People have used Time Series to predict a lot of things!

Stock Market Contigion

Store Inventory

Police Shootings

Electricity

Asset Price and Voltility

Sector Rotation

Rainfall and Temprature

Champagne Sales

But one thing dictates all of them!

Who is the top ~~dog~~ cat?



That’s what we want to predict!

US Democracy



The United States electoral cycle has 3 major elections:

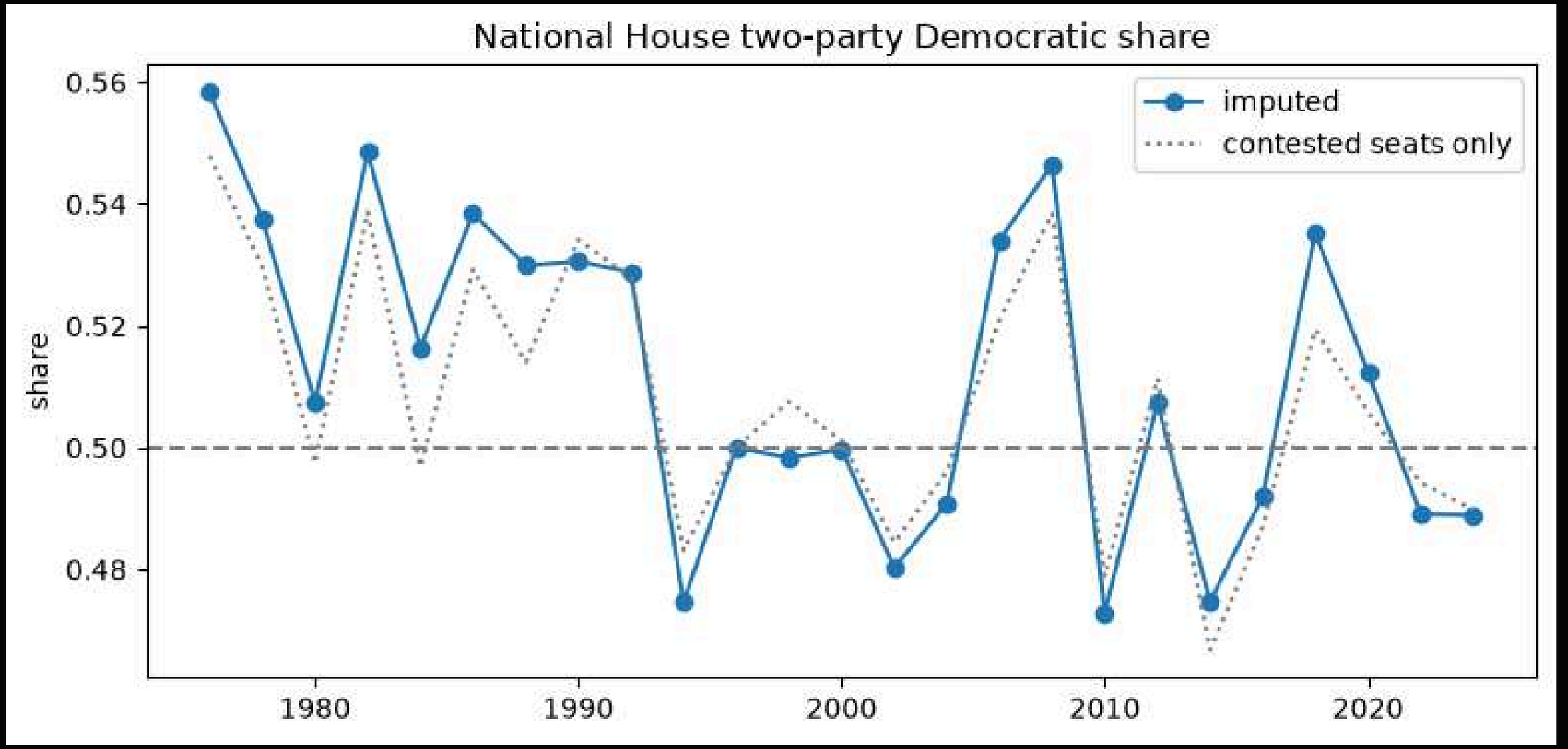
- Presedential Elections - every 4 yrs
- House Elections - every 2 yrs (all 435 seats)
- Senete Elections - every 2 yrs (about 1/3 house)

We were able to obtain statewise data on the same from **MIT Election + Data Sciece Lab (MEDSL).**

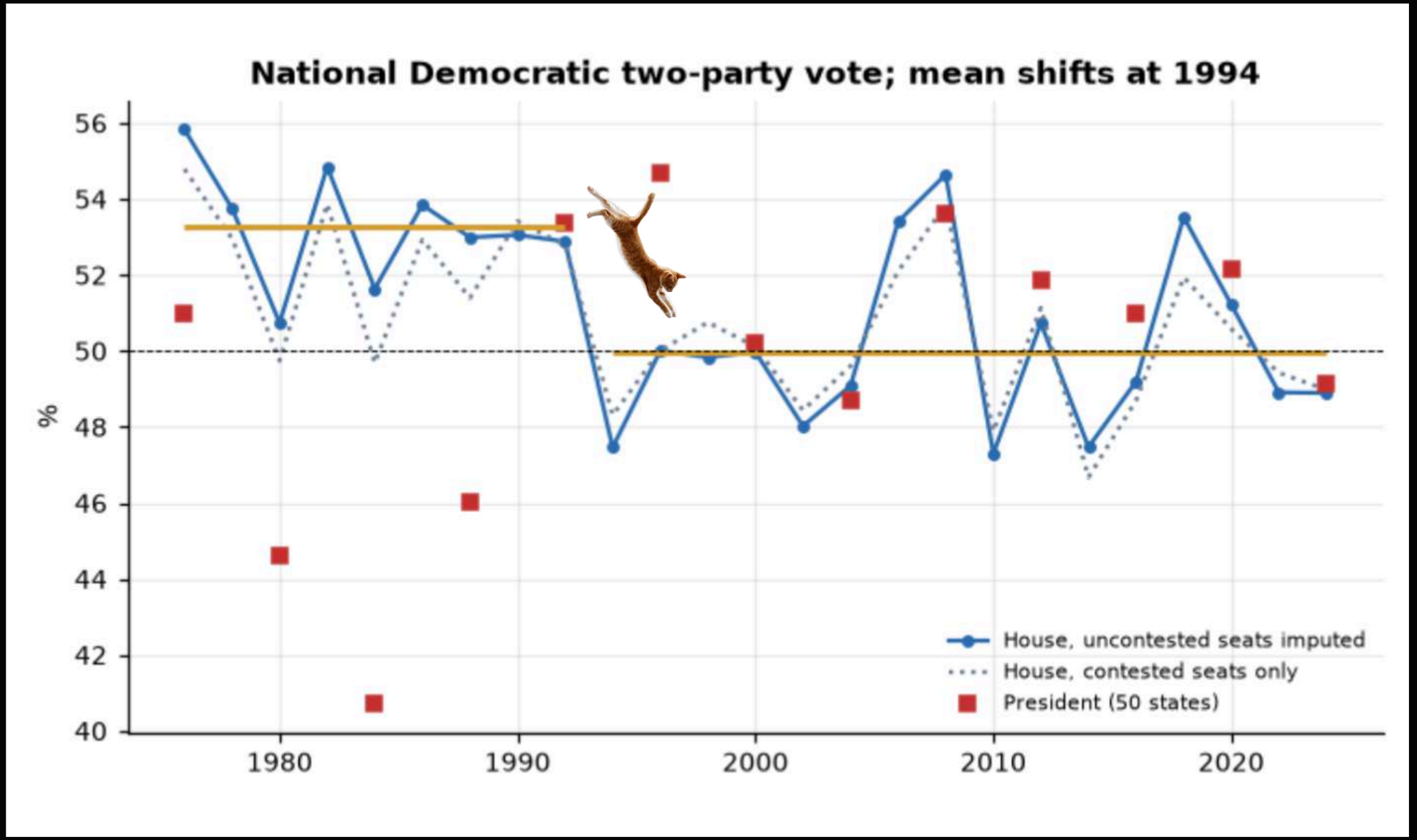


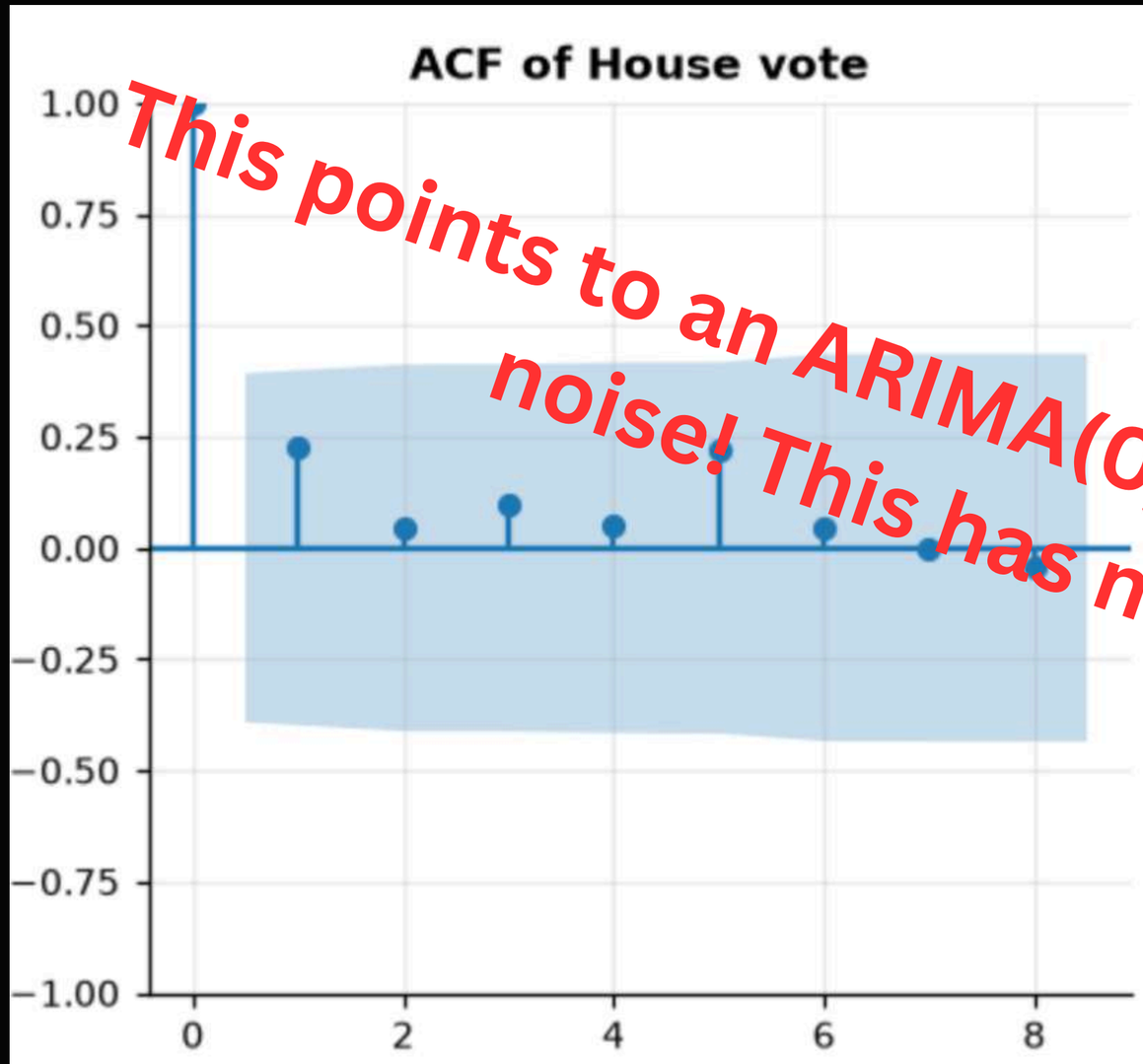
There are also some special elections which MEDSL included with the Senete Elections.

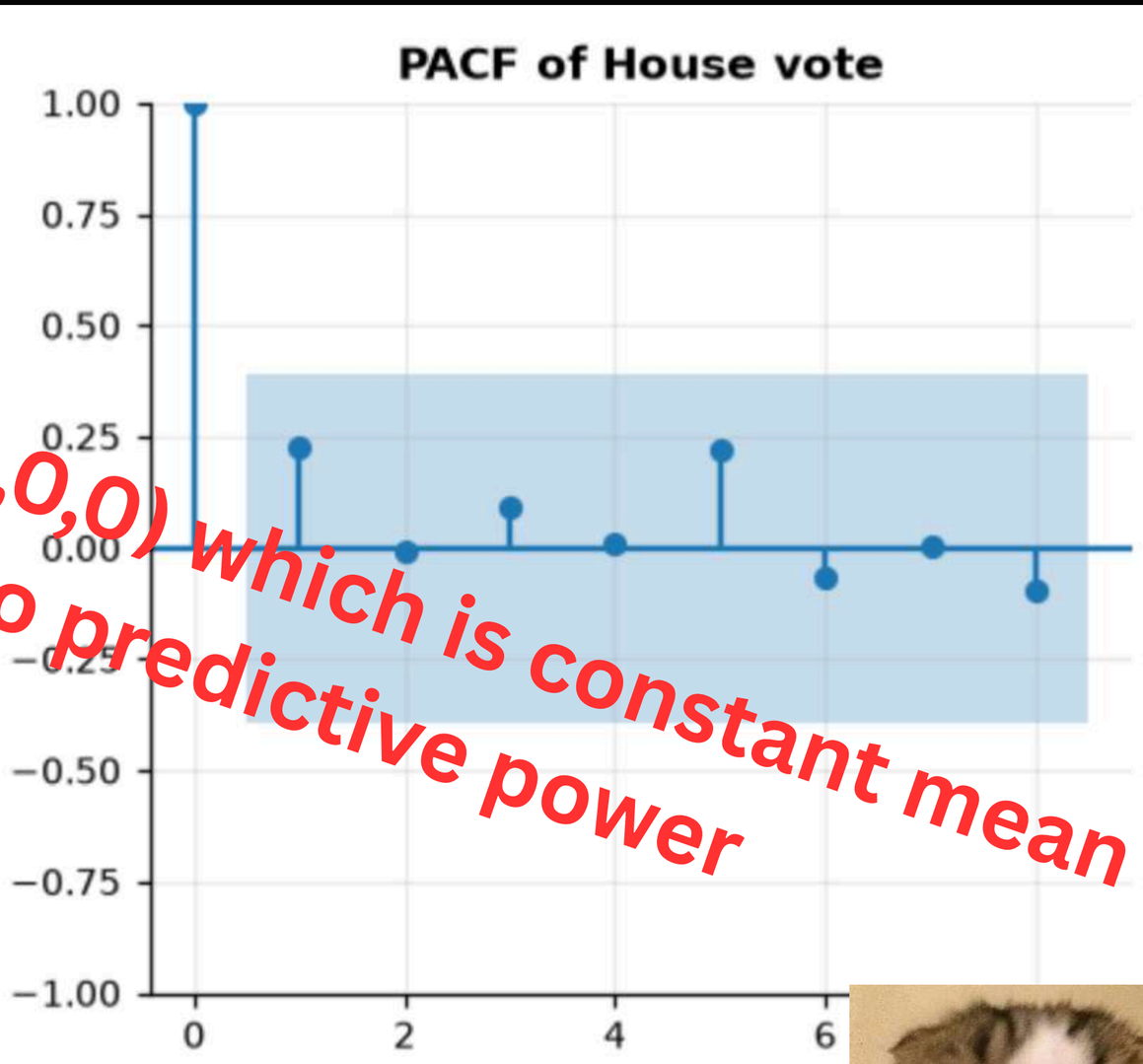
We have not considered Mayoral elections as the voting method in some cities is not plurality/FPP (which is the method in above 3).



Notice something strange?


National Democratic two-party vote; mean shifts at 1994
%
56
54
52
50
48
46
44
42
40
1980
1990
2000
2010
2020
House, uncontested seats imputed
House, contested seats only
President (50 states)





This points to an ARIMA(0,0,0) which is constant mean + noise! This has no predictive power

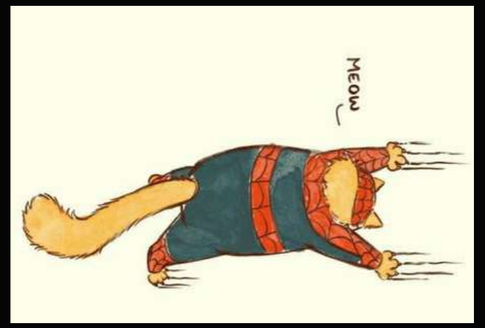
```
ADF  statistic −3.916, p = 0.0019   (H0: unit root)
KPSS statistic 0.559, p = 0.0284    (H0: stationary)
```



Spatial Catalysts
(aka SpaCats)



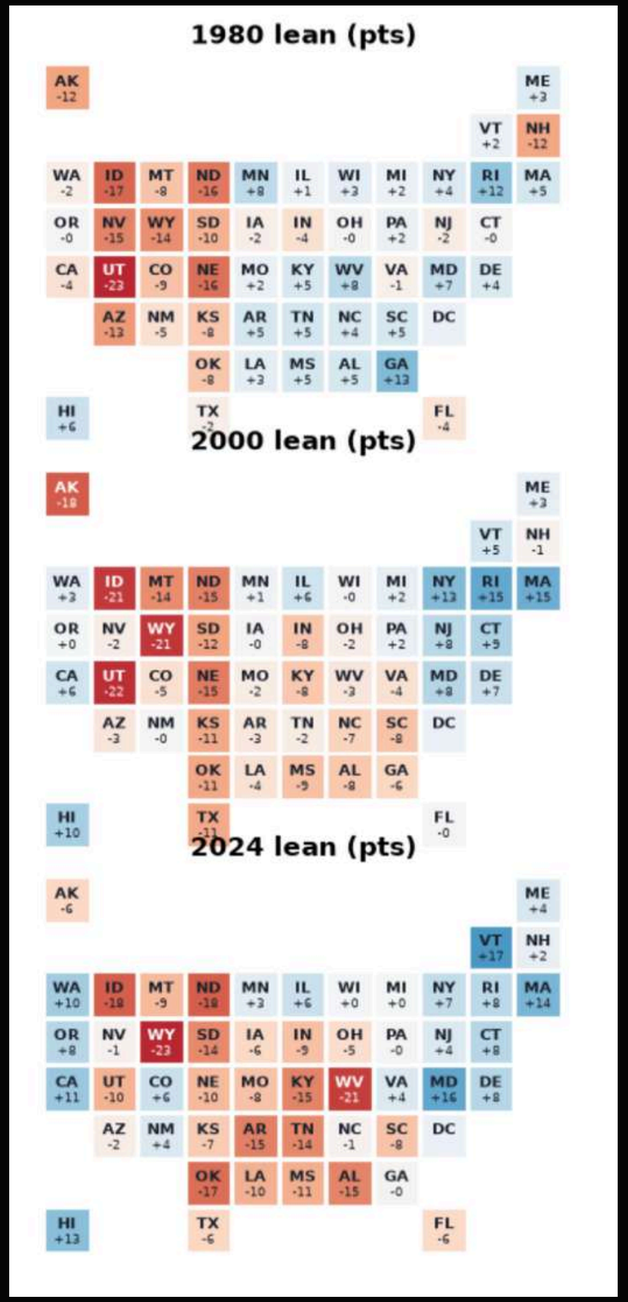






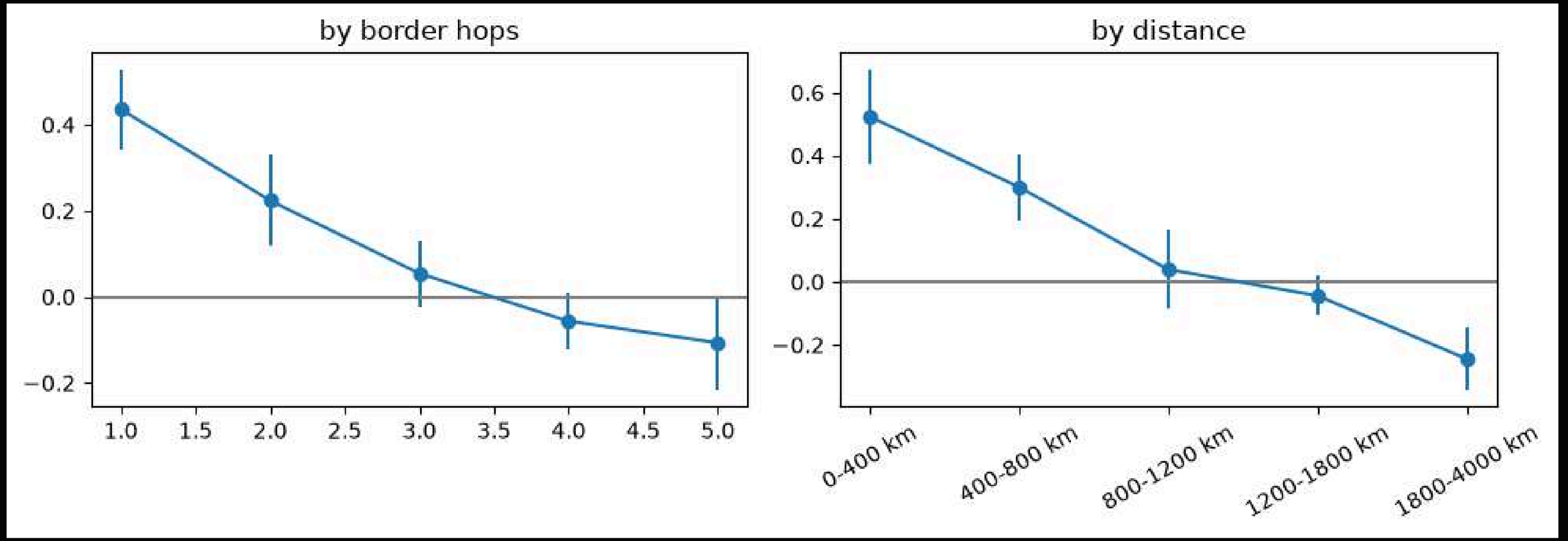
Time Series Analysis project by

Arjun Maneesh Agarwal **Urjaswi Chakraborty**



Can you notice something here?

Neighbouring states seem to be voting similar to each other!



We can quantify this by Moran's I test!



Global Moran's I [edit]

Global Moran's I is a measure of the overall clustering of the spatial data. It is defined as

$$I = \frac{N}{S_0} \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i=1}^{N} (x_i - \bar{x})^2}$$

where

- N is the number of spatial units indexed by i and j;
- x is the variable of interest;
- $\bar{x}$ is the mean of x;
- w_{ij} are the elements of a matrix of spatial weights with zeroes on the diagonal (i.e., $w_{ii} = 0$);
- and S_0 is the sum of all w_{ij}, $S_0 = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}$.[3][4]



This tells us that if we can forcast some states, we can forcast the election!



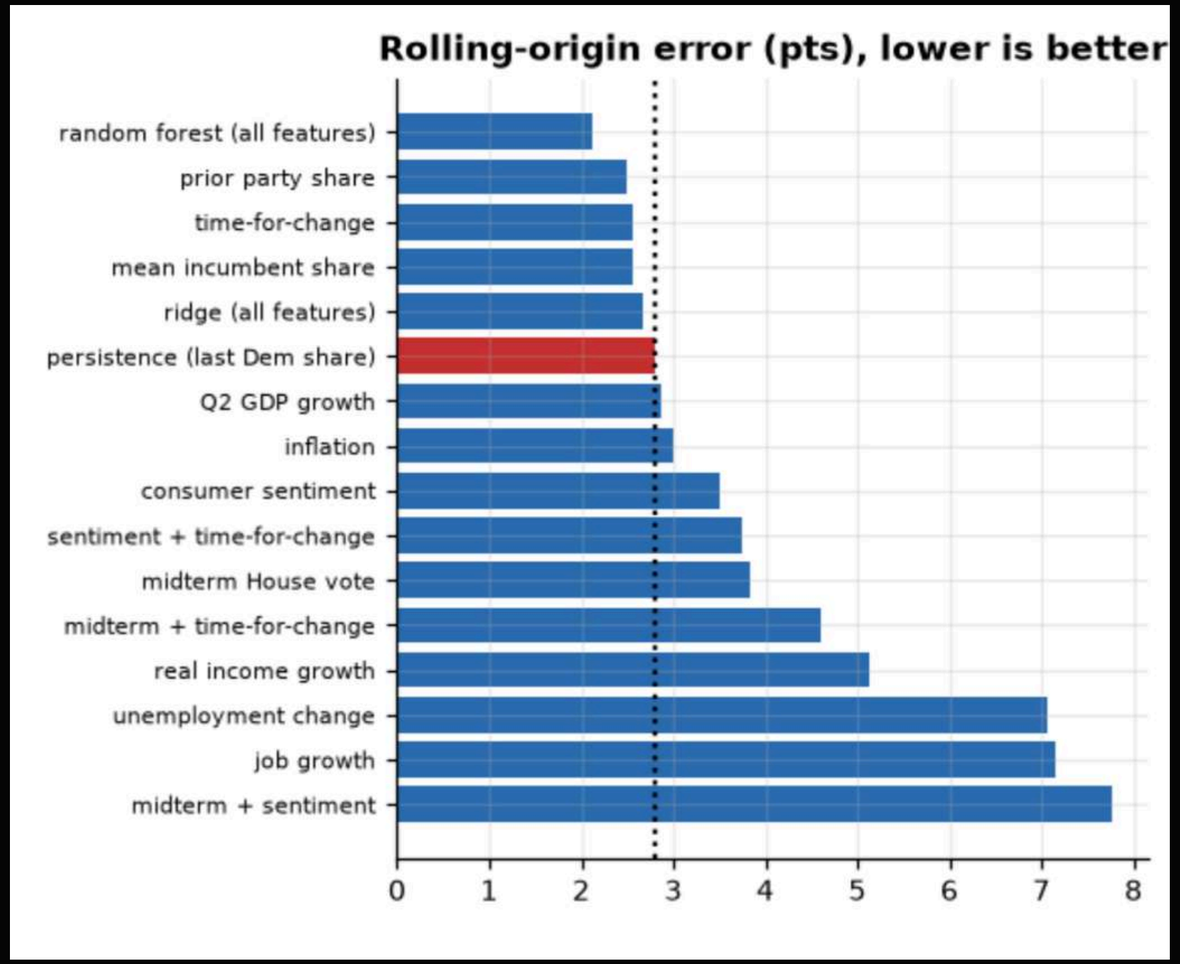
Other factors!

As we noted on our first slide, your votes effect everything around you!

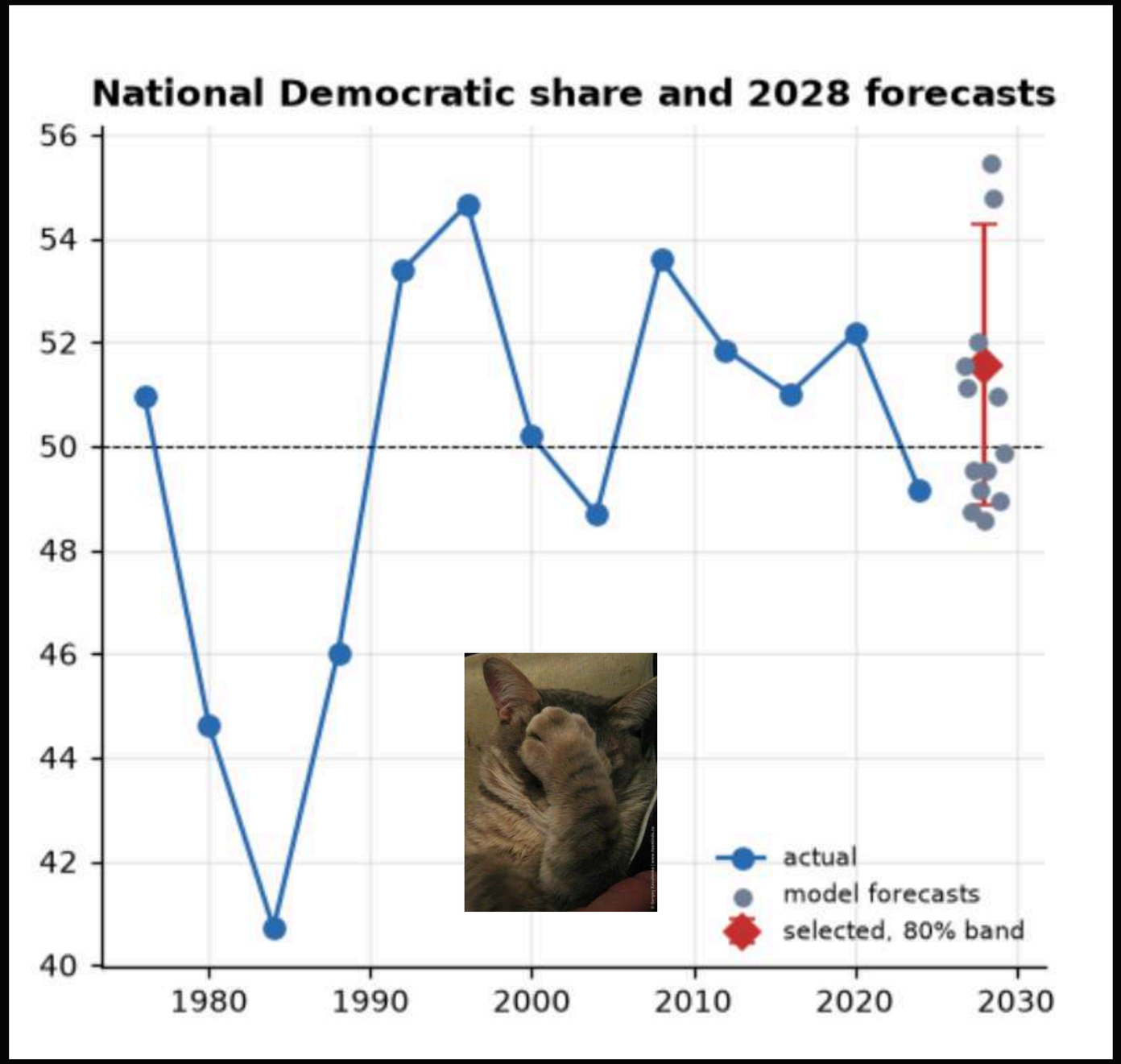
But so does everything else.

Federal Reserve Bank of St. Louis (FRED) allows us to get data about

- GDP
- Unemployement
- Income
- Education
- Consumer Sentiment
- Inflation



For this, we feeded data year by year and tried to predict the comming year. There is no data leakage!



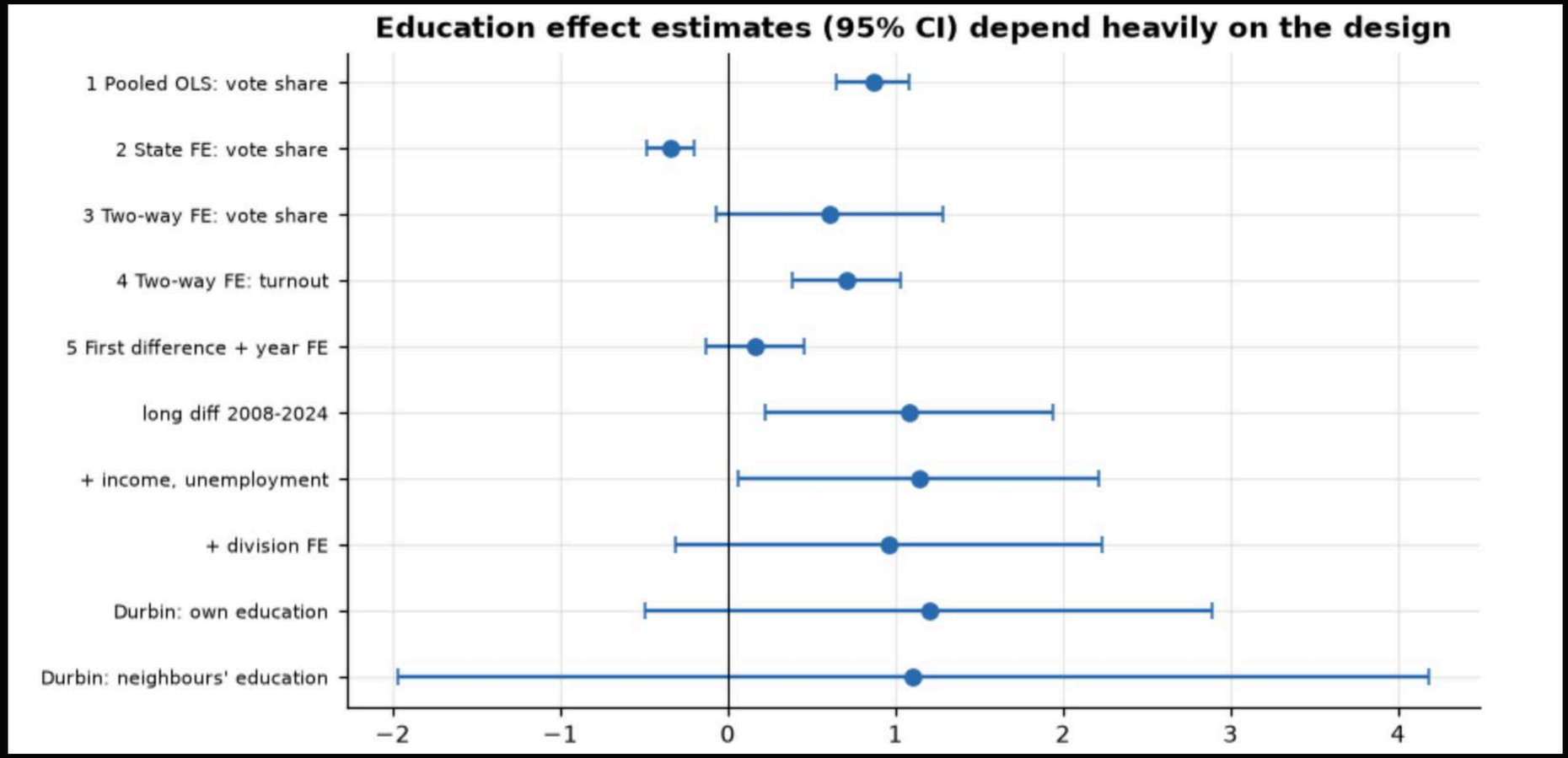
This is the issue with election forcasting!

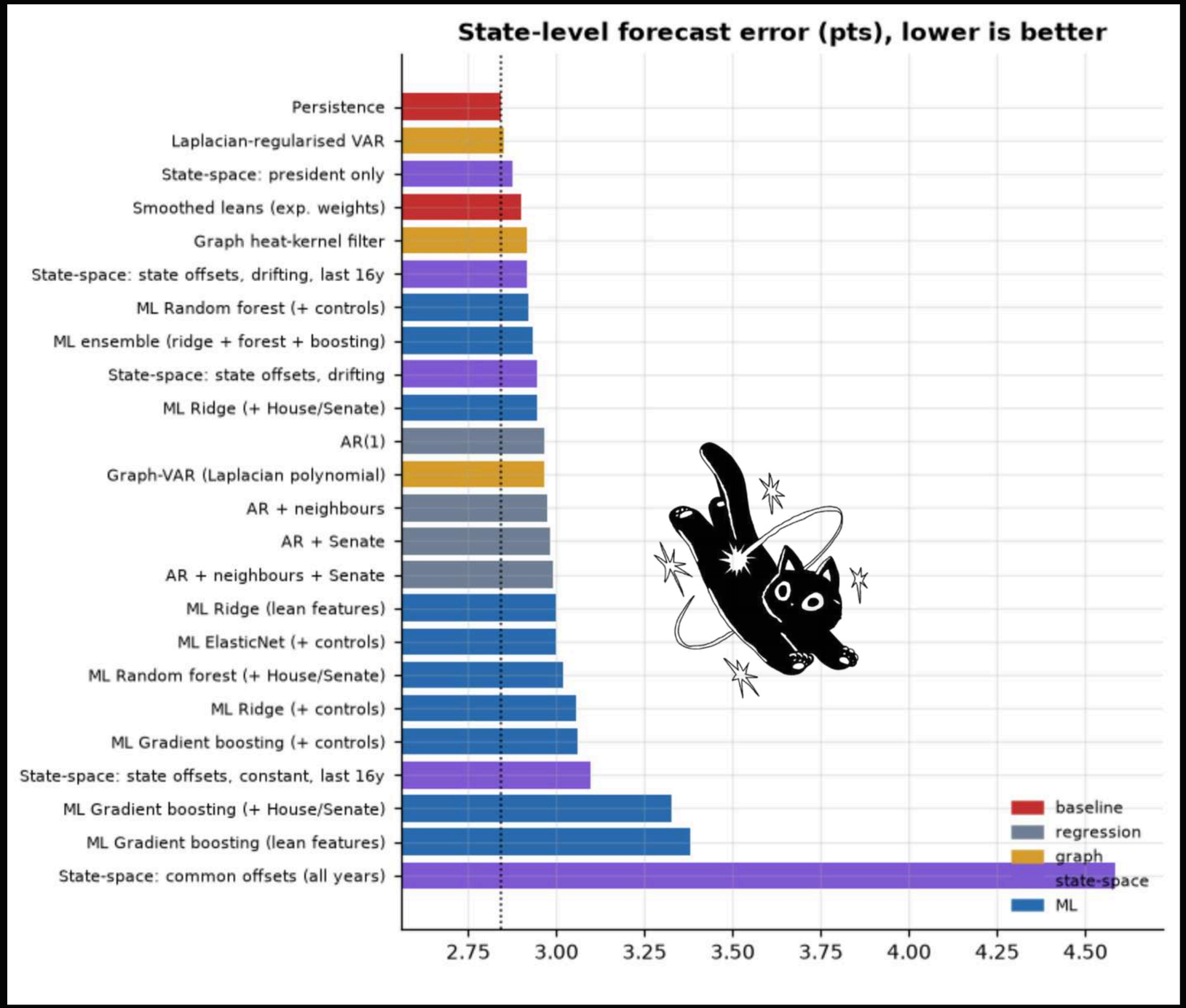
A good model just doesn’t cut it when 2% is all it takes to flip the house!

Furthermore, ML models are unable to pindown what matters!

For example, it is well established that education is corelated with dem vote. However, our model may not use it!







Furthermore, at the state level, ML starts to collapse!

So much ado about nothing?

Not really! Moran’s I told us that states vote similar to each other. This **does not** imply states affect each other’s vote (they indeed don’t).

Coefficient on neighbours' last lean, beyond own persistence (lean_t = a + b lean_(t−1) + g W lean_(t−1)):

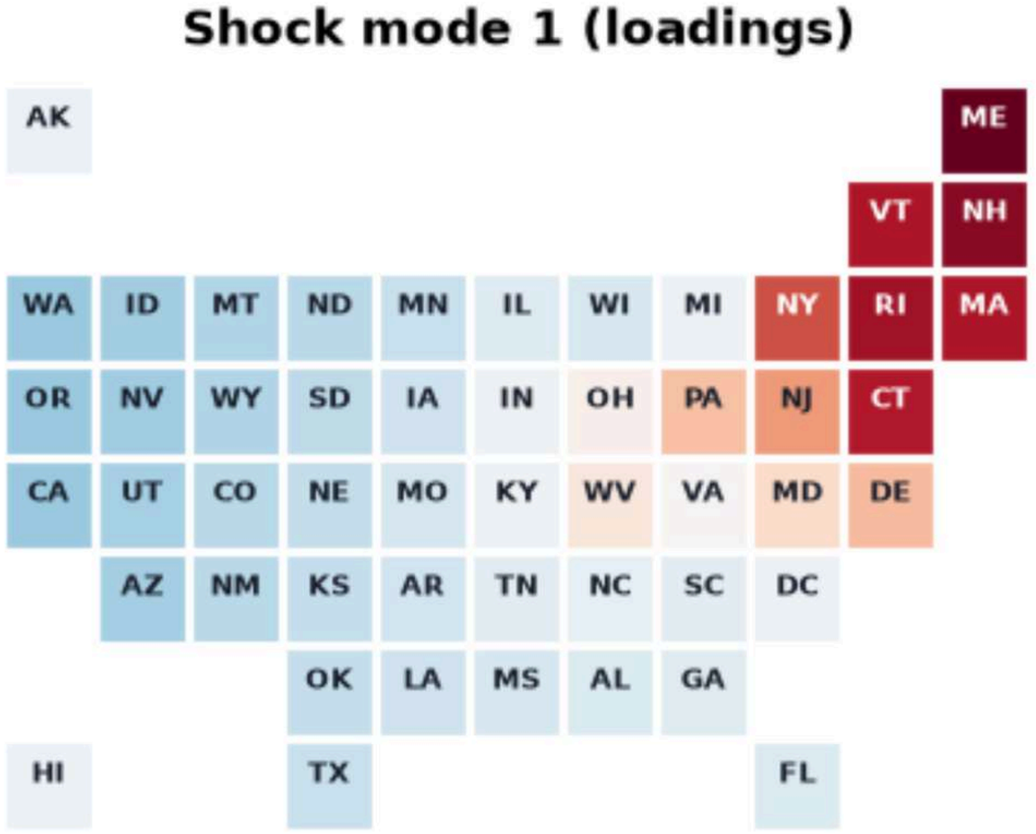
W	gamma	se_state	se_division	placebo_mean	p_vs_placebo
contiguity	−0.0165	0.0348	0.0586	−0.0185	0.5112
2nd order	−0.0216	0.0350	0.0670	−0.0311	0.4065
5-nearest	−0.0170	0.0356	0.0615	−0.0278	0.4214
census division	0.0205	0.0307	0.0483	−0.0259	0.1222

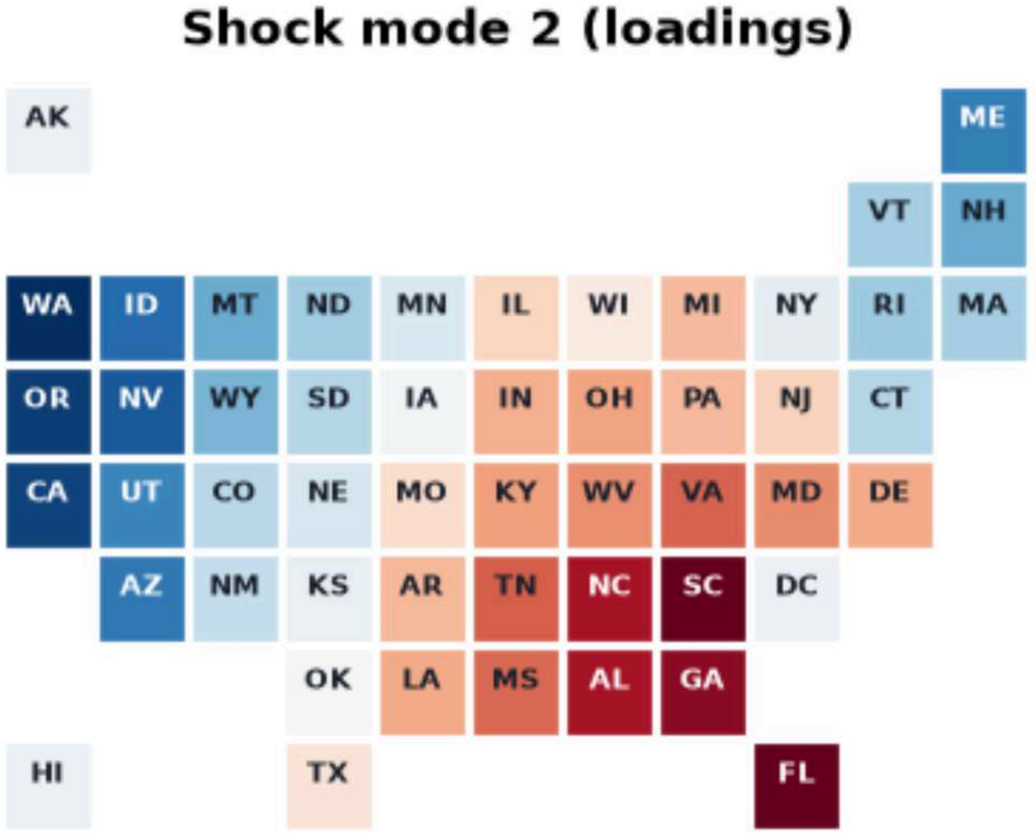
Reading: g near 0 means the map is spatially clustered (Moran) but states do NOT converge toward neighbours over time;

We can seperate states that ‘reprasent’ these clusturs using the Laplacian (and minimizing the energy).

The gore-y math details have been spared in the interest of time.







These are just the dominent eiganvectors!

in typical algorithms fashion, let’s talk about a tangatially related problem...

Election Night Problem

Say you want to forcast the national elections and you are given the results state by state.

In what order does one ask for the states to minimize error?

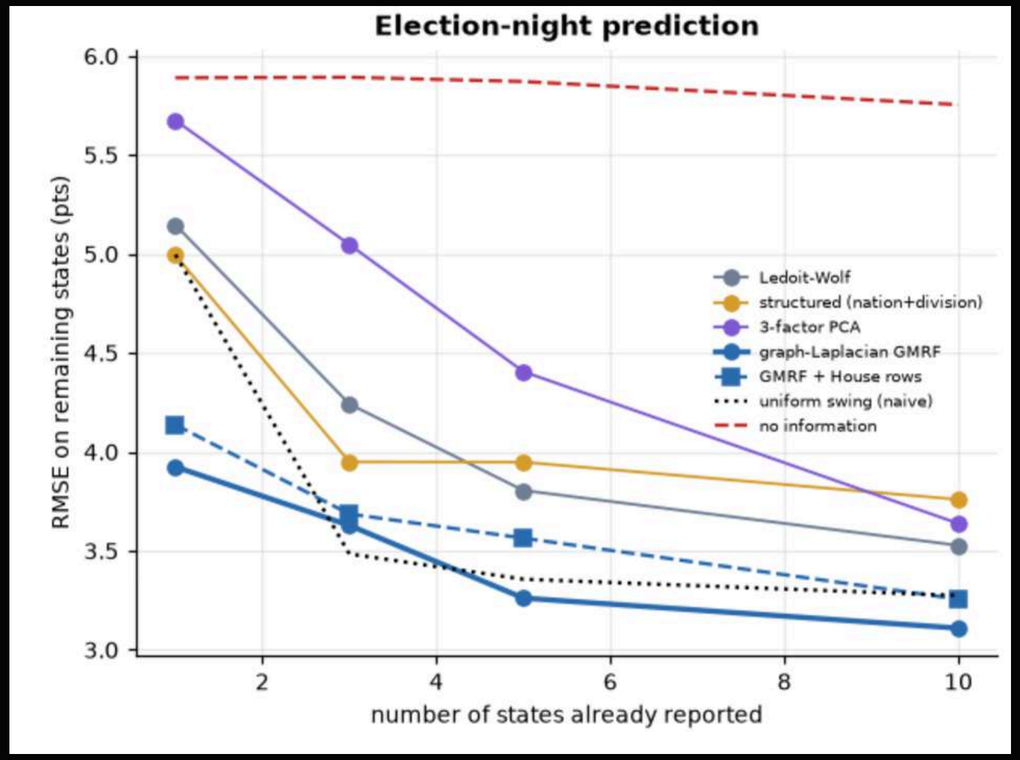
This needs us to basically get the covarience matrix! However, estimating 1,275 distinct covariance parameters from 12 observations is non-trivial!

So how to solve it?



Usually, we impose some structure on the covarience matrix.

Ledoit and Wolf gave a very general way of doing it in their semminal (and honestly, brilliently titled) paper: “Honey, I Shrunk the Sample Covariance Matrix”



How much ever I love their work, there is a simpler and much more suitable structure we could impose... the graph laplacian!

(techincally, the induced Markov random field)

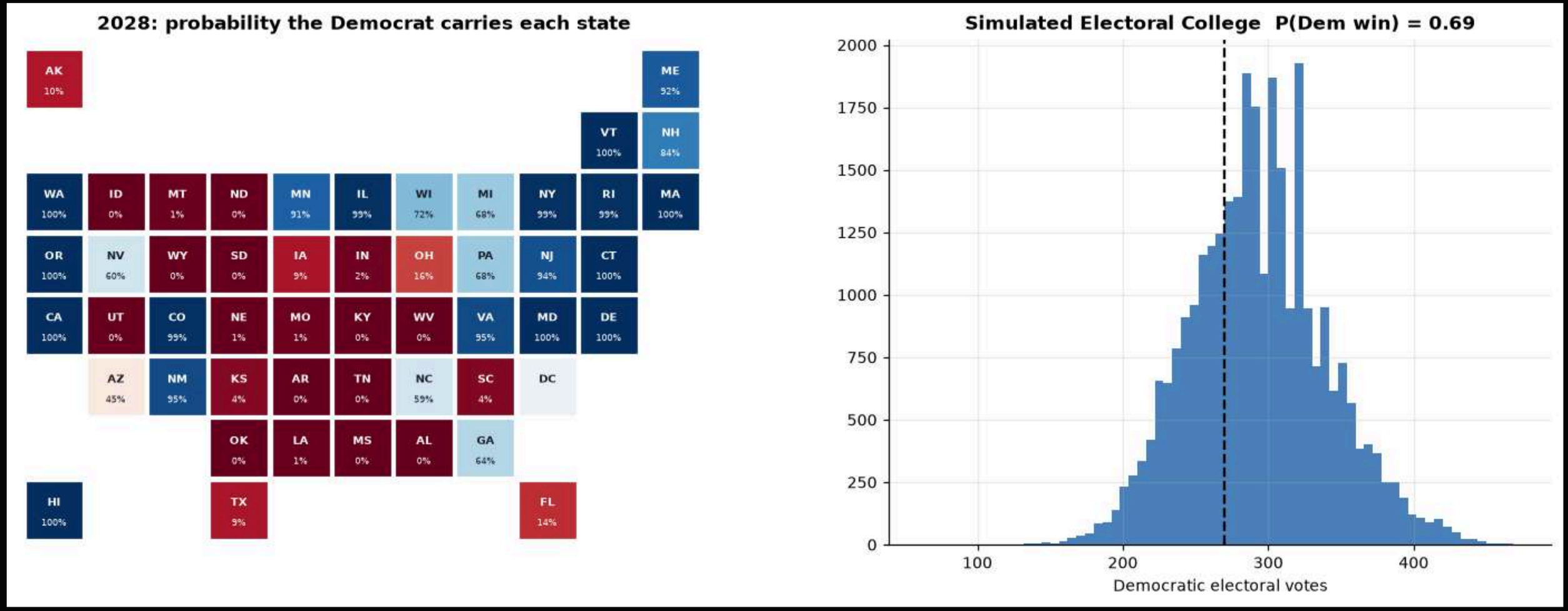
Our (greedy) catalysts were: Missouri, Pennsylvania, Idaho, Tennessee and Massachusetts

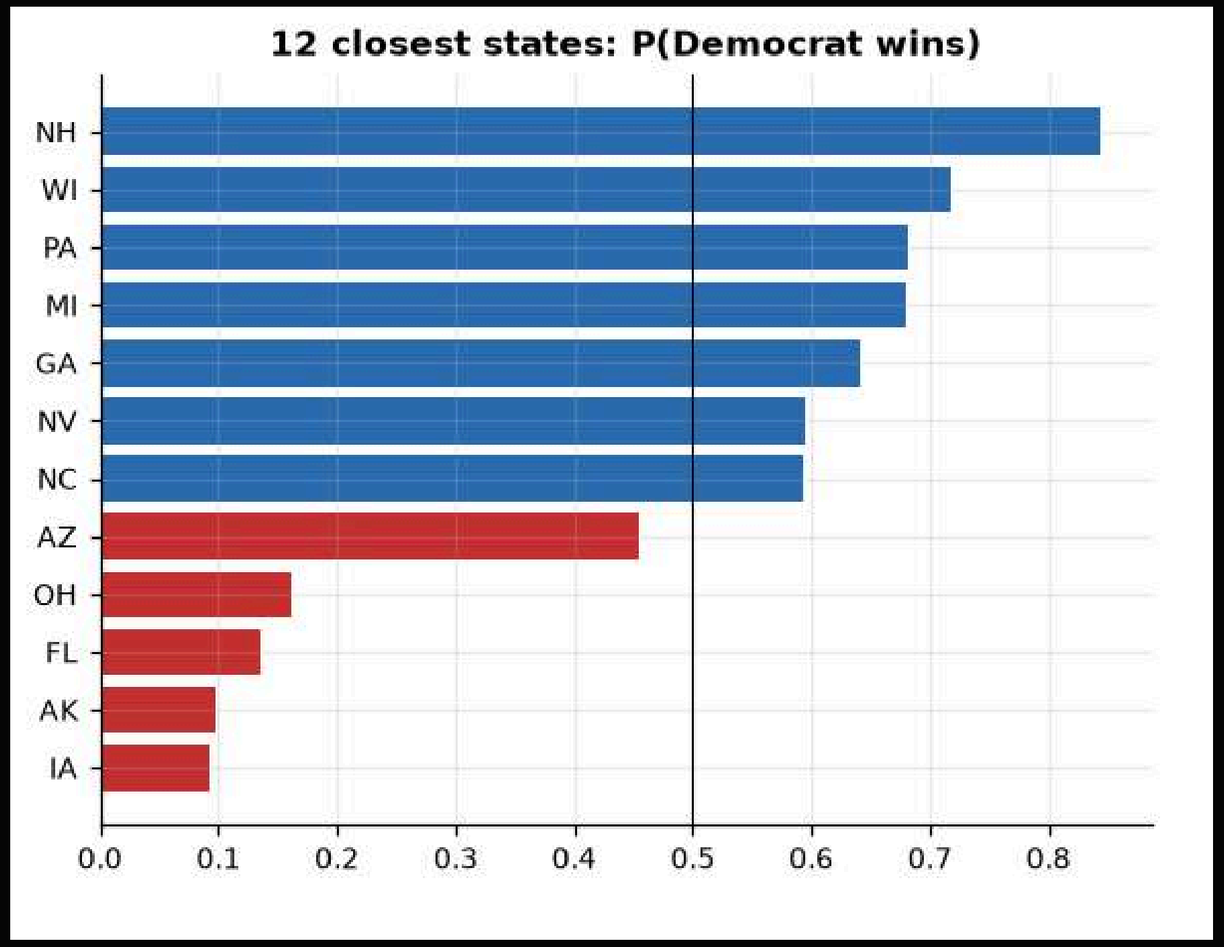
We can now combine all of these!

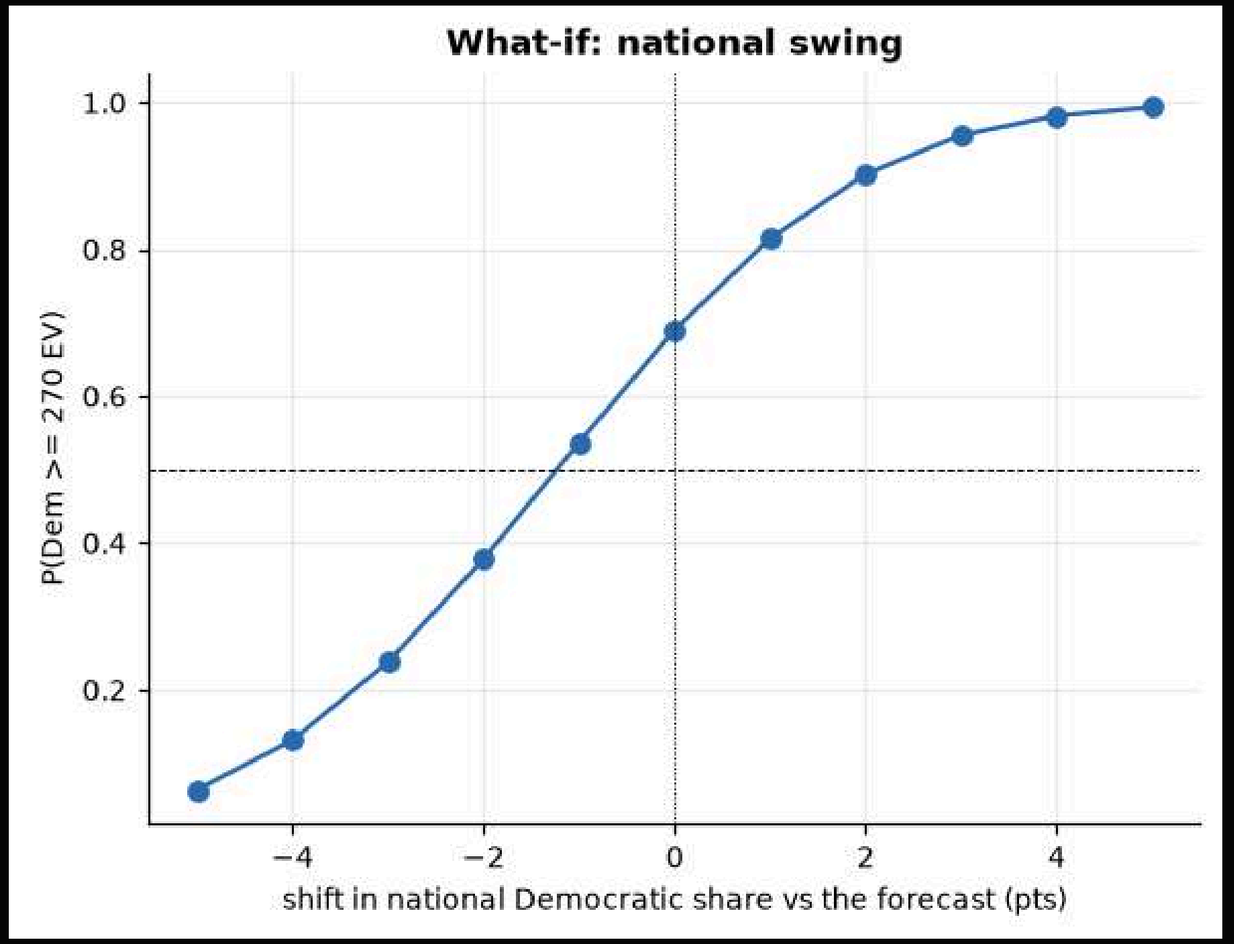
Random Forest for the national trend

Perstinance for state lean predictions
(of some catalyst states)

And Laplace GMRF to extend
it to all the other states!


2028: probability the Democrat carries each state
AK 10%
ME 52%
VT 100%
NH 84%
WA 100%
ID 0%
MT 1%
ND 0%
MN 91%
IL 99%
WI 72%
MI 68%
NY 99%
RI 99%
MA 100%
OR 100%
NV 60%
WY 0%
SD 0%
IA 9%
IN 2%
OH 16%
PA 68%
NJ 94%
CT 100%
CA 100%
UT 0%
CO 99%
NE 1%
MO 1%
KY 0%
WV 0%
VA 95%
MD 100%
DE 100%
AZ 45%
NM 95%
KS 4%
AR 0%
TN 0%
NC 59%
SC 4%
DC
OK 0%
LA 1%
MS 0%
AL 0%
GA 64%
HI 100%
TX 9%
FL 14%
Simulated Electoral College P(Dem win) = 0.69
2000
1750
1500
1250
1000
750
500
250
0
100
200
300
400
Democratic electoral votes


12 closest states: P(Democrat wins)
NH
WI
PA
MI
GA
NV
NC
AZ
OH
FL
AK
IA
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8


What-if: national swing
P(Dem >= 270 EV)
1.0
0.8
0.6
0.4
0.2
−4
−2
0
2
4
shift in national Democratic share vs the forecast (pts)

For 2028, we predict:

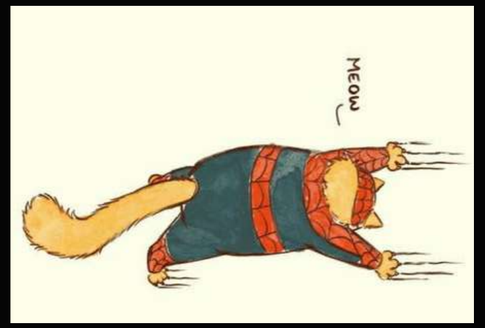
51.6% Democratic two-party with P(Democrat >= 270 EV) = 0.69 (mean 293 EV)


u good?
no
THANK YOU FOR YOUR ATTENTION
[screams internally]
go ahead
ANY QUESTIONS?
im listening

Spatial Catalysts
(aka SpaCats)









Time Series Analysis project by

Arjun Maneesh Agarwal **Urjaswi Chakraborty**