Did Obamacare enrollees vote for Trump?

Oren S. Savir

Introduction

The purpose of this paper is to determine whether enrollment in the Affordable Care Act ("ACA") can be used an effective predictor of President Donald Trump's electoral successes in 2016 on a county-by-county basis.

In December of 2016, Vox's Sarah Kliff traveled to Corbin, Kentucky to interview a unique class of voters — supporters of Donald Trump who were also enrolled in the Affordable Care Act, also known as Obamacare. Among Kliff's interviewees was an Obamacare enrollment counselor — someone who's job it is to enroll people in Obamacare — who herself voted for Donald Trump. Kliff's insight from these voters was that they were glad to have health insurance, but they were upset that their unemployed neighbors qualified for better healthcare through Medicaid, and they desired change in Washington. Desiring change in Washington, they cast their votes for Donald Trump on the gambit that his calls for repeal of Obamacare were empty rhetoric.

But Kliff only interviewed a handful of voters, and the insight she brought back from Corbin can hardly be assumed to be statistically significant. As such, the aim of this paper is to analyze the relevant data and determine whether there is a statistically significant correlation between the volume of enrollment in Obamacare and Trump's share of the vote in each county. This hypothesis rests on the assumed logic that Obamacare enrollees largely fit the demographics that supported trump — older, white inhabitants of rural or rust-belt regions with no college education.

Methods and Data

This paper relies on data obtained from the Department of Health and Human Services ("HHS") regarding the volume of Affordable Care Act plan selections as of March 2016 by county; the 2016 Presidential election results by county; demographic data obtained from the Census Bureau's 2015 American Community Survey; and unemployment rate data from the Bureau of Labor Statistics ("BLS").

It is important to note that only 38 U.S. states currently utilize the Affordable Care Act's marketplace exchange; the remaining states have all implemented their own exchanges, and therefore these 12 state are not included in the HHS plan selections data. Among these 12 states is Kentucky, the subject of the Vox story — so, unfortunately, no direct affirmation of Kliff's insight into the Corbin, KY voters is possible except by extrapolation from the 38 states for which ACA data is available.

The 2016 election results data was cleaned and adjusted to reflect the state, county, and FIPS code of each respective data point; the number of votes cast for the Democratic ticket, the Republican ticket, and for all other tickets combined in both the 2012 and 2016 elections; the winner of each of these elections ("Obama" or "Romney"; "Clinton" or "Trump"); the percentage of the vote won by Donald Trump; whether or not the county in question flipped — that is, voted for Barack Obama in 2012 and Donald Trump in 2016, or voted for Mitt Romney in 2012 and Hillary Clinton in 2016.

The HHS Plan Selections data was left untouched except to include the complete 5-digit FIPS code for each county.

The ACS demographic data was adjusted to include the percentage of each county's population with a high school education or less; the net migration into the county expressed as a positive (net immigration) or negative (net emigration) percent of the population; the total population of the county; the median age of the county; and the percentage of the county population that was caucasian.

The BLS unemployment data was adjusted to be expressed as a percentage of the total population.

Spatial Dependence

This data being geographical in nature, spatial dependence must be considered as a possible spoiler of the test results. In order to account for spatial dependence, I measured the spatial relationships of individual counties based on contiguity — counties that share borders were considered to be neighbors, "shared borders" defined as borders that share at least a single point ("Queen" configuration).

While some counties in the U.S. are virtual squares, in no state are counties arrayed in a perfect grid, and most counties are irregular polygons. Therefore, most counties that share a border share a *significant* border — in other words, there are few if any pairs of counties that are contiguous but share such a small border that their neighboring one another does not reflect any shared characteristics (an argument can be made that larger counties do in fact share borders without sharing characteristics; however, by the same token, since county lines are drawn politically or arbitrarily, a point on the border of a large county may just as well not share any characteristics with a point in its center or on an opposite border).

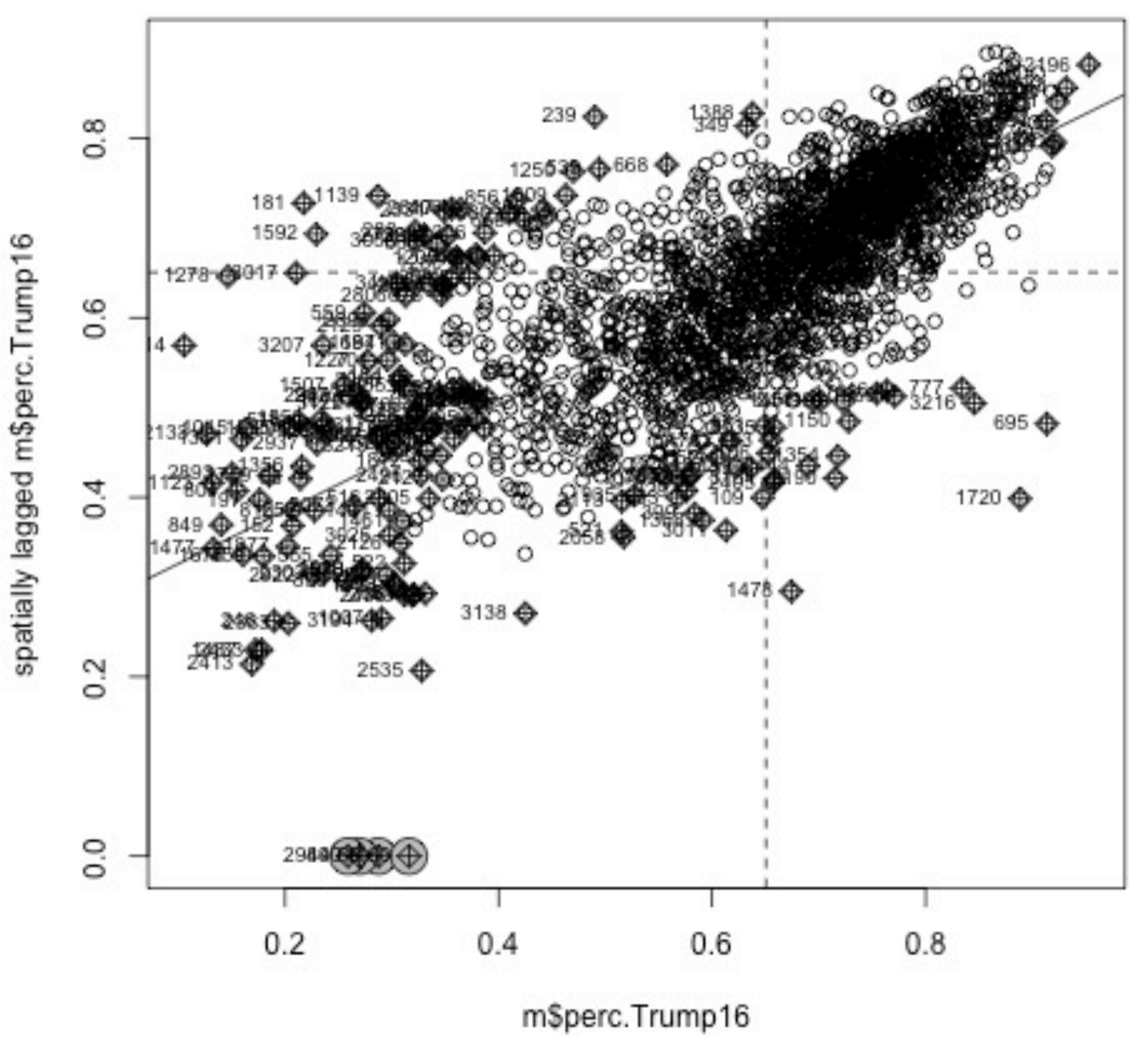
In constructing spatial weight matrices, I opted for row standardized weighting. Due to the irregular shape and lack of consistency in size across most counties, it is assumed there may be some counties that are overrepresented by binary weighting.

Moran's I testing was utilized to test for the influence of spatial dependence upon the value of Trump's percentage of the vote per county. As we will see, it was found that spatial dependence did impact the result, and spatial lag regression as well as spatial error regression were both used, with a Breusch-Pagan test was to determine whether or not the lag and error models still retained any spatial dependence within the final model. All regressions were performed with $\alpha = .05$.

Results

The Moran’s I test revealed results as follows:

Moran’s I Statistic Standard Deviation	45.88
Moran's I p-value	2.2E-16
Moran I Statistic	0.5721566101
Moran’s I Expectation	-0.0004156276
Moran’s I Variance	0.0001557470



The result shows that spatial dependence was significant. Before moving forward, a linear model was constructed to test the significance of the individual variables, yielding the following results:

Linear Regression Model

Adjusted R-Squared	0.465
R-squared p-value	2.2E-16
Plan Selections p-value	0.072336
Population p-value	2E-16
College education p-value	2E-16
Median age p-value	0.000551
Caucasian share of population p-value	2E-16

These results show that, among the chosen variables, the only one which was not significant was the volume of plan selections in the Affordable Care Act per county. In order to account for the impact of spatial dependence, spatial lag regression and spatial error regression resulted in the following:

Spatial Lag Regression

Results	With Plan Selections in model	Plan Selections removed
AIC	-5,059.7	-5,059.7
AIC for Linear Model	-3,819.2	-3,818
p-value	2.22E-16	2.22E-16
Plan Selections p-value	0.146	N/A
Population p-value	2E-16	2.2E-16
College education p-value	2E-16	2.2E-16
Median age p-value	2E-16	2.2E-16
Caucasian share of population p-value	2E-16	2.2E-16

Spatial Error Regression

Result	With Plan Selections in model	Plan Selections removed
AIC	-6,197.1	-6,197.4
AIC for Linear Model	-3,819.2	-3,818
p-value	0.1897	2.22E-16
Plan Selections p-value	2.2E-16	N/A
Population p-value	2.2E-16	2.2E-16
College education p-value	2.2E-16	2.2E-16
Median age p-value	2.2E-16	2.2E-16
Caucasian share of population p-value	2.2E-16	2.2E-16

These results show that, with both models, spatial dependence was a significant influencing factor. The unequivocally lower AIC values show that removing spatial dependence significantly detracted from both models.

A final Breusch-Pagan test shows that, even then, the resulting model still retains significant influence from spatial dependence.

Breusch-Pagan Test

Spatial Lag Model BP Test	**Plan Selections in model**	**Plan Selections removed**
BP value	157.07	156.34
p-value	2.2E-16	2.2E-16
Spatial Error Model BP Test	**Plan Selections in model**	**Plan Selections removed**
BP value	99.541	98.95
p-value	2.2E-16	2.2E-16

These results demonstrate that both models still retain evidence of significant spatial dependence. Even removing the Plan Selections variable from the model yields little change to the ultimate results.

Conclusion

The results show that spatial dependence cannot be fully separated from the results of the model. However, the Plan Selections variable was the only variable to fail the confidence test both before and after accounting spatial dependence: therefore, based on the data utilized, there is no statistical evidence that enrollment in the Affordable Care Act is an effective predictor of Donald Trump's county-level electoral successes in the 2016 election.

Citations

R Core Team (2016). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
URL https://www.R-project.org/.

Kliff, Sarah. "Why Obamacare enrollees voted from Trump." Vox.com
http://www.vox.com/science-and-health/2016/12/13/13848794/kentucky-obamacare-trump

Presidential election results for 2012 and 2016 from The Guardian and townhall.com. (Accessed February 20, 2017); Retrieved from:
https://github.com/tonmcg/County_Level_Election_Results_12-16

Plan Selections by ZIP Code and County in the health insurance marketplace: March 2016 (Accessed February 21, 2017) Retrieved from:
https://aspe.hhs.gov/basic-report/plan-selections-zip-code-and-county-health-insurance-marketplace-march-2016

Labor force data by county, not seasonally adjusted, latest 14 months (Accessed March 3, 2017) Retrieved from:
https://www.bls.gov/lau/#tables

U.S. Census Bureau, 2015 American Community Survey 5-Year estimates, B16010 (Accessed February 25, 2017) Retrieved from:
http://factfinder.census.gov

U.S. Census Bureau, 2015 American Community Survey 5-Year estimates, DP05 (Accessed February 25, 2017) Retrieved from:
http://factfinder.census.gov

U.S. Census Bureau, 2015 Population Estimates, PEPTCOMP (Accessed February 25, 2017) Retrieved from:
http://factfinder.census.gov