Econ 8710. Advanced Economic Analytics for Business

Assignment 2. Elections

Objective. We are interested in learning how different voter characteristics or features are related to voters choices in presidential elections.

Data. The data is a series of election surveys from 1990 to 2004 by the American National Election Studies. These surveys collect socioeconomic, demographic characteristics, political inclinations and voting patterns before and after Congressional and Presidential elections. The data only contains observations where information after the election is not missing. Use **elections.dta** on the assignment data folder. Use the codebook for the exact coding of the variables.

- **Dependent variable.** Create a party of presidential vote using with (VCF0704A), properly converted to a dummy variable.

$$y_i = \begin{cases} 1 & \text{if } i \text{ voted democrat} \\ 0 & \text{if } i \text{ voted republican} \end{cases}$$

- **Explanatory variables.** Choose **demographic** variables that you believe might be related to vote. Some variables only are available for some years. Note that the numeric missing values (e.g., -9, 99). The choice of variables in the model should take into account how many observations are "lost" by the inclusion of variables with missing values. Make an specific choice of the omitted category for each categorical variable set. Be careful to overfitting the model, i.e. choosing variables that only work on the dataset used to estimate the model do not generalize well.

- **Data Cleaning.** In general you want to either mark the observations to use or drop observations that are never going to be used. Below are some suggestions.

  - Keep only observations from years with a presidential election.
  - Coding of variables. Be careful about different coding of missing values.
  - Missing observations. Drop missing observations of the dependent variable. In general, you should not drop missing observations of missing explanatory variables as that will inform which variables to include.
  - Rename the variables to use (or use clonevar command to create new variables).
  - Label variable and the label values of the variables you are going to use.

Report

Linear Probability Model. The objective is to estimate a linear regression model (OLS) for the 1990 to 2004 presidential elections.

1. A first step in any analysis is to get a sense of the data. **Report the summary statistics of all the variables included in the regression.**

   (a) If the variable is continuous, e.g. age, report the mean, standard deviation, minimum, maximum and number of observations.

   (b) If the variable is categorical, e.g. race, report the distribution of the variable. A good idea is to create dummy variables for every categorical variable ((e.g. race, religion, etc.; you will need them in the next step as well). Report mean, standard deviation, minimum, maximum and number of observations. Notice that the mean of a dummy variable is the proportion of people in that category.

2. Interpret the coefficients of the regression. Specifically, I want to know what is the relationship between demographic characteristics and the political inclination of voters.

3. Interpret the regression as a whole, F-statistic and R^2.

4. Repeat your estimation in part (3) including a dummy variable for each year of the sample (i.e. if the observation is for 1994, $D_{1994} = 1$, and 0 otherwise). Omit the dummy for 2000. Interpret only the coefficients for 1996 and 2004.

5. Obtain the predicted y vector for your sample $\hat{y}$. Since y only takes values of 1 or 0, what is the interpretation of this prediction? What potential issues do you observe?

6. Perform analysis of residuals by obtaining the regression residual plots (histogram, residuals versus fit). What potential issues do you observe?

7. **Identification** Are we estimating the causal effect of these variables? What are the shortcomings with this estimation. What methodology do you suggest?

Variable	Description		Variable	Description
VCF0004	Year of Study		VCF9001	Length of Residence in Community
VCF0101	Respondent Age		VCF9002	Length of Residence in House
VCF0102	Respondent Age Group		VCF9009	Strength - Approve President Perfor
VCF0103	Respondent Cohort		VCF9013	Should Society Insure Equal Opportu
VCF0104	Respondent Gender		VCF9014	Have Gone Too Far Pushing Equal Rig
VCF0106A	Respondent Race 6-category		VCF9015	Big Problem If Chances Not Equal
VCF0110	R Education 4-category		VCF9016	Ok If Some Have More Equal Chance
VCF0111	Urbanism		VCF9017	Should Worry less about Equality
VCF0112	Census Region		VCF9018	Fewer Problems If More Equal Treatm
VCF0113	Political South/Nonsouth		VCF9026	How Did R Vote in Presidential Prim
VCF0114	Family Income		VCF9027	Party of Vote in Previous Pres Elec
VCF0115	Respondent Occupation Group [1 of 2]		VCF9028	Which Pres Candidate Will Carry Sta
VCF0116	R Work Status 7-category		VCF9029	Will Pres Elction Be Close in State
VCF0118	R Work Status 5-category		VCF9030	▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯
VCF0127	Household Union Membership		VCF9030A	Did Major Party Contact R
VCF0128B	Religion of R 8-category		VCF9030B	Did Democratic Party Contact R
VCF0138	Number of Children in Household		VCF9030C	Did Republican Party Contact R
VCF0142	▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯▯		VCF9031	Did Anyone Besides Parties Contact
VCF0143	R Parents Native Born		VCF9032	Radio Programs about Campaign -Numb
VCF0146	Home Ownership by R Family		VCF9035	How Many Days R Watched National TV
VCF0147	Marital Status of R		VCF9036	R Know Senate Majority Before Elect
VCF0148	Respondent Social Class 8-category		VCF9037	Should Govt Ensure Fair Treatment -
VCF0148A	Respondent Social Class 6-category [1 of 2]		VCF9039	Conditions Make it Difficult for Bl
VCF0149	Respondent Social Class 6-category [2 of 2]		VCF9040	Blacks Should Not Have Special Favo
VCF0150	Respondent Work Status 16-category		VCF9041	Blacks must Try Harder
VCF0151	Respondent Occupation Group [2 of 2]		VCF9042	Blacks Gotten less than They Deserv
VCF0152	Religion of R Full Code [2 of 2]		VCF9045	Is Position of the U.S. Weaker/Stro
VCF0154B	R Occupation 14-category		VCF9048	Space/Science/Techn - Federal Spend
			VCF9049	Social Security - Federal Spending
VCF0702	Did R Vote in Election		VCF9055	Type of Senate Race
VCF0704	Party of Pres Vote- All Major Candi			
VCF0704A	Party of Pres Vote- 2 Major Parties			
VCF0705	Party of Pres Vote- Major Parties,			
VCF0706	Party of Pres Vote and Nonvote			
VCF0707	Party of Vote for Congressman			
VCF0708	Party of Vote for Senator			