Markets and Models

Kiernan Nicholls

Spring, 2019

Why predict elections?

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

How to Predict Elections

1. Opinion Polls
2. Poll Aggregation
3. Forecast Models
4. Prediction Markets

Onion Polling

Ex: *Washington Post/ABC*

- ▶ Simple random sampling
- ▶ Response rates
- ▶ Sample size
- ▶ Statistical bias
- ▶ Partisanship

In 1824, *The Harrisburg Pennsylvanian* had Jackson over Adams, 335 to 169.

Literary Digest starting polling nationally in 1916. Infamously "sampled" 2.3 million readers in 1936 and bias caused them to predict Landon over over Roosevelt.

Polling Aggregation

Ex: *RealClearPolitics.com*

- ▶ 21st century invention
- ▶ Average out all polls
- ▶ Law of large numbers
- ▶ Minimize errors and reduce bias

Forecasting Models

(Forecasting models) take lots of polls, perform various types of adjustments to them, and then blend them with other kinds of empirically useful indicators... to forecast each race. Then they account for the uncertainty in the forecast and simulate the election thousands of times.

Forecasting Models

Most election models work in something like the following way: First, they calculate the most likely outcome in a particular state ("The Republican wins by 1 point") and then they determine the degree of uncertainty around that estimate. Most models do this by means of a normal distribution or something similar to it.

Historic indicators of *greater* uncertainty:

1. The election is further away in time
2. There are fewer polls
3. Those polls disagree more with one another
4. The polling average disagrees more with the state fundamentals
5. There are more undecideds or third-party voters in the polls
6. The race is more lopsided

Model Inputs

1. **Polling:** District level polling, weighted for historical accuracy.
2. **CANTOR:** A proprietary algorithm to identify similar districts to infers results for polling-sparce districts.
3. **Expert forecasts:** Ratings published by the historically accurate experts.
4. **Fundamentals:** Non-polling factors, historically useful factors:
   - Incumbency
   - Partisanship
   - Previous margin
   - Generic ballot
   - Fundraising
   - Scandals

Model Outputs

Table 1: Model Data (299,760 observations with 7 of 12 variables)

Date	State	District	Party		Incumbent	Prob	Share
2018-08-01	AK		1	R	TRUE	0.72	49.35
2018-08-01	AK		1	D	FALSE	0.28	44.11
2018-08-01	AL		1	R	TRUE	1.00	64.90
2018-08-01	AL		1	D	FALSE	0.00	35.10
2018-08-01	AL		2	R	TRUE	0.97	58.23
2018-08-01	AL		2	D	FALSE	0.03	41.77
2018-08-01	AL		3	R	TRUE	1.00	62.27
2018-08-01	AL		3	D	FALSE	0.00	37.73
2018-08-01	AL		4	R	TRUE	1.00	76.32
2018-08-01	AL		4	D	FALSE	0.00	23.68

Prediction Markets

- Exchange-traded binary options markets
- Contract price reflects probability
- Crowd-sourcing information
- Efficient market hypothesis
- Price discovery through equilibrium
- Risk aversion overcomes bias
- Dubious legality in the United States

In 1503, traders bet on Papal successor.

Iowa Election Market founded in 1988.

PredictIt

PredictIt is a unique and exciting real money site that tests your knowledge of political events by letting you trade shares on everything from the outcome of an election to a Supreme Court decision to major world events... PredictIt is run by Victoria University of Wellington, New Zealand, a not-for-profit university, for educational purposes

PredictIt Contracts

- Real money, $850 limit imposed by CFTC
- Elections, Justice, Administration, World
- Futures contracts, executes at time or condition
- Two buyers on either side
- Execute for $1 or $0 based on outcome
- Traders can sell at any time, price change reflects information

PredictIt Markets

- Who will win the 2020 Democratic presidential nomination?
- Will Donald Trump be impeached in his first term?
- Will Congress ratify the USMCA by year-end 2019?
- Will Facebook's Mark Zuckerberg run for president in 2020?
- How many tweets will @realDonaldTrump post from noon Mar. 22 to noon Mar. 29?
- Will Theresa May be prime minister of the United Kingdom on 6/30?

PredicIt Data

Table 2: Market Data (44,711 observations with 6 of 11 variables)

	ID	Ticker	Date	Price	Volume
3480	HEIT.NDSENATE.2018		2017-07-18	0.58	1452
3863	TX05.2018		2018-04-28	0.14	0
4271	PA17.2018		2018-06-09	0.21	0
3531	CURB.FL26.2018		2018-06-12	0.34	0
3767	NH01.2018		2018-07-25	0.25	0
2999	NELSON.FLSENATE.2018		2018-08-05	0.50	122
3886	VA02.2018		2018-08-06	0.31	51
3886	VA02.2018		2018-09-02	0.32	43
3886	VA02.2018		2018-10-10	0.78	272
4304	WI01.2018		2018-10-23	0.23	197

Messy Data

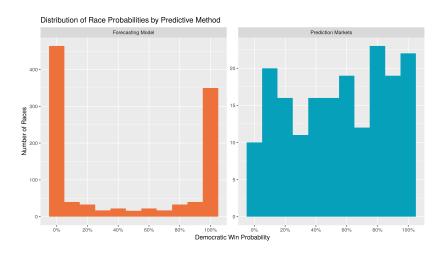Table 3: Messy Combined (9,200 observations of 4 variables)

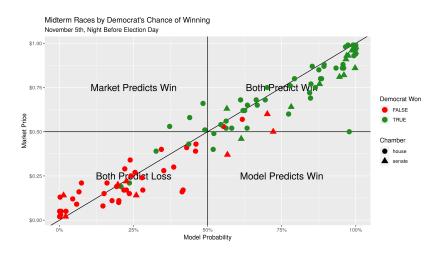| Date	Race	Market Price	Model Probability
2018-08-01	AZ-S1	0.66	0.738
2018-08-01	CA-12	0.91	1.000
2018-08-01	CA-22	0.30	0.049
2018-08-01	CA-25	0.61	0.745
2018-08-01	CA-39	0.61	0.377
2018-08-01	CA-48	0.72	0.666
2018-08-01	CA-49	0.74	0.795
2018-08-01	CA-S1	0.94	1.000
2018-08-01	CO-05	0.06	0.027
2018-08-01	CO-06	0.58	0.648

Tidy Data

Table 4: Tidy Combined (18,400 observations of 4 variables)

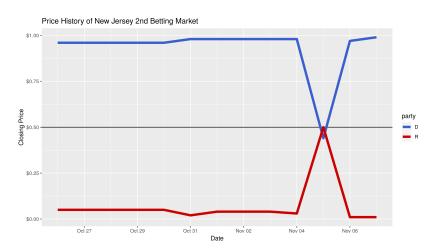| Date	Race	Predictive Method	Probability
2018-08-01	AZ-S1	market	0.660
2018-08-01	AZ-S1	model	0.738
2018-08-01	CA-12	market	0.910
2018-08-01	CA-12	model	1.000
2018-08-01	CA-22	market	0.300
2018-08-01	CA-22	model	0.049
2018-08-01	CA-25	market	0.610
2018-08-01	CA-25	model	0.745
2018-08-01	CA-39	market	0.610
2018-08-01	CA-39	model	0.377

Race Distributions
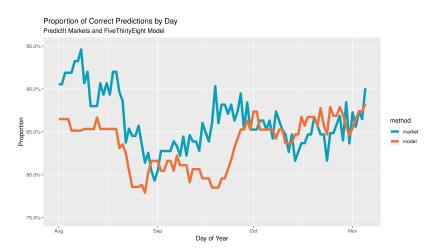

Distribution of Race Probabilities by Predictive Method

Method Similarities

Midterm Races by Democrat's Chance of Winning
November 5th, Night Before Election Day

Market Manipulation?

Price History of New Jersey 2nd Betting Market

Method Accuracy

Proportion of Correct Predictions by Day
PredictIt Markets and FiveThirtyEight Model

Method Accuracy

Proportion of Correct Predictions by Month
PredictIt Markets and FiveThirtyEight Model