

What happened on November 8, 2016 ?: A Study of Republican Vote Share and County-level Economics at US Presidential Elections

Raju Adhikari

March 16, 2017

Contents

Acknowledgement:	3
Abstract:	4
Table of Contents:	5
Executive Summary:	6
Introduction:	7
Literature Review:	8
Economic Voting Theory:	8
Macro-level analysis vs Regional Analysis (A Case for Counties):	10
Political and Demographic Controls:	13
Some Theories for 2016 US Presidential election outcome:	14
Timeframe, technique, and other peculiarities:	17
Methodology:	19
Step I:	19
Step II:	21
Step III:	22
Data sources and cleaning:	25
Result and Analysis:	27

First Part: Creating and analyzing a model to explain the Republican two-party vote share between 1992 and 2012	27
Second Part: Using the model to make a forecast for 2016 Presidential Election	30
Third Part: Creating and analyzing a model to explain the residual obtained from the forecasting exercise i.e. “Trump effect”	31
Conclusion:	38
Limitations and Further Research:	39
Appendix:	40
Rough Notes:	41

Acknowledgement:

write once everything is written

Abstract:

write once everything is written

Table of Contents:

Executive Summary:

"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 and Richard Nadeau (2011)

Introduction:

give basic intro

point out the two main questions of research Two main questions: We start by analyzing the question from a party-perspective and then move to a specific candidate, in this case being Donald Trump....

- Do county-level economic conditions influence presidential election results in the US? - effect on specific party vote share.
- Did county-level economic conditions play a role in the victory for the Republican party in 2016 elections (i.e. the victory of Donald Trump) and what can they say about Trump- specific effects (Does this show that Trump was a generic Republican or not)?

outline section-by-section details

Literature Review:

In the literature review section, I discuss various authors, theories and analysis used in the past on topics similar to the research that I am performing in this paper. Discussing the similarities and differences that occur among authors and the evolution of various theories and techniques will help me design my research hypothesis and the methods I choose. The first section of the literature review discusses economic voting theory, which is the main driver for the analysis in this paper. I choose important dependent and independent variables for my research based on the theories discussed under this section. The second section compares macro-level analysis of economic voting with micro-level analysis. I choose the geographical unit of analysis for my research 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 me choose my control variables. The fourth section discusses various theories that could be attributed as factors contributing to the dominance of Republican party in 2016 election. The reason these theories deserve a separate section and are not included in the first section is to separate them from the traditional economic variables to create a premise for their exploration when analysing the 2016 election. The theories mentioned here come from academic as well as non-academic sources, since traditional models failed in forecasting the election in 2016. The fifth section compares the techniques used by authors in the past, the timeframe of their analysis and different peculiarities of their theories.

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. ? lists several authors, such as “Kramer(1971), Fair(1978), Tufte(1978), Rosenstone(1983), Hibbs(1987), Erikson(1989), and Holbrook(1991)”, who have studied and shown the connection between national macro-economic conditions and election results in the United States for decades now.¹ The macro-economic variables that are widely used to determine election outcomes are economic growth, disposable income, (un)employment, job growth,

¹The citations for the listed authors obtained from Ebeid and Rodden(2006) are as follows: Gerald H. Kramer, ‘Short-Term Fluctuations in U.S. Voting Behavior’, American Political Science Review, 65 (1971), 131-43; Ray Fair, ‘The Effect of Economic Events on Votes for President’, Review of Economics and Statistics, 60 (1978), 159-72; Edward Tufte, Political Control of the Economy (Princeton, N.J.: Princeton University Press, 1978); Steve J. Rosenstone, Forecasting Presidential Elections (New Haven, Conn.: Yale University Press, 1983); Douglas A. Hibbs, The American Political Economy (Cambridge, Mass.: Harvard University Press, 1987); Robert S. Erikson, ‘Economic Conditions and the Presidential Vote’, American Political Science Review, 83 (1989), 567-73; Thomas M. Holbrook, ‘Presidential Elections in Space and Time’, American Journal of Political Science, 35 (1991), 91-109.

economic volatility, inflation, etc. [?]. Even in the area of election forecasting, which deals rather with predicting the future than analyzing the past, fundamental statistical models are widely used to determine which party will win the elections. 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, ? lists the many economic variables used in forecasting using fundamental models as they include a range of measures such as: “GDP growth (Abramowitz 2004; Campbell 2004b); GNP growth (Lewis-Beck and Tien 2004); perception of personal finances (Holbrook 2004); prospective personal finances (Lockerbie 2004); leading economic indicators (Wlezien and Erikson 2004); income growth (*ibid.*); job growth (Lewis-Beck and Tien 2004)”². Moreover, the realm of economic voting has expanded rapidly over the years with new variables, definitions, and processes attached to it. While classical economic voting theories have viewed economy as a valence issue, ? examine positional and patrimonial economic voting. *in my own words* In their examination, they analyze the different preferences voters have on different economic policy issues and also the economic status of the voters [?]. Similarly, seeing the rise of globalization all over the world and its impact on the economy, ? 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 proponents of economic voting agree and disagree with each other on a few aspects. Some of the main questions that still float around in the area of voting behavior 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 voteshare of the incumbent or the total voteshare of a specific party or the two-party voteshare of a specific party. The idea of choosing the total voteshare of the incumbent is embedded in one of the assumptions of the economic voting theory that claims that voters reward or punish the incumbent party or its candidates. ? use Tuftes’s³ “election-as-a-referendum-on-the-government”

² List all the authors like how it is done in the first one.

³Tufte(1978) is cited under Lewis-Beck(2005) as: Tufte, E. (1978) Political Control of the Economy (Princeton: Princeton University Press).

idea to incorporate incumbency in their model to predict US elections. Moreover, ?'s “cost of ruling” designation to incumbency showcases the effects that incumbency can have when it comes to German elections. They both use incumbent voteshare as a dependent variable, and attribute this to the party's voteshare, without elaborating if the results would be different if the candidate is incumbent as well. While most studies place candidates and parties under the same umbrella, ? 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[?]. Another way of measuring the election result is by looking into 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 particular issue [?]. *give example* In this paper, 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 affects the way voters vote for the specific party. *write better here* Since, the purpose of the paper is to see from the perspective of not just a voter but a Republican voter, this route makes more sense for the analysis.

The second question by Abrams is already touched upon above by listing a plethora of explanatory economic variables mentioned by various others as indicative of election outcome. While all of them have their merits, depending on the nature of analysis, unemployment is the one that sticks out specifically in the context of this paper. ?, while describing the theory of issue-ownership, claims that the most central problem that voters keep in mind to make their election decisions is whether they, their families, friends, and 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.

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 Counties):

Study of voting behavior has usually consisted of analysis at the national level. Scholars have tried to connect national economic indicators to election outcomes over a period of years. Even individual-level

surveys such as the American National Election Survey (ANES) have been connected to national economic conditions to make generalized inferences about voting behavior. One of the pros of national-level analysis is that it makes the analysis easy to carry out by removing many nuances of geographical differences across different sub-national levels. For example, it is easy to attribute the change in national unemployment rate over the years to the victory of a specific political party. 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 subnational analysis at county-level although some level of analysis have been done on a state-level. O'Laughlin et al. (1994)⁴ and Owens and Wade (1988)⁵ provide some evidence of sub-national level economic voting scenario in Germany and the United Kingdom, while in terms of the United States Archer and Taylor (1981)⁶ 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 disaggregation hypothesis - which suggests the disaggregation of the economic-conditions variables by geographical factor - and claims that state-level economic conditions impact voting outcomes [Abrams, 1980]. He provides 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 W. Bush ?. While these studies do not make any claims on the county-level, they do provide the basis to further disaggregate the economic-conditions variable at a more lower geographical level and study the resulting impact. Following up on Abrams' work, ?'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 [?] since people's welfares are deeply intertwined in today's world of vibrant and complex economy. Besides, it is safe to assume that such self-interest is stronger at a smaller geographical concentration than at a national-level where connection between individuals is much more complex and much less tangible.

The voteshare of an incumbent presidential candidate in a state is regressed against the 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 [Abrams, 1980]. He claims that voters hold presidents accountable for changes in state-level economic conditions therefore cutting certain programs

⁴Cited in Kim(2003) as: O'Laughlin, J., Flint, C., Anselin, L., 1994. Annals of the Association of American Geographers 84, 351-380.

⁵Cited in Kim(2003) as: Owens, J.R., Wade, L.L., 1988. Economic conditions and constituency voting in Great Britain. Political Studies 35, 30-51.

⁶Cited in Kim(2003) as: Archer, J.C., Taylor, P.J., 1981. Section and Party. John Wiley and Sons Ltd, New York.

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 outcomes can outweigh the efforts of such state-level redistributive policy [Abrams, 1980]. This state-level analysis from Abrams also provides a strong theoretical foundation to carry a county-level analysis. ? 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, ? 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 using a more general national-level analysis. ? go a step further in their analysis of the effect 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. ? also use U.S. county-level measures of economic indicator, such as employment, in order to assess the effect of trade on presidential voting. A study by ? 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 paper as it aims to understand the effects of the regional economic condition on election outcomes in the past and more specifically on the recent presidential election of 2016. This paper follows the framework by ? who claim to present “the first county-level analysis of economic voting in presidential elections.” The difference in this paper, 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 Presidential election to explain the residual that is not covered by the first analysis. This type of two-faceted 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 main aim of the paper is to analyze the impact of economic variables on Republican voteshare, controlling for other political and demographic factors is extremely important in order 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 [?]. States are known for their identity as red, blue, or swing and that the political identity of majority of voters in those states usually stays the same over time. To analyze what affects Republican voteshare in a red county in Louisiana without taking into account the party identification or partisanship indicator of that county will be erroneous. While survey data, that asks prospective voters about their preferences, is useful in understanding the party affiliation of voters, in this paper we use the actual voteshare numbers, since the analysis is done on a county-level. Moreover, the voteshare numbers can be attributed to party affiliation as they give a real picture of voters' preference through actual casted votes. Based on how the result in a certain county turned out, we can get an idea about the actual preference of the majority of voters of that county or atleast understand what percent of people in a county voted a certain way and what that means.

The purpose of this paper is not to verify or repudiate the party identification theory but to simply control for it in order to achieve the best estimates of the effects of economic indicator. The operationalization of party ID as a control can be done in several ways but the one that stands out is taking into account the voteshare of the selected party in the previous election. Previous vote share for the incumbent party is introduced as an independent variable by Abrams [1980] in his model as well. *find some sources* In one example, ? 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. *check how they measure partisanship not so strong this section*

Although, this paper will analyze economic voting from the persepective of the Republican party and not from the perspective of an incumbent party, it is extremely important to control for incumbency effect. Many political scientists have agreed to the idea of analyzing the economic vote through the perspective of an incumbent. The main idea behind this is that voters assess the performance of the incumbent party or candidate and use their vote as a means to determine either to reelect them or to choose the opposition candidate or party. Key's (1970)⁷ 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 [?]. In economic voting, incumbents are given utmost importance because people often relate economic performance with the incumbent party or candidate. Hence, if

⁷Cited in Kim(2003) as: Key, V.O., Munger, F., 1970. Social Determinism and electoral Decision: The Case of Indiana. In: Crotty, W.J. (Ed.), Public Opinion and Politics. Holt, Rinehart and Winston, New York.

the economy is doing well, it is believed to help the incumbent party or candidate and if it is doing bad it is believed to hurt them. With an incumbent dummy we can control for the effect of the incumbency on the voteshare of a specific party. Moreover, at a county-level analysis, where most variables are at county-level, the incumbency variable serves as a representative of the national effect on the county, since the incumbent candidate or party is the same nationally for all counties.

Demographic controls are also an important aspect of economic voting model, as they are known to significantly impact the voteshare of a party at the presidential election. Including them into the model will reduce the possibility of omitted variable bias and give more precise estimations. Race, age, education, religion, gender are some of the main demographic variables often used by authors. Republican voters are often characterized as white *source* and following ?'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 more people with fewer years of formal education can also be controlled for the understanding the republican voteshare. 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 voteshare 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 the one gender on the voteshare unless an individual survey-type analysis is done. *cite more* 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 subnational level analysis, this measure of rurality captures a lot of voting characteristics of a county, which this paper aims to conduct. *cite about rurality*

Some Theories for 2016 US Presidential election outcome:

The United States Presidential Election of 2016 has stood out for many reasons- the most significant of them being the failure of scientific forecasts 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?" This is a question which still hasn't been convincingly answered even months after the election. This paper does not claim to answer it either, but does aim to incorporate some new voting theories that have been floating around 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 FBI Director, 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 limitation on the scope of this paper, which focuses on county-level analysis and hence requires county-level data to do so, I will touch upon only a few of these theories and test them using a statistical model. Through elimination of one or two of these theories through an analysis at a subnational level, we could get one step closer to finding out what may have actually happened.

The first theory that I aim to test is what I will call the “employment theory”⁸, 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.[?] 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 [?]. In this paper, 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 at 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 time. *cite this* 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 the population that is either employed or unemployed (that is, either working or actively seeking work).”⁹ 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 fact that the unemployment rate between July and August didn’t change much despite a decline in the labor force participation by about 600,000 people [?]. This led me to include labor force participation rate as one of the variables to describe republican voteshare in the 2016 election. I base this on the recommendation made by [?] who suggest using other labor market statistics before drawing any inferences about the labor market conditions- something that plays an extremely important role in voting decisions as seen through literature discussed previously. Back in

⁸The nomenclature is given by the author for easiness to refer to the theory again throughout the paper. This does not mean that the author claims to have proposed the theory. The theory already exists and hence it is discussed in the Literature Review section in the first place.

⁹Definition obtained from the Bureau of Labor Statistics page: https://www.bls.gov/bls/cps_fact_sheets/lfp_mock.htm.

2010 they claimed that “there are important, permanent, discouraged-worker effects among males in the US economy” and suggested that measures such as employment rate or some other measure includes groups suspected of being discouraged workers in addition to the ones unemployed [?]. Following up on the work of Gustavsson and Österholm, ? 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. *check if graphs can be taken from Bullard2014*

The second theory that I aim to test for the 2016 election result is what I will call the “wage theory”.¹⁰ ? use average pay of the workers during the election year as an important regressor for the incumbent voteshare. More importantly they include it in their county-level analysis. Voters are expected to reward the incumbent in the event of higher average wage. Following up on the same idea, I aim to include the change in average pay between the 2016 and the 2012 election. The change seems more logical than the absolute value as it can capture the difference in economic condition between two election years and hence help us explain the change in Republican voteshare. 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.¹¹ *why does he use natural log*

The third theory is the “inequality theory,” which can help understand if the growing inequality in the US has any explanatory power when it comes to voting behavior. Afterall, Bernie Sanders’ whole campaign was based on reducing the inequality in the US and expanding the middle class, 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 inequality. ¹² Schneider-Petsinger (2016) ¹³ 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. Darvas and Efstathiou carry 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.¹⁴ Their analysis and results provided with the necessary motivation to test the findings at a county-level.

The fourth and one of the most important theories is non-economic in nature. I call this the “education

¹⁰The nomenclature is given by the author for easiness to refer to the theory again throughout the paper. This does not mean that the author claims to have proposed the theory. The theory already exists and hence it is discussed in the Literature Review section in the first place.

¹¹Source: <http://money.cnn.com/2016/12/22/news/economy/us-inequality-worse/>

¹²Source: <http://www.independent.co.uk/news/business/news/wealth-gap-global-income-inequality-brexit-donald-trump-victory-threat-a7521.html>

¹³Source: http://www.realclearworld.com/articles/2016/10/18/inequalityDefines_the_american_election_112094.html

¹⁴Source: <http://bruegel.org/2016/11/income-inequality-boosted-trump-vote/>

theory”¹⁵ and it is claims that the election outcome was influenced by difference in education level among the voters. The education theory indirectly also validates the idea that people who were less educated were more vulnerable to fake news, false economic reports, closed to diversity, 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 or “correctness of vote” is depended on their education level. However, the point of this paper is not to do so. 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.¹⁶ Hendrickson and Galston 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 in itself.¹⁷

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 the economic indicators to make decisions on their vote. Many others have answered this question by testing economic voting theory with available economic data (at that period of time). *cite* Many have included unconventional measures such as the indicators of the perception of the economy instead of the indicators of the actual economy. *cite* With the growing influence of media on voters and information available 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 data. *? claim* that relevant academic literature 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. *? 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, past papers lack such privilege. *This is evident in the fact that XXXX in year XXXX uses the absolute unemployment level while YYYY in year YYYY uses something else, to do the exact same type of analysis.* Another factor that is worth

¹⁵The nomenclature is given by the author for easiness to refer to the theory again throughout the paper. This does not mean that the author claims to have proposed the theory. The theory already exists and hence it is discussed in the Literature Review section in the first place.

¹⁶Source: <http://fivethirtyeight.com/features/education-not-income-predicted-who-would-vote-for-trump/>

¹⁷Source: <https://www.brookings.edu/blog/fixgov/2016/11/18/educational-rift-in-2016-election/>

noting is that due to the growing influence of partisan media on voters, there is a possibility that voters may not even base their preference based on actual economic performance but rather on their perception or forced perception of the economy. cite something from the article about fox news effect. Candidates on both sides use the media to put strong emphasis on campaigns, advertisements, conventions, debates, 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 ? 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 [?]. While analyzing whether people vote correctly, ? 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. They also conclude, through their analysis of US presidential elections, that most of time citizens do vote correctly. Combining Gelman's analysis with Lau and Redlawsk's gives a premise for this paper to analyze the US Presidential elections through the angle of economic voting.

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. When analyzing cross-sectional data, pooled ordinary least squares measures are used but if the data is panel (cross sections over different time periods), one can choose between pooled ordinary least squared method, fixed effects or random effects method, depending on the specification. *cite what they have used, FE, OLS, logit* This paper combines three ideas- the first idea is of looking at past elections from 1992 to 2012 at county-level to see how economic indicators behave over that span of years. The use of the appropriate estimation method is explained in the methodology section. 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. *cite all fundamental models papers from Simon's class* The third idea is of looking at 2016 Presidential election specifically to see how much this election varies from the first model. Using the residual of the predicted and actual voteshare for 2016, I will perform another regression analysis on this residual (difference in actual and predicted vote share) with economic and other non-economic theories described in the section above. This paper 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 voteshare with voteshare of past Republican candidates. Hence, the Trump effect.

Methodology:

This paper analyzes the research questions using quantitative regression 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 county-level economic variables 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 voteshare of 2016 that is not explained by the model in the first part. These two researches are carried out in three steps as illustrated in the diagram below:

Diagram

Step I:

Model Specification:

research question should be mentioned in the intro as well already.

Research Question: What are the effects of county-level economic variables on the two-party vote share of the Republican party in the Presidential election?

Hypothesis: A rise 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-ownership theory.

Regression Equation:

$$rep.share_{i,t} = \alpha + \beta_1(unemp_gro_{i,t}) + \beta_3(controls) + \epsilon$$

Dependent Variable:

rep.share_{i,t}: The dependent variable is the two-party vote share of the Republican party at the United States Presidential Election for county (i) in election year (t). Considering the effect of the third party in the United States Presidential election to be approximately equal on both the Democratic and the

Republican party and insignificant in relation to the voteshare of these two major parties *cite someone who has given this logic, Eisenberg and Ketcham 2004*, I have decided to eliminate the third-party vote share. Many others such as *XYZ* and *XYZ* have done the same thing in their analysis. Hence, the republican voteshare 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 voteshare of the incumbent party however, in this case, the research question aims to explore the effect of economic variables on the Republican party specifically. Hence, the usage of republican vote share becomes logical.

Independent Variable:

unemp_gro_{i,t}: The first independent variable, which is also the explanatory economic variable, is the growth in unemployment rate for county (i) between election year (t) and the preceeding 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 $(3.8 - 5.7) / 5.7 = -0.33$. ? in their county-level analysis of voting behavior claim that in order 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. *cite someone who has used this* In the United States, many states and counties within them are known to be historically *another term: perennially* 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 ommitted variable bias, result 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.

Pop_thou_{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. *explain why*

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. *why white*

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 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 any Fixed Effect model, 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 Figure X 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 specific 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). *some theory behind Fixed Effects*

Step II:

In the second step, I carry out a forecasting exercise by predicting the Republican voteshare for 2016 using the regression equation and coefficient estimates obtained from Step I. Because of the fixed effects model, the coefficient estimates remain the same for all the counties but each county will have a different

intercept. The purpose of the forecasting exercise is to analyze how well the model explains the election outcome of 2016. Once the forecasting is done, I plot the predicted voteshare of the Republican party against their actual voteshare 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.

? 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 lays the perfect premise for me to introduce some new rival theories concerning 2016 election (which I mention in the Literature review section above), operationalize them, and use them as explanatory variables in my second 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. I choose the variables keeping that in mind, although I am also tempted 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, I have followed ?'s suggestion who assert that,

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:

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 voteshare for the Republican party from the predicted voteshare based on a traditional model?

Hypothesis: Keeping in mind that Donald Trump's victory was unprecedented and he was touted as a non-regular candidate (i.e. different than a regular Republican), the effects of county-level variables should be different in 2016 election than the previous election. The additional variables that may have had an effect on 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 “Making America Great Again” and his promises to raise the lives of the poor who had suffered from the wraths of globalization and the neglect from Washington.

Research Equation:

$$resid = \alpha + \beta_1(manu_share_gro) + \beta_2(av_wage_gro) + \beta_3(lfpr_male_gro) + \beta_4(gini_gro) + \beta_5(uneduc) + \epsilon$$

Dependent Variable:

resid: The dependent variable is the difference in the actual two-party vote share of the Republican party in the 2016 United States Presidential Election from the predicted vote share based on the model in Step I of this paper. Substituting the values of different explanatory variables in equation 1, pertaining to election year 2016, we will get the value of the predicted republican vote share for each county for 2016. For example, if the predicted Republican voteshare for a county is 0.55 but the actual Republican voteshare 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 voteshare is not explained by the model represented by equation 1 and therefore, we would need to test a bunch of new explanatory variables, which could possibly explain this residual. A positive residual means that the Republican party outperfomed our forecasting model in those counties, a negtive residual means that the Republican party underperfomed, a value of 0 means that our forecasting model was able to accurately predict the actual voteshare of the party for 2016. The difference in using residual instead of simply voteshare 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 the ratio of manufacturing jobs to total private non-farm jobs between 2012 and 2015. An ideal measurement would have been to see the change 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. *show using evidence from data*

av_wage_gro: The second independent variable is the change in average wage between 2012 and 2015, measured in *dollars*. A positive value means that the average wage has increased in the county while a negative represents 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, I choose the change in labor force participation rate among men in county during the time frame ?. A positive value means that more people are getting back in the labor force while a negative value means that more are being 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 change between 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 lack thereof). This variable is calculated by dividing the total number of people above 25 years of age with less than high school degree (i.e. less than 12 years of formal education) by the total population of that county in 2015. *see if the same justification is used elsewhere justify what a Trump voter is expected to be and hence control for that check if it is 25 years and above*

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 effect of the explanatory variables on the dependent variable for each specific county between 2012 and 2015. Like mentioned earlier, I do not include the same variables already used in equation 1 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). More importantly, I also remove the intercept values for each county since *explain in better words*.

Data sources and cleaning:

The data necessary to carry out the quantitative analysis in this paper 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 paper is also written on R markdown. 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.¹⁸ To calculate the two-party vote share, the voteshare for Republican party was divided by the sum of vote shares for Republican and Democratic party. The lag of republican vote share used in the first part of the research was calculated based on the same election dataset. The data on unemployment from 1992 to 2015 was obtained from the Bureau of Labor Statistics (BLS).¹⁹ To calculate the growth in unemployment rate between two years, the unemployment rate for previous year was subtracted from the unemployment rate of the current year, and then the result was divided by the unemployment rate of the previous year. 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.²⁰ Data used to construct the incumbency dummy was created manually based on common knowledge and merged into the final dataframe. *Put table on Appendix with president's name and years and dummy* Data on rural population was obtained from the United States Census Bureau.²¹ 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.²² The race data that included white population in each county between 1992 to 2015 was obtained from the database of the National Cancer Institute.²³ The counts of race and education data were converted into percentages by combining them with the population data obtained from BEA.

¹⁸Dave 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 subnational level.

¹⁹The 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>).

²⁰The 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>).

²¹The 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).

²²The 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).

²³The 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>).

The data on Labor Force Participation rate²⁴ and Gini coefficient²⁵ for 2012 and 2015 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.²⁶

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. These processes resulted in a total of 3112 counties. The final merged panel dataset, which was used for the first part of the research, hence contained 3112 counties x 6 election years = 18672 observations. For the second part of the paper, the final cross-sectional dataframe consists of 3112 counties.

A list of all the variables and other specific details on them are available in *table X* under the Appendix section.

²⁴The 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)

²⁵The 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)

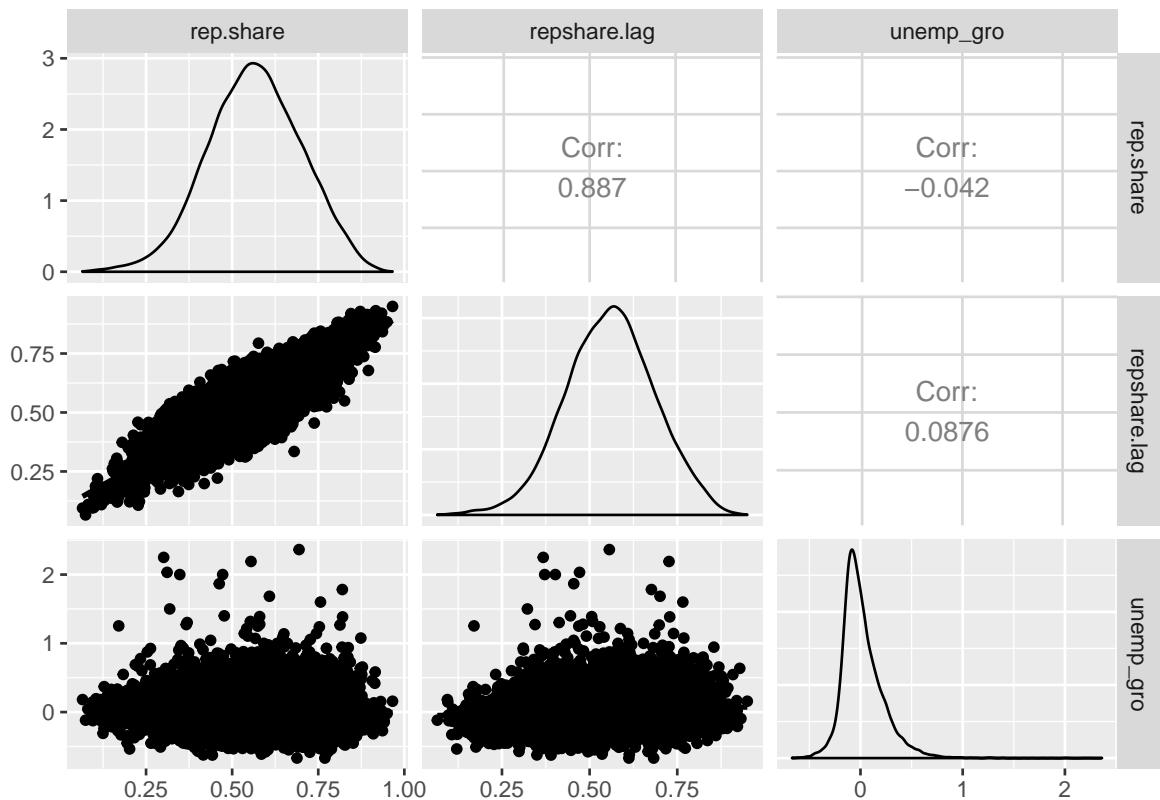
²⁶The data on manufacturing jobs for each county is obtained from the Bureau of Economic Affairs' API call for "RegionalIncome dataset, table CA25N, Linecode 500." The data on total private non-farm employment for each county is obtained from the same API call for "RegionalIncome dataset, table CA25N, 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:

First Part: Creating and analyzing a model to explain the Republican two-party vote share between 1992 and 2012

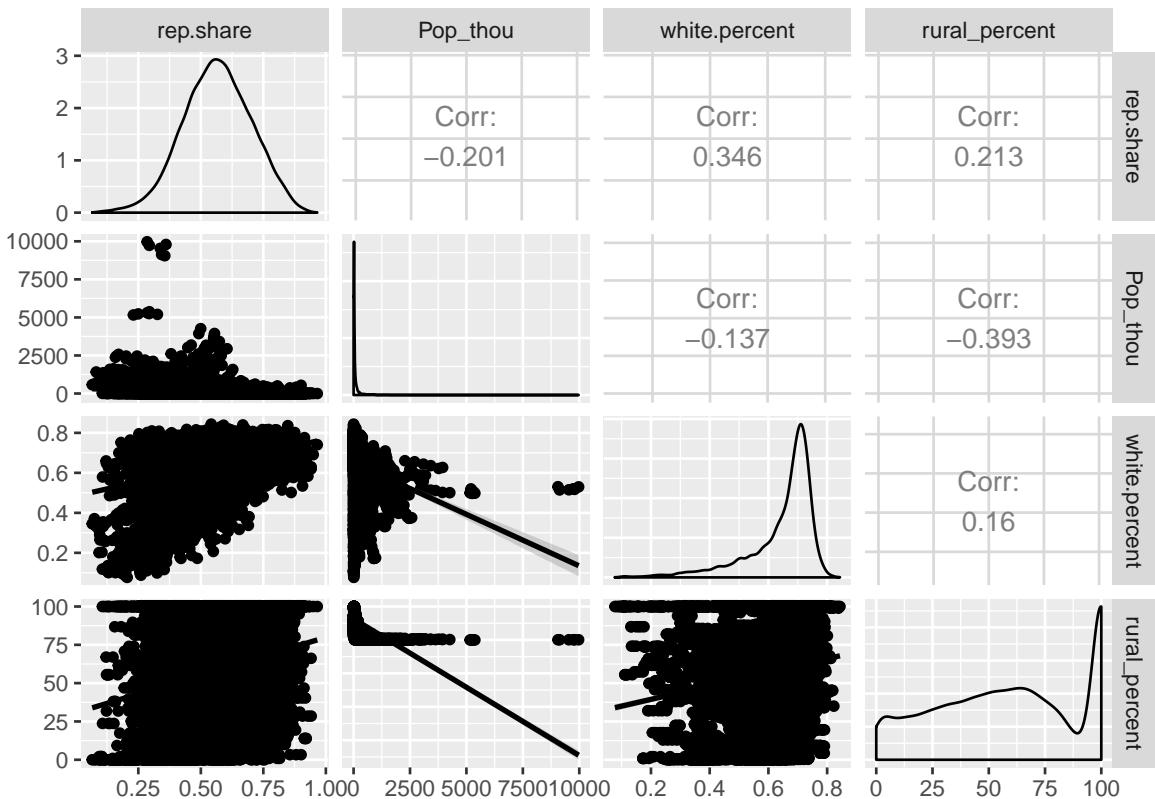
Descriptive Statistics:

FigX shows the how the main independent variable, the lag dependent variable and the dependent variable are correlated to one another, and therefore provide some basic understanding on how their causal relationship could turn out to be. We can see that growth in unemployment is negatively, yet not so strongly, correlated ($r = -0.042$) with the Republican two-party vote share. More importantly the Republican vote share in the previous election is positively, and strongly, correlated ($r = 0.887$) with the present voteshare. Both of these correlations are in line with the theory established earlier in the paper. The distribution plots of each of these variable 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, *we would not want to remove it, as it may remove important counties whose unemployment situation may be pertinent in our analysis.*



FigX shows the correlation between the dependent variable with other control variables. Republican two-party voteshare 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 are considered to be homogenous while the urban areas more heterogenous. It is important to note that counties with larger population have a negative correlation with Republican vote share($r = -0.201$) and also show that these counties are less rural ($r = -0.393$) and have lower white percentage($r = -0.137$).

The distribution plot shows an extreme right skewness with population due to the presence of some extremely populous county like the Los Angeles county in California that has a population of more than 9 million. *how to correct remove or log or categorical variable* Similarly, the plot of rural_percent shows *a weird shape why explain changing them into categorical will take away the fixed effects*



The descriptive summary table (see *figX* under Appendix) gives a complete overview of the central tendency and the variation of all the variables.

Regression Output and Results:

FigX shows the regression output with Fixed Effects estimate, showing the relationship between unemployment growth and other political and demographic controls for US Presidential elections between 1992 and 2012.

Table 1: Fixed Effects Estimate Model

	<i>Dependent variable:</i>			
	rep.share			
	<i>panel linear</i>		<i>coefficient test</i>	
	(1)	(2)	(3)	(4)
unemp_gro	-0.08*** (0.002)	-0.02*** (0.003)	-0.08*** (0.003)	-0.02*** (0.004)
repshare.lag	0.62*** (0.01)	0.73*** (0.01)	0.62*** (0.01)	0.73*** (0.01)
Pop_thou		0.0001*** (0.0000)		0.0001*** (0.0000)
white.percent		0.19*** (0.05)		0.19*** (0.04)
as.factor(rep_incumb)1		-0.04*** (0.001)		-0.04*** (0.001)
unemp_gro:as.factor(rep_incumb)1		-0.04*** (0.004)		-0.04*** (0.005)
white.percent:as.factor(rural)1		0.52*** (0.06)		0.52*** (0.05)
Observations	18,350	18,322		
R ²	0.36	0.46		
Adjusted R ²	0.24	0.35		

Note:

*p<0.1; **p<0.05; ***p<0.01

Due to the inclusion of a lag dependent variable *repshare.lag*, an Arellano-Bond test is conducted to account for nickell bias. Therefore, the estimates in the third column, with corrected standard errors, are used for interpretation. Growth in unemployment has a negative effect on Republican two-party vote share with stastical significance at 99 percent. One percent growth in unemployment rate between the year preceeding the election and the election year will result in a fall of Republican two-party vote share by 0.02 percentage points. This result is in line with the issue-priority theory, which says that

higher unemployment 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. While the increase in the percentage of white people in a county increases the republican vote share, the effect is even bigger if we look at rural white population.

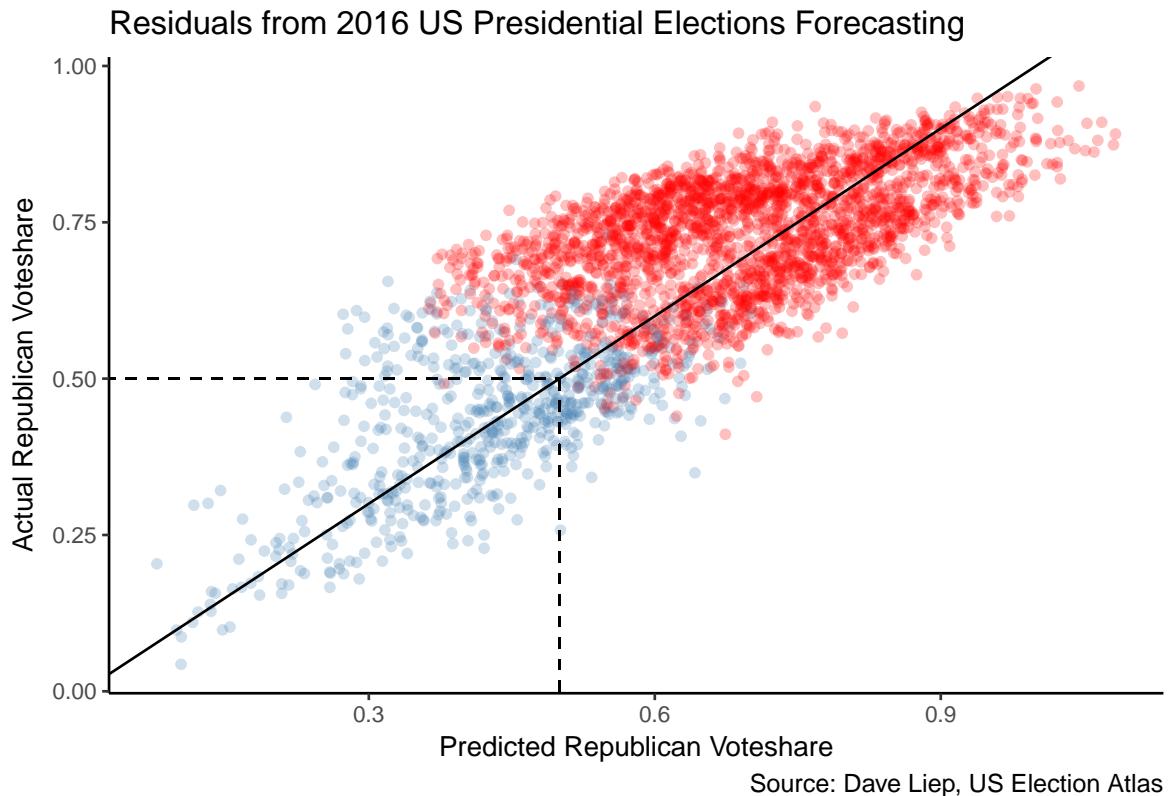
explain marginal effect Validating Norpoth's cost of ruling theory, the incumbency factor shows that if the incumbent is a Republican, then the Republican voteshare goes down, suggesting that voters do not like to keep the same party in power for many terms *source*. Moreover, the effect of unemployment growth on the Republican vote share is even higher, meaning the negative impact of unemployment growth is exacerbated, when the Republican party is incumbent. This result is in line with the theory that voters punish the incumbent for difficult economic condition.

In this case, they do so, even though they believe that the Republican party is less suited to solve the issue of unemployment. *explain marginal effect* 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 it’s marginal effect. The model explains 46 percent of the variation in the Republican voteshare 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.

Second Part: Using the model to make a forecast for 2016 Presidential Election

From the model above, we have unique coefficient estimates for each variable and a unique intercept for each county. By substituting the values of these variables for 2015 in the equation, we can forecast the republican two-party vote share for each county in the 2016 election. The graph below shows the accuracy of the prediction as we compare the predicted Republican vote share with actual vote share obtained during the election. *explain how far the model is* The counties that fall on the 45-degree line represent the ones where the model was able to accurately predict the actual voteshare. As we can see there are many counties (above and below the 45 degree line), where there are residuals, i.e. differences in the predicted and actual vote share. The Democratic counties where the Republican voteshare is below 50 percent, the actual voteshare was much less than what our model predicted to be. However, in the Republican counties, the actual voteshare is much higher than what our model predicted showing a big cluster of counties above the 45 degree line. What caused the Republican-leaning counties to vote significantly higher for Donald Trump and what led the Democratic -leaning counties to vote lower for

him than what the traditional model would expect to is a question that needs more exploration. In order to understand this difference (residual), we will need to test various rival hypothesis with the aim of explaining the effect that Donald Trump had that other traditional generic Republicans didn't have in the past.



Third Part: Creating and analyzing a model to explain the residual obtained from the forecasting exercise i.e. “Trump effect”

Descriptive Statistics:

The 2016 Presidential election took everyone by surprise and to explain it has been an arduous task to say the least. Since, we have established that a traditional model doesn't suffice when it comes to exploring the nuances of this year's election, it is important to resort to some rival theories and test their relevance. In the table below, we see the correlation matrix of some of the rival variables and the residual obtained from the forecasting exercise to create a premise for further exploration.

```
#> ## Warning: Removed 14 rows containing non-finite values (stat_density).
#> ## Warning in function (data, mapping, alignPercent = 0.6, method =

```

```

## "pearson", : Removed 431 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 20 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 14 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 14 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 14 rows containing missing values

## Warning: Removed 431 rows containing non-finite values (stat_smooth).

## Warning: Removed 431 rows containing missing values (geom_point).

## Warning: Removed 419 rows containing non-finite values (stat_density).

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 419 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 419 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 419 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 419 rows containing missing values

## Warning: Removed 20 rows containing non-finite values (stat_smooth).

## Warning: Removed 20 rows containing missing values (geom_point).

## Warning: Removed 419 rows containing non-finite values (stat_smooth).

## Warning: Removed 419 rows containing missing values (geom_point).

## Warning: Removed 6 rows containing non-finite values (stat_density).

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 6 rows containing missing values

## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 6 rows containing missing values

```

```
## Warning in (function (data, mapping, alignPercent = 0.6, method =
## "pearson", : Removed 6 rows containing missing values

## Warning: Removed 14 rows containing non-finite values (stat_smooth).

## Warning: Removed 14 rows containing missing values (geom_point).

## Warning: Removed 419 rows containing non-finite values (stat_smooth).

## Warning: Removed 419 rows containing missing values (geom_point).

## Warning: Removed 6 rows containing non-finite values (stat_smooth).

## Warning: Removed 6 rows containing missing values (geom_point).

## Warning: Removed 14 rows containing non-finite values (stat_smooth).

## Warning: Removed 14 rows containing missing values (geom_point).

## Warning: Removed 419 rows containing non-finite values (stat_smooth).

## Warning: Removed 419 rows containing missing values (geom_point).

## Warning: Removed 6 rows containing non-finite values (stat_smooth).

## Warning: Removed 6 rows containing missing values (geom_point).

## Warning: Removed 14 rows containing non-finite values (stat_smooth).

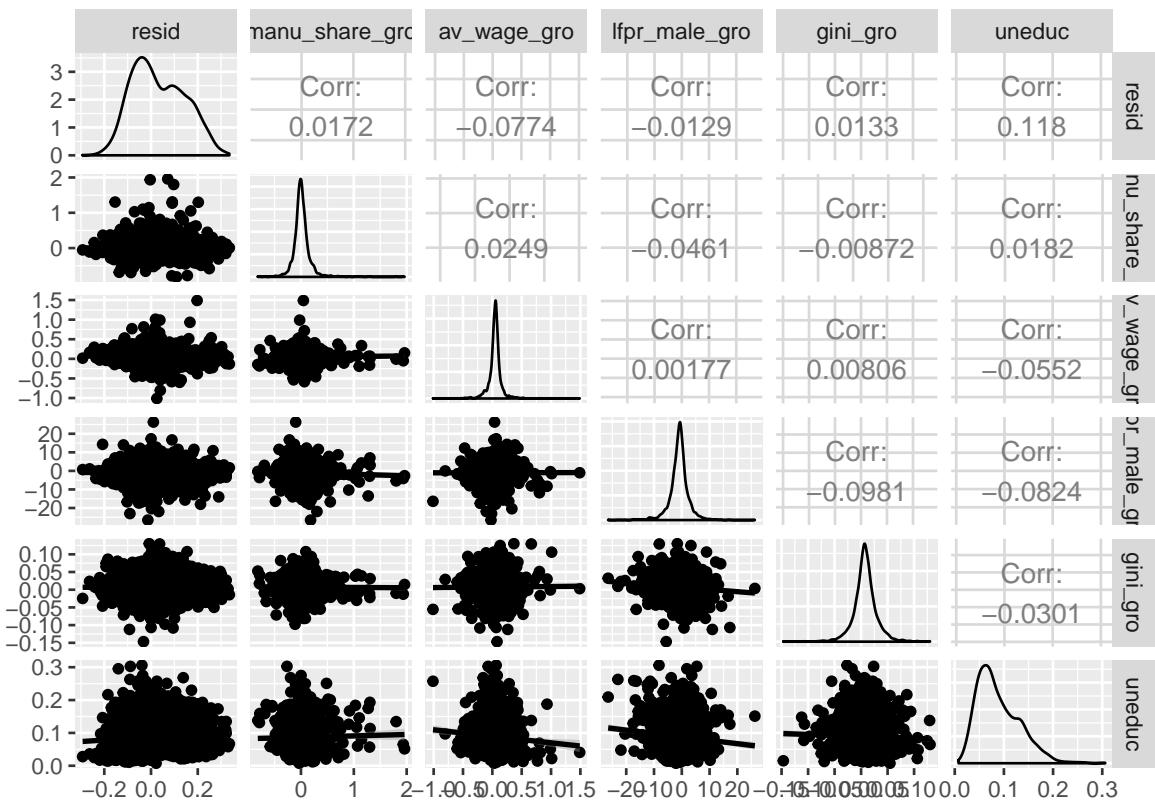
## Warning: Removed 14 rows containing missing values (geom_point).

## Warning: Removed 419 rows containing non-finite values (stat_smooth).

## Warning: Removed 419 rows containing missing values (geom_point).

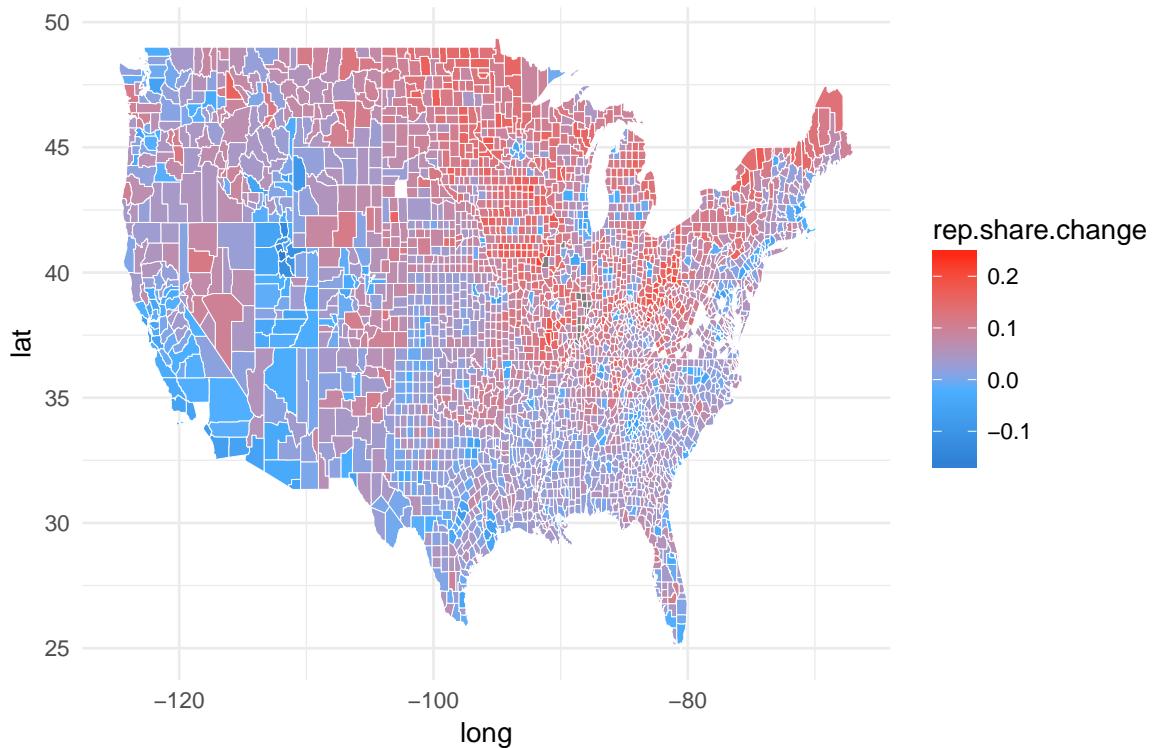
## Warning: Removed 6 rows containing non-finite values (stat_smooth).

## Warning: Removed 6 rows containing missing values (geom_point).
```

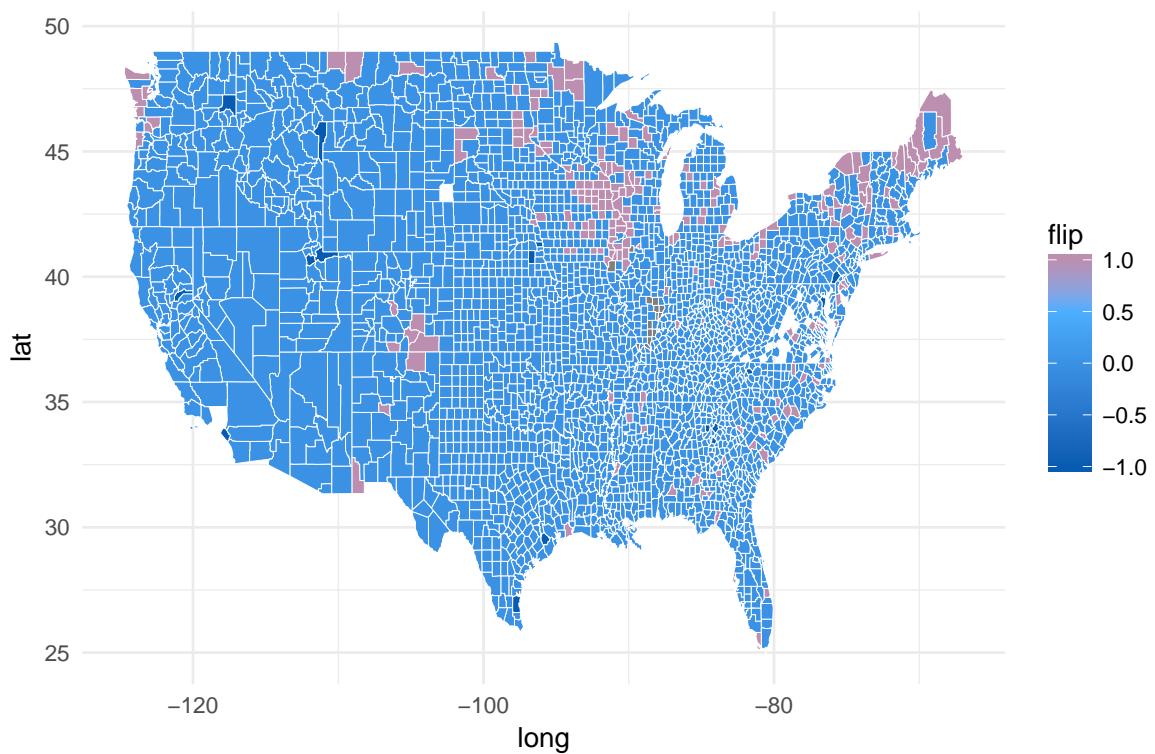


US situation in these variables

Change in voteshare for Republican party 2012 to 2016



Counties that flipped from Democrat to Republican 2012 to 2016



Regression Output and Results:

The regression output below shows the effect various factors on all the counties, on counties in swing states, and on counties in rust-belt states respectively. 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. Here is the table showing the swing states how they acted in the past few elections and same goes for rust-belt states. *make a table and put it*

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

Dependent variable: resid			
	(1)	(2)	(3)
manu_share_gro	0.01 (0.01)	-0.02 (0.04)	0.03 (0.03)
av_wage_gro	-0.09*** (0.02)	0.03 (0.06)	-0.09 (0.06)
lfpr_male_gro	-0.0000 (0.001)	-0.0005 (0.002)	-0.004* (0.002)
gini_gro	0.10 (0.12)	0.13 (0.29)	0.11 (0.33)
uneduc	0.42*** (0.03)	0.50*** (0.07)	1.10*** (0.08)
Observations	2,627	608	668
R ²	0.10	0.13	0.31
Adjusted R ²	0.10	0.12	0.31
Residual Std. Error	0.12 (df = 2622)	0.12 (df = 603)	0.12 (df = 663)
F Statistic	60.14*** (df = 5; 2622)	17.68*** (df = 5; 603)	60.15*** (df = 5; 663)

Note:

*p<0.1; **p<0.05; ***p<0.01

The first hypothesis that change is employment scenario had a role to play in increasing votes for Trump can be rejected when we look at the whole county's example. Neither change in manufacturing jobs nor the labor force participation rate among male 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 of labor-intensive and manufacturing-based works have been accepted by the workforce. Inclusion of women in the workforce and their contribution to the economy is also 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 labor force participation

rate among males to be moderately significant (90 percent) in explaining the rise in Republican vote share in those counties. The negative effect of the change in labor force participation rate among male on Republican vote share suggests that the rise in Republican vote share in these rust-belt counties were due to a decline in participation by men. 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 some what 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 resentful 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. *why*

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 into non-voters. This assumption would require testing which is not carried out in this paper.

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 this year's election. A one percent increase in the number of uneducated people shows almost a one-third 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.

show the result in maps next to each other with heat map for the variables that are significant

Put the hand made table from the Lit. review note book

Conclusion:

Limitations and Further Research:

- Can these economic conditions be found on individual-level survey data for 2016 to substantiate or repudiate the claims made above? *depending on time, we can skip this and put it under further research possibilities at the end*

Appendix:

(Test results of all the tests, graphs, tables, plots, etc. that don't go in the main body of the paper.)

Rough Notes:

a dummy for Southern States was taken [Abrams, 1980].

Three main inferences made by Bartels[Bartels, 2005] are: - The white working class has not abandoned the Democratic Party. - The white working class has not become more conservative. - Working class moral values do not trump economics.

Bartels [Bartels, 2005] uses family income levels to categorize voters in terms of their economic status. He places families with incomes in the bottom third of the income distribution under “low income” or “working class” thus differentiating them from the middle and upper class families. *how do we operationalize this in terms of county-level economic data.*

Bartels [Bartels, 2005] illustrates the Democratic share of the two-party presidential vote among white voters in the bottom third of the income distribution and in the top third of the income distribution. He attributes the loss of support from white voters to Democratic candidates to the middle and upper income groups, while claiming that support for Democratic candidates has increased among low-income white voters.

good point below Bartels [Bartels, 2005] argues that “Democratic identification declined by 18% among low-income whites (from 22% in 1952 to 4% in 2004) and by 29% among high-income whites (from 11% to 18%).” But he mentions that “However, it seems odd to attribute the Democrats’ problems to the white working class when the corresponding decline among more affluent whites is so much larger.”

Although one individual has one vote, the value of that vote may be different. *can we see if the 18 percent decline among low-income whites (and seeing where they are concentrated) has more impact on the electoral outcome than the 29 percent decline among high-income whites who may be concentrated in bigger cities and urban areas and democratic states.*

Bibliography:

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

Larry M Bartels. What's the matter with what's the matter with kansas ? 2005.