A Critique of *Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens* [1]

Michael T Da Silva*, Peter Cramer

July 9, 2024

Abstract

The conclusion drawn in *Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens* – the legislation passed in the U.S. is uninfluenced by the opinion of those not in the economic elite or business-oriented special interest groups – is supported by figures and tables which misrepresent the data. Upon reevaluation, it appears that there are insufficient data for the multivariate model used to differentiate between the three groups examined.

Additionally, there are critical defects with figures and statistics which, taken together with the failure to replicate the multivariate model, lead us to the conclusion that the original paper should be retracted.

1 Motivation and Outline of Critique

Rhetoric surrounding voting is becoming increasingly heated in the U.S. [3]. As such, an influential paper which has a figure which can be interpreted to mean that voting has no impact on policy adoption if one is not in the top 90th percentile of income or aligned with powerful interest groups is a potential political flash-point.

This analysis will begin with a close inspection of the model used by examining the tables given in the original paper. Then, we will attempt to regenerate the model in question to elucidate what methods were used to create it and what assumptions were baked in. Finally, the narrative and conclusions will be examined in light of the model and tables.

The Python and data used in this analysis can be found on Github at testing-theories-of-american-politics-data-analysis. Additionally found there is an R script provided by the authors (though used in a separate paper for a similar

---

*mdasilva@callutheran.edu

analysis but given as an example of their process). The software used for the authors' analysis was IBM's Amos package such that more tailored statistical analysis may have been performed but if so, we will here show that it was not properly communicated.

2 Acknowledgments

3 Tables 3 & 4 - p-scores and R^2

To begin, we will explore tables 3 and 4 which relate to both theoretical economic models outlined in the paper and the data. For more information on the economic models themselves, refer to the paper but the criticism contained herein is comprehensible without an understanding of them. For a brief refresher on p-scores and R^2 values, please refer to sections A and B respectively.

Table 3
Policy outcomes and the policy preferences of average citizens, economic elites, and interest groups

	Model 1	Model 2	Model 3	Model 4
Preferences of average citizens	.64 (.08)***	—	—	.03 (.08)
Preferences of economic elites	—	.81 (.08)***	—	.76 (.08)***
Alignment of interest groups	—	—	.59 (.09)***	.56 (.09)***
R-sq	.031	.049	.028	.074

***p<.001

Note: All predictors are scaled to range from 0 to 1. The dependent variable is the policy outcome, coded 1 if the proposed policy change took place within four years of the survey date and 0 if it did not. Predictors are the logits of the imputed percent of respondents at the fiftieth ("average citizens") or ninetieth ("economic elites") income percentile that favor the proposed policy change, and the Net Interest-Group Alignment Index described in the text. Standard errors are asymptotically distribution-free, and all analyses reflect estimated measurement error in the predictors, as described in Appendix 2. The standardized coefficients for model 4 in this table are .01, .21, and .16 for average citizens, economic elites, and interest groups, respectively. N=1,779.

Figure 1: Table 3 from *Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens* is here shown. The same criticism will be levied against both tables 3 and 4 such that repeating it twice is counterproductive. Please refer to table 4 (figure 2) for the full critique but it applies equally here.

Table 4
The separate policy impact of business-oriented and mass-based interest groups

Average citizens' preferences	.05
	(.08)
Economic elites' preferences	.78
	(.08)***
Mass-based interest groups	.24
	(.07)***
Business interest groups	.43
	(.08)***
R-sq	.07

***p<.001

Note: All predictors are scaled to range from 0 to 1. The dependent variable is the policy outcome, coded 1 if the proposed policy change took place within four years of the survey date and 0 if it did not. Predictors are the logits of the imputed percent of respondents at the fiftieth ("average citizens") or ninetieth ("economic elites") income percentile that favor the proposed policy change, and the Net Interest-Group Alignment Indices described in the text. Standard errors are asymptotically distribution-free, and all analyses reflect estimated measurement error in the predictors, as described in Appendix 2. N=1,779.

Figure 2: Table 4 from *Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens* is here shown. In particular we should focus on two numbers on this table: $p < .001$ and R-sq (R^2) = 0.07. Briefly, p-score is a measure of how confident one can be that their measurement is within a distribution (the lower, the more likely it is from a *different* distribution; this is the desired condition to prove significance), and R^2 is a measure of how well a model fits data (the closer to 1 the better the fit). As such, having a p-score so low (which indicates high confidence) and an R^2 of nearly zero (which indicates no correlation), we can be confident that the model used to generate this table fails to capture trends in the data.

With that, we cannot draw any further conclusions from tables 3 or 4. Unfortunately, these tables are critical for the authors' arguments such that they cannot be used as proof and may even contradict the authors' conclusion.

To further emphasize this point, figure 3 demonstrates data that has an R^2 of a similar value as these tables but is obviously a poor fit.
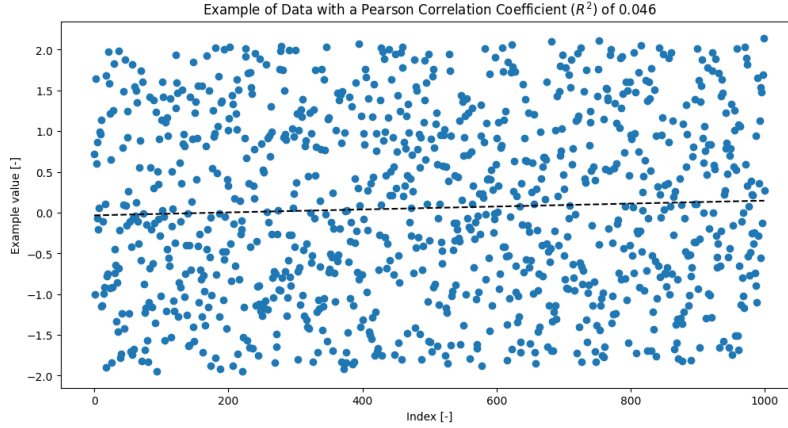
4

Figure 3: Example randomly generated data with an R^2 on par with those found in tables 3 and 4 are here shown. Hopefully it is plain to see that concluding the fitted line (black dotted) is a good representation of the data is false. For further info on R^2 values, their meaning, and calculation, please refer to section B.

4 Attempted Model Regeneration

4.1 Regenerating Exactly as Defined in Author-Provided R Script

In communicating with Dr. Martin Gilens, an R script was provided which includes a Structural Equation Model (SEM) which ostensibly reproduces the results obtained in the original paper. This model was run in both R (using Lavaan [8]) and Python (using semopy [9]) and the result obtained was nothing close to that expressed in tables 3 and 4 of the original paper.

The model used in both cases is the following (here in python):

```python
from semopy import Model
model = Model(
  description="""
    # measurement
    L5 =~ 1 * rangepred50
    L9 =~ 1 * rangepred90
    IG_bal =~ 1 * rangeinterest

    # regressions
    policy_change  ~ L5 + L9 + IG_bal

    # cov and var
    rangepred50  ~~ 0.0026380 * rangepred90
    L5  ~~ IG_bal
    L9  ~~ IG_bal
    L5  ~~ L9
```

```
    """ ,
    mimic_lavaan=True ,
)
```

Listing 1: SEM model with assumed, manual covariance between rangepred50 and rangepred90

where

1. L5 is a name substitution for rangepred50 which is the normalized and logistical quartile probability of policy adoption for the 50th percentile given by the following transformations:

   ```
   rangepred50 = normalize ( qlogis ( raw_data [ " pred50_sw " ] ) )
   ```

   where "normalize" is given by the simple equation $\tilde{x} = \frac{x - \min(x)}{\max(x) - \min(x)}$, "qlogis" [10] is the slightly more complex

   $$\text{qlogis}(x) = \frac{1}{1 + e^{-\frac{x - \mu}{\sigma}}}$$

   and "pred50_sw" is the proportion of the 50th economic percentile in favor of the piece of legislation passing – calculated per-piece of legislation. In this case, μ and σ are assumed to be 0 and 1 respectively[1].

2. L9 is a name substitution for rangepred90 which is transformed identically to L5.

3. IG_bal is a name substitution for rangeinterest which is calculated using the following equation[2]:

   $$\text{net interest group alignment} = \ln \left(\frac{\text{\# Strongly favor} + \frac{1}{2}\text{\# Weakly favor} + 1}{\text{\# Strongly oppose} + \frac{1}{2}\text{\# Weakly oppose} + 1} \right)$$

   This "net interest group alignment" is then normalized between 0 and 1 to yield IG_bal:

   $$\text{IG\_bal} = \text{normalize}(\text{net interest group alignment})$$

4. policy_change is declared to be a function of L5, L9, and IG_bal

5. rangepred50 is declared to have an artificially small dependence on rangepred90 (e.g. if rangepred90 changes by 1, this line asserts that rangepred50 should

---

[1]This may be a bad decision because it means all data vectors to which this is applied will be mapped to something gaussian-like and then normalized which sounds like it introduces artificial similarity. Normalization is necessary for the multivariate analysis but the additional use of "qlogis" with $[\mu, \sigma] = [0, 1]$ may be erroneous.

[2]This equation visually differs from that defined in the original paper, but using the log rule $\ln(a/b) = \ln(a) - \ln(b)$, we can obtain this slightly shorter equation.

change 0.0026380)[3]. This is asserting the conclusion that rangepred90 is the more significant factor. Despite that artificial assumption, we will see in figure 4 that it still does not regenerate the result obtained in the original paper's tables 3 and 4 and the optimizer finds the contradictory conclusion.

6. L5 is defined to be correlated with IG_bal

7. L9 is defined to be correlated with IG_bal

8. L5 is defined to be correlated with L9

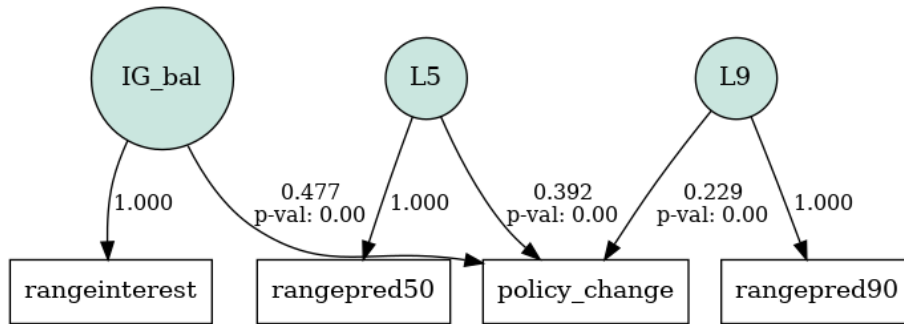When this is run, the following relationships between the three variables (L5, L9, IG_bal) are obtained:

Figure 4: An attempt at regenerating the multivariate model used in *Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens* is here shown. Please see listing 1 for a definition of the model and the corresponding explanation for variable name descriptions. The arrows, however, signify the following:
- IG_bal (i.e. interest groups) has the largest influence on the probability of policy change adoption (0.477)
- L5 (i.e. the 50th income percentile) has middling influence on the probability of policy adoption (0.392)
- L9 (i.e. the economic elite) has the smallest influence on the probability of policy adoption (0.229)
This contradicts the conclusion of the original paper.
It should be noted, however, that wherever an optimizer is involved, many results may follow. It is unclear how IBM's AMOS package may have optimized this. It does, however, show that the result is likely an artifact of how few data are present.

---

[3]At least the optimizer should start with this assumption; it is allowed to optimize away from this relationship).

A more complete look at the relationships between these three variables also shows the most critical feature:
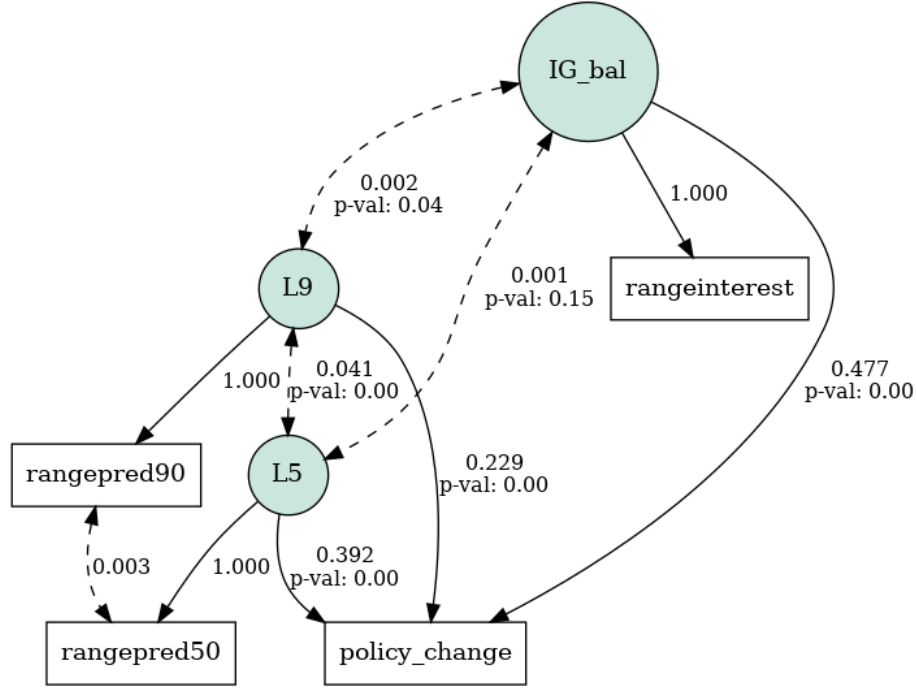

Figure 5: An more complex relationship plot for the model generated in figure 4 is here shown. Please see listing 1 for a definition of the model and the corresponding explanation for variable name descriptions. **The critical observation to make here is that the p-val for the relationship between L5 and IG_bal is quite large (0.15) such that we cannot conclude that they are separate distributions.** As such, this model fails to differentiate between the policy preferences of the 50th percentile and interest groups (at least in the way the data have been preprocessed).

In conclusion, the model as defined in the R script provided by Dr. Martin Giles fails to reproduce the model used in the generation of figure 1 and tables 3 and 4 of the original paper. Additionally, the model here developed has insufficient data to differentiate between the groups under examination such that it is unlikely that, even with RNG cherry-picking, it is possible to draw a substantial conclusion

Without that model, the result of the paper is unsubstantiated.

4.2 Regenerating without Assumed Covariance

Removing the manually assumed covariance from listing 1, we obtain the following model:

```python
from semopy import Model
model = Model(
  description="""
    # measurement
    L5 =~ 1 * rangepred50
    L9 =~ 1 * rangepred90
    IG_bal =~ 1 * rangeinterest

    # regressions
    policy_change ~ L5 + L9 + IG_bal

    # cov and var
    L5 ~~ IG_bal
    L9 ~~ IG_bal
    L5 ~~ L9
  """,
  mimic_lavaan=True,
)
```

Listing 2: SEM model without assumed covariance between rangepred50 and rangepred90

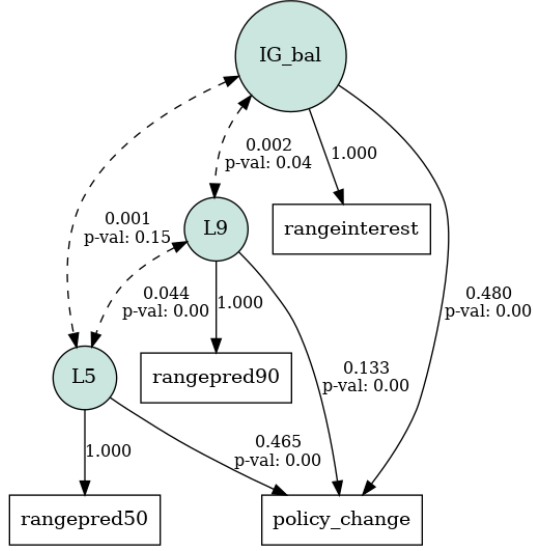Using this model, a similar result to figure 5 is obtained:

Figure 6: The model result from listing 2 is here shown. Please see listing 2 for a definition of the model and the corresponding explanation for variable name descriptions from listing 1. **Here we can see again (w.r.t. figure 5), that L5 and IG_bal cannot be differentiated and they both have a much more significant impact on the probability of policy adoption than L9.**

In conclusion, even the assumed covariance between rangepred50 and rangepred90 cannot explain the model obtained by the authors in the original paper; it is possible that manipulating the RNG seed could lead to the authors' original conclusion but even if such RNG-hacking were used, because L5 and L9 are so strongly correlated in the original dataset, it is likely that the p-value which currently marks IG_bal and L5 as undifferentiable would instead show IG_bal and L9 to be undifferentiable.
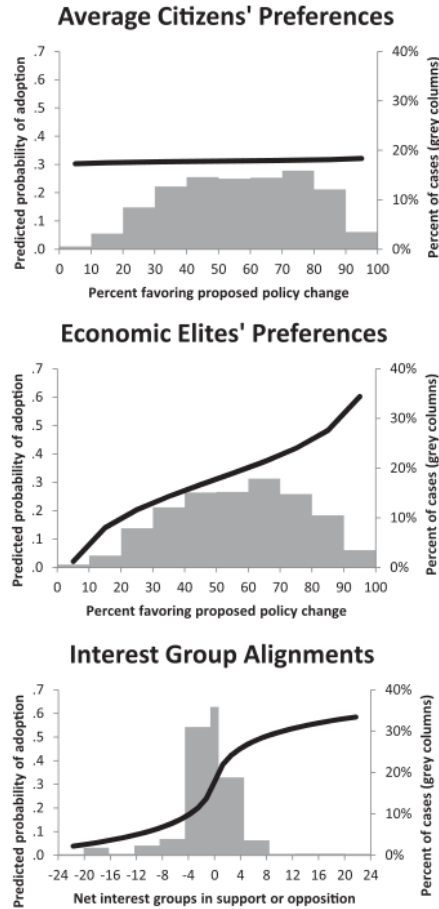
10

5 Inspection of Figure 1

Figure 7: Figure 1 from *Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens* is here shown. From a clarity standpoint, figures should, ideally, be able to stand by themselves because they are most likely to be taken out of context. As such, a more comprehensive figure caption would be appreciated such that the narrative is clearly explained along with the analysis method performed to generate these plots.

To begin, we must understand how this figure was generated. Based on page 572R, these figures were created using the multivariate model. In the previous sections, however, we have shown that the multivariate model is either deeply flawed or misleading. As such, we should be able to discount figure 1 immediately but upon closer inspection, additional concerns are raised.

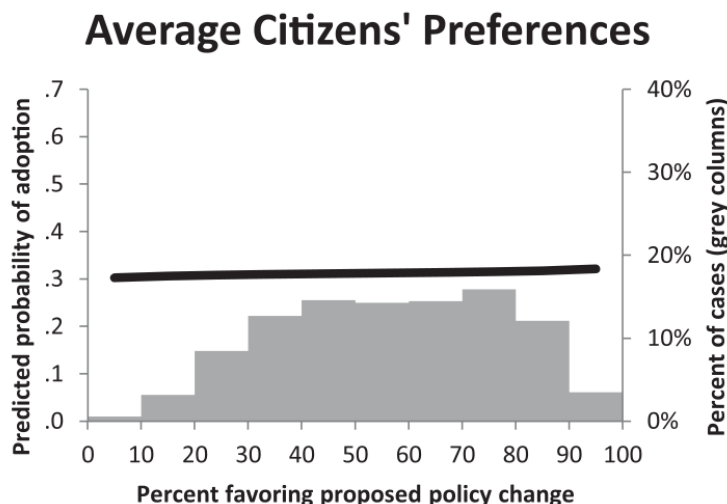## 5.1   Figure 1a - Average Citizens' Preferences

Figure 8: Figure 1a from *Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens* is here shown. An explanation of the axes follows: The x-axis is the proportion of people in this income percentile (here the 50th percentile) who want the legislation to pass; the 90-100 bracket therefore indicates that 90-100% of survey respondents in the 50th income percentile want the legislation to pass.
The left axis indicates the probability that the legislation will pass for that preference bracket; if 90-100% of 50th income percentile want legislation to pass, this plot indicates there is a slightly greater than 30% chance of it passing.
The right axis indicates the number of cases examined in this dataset which have that proportion of 50th income percentile respondents favoring the policy being adopted; out of these ≈1800 cases examined, about 4% of them had 90-100% of the 50th income percentile favoring their adoption.

Nothing about this figure immediately stands out as problematic beyond the model that is used to generate it. There is a reasonable distribution of cases (see figure 8 for an explanation of axes) across percent favoring brackets and it is approximately normal.

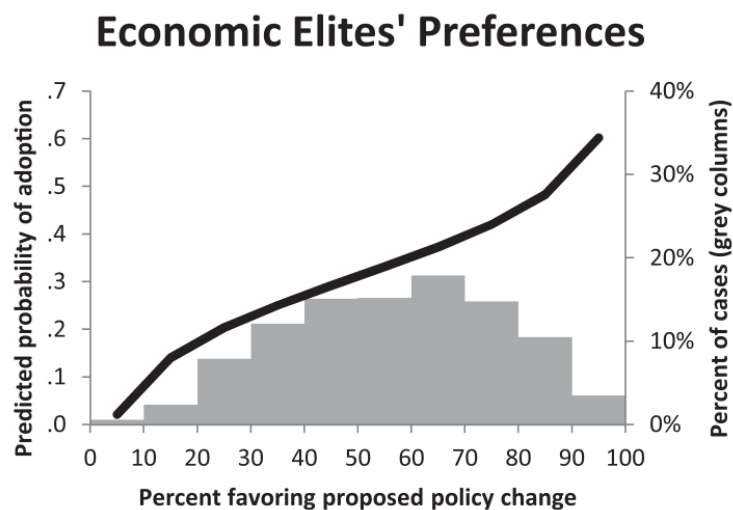## 5.2   Figure 1b - Economic Elites' Preferences

Figure 9: Figure 1b from *Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens* is here shown. Here we can plainly see (in contrast to figure 8) that as more economic elites prefer adopting a policy, the greater the chance that policy is adopted.

Nothing about this figure immediately stands out as problematic beyond the model that is used to generate it.
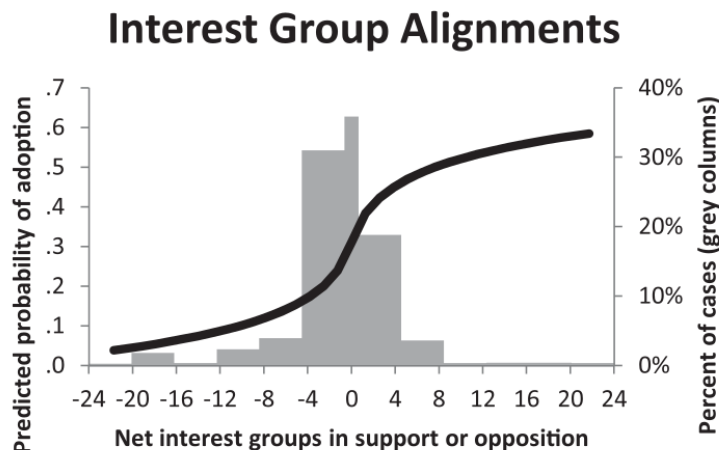
5.3 Figure 1c - Interest Group Alignments

Figure 10: Figure 1c from *Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens* is here shown. Without delving into the text, several issues stand out.

The data are approximately normal about 0 but the plot extends beyond where there are data. A plot between [-6, 6] would have been sufficient to hypothesize that interest group alignment is correlated with legislative outcome without extrapolating into a sparse or absent data region.

Given how few data are in some regions, even though this is the output of a multimodal model, this is clearly impossible. The choice of a logistic/sigmoid function for this interpolation/extrapolation additionally encodes an assumption which may not be valid.

Additionally, the equation for net interest groups as given on page 569R of the original paper (shown in section 4 and included again below this caption), cannot have been here used. That equation simplifies to Net Interest Group Alignment $= \ln(a/b)$ where a and b are functions of how many interest groups support, a, or oppose, b, a policy. If we examine the extrema of the axes, $[e^{-24}, e^{24}] \approx [0, 2.6E^{10}]$. There are not $2.6E^{10}$ humans on Earth, nevermind interest groups even if we double them for being strongly in favor or opposed. Even within the center of highest density, $e^4 \approx 54$ is a greater additive favor score, a, than any policy received in the dataset.

As a final sanity check for the data distribution, please find a histogram of the net interest group calculated as a simple sum in figure 11.

The equation ostensibly used in the generation of the x-axis for figure 10 follows:

$$\text{Net Interest Group Alignment} = \ln(\#\text{ Strongly Favor} + 0.5 * \#\text{ Somewhat Favor} + 1)$$
$$- \ln(\#\text{ Strongly Oppose} + 0.5 * \#\text{ Somewhat Oppose} + 1)$$
$$= \ln\left(\frac{\#\text{ Strongly Favor} + 0.5 * \#\text{ Somewhat Favor} + 1}{\#\text{ Strongly Oppose} + 0.5 * \#\text{ Somewhat Oppose} + 1}\right)$$
$$= \ln(a/b)$$

See figure 10's caption for an explanation of why this cannot have been used in figure 10.
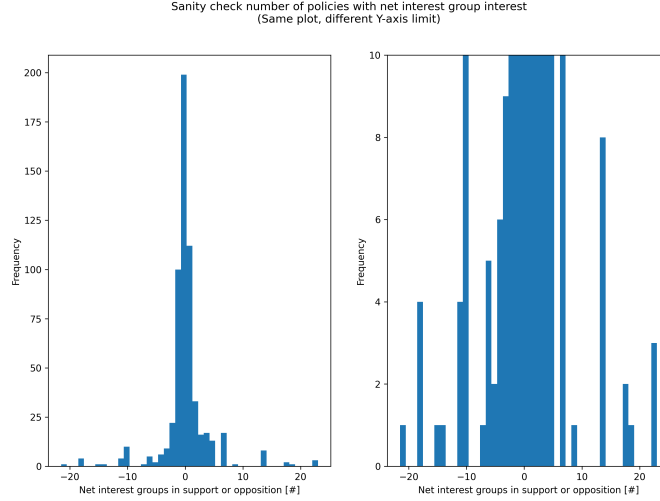

Figure 11: As an additional demonstration of the sparsity of net interest group data, a finer histogram is here shown. The left and right plots in this figure contain the same data but different y-axis limits to allow us to examine the extrema of the x-axis.
From this, we can clearly see there are insufficient data outside of the central spike to make a reasonable claim to statistical significance.

Finally, as a logical critique, there is little reason to believe that the more interest groups which are in favor or oppose a piece of legislation, the greater the chance it will pass. If an interest group with the weight and coffers of the NRA had a strong opinion about a piece of legislation such that the riches of Timbuktu were spent getting it passed, it likely matters little that a small interest group opposes them. This metric (net interest group alignment) gives both interest groups the same weight, differing only if one "strongly" or "somewhat"

agrees/disagrees[4]. Without a deeper understanding of why interest group alignment is a valid metric, it is more harmful than helpful; a more thorough study would need to establish that alignment is more important than something like inflation-adjusted lobbying or advertising.

5.4   Figure 1 as a Whole

All the figures are called into question because of their dependence on a faulty multivariate model, but figure 1c in particular has serious omissions of methodology and unfounded extrapolation which contaminate its use in the narrative.

6   Conclusion

Given the flaws with the multivariate model attempting to connect economic group, interest group alignment, and probability of policy adoption – R^2s of nearly zero, a $p - value > 0.05$ to differentiate between groups, and an overall failure to replicate – the argument presented in *Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens* is at best sentimental. There are data present which may, upon a more thorough analysis with more data in which groups disagree, support their claim, but there is currently a dearth of data for the SEM method provided by the authors to draw a conclusion.

Additionally, there are critical flaws in figure 10 which, even without relying on a potentially invalid model, call into question the validity of the analysis performed. Namely, extrapolating into regimes in which there are no data present and falsely claiming a formula was used for figure axes.

As such, we recommend that *Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens* be retracted for issues with its figure, analysis, and failure to replicate.

---

[4]And yet even this critique is assuming that money is a more powerful predictor of adopting legislation which is also not here proven. It may be so shown in other papers but it is not here a given but is unnecessary to complete the critique of net interest group alignment as a valid metric.

References

[1] Gilens, M. and Page, B.; *Testing Theories of American Politics: Elites, Interest Groups, and Average Citizens*; Perspectives in Politics, 2014; doi:10.1017/S1537592714001595

[2] Kvalseth, T.; *Cautionary Note about R^2*; The American Statistician, 1985; doi:10.2307/2683704

[3] LastWeekTonight; *Voting Rights: Last Week Tonight with John Oliver (HBO)*; Video Media; 2021-09-26; Accessed 2024-05-23; link

[4] OpenStax College; *Introductory Statistics*, Chapter 9; OpenStax College; 19 September 2013; link; ISBN-13 978-1-938168-20-8

[5] StatQuest with Josh Starmer; *p-values: What they are and how to interpret them*; Video Media; 2020-04-22; Accessed 2024-05-13; link

[6] StatQuest with Josh Starmer: *How to calculate p-values*; Video Media; 2020-04-22; Accessed 2024-05-13; link

[7] Wikipedia; *Coefficient of Determination*; Wikimedia; Edited 2024-04-19; Accessed 2024-05-23; link

[8] Rosseel, Y.; *Lavaan: Latent Variable Analysis*; CRAN; Version 0.6-18, published 2024-06-07; 10.32614/CRAN.package.lavaan

[9] Igolkina, A and Meshcheryakov, G.; *semopy: A Python Package for Structural Equation Modeling*; Structural Equation Modeling: A Multidisciplinary Journal; 2020; 10.1080/10705511.2019.1704289

[10] R Documentation; *Logistic: The Logistic Distribution*; DataCamp; stats, Version 3.6.2; link

A   p-score

Colloquially, p-scores (or p-values) are a measure of the confidence that a measurement is **not** significant. A more strict definition is "the probability of getting a test result [given an assumed distribution]". If the probability is very small, it is unlikely a value comes from that assumed distribution such that you can claim the value must come from a different distribution. In statistics parlance, this is "rejecting the null hypothesis" [4, 5, 6].

Most importantly, what it does not do is tell one how different two groups are. This tells us confidence, not difference [5, 6].

To calculate the (double-sided) p-score, we consider two things:

1. Probability that this happened (or more extreme) given an assumed distribution

2. Probability that something equally rare (or more extreme) happened given an assumed distribution

If the p-score is below some arbitrary threshold, α (usually 0.05 but again, this is **arbitrary**), then we "reject the null hypothesis" and conclude that the datum in question must come from a different probability distribution.

This is easier to see given an example; two will here be explored.

A.1   Coins

To begin, let us consider the case of flipping 3 heads on a **presumably fair** (this is the assumed probability distribution: a 50/50 H/T split) coin[5].

1. Probability that this happened (or more extreme) given an assumed distribution $\Rightarrow \frac{N_{HHH}}{N_{2HT}+N_{H2T}} = \frac{1}{3+3} = \frac{1}{6}$

2. Probability that something equally rare (or more extreme) happened given an assumed distribution $\Rightarrow \frac{N_{TTT}}{N_{2HT}+N_{H2T}} = \frac{1}{3+3} = \frac{1}{6}$

As such, $p = \frac{1}{3} = 0.3333\bar{3}$. This is not below our arbitrary threshold of 0.05 such that we cannot reject the null hypothesis such that we cannot conclude this coin follows a different probability distribution than 50/50 H/T.

How many heads (or tails) in a row are necessary to draw that conclusion?

One may recognize "1 3 3 1" as a row in Pascal's triangle and for a fair coin you would be correct! We can use that to help us here. Knowing that the numerator of p will, for a fair coin flipping all heads or tails, be 2, we just need to find the sum of the internal rows such that $p = \frac{2}{\sum_2^{N-1} P_i} < 0.05$.

This yields the 7th row of Pascal's triangle such that we need 6 heads or 6 tails in a row to conclude that our coin is not fair (using an arbitrary threshold of $p < 0.05$ which is again **arbitrary**).

---

[5]For all statistics must be taught with coins, cards, and dice to... discourage gambling?

A.2   Normal Distributions

Consider a normal distribution with $\mu = 155.7$ and $\sigma = 6.65$, if one measures a value of 142.4, does this value come from this distribution?

142.4 is exactly 2σ below the mean such that we know 2.5% of the distribution is at this value or below (more extreme). Additionally, the equally extreme value is 2σ above the mean (169) which also contains 2.5% of the distribution at or greater than 169.

As such, $p = 2.5\% + 2.5\% = 0.025 + 0.025 = 0.05$ which is exactly our p-score requirement. In an edge case like this, no conclusion can be drawn using our arbitrary threshold choice[6].

B R^2

R^2 is a measure of the "goodness of fit" for a model.

R^2 usually lies in the range $[0, 1]$ although there are several definitions, some of which can be negative [2]. All have the interpretation that a low number represents that the model fails to capture a correlation in the data and a high number represents that it successfully captures a correlation.

The most common definition of R^2 is given as the following [7]:
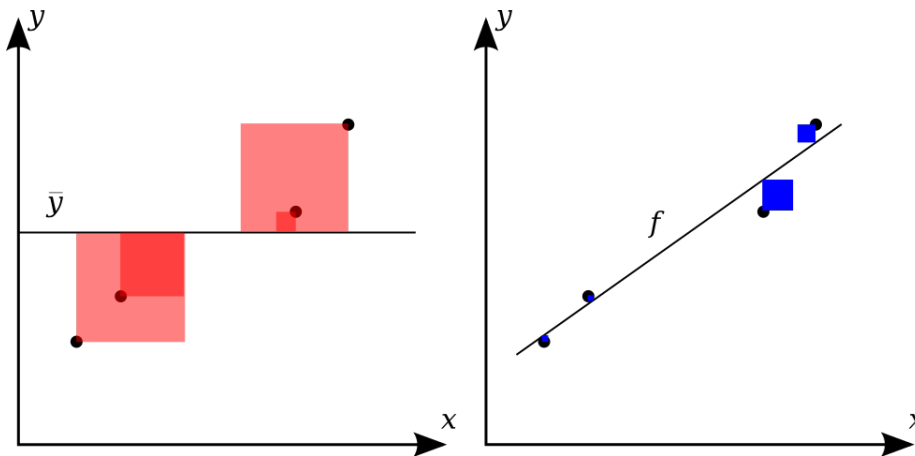
$$R^2 = 1 - \frac{\sum_i (y_i - f(x_i))^2}{\sum_i (y_i - \bar{y})^2}$$

---

[6] This scenario was chosen in particular because it shows from where the 0.05 arbitrary threshold comes; it is the 2σ threshold. If one needs even greater confidence, making the p-score requirement smaller implies your datum is even farther out on the fringes of a normal distribution and is therefore more likely part of a different distribution.

As we can see, if the blue term goes to zero, then all the data lie on the line $f(x)$ such that the fit captures the correlation well. If the blue term goes to infinity, we see that the fit becomes progressively worse and, in this case, R^2 becomes negative.