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

Acknowledgement:

write once everything is written

Abstract:

write once everything is written

"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 incumbent vote-share? 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 compares the techniques used by authors in the past, and the timeframe of their analysis.

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 proven 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, 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] The 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.

[2] _List all the authors like how it is done in the first one.

> (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 sets 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's[3] "election-as-a-referendum-on-the-government" 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, without elaborating if the resuts 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 Unites 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 and per capita income are the ones that stick out specifically in the context of this paper. *more on justifying unemployment per capita income rep vote share*

The third question by Abrams 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 generalize 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

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

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

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

voting scenario in Germany and the United Kingdom, while in terms of the United States Archer and Taylor (1981)[6] has 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 - and claims that state-level economic conditions impact voting outcomes (Abrams, 1980). *describe in own words what disaggregation means* (**???**) 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 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 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. *describe in own words* Moreover, what is considered as the self-interest of an individual can often encompass the welfare of others (**???**). 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.

(**???**) 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 crevaces 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 election results from the past (1992 to 2012) and then carry a separate analysis with the same variables for years 2012 and 2016 to capture the highly touted pecualiarity of the 2016 election. This type of two-faceted county-level analysis will first test the hypothesis with the assumption that the Republican candidate is a generic Republican and second test the hypothesis if Trump is a generic Republican based on how the 2016 model compares to the model from previous years.

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 in strongly dominated states and moving them to closely contested states may improve the chances of reelection for the

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

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). *can this model be applied at a county level, since it is simply a smaller geographical setting and more representative since it is closer to the individual level change position of this para in this section*

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 to be red, blue, or swing for a specific reason - 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 is useful in forecasting, the actual data we use can give a better picture of voter behavior since it captures the reality of after the election has taken place. However, based on the actual data, we are not able to measure party identification, since there is no way to know the individual preferences of voters. Nevertheless, 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.

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 indicators. 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*

many say lagged DV is not good: shit!!! (**???**) suggests that a lagged dependent variable would not be appropriate in this case since the relationship between the economy and the vote for president is not theorized as dynamic. He writes, "the data points are four years apart so it is unlikely that the performance of the economy at time t would be strong enough to significantly influence the vote for president four years in the future at time t + 1 or eight years in the future at time t + 2" (**???**). *how to counter this*

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 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. Meaning, it is not a variable unique to each county and in simple terms, the variable represents the existing national government and not the local government. *find some sources*

[7]Cited 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.

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 Republican voteshare. Gender, although being a strong control variable at national-level analysis, has sound theoretical argument to not be included in a county-level analysis. In a subnational 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 diffifcult to assess the effect the one gender on the voteshare unless a 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*

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 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 through different campaign events 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. *cite what they have used, FE, OLS, logit* This paper combines two ideas- the first idea is of looking at past elections from 1992 to 2012 to see how economic indicators behave over that span of years, like how does *cite* with his analysis from *xxx to xxx* and *cite* does his from *xxx to xxx*. The second idea is of looking at 2016 election specifically to see if this election varies or holds true to the analysis obtained from the first analysis. *cite* perform such one-year analysis of economic voting. However, this paper is unique in its approach of combining the two studies with a clear research agenda of comparing the generic republican voteshare from Donald Trump's voteshare. *Why start at 1992*

why separate study for 2016. forecasts failed, etc. what was the turnout find out

Background:

US Election Peculiarities:

electoral college swing states term limits media debates difficult to incorporate all of them, but in this paper, we will take on a few of them such as swing and non-swing states and other demographic variables that we will discuss later.

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 perfomed robustness tests, which are included in the Appendix.

The research question is divided into two sections: the first part studies the effect of county-level economic variables on Republican vote share from 1992 to 2012, and the second part studies the effect of county-level economic variables on the change in Republican voteshare from 2012 to 2016.

Part 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?

Hyopthesis: Depending on the type of economic variable, its effect on the Presidential two-party vote share of the Republican party can be different. 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 caused by the rise in unemployment. On the other hand, a rise in Per capita income should have a positive impact on Republican vote share because of the economic benefits through reduced taxation for rich that the party is known for *rephrase in better way*

Regression Equation:

$$rep.share_{i,t} = \alpha + \beta_1(unemp\_gro_{i,t}) + \beta_2(pci\_gro_{i,t}) + \beta_3(controls) + \epsilon$$

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 U.S. 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. Hence, the republican voteshare shown here is the republican proportion of the "two-party vote share." *For example from the Dataset* If, in a county, Republicans obtained 55 percent of the votes, Democrats obtained 40 percent, and the third party obtained 5 percent, then the value of *rep.share* will be 0.578 instead of 0.55. 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 Variables:

$unemp\_gro_{i,t}$: 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 year before (t-1). *For example* and *Justify the usage of growth instead of absolute value* and *justify the usage of the value from year before* and *cite*

$pci\_gro_{i,t}$: The second independent variable, is the growth in per capita income for county (i) between election year (t) and the year before (t-1). It is calculated using the same method as the growth in unemployment rate. *justify the usage* and *cite*

collinearity between these two variables and *justify the usage of both two together* and *list some other ones that they replace, for example unemployment substitutes output growth, etc.*

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).

$Pop_{i,t}$: 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. *explain why this is important*

$white.percent_{i,t}$: 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 in a county by the total population of that county. *why white is used instead of black was any age group chosen since the data is on all age groups*

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 measures the effect of the national political phenomemon at a subnational level. *for example* and *cite*

$rural_i$: A dummy variable for rural is introduced as another demographic control variable. The value of *rural* is 1 if the county is rural and 0 if the county is urban. This designation is based on a 2010 US Census designation that places counties with 50 percent or more people living in rural areas as a rural country and less than 50 percent of people living in rural areas as a urban county. This variable is fixed for a given county across all years.

collinearity of control variables with the independent variable or the dependent variable show it in a matrix in the appendix and present the result in a line or two here to justify their inclusion

KISS for variables *describe here, look on Mendeley*

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 in 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* is not accounted for in the FE estimation, although it maybe introduced as an interaction term (which will be explained in the Analysis section below). *some theory behind Fixed Effects one or two lines*

Part II:

Model Specification:

Research Question: What are the effects of county-level economic variables on the difference in two-party vote share of the Republican party between 2012 and 2016?

Hypothesis: Keeping in mind that Donald Trump's victory was unprecendented and he was touted as a non-regular candidate (i.e. different than a regular Republican), the effects of county-level economic variables should be different in 2016 election than the previous election. A rise in unemployment should have a positive impact on Republican vote share since Trump's campaign was revolved around "taking jobs back". However, in terms of per capita income, a lower per capita income should have a positive impact since his campaign also claimed to raise the lives of the poor people who had suffered from the wraths of globalization and the neglect from the political elites.

Research Equation:

$$rep.share.gro = \alpha + \beta_1(unemp\_gro) + \beta_2(pci\_gro) + \beta_3(controls) + \epsilon$$

Dependent Variable:

rep.share.gro: The dependent variable is the difference in the two-party vote share of the Republican party at the United States Presidential Election, for each county between election year 2012 and 2016. *For example from the Dataset* If, in a county in 2012, the two-party vote share for a Republican party is 0.55 and in 2016 it is 0.60, then the value of *rep.share.gro* will be 0.60-0.55 = 0.05. This indicates a growth of 5 percentage point in voteshare for Donald Trump compared to what the Republican nominee Mitt Romney obtained in 2012. A positive value means that Donald Trump received more votes in that county compared to Mitt Romney. The difference in using this variable, compared to the one used for the model in Part I, is that we are trying to measure Trump-specific effect here. Depending on how the dependent variable reacts to different explanatory variables in this model, we can come to conclusions on whether Trump was a generic Republican or not and if he was affected by the same factors that affected Republican candidates in the past.

Independent Variables:

unemp_gro: The first independent variable, which is also the explanatory economic variable, is the growth in unemployment rate for each county between 2015 and 2012. An ideal measurement would have been to see the growth in unemployment rate between 2012 and 2016, since both are election years. A positive value means that the unemployment rate in 2015 was greater than in 2012 and vice versa. However, economic data for 2016 is not available yet, which limited our ability to use the most recent data. However, the usage of 2015 unemployment rate should not be problematic and is infact considered an appropriate measure to conduct an analysis as such in the event of no other optimal option. *For example* and *Justify the usage of 2015.*

pci_gro: The second independent variable is the growth in per capita income for each county between 2015 and 2012. It is calculated using the same method as the growth in unemployment rate and with similar limitation in terms of the data for 2016. A positive value means that the per capita income in 2015 was greater than in 2012.

collinearity between these two variables

Control Variables:

pop: The population of a county in 2015 is used as a demographic control variable. Like in the previous model, this variable is introduced to control for the size of the counties.

educ: 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*

white.percent: The percentage of white people (both male and female combined) in a county in 2015 is used as another demographic control variable. Like in the previous model, this variable is introduced to control for race and the percentage is calculated by dividing the total number of white people in a county in 2015 by the total population of that county for the same year. *why white is used instead of black*

rural: A dummy variable for rural is used, similar to the one in the model in Part I. A value of 1 means a county is rural and a value of 0 means it is urban.

collinearity of control variables with the independent variable or the dependent variable show it in a matrix in the appendix and present the result in a line or two here to justify their inclusion

Estimation Technique

The regression estimation technique used for this model is a Ordinary Least Squares Model of estimation (called OLS estimation, hereafter). OLS estimation is ideal here due to the presence of a Cross-Sectional Dataset with no time-wise variation. The measures of the previous election(i.e. 2012) is incorporated in the cross section as a first-differencing. The OLS estimation explains the effect of the explanatory variables on the dependent variable for each specific county between 2012 and 2015. *Something more on OLS and what is does here*

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 dependent variables on both parts of the research were obtained from election data that consisted of the Republican vote-share in the presidential election between 1992 and 2016. The dataset came from Dave Leip's Atlas of U.S. Presidential Elections.[8] To calculate the two-party voteshare, the voteshare for Republican party was divided by the sum of voteshares for Republican and Democratic party. The lag of republican voteshare 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) website.[9] 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 to get the growth rate. The data on other explanatory economic variable, i.e. per capita income, and the population of each county from 1992 to 2015 were called from the Bureau of Economic Affairs (BEA) API directly into R studio.[10] 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 used to create the rural dummy

[8]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 subnational level.

[9]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).

[10]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).

was obtained from the United States Census Bureau.[11] 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.[12] The race data that included the number of whites in each county between 1992 to 2015 was obtained from the database of the National Cancer Institute.[13] The counts of race and education data were converted into percentages by combining them with the population data obtained from BEA. All the dataframes were merged based on county fips (unique identification for each county set 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 selected 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 Hawaaii 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. *remember if we do logit, we will do it for 3086 counties and if we do ols we will do it for 3112 counties*

[11]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).

[12]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).

[13]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).

Analysis:

First Question:

Second Question:

Testing the hypothesis:

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).

it would be interesting to see how people assign blame for their economic conditions on the President even if their regional governor or majority representation in the house may be from the non-incumbent's party. Moreover, Curry (**???**) writes that economic influence on the vote for president has been continuous and is not necessarily dependent on federal government control over the economy.

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 is 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.

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:

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

Bartels, L. M. (2005). What's the matter with what's the matter with kansas ?