MACS 30150
Problem Set 1
Shuyi Yang
Jan, 11th, 2020

1.

a) A paper from the latest issue of American Economic Review: *Estimating the Production Function for Human Capital: Results from a Randomized Controlled Trial in Colombia*

b) Citation: Attanasio, Orazio, Sarah Cattan, Emla Fitzsimons, Costas Meghir, and Marta Rubio-Codina. 2020. "Estimating the Production Function for Human Capital: Results from a Randomized Controlled Trial in Colombia." American Economic Review, 110 (1): 48-85. DOI: 10.1257/aer.20150183

c) The basic idea of the model is pretty much an AR1 on θ_t, which is a child's skill at period t:

$$\theta_{t+1} = f_{t+1}(\theta_t, I_{t+1}, P_t, X_t, \eta_{t+1})$$

where I_{t+1} are parental investments occurring between the two periods, P_t the maternal skills and X_t other arbitrary household characteristics. eta_{t+1} is the error term. So the model shows how a child's skill grows temporally given various conditions and thus the human capital is being built up.

The actual statistical model is based on the assumption that the production function is Cobb-Douglas, and therefore applies a log-linear model:

$$\ln(\theta_{idt+1}^k) = A_d^k + \gamma_{1d}^k \ln(\theta_{it}^C) + \gamma_{2d}^k \ln(\theta_{it}^S) + \gamma_{3d}^k \ln(P_{it}^C) + \gamma_{4d}^k \ln(P_{it}^S)$$

$$+\gamma_{5d}^k \ln(I_{idt+1}^M) + \gamma_{6d}^k \ln(I_{idt+1}^T) + \gamma_{7d}^k \ln(n_{it+1}) + \eta_{it+1}^k, \ k \in \{C, S\}$$

where C, S respectively denote cognitive and social-emotional skills, d indexes choice of measurement as the researchers devised multiple measures for skills, M, T show maternal and time investments respectively, and i certainly indexes the observations.

d) All these observable or measurable ones: $\theta_{t+1}, \theta_t^C, \theta_t^S, I_{t+1}^M, I_{t+1}^T, P_t^C, P_t^S, X_t$ are exogenous variables; the endogenous variables are: $\gamma_1, \ldots, \gamma_7, \eta_{t+1}$

e) This is a rather simple linear and deterministic model. No randomness is inputted into the model.

f) This model was made quite inclusive; essentially any additional variables can be added into X_t the vector of variables that shows all other parenting characteristics. In this paper the

authors focused on a program conducted in Columbia so they picked certain variables unique in this country. For instance, there are "conflict" and "terrorism" alongside ubiquitous ones such as medical treatment, toy price and food price. Based on the specific region that one applies the model to, the choice of the additional variables can vary greatly.

2.

a) b) c) Consider that marriage is a significant decision for any person's entire life, we can safely assume that usually such decision is not made whimsically. Therefore, rather than simply determining endogenous variable *get married* or *not get married*, we use if one is going to marry in the next period, where a period is six month. So clearly this is going to be a dynamic model which builds itself upon some past records of marriage. Some potentially influential factors leading to marriage in six month include one's economic characteristics such as income, housing or property status, and career prospects, and family characteristics such as parental pressure, sibling situations, and whether having a dating partner (there are some chances that one can marry even without having dated the companion). There are also personal characteristics reflecting the person's outgoingness, attitude towards marriage and parenting. The data generating process is to survey people that just got marries about these above mentioned variables. Since this is just one six-month period ago, we hope to safely assume that these traits, such as income/career, or attitude towards marriage did not change greatly over such a short period of time. Most of these variables are easy to quantify into float or categorical variables. Of course, the accuracy of the prediction will have to depend on the respondents' honesty in answering these somewhat personal questions. All these surveyed variables as the respondents marital status will be initially inputted as exogenous variables and the endogenous output are the effects of these variables. Since the marry or not is a binary choice, we can run a linear logistic regression model of all these variables. Then giving new exogenous variables of currently unmarried people, we can use these results to predict if they would marry in six month.

d) e) In fact, all of these three category of factors have huge impacts on the decision of marriage. The economic characteristics determine one's ability to get married; the family characteristics determine one's social urgency to get married; and finally the personal characteristics determine one's own willingness to get married. For a rational personal, all three categories are necessary but insufficient conditions. However, if one category has to be picked for being more key than others, it has to be the personal characteristics. One may

always choose to marry, against his economic interest, defying family expectations, just to satisfy personal preference. Still, frankly this category is least objective and thus ambiguous to measure; therefore using all three types of factors altogether is quite inclusive and is more predictive.

f) We can simply run a test with the original generated data, which comes from real life in the first place. Using the familiar train-test split technique, we can generate the model and test it with ease.