Election Forecasting

Mentor: Peter Gao
Mentee: Andy Qin

What is election forecasting?

- The business of predicting the outcome of an election
  - Binary outcome or vote share
- Built on statistical modelling, taking into account a variety of different factors
- Modern election forecasting is relatively young

Motivation: Should we be forecasting elections?

- Goal: to better understand presidential election forecasting and how people interact with it
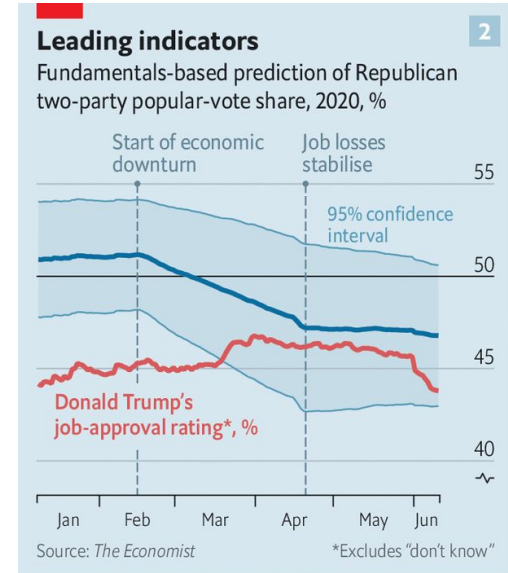
More open ended questions:

- How is election forecasting different from other applications of statistics?
- How are the models built? Can we really trust them?
- Is the information we get from the models really meaningful?

Overview of a Model

- Lots of different approaches to modelling
- Usually balances polling with fundamentals to predict vote share
- **Polling:** Surveys that attempt to capture information about a population within predictable margins of error
- **Fundamentals:** Non-polling factors
  - State of the economy, unemployment, demographics, incumbency, etc

Leading indicators
Fundamentals-based prediction of Republican two-party popular-vote share, 2020, %

The Economist

Example: The Economist Model

- Really complex and not everything is transparent!
- Balances polling averages and aggregation with a wide range of fundamental factors
- Adjusts relative weight of factors as it gets closer to the election date

The Economist

Our final pre-election forecast is that **Joe Biden is very likely to beat Donald Trump** in the electoral college.

|	Chance of winning the electoral college	Chance of winning the most votes	Predicted range of electoral college votes (270 to win)
Joe Biden Democrat	better than 19 in 20 or 97%	better than 19 in 20 or >99%	259-415
Donald Trump Republican	less than 1 in 20 or 3%	less than 1 in 20 or <1%	123-279

The probability of an electoral-college tie is <1%

https://projects.economist.com/us-2020-forecast/president
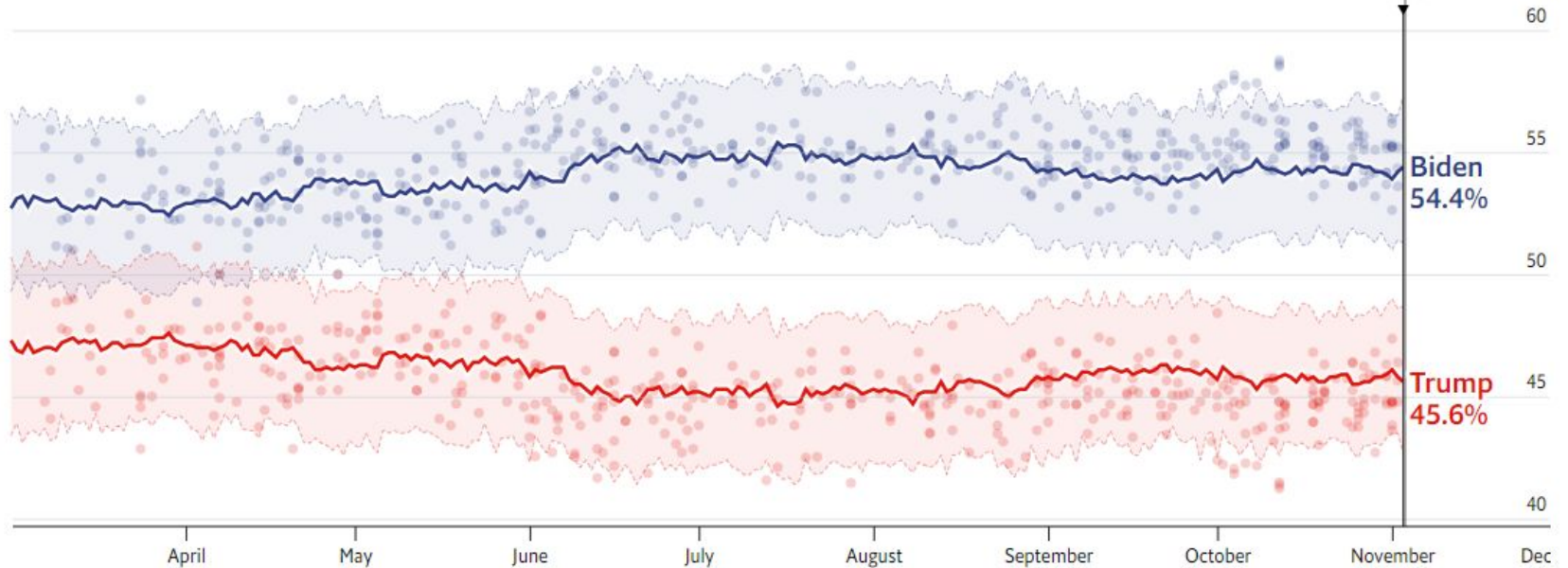
Electoral-college simulations

Our model works by simulating 20,000 paths for the election, each time varying candidates' vote shares to account for polling error, changes in turnout or the political environment and the effects of campaigning. The bars below represent the predicted likelihood of every plausible electoral-vote outcome.

■ Biden wins 97%　■ Tie <1%　■ Trump wins 3%
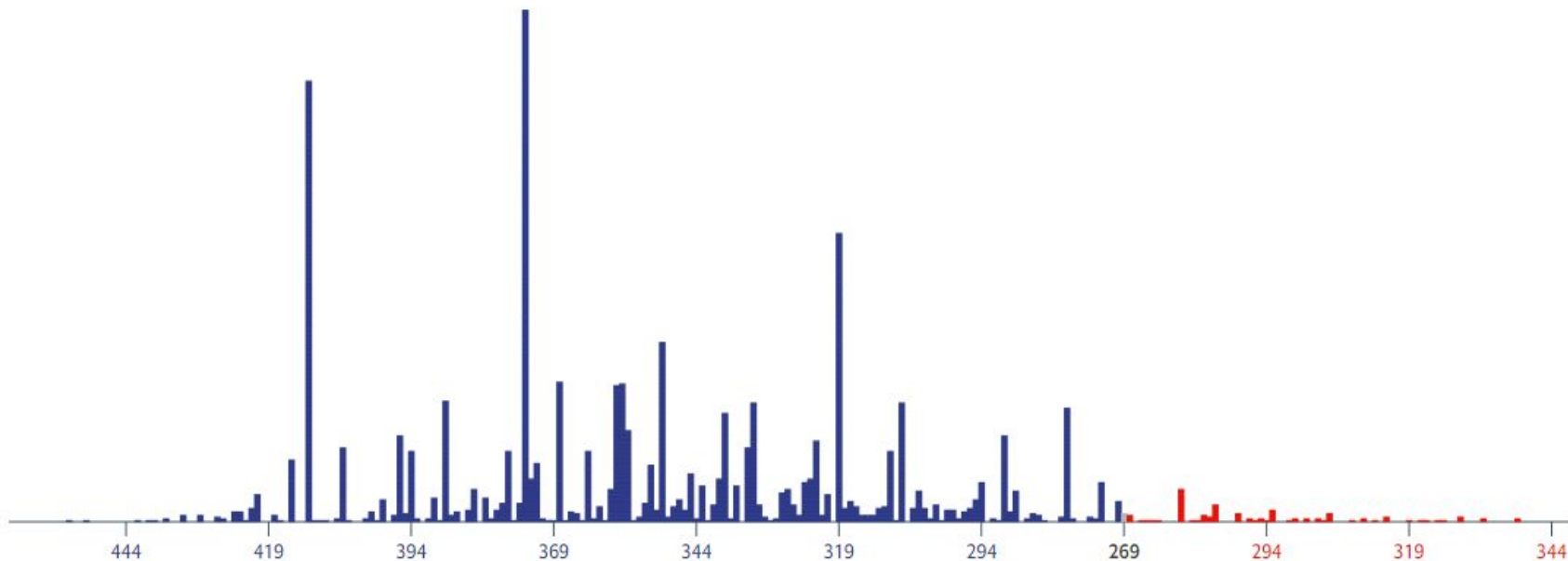
https://projects.economist.com/us-2020-forecast/president

This all seems pretty reasonable… right?

- The Economist, FiveThirtyEight, and other popular models operate relatively similarly
- Balances a lot of different kinds of information
- Attempts to address overfitting and uses resampling techniques
- Simulates multiple different possible outcomes

This all seems pretty reasonable... right?

- The Economist, FiveThirtyEight, and other popular models operate relatively similarly
- Balances a lot of different kinds of information
- Attempts to address overfitting and uses resampling techniques
- Simulates multiple different possible outcomes

Let's be critical!

Questions about fundamentals

- As with any kind of model building, tuning different factors is hard!
- Lots of arbitrary decisions that we're not sure make sense
- What factors are these models choosing to consider?
  - FiveThirtyEight model measures a "news volume index," the length of newspaper headlines
  - Don't account for special circumstances like Covid
  - Both models include a "polarity index"?
- Are these factors really predictive of human behavior?

Polling, polling, polling!

- Polling has changed!
- More difficult than to get a representative sample these days
  - Death of landlines/answering random numbers
  - Voluntary surveys are prone to bias
- Use weighting to address these issues

Polling, polling, polling!

- Polling has changed!
- More difficult than to get a representative sample these days
  - Death of landlines/answering random numbers
  - Voluntary surveys are prone to bias
- Use weighting to address these issues
- … but that introduces its own problems
  - What groups do you weight by? (2016 election)
  - Can disproportionally amplify individuals who are not representative of their groups
- Is it even effective to use widespread polling in this way?

Widespread polling aggregation (FiveThirtyEight)

UPDATED MAY 19, 2020 AT 8:00 AM

FiveThirtyEight's Pollster Ratings

Based on the historical accuracy and methodology of each firm's polls.

Read more   Download the data   See the latest polls

Ratings   Definitions

Search for a pollster

POLLSTER	METHOD	LIVE CALLER WITH CELLPHONES	NCPP/ AAPOR/ ROPER	POLLS ANALYZED	SIMPLE AVERAGE ERROR	RACES CALLED CORRECTLY	ADVANCED +/-	PREDICTIVE +/-	538 GRADE	BANNED BY 538	MEAN-REVERTED BIAS
SurveyUSA	IVR/ online/ live	◑	●	787	4.7	89%	-1.1	-0.8	A		D+0.1
Rasmussen Reports/ Pulse Opinion Research	IVR/ online			722	5.3	78%	+0.2	+0.8	C+		R+1.5
Zogby Interactive/JZ Analytics	Online			473	5.4	77%	+0.4	+0.9	C+		R+0.6
Mason-Dixon Polling & Strategy	Live	●		433	5.1	87%	-0.6	-0.3	B+		R+0.6
Public Policy Polling	IVR/ text			423	5.0	80%	-0.4	+0.1	B		D+0.3
YouGov	Online			416	4.9	88%	-0.2	+0.3	B		D+0.4
Research 2000	Live*			280	5.5	88%	-0.1	+0.3	F	✗	D+1.3
American Research Group	Live	●	●	273	7.4	75%	+0.3	+0.2	B		R+0.2
SurveyMonkey	Online			210	7.1	84%	+2.3	+2.6	D-		D+5.0

The closer to Election Day, the more our model relies on polls

Share of the weight assigned to polls and the "fundamentals," by number of days until the election

DAYS UNTIL ELECTION	POLLS	FUNDAMENTALS
0	100%	0%
5	97	3
10	94	6
25	89	11
50	84	16
75	79	21
100	74	26
150	65	35
200	57	43
250	47	53

"Accurate"?

- There's only one 2020 election
- How can we evaluate whether a given election forecast is effective?
- Considering sample size, does evaluating accuracy even make sense?
- How do different incentives influence how forecasts present themselves?
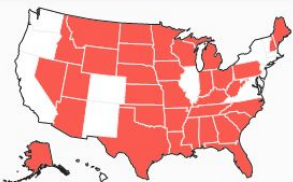   - News organization incentives, competing forecasts

So what do we make of it?

- Election forecasting isn't going away anytime soon! People want it, so it's here!
- Approach it with uncertainty! There are a lot of good reasons to be skeptical and ask questions
- Your individual political contributions matter! Your community matters! Participate! Vote!

Sources

- https://projects.economist.com/us-2020-forecast/president/how-this-works
- https://www.cambridge.org/core/journals/ps-political-science-and-politics/article/political-economy-model-a-blue-wave-forecast-for-2020/5D1FDCC27BA1FB66A915A6F2F5419BCE
- https://fivethirtyeight.com/features/how-fivethirtyeights-2020-presidential-forecast-works-and-whats-different-because-of-covid-19/
- https://www.pewresearch.org/fact-tank/2020/08/05/key-things-to-know-about-election-polling-in-the-united-states/
- http://www.stat.columbia.edu/~gelman/research/published/jdm200907b.pdf

A TRUMP WIN! | A TRUMP WIN! | A BIDEN WIN! | A BIDEN WIN! | A BIDEN WIN! | A BIDEN WIN!

A BIDEN WIN! | A BIDEN WIN! | | | A BIDEN WIN! | A BIDEN WIN!

Biden is *favored* to win the election

UPDATED 4 WEEKS AGO

A BIDEN WIN! | A BIDEN WIN! | A BIDEN WIN! | A BIDEN WIN! | A BIDEN WIN! | A BIDEN WIN!

A BIDEN WIN! | A BIDEN WIN! | A BIDEN WIN! | A BIDEN WIN! | A BIDEN WIN! | A BIDEN WIN!

Hey there! I'm Fivey Fox, and I'm here to show you around. Each of these maps is an example of how things might shake out on Election Day.

The economy is a noisy predictor of presidential success

FiveThirtyEight's economic index as of Election Day, since 1880,* where a score of zero reflects an average economy, a positive score a strong economy and a negative score a weak one

YEAR	ECONOMIC INDEX	YEAR	ECONOMIC INDEX
1880	+1.37	1948	-0.29
1884	-0.18	1952	+0.21
1888	-0.25	1956	+0.07
1892	+0.71	1960	-0.01
1896	-0.15	1964	+0.70
1900	+0.56	1968	+0.23
1904	-0.23	1972	+0.46
1908	-1.03	1976	+0.26
1912	+0.13	1980	-1.71
1916	+0.75	1984	+0.86
1920	-1.52	1988	+0.09
1924	+0.44	1992	-0.29
1928	+0.15	1996	+0.36
1932	-2.34	2000	+0.36
1936	+1.55	2004	+0.01
1940	+0.77	2008	-1.34
1944	+1.01	2012	-0.10
		2016	+0.08

Questions?

The closer to Election Day, the more our model relies on polls

Share of the weight assigned to polls and the "fundamentals," by number of days until the election

DAYS UNTIL ELECTION	POLLS	FUNDAMENTALS
0	100%	0%
5	97	3
10	94	6
25	89	11
50	84	16
75	79	21
100	74	26
150	65	35
200	57	43
250	47	53

Example: The Political Economy Model

$$\text{Incumbent Vote} = \text{Presidential Popularity} + \text{Economic Growth},$$

$$\text{Vote} = 37.50 + .26*\text{Popularity} + 1.18*\text{Growth}$$
$$(15.37) \qquad (4.73) \qquad\qquad (2.25) = \text{t-ratio}$$

Example: The Political Economy Model

$$Incumbent\ Vote = Presidential\ Popularity + Economic\ Growth,$$

- Seems very simple, but what can we learn from this?
- Relies solely on fundamentals
- Intro-stats level: Regression of two variables!
- Small sample size (18 elections!!)
  - Correctly predicts winner for 15/18
- A little silly :(

Year	Popular Two-party Vote for Incumbent Party	Jack Knife Forecast	Forecast Error	Popular Vote Winner Correctly Predicted?
1948	52.4	50.4	2.0	Yes
1952	44.6	46.3	−1.7	Yes
1956	57.8	54.2	3.6	Yes
1960	49.9	52.3	−2.4	No
1964	61.3	60.4	0.9	Yes
1968	49.6	51.8	−2.2	No
1972	61.8	55.8	6.0	Yes
1976	49.0	52.4	−3.4	No
1980	44.7	39.1	5.6	Yes
1984	59.2	55.0	4.2	Yes
1988	53.9	53.2	0.7	Yes
1992	46.6	47.8	−1.3	Yes
1996	54.7	54.7	0.0	Yes
2000	50.0	56.8	−6.7	Yes
2004	51.2	52.9	−1.7	Yes
2008	46.3	46.7	−0.4	Yes
2012	52.0	50.0	2.0	Yes
2016	51.1	51.0	0.1	Yes