# Technical Appendix for "The Politics of Need"

Charles Barrilleaux

Carlisle Rainey

## 1. Mathematical Representation of the Statistical Model

Let $y\_i$ be an indicator that equals one if the governor of state i opposes the Medicaid expansion and zero otherwise. Then we model the probability of opposition as $\Pr(y_i = 1) = \text{logit}^{-1}(X_i\beta)$, where

$$X_i\beta = \beta_{cons} + \beta_{GOP\ gov.}\text{GOP Governor}_i + \beta_{opinion}\text{Opinion}_i + \beta_{GOP\ leg.}\text{GOP Legislature}_i + \beta_{uninsured}\text{Percent Uninsured}_i + \beta_{income}\text{Income}_i + \beta_{nonwhite}\text{Percent Nonwhite}_i + \beta_{metro}\text{Percent Metropolitan}_i$$

and Cauchy(2.5) priors are placed on the coefficients and a Cauchy(10) prior is placed on the intercept.

The variables are defined as follows: GOP Governor is an indicator that equals one if the governor is a Republican and zero if the governor is a Democrat; Opinion is an estimate of the proportion of a state's population that has a favorable opinion of the Affordable Care act (see Section 2 of this Appendix for the details); GOP Legislature is an indicator variable that equals one if Republicans control both branches of the state legislature and zero otherwise; Income is the total personal income in the state per capita; Percent Nonwhite is the fraction of the state's population that identify as non-white Hispanic or African-American; Percent Metropolitan is the percent of the state that resides in a metropolitan area. Prior to estimation, the continuous variables are standardized by subtracting the mean and dividing by two standard deviations and binary variables are centered by subtracting the mean (Gelman 2008).

We place weakly informative Cauchy(2.5) prior distributions on the coefficients for the (standardized) explanatory variables and a more diffuse Cauchy(10) prior on the intercept. We use JAGS called through R to perform a 50,000 iteration burn-in and 50,000 additional posterior simulations for three MCMC chains. We combined these simulations to perform the inferences.

## 2. State-Level Estimates of ACA Favorability

In the main text, we rely on a state-level estimate of the favorability of the 2010 Affordable Care Act as our measure of public opinion. As robustness checks, we also use state-level estimates of support for the Medicaid expansion and support for the Tea Party. The estimate state-level public opinion on the Affordable Care Act, we use multilevel regression with post-stratification (MRP; Lax and Phillips 2009) to combine the Kaiser Family Foundation Health Tracking Polls from January to November of 2013 with census data from 2000 (with 2008 weights).1 Where $y\_i$ is an indicator for whether the respondent views the ACA favorably, supports the Medicaid expansion, or supports the Tea Party,

$$\Pr(\mathrm{y_i} = 1) = \text{logit}^{-1}(\alpha^{cons} + \alpha^{race}_{j[i]} + \alpha^{gender}_{k[i]} + \alpha^{race\times gender}_{l[i]} + \alpha^{age}_{m[i]} + \alpha^{income}_{p[i]} + \alpha^{education}_{q[i]} + + \alpha^{state}_{s[i]})$$

In our model, $\alpha^{cons}$ is a fixed intercept; $\alpha^{race}_j$, $\alpha^{gender}_j$, $\alpha^{race\times gnder}_j$, $\alpha^{age}_j$, $\alpha^{income}_j$, and $\alpha^{education}_j$ are unmodeled random intercepts by the (categorical) variables indicated by the superscripts and shown in Table 4; and $\alpha^{state}_s$ is a random intercept modeled as

$$\alpha^{state}_s \sim N(\alpha^{region}_{r[s]} + \alpha^{Obama}\text{Obama Vote Share in 2012}_s),$$

where $\alpha^{region}_{r[s]}$ is an unmodeled random effect and $\alpha^{Obama}$ is a fixed effect. We fit the model and performed the

[1] Data for the support for the Medicaid expansion are available only from the July 2012 Tracking Poll.

poststratification in R using the MRP package available at github.com/malecki/mrp. Our state-level estimates are given below in Figure 1 and available for download at github.com/carlislerainey/ACA_Opinion.

| Variable | Categories |
|---|---|
| Race | • White<br>• Black<br>• Hispanic<br>• Other |
| Gender | • Male<br>• Female |
| Race and Gender Interaction | • White male<br>• White female<br>• Black male<br>• Black female<br>• Hispanic male<br>• Hispanic female<br>• Other male<br>• Other female |
| Age | • 18-29<br>• 30-44<br>• 45-64<br>• 65+ |
| Income | • Less than $20,000<br>• $20,000 to $40,000<br>• $40,000 to $75,000<br>• $75,000+ |
| Education | • Less than high school<br>• High school graduate<br>• Some college<br>• College graduate<br>• Postgraduate |
| State | 50 states plus the District of Columbia, though the latter is not included in our model of gubernatorial opposition. |
| Region | • South<br>• Northeast<br>• Midwest<br>• West |

Table 1: This table provides individual-level categorical variables used in the MRP process to estimate ACA favorability in the states.

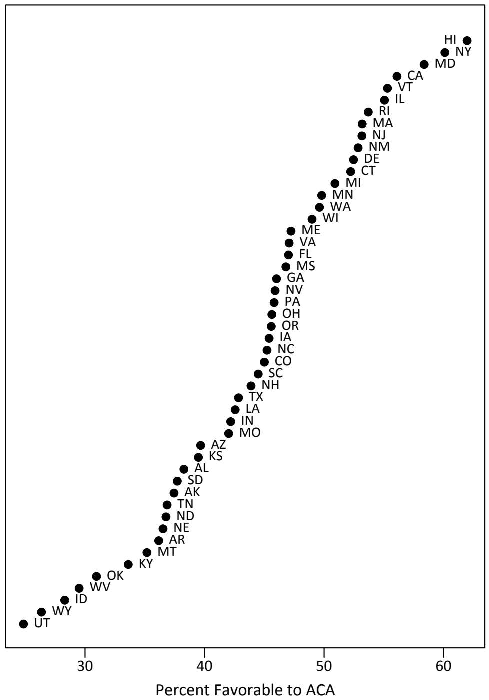


Figure 1: This figures shows the estimates of ACA favorability by state used in the main analysis.

# 3. Robustness Checks

We examine a variety of robustness checks, including the choice of prior scale for the Cauchy family, the choice of prior family, the choice of estimation technique, model specification, and case selection.

## *3.1. Prior Scale*

First, we demonstrate that the substantive conclusions are similar regardless of the prior scale. We re-estimated the main model with an alternative prior specifications to the default Cauchy(2.5) prior specification suggested by Gelman et al. (2008). **Error! Reference source not found.** shows the posterior medians and 90% credible intervals. Notice that the substantively conclusions do not change as the prior scale varies. Notice, though, that the prior scale mainly affects the upper bound of the 90% credible interval for the constant term and the coefficient for *GOP Governor*. This makes sense, as the likelihood is highly informative that the coefficient for *GOP Governor* is "not negative" but the prior is required to provide the information that the coefficient is "not positive infinity."

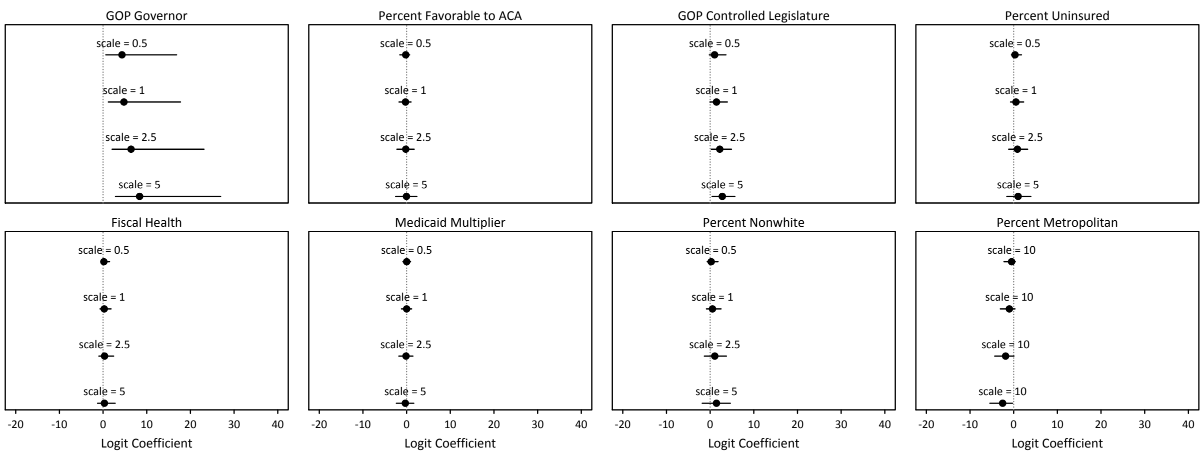


Figure 2: This figure shows how the inferences change as the scale on the Cauchy family of prior distributions changes. Notice that while the uncertainty around the estimate increases (mainly the upper bound), the substantive conclusions remain relatively unaffected.

## *3.2. Prior Family*

In addition to the results being robust to the prior *scale*, they are also robust to the prior *family*. To demonstrate this, we considered two alternative prior families, the normal and the $t_{10}$. We use the normal family because of its familiarity and the $t_{10}$ because the $t_{10}$ with scale 2.5 made the most sense to us as an informative prior based on a simple heuristic analysis.[2] Figure 3 shows the estimated using the normal prior with various scales (standard deviations). Notice that the normal family pools the coefficients much more strongly toward zero, but the results are otherwise similar.

[2] We arrived at the $t_{10}$ with scale 2.5 by simulating from the prior predictive distribution, computing quantities of interest, and choosing the simulations that seem to best reflect the prior uncertainty in the quantity of interest. The Cauchy(2.5) treats first differences of nearly one (or nearly negative one) as overly probable. On the other hand, the Normal priors do not give these possibilities enough weight. The $t_{10}$ prior with scale 2.5 allows these large shifts without placing undue prior weight on them. However, this is not to suggest that the $t_{10}$ with scale 2.5 is the best informative prior distribution, but simply useful as a robustness check.

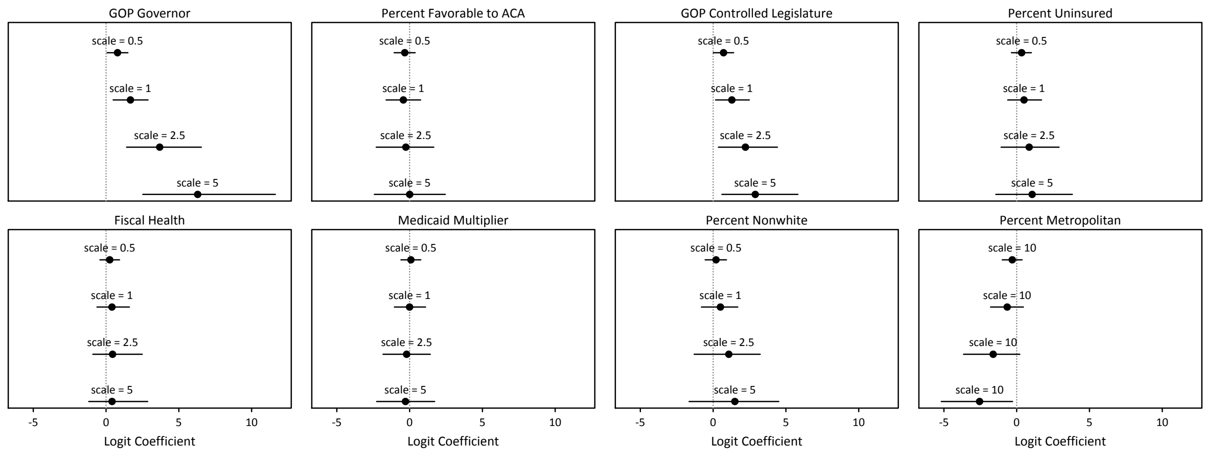


Figure 3: This figure shows how the inferences change as the scale (standard deviation) on the normal family of prior distributions changes. Notice that while the uncertainty around the estimate increases (mainly the upper bound), the substantive conclusions remain relatively unaffected. Compare to **Error! Reference source not found.** to see that while the normal family pools the coefficients more strongly toward zero, the substantive conclusions do not change.

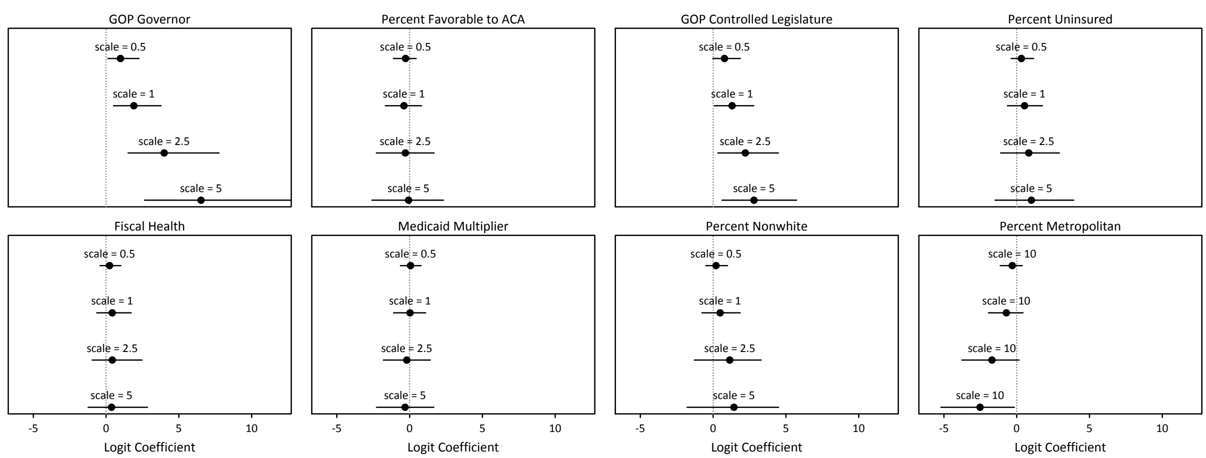


Figure 4: This figure shows how the inferences change as the scale (standard deviation) on the $t_{10}$ family of prior distributions changes. Notice that while the uncertainty around the estimate increases (mainly the upper bound) compared to the Cauchy family in **Error! Reference source not found.** and decreases compared to the normal family shown in Figure 3, the substantive conclusions remain relatively unaffected.

### *3.3. Choice of Estimation Technique*

In addition to the key substantive conclusions being robust to a range of prior specifications, they are also robust to several alternative estimation techniques. We consider several here. First, we compute the MLE estimates in spite of the separation. Second, we use Firth's penalty (Zorn 2005) and compute standard errors using both asymptotic approximations, likelihood profiling, and bootstrapping. We also use Gelman et al.'s (2008) approximation to the posterior mode and asymptotical approximate confidence intervals and bootstrapping. Finally, we compare each of these to our preferred approach of MCMC with a Cauchy(2.5) prior and include the results from the Normal(1) prior as a basis of comparison. The key point to take away from these results is the consistency in the estimated coefficients and uncertainty across the various estimation strategies.

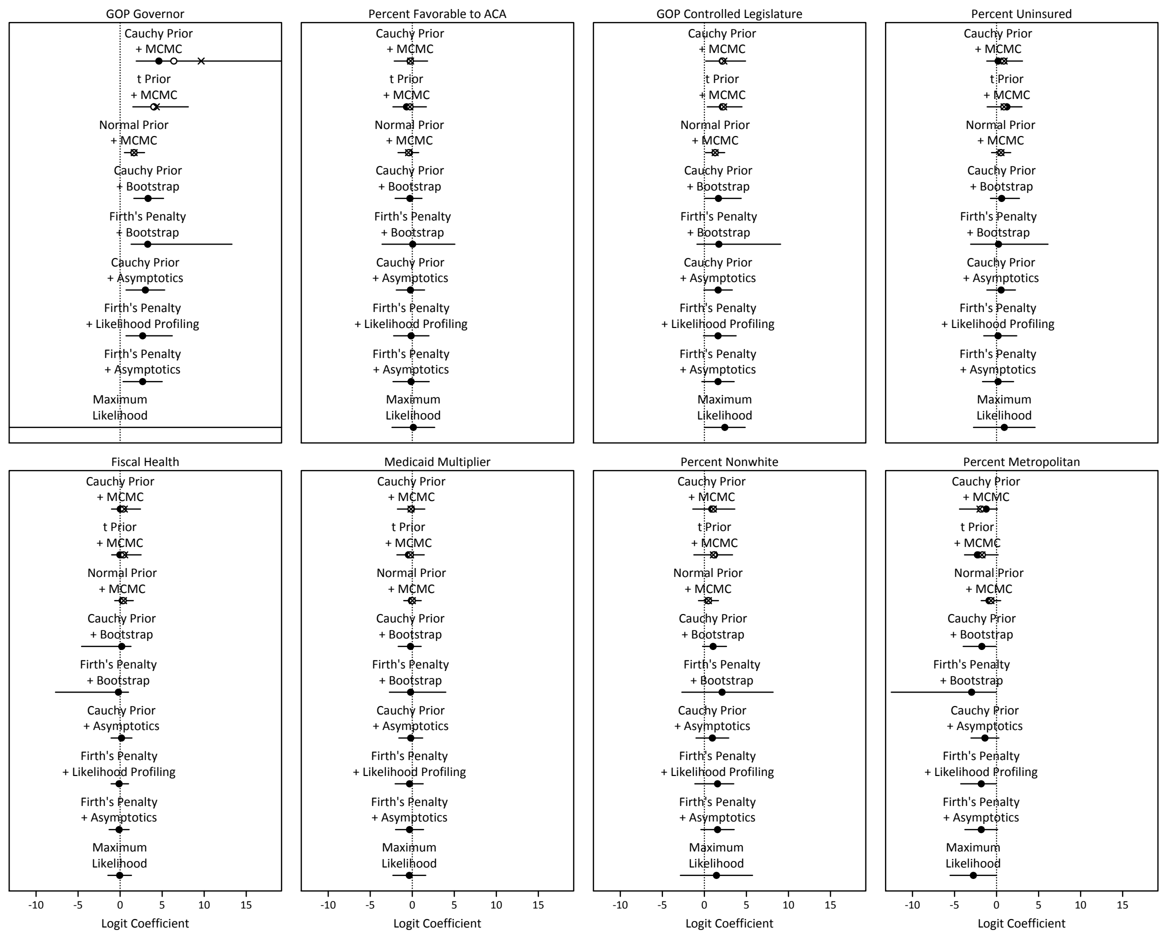


Figure 5: This figure shows how the inferences change across various estimation procedures for numeric explanatory variables transformed to have mean zero and standard deviation 0.5 and binary explanatory variables simply centered at their mean. The intercept is omitted to save space. For the models estimated with MCMC, the "x" represents the posterior mean, the circle represents the posterior median, and the dot represents the posterior mode. This distinction is consequential only for highly skewed posterior distributions. The substantive effects reported in the main text rely on MCMC approach with Cauchy priors, but the results do not change if we rely on the analytical standard errors or Firth's penalty. Maximum likelihood clearly fails in this case. Though the maximum likelihood estimates depend on the convergence criteria, the default criteria in R suggests that the intercept is about -7.7 with a 90% confidence interval from 1,345 to 1,360. For the coefficient for GOP Governor, the defaults lead to a coefficient estimate of 18.3 and a 90% confidence interval from 1,334 to 1,371. Notice that the Cauchy prior and Firth's penalty regularizes these quantities, providing more reasonable estimates.

### *3.4. Case Selection*

With such a small number of cases included in the analysis, we leave open the possibility that a particular case drives the results. To address these concerns, we refit the models dropping one state at a time.[3] Figure 6 shows the results. Notice that although some states are more influential than others, the results are similar regardless of which state is dropped from the analysis.

[3] For the sake of computational tractability, we compute these estimates and standard errors using Gelman et al.'s (2008) EM approximation to the posterior mode and asymptotic confidence intervals.

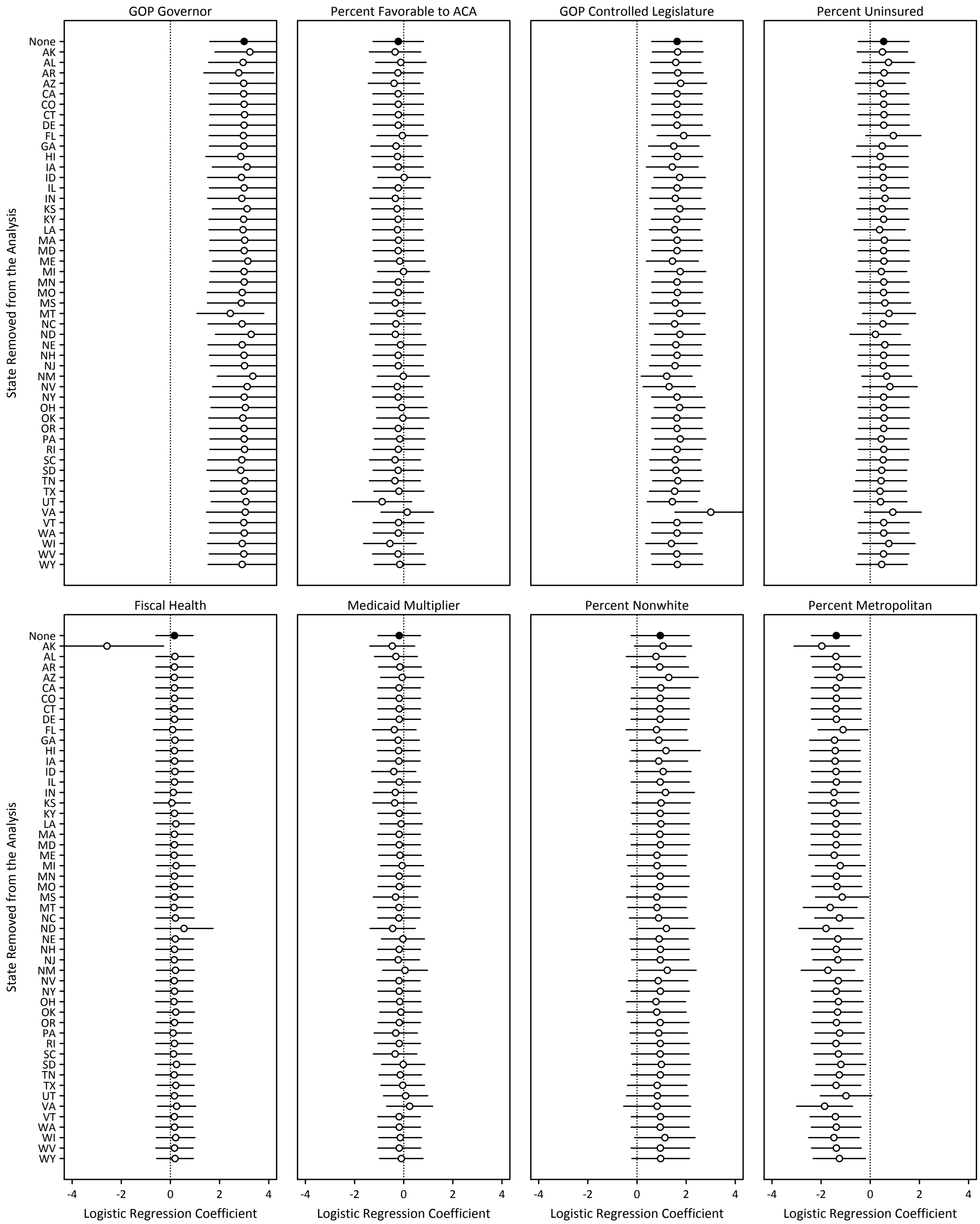


Figure 6: This figure shows how the estimates change when each state is removed from the analysis. The top estimate in each panel presents the estimate using the full data set. Though sometimes dropping a state causes the confidence interval to overlap zero, notice how little the estimates change when each state is dropped. The exception to this pattern is that dropping Alaska, which has a unusual amount of reserves given their spending, dramatically and substantively changes the estimated coefficient of fiscal health. However, this change does not affect the inference about the key variables in our analysis.

# 4. Evaluating the Possibility of Overfitting and an Overly Informative Prior

Related to the robustness checks above, we now demonstrate that (1) the Cauchy(2.5) prior is not overly informative and (2) the model does not overfit the data. We do this by showing the more informative priors predict out-of-sample observations slightly better than the prior that we rely on. To do this, we computed Brier scores using leave-one-out cross validation for a range of prior families and scales. The procedure works as follows:

1. Start by dropping a single observation (say, Alabama).

2. Use the rest of the data (all states except Alabama) to estimate the model and the probability that the governor of the left-out state opposes the expansion. Save the estimated probability.

3. Do this for all 50 states.

4. Calculate the Brier score B using the equation $B = \sum_{i=1}^{50}(y_i - p_i)^2$ where $y_i$ equals one if the governor of state $i$ opposes expansion and $p_i$ represents the out-of-sample estimated probability that $y_i$ equals one.

We repeat the procedure above for a variety of different prior families, including the Cauchy, normal, and t, and vary the scale parameter across each of the distributions. Figure 3 shows how the Brier scores change as the prior family and scale vary.

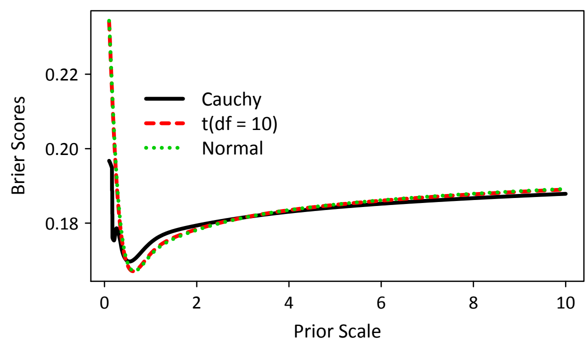


Figure 7: This figure shows the out-of-sample performance of three prior distributions as the scales vary. Lower Brier scores indicate improved out-of-sample predictions. Notice that informative priors (e.g. Normal(1) and Cauchy(2.5)) only slightly outperform weaker prior distributions. Because the more and less informative priors perform similarly, we are not as worried about overfitting the data. Because we would rather supply too little prior information than too much, we choose use the less informative Cauchy(2.5) prior distribution in the main analysis.

Notice that the Brier scores are lower (i.e. the out-of-sample prediction is better) when the prior is more informative that the prior we rely on. However, the improvement is not dramatic, so we prefer to use too little prior information rather than too much. To err on the side of caution, we use the default Cauchy(2.5) prior suggested by Gelman et al. (2008).

Another potential concern with the small sample is that we are over-fitting the data. (The problem of separation that we identify in the main text is just one example of over-fitting.) However, the Brier scores shown in Figure 7should alleviate any concerns about over-fitting. Notice that the Brier score for our preferred Cauchy(2.5) is about 0.17. The lowest Brier score among the parameterizations shown is about 0.15 for the much-more-informative Normal(0.8) prior distribution. Slightly better performance from a more informative prior is expected. Indeed, Gelman et al. (2008) find a very similar pattern across their large corpus of data sets (see especially the right panel of their Figure 6). If a more informative prior fit the data substantially better, then over-fitting would be a concern. However, the pattern shown in Figure 7 suggests that over-fitting is not a concern.

## References


Gelman, Andrew. 2008. "Scaling regression inputs by dividing by two standard deviations." *Statistics in Medicine* 27(15): 2865-2873.

Gelman, Andrew, Aleks Jakulin, Maria Grazia Pittau, and Yu-Sung Su. 2008. "A Weakly Informative Default Prior distribution for Logistic and Other Regression Models." *The Annals of Applied Statistics* 2(4): 1360-1383.

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

Mooney, Christopher Z., Robert D. Duval, and Robert Duvall. 1993. *Bootstrapping: A Nonparametric Approach to Statistical Inference*. SAGE Publications, Inc.

Zorn, Christopher. 2005. "A Solution to Separation in Binary Response Models." *Political Analysis* 13(2): 157.