# Reevaluating the Substantive Representation of LGB Americans with Multiverse Analysis

Joseph Saraceno
jsaracen@usc.edu
Department of Political Science
University of Southern California

Eric R. Hansen
ehansen4@luc.edu
Department of Political Science
Loyola University Chicago

Sarah A. Treul
streul@unc.edu
Department of Political Science
University of North Carolina at Chapel Hill

May 18, 2020

**Abstract**

Social scientists are facing a crisis of confidence in quantitative results. Multiverse analysis (Steegen et al. 2016) provides concerned scholars a tool for verifying the robustness of findings. This article introduces political scientists to multiverse analysis through an application. It identifies how differing approaches to data processing led to divergent conclusions about the representation of LGB Americans in Congress in a 2015 *Journal of Politics* article. The analysis casts doubt on the original conclusion that the size of the LGB population in a district is significantly associated with the bill sponsorship activity of its representative. More broadly, it demonstrates how researchers can keep a running tally of data processing decisions and parsimoniously present the consequences of those decisions for the findings. Multiverse analysis can help scholars publish replicable results in original work as well as replicate and extend previously published work.

**Keywords**: representation, multiverse analysis, replication, open science

Scholars have long sought to establish the factors motivating legislators to provide substantive representation, or activities that make meaningful changes to public policy, to minority constituents.[1] In a *Journal of Politics* article, Hansen and Treul (2015) argued that the size of the minority population of a district influences the substantive representation provided to it by the district's elected representative. Using the case of lesbian, gay, and bisexual (LGB) Americans, they found that representatives of districts with larger LGB populations were more likely to sponsor Congressional bills advancing gay rights, as were descriptive representatives. The results imply that even small minority populations can influence their representation in Congress, though the influence may not be strong enough to sway skeptical representatives.

In a replication submitted to the *Journal of Politics*, Saraceno found contradictory results. His analysis of congressional sponsorship data showed that the size of the LGB population in a district does not have a meaningful independent association with a legislator's sponsorship of gay rights legislation. Across a series of model specifications, he reported that (1) the coefficients for the LGB population variable were consistently indistinguishable from zero and (2) the predicted number of pro-LGB bills sponsored by members of Congress remained largely unchanged across values of LGB population share.

Having arrived at different conclusions despite using similar data, the editors of the *Journal of Politics* tasked us with teaming up to produce a collaborative replication, focusing primarily on Model 9, presented in Table 3 of the original article. We see our disagreement through the framework of a "garden of forking paths" (Gelman and Loken 2014), in which the data processing choices made at researchers' discretion ultimately influenced the conclusions. The two principal areas of disagreement fell on 1) how pro-LGB bills were coded and 2) how district partisanship was measured. However, exchange between the authors and helpful feedback from anonymous reviewers prompted us to expand our scrutiny to other data processing choices.

We come together in this article to show the reader how different decisions led us to different conclusions. To this end, we use multiverse analysis (Steegen et al. 2016). In total, we estimate and present the results from 288 model specifications. We explain what happens to the original

[1]The authors are grateful to the editor for the assignment and two anonymous reviewers for their careful and insightful comments.

results as we make measurement and modeling changes in a piecemeal fashion.

We find that the balance of the evidence across models would not allow us to conclude that a larger LGB population in a district is related to a larger number of bill sponsorships. Considering the distribution of $p$ values for the independent variable of interest across alternative models, the original finding fell among the minority of model specifications that yielded significant results. Roughly 11% of models show a $p$ value smaller than the conventional .05 threshold. However, the large majority of plausible models produce results that would not allow us to reject the null hypothesis.

This article contributes to the literature on minority representation by casting doubt on the finding of a connection between district presence of LGB Americans and their substantive representation in Congress. Our broader contributions are introducing political scientists to the multiverse analysis and providing a model for its application.[2] We conclude by discussing how multiverse analysis might be applied by scholars verifying the robustness of their own work and by scholars replicating and extending prior work.

## Defining the Multiverse

For the multiverse analysis, we constructed a data set that allows us to compare the original results with results obtained using alternative measures and modeling choices. Table 1 lists the constitutive elements we consider. Entries denoted with an asterisk indicate the choices made by Hansen and Treul in their original article. All other options represent alternative ways that the data could have been processed. Below, we provide additional details about these alternatives.

***Sponsorships***. The key dependent variable used to assess the substantive representation of LGB Americans in Hansen and Treul (2015) is the number of pro-LGB bills sponsored by a

[2]The underlying logic of multiverse analysis is comparable to extreme bounds analysis (EBA) (e.g. Leamer 1985; Miller et al. 2018). The primary difference lies in how alternative models are generated. Multiverse analysis compares results from models that differ by the measurement of variables. In contrast, EBA has traditionally been applied to assess the robustness of the relationship between the main independent variable and dependent variable while varying the number of included controls. Because coefficient estimates may not be directly comparable in multiverse analysis, researchers focus interpretation of results on the distribution of $p$ values for the variable(s) of interest. Researchers can use the common framework to compare models that vary in both measurement choices and the number of included controls, as we have done here.

Table 1: Model Specifications

1. ***Sponsorship (S)***
   (a) S1: Hansen and Treul original counts*
   (b) S2: Saraceno counts
   (c) S3: Revised Hansen and Treul counts
2. ***District Partisanship (D)***
   (a) Congressional Vote-share:
   D1: Democrat in unopposed races coded 100*
   D2: Democrat in unopposed races coded 90
   D3: Democrat in unopposed races coded 80
   D4: Democrat in unopposed races coded 70
   (b) D5: Democratic Presidential Vote-share from Jacobson
   (c) D6: Cook PVI
3. ***Public Opinion (O)***
   (a) O1: Excluded as a control variable*
   (b) O2: Included as a control variable
4. ***LGB Population (P)***
   (a) P1: ACS estimates used in Hansen and Treul*
   (b) P2: ACS 1-yr estimates
5. ***Campaign Contributions (C)***
   (a) C1: Included as a control variable*
   (b) C2: Excluded as a control variable
6. ***Lagged DV and Random Effects (L)***
   (a) L1: Lagged dependent variable without random effects*
   (a) L2: No lagged dependent variable with random effects

member in a given term. However, we identified several instances in which the original sponsorship counts appeared to be improperly attributed to members of Congress. This included counting anti-LGB rights bills as if they were pro-LGB, counting bills that only addressed gender discrimination, and counting individual bills multiple times. To address the errors, Saraceno independently gathered and coded bill sponsorship data, while Hansen and Treul additionally recoded their own measure, adhering strictly to the sources and stipulations described in the manuscript.[3] Sponsorship counts by Saraceno are identified by the S2 variable while Hansen

[3]This process included identifying the number of bills sponsored by each member that sought to extend the legal rights or privileges to LGBs in one term, excluding resolutions and bills addressing only gender discrimination. Following the original manuscript, the sources relied upon were the THOMAS database and the

and Treul's revised counts are identified by the S3 variable.[4]

***District Partisanship***. In the original manuscript, a district's partisanship was measured as the Democratic vote-share from the most recent congressional election.[5] Legislators who ran unopposed were coded as hailing from districts with 100% vote-share for their party. That is, if a Democrat ran uncontested in her election, the Democratic vote-share in that election was imputed to be 100%. Districts in which Republicans ran unopposed, conversely, were recorded to have 0% Democratic vote-share.

There are reasonable alternatives for operationalizing district partisanship. We construct three variables in which districts where Democrats ran unopposed were recorded as having 90%, 80%, or 70% Democratic vote-share. (Likewise, districts where Republicans ran unopposed were imputed to have 10%, 20%, and 30% Democratic vote-share respectively.) These cutpoints reflect a series of arbitrary but plausible values of partisanship in districts with uncontested races. Additionally, we include two other common measures of district partisanship: Democratic vote-share from the most recent *presidential* election (e.g. Ansolabehere et al. 2001; Canes-Wrone et al. 2002) and the Cook Partisan Voting Index (e.g. Aldrich et al. 2017; Ellis 2013)
.

***Public Opinion***. Model 9 in the original article did not control for district-specific estimates of public opinion on same-sex marriage, while Model 11 did. We include possible combinations including and excluding this control.

***LGB Population***. Hansen and Treul measured district-level LGB population using estimates of same-sex unmarried[6] households from the U.S. Census' American Community Survey (ACS). The ACS provides district-level estimates using one year, three years, or five years of data. One-year estimates are based solely on data collected in the year of interest, while five-year estimates combine data across years to create more accurate figures, particularly in

---

Congressional Bills Project.

[4]In the original bill count measure S1, Hansen and Treul identified two Republican sponsors and thus included a control for the party of the bill sponsor. In the new measures S2 and S3, no Republican sponsors are identified. Therefore, a control for the sponsor's party is included in models containing S1, but excluded in models containing S2 and S3.

[5]In the original manuscript, this variable was misidentified as Democratic vote-share from the most recent presidential election.

[6]During the period of study, 2005-2010, only six states had legalized same-sex marriage for any period of time. Only Massachusetts allowed it before 2008.

small jurisdictions. The original article was inconsistent in employing one-year, three-year, or five-year estimates. We include an alternative measure using 1-year estimates for two reasons: (1) it is the only estimate type available for all terms in the data and (2) Census guidance recommends one-year estimates for jurisdictions with more than 65,000 people.

***Campaign Contributions***. The original article included lagged estimates of donations from a prominent national LGBT+ rights organization, the Human Rights Campaign. Such a control might imply a data-generating process wherein donations incentivize members of Congress to work on pro-LGB legislation. Elsewhere, literature on campaign finance implies that contributions flow to members who for other reasons are already working on legislation favorable to donors (see Dawood 2015). Following this logic, controlling for contributions would misspecify the data-generating process. We estimate models in which this control variable is excluded.

***Lagged DV and Random Effects***. Models reported in the main text of the original article pooled observations and relied upon term fixed effects and a lagged dependent variable to correct for autocorrelation. However, pooling the data ignores the clustering of observations within districts, perhaps biasing coefficient estimates. Another reasonable model specification would be to include district random effects with the lagged dependent variable removed to allow for proper specification. We include estimates from both model specifications.

# Results

In this section, we present the results from a series of negative binomial regression models using the variables and estimation techniques described in Table 1. The unit of analysis is member-term and the key outcome of interest is the count of pro-LGB bills sponsored in one term. In total, we consider 288 models, representing all combinations of the data processing choices ($3S \times 6D \times 2O \times 2P \times 2C \times 2L = 288$ possible models).[7]

The histogram in Figure 1 displays the $p$ values for the *LGB Population* variable across all model specifications.[8] The solid line corresponds to the conventional threshold for statistical

[7] In one instance, the negative binomial regression model failed to converge.

[8] As an alternative to $p$ values, analysts might consider reporting $t$-statistics. Unlike $p$, $t$-statistics can take

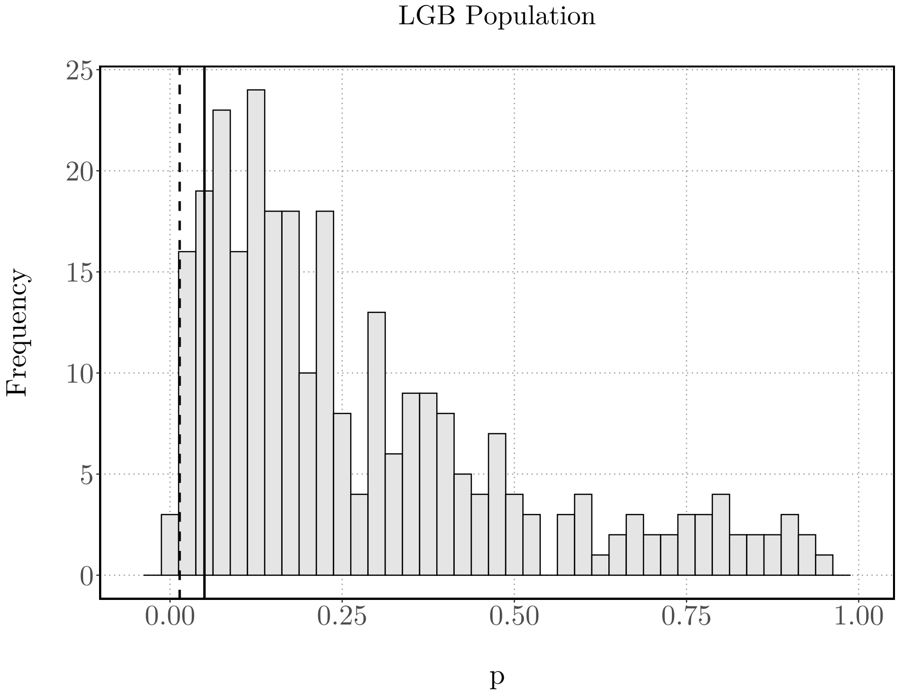


Figure 1: Histogram displaying the $p$ values for the *LGB Population* variable across model specifications. The solid line corresponds to $p = .05$ and the dashed line corresponds to the $p$ value of the *LGB Population* variable in Hansen and Treul (2015).

significance ($p = .05$) and the dashed line reflects the level of significance reached in the Hansen and Treul study ($p = .014$). We observe that the statistical significance of *LGB Population* varies considerably across the multiverse. In 31 cases (10.8%), the variable is found to be associated with pro-LGB bill sponsorship at the .05 level of significance. The coefficients from these models range from 1 to 1.588 with a mean of 1.28. However, in the vast majority of the specifications (89.2%), a district's lesbian, gay, and bisexual population is not significantly associated with pro-LGB bill sponsorships. Rather, we find a wide array of insignificant $p$ values ranging from slightly greater than .05 to nearly 1.0. The accompanying coefficients range from -0.34 to 1.37 with a mean of 0.6.

Turning to the other explanatory variables, we find that some follow the pattern described above while others are largely stable across the multiverse. As displayed in panels 2A and 2B, neither the public's opinion of same-sex marriage or campaign contributions from the Human Rights Campaign are consistently associated with LGB bill sponsorship. Whereas the

both positive and negative values, reflecting the signs of the coefficient estimates. Additionally, reporting $t$-statistics may more intuitively illustrate variation in significance levels in cases where $p$ varies by orders of magnitude (e.g. 0.001 vs. 0.1). We thank an anonymous reviewer for this insight.

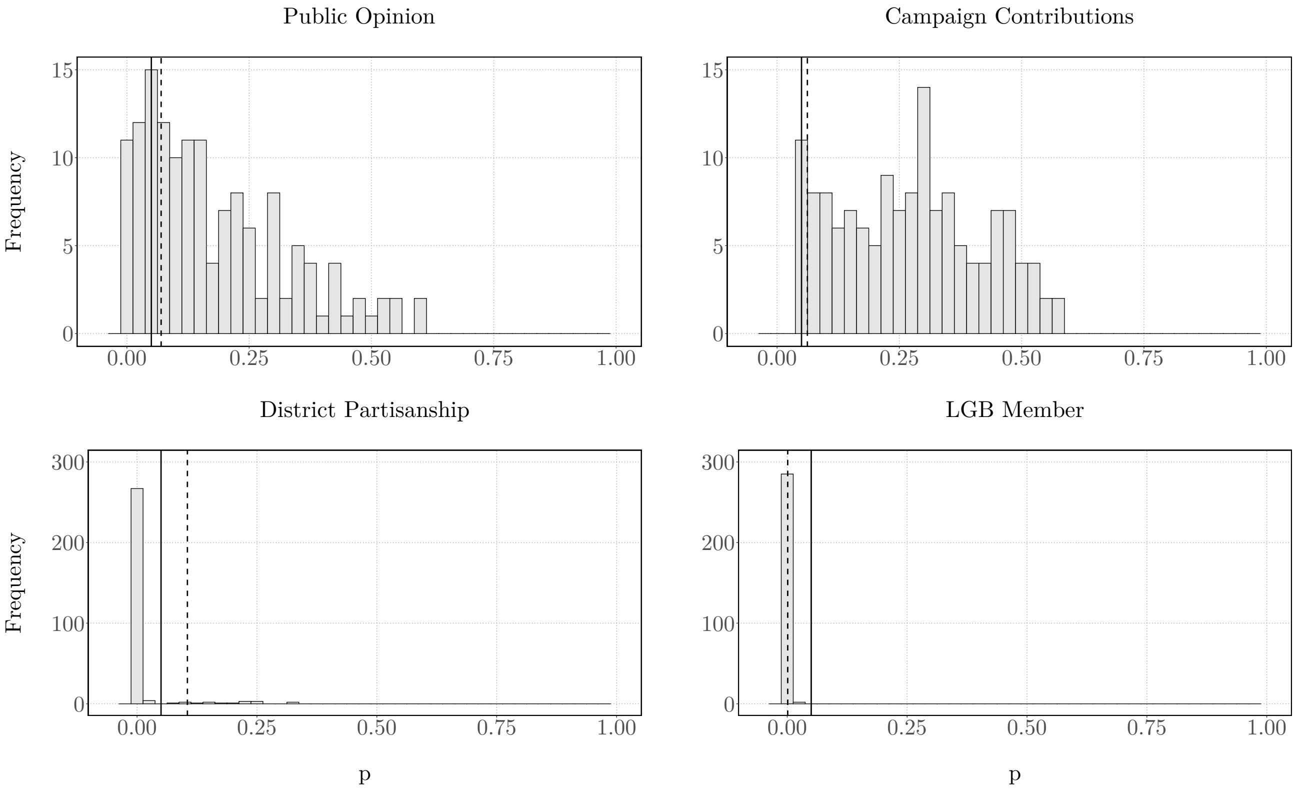


Figure 2: Histogram displaying $p$ values associated with the 287 specifications used to estimate the relationship between *Public Opinion* and *Sponsorships* (top left); *Campaign Contributions* and *Sponsorships* (top right); *District Partisanship* and *Sponsorships* (bottom left); and *LGB Member* and *Sponsorships* (bottom right). The solid lines correspond to $p =$ .05 and the dashed lines correspond to the $p$ value of each variable in Hansen and Treul (2015).

coefficient estimates for *Public Opinion* are statistically significant in 22% of our models, the estimates for *Campaign Contributions* reach significance in less than 5% of our models. In contrast, the results presented in panels 2C and 2D suggest that both the sexual orientation of the member and the partisanship of their district correlate with the sponsorship of LGB-rights bills. Specifically, *LGB Member* is significant in 100% of our models and *District Partisanship* is significant in 95.5% of them.

Figure 3 provides a more granular view of the results. Each line segment represents one of the data processing choices described in Table 1. As such, each cell is the unique combination of six choices: how the dependent variable, *Sponsorships*, is measured (**S**); how *LGB Population* is measured (**P**); how *District Partisanship* is measured (**D**); whether or not *Public Opinion* is included as a control (**O**); whether or not *Campaign Contributions* is included as a control

(**C**); and whether or not we include *Random Effects* (**L**).[9] For instance, the cell at [**S1**, **P1**, **C1**, **D1**, **O1**, **L1**] corresponds to the main model specifications from Hansen and Treul (2015). The color of each cell reflects whether the coefficient of the variable of interest is statistically significant (white) or insignificant (gray).[10]

This approach allows us to examine systematically how measurement choices correspond to statistical outcomes. Focusing first on *LGB Population* in panel 3A, we observe a cluster of significant coefficients in the top left corner of the diagram. Each of these corresponds to a regression estimate based on **S1**—the imprecisely measured dependent variable, *Sponsorships*, from the original study. Given that both the Saraceno (**S2**) and revised Hansen and Treul (**S3**) measures yield fewer significant results, it seems that the particular measurement of **S1** was more likely to produce a significant result for the *LGB Population* variable. That said, significant results using either **S2** or **S3** were obtained in 16 instances. The data also reveal that more than 83% of significant coefficients included either the **D1** or **D2** measure of *District Partisanship*. Using the alternative measures, **D3-D6**, overwhelmingly resulted in insignificant outcomes, regardless of other processing decisions. On balance, this evidence suggests that a district's LGB population is *not* associated with more bill sponsorships.

We also find that, when it was included in the model, *Public Opinion* was associated with sponsorships primarily in the random effects models (**L2**) that used the **S1** measure of *Sponsorships*. However, these results are not robust using the alternative measurements of sponsorships, **S2** and **S3**. Almost none of the models including the *Campaign Contributions* control show it to be a statistically significant predictor of pro-LGB bill sponsorship. In contrast, *District Partisanship* is insignificant only when it is measured as **D1** or **D2** in models of the **S1** dependent variable. Finally, the data show that descriptive representation, captured by the indicator *LGB Member*, is positively associated with the sponsorship of LGB bills across all models.

In the online appendix, we also model the relationship between these variables and pro-LGB bill sponsorships using logistic regression. The dependent variables are collapsed to binary indicators, with values of 1 indicating a member sponsored at least one pro-LGB bill and values

[9] We include a lagged dependent variable in all but the random effects models.

[10] The model that failed to converge is colored black.

of 0 indicating a member sponsored none. The logistic regression models yield results largely consistent with those above. *LGB Member* and *District Partisanship* are generally found to be significantly associated with *Sponsorships*, while *LGB Population*, *Public Opinion*, and *Campaign Contributions* are not.

# Discussion

Looking at the balance of evidence, the authors agree we should no longer conclude that there is a significant association between the size of the LGB population in a district and the bill sponsorship activity of its representative. Other factors like descriptive representation and district partisanship are more consistently associated with substantive representation, at least for this one minority group. The null findings raise questions about the power that small subconstituencies wield within political districts, questions that can be explored in future research.

The analysis above provides an example to political scientists of how multiverse analysis (Steegen et al. 2016) can be applied to assess the validity and replicability of findings. As researchers construct their models, they can keep a running tally of data processing decisions that might affect their conclusions. Multiverse analysis can then be employed to demonstrate the robustness of findings. Many researchers already provide a series of appendix tables that illustrate how different choices affect the results. Multiverse analysis provides a streamlined, elegant method of presenting robustness checks. It is perhaps a more comprehensive yet more parsimonious alternative to a lengthy appendix of results tables. Editors and reviewers might consider asking contributors to produce multiverse analyses of the principal findings for multiple regression models prior to publication.

Finally, we view multiverse analysis as a valuable tool in enabling the publication of replication and extension of prior work. Multiverse analysis allows authors and readers the opportunity to view all possible outcomes from data processing decisions, rather than the narrow outcomes produced by replicators who are, in turn, making their own (sometimes arbitrary) decisions. When disputes over findings arise after replication efforts contradict published work, journals may publish a small symposium featuring the replication analysis and a response from the original authors. However, an adversarial format aligns incentives for each side of the dispute to defend its own analysis, foregoing the opportunity to resolve issues jointly. Multiverse analysis provides a methodological framework that allows for collaboration between original authors and replicators. Moreover, if journal editors decide to publish such analyses, it offers a professional incentive for the original authors to participate in making their published results more transparent. Multiverse analysis is a useful tool to illustrate the robustness of findings, replicate prior work, and increase the transparency of quantitative findings in the field.

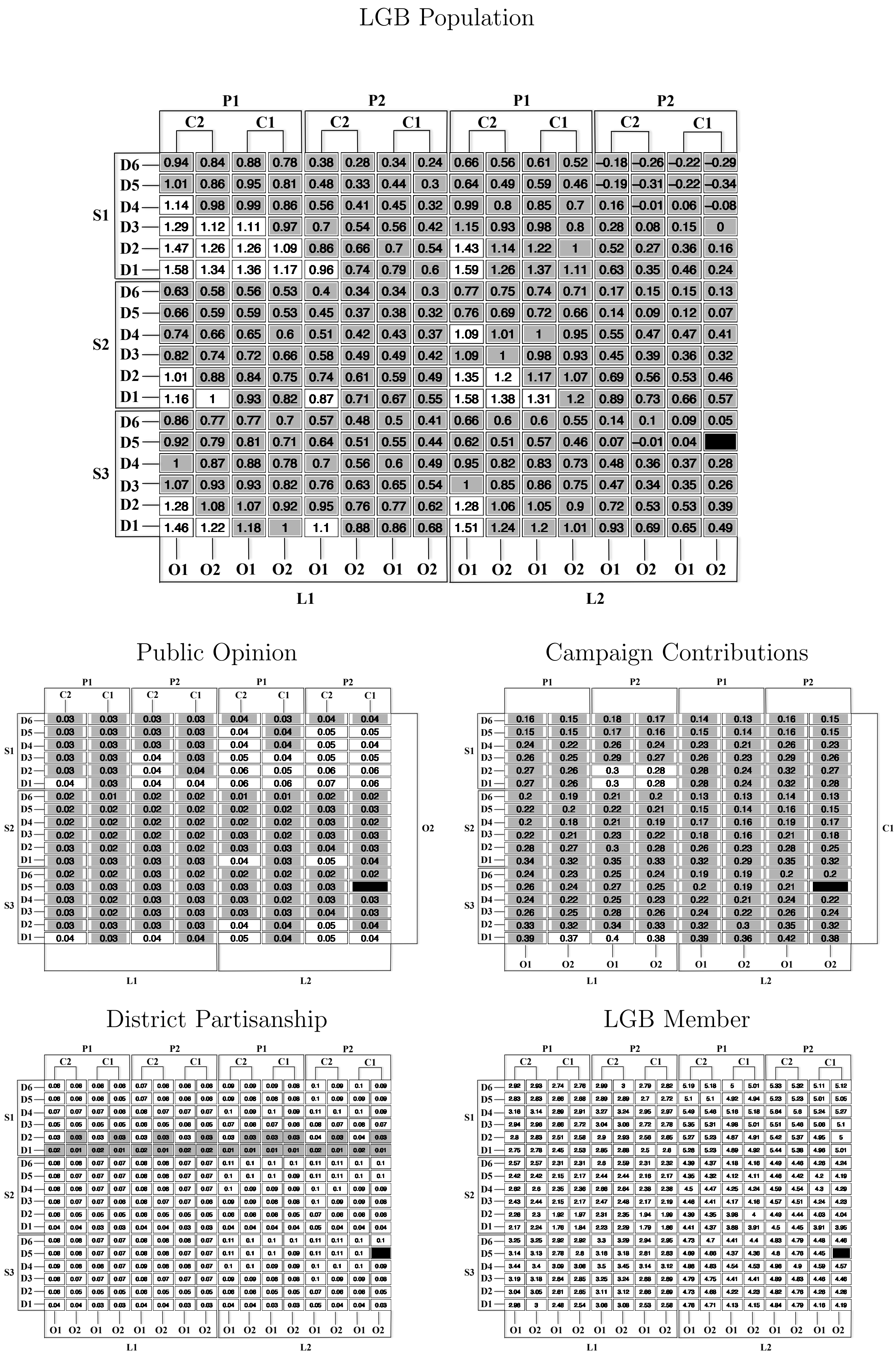


Figure 3: Multiverse of coefficient estimates for each independent variable. White cells indicate an accompanying $p$ value less than .05 while gray cells indicate a $p$ value greater than .05. Table 1 provides a list of variables and their abbreviations.

# References

Aldrich, John H., Andrew O. Ballard, Joshua Y. Lerner, and David W. Rohde. 2017. "Does the Gift Keep on Giving? House Leadership PAC Donations before and after Majority Status." *Journal of Politics* 79(4):1449–1453.

Ansolabehere, Stephen, James M. Snyder Jr., and Charles Stewart III. 2001. "Candidate Positioning in U.S. House Elections." *American Journal of Political Science* 45(1):136–59.

Canes-Wrone, Brandice, David W. Brady, and John F. Cogan. 2002. "Out of Step, Out of Office: Electoral Accountability and House Members' Voting." *American Political Science Review* 96(1):127–40.

Dawood, Yasmin. 2015. "Campaign Finance and American Democracy." *Annual Review of Political Science* 18:329–48.

Ellis, Christopher. 2013. "Social Context and Economic Biases in Representation." *Journal of Politics* 75(3):773–786.

Gelman, Andrew, and Eric Loken. 2014. "The Statistical Crisis in Science." *American Scientist* 102(5):460–5.

Hansen, Eric R., and Sarah A. Treul. 2015. "The Symbolic and Substantive Representation of LGB Americans in the U.S. House." *Journal of Politics* 77(4):955–67.

Leamer, Edward E. 1985. "Sensitivity Analyses Would Help." *American Economic Review* 75(3):308–13.

Miller, Michael K., Michael Joseph, and Dorothy Ohl. 2018. "Are Coups Really Contagious? An Extreme Bounds Analysis of Political Diffusion." *Journal of Conflict Resolution* 62(2):410–41.

Steegen, Sara, Francis Tuerlinckx, Andrew Gelman, and Wolf Vanpaemel. 2016. "Increasing Transparency Through a Multiverse Analysis." *Perspectives on Psychological Science* 11(5):702–12.