MACS 30150 Problem Set 1

Daniel Tracht

January 14, 2019

1 Classify a model from a journal

For this section, I consult a recent paper by Carrell, Hoekstra, Kuka, (Carrell, Scott E., Mark Hoekstra, and Elira Kuka. 2018. "The Long-Run Effects of Disruptive Peers." *American Economic Review*, 108 (11): 3377-3415.), in which they present a model a statistical model that seeks to examine the effect of having disruptive classmates using linkages to domestic violence as a method to remove confounding variables. In a single equation, their model is

$$y_{igst} = \theta_0 + \theta_1 \frac{\sum_{k \neq i} DV_{kgst}}{n_{gst} - 1} + \theta_2 X_{igst} + \lambda_{gs} + \sigma_{gt} + \varepsilon_{isgt}$$

where i, g, s, and t represent respectively the individual, grade, school, and academic year. Furthermore, y represents the outcome variables of interest (test scores, college enrollment, college graduation, labor force participation, and earnings), $\lambda$and σ are the grade-school and grade-year fixed effects. The coefficient of interest for the authors is θ_1 which is the coefficient on the proportion of peers from families linked to domestic violence, excluding children who are themselves linked to domestic violence. The vector X is additional controls for both individuals and cohorts and has a coefficient θ_2.

In this model, exogenous variables include the entire right-hand side of the equation. In this model, students do not choose their peers or any of the controls in X, nor do they choose the coefficients. Endogenous variables are the outcome variables of interest.

While this model is estimated over periods, agents are not making choices across them, and so this model may well be considered static. If previous outcomes were added to the controls, we might classify this as dynamic. This model is linear in terms. This model is deterministic, when assuming that ε represents just the unobserved variables that effect the outcome of interest.

Missing from this model would be a teacher fixed-effect. This should simple enough to harvest from the data and so would be feasible to include in the model. It would naturally fit with the grade-school and grade-year fixed effects and would not introduce a fixed-effect approach where it did not exist before. It would not be unreasonable to think that there would be significant relationship between outcomes and particular teachers who might be remarkably effective.

2 Make your own model

I want a model of whether an agent decides to get married. At the very least, one endogenous choice in the model is $GetMarried$ which takes a value of 1 if the agent decides to get married and a value of 0 if the agent decides not to get married. To keep this under a page, I make some extremely strong assumptions and work with a reduced form linear probability model that can easily be taken to data:

$$\begin{aligned} ProbGetMarried_{i,t} =& \beta_0 Married_{i,t} + \beta_1 logMarketSize_t + \beta_2 Age_{i,t} + \beta_3 logIncome_{i,t} \\ &+ \gamma_0 Married_{j,t} + \gamma_2 Age_{j,t} + \beta_4 DivorceCost_{i,t} + \varepsilon_{i,j,t} \end{aligned}$$

where

$$GetMarried_{i,t} = \begin{cases} 1 & ProbGetMarried_{i,t} \geq \frac{1}{2} \\ 0 & ProbGetMarried_{i,t} < \frac{1}{2} \end{cases}$$

The key factors of the choice are

1. Whether or not the agent and the potential partner is already married, as polygamy is illegal in the United States and divorce is costly
2. The size of the marriage market, which gives a rough sense of the outside options available to the agent
3. The age of the agent and the potential partner, as humans still have finite lives and so any forward-looking agent would consider the time left with that partner until T is hit
4. The income of the agent, as the benefits of marriage that I might want to see in a utility function can be more easily acquired on the open-market with greater income
5. The cost of leaving the marriage

These factors preferred to others for several reasons. One is the relative ease with which information on these variables can be obtained. Marriage records, demographics about locations, age, and income can all be obtained from administrative records, sidestepping the issue of survey data. The most difficult factor to pin down will be the cost of leaving the marriage, but will be an extremely important for this model. From the literature, making no-fault divorce legal in the United States significantly reduced the number of suicides for married women (see Stevenson and Wolfers, QJE, 2006).

I left out some of the other demographic variables, such as race and education, which might be used to define the scope of homophily and the bounds of the marriage market. While these factors are likely important in partner formation, that is dating, I don't believe them to be important for marriage conditional on already dating.

As for preliminary tests, taking this model to the administrative data would allow us to see significant and possible effect sizes.