What Happened on November 8, 2016?

A County Level Study of U.S. Presidential Elections

Raju Adhikari

Master of Public Policy

Class of 2017

Completed under the advisement of Prof. Mark Kayser, PhD

April 2017

Acknowledgement

I would first like to express my gratitude to my wife Anna Marie Walter, for being a partner, friend, mentor, and critic during the process of this thesis, brainstorming with me even during odd hours of the night, and sacrificing outdoor activities on rare sunny days in Berlin.

I would like to express my sincere gratitude to my thesis advisor, Prof. Dr. Mark Kayser, who provided great support, detailed guidance, and positive criticisms that inspired me throughout the process to take on a challenging quantitative endeavor like this. I would also like to express my special appreciation to Dr. Simon Munzert (lecturer of Election Forecasting at the Hertie School), and friends Devvart Poddar, Md. Mujahedul Islam, and Alvaro Lopez for helping me with statistical methods, brainstorming, and R programming.

I would also like to thank all my colleagues at the Hertie School, for sharing two rich years full of friendship, laughter, and knowledge with me - in classrooms, in ping-pong rooms, and in out-of-country trips. To my family and friends in Nepal, in the United States, in Honduras, and everywhere around the world, I am deeply grateful for their support.

Last but not least, I would like to thank my dog Mori for her unconditional love and stress-relieving presence, without which neither the time in Berlin nor the thesis writing process would have been possible.

Executive Summary

The purpose of this thesis is to compare the U.S. Presidential election of 2016 to the ones from the past and assess if they fit the same model of voting behavior. To carry out this objective, the thesis is divided into two unique research questions and three steps of quantitative methods. The first research question aims to see the effect of the change in county level unemployment rate on Republican vote share. A fixed effects estimation method is used to analyze the elections between 1992 and 2012 alongside economic, political, and demographic explanatory variables to answer the first research question. Literature on economic voting guides the selection of variables, theories, methods, and the justification for using county level analysis over a state or a national level. The second step is a forecasting exercise that predicts the Republican vote share for 2016 election based on the model created in the first step. The difference in the predicted and actual vote share of the Republican party presents itself as a perfect transition to explore the second research question. The second research question aims to test several rival theories regarding an unexpected and under-predicted Republican vote share in the 2016 election which led Donald Trump to the White House. An Ordinary Least Squares estimation method is used to analyze the election of 2016 using explanatory variables such as employment in manufacturing sector, labor force participation rate among men, change in average wage, inequality, and educational attainment level, for all counties.

Three noteworthy results can be summarized at the end of the thesis. First, the analysis of past elections shows that Republican vote share is dependent on the party identity of specific counties more than anything else. In addition, lower unemployment does tend to favor Republicans more than Democrats, although very weakly; and counties with larger percentages of white population tend to lean Republican. Second, the forecasting of 2016 election shows that the 2016 Republican vote share was unexpected based on a traditional model and there was a stark similarity in the counties where the votes were under-predicted and where the vote share was unexpectedly higher. This proves that Donald Trump was not a generic Republican and hence a revised model would be necessary to explain his effect in 2016. Third, the analysis of various rival theories used to measure Trump effect in 2016 shows that only education was a significant and highly strong explanatory factor.

Contents

Introduction

> The economy, as an issue, is one of consensus. All voters want a good economy, no voters want a bad economy. No distribution of opinion occurs. Everyone values prosperity. When they see prosperity, they vote for the ruling party, otherwise not.
>
> Michael Steven Lewis-Beck & Nadeau (2011)

Around the weeks and months surrounding the United States Presidential election of 2016, social media platforms were on fire, where people from all background were berating each other for supporting or not supporting certain candidates. Academics predicted the outcome based on their historical knowledge of elections and voting behavior, pollsters used randomized surveys simulated thousands of times to predict the correct winner with highest confidence levels, and media outlets showered people with news -real and fake- to turn this into the most exciting reality show of all times. Having lived in the South of the United States, in a historically red state for many years, and being interested in finding the connection between political and economic factors, I was immediately attracted by the idea of exploring the cause behind the 2016 election outcome. In this thesis, I decided to analyze the 2016 elections from the perspective of the Republican side and tried to draw comparison between the Republican vote share in past elections and the recently concluded one.

To better understand voting behavior in American political scenario, I narrow my focus to county level social, economic, and demographic factors and connect them to Republican support. In the first part of this thesis, I analyze the effect of county level factors on Republican two-party vote share. Following the framework of economic voting, while controlling for political factors such as previous election year vote share and incumbency, and demographic factors like white population and rurality, I use Fixed Effects estimation to analyze the effects of unemployment rate change on Republican vote share for elections between 1992 and 2012. The purpose of this analysis is not just to see the effects of these variables but also to create a sound statistical model that can be used to understand the 2016 election. With that purpose, in the second part of this thesis, I use the equation obtained from the first part to predict Republican vote share for 2016 election. While, the process of forecasting an election that already happened may seem uncharacteristic of election forecasting practice, there are some benefits to it. Forecasting result is used to measure the difference in the actual vote share and the predicted vote share and to assess if the Republican candidate in 2016 fits the definition of a generic Republican as outlined by the previous model. In the third part of the thesis, a new model is formulated that tests several rival theories surrounding the outcome of 2016 election and tests their statistical significance in explaining the difference (i.e. the residual) obtained from the forecasting exercise. At the end, the goal is to go back to the original question of what affects a Republican vote and how it is similar or

different in 2016 compared to the years before. This will help understand the Trump-specific effects that influenced the outcome in 2016.

Literature Review

In the literature review section, I discuss various authors, theories, and analysis used in the past on topics pertaining to the research conducted in this thesis. Discussing the similarities and differences that occur among authors and the evolution of various theories and techniques help design the research hypothesis and the appropriate methods. The first section of the literature review discusses economic voting theory, which is the main driver for the analysis in this thesis. Important dependent and independent variables are chosen based on theories discussed under this section. The second section compares macro-level analysis of economic voting with micro-level analysis. The geographical unit of analysis for the research is chosen based on the discussion under this section. The third section discusses the importance of various political and demographic factors that complement a good economic voting model and hence help choose the appropriate control variables. The fourth section discusses various theories that could be attributed as factors contributing to the dominance of the Republican party in the 2016 election. The reason these theories are not included in the first section is to separate them from the traditional model and also due to the unavailability of historical data for the variables explaining these theories. The theories mentioned here come from both academic as well as non-academic sources, since academic literature is limited in its exploratory power for an election that occurred very recently and with many peculiarities that made its forecasting extremely difficult. The fifth section compares the techniques used by authors in the past, the time-frame of their analysis, and other peculiarities.

Economic Voting Theory

The effect of economic factors on election outcomes has been widely studied by scholars all around the world and makes the basis for economic voting theory. When it comes to the United States, this idea of economic voting has been explored time and again through various types of statistical analysis carried out over the period of many election years. Ebeid & Rodden (2006) mention several authors, such as Gerald H. Kramer, Ray Fair, Edwards Tufte, Steve J. Rosenstone, Douglas A. Hibbs, Robert S. Erikson and Thomas M. Holbrook, who have studied and shown the connection between national macro-economic conditions and election outcomes in the United States. The macro-economic variables that are widely used to determine election outcomes are economic growth, disposable income, (un)employment, job growth, economic volatility, inflation, etc. (Jensen, Quinn, & Weymouth, 2016).

Even in election forecasting, which deals with predicting the future than analyzing the past, fundamental statistical models are widely used to determine election outcomes. These fundamental models also follow economic voting theory in the sense that they use economic variables, alongside various social and political factors, to predict the outcome of an election. Like Jensen, Michael S. Lewis-Beck (2005) lists the many economic variables used by authors in fundamental models of election forecasting as: GDP growth, GNP growth, perception of personal finances, prospective personal finances, leading economic indicators, income growth, and job growth. Moreover, the realm of economic voting has been expanding rapidly over the years with new variables, definitions, and processes attached to it. Some economic voting theories view the economy as a valence issue- "issues on which no one disagrees about the desired outcomes" (Vavreck, 2009, p. 22). Michael Steven Lewis-Beck & Nadeau (2011), on the other hand, examine positional and patrimonial economic voting, where positional voting means that voters use their vote with a policy aim regardless of the incumbency and patrimonial voting means that voters' own economic status plays a role in their voting behavior. According to Kiewiet, positional voting hypothesis hasn't been empirically tested as much as the incumbency-oriented hypothesis- where an incumbent is either rewarded or punished by the voters during elections- and therefore provides strong motivation to explore Presidential elections in the U.S. from a mix of both positional and patrimonial perspective (as cited in Michael Steven Lewis-Beck & Nadeau, 2011, p. 6).

The proponents of economic voting agree and disagree with each other on a few topics. Some of the main questions that still float around were clearly laid out by Abrams (1980) in what he labeled as the three main research questions of 'political business cycle' that needed clarification:

> (1) to determine which politicians, if any, are held responsible by the electorate for changes in general economic conditions; (2) to identify which general economic conditions influence the electorate's voting; and (3) to establish the time period that the electorate uses to assess economic policies. (p. 1)

The first two questions asked by Abrams set the stage to determine which indicator captures the election result (dependent variable) and which economic variables explain it (independent variables). To answer the first question regarding the basis for measuring the election result, one may choose to look at the total vote share of the incumbent or the total vote share of a specific party or the two-party vote share of a specific party. Tuftes's "election-as-a-referendum-on-the-government" idea has been used to incorporate incumbency in forecasting models to predict U.S. Presidential elections (as cited in Michael S. Lewis-Beck, 2005, p. 150). Moreover, Norpoth & Gschwend (2010)'s "cost of ruling" designation to incumbency showcases its effects on election outcomes. They both use incumbent vote share as a dependent variable but without elaborating if the results would be different if the candidate is incumbent as well and not just the party. While most studies place candidates and parties under

the same umbrella, Eisenberg & Ketcham (2004) separate the two entities and claim that voters do not hold incumbent candidates "additionally"" accountable even though they hold incumbent parties responsible. In doing so, they refute previous claims that economic voting is more significant for incumbent candidates than for incumbent parties with new candidates (Eisenberg & Ketcham, 2004). Another way of measuring the election result is by considering a specific party than the incumbent party. The rationale behind this approach is to understand if voters "always" assign their economic vote for or against the incumbent or if they do so based on a specific party, regardless of the incumbency. This happens under issue-priority theory, where voters relate certain economic policies with certain parties and vote for the party that is concerned with solving that issue even if the country is not performing so well under that party in that issue (J. Kim, Elliott, & Wang, 2003). In this thesis, we will follow Kim's theory in analyzing if voters in the United States take certain economic indicators into account when voting for the Republican party. We, will nevertheless, keep incumbency as a dummy control variable to check if it has any conditional effect. Meaning, while voters may vote based on certain policy issues, the effect of their examination of the policy on the vote share may be dependent on whether the party is an incumbent or not. The purpose of the thesis is to see from the perspective of not just a voter but a Republican voter.

The second question by Abrams is already touched upon above by listing a plethora of explanatory economic variables mentioned by various authors as indicative of an election outcome. While all of them have their merits, depending on the nature of analysis, unemployment is the one that sticks out the most specifically in the context of this thesis. Wright (2012), while describing the theory of issue-ownership, claims that the most central problem that voters keep in mind making their election decisions is whether they, their families, their friends, and their neighbors have jobs or not. Unemployment is also a variable that is more relatable among voters who may find it difficult to assess the performance of the government based on other complex economic indicators that look important on paper. Similarly, seeing the rise of globalization all over the world and its impact on the economy, Jensen et al. (2016) claim that trade contains information that growth and employment do not explain and hence include macro-economic indicator such as U.S. trade balance as an explanatory variable for national-level voting.

The third question by Abrams which discusses about the time frame of different variables will be discussed under the fourth section of the literature review.

Macro-level analysis vs Regional Analysis - a case for studying counties

Study of voting behavior has usually consisted of analysis at the national level as scholars have tried to connect national economic indicators to election outcomes. One of the advantages of national level analysis is that it is easy to carry out since it removes many nuances of geographical differences across different sub-national levels. However, the same easiness also presents itself as a big impediment to understand the real reasons behind an election outcome. A barely chartered territory is the sub-national analysis at the county level although some studies have been done at the state level. O'Laughlin et al. (as cited in J. Kim et al., 2003) and Owens and Wade (as cited in J. Kim et al., 2003) provide some evidence of sub-national economic voting scenario in Germany and the United Kingdom, while in terms of the United States Archer and Taylor (as cited in J. Kim et al., 2003) have sought to shed some light on the role of sectionalism in American politics. Abrams (1980) presents the rationality behind the idea of testing whether state-level economic conditions are used by voters to assess presidential policies and if they influence electoral outcomes at the national level. He validates the dis-aggregation hypothesis - which suggests the dis-aggregation of the economic-conditions variables by geographical factor - and claims that state-level economic conditions impact voting outcomes (Abrams, 1980). Abrams & Butkiewicz (1995) provide further evidence through a separate study of the 1992 U.S. Presidential election underlining the significance of state-level economic conditions in the defeat of George H.W. Bush. While these studies do not make any claims at the county level, they do provide the basis to further dis-aggregate the economic-conditions variables at a smaller geographical level and study the resulting impact. Following up on Abrams' work, Blackley & Shepard III (1994)'s study of the 1992 U.S. Presidential election provides evidence that local economic conditions have significant effects on presidential voting and therefore are consistent with both the self-interest and local altruism hypothesis. Meaning, voters are concerned about their issues and the issues of people around them, i.e. neighbors, friends, and residents of same geographical areas. Moreover, what is considered as the self-interest of an individual can often encompass the welfare of others (Anonymous, 1998) since people's welfare are deeply intertwined in today's vibrant and complex economy. Besides, it is safe to assume that such self-interest is stronger at a smaller geographical level than at a national level where connection between individuals is much more complex and much less tangible.

Abrams (1980) uses the vote share of an incumbent presidential candidate in a state and regresses it against general economic conditions of the state, which can be measured by the change in a state's unemployment rate and the percentage changes in real per capita personal income. He claims that voters hold Presidents accountable for changes in state-level economic conditions therefore cutting certain programs in strongly dominated states and moving them to closely contested states may improve the chances of reelection for the incumbent (Abrams, 1980). He also claims that abolishing the electoral

college and permitting popular votes to determine presidential election outcomes can outweigh the efforts of such state-level re-distributive policy (Abrams, 1980). This state-level analysis from Abrams also provides a strong theoretical foundation to carry out a county level analysis. Sartorius (2015) claims that the large number of observations available from more than 3000 counties of the United States and their unique economic and electoral characteristics make for an intriguing analysis of voting behavior at a sub-national level. In addition to this, J. Kim et al. (2003) further argue on behalf of county level analysis by saying,

> not only is the problem of ecological fallacy much less severe with county-level than state-level data, but from a practical standpoint, the county may be the smallest spatial unit of analysis for testing the partisanship thesis that requires the availability of macroeconomic variables such as unemployment rate. (p. 744)

The electoral college phenomenon that exists in the United States and the effects of gerrymandering in allowing various counties to exert a bigger weight on the outcome of the Presidential election also suggest that studying elections from county level perspective can shed light on many questions that have been answered unsatisfactorily in the past with a general national level analysis. J. Kim et al. (2003) go a step further in their study of the effects of spatial patterns in American politics by suggesting that increasing concentrations of geographical support for the parties will mean more ideological polarization and more demographic distinction between them. Their analysis seems nothing less than prophetic in the aftermath of the 2016 Presidential election that saw an extremely polarized and divided America fall into the spatial crevices of electoral college system. Jensen et al. (2016) also use U.S. county level measures of economic indicator, such as employment, to assess the effect of trade on presidential voting. A study by Wright (2012) uses county-level employment data to claim that Democratic vote share at gubernatorial and presidential elections is directly proportional to unemployment rate regardless of the incumbency. All these studies provide a perfect premise for this thesis as it aims to understand the effects of regional economic conditions on election outcomes in the past and more specifically on the recent election of 2016. This thesis also follows the framework by Eisenberg & Ketcham (2004) who claim to present "the first county-level analysis of economic voting in presidential elections." The difference in this thesis, however, is that I aim to first analyze the effect of county-level economic conditions on presidential election results from the past (1992 to 2012) and then carry a separate analysis for the 2016 election to explain the residual that is not explained by the first analysis. This type of county-level analysis will first test my hypothesis with the assumption that the Republican candidate is a generic Republican and secondly test if Trump is a generic Republican based on how the 2016 result fits into the model from previous years.

Political and Demographic Controls

While the first aim of this thesis is to analyze the impact of economic voting on Republican vote share, controlling for other political and demographic factors is extremely important to find the accurate estimation of the effect. Elections are not one dimensional phenomenon and voting behavior is influenced by a multitude of factors. In the United States, one of the important factors that stands out is partisanship driven by party identification (Campbell, Converse, Miller, & Donald, 1966). States are known for their identity as red, blue, or swing. And the political identity of majority of voters in those states usually stays the same over time. To analyze what affects Republican vote share in a blue parish in Louisiana (parish is the equivalent of county in Louisiana) without considering the party identification or partisanship indicator of that county will be erroneous. In surveys, individual voters are asked about their party preferences, but we do not have the luxury to assign these preferences to a whole county. Therefore, in this thesis, the previous vote share of the Republican party is used as a yardstick to measure the party identification of each specific county. Previous vote share for the incumbent party is introduced as an independent variable by Abrams (1980) in his model as well. Glaeser & Ward (2006) debunk the myth that America's political geography has been more stable lately than in the past by claiming that it has always been quite stable as suggested by the high correlation between the current and lagged vote share of the major parties. In one example, Kayser & Wlezien (2011) claim that in the event of strong partisan effects (i.e. when voters are strongly tied to a particular party), the effect of the government's performance (i.e. its economic performance) has little effect on vote shares.

Many political scientists have agreed to the idea of analyzing the economic vote through the perspective of an incumbent. Key's retrospective model uses bounded rationality theory to claim that voters use their vote as either a reward or a punishment depending on how the incumbent government performed (as cited in J. Kim et al., 2003). In economic voting, incumbents are given utmost importance because people often relate economic performance with the incumbent party or candidate. With an incumbent dummy we can control for the effect of incumbency on vote share. Moreover, when most variables are at county level, the incumbency variable serves as a representative of the national effect on the county, since the incumbent Presidential candidate or party is the same for all counties.

Demographic controls are also an important aspect of economic voting model, as they are known to significantly impact the vote share of a party at presidential elections. Including them into the model will reduce the possibility of omitted variable bias and gives more precise estimations. Race, age, education, and gender are some of the main demographic variables often used by authors. Republican voters are often characterized as white and following Campbell et al. (1966)'s "group interests" dimension

of party ID, one can hypothesize that counties with majority whites vote Republican, making it a strong control variable. Counties with population that has fewer years of formal education can also be controlled for the understanding of the Republican vote share. Gender, although being a strong control variable at national level analysis, doesn't fit a county level analysis very well. In a county level analysis, the role of gender on vote share is difficult to measure since most counties will have around the same number of men and women and it is difficult to assess the effect of one gender on the vote share unless an individual survey-type analysis is done. Rurality of a county is an important control that is indicative of a plethora of characteristics about the voters. Rural voters are usually considered to be poor, uneducated, and more conservative due to the lack of dynamism and exposure that voters in urban areas experience. The effect of religion and the perception about immigration can also be different in rural and urban areas. When doing a sub-national level analysis, this measure of rurality captures a lot of voting characteristics of a county. Galbraith & Hale (2008) in their state-level analysis claim that a high proportion of people living in urban areas, and higher minority population are associated with a greater Democratic vote.

Some rival theories regarding the 2016 U.S. Presidential election outcome

The U.S. Presidential election of 2016 has stood out for many reasons- the most significant of them being the failure of scientific forecasters and political pundits to accurately predict the winner. With years of research, theories, and models at their disposal, how did they fail to get it right is a question which still hasn't been convincingly answered even months after the election. This thesis does not claim to answer it either, but does aim to incorporate some new rival theories and see if and how they might have played a role. There are a multitude of rival theories on what might have influenced the 2016 election: economy, racism, sexism, Russian influence, anti-elite message from a supposedly non-political candidate, social media's impact- be it via fake news or late-swing generating tweets from influencing people, resentment from Bernie Sanders' supporters, shaming of the Republican voters by liberal elites, third-party effects, gerrymandering, and the list goes on and on. Considering the limited scope of this thesis, which focuses on county level analysis and hence requires county level data to do so, only a few of these theories will be tested with a statistical model.

The first theory that I aim to test is what I will call the "employment theory"[1], which focuses on employment rather than unemployment. The underlying idea behind this theory is the impact of globalization on workers from sectors that have suffered the most due to changing economic dynamic.

[1]The nomenclature is given by the author for easiness to refer to the theory again throughout the thesis. This does not mean that the author claims to have proposed the theory. The ideas mentioned under this theory already exist and hence they are discussed under the Literature Review section in the first place.

Jensen et al. (2016) claim that the traditional measures such as unemployment and economic growth alone do not fully capture the effect of trade (by which they mean globalization) on workers' voting behavior. In their county level analysis, they measure employment in firms that have been affected by trade liberalization, and in doing so they introduce variable to measure employment in manufacturing and service sector (Jensen et al., 2016). In this thesis, I will introduce a similar variable that will measure the change in the proportion of manufacturing jobs since the last election. The reason the change is measured in relation to the last election is embedded in the assumption that voters view the situation they are facing during this election with the situation they were facing at approximately the same time during the last election, explaining the change in their voting behavior from the last election. Another variable that I found extremely relevant in understanding the 2016 election is the labor force participation rate. According to the Bureau of Labor Statistics (BLS), "the labor force participation rate is the percentage of population that is either employed or unemployed (that is, either working or actively seeking work)."[2] The importance of this variable lies in the fact that there are many people in their prime working age who may not be participating in the labor force and hence reflecting certain peculiarities about the economy and its effect on them that could be translated into their voting behavior. A simple measure of unemployment rate doesn't capture what the labor force participation rate does, meaning a county with fairly low unemployment rate could be suffering from a low labor force participation rate as well. This is well evident in the example presented by the BLS in 2007 where the unemployment rate between July and August didn't change much despite a decline in the labor force participation by about 600,000 people (as cited in Gustavsson & Österholm, 2010). This led me to include labor force participation rate as one of the variables to describe Republican vote share in the 2016 election. I base this on the recommendation made by Gustavsson & Österholm (2010) who suggest using other labor market statistics before drawing any inferences about the labor market conditions. Back in 2010 they claimed that "there are important, permanent, discouraged-worker effects among males in the US economy" and suggested measures such as employment rate or some other measure that includes groups suspected of being discouraged workers in addition to the ones unemployed (Gustavsson & Österholm, 2010). Following up on the work of Gustavsson and Österholm, Liu (2014) argue that unemployment alone cannot explain the economic condition of a region and therefore region-specific employment policy should be driven in such a way so that the discouraged workers are brought back into the labor force.

[2]Definition obtained from the Bureau of Labor Statistics page: https://www.bls.gov/bls/cps_fact_sheets/lfp_mock.htm.

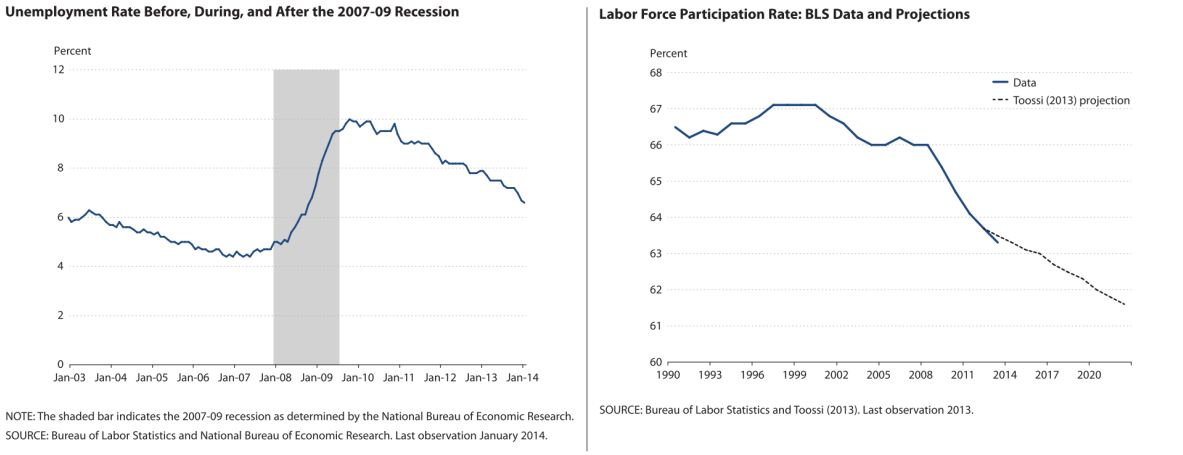

Figure 1: Unemployment and Labor Force Participation Rate in the United States

Source: (Bullard, 2014)

The second theory that I aim to test for the 2016 election is what I will call the "wage theory."[3] Jensen et al. (2016) use average pay of workers during the election year as an important regressor for the incumbent vote share. More importantly they include it in their county level analysis. Voters are expected to reward the incumbent in the event of higher average wage. The change in average wage between 2012 and 2016 seems more logical than the absolute value as it can capture the difference in economic condition between two election years and as explained above. The fact that wages for the poor have stagnated for many years could explain the frustration among voters and hence its expression through their votes (Long, 2016). Grossman and Rossi-Hansberg, for instance, claim that in addition to the loss of many jobs that off-shoring causes, it also suppresses wages of certain low-skilled sectors in a large economy like the United States (as cited in Margalit, 2011). Charles & Stephens (2013) also study the effect of local wages on voter turnout and analyze if they hold any effect on elections at different levels in the U.S.

The third theory is the "inequality theory,"[4] which can help understand if the growing inequality in the U.S. has any explanatory power when it comes to voting behavior. After all, Bernie Sanders' whole campaign was based on reducing inequality in the U.S. and it was able to garner immense support among young voters and Democratic states with majority white population. A report by World Economic Forum has also attributed the victory of Donald Trump and Brexit outcome to rising

[3]The nomenclature is given by the author for easiness to refer to the theory again throughout the thesis. This does not mean that the author claims to have proposed the theory. The ideas mentioned under this theory already exist and hence they are discussed under the Literature Review section in the first place.

[4]The nomenclature is given by the author for easiness to refer to the theory again throughout the thesis. This does not mean that the author claims to have proposed the theory. The ideas mentioned under this theory already exist and hence they are discussed under the Literature Review section in the first place.

inequality (Rodionova, 2017). Schneider-Petsinger (2016) claimed that inequality would be the largest concern among voters in the 2016 election and that 52 percent of Americans underline the gap between the rich and the poor being a very big problem. Efstathiou & Darva (2016) carried out a scientific analysis showing income inequality had a positive and statistically significant effect on Trump's victory in the election, meaning that more unequal states were more likely to vote for him. Their analysis and results provided with the necessary motivation to test the findings at a county level. Moreover, J. S. Partridge, Partridge, & Riskman (1998) study the causes behind income inequality in different regions in the United States and connect them, albeit superficially, to voting behavior and policy implications.

The fourth and one of the most important theories is non-economic in nature. I call this the "education theory"[5] and it claims that the election outcome was influenced by difference in education level among the voters. The education theory also runs parallel to the idea that people who are less educated are also more vulnerable to fake news, false economic reports, and non-scientific claims made during the campaign. Assessing that relationship would take another project in itself as it would involve analyzing whether voters vote rationally or not and how much of that rationality is dependent on their education level. Nate Silver, one of the well-known political pundits and forecasting guru, claimed that it was in fact education and not economy that predicted who would vote for Trump (Silver, 2016). Hendrickson & Galston (2016) suppose that educational attainment can have influences on both the economic conditions and cultural positions of the voters, making it a uniquely important explanatory variable. Caplan explains that congressional districts with higher educational achievements would support economic arguments such as free trade and comparative advantage (as cited in Georgia, 2004). McGill (2016) also underlines that concentration of college degrees in a county is a strong indicator to whether the county would back Republican or not and goes further to claim that education mattered more than economy or race during the recent Presidential election. Knuckey & Kim (2015), Kriner & Reeves (2012), and Franklin & Jackson (1983), all have introduced educational attainment as a control in their model to explain voting behavior, hence setting the tone for its inclusion in this thesis.

Timeframe, technique, and other peculiarities

Going back to the third question posed by Abrams (1980), it is important to understand which time-period do voters have in mind when they assess economic indicators to make voting decisions. Many authors have answered this question by testing economic voting theory with available economic data at that specific year. With the growing influence of media on voters and information available

[5]The nomenclature is given by the author for easiness to refer to the theory again throughout the thesis. This does not mean that the author claims to have proposed the theory. The ideas mentioned under this theory already exist and hence they are discussed under the Literature Review section in the first place.

to people instantly at their palms, one could make two contrasting claims- first, people utilize their ability to easily extract information and therefore build up on their knowledge about the economy over time, and second, people are influenced by the last-minute information thrown at them and base their voting decisions on events happening around the election time. Ebeid & Rodden (2006) claim that relevant academic literatures operationalize real per capita income (RPCI) as a one-year growth rate, and unemployment as an absolute level or first difference when constructing economic voting models. What this means is that voters compare the economic condition of the election year to that of the previous year. Eisenberg & Ketcham (2004) carry out a scientific examination of each of the four years in a presidential term against vote share and conclude that the most recent year explains a vast majority of the impact of economic performance. A lot of this also depends on the availability of the data. While more recent academic papers have been able to utilize the plethora of data available via various government and private agencies to carry out in-depth analyses, old papers lack such privilege. Another factor that is worth noting is that due to the growing influence of partisan media on voters, there is a possibility that voters may not even base their decision on actual economic performance but rather on their perception or forced perception of the economy. DellaVigna & Kaplan (2007) studied the effect of media bias on voting and detected the presence of Fox News Effect- that Republicans gained 0.4 to 0.7 percentage points in the towns that broadcast Fox News. Candidates on both sides use the media to put strong emphasis on campaigns, advertisements, conventions, debates, and controversies, all with the hope of generating a late swing in their favor. This suggests that the absolute or relative, or actual or perceived, economic effects can shape a voter's behavior much nearer to the election date. However, this is a debated topic because Gelman & King (1993) claim that voters cast their votes based on their "enlightened preference." They assert that although voters have incomplete information, their knowledge does get expanded over time leading up to the election day, and, throughout this period they inform themselves with the true values of the fundamental variables and their appropriate weights (Gelman & King, 1993). While analyzing whether people vote correctly, Lau & Redlawsk (1997) place "correct" voting as a middle ground between individual choice and social choice and therefore define it as the individual choice made under conditions of full information.

The types of techniques that authors use when modeling economic voting depends on many factors such as the range of the data, the variables of interest, and the research question. This thesis combines three ideas- the first idea is of looking at past elections from 1992 to 2012 to see how county level economic indicators affect Republican vote share. The second idea is to make a forecasting for 2016 Presidential election based on the regression estimation obtained from the first analysis of panel data. The election forecasting technique carried here is based on the fundamentals model that uses statistical estimation method and utilizes economic and sociopolitical data to test actual voting theories (Jerôme

& Jerôme-Speziari, 2012). The third idea is of looking at 2016 election specifically to see how much this election varies from the first model. Using the predicted and actual vote share for 2016, another regression analysis is performed on the residual (i.e. the difference in actual and predicted vote share) with economic and other non-economic rival theories described in the section above. This thesis is unique in its approach of combining the estimation models with a forecasting model and again coming back to the estimation of the residual with a clear research agenda of comparing Donald Trump's vote share with that of past Republican candidates, i.e. the Trump effect.

Methodology

This thesis analyzes the research questions using quantitative models. The specifications of the models are based on literature review and checked for biases and performed robustness tests, which are included in the Appendix. The research question is divided into two sections: the first section studies the effect of change in county level unemployment rate on Republican vote share from 1992 to 2012, and the second section studies the effect of county level economic and non-economic variables on the Republican vote share of 2016 that is not explained by the model in the first part. These two researches are carried out in three steps as described below:

Step I: Understanding U.S. Presidential elections between 1992 to 2012

Model Specification

Research Question: What is the effect of the change in county level unemployment rate on the two-party vote share of the Republican party in the Presidential election between 1992 to 2012?

Hypothesis: A growth in unemployment rate will have a negative impact on Republican vote share, since voters identify the Democratic party as the one more capable for tackling the issue of unemployment and for providing welfare and other benefits necessary in the event of high unemployment. The hypothesis is drawn from the discussion of unemployment in the literature review that presents it as a key explanatory variable of the issue-priority theory.

Regression Equation

$$rep.share_{i,t} = \alpha + \beta_1(unemp\_gro_{i,t}) + \beta_3(controls) + \epsilon \quad (1)$$

Dependent Variable

$rep.share_{i,t}$: The dependent variable is the two-party vote share of the Republican party in the U.S. Presidential election for county (i) in election year (t). Considering that it is not clear how the third-party vote share affects any of the major parties in the United States Presidential election, I have decided to eliminate the third-party vote share, as other authors have done in the past (Eisenberg & Ketcham, 2004). Hence, the Republican vote share shown here is the Republican proportion of the two-party vote share. For example, in Abbeville county in South Carolina in 2000 Presidential election, Republicans obtained 53.1 percent, Democrats obtained 45 percent, and other parties obtained 1.9 percent of the votes. Therefore, the value of *rep.share* for Abbeville in 2000 will be the ratio of Republican share to the total share of Democrats and Republicans combined, i.e. 0.541. Different scholars have used different versions of this variable in their analysis. Most of them use the vote share of the incumbent party. However, in this case, the research question aims specifically to explore the effect of county level variables on the Republican party. Hence, the usage of the Republican two-party vote share becomes logical.

Independent Variable

$unemp\_gro_{i,t}$: The main independent variable, which is also the explanatory economic variable, is the growth in unemployment rate for county (i) between election year (t) and the preceding year (t-1). For example, Abbeville county has an unemployment rate of 3.8 percent in 2000 and an unemployment rate of 5.7 percent in 1999. Therefore, the value of *unemp_gro* for Abbeville for election year 2000 will be (3.8-5.7)/5.7 = -0.33. Ebeid & Rodden (2006) in their county-level analysis of voting behavior claim that to maintain consistency and for the ease of interpretation, unemployment and other economic variables can be operationalized using a percentage change i.e. one-year growth rates, like the one done in this case.

Control Variables

$repshare.lag_{i,t-4}$: The two-party vote share of the Republican party in a county (i) in the previous Presidential election (t-4) is used as a political control variable. This variable is expected to measure the party affiliation of a specific county. In the United States, many states and counties within them are known to be historically blue or red. What this means is that the voting behavior of counties are often pre-determined based on how they voted in the election before. Not accounting for this variable could cause omitted variable bias, resulting in biased estimates for other explanatory variables, and

deteriorate the explanatory power of the model by affecting the goodness of fit (i.e. R-squared). The use of this lag-dependent variable does come with some caveats depending on the estimation method and it will be discussed in the analysis section later.

$log(Pop_{i,t})$: The population of a county (i) in the election year (t) is used as a demographic control variable. This variable is introduced to control for the size of the counties. The population variable is highly skewed due to the presence of some extremely populous counties, hence a natural log of this variable is taken to account for skewness.

$white.percent_{i,t}$: The percentage of white people (both male and female combined) in a county(i) in the election year(t) is used as another demographic control variable. This variable is introduced to control for race. The percentage is calculated by dividing the total number of white people above the age of 18 (since voting age in the United States is 18) in a county by the total population (all ages) of that county.

rep_incumb_t: A dummy variable for incumbency is introduced as another political control variable. The value of *rep_incumb* is 1 for election years that had the Republican party as the incumbent and 0 for election years that had the Democratic party as the incumbent. This variable does not change for counties in a given election year and it is extremely important in analyzing the incumbent effect discussed in the literature review section above.

$rural.percent_i$: The percentage of population living in rural areas of a county is introduced as another demographic control variable. Higher the percentage of people living in rural areas, the more rural a county is considered. This designation is based on the United States Census of 2010 that considers counties with 50 percent or more people living in rural areas as a rural county and less than 50 percent of people living in rural areas as an urban county. This variable is fixed for a given county across all years and does not appear in Fixed Effects estimate, unless interacted with other variables.

Estimation Technique

The regression estimation technique used for the model above is a Fixed Effects Model of estimation (called FE estimation, hereafter). FE estimation is ideal here due to the presence of a panel dataset and it has been preferred over Random Effects Model of estimation after performing Hausman test (See under Appendix). The FE estimation explains the 'within' county variation i.e. it explains the effect of the explanatory variables on the dependent variable for each county over a period of time. Time invariant fixed effects, such as the variable *rural.percent* is not accounted for in the FE estimation, although it may be introduced as an interaction term (which will be explained in the Analysis section below). This follows the framework by Stock and Watson who propose FE estimation to control for

omitted variables in panel data when these variables vary across entities but not over time (as cited in Sartorius, 2015).

Step II: Forecasting the Republican vote share for 2016 election

In the second step, I carry out a forecasting exercise by predicting the Republican two-party vote share for 2016 using the regression equation (1) and its coefficient estimates. Because of the FE estimation, the coefficient estimates remain the same for all the counties but each county will have a different intercept (constant). The purpose of the forecasting exercise is to analyze how well the model predicts the election outcome of 2016. Once the forecasting is done, I plot the predicted two-party vote share of the Republican party against their actual vote share in 2016 for all the counties to obtain a residual. The residual is the difference in the votes obtained by the Republican party from what it was predicted to obtain, i.e. the share of Republican party vote share that is not explained by the model in Step I. This demands the creation of another model, which aims to explain the residual and hence the Trump effect, leading us to the second research question.

Michael S. Lewis-Beck (2005) writes that forecasters revise their model specification based on trials and that one would be hard pressed to find a model that hadn't been changed at least once. This idea presents the perfect premise to introduce some new rival theories concerning 2016 election (which have been mentioned in the Literature review section above), operationalize them, and use them as explanatory variables in the next model. Since there are many possible variables that could be included in the new model, one must be careful in choosing the appropriate ones that are embedded in strong theory and do not have negative consequences on the model itself. The variables are chosen keeping that in mind, although there is a temptation to test a few variables despite a lack of strong academic theory behind them, since the election was very recent and the failure of many forecasting models pre-election suggest that sticking to only traditional fundamental variables may not do the job. In that regard, the following suggestion by Michael S. Lewis-Beck (2005) is taken into account:

> with respect to model revision, I favor the KISS strategy advocated by Arnold Zellner: Keep It Sophisticatedly Simple (García-Ferrer 1998). That is, judiciously introduce one or two theoretically potent new variables (or new measures of old variables), when the cumulating evidence suggests it. (p. 155)

Step III: Understanding the election of 2016

Model Specification

Research Question: Which theories (economic and non-economic) regarding the outcome of 2016 Presidential election, when analyzed at county level, explain the deviation in the actual vote share for the Republican party from the predicted vote share based on a traditional model?

Hypothesis: Keeping in mind that Donald Trump's victory was unprecedented and he was touted as a non-generic candidate (i.e. different than a generic Republican), the effects of county level variables should be different in 2016 election than the previous election. The additional variables that may have influenced Donald Trump's victory include the change in number of manufacturing jobs, change in average wage, income inequality, change in labor force participation rate, or level of educational attainment among voters. These variables seem to have some merit considering his campaign slogan "Make America Great Again" and his promises to raise the lives of the poor who had suffered from the wrath of globalization and the neglect of Washington.

Research Equation

$$resid = \alpha + \beta_1(manu\_share\_gro) + \beta_2(av\_wage\_gro) + \beta_3(lfpr\_male\_gro) + \beta_4(gini\_gro) + \beta_5(uneduc) + \epsilon \quad (2)$$

Dependent Variable

resid: The dependent variable is the difference in the actual two-party vote share of the Republican party in the 2016 U.S. Presidential election from their predicted vote share based on the model in Step I. Substituting the values of different explanatory variables pertaining to election year 2016 in equation (1), we will get the value of the predicted Republican vote share for each county for 2016. If the predicted Republican vote share for a county is 0.55 but the actual Republican vote share is 0.60, then that means the *residual* value for that county will be 0.60-0.55 = 0.05. This means that 5 percentage point increase in the Republican vote share is not explained by the model represented by equation (1) and therefore, we would need to test new explanatory variables, which could possibly explain this residual. A positive residual means that the Republican party outperformed our forecasting model in those counties, a negative residual means that the Republican party under-performed, and a value of 0 means that our forecasting model was able to accurately predict the actual vote share of the party for 2016. The difference in using the residual instead of the vote share for 2016, like the one used

for the model in Part I, is that we are trying to measure Trump-specific effect here. Depending on how the forecasting performs and how the residual reacts to different explanatory variables in this model, we can come to conclusions on whether Trump was a generic Republican or not and which new factors played significant role in his victory.

Variables from Rival Theories

manu_share_gro: The first independent variable is the change in manufacturing employment. It is obtained by the difference in the ratio of manufacturing jobs to total private non-farm jobs between 2012 and 2015. An ideal measurement would have been to see the difference between 2012 and 2016, since both are election years. However, the data for 2016 is not yet available and the usage of 2015 values should not be problematic considering that the explanatory variable is a slow-changing variable.

av_wage_gro: The second independent variable is the difference in average wage between 2012 and 2015. A positive value means that the average wage has increased in the county while a negative means that it has decreased.

$lfpr\_male\_gro$: The third independent variable is the change in labor force participation rate among males between 2012 and 2015. Based on the theory that the demographic group that has been discouraged to participate in the labor force is mostly men, the change in labor force participation rate among men in a county during the time frame is chosen Gustavsson & Österholm (2010). A positive value means that more men are getting back in the labor force while a negative value means that more are discouraged and not participating in the labor force.

gini_gro: The fourth variable is the change in gini coefficient for each county between 2012 and 2015. A gini coefficient is the most widely used measure of income inequality. A gini coefficient of 1 represents maximum inequality and a gini coefficient of 0 represents no inequality. In this case we will look at the difference between these two years. A positive value indicates more inequality while a negative value indicates less inequality within the county since 2012.

uneduc: The percentage of people in a county who are 25 years or older with less than high school degree in 2015 is used as a social/demographic control variable to account for the impact of education (or its lack thereof). This variable is calculated by dividing the total number of people above 25 years of age with less than high school degree by the total population of that county in 2015.

Estimation Technique

The regression estimation technique used for this model is a Ordinary Least Squares Model of estimation (called OLS estimation, hereafter). The OLS estimation is ideal here due to the presence of a cross-sectional dataset with no time-wise variation. The OLS estimation explains the average effect of the explanatory variables on the dependent variable for all the counties between 2012 and 2015. Like mentioned earlier, I do not include the same variables already used in the first model in Step I since this time we are only concerned with identifying the causes of the residual (i.e. explanatory factors additional to what had already been incorporated in the first model are sought). More importantly, I also remove the intercept values for each county since they represent county-fixed effects which have already been accounted for by the first FE estimation in Step I. Regards to the usage of the 2012 values instead of 2014 values, my assumption is based on the idea that voters compare the current economic situation with the one during the same time in the past election (See Table 6 under Appendix for an alternate analysis using 2014 data instead of 2012 data).

Data sources and cleaning

The data necessary to carry out the quantitative analysis in this thesis were obtained from multiple sources and rigorously cleaned and merged together using open source software R studio to extract the variables described above. The entire thesis is also written on R markdown and formatted using Latex. The dependent variable was obtained from election data that contained both Republican and Democratic vote share in Presidential elections between 1992 and 2016. The dataset came from Dave Leip's Atlas of U.S. Presidential Elections.[6] The lag of Republican vote share used in the first part of the research was obtained from the same election dataset. The data on unemployment from 1992 to 2015 was obtained from the Bureau of Labor Statistics (BLS).[7] The data on population for each county from 1992 to 2015 were called from the Bureau of Economic Affairs (BEA) API directly into R studio.[8] Data used to construct the incumbency dummy was created manually based on common knowledge and merged into the final dataframe (See Table 1 under Appendix to see which years had which incumbent parties and Presidents). Data on rural population was obtained from the United

[6]Dave Leip's Election Data is obtained from a third party, i.e. Çilek Agaci's github repository (Link: https://github.com/cilekagaci/us-presidential-county-1960-2016). The repository data is matched and confirmed with the data on Dave Leip's website. His dataset is often used by researchers in academia to carry out analysis involving US election results both at national and sub-national level.

[7]The data comes from the Bureau of Labor Statistics' program called "Local Area Unemployment Statistics (LAUS)" that contains annual averages of unemployment rate for each county in the United States (Link: https://www.bls.gov/lau/#cntyaa).

[8]The data on population for each county is obtained from the Bureau of Economic Affairs' API call for "RegionalIncome dataset, table CA1, Linecode 2." The data on per capita income for each county is obtained from the same API call for "RegionalIncome dataset, table CA1, Linecode 3" (Link: https://www.bea.gov/API/signup/index.cfm).

States Census Bureau.[9] Demographic data on educational attainment was not available for several years between 1992 and 2012, hence education related variable wasn't included in the model in the first part of the research. However, for the second part of the research, educational attainment data for 2015 was obtained from the United States Census Bureau as well.[10] The race data that included white population in each county between 1992 to 2015 was obtained from the database of the National Cancer Institute.[11] The counts of race and education data were converted into percentages by combining them with the population data obtained from BEA. The data on Labor Force Participation rate[12] and Gini coefficient[13] were obtained from the United States Census Bureau. The data on manufacturing jobs, total number of private non-farm jobs, and average wage were called directly from the BEA API.[14]

All the dataframes were merged based on county fips (unique identification for each county assigned by the US Census Bureau). There are a total of 3142 counties (or county-equivalents) in the United States. However, due to inconsistent nomenclature of counties and county-equivalents, and the matching problem it created since many counties have been reshaped over the years, some counties were removed from the dataset. All the 29 boroughs of Alaska were removed from the final dataset due to difficulty in matching them to the county fips across multiple datasets. Maui and Kalawao counties of Hawaii were merged together into the same county called Maui to make it consistent over multiple datasets.

[9]The County Rurality Level in 2010 was obtained from U.S. Census Bureau, 2010 Census, Summary File 1, Table P2. The data is based on 2010 Census while the county fips codes and names are updated as of 2015 (Link: https://www2.census.gov/geo/docs/reference/ua/County_Rural_Lookup.xlsx).

[10]The data on educational attainment is obtained from U.S. Census Bureau, 2011-2015 American Community Survey 5-Year Estimates, Table B15003, Educational Attainment for the Population 25 Years and Over. (Link: https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_15_5YR_B15003&prodType=table).

[11]The data on race is obtained from National Cancer Institute's "U.S. Population Data" dataset, which was made based on the Census Bureau estimates. (Link: https://seer.cancer.gov/popdata/download.html).

[12]The data on Labor Force participation rate is obtained from U.S. Census Bureau, 2011-2015 American Community Survey 5-Year Estimates, Table S2301, Employment Status. The data on 2012 is obtained from 2008-2012 American Community Survey 5-Year Estimates. (Link: https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_15_5YR_S2301&prodType=table.

[13]The data on Gini coefficient for 2015 is obtained from U.S. Census Bureau, 2011-2015 American Community Survey 5-Year Estimates, Table B19083, Gini Index of Income Inequality. The data for 2012 is obtained from the 2008-2012 American Community Survey 5-Year Estimates, Table B19083. (Link: https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?pid=ACS_15_5YR_B19083&prodType=table)

[14]The data on manufacturing jobs for each county is obtained from the Bureau of Economic Affairs' API call for "RegionalIncome dataset, table CA25N, Linecode 500." The data on total private non-farm employment for each county is obtained from the same API call for "RegionalIncome dataset, table CA25N, Linecode 90." The data on average wage was obtained from "RegionalIncome dataset, table CA30, Linecode 290." (Link: https://www.bea.gov/API/signup/index.cfm).

Result and Analysis

Step I: Understanding U.S. Presidential elections between 1992 to 2012

Descriptive Statistics

Figure 2 shows how the main independent variable, the lag dependent variable, and the dependent variable are correlated to one another, and therefore provide some basic understanding about them. The growth in unemployment is negatively, and weakly, correlated (r = -0.042) with the Republican two-party vote share. However, the Republican vote share in the previous election is positively, and strongly, correlated (r = 0.887) with their present vote share. Both correlations are in line with the theory established earlier in the thesis. The density plots of each of these variables show a normal distribution with no noticeable skewness for both the Republican vote share and the lag of Republican vote share. Although, the growth in unemployment shows a right-skew, it doesn't affect the analysis intended to be carried out since the normality assumption of the residuals is not violated and we will correct for the biased standard errors using robustness tests below.

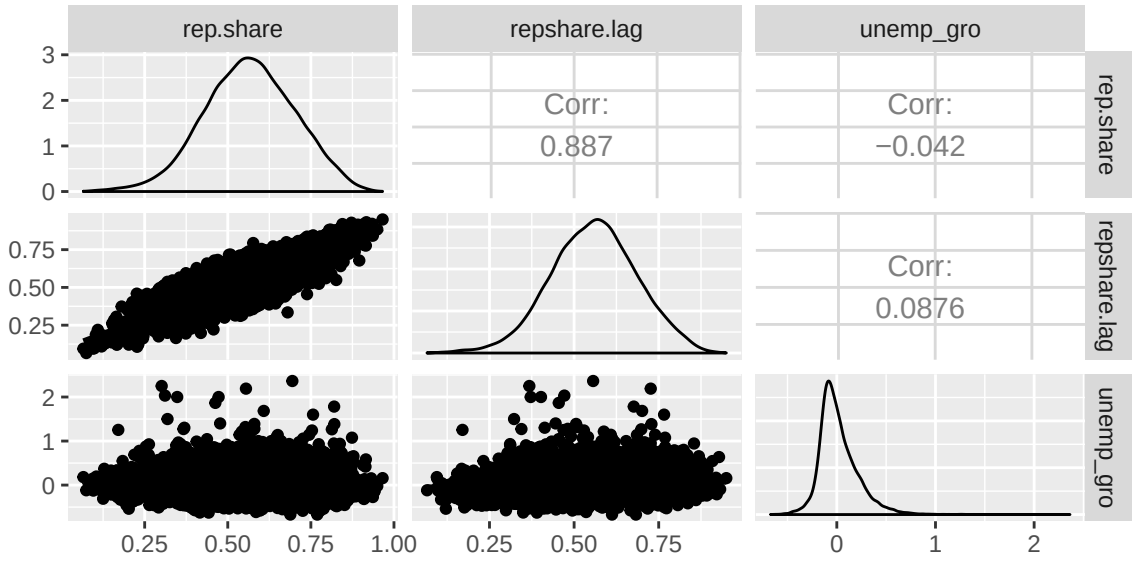

Figure 2: Correlation Matrix for Main Independent Variables in Step I

Figure 3 shows the correlation between the dependent variable and other control variables. Republican two-party vote share has moderately strong, and positive, correlation with both the percentage of white people living in a county (r = 0.346) and the percentage of people living in rural areas of the county (r = 0.213). We also see a positive correlation, yet not very strong, between rural population and white population (r = 0.16), depicting the situation in the United States where most rural areas

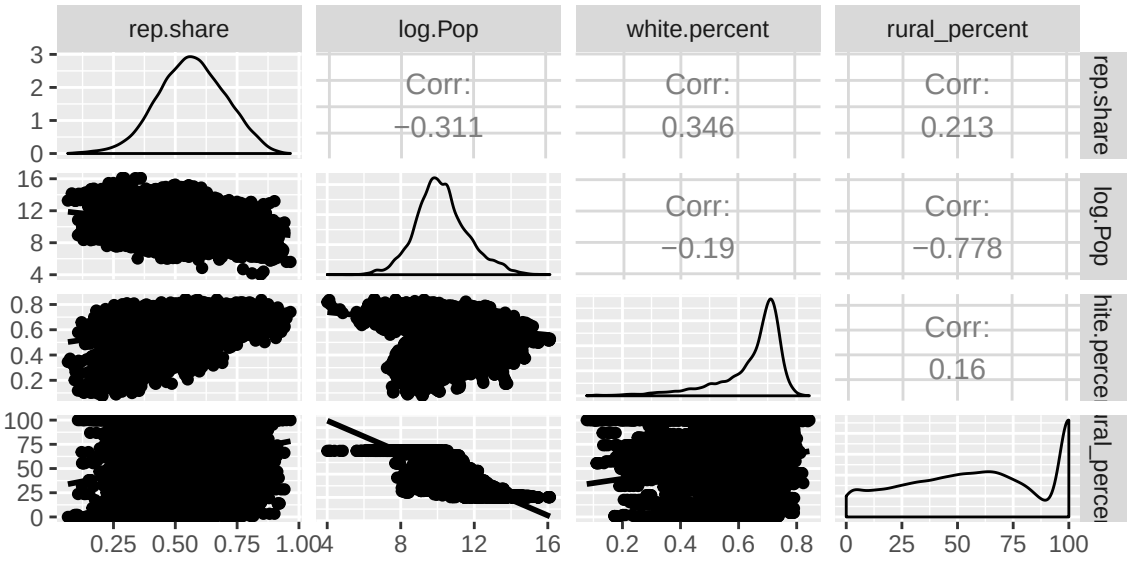

Figure 3: Correlation Matrix of Control Variables in Step I

are considered to be homogeneous while the urban areas are more heterogeneous. It is important to note that counties with larger population have a negative correlation with Republican vote share (r = -0.311) and also show that these counties are less rural (r = -0.778) and have lower white percentage(r = -0.19).

Initially, the distribution plot showed an extreme right skewness for population due to the presence of some extremely populous counties like the Los Angeles county in California that has more than 9 million people. This is corrected for by taking a natural log of the population. Similarly, the plot of rural_percent shows a unique phenomenon for the United States. It shows that there are approximately same number of counties in the U.S. at different levels of rurality. Meaning, the number of counties that are highly urban (when *rural_percent* is at 0) is similar to the number of counties that are moderately urban (when *rural_percent* is at 50). But once the rurality reaches to around 80 percent or above, there is a sharp increase in the number of counties. This shows that the extremely-rural counties are disproportionately high in number than counties at any other rurality level. If this somehow influences the way people are voting in the U.S. is an interesting question to explore. The descriptive summary table (see Table 2 under Appendix) gives a complete overview of the central tendency and the variation of all the variables.

Regression Output and Results

Table 1 shows the regression output with FE estimate, showing the relationship between unemployment growth and other political and demographic controls for U.S. Presidential elections between 1992 and 2012.

Table 1: FE Estimate to explain Republican two-party vote share(1992-2012)

	(1)	(2)	(3)	(4)	(5)	(6)
Unemp gro	-0.08^{***}	-0.08^{***}	-0.02^{***}	-0.02^{***}	-0.01^{***}	-0.01^{***}
	(0.002)	(0.003)	(0.003)	(0.004)	(0.002)	(0.002)
Prev voteshare	0.62^{***}	0.62^{***}	0.71^{***}	0.71^{***}	0.67^{***}	0.67^{***}
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Pop logged			0.13^{***}	0.13^{***}	0.01^{***}	0.01^{***}
			(0.004)	(0.01)	(0.004)	(0.005)
White			0.27^{***}	0.27^{***}	0.44^{***}	0.44^{***}
			(0.05)	(0.05)	(0.04)	(0.04)
Incumbent			-0.04^{***}	-0.04^{***}	-0.09^{***}	-0.09^{***}
			(0.001)	(0.001)	(0.001)	(0.002)
Unemp groxIncumbent			-0.05^{***}	-0.05^{***}	0.003	0.003
			(0.004)	(0.005)	(0.004)	(0.004)
WhitexRural			0.004^{***}	0.004^{***}	-0.004^{***}	-0.004^{***}
			(0.001)	(0.001)	(0.001)	(0.001)
Time Effects	N	N	N	N	Y	Y
Arellano Bond	N	Y	N	Y	N	Y
Observations	18,662		18,328		18,328	
R^2	0.36		0.49		0.71	
Adjusted R^2	0.24		0.38		0.66	
Note:						*p<0.1; **p<0.05; ***p<0.01

Due to the inclusion of a lag dependent variable, an Arellano-Bond test is conducted to account for nickell bias. Moreover, to correct for time-fixed effects, we add year dummies on the model. Therefore, the estimates in the sixth column, with corrected standard errors, and with all relevant control variables, are used for interpretation. The model explains 71 percent of the variation in the Republican vote share (given by R-square value) over the period of 1992 and 2012, which is a substantial result considering the plethora of factors that play a role in determining voting behavior. Growth in unemployment has a negative effect on Republican two-party vote share with high statistical significance. One percent increase in unemployment rate between the year preceding the election and the election year will result in a fall of Republican two-party vote share by 0.01 percentage points. This result is in line with the

issue-priority theory, which says that rise in unemployment rate benefits the Democratic party. It is important to note that the effect of the growth in unemployment is obtained by controlling for the Republican vote share from previous election which shows an extremely high effect on the current vote share with high statistical significance. The estimates also show that a 1 percentage point increase in white population in a county increases the Republican vote share by 0.44 percentage points.

Validating Norpoth & Gschwend (2010)'s cost of ruling theory, the incumbency factor shows that if the incumbent is a Republican, then the Republican vote share goes down, suggesting that voters do not like to keep the same party in power for many terms. However, the marginal effect of unemployment growth on the Republican vote share, in the event of a Republican incumbency, is insignificant. This result is in line with the idea that the effect of unemployment on Republican vote share is not affected by incumbency.

This model explains the "within variation" for each county in the United States and therefore excludes any time-invariant effects. Since, rurality is a time-invariant variable, we interact it with *white.percent* to capture its marginal effect. Surprisingly the marginal effect of the rise in white population on Republican vote share slightly falls if the county is more rural. This could be explained by the fact that rural counties have mostly white people anyway; therefore, the dominance of their "white vote" decreases in favor of other local issues, in turn decreasing the marginal effect.

Step II: Forecasting the Republican vote share for 2016 election

From the model above in Step I, unique coefficient estimates for each variable and a unique intercept for each county are obtained. By substituting the values of these variables for 2015 (see above under Methodology on why 2015 values are used instead of 2016 values) in equation (1), the Republican two-party vote share for each county in the 2016 election is forecasted. Figure 4 below compares the predicted Republican two-party vote share with their actual two-party vote share obtained during the election.

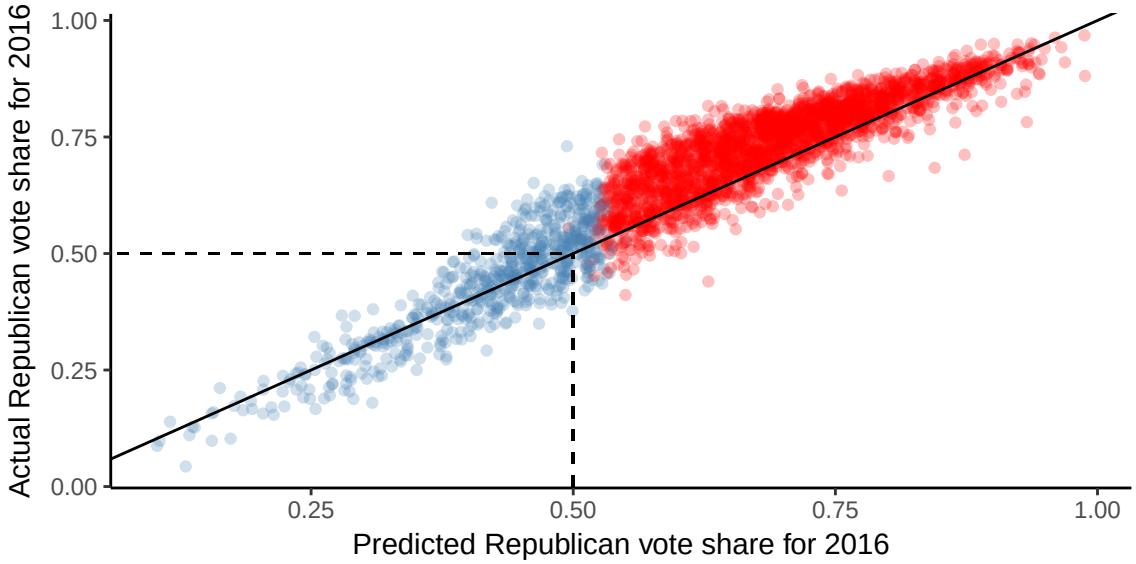

Figure 4: Residuals from 2016 U.S. Presidential Elections Forecasting

The counties that fall on the 45-degree line represent the ones where the model was able to accurately predict the actual vote share for 2016. There are many counties (above and below the 45-degree line), where there are residuals, i.e. difference in the predicted and actual vote share. The graph shows that there are higher number counties where the model under-predicted the Republican vote share than overpredicted. The red dots represent the counties that had voted Republican in the 2012 election. In 2016, almost all of these counties stayed Republican. However, interestingly many counties that voted Democrat in the 2012 election turned Republican in 2016 (these are the blue dots outside of the dashed lines).

Table 4 (under Appendix) shows the states where the previously Democratic counties turned Republican and in how many of those counties the model made an underprediction of the Republican vote share. 221 counties changed from Democrat to Republican between 2012 and 2016. Out of them, the Republican vote share for 212 counties were under-predicted by the model. Table 5 (under Appendix) shows the

states where the counties stayed Republican between 2012 and 2016 and in how many of those counties the model made an underprediction of the Republican vote share. 2361 counties stayed Republican between 2012 and 2016. Out of them, the Republican vote share of 1945 counties were under-predicted by the model.

In a nutshell, what that means is that the model under-predicted the Republican vote share in most of the counties that were won by Donald Trump in 2016. This may show that the forecast was a failure. However, the maps below showcase an interesting observation. They show that the areas where the Republican vote share in 2016 was higher than in 2012 were also the areas where the residual was higher. The color of the maps are flipped to emphasize this observation. The lighter shade areas in Figure 5 (around the Rust-belt and previously Democratic states) show a high change in Republican vote share since 2012. The same areas in Figure 6 show a higher residual (i.e. an underprediction of the Republican vote share by the generic model). This shows that the model unpredicted exactly the way the voting shift happened. The inference that can be drawn is that Donald Trump was not a generic Republican. The model would have performed well for a generic Republican as it shows stark similarity with the vote share from 2012.

What caused the Republican-leaning counties to vote significantly higher for Donald Trump and what led the Democratic -leaning counties to switch to Republican is a question that needs more exploration and this forecasting exercise provides the perfect premise to do so. To understand this difference (residual), various rival hypothesis are tested with the aim of explaining the effect that Donald Trump had that other generic Republicans didn't have in the past.

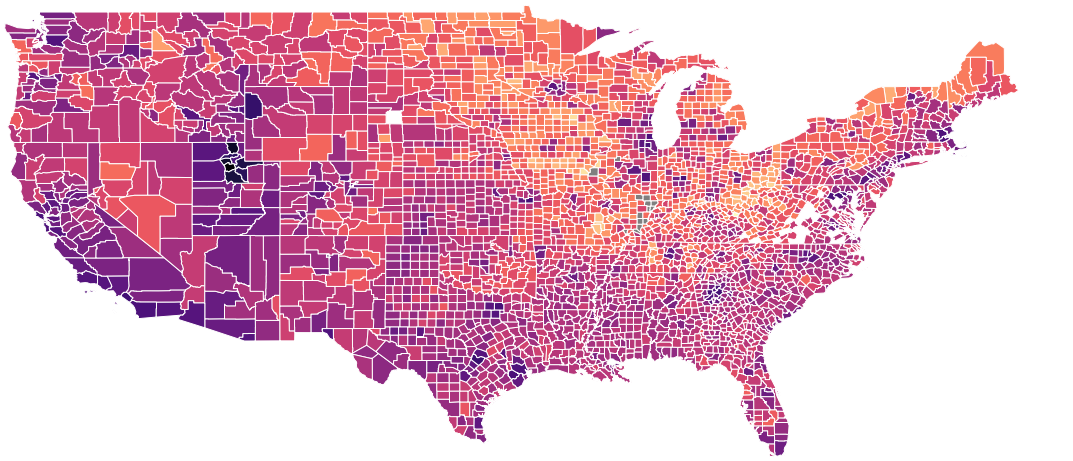

Figure 5: Change in voteshare for Republican party 2012 to 2016

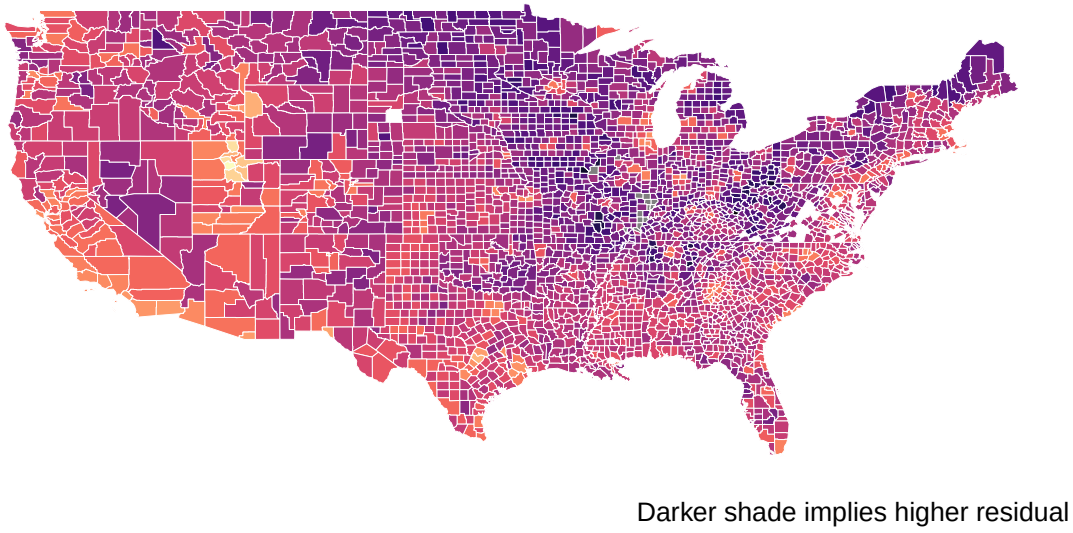

Figure 6: Residual from the forecasting of 2016 election

Step III: Understanding the election of 2016

Descriptive Statistics

The 2016 Presidential election took everyone by surprise and to explain it has been an arduous task. Since, it has been established that a traditional model doesn't suffice when it comes to accurately exploring the 2016 election, it is important to resort to some rival theories and test their relevance. In Figure 7 below, the correlation matrix of some of the rival variables is presented alongside the residual obtained from the forecasting exercise to create a premise for further exploration. Increase in the share of manufacturing jobs and labor force participation rate among men are positively correlated with the residual, which is surprising considering the hypothesis. However, the decrease in wage and increase in inequality are both positively correlated with the residual and are in line with the proposed hypothesis. More importantly, lower education level is also strongly, and positively, correlated with the residual. The density plots show a normal distribution for all the variables with a weak skewness in the case of education level. The descriptive summary table (see Table 3 under Appendix) gives a complete overview of the central tendency and the variation of all the variables.

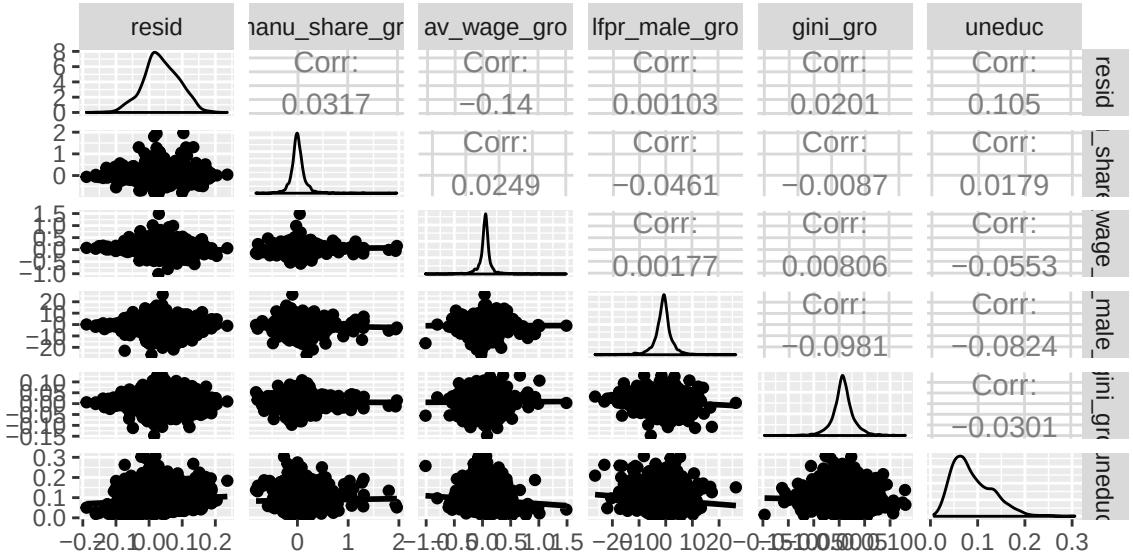

Figure 7: Correlation Matrix of Rival Theories for 2016 election

Regression Output and Results

The regression output below in Table 4 shows the effect of various variables from rival theories on county level voting of Presidential elections. The output compares all U.S. counties with the counties in swing states, and with the counties in rust-belt states. The reason for this differentiation is driven by the way swing states acted during the Presidential election and the how the rust-belt states, that had previously voted for Obama favored Trump this time (See Appendix for the list of swing and rust-belt states).

Table 2: Testing Rival Theories for 2016 Residuals using OLS Estimate

	resid		
	All counties	Swing-state counties	Rust-belt counties
	(1)	(2)	(3)
manu_share_gro	0.01*	0.001	0.01
	(0.01)	(0.02)	(0.01)
av_wage_gro	−0.06***	0.07**	−0.03
	(0.01)	(0.03)	(0.02)
lfpr_male_gro	0.0002	−0.0004	−0.001*
	(0.0003)	(0.001)	(0.001)
gini_gro	0.16***	0.24*	0.01
	(0.06)	(0.14)	(0.13)
uneduc	0.38***	0.42***	0.96***
	(0.01)	(0.04)	(0.03)
Observations	2,628	608	668
R^2	0.31	0.36	0.69
Adjusted R^2	0.31	0.35	0.69
Residual Std. Error	0.05 (df = 2623)	0.06 (df = 603)	0.05 (df = 663)
F Statistic	234.16*** (df = 5; 2623)	67.56*** (df = 5; 603)	300.49*** (df = 5; 663)
Note:			*p<0.1; **p<0.05; ***p<0.01

The first hypothesis that the change in employment scenario had a role to play in increasing votes for Trump can be rejected when we look at the whole country's example. Neither the change in manufacturing jobs nor the labor force participation rate among men show any significant effect on the increased Republican support in 2016. The benefits of globalization and a boom in service industry has helped many counties around the country as the loss in labor-intensive manufacturing-based jobs have been accepted by the labor force. Inclusion of women in the workforce and their contribution to the economy can also be seen as a reason why the loss of specific types of jobs do not show a big effect on

the national level. However, if we look at the effect only in rust-belt states, we see that the labor force participation rate among men to be moderately significant, although with weak influence, in explaining the rise in Republican vote share in those counties. The effect of the decrease in labor force participation rate among men on the rise of Republican vote share suggests a growing inclination of discouraged white workers to look at the Republican party (or Trump in this case) for support. Their frustration is justified by the changing employment scenario where manufacturing jobs are being replaced by cheap labor in foreign countries and that these people aren't provided with enough opportunities to make a good living through another sector.

The second theory regarding the influence of wage changes in the rise of Republican vote share holds partially true and in somewhat ambiguous way. Overall in the United States, we see that a decline in average wage benefits the Republican vote share in a highly significant way, meaning people who are resented that their wages have not gone up are looking for a change in the government. However, as we change the scenario to just swing states, the significance of this variable declines and even changes its nature. In swing states, as average wage goes up, so does the support for Republican vote share and this is the ambiguity that needs more exploration.

The third theory regarding inequality is also significant when analyzing the counties all over the country as we can see that a rise in gini coefficient, meaning high inequality, has a positive effect on Republican vote share. This is also surprising at some level because it was Bernie Sanders' campaign that was more dedicated to tackling this issue. But one can assume that his loss in the Democratic Primaries could have turned many of his supporters either to the Republican side or turned them into non-voters. This assumption would require testing which is not carried out in this thesis.

The theory that holds true unanimously, in overall analysis of all counties in the US, and also in more focused analysis of swing states and rust-belt states is the "education theory." The higher the number of people in a county with less than high school degree, the bigger their effect in the rise of the Republican vote share in the election. A one percent increase in the number of uneducated people shows a 0.38 percentage point increase in the vote share for the Republican party. In counties of rust belt states, the effect is even more with 1 percent increase in uneducated people increasing the residual by almost one percentage point.

Conclusion

Voting behavior is one of the highly-explored subjects in political science yet one of the unpredictable ones. If we knew who would win an election based on historical data alone, then politics would not be as interesting as it is now. This uncertainty, yet the necessity, in understanding election outcomes and their social, political, and economic causes make the study of voting behavior extremely important in today's rapidly changing world. This thesis was motivated with the idea of understanding a simple question- "why people vote the way they do?" To answer it, the best subject for exploration was provided by the recently concluded Presidential election in the United States that saw the rise of an unprecedented candidate in Donald Trump. The election was unique in two ways - in the outcome and in the factors leading up to it. In this thesis, I tried to connect the two to get a step closer in understanding what influenced it. To do this, first I questioned what factors influence the vote share of the Republican party in the United States in general. Regression analysis of historical data, spanning between 1992 and 2012 and consisting of various economic, political, and demographic variables at county level suggested that the two-party vote share of the Republican party is highly affected by how the counties voted in previous election. This put the biggest share of burden on the political identities that counties have preserved historically and how they are hard to change. Economic variable like change in unemployment showed some significant effect, yet very weak. However, demographic identity of a county, i.e. race, showed strong effect on whether the county would lean Democratic or Republican. These findings weren't new or difference-making in the study of voting behavior. However, they provided me with a statistically sound model to use for the forecasting of the 2016 election and get closer to answering the Trump question that I began with.

Forecasting an election that had already occurred may not seem such a fascinating idea. In this case though, the relevance of such an endeavor comes from the fact that it helped us understand whether the Republican candidate in 2016 was a generic Republican or not. A generic Republican would be the one that would fit the model created before. The unique similarity between the residuals obtained from the difference in actual and predicted vote share, and the difference in vote share between 2012 and 2016 elections showed that Trump was not a generic Republican and his effect could not have been completely measured by the traditional model. The second model that was created to explain this residual using some, but not all, rival theories shed some light on the significance of their effect. Through a county level analysis, and separated even to account for swing states and rust-belt states, we find that only education stands as a variable with both strong and significant explanatory power in describing the 2016 election. In this process, we reject many theories regarding the effect of the change in manufacturing employment, labor force participation rate among men, inequality, or changes

in wages, on Republican vote share. The policy implication of the study is huge, as it not only provides an insight into the voting behavior of different demographic groups but also the motivations behind their vote. It is scary, yet insightful, to realize that education, or its lack thereof, in a small county in the United States can carry significant implications on events around the world.

References

Abrams, B. A. (1980). The influence of state-level economic conditions on presidential elections. *Public Choice*, *35*(5), 623–631.

Abrams, B. A., & Butkiewicz, J. L. (1995). The influence of state-level economic conditions on the 1992 U.S. presidential election. *Public Choice*, *85*(1), 1–10.

Anonymous. (1998). The benevolence of self-interest. *The Economist*. Retrieved from http://www.economist.com/node/179495{\#}print

Blackley, P. R., & Shepard III, E. M. (1994). A Statistical Analysis of the Effects of State-Level Economic Conditions on the 1992 Presidential Election. *Public Finance Quarterly*, *22*(3), 366–382.

Bullard, J. (2014). *The Rise and Fall of Labor Force Participation in the United States* (pp. 1–12). St. Louis: Federal Reserve Bank of St. Louis.

Campbell, A., Converse, P. E., Miller, W. E., & Donald, E. (1966). Stokes. 1960. *The american voter*, *8*.

Charles, K. K., & Stephens, M. J. (2013). Employment, wages, and voter turnout. *American Economic Journal: Applied Economics*, *5*(4), 111–143.

DellaVigna, S., & Kaplan, E. (2007). The Fox News Effect: Media Bias and Voting. *The Quarterly Journal of Economics*, *3*(122), 1187–1234.

Ebeid, M., & Rodden, J. (2006). Economic Geography and Economic Voting: Evidence from the US States. *British Journal of Political Science*, *36*(03), 527–547.

Efstathiou, K., & Darva, Z. (2016). Income inequality boosted Trump vote. *Bruegel*. Retrieved from http://bruegel.org/2016/11/income-inequality-boosted-trump-vote/

Eisenberg, D., & Ketcham, J. (2004). Economic Voting in U.S. Presidential Elections : Who Blames Whom for What. *Topics in Economic Analysis and Policy*, *4*(1).

Franklin, C. H., & Jackson, J. E. (1983). The Dynamics of Party Identification. *The American Political Science Review*, *77*(4), 957–973.

Galbraith, J. K., & Hale, J. T. (2008). State Income Inequality and Presidential Election Turnout and Outcomes. *Social Science Quarterly*, *89*(4), 887–901.

Gelman, A., & King, G. (1993). Why Are American Presidential Election Campaign Polls so Variable

When Votes Are so Predictable? *British Journal of Political Science*, *23*(4), 409–451.

Georgia, P. J. (2004). *The Political Economy of Trade Policy in the United States: 1990 to 2000* (Thesis). George Mason University.

Glaeser, E. L., & Ward, B. A. (2006). Myths and Realities of American Political Geography. *Journal of Economic Perspectives*, *20*(2), 119–144.

Gustavsson, M., & Österholm, P. (2010). *Labor-Force Participation Rates and the Informational Value of Unemployment Rates in US: Evidence from Disaggregated US Data*. Uppsala: Department of Economics, Uppsala University.

Hendrickson, C., & Galston, W. A. (2016). The educational rift in the 2016 election. *Brookings Institution*. Retrieved from https://www.brookings.edu/blog/fixgov/2016/11/18/educational-rift-in-2016-election/

Jensen, J. B., Quinn, D. P., & Weymouth, S. (2016). *Winners and Losers in International Trade: The Effects on U.S. Presidential Voting*. Cambridge: National Bureau of Economic Research.

Jerôme, B., & Jerôme-Speziari, V. (2012). Forecasting the 2012 US Presidential Election: Lessons from a State-by-State Political Economy Model. *Political Science & Politics*, *45*(4), 663–668.

Kayser, M. A., & Wlezien, C. (2011). Performance pressure: Patterns of partisanship and the economic vote. *European Journal of Political Research*, *50*(3), 365–394.

Kim, J., Elliott, E., & Wang, D.-M. (2003). A spatial analysis of county-level outcomes in US Presidential elections: 1988-2000. *Electoral Studies*, *22*(4), 741–761.

Knuckey, J., & Kim, M. (2015). Racial Resentment, Old-Fashioned Racism, and the Vote Choice of Southern and Nonsouthern Whites in the 2012 U.S. Presidential Election. *Social Science Quarterly*, *96*(4), 905–922.

Kriner, D. L., & Reeves, A. (2012). The Influence of Federal Spending on Presidential Elections. *American Political Science Review*, *106*(2), 348–366.

Lau, R. R., & Redlawsk, D. P. (1997). Voting Correctly. *The American Political Science Review*, *91*(3), 585–598.

Lewis-Beck, M. S. (2005). Election Forecasting : Principles and Practice. *The British Journal of Politics and International Relations*, *7*(2), 145–164.

Lewis-Beck, M. S., & Nadeau, R. (2011). Economic voting theory: Testing new dimensions. *Electoral*

Studies, *30*(2), 288–294.

Liu, D.-C. (2014). Labor-Force Participation Rates and the Informational Value of Unemployment Rates in US: Evidence from Regional Data. *Social Indicators Research*, *116*(2), 447–455.

Long, H. (2016). U.S. inequality keeps getting uglier. *CNN*. Retrieved from http://money.cnn.com/2016/12/22/news/economy/us-inequality-worse/

Margalit, Y. (2011). Costly Jobs: Trade-related Layoffs, Government Compensation, and Voting in U.S. Elections. *American Political Science Review*, *105*(1), 166–188.

McGill, A. (2016). America's Educational Divide Put Trump in the White House. *The Atlantic*.

Norpoth, H., & Gschwend, T. (2010). The chancellor model: Forecasting German elections. *International Journal of Forecasting*, *26*(1), 42–53.

Partridge, J. S., Partridge, M. D., & Riskman, D. S. (1998). State Patterns in Family Income Inequality. *Contemporary Economic Policy*, *16*(3), 277–294.

Rodionova, Z. (2017). Rising inequality behind Brexit and Donald Trump victory is top global threat for the next decade, says WEF. *The Independent*. Retrieved from http://www.independent.co.uk/news/business/news/wealth-gap-global-income-inequality-brexit-donald-trump-victory-threat-a7521286.html

Sartorius, M. R. (2015). *County Level Economic Voting in U.S. Presidential Elections* (CMC Senior Theses. Paper 1172). Claremont McKenna College.

Schneider-Petsinger, M. (2016). Inequality Defines the American Election. *RealClearWorld*. Retrieved from http://www.realclearworld.com/articles/2016/10/18/inequality{_}defines{_}the{_}american{_}election{_}112

Silver, N. (2016). Education, Not Income, Predicted Who Would Vote For Trump. *FiveThirtyEight*. Retrieved from http://fivethirtyeight.com/features/education-not-income-predicted-who-would-vote-for-trump/

Vavreck, L. (2009). *The Message Matters*. Princeton: Princeton University Press.

Wright, J. R. (2012). Unemployment and the Democratic Electoral Advantage. *American Political Science Review*, *106*(4), 1–18.

Appendix:

Table 1: Incumbents and Winners in the U.S. election (1992-2016)

	Election Year	Incumbent President	Incumbent Party	Winner	Winner Party
1	1,992	George H.W. Bush	Republican	Bill Clinton	Democratic
2	1,996	Bill Clinton	Democratic	Bill Clinton	Democratic
3	2,000	Bill Clinton	Democratic	George W. Bush	Republican
4	2,004	George W. Bush	Republican	George W. Bush	Republican
5	2,008	George W. Bush	Republican	Barack Obama	Democratic
6	2,012	Barack Obama	Democratic	Barack Obama	Democratic
7	2,016	Barack Obama	Democratic	Donald Trump	Republican

Table 2: Step I Variables: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
rep.share	18,663	0.57	0.14	0.07	0.97
repshare.lag	18,662	0.56	0.13	0.07	0.95
rep_incumb	18,672	0.50	0.50	0	1
unemp_gro	18,672	0.02	0.20	−0.67	2.36
rural_percent	18,672	58.49	31.44	0.00	100.00
white.percent	18,329	0.65	0.11	0.08	0.84
Pop	18,363	92,431.41	299,329.00	55	9,970,436

Table 3: Step III Variables: Summary Statistics

Statistic	N	Mean	St. Dev.	Min	Max
resid	3,045	0.04	0.05	−0.19	0.24
rep.share	3,045	0.67	0.16	0.04	0.97
repshare.lag	3,045	0.61	0.15	0.07	0.97
pred_repshare	3,045	0.63	0.15	0.10	0.99
is.rep.2012	3,045	0.78	0.41	0	1
is.rep.2016	3,045	0.85	0.36	0	1
pop	3,059	103,670.00	332,792.50	112	10,170,292
unemp_gro	3,059	−0.28	0.13	−0.70	1.00
manu_share_gro	2,640	0.01	0.17	−0.82	1.96
lfpr_male_gro	3,059	−0.98	3.46	−26.50	26.30
av_wage_gro	3,053	0.05	0.12	−1.01	1.49
gini_gro	3,059	0.01	0.02	−0.15	0.13
uneduc	3,059	0.09	0.04	0.01	0.31

Table 4: Number of counties that turned Republican from Democrat between 2012 and 2016 and were underpredicted

	State	Counties D to R	Counties Underpredicted
1	AL	2	2
2	AR	1	1
3	CO	5	4
4	CT	1	1
5	DE	1	1
6	FL	4	4
7	GA	6	6
8	IA	33	33
9	IL	10	10
10	IN	5	5
11	KY	2	2
12	MD	1	1
13	ME	8	8
14	MI	12	12
15	MN	19	19
16	MS	4	3
17	MT	3	3
18	NC	7	4
19	ND	4	4
20	NE	1	1
21	NH	3	2
22	NJ	2	2
23	NM	3	3
24	NY	20	20
25	OH	10	10
26	OR	2	2
27	PA	3	3
28	RI	1	1
29	SC	6	6
30	SD	5	5
31	TN	1	1
32	TX	1	
33	VA	6	5
34	VT	1	1
35	WA	5	4
36	WI	23	23

Table 5: Number of counties that stayed Republican between 2012 and 2016 and were underpredicted

	State	Counties R to R	Counties underpredicted
1	AL	52	42
2	AR	66	63
3	AZ	11	4
4	CA	25	7
5	CO	37	31
6	CT	1	1
7	DE	1	1
8	FL	54	42
9	GA	122	94
10	IA	60	58
11	ID	42	27
12	IL	66	62
13	IN	83	74
14	KS	103	71
15	KY	116	107
16	LA	54	46
17	MD	16	12
18	ME	1	1
19	MI	63	57
20	MN	59	57
21	MO	111	109
22	MS	51	43
23	MT	47	38
24	NC	69	51
25	ND	47	47
26	NE	90	84
27	NH	3	2
28	NJ	7	5
29	NM	16	13
30	NV	15	14
31	NY	26	25
32	OH	70	65
33	OK	77	69
34	OR	26	20
35	PA	53	49
36	SC	25	16
37	SD	56	53
38	TN	91	88
39	TX	226	126
40	UT	27	3
41	VA	60	50
42	WA	22	13
43	WI	37	34
44	WV	55	54
45	WY	22	17

Alternate analysis for Step III:

Testing Rival Theories for 2016 using the difference in 2015 economic vaiables against 2014 economic variables

Table 6: Testing Rival Theories for 2016 Residuals using OLS Estimate

	resid		
	All counties	Swing-state counties	Rust-belt counties
	(1)	(2)	(3)
manu_share_gro	0.003	−0.005	0.03
	(0.01)	(0.03)	(0.03)
av_wage_gro	−0.16***	−0.11*	−0.19***
	(0.02)	(0.06)	(0.04)
lfpr_male_gro	0.0004	−0.001	−0.001
	(0.001)	(0.002)	(0.002)
gini_gro	−0.02	−0.01	−0.14
	(0.09)	(0.23)	(0.20)
uneduc	0.39***	0.52***	0.99***
	(0.01)	(0.03)	(0.03)
Observations	2,670	617	673
R^2	0.34	0.38	0.71
Adjusted R^2	0.33	0.37	0.71
Residual Std. Error	0.05 (df = 2665)	0.06 (df = 612)	0.05 (df = 668)
F Statistic	269.86*** (df = 5; 2665)	74.27*** (df = 5; 612)	333.83*** (df = 5; 668)
Note:			*p<0.1; **p<0.05; ***p<0.01

Table 7: Swing States and their Counties

	State	Number of counties
1	CO	64
2	FL	67
3	IA	99
4	NC	100
5	NH	10
6	NV	17
7	OH	88
8	PA	67
9	VA	81
10	WI	72

Source: https://en.wikipedia.org/wiki/Swing_state

Table 8: Rust-belt States and their Counties

	State	Number of counties
1	IA	99
2	IL	102
3	IN	92
4	MI	83
5	NY	62
6	OH	88
7	PA	67
8	WI	72
9	WV	55

Source: https://en.wikipedia.org/wiki/Rust_Belt

Hausman Test for Step I:

```
Hausman Test
```

data: rep.share ~ unemp_gro + repshare.lag + log(Pop) + white.percent + ... chisq = 2448.9, df = 7, p-value < 2.2e-16 alternative hypothesis: one model is inconsistent

Time-Fixed Effects Test for Step I:

```
Lagrange Multiplier Test - time effects (Breusch-Pagan) for
unbalanced panels
```

data: rep.share ~ unemp_gro + repshare.lag + log(Pop) + white.percent + ... chisq = 3803100, df = 1, p-value < 2.2e-16 alternative hypothesis: significant effects