Models & Markets

Comparing 2018 Midterm Predictions

Question

- What is the best way to predict elections?
- When does each method perform best? Why?

Hypothesis

- Prediction **markets** will outperform **models**

Why Predict Elections?

- Resource allocation
- Strategy adjustment
- Quantitative journalism
- Uncertainty is scary!

How to Predict Elections?

1. Opinion polling
2. Poll aggregation
3. Forecasting models
4. Prediction markets

Forecasting Models

Forecasting Models

- 21st Century Invention
- Quantitative inputs
- Express uncertainty
- Probabilistic
- Proprietary

Who will win the presidency?

Chance of winning

71.4%   **28.6**%

Electoral votes

🟦 Hillary Clinton	302.2
🟥 Donald Trump	235.0
🟪 Evan McMullin	0.8
🟨 Gary Johnson	0.0

Popular vote

🟦 Hillary Clinton	48.5%
🟥 Donald Trump	44.9%
🟨 Gary Johnson	5.0%
⬜ Other	1.6%

PRESIDENT SENATE

By Natalie Jackson and Adam Hooper
Additional design by Alissa Scheller

CLINTON
98.0%

TRUMP
1.7%

Model Process

> *Most election models work in something like the following way: [1] they calculate the <u>most likely outcome</u> in a particular state... and then [2] they determine the <u>degree of uncertainty</u> around that estimate.*

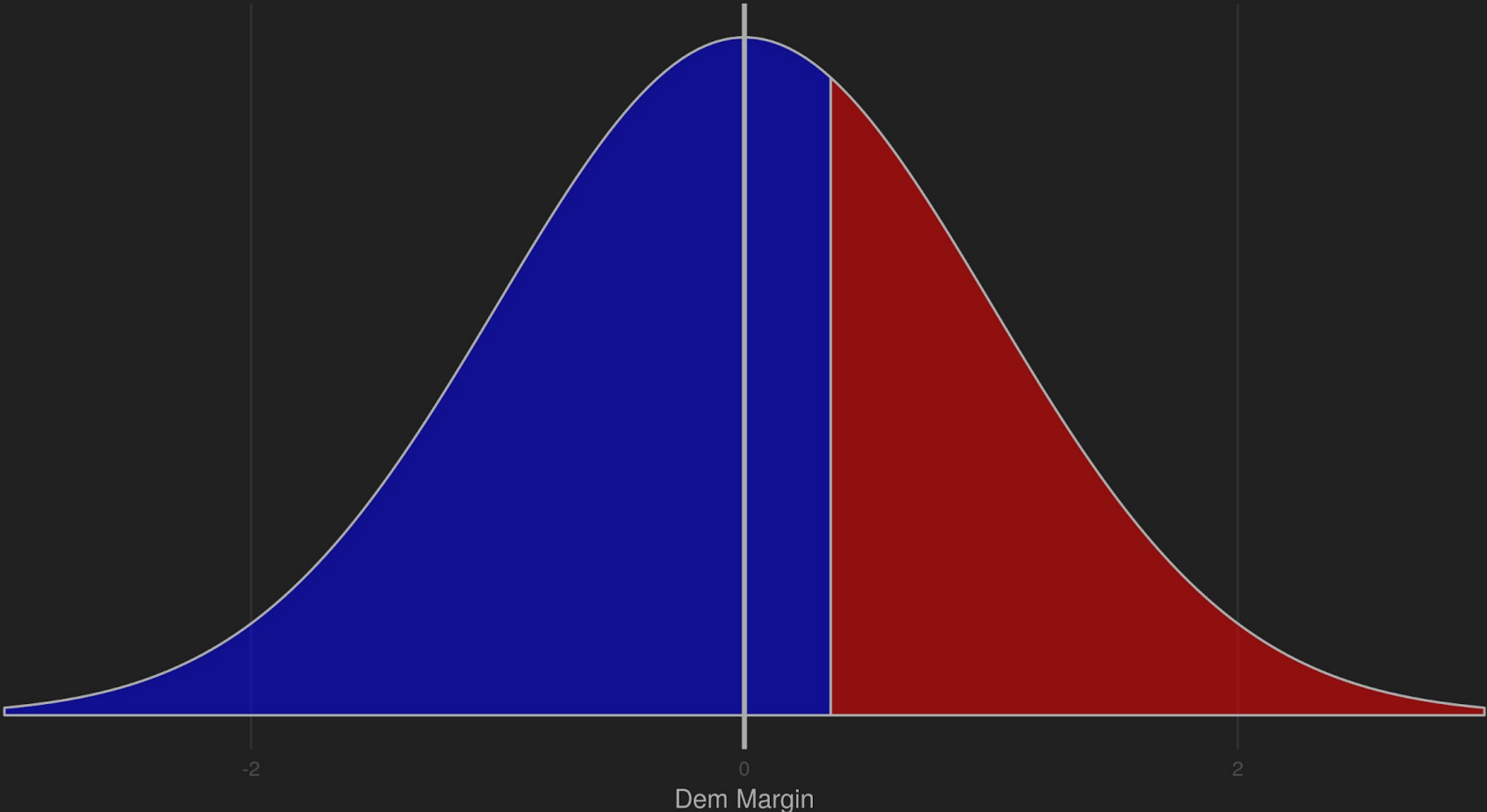– Nate Silver

[1] Model Inputs

1. Weighted poll aggregation
2. CANTOR imputation
3. Fundamentals
   - incumbency
   - fundraising
   - previous election
   - scandals, etc

[2] Model Uncertainty

- Fewer polls
- Lopsided race
- Further away
- Polls disagree
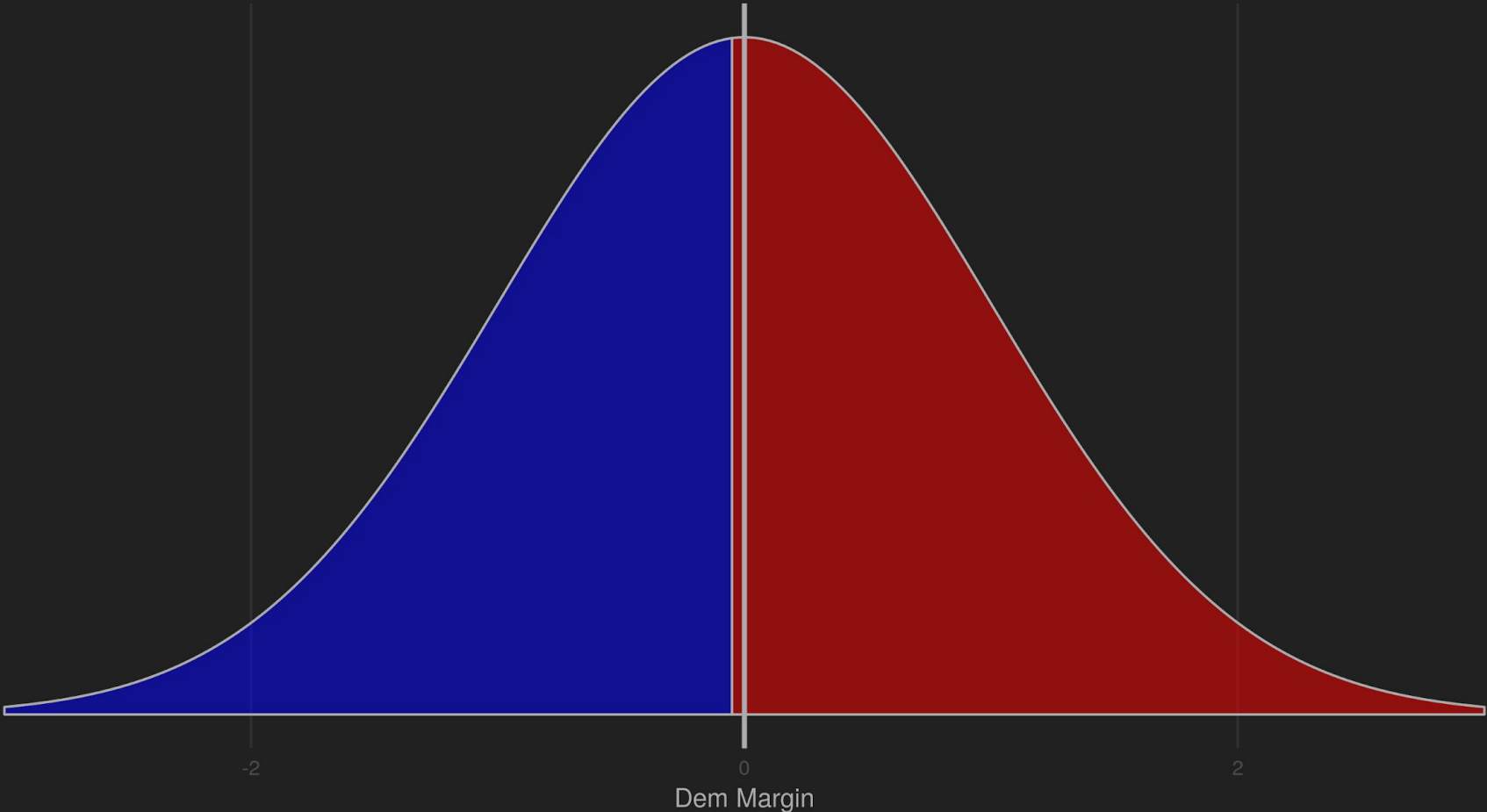 - With one another
  - With fundamentals
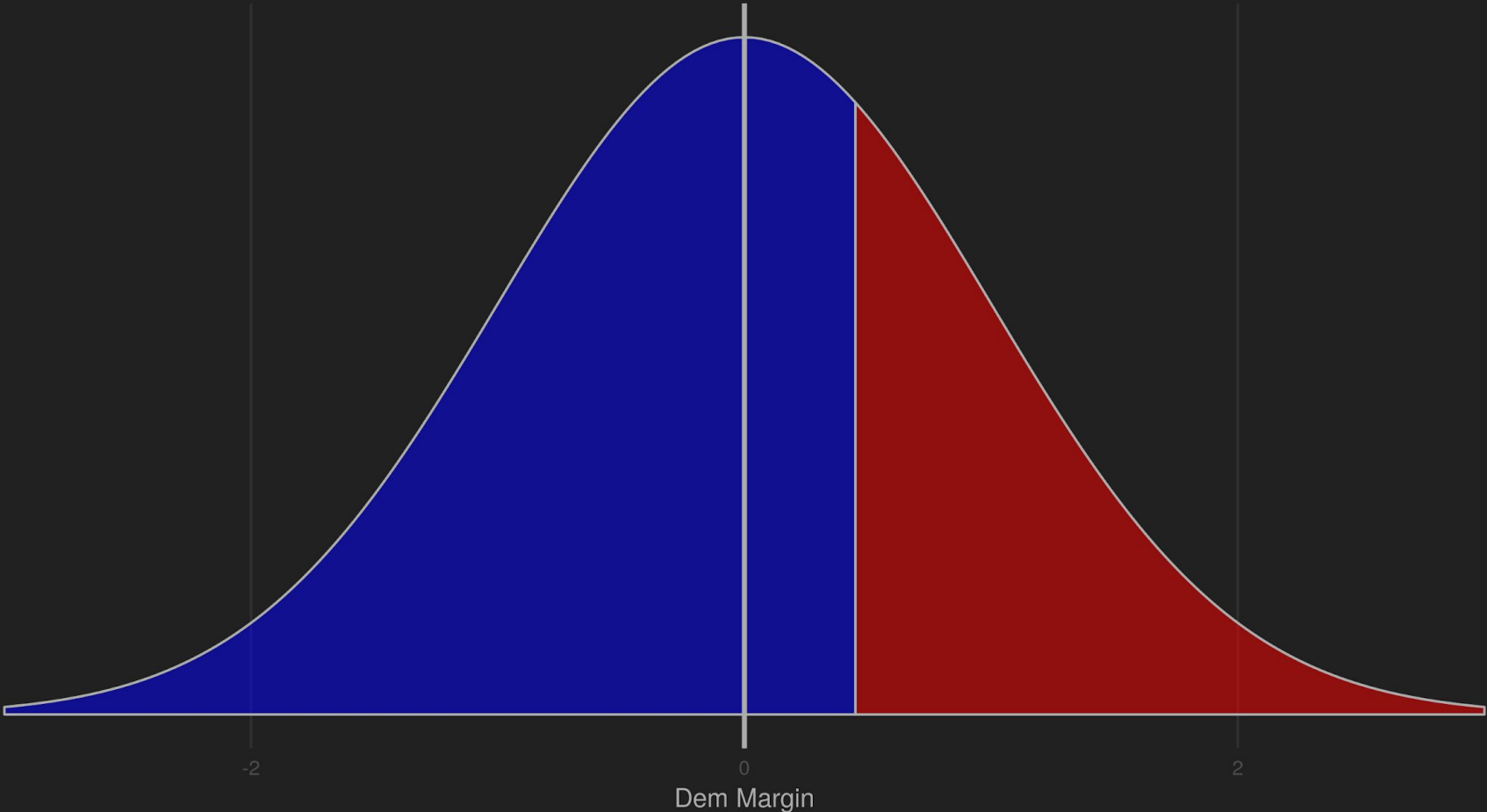
# Model Probability Distribution
Winner = DEM

Dem Margin

# Model Probability Distribution
Winner = GOP

Dem Margin

Model Probability Distribution

Winner = DEM

Dem Margin

Model Variables

1. Date
2. Election
3. Party
4. Special
5. Incumbency
6. Voteshare(s)
7. **Probability**

Prediction Markets

Prediction Markets

- Exchange-traded binary options
- Efficient market hypothesis
- Elections, terrorism, shipwrecks
- Dubious legality
  - Academic waiver from CFTC

Market Contracts

- $850 limit
- Buyers on either side
- $1 or $0 based on outcome
- Sell at any time
- Price change

Contract	Latest Yes Price	Best Offer		Best Offer
Joe Biden	24¢ 1¢▲	25¢	Buy Yes / Buy No	76¢
Bernie Sanders	22¢ 1¢▼	22¢	Buy Yes / Buy No	79¢
Pete Buttigieg	18¢ NC	19¢	Buy Yes / Buy No	82¢
Kamala Harris	14¢ NC	15¢	Buy Yes / Buy No	86¢
Andrew Yang	9¢ 1¢▼	10¢	Buy Yes / Buy No	91¢
Elizabeth Warren	8¢ 1¢▲	9¢	Buy Yes / Buy No	93¢

Market Variables

1. ID
2. Question
3. Name
4. Contract
5. Date
6. Volume
7. **Price**

Comparing Methods

Reading data

```
"https://fivethirtyeight.com/data/..." %>%
  read_memento(timestamp = "2018-11-06") %>%
  read_csv(col_types      = cols(
    Democrat_Won   = col_logical(),
    Republican_Won = col_logical(),
    uncalled       = col_logical(),
    forecastdate   = col_date("%m/%d/%y"),
    category       = col_factor())
```

Formatting data

```r
model_district %>%
  bind_rows(model_seat, .id = "chamber") %>%
  unite(state, district,
        col = race,
        sep = "-") %>%
  rename(date = forecastdate,
         prob = win_probability) %>%
  filter(name != "Others") %>%
  arrange(date, name)
```

Models

1. Date
2. Election
3. Party
4. Voteshare
5. **Probability**
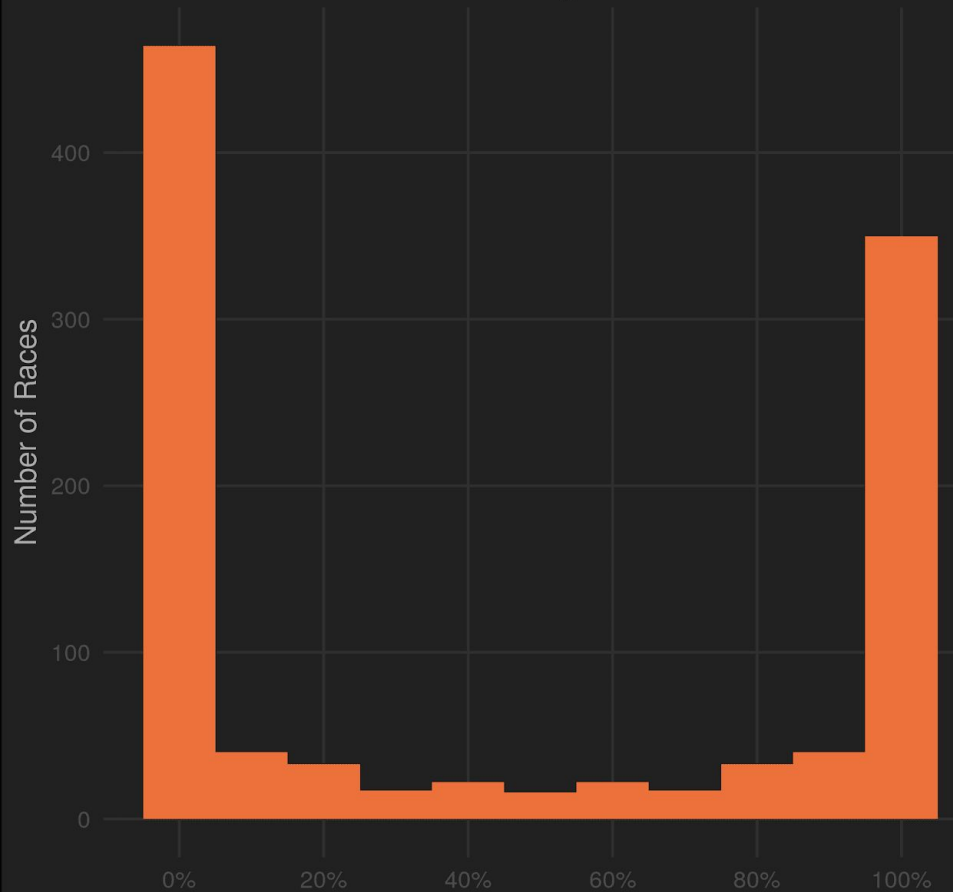
Markets

1. Date
2. Market
3. Contract
4. Volume
5. **Price**

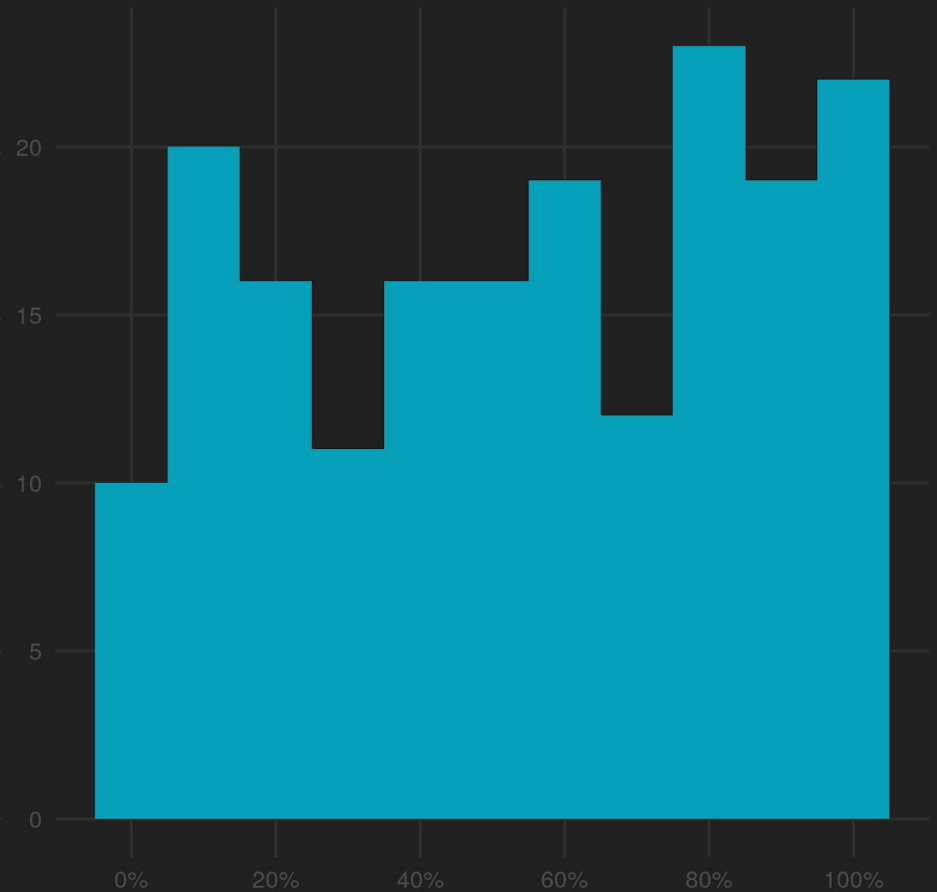# Comparing data

```
markets_dem %>%
  inner_join(model_dem) %>%
  rename(model  = prob,
         market = close) %>%
  gather(key   = method,
         value = prob) %>%
  inner_join(results) %>%
  mutate(brier = (winner - prob)^2)%>%
  mutate(hit = (prob > 0.5) == winner)
```

Distribution of Race Probabilities by Predictive Method

Forecasting Model

Prediction Markets

Number of Races
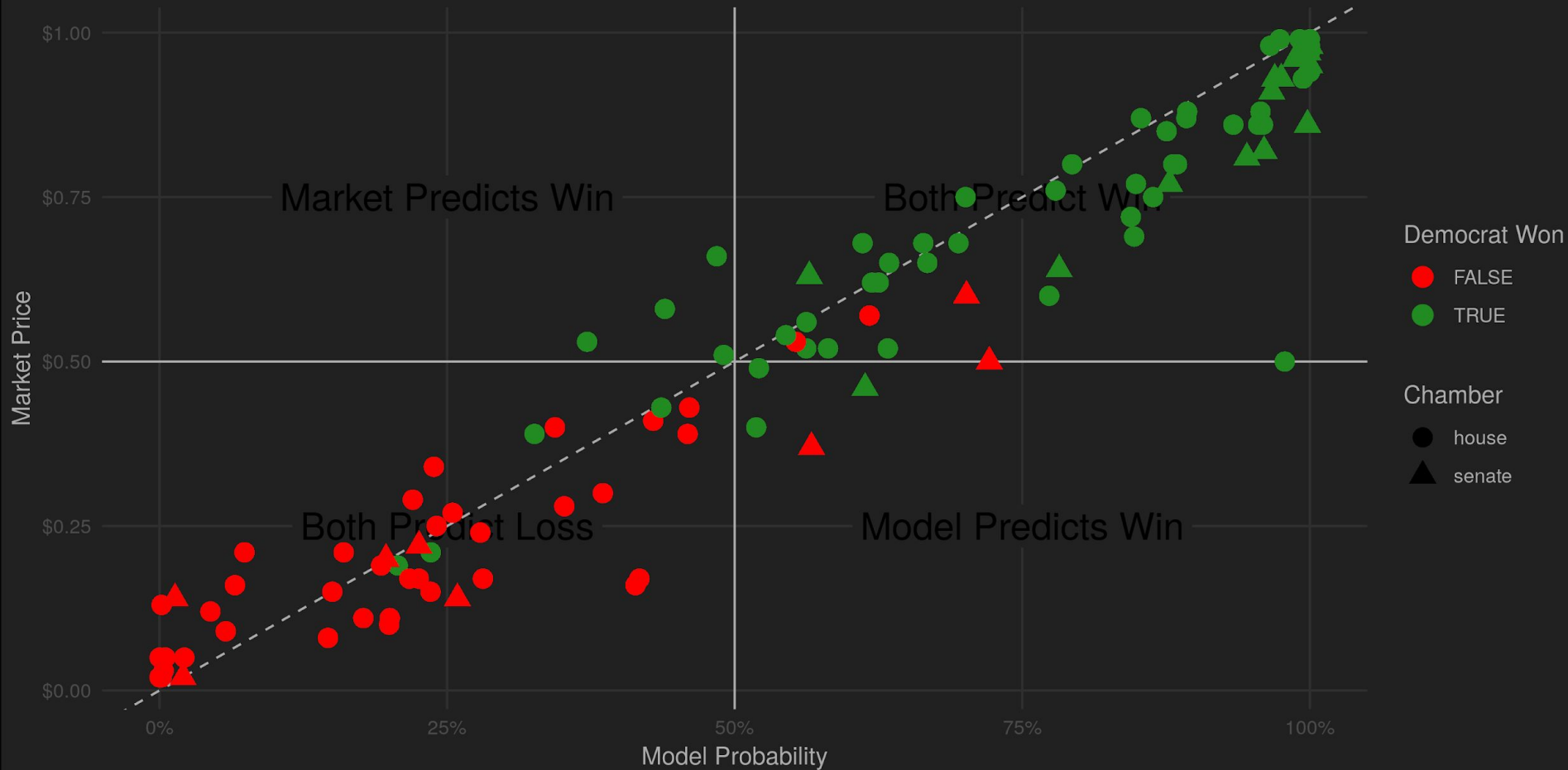
Democratic Win Probability

Midterm Races by Democrat's Chance of Winning

November 5th, Night Before Election Day

Market Predicts Win

Both Predict Win

Both Predict Loss

Model Predicts Win

Market Price

Model Probability

Democrat Won
FALSE
TRUE

Chamber
house
senate

Date	Race	Method	Prob	Pred	Win	Hit
8/1	AZ-S1	Market	66	1	1	1
8/1	AZ-S1	Model	74	1	1	1
8/1	CA-12	Market	91	1	1	1
8/1	CA-12	Model	100	1	1	1
8/1	CA-22	Market	30	0	0	1
8/1	CA-22	Model	5	0	0	1
8/1	CA-39	Market	61	1	1	1
8/1	CA-39	Model	38	0	1	0

Proportion of Correct Predictions by Week

PredictIt Markets and FiveThirtyEight Model
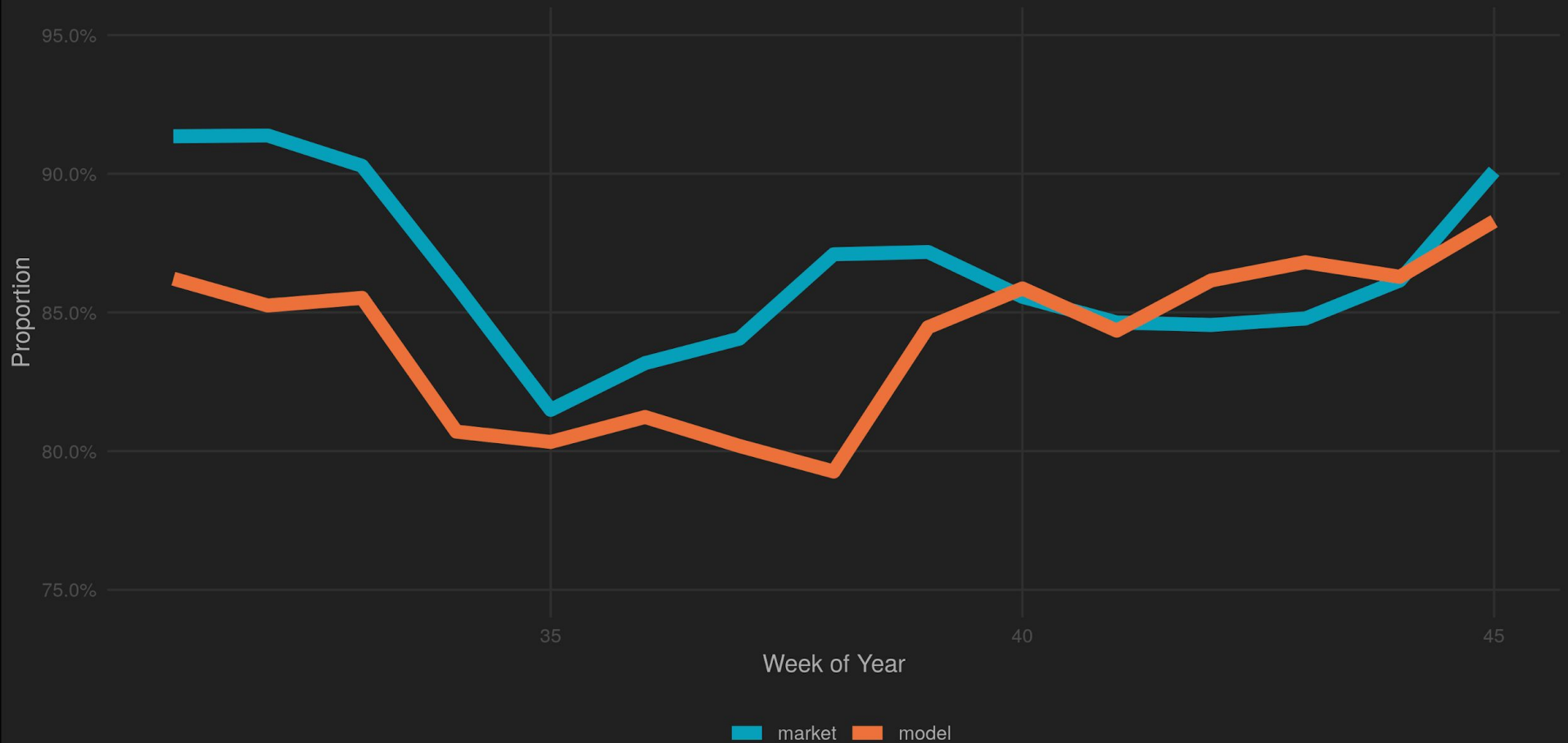
market model

Proportions Test

- Market proportion: 86.0343%
- Model proportion: 83.8057%
- Lower bound: 1.158%
- Upper bound: 3.2999%
- X-squared = 16.794
- p-value = 0.000042
- Alternative hypothesis: Two sided

	TRUE		FALSE	
	Markets	Model	Markets	Model
TRUE	**79.5** (51.2)	**84.5** (51.2)	**40.6** (9.7)	**36.5** (9.7)
FALSE	**59.3** (4.3)	**63.7** (6.5)	**23.0** (34.3)	**16.8** (32.1)

Brier Score

- Democrat has has 70% chance
- Democrat wins on election day
- Brier Score = $(0.70-1)^2 = 0.09$
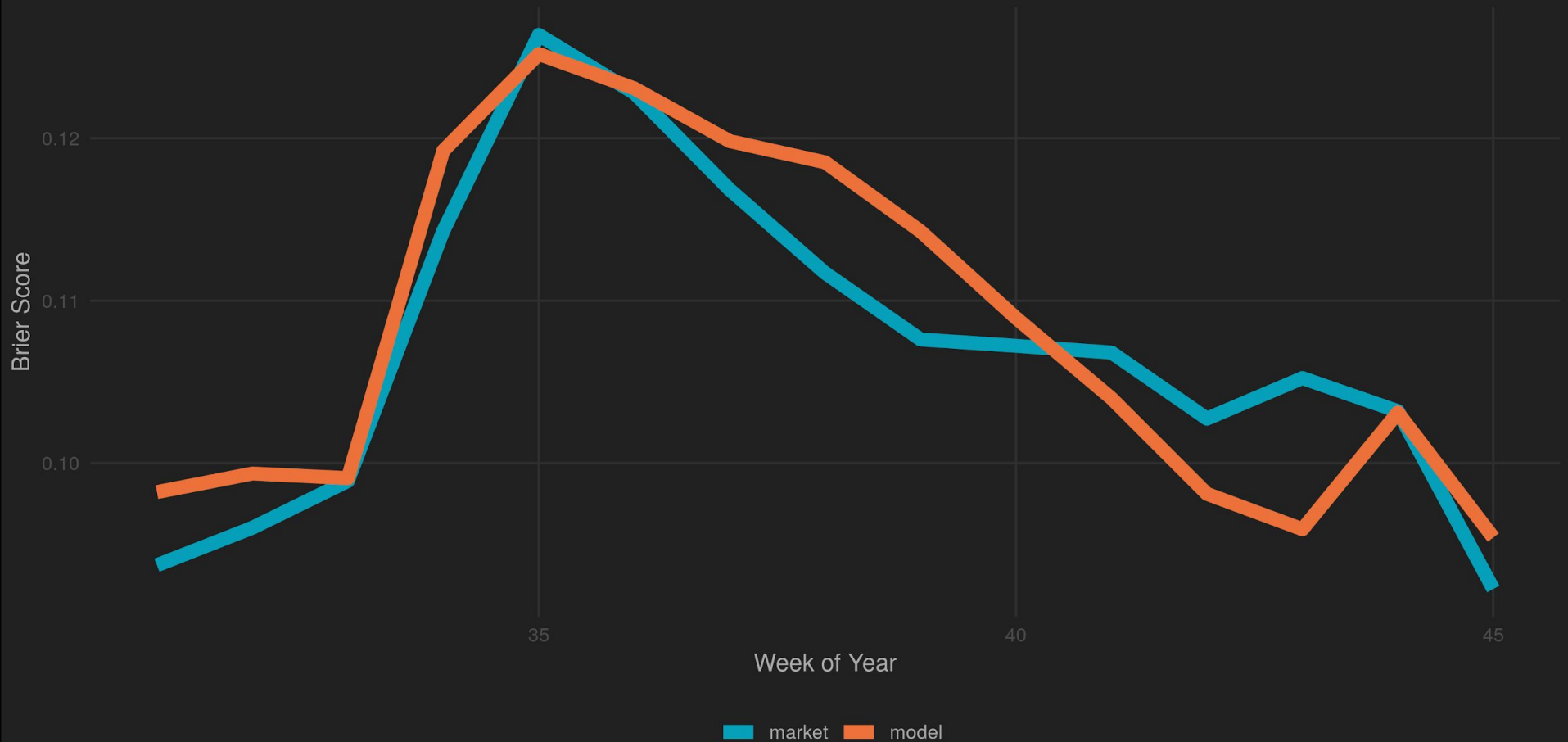
$$BS = \frac{1}{N}\sum_{t=1}^{N}(f_t - o_t)^2$$

Brier T-Test

- Market mean: 0.1084
- Model mean: 0.1091
- Lower bound: -0.005
- Upper bound: 0.0035
- t =   -0.33902
- p-value =   0.7346
- Alternative hypothesis: Difference

Brier Score Predictions by Week
PredictIt Markets and FiveThirtyEight Model

Expected Probabilities and Actual Proportions of Democratic Victory

Expected probabilities binned by rounding to the nearest 10%