Methods Appendix for "G.O.P. Senators Might Not Realize It, but Not One State Supports A.H.C.A."

Christopher Warshaw (MIT) and David Broockman (Stanford)

In order to estimate state-level public opinion on the American Health Care Act (AHCA), we use a technique called multi-level regression with post-stratification (MRP). This technique combines respondents' demographic characteristics, their state of residence, and their views of the AHCA to estimate the probability that a voter of a certain age, race, and sex in each state would support the proposal. By building support within a state based on the composition of the state's population we are able to calculate average levels of support for the bill in each state (for more background on MRP models, see, Park, Gelman, and Bafumi, 2004; Lax and Phillips, 2009; Warshaw and Rodden, 2012). Models like this have been used to estimate public opinion on other policy issues in the United States (e.g., Lax and Phillips, 2009, 2011). They were also used to successfully predict the recent UK election.[1]

1 Data

Even though very few state polls have been done on support for the AHCA, a few pollsters have measured nationwide support for the bill. We use data from 8 national polls between March and June 2017 from the Kaiser Family Foundation, YouGov, and Public Policy Polling. In all, we have 9,390 respondents.

Using these data, we run two separate models. First, we use an MRP model to estimate support for the AHCA (see details below). For this model, we code supporters as a 1 and both opponents and 'don't knows' as 0. Second, we estimate opposition for the AHCA. For this model, we code opponents as a 1 and both supporters and 'don't knows' as 0.

2 Multi-level Regression Model

There are two stages to the MRP model. In the first stage, we estimate the opinion of each geographic-demographic group (e.g., female, white Alabamans). Following Caughey and Warshaw (2015), we do this by treating individual citizens as having been randomly sampled from a given subpopulation g defined by demographic and geographic characteristics. We then model group g's total number of liberal (or conservative) responses on the AHCA as

$$s_g \sim \text{Binomial}(n_g,\ p_g), \tag{1}$$

where n_g is group g's total number of responses and s_g is the number of those responses that are liberal (or conservative).[2]

We then model p as a function of each group's demographics and and state. This approach allows both demographic factors and geography to contribute to our understanding of state

[1] See https://today.yougov.com/news/2017/06/09/the-day-after/

[2] Following Ghitza and Gelman (2013) and Caughey and Warshaw (2015, 202–3), we adjust the raw values of s_g and n_g to account for national survey weights.

opinion. We facilitate greater pooling across states by including in the model several state-level variables that are plausibly correlated with public opinion. We incorporate all of this information with the following hierarchical model for each group's responses:

$$
\begin{aligned}
p_g = \Phi(\gamma_0 + \alpha^{race}_{r[g]} + \alpha^{female}_{f[g]} + \alpha^{age}_{a[g]} + \alpha^{state}_{s[g]}) \\
\text{where:} \\
\alpha^{race}_{r[g]} \text{ for r = 1, 2, 3} \\
\alpha^{female}_{f[g]} \text{ for s = 1, 2} \\
\alpha^{age}_{a[g]} \text{ for a = 1, . . . , 4}
\end{aligned}
\tag{2}
$$

where the normal CDF Φ maps p to the $(0, 1)$ interval.

That is, each individual-level variable is modeled as drawn from a normal distribution with mean zero and some estimated variance. Following previous work using MRP, we assume that the effect of demographic factors does not vary geographically. We allow geography to enter into the model by adding a state level to the model, and giving each state a separate intercept.

The state effects are modeled as a function of the region into which the state falls, the two-party presidential vote in 2016, the state's average income, and whether the state adopted the Medicaid expansion.

$$
\begin{aligned}
\alpha^{state}_{s} \backsim N(\alpha^{region}_{z} + \beta_1 * pvote_s + \beta_2 * income_s + \beta_3 * medicaid_s, \sigma^2_s) \\
\text{for s = (1, . . . , 51)}
\end{aligned}
\tag{3}
$$

The region variable is, in turn, another modeled effect. We group states into regions based upon the BEA's definition of regions.

$$
\begin{aligned}
\alpha^{region}_{z[g]} \backsim N(0, \sigma^2_z) \\
\text{for z = (1, . . . , 8)}
\end{aligned}
\tag{4}
$$

We estimate the model using the `dgmrp` function in the `dgo` package in R (Dunham, Caughey, and Warshaw, 2016).

3 Poststratification

For any set of individual demographic and geographic values, cell g, the results above allow us to make a prediction of public opinion on the AHCA. Specifically, our model estimates cell g based on the combination of the relevant predictors and their estimated coefficients as well as the weighted raw data from that cell. Next, we weight these estimates of opinion in each cell g by the percentages of each type in the actual state populations in order to estimate opinion in each state. We calculate the necessary population frequencies using PUMS "5-Percent Public Use Microdata Sample" from the Census, which has demographic information for 5 percent of each state's voting- age population (Park, Gelman, and Bafumi, 2004; Lax and Phillips, 2009).

References

Caughey, Devin, and Christopher Warshaw. 2015. "Dynamic Estimation of Latent Opinion Using a Hierarchical Group-Level IRT Model." Political Analysis 23(2): 197–211.

Dunham, James, Devin Caughey, and Christopher Warshaw. 2016. "dgo: Dynamic Estimation of Group-Level Opinion. R package version 0.2.3. ." https://jamesdunham.github.io/dgo/.

Ghitza, Yair, and Andrew Gelman. 2013. "Deep Interactions with MRP: Election Turnout and Voting Patterns Among Small Electoral Subgroups." American Journal of Political Science 57(3): 762–776.

Lax, Jeffrey R., and Justin H. Phillips. 2009. "How Should We Estimate Public Opinion in The States?" American Journal of Political Science 53(1): 107–121.

Lax, Jeffrey R., and Justin H. Phillips. 2011. "The Democratic Deficit in the States." American Journal of Political Science 56(1): 148–166.

Park, David K., Andrew Gelman, and Joseph Bafumi. 2004. "Bayesian Multilevel Estimation with Poststratification: State-Level Estimates from National Polls." Political Analysis 12(4): 375–385.

Warshaw, Christopher, and Jonathan Rodden. 2012. "How Should We Measure District-Level Public Opinion on Individual Issues?" Journal of Politics 74(1): 203–219.