6-2016

[Revised] Econometrics Review of recent Congressional District Trends Tied to Voting Patterns

Kathleen Ryan Redpath

kredpath@vols.utk.edu

Kathleen Redpath
May 30th, 2016
UNHO 498
Honors Project

<u>Introduction</u>

Political economics is a tricky field to pin down, especially with approximately 196 countries, each containing their own cultures, customs, people groups, and governments. The political groups we choose to align ourselves can help to transfer and explain the identities of a group of voters. With a two-party system at the national level, the U.S. would seem to be able to employ a more straightforward methodology than other developed countries for predicting voting patterns, but when Presidential voting is broken down by Congressional district, it becomes clear that the states are more than simply "red or blue". Many other countries function under a multi-party system where no one group can gain power, leaving parties to form coalitions. The closest example of similar party divides seems to be in the U.K., where the multi-party system has been nicknamed the "two and a half party system", thanks in part to the majority heads of the Conservative and Labour Parties, with the Liberal-Democrats having a strong presence able to shift a majority. In fact, many in the UK considered their legislative process to be a two party system when the Liberal-Democrats formed a coalition with the Conservative Party to turn the tide in 2010. Since then, the U.K. has had to change the way they forecast elections, understanding that there are not simply two candidates for power anymore. In one sense, forecasting in the U.S. is simple enough; when voters shift, it is towards one majority or another. However, when the primary candidates and voter majorities are affiliated with only one of two parties, these parties will encompass a broader spectrum of people, for whom key patterns can be more difficult to distinguish. It can be especially tough to predict the type of person who will select party candidate because the unknown factors can be so broad as well. Why a certain district picks a political party it finds suitable is complex.

Near the end of 2015, the online group FiveThirtyEight published an article titled, "How Demographics Will Shape The 2016 Election"[1]. In it, author David Wasserman delves into what he describes as the main demographics that have held sway over past Presidential elections and the kind of voter turnout that would shift the balance of power. There is even a "Swing-O-Matic", inspired by BBC's Parliamentary voting tracker, which allows users to increase or decrease the percentage of key demographic voter turnout and, based on prediction of these demographics, establish the tipping point from a red state to a blue state and vice versa. Their primary focuses of race and education certainly make it appear that a state really could go either way, depending on who shows up to the polls. However,

---

[1] Bycoffe, Aaron and Wasserman, David. *How Demographics Will Shape The 2016 Election.* "What Would It Take To Turn Red States Blue?". FiveThirtyEight.com. Dec 3, 2015.

as much as this can help explain large shifts in voting trends, there are still discrepancies having to do with the make up of districts, particularly with how each state draws new ones every ten years. The practice whereby state politicians redraw their districts with the intent of influencing the vote share in future elections is known as gerrymandering,

Briefly, there must be some discussion of gerrymandering and its impact on Congressional districts. Earlier this year, in his State of the Union speech, President Obama addressed gerrymandering and called for an end to "the practice of drawing our Congressional districts so that politicians can pick their voters and not the other way around."[2] Certain states, such as Maryland and North Carolina, are particularly harsh offenders when it comes to redistricting in order to gain Congressional advantages.

Pictured left: Maryland districts. Pictured right: North Carolina districts.

In the figures above, the districts have been drawn to give a skewed presence to Democrats in Maryland and Republicans in North Carolina. There have been many proposed solutions to this problem, such as letting a computer generate unbiased population-based districts. However, this process does not adequately factor in minority groups and may still ostracize those minorities. According to the Voting Rights Act, in the process of redistricting, there must be assurance of effort towards minority representation in Congress. Some states, though, such as Florida, have used this to pack minorities into a single district and actually reduce the power of certain minorities[3]. Because of this, it is hard to say whether or not the districts would "fairly" elect a president, even if everyone turned out to vote. This is the reason for detailed analysis of each Congressional district and their choice of Congressmen, to further examine if the trends of the district follow predictable patterns, or if gerrymandering bias has enough of a noticeable effect on both Presidential and Congressional elections to limit our ability to accurately forecast.

Many attempts have been made to examine the causal structure of what modern economics refers to as "the popularity function". Here, using multiple angles to attempt to foresee short-term outcomes, there is a sharp focus on autoregressive

---

[2] Ingraham, Christopher. "This is actually what America would look like without gerrymandering". The Washington Post Online. Jan 13, 2016.
[3] Ingraham, Christopher. "".

models (Frey)[4]. Although we will be focusing on a linear cross-sectional equation, it is important to note that a more in-depth regression would include the dynamic function of an autoregressive model, which utilizes time series data to show how the values influence each other over the course of the model. The questions I am asking include, "Which microeconomic variables most coincide with the movement in a Congressional district's party affiliations?"

In examining factors from primary party power to races and socioeconomic status of a Congressional districts, I am looking for any factors that appear to correlate with specific Party Powers in the House and Presidential races. For the dependent-variable, I am using Senate Party affiliations in 2010. For the independent-variables, I am using cross-sectional data for each of the 435 Congressional districts as recorded in the 2010 census, with Presidential vote share for 2008 and 2012 on the right-hand side (since there was no election in 2010 when the rest of the data is specified). The 436th observation for the Congressional Representation of the District of Columbia, and includes Presidential voting percentages, even though D.C. does not get voting shares equivalent to the other states, has very limited power in the House of Representatives, and has no representation in the Senate.

Models and Methodology
Multiple trains of thought persist in the realm economic explanations of elections. Primarily, "party power" is considered as one of the long-time front-runners of influence. However, studies have shown that Congressional parties are less likely to follow the party of the district's Presidential vote than they once were (Bibby)[5]. This, however, does not mean that voters do not still respond in predictable ways to changing economic conditions. Here, I lean on Tufte's Economic Explanation of the Elections model, which guided much of the modern work from which I drew my assumptions. It leans on three primary ideas:

(1) "that national economic conditions are manipulated by politicians for political benefits."
(2) "that national economic conditions have a direct and predictable effect on the Presidential and Congressional vote."
(3) "that voters are motivated essentially by their personal, short-run economic welfare and so reward or punish an incumbent administration depending on the state of the economy."[6]

Tufte's work as a Statistician has preceded his economic theories, but much of the literature from the past ten years referenced Tufte's model as a direct

---

[4] B.S. Frey, *Modern Political Economy*, Martin Robertson (London 1978); B.S. Frey, 'Politico-Economic Models and Cycles', *Journal of Public Economics* 9, 1978: 203-220.

[5] Bibby, John E., Thomas E. Mann, and Norman J. Ornstein. 1980. Vital Statistics on Congress. Washington, DC: American Enterprise Institute.

[6] Tufte, Edward. 1978. *Political Control of the Economy*. Princeton: Princeton University Press.

influence. I mention Tufte's model simply because much of the literature that I drew from was based on it. Each subsequent researcher had their own reasons and interpretations of the effects and motivations, but they each steadily assured that Tufte's ideas about the economic explanations of elections allowed us to assume that there was a causal relationship, as opposed to simple correlation, to be explored between the way people vote and the state of their local economies.

The main model I used in shaping my theories was configured by Hummel/Rothschild (2013)[7]. Similarly, I am looking for predictive values to forecast elections, except that they used panel data formatting to combine multiple states' cross-sectional data, as well as time-series data across multiple Presidential elections at the state and national levels. They acknowledge that using state level data is "quite rare for forecasting models, but it confers several advantages." Namely, it allows for a higher accuracy probability, since we have a much wider range of data points. It also helps to focus on better finding which data is fundamental. **MOST IMPORTANTLY**, Hummel and Rothschild use a linear regression of the fraction of major party vote received by one party, based on multiple economic and political dividers. Interestingly enough, they note how changes in state data have more of an impact than the absolute or annual data (pg. 17). Their work also recognizes how different quarters of the year are statistically more or less significant in their predictions, such as forecasting gubernatorial elections using unemployment.

The reason I chose not to pursue a panel-series data set was mainly because when trying to find state-level data for the Congressional districts, you quickly figure out that the districts change. In the ten years between census data, new districts have emerged, old ones have disappeared, and many have shifted their borders. While we could have included dummy variables to attempt to rectify the differences, it would require a model of a higher level. What stands out in the Hummel/Rothschild model is the way they check for errors for predicting election results by running and calculating regression coefficients, while excluding the given year (pg. 19). I shaped my Stata regressions to attempt to do this for the Congressional districts but comparing districts instead of separate time periods.

In deciding whether to use the Electoral College votes, especially since our two-party system gives us a very broad outlook of the voters, I read through the work of Grofman and Feld (2005), who examined the Electoral College and the claims both for and against it[8]. They looked less at the overall Electoral votes and more at the factors affecting those votes, which they called the "electoral responsiveness". They based this on Tufte's framework in 1973, which methodized how to calculate the swing ratio and partisan bias. In understanding these tools, I

[7] Hummel, Patrick and Hummel, David. "Fundamental Models of Forecasting Elections". Electoral Studies (2013) 35.

[8] Grofman, Bernard and Feld, Scott. "Thinking About the Political Impacts of the Electoral Colelge". Public Choice (2005) 123: 1-18. Springer 2005.

am better able to explain the factors determining which political party a Congressional district has chosen. Grofman and Feld used weighted voting games and showed how they might be applied to the Electoral College to understand the nature of large-state versus small-state biases. They used this method to show how it led them to rely on a "multiplicative model of optimal campaign investment at the state level." Multiplicative models typically use time-series data and focus on the percentage changes rather than being primarily interested in the absolute changes. This appears to be consistent with how Hummel approached his election forecasting, and it gives our linear model a better backing, especially in choosing to use log function for certain beta coefficients. Mainly, I can understand why the Electoral College is a reliable Y-variable in my model.

Assembling Data Set Variables

In the midst of attempting to better understand federal election procedures, it is still vital to remember the types of bias that influence the data. As well as addressing bias, I wanted to point to those who have better assessed the difficulties of the model I am working with. For example, our model does not account for potential projection bias, in which we assume that future voters will resemble the voters that participated in the 2010 census and the '08/'12 elections. There are economists, such as Muûls and Petropoulou (2007), who developed an infinite horizon, political agenda model.[9] With a continuum of political districts, they managed to examine the Electoral College from an angle that looks at tariffs and trade protectionism influencing both sides of the relationship between both "swing" and "decided" states. Their work had no direct influence on my economic exercise, but it does carry weight for the broader implications of this work.

In understanding the selection bias that might occur, I looked to Larry M. Bartels, who writes about Economic Inequality and Political Representation. Bartels' model is similar to my own, in which he also conducts a lot of cross-sectional work. He mainly focuses on how Senators vote once in office regarding matters of controversy, predicting who they're more likely to listen to based on income distribution. He acknowledges that "wealthier and better-educated citizens are more likely than the poor and less-educated to have well-formulated and well-informed preferences, significantly more likely to turn out to vote, much more likely to have direct contact with public officials, and much more likely to contribute money and energy to political campaigns".[10] With substantial widening of the margins between the rich and not-so-rich, the significance of these gaps has been increasingly magnified. Increasingly expensive campaigns, decreased participation in labor unions, and accelerating lobbying by corporations and businesses has only

[9] Muûls, Mirabelle and Petropoulou, Dimitra. "A Swing-State Theory of Trade Protection in the Electoral College". University of Oxford: Department of Economics Discussion Paper Series. 2007.

[10] Bartels, Larry M. *Economic Inequality and Political Representation*. Department of Politics and Woodrow Wilson School of Public and International Affairs. Aug 2005.

caused the political campaign process to further shift in what appears to be the wealthy class's favor.

My research acknowledges bias in that it can't take account of which wealth classes are voting. It can only look at the income of a district, who they voted for, and then draw conclusions. However, Bartels (as well as his sources: Verba, Schlozman, and Brady 1995) drive home the point that there is an increasing disparity between rich and poor citizens (pg. 3). While Bartels expands his ideas into the implications for Senators' patterns once elected, he starts with how these factors influence voter electoral turnout, something I want to acknowledge as potential bias in my tests.

For incidental reasons of data availability, my research focuses on the different economic factors for each district and who the Representatives of each district are. I examine which factors vary or remain consistent between those who vote for a Republican or Democratic Congressman. This doesn't account for instances where candidates ran unopposed. The Congressional party affiliations appear on the left side of the equation (as a dummy variable), with Presidential vote-share on the right side. This will show if a difference in vote-share for President Obama compared to both McCain and Romney in the '08/'12 Presidential elections coincides with a change in Congressional party affiliation. The variables in brief are outlined below.

Listing Variables with Descriptions

Variable	Description
State:District	Preceded by the state for clarification purposes, this divvies the variables by Congressional district
District Presidential Party Majority	District party majority for most 2008 and 2012 Presidential elections.
Congressional Party Majority	Either Republican or Democratic, as of 2010.
Total Population	Total number of persons counted for a district, male and female, with ages ranging from under 5 years to above 85 years, with the Median Age.
Race	Persons identifying as White, Black or African American, American Indian or Alaska Native, Asian, Native Hawaiian or Other Pacific Islander, as none of the stated races, or a combination of two or more races.
Hispanic or Latino Origin by Type	Originating from Mexico, Puerto Rico, Cuba, or an unspecified Country.
Place of Birth	Native, State or territory of birth and residence, and Foreign Born

Ancestry	"Ancestry" Assumed Given.
Listing Variables with Descriptions [Cont.]	
Veteran Status	Civilian population 18 and older and Civilian Veterans
Residence 1 Year Ago	Whether in the US or abroad
Workers - Employment Status	Employed, Unemployed, Armed Forces, In and out of the Labor Force
Commuting to Work	Over 16, drivers, carpoolers, public transportation, walkers, and other
Occupation	Specified and broken down below
Industry	Thirteen Types of Industry Classification
Class of Worker	Private wage, Gov. workers, self-employed, and unpaid family workers
Housing Occupancy	Total, occupied, and vacant. Homeowner and Rental Rates
Housing Tenure	Occupational Housing Units with average household size of owners/renters
Year Householder Moved into Unit	In ten year increments
Value	Ranging from less than $50,000 to $1million or more. Also, median in dollars.
Mortgage Status	Housing units with and without a mortgage
Selected Monthly Owner Costs(SMOC)	Mortgage rates impact on housing units
Gross Rent	Prices on Occupied Units paying Rent
Socioeconomic - Income (In 2014 inflation-adjusted dollars)	Total Household Income, Median and Mean included.
Health Insurance Coverage	For civilians and minors with private, personal, and public healthcare. Or none
Percentage of Families and People whose Income in the Past 12 Months is Below the Poverty Level	With children under 18, children only under 5, married and unmarried, female-led no-husband-present, single-majority age percentage, and households with unrelated individuals.
Education - School Enrollment	Those 3 years and older enrolled, and what level of schooling
Educational Attainment	In population 25 years and older
Business - Paid employees for pay period including March 12	Total and for each sector
Annual payroll ($1,000)	Total and for each sector
Total establishments	Total and for each sector

Table 1
Demographic Regression by Party Affiliation - 1

Dependent Variable: Republican

Independent Variable	Beta Coefficient	t-stat	P>|t|
Voters for Obama in 2012	-0.026	-3.98	0.000
Voters for Obama in 2008	0.001	0.14	0.886
Females	3.34e-08	0.03	0.976
Ages 18 years and older	5.57e-08	0.08	0.940
Whites	7.06e-08	0.44	0.658
English Ancestries	-2.61e-07	-0.39	0.697
Constant	1.767	6.39	0.000

Notes: I ran the regression against the Y-value "Republican". This is a dummy variable that equals 1 for a Republican district and 0 for a Democratic district. The R-squared is 0.6461. There are 436 observations, and these are made from the 435 official Congressional districts and their political party affiliation with the average demographics of the District of Columbia acting as the 436th. *The District of Columbia has a single Congressman and 3 electoral votes, but they have very limited power in the House*

An initial attempt at running a regression of some of the variables: Previous Election voting, Gender (female), Older than eighteen years, White, English heritage.

Table 2
Logistic Demographic Regression by Party Affiliation

Dependent Variable: Republican

Independent Variable	Beta Coefficient	z-value	p>|z|
Voters for Obama in 2012	-0.235	-2.30	0.022
Voters for Obama in 2008	-0.191	-1.65	0.098
Constant	22.477	8.23	0.000

Notes: The Log Likelihood is -70.45, holding the number of observations constant at 436. The Y-value remains our dummy variable of "Republican".

It's a basic Logit function, to make sure my variables are working. Here it tells us that the likelihood of a Republican district having voted for Obama in 2008 and 2012 is extremely low.

Table 3
Demographic Regression by Party Affiliation - 2

Dependent Value: Republican

Independent Variables	Beta Coefficient	t-value	P>|t|
Ages 18 years and older	2.12e-07	0.26	0.794
Government Workers	-3.80e-06	-2.11	0.036
Households $500,000~$999,999*	-8.23e-06	-4.12	0.000
Incomes $200,000 and more**	1.94e-06	0.43	0.668
Married Couple Families	-4.614236	-5.64	0.000
Bachelors degrees	-9.13e-07	-0.56	0.575
Constant	1.079	2.64	0.009

Notes: The R-squared is 0.1377. *Households are referring to the approximate valuation of currently owned houses. **Income is each approximate household income.

This regression takes into account some of the more "obscure" factors belong to each district, such as how many individuals have their Bachelor's degree, belong to a married couple household, own a house paid for at an amount more than $500,000 and less than $1million (FZ), or make more than $200,000 a year (HU).

Table 4

Ordinary Least Squares Regression Analysis

Dependent Variable: Republican

Independent Variables	VIF	1/VIF
Incomes $200,000 and more	5.57	0.180
Bachelors degrees	5.39	0.186
Households $500,000~$999,999	2.62	0.381
Married Couple Families	1.50	0.668
Ages 18 years and older	1.47	0.682
Government Workers	1.14	0.876

Notes: The Mean VIF or Variance Inflation Factor is 2.95.

In checking for Multicollinearity, one way if to see if the Variance Inflation Factor (VIF) is greater than 5. Above, we see that having a Bachelor's degree, earning more than $200,000 a year (HU), or having private insurance are inflated variables and lend themselves to a more severe multicollinearity. To find out which variables are highly correlated, I ran a correlation matrix.

Table 5
Pearson's Correlation Coefficient Matrix - 1

Ind. and Dep. Variables	(R)*	18 yrs and up*	Gov. workers*	Homes $500k to $999k*	Inc. $200k and up*	Marr. Coup. Homes *	Bach. Deg.*
(R)	1.0	–	–	–	–	–	–
18 yrs and up	-0.04	1.00	–	–	–	–	–
Gov. workers	-0.07	0.21	1.00	–	–	–	–
Homes $500k to $999k	-0.23	0.28	0.13	1.00	–	–	–
Inc. $200k and more	-0.11	0.32	0.07	0.77	1.00	–	–
Marr. Coup. Homes	-0.20	-0.20	-0.23	-0.25	-0.42	1.00	–
Bach. Deg.	-0.06	0.49	0.08	0.66	0.86	-0.51	1.00

Notes: *(R) is Republican. 18 yrd and up is Ages 18 and Older.
Gov. workers is Government Workers. Homes $500k to $999k is
Households valued at $500,000 to $999,999. Inc. $200k and up is
Household Incomes $200,000 and more. Marr. Coup. Homes is Married
Couple Households. Bach. Deg. is Bachelor's degrees.

Here I can get a better understanding of what the VIF indicated. Having a
Bachelor's degree is tied closely to being employed by the government. Making over
$200,000 a year is closely related with owning a house worth between $500,000
and $1million and with having your Bachelor's degree. A household that consists of
a married couple is strongly negatively related to having a household income of
more than $200,000, even though they potentially have combined salaries.

Table 6
Pearson's Correlation Coefficient Matrix - 2

| Ind. and Dep. Var's	(R)*	Vote for Obama 2012*	Vote for Obama 2008*	18 yrs and up*	Blck or Afr Am*	Born in USA*	Civ. Vets*	In Lbr Frc*	W/ Hlth Ins Cov*
(R)*	1.00	-	-	-	-	-	-	-	-
Vote for Obama 2012*	-0.80	1.00	-	-	-	-	-	-	-
Vote for Obama 2008*	-0.79	0.99	1.00	-	-	-	-	-	-
18 yrs and up*	-0.04	0.06	0.07	1.00	-	-	-	-	-
Blck or Afr Am*	-0.25	0.37	0.35	0.03	1.00	-	-	-	-
Born in USA*	0.37	-0.41	-0.38	0.20	-0.00	1.00	-	-	-
Civ. Vets*	0.35	-0.40	-0.39	0.18	-0.13	0.69	1.00	-	-
In Lbr Frc*	-0.17	0.19	0.21	0.63	-0.03	-0.04	-0.16	1.00	
W/ Hlth Ins Cov*	0.04	-0.03	0.01	0.69	-0.16	0.34	0.15	0.67	1.00

Notes: *(R) is Republican. Vote for Obama 2012 is Voters for Obama in 2012. Vote for Obama 2008 is Voters for Obama in 2008. 18 yrs and up is Ages 18 and Older. Blck or Afr Am is Blacks or African Americans, hereby referred to as Blacks. Born in USA is Born in the United States, excluding non-voting districts and

territories. Civ. Vets is Civilian Veterans. In Lbr Frc is In the
Labor Force. W/ Hlth Ins Cov is With Health Insurance Coverage.

Again, checking for multicollinearity with a different set of variables by creating a correlation matrix, I can understand the influence that the different variables might have on one another and the Y-value. The reason our best option is to do nothing here, is that the large correlation I see between certain variables is not necessarily indicative of incorrect information. The idea that the majority of citizens that voted for Obama in 2008 voted for Obama in 2012 is not a cause for alarm. There are stacks of literature that attempt to explain many of the strongest correlations and much speculation about outside causes or reasons.

In most cases, I remove the Presidential elections of 2008 and 2012 because they are often redundant when paired with the Representative's party affiliation. However, we are looking to see how these districts might vote the same way and yet have completely different socioeconomic makeups. In knowing that, I can run models with the Presidential election cycles, and recognize other significant correlations.

In checking for heteroskedasticity, I started with a quick visual test.
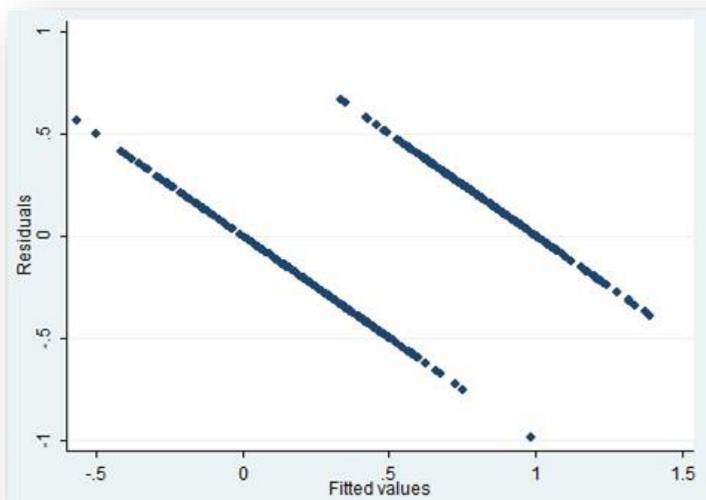
With the suspicion of heteroskedasticity, I ran a Breusch-Pagen / Cook-Weisberg test, which specifically tests for linear heteroskedasticity. It tests the null hypothesis of equal error variances against the alternative hypothesis that the error variances are unequal and thus a possible multiplicative function of one or more variables.

Table 7
Breusch-Pagan Cook-Weisberg test for Heteroskedasticity

        Chi-squared (1)   3.70

```
        Prob > chi2                              0.0545
```

Notes: The Null Hypothesis being that there is a constant
variance among the variables using the fitted values of
Republican.

      The high chi-square value lets us know that heteroskedasticity is present. To try to balance the variables' error variance, I wanted to see what would happen if I ran them in natural log form. I took some of the variables I initially looked at and defined them in logarithm form. Then I ran them to see if there was a better fit against our Y-variable, and if it helped the issue of heteroskedasticity. However, after log-ing some of our variables, I then have to use a new method to test for heteroskedasticity. This is because the standard Breusch-Pagen test does not work for non-linear forms of heteroskedasticity, where error variances are measured exponentially through other functions.

      So instead we use White's general test for heteroskedasticity. This extension of the Breusch-Pagen test is specially designed for the case in which errors are not normally distributed. The results of running it through Stata are below.

Table 8*
White's test for Homoskedasticity

```
    Chi-squared (27)              89.96
      Prob > chi2                 0.0000
```

Cameron & Trivedi's Decomposition of IM-test

Source	Chi-squared	df	p-value
Heteroskedasticity	89.96	27	0.0000
Skewness	29.44	6	0.0001
Kurtosis	4.95	1	0.0261
Total	124.35	34	0.0000

Notes: The Null Hypothesis being that there is homoskedasticity,
which we are testing against the Alternative Hypothesis that
there is unrestricted heteroskedasticity. *This is a two part
updated version of the Breusch-Pagen test, which is why we have
kept it as one table, not two.

When I run the White Test in Stata, I can clearly see that heteroskedasticity is an issue. There are multiple reasons for the variance in error terms. Mainly, I know that the differences in the variables attributed to the individual districts do not vary linearly with whether the district is Republican or Democrat. To account for these exponential changes in error variance I can shape our linear function and transform the variables. I can try this assuming that heteroskedasticity might be a model misspecification before assuming severe biases in the standard errors, though I do not want to completely forget that risk.

To transform our variables, I looked to the logarithm forms.

Table 9
Demographic Regression by Party Affiliation - 3

Dependent Variable: Republican

Independent Variables	Beta Coefficient	t-value	P>|t|
Voters for Obama in 2012*	-0.724	-2.64	0.008
Voters for Obama in 2008*	-0.524	-1.71	0.088
Whites*	0.038	0.63	0.529
Ages 18 years and older*	0.108	0.42	0.675
Born in the United States*	0.144	1.27	0.206
Constant	1.591	0.47	0.641

Notes: The R-squared is .6222. Again, as I have kept it throughout this exercise, the Y-value is the Republican dummy value. *Logarithm form.

Our R^2 is the same, but our coefficients have improved. By using the natural log functions for our model, I can understand that the exponential fit can better explain some of our variables. Look at the coefficient for the "*log_White*" variable. Previously, its coefficient reflected no significant relationship, but now it can tell us that a Congressional district that is Republican is likely to have 3.8% more white citizens than one that is Democrat.

I also try to correct our functional form, to guard against not only heteroskedasticity but also serial correlation. By finding a form that better fits, I am able to fight against type-II error probability, while possibly increasing our probability of type-I error. I will allow for that possibility, though, because I would rather not risk biasing my betas.

Summary

What is there to learn about what makes a Congressional district Republican or Democrat? I could expand the data indefinitely, exploring many different connections between race, income, households, and careers. Some of the main factors that seemed to highly differ between Congressional district party powers were: how many people voted for Obama or not in 2008 and in 2012, how much income the majority of them earn annually, race, and education. When running a district's Congressional party against demographic factors, I found that:

- Republican-held districts are made up of less-educated, aging, and predominantly white populations.
- Democrat-held districts tended to have higher percentages of black voters, higher rates of educated females, and other minority voters, such as Hispanics and Asians.

As I discussed in understanding the risks of multicollinearity, there are factors with particularly high correlation rates. One of the strongest demographic variable connections is a Democratic district and the percentage of black citizens. Now, while a Democratic district does not always have a large black population percentage, a district with a large black population percentage is almost always Democratic. In diving into the history of why this is so, many writers point less to the socioeconomic factors of today and more towards the history of the Democratic Party and race relations. Historically, blacks or African Americans have not always sided primarily with the Democratic party, but the continual advances of politicians like Harry Truman and Lyndon B. Johnson in the arena of civil rights have formed what many consider to be the main force to vote Democrat. From examining the way similar districts vote, I propose that we are, in fact, less divided as a nation on where we stand and more separated by how we appear.

The linear approach works to encompass multiple moving variables, while taking into consideration the exponential logarithms. The theory that Republican and Democratic Congressional districts look very different from each other holds together on certain accounts, but there are more similarities in the data than there are differences. This might change, and political parties might have become more polarizing in the last decade, but the district lines do not necessarily have to divide groups in which participants have more in common than they might realize.

Analyzing Democrats and Republicans during the 2016 Presidential election cycle has been a fascinating exercise. The party frontrunners are more caricatures than candidates, and there is significant controversy within party ranks about whether Donald Trump and Hillary Clinton are accurate and sufficient representations of the Republican and Democratic parties. Interestingly enough, the numbers indicate that both candidates share key characteristics with the individuals who are most likely to support their party. Donald Trump is aging, white, with some college education, and he values consumerism. Hillary Clinton is female, post-graduate, has liberal views, and she is working to appeal to the minority vote.

Despite the fact that they mirror many of the individuals that seem to make up their parties, both have been widely criticized, most surprisingly from within their parties. This is suggesting, at least from my outlook, that the US political system is fertile ground for another party to rise. With Republicans and Democrats looking increasingly polarized and many people finding themselves stuck between candidates this year, perhaps we will see the rise of third party power. If so, who will make up this third party? Will it be an already-formed group or some unknown affiliation?

Already, people have begun to look at the Libertarian party for their contribution, with former New Mexico Governor Gary Johnson having been selected for their ticket.[11] The Libertarian party is also the only third party with ballot access in all 50 states. A national poll from May 16th this year, conducted by Fox News, showed Johnson receiving 10% of the vote from registered voters. Johnson ran on the ballot in 2012, scoring almost 1% of the general election vote, but he, as well as his party, believes that the bitter dissension among the Republican and Democratic parties has provided the Libertarians with a strong chance of breaking through. This might be the beginning of a major shift in American politics and how we, as American citizens, choose to align ourselves.

---

[11] Watkins, Eli. "Libertarians pick ticket, Slam Trump". CNN Politics. May 29, 2016.