

# PREDICTIVE ELECTION MODELS

Data Analysis for Journalism and Political Communication  
(Spring 2025)

Prof. Bell

# DART-THROWING CHIMPANZEES



# EXPERT POLITICAL JUDGMENT



VS.



# WISDOM OF THE CROWDS



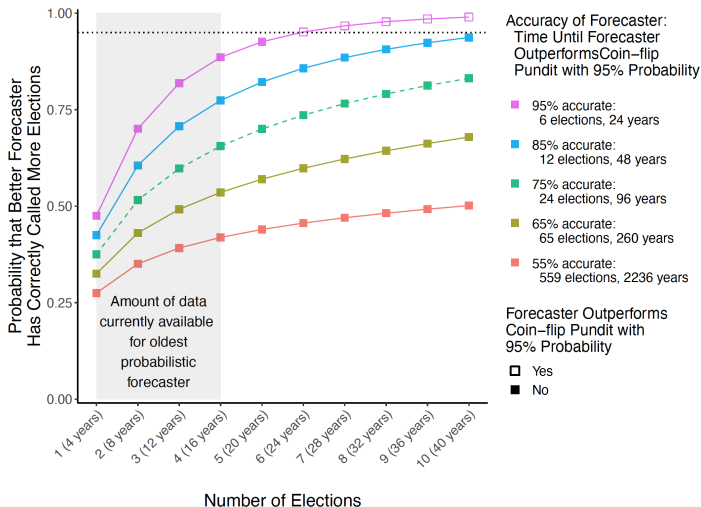
# WISDOM OF THE CROWDS



# 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	11
Split Ticket	12
Decision Desk HQ	12
Sabato's Crystal Ball	12
Inside Elections	13
JHK Forecasts	14
Race to the WH	14
The Economist	17
Elections Daily	17
Cook Political Report	17.5
CNalysis	18

\*525 = 435 House seats + 34 Senate races + 56 states/districts voting for Pres

\*\* Toss Ups are counted as 0.5 incorrect

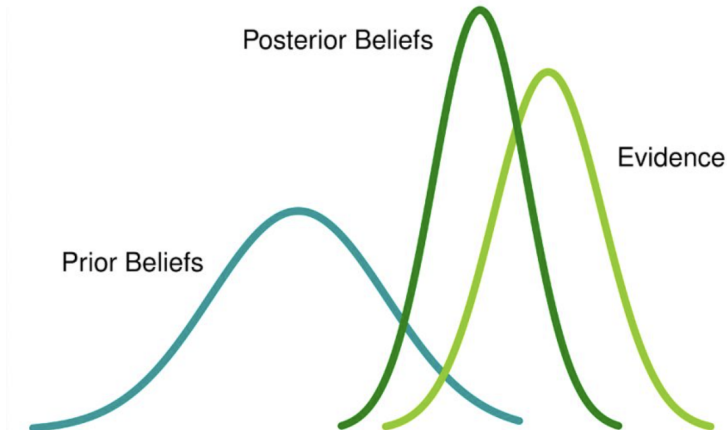
Forecast	Score
Sabato's Crystal Ball	80.0
Cook Political Report	82.0
Inside Elections	87.0
Split Ticket	89.0
FiveThirtyEight	90.5
Elections Daily	93.0
Decision Desk HQ	93.5
The Economist	117.5
JHK Forecasts	118.5
Race to the WH	121.0
CNalysis	124.5

Lower scores indicate higher accuracy

Source: Ethan Chen (@ECaliberSeven)



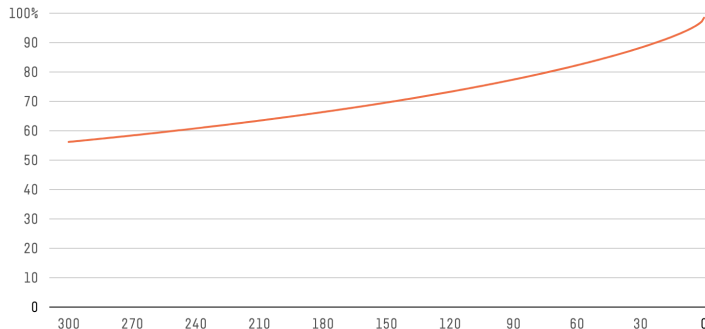
# HOW PREDICTION MODELS WORK



# 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

$$I_{0, Support_t} \sim N(\mu_t, \sigma)$$
 WHERE  $\mu_t \sim N(\mu_{t-1}, T_t)$  &  $T_t = N(t^{\frac{1}{\alpha}}, (t+1))$   
 or  $I_t \sim N(0, \sigma_{true})$

AND "X..."

$$Full\ obs_i \sim N(\mu_i + (pollster_i + media_i + third\ party_i + bullet_i), obs\_sigma + \sigma_{US})$$

$$X_{...} \sim N(0, 1) \times \sigma_{US} \text{ & } \sigma_{US} \sim N(0, 1)$$

$$obs\_sigma \sim \frac{1}{\sqrt{2}}$$

$$obs \sim N(0, 1), [0, 2]$$

one party, one geo  
 2 parties  
 20 years

And with the following trend  
 expansion for states  $S \in \{1 \dots S\}$   
 and parties  $P \in \{1 \dots P\}$ :

$$M[P, P, S, S] \sim MN(\mu_{P, t+1, S}, \Sigma_P)$$

&  $\mu_P: \text{a } S \times S \text{ for } P \in \{1 \dots P\}$

$$\mu_{P, t+1, S} \sim MN(\mu_{P, t+1, S}, \Sigma_P)$$

For pulling together  $\Sigma_P$  as a prior of  $\Sigma_P$  as a prior of  $\Sigma_P$   
 $\Sigma_P$  and random  $\Sigma_P$  with  $\Sigma_P \sim \text{diag}(S)$

$$\begin{matrix} P1/P2/P3 \\ 1/0.5/-0.1 & P1 \\ 0.5/1/-0.1 & P2 \\ -0.1/-0.1/1 & P3 \end{matrix}$$

$$S: 1 \dots S$$

# 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:

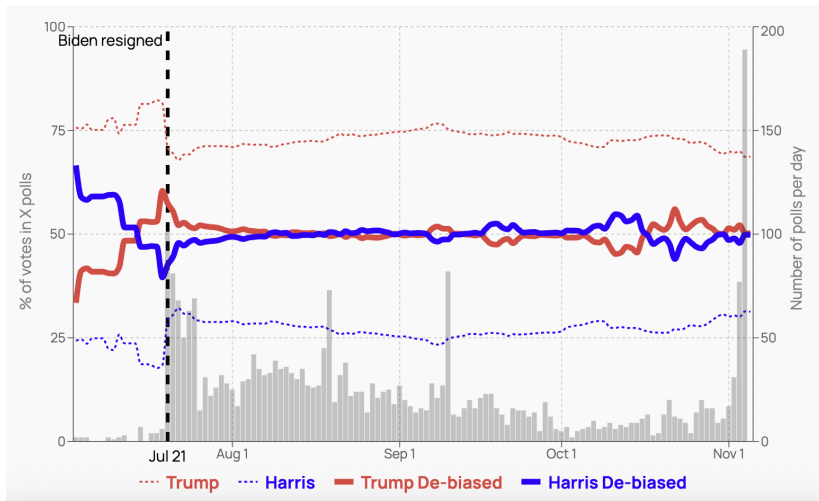
- What polls to include (and how much to weight them): quality, quantity, sample size, time, etc.
- How to adjust polls for house effects and/or mode effects
- How to quantify the uncertainty in poll results
- How to model election outcomes (e.g, intra-state correlation)
- How to communicate probabilities



# 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

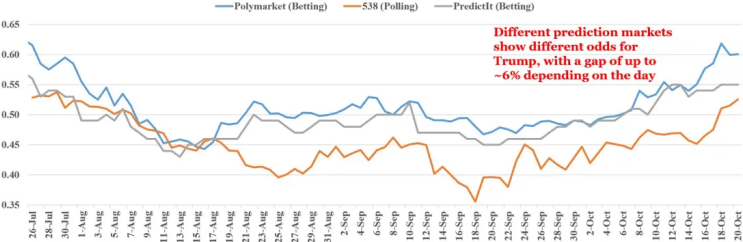
# NEXT FRONTIERS

- 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?