Empirical Analysis of Socioeconomic Factors in the 2016 Presidential Election

Aaron Pollack

Abstract
This report presents a county-level analysis of the relationship between various socioeconomic indicators and voting behavior in the 2016 United States presidential election. We show, empirically, that a statistically significant relationship exists between these indicators – education, unemployment rate, unemployment change, and poverty rate – and voter preference at the county level. We show that no such relationship is observed for median income. Finally, we demonstrate that these relationships, observed across all US counties, are robust to county-level median income.

Keywords
Econometrics — Election Analysis — Regression Analysis

Contents

Introduction

In the wake of Donald J. Trump's electoral victory in the 2016 US presidential election, various sources sought to explain what pundits and commentators has previously thought to be a vanishingly unlikely outcome – a Trump presidency. Explanations for Trump's victory ranged from the fact of changing electoral demographics, to cultural realignments in America's electoral landscape, to a deep and growing economic divide along party lines that allowed such an unconventional candidate to clinch victory.

Although numerous explanations were offered, discussions of economic and social factors were particularly prominent in the nation's post-election analysis. In this paper, we will examine to what degree differences in economic and social indicators – unemployment, median income, educational attainment, and poverty – correlate with electoral outcomes at the county level in the 2016 presidential election. This report is not aimed at providing rigorous statistical proof of a causal relationship between certain indicators and electoral outcomes, nor, necessarily, at empirically verifying the veracity of any particular electoral theory, but rather at describing how socioeconomics interacted with the political will expressed by the electorate in 2016.

1. Motivation

The 2016 election pitted Republican Donald J. Trump against Democrat Hillary Clinton. Trump triumphed in a stunning electoral upset wherein he clinched the all-important electoral college vote though he very notably lost the popular vote by by over 3 million votes, over two percent of the electorate. Analysis of the results began immediately after the election's outcome was known. The unexpected nature of Trump's electoral triumph, and the grim portents that the result represented in the minds of many, heightened the urgency with which pundits, journalists, analysts, and political onlookers sought an answer to the question of what could explain Trump's victory.

Over the coming months, a narrative emerged (in this study, we refer to it as "the economic anxiety narrative") which argued that Trump's victory was decisively influenced by the economic anxiety felt by (white) working-class voters, in particular those without a college education who were most effected by national and global trends in technology and glottalization. Discussion also surrounded the degree to which cultural anxiety and racial resentment on the part of white voters might have contributed to Trump's electoral appeal.

This study will not attempt to address the cultural and racial dimensions of the 2016 election, however our analysis will attempt to study, empirically, the degree to which education, income, unemployment, and poverty can be related to electoral outcome. The common media narrative surrounding the 2016 election leads us to expect a *positive* correlation between Republican vote percentage and both unemployment and poverty. The intuition being that both poverty and unemployment signify economic instability that could make Trump's populist message appealing to a voting populous disillusioned with the current political and economic regime.

In the profile of the Trump voter, relatively low levels of education are frequently noted. Therefore, we expect to observe a *negative* relationship between education and Republican vote percentage. Intuitively, Voters with relatively low levels of educational attainment may feel a heightened sense of economic anxiety as the economy increasingly moves towards high-skilled information-intensive jobs while occupations that once offered a decent standard of living to those without college degrees – most notably those in manufacturing – are automated or outsourced away.

Finally, we expect to see a *negative* relationship between median income and Republican vote percentage. The Republican party has, historically, been the party of the rich, while the Democratic party has historically been the party of the working and middle classes. The 2016 election saw some realignment of these decades-old electoral coalitions, nonetheless both Clinton and Trump counted rich and poor alike in their constituencies. Additionally, the economic anxiety narrative, with its focus on the political effects of economic dislocation, lead us to expect a slight negative relationship between median income and Republican vote percentage as communities with more economic vulnerability – likely those with more low-skilled, low-paying jobs – may be most sympathetic to Trump's brand of Republican populism.

2. Data

Our analysis makes use of four data sets. The first dataset comes from the Economic Research Service at the US Department of Agriculture (USDA-ERS) and compiles a panel of county-level unemployment and income data. [1] In this study, we used data from both 2016 and 2012.

The second dataset comes from the US Census bureau and tabulates the poverty rate on a county-level basis. [2] We, again, make use of data from both 2016 and 2012.

The third dataset, also prepared by the USDA-ERS tabulates educational attainment for each country, specifically, the percentage of adults attaining a bachelor's degree. [3] The percentage reported for each county is a 5-year average of degree attainment as reported to the US Census bureau's American Community Survey. Only education data for 2016 – estimated from the years 2016-2012 – was used for this study.

Finally, this study makes use of a dataset compiling county-level results of the 2012 and 2016 presidential ellections. The data itself was scraped from from townhall.com. The dataset was prepared by Tony McGovern (github.com/tonmcg) [4]

Because out approach makes user of county-level data, not individual population surveys, the explanatory power of our analysis may be limited. However, we feel that the large number of data points used – data from 3111 of the United States' 3124 counties could be retrieved for all indicators – offers a high degree of statistical validity to our approach. Even so, our methodology and results should not be confused with those of an individual-level analysis.

3. Methods

The empirical methodology for this study centers around an ordinary least square (OLS) model which measures weather and how county-level electoral outcomes are related socioeconomic factors. In this study, we measure relationships relative to unemployment rate, median income, poverty rate, educational attainment, and unemployment trend. Our model is specified as follows.

$$\begin{aligned} votePercentage = \beta_0 + \beta_1 * unemployment + \beta_2 \\ * unemploymentChange + \beta_3 * income \\ * \beta_4 * education + \beta_5 * poverty \end{aligned} \tag{1}$$

In the above model **votePercentage** represents the percentage of vote's cast for Donald Trump in the 2016 presidential election as a percentage of votes cast in a given county. The variable **unemployment** measures the percentage of the population which was unemployed in 2016. Note that unemployment is defined as the percentage of working-age adults without of job who are
textitcurrently seeking one. Importantly, this means that disgruntled workers, part-time workers, and other persons marginally attached to the labor force are not included in the measure. The variable **unemploymentChange** measures the percentage-point change in the unemployment rate, for the given county, between the years 2012 and 2016. The variable **income** measures the median income in 2016 US Dollars for the given county. The variable **poverty** measures the poverty rate, as a percentage, in the given county for 2016. Finally, **education** measures the percentage of people age 25 or older in a given county who have attained a bachelors degree or higher.

Taken together, the above models allows us to empirically the relationship between the electoral outcome in the 2016 presidential election and a collection of socioeconomic measures which paint a rough snapshot of the economic situation in a given county. The model will allow us to explore the degree to which the commonly-offered explanations – centering around wrenching economic dislocation and deeply-felt economic anxiety – can actually explain voter behavior in 2016.

4. Results

Our results (reported in Table 2) show highly statistically significant relationships between many of our socioeconomic indicators and the percentage of votes cast for Donald Trump. We observe a negative relationship between unemployment and Trump vote percentage. This suggests that as the unemployment rate *increases* across counties, the percentage

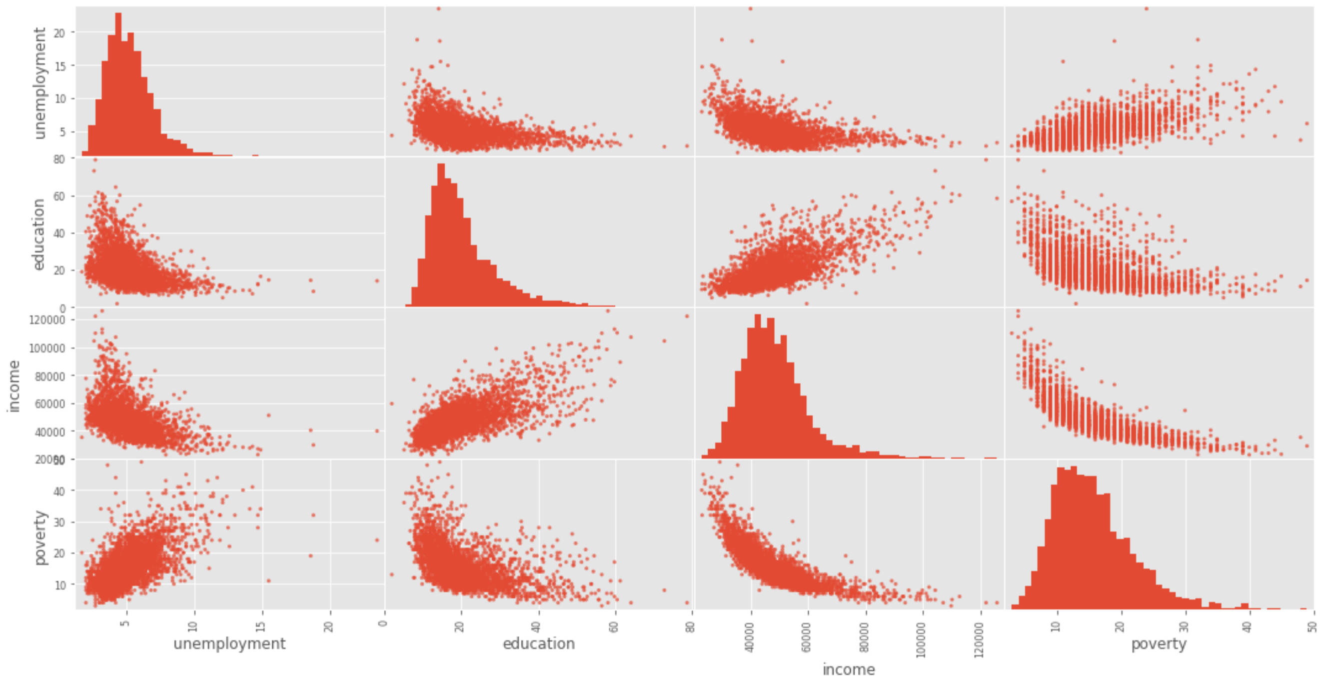

Figure 1. A scatter matrix of the principal socioeconomic data sets used in this study. See the **Methods** section for a more thorough explanation of the meaning of each variable

of votes cast for Trump *decreases*. We observe a positive relationship between an increase in the unemployment rate between 2012 and 2016 and the number of votes cast for Trump. This indicates that counties which saw *increases* in unemployment between 2012 and 2016 more readily voted for Trump. The relationship between voting behavior and county median income was much more tenuous. Although a statistically significant relationship is suggested by our model, an economically significant relationship is not observed. Education saw the most pronounced relationship. Our model suggests a *negative* relationship between the median level of educational attainment in a given county and support for Donald Trump in that county. Finally, we observe a negative relationship between votes cast for Trump and the level of poverty in a given county. That is to saw that, across the counties, as poverty increased the percentage of votes cast for Trump decreased. In all cases, except income, the relationship between the relevant socioeconomic indicator and electoral support for Trump was seen to be highly statistically significant – significant at the $p < .01$ level.

4.1 Income Robustness

Table 1. Income Groups

Group Name	Range	Count
Low	$x < \$40000$	734
Low-mid	$\$40000 < x < \$50000$	1180
High-Mid	$\$50000 < x < \$60000$	780
High	$x > \$60000$	417

To enhance the empirical validity of our conclusion and to further explore the relationship between socioeconomic factors and electoral outcome, we examined the degree to which the above effects are robust to income. That is to say, we ran our model with various subsets of the collection of counties to examine whether the statistical relationships observed across the entire country would be present in a more isolated sample. We divided the counties into groups based on county median-income. Table 1 describes the income groups used. Note that the boundaries chosen for the groups are arbitrary and do not reflect any particular economic significance. Note further that the group names (”Low Income”, ”Low-Mid Income”, etc) are purely for convenience and do not represent an attempt to characterize relative income levels outside the context of our analysis.

We observe that the relationships seen in the regression of our entire sample are present, with a high degree of statistical significance, in each of our income-based sub-sampled. This suggests that the statistical and economic relationships described for the entire sample are robust to income. The sole exception is the effect of unemployment among high income counties (counties with median income greater than \$60,000). Although a potentially economically significant relationship is observed (a decrease of .79 percentage points in Trump's vote percentage for each percentage point increase in unemployment) a statistically significant relationship is not. In all other cases, the relationships observed in the income-based subsets of our sample were seen to be highly statistically significant.

Table 2. General Regression Results

	All	Low Income	Low-Mid Income	High-Mid Income	High Income
unemployment	-1.6658***	-1.8184***	-1.5858***	-2.1093***	-0.7965
	(0.1451)	(0.2936)	(0.2047)	(0.3163)	(0.5978)
unemployment_change	2.1073***	2.1208***	1.9339***	2.4124***	3.3246***
	(0.1178)	(0.2685)	(0.1683)	(0.2400)	(0.3497)
income	-0.0002***	0.0002	-0.0007***	0.0000	-0.0000
	(0.0000)	(0.0002)	(0.0001)	(0.0001)	(0.0001)
education	-1.1339***	-1.3454***	-1.0070***	-1.2246***	-1.0404***
	(0.0325)	(0.1130)	(0.0476)	(0.0547)	(0.0723)
poverty	-0.8460***	-1.1327***	-0.9853***	-0.4665***	-1.7859***
	(0.0589)	(0.1361)	(0.1057)	(0.1640)	(0.3133)
Observations	3111	734	1180	780	417
R-Squared	0.463	0.366	0.405	0.528	0.610

Standard errors in parentheses
* $p < 0.1$, ** $p < 0.05$, *** $p < .01$

5. Discussion

The statistical relationships described above offer potentially valuable insights into the socio-political forces at work in the American electorate.

5.1 General Regression

Applying our regression model to the entire data set yields some surprising results. Much of the political commentary surrounding Trump's victory in 2016 painted the vote of the white, working-class as decisive. While this may, of course, still be the case, we see many trends in the data that suggest high levels of economic vulnerability or economic dislocation are more readily associated with Democratic support than with Republican support.

Specifically, we might expect the level of unemployment to be positively related to Republican support, however, this is not observed in our dataset. In fact, we see that unemployment is negatively related to support for Donald Trump; as unemployment increase, Republican vote percentage decreases. Similarly, we observe that the level of poverty is negatively related to Republican vote percentage; as the average number of people in poverty increases, the percentage of Trump voters decreases.

Together these two statistical relationships would seem to paint a different story than the one of economic vulnerability and dislocation being strongly related to support for Trump's disruptive brand of nationalistic populism. One might expect the the most economically troubled counties – represented by those with high unemployment and poverty rates – to be the most supporting of Trump. The fact that this is not observed in our data can be explained one of several ways. For one, in spite of these statistical relationships it must be said that economic anxiety, economic dislocation, and feelings of economic vulnerability can still be decisive factors in tipping the electoral balance in favor of Trump. Although Trump's most vocal base is often characterized by economic and cultural anxiety, he would been a far throw from the white house had this base been his only electoral support. These statistical relationships indicate that Trump's electoral coalition is broader than the ”angry, white, working-class voter” pundits often point to, or that that the political motivations of Trump's coalition cannot be explained by economic anxiety alone. Additionally, recall that the Democratic party has, historically, counted economically vulnerable communities – particularly communities of color – as among it's most loyal electoral constituencies. The observed statistical relationships with respect to unemployment and income could simply reflect the strong support enjoyed by the democratic party of many economically and culturally marginalized communities.

In contrast to the statistical relationships described above, the other relationships paint a picture more in line with the ”economic anxiety” narrative surrounding Trump's victory. Specifically, the level of bachelors degree attainment in a given county is observed to have a negative relationship with the level of electoral support for Trump while a rise in unemployment in a given county between 2012 and 2016 is observed to have a positive relationship with electoral support for Trump.

The statistical relationship observed for education could be interpreted as supporting the ”economic anxiety” narrative surrounding Trump's victory. In theory counties with high attainment of post-secondary education would be more insulated from the economic transofrmations that have destabilized industries and shunted once stable careers overseas or out of the workforce entirely. Jobs like manufacturing are the most vulnerable to outsourcing or to automation. By contrast, many of the jobs and career paths offered to holders of post-secondary degrees are, for the moment, more economically secure. Under this line of reasoning, Trump's brand of strident populism would be most resonant with communities feeling the most economically vulnerable – those with low post-secondary education – while more highly educated communities are less drawn to Trump's message.

It could also be that Trump's political brand and political message are less appealing to highly educated communities. Educational attainment is often pointed to as a correlate of liberalism. One could imagine that communities with high post-secondary attainment could find the turbulent, unconventional nature of Trump's candidacy to be less appealing.

The statistical relationship observed for unemployment change paints a similar story as that of education. An increase in unemployment in a given country between 2012 and 2016 is observed to have a positive relationship with support for Trump in 2016. One imagines that the communities most hard hit by economic dislocation would be those with the greatest rises in unemployment. Such economic turmoil could fuel a demand for the political change Trump promised his supporters in 2016. Additionally, since the national unemployment rate trended downwards between 2012 and 2016, seeing unemployment rise in one's own community would likely create feelings of resentment which would make one especially sympathetic to Trump's message.

We observe no economically significant relationship between median level of income and support for Trump in a given county. The fact that a community's level of income – and by proxy a community's social class – is a poor predictor of a community's support for Donald Trump would seem, at first, to challenge the narrative of Trump's working-class support. However, as this study does not include demographic characteristics such as race, we feel that this conclusion is too broad. A more nuanced interpretation of this statistical relationship is that both Trump's and Clinton's electoral coalitions induced people from across the spectrum of income. Specifically, one imagines that low income communities of color would be much more likely to support Clinton than would low income white communities. In short, this statistic could indicate that low income white voters may have voted for Trump while low income voters of color may have voted for Clinton *or* that income is uncorrelated with party alignment. As a final note, the weak relationship may reflect a partial shift in the traditional base of upper-class support historically enjoyed by Republicans as some of these voters, disheartened by Trump's turbulent candidacy, shifted their support to Clinton's more conventional campaign.

Finally, we wish to point out that, although all relationships observed in our whole-sample regression are highly statistically significant ($p > 0.01$), the regression itself is a poor predictor of electoral outcome overall. The Relatively low R-Squared statistics ($R^2 = 0.463$) suggests that our linear model poorly fits the data. Although low R-Squared values are not inherently bad, we must acknowledge the fact that the indicators used in this study can, themselves, explain only part of the story of the 2016 electoral outcome.

5.2 Income Subgroup Regressions

The statistical story told by the income subgroup regressions largely mirrors that of the full-sample regression. The robustness of our findings to income level reinforces the statistical narratives articulated above.

The only deviation between the statistical relationships observed in the entire sample and those observed in the income subgroups is seen in the relationship between trump support and unemployment for counties with median income greater than $\$60,000$. Although the relationship is negative – mirroring the result seen across the entire sample – the relationship is not statistically significant. This lack of statistical significance could reflect the notion that the voting behavior of more affluent communities is less readily predicted by economic indicators. Alternatively, it could reflect, simply, the low sample size ($n = 417$) which would cause a potentially economically significant relationship ($\beta_1 = -0.7956$) to fail a test of statistical significance.

We see considerable variability in the R-squared scores across the various income subgroups. Specifically, we observe that the R-squared score is lowest for the "Low Income" subgroup (counties with median income below $\$40,000$) at $R^2 = .366$ and at its highest for "high income" counties (median income above $\$60,000$) at $R^2 = .610$. This indicates that the socioeconomic indicators used in this study become better predictors of electoral outcome for higher income brackets. Since this study is operating on county level data not individual data, one interpretation of this finding is that counties with higher income become more partisan, though not necessarily in one direction or another. In other words, higher income counties could be more politically predictable even as the collection of counties themselves doesn't necessarily align with one party or another (as suggested by the weak relationship between electoral outcome and median income).

6. Caveats

This study is motivated by an attempt to examine, empirically, the socioeconomic narratives surrounding Trump's unexpected victory in the 2016 presidential election. However, our use of county-level date instead of individual-level data means that we are unable to examine voting patterns across populations of individuals. Examination of average tenancy within counties may not paint an accurate picture of voter behavior or the factors that predict it. Readers are therefore cautioned not to confuse the statistical relationships outlined in this study – or the social, political, and cultural narratives offered to explain them – as describing the behavior of individuals. Further study is needed if we wish to describe the voter behavior of individual voters with any degree of statistical authority.

Our study intentionally did not examine the degree to which demographic characteristics such as race and ethnicity are predictive of electoral outcome. This isolation of scope reflects a desire to analyze, specifically, the economic narratives surrounding the 2016 election. However, given the unconventional nature of Trump's politics as well as the current, fractious state of American politics, factors like race and ethnicity – which always play a role in predicting voter behavior – likely played an even greater role in 2016. For this reason, we must acknowledge that the statistical relationships described in this study and the conclusions which we draw from them can paint only a partial empirical picture of the narratives surrounding Trump's victory and American politics generally.

Finally, it must, of course, be noted that the "economic anxiety" narrative discussed in this study is only one of many proposed explanations for the outcome of the 2016 election. Although our analysis centers around this narrative, we do not wish to imply that it is the only valid explanation of events in 2016 nor that the 2016 election can be completely explained by it.

7. Conclusion

Trump's surprise electoral victory in 2016 capped a tortuous, bitter, and divided political campaign season. Some felt elation and Trump's insurgent triumph, others felt deep dismay. Everyone was searching for an answer as to what could have led to an electoral outcome which seemed so unlikely even day before.

In the wake of the 2016 election, numerous explanations were offered in an attempt to articulate a unified, cohesive narrative regarding the outcome of the 2016 election and it's broader significance in the country and world. Many pointed to the economic anxiety and financial marginalization felt by communities of predominantly white, working-class voters across the country. Many noted the inability of Democrats, particularly Hillary Clinton, to reach these voters with a message that could draw their support. Others pointed to cultural anxiety felt by many in response to shifting national demographics and changing times. Others noted the decisive role that foreign intervention could have played in the election. Still others pointed to the role that the scandal-ridden nature of the election itself could have had in determining who showed up to the polls and which candidate they chose.

A complete and nuanced truth, of course, resists simplicity. Such a complicated event as the 2016 presidential election no doubt had many more causes than can be listed here. However, the narrative describing a collection of economically vulnerable communities being draw to Trump's populist rhetoric and promises of change seems particularly important.

This study examines how socioeconomics relates to voting patterns and in so doing, explores the degree to which this economic and cultural narrative can be observed empirically. We examined the degree to which socioeconomic indicators such as unemployment, income, education, and poverty are predictive of electoral support for Donald Trump in 2016. We found both economically and statistically significant relationships for many of these indicators. However, some of the observed relationships seem to challenge the narrative surrounding the role economic anxiety plays in support for Trump. We find that the observed relationship between socioeconomics and voting patterns are robust to income even as income, itself, is a poor predictor of a community's voting pattern.

We conclude that, although the "economic anxiety" narrative is an incomplete explanation of Trump's electoral victory, it is likely a valid one in part. However more generally, we conclude that socioeconomic factors, though relevant in predicting voting patterns, are only a small part of the social, political, cultural, and economic story of American politics and American society.

References

[1] Bureau of Labor Statistics. Unemployment and median household income for the u.s., states, and counties, 2007-16. http://www.bls.gov/lau/, 2017.

[2] United States Census Bureau. Poverty estimates for the u.s., states, and counties, 2016. http://www.census.gov/did/www/saipe/, 2017.

[3] US Department of Agriculture. Educational attainment for adults age 25 and older for the u.s., states, and counties. https://www.ers.usda.gov/data-products/county-level-data-sets/, 2017.

[4] McGovern Tony. United states general election presidential results by county from 2008 to 2016. https://github.com/tonmcg, 2017.