ELECTION METHODOLOGIES
FML PROJECT 2016

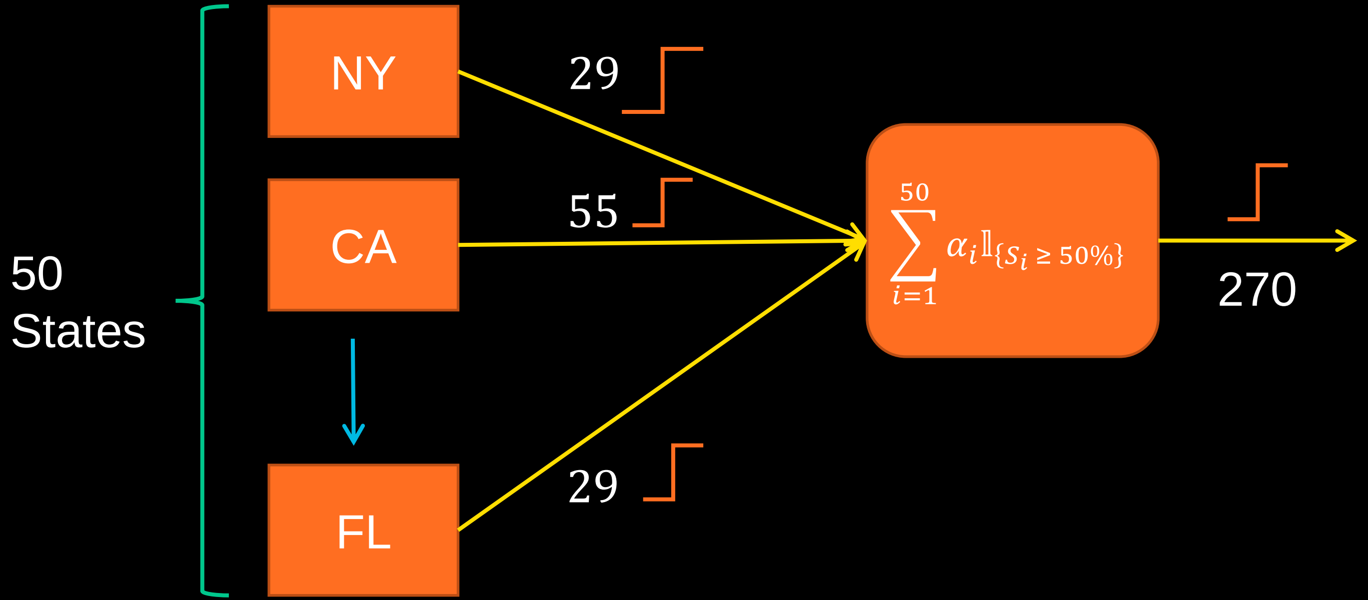
Dhruv Madeka

TOPICS

- Modeling the Election
- Forecaster Evaluation – Ex-Post
- Forecaster Evaluation – Ex-Ante
- Online Learning of Forecasters

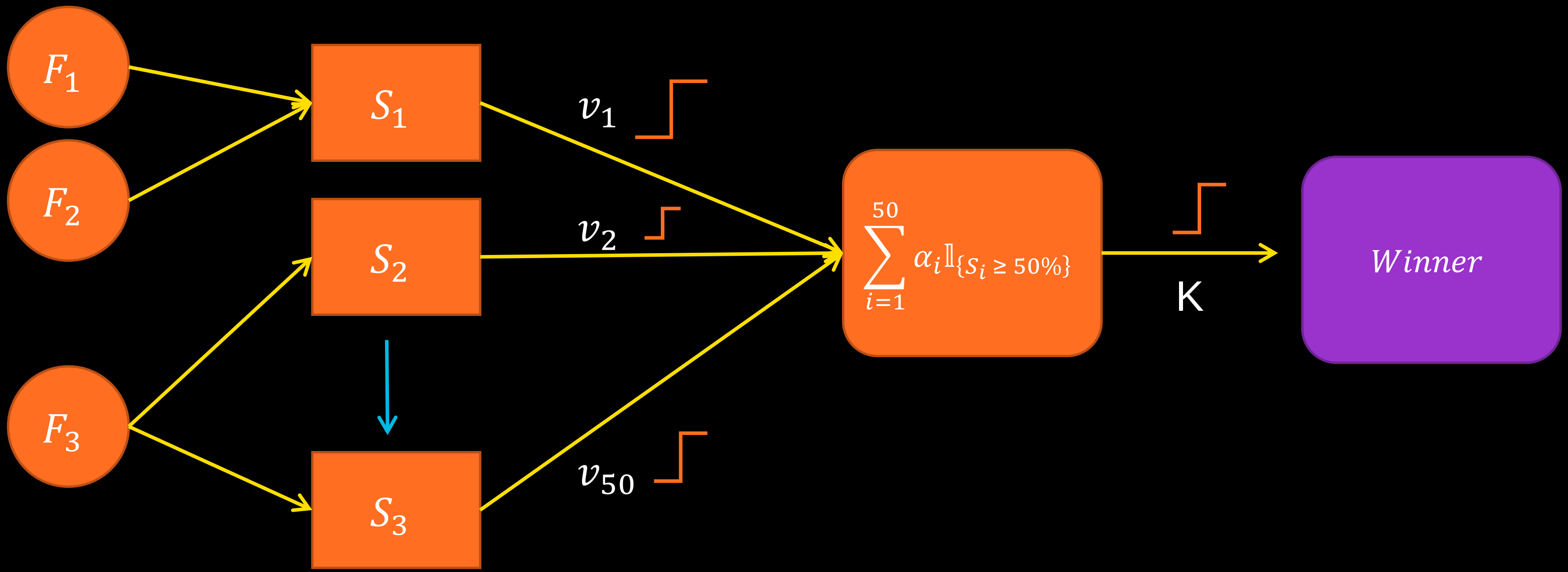
PRESIDENTIAL MECHANICS

In the US, the Presidential Election isn't decided by the popular vote but rather by the total number of Electoral College Votes.



PRESIDENTIAL MECHANICS

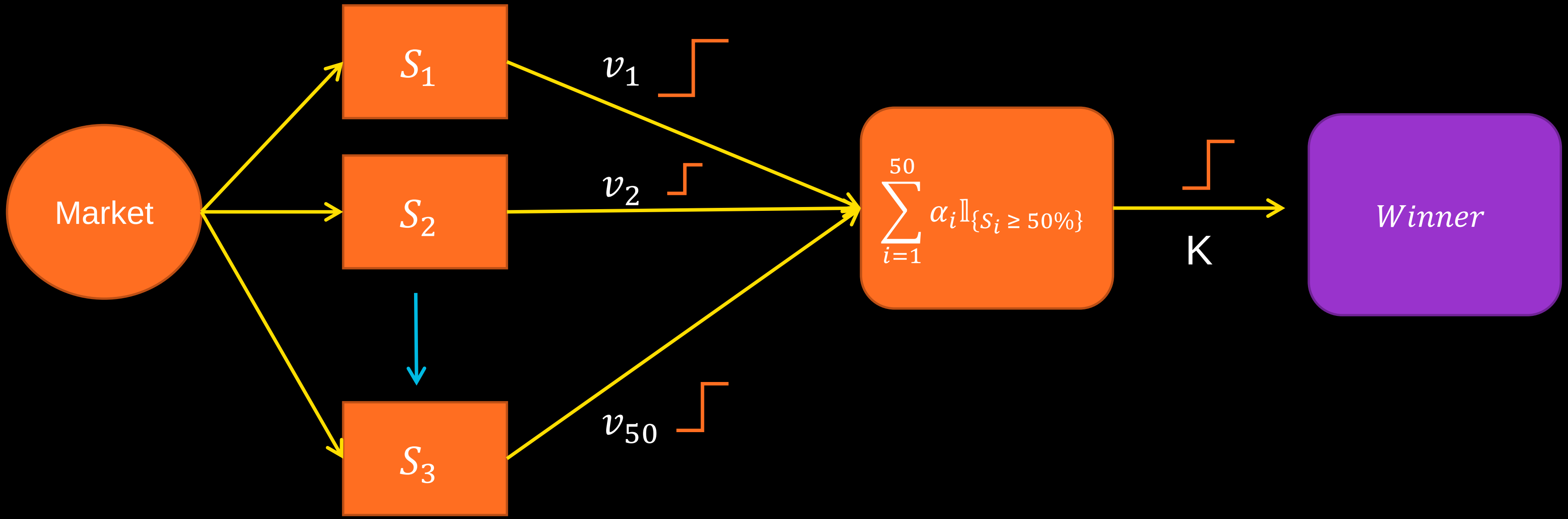
In the US, the Presidential Election isn't decided by the popular vote but rather by the total number of Electoral College Votes.



CAPM MODEL

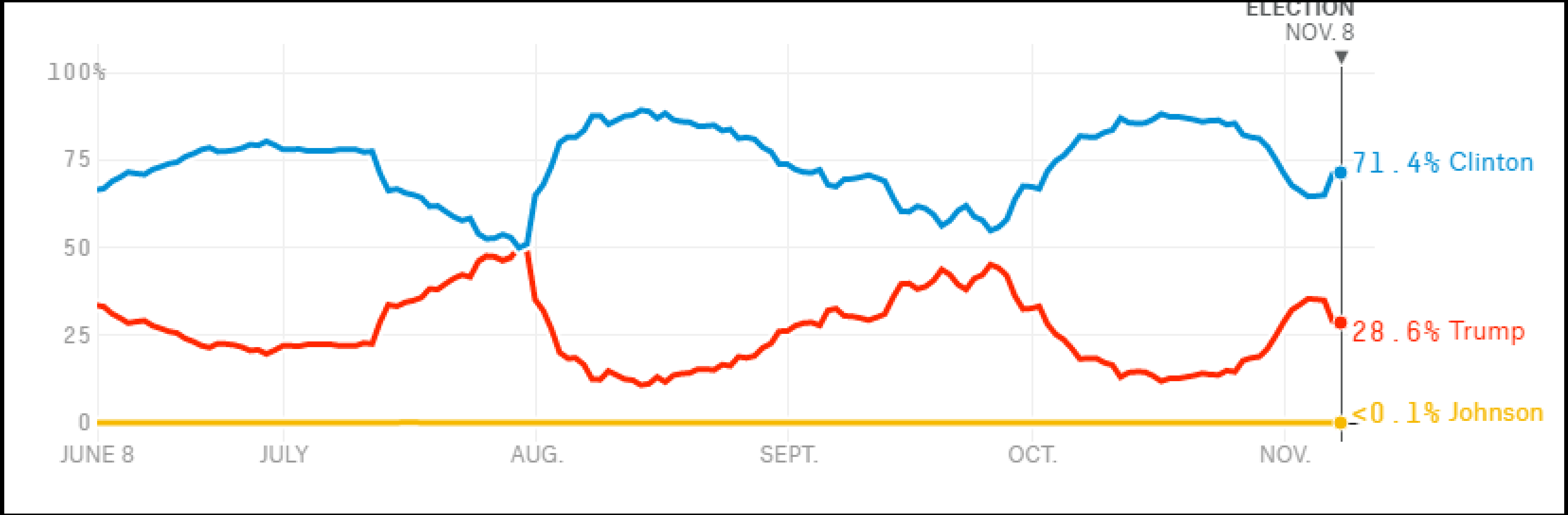
PRESIDENTIAL MECHANICS

In the US, the Presidential Election isn't decided by the popular vote but rather by the total number of Electoral College Votes.



MODELLING ISSUES

- Not enough noise! If you truly believe probabilities move this much (uncertainty is very high) then the reported probability should be 50%



CAPM MODEL

- Each state is a stock, the popular vote is the Market
- Assume each state evolves like:

$$S_t^i = \alpha^i + \beta^i M_t + \epsilon^i$$

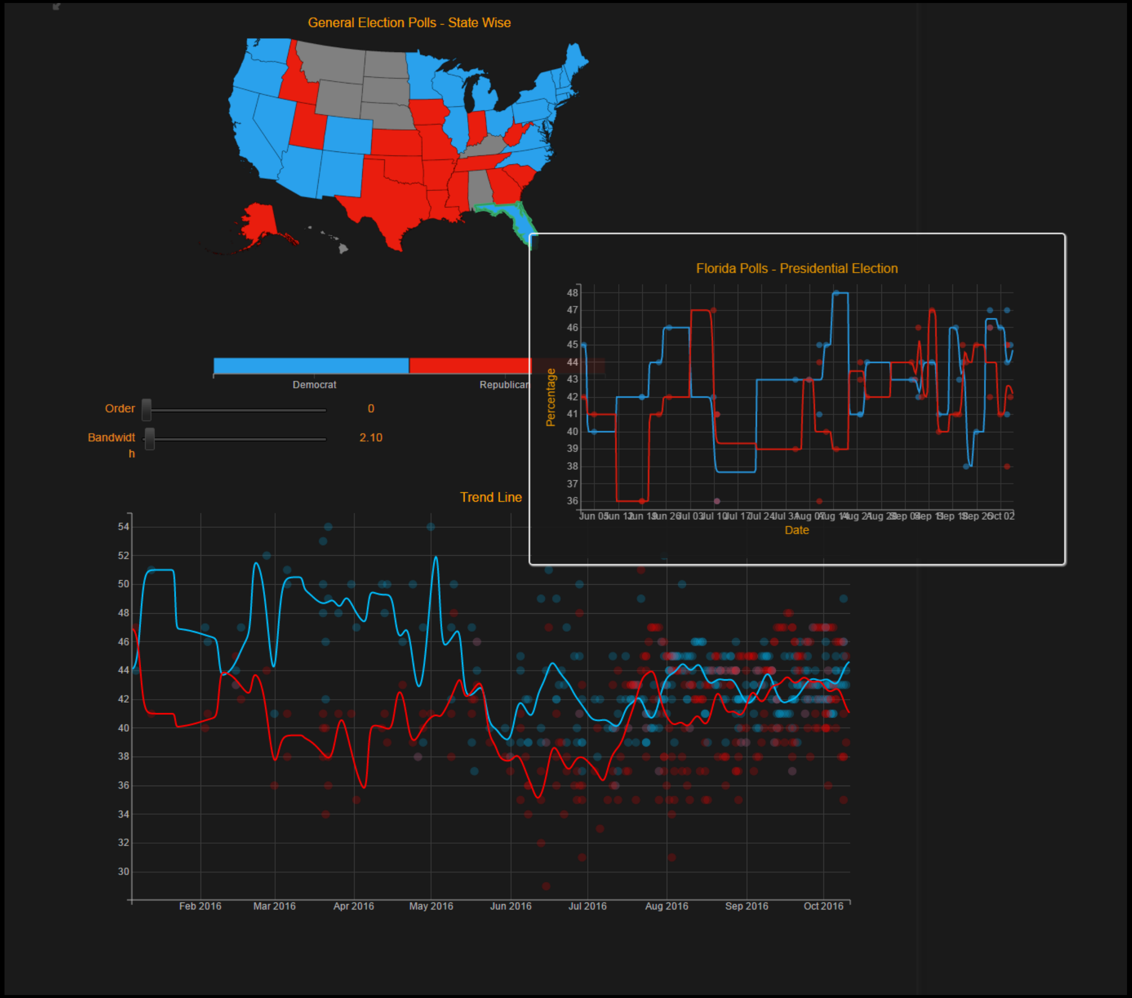
- Assume Market follows a Bachelier Process:

$$dM_t = \sigma^m dW_t$$

- Calibrate to polling data (Source: RealClearPolitics)

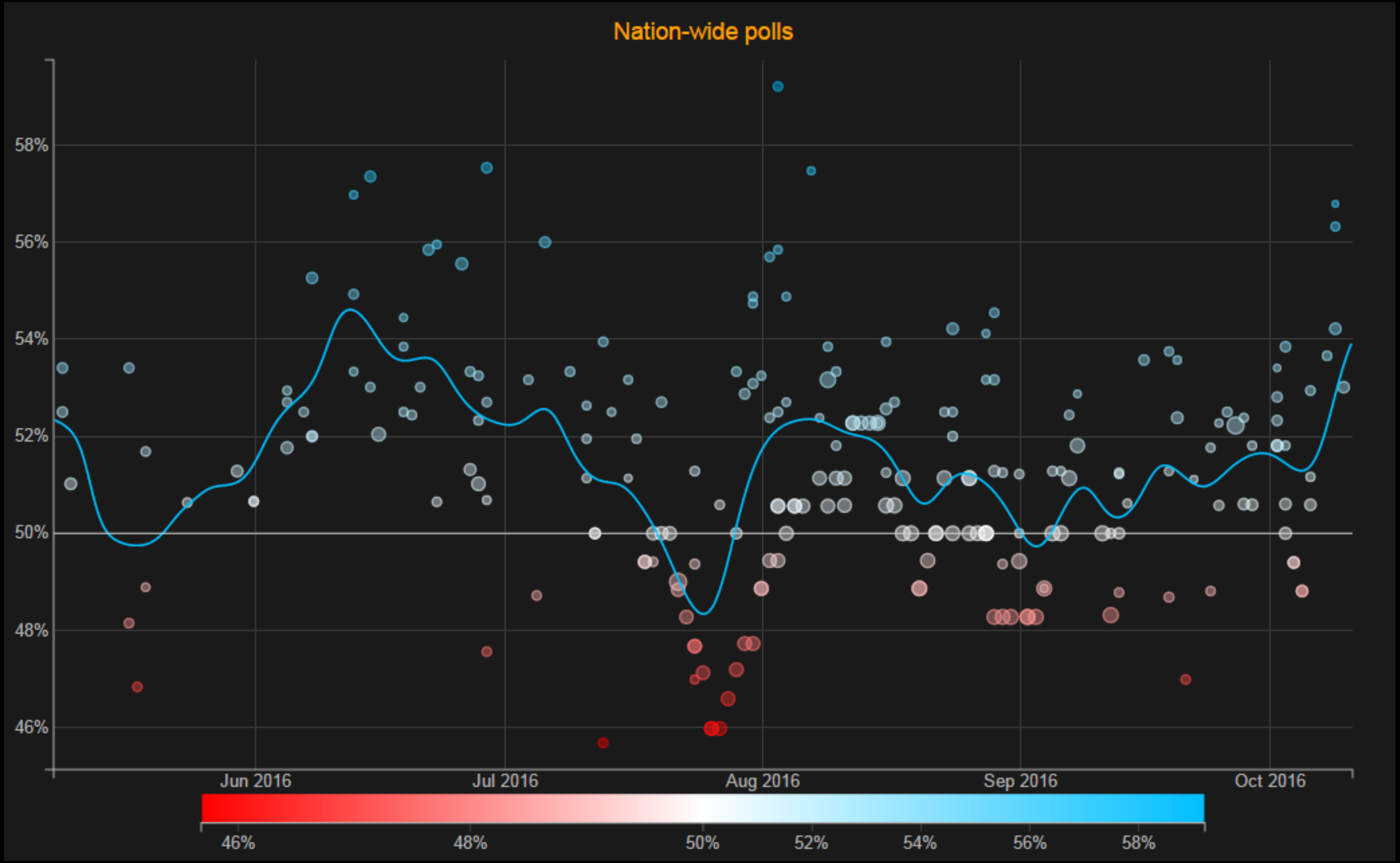
CALIBRATION

- Calibrate M to national polls
- Calibrate S to state polls



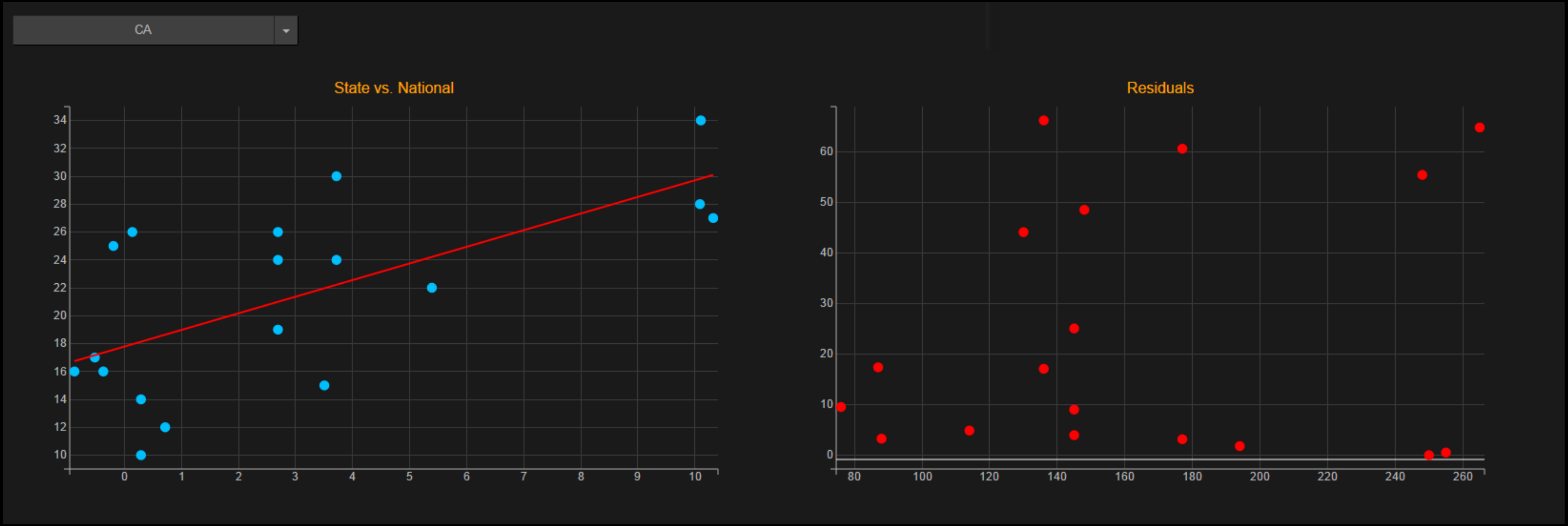
CALIBRATION

- Calibrate M to national polls

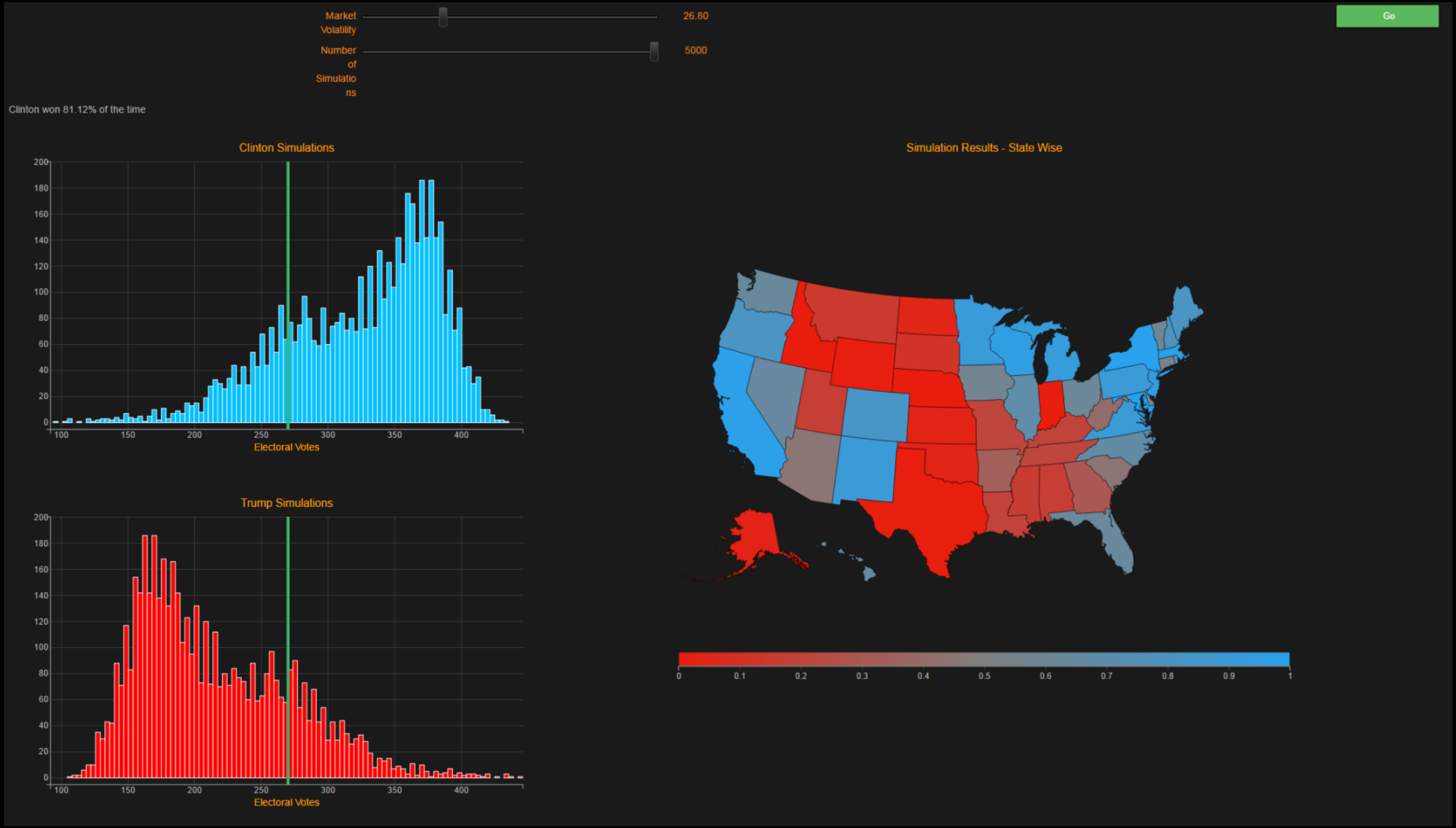


CALIBRATION

- Calibrate M to national polls
- Calibrate S to state polls



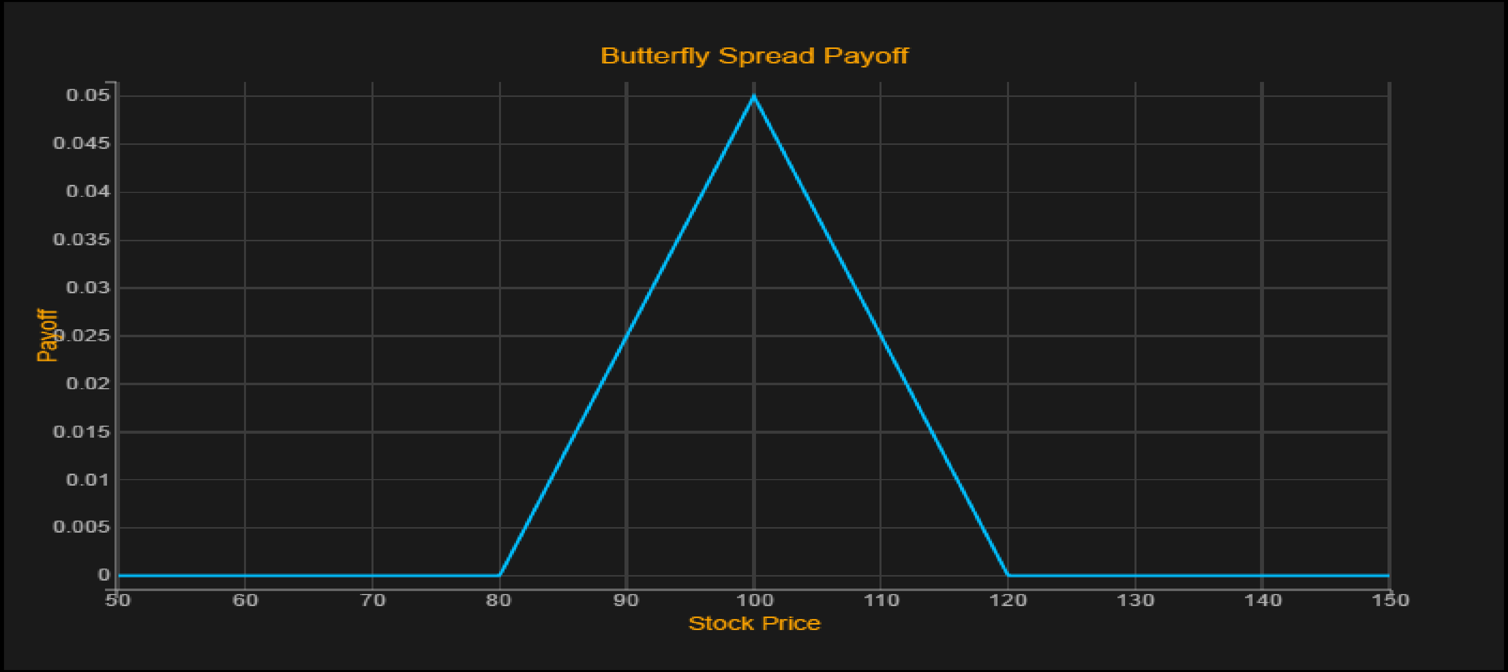
FINAL STEP: SIMULATE



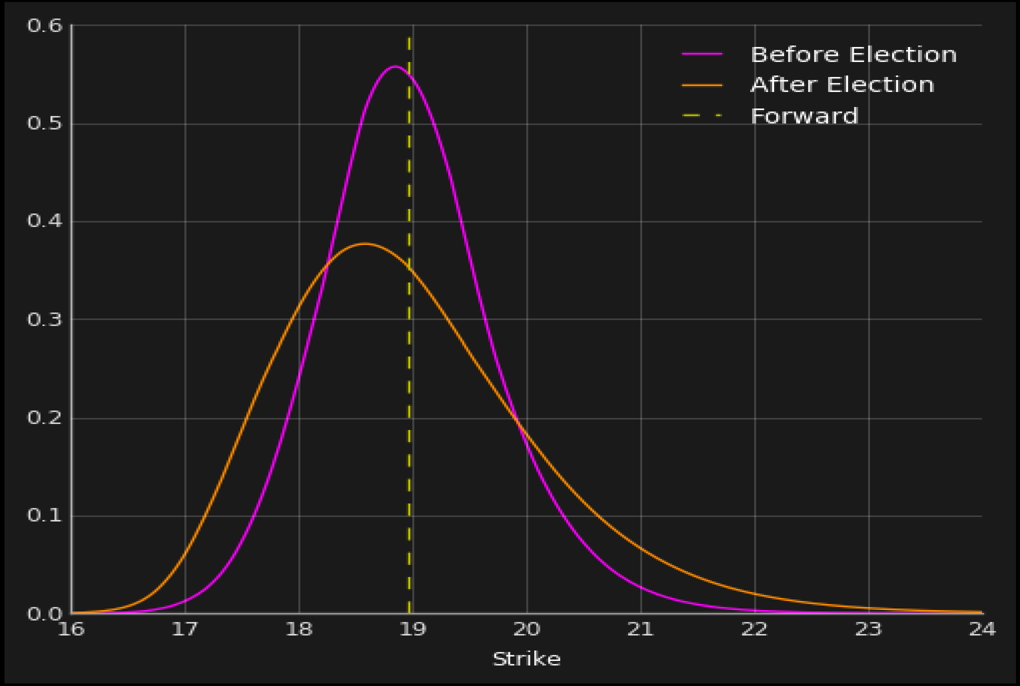
OPTION MODEL

RISK NEUTRAL DENSITY (NO RATES)

Butterfly Spread: $(\varepsilon > 0)$ $\dfrac{C(K-\varepsilon)-2C(K)+C(K+\varepsilon)}{\varepsilon^2}$



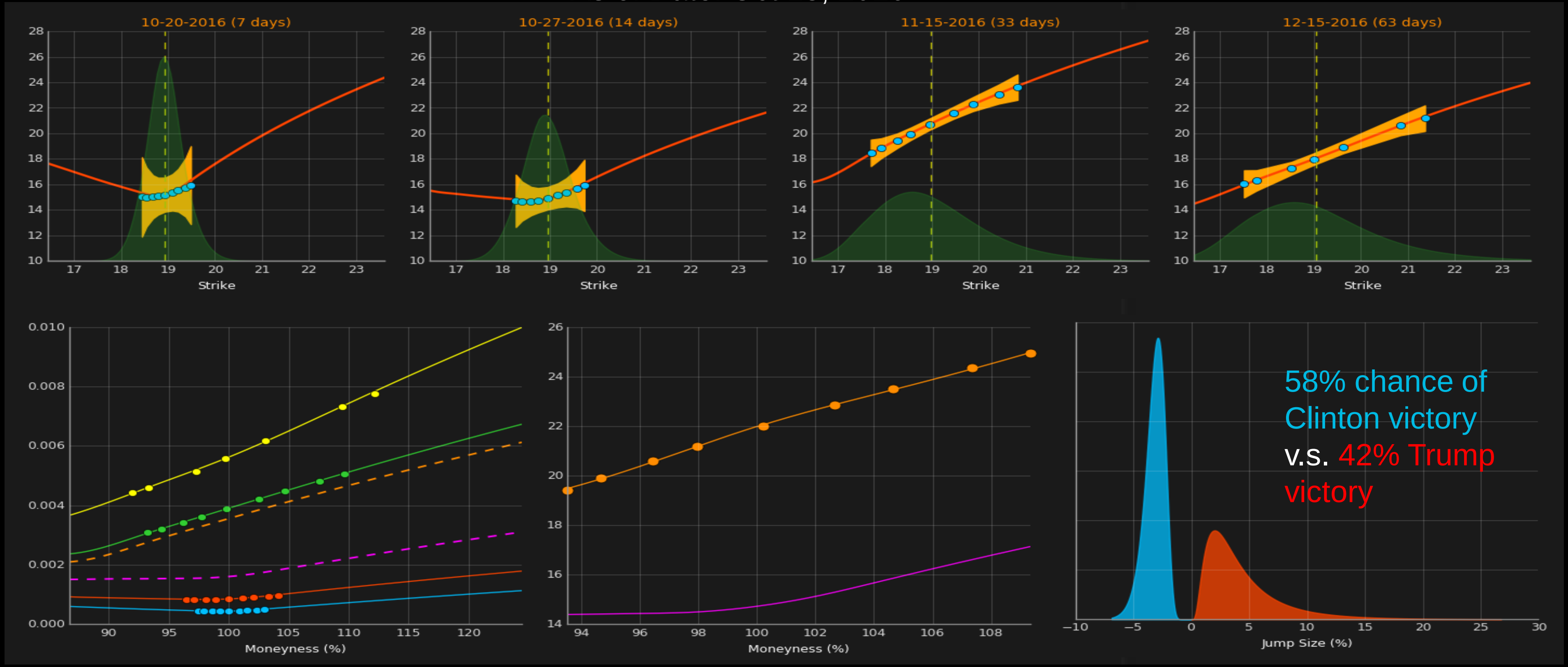
OPTION MARKET ANALYSIS



As of Date: Oct 13, 2016

OPTION MARKET ANALYSIS

As of Date: Oct 13, 2016



OTHER MODELS

ROBUST CALIBRATION OF S

- Calibrate S to state polls

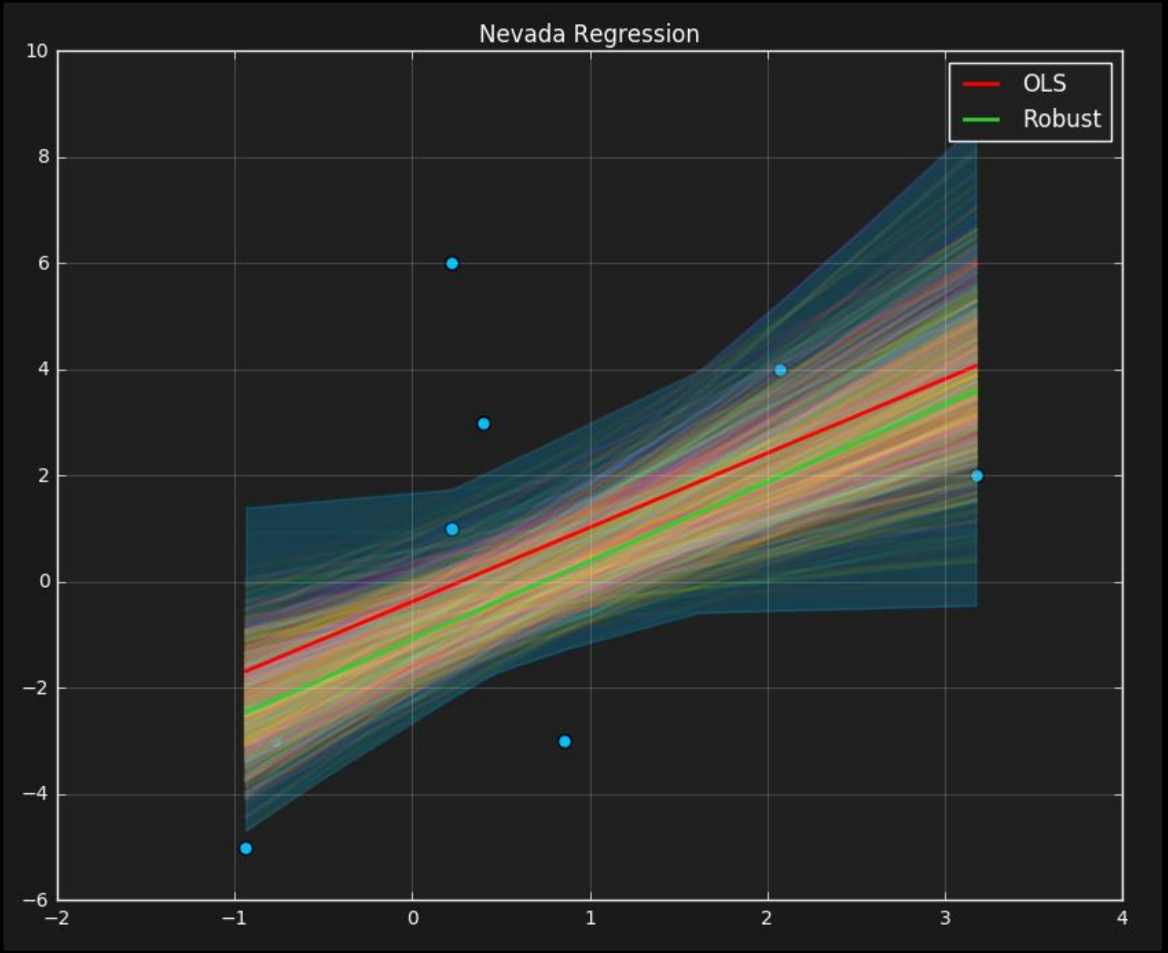
$$\alpha^i \sim N(\alpha^{OLS}, \sigma^{\alpha})$$

$$\beta^i \sim N\left(\beta^{OLS}, \sigma^{\beta}\right)$$

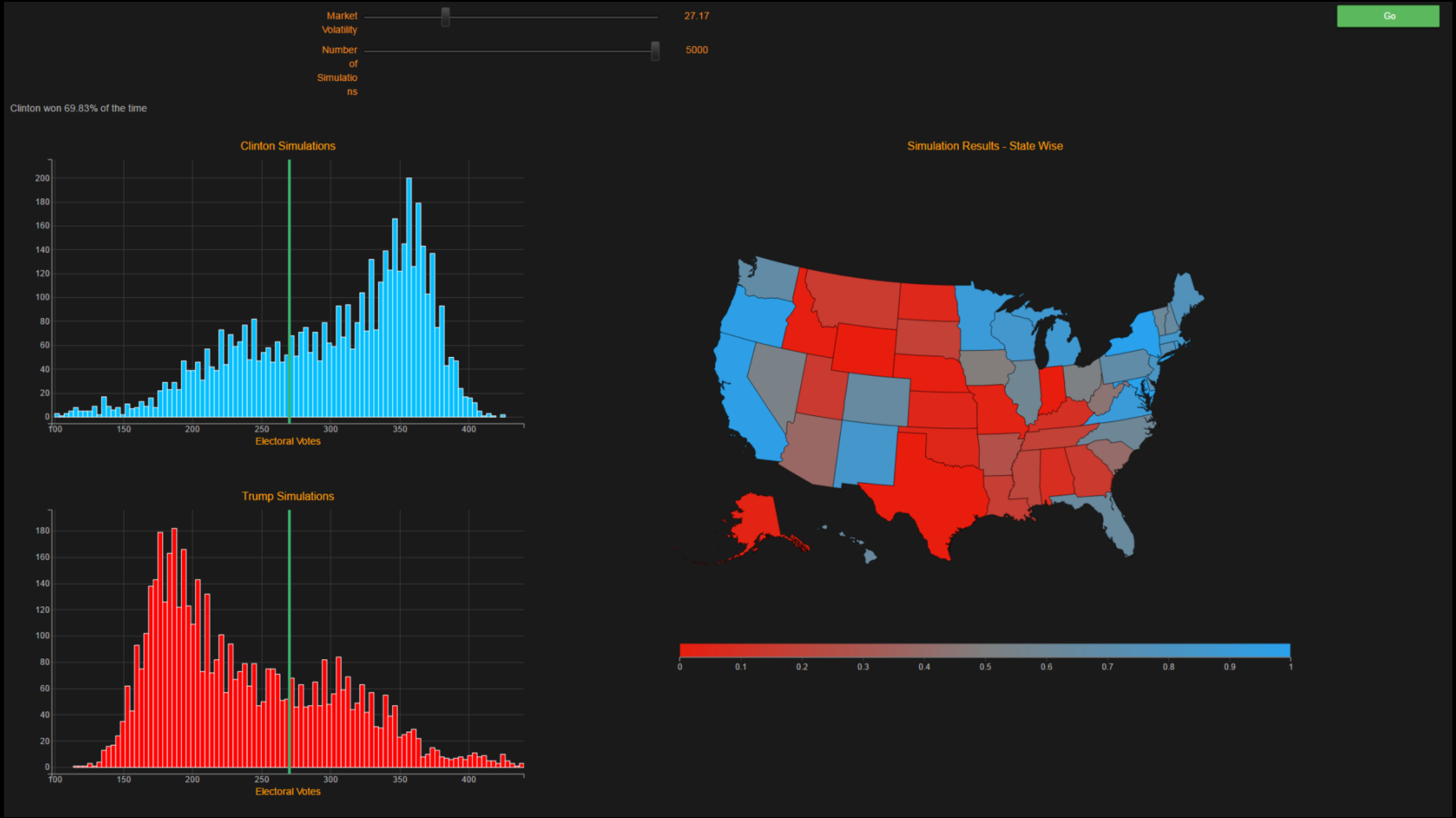
$$\sigma^i \sim \left|N\left(0, \epsilon^i\right)\right|$$

$$\mu^i = \alpha^i + \beta^i M_T$$

$$S_T^i \sim StudentT(\mu^i, \sigma^i, \nu = 3)$$



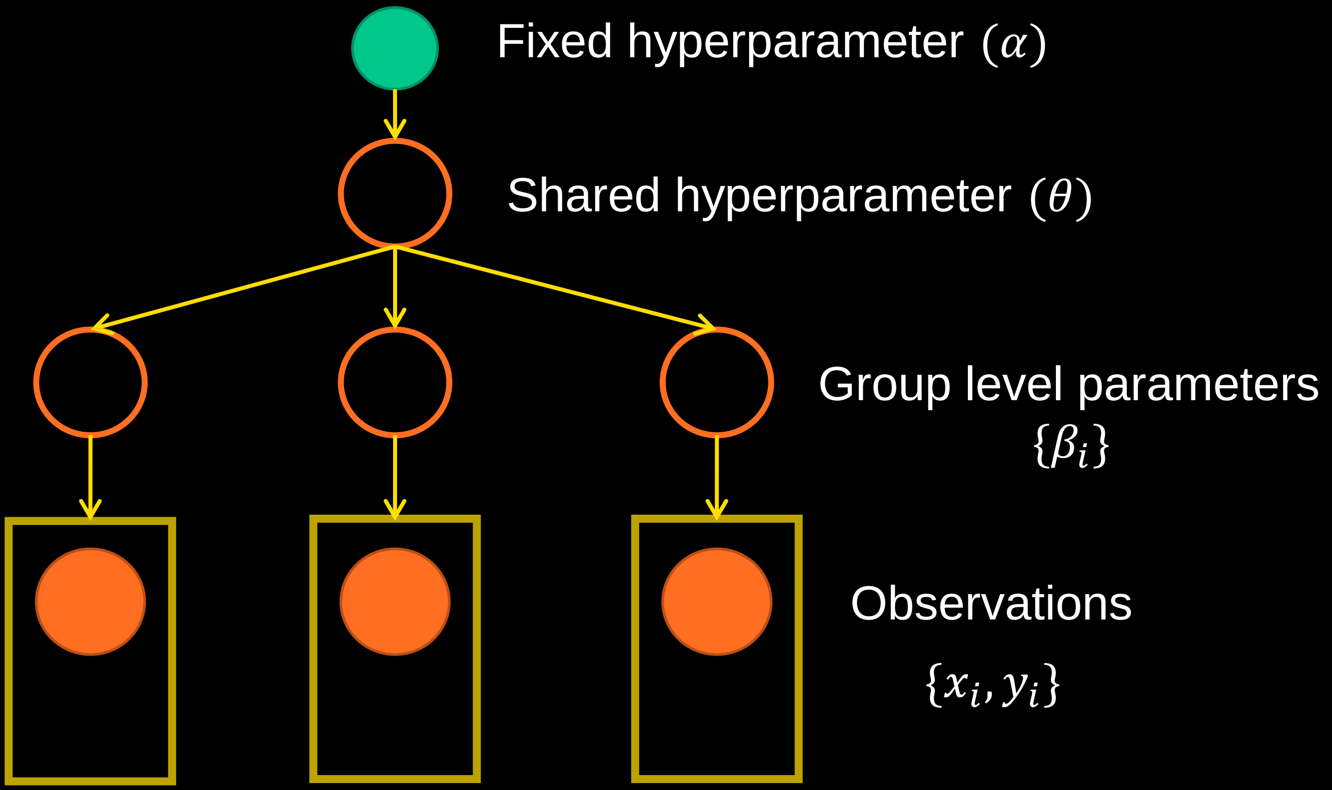
FINAL STEP: SIMULATE



HIERARCHICAL MODEL

FACTORING IN JOINT DYNAMICS

- Hierarchical Regression to incorporate joint dynamics between states



COMPARING FORECASTERS

WHO WAS BETTER?

- In general, given a sequence of forecasts and a single (or multiple realizations, how do we evaluate efficacy?

- Consider a canonical probability space endowed with a filtration
-

$$(\Omega, \mathcal{F})$$
$$\mathcal{P} \sim Class\ of\ Prob. Measures$$

- Scoring Rules: Map Probability Measure and Realized Event to Real Number

$$S: \Omega \times \mathcal{P} \to \overline{\mathbb{R}}$$

- What's the issue with using any function (like say $p(\omega)$)?

PROPER SCORING FUNCTIONS

- Scoring function should incentivize scorer to publicize his **true** probability. This means that:

$$\max_{q} \mathbb{E}^{\mathbb{P}}[S(\omega, q)] = p$$

It is strictly proper if p is the unique maximizer (since a constant scoring function which assigns 1 to all events is technically proper).

WHO WAS BETTER?

There are many ways to compare forecasters:

- Brier Score

$$Brier\ Score = \sum_{t=1}^{T} (p_i - o_i)^2$$

- Lower score is better
- Proper (Honest Scoring Rule)

CONNECTION TO MACHINE LEARNING

Condition for a proper scoring function:

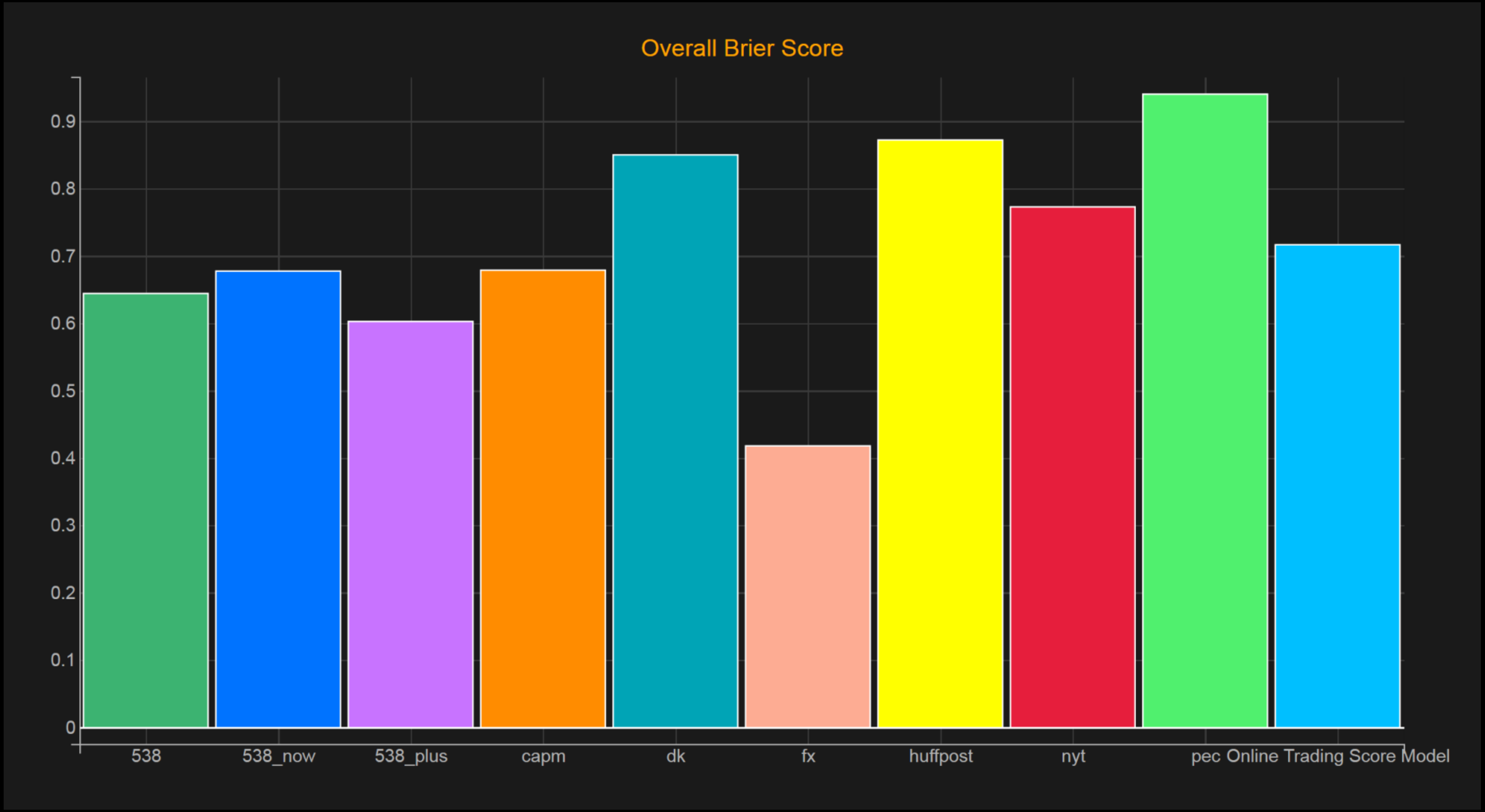
$$\operatorname*{argmin}_{q} \mathbb{E}^{\mathbb{P}}[S(\omega, q)] = \mathbb{E}^{\mathbb{P}}[\omega]$$

imply (with some regularity conditions [1] that $S(x, y) = D_{\phi}(x, y)$ for some convex, differentiable function $\phi: \mathbb{R} \to \mathbb{R}$

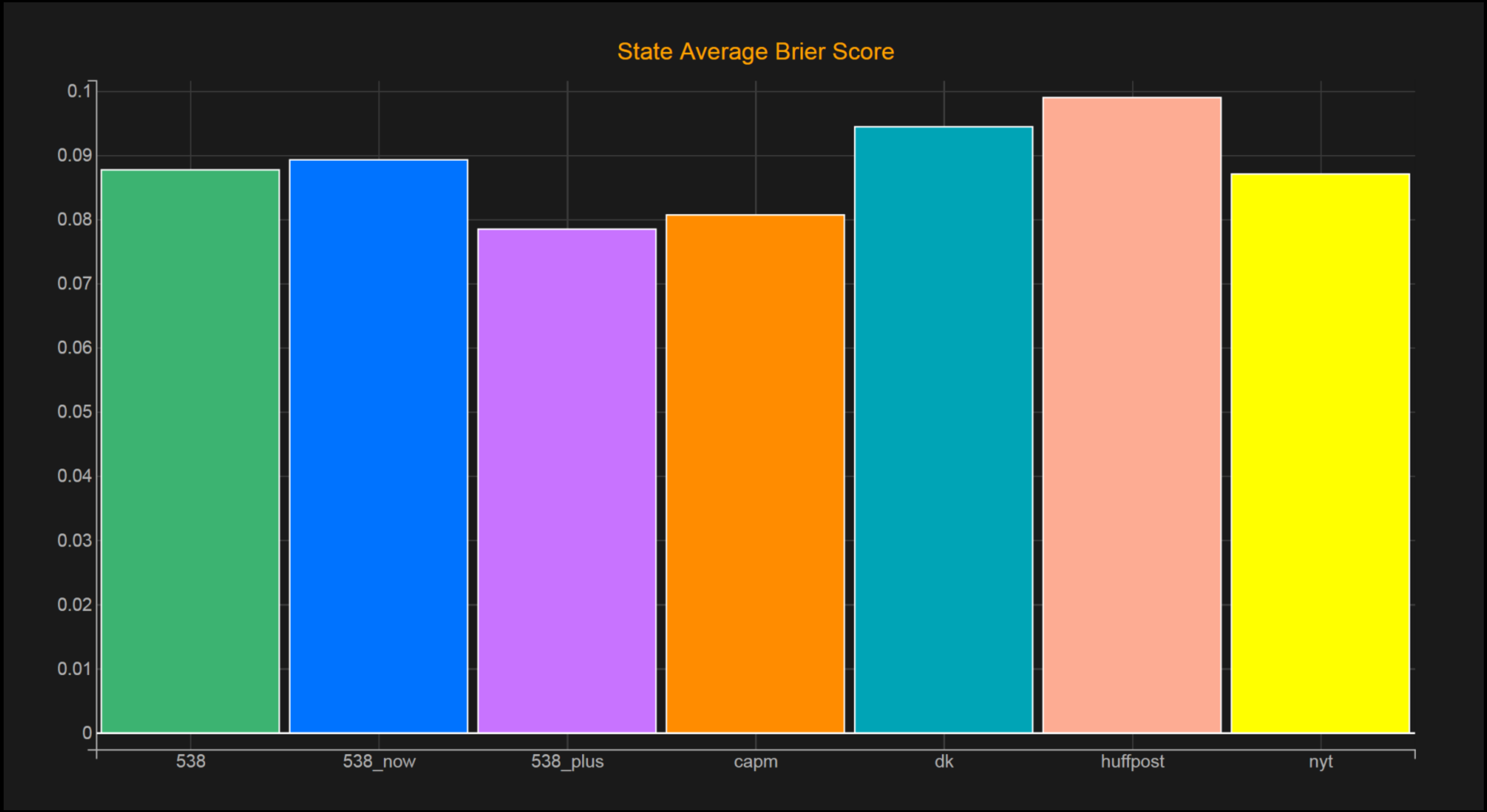
$$D_{\phi}(x, y) = \phi(x) - \phi(y) - \phi'(y)(x - y)$$

Proper Scoring Functions ↔ Bregman Divergences

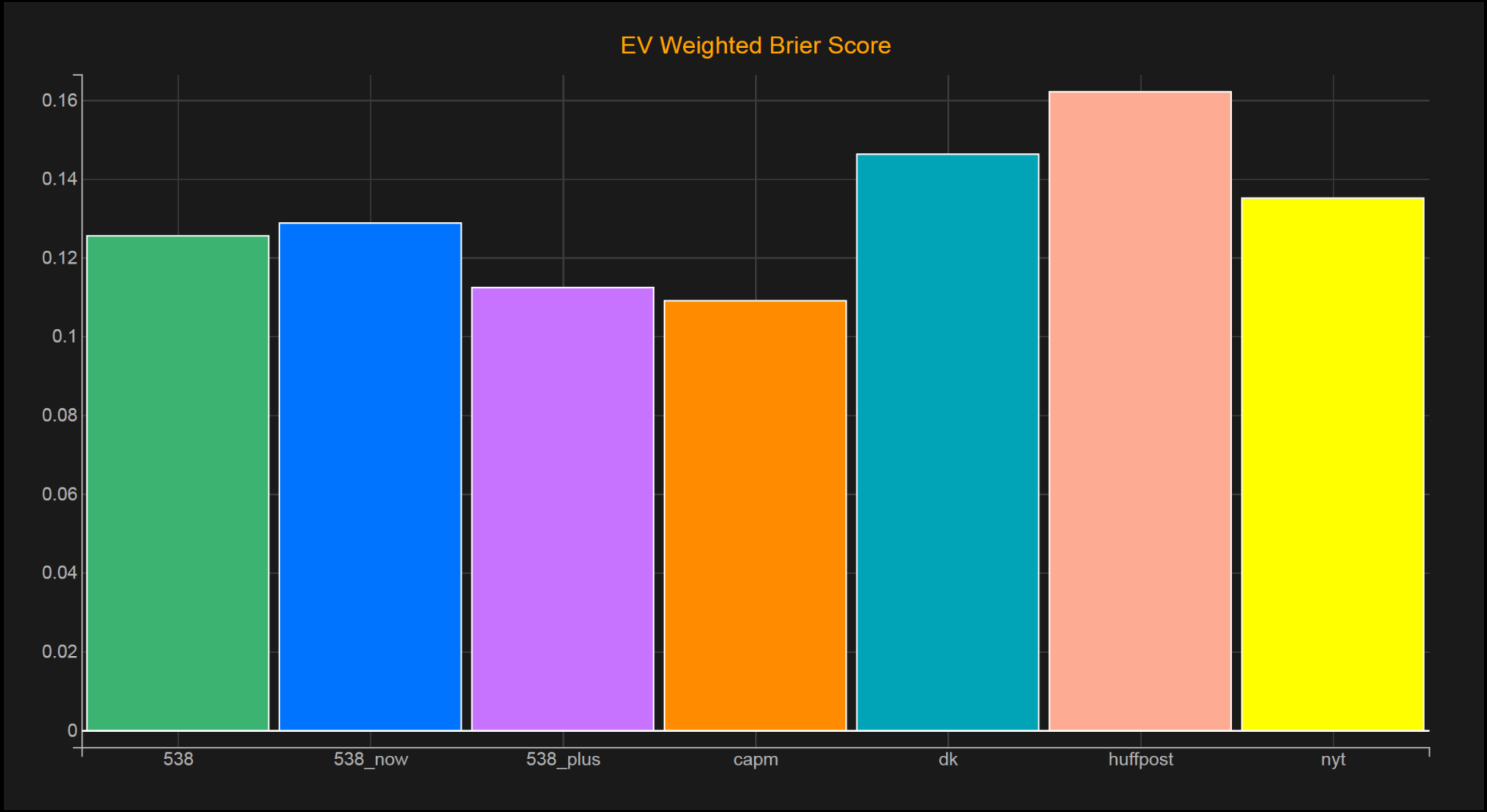
BRIER SCORES



BRIER SCORES – STATEWISE AVERAGE



BRIER SCORES – EV WEIGHTED AVERAGE



WHO WAS BETTER?

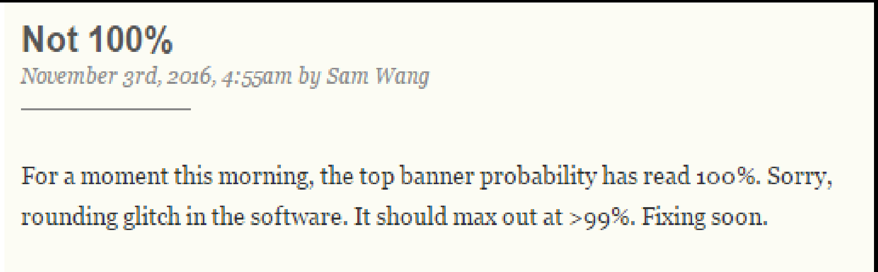
There are many ways to compare forecasters:

- Log Likelihood

Imagine we were trying to forecast rain or not every month.

- True Probability $\rightarrow \frac{1}{10}$
- Forecaster A $\rightarrow \frac{1}{4}$
- Forecaster B $\rightarrow 0$

SILVER VS WANG 2016



Not 100%

November 3rd, 2016, 4:55am by Sam Wang

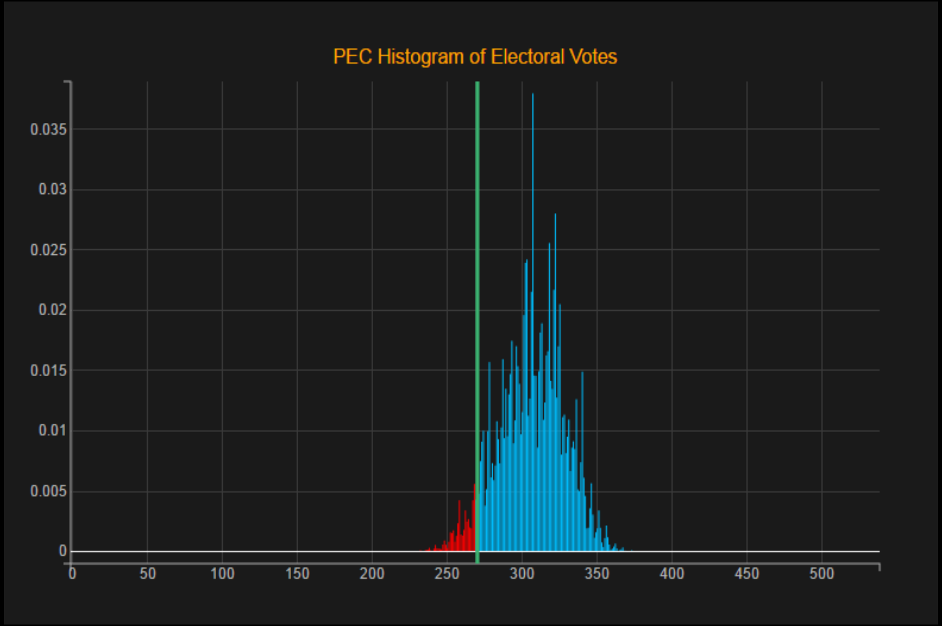
For a moment this morning, the top banner probability has read 100%. Sorry, rounding glitch in the software. It should max out at >99%. Fixing soon.

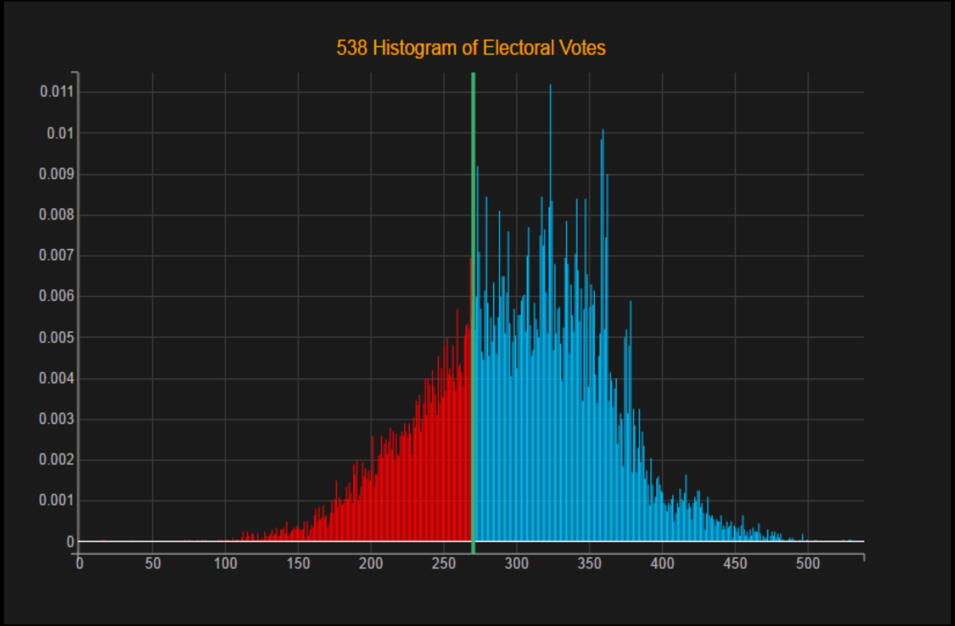
What if Clinton had won in a landslide? Who would have been better?

SILVER VS WANG 2016

What if Clinton had won in a landslide? Who would have been better?

Nate Silver

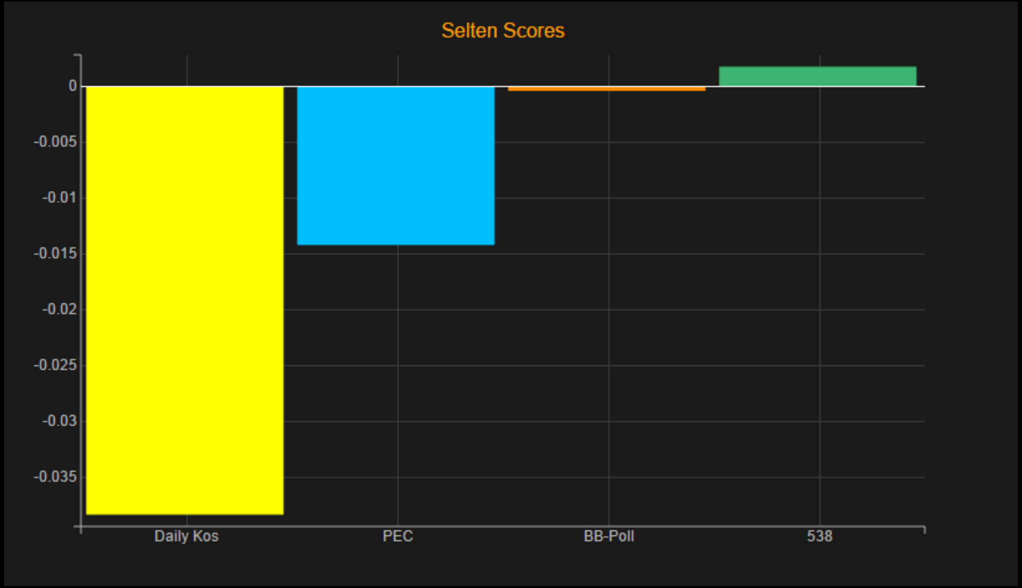




WHO WAS BETTER?

There are many ways to compare forecasters:

- Selten Score – Compute Brier Score for each bin of the Histogram

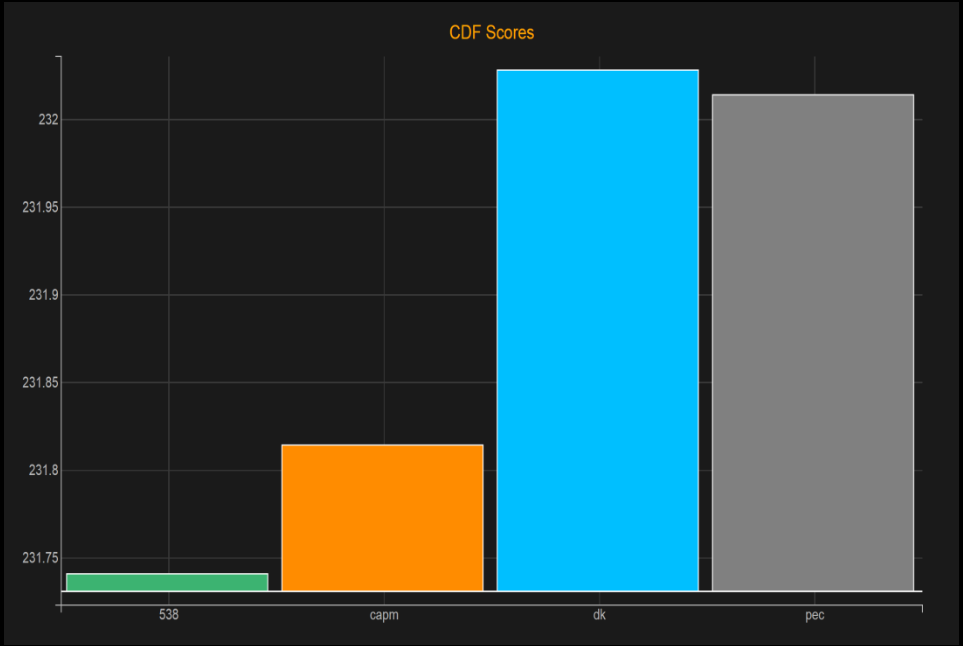


$$Selten = 2p_i^* - \sum_{i=1}^{N} p_i^2$$

- Proper (Honest Scoring Rule)
- Highest value is best
- Does not take into account topology of bins

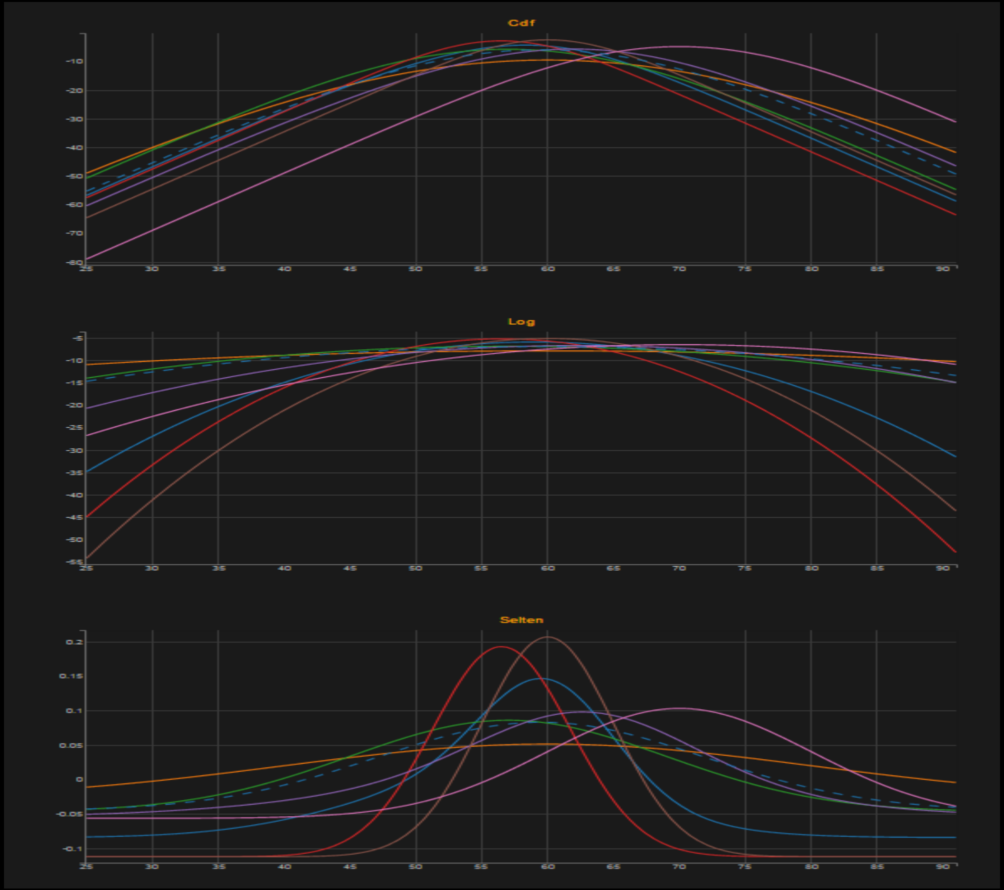
WHO WAS BETTER?

Selten has no notion of the topology of bins. Guessing 302 EV, when the result was 303 EV is no different from guessing 538 EV. We present the CDF score which factor this into the scoring function, by taking the Brier at each level for the CDF.



$$CDF\ Score = \sum \left(F(x) - 1_{\{x \geq \omega\}}\right)^2$$

WHO WAS BETTER?

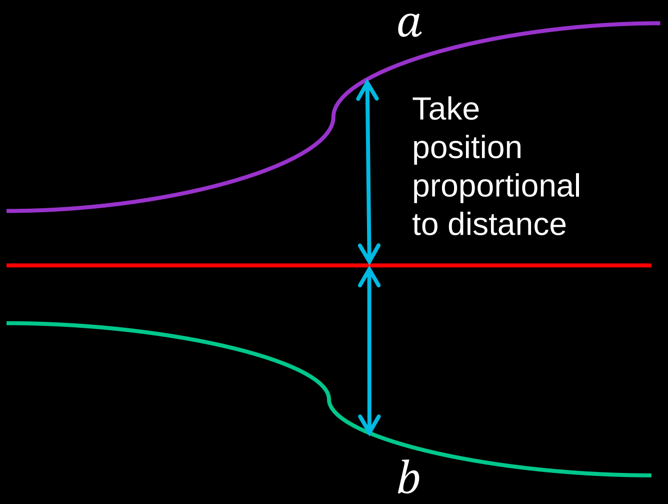


WHO WAS BETTER?

There are many ways to compare forecasters:

- All these are ex-post judgements. How can we decide who is better before the event? (or how do we perform online learning on forecasters)
- Take position at time t proportional to their distance from mid_t or the betting market – Settle at realization
- Compute P&L of the holdings at current mid
- Proper Scoring Rule
- Best P&L so far is best guess of best forecaster

TRADING SCORE



$$Position\ at\ time\ t\!:\ (a_t - b_t)$$

$$Price\!:\ \frac{a_t + b_t}{2}$$

2 Date, 1 Period Model:

$$\mathrm{PNL(T)} = 1_{\{\omega\}}(a_0 - b_0) + \frac{{a^2}_0 - b_0^2}{2}$$

$$\Rightarrow \underset{a_0}{\operatorname{argmax}} PNL(T) = \mathbb{E}^{\mathbb{P}}\left[1_{\{\omega\}}\right]$$

COMBINING FORECASTERS

Minimize regret:

$$Regret(T) := \sum_{t=1}^{T} L(\hat{y}_t, y_t) - \min_i \sum_{t=1}^{T} L(\hat{y}_{t,i}, y_t)$$

Weight Update:

$$w_{t+1,i} \leftarrow w_{t,i} e^{-\eta L\left(\widehat{y_{t,i}}, y_t\right)}$$

Prediction:

$$\widehat{y_t} = \frac{\sum_{i=1}^{N} w_{t,i} y_{t,i}}{\sum_{i=1}^{N} w_{t,i}}$$

PREDICTION AND P&L