Problem Set #1

MACS 30150, Dr. Evans

Ying Sun

Problem 1: Classify a model from a journal

Part (a). The paper I choose is *"How Does Household Income Affect Child Personality Traits and Behaviors?"* in the *American Economic Review*.

Part (b). The detailed citation of the article:

Akee, Randall, et al. "How does household income affect child personality traits and behaviors?" *American Economic Review*, 108.3 (2018): 775-827.

Part (c). The model is as follows:

$$Y_{it} = \alpha_0 + \beta_1 YoungestCohorts_i + \beta_2 After_t + \beta_3 AmericanIndian_i +$$
$$\delta_1 YoungestCohorts_i \times After_t + \delta_2 YoungestCohorts_i \times AmericanIndian_i +$$
$$\lambda YoungestCohorts_i \times After_t \times AmericanIndian_i +$$
$$X'\mu + \epsilon_{it}$$

(ϵ_{it} is the error term.)

Part (d). The endogenous variable is Y_{it} which represents the child i's outcomes, including "psychopathologies (emotional/behavior disorders) and personality traits (conscientiousness, agreeableness and neuroticism) at year t" (Akee et al., 2018, p.777).

The exogenous variables in this model are as follows:

$YoungestCohorts_i$: a dummy indicator for whether the child belongs to the youngest or the second youngest;

$After_t$: a dummy indicator for whether the year is after the start of the casino transfer payment;

$AmericanIndian_i$: a dummy indicator for whether the child is American Indian race;

X': a vector which includes "a control for child age, the interaction between age and American Indian race, calendar-month-specific dummies and a count variable for the number of children younger than six in the household" (Akee et al., 2018, p.789)

Part (e). The model mentioned above is dynamic, nonlinear and stochastic. First, there is time component that influences the performance of this model, so it is dynamic; Second, the model is nonlinear, it includes two double-interaction terms and a triple-intersection term; Third, it is stochastic because an error term ϵ_{it} is included in this model.

Part (f). A missing variable that might be valuable here is the initial household income. According the data in the pre-intervention period, we can find that there is a strong positive relationship between initial household income and initial personality skill endowment across both racial groups. A lower household income may cause a lower investment in child skills. So, it is possible that children from households with lower income (a lower investment in child skills) would exhibit greater human

capital gains from the unconditional household income transfer if the marginal effect is decreasing.

Reference:

Akee, Randall, et al. "How does household income affect child personality traits and behaviors?" *American Economic Review*, 108.3 (2018): 775-827.

Problem 2: Make your own model

    Part (a). A logistic regression model for marriage decision:

$$Pr(M) = \frac{exp(\beta_0 + \beta_1 Age + \beta_2 Gender + \beta_3 Education + \beta_4 Income + \beta_5 Religion) + \epsilon}{1 + exp(\beta_0 + \beta_1 Age + \beta_2 Gender + \beta_3 Education + \beta_4 Income + \beta_5 Religion) + \epsilon}$$

$$Marriage = \begin{cases} 1 = \text{get married, if } \Pr(M) >= 0.5 \\ 0 = \text{not get married, if } \Pr(M) < 0.5 \end{cases}$$

    Marriage: a dummy variable of whether the individual decides to get married, it equals to 1 when the individual decides to get married, otherwise, it equals to 0;

    Age: the age of the individual;

    Gender: a dummy variable of whether the individual is male, it equals to 1 when the individual is male and it equals to 0 when the individual is female;

    Education: a categorical variable to indicate the education level of the individual;

    Income: the Logarithmic form of the individual's income;

    Religion: a categorical variable to indicate the religious belief of the individual

    Part (b). The endogenous variable is *Marriage*, which represents whether the individual decides to get married(1 = get married and 0 = not get married).

    Part (c). Based on this model, we could use the collected data to estimate the value of β. Then we could simulate data from this model given all the parameters and relationships. So the model is a complete data generating process.

    Part (d). I think the key factors that influence this outcome are *Age*, *Education* and *Income*. In terms of age, many countries have legal provisions on the age of marriage and people cannot get married before this legal age. Besides, for a particular age stage, people are more likely to get married as they get older. As for education level, there is no doubt that education level has important influence on marriage. First, people with different education levels have different requirements and expectations for marriage. Second, it takes time to get a higher education, which may delay the marriage plan of these people. In a sense, income can be regarded as a kind of condition for marriage decisions. In fact, many people set a certain income level as a prerequisite for marriage to ensure that their incomes can cover the necessary living expenses.

    Part (e). The reasons that I decide on these factors are as follows:

Age: The age of the individual obviously influences the marriage decision. Generally, at certain ages, people are more likely to get married as they get older.

    Gender: Males and females often hold different views on marriages. Compared with males, females are more willing to get married.

    Education: The level of education has several effects on marriage. First, people with different education levels may have different views on marriage. Usually they have different requirements and expectations for marriage. Besides, it takes time to get a higher education, which may delay the marriage plan of these people.

    Income: For many people, a certain level of economic base or income may be a necessary condition for marriage. In this sense, income may influence people's marriage decision.

Religion: People with different religious beliefs often have different views on marriage.

Part (f). There are several steps in a preliminary test to check whether the factors above are significant in the real life:

First, we can conduct a survey to collect relevant data or just extract relevant information we need from previous datasets. Second, we need to divide our dataset into training data and test data. Then we can use the training data to estimate the parameters in this model and check whether these estimators are significant. Finally, we need use the test data to check the robustness of our model. More specifically, we put the values of the explanatory variables into this model to get the predicted outcome and compare the predicted outcome with the real outcome.