Forecasting Elections
Session 10

Professor Devin Caughey

MIT Department of Political Science
17.263: American Elections

Roadmap

News

Fundamentals

Polls

Forecasts

News

The New York Times

2 0 2 0 | **LIVE** Latest Updates   Debate Fact Check   Today's Polls   How to Vote   Paths to 270   Tracking Misinformation

With Cross Talk, Lies and Mockery, Trump Tramples Decorum in Debate With Biden

Interrupting Joe Biden nearly every time he spoke, President Trump made little attempt to reassure swing voters about his leadership. Mr. Biden hit back: "This is so unpresidential."

Aides to President Trump said Wednesday there was nothing to clarify about his debate remarks about the far-right group Proud Boys.

Fundamentals

The Predictability of Elections

▶ The economy

▶ National security

▶ Incumbency

Income Growth and Incumbent Party Performance
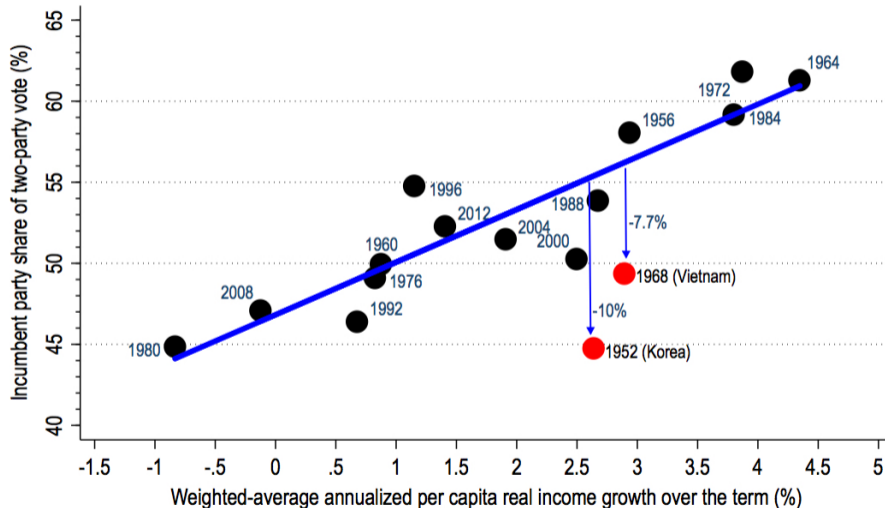
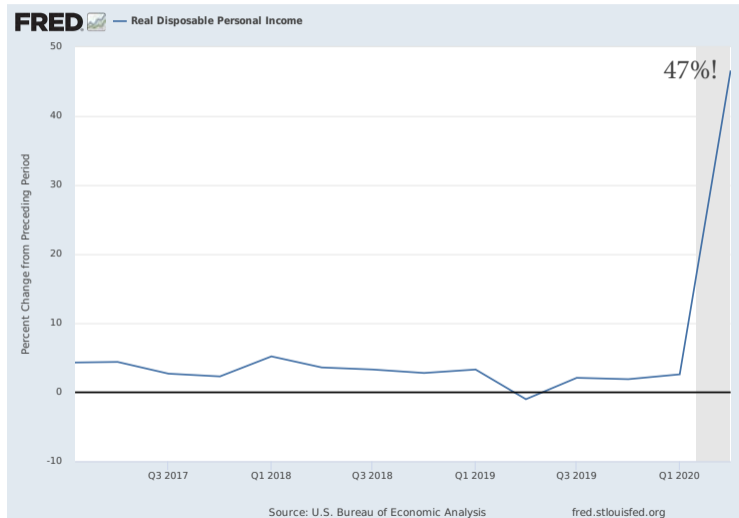# The "Bread and Peace" Prediction for 2016
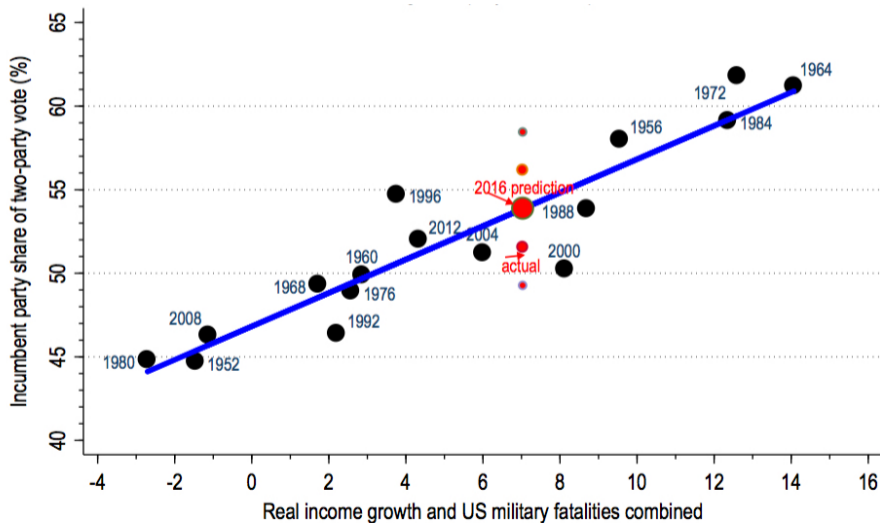
Income Growth

FRED — Real Disposable Personal Income

47%!

Source: U.S. Bureau of Economic Analysis   fred.stlouisfed.org

COVID Relief

Chart 1. Dollar Change in Personal Income and Select Components, United States: 2020 Q1-Q2
(Millions of Dollars)

1,446,296	Personal Income
(COVID relief) 2,450,566	Transfer receipts
-920,386	Net earnings
-83,884	Property income
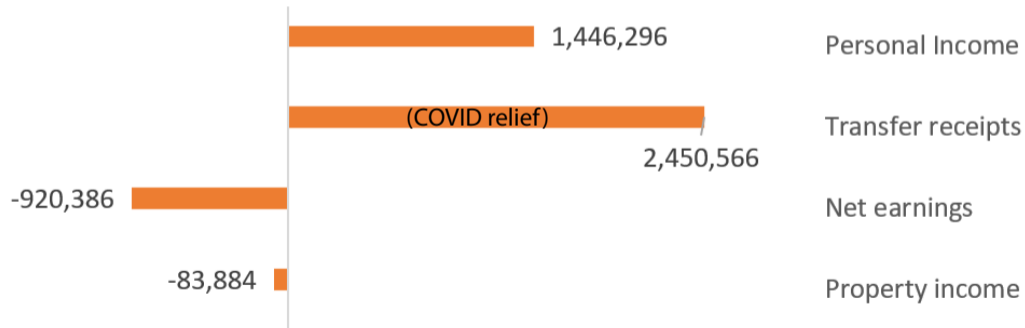
GDP Growth

Real GDP: Percent change from preceding quarter

	2016		2017				2018				2019				2020	
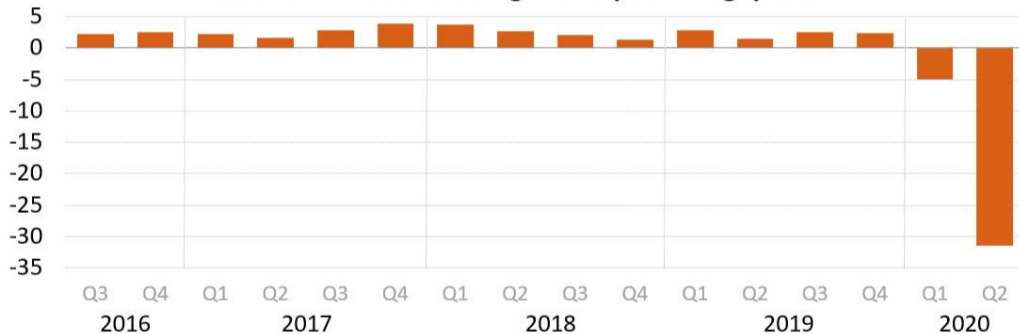
U.S. Bureau of Economic Analysis

Seasonally adjusted at annual rates

Limits of Fundamentals-Only Forecasts

- ▶ Small sample size (and range)
  - ▶ Large sampling error
  - ▶ Easy to over-fit
  - ▶ Out-sample-predictions particularly unreliable
- ▶ Doesn't incorporate all information (bug or feature?)
  - ▶ In particular, polls

Polls

The Evolution of Election Polling

The Quota-Sampling Era (1930s–1940s)

▶ Interviewers sent to locations across the country and instructed to select interview subjects according to specific demographic quotas (e.g., 70% men).

▶ Pollsters such as George Gallup predicted the 1936 election when much larger (but non-representative) samples failed, but had their own failure in 1948.

The Evolution of Election Polling

The Quota-Sampling Era (1930s–1940s)

The Random-Digit-Dialing Era (1950s–1990s)
- ▶ Call randomly sampled phone numbers and interview survey respondents remotely.
- ▶ High phone ownership and response rates $\implies$ representative samples.

The Evolution of Election Polling

The Quota-Sampling Era (1930s–1940s)
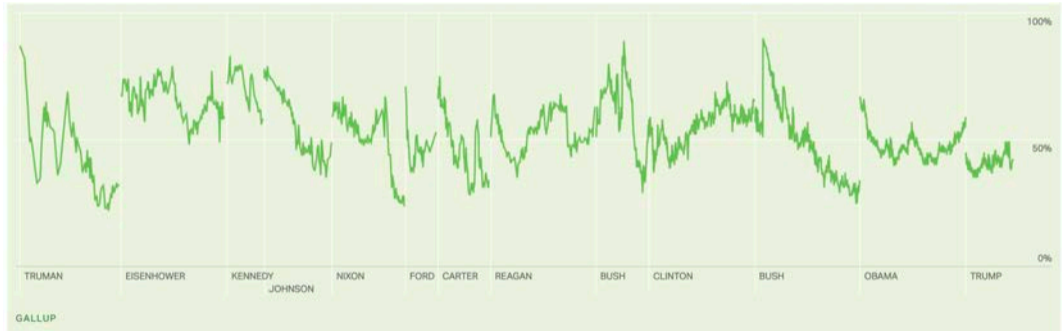
The Random-Digit-Dialing Era (1950s–1990s)

The Cell Phone and Internet Era (2000s)
- ▶ Collapsing telephone response rates ($< 10\%$)
- ▶ Shift to internet-based polls, many sampled using techniques similar to quotas (and heavily weighted).

Presidential Approval

Trial Heats

Poll	Date	Sample	MoE	Biden (D)	Trump (R)	Spread
RCP Average	**9/16 - 9/30**	--	--	**49.7**	**43.1**	**Biden +6.6**
Economist/YouGov	9/27 - 9/30	1350 LV	4.0	50	42	Biden +8
Reuters/Ipsos	9/25 - 9/29	864 LV	3.8	51	42	Biden +9
Rasmussen Reports	9/23 - 9/29	3000 LV	2.0	51	43	Biden +8
Monmouth*	9/24 - 9/27	809 LV	3.5	50	45	Biden +5
JTN/RMG Research*	9/24 - 9/26	752 LV	3.6	51	45	Biden +6
Harvard-Harris	9/22 - 9/25	LV	–	47	45	Biden +2
NY Times/Siena	9/22 - 9/24	950 LV	3.5	49	41	Biden +8
The Hill/HarrisX	9/22 - 9/25	2768 RV	1.9	45	40	Biden +5
ABC News/Wash Post	9/21 - 9/24	739 LV	4.0	54	44	Biden +10
Emerson	9/22 - 9/23	1000 LV	3.0	48	44	Biden +4
USC Dornsife	9/16 - 9/29	5096 LV	–	51	43	Biden +8

Aggregating Polls

Figure: Real Clear Politics poll average
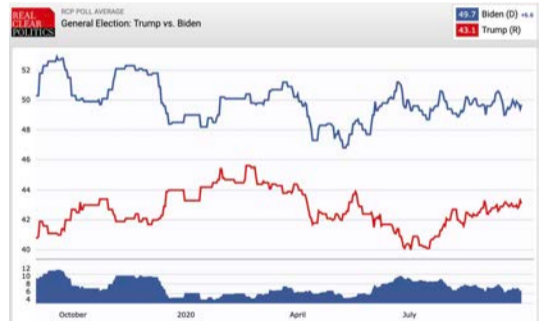
Poll aggregators such as RealClearPolitics and FiveThirtyEight summarize the results of many polls. But how to summarize?

► RCP: Simple (rolling) average

Aggregating Polls

Poll aggregators such as RealClearPolitics and
FiveThirtyEight summarize the results of many
polls. But how to summarize?

▶ RCP: Simple (rolling) average
▶ 538: Local polynomial trend, weighting by
poll quality and sample size and adjusting
for house effects.

Figure: Five Thirty Eight poll average
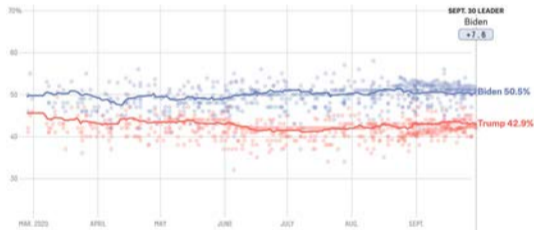
Forecasts

Polls versus Fundamentals

Polls (and poll averages) are not necessarily good forecasts. They are an estimate of where the race stands now (a "nowcast"), not where it is heading.

Sources of error in poll-based forecasts:

▶ Sampling variability (random)

▶ Sampling bias (e.g., undersampling of low-education white men in 2016)

▶ Real opinion change (semi-systematic, e.g., races tend to tighten)

Early in the campaign, fundamentals-based forecasts are often more predictive than polls-based. Polls became more predictive over time, partly because they move toward the fundamentals.
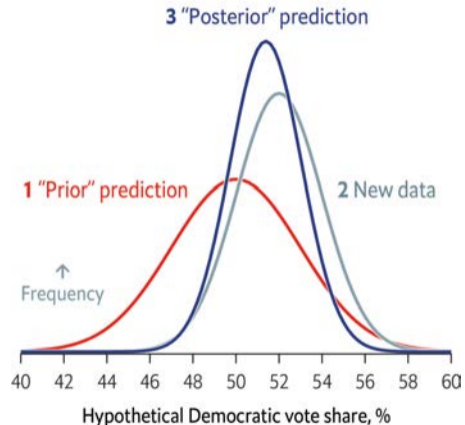
Combining Information

This suggests that we can obtain more accurate predictions by combining information derived from polls with that from the fundamentals.

Bayes' rule:

$$\underbrace{\Pr(O \mid P, F)}_{\text{prediction}} \propto \underbrace{\Pr(O \mid F)}_{\text{fundamentals}} \times \underbrace{\Pr(P \mid O, F)}_{\text{polls}}$$
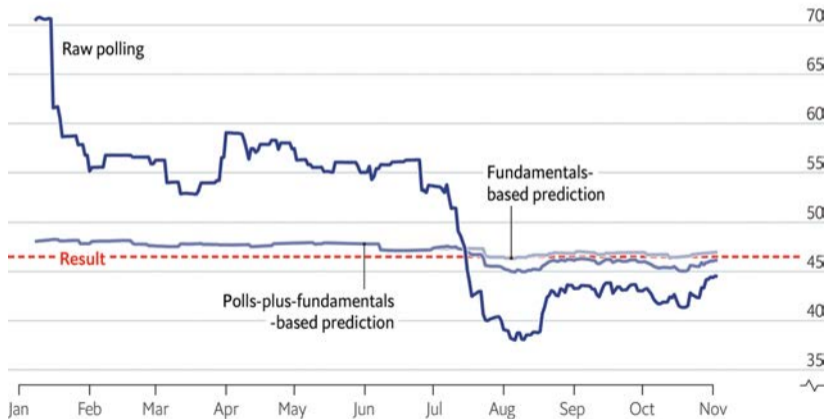

Three steps of Bayesian inference

3 "Posterior" prediction

1 "Prior" prediction

2 New data

↑ Frequency

40 42 44 46 48 50 52 54 56 58 60

Hypothetical Democratic vote share, %

The Economist

Example: Bush vs. Dukakis (1992)

Predictions in 1992

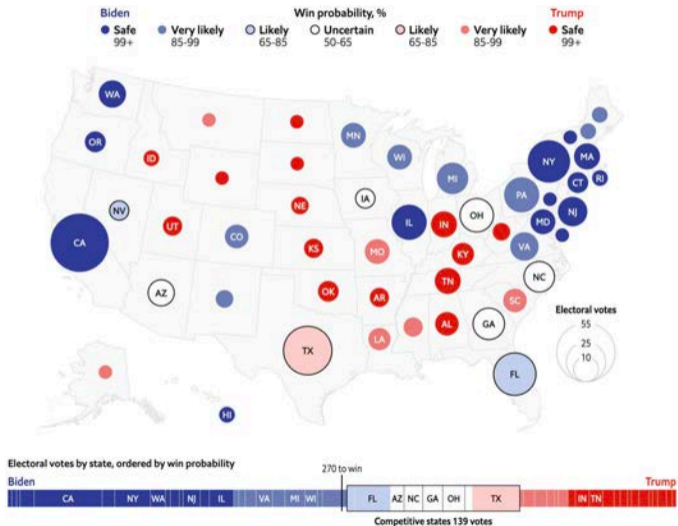Two-party vote share for George H. W. Bush, %

The *Economist*'s Popular Vote Forecast for 2020
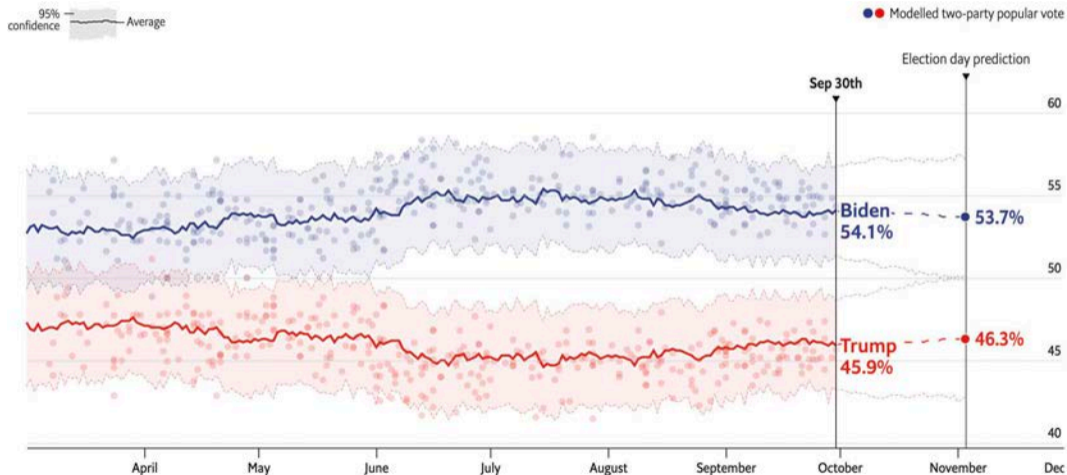
The *Economist*'s Electoral College Forecast

The Problem of Uncertainty

- ▶ Forecasts have uncertainty, and ideally that uncertainty should be well calibrated.
  - ▶ Forecasts with 80% certainty should be wrong 20% of the time.
- ▶ A big challenge with election forecasts is there is not enough data to calibrate uncertainty, especially in the tails of the distribution.
- ▶ States and other lower-level elections add additional information, but it's hard to say how much since forecast errors are likely to be correlated across states.
  - ▶ e.g., Trump's overperformance in Wisconsin AND Michigan AND Pennsylvania

Correlation across States
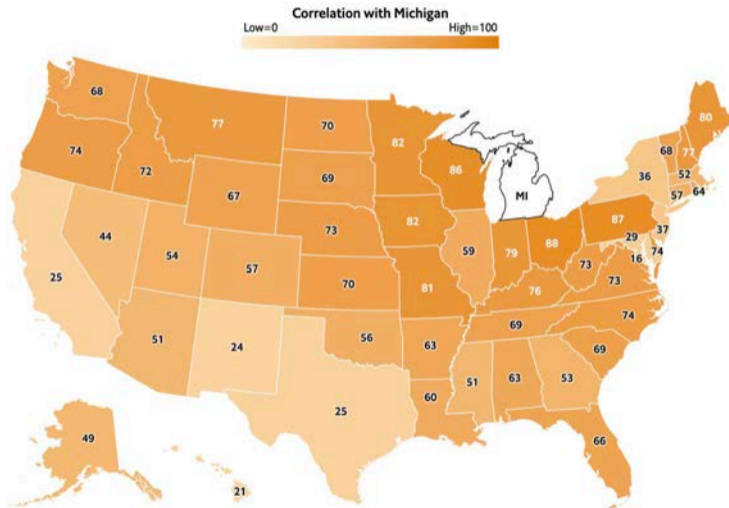
17.263 American Elections
Fall 2020