Predictive Analytics
Homework 3

Due August 6

Your submission should consist of a neatly formatted PDF report and Stata do files that document what you did for each problem.

1 Minimum Wage (30 Points)

Use the dataset *minwage.dta* for this problem. This dataset comes from the Card and Krueger study of minimum wage and employment. Between February and November, New Jersey increased its minimum wage, while Pennsylvania did not.

1.a

Use a differences-in-differences regression to estimate the effect of the minimum wage change on total employment. We define total employment as the sum of full time employees, managers, and 1/2 times part-time employees. Report and explain the meaning of the results.

1.b

Construct a graph showing the change in total employment for both states and the counterfactual. Indicate the DD estimator on your graph.

1.c

Use a differences-in-differences regression to estimate the effect of the minimum wage change on the price of a full meal (entree, fries, and soda). Report and explain the meaning of the results.

1.d

Construct a graph showing the change in the price of a full meal for both states and the counterfactual. Indicate the DD estimator on your graph.

2 Airlines (35 Points)

Use the dataset *airfare.dta* for this problem. Each observation is a passenger air route between a pair of cities in a particular year. Consider the model

$$log(fare_{it}) = \beta_0 + \beta_1 concen_{it} + \delta_{98} y98_t + \delta_{99} y99_t + \delta_{00} y00_t + a_i + \varepsilon_{it}$$

where *concen* is the market share of the largest carrier on the route and $y98, y99, y00$ are dummies for years 1998, 1999, and 2000.

2.a

Estimate the model using pooled OLS (ignoring unobserved route heterogeneity) and report the results.

2.b

Estimate the model using fixed effects with the within estimator and report the results.

2.c

Estimate the model using first differences and report the results.

2.d

Discuss the meaning of your estimate for β_1. How have your three estimates of this parameter differed? Do any of these estimates call into question the model that produced it? Give an economic explanation for the sign of this variable.

2.e

Explain two characteristics of a route that you think are captured by a_i. Would these have contributed to *concen* being endogenous?

3 Elections (35 Points)

Use the dataset *voting.dta* for this problem. The data in this problem come from the 1988 U.S. House of Representatives elections.

3.a

Use the linear probability model to estimate the probability of candidate A winning the election using the model

$$win = \beta_0 + \beta_1 expendA + \beta_2 prtystrA + \beta_3 democA + \varepsilon$$

Report the results and discuss the interpretation of your estimate of β_1.

3.b

Explain why not including $expendB$ in the model might bias your estimate of β_1. Be as specific as possible about the mechanism of the bias.

3.c

Add $expendB$ to the linear probability model and report the estimates. How has your estimate of β_1 changed from the previous model?

3.d

Use a probit regression to estimate the probability of candidate A winning. Include $expendA$, $expendB$, $prtystrA$, and $democA$ as explanatory variables in the model. Report the results.

3.e

Estimate the change in win chance due to spending an additional \$50,000. Use the sample means of the explanatory variables as the starting point.