Determinants of State-Level Voter Choice

Kenneth Rios*

December 30, 2018

Abstract

We estimate the marginal effects of minority populations, income inequality, and unemployment at the state-level on the probability of a given state voting for the Democratic Presidential candidate using a random-effects logistic panel regression. We utilize a balanced panel consisting of state-level voting outcomes for five Presidential elections from 2000 through 2016. We find a statistically significant, positive relationship between the probability of a given state voting for the Democratic candidate and the proportion of the Asian population in the state, the proportion of the female population the state, and the state unemployment rate.

I. Introduction

This paper examines the relationship between characteristics of US states and their propensity to vote either Democrat or Republican in Presidential elections. In general, the nature of the economic climate of a given state and the evolution of the demographics of its constituents will determine the preferences that the population has for candidates from either the Republican or Democratic party. Based on the articles by Wright (2012) and Galbraith & Hale (2006), we have reason to hypothesize that the probability of a given state voting for the Democratic Presidential candidate is

i) positively related with the proportion of minority populations in a state,

ii) positively related with income inequality in a state, and

iii) positively related with unemployment in a state.

*NYU, Department of Economics

Our paper aims to test this hypothesis.

In order to test this hypothesis empirically, we use annual data for each state and the District of Columbia with a frequency of four years, which corresponds to every Presidential election from 2000 to 2016 (2000, 2004, 2008, 2012, and 2016). The dataset contains 51 ids $\times$ 5 Presidential election years = 255 observations. Our dependent variable of analysis is a binary variable which takes the value 1 if the Democratic candidate is elected by the state and 0 if the Republican candidate is elected. Since our dependent variable is binary, we estimate our model using logistic regression in order to ultimately estimate the marginal effects that the explanatory variables have on the probabilities of a state voting for the Democratic Presidential candidate. This methodology is much in the spirit of the methodology employed by Carsey & Wright (1998).

Our empirical results suggest, via the Hausman test, that the random-effects version of our logistic panel regression produces the efficient estimator as compared with the fixed-effects model. Using the estimated coefficients from the random-effects model we calculate the average marginal effects on the probability of a state voting for Democratic Presidential candidate. Our estimated average marginal effects provide evidence of a positive association between the proportion of several minority populations in a given state and the state unemployment rate with the probability of voting for the Democratic Presidential candidate. We, however, do not find evidence to support a relationship between the voting probability and state-level income inequality.

II. Literature Review

Wright (2012) finds that the Democratic candidate can engage in effective political campaigning by taking advantage of high unemployment time periods. The Democratic platform has been traditionally favored by minority groups suffering from high unemployment and poverty levels due to the platform's emphasis on welfare programs and progressive legislation. Wright conducts a panel regression of Democratic vote shares on various country-level characteristics using county-level US data from 1994 to 2010. Wright finds that, on average, a one percentage point increase in the county-level unemployment rate increase vote shares for the Democratic candidate by .515 percentage points. Our analysis will test this relationship across states instead of counties and incorporate the two most recent elections.

Galbraith & Hale (2006) undertake a comprehensive analysis that analyzes state-level inequality and its effect on the Democratic share of the voting population even after controlling for race and state and time effects. In particular they find that, for Presidential elections from 1992 to 2004, a 2% increase in inequality led to a 1.5% increase in the share of the Democratic votes. Our analysis attempts to estimate these effects using more recent election data and to include additional explanatory variables representing minority population sizes. Galbraith & Hale use the Gini Index as their measure of income inequality. In the spirit of their analysis and in many other works on income inequality, we also use the Gini Index as our measure of income inequality. Galbraith & Hale also discover that higher levels of minority populations in a given state give rise to a larger Democratic vote.

Carsey & Wright (1998) attempt to model the interactions between Presidential approval and voting outcomes in gubernatorial and senatorial elections for both the Republican and Democratic candidates. In their methodology, Carsey & Wright use a balanced panel dataset from ANES (American National Election Studies) across states for gubernatorial and senatorial elections from 1986 to 1990. They estimate their model using logistic regression, with voting outcomes as their dependent variable of analysis. In our analysis, we employ similar methodology as we also take voting outcome as the dependent variable and estimate a logistic regression on a balanced panel. Carsey & Wright also control for the clustered nature of the panel data, which we similarly do in our logistic panel regression estimation in order to account for the presence of autocorrelation in our panel data.

III. Theoretical Model

Using results from Galbraith & Hale (2008) and Wright (2012) as motivation, our paper hypothesizes that the the probability of a given state voting for the Democratic Presidential candidate is

i) positively related with the proportion of minority populations in a state,

ii) positively related with income inequality in a state, and

iii) positively related with unemployment in a state.

We now discuss the economic theory behind each point of our hypothesis using works discussed in the literature. Wright (2012) finds evidence to support that counties with higher levels of unemployment tend to generate a greater share of votes for the gubernatorial and Presidential Democratic candidate. Wright explains this finding theoretically by claiming that the Democratic Party essentially "owns" the issue of unemployment. Unemployment is an issue typically found to be a partisan issue among voters. Galbraith & Hale discuss reasons for why it is observed that greater income inequality tends to lead to a greater share of Democratic votes. In particular, they hypothesize that "economic inequality serves as a proxy for some of the real and imagined differences that divide so-called 'red and 'blue' states" (Galbraith & Hale, p. 9). Minority populations are then hypothesized to be more likely to vote for the Democratic candidate as they view the various progressive policies under the Democratic platform as more beneficial to them than the Republican alternative.

We now posit that the conditional probability of a given state voting for the Democratic Presidential candidate is an increasing, logistic function of the independent variables of interest M_p, I, and U:

$$Pr(Vote = 1|M_p, I, U) = \frac{1}{1 + exp\left(-(\alpha_0 + \sum\limits_{p \in P} \beta_p M_p + \alpha_1 I + \alpha_2 U)\right)},$$

where $Vote$ is a binary variable for the state-level voting outcome (=1 if the Democratic candidate chosen, =0 for the Republican candidate), M_p represents a minority (or relevant comparison group) population p in the state, I is income inequality in the state, and U is unemployment in the state. $\beta_p \; \forall p \in P, \alpha_0, \alpha_1, \alpha_2 \in \mathbb{R}$ are the population parameters.

The papers that we considered as our theoretical impetus investigated the relationship between only a subset of the variables in our theoretical model with voting probabilities. Therefore, our work builds upon these papers by hypothesizing a model that jointly considers the relationships of all of these variables with voting probabilities. Our model is similar to those discussed in the literature review as we are concerned with binary voting outcomes as our dependent variable, although other works neither focus on state-level outcomes for Presidential elections nor looked at recent data.

Our dependent variable of analysis is state-level voting outcomes from Presidential elections. Our independent variables of interest are state-level minority populations M_p, income inequality I, and unemployment U. Since the inclusion of minority populations will introduce a substantial amount of demographic data into the analysis, we do not incorporate additional state-level variables as controls.

Our theoretical model posits that the probability of a state voting for the Democratic Presidential candidate is a logistic function of our theoretical variables of interest: M_p, I, and U. The theoretical underpinning of the inclusion of these variables was discussed earlier in this section and is derived from theoretical interpretations of results derived in papers by Wright and Galbraith & Hale. We decide to model probabilities using a logistic function as this methodology is consistent with similar empirical estimation performed by Carsey & Wright. They estimate models that include binary voting outcomes as the dependent variable using logistic regression.

In our theoretical model, we posited a relationship between the various independent variables of interest M_p, I, and U, and the dependent variable of analysis, $Vote$. We hypothesized that the relationships between minority populations in a state, income inequality in a state, and unemployment in a state are all positively associated with the probability of a state voting for the Democratic Presidential candidate. We wish to test this hypothesis using our data. We expect that our hypothesis holds and that the average marginal effects on voting probabilities that are generated from our logistic regression estimates are all positive.

IV. Empirical Model

As per Carsey & Wright (1998), we estimate the population parameters in our theoretical model of logit probabilities by maximizing the likelihood function characterized by the multivariate regression of the log-odds on the independent variables of interest:

$$ln\left(\frac{Pr(Vote = 1|M_p, I, U)}{1 - Pr(Vote = 1|M_p, I, U)}\right) = \alpha_0 + \sum_{p \in P} \beta_p M_p + \alpha_1 I + \alpha_2 U + \varepsilon.$$

For the minority and comparison group population variables M_p we include the following groups $p \in P$:

1) Whites ($White_{it}$)

2) Blacks ($Black_{it}$)

3) Asians ($Asian_{it}$)

4) Females ($Female_{it}$)

5) Hispanics ($Hispanic_{it}$)

Furthermore, we will express these population variables in terms of the *percentages of the total state population*, for each state in each Presidential election year. It is important to emphasize that the various population groups are *not* mutually exclusive – for instance, White Hispanics would be pooled in both $White_{it}$ and $Hispanic_{it}$ variables. As a measure of income inequality we will use state-level Gini indices, $Gini_{it}$. To measure unemployment we simply use the reported October unemployment rates in each state for each Presidential election year $Unemp_{it}$, as these unemployment rates are likely to be the most relevant ones in forming voter's attitudes of candidates on Election Day, which takes place on the first Tuesday after November 1$^{\text{st}}$. Since we have observations for each state (and District of Columbia) i across Presidential elections spanning election years t, we treat our data as a (balanced) panel and run two versions of our logistic regression: a specification assuming conditional fixed effects and a specification assuming random effects. Our full econometric model incorporating panel data effects is therefore

Fixed Effects

$$ln\left(\frac{Pr(Vote_{it}=1|X)}{1-Pr(Vote_{it}=1|X)}\right) = \alpha_i + \beta_1 White_{it} + \beta_2 Black_{it} + \beta_3 Asian_{it} + \beta_4 Female_{it}$$
$$+\beta_5 Hispanic_{it} + \beta_6 Gini_{it} + \beta_7 Unemp_{it} + \varepsilon_{it}$$

vs.

Random Effects

$$ln\left(\frac{Pr(Vote_{it}=1|X)}{1-Pr(Vote_{it}=1|X)}\right) = \alpha + \beta_1 White_{it} + \beta_2 Black_{it} + \beta_3 Asian_{it} + \beta_4 Female_{it}$$
$$+\beta_5 Hispanic_{it} + \beta_6 Gini_{it} + \beta_7 Unemp_{it} + u_i + \nu_{it}$$

for $i = 1, 2, \ldots, 51$ and $t = 2000, 2004, 2008, 2012, 2016$; where X is the data matrix. $White_{it}$, $Black_{it}$, and $Asian_{it}$ do not sum to 100% as we have omitted a small population group consisting of Native Americans/Pacific Islanders. Therefore we do not have issues of perfect collinearity.

We consider both logistic panel regression specifications assuming either fixed effects or random effects and then decide the appropriate specification to estimate average marginal effects based on the conclusion of a Hausman test of the null hypothesis that the random effects estimator is the efficient estimator. For sake of consistency with respect to previous works mentioned in the literature review, we decided to carry on with logistic regression instead of probit regression with an understanding that there is little difference between both estimation procedures in practice. We did not consider running an ordinary least squares regression, which would correspond to the linear probability model, due to well known issues with the LPM. Among these are fitted probabilities that lie outside of $[0, 1]$ and issues with heteroscedasticity.

For each data series listed in Table 4.1 below, we use annual data for each state and the District of Columbia with a frequency of four years, which corresponds to every Presidential election from 2000 to 2016 (2000, 2004, 2008, 2012, 2016). The dataset contains 51 ids $\times$ 5 election years = 255 observations. We use percentages instead of levels for the demographic variables $White_{it}, Black_{it}, Asian_{it}, Female_{it}$, and $Hispanic_{it}$ as percentages offer a more intuitive and useful interpretation of the marginal effects. Summary statistics for these variables are provided in Section V.

Table 4.1: Data Series Used in Logistic Regression Analysis

Variable	*Data Series*
$Vote_{it}$	A binary variable that equals 1 if state i in election year t voted for the Democratic Presidential candidate and equals 0 if state i in election year t voted for the Republican Presidential candidate. We obtained the data used to construct this binary variable from electoral college results provided by the biennial *Federal Elections* releases published by the Federal Election Commission available on `https://transition.fec.gov/pubrec/electionresults.shtml`. This site provides links for each election year. For example, to access election year 2000 data, we would visit `https://transition.fec.gov/pubrec/fe2000/tcontents.htm` and then click on the *Official General Elections Results by State* link.
$White_{it}$	The number of whites in state i in each election year t expressed as a percentage of the total state population. That is, the number of whites in a state divided by the total state population, multiplied by 100. Data for election years 2000, 2004, and 2008 were obtained from the US Census Bureau 2000-2010 State Characteristics Intercensal Population Estimates released in October 2012, available by visiting `https://www.census.gov/data/datasets/time-series/demo/popest/intercensal-2000-2010-state.html` and clicking on the *Intercensal Estimates of the Resident Population by Sex, Race and Hispanic Origin for States and the United States: April 1, 2000 to July 1, 2010* spreadsheet. Data for election years 2012 and 2016 were obtained from the yearly US Census Bureau American Community Survey, available at `https://www.census.gov/programs-surveys/acs/news/data-releases/2016/release.html`, which directs the reader to *American FactFinder*.

$Black_{it}$	The number of blacks in state i in each election year t expressed as a percentage of the total population. That is, the number of blacks in a state divided by the total state population, multiplied by 100. Data for election years 2000, 2004, and 2008 were obtained from the US Census Bureau 2000-2010 State Characteristics Intercensal Population Estimates released in October 2012, available by visiting https://www.census.gov/data/datasets/time-series/demo/popest/intercensal-2000-2010-state.html and clicking on the *Intercensal Estimates of the Resident Population by Sex, Race and Hispanic Origin for States and the United States: April 1, 2000 to July 1, 2010* spreadsheet. Data for election years 2012 and 2016 were obtained from the yearly US Census Bureau American Community Survey, available at https://www.census.gov/programs-surveys/acs/news/data-releases/2016/release.html, which directs the reader to *American FactFinder*.
$Asian_{it}$	The number of Asians in state i in each election year t expressed as a percentage of the total population. That is, the number of Asians in a state divided by the total state population, multiplied by 100. Data for election years 2000, 2004, and 2008 were obtained from the US Census Bureau 2000-2010 State Characteristics Intercensal Population Estimates released in October 2012, available by visiting https://www.census.gov/data/datasets/time-series/demo/popest/intercensal-2000-2010-state.html and clicking on the *Intercensal Estimates of the Resident Population by Sex, Race and Hispanic Origin for States and the United States: April 1, 2000 to July 1, 2010* spreadsheet. Data for election years 2012 and 2016 were obtained from the yearly US Census Bureau American Community Survey, available at https://www.census.gov/programs-surveys/acs/news/data-releases/2016/release.html, which directs the reader to *American FactFinder*.
$Female_{it}$	The number of females in state i in each election year t expressed as a percentage of the total population. That is, the number of females in a state divided by the total state population, multiplied by 100. Data for election years 2000, 2004, and 2008 were obtained from the US Census Bureau 2000-2010 State Characteristics Intercensal Population Estimates released in October 2012, available by visiting https://www.census.gov/data/datasets/time-series/demo/popest/intercensal-2000-2010-state.html and clicking on the *Intercensal Estimates of the Resident Population by Sex, Race and Hispanic Origin for States and the United States: April 1, 2000 to July 1, 2010* spreadsheet. Data for election years 2012 and 2016 were obtained from the yearly US Census Bureau American Community Survey, available at https://www.census.gov/programs-surveys/acs/news/data-releases/2016/release.html, which directs the reader to *American FactFinder*.
$Hispanic_{it}$	The number of Hispanics in state i in each election year t expressed as a percentage of the total population. That is, the number of Hispanics in a state divided by the total state population, multiplied by 100. Data for election years 2000, 2004, and 2008 were obtained from the US Census Bureau 2000-2010 State Characteristics Intercensal Population Estimates released in October 2012, available by visiting https://www.census.gov/data/datasets/time-series/demo/popest/intercensal-2000-2010-state.html and clicking on the *Intercensal Estimates of the Resident Population by Sex, Race and Hispanic Origin for States and the United States: April 1, 2000 to July 1, 2010* spreadsheet. Data for election years 2012 and 2016 were obtained from the yearly US Census Bureau American Community Survey, available at https://www.census.gov/programs-surveys/acs/news/data-releases/2016/release.html, which directs the reader to *American FactFinder*.

$Gini_{it}$	An index ranging from 0 to 1, where 0 expresses perfect equality and 1 expresses maximal inequality, used as our measure of income inequality in state i in each election year t. The Gini index can be defined as one-half of the relative mean of absolute differences of incomes in a population. See Figure 4.1 below for the precise formula used to calculate the Gini index. We used state-level Gini indices for election years 2000 through 2012 calculated and published by Prof. Mark W. Frank in his state-level income inequality dataset, which is accessible by visiting `http://www.shsu.edu/eco_mwf/inequality.html` and downloading the spreadsheet labeled *Update on Other Measures of Income Inequality (Atkinson Index, Gini Coefficient, Relative Mean Deviation, Theil Index), 1917-2015:*. Figures for election year 2016 were unavailable hence they were imputed using 2015 figures.
$Unemp_{it}$	The October unemployment rate in state i in each election year t; that is, the percentage of the labor force in October that was not employed in state i in each election year t. This data was obtained from the *State Employment and Unemployment* releases from the Federal Reserve Bank of St. Louis available by visiting `https://research.stlouisfed.org/pdl/337` and clicking on *Download Data*.

Prof. Mark Frank calculated his estimates of state-level Gini indices using the formula below.

Figure 4.1: Gini Index Formula

The Gini index G_{state} is estimated as

$$G_{state} = \frac{\sum_{i=1}^{N}\sum_{j=1}^{N}|y_i - y_j|}{2\sum_{i=1}^{N}\sum_{j=1}^{N} y_j},$$

where y_i is the income of individual i in the state sample and N is the size of the sample of the state population.

We estimate our logistic panel regression model using maximum likelihood estimation. As previously mentioned, we do not use ordinary least squares estimation because the fitted values can correspond to "probabilities" outside the $[0, 1]$ interval. There is also the issue of heteroscedasticity being introduced into the errors when fitting a linear probability model. We decide whether to use the fixed effects or the random effects specification for interpretation and analysis by performing a Hausman test of the null hypothesis that the random-effects estimator is efficient (and consistent).

V. Empirical Results

The econometric model that we wish to estimate is

Fixed Effects

$$ln\left(\frac{Pr(Vote_{it}=1|X)}{1-Pr(Vote_{it}=1|X)}\right) = \alpha_i + \beta_1 White_{it} + \beta_2 Black_{it} + \beta_3 Asian_{it} + \beta_4 Female_{it}$$
$$+\beta_5 Hispanic_{it} + \beta_6 Gini_{it} + \beta_7 Unemp_{it} + \varepsilon_{it}$$

vs.

Random Effects

$$ln\left(\frac{Pr(Vote_{it}=1|X)}{1-Pr(Vote_{it}=1|X)}\right) = \alpha + \beta_1 White_{it} + \beta_2 Black_{it} + \beta_3 Asian_{it} + \beta_4 Female_{it}$$
$$+\beta_5 Hispanic_{it} + \beta_6 Gini_{it} + \beta_7 Unemp_{it} + u_i + \nu_{it}$$

for $i = 1, 2, \ldots, 51$ and $t = 2000, 2004, 2008, 2012, 2016$; where X is the data matrix. Both specifications are estimated using maximum likelihood estimation. We decide the appropriate specification to estimate average marginal effects with based on the conclusion of a Hausman test of the null hypothesis that the random-effects estimator is the efficient estimator. Table 5.1 below describe our Hausman test:

Table 5.1: Hausman Test – Fixed vs. Random Effects

H_0	*Statistic*	*P-Value*	*Conclusion**
Random Effects is efficient	$\chi^2(7) = 13.69$	0.0570	Fail to reject H_0

*At 5% significance level

Since we fail to reject the null hypothesis of the Hausman test at the 5% significance level, we have evidence to conclude that the random effects estimator is the efficient and consistent estimator of our population parameters. **Therefore, we only report the coefficients and interpret the average marginal effects estimated from only the random-effects specification.**

Diagnostic Tests

Since we are conducting a logistic regression, we are not compelled to test for heteroscedasticity as logistic regression is, by design, heteroscedastic given the predictors, and therefore obviates the need to perform any sort of heteroscedasticity testing.

We do, however, explicitly test for serial correlation in the panel data. In order to test for serial correlation we conduct a Woolridge test of the null hypothesis that there is no first-order serial correlation. Table 5.2 below illustrates the results of the Woolridge test:

Table 5.2: Woolridge Test for First-Order Serial Correlation in Panel

H_0	*Statistic*	*P-Value*	*Conclusion**
No first-order serial correlation	$F(1, 50) = 5.420$	0.0240	Reject H_0

*At 5% significance level

Based on the Woolridge test, we reject the null hypothesis at the 5% significance level and we therefore have evidence to conclude that there is serial correlation in the panel data. Therefore, in order to account for this serial correlation present in the data, we use the *autocorrelation-consistent variance-covariance* estimator when estimating the logistic regression under random-effects (using the *vce*(*cluster panelvar*) option in *Stata*).

Summary Statistics

Table 5.3 on the following page provides summary statistics of the variables listed in Table 4.1. Table 5.4 on the following page provides a correlation matrix of the variables listed in Table 4.1.

Table 5.3: Summary Statistics of Variables

Variables	Variation	Mean	Std. Dev.	Min	Max	Observations
$Vote_{ij}$	*Overall*	0.47	0.50	0	1	$N = 255$
	Between		0.48	0	1	$n = 51$
	Within		0.23	-0.33	1.27	$T = 5$
$White_{ij}$	*Overall*	79.31	13.86	24.90	97.26	$N = 255$
	Between		13.64	25.54	95.83	$n = 51$
	Within		2.60	68.84	87.17	$T = 5$
$Black_{ij}$	*Overall*	11.26	11.23	0.31	61.05	$N = 255$
	Between		11.29	0.39	53.90	$n = 51$
	Within		0.84	4.45	18.40	$T = 5$
$Asian_{ij}$	*Overall*	3.60	5.65	0.54	42.27	$N = 255$
	Between		5.66	0.63	39.76	$n = 51$
	Within		0.56	1.61	6.11	$T = 5$
$Female_{ij}$	*Overall*	50.72	0.82	47.40	52.91	$N = 255$
	Between		0.81	47.94	52.76	$n = 51$
	Within		0.17	49.85	51.43	$T = 5$
$Hispanic_{ij}$	*Overall*	9.91	9.66	0.68	48.5	$N = 255$
	Between		9.62	1.11	45.44	$n = 51$
	Within		1.53	4.94	13.65	$T = 5$
$Gini_{ij}$	*Overall*	0.60	0.04	0.53	0.71	$N = 255$
	Between		0.03	0.56	0.67	$n = 51$
	Within		0.02	0.54	0.68	$T = 5$
$Unemp_{ij}$	*Overall*	5.39	1.69	2.10	10.70	$N = 255$
	Between		1.00	3.16	7.16	$n = 51$
	Within		1.36	2.27	9.57	$T = 5$

Table 5.4: Correlation Matrix of Variables

Variables	$Vote_{ij}$	$White_{ij}$	$Black_{ij}$	$Asian_{ij}$	$Female_{ij}$	$Hispanic_{ij}$	$Gini_{ij}$	$Unemp_{ij}$
$Vote_{ij}$	1.00							
$White_{ij}$	-0.15	1.00						
$Black_{ij}$	-0.02	-0.65	1.00					
$Asian_{ij}$	0.34	-0.63	-0.09	1.00				
$Female_{ij}$	0.26	-0.24	0.66	-0.17	1.00			
$Hispanic_{ij}$	0.18	-0.13	-0.11	0.19	-0.15	1.00		
$Gini_{ij}$	-0.01	-0.19	0.17	0.01	0.03	0.41	1.00	
$Unemp_{ij}$	0.15	-0.26	0.28	0.00	0.16	0.20	0.33	1.00

Regression Results

Table 5.5 below reports the estimated logistic regression coefficients for the random-effects specification using the autocorrelation-consistent variance-covariance estimator to generate robust standard errors.

Table 5.5: Random-Effects Logistic Regression Output

VARIABLES	*Estimated Coefficients*	*Robust Std. Errs.*	*P-Values*
$White_{it}$	0.083	0.145	0.565
$Black_{it}$	-0.249	0.167	0.136
$Asian_{it}$	2.419***	0.491	0.000
$Female_{it}$	6.104***	1.233	0.000
$Hispanic_{it}$	0.015	0.062	0.811
$Gini_{it}$	-5.691	13.605	0.676
$Unemp_{it}$	0.593***	0.178	0.001
Constant	-320.461***	63.187	0.000
Observations	255		
$\chi^2(7)$	47.51		
$P > \chi^2(7)$	0.000		

*** = significant at 1% significance level

According to the table above, the p-value of the chi-squared statistic $\chi^2(7) = 47.51$ is 0. Therefore, even at the 1% significance level, we reject the null hypothesis that our independent variables of interest are jointly statistically insignificant.

Of our independent variables of interest, only $Asian_{it}$, $Female_{it}$, and $Unemp_{it}$ have estimated coefficients that are statistically significant at the 1% level. The sign of these coefficients are in-line with our theoretical hypothesis; that is, each of these coefficients are positive. Hence we have evidence to support our hypothesis that an increase in the female and Asian proportion of the state population is associated with an increase in the probability of the given state voting for the Democratic Presidential candidate. We also have evidence to conclude that an increase in the unemployment rate in the state is associated with an increase in the probability of the given state voting for the Democratic Presidential candidate.

The coefficients on $White_{it}$, $Black_{it}$, $Hispanic_{it}$, and $Gini_{it}$ are not statistically significant and therefore go against our hypothesis. We believe, however, that including a greater number of Presidential elections in our analysis would have helped to improve coefficient significance. Indeed, we considered including earlier elections in our analysis but it became prohibitively difficult and time-consuming to find all of the appropriate data. We also note that there is very little variation in $Gini_{it}$ across the entire panel, and we believe that including data from elections in the 1900s would have given a more accurate representation of how inequality affects voting outcomes. The necessary datawork required to undertake our analysis across a longer time-span of Presidential elections could be implemented in future versions of this research.

In order to provide an interpretable estimate of how $Asian_{it}$, $Female_{it}$, and $Unemp_{it}$ affect the probability of the state voting for the Democratic Presidential candidate, we calculate the average marginal effects that each variable has on the probability. Table 5.6 below reports the **average marginal effects** predicted from the logistic regression coefficient estimates in Table 5.5.

Table 5.6: Average Marginal Effects

VARIABLES	*Marginal Effect*	*Robust Std. Errs.*	*P-Values*
$White_{it}$	0.005	0.009	0.563
$Black_{it}$	-0.016	0.011	0.134
$Asian_{it}$	0.155***	0.018	0.000
$Female_{it}$	0.392***	0.056	0.000
$Hispanic_{it}$	0.001	0.004	0.814
$Gini_{it}$	-0.365	0.891	0.682
$Unemp_{it}$	0.038***	0.011	0.001
Observations	255		

*** = significant at 1% significance level

According to our point estimates of the average marginal effect that each statistically significant independent variable has on the probability of a given state voting for the Democratic Presidential candidate, we can make the following three conclusions:

i) on average, a one percentage point increase in the proportion of Asians in a given state population is associated with a 15.5 percentage point increase in the probability of the state voting for the Democratic Presidential candidate;

ii) on average, a one percentage point increase in the proportion of females in a given state population is associated with a 39.2 percentage point increase in the probability of the state voting for the Democratic Presidential candidate; and,

iii) on average, a one percentage point increase in the unemployment rate of a given state is associated with a 3.8 percentage point increase in the probability of the state voting for the Democratic Presidential candidate.

VI. Conclusion

In summary, we hypothesized that the probability of a given state voting for the Democratic Presidential candidate is positively related with the proportion of minority populations of that state, positively related with income inequality in that state, and positively related with unemployment in that state. After estimating our model using a panel logit under random effects, we have found evidence that out of the various minority groups that we have considered in our analysis, only the proportion of females and Asians in a given state have a statistically significant, positive relationship with the probability of voting for the Democratic candidate for the given state. We also have found evidence that state unemployment is positively related with the probability of voting for the Democratic candidate. We find no evidence, however, to conclude that greater state-level income inequality makes the state more likely to vote for the Democratic Presidential candidate.

We believe that the discrepancies between some of our results and our initial hypothesis may be resolved by the inclusion of additional election years in our analysis. This is one avenue for further research. Alternatively, future research avenues may seek to include additional explanatory variables that model more nuanced features of the economic climate of a given state in a given election year – such as inflation, per-capita income growth, and the proportion of state production derived from manufacturing.

References

[1] Carsey, T.M. and Wright, G. C. (1998), State and National Factors in Gubernatorial and Senatorial Elections. *American Journal of Political Science*, Vol. 42, No. 3 (Jul., 1998), 994-1002.

[2] Galbraith, J. K. and Hale, J. T. (2008), State Income Inequality and Presidential Election Turnout and Outcomes. *Social Science Quarterly*, 89: 887901.

[3] Wright, J. (2012). Unemployment and the Democratic Electoral Advantage. *The American Political Science Review*, 106(4), 685-702.

[4] US Federal Election Commission. *Federal Elections.* (2017). https://transition.fec.gov/pubrec/electionresults.shtml

[5] United States Census Bureau. *2000-2010 State Characteristics Intercensal Population Estimates.* (2012). https://www.census.gov/data/datasets/time-series/demo/popest/intercensal-2000-2010-state.html

[6] United States Census Bureau. *American Community Survey.* (2016). https://www.census.gov/programs-surveys/acs/news/data-releases.2016.html

[7] Frank, Mark W. *U.S. State-Level Income Inequality Data*. (2015). http://www.shsu.edu/eco_mwf/inequality.html

[8] Federal Reserve Bank of St. Louis. *State Employment and Unemployment.* (2018). https://research.stlouisfed.org/pdl/337