**Thoughts initial**

A brief (and copycat) go at **modeling roll call voting behavior** in the US House of Representatives using (1) constituency demographics, (2) house member party affiliation, and (3) house member characteristics. This post is based directly on work presented in McCarty, Poole, and Rosenthal ([2016](https://jtimm.net/2020/02/11/modeling-roll-call-voting-behavior/#ref-mccarty2016polarized)), specifically chapter 2 (‘Polarized politicians’). Here, we make **transparent & reproducible** some methods using R & open source data sets – with some fairly comparable results.

**Data sets**

The basic model we consider here is ~formalized as follows:

* Political ideologyi = constituency demographicsi + member partyi + member characteristicsi

**Can we predict how house members vote** based solely on the characteristics of their constituents, their party affiliation, and their age/gender/race? Variables comprising these four model components (for district ***i*** and the representative of district ***i***) are summarized in the table below. Variables included here mostly align with those presented in McCarty, Poole, and Rosenthal ([2016](https://jtimm.net/2020/02/11/modeling-roll-call-voting-behavior/#ref-mccarty2016polarized)). We will describe the details of variables & sources in the following sections.

| **group** | **var** | **source** |
| --- | --- | --- |
| political ideology | dw\_NOMINATE\_DIM\_1 | Rvoteview |
| constituency demographics | cd\_BACHELORS | ACS 2018 1-Year estimates |
| constituency demographics | cd\_HH\_INCOME | ACS 2018 1-Year estimates |
| constituency demographics | cd\_BLACK | ACS 2018 1-Year estimates |
| constituency demographics | cd\_HISPANIC | ACS 2018 1-Year estimates |
| constituency demographics | cd\_IS\_SOUTH | Dixie + KE + OK |
| member party | member\_PARTY | Rvoteview |
| member characteristics | member\_GENDER | CivilServiceUSA |
| member characteristics | member\_ETHNICITY | CivilServiceUSA |
| member characteristics | member\_AGE | CivilServiceUSA |

**+Political ideology**

The **dependent variable** in the model is roll call voting behavior, or political ideology. DW-NOMINATE scores from the [VoteView](https://voteview.com/) project are accessed here via the Rvoteview package. Our focus here, then, is first dimension scores from the 116th House. Scores range from -1 (the most liberal) to +1 (the most conservative). A score of 0, then, would represent a more moderate voter.

nominates <- Rvoteview:: member\_search(chamber= 'House', congress = 116) %>%

select(state\_abbrev, district\_code, nominate.dim1)

**+Member party & member characteristics**

The CivilServicesUSA makes available a collection of characteristics about house members, including **age, gender, and race/ethnicity**. I have cached these data in a package dubbed uspoliticalextras, available.

library(uspoliticalextras)

reps <- uspoliticalextras::uspol\_csusa\_house\_bios %>%

filter(congress == 116) %>%

select(state\_fips:district\_code, party, last\_name, gender, ethnicity, date\_of\_birth) %>%

mutate(age = round(lubridate::interval(date\_of\_birth,

Sys.Date())/lubridate::duration(num = 1,

units = "years"))) %>%

select(-date\_of\_birth)%>%

mutate(district\_code = ifelse(district\_code == 0, 1, district\_code),

ethnicity = ifelse(grepl('middle|multi|nativ|pacific', ethnicity),

'other-race',

ethnicity),

ethnicity = gsub('-american', '', ethnicity),

ethnicity = gsub('african', 'black', ethnicity))

**+Constituency demographics**

Some demographic characteristics for US congressional districts are also included uspoliticalextras. These were accessed via tidycensus, and included in the package out of convenience. **2018 1-year ACS estimates for 12 variables**, which include:

unique(uspoliticalextras::uspol\_dems2018\_house$variable)

## [1] "Median\_HH\_Income" "Per\_BachelorsHigher"

## [3] "Per\_BachelorsHigher\_White" "Per\_Black"

## [5] "Per\_clf\_unemployed" "Per\_ForeignBorn"

## [7] "Per\_Hispanic" "Per\_LessHS"

## [9] "Per\_LessHS\_White" "Per\_VotingAge"

## [11] "Per\_White" "CD\_AREA"

For modeling purposes, constituency demographics are characterized in terms of % population that is Black, % population that is Hispanic, and % population that has obtained a bachelor’s degree or higher. Median household income for the district is also considered. Lastly, each district is identified as being a part of the south or not, where southern states are defined as the eleven states of the Confederacy, plus Oklahoma & Kentucky.

south <- c('SC', 'MS', 'FL',

'AL', 'GA', 'LA', 'TX',

'VA', 'AR', 'NC', 'TE',

'OK', 'KE')

dems <- uspoliticalextras::uspol\_dems2018\_house %>%

spread(variable, estimate) %>%

mutate(district\_code = ifelse(district\_code == 0, 1, district\_code),

is\_south = ifelse(state\_abbrev %in% south, 'Yes', 'No'))

We then join the three data sets, and are good to go. This full data set is available [here](https://github.com/jaytimm/uspoliticalextras/blob/master/clean-data-sets/modeling-roll-call-voting-behavior.csv).

full <- reps %>% left\_join(dems) %>% left\_join(nominates) %>%

mutate(ethnicity = as.factor(ethnicity),

gender = as.factor(gender),

party = as.factor(party),

is\_south = as.factor(is\_south))

**Modeling political ideology in the 116th House**

keeps <- c('nominate.dim1', 'Per\_BachelorsHigher',

'Median\_HH\_Income', 'Per\_Black',

'Per\_Hispanic', 'is\_south', 'party',

'gender', 'ethnicity', 'age')

full1 <- full[, c(keeps)]

colnames(full1) <- meta$var

full1 <- within(full1, member\_ETHNICITY <- relevel(member\_ETHNICITY, ref = 4))

full1 <- within(full1, member\_GENDER <- relevel(member\_GENDER, ref = 2))

full1 <- within(full1, member\_PARTY <- relevel(member\_PARTY, ref = 2))

full1 <- within(full1, cd\_IS\_SOUTH <- relevel(cd\_IS\_SOUTH, ref = 2))

**+Three models**

Per McCarty, Poole, and Rosenthal ([2016](https://jtimm.net/2020/02/11/modeling-roll-call-voting-behavior/#ref-mccarty2016polarized)), and largely for good measure here, we investigate the utility of three models in accounting for variation in DW-NOMINATE scores: (1) only constituent demographics, (2) constituent demographics and house member party, and (3) constituent demographics, house member party, and house member characteristics.

modA <- lm(dw\_NOMINATE\_DIM\_1 ~

cd\_BACHELORS +

cd\_HH\_INCOME +

cd\_BLACK +

cd\_HISPANIC +

cd\_IS\_SOUTH,

data = full1)

####

modB <- lm(dw\_NOMINATE\_DIM\_1 ~

cd\_BACHELORS +

cd\_HH\_INCOME +

cd\_BLACK +

cd\_HISPANIC +

cd\_IS\_SOUTH +

member\_PARTY,

data = full1)

####

modC <- lm(dw\_NOMINATE\_DIM\_1 ~

cd\_BACHELORS +

cd\_HH\_INCOME +

cd\_BLACK +

cd\_HISPANIC +

cd\_IS\_SOUTH +

member\_PARTY +

member\_GENDER +

member\_ETHNICITY +

member\_AGE,

data = full1)

**+Adjusted r-squared per model**

Values are similar to those presented McCarty, Poole, and Rosenthal ([2016](https://jtimm.net/2020/02/11/modeling-roll-call-voting-behavior/#ref-mccarty2016polarized)).

data.frame(modA = round(summary(modA)$adj.r.squared, 3),

modB = round(summary(modB)$adj.r.squared, 3),

modC = round(summary(modC)$adj.r.squared, 3)) %>%

knitr::kable()

| **modA** | **modB** | **modC** |
| --- | --- | --- |
| 0.533 | 0.918 | 0.926 |

**+Coefficients: full model**

td <- broom::tidy(modC) %>%

mutate\_if(is.numeric, round, 3)

colors <- which(td$`p.value` < .05)

td %>%

knitr::kable(booktabs = T, format = "html") %>%

kableExtra::kable\_styling() %>%

kableExtra::row\_spec(colors,

background = "#e4eef4") #bold = T, color = "white",

| **term** | **estimate** | **std.error** | **statistic** | **p.value** |
| --- | --- | --- | --- | --- |
| (Intercept) | 0.686 | 0.057 | 11.985 | 0.000 |
| cd\_BACHELORS | -0.003 | 0.001 | -2.748 | 0.006 |
| cd\_HH\_INCOME | 0.000 | 0.000 | 1.462 | 0.145 |
| cd\_BLACK | 0.000 | 0.001 | -0.142 | 0.887 |
| cd\_HISPANIC | -0.001 | 0.001 | -2.149 | 0.032 |
| cd\_IS\_SOUTHNo | -0.077 | 0.016 | -4.825 | 0.000 |
| member\_PARTYdemocrat | -0.776 | 0.017 | -45.029 | 0.000 |
| member\_GENDERfemale | -0.036 | 0.016 | -2.303 | 0.022 |
| member\_ETHNICITYasian | -0.086 | 0.060 | -1.427 | 0.154 |
| member\_ETHNICITYblack | -0.166 | 0.052 | -3.213 | 0.001 |
| member\_ETHNICITYhispanic | -0.085 | 0.053 | -1.605 | 0.109 |
| member\_ETHNICITYwhite | -0.012 | 0.046 | -0.253 | 0.801 |
| member\_AGE | -0.001 | 0.001 | -2.680 | 0.008 |

**+A visual summary**

jtools::plot\_summs(modC, scale = TRUE)

**+Some interpretations**

**In terms of constituency demographics**, then, house members representing districts with higher percentages of college grads and Hispanics tend to have lower NOMINATE scores, ie, are more liberal in voting behavior. Also, house members representing non-Southern districts have lower scores.

**In terms of member characteristics**, Black house members, female house members, and older house members all have lower scores as well. Party affiliation is the strongest predictor – simply getting elected as a Democrat amounts to a 0.776 *decrease* in NOMINATE scores (on average) relative to a Republican (again, on a scale from -1 to +1).

For the average present day American, model results are in no way surprising. However, as McCarty, Poole, and Rosenthal ([2016](https://jtimm.net/2020/02/11/modeling-roll-call-voting-behavior/#ref-mccarty2016polarized)) demonstrate (see Chapter 2 appendix), constituency demographics & member party have become increasingly more predictive of roll call voting behavior since the early 70s. As voting behavior in the house has become more extreme & ideologically divided.

**A final thought**

So, this post has focused largely on aggregating pieces of a model puzzle as presented in McCarty, Poole, and Rosenthal ([2016](https://jtimm.net/2020/02/11/modeling-roll-call-voting-behavior/#ref-mccarty2016polarized)). We have assumed quite a bit of knowledge wrt the NOMINATE research paradigm, without contextualizing or motivating model composition or results in any real way. Read the reference for this – it tells the story of **increasing polarization in American politics** over the last 40 years or so. A story that becomes more relevant by the day.

**References**

McCarty, Nolan, Keith T Poole, and Howard Rosenthal. 2016. *Polarized America: The Dance of Ideology and Unequal Riches*. mit Press.