

Problem Set #1

MACS 30100, Dr. Evans

Zunda Xu

Problem 1 Classify a model from a journal.

Part (a). The published article I found is from the *American Economic Review* 2017,107(6), which is named *Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets*.

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

Andrabi, Tahir, Jishnu Das, and Asim Ijaz Khwaja. 2017. "Report Cards: The Impact of Providing School and Child Test Scores on Educational Markets." *American Economic Review*, 107(6): 1535-63.

DOI: 10.1257/aer.20140774

Part (c). The equation of the model is as follows:

$$\begin{aligned} Y_{mi2} = & \alpha_d + \beta_0 RC_m + \beta_1 GOV_{mi} + \beta_2 HIGH_{mi1} + \beta_3 RC_m \cdot GOV_{mi} \\ & + \beta_4 RC_m \cdot HIGH_{mi1} + \beta_5 RC_m \cdot GOV_{mi} \cdot HIGH_{mi1} + \gamma \cdot Y_{mi1} \\ & + \delta \cdot X_{m1} + \epsilon_{mi} \end{aligned}$$

Part (d). The endogenous variable in this model is Y_{mi2} , which is the outcome of interest such as fees or enrollment) for school i in village m in time period 2 (post-intervention year), while the exogenous variables in this model are as follows:

RC_m : the treatment dummy assigned to village m, which means if the village received the treatment, $RC_m = 1$

α_d : district fixed-effects

Y_{mi1} : the baseline measurement of the outcome variable

X_{m1} : a vector of village-level baseline controls (size, wealth, adult literacy, and Herfindahl measure of school competition)

GOV_{mi} : a dummy indicator for whether the school is a public school

$HIGH_{mi1}$: an indicator for whether the school baseline score was above a predefined baseline test score threshold

Part (e). The model mentioned above is a static, nonlinear and deterministic model. Since the model does not contain an element of time or the interactions between variables over time, thus the model is static, also the output of the model is fully determined by the parameter values and the initial conditions, thus it is a static, nonlinear and deterministic model.

Part (f). A valuable that the model may be missing is that the amount of education funds those schools received during the experiment time, since the endogenous variable Y_{mi2} , which represents the outcome of interest such as fees may also influenced by the education funds that those schools received, for those schools which obtaining higher education funds, they may also choose to reduce tuition fees.

Problem 2 Make your own model.

Part (a)(b). Write down a model of whether someone decides to get married.

The model I designed is a Probit Model and the equation of the model are as follows:

$$Pr(Y_i = 1|X_i) = \Phi(X_i^T \beta),$$

or

$$Y_i = \begin{cases} 1, & Y_i^* > 0 \\ 0, & \text{otherwise} \end{cases}, \quad Y_i^* = X_i^T \beta + \epsilon_i.$$

, where Pr denotes probability, and Φ is the Cumulative Distribution Function (CDF) of the standard normal distribution. The parameters β are typically estimated by maximum likelihood. We can also suppose there exists an auxiliary random variable Y_i^* , then Y_i can be viewed as an indicator for whether this latent variable is positive, X_i represents the vector of independent variables and $\epsilon_i \sim N(0, 1)$.

To be specific, the equation of Y_i^* can be described as follows:

$$Y_i^* = \beta_0 + \beta_1 \cdot AGE_i + \beta_2 \cdot GEN_i + \beta_3 \cdot INC_i + \beta_4 \cdot EDU_i + \beta_5 \cdot RACE_i \\ + \beta_6 \cdot REG_i + \beta_7 \cdot REL_i + \beta_8 \cdot SEX_i + \beta_9 \cdot PRES_i + \epsilon_i.$$

The endogenous variable in this model is Y_i , which is a dummy variable for whether the individual i decides to get married, in other word, if $Y_i = 1$, means the individual i decides to get married.

The exogenous variables in this model are as follows:

AGE_i : the age of individual i

GEN_i : the gender of individual i , which is a dummy variable, if i is male, $GEN_i = 1$

INC_i : the income level of individual i , which is a categorical variable

EDU_i : the education level of individual i , which is a categorical variable

$RACE_i$: the race of individual i , which is a categorical variable

REG_i : the region that individual i lives in, which is a categorical variable

REL_i : the religious of individual i , which is a categorical variable

SEX_i : the sexual orientation of individual i , which is a dummy variable, if i is heterosexual, then $SEX_i = 1$

$PRES_i$: the pressure of getting married received from families or friends , which is a categorical variable

Part (c). Make sure that the model is a complete data generating process.

In this model, we could use the sample data, for example, a sample with n individuals, to estimate the value of parameters $\beta_0, \beta_1, \dots, \beta_9$. When the parameters β and the vector of exogenous variables X_i is known (fixed), a sample of size n_0 is created by obtaining n_0 values of the random variable ϵ and then using these values, in conjunction with the rest of the model, to generate n_0 values of Y_i , which make sure that the model is a complete data generating process.

Part (d). The key factors that influence this outcomes are AGE_i , REL_i and $PRES_i$, which represent people's age, religious and pressure of getting married.

Part (e). The reasons I decide on these three factors are as follows: 1) Generally speaking, the decision of getting married is obviously influenced by people's age,

which means in some specific age groups, older people are more likely to decide to get married than younger people. **2)** People's decisions of getting married are also influenced by their religious, since people with some specific religious may not be allowed to get married, also people with different religious may have different attitudes towards marriage. **3)** The pressure people received of getting married from families, relatives or friends is also a key factor, as far as I am concerned, people are more likely to choose to get married when they receive more pressure.

Part (f). The steps of the preliminary test are as follows: **1)** Conduct the anonymous online survey to create the dataset used for the test, the content of the online survey including interviewers' age, gender, income level, education level, race, region, religious, sexual orientation, pressure of getting married (high, medium or low) and whether they decide to get married (Yes or No). After collecting those survey results, we change those categorical variables into numeric ordinal variables based on our benchmark to establish the final dataset. **2)** Randomly choose a sample set with size n and use this sample as training data to estimate the parameters in this model. **3)** The rest of the data will be used as test data, we input the value of those independent variables into the model to estimate the outcome, which is whether the individual i decide to get married, and compare the outcomes with the real outcome in the dataset to check whether those factors in this model are significant in real life.