Problem Set #1

MACS 30100, Dr. Evans
Xinzhu Sun(Clara)

Problem 1

Part (a). *All for One? Family Size and Children's Educational Distribution under Credit Constraints:* This article examines the possibility that a child's schooling can be increased by having more siblings, instead of being diminished by competition for parents' resources: if unable to finance the education of their younger children, parents may do so through their older children's labor income. The authors examine this possibility in a model combining convex returns to education and credit constraints. This model predicts correlations among family size, years of schooling and birth order, which would not exist when either of these two elements is absent. Empirical patterns shown in the United States, Mexico, and South Korea support the model predictions.

Part (b). Lafortune, Jeanne and Soohyung Lee. 2014. "All for One? Family Size and Children's Educational Distribution under Credit Constraints." American Economic Review, 104(5): 365-69. DOI: 10.1257/aer.104.5.365

Part (c). First, the authors examine the relationships among family size, education, birth order and years of schooling in a model combining convex returns to education and credit constraints.

Parents j with n_j number of children value their children's schooling as:

$$\sum_{k=1}^{n_j}(h(S_{kj}) - (1 + c_{kj})S_{kj}) \tag{1}$$

Where k denotes birth order(1 for the eldest child, and n_j for the youngest), S_{kj} denotes years of schooling of the k^{th}, c_{kj} is an idiosyncratic utility cost of schooling that is assumed independent of birth order. Function h governs the returns to schooling, assumed such that $h(0) = 1$, $h' > 0$, and $h''(S) > 0$ for any $0 \leq S < a$ and $h'' < 0$ for any $S \geq a$.

Parents cannot borrow to finance the education of their children. The monetary cost of schooling, normalized as 1 per unit, must be financed out of parental assets A_j and labor earnings of siblings working as youth at a wage rate of w_k when they are not attaining the maximum education level $\overline{S}$. The wage rate is assumed to decrease in birth order. Then the budget set is

$$\sum_{k=1}^{n_j} S_{kj} \leq A_j + \sum_{k=1}^{n_j} w_k(\overline{S} - S_{kj}) \tag{2}$$

Unconstrained parents will select S^*_{kj} such that $h'(S^*_{kj}) = 1 + c_{kj}$. When parents are credit constrained, the first- and second-order conditions ensure that for any two children k and k', receiving a positive education, we must have $h''(S^*_{kj})$, $h''(S^*_{k'j}) < 0$,

$$\frac{h'(S^*_{kj}) - (1 + c_{kj})}{h'(S^*_{k'j}) - (1 + c_{k'j})} = \frac{1 + w_k}{1 + w_{k'}} \tag{3}$$

and that the monetary cost of their education is equal to resources:

$$\sum_{k=1}^{n_j} S_{kj}(1+w_k) = A_j + \overline{S}\sum_{k=1}^{n_j} w_k \tag{4}$$

Depending on asset level, parents' schooling decision follows one of the three regimes: (i) For assets above $A^*(n_j)$, parents are unconstrained and select S^*_{kj} such that $h'(S^*_{kj}) = 1 + c_{kj}$. (ii) For parents with assets below $A^*(n_j)$ but above $\widetilde{A}(n_j)$, all children receive a positive investment driven by the first-order conditions(3). (iii) For the rest of parents, at least one child acquires no schooling.

Then they examine the model implications using datasets from the United States, Mexico and South Korea. First, they empirically test the relationship between family size (n_j) and the years of schooling of the most educated child in a family (y_j) as follows:

$$y_j = \alpha n_j + \beta A_j + \gamma n_j \times A_j + \rho X_j + \epsilon_j \tag{5}$$

where A_j is proxied with the father's educational attainment. In the model, $\alpha > 0$ but $\gamma < 0$; that is, for low income families, family size has positive impact on the most educated child's schooling, but that impact decreases as A_j increases. Additional control variables (X_j) include the cohort of the first child born interacted with A_j to capture differential schooling trends over time by father's educational level.

Next the authors use the following equation to examine the birth order effect:

$$y_{kj} = \alpha k + \beta A_j + \gamma k \times A_j + \rho X_{kj} + \epsilon_{kj} \tag{6}$$

where y_{kj} measures the educational attainment of the k^{th} birth order child in family j, and X_{kj} includes a gender dummy in addition to the controls mentioned previously. In the model, $\alpha > 0$ but $\gamma < 0$. The standard errors are clustered at family level.

Finally, to study the gender effect, we examine not only birth order but also siblings' gender composition within a similar frame work.

Part (d).

- Endogenous variables:
  Family size effect (years of schooling of the most educated child in a family): y_j
  Birth order effect (the educational attainment of the k^{th} birth order child in family j): y_{kj}
  Gender effect

- Exogenous variables:
  Equation(5) :family size n_j ; parental assets A_j, which is proxied with the father's educational attainment; additional control variables (X_j)
  Equation(6): birth order k ; parental assets A_j, which is proxied with the father's educational attainment; additional control variables X_{kj} which includes a gender dummy
  Gender effect: gender dummy g ; birth order k ; parental assets A_j, which is proxied with the father's educational attainment; additional control variables X_j

Part (e).

- Static: This model describes a static relationship between the variables; there is no time dimension in the model
- Linear: all empirical regressions are in linear formation
- Stochastic: there is an error term following $N(0, \sigma^2)$ in all models

Part (f) I think the model could potentially be more powerful if it could include a vector of characteristic variables of the social environment/policy of the a specific country. That is, for those middle-income country like Mexico the human capital investment decision may display different pattern from South Korea and USA. For example, different countries may have different years of compulsory education, different financial support out of family and different valuation towards education. Thus, including a social environment variable may be reasonable.

Problem 2

Part (a)(b)(c).

My model of how long popular musicians live is :

$$\begin{aligned} Predicted\,Lifespan =& \beta_0 + \underbrace{\beta_{1,1}Parents'\,lifespan + \beta_{1,2}Gender +}_{\text{genetic}} \\ & \underbrace{\beta_{2,1}Race + \beta_{2,2}Years\,of\,education + \beta_{2,3}Weight/Height +}_{\text{lifestyle}} \\ & \underbrace{\beta_{3,1}Married\,or\,not + \beta_{3,2}Num\,of\,children +}_{\text{family}} \\ & \underbrace{\beta_{4,1}Year\,of\,born + \beta_{4,2}Nationality +}_{\text{social influence}} \\ & \underbrace{\beta_{5,1}Income + \beta_{5,2}Music\,genre}_{\text{career}} + \epsilon \end{aligned} \tag{7}$$

Part (d). The key factors that influence the outcome are genetic reasons (Parents' lifespan, Gender), lifestyle reasons (Race, Years of Education and Weight-height ration) and career reasons(Music genre). This is because: 1) Parents' lifespan is a good indicator of genetic disease that is inheritable and women tend to live longer than men; 2) Years of education and Race are a good proxy for self-discipline. They can indicate whether the musician is drug abuse/alcohol abuse/smoking in some degree ; 3) weight and height ratio is a good proxy for exercise and diet; 4) Music genre also indicates the musicians lifestyle and stress level. Country musicians tend to live longer than rock musicians.

Part (e)

Besides the reasons I stated in part(d), the reasons that I choose other variables are:1) the family reasons reflect the happiness of the musician. The scientific researches imply that the quality of close relationship of a person can influence the lifespan; 2)

the social influence factors reflect the technology level especially the medical level of the society, besides, they also reflect the tolerance and protection of privacy of the musician which influence the stress level that the musician faced up with; 3) Income level also indicates the achievement of the musician and his/her lifestyle.
What's more, comparing to other factors, those variables are more crucial and easy for qualification. Those data are acquirable and easy to get access to.

Part (f).

- Step 1. Build the Dataset. Collect a list of passed-away popular musicians from the Internet and collect data of each musician.
- Step 2. Explore the dataset to seethe correlation between life expectancy and each variable. We can answer whether there is a correlation and if there is, whether the correlation is negative or positive through our exploring. Furthermore, we can see the size of the influence and then decide whether to drop some of the variables.
- Step 3. Do linear regression on the training set to see the R^2. If the data size is sufficient, a regression model can be built according to it. If it is not, then consider adding more dependent variables into it.
- Step 4. Train the model. Take the parameters from the training set to test the fit of the model in the test set.