# Problem Set #1

Yongfei Lu

Jan.13, 2019

## 1. Classify a model from a journal (5 points).

(b) 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)

$$ln(\theta^k_{idt+1}) = A^k_d + \gamma^k_{1d} ln(\theta^C_{it}) + \gamma^k_{2d} ln(\theta^S_{it}) + \gamma^k_{3d} ln(P^C_{it}) + \gamma^k_{4d} ln(P^S_{it}) + \gamma^k_{5d} ln(I^M_{idt+1}) + \gamma^k_{6d} ln(I^T_{idt+1}) + \gamma^k_{7d} ln(n_{it}) + \eta^k_{it+1}, k = C, S$$

The production function for human capital regresses log child development (2 dimensions: cognitive and socio-emotional skills) in period t+1 on log initial conditions of child development and mother's (cognitive and socio-emotional) skills, log parental investment of materials and time, and log number of children in the household. The model aims to prove that a randomized early childhood intervention in Colombia significantly improves the cognitive and socio-emotional skills of a sample of disadvantaged children aged 12 to 24 months at baseline.

$$ln(\theta^k_{idt+1}) = A^k_d + \gamma^k_{1d} ln(\theta^C_{it}) + \gamma^k_{2d} ln(\theta^S_{it}) + \gamma^k_{3d} ln(P^C_{it}) + \gamma^k_{4d} ln(P^S_{it}) + \gamma^k_{5d} ln(I^M_{idt+1}) + \gamma^k_{6d} ln(I^T_{idt+1}) + \gamma^k_{7d} ln(n_{it}) + \eta^k_{it+1}, k = C, S$$

Parental investment model regresses log investment on log child and mother's baseline skills, log number of children in the household, and a vector of variables (including log toy and food price, maternal childhood exposure to conflict). This model, on the other hand, aims to isolate the mechanism of how the intervention improves children's development, which is an increase in parental investment of time and materials.

(d)

**Production Function for Human Capital**

**Exogenous:**

$A^k_d$: a factor-neutral productivity parameter or TFP

$ln(\theta^C_{it}), ln(\theta^S_{it})$: the stock of cognitive and socio-emotional skills of child i in period t

$ln(P^C_{it}), ln(P^S_{it})$: the stock of cognitive and socio-emotional skills of mother i in

period t

$ln(I^M_{idt+1}), ln(I^T_{idt+1})$: material and time investments in period t+1 by parent i

$ln(n_{it})$: number of children in the household in period t

$\eta^k_{it+1}$: unobserved shocks in period t+1

**Endogenous:**

$ln(\theta^k_{idt+1})$: the stock of cognitive or socio-emotional skills of child i in period t+1

**Parental Investment model**
**Exogenous:**

$ln(\theta^C_{it}), ln(\theta^S_{it})$: the stock of cognitive and socio-emotional skills of child i in period t

$ln(P^C_i), ln(P^S_i)$: mother's stock of cognitive and socio-emotional skills

$u^\tau_{it+1}$: number of children in the household in period t

$Z_{it}$: a vector of variables, including prices of food and toy, and Maternal childhood exposure to conflict

**Endogenous:**

$ln(I^\tau_{idt+1})$: material ot time investments in period t+1 by parent i

(e) the two models are dynamic, linear and deterministic
(f) As studies have shown that gender discrimination can exist in families, especially in disadvantaged areas and that returns to human capital investment differ among males and females, gender difference can lead to different production functions for human capital and impact parents' investment behaviors. Therefore, the model is missing child gender that might be valueable.

## 2. Make your own model (5 points).

(a) I will use a nonlinear probability model to explain whether someone decides to get married:

$$getmarried = \beta_0 + \beta_1 age + \beta_2 educ + \beta_3 gender + \beta_4 fincome + \beta_5 workinghours + \beta_6 orientation + \beta_7 lawsupport + \beta_8 genderratio + \beta_9 housingprice + \beta_{10} favorablepolicy + \beta_{11} orientation * lawsupport$$

**Endogenous:**
getmarried: probability of individual i getting married
**Exogenous:**
age, educ, gender, fincome, workinghours: age, highest education degree, family

income, average weekly working hours of individual i, numerical
genderratio, housingprice: gender ratio of the country, average house price of the city where individual i lives, numerical
Orientation, lawsupport, favorablepolicy: sexual orientation of individual i, law support of getting married and divorced, favorable policy for getting married (like tax reduction for married couples), dummy variables
orientation * lawsupport: for individual i with certain sexual orientation, whether his/her country supports their desired marriage (some countries don't support same-sex marriage), interaction term

(b) $y_i = 1, getmarried > 0.5$
$y_i = 0, otherwise$

(c) Data generating process: with the model above and data for all the exogenous variables in the model, we can predict the values for the endogenous variables getmarried and $y_i$.

(d) Key factors: while all the independent variables can exert influence on one's decision to get married, I think age, education, family income and house price are especially important factors (in particular in Asian countries).

(e) According to relevant literature, my observation and intuition, if a country owns a population with significantly biased gender ratio, its housing price is too high (which means high cost of marriage), and the individual living there is not financially attractive enough, he would probably lose in the marriage market and postpone his marriage. Also, studies have shown that people with higher educational degrees tend to get married later than those with lower educational degrees.

(f) Preliminary test: I would extract the data for the independent variables in my model from the national censuses, and search the information on law and policy I need for the dummy independent variables. After preliminarily cleaning the data and form the dataset, I would run the regression model to see whether these variables are significant or not.
Or from a statistical view, I can randomly divide the sample into two parts, training and testing datasets. I would run the regression using the training dataset and then check the accuracy of the predicted outcome using the testing dataset.