ECON 0150 Written Report

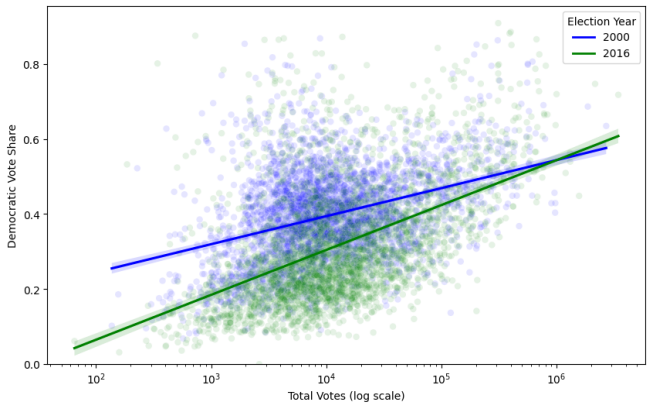
Nick Sobolewski and Rylie Teets

Part 1:

Our question concerned how partisan alignment has changed in U.S. counties over the past couple of decades. Specifically, we decided to look at the years 2000 and 2016 so we could see a more dramatic shift in voting patterns. The question we settled on was "How has the Democratic vote share in US Presidential Elections changed across counties from 2000 to 2016, and do these trends point to widening urban–rural polarization?"

Part 2:

We got all of our data from the County Presidential Election Returns 2000-2024 dataset in the MIT Election Lab - a very credible source. The dataset is very large, so we needed to do some work on it before we could use it. First, we filtered for only Democratic votes because we realized that we only needed to look at the data for one of the major parties to see shifts in partisan alignment. Then, we created a column called "d_share" that stored the proportion of votes that went to the Democratic candidate by dividing the Democratic votes by the total votes in the county. After that, we filtered for only the years 2000 and 2016 and created a column called "now," which had a 1 if the year was 2016 and a 0 for 2000. Finally, we created a column called "log_votes" that took the log of total votes to make the data easier to see. Ultimately, we looked at the relationship between total votes and Democratic share, with a regression line for each year to see changes between them.



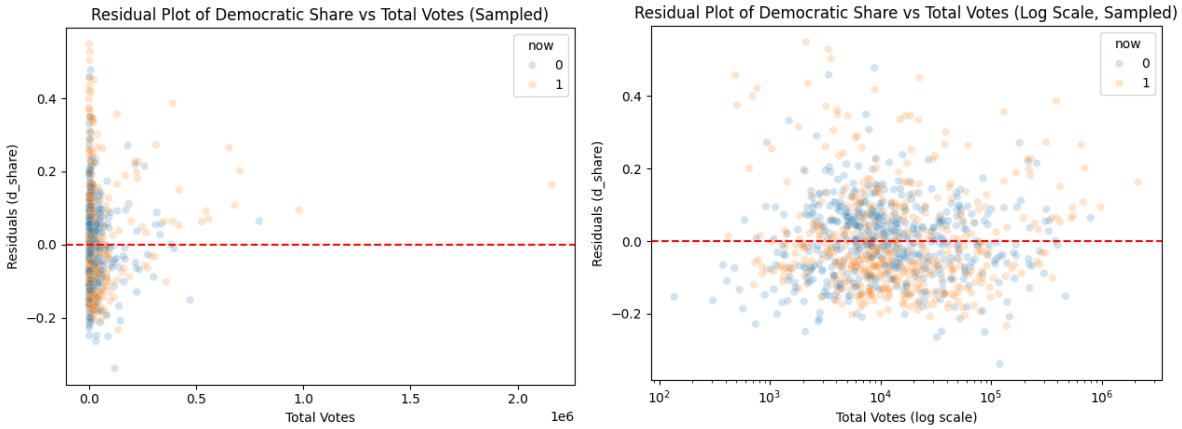
Part 3:

Our GLM uses the year (2000 or 2016) and the log votes to show a difference in Democratic vote share. Our β0 (0.0955) represents the baseline Democratic share for a small county in 2000. Next, our β1 (0.0324) represents the slope of our line (Democratic vote share against log votes) in 2000, which was small but still illustrated a clear trend that larger, urban counties vote more heavily for the Democratic candidate than smaller, rural counties. Then, our β2 (-0.2697) shows how the intercept shifted in 2016 - a huge difference, which shows that in 2016, counties were on average 27 percentage points less Democratic. Finally, our β3 (0.0195) represents the change in slope between the 2 years. Since the number is moderately sized and positive, it reveals that urban counties became more Democratic relative to rural ones - even though on average, all counties voted less Democratic in 2016. In addition, all of our P-values were near-zero, meaning our data is statistically significant. This shows a clear widening of urban-rural polarization.

Part 4:

The results of our model were significant and showed that voting trends in 2000 and 2016 were significantly different. The intercept of the regression line for 2016 (β_2) was 0.2697 less than the intercept (β_0) for 2000 (0.0955), with a p-value < 0.001 showing these to be significantly different. The additional slope for 2016 (β_3) was 0.0195, which also had a p-value < 0.001, making it significantly different from the slope of 2000. This may seem like a very minimal difference, but in comparison to the slope of 2000 (β_3), which is 0.0234, this is a substantial portion of the original slope.

The original residual plot is clearly heteroscedastic, so a log scale transformation was applied to the total votes to get a clearer view of the residuals. The transformation was successful, and the plot is now homoscedastic. This transformation overall helped to make the data interpretable, as before the points were too close together to draw any conclusions.



Part 5:

When we observed the relationship between Democratic vote share and total county votes for the years 2000 and 2016, we saw a clear shift in partisan alignment between the two years. Both the range of vote share and the slope were larger in 2016. The large difference between the highest and lowest Democratic vote shares signals increasing political polarization, while the difference in slopes shows a much larger difference between the voting trends of rural voters (left of the graph) and urban voters (right of the graph). Finally, our P-values for all of our calculations were nearly 0, suggesting a high confidence level in our findings.

References:

https://dataverse.harvard.edu/dataset.xhtml?persistentId=doi:10.7910/DVN/VOQCHQ