**Assignment #1**

**Using data on taxicabs to think about selection and treatment effects**

Imagine you are interested in how technology adoption influences firm boundaries. You are aware of a survey of taxicab firms that contains data on both of these phenomena. You contact the scholar who published the survey and she kindly gives you access to the data. The file is called: taxi\_phd.dta

1. Look at the summary stats, cross tabs and kernel density plots. What do notice about the relationship between computerized dispatching (evercmptr==1) and firm size (logtaxis and logtaxisq) and demographics (by msa): population density (logdensity), regional dummies (southdum, northeastdum, midwestdum), and , median income in 1989 (medianinc89), percentage of workers who take transit to work (worktransit).

2. You go to a taxi industry trade show where you talk to some technology vendors. They say they have the hardest time selling systems in the Northeast and the South. The Northeast has older cities with idiosyncratic streets and traffic patterns and busy taxi traffic in the city center. Hails are common and computerized dispatching is not much more effective than traditional radio. They characterize the South as a friendlier environment where the anonymity of computerized dispatching is a turnoff. Verify whether these factors appear to influence adoption of computerized dispatching by running a logit that regresses computerized dispatching on all the observables. After finding the marginal effects of the regressors, interpret the magnitude and significance of the key variables that appear to drive selection.

1. Now let’s think about treatment effects. Regress driver ownership rate (%, downrt) on computer adoption (evercmptr) controlling for measures of firm size (logtaxis and logtaxisq).

a. What is the interpretation of the coefficient on computer adoption?

b. What happens if you use robust standard errors? How about robust and clustered standard errors, clustering at the msa-level or county-level (fips). Why doesn’t clustering matter much in this context?

c. Try including MSA fixed effects. What is the coefficient estimate now? How important is local market variation in explaining the cross-section of driver ownership rates?

d. Interpret the effect of computer adoption on firm boundaries if adoption is random.

e. Why might adoption be non-random (endogenous) with respect to firm boundaries?

4. Assume adoption and boundary decisions are jointly determined, but that the region the firm operates within has no direct effect on asset ownership decisions. Run a 2SLS regression with computer adoption as the endogenous regressor using regional dummies as instruments. You also believe that computerized dispatching is more valuable when people commute via public transportation to work, and need reliable “last mile” service on the job and to get to/from transit hubs to work locations, and yet transit use would seem to satisfy the exclusion restriction, so you include worktransit as an instrument as well.

a. How strong are the exogenous regressors in the first-stage? Look at the t-statistics on the instruments, the R-squared of the first stage and the F-test for weak identification. What do these statistics tell us about the instruments?

b. What is the interpretation of the second-stage coefficient estimate on computer adoption? How does that compare to the OLS estimates?

c. Given the data we have, but can we say about the impact of computer adoption on firm boundaries? What would the ideal experiment looks like? How could we get closer to the ideal?