

Shocks

Kiara Hernandez

October 20, 2022

Harvard University

Today's agenda

- ****Review: prediction with GLM**
 - Blog tip: remember to include prediction intervals! We care about uncertainty in our predictions
- **Prediction: why use demographic variables?**
 1. to simulate **surges** from a particular bloc
 2. to capture correlations between districts (i.e. build a **pooled model**)
 3. think about the relationship between shocks and surges - Mobilization vs. persuasion
- **Discussion: do shocks really affect election outcomes?**
 - What can we learn from the Dobbs Supreme Court decision (current)?
 - Lessons learned from shark attacks (historical)
- **Discussion: if shocks have effects, does that mean voters are irrational and incompetent?**
- **Discussion: Shocks blog brainstorm**
 - Discussion in small groups

Prediction: why use demographic variables?

Why use demographic variables?

Certain demographic blocs have either historically (i.e. predictably) and/or contextually (i.e. unpredictably) mobilized in elections:

- **African-Americans** historically vote “steadfastly” for Democrats
- **Working class whites** surged for Trump in 2016
- **Religious whites** surged for Reagan in 1980
- **Latinos** historically vote for Democrats, but especially surged during 2018 midterms
- ... what groups, if any, do we expect to be mobilized, this election cycle? And what events may be mobilizing them (think back to lecture)

Election observers tend to call these deviations that can't be predicted from past behavior **surges**.

A linear regression model specifically estimates coefficients for variables to best predict historical (in-sample) behavior, but we can then perturb those coefficients to simulate surges.

Why use demographic variables: surges [1/2]

Fitting a predictive model of elections using changes in demographic blocs.

```
mod_demog_change <- lm(DemVotesMajorPercent ~ Black_change + Hispanic_change + Asian_change +  
  Female_change +  
  age3045_change + age4565_change + age65_change  
  , data = dat_change)
```

Table 1: The electoral effects of demographic change (across districts)

	<i>Dependent variable:</i>
	DemVotesMajorPercent
Black_change	38.392** (17.629)
Hispanic_change	2.900 (15.744)
Asian_change	31.727 (27.846)
Female_change	−175.070 (129.934)
age3045_change	231.542*** (74.899)
age4565_change	129.585** (57.621)
age65_change	−19.453 (63.342)

Why use demographic variables: surges [1/2]

Now, how would our forecast change if there's a Latino surge for Democrats in 2022? Suppose the effect of Latino bloc growth on Dem vote is doubled this year (across districts).

- Historically estimated coefficient for % Hispanic change: **2.9**

```
predict(mod_demog_change, newdata = demog_2022_change)
```

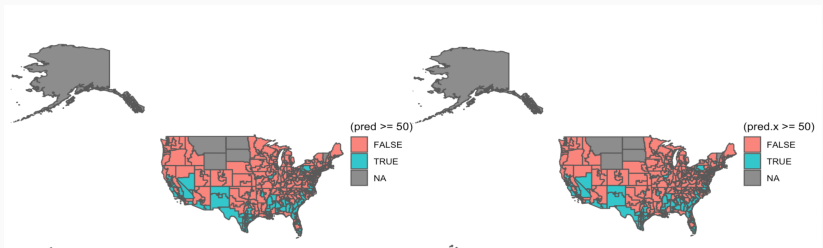
- Hypothetical 2022 coefficient** for % Hispanic change (2x surge): **5.8**

```
predict(mod_demog_change, newdata = demog_2022_change) +  
(5.8-2.9)*demog_2022$hispanic
```

Why use demographic variables: surges [1/2]

```
## Reading layer `districts114' from data source
##   `/private/var/folders/rm/t20t932n257f8ysr6s45ngdw0000gr/T/R
##   using driver `ESRI Shapefile'
## Simple feature collection with 436 features and 15 fields (wi
## Geometry type: MULTIPOLYGON
## Dimension:      XY
## Bounding box:   xmin: -179.1473 ymin: 18.91383 xmax: 179.7785
## Geodetic CRS:   NAD83
```

```
plot_grid(plot_original, plot_1)
```



Why use demographic variables: surges [1/2]

Previously, we fit a separate poll-based regression for each district \rightsquigarrow
assumes separate parameters for each district:

$$DemPV_{district} = f(\alpha_{district} + \beta_{1,district}x_1 + \dots + \beta_{k,district}x_{k,district})$$

With our demographic bloc model, we fit one regression across districts
 \rightsquigarrow *assumes one set of parameters for all districts:*

$$DemPV_{district} = f(\alpha + \beta_1x_1 + \dots + \beta_kx_k)$$

Q: Why are some advantages of the latter?

Why *else* use demographic variables?

Previously, we fit a separate poll-based regression for each district \rightsquigarrow
*assumes separate parameters for each district*¹ (**unpooled model**):

$$DemPV_{district} = f(\alpha_{district} + \beta_{1,district}x_1 + \dots + \beta_{k,district}x_{k,district})$$

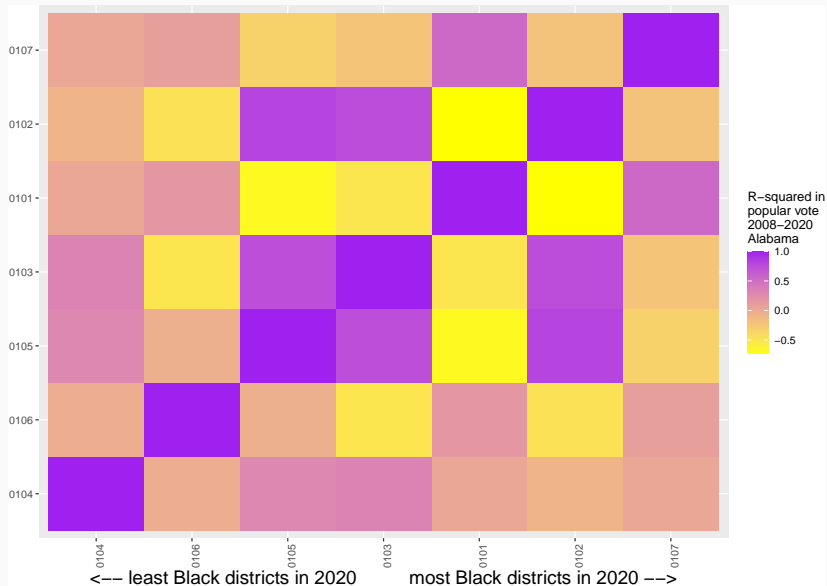
With our demographic bloc model, we fit one regression across district \rightsquigarrow
assumes one set of parameters for all districts (**pooled model**):

$$DemPV_{district} = f(\alpha + \beta_1x_1 + \dots + \beta_kx_k)$$

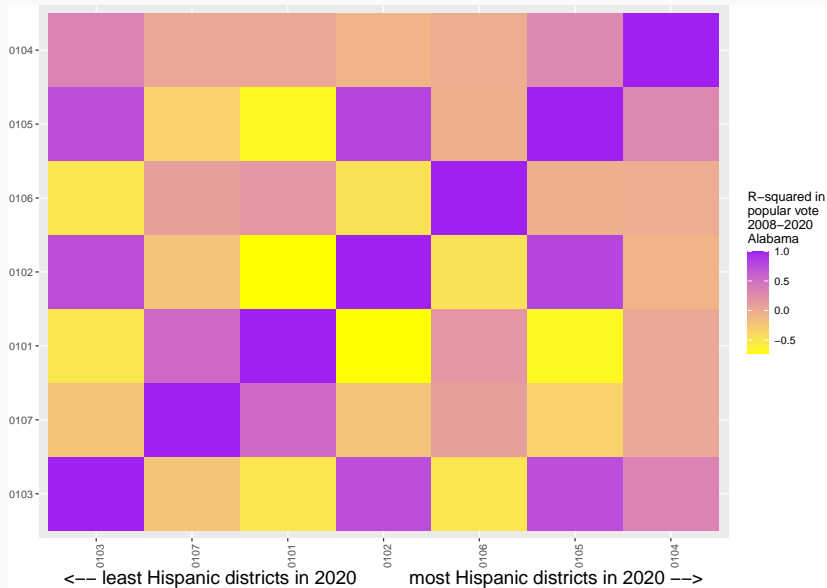
Q: Why are some advantages of the latter?

¹We lied to you last week! This is technically the proper notation for district-by-district models.

Why use demographic variables: across-district correlations [2/2]



Why use demographic variables: across-district correlations [2/2]



Why use demographic variables: across-state correlations [2/2]

To recap:

- A pooled model captures correlation across districts by using the same α and β_1, \dots, β_k to predict each district's outcome.
- A pooled model relies on less data from each district, thus “drawing strength” on less data-sparse district
- **You can think about this as the "data borrowing" concept we've talked about in past weeks - you can use these correlations to update district's A's prediction *if* district B's outcome is hypothetically known**
 - ex: ricardofernholz.com/election - an example to read through

Q: Can I combine a *pooled (i.e. correlated) model* with *unpooled (i.e. idiosyncratic) models* fit to each district?

A: Yes, through the power of ensembling! **And with appropriate checks on the model performance!**

```
for (s in st_cd_fips) {  
  DemPV_s <- 0.5 * predict(mod_demog_change, demog_2020_change %>% filter(st_cd_fips==s)) +  
    0.5 * predict(mod_state_poll_glm[s], poll_pvstate_vep_df %>% filter(st_cd_fips==s, party=="democrat"))  
}
```

**Discussion: Do shocks really
affect election outcomes?**

Shocks?

Q: What are *examples* of **apolitical shocks** that would affect elections?

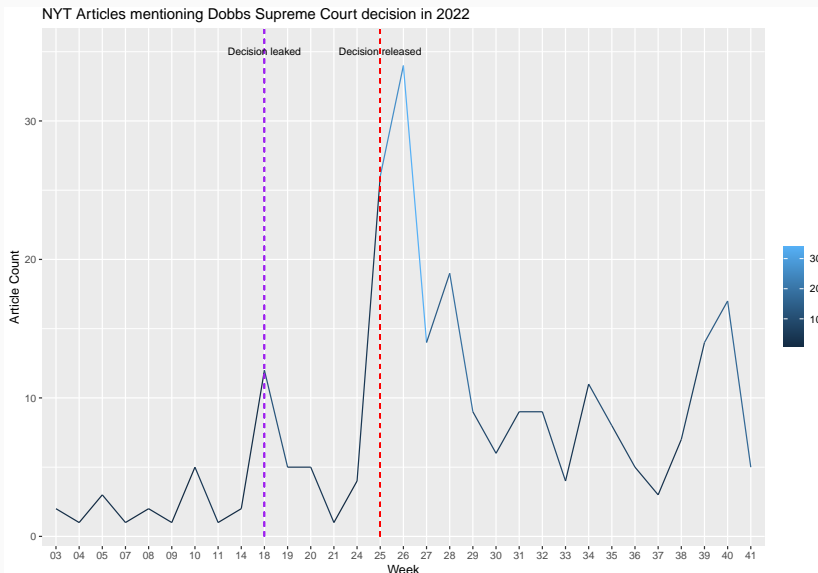
- **Natural disaster:** Shark attack (Achen and Bartels 2017), tornado (Healy and Malhotra 2010) \rightsquigarrow **decreased** support for incumbents
- **Sports** (Healy, Mo, Malhotra 2010): college football team losing their game \rightsquigarrow **decreased** support for incumbents
- **Lottery** (Bagues and Esteve-Volart 2016): Towns winning Spanish Christmas lottery \rightsquigarrow **increased** support for incumbents

Q: Why do we care about apolitical shocks?

Q: What are *examples* of **political shocks** that would affect elections?

- **Natural disaster:** Shark attack (Achen and Bartels 2017), tornado (Healy and Malhotra 2010) \rightsquigarrow **decreased** support for incumbents
- **Sports** (Healy, Mo, Malhotra 2010): college football team losing their game \rightsquigarrow **decreased** support for incumbents
- **Lottery** (Bagues and Esteve-Volart 2016): Towns winning Spanish Christmas lottery \rightsquigarrow **increased** support for incumbents

The Dobbs Supreme Court decision (Roe v. Wade)

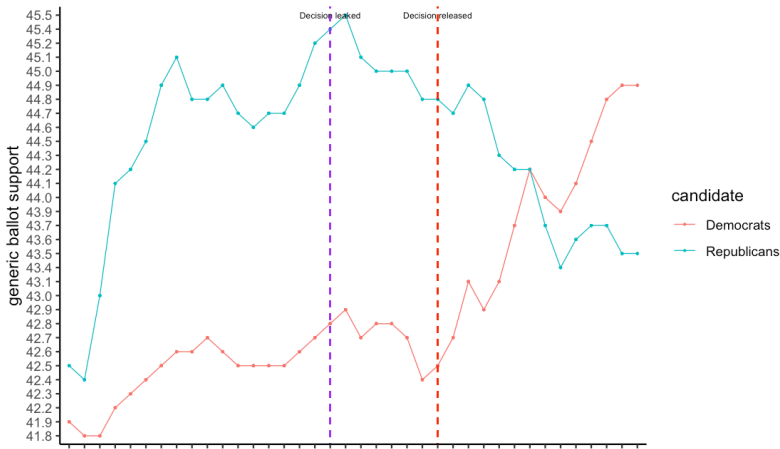


Questions

Q: What do you notice about the number of articles published before and after the Dobbs decision was leaked and released?

Q: Are these “shocks,” in terms of article count, long-lasting? What seems like a reasonable metric for shock longevity?

Q: What can we infer, if anything, about the duration and strength of this particular shock, especially with regard to voter opinion/behavior?



Questions pt.2

Q: What changes do we see in the generic ballot in response to the Dobbs decision?

Q: What additional evidence would be helpful in understanding Dobbs' effect on voter behavior?

Q: Do you think we will see a **surge** in turnout or registration amongst certain demographic groups this cycle?

HEALTH CARE

Kansas' abortion vote kicks off new post-Roe era

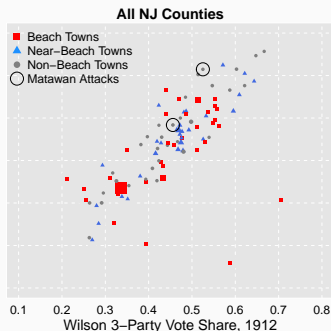
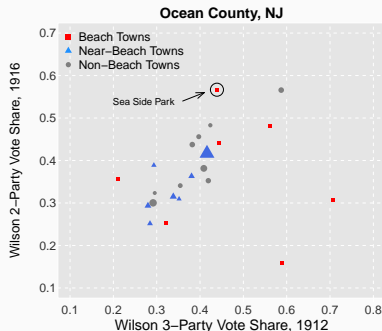
Canvassers are going door to door to lay out the stakes of Tuesday's referendum.

POLITICO

Despite those headwinds, abortion-rights groups feel emboldened by the response on the ground. The co/efficient poll found that significantly more Democrats than Republicans said they were motivated to vote because of the measure — 94 percent to 78 percent. And data from the secretary of state's office shows early in-person voter turnout is nearly **250 percent higher** than the last primary midterm election in 2018, while the number of mail-in ballots is more than double.

A canonical debate: shark attacks (NJ, 1916)

- Achen and Bartels say: Beach towns in NJ affected by shark attacks voted significantly less for the incumbent than non-beach towns
- Fowler and Hall respond:



- Including Sea Side Park (that was omitted) dampens the effect
- Including all counties shows that (1) the two beach counties were very unusual outliers (2) beach towns and near-beach/non-beach towns voted similarly

Discussion: If shocks have effects, does that mean voters are irrational and incompetent?

Q: How do voters *rationally* consider the impact of shocks in their vote?

- **Consider what gov't could have done better** (Many seemingly apolitical shocks are not apolitical):
 - Better **response** (Healy and Malhotra 2010): effects depend on whether the disaster declaration took place
 - Better **relief** (Gasper and Reeves 2011): when presidents reject governors' requests for federal aid \rightsquigarrow punish presidents and reward governors
 - Better **preparation**

Q: How do voters *rationally* consider the impact of shocks in their vote?

- **Global comparison** (Powell & Whitten 1993): Punish **only when** the gov't is doing **worse than other gov'ts**
 - Ex: the speed of recovery from global economic crisis compared to other countries
- **Forward-looking voting** (Leininger and Schaub 2020): Germans **vote to align local incumbents with higher levels of government**—with expectation that this will help them through the COVID crisis

Irrational?

Q: How do voters *irrationally* consider the impact of shocks in their vote?

- **Partisan bias** (Malhotra and Kuo 2008; Healy, Kuo, Malhotra 2014).
 - Survey question: "How much is [Official X] to blame for making American vulnerable to the attacks on September 11?"
 - Experimental variation: Randomly change the partisanship of [Official X]
 - Finding: The answer depends on the partisanship
- **Myopic voting:** do not respond to policies related to disaster preparedness.
 - Respond only to policies and relief right after the crisis has already happened (Healy and Malhotra 2010) \rightsquigarrow create **perverse incentives** to (i) invest less on (more effective + less expensive) preventive measures and (ii) claim credit on performative ex-post measures

Breakout room: Shocks blog brainstorm

Small group discussion: How do you think shocks will factor into the election this year? Do you expect shocks to matter more for certain (groups of) voters or in certain races?

- **NYT scrape and anticipated shocks:** Replicate the Dobbs NYT example from section, but with a different topic/issue that you believe has been a “shock” during the current cycle. Do your findings alter your predictions (hypothetically)? Explain why/ why not.

To replicated the section example, you will need to register for a NYT developer account and acquire an API key.

See additional documentation on replicating the section example:

<https://developer.nytimes.com/get-started>

<https://developer.nytimes.com/docs/articlesearch-product/1/overview>

<https://rpubs.com/justinmorgan/nytimesarticleapi>

- **Incorporate a pooled model into your prediction this week** If you want to just continue refining your current model, that is also fine, but write up a brief explanation as to why you’ve chosen not to incorporate a pooled model.