end{cases}
$$

D. Calculating the *scaled* distributions

The winner and runner-up vote shares and specific margins are random variables scaled by the voter turnout T. However, through a simple change of variable, $Y = yT$, we can obtain the conditional distributions of the unscaled variables as,

$$\mathcal{P}(Y|T) = \frac{1}{T}P_y(Y/T), \tag{S42}$$

where y can be v_w, v_r, and μ and Y represents unscaled variables V_w, V_r, and M respectively. The distribution of Y for an arbitrary turnout distribution $g(T)$ can be obtained as,

$$Q_Y(Y) = \int g(T)\ \mathcal{P}(Y|T)dT, \tag{S43}$$

with $\langle Y \rangle$ defined as,

$$\langle Y \rangle = \int Y\ Q_Y(Y)dY. \tag{S44}$$

Finally the distribution of *scaled* Y, defined as $\widetilde{Y} = Y/\langle Y \rangle$, can be obtained as follows,

$$Q_{\widetilde{Y}}(\widetilde{Y}) = \langle Y \rangle\ Q_Y(\widetilde{Y}\langle Y \rangle) \tag{S45}$$

Again, the dummy random variable Y can be either, V_w, V_r, and M.

E. The universal distribution of scaled specific margins

The distribution of scaled specific margins ($\widetilde{\mu} = \mu/\langle\mu\rangle$) is universal and can be obtained through a simple scaling of the distribution of specific margins from 3-candidate RVM by its mean. The distribution $Q_{\widetilde{\mu}}(\widetilde{\mu})$ has the following form:

$$Q_{\widetilde{\mu}}(\widetilde{\mu}) = \langle\mu\rangle\ Q_\mu(\widetilde{\mu}\langle\mu\rangle) = \frac{\langle\mu\rangle(1 - \widetilde{\mu}\langle\mu\rangle)(5 + 7\ \widetilde{\mu}\langle\mu\rangle)}{(1 + \widetilde{\mu}\langle\mu\rangle)^2(1 + 2\ \widetilde{\mu}\langle\mu\rangle)^2}, \tag{S46}$$

with $\langle\mu\rangle = \frac{1}{2} + \ln\left(\frac{9\sqrt[4]{3}}{16}\right)$.

S3. SIMULATION DETAILS

The Random Voting Model, $\mathcal{V}(T, n^c)$, with only voter turnouts and number of candidates as an input, can predict the distributions of the winner, runner-up votes, and the margins, when scaled appropriately. For the purpose of this model, the length of the voter turnout array can be assumed to be the total number of electoral units. The model can be simulated by drawing n^c random numbers from a uniform distribution $\mathcal{U}(0,1)$ for each electoral unit. These random numbers are further normalized using Eq. ?? to find the probabilities for attracting votes for each candidate. Next, each of the T_i (voter turnout in ith electoral unit) electors cast their votes according to the previously calculated probabilities. Finally, all the votes in that unit are counted, and votes received by the winner and the runner-up, along with the margin of victory, are stored. This is repeated for all the electoral units to obtain arrays of the winner votes, runner-up votes, and margins.

S4. PREDICTION OF SCALED MARGIN DISTRIBUTION AT DIFFERENT SCALES
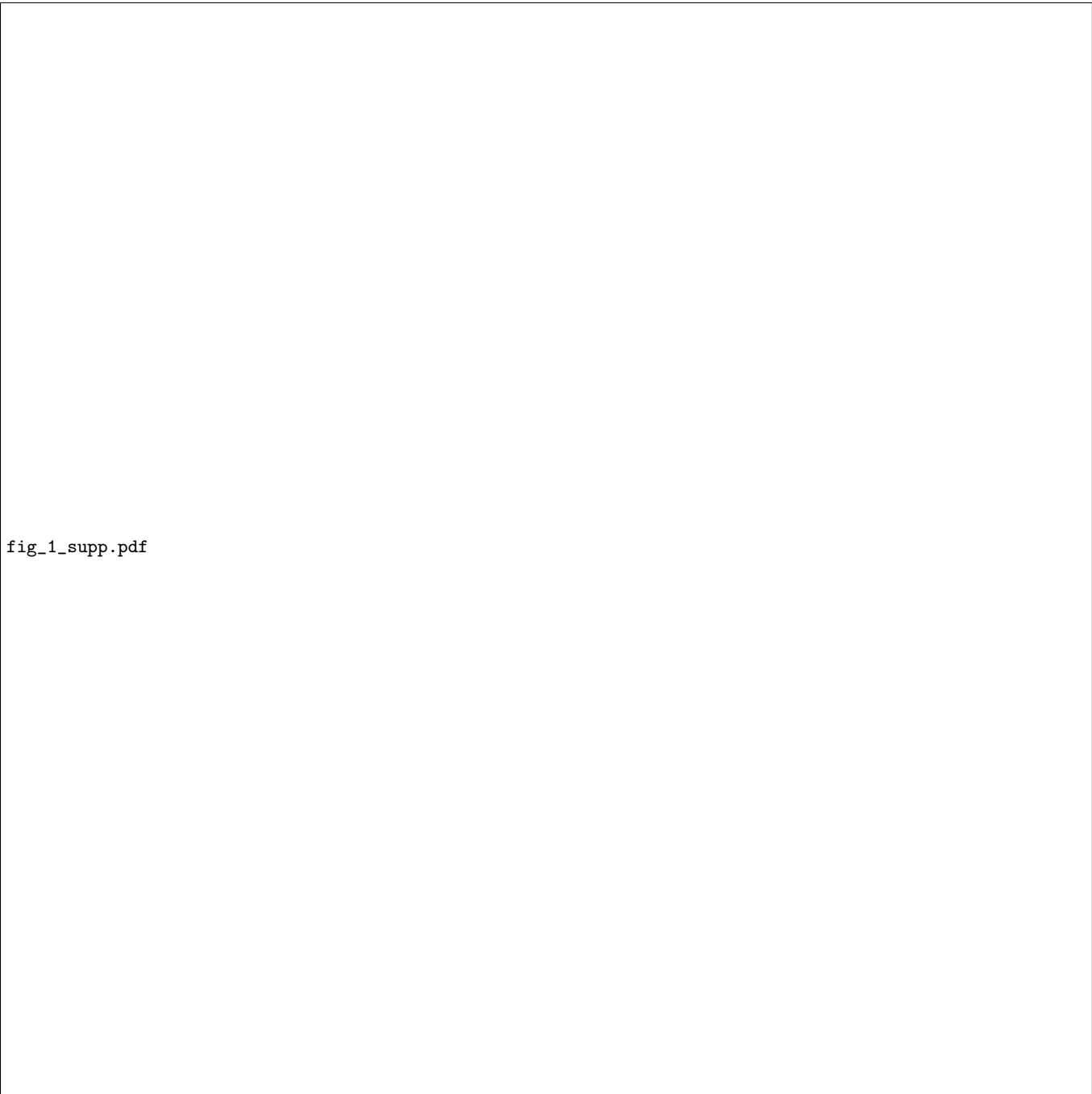
FIG. S1. Scaled distribution of margins at different electoral scales. Panel (a) - (c) demonstrates the excellent prediction of scaled margin distribution for Indian General Elections at the polling booth, assembly constituency, and parliamentary constituency levels, respectively. Panel (d) demonstrates the same for state Assembly Elections at the assembly constituency level. The open circles denote the empirical distributions. The dashed and solid lines denote the prediction from RVM simulations and analytical calculations, respectively.

S5. DATA SUMMARY

The parliamentary and assembly constituency level data of Indian General elections, along with the assembly constituency level data of the Assembly election, were obtained from the election data repository of Lok Dhaba [**? **]. The polling booth level data for the general elections were collected from the websites of chief electoral officers of different states in India [**? **]. The following table contains the summary of Indian election data.

Election Type	General Election	General Election	General Election	State Election
Electoral Scale	Parliamentary Constituency	Assembly Constituency	Polling Booth	Assembly Constituency
Time Span	1962-2019	1999-2019	2004-2019	1961-2023
Number of Elections	52 (including bye-elections)	5	4	61
Average Turnout	587329	116577	583	86484
Average Winner Votes	286807	56874	348	39884
Average Runner-up Votes	201281	38887	159	28562
Average Margin	85526	17987	189	11322

TABLE S1. The table presents typical values of voter turnout and winner votes, runner-up votes, and winning margins across different electoral levels for various types of Indian elections. The available data for the mentioned time spans were consolidated for each country and used to calculate the respective averages. The data for an electoral unit is considered valid if it meets the following criteria: (a) a complete list of votes received by all candidates was available, (b) at least two candidates contested the election, and (c) the turnout was non-zero.