

# PREDICTIVE ELECTION MODELS

Data Analysis for Journalism and Political Communication  
(Spring 2026)

Prof. Bell

# DART-THROWING CHIMPANZEES



# EXPERT POLITICAL JUDGMENT



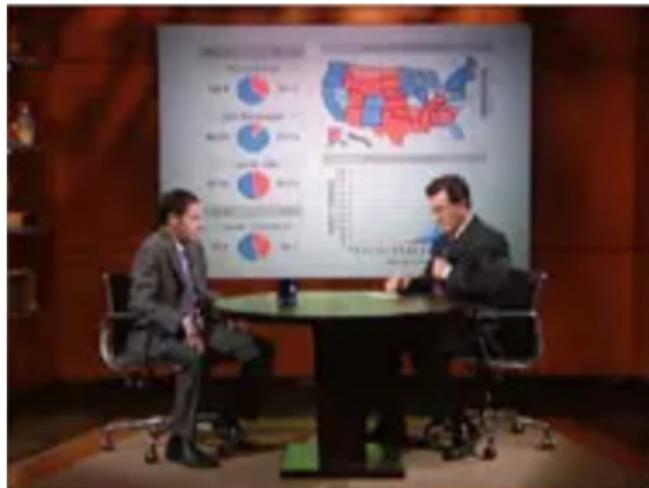
VS.



# WISDOM OF THE CROWDS



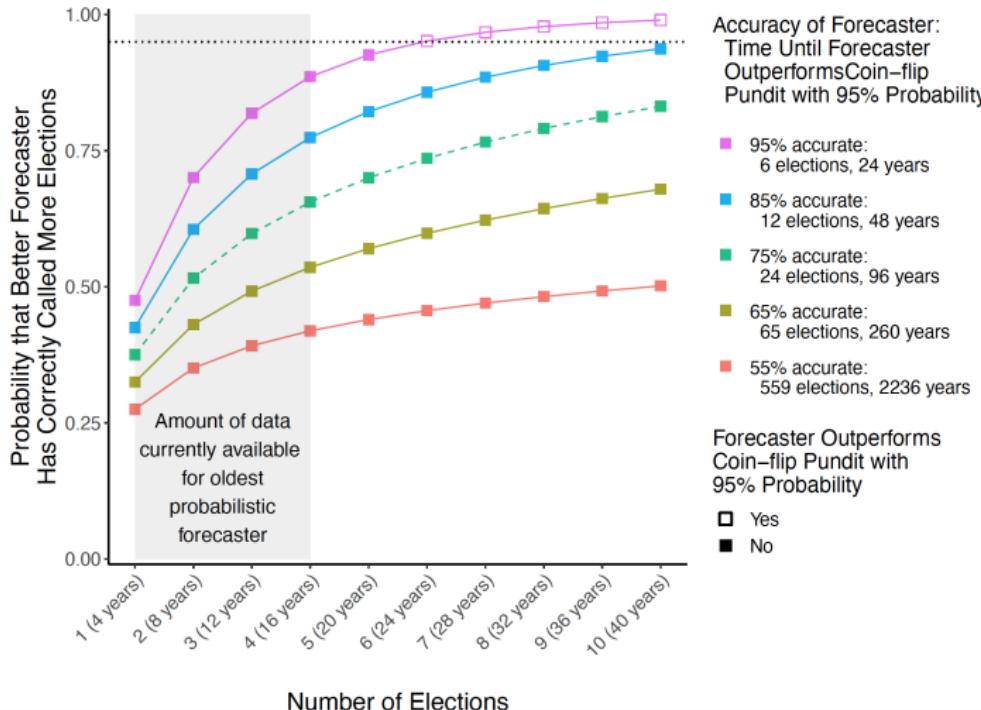
# WISDOM OF THE CROWDS



# WISDOM OF THE CROWDS

Polling aggregation	Clinton	Trump	Clinton	Trump	Clinton	Trump
	Probabilities		Electoral college		Vote share	
FiveThirtyEight	70.3	29.6	299	238	48.6	45.1
The Upshot	84.0	16.0	322	216		
RCP average of polls			301	237	47.2	44.3
The Daily Kos	88.0	12.0	313	225		
Princeton EC	99.0	1.0	312	226	51.3	48.8
HuffPost	98.1	1.6	323	215		
PollyVote			323	215	52.6	47.4
<b>Mean</b>	<b>87.9</b>	<b>12.0</b>	<b>313.3</b>	<b>224.6</b>	<b>49.0</b>	<b>46.1</b>

# WISDOM OF THE CROWDS



Source: Grimmer, Justin, Dean Knox, and Sean Westwood. 2024. "Assessing the Reliability of Probabilistic US Presidential Election Forecasts May Take Decades." OSF Preprints. August 26. doi:10.31219/osf.io/6g5zq.

# WISDOM OF THE CROWDS

Forecast	Incorrect calls out of 525
FiveThirtyEight	<b>11</b>
Split Ticket	<b>12</b>
Decision Desk HQ	<b>12</b>
Sabato's Crystal Ball	<b>12</b>
Inside Elections	<b>13</b>
JHK Forecasts	<b>14</b>
Race to the WH	<b>14</b>
The Economist	<b>17</b>
Elections Daily	<b>17</b>
Cook Political Report	<b>17.5</b>
CNalysis	<b>18</b>

\*525 = 435 House seats + 34 Senate races +

56 states/districts voting for Pres

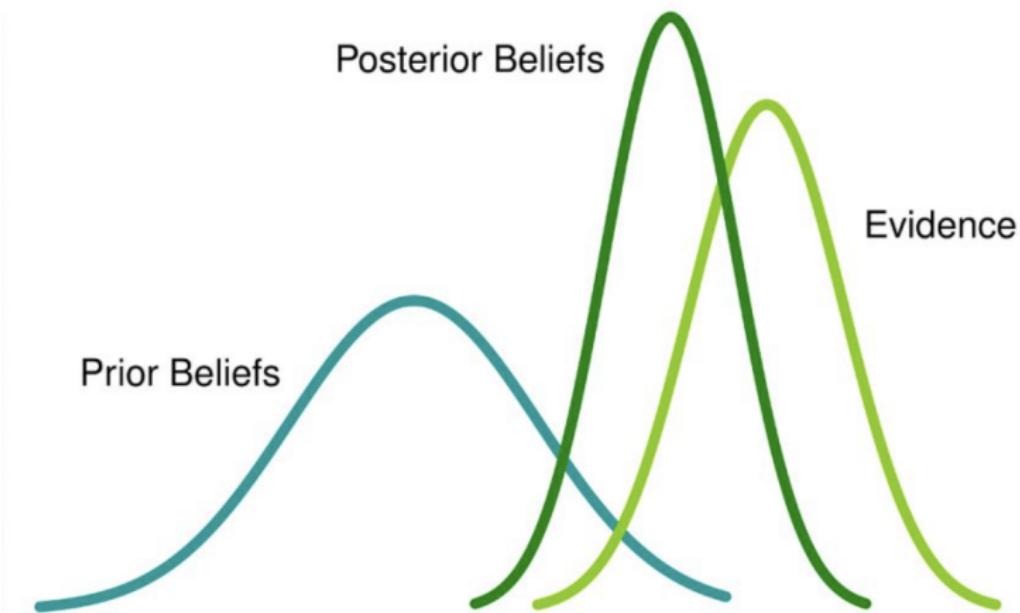
\*\* Toss Ups are counted as 0.5 incorrect

Forecast	Score
Sabato's Crystal Ball	<b>80.0</b>
Cook Political Report	<b>82.0</b>
Inside Elections	<b>87.0</b>
Split Ticket	<b>89.0</b>
FiveThirtyEight	<b>90.5</b>
Elections Daily	<b>93.0</b>
Decision Desk HQ	<b>93.5</b>
The Economist	<b>117.5</b>
JHK Forecasts	<b>118.5</b>
Race to the WH	<b>121.0</b>
CNalysis	<b>124.5</b>

Lower scores indicate higher accuracy

Source: Ethan Chen (@ECaliberSeven)

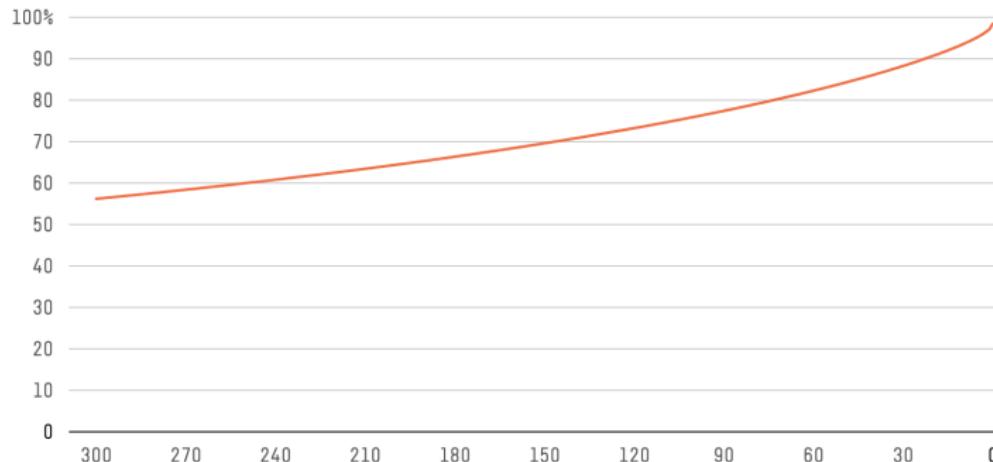
# How PREDICTION MODELS WORK



# How PREDICTION MODELS WORK

## We put more weight on the polls as Election Day approaches

Estimated\* share of the overall 538 presidential forecast that is based on polls (as opposed to non-polling historical "fundamentals"), by day before the election



\*As of Aug. 23, this estimate uses a standard deviation of 9 percentage points in 538's fundamentals model and a daily standard deviation of 0.35 points in polls, plus overall uncertainty of 1 point about the polling average. Real values will depend on how many polls we have and can differ from this estimate by a few points.

538

# How PREDICTION MODELS WORK

Election prediction models must decide:

- What polls to include (and how much to weight them): quality, quantity, sample size, time, etc.

# How PREDICTION MODELS WORK

Election prediction models must decide:

- What polls to include (and how much to weight them): quality, quantity, sample size, time, etc.
- How to adjust polls for house effects and/or mode effects

# How PREDICTION MODELS WORK

Election prediction models must decide:

- What polls to include (and how much to weight them): quality, quantity, sample size, time, etc.
- How to adjust polls for house effects and/or mode effects
- How to quantify the uncertainty in poll results

# How PREDICTION MODELS WORK

$\pi_{\text{support}} = N(\mu_s, \sigma)$  WHERE  $\mu_s \sim N(\mu_{s-1}, \tau_s)$  &  $\tau_s = N(t_s^{\beta}, \delta_{\text{true}})$   
or  $\tau_s = N(0, \delta_{\text{true}})$

AND "X"

1. Poll obs:  $\sim N(\mu_p, \sigma_p)$  (pollster's model) where  $\mu_p = \frac{\text{population} + \text{third party}}{\text{population} + \text{third party}}$ ,  $\sigma_p = \sqrt{\delta_{\text{true}}}$

2. Regressions:  $X \sim N(0, 1) \sim \text{obs. sig} + \delta_{\text{true}} \sim N(0, 1)$

Obs. sig:  $\sim N(0, 1)$  where  $\mu_{\text{sig}} = \frac{\text{population}}{\text{population} + \text{third party}}$ ,  $\sigma_{\text{sig}} = \sqrt{\delta_{\text{true}}}$

3.  $\mu_s = N(\mu_s, \sigma_s)$  [one party, one geo]

4.  $\mu_p = N(\mu_p, \sigma_p)$  [two parties, two geo]

And with the following trend:  
Expansion for states  $S_1, S_2, \dots, S_n$  and parties  $P_1, P_2, \dots, P_m$

$M[\mu_i, \mu_j, \sigma_i, \sigma_j] \sim MN(\mu_i, \mu_j, \sigma_i^2, \sigma_j^2, \Omega^2)$   $\Omega^2 = \begin{bmatrix} P_1 & P_2 & P_3 \\ P_2 & P_1 & P_3 \\ P_3 & P_3 & P_1 \end{bmatrix}$

&  $\Omega^2 = \begin{bmatrix} P_1 & P_2 & P_3 \\ P_2 & P_1 & P_3 \\ P_3 & P_3 & P_1 \end{bmatrix}$

For polling model:  $\Omega^2$  as a function of  $\Omega^2$  and random  $S_1, S_2, \dots, S_n$  with  $\Omega^2 \sim \text{Wish}(S)$

# How PREDICTION MODELS WORK

Election prediction models must decide:

- What polls to include (and how much to weight them): quality, quantity, sample size, time, etc.
- How to adjust polls for house effects and/or mode effects
- How to quantify the uncertainty in poll results
- How to model election outcomes (e.g, intra-state correlation)

# How PREDICTION MODELS WORK

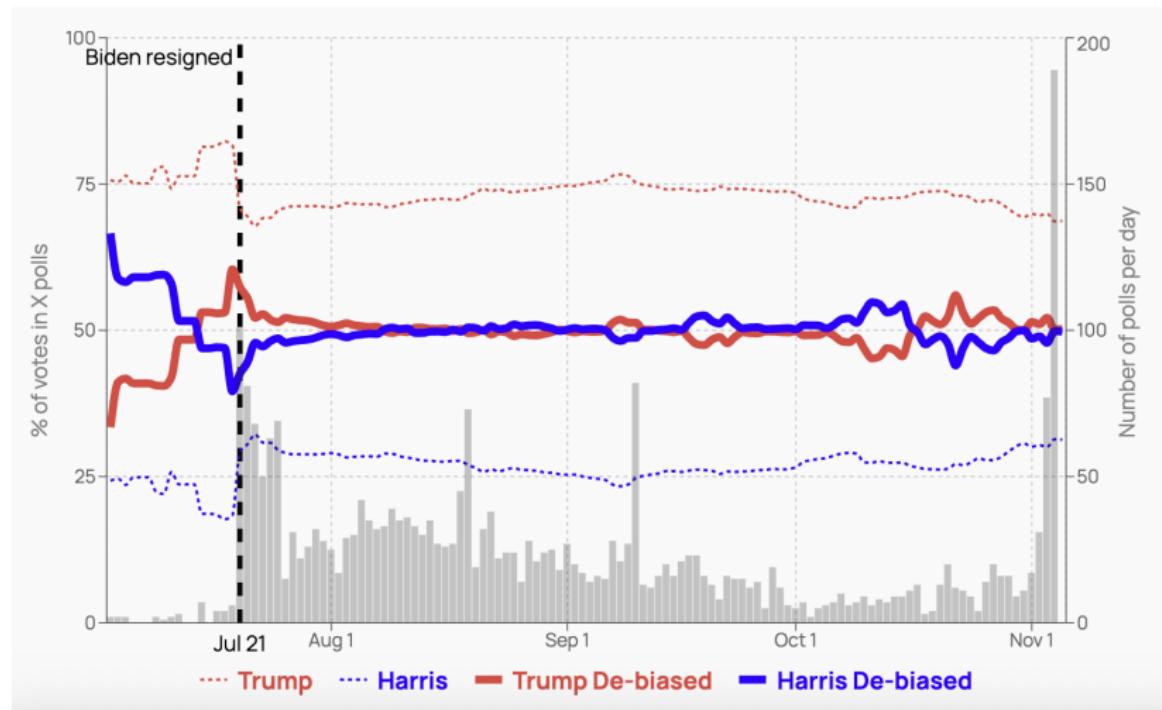
Election prediction models must decide:

- What polls to include (and how much to weight them): quality, quantity, sample size, time, etc.
- How to adjust polls for house effects and/or mode effects
- How to quantify the uncertainty in poll results
- How to model election outcomes (e.g, intra-state correlation)
- How to communicate probabilities

# NEXT FRONTIERS

- New sources of data (e.g., social media behavior)

# NEXT FRONTIERS



# NEXT FRONTIERS

- New sources of data (e.g., social media behavior)
- AI digital personas

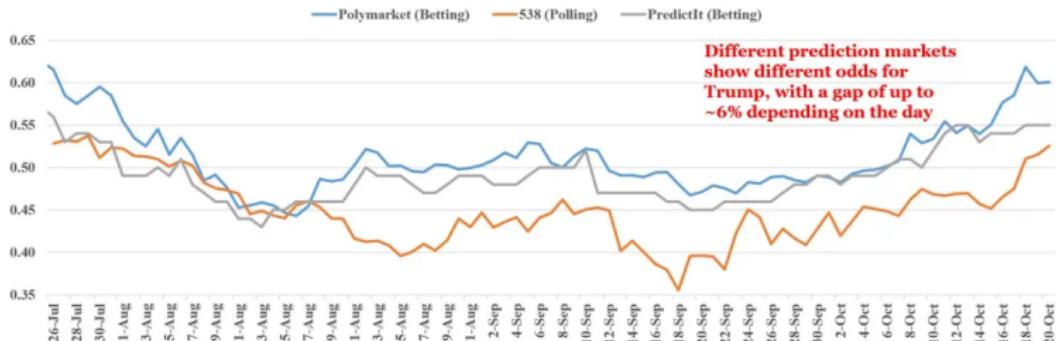
# NEXT FRONTIERS

- New sources of data (e.g., social media behavior)
- AI digital personas
- Prediction markets



## Prediction Markets Tell A Different Story From The Polls – And A Different Story From Each Other

Odds For Donald Trump Winning The 2024 Presidential Election Across Polymarket, PredictIt, and 538  
(from July 26, 2024, through October 20, 2024)



Sources: Yale Chief Executive Leadership Institute, 538, Polymarket, PredictIt



# NEXT FRONTIERS

- New sources of data (e.g., social media behavior)
- AI digital personas
- Prediction markets
- Do complex models outperform the fundamentals?