Nationscape Insights:
Estimating Policy Support by Congressional District

PUBLISHED JUNE 2021

The document describes how we use the Nationscape survey data to estimate public support for various policies at the congressional-district level. The central challenge in estimating public opinion at the congressional district level is that the number of survey respondents per district, even when pooled across waves in a very large national survey such as Nationscape, is relatively small (between 650 and 3,431). A second challenge is that national survey data are not expected to be representative at the congressional-district level. To produce an accurate picture of variation in public opinion across congressional-districts, both of these challenges must be overcome.

To estimate opinon at the congressional-district level, we employ Gelman's well-known approach of multilevel regression with poststrafication (MrP). Full explainations of MrP can be found on this webpage and in this research paper. In short, we first fit a (logistic) regression model that predicts each survey respondent's answer to a given policy question as a function of a number of sociodemographic and political variables. The regression model specification is flexible and many of its parameters are allowed to vary across congressional districts. Each of these district-specific parameter values are assumed to be drawn from common national distributions of values across districts for each of these parameters. In this way, the multilevel regression model characterizes both individual-level relationships between demographics and policy opinions, and variation in those relationships across districts. After fitting this regression model, we then estimate the share of respondents supporting and opposing the given policy in a number of demographic strata for each district. Each stratum is defined by a combination of socio-demographic characteristics; an example of such a stratum could be something like "young, highly educated, Black, female, Democrat." These district-level estimates are informed directly by the respondents who reside in each district and, because of the multilevel nature of the model, are indirectly informed by the opinions of demographically similar respondents living in other congressional districts. By using the multilevel regression model to smooth variation in opinion across districts within-demographic strata in this way, we arrive at more reliable and stable estimates of voter opinion within each strata and district than could be obtained by simply tabulating the responses of the survey respondents within each strata and district directly.

Assuming that the multilevel regression model provides a reasonable approximation of the cross-district variation in public opinion on the policy question within each strata, we now have what we need to overcome the first of the two challenges to estimating the congressional district-level public opinion. In particular, we have (relatively) reliable district-level estimates of public opinion within a large number of demographic strata despite having only

a relatively small number of respondents in each district. To overcome the second problem of district-level representatives, we employ post-stratification weighting. We take the multilevel-regression-estimated rates of support and opposition to each policy question within each stratum in each district. We then weight them by the fraction of the population in each district that belongs to that stratum. This yields district-wide estimates of policy support and opposition that are demographically representative. Assuming that we are weighting on the right demographics, this poststratification weighting approach results in district-wide estimates that are properly adjusted to account for the non-representativeness of the survey respondents at the district level.

Specific Implementation Details

The implementation of MrP requires a number of specific modeling choices and auxiliary data sources. Given the size and complexity of the Nationscape data and the nature of the policy position questions, there are additional wrinkles to the application of MrP that require further elaboration. In what follows, we provide these details.

The MrP Model

The regression's demographic covariates includes: age (four categories), education (two categories), income (nine categories), race/ethnicity (three categories), gender (two categories), party identification (three categories), urban/rural (three categories), and 2016 presidential vote (five categories including "did not vote" and "not eligible to vote"). Each of these demographics enter into the model as categorical variables. Additionally, district-level 2016 Trump vote interacted with party identification is included as a predictor in the model. The effects of: education and urban/rural, age and race/ethnicity, and 2016 vote and race/ethnicity are interacted in the model. The effects of party and race/ethnicity are allowed to vary by district as is the regression intercept.

The multilevel regression is estimated using the glmer function from the lme4 package in R. Because the policy questions each have three response categories "agree," "disagree," "unsure," it is modeled as multinomial. The lme4 package does not provide a multinominal model estimator, so the multinominal logistical model is instead estimated via two binary logits (see Begg and Gray, 1984). Sampling distributions for the district-stratum-level estimates are arrived at using simulation methods provided in R's merTools package.

Poststratification

In order to poststratify the estimates from the multilevel regression, we first estimate the population size of each demographic stratum (individuals sharing the same combination of age, education, income, race/ethnicity, gender, party identification, urban/rural, and 2016 vote characteristics) within each congressional district. To do this, we employ the Census's 2014–2018 American Community Survey (ACS), a very-large, high-quality, nationally-representative sample of the U.S. population. Because the ACS does not include measures of party identification and 2016 presidential vote, those variables are probabilistically appended to the ACS using auxiliary multilevel regression models estimated using Nationscape data. Finally, after the poststratification, the ACS strata are reweighted to match the known 2016 presidential vote tallies for the district.