Re-imagining the 2022 Midterm Elections

A Counterfactual Vote Share Model

Lawrence Lessig and Maia Cook *

The results of any election depend upon which voters turnout. In this paper, we model the 2022 election and ask what the results would have been if people under 40 had voted at the same rate as people 65 years and older. 35% of voters under 40 participated in the 2022 election, while 72% of voters 65 and older participated. Had younger voters participated at the same rate, we find there would have been a substantial change in the make-up of the House of Representatives and an uncertain change in the Senate.

1 Data

To estimate the results, we built a model that drew on data provided by Catalist. The dataset included the turnout rates, registration rates, and partisan vote share (liberal, moderate[1], and conservative) in all 435 congressional districts. Each variable was divided into five age groups: *18-28, 29-39, 40-50, 51-64*, and *65+*. Catalist drew its data from state voter files from the 2022 election.

From these data, we calculated a handful of new variables: $Available\ Votes_{i,j}$, which is the difference between the total number of people that registered and the total number of people that ultimately voted, i denoting congressional district, and j denoting age group; $Participation\ Rate_{i,j}$, which is the quotient of the total number of people who voted and total registered; $New\ Votes_{i,j}$ which calculates how many new votes would have been introduced had voters under the age of 40 voted at the same rate as those in the 65+ group.

$$\textbf{\textit{Available Votes}}_{i,j} = Registered_{i,j} - Total\ Voted_{i,j}$$

$$\textbf{\textit{Participation Rate}}_{i,j} = Total\ Voted_{i,j} / Registered_{i,j}$$

$$\textbf{\textit{New Votes}}_{i,j} = (Participation\_Rate_{i,65+} * Registered_{i,j}) - Voted_{i,j}$$

$$\begin{cases} \textbf{\textit{New DEM Votes}}_{i,j} = New\ Votes_{i,j} * (DEM\ Votes_{i,j} / Total\ Voted_{i,j}) \\ \textbf{\textit{New GOP Votes}}_{i,j} = New\ Votes_{i,j} * (GOP\ Votes_{i,j} / Total\ Voted_{i,j}) \\ \textbf{\textit{New MID Votes}}_{i,j} = New\ Votes_{i,j} * (MID\ Votes_{i,j} / Total\ Voted_{i,j}) \end{cases}$$

*Lawrence Lessig, Roy L. Furman Professor of Law and Leadership Harvard Law School;
Maia Cook, Ethics, Politics, & Economics at Yale University.

[1]Catalist defines "moderate" voters as cross-pressure, 3rd-party, and Independent voters. Catalist also placed voters in this category if there was insufficient data to make a determination about how they voted.

2 Building the House Model

Using these variables, we built an equation to calculate the new vote shares between the Republican and Democratic candidates, assuming people under 40 had voted at the same rate as those 65 and older. To do so, we made three important assumptions. First, we excluded unchallenged,[2] independent-won,[3] and ranked-choice races[4] from this analysis. That removed 33 congressional districts. Second, we assumed new Democratic and Republican voters would have voted with their party.

Third, we ran two scenarios for dividing "moderate" voters into a binary ballot. First, we tested a model that assumed moderate voters would have divided according to the split between Republicans and Democrats in each district. For instance, if 55% of the total Republican and Democratic vote in a district voted Republican, we allocated 55% of the new moderate voters for Republicans. Second, we tested a model that assumed that moderate voters would vote for Democrats at a fixed rate. In the notation below, this coefficient is denoted by γ.

$$\forall\, i \notin \{\text{RCV elections},\ \text{Unchallenged elections},\ \text{Independent-won elections}\},$$

When γ is a fixed rate:

$$\textbf{\textit{Model GOP Votes}}_i = (\textit{Total GOP Votes}_i) + (\textit{New GOP Votes}_{i,18-28}) + (\textit{New GOP Votes}_{i,29-39}) + (1-\gamma)\cdot(\textit{New MID Votes}_{i,18-28} + \textit{New MID Votes}_{i,29-39})$$

$$\textbf{\textit{Model DEM Votes}}_i = (\textit{Total DEM Votes}_i) + (\textit{New DEM Votes}_{i,18-28}) + (\textit{New DEM Votes}_{i,29-39}) + \gamma(\textit{New MID Votes}_{i,18-28} + \textit{New MID Votes}_{i,29-39})$$

When γ is proportional:

$$\textbf{\textit{Model GOP Votes}}_i = (\textit{Total GOP Votes}_i) + (\textit{New GOP Votes}_{i,18-28}) + (\textit{New GOP Votes}_{i,29-39}) + (\textit{New MID Votes}_{i,18-28}\cdot \%\ \textit{New GOP Votes}_{i,18-28}) + (\textit{New MID Votes}_{i,29-39}\cdot \%\ \textit{New GOP Votes}_{i,29-39})$$

$$\textbf{\textit{Model DEM Votes}}_i = (\textit{Total DEM Votes}_i) + (\textit{New DEM Votes}_{i,18-28}) + (\textit{New DEM Votes}_{i,29-39}) + (\textit{New MID Votes}_{i,18-28}\cdot \%\ \textit{New DEM Votes}_{i,18-28}) + (\textit{New MID Votes}_{i,29-39}\cdot \%\ \textit{New DEM Votes}_{i,29-39})$$

To identify what party would have won under this model, we created a binary dummy variable called *Modeled Results*. If the total modeled Democrat votes for a district was larger than the total modeled Republican votes, then the dummy variable was assigned a value of 1.

$$\textbf{\textit{Modeled Results}}_i = \begin{cases} 1 & \text{if } \textit{Model DEM Votes}_i > \textit{Model GOP Votes}_i \\ 0 & \text{otherwise} \end{cases}$$

[2]Unchallenged: AL-1, AL-6, FL-5, FL-6, FL-18, LA-4, LA-6, PA-13, PA-14, SC-3, SC-4, SD-1, TX-6, TX-11, TX-25, TX-26, TX-31, WI-6, CA-10, CA-15, CA-16, CA-29, CA-30, CA-34, CA-37, IL-7, MA-4, NY-13, PA-3.
[3]Independent-Won: ND-1, TX-19.
[4]Ranked Choice Vote: AK-1, ME-2.

3 Building the Senate Model

The counterfactual Senate model followed the same pattern as the House. We created the aggregate Senate totals by summing the state congressional districts for the 33 states holding elections for the Senate. Because Oklahoma elected two senators in 2022, we modeled 34 elections. A list of the states with senators up for re-election can be found in the table below.

Table 1: **States with Senate Elections in 2022 Midterms**

AL	AR	AZ	CA	CO	CT	FL	GA	HI
IA	ID	IL	IN	KS	KY	LA	MD	MO
NC	ND	NH	NV	NY	OH	OK (2)	OR	PA
SC	SD	VT	WA	WI				

We used the same equations as the House model to calculate $Available\ Votes_{i,j}$, $Participation\ Rate_{i,j}$, and $New\ Votes_{i,j}$ — only now the subscript i denotes "state," not "congressional district." We retained the three assumptions from the House model to calculate the vote shares for the Senate.

We recreated a binary dummy variable $Modeled\ Results_i$ for our Senate model. Our findings are discussed on the next page.

(See page below for results)

4 Results

Under both assumptions for how new "moderate" votes would be divided, Democrats would have achieved substantial gains in the House. If $\gamma = 0.80$, House Democrats would have gained 20 seats. If new moderate votes are proportional, House Democrats would have won 17 seats. In the Senate, had new "moderate" votes broken Democratic ($\gamma = 0.80$), Republicans would have lost 2 seats; if the vote had split proportionally, the balance in the Senate would not have changed.

Table 2: **Modeled Election Results when γ = 0.80**

	Observed		**Modeled**	
	House	Senate	House	Senate
GOP	222	49	202	47
DEM	213	51	233	53

**Ohio and Missouri's Senate election would have gone to Democrats*

Table 3: **20 Districts with seat changes in House model when γ = 0.80**

AZ-1	AZ-6	CA-13	CA-22	CA-27	CA-41	CA-45
CO-3	IA-1	IA-3	MI-10	MT-1	NE-2	NJ-7
NY-17	NY-19	OR-5	TX-15	VA-2	WI-3	

Table 4: **Modeled Election Results when γ = Proportional**

	Observed		**Modeled**	
	House	Senate	House	Senate
GOP	222	49	205	49
DEM	213	51	230	51

**No Senate election outcomes would have changed.*

Table 5: **17 Districts with seat changes in House model when γ = Proportional**

AZ-1	AZ-6	CA-13	CA-22	CA-27	CA-45
IA-3	MI-10	MT-1	NE-2	NJ-7	NY-17
NY-19	OR-5	TX-15	VA-2	WI-3	

5 Discussion

The substantial difference in participation between voters of different age brackets is well known. Our model attempts to estimate the consequence of that difference in a particular election cycle. Assuming the participation rates were equivalent for voters under 40 and 65 and older, the resulting Congress would have been substantially different. That different Congress, with a Democratic President, could have passed a wide range of legislation that did not get passed under Republican control of the House.

This difference in legislation would have been particularly salient for younger voters. Much of the Democrats' agenda would have substantially benefited younger voters—including legislation to restore the right to choose, climate change legislation, student debt relief, increasing the minimum wage, and lowering drug prices. We conclude that younger voters could secure substantially different legislative outcomes that would benefit their age demographic if they participated in the election at rates similar to those of older voters.